Maestría en

ADMINISTRACIÓN DE EMPRESAS

Tesis previa a la obtención de título de Magister en Administración de Empresas

AUTORES:

Ing. Daissy Andreina Páez Mora

Ing. Bryan Javier Mayorga Gamboa

MSc. Cristian Rene Viracucha Pila

Lic. César Andrés Sarmiento Rojas

Psic. Luis Andrés Urbina Bastidas

TUTOR: Lic. José Luis Pérez Galán

Propuesta de implementación de un departamento de Business Intelligence y Big Data para la predicción de la demanda y la optimización de inventarios en Walmart Inc.

PROYECTO DE TITULACIÓN – FIN DE MÁSTER
MAESTRÍA EN ADMINISTRACIÓN DE EMPRESA ENL

Propuesta de implementación de un departamento de Business Intelligence y Big Data para la predicción de la demanda y la optimización de inventarios en Walmart Inc.

Por

Ing. Daissy Andreina Páez Mora
Ing. Bryan Javier Mayorga Gamboa
MSc. Cristian Rene Viracucha Pila
Lic. César Andrés Sarmiento Rojas
Psic. Luis Andrés Urbina Bastidas

Octubre 2026

Aprobado



Cristian Melo
Presidente(a) del Tribunal
Universidad Internacional del Ecuador

Yo, Cristian Javier Melo González e Ignacio Maroto, declaramos que, personalmente, conocemos que los graduandos: Ing. Daissy Andreina Páez Mora, Ing. Bryan Javier Mayorga Gamboa, MSc. Cristian Rene Viracucha Pila, Lic. César Andrés Sarmiento Rojas y Psic. Luis Andrés Urbina Bastidas, son los autores exclusivos de la presente investigación y que esta es original, auténtica y personal de ellos.



Cristian Melo
Coordinador MBA UIDE



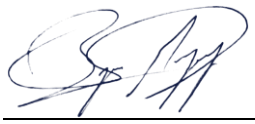
Ignacio Maroto
Provost WBS

Autoría del Trabajo de Titulación

Nosotros, Ing. Daissy Andreina Páez Mora, Ing. Bryan Javier Mayorga Gamboa, MSc. Cristian Rene Viracucha Pila, Lic. César Andrés Sarmiento Rojas y Psic. Luis Andrés Urbina Bastidas, declaramos bajo juramento que el trabajo de titulación titulado **Propuesta de implementación de un departamento de Business Intelligence y Big Data para la predicción de la demanda y la optimización de inventarios en Walmart Inc.** es de nuestra autoría y exclusiva responsabilidad legal y académica; que no ha sido presentado anteriormente para ningún grado o calificación profesional, habiéndose citado las fuentes correspondientes y respetando las disposiciones legales que protegen los derechos de autor vigentes.



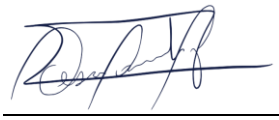
Ing. Daissy Andreina Páez Mora
Correo electrónico: daissy.damp@gmail.com



Ing. Bryan Javier Mayorga Gamboa
Correo electrónico: bryanjvr93@gmail.com



MSc. Cristian Rene Viracucha Pila
Correo electrónico: tradexc@hotmail.com



Lic. César Andrés Sarmiento Rojas
Correo electrónico: cesar.sarmiento1999@gmail.com



Psic. Luis Andrés Urbina Bastidas
Correo electrónico: luisurbina6913@gmail.com

Autorización de Derechos de Propiedad Intelectual

Nosotros, Ing. Daissy Andreina Páez Mora, Ing. Bryan Javier Mayorga Gamboa, MSc. Cristian Rene Viracucha Pila, Lic. César Andrés Sarmiento Rojas y Psic. Luis Andrés Urbina Bastidas, en calidad de autores del trabajo de investigación titulado **Propuesta de implementación de un departamento de Business Intelligence y Big Data para la predicción de la demanda y la optimización de inventarios en Walmart Inc.**, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden se rigen por lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, octubre de 2026

Ing. Daissy Andreina Páez Mora
Correo electrónico: daissy.damp@gmail.com

Ing. Bryan Javier Mayorga Gamboa
Correo electrónico: bryanjvr93@gmail.com

MSc. Cristian Rene Viracucha Pila
Correo electrónico: tradexc@hotmail.com

Lic. César Andrés Sarmiento Rojas
Correo electrónico: cesar.sarmiento1999@gmail.com

Psic. Luis Andrés Urbina Bastidas
Correo electrónico: luisurbina6913@gmail.com

Dedicatorias y Agradecimientos

Agradecemos a la Universidad Internacional del Ecuador, a Westfield Business School y a EIG Business School por la formación recibida a lo largo de la Maestría en Administración de Empresas, y en especial a nuestro tutor, José Luis Pérez Galán, por su guía durante el desarrollo de este proyecto.

Dedicamos este trabajo a nuestras familias, por su paciencia y apoyo constante durante estos meses de estudio, y a nuestros compañeros de grupo, cuyo compromiso semanal hizo posible integrar diez entregas en un solo proyecto.

INDICE DEL DOCUMENTO

INTRODUCCIÓN Y OBJETO DEL PROYECTO 12
Contexto 12
Planteamiento del problema 12
Objeto del proyecto 12
Enfoque metodológico y alcance 13
Estructura del documento 14
PARTE I: BUSINESS INTELLIGENCE 15
CAPÍTULO 1. Walmart Inc.: perfil corporativo y definición del proyecto 16
1.1 Perfil corporativo de Walmart Inc. 16
1.2 Walmart, la inteligencia de negocios y el Big Data 19
1.3 Objetivo del proyecto y problema a resolver 21
1.4 Áreas y personas que demandan información y uso actual de la misma 23
1.5 Objetivos de cada área con la información actual y necesidades detectadas 25
1.6 Indicadores clave de desempeño (KPIs) 29
CAPÍTULO 2. Estructura de la gestión de datos del proyecto 33
2.1 Origen de la información y recurrencia de carga: interna y externa 33
2.2 Sistemas y destino de la carga 34
2.3 Almacenamiento en Data Warehouse y Data Marts 35
2.4 Carencias y problemas en la carga de datos 36
2.5 Informes que se generan actualmente, necesidades detectadas y no cubiertas 37
CAPÍTULO 3. Arquitectura de datos, sistemas implicados y recursos de IT 39
3.1 Procesos de carga de datos (ETL y Data Lake) 39
3.2 Información y calidad de la información disponible 41
3.3 Variables disponibles y acceso al Data Warehouse 44
3.4 Sistemas implicados en los procesos 46
3.5 Recursos de IT o externos empleados 49
CAPÍTULO 4. Visualización de datos: herramienta, perfiles de usuario y accesos 51

4.1 Análisis de herramientas y selección de la solución acorde con el ecosistema de Walmart....51
4.2 Usuarios y perfiles de uso de cada dashboard....54
4.3 Necesidades de información y accesos por departamento y perfil....56
4.4 Revisión y cumplimiento de los objetivos iniciales....57
CAPÍTULO 5. El departamento de Analítica de Datos y Business Intelligence 59
5.1 Situación actual del departamento 59
5.2 Carencias en formación y recursos 60
5.3 Coordinación de los flujos con otras áreas 61
5.4 Organigrama, dependencias y solución de problemas 64
5.5 Recursos externos: subcontratación 67
5.6 Caso de negocio: costos y beneficios....68
5.7 Conclusiones del capítulo 70
PARTE II: BIG DATA E INTELIGENCIA ARTIFICIAL....71
CAPÍTULO 1. Datos no estructurados y fuentes Big Data para la predicción de la demanda....72
1.1 Fuentes de datos internas y externas utilizadas por Walmart 72
1.2 Relación de los datos con la propuesta inicial....73
1.3 Casos de uso que permitirán incorporar estos datos al modelo de BI 79
1.4 Impacto del Big Data por áreas, departamentos y procesos de negocio 81
CAPÍTULO 2. Arquitectura Big Data en la nube: proveedor, modelo de servicio e integración....90
2.1 Análisis de alternativas de nube y arquitectura Big Data....90
2.2 Selección de la mejor alternativa de proveedor Cloud....93
2.3 Modelo de cloud computing: PaaS....97
2.4 Flujo de integración de la arquitectura Big Data con el data warehouse y la estructura de Business Intelligence....99
CAPÍTULO 3. Framework de procesamiento y ecosistema tecnológico Big Data....105
3.1 Análisis y selección crítica de frameworks 105
3.2 Elementos complementarios del ecosistema Big Data 107

CAPÍTULO 4. Data Science y Machine Learning: modelo predictivo y caso de negocio 112
4.1 Identificación y definición del objetivo de aplicación del modelo de Data Science y Machine Learning 112
4.2 Fuentes de datos utilizadas 115
4.3 Área de aplicación y departamentos de negocio implicados 116
4.4 Identificación de la mejora esperada en la aplicación de inteligencia artificial 122
4.5 Caso de negocio: cuantificación de la mejora esperada 123
CAPÍTULO 5. Analítica en tiempo real (Real Time) aplicada a la demanda 131
5.1 Objetivo de aplicación del modelo Real Time 131
5.2 Fuentes de datos utilizadas 132
5.3 Área de aplicación y departamentos de negocio implicados 133
5.4 Mejora esperada 133
5.5 Caso de negocio del modelo Real Time 134
CONCLUSIONES Y APLICACIONES 136
Síntesis del proyecto 136
Conclusiones 136
Aplicaciones 137
Limitaciones y líneas futuras 137
REFERENCIAS 139

INDICE DE FIGURAS Y TABLAS

Índice de figuras

Figura 1. *Arquitectura de Business Intelligence propuesta para la predicción de demanda en Walmart* 48
Figura 2. *Flujo de información entre el Departamento de BI y las áreas del negocio* 62
Figura 3. *Del dato crudo al dashboard: secuencia de coordinación del flujo de datos* 63
Figura 4. *Organigrama propuesto de la Gerencia de Analítica e Inteligencia de Demanda* 65
Figura 5. *Flujo de integración de datos para la predicción de demanda y la optimización de inventarios* 76
Figura 6. *De los datos a las decisiones: evolución de la analítica descriptiva a la prescriptiva* 77
Figura 7. *Ejemplo de aplicación: detección anticipada de un riesgo de stock-out* 78
Figura 8. *Proceso de marketing personalizado basado en datos* 83
Figura 9. *Variables integradas en el modelo predictivo de inventario* 84
Figura 10. *Recorrido de la cadena de suministro analizado con Big Data* 85
Figura 11. *Integración de datos financieros para decisiones estratégicas* 87
Figura 12. *Impacto global esperado del Big Data en Walmart* 89
Figura 13. *Solución propuesta para Walmart: datos integrados en la nube* 94
Figura 14. *Del dato a la decisión: aplicación de Azure al modelo predictivo* 96
Figura 15. *Arquitectura de datos propuesta sobre Microsoft Azure* 97
Figura 16. *Flujo de integración de la arquitectura Big Data con el data warehouse y la capa de BI* 101
Figura 17. *Funcionamiento del modelo predictivo de demanda y riesgo de inventario* 114
Figura 18. *Flujo de aplicación del modelo de pronóstico de demanda y áreas que intervienen* 118

Índice de tablas

Tabla 1. *Ficha de identificación de Walmart Inc.* 16
Tabla 2. *Segmentos de negocio reportables de Walmart Inc.* 17
Tabla 3. *Datos financieros clave de Walmart Inc. (FY2024–FY2026)* 18
Tabla 4. *Principales competidores de Walmart en Estados Unidos* 19
Tabla 5. *Áreas que demandan información y uso actual de la misma* 23
Tabla 6. *KPIs de Walmart por categoría* 30

Tabla 7. *Variables disponibles en el Data Warehouse de Walmart*.....44
Tabla 8. *Sistemas involucrados en el modelo de Business Intelligence*.....47
Tabla 9. *Análisis comparativo de herramientas de visualización frente al ecosistema de Walmart*.....52
Tabla 10. *Necesidades de información y niveles de acceso por área*.....56
Tabla 11. *Síntesis de las carencias identificadas en el área de analítica*.....61
Tabla 12. *Dependencias de la Gerencia de Analítica e Inteligencia de Demanda con otras áreas*.....66
Tabla 13. *Modelo de solución de problemas del departamento*.....67
Tabla 14. *Costos anuales estimados de la solución (escenario académico)*.....68
Tabla 15. *Relación entre las fuentes de datos, la demanda y su aplicación en el proyecto*.....74
Tabla 16. *Ejemplo de detección de sobrestock*.....80
Tabla 17. *Mapa de impacto del Big Data por área o departamento*.....89
Tabla 18. *Cuadro comparativo de proveedores de nube y Big Data*.....90
Tabla 19. *Criterios considerados para la selección de Microsoft Azure*.....93
Tabla 20. *Capas de la arquitectura de integración*.....100
Tabla 21. *Recorrido de un dato de venta en el flujo de pronóstico de demanda*.....103
Tabla 22. *Ecosistema tecnológico seleccionado*.....110
Tabla 23. *Escenarios de inventario que anticipa el modelo*.....113
Tabla 24. *Alcance de aplicación del modelo de pronóstico de demanda en Walmart*.....117
Tabla 25. *Tratamiento del pronóstico por segmento de artículos*.....118
Tabla 26. *Departamentos implicados y su rol frente al modelo*.....119
Tabla 27. *Matriz RACI del modelo de pronóstico de demanda*.....120
Tabla 28. *Base financiera utilizada (año fiscal 2026)*.....124
Tabla 29. *Supuestos del escenario*.....124
Tabla 30. *Proyección simplificada a tres años*.....128
Tabla 31. *Análisis de sensibilidad del beneficio anual*.....128
Tabla 32. *Indicadores para evaluar el proyecto*.....129
Tabla 33. *Caso de negocio del modelo Real Time (escenario académico)*.....134

INTRODUCCIÓN Y OBJETO DEL PROYECTO

Contexto

Walmart Inc. es el minorista más grande del mundo por ingresos: en el año fiscal 2026 alcanzó ingresos totales de 713 200 millones de dólares, con un crecimiento de 4,7 % frente al ejercicio anterior, y atiende cada semana a cerca de 270 millones de clientes a través de más de 10 750 tiendas en 19 países y de sus canales digitales (Walmart Inc., 2026). Más allá de su escala, la compañía es un caso de estudio privilegiado para la inteligencia de negocios: fue pionera en la construcción de grandes almacenes de datos empresariales en la década de 1990 y hoy procesa del orden de 2,5 petabytes de información por hora para apoyar decisiones de abastecimiento, precios y experiencia del cliente (Marr, 2017).

Sin embargo, el volumen de datos no garantiza por sí mismo mejores decisiones. Como señalan Chen et al. (2012), el valor de la analítica surge cuando los datos se integran, se gobiernan y se convierten en conocimiento accionable para quienes toman decisiones. En el comercio minorista de consumo masivo, ese reto se concentra en un punto crítico de la operación: anticipar cuánto se venderá de cada producto, en cada tienda y en cada período.

Planteamiento del problema

El problema que aborda este proyecto es la coexistencia de sobrestock y desabastecimiento en la red de Walmart. El sobrestock inmoviliza capital de trabajo y genera costos de almacenamiento, manipulación, deterioro y obsolescencia; el desabastecimiento provoca ventas perdidas, deteriora la experiencia del cliente y lo empuja hacia la competencia. Ambos fenómenos tienen una causa común: pronósticos de demanda que no capturan con suficiente precisión la variabilidad del consumo, influida por promociones, estacionalidad, días festivos, clima, eventos locales y tendencias digitales.

En la situación de partida, la información relevante para planificar la demanda está distribuida entre sistemas operacionales y reportes especializados de cada área, y el análisis es mayoritariamente descriptivo: explica lo que ocurrió, pero no anticipa lo que ocurrirá. La consecuencia es una gestión reactiva del inventario, en la que los problemas se detectan cuando ya afectan a la venta o al costo.

Objeto del proyecto

El objeto de este trabajo es diseñar la propuesta de implementación de un departamento de Business Intelligence y Big Data para Walmart Inc., cuyo caso de uso central es un modelo de predicción de la demanda por producto (SKU), tienda y período, orientado a reducir el sobrestock y el desabastecimiento y a mejorar la planificación de compras, reposición y distribución.

Objetivo general

Proponer una solución integral de Business Intelligence, Big Data e Inteligencia Artificial que permita a Walmart evolucionar de una gestión descriptiva y reactiva del inventario hacia un modelo predictivo y prescriptivo, sustentado en una arquitectura de datos escalable, una estructura organizativa definida y un caso de negocio cuantificado.

Objetivos específicos

- Caracterizar a Walmart, su modelo de negocio y su trayectoria en el uso de datos, e identificar las áreas que demandan información para la planificación de la demanda.
- Diagnosticar la estructura actual de gestión de datos: fuentes, sistemas, almacenamiento, carencias e informes disponibles frente a las necesidades no cubiertas.
- Describir la arquitectura de datos, las variables disponibles en el Data Warehouse y los recursos tecnológicos internos y externos que soportan el proyecto.
- Seleccionar la herramienta de visualización y definir los perfiles de usuario, las necesidades de información y los niveles de acceso por departamento.
- Diseñar el departamento de analítica: carencias de formación y recursos, organigrama, dependencias y modelo de solución de problemas.
- Incorporar fuentes de datos no estructurados y definir la arquitectura Big Data: proveedor de nube, modelo de servicio, flujo de integración y framework de procesamiento.
- Definir el modelo de Machine Learning, sus áreas de aplicación y los departamentos implicados, y cuantificar el beneficio esperado mediante un caso de negocio.
- Analizar la aplicación de modelos de procesamiento en tiempo real como evolución de la arquitectura propuesta.

Enfoque metodológico y alcance

El proyecto se desarrolló de forma incremental a lo largo de diez entregas semanales, siguiendo la secuencia propia de la cadena de valor del dato: primero se diagnostica y estructura la información (Business Intelligence), luego se amplía la capacidad de captura y procesamiento a datos masivos y no estructurados (Big Data) y, finalmente, se aplica la Inteligencia Artificial para pasar de la analítica descriptiva a la predictiva y prescriptiva. Como marco de referencia para el tratamiento de los datos se adopta el proceso de descubrimiento de conocimiento en bases de datos (KDD), que articula la selección, el preprocesamiento, la transformación, la minería de datos y la interpretación de resultados (Fayyad et al., 1996).

Es necesario precisar dos consideraciones de alcance que aplican a todo el documento. En primer lugar, Walmart no publica el detalle de su arquitectura

tecnológica ni de sus datos internos por razones de seguridad y propiedad intelectual; por ello, el análisis se construye a partir de fuentes públicas (informes anuales y formularios ante la SEC, comunicados corporativos, publicaciones especializadas y literatura académica), y las descripciones de procesos internos se presentan como propuestas técnicas fundamentadas, no como hechos confirmados por la compañía.

En segundo lugar, se distingue entre la arquitectura objetivo y el prototipo académico. La arquitectura objetivo describe cómo operaría la solución en producción dentro de Walmart, con datos internos (punto de venta, inventarios, logística, comercio electrónico) e ingesta continua en tiempo casi real. El prototipo académico, en cambio, se plantea con datos históricos y fuentes de dominio público, que son los únicos disponibles para el grupo. Del mismo modo, las cifras del caso de negocio se presentan como un escenario académico con supuestos explícitos, que debería validarse mediante un piloto controlado con datos reales.

Estructura del documento

El documento se organiza en dos partes, que corresponden a los dos bloques del programa. La Parte I, dedicada a Business Intelligence, presenta a Walmart y define el proyecto (Capítulo 1), analiza la estructura de gestión de datos (Capítulo 2), describe la arquitectura de datos y los sistemas implicados (Capítulo 3), justifica la herramienta de visualización y los perfiles de usuario (Capítulo 4) y diseña el departamento de analítica (Capítulo 5). La Parte II, dedicada a Big Data e Inteligencia Artificial, incorpora los datos no estructurados (Capítulo 1), la arquitectura en la nube (Capítulo 2), el framework de procesamiento (Capítulo 3), el modelo de Machine Learning con su caso de negocio (Capítulo 4) y el análisis en tiempo real (Capítulo 5). El trabajo cierra con las conclusiones y aplicaciones del proyecto.

PARTE I: BUSINESS INTELLIGENCE

La primera parte del proyecto corresponde al bloque de Business Intelligence. Su propósito es construir la base informacional sobre la que se apoya todo el trabajo: conocer a la empresa y el problema, diagnosticar cómo se gestiona hoy la información, describir la arquitectura de datos disponible, seleccionar la herramienta con la que se entregará la información a cada usuario y diseñar el departamento que será responsable de la solución.

CAPÍTULO 1. Walmart Inc.: perfil corporativo y definición del proyecto

Este capítulo presenta a Walmart Inc. como caso de estudio: su perfil corporativo, su modelo de negocio, su desempeño financiero y su trayectoria como pionera en el uso de datos. Sobre esa base se define el objetivo del proyecto, se identifican las áreas que demandan información para la planificación de la demanda y se establecen los indicadores clave de desempeño que servirán de referencia en los capítulos siguientes.

1.1 Perfil corporativo de Walmart Inc.

Walmart Inc. es el minorista más grande del mundo por ingresos y uno de los mayores empleadores privados del planeta. En el año fiscal 2025 (cerrado el 31 de enero de 2025) generó ingresos totales por 681,0 mil millones de dólares, y en el año fiscal 2026 (cerrado el 31 de enero de 2026) superó los 713 mil millones de dólares, atendiendo cada semana a cerca de 270–280 millones de clientes y miembros a través de más de 10.750 tiendas en 19 países y sus plataformas de comercio electrónico.

Para un proyecto de Business Intelligence y Big Data, Walmart es un objeto de estudio excepcional por una razón concreta: es, desde hace décadas, uno de los pioneros mundiales en el uso intensivo de datos para competir. Construyó el data warehouse comercial más grande del mundo, fue de las primeras compañías en alcanzar un terabyte de datos (1992) y hoy captura y procesa petabytes de información por hora para pronosticar demanda, optimizar su cadena de suministro y personalizar la experiencia del cliente con inteligencia artificial. Su ventaja competitiva —precios bajos a escala masiva— se sostiene, en buena parte, sobre su capacidad analítica.

1.1.1 Ficha de identificación

Tabla 1. ***Ficha de identificación de Walmart Inc.***

Concepto	Dato
Razón social	Walmart Inc.
Símbolo bursátil	WMT (NASDAQ, desde diciembre de 2025; antes NYSE)
Fundación	1962, por Sam Walton (primera tienda en Rogers, Arkansas)
Sede central	Bentonville, Arkansas, Estados Unidos
Sector / Industria	Consumo defensivo / Comercio minorista de descuento
CEO	John Furner (desde febrero de 2026)
Presidente del Consejo	Gregory B. Penner
Asociados (empleados)	≈ 2,1 millones a nivel mundial
Presencia	Más de 10.750 tiendas en 19 países
Clientes por semana	≈ 270–280 millones a nivel mundial
Ingresos FY2026	≈ 713 mil millones de USD

Nota. Elaboración propia con base en Walmart Inc. (2025a, 2026).

1.1.2 Perfil e historia de la empresa

Walmart fue fundada en 1962 por **Sam Walton**, quien abrió su primera tienda de descuento en Rogers, Arkansas, bajo una premisa simple y perdurable: ofrecer una amplia variedad de mercancía general a precios bajos. La empresa se listó en la Bolsa de Valores de Nueva York en 1972 y, hacia finales de la década de 1980, ya era el minorista más rentable y de mayores ingresos de Estados Unidos. En diciembre de 2025 trasladó su cotización al índice Nasdaq.

La familia Walton mantiene más del 50 % de la propiedad de la compañía a través de Walton Enterprises y sus tenencias individuales, lo que le confiere un carácter de empresa familiar controlada pese a su escala global. En 2025 Walmart inauguró un nuevo campus corporativo ("Home Office") de 350 acres en Bentonville. La compañía se refiere internamente a sus empleados como asociados, reflejo de una cultura organizacional que históricamente ha promovido el ascenso interno: su actual CEO, John Furner, comenzó como asociado por hora en 1993.

1.1.3 Propósito y modelo de negocio

Walmart se define hoy como un minorista omnicanal "impulsado por las personas y potenciado por la tecnología", cuyo propósito declarado es ayudar a las personas a ahorrar dinero y vivir mejor. El modelo descansa sobre tres pilares que se refuerzan entre sí:

- **Volumen y escala:** su enorme poder de compra le permite negociar precios bajos con los proveedores y trasladarlos al cliente.
- **Eficiencia operativa:** una cadena de suministro considerada de las más avanzadas del mundo, con minimización de costos de inventario y logística.
- **Omnicanalidad:** integración de tiendas físicas, e-commerce, aplicación móvil, recogida y entrega, más nuevos negocios de alto margen como publicidad (Walmart Connect), membresías (Walmart+) y monetización de datos.

1.1.4 Estructura y segmentos de negocio

La compañía organiza sus operaciones en tres grandes segmentos reportables, cada uno con su propio equipo directivo:

Tabla 2. ***Segmentos de negocio reportables de Walmart Inc.***

Segmento	Descripción	Formatos / Alcance
Walmart U.S.	El segmento más grande y rentable. Opera la mayor operación minorista de Estados Unidos.	Más de 4.600 tiendas: Supercenters, tiendas de descuento y Neighborhood Markets.
Walmart International	Operaciones fuera de EE. UU. en aproximadamente 17–19 países.	Incluye mercados como México (Walmex), Canadá, China, Centroamérica y otros.

Segmento	Descripción	Formatos / Alcance
Sam's Club	Cadena de clubes de membresía por suscripción; compite directamente con Costco.	≈ 600 clubes en EE. UU.; venta de productos al por mayor y combustible con descuento.

Nota. Elaboración propia con base en Walmart Inc. (2025a).

Además de las ventas de mercancía, Walmart ha desarrollado unidades de negocio de rápido crecimiento y mayor margen que son especialmente relevantes desde una óptica de datos: Walmart Connect (medios minoristas y publicidad), Walmart+ (membresía tipo suscripción), Walmart Marketplace (vendedores terceros) y Walmart Data Ventures, que monetiza los datos de compra convirtiéndolos en información comercial para proveedores.

1.1.5 Datos financieros clave

El año fiscal de Walmart cierra el 31 de enero. Las cifras siguientes muestran la evolución reciente de sus ingresos y rentabilidad. Nótese el crecimiento sostenido en un solo dígito, propio de un negocio maduro pero de escala colosal: cada punto porcentual representa miles de millones de dólares.

Tabla 3. ***Datos financieros clave de Walmart Inc. (FY2024–FY2026)***

Métrica	FY2024	FY2025	FY2026
Ingresos totales	$648,1 mil M	$681,0 mil M	≈ $713,2 mil M
Ventas netas	—	$674,5 mil M	—
Utilidad operativa	$27,0 mil M	$29,3 mil M	—
Margen operativo	4,2 %	4,3 %	—
Crecimiento de ingresos	≈ 6,0 %	≈ 5,1 %	≈ 4,7 %
Flujo de caja operativo	—	$36,4 mil M	—

Nota. Elaboración propia con base en Walmart Inc. (2025a, 2026). El año fiscal cierra el 31 de enero.

En los doce meses terminados en abril de 2026, los ingresos alcanzaban ya cerca de 725 mil millones de dólares. Para dimensionar la escala: los $681 mil millones del FY2025 equivalen a alrededor de 1,87 mil millones de dólares en ventas por día. El comercio electrónico ha sido el principal motor del crecimiento reciente, con expansiones trimestrales del orden del 20 % al 25 %, impulsado por la recogida y entrega surtidas desde tienda.

Un dato estratégico: el margen operativo pasó de 3,3 % (FY2023) a 4,3 % (FY2025). Esa mejora no proviene tanto de las ventas de mercancía, de márgenes muy delgados, como de los negocios de datos y servicios: publicidad, servicios de farmacia y compras asistidas por IA. Es decir, la analítica y los datos no solo hacen a Walmart más eficiente, sino estructuralmente más rentable.

1.1.6 Panorama competitivo

Walmart controla cerca del 20,6 % del mercado minorista de Estados Unidos, una posición dominante en el comercio físico. Sin embargo, enfrenta una competencia intensa y diferenciada según el canal y el segmento:

Tabla 4. *Principales competidores de Walmart en Estados Unidos*

Competidor	Perfil y punto de fricción con Walmart	Escala aproximada
Amazon	Principal rival en comercio electrónico y logística; líder indiscutido del canal digital. Presiona a Walmart en entrega rápida y tecnología.	Ventas minoristas ≈ $391 mil M (2024); capitalización ≈ $2,56 billones (ene. 2026).
Costco	Club de almacén; rival directo de Sam's Club. Fuerte lealtad de miembros y ganancias de cuota superiores a las de Walmart.	≈ $270–275 mil M (fiscal 2025); ≈ 7,8 % del mercado minorista de EE. UU.
Target	El competidor de "gran superficie" más parecido; posicionamiento más aspiracional. Ha perdido cuota frente a Walmart y Costco.	≈ $106–107 mil M (2024); tráfico en tiendas bajo presión.
Kroger	Mayor cadena de supermercados de EE. UU.; competencia directa en el negocio de abarrotes.	≈ 7,0 % del mercado minorista de EE. UU.
Dollar General / Dollar Tree	Cadenas de descuento profundo; capturan al consumidor que busca precio en formatos pequeños.	Miles de tiendas de conveniencia de bajo costo.

Nota. Elaboración propia con base en fuentes públicas del sector.

La lectura estratégica es clara: Walmart defiende su liderazgo en el comercio físico mientras persigue a Amazon en lo digital. En ambos frentes, la batalla se libra cada vez más con datos —pronóstico de demanda, personalización, optimización de rutas y precios dinámicos—, lo que conecta directamente con el enfoque de este proyecto.

1.2 Walmart, la inteligencia de negocios y el Big Data

Este apartado constituye el núcleo del caso de estudio. Walmart no es solo una empresa grande que "además usa datos": es uno de los referentes históricos y actuales del uso de datos como ventaja competitiva.

1.2.1 El legado: un pionero del data warehousing

Walmart construyó, de la mano de Teradata, uno de los primeros grandes data warehouses empresariales (EDW) del mundo. Fue de las primeras organizaciones en alcanzar un terabyte de datos en 1992 y, durante años, operó el data warehouse comercial más grande del mundo: pasó de 7,5 TB a más de 24 TB a finales de los años noventa y llegó a superar los 30 petabytes. Este sistema centralizaba los datos de punto de venta (POS) de miles de tiendas, permitiendo analizar el comportamiento por artículo, por tienda y por día.

De ese legado nació Retail Link, una plataforma que comparte datos de ventas e inventario en tiempo casi real con los proveedores, de modo que ellos mismos sepan cuándo reabastecer. Más tarde, WalmartLabs desarrolló soluciones como Social Genome, que analizaba millones de menciones en redes sociales, y la compañía migró a clústeres Hadoop de cientos de nodos para procesar datos no

estructurados. Este recorrido convierte a Walmart en un caso clásico en la literatura de BI y Big Data.

1.2.2 La escala de datos hoy

La magnitud de datos que maneja Walmart es difícil de exagerar. Se estima que captura del orden de 2,5 petabytes de datos por hora a partir de las transacciones, la actividad en línea, la aplicación móvil y las interacciones en redes sociales. Cada semana registra cientos de miles de menciones sociales sobre sus productos. Para gestionar esta escala, la compañía firmó un acuerdo con Microsoft Azure y ha invertido en una de las infraestructuras de nube privada más grandes del sector, combinando data warehouse estructurado, data lakes para datos no estructurados y capacidades de analítica avanzada.

1.2.3 Casos de uso actuales de analítica e IA

Hacia 2025–2026, Walmart evolucionó de modelos de IA aislados a una arquitectura "centrada en el sistema": un ecosistema de agentes de inteligencia artificial que automatizan flujos de trabajo de punta a punta. Algunos ejemplos concretos y medibles:

- **Pronóstico de demanda predictivo:** modelos de machine learning que analizan ventas históricas, estacionalidad, clima y tendencias para mantener el stock óptimo y reducir tanto quiebres como sobreinventario.
- **Route Optimization (optimización de rutas):** analiza en tiempo real capacidad de camiones, tráfico, clima y horarios. Permitió evitar 30 millones de millas innecesarias y 94 millones de libras de emisiones de CO_2, y ganó el prestigioso premio Franz Edelman de analítica.
- **Trend-to-Product:** motor multi-agente de IA generativa que rastrea tendencias en redes sociales y búsquedas, genera conceptos de producto y los lleva a diseño y abastecimiento en tan solo seis semanas.
- **Enterprise Inventory e inventario "auto-reparable":** una vista única y unificada del stock en tiendas, centros de distribución y canal en línea, con capacidad de corregirse a sí mismo.
- **Agentes conversacionales:** "Ask Sam" para que los asociados consulten datos operativos en segundos, y "Sparky" como asistente de compra para los clientes.
- **Precios y publicidad basados en datos:** optimización de precios según competencia y sensibilidad del cliente, y monetización de la información de compra a través de Walmart Connect y Walmart Data Ventures.

La ambición declarada por el nuevo CEO John Furner es una estrategia "AI-first": automatizar cerca del 65 % de los servicios en tienda y más de la mitad del volumen de cumplimiento (fulfillment) mediante robótica e IA, con sistemas de tiempo real ya activos en mercados como Costa Rica, México y Canadá.

En síntesis, Walmart demuestra cómo la inteligencia de negocios y el Big Data dejaron de ser un área de soporte para convertirse en el sistema nervioso de una

corporación que factura más de 700 mil millones de dólares al año. Por ello, constituye el punto de partida del presente proyecto.

1.3 Objetivo del proyecto y problema a resolver

El presente proyecto tiene como finalidad optimizar la gestión de inventarios y mejorar el desempeño de las ventas de Walmart mediante la implementación de un sistema de predicción de la demanda basado en analítica de datos y modelos de Machine Learning. La propuesta surge de la necesidad de fortalecer la toma de decisiones dentro de la cadena de suministro, utilizando grandes volúmenes de información para anticipar el comportamiento de la demanda de los consumidores y planificar de manera más eficiente el abastecimiento de productos en cada uno de los puntos de venta.

Walmart es una de las empresas minoristas más grandes del mundo, con operaciones en múltiples países y un portafolio conformado por miles de productos pertenecientes a diferentes categorías. La magnitud de sus operaciones implica gestionar diariamente enormes cantidades de información relacionadas con ventas, inventarios, logística, proveedores, promociones, comportamiento de los clientes y tendencias del mercado. Debido a esta complejidad, resulta indispensable contar con herramientas analíticas que permitan transformar dichos datos en información útil para la toma de decisiones estratégicas y operativas.

El objetivo principal del proyecto consiste en desarrollar un modelo predictivo capaz de estimar con mayor precisión la demanda futura de los productos comercializados por la empresa. Para ello, se emplearán datos históricos de ventas junto con variables que influyen directamente en el comportamiento del consumidor, como la estacionalidad, las promociones comerciales, los días festivos, las condiciones climáticas, la ubicación geográfica de las tiendas, las tendencias de consumo y otros factores externos que pueden afectar el nivel de demanda. La integración de estas variables permitirá generar pronósticos más confiables que sirvan de apoyo para la planificación de compras, la distribución de mercancías y la administración eficiente del inventario.

El problema principal que busca resolver este proyecto es la presencia de sobrestock y desabastecimiento, dos situaciones que representan importantes desafíos para cualquier empresa del sector retail. El sobrestock ocurre cuando la empresa mantiene mayores cantidades de inventario de las necesarias para satisfacer la demanda real del mercado. Esta situación genera costos adicionales asociados al almacenamiento, manipulación, seguros, deterioro, obsolescencia y mantenimiento de los productos, además de inmovilizar recursos financieros que podrían destinarse a otras actividades estratégicas del negocio. Por otra parte, el desabastecimiento se presenta cuando la demanda supera la disponibilidad de inventario, ocasionando la pérdida de ventas, afectando la experiencia del cliente y favoreciendo que los consumidores opten por adquirir los productos en empresas competidoras.

Estas problemáticas suelen originarse por pronósticos de demanda poco precisos o por la utilización de métodos tradicionales que no logran capturar la complejidad y variabilidad del comportamiento de los consumidores. En un entorno altamente dinámico como el retail de consumo masivo, la demanda puede modificarse constantemente debido a campañas promocionales, cambios en los hábitos de compra, eventos especiales, temporadas del año, condiciones económicas e incluso factores externos inesperados. Como consecuencia, los modelos convencionales de planificación pueden resultar insuficientes para responder oportunamente a estas variaciones.

Frente a este escenario, la incorporación de técnicas de analítica predictiva y Machine Learning permite analizar grandes volúmenes de datos de manera automatizada, identificar patrones de comportamiento y generar pronósticos con un mayor nivel de precisión. Estos modelos tienen la capacidad de aprender continuamente a partir de la información histórica y adaptarse a nuevas condiciones del mercado, ofreciendo una herramienta de apoyo para la toma de decisiones basada en evidencia y no únicamente en la experiencia o el criterio humano.

La implementación de este proyecto permitirá optimizar los niveles de inventario en toda la cadena de suministro, asegurando que cada tienda disponga de la cantidad adecuada de productos en el momento oportuno. Esto contribuirá a reducir significativamente los costos derivados del exceso de inventario, minimizar las pérdidas ocasionadas por la falta de productos, mejorar la eficiencia logística y optimizar la utilización de los recursos disponibles. Asimismo, facilitará una mejor coordinación entre las áreas de compras, abastecimiento, logística, almacenamiento y ventas, fortaleciendo la capacidad de respuesta de la organización frente a las variaciones del mercado.

Desde una perspectiva estratégica, la solución también favorecerá el incremento de las ventas y la rentabilidad de Walmart, ya que una mejor disponibilidad de productos reduce las oportunidades de venta perdidas y mejora la satisfacción de los clientes al encontrar los artículos que buscan cuando los necesitan. De igual manera, disponer de información predictiva confiable permitirá realizar negociaciones más eficientes con proveedores, planificar campañas comerciales con mayor precisión y optimizar la distribución de mercancías entre centros de distribución y tiendas.

En conjunto, este proyecto busca demostrar cómo la analítica de datos puede convertirse en un elemento diferenciador dentro del sector retail, transformando la información generada diariamente por la operación en una ventaja competitiva sostenible. La aplicación de modelos predictivos no solo contribuirá a resolver los problemas asociados al sobrestock y al desabastecimiento, sino que también permitirá fortalecer la eficiencia operativa, mejorar la calidad de las decisiones empresariales y consolidar una cultura organizacional orientada al uso estratégico de los datos para generar valor y mantener el liderazgo competitivo de Walmart en el mercado global.

1.4 Áreas y personas que demandan información y uso actual de la misma

La gestión eficiente de inventarios y la planificación de la demanda en Walmart requieren la participación coordinada de múltiples áreas funcionales, cada una con necesidades específicas de información para apoyar la toma de decisiones. Actualmente, la empresa dispone de grandes volúmenes de datos provenientes de ventas, inventarios, logística, compras, comercio electrónico y comportamiento de los clientes; sin embargo, esta información suele encontrarse distribuida en distintos sistemas operacionales y reportes especializados.

Como consecuencia, cada área utiliza la información de acuerdo con sus responsabilidades operativas y estratégicas, apoyándose principalmente en datos históricos, indicadores de desempeño y reportes periódicos. Aunque este enfoque permite controlar la operación diaria, presenta limitaciones para anticipar cambios en la demanda y responder de manera proactiva a las condiciones del mercado.

Las principales áreas y los responsables que demandan información son las siguientes:

Tabla 5. *Áreas que demandan información y uso actual de la misma*

Área	Principales usuarios	Información utilizada actualmente	Uso actual de la información
Dirección General	CEO, Vicepresidentes, Directores Ejecutivos	Ventas, rentabilidad, costos operativos, indicadores financieros e inventarios.	Supervisar el desempeño global del negocio, evaluar resultados y definir estrategias corporativas.
Compras y Abastecimiento	Gerente de Compras, Planificadores de Demanda, Compradores	Históricos de ventas, niveles de inventario, órdenes de compra, proveedores y promociones.	Planificar compras, definir cantidades de abastecimiento y coordinar la reposición de productos.
Gestión de Inventarios	Gerente de Inventarios, Analistas de Inventario	Inventario disponible, inventarios de seguridad, productos de alta y baja rotación.	Controlar existencias, monitorear niveles de inventario y minimizar pérdidas por sobrestock o agotados.
Logística y Distribución	Gerente Logístico, Coordinadores de Centros de Distribución	Inventarios por almacén, rutas de distribución, pedidos y tiempos de entrega.	Coordinar la distribución de mercancías y optimizar el abastecimiento entre centros de distribución y tiendas.
Área Comercial y Ventas	Gerente Comercial, Gerentes Regionales, Jefes de Tienda	Ventas por tienda, categorías, promociones y comportamiento del consumidor.	Analizar tendencias comerciales, evaluar campañas y gestionar el cumplimiento de metas de ventas.

Área	Principales usuarios	Información utilizada actualmente	Uso actual de la información
Comercio Electrónico (E-commerce)	Gerente de E-commerce, Analistas Digitales	Pedidos en línea, inventario disponible y comportamiento de navegación.	Gestionar la disponibilidad de productos en los canales digitales y mejorar la experiencia del cliente.
Marketing	Gerente de Marketing, Analistas de Mercado	Promociones, campañas, comportamiento de compra y segmentación de clientes.	Diseñar campañas comerciales y evaluar el impacto de las promociones sobre la demanda.
Tecnología y Business Intelligence	CIO, Arquitectos de Datos, Ingenieros de Datos, Analistas BI	Bases de datos transaccionales, Data Warehouse, reportes corporativos e indicadores.	Integrar información, desarrollar modelos analíticos y generar reportes para la toma de decisiones.
Finanzas	CFO, Gerente Financiero, Analistas Financieros	Costos operativos, flujo de caja, inversión en inventarios y márgenes.	Evaluar el impacto financiero de las decisiones relacionadas con inventarios y abastecimiento.

Nota. Elaboración propia.

1.4.1 Flujo actual de utilización de la información

Actualmente, la información se genera principalmente a partir de los sistemas transaccionales de Walmart, incluyendo las ventas en tiendas físicas y canales digitales, la gestión de inventarios, las compras a proveedores y las operaciones logísticas. Estos datos son consolidados mediante reportes periódicos que permiten a cada área monitorear el desempeño de sus procesos y apoyar la toma de decisiones operativas.

No obstante, el análisis se basa en gran medida en información histórica y descriptiva, lo que dificulta anticipar cambios en el comportamiento de la demanda y responder oportunamente a factores externos como promociones, estacionalidad, eventos especiales o variaciones en las preferencias de los consumidores. Esta situación limita la capacidad de optimizar los niveles de inventario y aumenta el riesgo de sobrestock o desabastecimiento.

Por ello, la implementación de una solución de Business Intelligence permitirá integrar la información proveniente de todas las áreas en una única plataforma analítica, facilitando la generación de indicadores en tiempo real, modelos predictivos de demanda y tableros ejecutivos que fortalezcan la toma de decisiones en toda la organización.

1.5 Objetivos de cada área con la información actual y necesidades detectadas

La implementación de un modelo de Business Intelligence para la optimización de inventarios en Walmart busca alinear todas las áreas estratégicas de la organización mediante el uso de información integrada, confiable y en tiempo real. Actualmente, gran parte de las decisiones dependen de datos históricos y reportes operativos, lo que limita la capacidad de anticipar cambios en la demanda y responder oportunamente a las variaciones del mercado. El proyecto propone transformar la toma de decisiones mediante modelos predictivos que permitan optimizar el abastecimiento y mejorar el desempeño de toda la cadena de suministro.

1.5.1 Dirección General

Información utilizada actualmente

La alta dirección monitorea indicadores financieros y operativos como ventas, rentabilidad, rotación de inventario, costos logísticos y nivel de servicio mediante reportes consolidados provenientes de distintas áreas.

Necesidades detectadas

- Información consolidada en tiempo real.
- Pronósticos confiables para la planificación estratégica.
- Identificación temprana de riesgos comerciales.
- Medición del impacto financiero de las decisiones sobre inventario.
- Simulación de escenarios futuros.

Objetivos

- Implementar un sistema integral de BI que permita visualizar el desempeño global del negocio.
- Mejorar la planificación estratégica utilizando modelos predictivos de demanda.
- Incrementar la rentabilidad mediante decisiones basadas en datos.
- Reducir el capital inmovilizado en inventarios.
- Mejorar la capacidad de respuesta ante cambios del mercado.

1.5.2 Área de Compras y Abastecimiento

Información utilizada actualmente

El área realiza pedidos considerando ventas históricas, niveles mínimos de inventario, promociones programadas y experiencia del personal.

Necesidades detectadas

- Predicciones más precisas por producto.
- Identificación automática de productos con riesgo de sobrestock.
- Alertas tempranas de desabastecimiento.
- Integración con información de proveedores.

- Priorización inteligente de compras.

Objetivos

- Automatizar la generación de órdenes de compra mediante modelos predictivos.
- Reducir las compras innecesarias.
- Optimizar la frecuencia de abastecimiento.
- Mejorar la disponibilidad de productos críticos.
- Disminuir costos de almacenamiento.

1.5.3 Gestión de Inventarios

Información utilizada actualmente

El control se realiza mediante niveles máximos y mínimos, inventarios físicos y reportes diarios de existencias.

Necesidades detectadas

- Visualización en tiempo real del inventario.
- Predicción de agotamientos.
- Clasificación dinámica de productos.
- Optimización del inventario de seguridad.
- Monitoreo automático de productos de baja rotación.

Objetivos

- Mantener niveles óptimos de inventario.
- Reducir el sobrestock.
- Disminuir pérdidas por vencimiento.
- Mejorar la rotación de inventario.
- Optimizar la utilización del espacio en centros de distribución.

1.5.4 Logística y Distribución

Información utilizada actualmente

Se utilizan planes de despacho, inventarios disponibles y programación de transporte.

Necesidades detectadas

- Priorización automática de rutas.
- Predicción de necesidades logísticas futuras.
- Balanceo de inventario entre centros de distribución.
- Optimización del transporte.
- Mejor utilización de la capacidad instalada.

Objetivos

- Reducir tiempos de entrega.
- Disminuir costos logísticos.

- Optimizar la distribución regional.
- Evitar traslados innecesarios entre almacenes.
- Mejorar el nivel de servicio.

1.5.5 Área Comercial y Ventas

Información utilizada actualmente

Analiza ventas por tienda, promociones, comportamiento histórico y participación por categoría.

Necesidades detectadas

- Predicción del comportamiento de compra.
- Identificación de tendencias emergentes.
- Evaluación del impacto de promociones.
- Segmentación avanzada de clientes.
- Recomendaciones comerciales basadas en IA.

Objetivos

- Incrementar las ventas.
- Reducir pérdidas por falta de inventario.
- Optimizar campañas promocionales.
- Personalizar las estrategias comerciales.
- Maximizar la satisfacción del cliente.

1.5.6 Comercio Electrónico (E-commerce)

Información utilizada actualmente

Monitorea pedidos en línea, disponibilidad de inventario y comportamiento de navegación.

Necesidades detectadas

- Sincronización completa entre inventarios físicos y digitales.
- Predicción de demanda por canal.
- Optimización del surtido online.
- Recomendaciones automáticas de productos.
- Integración omnicanal.

Objetivos

- Garantizar disponibilidad en tiempo real.
- Reducir cancelaciones por falta de stock.
- Mejorar la experiencia del cliente digital.
- Incrementar la conversión de ventas.
- Optimizar la logística del canal online.

1.5.7 Marketing

Información utilizada actualmente

Utiliza información de promociones, campañas y comportamiento histórico de ventas.

Necesidades detectadas

- Medición del retorno de inversión de campañas.
- Segmentación predictiva de clientes.
- Pronóstico del impacto de promociones.
- Integración de información de clientes.
- Modelos de comportamiento del consumidor.

Objetivos

- Diseñar campañas basadas en datos.
- Mejorar la efectividad promocional.
- Incrementar la fidelización.
- Optimizar la inversión publicitaria.
- Aumentar las ventas por campañas dirigidas.

1.5.8 Tecnología y Business Intelligence

Información utilizada actualmente

Administra múltiples bases de datos provenientes de ventas, inventarios, logística y proveedores.

Necesidades detectadas

- Integración de fuentes de datos.
- Automatización del procesamiento.
- Calidad y gobierno de datos.
- Desarrollo de modelos predictivos.
- Dashboards ejecutivos en tiempo real.

Objetivos

- Implementar un Data Warehouse corporativo.
- Automatizar procesos ETL.
- Desarrollar modelos de Machine Learning para predicción de demanda.
- Garantizar la calidad y consistencia de la información.
- Proveer dashboards dinámicos para todas las áreas.

1.5.9 Finanzas

Información utilizada actualmente

Analiza costos operativos, flujo de caja, inversión en inventarios y rentabilidad.

Necesidades detectadas

- Proyección financiera basada en demanda futura.
- Cálculo del costo del sobrestock.
- Evaluación del retorno de inversión del proyecto BI.
- Optimización del capital de trabajo.
- Simulación de escenarios financieros.

Objetivos

- Reducir el capital inmovilizado en inventarios.
- Disminuir costos operativos.
- Mejorar el flujo de caja.
- Incrementar el margen operativo.
- Medir el impacto económico de las decisiones de abastecimiento.

1.6 Indicadores clave de desempeño (KPIs)

Walmart Inc. es una corporación minorista multinacional con sede en Bentonville, Arkansas, que opera a través de los segmentos Walmart U.S., Walmart International y Sam's Club. Para dirigir su desempeño y comunicar resultados a inversionistas, Walmart organiza sus indicadores clave de desempeño (KPIs) en tres prioridades financieras estratégicas: Crecimiento (Growth), Margen (Margin) y Retornos (Returns), tal como lo describe en sus reportes anuales presentados ante la Comisión de Bolsa y Valores de Estados Unidos (SEC). A esto se suman KPIs operativos específicos para la cadena de suministro, los proveedores y los vendedores del Marketplace.

1.6.1 Marco estratégico de KPIs corporativos

Walmart define explícitamente tres prioridades financieras que agrupan sus principales métricas de desempeño:

- **Crecimiento (Growth)**: servir a los clientes mediante una experiencia omnicanal fluida, medida principalmente con las ventas comparables (comparable sales) y el crecimiento de las ventas de comercio electrónico
- **Margen (Margin)**: mejorar el margen de utilidad operativa a través de iniciativas de productividad y de la combinación de categorías y negocios
- **Retornos (Returns)**: mejorar el Retorno sobre la Inversión (ROI) mediante la mejora del margen y una inversión de capital disciplinada

En el año fiscal 2026, Walmart reportó ventas netas de 706 413 millones de dólares (crecimiento de 4,7 % respecto al año anterior), un margen bruto de 24,2 % sobre ventas netas y un ingreso operativo de 29 825 millones de dólares, equivalente a un margen operativo de 4,2 %.

1.6.2 Tabla resumen de KPIs por categoría

Tabla 6. ***KPIs de Walmart por categoría***

Categoría	KPI	Qué mide
Crecimiento	Ventas comparables (Comparable Sales)	Variación de ventas en tiendas/clubes abiertos hace más de 12 meses, incluyendo e-commerce
Crecimiento	Crecimiento de ventas de e-commerce	Ritmo de expansión de las ventas digitales y omnicanal
Crecimiento	Membresías (Sam's Club, Walmart+)	Número y crecimiento de miembros activos
Margen / Rentabilidad	Margen bruto (Gross Profit Rate)	Utilidad bruta como porcentaje de las ventas netas
Margen / Rentabilidad	Margen operativo (Operating Income Margin)	Utilidad operativa como porcentaje de las ventas netas
Margen / Rentabilidad	Apalancamiento de gastos (Expense Leverage)	Que las ventas netas crezcan más rápido que los gastos operativos
Retornos financieros	Retorno sobre Activos (ROA)	Utilidad neta dividida entre el promedio de activos totales
Retornos financieros	Retorno sobre la Inversión (ROI)	Métrica no-GAAP que evalúa qué tan eficientemente se despliega el capital invertido
Retornos financieros	Flujo de caja libre (Free Cash Flow)	Efectivo generado por operaciones menos gastos de capital
Cadena de suministro / Proveedores	OTIF (On Time, In Full)	Porcentaje de pedidos entregados a tiempo y con la cantidad completa
Cadena de suministro / Proveedores	Fill Rate	Porcentaje del pedido que el proveedor logra surtir
Cadena de suministro / Proveedores	GMROII	Retorno bruto sobre la inversión en inventario
Cadena de suministro / Proveedores	AUR (Average Unit Retail)	Precio de venta promedio por unidad
Cadena de suministro / Proveedores	Semanas de suministro (Weeks of Supply)	Nivel de inventario disponible expresado en semanas de venta
Marketplace / Vendedores	Tasa de cancelación del vendedor	Porcentaje de pedidos cancelados por el vendedor (meta: menor a 2%)
Marketplace / Vendedores	Tasa de Pedidos Defectuosos (ODR)	Pedidos con cancelación, devolución o defecto de envío sobre el total de compras
Marketplace / Vendedores	Valid Tracking Rate	Porcentaje de pedidos con información de rastreo válida y a tiempo

Nota. Elaboración propia con base en Walmart Inc. (2026) y documentación pública de Walmart para proveedores.

1.6.3 KPIs de crecimiento

La métrica central de crecimiento es la venta comparable ("same-store sales"), que mide el cambio en las ventas de tiendas y clubes abiertos durante los últimos 12 meses, incluyendo remodelaciones, reubicaciones, expansiones y conversiones, así como las ventas de comercio electrónico. Walmart también da seguimiento al crecimiento de membresías en Sam's Club y Walmart+, y a la expansión de su ecosistema de publicidad digital.

1.6.4 KPIs de margen y rentabilidad

Walmart mide su rentabilidad principalmente con el margen bruto (gross profit rate), el margen operativo (operating income as a percentage of net sales) y el apalancamiento de gastos (expense leverage), definido como el crecimiento de las ventas netas a un ritmo mayor que el de los gastos operativos. La compañía también busca el "operating income leverage", es decir, que el ingreso operativo crezca más rápido que las ventas netas.

1.6.5 KPIs de retorno financiero

Para evaluar el retorno sobre sus activos y capital invertido, Walmart utiliza el Retorno sobre Activos (ROA), calculado conforme a los principios contables generalmente aceptados (GAAP), y el Retorno sobre la Inversión (ROI), una medida no-GAAP que la administración considera más representativa de la eficiencia con la que despliega sus activos clave. Estas métricas se complementan con el flujo de caja libre y la política de recompra de acciones y dividendos.

1.6.6 KPIs de cadena de suministro y proveedores

A través de su plataforma de datos Scintilla, Walmart comparte con sus proveedores un conjunto amplio de métricas operativas. Entre las más relevantes destacan:

- **OTIF (On Time, In Full)**: el indicador insignia de la cadena de suministro; mide si los envíos llegaron a tiempo (On Time) y con la cantidad completa (In Full). Un OTIF deficiente puede generar multas, reducción de pedidos o pérdida de privilegios de entrega directa a tienda
- **AUR (Average Unit Retail)**: el precio de venta promedio por unidad, resultado de dividir las ventas totales entre las unidades vendidas
- **GMROII**: el retorno bruto de margen sobre la inversión en inventario
- **Fill Rate y semanas de suministro**: indicadores de disponibilidad de producto y nivel de inventario

1.6.7 KPIs del Marketplace (vendedores externos)

Para los vendedores externos que operan en Walmart Marketplace, la compañía evalúa el desempeño mediante un "Scorecard" enfocado en la experiencia del cliente. Los indicadores principales incluyen la tasa de cancelación del vendedor (con una meta inferior al 2%), la Tasa de Pedidos Defectuosos o ODR (Order

Defect Rate, que combina cancelaciones, devoluciones y defectos de envío) y el Valid Tracking Rate, que mide la disponibilidad de información de rastreo válida para los pedidos.

CAPÍTULO 2. Estructura de la gestión de datos del proyecto

Una vez definido el problema, el siguiente paso es comprender cómo fluye la información que alimentará el modelo de predicción de la demanda. Este capítulo describe el origen de los datos y su recurrencia de carga, los sistemas que los generan y su destino, la estructura de almacenamiento en Data Warehouse y Data Marts, las carencias detectadas en la carga y la brecha entre los informes actuales y las necesidades del negocio.

2.1 Origen de la información y recurrencia de carga: interna y externa

El éxito de un proyecto de predicción de la demanda depende, en gran medida, de la disponibilidad y calidad de la información utilizada para alimentar los modelos analíticos. En el caso de Walmart, una empresa que opera a escala global y administra millones de transacciones diariamente, la información constituye uno de los activos más importantes para la toma de decisiones estratégicas y operativas.

La información empleada para este proyecto proviene de una combinación de fuentes internas y externas que permiten comprender de manera integral el comportamiento del negocio y de sus consumidores. Las fuentes internas representan la mayor parte del volumen de datos disponible y se generan como resultado de las operaciones diarias de la compañía. Entre las principales fuentes se encuentran los sistemas de punto de venta (POS), que registran cada transacción realizada en las tiendas físicas; los sistemas ERP, encargados de la administración financiera y operativa; los sistemas de gestión de almacenes (WMS); las plataformas de comercio electrónico; los registros de inventarios; los sistemas de abastecimiento y compras; la información proveniente de proveedores; y los programas de fidelización y análisis del cliente.

Cada una de estas fuentes aporta información valiosa para el análisis. Por ejemplo, los sistemas POS permiten conocer qué productos se venden, en qué momento, en qué tienda y bajo qué condiciones comerciales. Por su parte, los sistemas de inventarios facilitan la identificación de niveles de stock, rotación de productos y puntos de reposición. Asimismo, la información logística permite evaluar tiempos de entrega, eficiencia de los centros de distribución y desempeño de la cadena de suministro.

Las fuentes externas complementan el análisis interno y permiten incorporar variables que influyen en la demanda, pero que no son generadas directamente por la organización. Entre ellas destacan los datos meteorológicos, calendarios de días festivos, indicadores económicos, índices de inflación, niveles de desempleo, tendencias de consumo, información demográfica, datos de movilidad urbana y análisis provenientes de redes sociales y motores de búsqueda. Por ejemplo, un aumento en las temperaturas puede incrementar la demanda de bebidas, mientras que una festividad nacional puede ocasionar un incremento significativo en la compra de determinados productos.

En cuanto a la recurrencia de carga, Walmart opera en un entorno donde la velocidad de actualización de la información es fundamental. Las ventas realizadas en las tiendas físicas y plataformas digitales pueden ser procesadas prácticamente en tiempo real, permitiendo monitorear el comportamiento de la demanda minuto a minuto. Sin embargo, otras fuentes, como los indicadores macroeconómicos o los estudios de mercado, suelen actualizarse de forma semanal, mensual o trimestral.

Para efectos de este proyecto, se propone un esquema híbrido de actualización. Los datos transaccionales relacionados con ventas, inventarios y logística se cargarán cada 24 horas, mientras que las variables externas serán actualizadas de acuerdo con su disponibilidad. Este enfoque garantiza que los modelos predictivos dispongan de información reciente y confiable, incrementando su capacidad para anticipar cambios en la demanda y generar recomendaciones oportunas para la gestión de inventarios.

2.2 Sistemas y destino de la carga

La infraestructura tecnológica de Walmart está compuesta por un ecosistema de sistemas empresariales altamente integrado que permite capturar, procesar y distribuir grandes volúmenes de información. Dada la complejidad de sus operaciones, la compañía requiere mecanismos robustos que aseguren la disponibilidad y consistencia de los datos en todas las áreas funcionales.

Entre los sistemas involucrados en la generación de información destacan los ERP corporativos, encargados de centralizar procesos financieros, administrativos y de abastecimiento; los sistemas POS utilizados en miles de establecimientos comerciales; los sistemas de gestión de almacenes (WMS); las soluciones SCM (Supply Chain Management) para la administración de la cadena de suministro; y las plataformas digitales que soportan las operaciones de comercio electrónico.

Una vez generados, los datos son sometidos a procesos ETL (Extract, Transform and Load), fundamentales dentro de cualquier arquitectura de Business Intelligence. Durante la fase de extracción, la información es recopilada desde las diferentes fuentes disponibles. Posteriormente, en la etapa de transformación, se ejecutan procesos de limpieza, normalización, validación y enriquecimiento de datos para garantizar su calidad. Finalmente, los datos son cargados en repositorios especializados destinados al análisis y la generación de conocimiento.

El destino de la carga se encuentra constituido por una arquitectura analítica compuesta por Data Warehouses, Data Lakes y Data Marts. Estos repositorios permiten almacenar grandes volúmenes de información histórica y facilitan su acceso para distintos usuarios dentro de la organización.

Adicionalmente, los datos son consumidos por herramientas de visualización y análisis como Power BI, Tableau y plataformas de Inteligencia Artificial que permiten desarrollar modelos predictivos y dashboards interactivos. De esta

manera, la información deja de ser un recurso aislado para convertirse en un activo estratégico que impulsa la toma de decisiones basada en evidencia.

Desde la perspectiva del presente proyecto, el principal destino de la información será un entorno analítico capaz de alimentar modelos de Machine Learning diseñados para predecir la demanda futura. Esto permitirá que las áreas de compras, logística y abastecimiento dispongan de información precisa para optimizar sus operaciones y responder de manera más eficiente a las necesidades del mercado.

2.3 Almacenamiento en Data Warehouse y Data Marts

La consolidación y almacenamiento adecuado de la información constituye uno de los pilares fundamentales de cualquier estrategia de analítica de datos. En organizaciones del tamaño de Walmart, donde diariamente se generan millones de registros, resulta indispensable contar con mecanismos que permitan integrar y organizar la información de manera eficiente.

El Data Warehouse representa el núcleo de la arquitectura analítica propuesta para este proyecto. Se trata de un repositorio centralizado diseñado para almacenar información histórica proveniente de múltiples fuentes y proporcionar una visión unificada del negocio. A diferencia de las bases de datos operacionales, cuyo objetivo principal es soportar las transacciones del día a día, el Data Warehouse está orientado al análisis y a la toma de decisiones.

Dentro de este entorno se almacenarán datos relacionados con ventas históricas, inventarios, movimientos logísticos, información financiera, comportamiento de clientes, desempeño de proveedores y métricas operacionales. La integración de estos elementos permitirá construir una base sólida para el desarrollo de modelos predictivos y análisis avanzados.

Además del Data Warehouse, se contempla la implementación de diversos Data Marts especializados, los cuales permiten segmentar la información según las necesidades específicas de cada área funcional. Esta estrategia mejora la eficiencia en las consultas y facilita el acceso a información relevante para cada unidad de negocio.

Entre los Data Marts considerados para el proyecto destacan:

- **Data Mart de Ventas:** enfocado en el análisis de ingresos, productos más vendidos, estacionalidad y desempeño por tienda.
- **Data Mart de Inventarios:** orientado al monitoreo de niveles de stock, rotación, productos obsoletos y desabastecimientos.
- **Data Mart Logístico:** encargado del análisis de tiempos de entrega, desempeño de centros de distribución y eficiencia operativa.
- **Data Mart Financiero:** destinado al seguimiento de costos, márgenes de utilidad y rentabilidad.
- **Data Mart de Clientes:** especializado en segmentación, hábitos de compra y análisis de comportamiento del consumidor.

La implementación de esta arquitectura permitirá disponer de información estructurada, consistente y fácilmente accesible, reduciendo significativamente los tiempos de procesamiento y mejorando la capacidad de respuesta de la organización frente a cambios en la demanda.

2.4 Carencias y problemas en la carga de datos

A pesar de los avances tecnológicos existentes en el sector retail, la gestión de datos continúa enfrentando importantes desafíos asociados al volumen, velocidad y variedad de la información. En el caso de Walmart, la magnitud de sus operaciones incrementa la complejidad de los procesos de integración y procesamiento de datos.

Uno de los principales problemas identificados es la heterogeneidad de las fuentes de información. Los datos pueden provenir de distintos sistemas, países y unidades de negocio, cada uno con estructuras y formatos particulares. Esta diversidad dificulta la consolidación de la información y aumenta el riesgo de inconsistencias.

Otro desafío importante corresponde a la calidad de los datos. La existencia de registros duplicados, campos incompletos, errores de captura y diferencias en los criterios de clasificación puede afectar directamente la precisión de los análisis y modelos predictivos. Incluso pequeñas inconsistencias pueden generar desviaciones significativas en los pronósticos de demanda.

Asimismo, la integración de información en tiempo real representa un reto considerable. La necesidad de procesar millones de transacciones diariamente exige una infraestructura tecnológica robusta capaz de garantizar la disponibilidad y actualización constante de los datos. Cualquier retraso en los procesos ETL puede impactar negativamente la capacidad de respuesta de la organización.

Entre las principales carencias identificadas se encuentran:

- Falta de estandarización en determinados conjuntos de datos.
- Retrasos en la disponibilidad de información externa.
- Dificultades para integrar datos históricos con información en tiempo real.
- Incremento en los costos de almacenamiento y procesamiento.
- Limitaciones relacionadas con la gobernanza y seguridad de los datos.
- Complejidad para consolidar información proveniente de operaciones internacionales.
- Dependencia de múltiples sistemas y proveedores tecnológicos.
- Riesgo de sesgos en los modelos predictivos derivados de datos incompletos.

Por este motivo, el proyecto contempla la implementación de mecanismos de control de calidad, políticas de gobierno de datos y procesos automatizados de

validación que permitan garantizar la integridad, consistencia y confiabilidad de la información utilizada en los análisis.

2.5 Informes que se generan actualmente, necesidades detectadas y no cubiertas

Actualmente, Walmart genera una amplia variedad de informes orientados a monitorear el desempeño de sus operaciones y facilitar la toma de decisiones. Estos reportes abarcan diferentes niveles organizacionales y son utilizados tanto por equipos operativos como por la alta dirección.

Entre los informes más relevantes se encuentran los reportes diarios de ventas, análisis de inventarios, indicadores logísticos, estados financieros, métricas de rentabilidad, desempeño de proveedores, análisis de comercio electrónico y reportes de satisfacción del cliente. La mayoría de estos documentos ofrecen una visión descriptiva del negocio, permitiendo comprender qué ocurrió durante un período determinado.

Sin embargo, la creciente complejidad del entorno empresarial ha puesto de manifiesto la necesidad de evolucionar desde una analítica descriptiva hacia una analítica predictiva y prescriptiva. En otras palabras, ya no basta con conocer lo sucedido; las organizaciones requieren herramientas capaces de anticipar eventos futuros y recomendar acciones concretas.

Entre las necesidades detectadas destacan:

- Pronósticos de demanda más precisos por producto, tienda y región.
- Identificación temprana de riesgos de desabastecimiento.
- Sistemas de alerta automática para productos con alta probabilidad de sobrestock.
- Visualización en tiempo real de indicadores críticos.
- Integración de variables externas en el análisis de la demanda.
- Automatización de decisiones relacionadas con la reposición de inventarios.
- Capacidad para simular escenarios futuros y evaluar su impacto.
- Mayor capacidad para identificar patrones ocultos en grandes volúmenes de información.

Estas necesidades evidencian una brecha entre la información actualmente disponible y el conocimiento requerido para enfrentar los desafíos del mercado moderno. En este contexto, el proyecto propone incorporar herramientas de analítica avanzada que permitan transformar los datos en información accionable.

La implementación de dashboards interactivos, modelos de Machine Learning y sistemas inteligentes de recomendación permitirá a Walmart disponer de una plataforma integral para la gestión predictiva de inventarios. Como resultado, la organización podrá mejorar la precisión de sus decisiones, incrementar la

eficiencia operativa, reducir costos y fortalecer su posición competitiva en el sector retail.

En conclusión, la adopción de una estrategia basada en datos representa una oportunidad para evolucionar desde una gestión reactiva hacia un modelo de negocio proactivo, capaz de anticiparse a las necesidades de los consumidores y responder de manera eficiente a un entorno cada vez más dinámico y competitivo.

CAPÍTULO 3. Arquitectura de datos, sistemas implicados y recursos de IT

Este capítulo profundiza en la arquitectura de datos de Walmart: los procesos de carga, la calidad de la información disponible, las variables que residen en el Data Warehouse y el modelo de acceso, los sistemas implicados en el flujo de información y los recursos tecnológicos internos y externos con los que cuenta la compañía.

Conviene precisar que Walmart opera bajo un modelo multinube (denominado «Triplet Model») que combina su nube privada con Microsoft Azure y Google Cloud Platform. Por ello, algunas capacidades descritas en este capítulo, como el uso de BigQuery para la gestión de inventario y los insights de clientes, corresponden a Google Cloud. Para el departamento propuesto, la plataforma de referencia es Microsoft Azure, decisión que se justifica en el Capítulo 2 de la Parte II.

3.1 Procesos de carga de datos (ETL y Data Lake)

La capacidad de Walmart para administrar una operación global con miles de tiendas, millones de clientes y una amplia red de centros de distribución depende, en gran medida, de su infraestructura tecnológica y de su capacidad para procesar grandes volúmenes de información. En sus informes anuales y comunicaciones corporativas, la empresa señala que la transformación digital constituye uno de los pilares de su estrategia de crecimiento, motivo por el cual ha incrementado de manera sostenida sus inversiones en inteligencia artificial, automatización, analítica avanzada y tecnologías para la cadena de suministro. Estas inversiones buscan convertir los datos generados diariamente por la organización en información útil para optimizar la toma de decisiones y mejorar la eficiencia operativa. Aunque Walmart no divulga públicamente la arquitectura detallada de sus procesos de integración de datos, sí confirma que utiliza tecnologías avanzadas para conectar sus operaciones físicas y digitales y respaldar decisiones críticas relacionadas con inventarios, abastecimiento, logística y experiencia del cliente.

De acuerdo con el **Annual Report 2025**, Walmart continúa desarrollando una infraestructura tecnológica capaz de integrar información proveniente de sus tiendas físicas, plataformas de comercio electrónico, centros de distribución, operaciones logísticas y servicios digitales. La empresa destaca que uno de sus principales objetivos es construir una organización "people-led, tech-powered", donde la tecnología permita mejorar la productividad de los colaboradores, optimizar la cadena de suministro y ofrecer una experiencia omnicanal consistente. Esta estrategia requiere la integración permanente de grandes volúmenes de información generados por millones de transacciones realizadas diariamente en diferentes países y canales de venta.

Entre las principales fuentes de información identificadas en los informes corporativos se encuentran las ventas registradas en los sistemas de punto de

venta (POS), las operaciones realizadas a través de Walmart.com y su aplicación móvil, los movimientos de inventario en tiendas y centros de distribución, la información logística relacionada con transporte y abastecimiento, los datos financieros, las operaciones de proveedores y la información generada por los servicios digitales ofrecidos a los clientes. Todos estos datos son utilizados para monitorear el desempeño del negocio, optimizar los niveles de inventario, mejorar la disponibilidad de productos y fortalecer la capacidad de respuesta frente a las variaciones de la demanda.

Asimismo, Walmart ha informado que está incorporando soluciones basadas en inteligencia artificial generativa y aprendizaje automático para automatizar procesos dentro de la cadena de suministro. En publicaciones recientes, la compañía explica que estas herramientas permiten analizar grandes volúmenes de información para anticipar cambios en la demanda, optimizar la distribución de mercancías entre centros logísticos, reducir desperdicios y mejorar la disponibilidad de productos en las tiendas. Estas capacidades demuestran que la empresa dispone de plataformas capaces de procesar información de forma continua y de convertir los datos operativos en decisiones accionables para distintas áreas del negocio.

A pesar de la amplia información estratégica que Walmart publica, la empresa **no revela** el funcionamiento interno de sus procesos ETL (Extract, Transform and Load), ni la arquitectura específica utilizada para integrar la información. Tampoco hace públicos los procesos mediante los cuales extrae los datos de sus sistemas operacionales, realiza la limpieza y validación de la información o la almacena en repositorios analíticos. Esta ausencia de información responde principalmente a razones de seguridad informática, protección de la propiedad intelectual y confidencialidad tecnológica, ya que estos procesos forman parte de las ventajas competitivas de la organización.

Sin embargo, considerando las mejores prácticas internacionales en Business Intelligence y el volumen de operaciones que Walmart administra diariamente, es razonable inferir que la empresa emplea procesos ETL altamente automatizados. Estos procesos tendrían como finalidad extraer información desde múltiples sistemas empresariales, transformarla mediante procesos de depuración, estandarización, integración y validación, y posteriormente cargarla en plataformas analíticas utilizadas por los equipos de negocio y los modelos de inteligencia artificial. Esta inferencia no representa una afirmación oficial de Walmart, sino una propuesta técnica sustentada en la literatura especializada y en las capacidades tecnológicas que la propia compañía describe en sus informes públicos.

De igual manera, Walmart no confirma públicamente el uso de un **Data Lake** específico ni detalla su arquitectura de almacenamiento de datos. No obstante, la empresa sí comunica que trabaja con enormes volúmenes de información provenientes de diferentes fuentes estructuradas y no estructuradas, además de utilizar inteligencia artificial y analítica avanzada para transformar dichos datos en conocimiento útil. Desde una perspectiva académica, estas características

son compatibles con el uso de arquitecturas modernas de almacenamiento, donde un Data Lake complementa al Data Warehouse permitiendo conservar información sin procesar para su posterior explotación mediante modelos de Machine Learning, analítica predictiva y sistemas inteligentes de apoyo a la decisión.

En conjunto, la información oficial disponible permite concluir que Walmart ha construido una infraestructura tecnológica altamente sofisticada para integrar, procesar y analizar datos a gran escala. Aunque la empresa no publica los detalles técnicos de sus procesos ETL ni de su arquitectura de datos, las inversiones reportadas en inteligencia artificial, automatización y transformación digital evidencian que la integración de datos constituye un componente estratégico para optimizar su cadena de suministro, mejorar la gestión del inventario y fortalecer la capacidad predictiva de la organización.

3.2 Información y calidad de la información disponible

Uno de los principales activos estratégicos de Walmart es la enorme cantidad de información que genera diariamente como resultado de sus operaciones globales. Con presencia en múltiples países, miles de establecimientos comerciales y millones de clientes que interactúan tanto en tiendas físicas como en plataformas digitales, la compañía produce un flujo constante de datos que sirve de base para la planificación operativa, la toma de decisiones estratégicas y el desarrollo de soluciones basadas en inteligencia artificial. Los informes anuales de Walmart destacan que el análisis de datos se ha convertido en un elemento fundamental para mejorar la experiencia del cliente, optimizar la cadena de suministro y aumentar la eficiencia de sus operaciones.

La información disponible dentro de la organización abarca prácticamente todas las áreas del negocio. Entre los principales conjuntos de datos se encuentran las ventas netas por segmento de negocio, las ventas comparables (Comparable Sales), los niveles de inventario, la rotación de productos, los costos de operación, los resultados financieros, la información proveniente del comercio electrónico, el desempeño de las operaciones internacionales, los indicadores relacionados con publicidad digital, las métricas de cumplimiento logístico y los resultados obtenidos por cada unidad de negocio. Además, Walmart recopila información derivada del comportamiento de compra de los clientes, las interacciones realizadas en sus plataformas digitales y el desempeño de su red global de proveedores y centros de distribución.

En sus comunicaciones corporativas, la empresa también explica que la inteligencia artificial desempeña un papel cada vez más importante en el procesamiento y análisis de estos datos. Walmart ha desarrollado herramientas capaces de analizar grandes volúmenes de información para optimizar la reposición de inventarios, mejorar la disponibilidad de productos, automatizar tareas dentro de la cadena de suministro y apoyar la toma de decisiones mediante modelos predictivos. La utilización de estas tecnologías permite identificar patrones de comportamiento, detectar tendencias de consumo y

anticipar cambios en la demanda antes de que estos impacten las operaciones comerciales.

La calidad de la información representa un requisito indispensable para el funcionamiento de estas soluciones tecnológicas. Aunque Walmart no publica indicadores específicos relacionados con la calidad de sus bases de datos, la compañía reconoce en sus informes que la confiabilidad, disponibilidad y seguridad de la información constituyen elementos esenciales para el éxito de su estrategia digital. Los sistemas de inteligencia artificial únicamente pueden generar resultados precisos cuando la información utilizada es consistente, completa, actualizada y correctamente integrada. Por esta razón, la empresa mantiene inversiones permanentes en infraestructura tecnológica, automatización y plataformas digitales que permitan gestionar de forma eficiente el enorme volumen de datos generado por sus operaciones.

Otro aspecto relevante es la incorporación de información proveniente de diversas fuentes para enriquecer los procesos analíticos. Además de los datos internos relacionados con ventas e inventarios, Walmart utiliza información generada por sus plataformas digitales, operaciones logísticas y servicios financieros. En publicaciones recientes, la compañía también señala que sus soluciones de inteligencia artificial consideran variables relacionadas con condiciones operativas, disponibilidad de productos y comportamiento de la demanda para optimizar la distribución de mercancías entre centros logísticos y tiendas. Esto demuestra que el análisis de datos dentro de Walmart no se limita únicamente a información financiera, sino que integra múltiples dimensiones del negocio para obtener una visión más completa del funcionamiento de la organización.

No obstante, es importante señalar que Walmart **no divulga públicamente** sus políticas internas de gobierno de datos, los procedimientos específicos utilizados para validar la calidad de la información ni las metodologías empleadas para depurar registros antes de ser utilizados por sus plataformas analíticas. Tampoco publica indicadores como porcentaje de registros inconsistentes, tiempos de validación, métricas de integridad o niveles de calidad alcanzados por sus bases de datos. En consecuencia, cualquier descripción detallada de estos procesos excedería la información disponible en las fuentes oficiales.

Desde una perspectiva académica, puede afirmarse que la elevada precisión operativa demostrada por Walmart, junto con el despliegue de soluciones avanzadas de inteligencia artificial y automatización, evidencia la existencia de procesos sólidos de gestión y control de la calidad de los datos. Sin una información confiable y correctamente integrada, sería inviable operar modelos predictivos a la escala que la empresa describe en sus informes corporativos. Por ello, aunque los mecanismos internos no sean públicos, las inversiones reportadas y los resultados obtenidos permiten concluir que la calidad de la información constituye un elemento estratégico dentro del modelo de negocio de Walmart y uno de los principales habilitadores de su transformación digital.

3.2.1 Horarios de carga y el modelo de cross-docking

Uno de los pilares del sistema logístico de Walmart es el llamado cross-docking, un método mediante el cual la mercancía que llega de los proveedores se transfiere casi de inmediato a los camiones de salida, minimizando el tiempo de almacenamiento (Vector, 2026). Este proceso exige una coordinación precisa entre la descarga de un camión proveedor y la carga de un camión de reparto en el mismo centro de distribución, generalmente dentro de un margen de un día (Vector, 2026). Dicha sincronización solo es posible gracias al intercambio de datos en tiempo real y al uso de tecnología de rastreo, como el seguimiento satelital y las etiquetas de identificación por radiofrecuencia (RFID), que permiten programar la recepción de productos de manera prácticamente instantánea (Vector, 2026).

Las etiquetas RFID pueden leerse a distancia mientras las tarimas y cajas se desplazan a través de un portal ubicado en el propio muelle de carga, sin necesidad de detener la mercancía para escanearla manualmente (Vector, 2026). Gracias a esta tecnología, los productos con RFID se reponen hasta tres veces más rápido que aquellos que solo utilizan código de barras, y los quiebres de inventario se han reducido en un 16% desde su implementación (Vector, 2026). A esto se suma el uso de sensores de Internet de las Cosas (IoT) e inteligencia artificial para el pronóstico de demanda, lo que permite ajustar de manera dinámica los horarios de reabastecimiento y optimizar los niveles de inventario en cada ubicación (LitCommerce, 2026).

3.2.2 Infraestructura tecnológica y accesibilidad

En el núcleo de la cadena de suministro de Walmart existe una infraestructura de IT centralizada que conecta tiendas, centros de distribución y proveedores mediante sistemas propietarios desarrollados internamente, lo cual permite una comunicación y sincronización fluida a lo largo de toda la red (LitCommerce, 2026). Esta plataforma es la que sostiene la programación de los horarios de carga, ya que integra la información de inventario, transporte y demanda en un solo entorno digital accesible para los distintos equipos operativos (LitCommerce, 2026).

El equipo de Walmart Global Tech también ha priorizado la accesibilidad digital como parte de su cultura corporativa. En sus perfiles de puestos de Operaciones Tecnológicas, la empresa señala explícitamente que busca candidatos con conocimiento de buenas prácticas de accesibilidad y con experiencia en la implementación de las pautas WCAG 2.2 AA, así como en tecnologías asistivas, con el fin de crear productos y servicios inclusivos alineados con los estándares de accesibilidad de la compañía (Walmart Inc., s. f.). Asimismo, el personal de soporte de Operaciones Tecnológicas brinda cobertura de nivel 2 (L2) durante las 24 horas del día para los sistemas de programación por lotes (batch schedulers) que se ejecutan tanto en mainframe como en sistemas abiertos, garantizando que los horarios de carga y los procesos logísticos no se interrumpan (Walmart Inc., s. f.).

3.2.3 Herramientas digitales para asociados y automatización

Walmart ha desarrollado aplicaciones móviles, como Me@Walmart, que permiten a los asociados consultar y administrar sus horarios de trabajo desde un dispositivo portátil, incluyendo la posibilidad de intercambiar turnos. Walmart Inc. (2018) documentó, además, la automatización del proceso de descarga de camiones mediante un sistema conocido como “Fast Unloader”, capaz de trasladar un tráiler completo de mercancía por una banda transportadora, clasificarla y enviarla a los carritos de reabastecimiento en una fracción del tiempo que tomaría hacerlo manualmente. En los centros de distribución, el flujo logístico incluye recepción de embarques, clasificación, cross-docking, almacenamiento y distribución de pedidos, apoyado en pronósticos de demanda impulsados por inteligencia artificial y sistemas de gestión de pedidos tipo B2B.

3.3 Variables disponibles y acceso al Data Warehouse

El Data Warehouse constituye el repositorio central donde se integra la información histórica de Walmart y desde el cual las distintas áreas del negocio obtienen los datos necesarios para el análisis y la toma de decisiones. Para que el modelo de Business Intelligence orientado al pronóstico de demanda sea efectivo, resulta indispensable definir tanto las variables disponibles dentro del repositorio como los mecanismos mediante los cuales los usuarios acceden a dicha información.

Walmart cuenta con una de las infraestructuras analíticas más grandes del mundo. Su centro de análisis, conocido como Data Café y ubicado en la sede de Bentonville, Arkansas, funciona sobre una nube privada capaz de procesar alrededor de 2,5 petabytes de información por hora, e integra cerca de 40 petabytes de datos transaccionales recientes que representan aproximadamente 200 mil millones de registros correspondientes a solo las últimas semanas de operación (Marr, 2017; Marr, s. f.). Esta capacidad convierte al Data Warehouse en una fuente de datos suficientemente amplia y detallada como para alimentar modelos predictivos de demanda a nivel de producto, tienda y región.

3.3.1 Variables disponibles en el Data Warehouse

Las variables que residen en el Data Warehouse pueden agruparse en categorías internas —originadas en la propia operación de Walmart— y externas —provenientes de fuentes de terceros que enriquecen el análisis—. La siguiente tabla resume las principales categorías de variables disponibles y su relación con el proyecto de pronóstico de demanda:

Tabla 7. *Variables disponibles en el Data Warehouse de Walmart*

Categoría de variable	Descripción	Ejemplos de variables
Variables transaccionales (POS)	Registran cada venta realizada en tiendas físicas y canales digitales; constituyen la fuente principal para el análisis de la demanda.	SKU, cantidad vendida, precio, fecha y hora, tienda, canal, ticket.

Categoría de variable	Descripción	Ejemplos de variables
Variables de inventario	Reflejan la disponibilidad y el movimiento de productos en tiendas y centros de distribución.	Stock disponible, inventario de seguridad, rotación, quiebres de stock, reposiciones.
Variables de producto y precio	Describen las características comerciales de cada artículo y sus condiciones de venta.	Categoría, marca, precio regular, precios promocionales, elasticidad.
Variables de tienda y ubicación	Contextualizan la demanda según la geografía y el formato de cada punto de venta.	Región, formato de tienda, ubicación geográfica, centro de distribución asociado.
Variables de cliente y comportamiento	Consolidan los hábitos de compra y la segmentación de los consumidores.	Historial de compra, segmento, canal preferido, comportamiento digital.
Variables temporales y de estacionalidad	Permiten identificar patrones de demanda a lo largo del tiempo.	Estacionalidad, festivos, tendencias, ciclos de reabastecimiento.
Variables externas	Datos de fuentes ajenas a la operación que enriquecen los modelos de pronóstico de demanda.	Datos meteorológicos, económicos, Nielsen, telecomunicaciones, redes sociales, precios de la gasolina, eventos locales.
Variables logísticas	Capturan el desempeño de la cadena de suministro y los flujos de mercancía.	Tiempos de reposición, movimientos de cross-docking, lecturas RFID, sensores IoT.

Nota. Elaboración propia con base en Marr (2017), Walmart Inc. (2025a) y fuentes públicas.

3.3.2 Fuentes de datos internas y externas

Además de la información transaccional generada por los sistemas operacionales (POS, WMS, ERP, comercio electrónico), el Data Café de Walmart incorpora datos provenientes de cerca de 200 fuentes externas e internas distintas. Entre las fuentes externas destacan los datos meteorológicos, económicos, de Nielsen, de telecomunicaciones, de redes sociales, los precios de la gasolina y las bases de datos de eventos locales (Marr, s. f.). La combinación de estas fuentes permite que los modelos de demanda no dependan únicamente del comportamiento histórico de ventas, sino que incorporen factores del entorno que afectan directamente el consumo, lo cual es especialmente relevante para anticipar picos de demanda asociados a fenómenos climáticos, festivos o eventos regionales.

3.3.3 Acceso al Data Warehouse

Antes de la implementación del Data Café, la infraestructura de datos de Walmart solo permitía a los gerentes obtener reportes semanales o del día siguiente, con formatos estandarizados y escaso margen de personalización, lo que impedía gestionar el negocio con base en las condiciones del mercado en tiempo real (McKendrick, 2016). El modelo actual transformó ese esquema al ofrecer información casi en tiempo real, con actualizaciones al menos cada hora, y al ampliar el acceso desde un grupo reducido de tomadores de decisiones

estratégicas hacia un amplio grupo de personal operativo que necesita resolver problemas cotidianos (McKendrick, 2016).

El acceso se apoya en un enfoque de autoservicio gobernado. Los equipos de cualquier área pueden llevar sus preguntas de negocio al centro de análisis y visualizar las soluciones sobre tableros interactivos, reduciendo de semanas a minutos el tiempo necesario para responder consultas complejas que dependen de múltiples variables internas y externas (Marr, 2017). En su evolución hacia la nube, Walmart ha incorporado plataformas de analítica de autoservicio gobernado que permiten a usuarios no técnicos consultar la información sin depender de los equipos de tecnología.

El acceso a la información está regulado mediante marcos de gobierno de datos que contemplan el control de accesos, la trazabilidad (lineage), la calidad y el cumplimiento normativo (Walmart Inc., s. f.). En entornos como BigQuery, este gobierno se materializa a través de catálogos unificados, políticas de acceso y enmascaramiento a nivel de columna y seguridad a nivel de fila, lo que garantiza que cada usuario acceda únicamente a las variables que le corresponden según su rol (Google Cloud, s. f.). De esta forma, se equilibra la democratización del acceso a los datos con la protección de la información sensible.

3.3.4 Relación con el proyecto de Business Intelligence

La amplitud de variables disponibles y el modelo de acceso gobernado en tiempo casi real constituyen habilitadores directos del proyecto de pronóstico de demanda. La disponibilidad de variables transaccionales, de inventario, de producto y externas permite construir modelos predictivos robustos, mientras que el acceso ágil y gobernado asegura que las áreas de compras, logística, inventarios y comercial trabajen sobre una única versión de la información. Así, la definición de variables y accesos descrita en esta sección se articula con los sistemas implicados en los procesos (sección 3.5) y con los recursos de IT (sección 3.6) para sostener el objetivo general del proyecto: optimizar la gestión de inventarios anticipando la demanda, reduciendo el sobrestock y minimizando los quiebres de inventario.

3.4 Sistemas implicados en los procesos

Para soportar el modelo de Business Intelligence propuesto, Walmart requiere la integración de múltiples sistemas operacionales y analíticos que permiten capturar, procesar, almacenar y explotar grandes volúmenes de información. Cada uno de estos sistemas cumple una función específica dentro del proceso ETL, garantizando que los datos generados en la operación diaria puedan transformarse en información útil para la toma de decisiones.

La arquitectura propuesta integra sistemas transaccionales, plataformas de almacenamiento, herramientas de integración de datos y soluciones analíticas, permitiendo disponer de una visión unificada del negocio y facilitando el desarrollo de modelos predictivos para la optimización de inventarios.

3.4.1 Sistemas involucrados

Tabla 8. *Sistemas involucrados en el modelo de Business Intelligence*

Sistema	Función dentro del proceso	Información gestionada
Point of Sale (POS)	Registra las ventas realizadas en cada tienda física en tiempo real. Constituye la principal fuente de información para el análisis de la demanda.	Ventas, productos, cantidades, precios, fecha, hora y tienda.
ERP (Enterprise Resource Planning)	Administra los procesos administrativos y financieros de la compañía, integrando compras, proveedores y costos.	Órdenes de compra, facturación, costos, proveedores y abastecimiento.
WMS (Warehouse Management System)	Gestiona la operación de los centros de distribución y el movimiento de inventarios.	Entradas y salidas de mercancía, almacenamiento, despachos y reposiciones.
Sistema de Inventarios	Controla la disponibilidad de productos en tiendas y centros de distribución.	Stock disponible, inventario de seguridad, productos agotados y rotación.
CRM (Customer Relationship Management)	Consolida la información relacionada con los clientes y sus hábitos de consumo.	Historial de compras, segmentación de clientes y comportamiento de compra.
Plataforma de Comercio Electrónico (Walmart.com y App)	Registra las transacciones realizadas a través de los canales digitales, complementando la información proveniente de las tiendas físicas.	Pedidos en línea, navegación, preferencias y comportamiento digital.
Data Lake	Repositorio inicial donde se almacenan los datos provenientes de todas las fuentes internas y externas, tanto estructurados como no estructurados.	Información operativa, archivos históricos, registros de sensores, datos climáticos y redes sociales.
Proceso ETL	Extrae, transforma y carga la información hacia el Data Warehouse, garantizando la calidad e integración de los datos.	Limpieza, homologación, validación y transformación de datos.
Data Warehouse	Consolida la información histórica integrada y constituye la fuente oficial para el análisis corporativo.	Datos históricos organizados para consultas analíticas.
Data Marts	Subconjuntos especializados del Data Warehouse orientados a las necesidades de cada área funcional.	Información específica para Comercial, Logística, Inventarios, Compras y Finanzas.
Power BI	Plataforma de visualización utilizada para generar dashboards e indicadores ejecutivos.	KPIs, reportes, tendencias e indicadores estratégicos.
Modelo Predictivo (Machine Learning)	Analiza la información histórica y genera pronósticos de demanda para apoyar la toma de decisiones.	Predicciones de demanda, riesgo de desabastecimiento y recomendaciones de reposición.

Nota. Elaboración propia con base en Marr (2017), Walmart Inc. (2025a) y fuentes públicas.

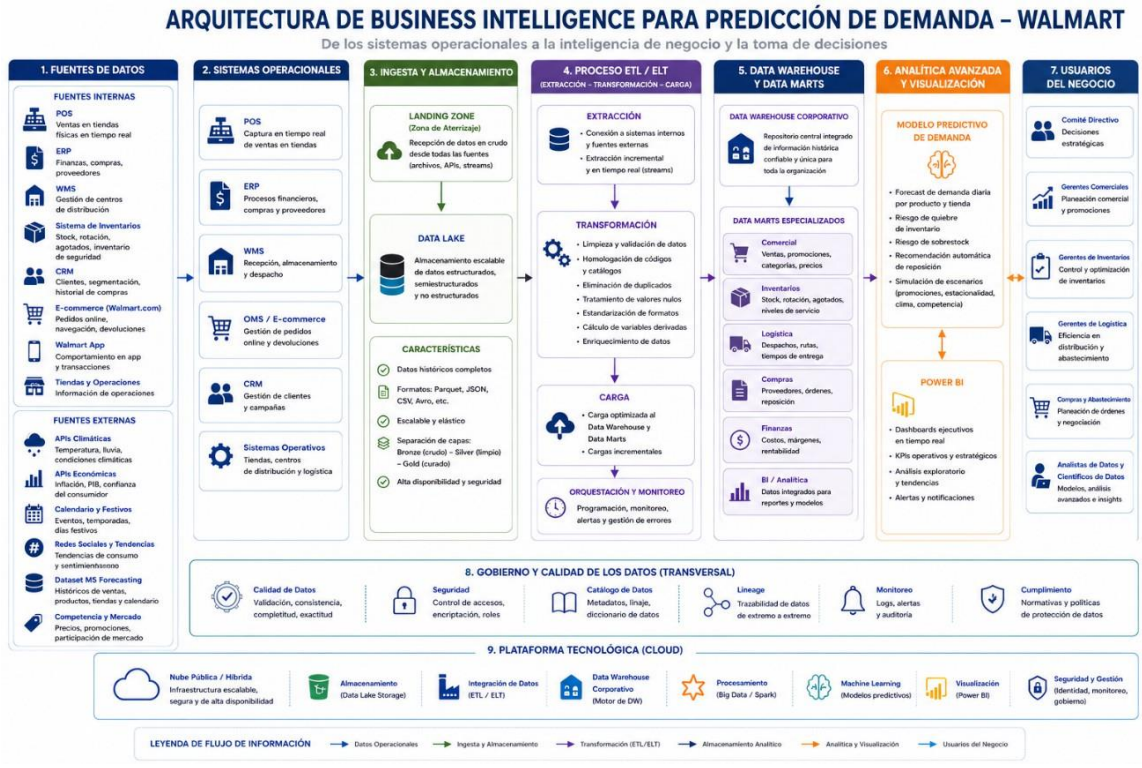
3.4.2 Integración de los sistemas

La integración de estos sistemas permite que la información fluya de manera continua desde la operación diaria hasta los procesos analíticos. Las ventas registradas en los puntos de venta, la información de inventarios, los movimientos logísticos y los datos de comercio electrónico son consolidados mediante procesos ETL y almacenados en un Data Warehouse corporativo.

Posteriormente, esta información es distribuida hacia diferentes Data Marts especializados, desde donde es utilizada por herramientas de visualización como Power BI y por modelos de Machine Learning encargados de generar pronósticos de demanda. Esta arquitectura garantiza que todas las áreas de la organización trabajen sobre una única versión de la información, reduciendo inconsistencias y mejorando la calidad de las decisiones.

3.4.3 Flujo de información propuesto

Figura 1. ***Arquitectura de Business Intelligence propuesta para la predicción de demanda en Walmart***



Nota. Elaboración propia.

3.4.4 Relación con el proyecto de Business Intelligence

La integración de estos sistemas constituye la base tecnológica del proyecto desarrollado en los capítulos anteriores. Mientras los sistemas transaccionales capturan la información operativa, el proceso ETL garantiza su calidad y

consolidación. Posteriormente, el Data Warehouse y los Data Marts permiten organizar la información para su análisis, mientras que Power BI facilita la visualización de indicadores y el modelo predictivo transforma los datos históricos en pronósticos de demanda.

De esta manera, la arquitectura propuesta soporta el objetivo general del proyecto: optimizar la gestión de inventarios mediante un modelo de Business Intelligence capaz de anticipar la demanda, reducir el sobrestock y minimizar los quiebres de inventario, fortaleciendo la toma de decisiones en las áreas de compras, logística, inventarios y comercial.

3.5 Recursos de IT o externos empleados

3.5.1 Recursos internos

Walmart Global Tech es la organización tecnológica interna que centraliza el desarrollo, se describe a sí misma como un equipo de más de 15,000 ingenieros de software, científicos de datos y profesionales de servicio dentro de Walmart, que entregan innovaciones para mejorar la experiencia de compra y potenciar a los 2.3 millones de asociados de la compañía.

- **Cultura "build in-house" como preferencia histórica:** Walmart típicamente prefiere desarrollar sus aplicaciones de retail internamente, y solo hacia finales de 2007 comenzó a comprar aplicaciones empaquetadas de proveedores de software líderes como Oracle, HP y SAP. Esto es clave: la base de Walmart no es "comprar todo hecho", sino construir su propia tecnología y usar terceros de forma selectiva.
- **Hubs tecnológicos distribuidos:** Walmart opera con 17 tech hubs alrededor del mundo (por ejemplo, en Atlanta se sumaron 140 roles enfocados en ciencia de datos e ingeniería de software, sumándose a una base local de 70,000 asociados), lo que muestra una estrategia de contratación directa y descentralizada de talento técnico, en lugar de concentrar todo en un solo campus o depender completamente de outsourcing.
- **Pipeline de talento propio:** Walmart forma a sus Data Engineers a través de programas de carrera temprana, incluyendo pasantías y oportunidades de ingeniería de software de nivel inicial, con onboarding técnico y bootcamps de desarrollo es decir, invierte en "fabricar" su propio talento en vez de solo contratarlo ya formado.

3.5.2 Recursos externos: alianzas y proveedores de nube

- **Modelo "Triplet":** la infraestructura de nube de Walmart combina su propia nube privada (Walmart Cloud Native Platform, WCNP, basada en Kubernetes y construida desde cero en 2020) con dos proveedores públicos externos: Microsoft Azure y Google Cloud Platform (GCP).
- **Exclusión deliberada de AWS:** Walmart trabaja con Microsoft Azure y Google Cloud como socios de nube, pero notablemente no con Amazon

Web Services, como sucede con muchos minoristas que compiten directamente con Amazon.

- **División de cargas por especialidad de cada proveedor:** Walmart usa Google Cloud para gestión de inventario impulsada por IA, y Microsoft Azure para aplicaciones empresariales como cargas de trabajo SAP; además usa BigQuery de Google Cloud para insights de clientes en tiempo real.
- **Regionalización de la nube híbrida:** Walmart dividió su infraestructura de nube distribuida en EE. UU. en tres nubes regionales (Pacífico Noroeste, y zonas cercanas a Texas y Virginia), lo que permite ubicar cargas de trabajo sensibles a la latencia cerca de los centros de datos de los proveedores cloud, e incluso en la misma instalación de colocación cuando es necesario. Este modelo regional también da la capacidad de absorber picos de tráfico en temporadas como el regreso a clases o Navidad.
- **Poder de negociación por escala:** según un VP de infraestructura de Walmart, atar toda la estrategia de TI a un solo proveedor (sea AWS, Azure o Google) probablemente genera costos más altos que tener la opción de correr cargas de trabajo en al menos dos plataformas, y el tamaño de Walmart le da capacidad de negociación — Microsoft y Google permiten a Walmart correr parte de su software en cualquiera de las tres nubes, concesión lograda después de que Walmart "aplicara algo de presión".
- **Contrato histórico multimillonario (2009):** Walmart seleccionó a tres proveedores de TI en India Infosys Technologies, Cognizant Technology Solutions y UST Global para contratos multianuales por más de $600 millones de dólares, un monto casi equivalente al valor de los bienes (textiles, artesanías y otros productos) que el minorista más grande del mundo importa de India cada año. Antes de este contrato formal, Infosys y Cognizant habían tenido proyectos piloto con Walmart unos cinco meses antes.

CAPÍTULO 4. Visualización de datos: herramienta, perfiles de usuario y accesos

Con la arquitectura de datos caracterizada, este capítulo aborda la capa de entrega de la información. Se analizan las herramientas de visualización disponibles en el mercado frente al ecosistema tecnológico de Walmart, se justifica la selección de Power BI, se definen los perfiles de usuario y sus dashboards, se establecen las necesidades de información y los niveles de acceso por área y se revisa el cumplimiento de los objetivos tecnológicos de la compañía.

4.1 Análisis de herramientas y selección de la solución acorde con el ecosistema de Walmart

El primer paso de la implementación consiste en analizar el conjunto de herramientas de visualización y entrega de información disponibles en el mercado y seleccionar aquella que mejor se integre con el ecosistema tecnológico y la arquitectura de datos que Walmart ya tiene en operación. Esta decisión no puede tomarse de manera aislada: debe responder a las necesidades del proyecto de pronóstico de demanda, que constituye el caso de uso central del proyecto, y, al mismo tiempo, apalancarse en la infraestructura existente para reducir los costos de integración, la curva de aprendizaje y el riesgo tecnológico.

4.1.1 Contexto: la arquitectura de datos de Walmart

Antes de comparar herramientas es necesario caracterizar el entorno en el que la solución de visualización debe insertarse. Walmart opera una arquitectura de datos híbrida y multinube de gran escala, cuyos componentes principales son:

- **Almacén de datos analítico:** Azure Synapse Analytics como plataforma central, que procesa más de 2,5 petabytes de datos por hora y unifica en un mismo entorno el data warehousing, el procesamiento de big data y el machine learning.
- **Ingesta y datos crudos:** flujos de streaming en tiempo real mediante clústeres de Apache Kafka, almacenamiento del dato crudo en Azure Data Lake Storage y orquestación de los procesos ETL en Azure Data Factory.
- **Gobierno de datos:** linaje, clasificación y control de accesos a través de Azure Purview, esenciales para el aseguramiento de la calidad y el cumplimiento regulatorio.
- **Analítica avanzada:** la plataforma propia “Element”, un entorno de machine learning multinube (nube privada, Google Cloud Platform y Microsoft Azure) que administra varios millones de modelos en paralelo y soporta lenguajes como Python, Scala, R y SQL.

Este conjunto define el marco de referencia: la herramienta seleccionada debe conectarse con naturalidad a esta pila, escalar al volumen de datos de Walmart

y ser capaz de consumir las salidas de los modelos analíticos, en particular las del modelo de pronóstico de demanda.

4.1.2 Criterios de evaluación

Siguiendo el Plan de Visualización planteado en las lecturas del módulo —qué información mostrar, cómo mostrarla, a quién va dirigida y con qué objetivo— y considerando las restricciones de la arquitectura descrita, se definieron los siguientes criterios de selección:

- Integración con la arquitectura de datos actual (Azure Synapse, Data Lake, Databricks) y autenticación corporativa.
- Escalabilidad a volúmenes de nivel petabyte sin degradación del desempeño.
- Capacidad de consumir y visualizar las salidas de modelos predictivos, requisito indispensable para el pronóstico de demanda.
- Actualización en tiempo real o near-real-time (right-time data).
- **Autoservicio y gobierno de datos:** seguridad, control de accesos por perfil y trazabilidad.
- Costo de licenciamiento y eficiencia operativa.
- Accesibilidad multidispositivo (Windows, iOS, Android) y trabajo colaborativo.
- Madurez y adopción interna previa dentro de la organización.

4.1.3 Análisis comparativo de herramientas

A partir del catálogo de herramientas de reportería y business intelligence revisado en el módulo (Power BI, Tableau, Qlik, MicroStrategy, SAP Business Objects, Domo y Pentaho, entre otras), se evaluó cada alternativa frente a los criterios anteriores. La tabla siguiente resume el análisis.

Tabla 9. *Análisis comparativo de herramientas de visualización frente al ecosistema de Walmart*

Herramienta	Descripción y consideraciones	Ajuste con el ecosistema
Power BI (Microsoft)	Suite de analítica de negocio con reportería interactiva, tableros de control y consultas ad hoc; conexión nativa a Azure Synapse, Data Lake y Databricks, autenticación con Microsoft Entra ID y consumo de modelos de ML.	Alto
Tableau (Salesforce)	Herramienta líder en visualización exploratoria, con gran capacidad gráfica y facilidad de uso. Se conecta a múltiples fuentes, aunque su integración con la pila Azure de Walmart no es nativa y su licenciamiento es más costoso.	Medio-alto
Qlik Sense / QlikView	Plataforma de análisis asociativo en memoria, potente para la exploración libre de datos. Requiere infraestructura y gobierno propios, con integración adicional al ecosistema multinube de Walmart.	Medio

Herramienta	Descripción y consideraciones	Ajuste con el ecosistema
MicroStrategy	Plataforma empresarial robusta orientada a reportería corporativa y grandes despliegues. Sólida en gobierno, pero con mayor complejidad de administración y costos de implementación.	Medio
SAP Business Objects	Suite tradicional de reportería estática y dinámica sobre fuentes SAP. Adecuada en entornos centrados en SAP, menos alineada con la arquitectura Azure/multinube de Walmart.	Medio-bajo
Domo	Plataforma SaaS de BI en la nube con tableros y conectores predefinidos. Ágil para despliegues rápidos, pero con menor control sobre la infraestructura y dependencia de un proveedor externo.	Medio
Pentaho	Suite open source de integración y BI. Flexible y de bajo costo de licenciamiento, pero exige mayor esfuerzo de desarrollo y soporte interno para operar a escala de petabytes.	Bajo

Nota. Elaboración propia con base en las lecturas del módulo de visualización.

4.1.4 Herramienta seleccionada: Power BI

Con base en el análisis anterior, la herramienta que mejor encaja en el ecosistema y la arquitectura de Walmart es Microsoft Power BI. La selección se sustenta en las siguientes razones:

- **Integración nativa con la arquitectura existente.** Dado que el núcleo analítico de Walmart se apoya en Azure Synapse Analytics y Azure Data Lake, Power BI se conecta de forma nativa a estas fuentes y accede a los datos curados mediante autenticación con Microsoft Entra ID, sin necesidad de credenciales adicionales ni capas intermedias. Esto minimiza el esfuerzo y el riesgo de integración.
- **Adopción y madurez interna.** La organización de finanzas de Walmart ya estandarizó su reportería y analítica sobre Power BI para lograr una solución ágil, escalable y costo-eficiente, lo que evidencia que la herramienta está probada dentro de la compañía y reduce la curva de adopción del proyecto.
- **Escalabilidad y tiempo real.** Power BI soporta el consumo de grandes volúmenes mediante modos como DirectQuery y Direct Lake, y puede alimentarse de los flujos de streaming (Apache Kafka), habilitando tableros con actualización en el tiempo correcto para el seguimiento de la demanda.
- **Conexión con la analítica predictiva.** Al integrarse con Azure y con los modelos gestionados en la plataforma Element, Power BI puede visualizar las salidas del modelo de pronóstico de demanda: demanda proyectada frente a demanda real, por tienda, región y categoría, cerrando el ciclo entre el modelo analítico y la decisión de negocio.

- **Autoservicio, gobierno y colaboración.** Ofrece tableros personalizables por perfil, control de accesos, y acceso desde Windows, iOS y Android en un mismo espacio colaborativo, en línea con las necesidades de las áreas de planeación, abastecimiento y operaciones.
- **Eficiencia de costos.** Al pertenecer al mismo ecosistema Microsoft ya contratado por Walmart, se aprovechan sinergias de licenciamiento y soporte frente a alternativas como Tableau o MicroStrategy.

En síntesis, aunque herramientas como Tableau o Qlik ofrecen capacidades gráficas destacadas, ninguna se alinea con la arquitectura de Walmart con la naturalidad de Power BI, cuya integración nativa con Azure Synapse, su adopción interna previa y su capacidad para consumir modelos predictivos lo convierten en la opción más adecuada para soportar el proyecto de pronóstico de demanda. Sobre esta herramienta se construirán los tableros descritos en las secciones siguientes, definiendo los usuarios y perfiles de uso, las necesidades de información por departamento y la revisión del cumplimiento de los objetivos iniciales.

4.2 Usuarios y perfiles de uso de cada dashboard

La correcta segmentación de los usuarios constituye un elemento fundamental para garantizar que la información llegue a las personas adecuadas dentro de la organización. Debido a la complejidad operacional de Walmart y al gran volumen de información que se genera diariamente, resulta indispensable definir perfiles diferenciados de acceso que permitan adaptar los dashboards a las necesidades específicas de cada departamento.

En el presente proyecto se han identificado cinco perfiles principales de usuario:

4.2.1 Alta Dirección

Los miembros de la alta dirección tendrán acceso a dashboards ejecutivos orientados al análisis estratégico del negocio. La información disponible incluirá indicadores relacionados con el crecimiento de las ventas, la rentabilidad, el comportamiento de la demanda, la disponibilidad de productos, el desempeño logístico y las proyecciones generadas por los modelos predictivos.

Su principal objetivo será apoyar la toma de decisiones estratégicas relacionadas con inversiones tecnológicas, planificación operativa y optimización de la cadena de suministro.

Entre los principales indicadores disponibles se encuentran:

- Ventas totales.
- Variación porcentual de las ventas.
- Productos con mayor demanda.
- Nivel global de inventarios.
- Pronósticos de demanda.
- Riesgo de desabastecimiento.

- Desempeño de la cadena logística.
- Indicadores financieros asociados al inventario.

4.2.2 Inventarios y Abastecimiento

Este perfil dispondrá de dashboards especializados en la gestión operativa del inventario. Los responsables del área podrán monitorear continuamente los niveles de stock, la rotación de productos, las necesidades de reposición y los productos que presentan riesgos de sobrestock o desabastecimiento.

Los principales indicadores serán:

- Inventario disponible por tienda.
- Productos próximos a agotarse.
- Inventario de seguridad.
- Tiempo promedio de reposición.
- Rotación del inventario.
- Productos con sobrestock.
- Predicciones de demanda por categoría y producto.

4.2.3 Logística y Cadena de Suministro

El área logística tendrá acceso a información relacionada con los procesos de distribución y abastecimiento de mercancías.

Los dashboards permitirán visualizar:

- Estado de los centros de distribución.
- Cumplimiento de entregas.
- Tiempos de transporte.
- Niveles de abastecimiento.
- Disponibilidad de mercancías.
- Costos logísticos.
- Proyecciones de demanda por región.

La utilización de modelos predictivos permitirá anticipar incrementos inesperados en la demanda y optimizar la distribución del inventario entre los diferentes puntos de venta.

4.2.4 Comercial y Ventas

El departamento comercial contará con información relacionada con el comportamiento del consumidor y el desempeño de las ventas.

Entre las variables disponibles destacan:

- Productos más vendidos.
- Ventas por categoría.
- Promociones con mayor impacto.

- Comportamiento histórico de la demanda.
- Tendencias de consumo.
- Ventas por región y canal comercial.
- Predicciones de crecimiento de las ventas.

Estos indicadores facilitarán el diseño de estrategias comerciales más eficientes y permitirán evaluar el impacto de las campañas promocionales sobre la demanda de los productos.

4.2.5 Equipo de Ciencia y Analítica de Datos

Finalmente, los científicos y analistas de datos dispondrán del mayor nivel de acceso dentro del entorno analítico, puesto que serán responsables del desarrollo, entrenamiento y evaluación de los modelos predictivos utilizados en el proyecto.

Sus dashboards incluirán:

- Variables utilizadas por los modelos predictivos.
- Calidad de los datos.
- Indicadores de desempeño de los algoritmos.
- Precisión de los pronósticos.
- Variables explicativas de la demanda.
- Monitoreo del comportamiento de los modelos implementados.

La segmentación de usuarios permitirá optimizar la experiencia de utilización de Power BI y fortalecer las políticas de seguridad relacionadas con el acceso a la información.

4.3 Necesidades de información y accesos por departamento y perfil

A partir de los perfiles definidos, se establecen la información que requiere cada área y el nivel de acceso que tendrá dentro de Power BI. Esta matriz es la base para configurar la seguridad a nivel de fila (row-level security) y los espacios de trabajo de la herramienta, de modo que cada usuario acceda únicamente a la información que necesita para su función.

Tabla 10. ***Necesidades de información y niveles de acceso por área***

Área / perfil	Información requerida	Nivel de acceso
Dirección corporativa	KPIs agregados (ventas, margen, tráfico, participación de mercado, indicadores ESG, comparativos geográficos).	lectura sobre cuadros de mando estratégicos; posibilidad de drilldown limitado a región/segmento; alta prioridad de calidad y disponibilidad (right-time más que realtime).

Área / perfil	Información requerida	Nivel de acceso
Merchandising / Comercial	detalle por SKU, categoría, tienda, promoción; elasticidad de precios, resultados de campañas, visibilidad en góndola, datos del Data Café.	escritura limitada (definición de parámetros de promociones dentro de los modelos), lectura detallada con filtros por zona, segmento y periodo; dashboards operacionales y tácticos.
Supply chain / logística	niveles de inventario, órdenes de reposición, OTIF, tiempos de tránsito, incidencias, datos de proveedores y rutas.	dashboards operativos en tiempo cercano a real, alertas y mapas de calor, con permisos de edición para configurar umbrales y reglas de negocio.
Finanzas y control de gestión	estados financieros consolidados, análisis de rentabilidad por formato y categoría, presupuestos, proyecciones y escenarios.	cuadros de mando tácticos y estratégicos; lectura completa de KPIs y capacidad para construir reportes ad hoc sobre el data warehouse financiero.
TI / data & analytics	calidad de datos (completitud, consistencia), performance de procesos ETL, errores de integración, uso de herramientas de BI.	lectura y escritura sobre dashboards técnicos, administración de modelos de seguridad y perfiles de acceso, explotación de SAS/Python/R para analítica avanzada.
Proveedores (Scintilla / Luminate)	ventas por producto, inventarios en tienda, participación de mercado, shopper insights, resultados de test en hogar (IHUT), percepción de marca.	lectura controlada vía suscripción (Basic/Charter), con filtros por categoría, banner y país; sin acceso directo a datos individuales de clientes.

Nota. Elaboración propia.

4.4 Revisión y cumplimiento de los objetivos iniciales

4.4.1 Objetivos iniciales de la división de comercio electrónico

En 2010, Walmart consolidó sus operaciones de comercio electrónico a nivel mundial mediante la creación de la Global eCommerce Division, con sede en San Bruno, California. Esta división se planteó tres objetivos iniciales: desarrollar y ejecutar una estrategia global de comercio electrónico, acelerar el crecimiento del canal en línea a nivel mundial y crear plataformas tecnológicas y aplicaciones para cada uno de los mercados en los que opera la empresa. El cumplimiento de estas metas fue progresivo: en 2012 la división renovó el motor de búsqueda de Walmart.com con Polaris, un motor de búsqueda semántica desarrollado internamente por @WalmartLabs, con el que la proporción de visitantes que completaban una compra tras buscar un producto aumentó aproximadamente entre un 10 % y un 15 % (Supermarket News, 2012).

4.4.2 Revisión y consolidación de plataformas tecnológicas

Con el paso del tiempo, Walmart identificó que la fragmentación tecnológica heredada de sus primeros objetivos generaba ineficiencias. Un ejecutivo de la compañía señaló que, en el caso específico de México, llegaron a operar hasta

cuatro plataformas distintas para impulsar el comercio electrónico, lo que evidenció la necesidad de contar con una plataforma común que permitiera escalar de manera más eficiente (National CIO Review, 2025). Como resultado de esta revisión, Walmart consolidó sus operaciones de comercio electrónico y mercado en una sola plataforma global, lo que le permite introducir nuevas funciones y actualizaciones con mayor rapidez, reducir costos operativos y realizar pruebas piloto de innovación en mercados más pequeños antes de expandirlas (National CIO Review, 2025). Bajo este enfoque de “probar y aprender”, la empresa ha implementado tecnología de transporte primero en países como Chile y en Centroamérica antes de llevarla a Estados Unidos, lo que le da flexibilidad para perfeccionar sus sistemas antes de escalarlos globalmente (National CIO Review, 2025).

4.4.3 Cumplimiento de metas y gobernanza de la información

El cumplimiento de los objetivos tecnológicos también se revisa formalmente a través de los reportes corporativos que Walmart presenta ante la Comisión de Bolsa y Valores de Estados Unidos (SEC). En su declaración de representación (DEF 14A) correspondiente al año fiscal 2019, la compañía documentó, entre sus prioridades estratégicas, la disciplina operativa y el objetivo de “hacer de la confianza una ventaja competitiva”, lo cual incluye modelar la excelencia en el cumplimiento normativo y la ética a nivel global, así como liderar en temas sociales y ambientales (Walmart Inc., 2019). Este tipo de revisión periódica evidencia que el manejo de la información tecnológica en Walmart no se limita a la innovación operativa, sino que también se supervisa bajo estándares de cumplimiento y gobernanza corporativa.

4.4.4 Resultados recientes frente a los objetivos planteados

Los resultados financieros más recientes permiten evaluar el grado de cumplimiento de los objetivos tecnológicos originales. En el año fiscal 2025, Walmart invirtió aproximadamente 22 000 millones de dólares en gasto de capital, de los cuales cerca de 19 000 millones se destinaron a Estados Unidos, financiando nuevas tiendas, automatización y tecnología (DCFmodeling, 2025). Las ventas globales de comercio electrónico crecieron un 20,8 % y alcanzaron los 121 000 millones de dólares, equivalentes a aproximadamente el 18 % de los ingresos totales de la compañía (DCFmodeling, 2025). Asimismo, un análisis estratégico reciente concluye que Walmart ha invertido más de 20 000 millones de dólares en la última década en tecnología, incluyendo adquisiciones, inteligencia artificial y automatización, integrando sus canales físico y digital en lugar de tratarlos como negocios separados (Poornaprajna International Journal of Management, Education & Social Science, 2025).

CAPÍTULO 5. El departamento de Analítica de Datos y Business Intelligence

Los capítulos anteriores mostraron que Walmart dispone de una infraestructura de datos de gran escala y que Power BI es la herramienta adecuada para entregar la información a cada perfil de usuario. Este capítulo se centra en el componente humano y organizativo de la solución: la situación actual del área de analítica, las carencias de formación y de recursos, la coordinación de los flujos de información con otras áreas, el organigrama del departamento, sus dependencias y su modelo de solución de problemas, el uso de recursos externos y el caso de negocio de costos y beneficios.

5.1 Situación actual del departamento

Para el proyecto de optimización de ventas y demanda de Walmart, el área central de análisis puede ubicarse en Business Intelligence y Analítica de Datos, trabajando de manera coordinada con las áreas de ventas, inventario, compras y cadena de suministro.

Actualmente, Walmart opera en un entorno caracterizado por un elevado volumen de productos, múltiples canales de venta y una amplia actividad comercial. Esta situación genera grandes cantidades de información relacionada con ventas, demanda, inventario, comercio electrónico, logística y comportamiento del consumidor. La principal dificultad para el proyecto no es la ausencia de información, sino la necesidad de integrar, procesar y convertir diferentes fuentes de datos en información útil para la toma de decisiones.

Desde la perspectiva del proyecto académico, se propone fortalecer el departamento mediante una solución basada en Big Data, Business Intelligence, análisis predictivo y procesamiento de datos en tiempo real. El propósito sería pasar de una gestión basada principalmente en la observación de resultados históricos a una gestión que también permita anticipar cambios en la demanda.

La situación actual que se plantea para el proyecto puede resumirse en tres necesidades principales. Primero, mejorar la capacidad para integrar diferentes fuentes de información. Segundo, aumentar la capacidad de predicción de la demanda. Tercero, proporcionar información actualizada a los responsables de inventario, compras, logística y ventas mediante dashboards de Power BI.

La solución propuesta utilizaría Microsoft Azure como proveedor cloud y PaaS como modelo de servicio, junto con Azure Data Factory para integración y preparación de datos, Azure Data Lake Storage para almacenamiento, Apache Spark y PySpark para procesamiento, herramientas de Machine Learning para los modelos predictivos y Power BI para la visualización de resultados.

Por tanto, la situación actual del departamento dentro del proyecto se caracteriza por la necesidad de evolucionar hacia una estructura de analítica más integrada, capaz de transformar datos públicos y otras fuentes disponibles en información que permita mejorar las decisiones relacionadas con ventas y demanda.

El cambio de fondo que persigue el departamento es el paso de una visión descriptiva a una visión predictiva. Hoy predominan los informes tradicionales de ventas e inventario de un período determinado, que responden a la pregunta «¿qué ocurrió?» y explican el comportamiento pasado del negocio. Con este enfoque persiste el riesgo de que unas tiendas acumulen exceso de producto mientras otras enfrentan desabastecimiento. La meta es responder a la pregunta «¿qué podría ocurrir?», utilizando datos históricos y variables relevantes para anticipar la demanda, mediante modelos de Machine Learning que proyecten escenarios y generen recomendaciones de inventario y reabastecimiento.

5.2 Carencias en formación y recursos

Una de las principales necesidades identificadas para implementar la solución es contar con personal que posea conocimientos tanto de negocio como de tecnología y análisis de datos.

En materia de formación, sería necesario fortalecer las competencias relacionadas con Big Data, Python, PySpark, Apache Spark, Machine Learning, Power BI, gestión de bases de datos y servicios cloud de Microsoft Azure. También sería importante que los profesionales comprendan el funcionamiento del negocio retail, especialmente los procesos de ventas, inventario, compras y cadena de suministro.

La formación no debería concentrarse exclusivamente en aspectos técnicos. Los responsables del área también necesitan comprender cómo interpretar los modelos y convertir sus resultados en decisiones empresariales. Por ejemplo, no basta con que un modelo indique que existe un incremento previsto de la demanda; el equipo debe saber cómo utilizar esa información para determinar si es necesario revisar el abastecimiento, el inventario o la distribución.

En cuanto a recursos tecnológicos, el proyecto requiere una infraestructura capaz de almacenar y procesar grandes volúmenes de información. Para ello se propone utilizar Azure Data Lake Storage como entorno de almacenamiento, Azure Data Factory para integración y preparación de datos, Apache Spark para procesamiento distribuido y Azure Machine Learning o Spark MLlib para desarrollar los modelos predictivos.

También es necesario disponer de herramientas de visualización y análisis. En este proyecto se ha seleccionado Power BI, debido a que permite presentar los principales indicadores mediante dashboards y facilitar su interpretación por parte de los responsables de negocio.

Finalmente, existe una necesidad relacionada con la disponibilidad y calidad de los datos. Debido a que nuestro proyecto utiliza información de dominio público, no se dispone de determinadas fuentes internas de Walmart. Por ello, una de las principales limitaciones del proyecto académico es la disponibilidad de datos transaccionales detallados de productos, tiendas e inventarios. Esta situación debe considerarse una limitación metodológica y no debe interpretarse como una afirmación de que Walmart carezca de esos recursos internamente.

Walmart ya ha dado pasos en esta dirección. En septiembre de 2025 anunció una alianza con OpenAI para ofrecer a sus asociados de tiendas y oficinas en Estados Unidos un programa de certificación en inteligencia artificial a través de Walmart Academy, como parte de su compromiso de invertir cerca de 1000 millones de dólares en capacitación de habilidades hasta 2026 (Retail Dive, 2025). El departamento propuesto puede apoyarse en estos programas corporativos y complementarlos con formación técnica especializada para sus perfiles analíticos.

La Tabla siguiente sintetiza las carencias identificadas en formación, tecnología, recursos humanos y gestión, junto con la necesidad que plantea cada una.

Tabla 11. ***Síntesis de las carencias identificadas en el área de analítica***

Ámbito	Carencia identificada	Necesidad
Formación	Power BI	Capacitación en dashboards, KPI y visualización
Formación	Analítica avanzada	Capacitación en análisis predictivo
Formación	Machine Learning	Desarrollo y evaluación de modelos de predicción
Formación	Estadística	Interpretación de tendencias y patrones
Formación	Negocio (retail)	Conocimiento de ventas e inventarios
Tecnología	Integración de datos	Mejorar la conexión entre fuentes
Tecnología	Procesamiento	Capacidad para manejar grandes volúmenes
Tecnología	Visualización	Implementación de Power BI
Tecnología	Calidad de datos	Procesos de validación y depuración
Recursos humanos	Especialistas	Ingenieros de datos, científicos de datos y analistas de BI
Gestión	Acceso a la información	Dashboards diferenciados por perfil
Gestión	Toma de decisiones	Evolucionar de la analítica descriptiva a la predictiva

Nota. Elaboración propia.

5.3 Coordinación de los flujos con otras áreas

La solución requiere una coordinación permanente entre las diferentes áreas involucradas, debido a que la información generada por el sistema tiene utilidad para varias unidades de negocio.

El área de Ventas proporciona el contexto comercial necesario para analizar el comportamiento de las ventas y determinar cuáles son las categorías o productos que presentan cambios relevantes en la demanda.

El área de Inventario utiliza los resultados del análisis para identificar productos con riesgo de sobrestock o desabastecimiento. Esta información permite priorizar aquellos productos que necesitan una revisión.

El área de Compras y Abastecimiento utiliza las previsiones de demanda como elemento de apoyo para determinar las necesidades futuras de productos y revisar las decisiones de reposición.

El área de Logística y Cadena de Suministro recibe información relacionada con las necesidades de distribución y disponibilidad de productos. Esto permite coordinar mejor los recursos logísticos con las necesidades estimadas de cada mercado o ubicación.

Por su parte, Business Intelligence y Data Science funcionan como áreas articuladoras. Su responsabilidad consiste en integrar los datos, procesarlos, construir los modelos predictivos, validar los resultados y proporcionar la información mediante Power BI.

Finalmente, la Gerencia utiliza los indicadores consolidados para supervisar el desempeño y tomar decisiones estratégicas.

La Figura siguiente representa este esquema: el Departamento de BI recibe la información funcional de las áreas de negocio y les devuelve indicadores y dashboards, al tiempo que entrega a la Alta Dirección los cuadros de mando ejecutivos.

Figura 2. ***Flujo de información entre el Departamento de BI y las áreas del negocio***

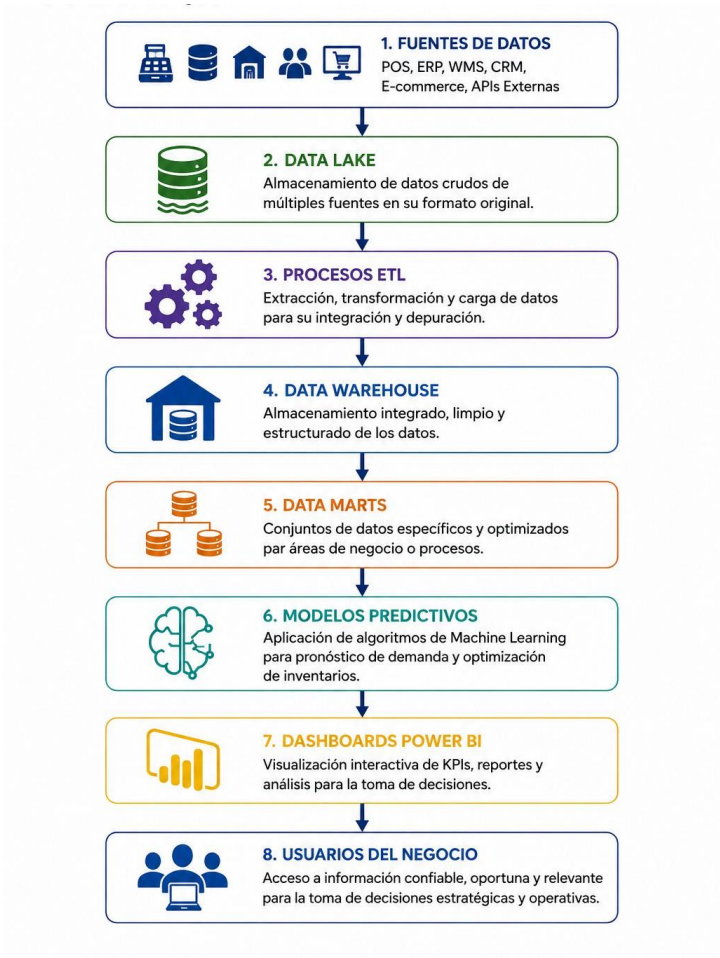


Nota. Elaboración propia.

La coordinación debe establecer un flujo continuo de información. Los datos disponibles son integrados y preparados por el equipo tecnológico, posteriormente son procesados por los modelos analíticos y sus resultados son transformados en indicadores y dashboards. Estos resultados son interpretados

por las áreas de negocio, que pueden tomar decisiones sobre inventario, abastecimiento, logística y ventas. Las decisiones y nuevos resultados generados por el negocio permiten posteriormente alimentar nuevos análisis.

Figura 3. ***Del dato crudo al dashboard: secuencia de coordinación del flujo de datos***



Nota. Elaboración propia.

Esta coordinación exige delimitar responsabilidades con el área de Tecnología. IT provee y mantiene la infraestructura en la nube y el Data Lake, monitorea los procesos ETL, garantiza la seguridad y el respaldo de los datos y asegura la disponibilidad de dashboards y reportes. El Departamento de BI, por su parte, es responsable de la calidad del dato, de la definición de los indicadores y de la evolución de los modelos predictivos. Esta división evita duplicidades y deja claro quién responde ante cada tipo de incidencia.

De esta manera, la solución no debe funcionar como una herramienta aislada del departamento de tecnología, sino como un sistema de apoyo transversal para toda la organización.

5.4 Organigrama, dependencias y solución de problemas

5.4.1 Estructura organizativa del departamento

Para el proyecto se propone una estructura organizacional en la que el equipo de Business Intelligence y Data Analytics tenga una función transversal y mantenga coordinación directa con las principales áreas involucradas en el proceso de ventas y demanda.

En la parte superior se encontraría la Gerencia o Dirección responsable del negocio, encargada de definir los objetivos estratégicos y aprobar las decisiones principales relacionadas con la solución.

Debajo se ubicaría el responsable de Business Intelligence y Analítica, quien coordinaría la ejecución del proyecto y actuaría como vínculo entre las necesidades empresariales y las capacidades tecnológicas.

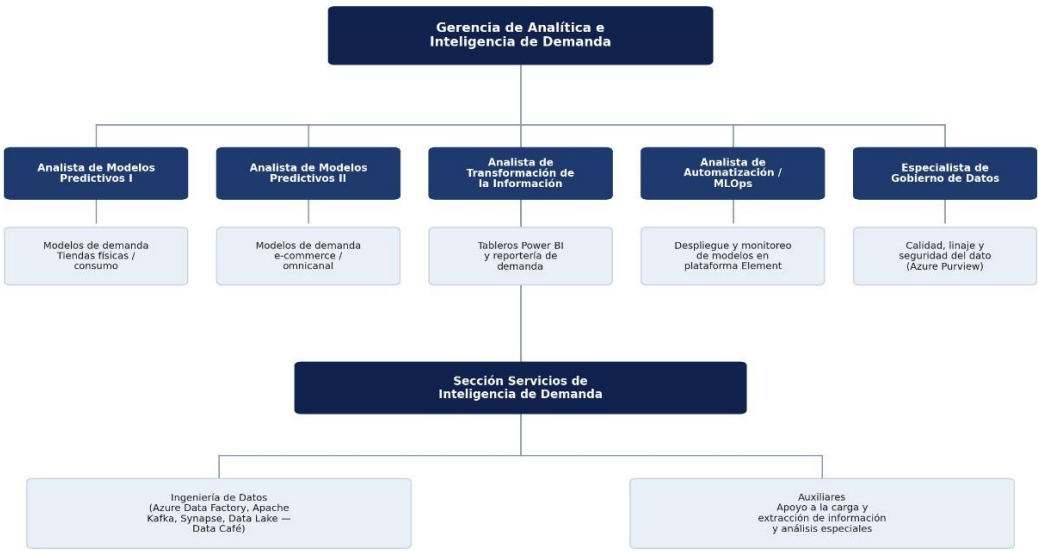
El equipo podría estar conformado por los siguientes perfiles:

- **Data Engineer:** responsable de integrar, preparar y organizar las fuentes de datos.
- **Data Scientist:** responsable del desarrollo, evaluación y actualización de los modelos predictivos.
- **Data Analyst / BI Analyst:** responsable del análisis de resultados y generación de indicadores.
- **Power BI Developer:** responsable del desarrollo y mantenimiento de dashboards.
- **Cloud/IT Specialist:** responsable de la infraestructura y servicios utilizados en Azure.
- **Representantes de negocio:** responsables de Inventario, Ventas, Compras y Logística.

5.4.2 Gerencia de Analítica e Inteligencia de Demanda

En concreto, esta estructura se materializa en la Gerencia de Analítica e Inteligencia de Demanda, dependiente de la Dirección de Datos y Analítica y con relación funcional directa con la Vicepresidencia de Cadena de Suministro, que es su principal cliente interno. Los perfiles descritos se organizan en cinco posiciones especializadas que dependen de la gerencia y en una sección de servicios de soporte, donde se ubica la ingeniería de datos.

Figura 4. ***Organigrama propuesto de la Gerencia de Analítica e Inteligencia de Demanda***



Nota. Elaboración propia.

- **Gerencia de Analítica e Inteligencia de Demanda:** define la estrategia analítica, prioriza la demanda de las áreas usuarias y responde por los resultados del modelo ante la Dirección de Datos y Analítica y la Vicepresidencia de Cadena de Suministro.
- **Analistas de Modelos Predictivos I y II:** desarrollan, entrenan y validan los modelos de demanda; el primero se especializa en tiendas físicas y consumo, y el segundo en comercio electrónico y omnicanalidad.
- **Analista de Transformación de la Información:** construye los tableros de Power BI y la reportería de demanda, y traduce los resultados analíticos al lenguaje del negocio.
- **Analista de Automatización / MLOps:** despliega y monitorea los modelos en la plataforma Element, automatiza los ciclos de reentrenamiento y vigila la degradación de su desempeño.
- **Especialista de Gobierno de Datos:** vela por la calidad, el linaje y la seguridad del dato con Azure Purview, y por el cumplimiento de la normativa de privacidad.
- **Sección de Servicios de Inteligencia de Demanda:** agrupa la Ingeniería de Datos, responsable de los pipelines sobre Azure Data Factory, Apache Kafka, Synapse y el Data Lake, y a los Auxiliares, que apoyan la carga y extracción de información y los análisis especiales.

5.4.3 Dependencias con otras áreas

Las principales dependencias se producirían con Ventas, Inventario, Compras, Logística, Tecnología y Gerencia. Cada área tendría una responsabilidad diferente dentro del funcionamiento de la solución.

La Tabla siguiente detalla la naturaleza de cada dependencia, distinguiendo entre áreas clientes, proveedoras y relaciones bidireccionales.

Tabla 12. *Dependencias de la Gerencia de Analítica e Inteligencia de Demanda con otras áreas*

Área	Tipo de relación	Naturaleza de la dependencia
Merchandising y Categorías	Cliente / bidireccional	Consume pronósticos; aporta contexto de categorías y calendario promocional.
Cadena de Suministro y Logística	Cliente principal	Los pronósticos alimentan el reabastecimiento y la planeación de la distribución.
Operaciones de Tienda	Bidireccional	Aporta datos locales; consume alertas de quiebre de stock.
Finanzas Corporativas	Bidireccional	Integra los pronósticos en el presupuesto; comparte el estándar de Power BI.
Tecnología / Nube (Azure)	Proveedor	Provee y mantiene la infraestructura (Synapse, Data Lake, Data Factory).
Gobierno de Datos (Data Café)	Proveedor / normativo	Habilita el acceso gobernado y define estándares de calidad.
Marketing y Precios	Proveedor de insumos	Aporta el calendario promocional y las variables de precio.

Nota. Elaboración propia.

5.4.4 Modelo de solución de problemas

En cuanto a la solución de problemas, el modelo organizacional permitiría asignar los incidentes de acuerdo con su naturaleza. Los problemas relacionados con datos serían gestionados por Data Engineering; los problemas relacionados con modelos predictivos serían revisados por Data Science; los problemas de dashboards e indicadores serían atendidos por BI; y los problemas de infraestructura serían gestionados por IT y los responsables de Azure.

Este esquema permitiría evitar que todos los problemas sean tratados por un único departamento y facilitaría una respuesta especializada.

El modelo de solución de problemas del departamento se apoya en el proceso KDD, que estructura el paso de los datos al conocimiento en etapas sucesivas de selección, preprocesamiento, transformación, minería de datos e interpretación (Fayyad et al., 1996). Este enfoque es especialmente relevante porque la preparación de los datos concentra entre el 50 % y el 80 % del esfuerzo total de un proyecto analítico (Pérez Galán, 2019): la mayor parte de los

problemas que enfrentará el departamento no estarán en el algoritmo, sino en la calidad, la oportunidad y la interpretación de los datos.

La Tabla siguiente asigna a cada problema previsible una respuesta y un responsable dentro del organigrama, de modo que ninguna incidencia quede sin dueño.

Tabla 13. ***Modelo de solución de problemas del departamento***

Problema previsible	Respuesta	Responsable
Datos faltantes o inconsistentes	Rutinas de limpieza y validación en el preprocesamiento (KDD)	Ingeniería de Datos / Gobierno de Datos
Degradación del modelo (drift)	Monitoreo continuo y reentrenamiento programado	Automatización / MLOps
Conflicto de prioridades entre áreas	Mesa de demanda con acuerdos de nivel de servicio (SLA) de priorización	Gerente de Analítica
Latencia en los flujos de datos	SLA de disponibilidad y redundancia en los pipelines	Ingeniería de Datos y Tecnología
Interpretación errónea de resultados	Documentación y storytelling de datos en Power BI	Analista de Transformación de la Información
Riesgos de seguridad y privacidad	Controles de acceso y linaje mediante Azure Purview	Especialista de Gobierno de Datos

Nota. Elaboración propia con base en Fayyad et al. (1996) y Pérez Galán (2019).

5.5 Recursos externos: subcontratación

Para la implementación del proyecto puede contemplarse la utilización de determinados recursos externos mediante subcontratación o servicios especializados.

La primera posibilidad corresponde a la contratación de especialistas en Microsoft Azure, especialmente durante las etapas iniciales de diseño e implementación de la arquitectura cloud. Estos profesionales podrían apoyar la configuración de servicios como almacenamiento, integración de datos, procesamiento y seguridad.

También podría contratarse apoyo especializado para el desarrollo o validación de modelos de Machine Learning, especialmente si la organización no cuenta inicialmente con suficientes Data Scientists especializados en modelos predictivos de demanda.

Otra posibilidad sería contratar servicios de consultoría en Business Intelligence y Power BI para el diseño inicial de los dashboards, definición de indicadores y establecimiento de estándares de visualización.

Sin embargo, la subcontratación no debería sustituir completamente las capacidades internas. La organización debería conservar internamente el conocimiento relacionado con el negocio, los indicadores estratégicos, la interpretación de resultados y la toma de decisiones.

Por esta razón, se propone que los proveedores externos participen principalmente en actividades especializadas o de implementación, mientras que las funciones críticas relacionadas con el negocio y la gestión de la información permanezcan bajo control interno.

Esta propuesta es coherente con la forma en que Walmart gestiona hoy sus recursos tecnológicos, que combina el desarrollo interno con alianzas externas seleccionadas:

- **Estrategia multinube:** el denominado «Triplet Model» combina la Walmart Cloud Native Platform (nube privada) con Microsoft Azure y Google Cloud Platform, evitando depender de un único proveedor y, deliberadamente, de Amazon Web Services, cuya matriz es su principal competidor (TechTarget, 2022).
- **Plataforma de IA «Element»:** desarrollo propio de Walmart que opera sobre la infraestructura de varios proveedores bajo el mismo esquema multinube y que administra el ciclo de vida de los modelos de Machine Learning.
- **Subcontratación de procesos (BPO):** Genpact apoya a Walmart en finanzas y contabilidad desde 2018, alianza que se amplió en 2023 a Norteamérica, América Latina y Sudáfrica (Stock Titan, 2023).
- **Outsourcing de desarrollo:** Walmart ha recurrido históricamente a talento técnico externo; en 2009, por ejemplo, firmó contratos multianuales con Infosys, Cognizant y UST Global, y mantiene relaciones con integradores como Accenture y TCS.

5.6 Caso de negocio: costos y beneficios

Para evaluar la viabilidad económica del proyecto se propone un Business Case basado en un escenario académico. Debido a que el proyecto utiliza información pública y no dispone de información financiera interna de Walmart relacionada con costos específicos de inventario, ventas perdidas o implementación tecnológica, los valores utilizados son supuestos de referencia y no representan resultados financieros reales de Walmart.

Se plantea un escenario en el que la implementación anual de la solución de Big Data, Machine Learning, Azure y Power BI tendría los siguientes costos estimados:

Tabla 14. ***Costos anuales estimados de la solución (escenario académico)***

Concepto	Costo anual estimado
Servicios cloud y procesamiento	USD 300.000
Desarrollo y mantenimiento de modelos	USD 180.000
Personal especializado adicional	USD 180.000
Power BI, herramientas y licencias	USD 60.000
Consultoría y soporte externo	USD 80.000
Costo total estimado	USD 800.000

Nota. Elaboración propia. Valores de referencia; no representan costos reales de Walmart.

En cuanto a los beneficios, se mantiene el escenario de referencia desarrollado previamente para nuestro proyecto.

Se considera un inventario analizado de USD 100 millones y un costo anual de mantenimiento equivalente al 20 %. Esto representa un costo de referencia de USD 20 millones.

Si la solución permite reducir en un 5 % el inventario innecesario, se liberarían aproximadamente USD 5 millones de inventario. Considerando el costo de mantenimiento, esto representaría un ahorro anual estimado de USD 1 millón.

Adicionalmente, se considera un volumen de ventas analizado de USD 200 millones. Si la mejora en la disponibilidad de productos permitiera recuperar el equivalente al 1 % de las ventas que potencialmente podrían perderse por desabastecimiento, se generarían aproximadamente USD 2 millones en ventas recuperadas.

El beneficio económico bruto estimado sería, por tanto, de:

USD 1.000.000 de ahorro en inventario + USD 2.000.000 de ventas recuperadas = USD 3.000.000.

Al descontar el costo anual estimado de la solución:

USD 3.000.000 − USD 800.000 = USD 2.200.000 de beneficio neto anual.

El ROI estimado sería:

ROI = (USD 2.200.000 / USD 800.000) × 100 = 275 %.

Por lo tanto, bajo este escenario académico, la implementación produciría un ROI estimado del 275 %.

Además del beneficio económico directo, existirían beneficios operativos y estratégicos difíciles de cuantificar inicialmente: mayor capacidad de anticipación de la demanda, reducción del tiempo necesario para identificar problemas de inventario, mejor coordinación entre áreas, mayor disponibilidad de información para los gerentes y una mejor utilización de los recursos tecnológicos.

En consecuencia, el Business Case indica que la solución puede resultar económicamente atractiva si los beneficios derivados de la reducción del sobrestock y del desabastecimiento superan los costos de implementación y operación. No obstante, para convertir este escenario en un Business Case financiero real de Walmart sería indispensable disponer de datos internos de costos, inventario, márgenes, ventas perdidas y operación tecnológica.

Conviene precisar la relación entre este escenario y el caso de negocio del Capítulo 4 de la Parte II. El escenario aquí presentado es preliminar: se plantea a la escala de una unidad de negocio (USD 100 millones de inventario y USD 200 millones de ventas analizadas) y considera únicamente costos operativos anuales. En la Parte II, una vez definido el modelo de Machine Learning, el caso se recalcula sobre un piloto equivalente al 10 % de Walmart U.S. con base en

sus cifras públicas del año fiscal 2026, e incorpora la inversión inicial de implementación. Por ello, el retorno del primer año resulta más conservador. Ambos escenarios coinciden en la conclusión principal: la solución es rentable cuando la reducción del sobrestock y del desabastecimiento supera los costos de implementación y operación.

5.7 Conclusiones del capítulo

La principal carencia identificada no es la ausencia de datos, sino la necesidad de fortalecer las capacidades para transformarlos en conocimiento útil. La Gerencia de Analítica e Inteligencia de Demanda responde a esa necesidad con una estructura que combina perfiles analíticos, de automatización y de gobierno del dato; con dependencias claramente definidas respecto de las áreas usuarias y proveedoras; y con un modelo de solución de problemas que asigna un responsable a cada riesgo previsible. Con este capítulo se completa la base de Business Intelligence del proyecto, que en la Parte II se amplía con las capacidades de Big Data e Inteligencia Artificial.

PARTE II: BIG DATA E INTELIGENCIA ARTIFICIAL

La segunda parte del proyecto corresponde al bloque de Big Data e Inteligencia Artificial. Si la Parte I construyó la base informacional y organizativa, esta parte amplía la capacidad de la solución en tres direcciones: incorporar datos de mayor volumen, velocidad y variedad, incluidos los no estructurados; desplegar una arquitectura en la nube capaz de procesarlos a escala; y aplicar modelos de Machine Learning y de procesamiento en tiempo real para anticipar la demanda. La literatura coincide en que el valor del Big Data no reside en el volumen de datos en sí, sino en su capacidad para mejorar decisiones concretas del negocio (Manyika et al., 2011; Wamba et al., 2015), y ese es el criterio que guía cada una de las decisiones de esta parte.

CAPÍTULO 1. Datos no estructurados y fuentes Big Data para la predicción de la demanda

Este capítulo identifica las fuentes de datos internas y externas, estructuradas y no estructuradas, que Walmart utiliza o podría utilizar para enriquecer el pronóstico de demanda; explica cómo se relacionan con la propuesta inicial del proyecto; define los casos de uso que incorporan esos datos al modelo de Business Intelligence; y justifica el impacto que el Big Data puede aportar en cada área, departamento y proceso de negocio.

1.1 Fuentes de datos internas y externas utilizadas por Walmart

Para identificar las fuentes de datos fue necesario revisar cómo Walmart USA identifica y selecciona las fuentes de datos que alimentan su operación diaria, tomando como referencia principal su centro de analítica conocido como Data Café, inaugurado en 2017 en las oficinas centrales de Bentonville, Arkansas. Este espacio reúne más de 200 flujos de datos internos y externos, junto con cerca de 40 petabytes de información transaccional reciente, que los equipos de la empresa pueden consultar para tomar decisiones casi en tiempo real (Marr, 2017). A partir de esa base, es posible clasificar las fuentes que utiliza Walmart en cuatro grupos: internas estructuradas, internas semiestructuradas, externas estructuradas y externas semiestructuradas o no estructuradas.

Fuentes internas estructuradas

El pilar de este grupo es la información que se genera dentro de la propia operación de Walmart y que ya llega organizada en tablas o bases de datos relacionales. El ejemplo más evidente son los datos de punto de venta (POS), capturados en cada transacción de cada caja registradora, los cuales constituyen la base sobre la que Walmart construyó su primera generación de analítica de inventarios (Ankashram, 2025). Esta misma información se distribuye a los proveedores a través de la plataforma Retail Link, una herramienta que permite consultar ventas, niveles de inventario y rotación de producto por tienda (18Knowledge, 2020). A esto se suman los registros de inventario, la información de la cadena de suministro y los datos financieros y operativos que la compañía administra internamente.

Fuentes internas semiestructuradas

Un segundo grupo corresponde a datos que se generan dentro de la empresa pero que no siempre siguen un formato tabular rígido, como los registros de sensores, cámaras y dispositivos del Internet de las Cosas (IoT) instalados en tiendas para monitorear anaqueles y flujo de clientes, un enfoque que Walmart comenzó a probar de forma más sistemática en su Intelligent Retail Lab a partir de 2019 (Digital Innovation and Transformation, 2020). También entran en esta categoría los registros de aplicaciones internas para asociados y los históricos de eventos de los sistemas logísticos, que combinan campos estructurados con notas o metadatos de formato más libre.

Fuentes externas estructuradas

Walmart también incorpora datos generados por terceros que llegan ya organizados en bases de datos, como los reportes de la firma Nielsen sobre comportamiento de consumo, la información meteorológica utilizada para anticipar la demanda de bebidas o ropa de temporada, los indicadores económicos y hasta los precios de la gasolina, todos ellos integrados como flujos de datos externos dentro del Data Café (Marr, 2017; Ankashram, 2025). Este tipo de fuente le permite a la compañía anticipar variaciones de demanda que no dependen de su propia operación, sino del entorno económico y climático en el que venden sus tiendas.

Fuentes externas semiestructuradas o no estructuradas

Finalmente, Walmart utiliza fuentes externas cuyo formato es más variable, como las menciones en redes sociales, que según estimaciones de la propia empresa alcanzan cerca de 300,000 interacciones semanales (ProjectPro, 2024), así como los datos de telecomunicaciones y los calendarios de eventos locales, que ayudan a explicar picos de demanda en zonas específicas (Retail IT Insights, 2017). Este tipo de información suele requerir procesamiento adicional de texto o geolocalización antes de poder cruzarse con los datos internos de ventas.

En conjunto, la selección de estas fuentes no responde a un criterio único, sino a la combinación de tres factores: la relevancia de la fuente para el problema de negocio que se quiere resolver (por ejemplo, quiebres de inventario o pronóstico de demanda), la posibilidad de integrarla en tiempo casi real dentro de la infraestructura de nube de la compañía, y el nivel de estructura del dato, que determina si puede cargarse directamente a un modelo analítico o si requiere un procesamiento previo. Esta combinación de fuentes internas y externas, estructuradas y semiestructuradas, es lo que le ha permitido a Walmart pasar de decisiones que antes tomaban semanas a análisis que hoy se resuelven en minutos (Marr, 2017).

1.2 Relación de los datos con la propuesta inicial

El objetivo inicial del proyecto de Business Intelligence aplicado a Walmart consiste en mejorar la predicción de la demanda y optimizar la gestión de inventarios, reduciendo tanto los quiebres de stock como los niveles de sobrestock. Para fortalecer esta propuesta, no es suficiente analizar únicamente las ventas históricas; es necesario relacionarlas con variables internas y externas que permitan comprender qué factores modifican la demanda, dónde se producen estos cambios y con qué anticipación pueden identificarse.

Como se identificó en la sección anterior, Walmart dispone de un ecosistema de información que combina datos internos estructurados —como ventas POS, inventarios y cadena de suministro— con información semiestructurada y fuentes externas relacionadas con clima, economía, telecomunicaciones, redes sociales y eventos locales. Esta diversidad de fuentes es especialmente relevante para nuestro proyecto porque permite enriquecer el comportamiento histórico de las

ventas con variables explicativas capaces de anticipar cambios futuros en la demanda.

El Data Café constituye un ejemplo de esta integración: Walmart ha utilizado cientos de flujos internos y externos junto con grandes volúmenes de información transaccional para realizar análisis prácticamente en tiempo real. Fuentes externas como datos meteorológicos, económicos, telecomunicaciones, redes sociales, precios de combustible y eventos cercanos a las tiendas pueden incorporarse a los datos transaccionales para explicar mejor las variaciones del negocio.

1.2.1 Relación entre las fuentes de datos y el modelo predictivo

Para integrar las diferentes fuentes se propone utilizar como eje común cuatro dimensiones:

Producto/SKU + Tienda/Ubicación + Fecha/Hora + Demanda

Estas dimensiones permiten relacionar la información histórica y operacional con factores externos y generar un dataset enriquecido para alimentar los modelos predictivos.

Tabla 15. *Relación entre las fuentes de datos, la demanda y su aplicación en el proyecto*

Fuente de datos	Relación con la demanda	Aplicación en nuestro proyecto
POS / ventas históricas	Identifica qué producto se vendió, cuándo, dónde y en qué cantidad.	Constituye la base histórica para entrenar el modelo predictivo.
Inventarios	Permite relacionar las ventas con la disponibilidad real del producto.	Ayuda a diferenciar una disminución real de demanda de una venta no realizada por falta de stock.
Supply Chain / Logística	Relaciona la demanda con tiempos de abastecimiento, distribución y reposición.	Permite calcular cuándo y cuánto reponer considerando demanda futura y lead time.
Retail Link	Integra información de ventas e inventarios disponible para proveedores.	Favorece la coordinación del abastecimiento y la planificación con proveedores.
IoT, sensores y cámaras	Generan señales relacionadas con flujo y comportamiento dentro de las tiendas.	Pueden complementar el análisis de tráfico, disponibilidad y operación física de cada establecimiento.
Clima	Temperatura, lluvia y otros fenómenos pueden modificar el consumo de determinadas categorías.	Permite anticipar variaciones de demanda por producto, categoría, tienda y región.
Datos de mercado/Nielsen	Permiten observar tendencias de consumo más allá de la operación interna.	Complementan el comportamiento histórico con cambios generales en las preferencias del consumidor.

Fuente de datos	Relación con la demanda	Aplicación en nuestro proyecto
Indicadores económicos	Variables macroeconómicas pueden afectar la capacidad y los patrones de consumo.	Permiten enriquecer los pronósticos de demanda de mediano plazo.
Precios de gasolina	Pueden aportar contexto sobre movilidad y condiciones económicas locales.	Funcionan como una variable externa complementaria para explicar variaciones geográficas.
Redes sociales	Permiten detectar tendencias, intereses y cambios en las conversaciones de los consumidores.	Mediante análisis de texto y sentimiento pueden utilizarse como señales tempranas de cambios en la demanda.
Eventos locales	Eventos deportivos, conciertos, festividades u otras actividades pueden producir picos extraordinarios de consumo.	Permiten anticipar incrementos temporales de demanda en determinadas tiendas o regiones.
Telecomunicaciones / geolocalización	Los datos agregados pueden aportar información sobre patrones de movilidad.	Complementan la estimación del flujo potencial de consumidores por ubicación.

Nota. Elaboración propia.

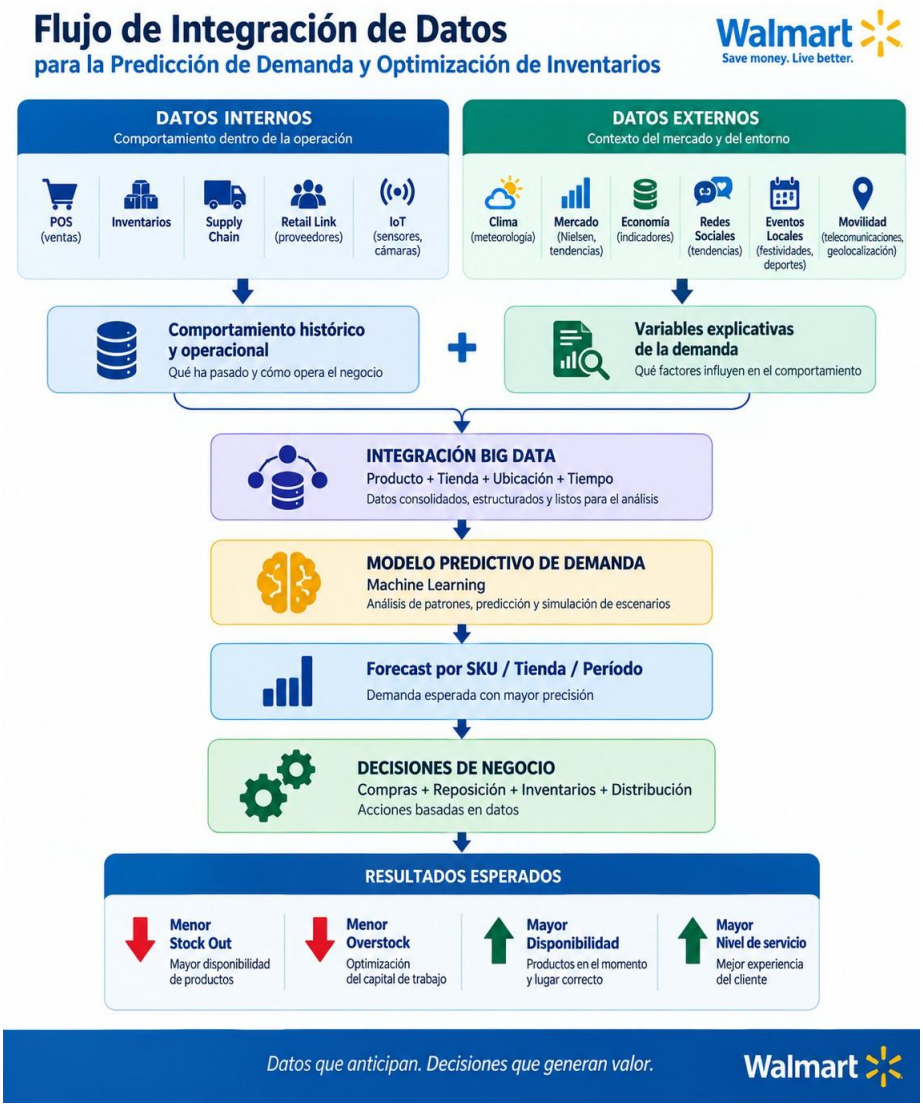
Las fuentes externas anteriores no reemplazan la información transaccional, sino que la enriquecen. El análisis desarrollado en la Parte I identifica precisamente información meteorológica, económica, precios de gasolina, redes sociales, telecomunicaciones y eventos locales como fuentes externas relevantes dentro del ecosistema analítico analizado.

Asimismo, Retail Link permite a proveedores consultar información relacionada con ventas e inventarios, lo que demuestra la importancia de compartir información operacional para mejorar la coordinación de la cadena de suministro.

1.2.2 Integración de los datos

La relación propuesta entre los diferentes datos puede representarse mediante el siguiente flujo:

Figura 5. ***Flujo de integración de datos para la predicción de demanda y la optimización de inventarios***



Nota. Elaboración propia.

1.2.3 ¿Cómo mejora esta integración nuestra propuesta inicial?

En las primeras etapas del proyecto se estableció un modelo de Business Intelligence orientado a utilizar información histórica y operacional para anticipar la demanda. La incorporación de las nuevas fuentes permite evolucionar progresivamente desde un enfoque descriptivo hacia uno predictivo y prescriptivo.

En una primera etapa, la analítica descriptiva permitirá responder:

- ¿Qué ocurrió con las ventas y los inventarios?

Posteriormente, al incorporar nuevas variables, la analítica predictiva permitirá responder:

- ¿Qué factores están modificando la demanda y cuánto esperamos vender?

Finalmente, la analítica prescriptiva permitirá responder:

- ¿Cuánto debemos reponer, dónde debemos hacerlo y en qué momento?

De esta manera, el sistema no se limitará a mostrar un posible quiebre de inventario una vez ocurrido, sino que podrá generar alertas anticipadas y recomendaciones de abastecimiento basadas en el comportamiento esperado de la demanda.

Aplicación al modelo

Podemos sintetizar esta evolución de la siguiente manera:

Figura 6. ***De los datos a las decisiones: evolución de la analítica descriptiva a la prescriptiva***



Nota. Elaboración propia.

1.2.4 Ejemplo aplicado al proyecto

Supongamos que para una determinada región el modelo detecta simultáneamente:

- un incremento significativo de temperatura para el próximo fin de semana

- un evento deportivo de alta concurrencia
- aumento de interés digital por determinadas categorías
- crecimiento histórico de ventas de bebidas y productos para reuniones bajo condiciones similares.

Esta información puede relacionarse con:

Ventas históricas + Inventario disponible + Clima + Evento local + Tendencias digitales + Lead time del proveedor El modelo predictivo podría identificar que determinadas tiendas presentan una probabilidad elevada de incremento de demanda y recomendar anticipadamente un aumento de inventario para categorías específicas.

Por ejemplo:

Figura 7. *Ejemplo de aplicación: detección anticipada de un riesgo de stock-out*



Nota. Elaboración propia.

1.2.5 Impacto esperado en la propuesta inicial

La incorporación de estas relaciones fortalecerá directamente cuatro componentes del proyecto:

Predicción de demanda: mayor cantidad de variables explicativas puede mejorar la capacidad del modelo para reconocer patrones y cambios en el comportamiento de compra.

Gestión de inventarios: la demanda proyectada podrá compararse continuamente con el stock disponible para detectar anticipadamente riesgos de agotamiento o exceso.

Abastecimiento: Compras podrá recibir recomendaciones de reposición basadas no solamente en ventas históricas, sino también en demanda esperada y tiempos de abastecimiento.

Toma de decisiones: los resultados podrán incorporarse a los dashboards de Power BI desarrollados en la Parte I, permitiendo visualizar alertas, forecast y recomendaciones.

En consecuencia, la relación entre estas fuentes completa el ciclo analítico diseñado durante el proyecto:

Datos → Integración → Análisis → Predicción → Recomendación → Decisión

1.3 Casos de uso que permitirán incorporar estos datos al modelo de BI

Una vez identificadas las fuentes y relaciones, debemos determinar **para qué utilizaremos la información dentro de Power BI**.

El modelo de BI tendrá como finalidad convertir los datos en información que permita tomar decisiones concretas.

1.3.1 Caso de uso 1: Predicción de demanda

Será el principal caso de uso del proyecto.

El modelo analizará el comportamiento histórico de las ventas y variables adicionales para generar una estimación de la demanda futura por:

- Producto.
- Categoría.
- Tienda.
- Región.
- Canal.
- Período.

El resultado podrá visualizarse en Power BI mediante gráficos de demanda histórica versus demanda proyectada.

1.3.2 Caso de uso 2: Detección de riesgo de desabastecimiento

El sistema identificará productos cuya demanda proyectada pueda superar el inventario disponible.

El dashboard podría mostrar:

Producto → Inventario actual → Demanda proyectada → Días de inventario → Nivel de riesgo.

Esto permitiría que los responsables de inventarios actúen antes de que el producto llegue a agotarse.

Este caso de uso está directamente relacionado con las soluciones que Walmart ya desarrolla para detectar desequilibrios de inventario.

1.3.3 Caso de uso 3: Detección de sobrestock

El modelo también identificará productos cuyo inventario sea significativamente superior a la demanda proyectada.

Por ejemplo:

Tabla 16. ***Ejemplo de detección de sobrestock***

Producto	Inventario	Demanda proyectada	Situación
Producto A	10.000	3.000	🔴 Sobrestock
Producto B	5.000	4.800	🟢 Normal
Producto C	2.000	3.500	🔴 Riesgo de quiebre

Nota. Elaboración propia.

Esto permitiría reducir costos de almacenamiento y evitar la inmovilización innecesaria de capital.

1.3.4 Caso de uso 4: Optimización de reposición

El sistema podrá recomendar cuándo y cuánto reponer.

La lógica sería:

Demanda proyectada + inventario disponible + tiempo de reposición = cantidad recomendada de reposición.

Este caso de uso conecta directamente el modelo predictivo con las decisiones de compras y abastecimiento.

1.3.5 Caso de uso 5: Análisis del impacto de promociones

Power BI permitirá comparar el comportamiento de las ventas antes, durante y después de una promoción.

Esto permitirá responder preguntas como:

- ¿Cuánto aumentaron las ventas?
- ¿Qué productos respondieron mejor?
- ¿En qué tiendas funcionó mejor la promoción?
- ¿La promoción generó sobrestock posterior?
- ¿Cuánto inventario adicional se necesitó?

1.3.6 Caso de uso 6: Optimización de distribución

La información de inventarios, demanda y ubicación permitirá identificar dónde existe exceso de producto y dónde existe necesidad.

Esto es particularmente coherente con la estrategia actual de Walmart, que utiliza sistemas de IA para **redirigir inventarios hacia los lugares donde son necesarios**.

1.3.7 Caso de uso 7: Dashboard ejecutivo

Finalmente, todos estos análisis podrán consolidarse en un dashboard ejecutivo de Power BI.

Los principales KPI podrían ser:

- Ventas totales.
- Crecimiento de ventas.
- Demanda proyectada.
- Exactitud del pronóstico.
- Nivel de inventario.
- Rotación de inventario.
- Tasa de desabastecimiento.
- Nivel de sobrestock.
- Productos críticos.
- Cumplimiento de abastecimiento.

De esta forma, Power BI se convierte en la capa de visualización que conecta los datos con los responsables de tomar decisiones.

1.4 Impacto del Big Data por áreas, departamentos y procesos de negocio

1.4.1 Justificación del impacto

En el caso de Walmart, el Big Data tiene un impacto estratégico porque permite integrar información procedente de múltiples fuentes y convertirla en conocimiento para anticipar comportamientos y optimizar procesos.

Walmart genera grandes cantidades de información a través de sus tiendas físicas, comercio electrónico, transacciones, inventarios, proveedores, logística y relación con los clientes. Si estos datos se analizan de manera independiente, su utilidad es limitada. Sin embargo, al integrarlos en un modelo de Business Intelligence (BI), es posible identificar relaciones y patrones que permitan tomar decisiones más precisas.

La diferencia fundamental sería:

Modelo tradicional

"¿Qué ocurrió?"

Business Intelligence

"¿Por qué ocurrió?"

Big Data + Analytics

"¿Qué probablemente ocurrirá y qué deberíamos hacer?"

Por ejemplo, en lugar de esperar a que un producto se agote para realizar una reposición, Walmart podría utilizar información histórica y actual de ventas, inventarios, ubicación, temporada, promociones y otros factores para anticipar la demanda.

1.4.2 Impacto en el área comercial y de ventas

El área de ventas es una de las principales beneficiadas.

Big Data permite analizar variables como:

- Historial de compras.
- Productos adquiridos.
- Frecuencia de compra.
- Horarios de compra.
- Ubicación de las tiendas.
- Comportamiento del cliente en canales digitales.
- Promociones utilizadas.
- Productos comprados conjuntamente.
- Tendencias de consumo.

Con esta información Walmart puede identificar patrones de comportamiento de los consumidores.

Ejemplo

Si el análisis identifica que determinados productos aumentan su demanda durante ciertos períodos, Walmart puede anticiparse y preparar inventario, promociones y personal.

El impacto sería:

Datos de clientes + datos de ventas → identificación de patrones → predicción de demanda → acciones comerciales → incremento de ventas.

Indicadores que podrían medirse

- Ventas por tienda.
- Ventas por producto.
- Ventas por categoría.
- Ticket promedio.
- Frecuencia de compra.
- Tasa de conversión.
- Margen por producto.

- Ventas generadas por promociones.

1.4.3 Impacto en Marketing

Big Data puede transformar el marketing de Walmart desde un enfoque masivo hacia uno segmentado y personalizado.

Por ejemplo, se pueden identificar grupos de clientes según:

- comportamiento de compra
- ubicación
- frecuencia
- categorías preferidas
- sensibilidad al precio
- respuesta a promociones
- comportamiento en canales digitales.

Esto permite diseñar promociones específicas para diferentes segmentos.

Ejemplo

En lugar de enviar la misma promoción a todos los clientes:

"20 % de descuento en productos seleccionados."

el sistema puede identificar qué productos tienen mayor probabilidad de interesar a cada segmento.

El proceso sería:

Figura 8. ***Proceso de marketing personalizado basado en datos***



Nota. Elaboración propia.

Impacto

- Mayor efectividad de las campañas.
- Menor desperdicio de inversión publicitaria.
- Mayor conversión.
- Mayor fidelización.
- Mejor conocimiento del consumidor.

1.4.4 Impacto en inventarios

Este es probablemente uno de los impactos más importantes para el proyecto de Walmart.

Uno de los problemas de cualquier empresa minorista es encontrar el equilibrio entre:

exceso de inventario / falta de inventario.

Demasiado inventario genera costos de almacenamiento y riesgo de productos sin rotación.

Muy poco inventario genera:

- productos agotados
- pérdida de ventas
- insatisfacción del cliente
- oportunidades perdidas.

Big Data permite relacionar:

Figura 9. ***Variables integradas en el modelo predictivo de inventario***



Nota. Elaboración propia.

Con ello se puede construir un modelo predictivo de inventario.

Ejemplo

Si el sistema detecta que una determinada tienda tendrá un incremento de demanda de un producto durante los próximos días, Walmart puede anticipar el abastecimiento.

Esto convierte el inventario de un proceso reactivo en uno predictivo.

Impacto

- Reducción de quiebres de stock.
- Reducción del exceso de inventario.
- Mayor rotación.
- Menores costos de almacenamiento.

- Mayor disponibilidad de productos.

1.4.5 Impacto en la cadena de suministro

Big Data también puede mejorar la Supply Chain de Walmart.

Se pueden analizar:

- proveedores
- tiempos de entrega
- cantidades solicitadas
- niveles de inventario
- rutas
- centros de distribución
- costos de transporte
- demanda por ubicación
- desempeño de proveedores.

Esto permite detectar cuellos de botella y anticipar problemas.

Ejemplo

Si el análisis identifica que determinados proveedores presentan recurrentemente retrasos, Walmart puede evaluar alternativas antes de que estos retrasos provoquen desabastecimiento.

Impacto

Figura 10. ***Recorrido de la cadena de suministro analizado con Big Data***



Nota. Elaboración propia.

El análisis de todo el recorrido permite optimizar la cadena completa, en lugar de analizar cada proceso aisladamente.

1.4.6 Impacto en comercio electrónico

El crecimiento del comercio electrónico genera otra fuente importante de datos.

Walmart puede analizar:

- búsquedas

- productos visualizados
- clics
- carritos abandonados
- compras
- frecuencia de navegación
- dispositivos utilizados
- ubicación
- tiempos de entrega
- productos recomendados.

Estos datos permiten entender qué quiere comprar el cliente incluso antes de que efectúe la compra.

Ejemplo

Un cliente busca varias veces un producto, lo agrega al carrito y posteriormente abandona la compra.

Ese comportamiento puede convertirse en información para:

- realizar una recomendación
- ofrecer una promoción
- mejorar la experiencia
- identificar problemas en el proceso de compra.

1.4.7 Impacto en atención y experiencia del cliente

Big Data también puede utilizarse para analizar información no estructurada, como:

- comentarios
- reseñas
- reclamaciones
- encuestas
- conversaciones
- redes sociales.

El material del módulo destaca precisamente el uso de Text Analytics para extraer palabras clave y transformar información desorganizada en información útil para el análisis.

En Walmart, por ejemplo, se podría analizar automáticamente miles de comentarios para detectar temas recurrentes:

"tiempo de entrega"

"producto agotado"

"precio"

"atención"

"calidad"

De esta manera, la empresa podría determinar cuáles son los principales problemas experimentados por sus clientes.

Impacto

- Mayor satisfacción.
- Reducción de reclamaciones.
- Identificación rápida de problemas.
- Mejor experiencia omnicanal.
- Mayor fidelización.

1.4.8 Impacto en Finanzas

El área financiera también puede beneficiarse de la integración de información.

El Big Data puede relacionar:

Figura 11. ***Integración de datos financieros para decisiones estratégicas***



Nota. Elaboración propia.

Esto permite analizar no solamente cuánto vende Walmart, sino qué ventas son realmente rentables.

Por ejemplo, una promoción puede incrementar las unidades vendidas, pero reducir significativamente el margen.

El análisis permite determinar:

¿La promoción generó realmente valor para Walmart?

Impacto

- Mejor análisis de rentabilidad.
- Control de costos.
- Evaluación de promociones.
- Proyecciones financieras.
- Mejor asignación de recursos.

1.4.9 Impacto en Recursos Humanos y operaciones

Big Data también puede utilizarse para optimizar la operación de las tiendas.

Por ejemplo, se pueden relacionar:

afluencia de clientes + ventas + horarios + días de mayor demanda + disponibilidad de personal.

Esto permitiría determinar cuántos trabajadores se necesitan en determinados horarios.

Ejemplo

Si una tienda presenta un incremento considerable de clientes durante determinadas horas, Walmart podría ajustar la programación del personal.

Impacto

- Mejor utilización del personal.
- Menores tiempos de espera.
- Mayor productividad.
- Mejor servicio al cliente.
- Reducción de costos operativos.

1.4.10 Impacto en la toma de decisiones gerenciales

Finalmente, uno de los mayores aportes del Big Data está en la dirección estratégica.

Los gerentes podrían disponer de dashboards de BI que integren:

Ventas + clientes + inventario + logística + marketing + finanzas.

Esto permite pasar de múltiples informes independientes a una visión integral del negocio.

Por ejemplo, un gerente podría identificar:

Una categoría está disminuyendo sus ventas.

Pero el análisis de Big Data podría revelar que el problema no es la falta de demanda, sino:

menor inventario → productos agotados → pérdida de ventas.

Por lo tanto, la solución no sería necesariamente aumentar la publicidad, sino mejorar el abastecimiento.

Esto demuestra por qué la relación entre datos es fundamental.

1.4.11 Mapa de impacto para el proyecto

Podemos resumir la propuesta de Walmart de la siguiente manera:

Tabla 17. ***Mapa de impacto del Big Data por área o departamento***

Área / Departamento	Datos analizados	Aplicación Big Data	Impacto
Ventas	Transacciones, productos, tiendas	Predicción de ventas	Aumento Ingresos
Marketing	Clientes, compras, comportamiento	Personalización	Aumento Conversión
Inventarios	Stock, ventas, demanda	Predicción de demanda	Reducción Quiebres de stock
Compras	Proveedores, demanda, precios	Planificación de compras	Reducción Costos
Logística	Rutas, entregas, inventarios	Optimización logística	Reducción Tiempo y costos
E-commerce	Búsquedas, clics, carritos	Recomendaciones	Aumento Ventas digitales
Servicio al cliente	Reseñas, reclamos, comentarios	Text Analytics	Aumento Satisfacción
Finanzas	Ventas, costos, márgenes	Rentabilidad predictiva	Aumento Margen
RR. HH.	Afluencia, horarios, ventas	Planificación de personal	Aumento Productividad
Gerencia	Todos los anteriores	BI y analítica predictiva	Aumento Calidad de decisiones

Nota. Elaboración propia.

1.4.12 Impacto global esperado

El impacto del Big Data para Walmart puede representarse mediante cinco niveles:

Figura 12. ***Impacto global esperado del Big Data en Walmart***



Nota. Elaboración propia.

CAPÍTULO 2. Arquitectura Big Data en la nube: proveedor, modelo de servicio e integración

Con las fuentes de datos identificadas, el siguiente paso es decidir sobre qué infraestructura se integrarán, almacenarán y procesarán. Este capítulo compara los principales proveedores de nube, analiza la estrategia multinube de Walmart y las alternativas de subcontratación de procesos, justifica la selección de Microsoft Azure y del modelo de servicio PaaS, y describe el flujo de integración entre la arquitectura Big Data, el data warehouse y la capa de Business Intelligence.

Continuando con el análisis del área de tecnología de Walmart USA, este capítulo aborda las alternativas de proveedores tecnológicos disponibles para dos de sus decisiones más relevantes: la infraestructura de nube / arquitectura Big Data y la subcontratación de procesos administrativos (BPO) con componente digital.

Walmart es un negocio particularmente intensivo en datos: combina tiendas físicas, comercio electrónico, cadena de suministro, inventarios, transacciones, comportamiento de clientes y operaciones omnicanal. El informe anual de Walmart describe a la compañía como un retailer omnicanal impulsado por tecnología, con operaciones en tiendas y comercio electrónico. Por ello, para la selección de tecnologías no solo se considera el precio, sino también la escalabilidad, el procesamiento de grandes volúmenes de datos, la integración con sistemas existentes, la analítica, la inteligencia artificial, la seguridad, el gobierno de datos y la capacidad de operar a escala global.

Walmart no ha elegido al proveedor más grande del mercado por defecto, sino que ha evaluado varias opciones frente a criterios propios de su estrategia de negocio, combinando el desarrollo interno con alianzas externas cuidadosamente seleccionadas.

2.1 Análisis de alternativas de nube y arquitectura Big Data

El mercado de nube pública está dominado por tres proveedores principales: Amazon Web Services (AWS), Microsoft Azure y Google Cloud Platform (GCP).

2.1.1 Cuadro comparativo de proveedores de nube y Big Data

Tabla 18. ***Cuadro comparativo de proveedores de nube y Big Data***

Criterio	AWS	Microsoft Azure	Google Cloud (GCP)
Escalabilidad	Muy alta	Muy alta	Muy alta
Big Data	Excelente	Excelente	Excelente
Data Lake	Amazon S3	Azure Data Lake Storage	Cloud Storage

Criterio	AWS	Microsoft Azure	Google Cloud (GCP)
Procesamiento	Glue, EMR, Athena	Data Factory, Synapse, Databricks	Dataflow, Dataproc
Data Warehouse	Redshift	Synapse / Fabric Warehouse	BigQuery
Machine Learning / IA	SageMaker	Azure ML / Azure AI	Vertex AI
BI	QuickSight	Power BI	Looker
Streaming	Kinesis	Event Hubs	Pub/Sub / Dataflow
Gobierno de datos	Lake Formation	Microsoft Purview / Fabric	Dataplex
Integración empresarial	Muy alta	Muy alta para Walmart	Muy alta
Experiencia previa con Walmart	Alta	Alta	Alta
Adecuación propuesta	Muy alta	Muy alta	Muy alta

Nota. Elaboración propia.

2.1.2 Evaluación detallada por proveedor

A. Amazon Web Services (AWS)

AWS fue el primero en lanzarse (2006) y sigue siendo el más utilizado entre grandes empresas gracias a su amplio catálogo de servicios y red global (más de 39 regiones y 700 puntos de presencia). Ofrece una arquitectura muy sólida para Big Data: Amazon S3 como Data Lake, AWS Glue para ETL y catálogo, Lake Formation para gobierno centralizado y Amazon Redshift como Data Warehouse.

- **Ventajas para Walmart:** Gran capacidad de almacenamiento, excelente ecosistema Big Data, amplia oferta de servicios, procesamiento distribuido, madurez en Data Lakes y amplias capacidades de Machine Learning.
- **Desventajas:** Ecosistema complejo que puede incrementar la carga administrativa y requerir la mantención de una mayor diversidad de tecnologías.
- **Evaluación Estratégica:** A pesar de ser el proveedor más consolidado, Walmart descartó a AWS como proveedor principal debido a que Amazon es su mayor competidor directo en el comercio minorista, lo cual representa un riesgo estratégico inaceptable.

B. Microsoft Azure

Azure es una alternativa sólida para entornos híbridos, usada por más del 95% de las empresas Fortune 500. Existe una relación tecnológica histórica entre Walmart y Microsoft. Walmart adoptó Azure para acelerar su transformación digital y utilizó Azure Cosmos DB en Walmart.com. Recientemente, ha integrado Azure OpenAI para personalizar experiencias de compra.

- **Componentes de arquitectura:** Azure Data Lake Storage, Azure Data Factory, Azure Synapse Analytics (que integra Data Warehouse, Big Data, Spark y pipelines), Azure Databricks, Azure Machine Learning, Power BI y Microsoft Fabric.
- **Ventajas para Walmart:** Integración directa con tecnologías que Walmart ya utiliza. Excelente capacidad Big Data, integración unificada de Data Lake, Data Warehouse, BI (Power BI) e IA, y escalabilidad global.
- **Desventaja principal:** Riesgo de dependencia (vendor lock-in) si toda la arquitectura depende exclusivamente de servicios propietarios de Microsoft.

C. Google Cloud Platform (GCP)

GCP se distingue por su capacidad en analítica de datos, aprendizaje automático y mayor claridad en sus costos. Su fortaleza radica en BigQuery, Dataflow (procesamiento en tiempo real/batch para eventos, clics y sensores), Dataproc, Cloud Storage y Vertex AI. Google Cloud cuenta con experiencia directa con Walmart; por ejemplo, en 2025 se destacó el uso de Spanner para cargas de trabajo de gran escala e iniciativas de unificación de datos.

- **Ventajas:** BigQuery es excepcionalmente potente para analítica a gran escala. Gran capacidad en procesamiento de datos en tiempo real, aprendizaje automático y experiencia previa comprobada con la empresa.
- **Desventaja:** Requiere integración con otros componentes existentes en Walmart, lo que puede dar lugar a una arquitectura más heterogénea.

2.1.3 La estrategia de nube de Walmart: el Triplet Model

En lugar de elegir un solo proveedor o quedarse por defecto con el más grande del mercado (AWS), Walmart adoptó una arquitectura basada en el Triplet Model, el cual combina:

- **Microsoft Azure:** Infraestructura en la nube pública de Microsoft.
- **Google Cloud Platform (GCP):** Capacidades analíticas y de datos avanzadas.
- **Nube Privada Propia:** Control interno y soberanía de datos críticos.

Según Kevin Evans, vicepresidente de infraestructura de Walmart, trabajar con un modelo híbrido y multi-nube evita la dependencia de un solo proveedor y otorga mayor poder de negociación sobre precios y condiciones contractuales. Esta decisión confirma que Walmart prioriza el ajuste estratégico y la independencia sobre la cuota de mercado del proveedor.

2.1.4 Alternativas de subcontratación de procesos (BPO)

En el ámbito de la subcontratación de procesos de negocio con componente digital, Walmart mantiene desde 2018 una alianza con Genpact para operaciones de finanzas y contabilidad, ampliada en 2023 a Norteamérica, América Latina y Sudáfrica.

Comparativa de mercado en BPO de Finanzas y Contabilidad

Evaluaciones independientes de la industria de BPO ubican a cuatro proveedores principales a la cabeza:

- **Genpact:** Recomendado cuando se requiere una gestión de flujos de trabajo configurable con controles de proceso bien gobernados.
- **TCS (Tata Consultancy Services):** Destaca por combinar modelos de operación estandarizados con un seguimiento riguroso de indicadores.
- **Cognizant:** Reconocido por sus capacidades de analítica e integración de servicios a gran escala.
- **Accenture:** Reconocido, junto con Cognizant, por la integración de servicios a gran escala y capacidades analíticas avanzadas.

Criterio de selección de Walmart

Frente a estas opciones, Walmart ha optado por mantener y expandir su relación con Genpact. Esto demuestra que la empresa valora la continuidad y la confianza acumulada con un socio que ya conoce sus procesos internos. Sustituir a un proveedor de finanzas y contabilidad consolidado resulta costoso y complejo, por lo que la continuidad prevalece sobre el cambio a competidores con capacidades similares.

2.2 Selección de la mejor alternativa de proveedor Cloud

Una vez analizadas las alternativas de proveedores de servicios en la nube, se propone seleccionar Microsoft Azure como proveedor cloud de referencia para la arquitectura de información del proyecto de Business Intelligence y Big Data de Walmart.

La selección responde a los requerimientos definidos durante las diferentes etapas del proyecto: integrar fuentes de datos internas y externas, almacenar información estructurada, semiestructurada y no estructurada, procesar grandes volúmenes de datos, implementar modelos predictivos de demanda y disponibilizar los resultados para la toma de decisiones.

Además, esta elección mantiene coherencia con las decisiones tecnológicas adoptadas en los capítulos anteriores, particularmente con la selección de Power BI como herramienta de Business Intelligence y visualización.

2.2.1 Criterios considerados para la selección

La evaluación de Microsoft Azure se realizó considerando los requerimientos funcionales y tecnológicos del proyecto:

Tabla 19. ***Criterios considerados para la selección de Microsoft Azure***

Criterio	Necesidad del proyecto Walmart	Valoración de Azure
Escalabilidad	Procesar el crecimiento de datos provenientes de ventas, inventarios, Supply Chain, IoT y fuentes externas.	Muy alta

Criterio	Necesidad del proyecto Walmart	Valoración de Azure
Integración de datos	Integrar POS, inventarios, logística, Retail Link, IoT, clima, mercado, redes sociales, eventos y movilidad.	Muy alta
Big Data	Gestionar datos estructurados, semiestructurados y no estructurados.	Muy alta
Machine Learning	Desarrollar modelos predictivos de demanda por SKU, tienda y período.	Muy alta
Seguridad y gobierno	Gestionar identidad, accesos, trazabilidad, calidad, monitoreo y protección de la información.	Muy alta
Integración con Power BI	Alimentar dashboards, KPIs, forecast y alertas definidos en etapas anteriores.	Muy alta
Flexibilidad Cloud	Adaptar recursos de almacenamiento y procesamiento a las necesidades de la solución.	Muy alta

Nota. Elaboración propia.

Estos criterios guardan relación con las capacidades establecidas en el material del módulo para una arquitectura Big Data, entre ellas almacenamiento y procesamiento distribuido, integración de bases de datos, autenticación y autorización, administración, auditoría, gestión de metadata, Machine Learning y procesamiento de streams.

2.2.2 Justificación de Microsoft Azure

La principal razón para seleccionar Microsoft Azure es su capacidad para soportar de manera integrada las diferentes etapas de la arquitectura propuesta, desde la incorporación y almacenamiento de los datos hasta su procesamiento, análisis y posterior utilización en Business Intelligence.

En nuestro proyecto, la plataforma cloud deberá recibir información procedente de múltiples fuentes:

Figura 13. ***Solución propuesta para Walmart: datos integrados en la nube***



Nota. Elaboración propia.

El material académico establece que una Landing Zone permite centralizar datos provenientes de fuentes internas y externas y ejecutar procesos de calidad, transformación y enriquecimiento antes de utilizarlos en Data Warehouses, Data Marts y soluciones de negocio.

Por lo tanto, Azure permitiría soportar conceptualmente esta arquitectura y proporcionar la escalabilidad necesaria para incorporar progresivamente nuevas fuentes y mayores volúmenes de información.

2.2.3 Coherencia con la selección previa de Power BI

Un criterio relevante para seleccionar Azure es la continuidad con la arquitectura tecnológica definida en los capítulos anteriores del proyecto, en las cuales se seleccionó Power BI como herramienta de Business Intelligence y visualización.

Power BI constituye la capa mediante la cual los usuarios de Walmart podrán visualizar los resultados del procesamiento y análisis de los datos a través de:

- dashboards ejecutivos y operacionales;
- indicadores y KPIs;
- forecast de demanda;
- alertas de riesgo de stock-out y sobrestock;
- recomendaciones para abastecimiento e inventarios.

Azure, por su parte, representa la infraestructura cloud encargada de soportar la integración, almacenamiento, procesamiento y preparación de los datos.

Por consiguiente, ambas tecnologías cumplen funciones diferentes pero complementarias:

AZURE = Integrar + Almacenar + Procesar + Analizar

POWER BI = Visualizar + Monitorear + Apoyar decisiones

La selección de Power BI no obliga técnicamente a utilizar Azure, ya que la herramienta puede conectarse con diferentes plataformas y fuentes de información. Sin embargo, para la arquitectura propuesta en este proyecto, mantener ambos componentes dentro del ecosistema Microsoft favorece una solución conceptualmente integrada y evita incorporar tecnologías adicionales sin una necesidad específica.

2.2.4 Aplicación de Azure al modelo predictivo de Walmart

La selección del proveedor debe contribuir directamente a resolver el problema de negocio planteado desde el inicio del proyecto.

Por ejemplo:

Figura 14. ***Del dato a la decisión: aplicación de Azure al modelo predictivo***



Nota. Elaboración propia.

De esta forma, la plataforma cloud no se limita a almacenar información. Su función dentro de la propuesta es proporcionar la infraestructura necesaria para relacionar múltiples fuentes de datos y convertirlas en información predictiva que pueda ser utilizada para tomar decisiones operativas.

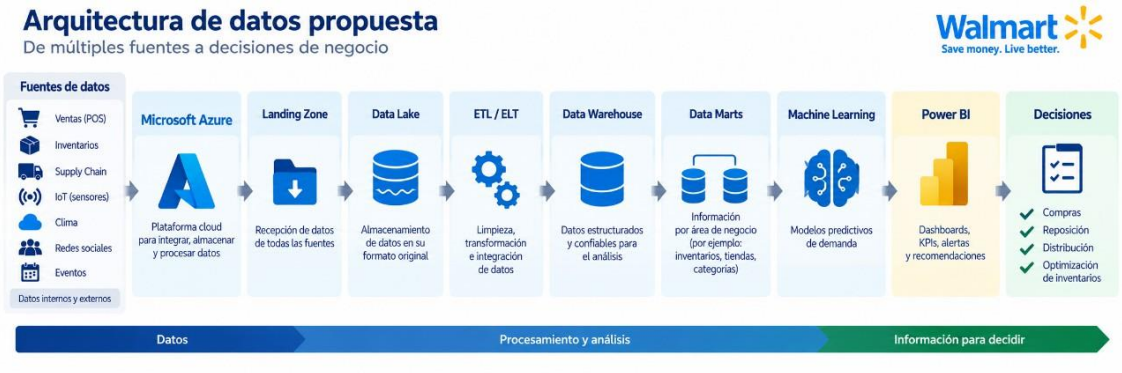
2.2.5 Decisión final

Proveedor cloud seleccionado: Microsoft Azure

Microsoft Azure se selecciona como la alternativa de referencia que mejor se ajusta a los requerimientos definidos para nuestro proyecto de Walmart debido a su capacidad de escalabilidad, integración de fuentes heterogéneas, procesamiento de Big Data, seguridad, gobierno del dato, analítica avanzada y soporte para Machine Learning.

Adicionalmente, su relación con Power BI, herramienta seleccionada previamente para la visualización y explotación de información, permite plantear una arquitectura tecnológica coherente de extremo a extremo:

Figura 15. ***Arquitectura de datos propuesta sobre Microsoft Azure***



Nota. Elaboración propia.

Esta arquitectura permitirá transformar grandes volúmenes y variedades de datos en pronósticos de demanda, alertas y recomendaciones, contribuyendo al objetivo central del proyecto: reducir los quiebres de stock y el sobrestock, mejorar la disponibilidad de productos y optimizar las decisiones de abastecimiento e inventario de Walmart.

Finalmente, la selección de Azure debe entenderse como una propuesta tecnológica desarrollada para este proyecto académico, y no como una afirmación de que Walmart utilice exclusivamente este proveedor en su operación real. El material de la asignatura solicita precisamente seleccionar la alternativa que el equipo considere más adecuada para su propia arquitectura.

2.3 Modelo de cloud computing: PaaS

Para el desarrollo del proyecto de Business Intelligence orientado a la optimización de las ventas y la demanda de Walmart, se selecciona **PaaS (Platform as a Service)** como el modelo principal de servicio en la nube. Esta elección responde a las características y necesidades del proyecto, debido a que la solución propuesta no requiere únicamente una herramienta para visualizar información, sino una plataforma tecnológica capaz de integrar diferentes fuentes de datos, procesarlos, almacenarlos, analizarlos y desarrollar modelos predictivos que permitan mejorar la planificación de inventarios y la toma de decisiones.

El proyecto requiere trabajar con información proveniente de diferentes fuentes públicas, como reportes financieros y operativos de Walmart, información sobre ventas, indicadores económicos, datos relacionados con temporadas y fechas especiales, información climática, tendencias de consumo y otras variables externas que pueden tener relación con el comportamiento de la demanda. Debido a esta diversidad de fuentes, es necesario contar con un entorno que facilite la integración y preparación de los datos antes de utilizarlos para el análisis. En este sentido, PaaS proporciona una plataforma administrada que permite desarrollar y ejecutar procesos de integración y procesamiento sin que el equipo tenga que encargarse directamente de la administración de servidores

físicos o de toda la infraestructura tecnológica que se encuentra detrás de estos servicios.

Otro aspecto importante es que el proyecto contempla el uso de **analítica predictiva**. El objetivo no es solamente conocer cuánto se ha vendido anteriormente, sino utilizar los datos históricos y las variables disponibles para identificar patrones y generar estimaciones sobre el comportamiento futuro de la demanda. Esto resulta especialmente relevante para Walmart debido a la magnitud de sus operaciones y a la necesidad de mantener niveles adecuados de inventario. Una plataforma PaaS permite incorporar servicios especializados para el desarrollo, entrenamiento y ejecución de modelos de aprendizaje automático, facilitando que los datos preparados puedan ser utilizados posteriormente para generar predicciones.

La selección de PaaS también está relacionada con la necesidad de **escalabilidad**. Walmart maneja una operación de gran magnitud y el proyecto está planteado bajo una perspectiva de Business Intelligence y Big Data. Por esta razón, la solución debe tener la capacidad de crecer en función del volumen de información procesada y de las necesidades futuras de análisis. PaaS permite utilizar recursos tecnológicos de manera flexible sin que sea necesario diseñar y administrar manualmente toda la infraestructura cada vez que aumentan los datos o los requerimientos del sistema.

Además, PaaS permite que el equipo del proyecto concentre sus esfuerzos en los aspectos que realmente aportan valor al negocio: la preparación de los datos, la construcción de indicadores, el desarrollo de modelos predictivos, la identificación de patrones de demanda y la generación de información útil para la toma de decisiones. La administración de elementos de infraestructura, como servidores, sistemas operativos y determinados componentes de red, queda en gran medida bajo responsabilidad del proveedor del servicio. Esto representa una ventaja frente a un modelo IaaS, en el cual el equipo tendría una mayor responsabilidad sobre la configuración y administración de la infraestructura.

En el caso de **IaaS (Infrastructure as a Service)**, aunque proporciona un alto nivel de control sobre los recursos tecnológicos, este nivel de control no constituye una necesidad principal para el proyecto. El objetivo no es administrar máquinas virtuales, sistemas operativos o infraestructura tecnológica desde un nivel bajo, sino disponer de un entorno que facilite el desarrollo de una solución de datos y analítica. Por lo tanto, utilizar IaaS implicaría asumir responsabilidades adicionales de infraestructura que no son esenciales para alcanzar los objetivos establecidos.

Por otra parte, **SaaS (Software as a Service)** resulta adecuado para consumir aplicaciones terminadas a través de Internet. Dentro del proyecto, por ejemplo, Power BI puede utilizarse como una solución de este tipo para desarrollar y consultar dashboards e indicadores. Sin embargo, SaaS por sí solo no representa completamente la arquitectura necesaria para el proyecto, porque la solución requiere realizar procesos de integración, almacenamiento, transformación de datos y desarrollo de modelos predictivos. Por esta razón,

aunque algunas herramientas utilizadas pueden funcionar bajo un esquema SaaS, el modelo general de servicio seleccionado para la arquitectura es PaaS.

Para este proyecto se ha seleccionado **Microsoft Azure** como proveedor de servicios en la nube. Dentro de este entorno pueden utilizarse diferentes servicios de acuerdo con las necesidades de cada etapa. Azure Data Factory puede apoyar la integración y preparación de las diferentes fuentes de información; Azure Data Lake Storage puede utilizarse como entorno de almacenamiento para organizar los datos; Azure Machine Learning puede apoyar el desarrollo de los modelos predictivos; y Power BI puede utilizarse para transformar los resultados del análisis en dashboards e indicadores comprensibles para los usuarios del negocio.

El funcionamiento de la solución comienza con la recopilación de las fuentes de datos públicas seleccionadas. Posteriormente, la información debe ser integrada y preparada para eliminar inconsistencias, organizar las variables y establecer una estructura adecuada para el análisis. Una vez que los datos se encuentran preparados, pueden almacenarse en un entorno centralizado y utilizarse para realizar diferentes tipos de análisis. A partir de estos datos se pueden desarrollar modelos predictivos capaces de identificar patrones y estimar el comportamiento futuro de la demanda. Finalmente, los resultados pueden ser presentados mediante Power BI a través de dashboards que permitan visualizar indicadores relacionados con ventas, demanda e inventarios.

De esta manera, la selección de PaaS es coherente con el enfoque general del proyecto, porque permite construir una solución tecnológica orientada al análisis de datos sin que el equipo tenga que asumir directamente la administración de toda la infraestructura. Además, proporciona un entorno adecuado para incorporar capacidades de Big Data, analítica avanzada y modelos predictivos, elementos fundamentales para el objetivo de optimizar las ventas y la gestión de inventarios de Walmart.

2.4 Flujo de integración de la arquitectura Big Data con el data warehouse y la estructura de Business Intelligence

2.4.1 Objetivo

Definido el proveedor de nube y el modelo de servicio en las secciones anteriores, este apartado responde la pregunta operativa que queda abierta: cómo circula el dato desde que se genera —una venta en el punto de pago, un clic en la aplicación, un movimiento de inventario o un archivo de un proveedor externo— hasta que un planificador de demanda o un sistema de reposición decide con él. Se conserva la lógica del caso de referencia, en el que la Landing Zone actúa como zona de exposición temprana y como fuente de datos de las bodegas y datamarts, pero instanciada sobre la pila tecnológica real de Walmart Inc. y orientada al caso de uso que articula el proyecto: el pronóstico de demanda.

2.4.2 Capas de la arquitectura de integración

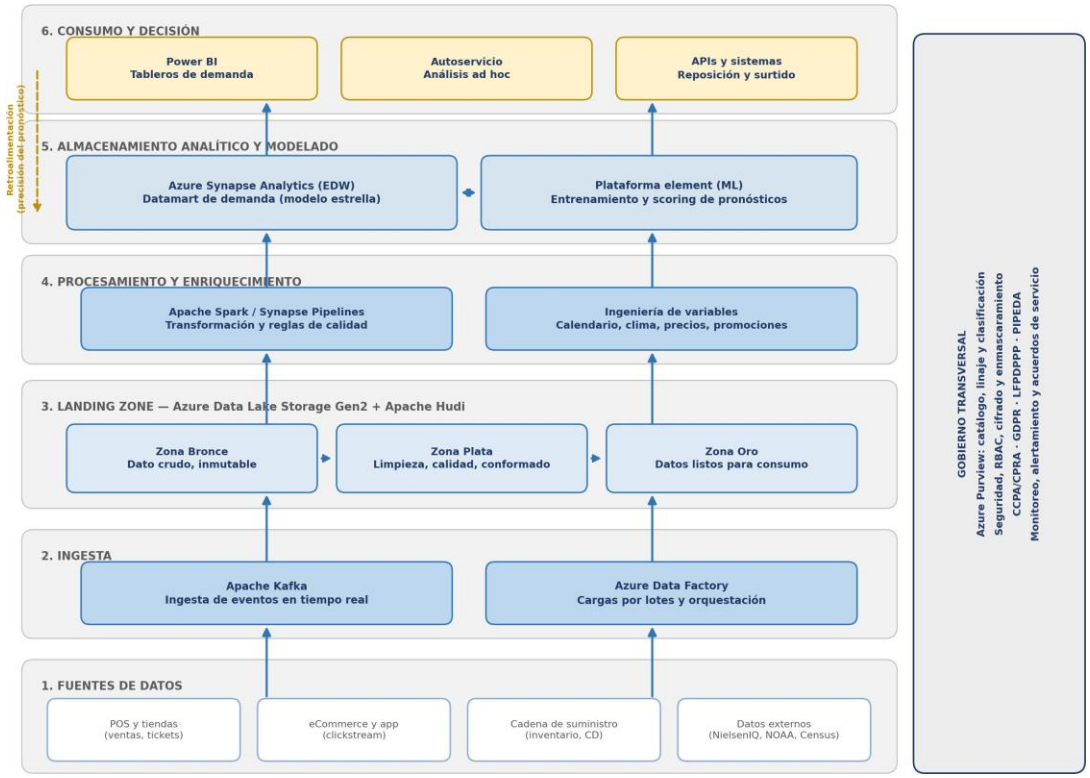
El flujo se organiza en seis capas funcionales más una franja transversal de gobierno que acompaña al dato en todo su recorrido:

Tabla 20. ***Capas de la arquitectura de integración***

Capa	Componentes	Función principal en el flujo
1. Fuentes de datos	POS de tiendas, eCommerce y aplicación móvil, sistemas de inventario y cadena de suministro, proveedores externos (NielsenIQ, NOAA, U.S. Census Bureau)	Originar el dato transaccional, de comportamiento y de contexto de mercado
2. Ingesta	Apache Kafka, Azure Data Factory	Capturar eventos en tiempo real y cargas masivas periódicas sin acoplar el origen al consumidor
3. Landing Zone	Azure Data Lake Storage Gen2 con formato de tabla Apache Hudi; zonas bronce, plata y oro	Aterrizar, conservar y conformar el dato crudo, y exponerlo tempranamente a toda la organización
4. Procesamiento y enriquecimiento	Apache Spark, canalizaciones de Azure Synapse	Aplicar reglas de calidad, transformar y construir las variables analíticas
5. Almacenamiento analítico y modelado	Azure Synapse Analytics (data warehouse y datamart de demanda), plataforma element de aprendizaje automático	Servir el modelo dimensional conformado y ejecutar el entrenamiento y la ejecución de los pronósticos
6. Consumo y decisión	Power BI, análisis de autoservicio, APIs hacia sistemas de reposición y surtido	Entregar el resultado al decisor humano y a los sistemas operativos
Transversal	Azure Purview, control de acceso basado en roles, cifrado, monitoreo	Catalogar, trazar, proteger y vigilar el cumplimiento de los acuerdos de servicio

Nota. Elaboración propia.

Figura 16. ***Flujo de integración de la arquitectura Big Data con el data warehouse y la capa de BI***



Nota. Elaboración propia.

2.4.3 Descripción del flujo

Ingesta y aterrizaje

La ingesta opera con dos mecanismos complementarios. El flujo continuo se apoya en Apache Kafka, columna vertebral de eventos de la compañía: los sistemas de punto de venta, de comercio electrónico y de inventario publican allí sus eventos y los consumidores analíticos se suscriben sin afectar al productor, de modo que un mismo evento de venta alimenta a la vez el pronóstico de demanda, el seguimiento de faltantes y los tableros operativos. El flujo por lotes se orquesta con Azure Data Factory y atiende lo que no se genera como evento: maestros de producto y tienda, jerarquías, precios, calendarios promocionales y fuentes externas contratadas.

Todo lo ingresado aterriza en la Landing Zone construida sobre Azure Data Lake Storage Gen2 y organizada en tres zonas sucesivas: bronce, donde el dato se conserva tal como llegó, inmutable y particionado por fecha; plata, donde se limpia, deduplica y conforma contra los identificadores corporativos de ítem y tienda; y oro, donde queda listo para consumo analítico. La gestión de estas tablas se realiza con Apache Hudi, formato que Walmart seleccionó para su lago de datos de siguiente generación tras evaluar disponibilidad, compatibilidad,

costo, desempeño y soporte. La elección importa por una razón concreta: permite actualizaciones incrementales sobre el lago, con lo que una corrección de venta no obliga a reprocesar el histórico completo, y conserva versiones anteriores para reconstruir el estado del dato en una fecha pasada.

Procesamiento y carga al data warehouse

Entre las zonas bronce y oro se ejecutan procesos distribuidos de Apache Spark orquestados desde Synapse, que aplican las reglas de calidad acordadas con los dueños del dato, consolidan la serie de ventas a la granularidad de ítem–tienda–día y construyen las variables predictivas: calendario fiscal y de festivos, precio propio y de la competencia, mecánica y profundidad de promoción, clima observado y pronosticado, e indicadores económicos locales. Este enriquecimiento con datos internos y externos es lo que separa un pronóstico basado en la inercia del histórico de uno capaz de anticipar quiebres de tendencia.

Desde la zona oro se ejecuta la carga incremental hacia Azure Synapse Analytics, donde reside el data warehouse y el datamart de demanda, modelado como esquema estrella con hechos de venta y de pronóstico rodeados de dimensiones conformadas de producto, tienda, tiempo, promoción y contexto externo. Conviene subrayar la dirección de la relación: el lago y el almacén no son alternativas que compitan, sino etapas encadenadas. La Landing Zone es la fuente de datos del data warehouse y aporta volumen, variedad y flexibilidad; el data warehouse aporta la semántica conformada, la consistencia y el desempeño de consulta que la capa de BI necesita.

Analítica avanzada y consumo

El entrenamiento y la ejecución de los modelos ocurren en element, la plataforma de aprendizaje automático propia de Walmart, que opera bajo un modelo de despliegue en tres entornos —nube privada, Google Cloud Platform y Microsoft Azure— y permite construir una solución una sola vez y desplegarla en cualquiera de ellos, administrando varios millones de modelos en paralelo. Para el pronóstico de demanda, el flujo reproduce el enfoque documentado por la propia compañía en sus patentes: un motor común de preparación que procesa la información de forma incremental y un motor de pronóstico que ejecuta un modelo principal de red neuronal recurrente, apoyado en modelos complementarios de tipo ARIMA o LOESS cuando la estacionalidad o los eventos de calendario lo requieren, combinados mediante un ensamble ponderado.

El resultado no se queda en la plataforma analítica: se escribe de vuelta en el datamart como tabla de hechos de pronóstico, versionada por modelo y fecha de ejecución, de modo que la capa de BI consuma histórico y predicción desde un mismo lugar y con las mismas dimensiones. La entrega final se realiza en Power BI, con un modelo semántico único que define las métricas certificadas, y por servicios hacia los sistemas de reposición y surtido, que convierten la predicción en órdenes sugeridas. Cierra el circuito la comparación semanal entre pronóstico y venta real, que alimenta el reentrenamiento de los modelos.

2.4.4 El flujo aplicado al pronóstico de demanda

La tabla siguiente traza el recorrido completo de un dato de venta, de modo que el flujo pueda leerse como una secuencia operativa y no solo como un esquema conceptual:

Tabla 21. ***Recorrido de un dato de venta en el flujo de pronóstico de demanda***

Paso	Componente	Frecuencia	Resultado
1. Captura de la venta	POS y eCommerce hacia Kafka	Continua	Evento de venta por ítem y tienda
2. Aterrizaje del dato crudo	Landing Zone, zona bronce	Microlotes cada 15 minutos	Archivo inmutable particionado por fecha
3. Cargas complementarias	Azure Data Factory	Diaria y semanal	Maestros, precios, promociones y fuentes externas actualizados
4. Calidad y conformación	Spark, zona plata	Diaria, ventana nocturna	Serie ítem–tienda–día validada
5. Ingeniería de variables	Spark, zona oro	Diaria	Tabla de variables predictivas
6. Carga al datamart	Azure Synapse Analytics	Diaria, incremental	Hechos de venta y dimensiones actualizados
7. Entrenamiento y ejecución	Plataforma element	Entrenamiento semanal, ejecución diaria	Pronóstico por ítem–tienda–día con horizonte de doce semanas
8. Publicación y consumo	Synapse, Power BI y APIs de reposición	Diaria, disponible a las 06:00	Tablero de demanda y órdenes sugeridas
9. Retroalimentación	Pronóstico contra venta real	Semanal	Métricas de error por categoría y región para reentrenamiento

Nota. Elaboración propia.

2.4.5 Gobierno y seguridad del flujo

El recorrido atraviesa datos transaccionales, de comportamiento de clientes y de terceros, por lo que el gobierno no puede ser una etapa final sino una franja transversal. Azure Purview mantiene el catálogo corporativo, la clasificación de la información sensible y el linaje extremo a extremo, de manera que ante cualquier cifra de un tablero sea posible reconstruir su origen y las transformaciones aplicadas; sobre esa base operan el control de acceso basado en roles, el cifrado y el enmascaramiento de atributos identificables. La condición multinacional de Walmart obliga además a políticas que satisfagan simultáneamente los marcos de California, la Unión Europea, México y Canadá, lo que se traduce en reglas de residencia por región, seudonimización temprana y retención diferenciada por tipo de dato. Un riesgo técnico merece mención: buena parte de los datos en flujo continuo viaja con el esquema definido en el

código y codificado en JSON, por lo que el flujo incorpora validación de esquema en la escritura, un registro de esquemas corporativo y contratos de datos con los equipos productores.

CAPÍTULO 3. Framework de procesamiento y ecosistema tecnológico Big Data

Este capítulo completa la arquitectura con la selección del framework de procesamiento distribuido y de los elementos complementarios del ecosistema: lenguaje de programación, almacenamiento, integración, Machine Learning y visualización. La elección de Apache Spark se apoya en su modelo de procesamiento en memoria, que permite combinar cargas por lotes, consultas y aprendizaje automático sobre un mismo motor (Zaharia et al., 2012), frente a alternativas como Hadoop MapReduce o Apache Flink (Kleppmann, 2017; White, 2015).

Debe recordarse la distinción planteada en la Introducción: la arquitectura objetivo contempla ingesta continua con Apache Kafka, mientras que el prototipo académico trabaja con datos históricos de dominio público. Por ello, en este capítulo Flink se descarta como framework principal del prototipo, sin perjuicio de que el procesamiento en tiempo real se retome en el Capítulo 5 como evolución de la arquitectura.

3.1 Análisis y selección crítica de frameworks

Para desarrollar el proyecto de optimización de ventas y demanda de Walmart es necesario seleccionar frameworks que permitan trabajar con grandes volúmenes de datos, realizar procesos de transformación y desarrollar modelos de análisis y predicción. La selección debe considerar no solamente la capacidad técnica de cada framework, sino también su compatibilidad con la arquitectura de Microsoft Azure, el tipo de información que utilizaremos y la facilidad de integración con Power BI.

Entre las principales alternativas que pueden considerarse dentro de un ecosistema Big Data se encuentran **Apache Hadoop, Apache Spark y Apache Flink**. Cada uno presenta características diferentes y, por tanto, su utilidad depende del tipo de procesamiento que se quiera realizar.

3.1.1 Apache Hadoop

Apache Hadoop es uno de los frameworks más conocidos del ecosistema Big Data. Su arquitectura fue diseñada para trabajar con grandes cantidades de información distribuida entre diferentes nodos. Una de sus principales características es **HDFS (Hadoop Distributed File System)**, utilizado para almacenar grandes volúmenes de datos de manera distribuida, junto con **MapReduce**, utilizado tradicionalmente para procesar dichos datos.

Hadoop puede ser útil cuando una organización necesita construir una infraestructura distribuida para almacenar y procesar grandes cantidades de información. Sin embargo, para nuestro proyecto presenta una desventaja importante: **su implementación y administración puede resultar más compleja de lo necesario** para la solución que estamos planteando.

Nuestro proyecto no pretende construir desde cero una infraestructura Big Data independiente. Ya hemos seleccionado Microsoft Azure como proveedor Cloud y PaaS como modelo de servicio. Por ello, utilizar Hadoop como componente central implicaría incorporar una capa tecnológica adicional que podría aumentar la complejidad de la arquitectura.

Por esta razón, aunque Hadoop es una alternativa válida dentro del ecosistema Big Data, **no será nuestro framework principal**.

3.1.2 Apache Spark

Apache Spark resulta especialmente interesante para nuestro proyecto porque está diseñado para el procesamiento distribuido de grandes volúmenes de datos y permite realizar diferentes tipos de análisis sobre una misma plataforma.

Spark permite trabajar con procesamiento por lotes y también con procesamiento de datos en tiempo casi real. Además, dispone de componentes orientados a diferentes necesidades analíticas, entre ellos **Spark SQL** para consultas estructuradas y **MLlib** para Machine Learning.

Esta última característica es especialmente relevante para nuestro proyecto porque el objetivo no es únicamente almacenar o consultar información. Queremos utilizar los datos históricos y las variables disponibles para **identificar patrones y desarrollar modelos de predicción de demanda**.

Por ejemplo, Spark podría utilizarse para procesar grandes cantidades de registros históricos de ventas y preparar las variables que posteriormente serán utilizadas por los modelos predictivos. Esto resulta coherente con nuestro objetivo de utilizar Big Data para pasar de un análisis descriptivo hacia un enfoque predictivo.

Además, Spark puede integrarse con servicios cloud y puede utilizarse dentro de ecosistemas de Azure orientados al procesamiento y análisis de datos.

Por estas razones, **Apache Spark será el framework seleccionado para nuestro proyecto**.

3.1.3 Apache Flink

Apache Flink es otro framework de procesamiento distribuido que destaca especialmente por sus capacidades de procesamiento de datos en tiempo real y por eventos.

Su utilización sería especialmente interesante en escenarios donde fuera necesario analizar continuamente información que llega de manera constante, por ejemplo, transacciones en tiempo real, sensores, eventos de aplicaciones o flujos de información.

Sin embargo, nuestro proyecto académico estará basado principalmente en **datos históricos y fuentes de dominio público**, y el objetivo principal será desarrollar un modelo de predicción de demanda. No estamos construyendo

actualmente una plataforma que necesite procesar millones de transacciones de Walmart en tiempo real.

Por lo tanto, aunque Flink podría ser una alternativa para una futura evolución de la arquitectura, **no será nuestra selección principal**.

3.1.4 Selección final: Apache Spark

Después de comparar las alternativas, seleccionamos **Apache Spark** como framework principal de procesamiento Big Data.

La decisión se fundamenta principalmente en cuatro aspectos:

- **Capacidad de procesamiento:** permite trabajar con grandes volúmenes de información de manera distribuida.
- **Analítica avanzada:** incorpora herramientas que facilitan el procesamiento y preparación de datos para modelos de Machine Learning.
- **Flexibilidad:** permite trabajar con datos estructurados y diferentes tipos de información, algo importante considerando que nuestro proyecto combinará información proveniente de distintas fuentes públicas.
- **Compatibilidad con nuestra arquitectura:** puede integrarse con entornos cloud y complementar los servicios que utilizaremos dentro de Microsoft Azure.

Por lo tanto, **Spark se adapta mejor a nuestra propuesta que Hadoop o Flink**, principalmente porque combina procesamiento de datos y capacidades analíticas que son directamente relevantes para nuestro objetivo de predicción de demanda.

3.2 Elementos complementarios del ecosistema Big Data

La selección de Apache Spark por sí sola no constituye un ecosistema Big Data completo. Para desarrollar nuestra propuesta necesitamos incorporar otros elementos que permitan capturar, almacenar, transformar, analizar y visualizar la información.

En este sentido, la arquitectura debe integrar diferentes tecnologías que cumplan funciones específicas.

3.2.1 Lenguaje de programación: Python

El lenguaje principal seleccionado será **Python**.

La elección de Python se debe a que es ampliamente utilizado en análisis de datos, estadística, Machine Learning y procesamiento de información. Además, permite trabajar con diferentes bibliotecas y herramientas que pueden complementar el funcionamiento de Spark.

En nuestro proyecto, Python podría utilizarse para:

- limpieza y preparación de datos
- análisis exploratorio

- transformación de variables
- desarrollo de modelos predictivos
- evaluación de modelos
- automatización de determinados procesos
- interacción con herramientas de Machine Learning.

Además, **PySpark** permite utilizar Apache Spark mediante Python, lo que facilita la integración entre el procesamiento distribuido de datos y los modelos analíticos.

Esto resulta conveniente porque podemos mantener un lenguaje común para diferentes etapas del proyecto.

3.2.2 Almacenamiento: Data Lake

Como parte de nuestra arquitectura utilizaremos el concepto de **Data Lake** para almacenar información proveniente de diferentes fuentes.

Esto es especialmente importante porque las fuentes públicas que utilizaremos pueden presentar diferentes estructuras y formatos.

Por ejemplo, podemos encontrar:

- archivos CSV
- archivos Excel
- información publicada en PDF
- datos estructurados
- información semiestructurada
- conjuntos de datos históricos.

El Data Lake permitirá conservar la información antes y después de determinados procesos de transformación, facilitando posteriormente su utilización para análisis.

Como hemos seleccionado **Microsoft Azure**, podemos plantear **Azure Data Lake Storage** como el componente de almacenamiento dentro de nuestro ecosistema cloud.

De esta manera, el almacenamiento estaría integrado con el resto de servicios de Azure.

3.2.3 Procesamiento: Apache Spark

Apache Spark será el componente encargado del procesamiento Big Data.

Una vez que los datos estén disponibles en el entorno de almacenamiento, Spark podrá utilizarse para realizar operaciones como:

- limpieza
- transformación
- integración

- agregación
- análisis
- preparación de variables
- procesamiento de grandes volúmenes de información.

Por ejemplo, si contamos con información histórica de ventas y variables externas, Spark puede utilizarse para integrar ambas fuentes y generar un conjunto de datos preparado para el análisis predictivo.

3.2.4 Machine Learning

El proyecto también requiere un componente de **Machine Learning**, debido a que nuestro objetivo es predecir la demanda.

Aquí podemos combinar **Apache Spark MLlib** con los servicios de Machine Learning disponibles en Azure.

El modelo podría utilizar información histórica para identificar patrones relacionados con la demanda.

Entre las variables potenciales podrían encontrarse:

- período
- ventas históricas
- categoría de producto
- comportamiento de la demanda
- estacionalidad
- variables económicas
- variables climáticas
- ubicación geográfica
- tendencias de consumo.

El resultado sería una estimación de la demanda futura que posteriormente podrá ser utilizada dentro de los dashboards.

3.2.5 Integración de datos: Azure Data Factory

Otro componente importante será **Azure Data Factory**, que puede utilizarse para organizar los procesos de integración y movimiento de datos.

Esto resulta útil porque nuestro proyecto utilizará diferentes fuentes públicas.

El proceso comienza recopilando la información desde las fuentes seleccionadas. Posteriormente, los datos pueden ser integrados y preparados para que tengan una estructura consistente. Una vez transformados, pueden almacenarse en el entorno correspondiente y quedar disponibles para Spark y los procesos analíticos.

Esto permite establecer una metodología organizada para las cargas de información y evita que cada fuente tenga que procesarse manualmente de manera independiente.

3.2.6 Base de datos y almacenamiento analítico

Además del Data Lake, nuestra arquitectura puede incorporar un repositorio estructurado para los datos que ya han sido procesados y preparados para consulta.

En este componente podemos considerar servicios de Azure orientados al almacenamiento y análisis de datos, dependiendo del nivel de implementación que adoptemos.

La diferencia conceptual sería que el **Data Lake** conservaría información proveniente de diferentes fuentes y formatos, mientras que el entorno analítico contendría información preparada para realizar consultas, análisis y visualizaciones.

Esto resulta importante para Power BI, ya que los dashboards deben trabajar con información organizada y estructurada.

3.2.7 Power BI

Finalmente, **Power BI** será la herramienta de Business Intelligence seleccionada para la presentación de los resultados.

Power BI permitirá convertir los resultados obtenidos mediante los procesos de Big Data y Machine Learning en información visual para los usuarios.

Los dashboards podrían mostrar indicadores como:

- evolución de ventas
- demanda histórica
- demanda proyectada
- crecimiento o disminución de determinadas categorías
- productos con mayor demanda
- productos con posible riesgo de sobrestock
- productos con posible riesgo de desabastecimiento
- comportamiento por período
- precisión del modelo predictivo.

De esta manera, Power BI representa la capa de visualización y consumo de información del proyecto.

3.2.8 Ecosistema tecnológico seleccionado

Tomando todas las decisiones anteriores, nuestro ecosistema puede quedar estructurado de la siguiente manera:

Tabla 22. *Ecosistema tecnológico seleccionado*

Componente	Tecnología seleccionada	Función dentro del proyecto
Proveedor Cloud	Microsoft Azure	Proporcionar el entorno cloud

Componente	Tecnología seleccionada	Función dentro del proyecto
Modelo Cloud	PaaS	Facilitar el desarrollo y ejecución de los servicios
Framework Big Data	Apache Spark	Procesamiento distribuido de datos
Lenguaje	Python	Análisis, transformación y Machine Learning
Procesamiento con Python	PySpark	Utilización de Spark mediante Python
Almacenamiento	Azure Data Lake Storage	Almacenamiento de información
Integración/ETL	Azure Data Factory	Integración y preparación de datos
Machine Learning	Spark MLlib / Azure Machine Learning	Desarrollo de modelos predictivos
Business Intelligence	Power BI	Visualización y análisis de resultados

Nota. Elaboración propia.

3.2.9 Síntesis de la selección

Para nuestro proyecto de Walmart seleccionaremos **Apache Spark como framework principal de Big Data**, debido a que proporciona capacidades de procesamiento distribuido y herramientas que pueden utilizarse en el análisis y Machine Learning.

Lo complementaremos con **Python como lenguaje de programación**, principalmente por su utilidad para análisis de datos y desarrollo de modelos predictivos. Para el almacenamiento utilizaremos el concepto de **Data Lake**, implementado dentro del ecosistema de Microsoft Azure mediante Azure Data Lake Storage. **Azure Data Factory** permitirá organizar los procesos de integración y preparación de las fuentes de información, mientras que los modelos predictivos podrán desarrollarse utilizando Spark MLlib y los servicios de Machine Learning de Azure.

Finalmente, **Power BI** será la herramienta encargada de presentar los resultados mediante dashboards e indicadores.

Esta combinación mantiene coherencia con todas las decisiones tomadas anteriormente. **Microsoft Azure proporciona el entorno Cloud y PaaS constituye el modelo de servicio; Spark aporta las capacidades de procesamiento Big Data; Python facilita el análisis y Machine Learning; el Data Lake permite gestionar las diferentes fuentes de información; y Power BI convierte los resultados en información útil para la toma de decisiones.**

Lo más importante es que el ecosistema está alineado con nuestro objetivo inicial: **utilizar información pública para construir una solución de análisis y predicción de demanda que permita identificar patrones de ventas y apoyar una mejor gestión de inventarios, reduciendo el riesgo de sobrestock y desabastecimiento en Walmart.**

CAPÍTULO 4. Data Science y Machine Learning: modelo predictivo y caso de negocio

Este capítulo es el núcleo analítico del proyecto. Define el modelo de Data Science y Machine Learning para la predicción de la demanda y el riesgo de inventario, las fuentes de datos que lo alimentan, el área de aplicación y los departamentos implicados, la mejora esperada y, finalmente, su cuantificación mediante un caso de negocio. Se sigue así la secuencia de cinco pasos propuesta en el material del módulo para implementar inteligencia artificial en la empresa (Pérez Galán, 2025), en línea con el enfoque de casos de uso acotados y medibles que recomiendan Davenport y Ronanki (2018).

El caso de negocio de este capítulo, con un ROI de 36,1 % en el primer año y una recuperación de la inversión en unos ocho meses, constituye la referencia económica principal del proyecto. Se construye sobre las cifras públicas del segmento Walmart U.S. del año fiscal 2026 (Walmart Inc., 2026) e incorpora la inversión inicial de implementación.

4.1 Identificación y definición del objetivo de aplicación del modelo de Data Science y Machine Learning

4.1.1 Contexto de aplicación

El proyecto de Business Intelligence y Big Data desarrollado para Walmart tiene como propósito central mejorar la predicción de la demanda y optimizar la gestión de inventarios, reduciendo situaciones de quiebre de stock y sobrestock.

Durante las etapas anteriores del proyecto se diseñó un ecosistema capaz de integrar, almacenar, procesar y visualizar información del negocio. La incorporación de Data Science y Machine Learning representa la evolución de esta propuesta hacia una capacidad predictiva, utilizando los datos disponibles para identificar patrones y anticipar comportamientos futuros.

El modelo permitirá pasar de analizar únicamente qué ocurrió con las ventas y los inventarios, a estimar qué podría ocurrir con la demanda futura y determinar si los niveles de inventario serán suficientes para atenderla.

4.1.2 Modelo propuesto

Se propone implementar un:

Modelo Predictivo de Demanda y Riesgo de Inventario

El modelo utilizará técnicas de Data Science y Machine Learning para analizar patrones históricos y relaciones entre las variables que influyen en el comportamiento de la demanda, con el propósito de estimar las unidades requeridas para cada combinación de:

Producto/SKU + Tienda + Período

Posteriormente, la demanda pronosticada se comparará con el inventario disponible y proyectado, permitiendo identificar anticipadamente tres posibles escenarios:

Tabla 23. ***Escenarios de inventario que anticipa el modelo***

Escenario	Interpretación	Acción potencial
Riesgo de Stock-Out	Inventario insuficiente frente a la demanda esperada.	Reponer, comprar o redistribuir inventario.
Inventario adecuado	Existencias suficientes para atender la demanda proyectada.	Mantener planificación.
Riesgo de Overstock	Inventario superior a las necesidades proyectadas.	Reducir reposición o redistribuir existencias.

Nota. Elaboración propia.

De esta manera, el modelo no se limitará a generar un pronóstico de ventas, sino que relacionará la **demanda esperada con la disponibilidad de inventario** para generar información accionable.

4.1.3 Objetivo general

Desarrollar un modelo predictivo basado en Data Science y Machine Learning que permita estimar la demanda futura por producto/SKU, tienda y período, identificando anticipadamente posibles riesgos de stock-out o sobrestock, con el propósito de apoyar la planificación de inventarios y las decisiones de reposición, compras y distribución de Walmart.

4.1.4 Objetivos específicos

Para alcanzar el objetivo general, el modelo buscará:

- Identificar patrones históricos de demanda, considerando el comportamiento de ventas por producto, tienda y período.
- Reconocer variables que influyen en las variaciones de la demanda, permitiendo enriquecer el análisis más allá del comportamiento histórico.
- Generar pronósticos de demanda por SKU, tienda y período.
- Comparar la demanda proyectada con el inventario disponible y esperado, identificando posibles desequilibrios.
- Detectar anticipadamente riesgos de stock-out y sobrestock, antes de que afecten la operación.
- Generar alertas y recomendaciones que apoyen las decisiones de compra, reposición y redistribución de inventario.
- Evaluar continuamente el desempeño del modelo, comparando las predicciones con la demanda real para mejorar progresivamente su precisión.

4.1.5 Problema que busca resolver el modelo

La información histórica permite conocer situaciones que ya ocurrieron:

- ¿Cuánto se vendió?
- ¿Qué productos tuvieron mayor demanda?
- ¿Dónde existió un quiebre de inventario?

Sin embargo, el modelo propuesto busca anticiparse y responder preguntas como:

- ¿Cuánto se espera vender?
- ¿En qué tienda y período se producirá esa demanda?
- ¿El inventario disponible será suficiente?
- ¿Existe riesgo de stock-out o sobrestock?
- ¿Es necesario comprar, reponer o redistribuir inventario?

Por tanto, la aplicación de Machine Learning permitirá evolucionar desde una gestión principalmente descriptiva y reactiva hacia una gestión predictiva y anticipativa.

4.1.6 Funcionamiento del modelo

La lógica general propuesta es:

Figura 17. *Funcionamiento del modelo predictivo de demanda y riesgo de inventario*



Nota. Elaboración propia.

4.1.7 Resultado esperado del modelo

El resultado esperado no será únicamente obtener una cifra de ventas futuras. El propósito es convertir la predicción en información útil para la gestión de inventarios:

Forecast de demanda + Riesgo de inventario + Alerta + Recomendación

Los resultados del modelo podrán posteriormente integrarse en **Power BI**, herramienta seleccionada previamente dentro del proyecto, para visualizar pronósticos, indicadores, alertas y recomendaciones que apoyen la toma de decisiones.

Así, el proceso evoluciona de:

Situación tradicional

Datos históricos → Análisis → Identificación del problema → Reacción

a:

Modelo propuesto

Datos → Machine Learning → Predicción → Alerta → Recomendación → Decisión anticipada

4.2 Fuentes de datos utilizadas

Las fuentes de datos de nuestro modelo estarán divididas principalmente en fuentes internas y externas desde el punto de vista de la arquitectura empresarial, pero para la implementación real de nuestro proyecto académico utilizaremos únicamente aquellas fuentes que sean de dominio público.

Esta distinción es importante.

En una implementación real de Walmart, el modelo podría utilizar información interna proveniente de sistemas de ventas, inventarios, compras, logística y comercio electrónico. Sin embargo, nosotros no disponemos de esas bases internas, por lo que no debemos presentarlas como si fueran datos disponibles para nuestro análisis.

4.2.1 Fuentes públicas relacionadas con Walmart

Podemos utilizar:

- informes anuales de Walmart;
- reportes financieros;
- información pública sobre ventas;
- información pública sobre comercio electrónico;
- información corporativa;
- información sobre segmentos de negocio;
- reportes de resultados;
- información pública sobre operaciones.

Estas fuentes permiten contextualizar el comportamiento general del negocio.

4.2.2 Fuentes externas

También podemos incorporar información externa que pueda ayudar a explicar variaciones en la demanda.

Entre ellas:

- **Datos económicos:** inflación, tasas de crecimiento, indicadores de consumo y otros indicadores macroeconómicos.
- **Datos climáticos:** temperatura, precipitaciones, eventos climáticos y otras variables que puedan afectar el comportamiento de determinadas categorías de productos.

- **Datos temporales:** días de la semana, meses, temporadas, feriados y períodos especiales.
- **Datos geográficos:** ubicación y características de las regiones en las que se analiza la demanda, siempre que dicha información esté disponible públicamente.
- **Datos de consumo:** estadísticas públicas relacionadas con tendencias de consumo y comportamiento del mercado.

La combinación de estas fuentes permitiría que el modelo no dependa exclusivamente del comportamiento histórico de las ventas. Por ejemplo, si una categoría presenta una determinada estacionalidad, el modelo podría aprender que la demanda cambia durante determinados períodos del año.

La lógica es similar a la planteada en el material del módulo, donde se destaca la utilización de diferentes fuentes y variables para construir una base que permita posteriormente realizar análisis y modelos predictivos.

4.3 Área de aplicación y departamentos de negocio implicados

El modelo de pronóstico de demanda definido en la sección 4.1 es un modelo de aprendizaje supervisado de regresión: aprende de la historia de ventas, en la que el resultado (unidades vendidas) es conocido, para estimar una variable continua, la demanda futura de cada artículo en cada punto de venta (Pérez Galán, s. f.-b). Definir dónde se aplica y quiénes participan es tan importante como el algoritmo, porque un pronóstico solo genera valor cuando se convierte en una decisión: cuánto comprar, cuánto mover al centro de distribución y cuánto exhibir en la góndola. En el caso de referencia, el modelo de propensión se aplicaba en el área de cobranza y corría en cada bucket de mora (Pérez Galán, s. f.-a); en Walmart, el modelo se aplica en la cadena de suministro y corre en cada ciclo de planificación sobre segmentos de artículos.

4.3.1 Área de aplicación principal: planificación de la demanda y reabastecimiento

El área de aplicación es la cadena de suministro de Walmart, específicamente el proceso de planificación de la demanda y reabastecimiento (Demand Planning & Replenishment). Es el punto de la organización donde se une lo que el cliente compra con lo que Walmart pide a sus proveedores y distribuye a través de su red de centros de distribución (CD), tiendas, clubes y canales digitales. Se eligió este proceso porque se ubica aguas arriba de casi todas las decisiones operativas: un error de pronóstico se propaga en cadena hacia órdenes de compra mal dimensionadas, capacidad ociosa o saturada en los CD, quiebres de stock en tienda y merma en perecibles. Además, es un frente en el que la compañía declara inversiones crecientes en tecnología, inteligencia artificial, automatización y cadena de suministro (Walmart Inc., 2025a).

Walmart no parte de cero. Walmart Global Tech opera una plataforma de pronóstico y reabastecimiento basada en machine learning que trabaja a nivel de combinación artículo-tienda (Walmart Inc., s. f.), y la compañía cuenta hoy

con herramientas como Self-Healing Inventory, que detecta desbalances de stock y redirige producto de forma automática, y Enterprise Inventory, que ofrece una vista unificada del inventario en tiendas, centros de fulfillment y canal online (Walmart Inc., 2025b). El departamento de analítica propuesto no busca sustituir esas capacidades, sino actuar como dueño técnico del modelo dentro de la arquitectura definida en los capítulos anteriores (ingesta con Azure Data Factory y Apache Kafka, almacenamiento en Azure Data Lake Storage, modelado en Azure Synapse Analytics, entrenamiento en Element y visualización en Power BI) y garantizar que su salida llegue de forma gobernada a cada área usuaria. La Tabla 3.1 delimita el alcance de aplicación.

Tabla 24. *Alcance de aplicación del modelo de pronóstico de demanda en Walmart*

Dimensión	Definición para Walmart
Unidad de negocio	Fase piloto en Walmart U.S.; expansión posterior a Sam's Club y Walmart International (México, Canadá, Centroamérica y Chile), aprovechando que la compañía ya escala plataformas comunes de cadena de suministro entre mercados (Walmart Inc., 2025b).
Canales	Tiendas, clubes, pickup, delivery y walmart.com, pronosticados de forma conjunta para capturar la sustitución de demanda entre canales.
Granularidad	Artículo (SKU) × tienda o club × semana para compras y asignación; SKU × tienda × día para perecibles; agregación por CD para la planificación logística.
Horizontes	Corto plazo (1 a 14 días): reabastecimiento y asignación. Mediano plazo (4 a 13 semanas): compras, promociones y capacidad de CD. Largo plazo (6 a 12 meses): temporadas (Back to School, Black Friday, Holiday) y presupuesto de inventario.
Categorías del piloto	Perecibles (frutas, verduras, lácteos y panadería), consumibles de alta rotación y mercancía estacional, por su alta sensibilidad al clima, las promociones y el calendario.
Frecuencia de ejecución	Generación diaria del pronóstico; reentrenamiento programado semanal y reentrenamiento extraordinario ante caída de la precisión o eventos atípicos.

Nota. Elaboración propia.

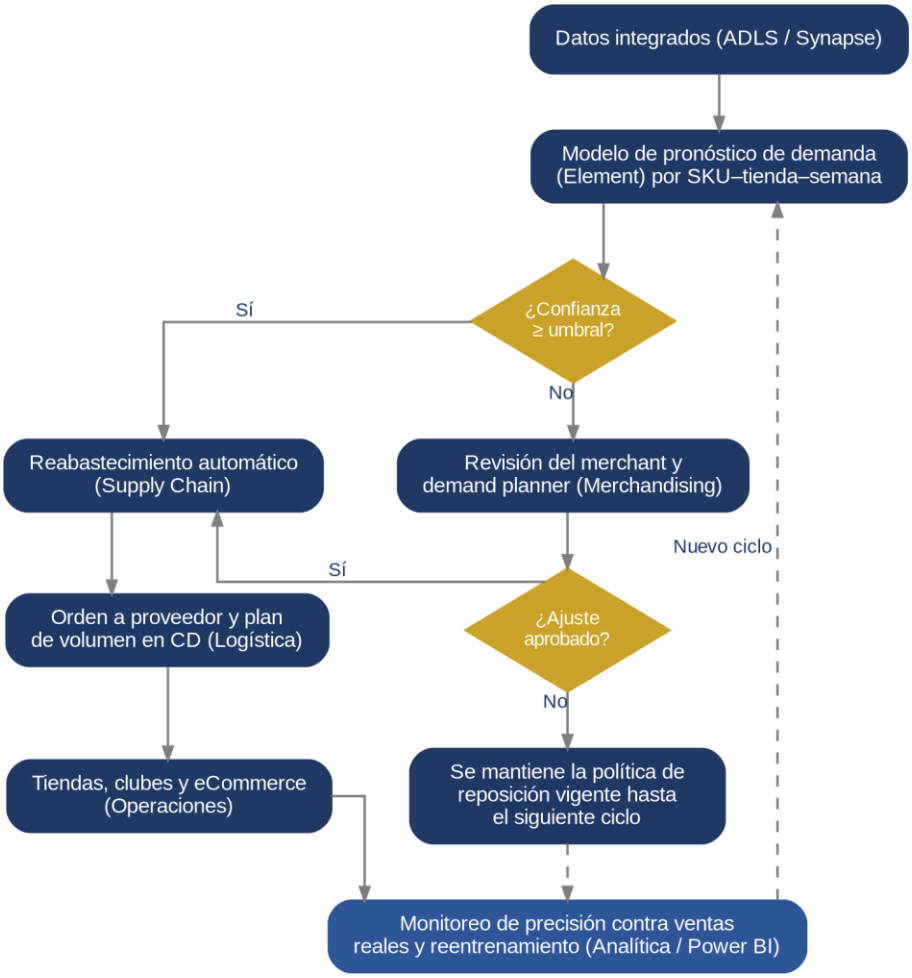
4.3.2 Lógica de aplicación del modelo por segmentos

Al igual que el modelo de referencia, que se ejecutaba en cada bucket de mora y trasladaba al siguiente ciclo a los clientes que aún no cumplían el perfil, el modelo de Walmart se ejecuta en cada ciclo de planificación sobre segmentos de artículos definidos por rotación, variabilidad y ciclo de vida. Estos segmentos pueden construirse con técnicas de agrupamiento, es decir, aprendizaje no supervisado aplicado a variables sin etiqueta previa (Pérez Galán, s. f.-b).

La regla de decisión central es la confianza del pronóstico. Cuando el intervalo de predicción es estrecho y el error histórico del segmento está dentro del umbral acordado con Supply Chain, la recomendación pasa directamente al reabastecimiento automático. Cuando no, se deriva al merchant y al demand

planner de la categoría para su revisión. Los artículos cuyo ajuste no se aprueba mantienen la política de reposición vigente y se vuelven a evaluar en el ciclo siguiente, ya con la venta real incorporada al entrenamiento. La Ilustración 3.1 resume el flujo y el área que interviene en cada paso.

Figura 18. ***Flujo de aplicación del modelo de pronóstico de demanda y áreas que intervienen***



Nota. Elaboración propia.

La Tabla 3.2 muestra cómo se trata cada segmento y qué área toma la decisión final sobre el pronóstico.

Tabla 25. ***Tratamiento del pronóstico por segmento de artículos***

Segmento	Características	Tratamiento del pronóstico	Área que decide
Alta rotación y demanda estable	Historia larga y baja variabilidad (abarrotes básicos, cuidado del hogar).	Reabastecimiento automático; revisión humana solo por excepción.	Supply Chain

Segmento	Características	Tratamiento del pronóstico	Área que decide
Perecibles	Vida útil corta; alta sensibilidad al clima y a la frescura.	Pronóstico diario por tienda y ajuste de pedidos a los CD de perecibles.	Supply Chain y Operaciones de tiendas
Promocionados y estacionales	Picos por precio, publicidad o calendario.	Pronóstico con variables de precio y calendario promocional; validación obligatoria del merchant.	Merchandising, con Marketing y Pricing
Nuevos productos y baja rotación	Sin historia o con ventas intermitentes.	Pronóstico por similitud con artículos análogos; revisión hasta acumular historia suficiente.	Merchandising
Eventos disruptivos	Picos no recurrentes (tormentas, huracanes, olas de calor).	Ajuste con datos climáticos externos y reasignación de inventario entre regiones.	Supply Chain, Logística y Operaciones

Nota. Elaboración propia.

4.3.3 Departamentos de negocio implicados

La aplicación del modelo involucra a diez áreas que cumplen seis roles: dueño del proceso, dueño técnico del modelo, usuario de la salida, proveedor de insumos de negocio, habilitador tecnológico y control. Esta clasificación es coherente con el organigrama y las dependencias interdepartamentales descritas en el Capítulo 5 de la Parte I, y se detalla en la Tabla 3.3.

Tabla 26. ***Departamentos implicados y su rol frente al modelo***

Departamento	Rol	Participación y decisiones que soporta
Supply Chain: planificación de la demanda y reabastecimiento	Dueño del proceso	Define umbrales de confianza y niveles de servicio; convierte el pronóstico en órdenes de reposición y asignación a tiendas y CD. Es el principal usuario de la salida del modelo.
Departamento de Analítica (propuesto)	Dueño técnico del modelo	Construye, valida y mantiene el modelo: preprocesamiento, validación cruzada, ajuste de hiperparámetros, selección y monitoreo de la precisión (Pérez Galán, s. f.-b).
Merchandising (merchants por categoría)	Usuario y proveedor de insumos	Decide compras, surtido y volúmenes con proveedores; revisa los pronósticos de baja confianza, nuevos productos y promociones; aporta conocimiento de la categoría.
Operaciones de tiendas y clubes	Usuario	Recibe cantidades y tiempos de reposición; planifica el personal de recepción y exhibición; reporta quiebres y merma que retroalimentan el modelo.
Logística, centros de distribución y transporte	Usuario	Dimensiona capacidad, turnos y rutas con el volumen proyectado. El Walmart Fulfillment Engine asigna pedidos online sobre inventario previamente posicionado con base en pronósticos (Walmart Global Tech, 2025).

Departamento	Rol	Participación y decisiones que soporta
eCommerce y omnicanal	Usuario y proveedor de insumos	Aporta búsquedas, visitas a páginas y pedidos online como señales anticipadas de demanda, que Walmart ya utiliza para entrenar sus modelos de temporada (Walmart Global Tech, 2023); consume el pronóstico para posicionar inventario de pickup y delivery.
Marketing, Pricing y Walmart Connect	Proveedor de insumos	Entrega el calendario promocional, los cambios de precio y las campañas de medios, variables que explican buena parte de la variación de la demanda; usa el pronóstico para estimar el volumen esperado de cada promoción.
Finanzas (FP&A)	Usuario	Traduce el pronóstico en presupuesto de inventario, capital de trabajo y proyección de ventas; consume los tableros de Power BI, herramienta estándar del área.
Walmart Global Tech	Habilitador tecnológico	Provee la infraestructura en Azure y Element, los pipelines de datos y la integración con los sistemas de reposición y de gestión de almacenes y transporte.
Gobierno de datos, Privacidad y Legal	Control	Administra el catálogo y el linaje de datos en Azure Purview y valida que cada variable cumpla finalidad, minimización y consentimiento bajo CCPA/CPRA, LFPDPPP, PIPEDA y GDPR.

Nota. Elaboración propia.

Fuera de la organización, los proveedores son el actor externo más impactado: reciben las órdenes derivadas del pronóstico y, a través de Scintilla, la plataforma de inteligencia comercial de Walmart Data Ventures (antes Walmart Luminate), acceden a información de desempeño para planificar su producción y actuar sobre los datos de venta (Chain Store Age, 2025). Compartir el pronóstico por este canal alinea la capacidad del proveedor con la demanda esperada y reduce el efecto látigo en toda la cadena.

4.3.4 Matriz de responsabilidades (RACI)

Para evitar vacíos o duplicidades entre áreas, la Tabla 3.4 asigna las responsabilidades a lo largo del ciclo de vida del modelo. Cada actividad tiene un único responsable final de la aprobación (A).

Tabla 27. *Matriz RACI del modelo de pronóstico de demanda*

Actividad	AN	SC	MER	OPS	LOG	ECO	MKT	FIN	GT	GOB
Definir requerimientos, segmentos y niveles de servicio	R	A	C	C	C	C	I	C	I	I
Ingesta e integración de datos (ADF, Kafka, ADLS)	C	I	I	I	I	C	C	I	A/R	C
Preparación de variables y control de privacidad	R	I	I	I	I	C	C	I	C	A
Entrenamiento, validación cruzada y selección del modelo (Element)	A/R	C	I	I	I	I	I	I	C	I

Actividad	AN	SC	MER	OPS	LOG	ECO	MKT	FIN	GT	GOB
Aprobación del paso a producción	R	A	C	I	I	I	I	I	C	C
Generación diaria del pronóstico y reabastecimiento automático	C	A	I	I	I	I	I	I	R	I
Revisión de pronósticos de baja confianza y promociones	C	C	A/R	I	I	C	C	I	I	I
Ejecución logística y en tienda	I	C	I	R	A/R	C	I	I	I	I
Monitoreo de precisión y reentrenamiento	A/R	C	I	I	I	I	I	I	C	I
Reporte de impacto financiero en Power BI	C	C	I	I	I	I	I	A/R	I	I

Nota. Elaboración propia.

R: responsable de ejecutar; A: aprueba y rinde cuentas; C: consultado; I: informado. AN: Analítica; SC: Supply Chain; MER: Merchandising; OPS: Operaciones de tiendas y clubes; LOG: Logística y CD; ECO: eCommerce; MKT: Marketing, Pricing y Walmart Connect; FIN: Finanzas; GT: Walmart Global Tech; GOB: Gobierno de datos, Privacidad y Legal.

4.3.5 Consideraciones éticas y de privacidad en el área de aplicación

El pronóstico de demanda trabaja principalmente con datos agregados (ventas por artículo y tienda, inventario, precios y clima) que no identifican a personas. Sin embargo, las señales de eCommerce y del programa de membresía Walmart+ sí pueden contener datos personales. La lectura del curso recuerda que los datos deben usarse para la finalidad con la que fueron recogidos, con información inequívoca al usuario y con posibilidad de acceso, rectificación y eliminación, a lo que el RGPD añadió los derechos al olvido y a la portabilidad (Aspectos legales y éticos del big data, s. f.). Por ello, Gobierno de datos, Privacidad y Legal participa desde el diseño: las variables de comportamiento digital ingresan al modelo anonimizadas y agregadas a nivel de artículo y zona, el acceso se controla y audita en Azure Purview, y cada fuente se revisa frente a CCPA/CPRA en Estados Unidos, LFPDPPP en México, PIPEDA en Canadá y GDPR donde aplique.

La segunda consideración es la transparencia y el control humano. La transparencia es una condición para que los usuarios de negocio confíen en el pronóstico y no traten el modelo como una caja negra (Shrestha et al., 2019). Por ello, en la propuesta cada pronóstico se entrega en Power BI junto con los factores que más lo explican (precio, promoción, clima y estacionalidad), el merchant conserva la decisión final sobre los casos de baja confianza y Supply Chain puede suspender el reabastecimiento automático de cualquier segmento, un mecanismo de intervención coherente con el principio de contar con un “interruptor de emergencia” en los sistemas automatizados que recoge la lectura del módulo (Pérez Galán, 2019).

4.3.6 Síntesis

El modelo se aplica en el proceso de planificación de la demanda y reabastecimiento de la cadena de suministro, con Supply Chain como dueño del proceso y el departamento de Analítica como dueño técnico. Su valor se materializa a través de las decisiones de Merchandising, Operaciones de tiendas, Logística, eCommerce y Finanzas, mientras que Marketing y Pricing aportan las variables que explican la demanda, Walmart Global Tech habilita la plataforma y Gobierno de datos asegura el uso legítimo de la información. Esta delimitación sienta la base para identificar la mejora esperada en cada área (sección 4.4) y cuantificarla en el caso de negocio (sección 4.5).

4.4 Identificación de la mejora esperada en la aplicación de inteligencia artificial

El documento de referencia sobre implementación de inteligencia artificial (Pérez Galán, 2025) plantea una estructura de cinco pasos para aplicar un modelo de ciencia de datos dentro de una empresa: definir el objetivo del modelo, identificar las fuentes de datos, delimitar el área de aplicación, identificar la mejora esperada y, finalmente, cuantificar esa mejora mediante un business case. En las secciones anteriores ya se abordaron los tres primeros pasos —el objetivo de la tecnología aplicada a inventario y reposición, las fuentes de datos internas y externas disponibles, y las áreas de negocio implicadas—, por lo que esta sección se concentra en el cuarto paso: la identificación de la mejora esperada.

4.4.1 Qué significa este punto en el modelo de Walmart

Así como el modelo descrito en el documento de referencia busca predecir qué clientes en mora aceptarán una alternativa de negociación para anticipar la gestión de cobranza (Pérez Galán, 2025), el modelo análogo que utiliza Walmart en su operación busca anticipar qué productos tienen probabilidad de agotarse en anaquel o de necesitar reposición, cruzando datos históricos de ventas, inventario y sensores de punto de venta. Identificar la mejora esperada significa, en ambos casos, definir de manera concreta qué cambio medible se espera obtener antes de construir el business case que lo cuantifique en dinero.

4.4.2 Mejoras esperadas identificadas para Walmart

- **Reducción de quiebres de inventario:** al anticipar la necesidad de reposición mediante el cruce de datos de RFID y ventas históricas, Walmart ha reportado una disminución del 16% en los quiebres de inventario desde la implementación de esta tecnología (Vector, 2026).
- **Mayor velocidad de reposición:** los productos etiquetados con RFID se reponen hasta tres veces más rápido que aquellos que dependen únicamente de código de barras, lo cual constituye una mejora esperada directamente atribuible al modelo de identificación anticipada (Vector, 2026).
- **Mejor precisión en el pronóstico de demanda:** el uso de inteligencia artificial y sensores IoT permite ajustar de manera dinámica los niveles de

inventario y anticipar variaciones de demanda antes de que se conviertan en desabastecimiento (LitCommerce, 2026).

- **Mayor especialización y priorización en los equipos operativos:** de forma similar a lo señalado en el documento de referencia para los equipos de cobranza (Pérez Galán, 2025), la segmentación predictiva de productos y tiendas permite que el personal de reposición concentre su esfuerzo en los puntos de mayor riesgo de quiebre, en lugar de revisar el inventario de manera uniforme.
- **Mejora en la experiencia del cliente:** la disponibilidad más consistente de producto en anaquel, resultado esperado de las secciones anteriores, se traduce en menos oportunidades de venta perdidas y una experiencia de compra más confiable (Ankashram, 2025).

4.4.3 De la mejora esperada al business case

Tal como lo plantea el documento de referencia, identificar la mejora esperada es el paso previo a cuantificarla económicamente. En el caso de Walmart, esa cuantificación ya cuenta con indicios documentados: la compañía ha invertido más de 20,000 millones de dólares en la última década en tecnología, inteligencia artificial y automatización (Poornaprajna International Journal of Management, Education & Social Science, 2025), mientras que sus ventas de comercio electrónico crecieron un 20,8 % en el año fiscal 2025, hasta alcanzar 121,000 millones de dólares (DCFmodeling, 2025). Estas cifras sugieren que las mejoras esperadas descritas anteriormente —menos quiebres de inventario, reposición más rápida y mejor pronóstico de demanda— efectivamente se han traducido en resultados financieros medibles, lo que validaría el business case que justificó la inversión inicial en estos modelos.

4.5 Caso de negocio: cuantificación de la mejora esperada

4.5.1 Objetivo del caso de negocio

A partir del modelo de Data Science y Machine Learning para la predicción de demanda y gestión del riesgo de inventario desarrollado en las secciones anteriores, se plantea un caso de negocio que permita estimar su impacto económico potencial para Walmart.

El objetivo es cuantificar cómo una mejora en la predicción de la demanda puede contribuir a:

- Reducir las pérdidas asociadas a productos agotados (stockouts).
- Recuperar ventas que potencialmente no se concretan por falta de inventario.
- Reducir el exceso de inventario.
- Liberar capital de trabajo.
- Disminuir los costos asociados al mantenimiento de inventarios.

- Evaluar la rentabilidad de implementar el modelo mediante indicadores como ROI y período de recuperación de la inversión.

Es importante diferenciar entre los datos reales publicados por Walmart y los supuestos utilizados para construir el escenario académico. Las cifras financieras de Walmart corresponden a información pública, mientras que variables como el porcentaje de ventas afectadas por *stockouts*, el alcance inicial del piloto, la reducción esperada del inventario y la inversión del proyecto son supuestos utilizados exclusivamente para elaborar el Business Case.

4.5.2 Base financiera utilizada

Para construir el escenario se utilizan como referencia los resultados financieros publicados por Walmart para el ejercicio fiscal 2026 (Walmart Inc., 2026).

Entre los principales datos utilizados se encuentran:

Tabla 28. *Base financiera utilizada (año fiscal 2026)*

Indicador	Valor
Ventas netas Walmart U.S.	US$482.975 millones
Margen bruto Walmart U.S.	27,50%
Inventario consolidado	US$58.851 millones

Nota. Elaboración propia con base en Walmart Inc. (2026). Ventas netas y margen bruto del segmento Walmart U.S.; inventario consolidado.

Estos valores permiten dimensionar económicamente el impacto potencial de aplicar el modelo predictivo dentro de Walmart.

Para efectos del Business Case se plantea inicialmente un piloto equivalente al 10 % de las ventas y del inventario de Walmart U.S. Esta proporción no representa una implementación real de Walmart, sino un supuesto académico que permite evaluar el proyecto de manera controlada antes de plantear una posible ampliación.

4.5.3 Supuestos del escenario

Para cuantificar los beneficios se establecen los siguientes supuestos:

Tabla 29. *Supuestos del escenario*

Variable	Supuesto
Cobertura inicial del piloto	10%
Ventas potencialmente afectadas por stockouts	0,50%
Reducción de stockouts	16%
Reducción esperada del inventario	1%
Costo anual de mantenimiento de inventario	12%
Inversión inicial	US$10 millones
Costo operativo anual	US$3 millones

Nota. Elaboración propia.

El 16 % de reducción de stockouts se toma como referencia de la sección 4.4, donde se presenta como un indicador asociado a las iniciativas de Walmart relacionadas con RFID y disponibilidad de productos.

El resto de los porcentajes constituyen supuestos del caso académico y deben interpretarse como variables que deberían validarse mediante un piloto real.

4.5.4 Estimación del beneficio por recuperación de ventas

El primer beneficio económico considerado corresponde a la recuperación de ventas que potencialmente se pierden cuando un cliente encuentra un producto agotado.

Paso 1. Ventas correspondientes al piloto

Las ventas de Walmart U.S. utilizadas como referencia son:

US$482.975 millones

Aplicando una cobertura inicial del 10 %:

$$\textbf{Ventas del piloto} = 482.975 \times 10\%$$

$$\textbf{Ventas del piloto} = \text{US\$48.297,5 millones}$$

Por lo tanto, el escenario considera aproximadamente US$48.297,5 millones de ventas dentro del alcance inicial del modelo.

Paso 2. Ventas potencialmente afectadas por stockouts

Se supone que el 0,5 % de las ventas del piloto podría estar relacionado con situaciones de falta de inventario:

$$48.297,5 \times 0,5\% = \text{US\$241,49 millones}$$

Esto representa aproximadamente US$241,49 millones de ventas potencialmente expuestas a pérdidas por falta de disponibilidad.

Paso 3. Ventas recuperadas

Si el modelo contribuye a reducir los *stockouts* en un 16 %:

$$241,49 \times 16\% = \text{US\$38,64 millones}$$

El escenario estima, por tanto, una recuperación potencial de aproximadamente US$38,64 millones en ventas anuales.

Sin embargo, no todo el valor de ventas recuperadas constituye beneficio económico directo. Para estimar el impacto sobre el resultado se utiliza el margen bruto de Walmart U.S.

Paso 4. Beneficio bruto generado

Aplicando el margen bruto del 27,5 %:

$$38,64 \times 27,5\% = \text{US\$10,63 millones}$$

El beneficio económico anual estimado por recuperación de ventas sería aproximadamente:

US$10,63 millones.

4.5.5 Beneficio asociado a la reducción del inventario

El segundo beneficio esperado corresponde a una utilización más eficiente del inventario.

Una mejor predicción de demanda permite reducir la probabilidad de mantener cantidades superiores a las necesarias, especialmente en productos con demanda variable, estacionalidad o cambios en el comportamiento del consumidor.

Paso 1. Inventario correspondiente al piloto

El inventario consolidado de Walmart utilizado como referencia es:

US$58.851 millones

Aplicando el 10 % correspondiente al alcance del piloto:

58.851×10%=US$5.885,1 millones

Por lo tanto, el inventario considerado dentro del escenario corresponde a aproximadamente US$5.885,1 millones.

Paso 2. Reducción potencial del inventario

Se plantea una reducción del 1 %:

5.885,1×1%=US$58,85 millones

Esto significa que el modelo podría contribuir potencialmente a liberar aproximadamente US$58,85 millones de capital asociado al inventario dentro del alcance considerado.

Es importante señalar que liberar capital no significa necesariamente obtener US$58,85 millones de utilidad. El beneficio económico se relaciona principalmente con la reducción de los costos de mantener dicho inventario.

Paso 3. Ahorro en costos de mantenimiento

Utilizando un costo anual de mantenimiento de inventario del 12 %:

58,85×12%=US$7,06 millones

Por lo tanto, el ahorro anual estimado por reducción de costos de inventario sería aproximadamente:

US$7,06 millones.

4.5.6 Beneficio económico total

Los dos principales beneficios cuantificados son:

- **Recuperación de ventas:** US$10,63 millones.
- **Ahorro por reducción de costos de inventario:** US$7,06 millones.

Por lo tanto:

Beneficio anual=10,63+7,06

Beneficio anual=US$17,69 millones

El modelo tendría, bajo los supuestos establecidos, un beneficio económico anual potencial de aproximadamente US$17,69 millones.

Este resultado representa un escenario académico y no una cifra reportada por Walmart.

4.5.7 Inversión y retorno esperado

Para implementar el piloto se establece una inversión inicial de:

US$10 millones

Adicionalmente, se considera un costo operativo anual de:

US$3 millones

Por lo tanto, durante el primer año el costo total sería:

10+3=US$13 millones

El beneficio económico anual estimado es de US$17,69 millones.

Por lo tanto, el resultado neto durante el primer año sería:

17,69−13=US$4,69 millones

ROI del primer año

ROI = Beneficio neto / Inversión total × 100

ROI = 4,69 / 13 × 100 ≈ 36,1 %

Bajo los supuestos definidos, el proyecto presenta un ROI estimado de 36,1 % durante el primer año.

Este porcentaje debe entenderse como el resultado del escenario construido para el trabajo académico y no como un ROI obtenido o publicado por Walmart.

4.5.8 Período de recuperación de la inversión

Para estimar el período de recuperación se considera la inversión inicial de US$10 millones y el beneficio anual después de costos operativos:

17,69−3=US$14,69 millones

Por lo tanto:

Payback = 10 / 14,69 ≈ 0,68 años

Expresado en meses:

0,68×12≈8,2 meses

El período estimado de recuperación de la inversión sería aproximadamente 8 meses.

Esto significa que, manteniéndose las condiciones del escenario, la inversión inicial podría recuperarse durante el primer año de operación.

4.5.9 Proyección simplificada a tres años

Considerando que la inversión inicial se realiza únicamente en el primer año y que el costo operativo anual permanece en US$3 millones, se obtiene la siguiente proyección:

Tabla 30. *Proyección simplificada a tres años*

Concepto	Año 1	Año 2	Año 3
Beneficio económico	17,69	17,69	17,69
OPEX	-3	-3	-3
Inversión inicial	-10	—	—
Resultado neto	4,69	14,69	14,69
Resultado acumulado	4,69	19,38	34,07

Nota. Elaboración propia. Valores en millones de US$.

Bajo este escenario, el proyecto podría generar aproximadamente US$34,07 millones de resultado neto acumulado durante tres años.

La proyección supone que los beneficios permanecen constantes, por lo que debe considerarse como una aproximación conservadora y simplificada. En una implementación real deberían incorporarse crecimiento de ventas, inflación, evolución del inventario, costos tecnológicos, depreciación, capacitación y posibles cambios en el comportamiento de los consumidores.

4.5.10 Análisis de sensibilidad

Debido a que algunas variables utilizadas son supuestos académicos, es necesario analizar cómo cambia el resultado ante diferentes niveles de ventas potencialmente perdidas por *stockouts*.

Se plantean tres escenarios:

Tabla 31. *Análisis de sensibilidad del beneficio anual*

Escenario	Ventas afectadas	Ventas recuperadas	Beneficio anual estimado
Conservador	0,25%	US$19,32 M	US$12,37 M
Base	0,50%	US$38,64 M	US$17,69 M
Alto	0,75%	US$57,96 M	US$23,00 M

Nota. Elaboración propia.

El escenario base utiliza el supuesto de 0,5 % de ventas potencialmente afectadas por falta de inventario.

La sensibilidad demuestra que el resultado económico depende significativamente de la magnitud real de las ventas que puedan recuperarse mediante una mejor disponibilidad de productos. Por esta razón, en una implementación real sería necesario medir el nivel actual de *stockouts* y establecer una línea base antes de desplegar el modelo.

4.5.11 Indicadores para evaluar el proyecto

Para comprobar si el modelo genera los resultados esperados, se proponen los siguientes KPI:

Tabla 32. *Indicadores para evaluar el proyecto*

KPI	Meta propuesta
Reducción de stockouts	≥ 16 %
Reducción de inventario	1%
Capital potencial liberado	US$58,85 M
Ventas potencialmente recuperadas	US$38,64 M
Beneficio bruto por ventas recuperadas	US$10,63 M
Ahorro en costos de inventario	US$7,06 M/año
Beneficio económico anual	US$17,69 M
ROI Año 1	36,10%
Payback	≈ 8 meses

Nota. Elaboración propia.

Además de estos indicadores financieros, será necesario medir la precisión del pronóstico, el error de predicción, la velocidad de reposición, el nivel de servicio y el porcentaje de recomendaciones automáticas aceptadas por los responsables de planificación.

4.5.12 Conclusión del caso de negocio

El análisis desarrollado permite transformar el modelo de Machine Learning planteado en las secciones anteriores en un caso de negocio cuantificable. La propuesta busca que Walmart pase de una gestión principalmente reactiva del inventario a una gestión predictiva basada en datos.

Bajo los supuestos establecidos, el piloto podría generar aproximadamente US$17,69 millones de beneficio económico anual, compuesto por US$10,63 millones asociados a la recuperación de ventas y US$7,06 millones derivados de la reducción de costos de inventario.

Considerando una inversión inicial de US$10 millones y costos operativos anuales de US$3 millones, el escenario estima un ROI de 36,1 % en el primer año y un período de recuperación aproximado de 8 meses.

No obstante, estos resultados deben considerarse como una estimación académica, ya que algunos parámetros fundamentales no son datos públicos de Walmart. La principal recomendación sería implementar primero un piloto controlado, medir la línea base de *stockouts*, precisión del pronóstico, inventario y ventas, y posteriormente recalcular el Business Case utilizando resultados reales.

De esta manera, el proyecto conecta los cinco pasos desarrollados: fuentes de datos → modelo de Machine Learning → áreas de aplicación → mejora esperada → impacto económico, demostrando cómo Big Data, Business Intelligence y

Data Science pueden integrarse para apoyar decisiones de inventario y generar valor empresarial.

CAPÍTULO 5. Analítica en tiempo real (Real Time) aplicada a la demanda

El último capítulo analiza la aplicación de un modelo de procesamiento en tiempo real (Real Time) como evolución natural de la arquitectura. Mientras el modelo del Capítulo 4 aprende de los datos históricos para predecir, el Real Time incorpora nuevas señales de demanda conforme ocurren, de modo que las decisiones de inventario y abastecimiento puedan ajustarse con mayor rapidez.

5.1 Objetivo de aplicación del modelo Real Time

El objetivo de aplicar un modelo de Real Time al proyecto de Walmart es desarrollar una capacidad de análisis que permita identificar y procesar inmediatamente nuevas señales relacionadas con la demanda de productos, con el propósito de mejorar la gestión de inventarios, las ventas y el abastecimiento. La finalidad principal es que Walmart pueda reaccionar rápidamente ante cambios inesperados en el comportamiento de la demanda y reducir situaciones de sobrestock y desabastecimiento.

En nuestro proyecto, el modelo de Real Time complementaría el modelo predictivo de demanda. Mientras que el análisis histórico permite estudiar el comportamiento de las ventas y generar predicciones, el procesamiento en tiempo real permitiría incorporar nuevas señales conforme estas se producen. Por ejemplo, un incremento repentino en las búsquedas de determinados productos, un cambio significativo en una variable climática, una fecha especial o una variación inesperada en determinados indicadores de consumo podrían convertirse en señales para actualizar el análisis de demanda.

La importancia de esta aplicación se encuentra en que el sector retail trabaja con productos de alta rotación y con una demanda que puede cambiar rápidamente. Una predicción realizada únicamente con información histórica puede quedarse desactualizada cuando ocurre un cambio inesperado en las condiciones del mercado. El procesamiento Real Time permitiría detectar estas variaciones con mayor rapidez y proporcionar información actualizada para apoyar las decisiones relacionadas con inventario, abastecimiento y ventas.

Por tanto, el objetivo específico del modelo consiste en detectar oportunamente cambios en las señales de demanda y proporcionar información actualizada que permita ajustar las decisiones comerciales y de inventario de Walmart, buscando reducir tanto el exceso de inventario como la falta de disponibilidad de productos.

Este planteamiento es coherente con el material del módulo, que define los datos en stream como información que fluye a medida que transcurre el tiempo y señala que el procesamiento Real Time resulta apropiado cuando se necesita una respuesta inmediata.

5.2 Fuentes de datos utilizadas

Para nuestro proyecto, las fuentes de datos deben mantenerse dentro del criterio que ya establecimos: utilizar información pública que podamos consultar, procesar y analizar, sin asumir acceso a las bases de datos internas de Walmart.

Una primera fuente corresponde a la información pública de Walmart, incluyendo sus informes anuales, resultados financieros, publicaciones corporativas, información sobre ventas, comercio electrónico, segmentos de negocio, evolución de operaciones e información pública relacionada con inventario y cadena de suministro. Estas fuentes permiten conocer el comportamiento general de la empresa y construir variables de contexto para el análisis.

Una segunda fuente corresponde a datos económicos públicos, como inflación, empleo, crecimiento económico y otros indicadores que puedan influir en el comportamiento del consumidor. Estas variables pueden utilizarse como factores externos para explicar variaciones en la demanda.

También se pueden utilizar datos climáticos y meteorológicos públicos, especialmente cuando las condiciones ambientales pueden modificar el comportamiento de compra. Por ejemplo, determinados productos pueden presentar incrementos de demanda durante periodos de temperaturas extremas, lluvias o determinados fenómenos climáticos.

Otra fuente corresponde a calendarios y fechas especiales, incluyendo festividades, temporadas comerciales, vacaciones y otros acontecimientos que puedan producir cambios en los patrones normales de consumo.

Además, pueden incorporarse datos públicos de tendencias de búsqueda y comportamiento del consumidor, siempre que sean accesibles y analizables, con el propósito de identificar cambios en el interés de los consumidores por determinadas categorías de productos.

Finalmente, pueden incorporarse datos geográficos y demográficos públicos, que permitan contextualizar diferencias de comportamiento entre regiones y mercados.

Es importante señalar que el proyecto no dispone de datos internos de Walmart, como transacciones individuales de clientes, inventario en tiempo real por tienda, información detallada de cada producto, datos de punto de venta, información privada de proveedores o registros internos de logística. Por esta razón, estas fuentes no deben presentarse como datos utilizados directamente en nuestro proyecto. El Real Time planteado constituye una arquitectura académica que demuestra cómo podrían procesarse señales actualizadas dentro de una solución empresarial.

El material del módulo establece que los datos en stream pueden proceder de diferentes fuentes y que pueden procesarse inmediatamente cuando se requiere una respuesta rápida.

5.3 Área de aplicación y departamentos de negocio implicados

El área principal de aplicación será la gestión de ventas, demanda e inventario de Walmart, debido a que el objetivo del proyecto es mejorar la capacidad de la empresa para anticipar el comportamiento de la demanda y tomar mejores decisiones de abastecimiento.

El primer departamento involucrado sería Inventarios, debido a que recibiría información actualizada sobre posibles cambios en la demanda. Esto permitiría identificar productos que podrían presentar un riesgo de desabastecimiento o, por el contrario, productos cuya demanda esté disminuyendo y puedan generar sobrestock.

El segundo departamento sería Compras y Abastecimiento, ya que las variaciones detectadas en la demanda pueden utilizarse como información para apoyar las decisiones sobre cantidades y momentos de reposición.

También participaría Logística y Cadena de Suministro, debido a que una variación en la demanda puede requerir modificar la distribución de productos entre diferentes ubicaciones. El análisis en tiempo real permitiría proporcionar información más actualizada para priorizar determinados productos o zonas.

El área de Ventas y Comercial también tendría un papel importante, ya que podría utilizar la información para identificar cambios en el comportamiento de los consumidores y evaluar posibles oportunidades comerciales o variaciones en la demanda de determinadas categorías.

Finalmente, estaría involucrada la Alta Dirección, mediante el uso de dashboards desarrollados en Power BI. Los responsables de gestión podrían visualizar indicadores actualizados relacionados con ventas, demanda, inventario, productos críticos y posibles riesgos de abastecimiento.

La participación conjunta de estas áreas permitiría que el modelo no se limite a generar información técnica, sino que produzca información útil para diferentes decisiones empresariales.

5.4 Mejora esperada

La principal mejora esperada es reducir el tiempo de reacción de Walmart ante cambios en la demanda. En lugar de depender únicamente de análisis históricos realizados periódicamente, el sistema podría identificar nuevas señales conforme aparecen y actualizar la información utilizada para la toma de decisiones.

Una segunda mejora sería la reducción del sobrestock. Si el sistema identifica que determinadas señales indican una disminución de la demanda, esta información puede utilizarse para evitar decisiones de abastecimiento excesivas y disminuir la acumulación innecesaria de productos.

También se espera una reducción del desabastecimiento. Cuando las señales disponibles indiquen un incremento inesperado de la demanda, el sistema podría

generar alertas para que las áreas de inventario, compras y logística revisen la disponibilidad del producto y actúen con mayor rapidez.

Otra mejora importante sería aumentar la precisión de las decisiones de abastecimiento, debido a que los responsables contarían con información histórica, predictiva y actualizada para complementar sus decisiones.

El Real Time también permitiría mejorar la capacidad de respuesta ante acontecimientos inesperados. Por ejemplo, una modificación importante en las condiciones climáticas, una tendencia de consumo o una fecha comercial podría producir una variación de demanda que no estuviera completamente reflejada en los datos históricos. La incorporación de estas señales permitiría actualizar el análisis con mayor rapidez.

En términos generales, la aplicación permitiría que Walmart avance desde una gestión principalmente basada en información histórica hacia una gestión que combine análisis descriptivo, predicción de demanda y señales actualizadas, mejorando la capacidad de anticipación de la empresa.

5.5 Caso de negocio del modelo Real Time

Para cuantificar la mejora se plantea un caso de negocio académico. Dado que el proyecto utiliza información pública, los valores empleados son supuestos que permiten demostrar la metodología de cálculo y no constituyen resultados financieros reales de Walmart. El escenario de referencia es el mismo que se planteó en el Capítulo 5 de la Parte I, ahora atribuido a la combinación de analítica predictiva y procesamiento en tiempo real.

Tabla 33. ***Caso de negocio del modelo Real Time (escenario académico)***

Concepto	Supuesto o cálculo	Resultado anual
Costo de mantener el inventario	USD 100 M de inventario × 20 %	USD 20.000.000
Inventario innecesario liberado	USD 100 M × 5 %	USD 5.000.000
Ahorro por menor inventario	USD 5 M × 20 % de costo de mantenimiento	USD 1.000.000
Ventas recuperadas por mejor disponibilidad	USD 200 M de ventas analizadas × 1 %	USD 2.000.000
Beneficio económico bruto	USD 1.000.000 + USD 2.000.000	USD 3.000.000
Costo anual de implementar y operar la solución	Supuesto académico	USD 800.000
Beneficio neto	USD 3.000.000 − USD 800.000	USD 2.200.000
ROI	(USD 2.200.000 / USD 800.000) × 100	275 %

Nota. Elaboración propia. Valores supuestos con fines metodológicos.

Bajo este escenario hipotético, cada dólar invertido en la solución generaría 2,75 dólares de beneficio neto. Conviene, sin embargo, leer esta cifra junto con el caso de negocio del Capítulo 4, que constituye la referencia principal del

proyecto. Aquel caso se construye sobre las cifras públicas de Walmart U.S. del año fiscal 2026, incluye una inversión inicial de 10 millones de dólares y arroja un ROI de 36,1 % en el primer año, con recuperación de la inversión en aproximadamente ocho meses. La diferencia entre ambos resultados no es una contradicción: el escenario de 275 % considera solo costos operativos anuales sobre una unidad de negocio de referencia, mientras que el de 36,1 % incorpora la inversión de implementación de un piloto a escala real. Los dos coinciden en la conclusión central: los beneficios de reducir el sobrestock y el desabastecimiento superan el costo de la solución.

Para convertir este ejercicio en un caso de negocio real, Walmart debería utilizar sus datos internos de inventario, costos de almacenamiento, márgenes, ventas perdidas por falta de producto y costos efectivos de implementación. En consecuencia, el caso demuestra que la combinación de Big Data, análisis predictivo y procesamiento en tiempo real puede generar valor económico, siempre que se valide mediante un piloto controlado con una línea base medida.

CONCLUSIONES Y APLICACIONES

Síntesis del proyecto

El proyecto ha diseñado una propuesta integral para que Walmart Inc. evolucione de una gestión descriptiva y reactiva del inventario hacia un modelo predictivo y prescriptivo, cuyo caso de uso central es la predicción de la demanda por producto, tienda y período. La propuesta se construyó de forma incremental y coherente: cada decisión se apoyó en la anterior y todas responden al mismo problema de negocio, la coexistencia de sobrestock y desabastecimiento.

En el bloque de Business Intelligence se diagnosticó que el reto de Walmart no es la falta de datos, sino su integración, su gobierno y su conversión en conocimiento accionable. Se describió una arquitectura que va de los sistemas transaccionales al Data Warehouse y a los Data Marts especializados, se seleccionó Power BI por su integración nativa con el ecosistema Azure y su adopción interna previa, se definieron cinco perfiles de usuario con sus necesidades de información y niveles de acceso, y se diseñó la Gerencia de Analítica e Inteligencia de Demanda, con un organigrama, unas dependencias y un modelo de solución de problemas basado en KDD.

En el bloque de Big Data e Inteligencia Artificial se incorporaron fuentes de datos externas y no estructuradas —clima, eventos locales, redes sociales, indicadores económicos— como señales tempranas de la demanda; se seleccionó Microsoft Azure con un modelo PaaS y una arquitectura en seis capas con gobierno transversal; se eligió Apache Spark con Python como motor de procesamiento; y se definió un modelo supervisado de regresión que cruza la demanda pronosticada con el inventario para anticipar riesgos de stock-out y de sobrestock, complementado con un modelo de tiempo real que incorpora las señales conforme ocurren.

Conclusiones

- **Primera:** el valor de la analítica no proviene de la tecnología aislada, sino de su alineación con una decisión de negocio concreta. Todas las elecciones tecnológicas del proyecto (Power BI, Azure, Spark, el modelo de Machine Learning) se justificaron por su contribución a reducir el sobrestock y el desabastecimiento, en línea con lo que señalan Chen et al. (2012) y Provost y Fawcett (2013b).
- **Segunda:** la coherencia del ecosistema es tan importante como la potencia de cada componente. La integración nativa entre Azure, Synapse, el Data Lake y Power BI reduce costos, riesgos de integración y curva de aprendizaje, y respeta la estrategia multinube de Walmart, que evita depender de un único proveedor y, en particular, de su principal competidor.
- **Tercera:** el componente humano y organizativo es decisivo. La principal carencia identificada no es tecnológica, sino de capacidades para interpretar los datos y convertirlos en decisiones. Por ello, la propuesta

combina perfiles técnicos y de negocio, responsabilidades claras frente a cada problema y control humano en las decisiones de baja confianza, como recomiendan Shrestha et al. (2019) y Ransbotham et al. (2017).

- **Cuarta:** el caso de negocio muestra que la solución es económicamente viable. El piloto sobre el 10 % de Walmart U.S. generaría un beneficio anual estimado de 17,69 millones de dólares, con un ROI de 36,1 % en el primer año y la recuperación de la inversión en unos ocho meses. El escenario preliminar de una unidad de negocio y el de tiempo real (ROI de 275 %) confirman la misma conclusión con supuestos distintos. En todos los casos, el resultado depende de forma crítica de la magnitud real de las ventas perdidas por falta de inventario, por lo que debe validarse con datos internos.
- **Quinta:** la evolución de la analítica es progresiva. El proyecto recorre las cuatro etapas —descriptiva, diagnóstica, predictiva y prescriptiva— y muestra que el tiempo real no sustituye al modelo predictivo, sino que lo complementa al reducir el tiempo de reacción ante cambios de la demanda (Iansiti y Lakhani, 2020).

Aplicaciones

La aplicación inmediata del proyecto es un piloto controlado en Walmart U.S., centrado en categorías de alta rotación (perecibles, consumibles y productos estacionales), con una línea base medida antes de la implementación y con indicadores de seguimiento definidos: precisión del pronóstico, tasa de desabastecimiento, días de inventario, nivel de servicio y ventas recuperadas. Los resultados del piloto permitirían recalcular el caso de negocio con datos reales y decidir su escalamiento a Sam's Club y Walmart International.

Más allá de Walmart, la metodología es transferible a cualquier organización que deba anticipar una demanda variable con recursos limitados: cadenas de supermercados y farmacias, empresas de consumo masivo, distribuidores, e incluso empresas de servicios cuyos ingresos dependen de transacciones recurrentes. En el contexto latinoamericano, donde muchas empresas aún gestionan su inventario con reportes descriptivos y hojas de cálculo, la secuencia propuesta —ordenar y gobernar los datos, visualizarlos por perfil, organizar un equipo responsable y, solo después, incorporar modelos predictivos— ofrece una hoja de ruta escalable que no exige la infraestructura de Walmart para empezar a generar valor.

Limitaciones y líneas futuras

La principal limitación del trabajo es su carácter académico: se basa en fuentes públicas y no en los datos internos de Walmart, de modo que la arquitectura y los procesos internos se presentan como propuestas fundamentadas y las cifras del caso de negocio como escenarios con supuestos explícitos. Como líneas futuras se proponen: validar el modelo con datos transaccionales reales mediante el piloto; incorporar técnicas de explicabilidad que faciliten la confianza

de los usuarios en el pronóstico; evaluar el procesamiento de flujos continuos con Apache Flink o Spark Structured Streaming en la fase de tiempo real; y profundizar en el marco ético y de privacidad (CCPA/CPRA, LFPDPPP, PIPEDA y GDPR) a medida que se incorporen datos de comportamiento digital de los clientes.

REFERENCIAS

18Knowledge. (2020). *Retail Link de Walmart: Cómo funciona y su impacto en la cadena de suministro*. 18Knowledge Insights.

Amazon Web Services. (s. f.). *What is big data?* https://aws.amazon.com/what-is/big-data/

Ankashram. (2025, 23 de agosto). *Walmart's inventory optimization: Data analytics rewrote the rules of retail*. https://www.ankashram.com/case-study/data-analytics-walmart-retail

Aspectos legales y éticos del big data: La privacidad y la seguridad de la información [Material del curso]. (s. f.). Westfield Business School.

CertLibrary. (2025, 24 de junio). *Why Walmart chose Microsoft Azure for their cloud transformation*. https://www.certlibrary.com/blog/why-walmart-choose-microsoft-azure-for-their-cloud-transformation/

Chain Store Age. (2025, 12 de septiembre). *Walmart expands customer insights available to suppliers*. https://chainstoreage.com/walmart-expands-customer-insights-available-suppliers

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188. https://doi.org/10.2307/41703503

DataCamp. (2026, 6 de agosto). *AWS vs. Azure vs. Google Cloud: Una comparativa completa*. https://www.datacamp.com/es/blog/aws-vs-azure-vs-gcp

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Davenport, T. H., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*, 24–42. https://doi.org/10.1007/s11747-019-00696-0

DCFmodeling. (2025, 25 de noviembre). *Mission statement, vision, & core values (2025) of Walmart Inc. (WMT)*. https://www.dcfmodeling.com/blogs/vision/wmt-mission-vision

Digital Innovation and Transformation. (2020, 14 de abril). *Walmart's Intelligent Retail Lab: AI in the aisles*. Harvard Business School Digital Initiative.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Gartner Peer Insights. (s. f.). *Top Cognizant finance and accounting business process outsourcing services alternatives & competitors 2026*.

Recuperado el 15 de septiembre de 2026, de https://www.gartner.com/reviews/

Google Cloud. (s. f.). *Introduction to column-level access control*. BigQuery documentation. https://cloud.google.com/bigquery/docs/column-level-security-intro

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3.ª ed.). Morgan Kaufmann.

Horses for Sources. (2016, 24 de abril). *Accenture, Genpact, IBM, EXL, TCS, Capgemini and WNS lead the first As-a-Service lens of finance and accounting*. https://www.horsesforsources.com/

Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press.

InfoWorld. (2025, 13 de mayo). *How Walmart abstracts its hybrid cloud for developers*. https://www.infoworld.com/article/2335955/how-walmart-abstracts-its-hybrid-cloud-for-developers.html

InvGate. (2022, 29 de julio). *Amazon AWS vs. Microsoft Azure vs. Google Cloud: ¿Qué proveedor en la nube es mejor?* https://blog.invgate.com/es/amazon-aws-vs-microsoft-azure-vs-google-cloud

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. O'Reilly Media.

LitCommerce. (2026, 10 de junio). *Walmart supply chain: Strategy, tech & fulfillment in 2026*. https://litcommerce.com/blog/walmart-supply-chain/

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Marquesme. (2022, 9 de mayo). *Los mejores servicios en la nube: Azure, AWS y Google Cloud*. https://www.marquesme.com/servicios-en-la-nube-azure-aws-google-cloud/

Marr, B. (2017, 23 de enero). Really big data at Walmart: Real-time insights from their 40+ petabyte data cloud. *Forbes*. https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/

Marr, B. (s. f.). *Walmart: Big data analytics at the world's biggest retailer*. Bernard Marr & Co. https://bernardmarr.com/walmart-big-data-analytics-at-the-worlds-biggest-retailer/

McKendrick, J. (2016). *Walmart's gigantic private cloud for real-time inventory control*. RTInsights. https://www.rtinsights.com/walmart-cloud-inventory-management-real-time-data/

Microsoft Azure. (s. f.). *Servicios de informática en la nube*. https://azure.microsoft.com/es-es

National CIO Review. (2025). *What CIOs can learn from Walmart's high-tech strategy*. https://nationalcioreview.com/articles-insights/leadership/what-cios-can-learn-from-walmarts-high-tech-strategy/

Navertia. (2024, 21 de mayo). *Comparativa servicios en la nube: AWS vs Azure vs Google Cloud*. https://www.navertia.com/blog/aws-vs-azure-vs-google-cloud/

Pérez Galán, J. L. (2019). *Artículo, caso y nota técnica de business intelligence y big data* [Material del curso]. EIG Business School.

Pérez Galán, J. L. (2025). *Implementación de inteligencia artificial (IA)* [Documento interno]. Westfield USA, LLC.

Pérez Galán, J. L. (s. f.-a). *Implementación de inteligencia artificial (IA)* [Material del curso]. Westfield Business School.

Pérez Galán, J. L. (s. f.-b). *Introducción al machine learning* [Material del curso]. Westfield Business School.

Poornaprajna International Journal of Management, Education & Social Science. (2025). *Strategic analysis of Walmart Inc.: Innovation, ethical challenges, and the future of AI-driven services*. https://www.researchgate.net/publication/395708661

ProjectPro. (2024). *How Walmart uses big data analytics: Real-world case study and architecture*. ProjectPro.

Provost, F., & Fawcett, T. (2013a). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Provost, F., & Fawcett, T. (2013b). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). *Reshaping business with artificial intelligence: Closing the gap between ambition and action*. MIT Sloan Management Review.

Retail Dive. (2025, septiembre). *Walmart taps OpenAI for employee training*. https://finance.yahoo.com/news/walmart-taps-openai-employee-training-112700756.html

Retail IT Insights. (2017, 3 de febrero). *Walmart turns to Data Café analytics hub to make sense of data*. https://www.retailitinsights.com/doc/walmart-turns-to-data-caf-analytics-hub-to-make-sense-of-data-0001

Sharda, R., Delen, D., & Turban, E. (2022). *Analytics, AI, and decision support*. Pearson.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Stock Titan. (2023, 27 de junio). *Genpact and Walmart extend partnership in North America*. https://www.stocktitan.net/

Supermarket News. (2012, 31 de agosto). *Wal-Mart unveils search engine for website*. https://www.supermarketnews.com/grocery-technology/wal-mart-unveils-search-engine-for-website

TechTarget. (2022, 12 de julio). *Walmart's multi-cloud strategy cuts millions in IT costs*. https://www.techtarget.com/searchcloudcomputing/news/252522631/

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media.

Vector. (2026, 18 de mayo). *Walmart's supply chain: A detailed look at how they manage it*. https://www.withvector.com/blog/walmarts-supply-chain-a-detailed-look-at-how-they-manage-it/

Walmart Inc. (2018, 30 de octubre). *When man and machine meet in retail, customers win*. https://corporate.walmart.com/news/2018/10/30/when-man-and-machine-meet-in-retail-customers-win

Walmart Inc. (2019). *Proxy statement for the 2019 annual shareholders' meeting* (Form DEF 14A). U.S. Securities and Exchange Commission. https://www.sec.gov/

Walmart Inc. (2025a). *Annual report on Form 10-K for the fiscal year ended January 31, 2025*. U.S. Securities and Exchange Commission. https://www.sec.gov/

Walmart Inc. (2025b, 17 de julio). *Walmart's U.S. supply chain playbook goes global — and it's reinventing retail at scale*. https://corporate.walmart.com/news/2025/07/17/walmarts-us-supply-chain-playbook-goes-global-and-its-reinventing-retail-at-scale

Walmart Inc. (2026). *Annual report on Form 10-K for the fiscal year ended January 31, 2026*. U.S. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/104169/000010416926000055/wmt-20260131.htm

Walmart Inc. (s. f.). *Walmart Careers* [Portal de empleo de Walmart y Walmart Global Tech]. https://careers.walmart.com/

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a

longitudinal case study. *International Journal of Production Economics, 165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

White, T. (2015). *Hadoop: The definitive guide* (4.ª ed.). O'Reilly Media.

Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M. J., Shenker, S., & Stoica, I. (2012). Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. En *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)* (pp. 15–28). USENIX.

ZipDo. (2026, 6 de agosto). *Top 10 best outsourcing business services: 2026 comparison*. https://zipdo.co/service/outsourcing-business/