

Maestría en

ADMINISTRACIÓN DE EMPRESAS

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

AUTORES:

Ing. Darío Estuardo Pallmay Macías

Econ. Eliana Micaela Guanoluisa Ruiz

Lic. Nathaly Haymeth Rivadeneira Lagos

Lic. Esteban Daniel Rosas Pazmiño

Ing. Dayana Estefanía Herrera Granda

TUTOR: José Antonio García Fernández

Big Data y Business Intelligence aplicados a PIGE: de la
información operativa a la decisión estratégica

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

Big Data y Business Intelligence aplicados a PIGE: de la información operativa a la decisión
estratégica

Por

Ing. Darío Estuardo Pallmay Macías

Econ. Eliana Micaela Guanoluisa Ruiz

Lic. Nathaly Haymeth Rivadeneira Lagos

Lic. Esteban Daniel Rosas Pazmiño

Ing. Dayana Estefanía Herrera Granda

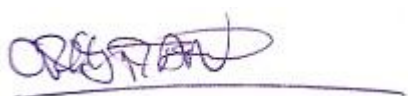
Octubre 2026

Aprobado

A handwritten signature in purple ink, appearing to read 'CRISTIAN', is written over a horizontal line.

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. Darío Estuardo Pallmay Macías, Econ. Eliana Micaela Guanoluisa Ruiz, Lic. Nathaly Haymeth Rivadeneira Lagos, Lic. Esteban Daniel Rosas Pazmiño y Ing. Dayana Estefanía Herrera Granda son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Cristian Melo
Coordinador MBA UIDE



Ignacio Maroto
Provost WBS

Autoría del Trabajo de Titulación

Nosotros, Ing. Darío Estuardo Pallmay Macías, Econ. Eliana Micaela Guanoluisa Ruiz, Lic. Nathaly Haymeth Rivadeneira Lagos, Lic. Esteban Daniel Rosas Pazmiño y Ing. Dayana Estefanía Herrera Granda, declaramos bajo juramento que el trabajo de titulación titulado **Big Data y Business Intelligence aplicados a PIGE: de la información operativa a la decisión estratégica** 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. Darío Estuardo Pallmay Macías

Correo electrónico:

pallmaydario@gmail.com



Econ. Eliana Micaela Guanoluisa Ruiz

Correo electrónico:

micaelaguanoluisa56@gmail.com



Lic. Nathaly Haymeth Rivadeneira

Lagos

Correo electrónico:

nathalyrivadeneira2010@gmail.com



Lic. Esteban Daniel Rosas Pazmiño

Correo electrónico:

esteban_rp24@outlook.com



Ing. Dayana Estefanía Herrera Granda

Correo electrónico: daya-

herrera@hotmail.com

Autorización de Derechos de Propiedad Intelectual

Nosotros, Ing. Darío Estuardo Pallmay Macías, Econ. Eliana Micaela Guanoluisa Ruiz, Lic. Nathaly Haymeth Rivadeneira Lagos, Lic. Esteban Daniel Rosas Pazmiño y Ing. Dayana Estefanía Herrera Granda, en calidad de autores del trabajo de investigación titulado «Big Data y Business Intelligence aplicados a PIGE: de la información operativa a la decisión estratégica», 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 son los establecidos 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. Darío Estuardo Pallmay Macías

Correo electrónico:

pallmaydario@gmail.com



Lic. Nathaly Haymeth Rivadeneira

Lagos

Correo electrónico:

nathalyrivadeneira2010@gmail.com



Ing. Dayana Estefanía Herrera Granda

Correo electrónico: daya-

herrera@hotmail.com



Econ. Eliana Micaela Guanoluisa Ruiz

Correo electrónico:

micaelaguanoluisa56@gmail.com



Lic. Esteban Daniel Rosas Pazmiño

Correo electrónico:

esteban_rp24@outlook.com

Agradecimientos

Agradecemos a la Universidad Internacional del Ecuador, a Westfield Business School y a EIG por la formación recibida; a nuestros docentes, por su orientación y su retroalimentación; y a la Gerencia y al equipo de PIGE, por abrir la información de la empresa para este trabajo.

Dedicatorias

A Dios, por ser mi guía y fortaleza.

A mi esposa e hijos, quienes fueron mi inspiración constante y la razón principal para seguir adelante. Gracias por comprensión y paciencia durante los momentos en los que mi tiempo y mi presencia estuvieron dedicados a esta meta académica.

A mis padres, por inculcarme valores, esfuerzo y responsabilidad. Gracias por creer siempre en mí.

A mis hermanos, por su cariño.

A mis compañeros de estudio, con quienes compartí esta etapa de aprendizaje, intercambio de experiencias y crecimiento académico. Gracias por enriquecer cada asignatura con sus conocimientos e ideas.

Finalmente, a cada uno de mis profesores, por su compromiso, dedicación y entrega. Gracias por compartir sus conocimientos, experiencias y valores.

A todos ustedes, mi más profunda gratitud por haber sido parte esencial de este logro que hoy culmina con orgullo y satisfacción.

Dario.

A mis padres, por sus esfuerzos incansables y por enseñarme con su ejemplo el verdadero significado de la superación. A mi hermano mayor y a mis tías, por su apoyo constante en cada paso de este camino. Este logro es tan suyo como mío.

Con amor y gratitud infinita,

Dayana

Primero a Dios, porque en los momentos que sentí que no podía más él me sostuvo y me dio la fuerza para seguir.

A mi hija, mi razón de ser; cada página de este trabajo lleva el tiempo que te quité y la esperanza de que algún día te sientas orgullosa de tu mamá.

A mi esposo, por caminar a mi lado, por cargar con lo que yo no alcanzaba a cubrir y por creer en mí aun cuando yo dudaba.

A mis padres, por enseñarme con su ejemplo que el esfuerzo siempre vale la pena y que todos los sueños se cumplen con constancia.

A toda la familia, por el cariño, las palabras de ánimo y por nunca soltarme en ningún momento.

A todos mis compañeros, con quienes compartí desvelos, aprendizajes, trabajos en equipo que hicieron más llevadero este camino.

A la universidad y a sus docentes, por la formación recibida y por abrirme las puertas para alcanzar esta meta.

Con profundo amor y gratitud,

Nathaly

A mis padres, por su amor, apoyo incondicional y por ser mi mayor inspiración para alcanzar esta meta. A mi hermana Marcela, por su cariño, compañía y por estar siempre a mi lado. Gracias por creer en mí y acompañarme en cada paso de este camino.

Micaela

A Dios, por darme la fortaleza, la sabiduría y la perseverancia para culminar esta etapa de mi formación académica.

A mis padres, Javier y Sashen, por ser mis guías y mi apoyo incondicional, incluso en las noches de desvelo. Me enseñaron con el ejemplo que el esfuerzo siempre vale la pena.

A mi abuelita Esthelita, por estar siempre a mi lado. Y a mi abuelito Vichi (†), que me enseñó el valor del aprendizaje; sé que desde donde estés celebras conmigo este logro. Ambos son mi ejemplo a seguir.

A mi familia, que hoy son grandes profesionales y, siendo mayores que yo, abrieron el camino que ahora recorro. De cada uno aprendí algo que me trajo hasta aquí.

A la universidad y a mis profesores, por acompañarme y aconsejarme durante todo este tiempo de estudio y crecimiento.

Este logro también es de ustedes.

Esteban

ÍNDICE DEL DOCUMENTO

INTRODUCCIÓN Y OBJETO DEL PROYECTO.....	13
PARTE I: DIAGNÓSTICO, INDICADORES Y DISEÑO DE LA SOLUCIÓN DE BUSINESS INTELLIGENCE.....	15
Capítulo 1. Definición de la propuesta y contexto de PIGE (Semana 1).....	15
1.1 Identificación de la empresa	15
1.2 Procesos y organización.....	16
1.3 Objetivo del proyecto, vinculación y fuentes de información	17
1.4 Valor del Big Data para un distribuidor de EPP	17
Capítulo 2. Fuentes de datos, KPI y diagnóstico de carencias (Semana 2).....	19
2.1 Objetivo y alcance	19
2.2 Consolidación y origen de la información	19
2.3 Destino de la carga y datamarts.....	20
2.4 Indicadores clave de desempeño (KPI)	21
2.5 Informes actuales, necesidades y soluciones propuestas.....	23
2.6 Carencias identificadas en la carga y gestión de datos.....	24
2.7 Componentes de Business Intelligence aplicados a PIGE	24
2.8 Protocolo de implantación	25
Capítulo 3. Proceso ETL e infraestructura de datos (Semana 3)	27
3.1 Objetivo y alcance	27
3.2 Proceso ETL	27
3.3 Calidad de la información disponible.....	28
3.4 Horarios de carga y accesibilidad.....	29
3.5 Variables disponibles y gobierno de acceso	29
3.6 Sistemas y recursos	30
3.7 Dimensionamiento de la infraestructura y su costo	30
Capítulo 4. Plan de visualización (Semana 4).....	32
4.1 ¿Qué información se mostrará?	32
4.2 ¿Cómo se mostrará?.....	32
4.3 ¿A quién va dirigida?	33
4.4 ¿Cuál es el objetivo de la información?	34
4.5 Ampliación del catálogo de KPI y período de análisis.....	35
Capítulo 5. Área de analítica y caso de negocio inicial (Semana 5).....	37

5.1 Posicionamiento y objetivo del área	37
5.2 Estructura y responsabilidades	37
5.3 Impacto esperado y presupuesto	38
5.4 Impacto en las iniciativas propuestas	39
5.5 Gestión del cambio y adopción	39
5.6 Estructura del trabajo y valoración	40
PARTE II: DATOS EXTERNOS, ARQUITECTURA TECNOLÓGICA E INTELIGENCIA ARTIFICIAL.....	41
Capítulo 6. Obtención de datos externos y no estructurados (Semana 6).....	41
6.1 Objetivo y alcance	41
6.2 Marco legal: protección de datos personales en Ecuador	41
6.3 Fuentes externas	42
6.3.1 Plataformas digitales: Meta y LinkedIn	43
6.3.2 Inteligencia de comercio exterior: Veritrade	44
6.4 Datos internos desorganizados	45
6.5 Solución para estructurar la información.....	46
6.6 Avance técnico en Power BI	47
Capítulo 7. Arquitectura de la información: Zona de Aterrizaje (Semana 7).....	49
7.1 Objetivo y situación actual.....	49
7.2 Modelo conceptual: Zona de Aterrizaje para PIGE	49
7.3 Capacidades aplicables	50
7.4 Estructura y codificación de los datos	51
7.5 Estructura propuesta del proyecto.....	51
Capítulo 8. Arquitectura de código abierto (Semana 8).....	53
8.1 Objetivo y criterio de selección	53
8.2 Hadoop frente a Spark	53
8.3 Lenguajes y herramientas complementarias.....	54
8.4 Ecosistema propuesto.....	55
8.5 Costos de implementación y operación	55
8.6 Estructura y valoración.....	57
Capítulo 9. Implementación de inteligencia artificial (Semana 9).....	58
9.1 Objetivo de aplicación del modelo	58
9.2 Fuentes de datos	60
9.3 Áreas de aplicación y departamentos implicados.....	61

9.4 Identificación de la mejora esperada	62
9.5 Caso de negocio.....	63
Capítulo 10. Caso de uso en tiempo casi real y valoración integral (Semana 10)	66
10.1 Objetivo y alcance	66
10.2 Marco teórico aplicado: qué significa tiempo real para PIGE	66
10.3 Fases del detector de quiebres	69
10.4 Diseño del repositorio de datos	69
10.5 Reglas y algoritmo de detección	70
10.6 Herramientas evaluadas y diseños propuestos.....	72
10.7 Fuentes de datos y departamentos implicados	73
10.8 Mejora esperada y costo incremental	73
10.9 Valoración integral del proyecto	75
10.10 Riesgos y supuestos críticos	77
10.11 Conclusión del capítulo.....	78
CONCLUSIONES Y APLICACIONES	79
REFERENCIAS.....	81

INDICE DE FIGURAS Y TABLAS

Tabla 1	<i>Correspondencia entre los capítulos y las entregas semanales</i>	14
Tabla 2	<i>Cifras del negocio, 2021-2023</i>	16
Tabla 3	<i>Procesos clave de PIGE y su medición actual</i>	16
Tabla 4	<i>Fuentes de datos y recurrencia de carga propuesta</i>	20
Tabla 5	<i>Tipos de datos requeridos por área</i>	21
Tabla 6	<i>KPI de Stocks y Bodega</i>	21
Tabla 7	<i>KPI de Compras</i>	22
Tabla 8	<i>KPI de Ventas</i>	22
Tabla 9	<i>KPI de Logística</i>	22
Tabla 10	<i>Informes actuales, necesidades detectadas y solución propuesta</i>	23
Tabla 11	<i>Carencias del proceso de información y su implicación</i>	24
Tabla 12	<i>Componentes de BI en PIGE</i>	25
Tabla 13	<i>Calidad actual de las fuentes de datos</i>	28
Tabla 14	<i>Horarios de carga propuestos</i>	29
Tabla 15	<i>Variables por datamart y responsable de autorizar el acceso</i>	29
Tabla 16	<i>Dimensionamiento y costo referencial de la infraestructura, primera etapa</i>	31
Tabla 17	<i>Matriz de audiencias, KPI y periodicidad</i>	33
Tabla 18	<i>KPI complementarios</i>	35
Tabla 19	<i>Plan de visualización de PIGE</i>	36
Tabla 20	<i>Estructura del área de analítica</i>	37
Tabla 21	<i>Presupuesto anual de creación del área de analítica</i>	38
Tabla 22	<i>Ahorro en la consolidación de informes</i>	39
Tabla 23	<i>Entidades ecuatorianas que podrían proveer información</i>	43
Tabla 24	<i>Herramientas de plataformas digitales</i>	44
Tabla 25	<i>Datos de Veritrade y su utilidad</i>	45
Tabla 26	<i>Capacidades de la Zona de Aterrizaje aplicables a PIGE</i>	50
Tabla 27	<i>Comparación de Hadoop y Apache Spark aplicada a PIGE</i>	53
Tabla 28	<i>Ecosistema de código abierto de PIGE, de las fuentes a Power BI</i>	55
Tabla 29	<i>Costos estimados de implementación y operación, primer año</i>	56
Tabla 30	<i>Definición del modelo predictivo de reabastecimiento y riesgo de quiebre</i>	59
Tabla 31	<i>Fuentes de datos del modelo</i>	60
Tabla 32	<i>Papel de cada departamento frente al modelo</i>	61
Tabla 33	<i>Situación actual y proyectada de los indicadores</i>	62
Tabla 34	<i>Beneficios y costos anuales con el modelo en régimen</i>	64
Tabla 35	<i>Análisis de escenarios, con el modelo en régimen</i>	64
Tabla 36	<i>Técnicas de análisis en tiempo real y su aplicación en PIGE</i>	67
Tabla 37	<i>Latencias de PIGE: situación actual y objetivo de diseño</i>	68
Tabla 38	<i>Flujo del detector, de la transacción a la alerta</i>	69
Tabla 39	<i>Estructuras de datos del detector</i>	70
Tabla 40	<i>Eventos que reconoce el detector</i>	71
Tabla 41	<i>Herramientas de Big Data y decisión para PIGE</i>	72
Tabla 42	<i>Situación actual y objetivo de los indicadores</i>	73
Tabla 43	<i>Costo incremental de la capa de tiempo real, primer año</i>	74

Tabla 44 <i>Sensibilidad: quiebre adicional evitado por la capa de tiempo real</i>	75
Tabla 45 <i>Resultado financiero del proyecto a cinco años</i>	75
Tabla 46 <i>Sensibilidad del valor actual neto</i>	76
Tabla 47 <i>Cronograma de implantación en cinco fases</i>	76
Tabla 48 <i>Dos trayectorias para PIGE</i>	77

INTRODUCCIÓN Y OBJETO DEL PROYECTO

PIGE es una empresa ecuatoriana de origen familiar, fabricante e importadora de equipos de protección personal (EPP), que opera desde Quito con cobertura nacional y que entre 2021 y 2023 elevó sus ventas de USD 1,15 a USD 1,63 millones. Ese crecimiento sostenido ha incrementado la complejidad de su operación y, con ella, la necesidad de decidir con evidencia (Davenport y Harris, 2007). Hoy la compañía cuenta con abundante información transaccional en su sistema SitacPlus, pero la utiliza de manera operativa y reactiva: los reportes responden preguntas puntuales y no ofrecen una visión agregada, gráfica ni predictiva del negocio.

El objeto de este proyecto es diseñar la evolución de PIGE desde el dato duro hacia la decisión informada, mediante la implementación progresiva de herramientas de Big Data y Business Intelligence (BI). El trabajo recorre el ciclo completo: definición del negocio y de sus indicadores, diseño del proceso de extracción, transformación y carga (ETL), plan de visualización, creación del área de analítica, incorporación de fuentes externas, arquitectura de información, selección de tecnologías de código abierto y, finalmente, la aplicación de un modelo predictivo de reabastecimiento, un detector de quiebres en tiempo casi real y la valoración financiera integral del proyecto.

El documento consolida las diez entregas semanales del Equipo 3 y está organizado en dos partes. La Parte I (capítulos 1 a 5) establece el diagnóstico, los indicadores, el proceso ETL, la visualización y el área responsable de la analítica. La Parte II (capítulos 6 a 10) amplía el alcance hacia datos externos y no estructurados, la arquitectura de información, el ecosistema tecnológico, la inteligencia artificial y el análisis en tiempo casi real, y cierra con la valoración financiera. Las cifras económicas incorporan las precisiones confirmadas por la Gerencia de PIGE: la empresa vende a crédito, el margen bruto global ronda el 30 %, no se

dispone de márgenes por línea y la tasa de descuento de 12 % equivale al costo de un crédito.

Cada capítulo incorpora, cuando correspondía, las observaciones formuladas por el docente durante el curso.

TABLA 1

CORRESPONDENCIA ENTRE LOS CAPÍTULO Y LAS ENTREGAS SEMANALES

Capítulo	Contenido	Semana
1	Definición de la propuesta y contexto de PIGE	1
2	Fuentes de datos, KPI y diagnóstico de carencias	2
3	Proceso ETL e infraestructura de datos	3
4	Plan de visualización	4
5	Área de analítica y caso de negocio inicial	5
6	Datos externos y no estructurados; avance en Power BI	6
7	Arquitectura de la información: Zona de Aterrizaje	7
8	Arquitectura de código abierto	8
9	Implementación de inteligencia artificial	9
10	Caso de uso en tiempo casi real y valoración integral	10

PARTE I: DIAGNÓSTICO, INDICADORES Y DISEÑO DE LA SOLUCIÓN DE BUSINESS INTELLIGENCE

La primera parte responde a la pregunta de dónde está PIGE y hacia dónde debe moverse. Describe el negocio, identifica las fuentes de información y los indicadores que importan, diseña el flujo que llevará los datos hasta los tableros, define cómo se presentarán a cada audiencia y propone la estructura organizacional que sostendrá el esfuerzo.

Capítulo 1. Definición de la propuesta y contexto de PIGE (Semana 1)

1.1 Identificación de la empresa

PIGE, Protección Industrial y Gestión de Equipos, se desarrolla en el sector comercial como importador y distribuidor directo de marcas reconocidas de EPP, además de fabricante de una parte de su portafolio. Atiende a los segmentos de construcción, minería e industria, cuenta con distribuidores, y su sede y única sucursal están en Quito. Inició actividades en 2014.

Visión. Ser líderes reconocidos en la provisión y fabricación de equipos de protección personal, pioneros en la adopción de tecnologías e innovación de producto que cumplan la normativa vigente, y convertirse en socios estratégicos de los clientes.

Misión. Fabricar e importar equipos de protección personal y, con el respaldo de un equipo humano comprometido, asesorar las necesidades de los proyectos de los clientes en todo el territorio nacional, a la vanguardia de la comercialización y distribución de marcas reconocidas.

El portafolio se organiza en once líneas de producto: equipos anticaídas, protección auditiva, corporal, para cabeza y rostro, para manos, respiratoria, para piernas, visual, remediación ambiental, vial y consumibles.

TABLA 2*CIFRAS DEL NEGOCIO, 2021-2023*

Indicador	Valor
Ventas 2021 / 2022 / 2023	USD 1,15 M / 1,32 M / 1,63 M
Crecimiento anual	14,8 % (2022) y 23,5 % (2023)
Tasa de crecimiento anual compuesta	≈ 19 %
Colaboradores	16
Sucursales	1
Flota vehicular	3 unidades
Modalidad de venta	A crédito
Margen bruto global	≈ 30 % (la empresa no dispone de márgenes por línea)

Nota. Elaboración propia a partir de la propuesta de proyecto y de las precisiones de la Gerencia de PIGE. La tasa compuesta corresponde al 19 % reportado por la empresa.

1.2 Procesos y organización

TABLA 3*PROCESOS CLAVE DE PIGE Y SU MEDICIÓN ACTUAL*

Proceso	Cómo opera hoy	Qué se mide
Compras	El reabastecimiento se activa cuando el stock cae bajo el mínimo, según el criterio de la jefatura de bodega, lo que genera una alerta a Compras o Importaciones.	Volúmenes de adquisiciones locales e importaciones.
Bodega	Recepción, estiba, preparación de pedidos, embalaje y despacho.	Rotación de inventarios.
Ventas	Publicidad en redes y web; atención presencial, en mostrador, por WhatsApp y teléfono; propuesta económica, seguimiento y cierre.	Volumen, márgenes, segmentos, ventas por sucursal y por vendedor.
Logística	Reparto local e interprovincial, con flota propia o servicio externo.	Pedidos despachados, ubicaciones y medios de transporte.

La organización prioriza la optimización de los tiempos de entrega y los precios preferenciales para cuentas clave, y gestiona a diario la cadena de abastecimiento para sostener un inventario óptimo que permita reaccionar con oportunidad. Su comunicación es

lineal, con política de puertas abiertas, y todo el personal es consciente de su papel dentro del proceso.

1.3 Objetivo del proyecto, vinculación y fuentes de información

Objetivo. Implementar herramientas de Big Data en las áreas de la empresa.

Actualmente PIGE solo dispone de datos duros provenientes de su sistema de información y del servicio contable externo; el proyecto busca aplicar el conocimiento de Big Data para analizar esa información de forma amigable, confiable y gráfica, de modo que la empresa sea más eficiente y decida mejor.

Vinculación. Uno de los integrantes del equipo se desempeña como Gerente de PIGE y mantiene un vínculo de doce años con la organización, por lo que conoce a todo el personal, las áreas y los procesos; valora su nivel de conocimiento de la empresa en 10 sobre 10. Esto facilita el acceso a la información y la validación de supuestos.

Fuentes. SitacPlus registra todas las transacciones comerciales y contables e integra el maestro de artículos, las bases de proveedores y clientes y el plan de cuentas. También se dispone de los estados financieros. El giro del negocio puede consultarse en www.pige.com.ec.

1.4 Valor del Big Data para un distribuidor de EPP

Para un distribuidor de EPP, el Big Data permite anticipar picos de demanda asociados a cambios normativos, optimizar el stock crítico y personalizar los contratos de suministro industrial. Al cruzar datos de siniestralidad, caducidad y empleo, el negocio deja de vender de forma reactiva y se convierte en un socio estratégico de seguridad para sus clientes. Las ventajas se agrupan en tres frentes:

- **Gestión de inventario y logística.** Alertas sobre faltantes de productos críticos, como respiradores y arneses, antes de que el mercado los agote; control automático de las fechas de vencimiento de lotes de cascos, líneas de vida y filtros químicos; y planificación de rutas de entrega según los turnos de planta de los clientes industriales.
- **Fidelización y venta B2B inteligente.** Alertas de reposición calculadas con la tasa histórica de uso de cada cliente (por ejemplo, guantes o calzado); venta cruzada técnica, como recomendar protección auditiva según la maquinaria recién adquirida; y tarifas dinámicas por volumen de consumo real y no por estimaciones.
- **Cumplimiento normativo y análisis de mercado.** Adaptación anticipada a los cambios en la normativa de prevención de riesgos para abastecer el equipo obligatorio antes que la competencia; uso de datos de siniestralidad sectorial para ofrecer productos que mitiguen los riesgos más frecuentes de cada cliente; y evaluación del cumplimiento y los tiempos de entrega de los fabricantes para privilegiar a los más confiables.

Estas oportunidades tienen un carácter de horizonte: su aprovechamiento pleno depende de contar con datos externos y de calidad, lo que se aborda en los capítulos 6 a 9 (Gandomi y Haider, 2015; Chen et al., 2012).

Capítulo 2. Fuentes de datos, KPI y diagnóstico de carencias (Semana 2)

2.1 *Objetivo y alcance*

Hoy la información de PIGE se emplea con un enfoque operativo y táctico: las ventas registradas, los niveles de inventario y las compras realizadas responden consultas puntuales, pero no ofrecen información agregada ni gráfica para el análisis. El proyecto propone evolucionar hacia cuatro tipos de producto informativo:

- **Informes corporativos:** tableros e indicadores que apoyen las decisiones de Gerencia.
- **Informes de procesos:** reportes detallados por área operativa (Compras, Bodega, Ventas y Logística).
- **Informes comerciales:** análisis desde la perspectiva del cliente, como segmentación y rentabilidad por canal.
- **Analítica y minería de datos (a futuro):** modelos predictivos de demanda y detección de patrones de compra.

Este capítulo cubre el levantamiento de las fuentes de datos, la definición de los KPI por proceso y la identificación de carencias en la carga de información. La implementación técnica del modelo de datos y de los tableros se desarrolla en los capítulos posteriores.

2.2 *Consolidación y origen de la información*

PIGE no cuenta con un proyecto formal de Data Warehouse ni con una dirección responsable de las políticas de información. Toda la información transaccional, comercial y contable, reside en SitacPlus, alojado en el servidor de la empresa, sin una función que gobierne la calidad y la estructura de los datos. Esta ausencia de gobierno se retoma en la sección 2.6. La extracción desde SitacPlus hacia hojas de cálculo se hace manualmente, sin un proceso automatizado de transferencia.

TABLA 4*FUENTES DE DATOS Y RECURRENCIA DE CARGA PROPUESTA*

Fuente	Tipo	Contenido	Recurrencia
SitacPlus (servidor de la empresa)	Interna	Transacciones comerciales y contables, maestro de artículos, proveedores, clientes, plan de cuentas	Diaria, al cierre de operaciones
Hojas de ruta de Logística	Interna	Pedidos despachados, rutas y medios de transporte	Diaria
Página web pige.com.ec	Interna	Solicitudes de contacto y cotización	Semanal
Superintendencia de Compañías	Externa	Estados financieros de mercado y competencia	Anual
Sitios web de la competencia	Externa	Precios y portafolio de productos	Mensual o bajo demanda

2.3 Destino de la carga y datamarts

Como destino se proponen hojas de cálculo para las consolidaciones intermedias y la validación de datos, y Power BI como capa de visualización final, con conexión directa o mediante archivos planos exportados desde SitacPlus, según lo permita el sistema (aspecto por validar con el proveedor del software). El único repositorio existente es el servidor de SitacPlus, sin segmentación. Se propone organizar la información en datamarts lógicos por proceso, alimentados desde la misma fuente:

- **Compras:** volúmenes y tiempos de reabastecimiento.
- **Bodega e Inventarios:** rotación, exactitud y ciclos de despacho.
- **Ventas:** desempeño por vendedor, sucursal y línea de producto.
- **Logística:** cumplimiento de entregas y costos de transporte.

TABLA 5*TIPOS DE DATOS REQUERIDOS POR ÁREA*

Área	Datos requeridos
Stocks y Bodega	Unidades por artículo al inicio y al final del período, unidades vendidas, resultados de conteos físicos, tiempos de preparación y despacho.
Compras	Monto y volumen de compras locales e importadas, tiempos de entrega por pedido, pedidos no satisfechos.
Ventas	Unidades y precio por producto, vendedor y sucursal; ingresos y costos de producción; cotizaciones emitidas frente a cerradas.
Logística	Pedidos entregados a tiempo, costo de transporte por pedido, número total de pedidos.

2.4 Indicadores clave de desempeño (KPI)

Los indicadores se definieron por proceso, con su fórmula y sus variables.

Constituyen la base de los tableros y, más adelante, de las variables de entrada del modelo predictivo del capítulo 9 (Parmenter, 2015).

TABLA 6*KPI DE STOCKS Y BODEGA*

KPI	Fórmula	Variables
Inventario promedio	$IP = (U_i + U_f) / 2$	U _i : unidades al inicio del período; U _f : unidades al final.
Rotación trimestral del artículo	$RA = UV / IP$	UV: unidades vendidas; IP: inventario promedio del artículo.
Exactitud de inventario (ERI)	$ERI = (AC / TAV) \times 100$	AC: artículos correctos; TAV: total de artículos verificados.
Tiempo de ciclo de preparación de pedidos	$TCPP = TT / UPP$	TT: tiempo total de trabajo; UPP: unidades o pedidos procesados.

TABLA 7*KPI DE COMPRAS*

KPI	Fórmula	Variables
% de compras locales	$\%CL = CL / (CL + CI) \times 100$	CL: compras locales; CI: compras importadas.
% de importaciones	$\%CI = CI / (CL + CI) \times 100$	CL: compras locales; CI: compras importadas.
Tiempo promedio de reabastecimiento	$TPR = \Sigma TEi / NP$	ΣTEi : suma de los tiempos de entrega de cada pedido, en días; NP: número total de pedidos.
Índice de rotura de stock	$IRS = (PNS / PT) \times 100$	PNS: pedidos no satisfechos; PT: pedidos totales.

Nota. La fórmula del IRS se corrigió respecto de la entrega original, que dividía entre 100 en lugar de multiplicar, lo que no expresa el indicador como porcentaje.

TABLA 8*KPI DE VENTAS*

KPI	Fórmula	Variables
Venta por vendedor	$VVe = CPv \times Pu$	CPv: cantidad de productos vendidos por el vendedor; Pu: precio por unidad.
Venta por sucursal	$VSu = CPs \times Pu$	CPs: cantidad de productos vendidos por la sucursal; Pu: precio por unidad.
Margen bruto por línea de producto	$MB = (I - CP) / I \times 100$	I: ingresos; CP: costos de producción.
Tasa de conversión	$TC = VC / CE \times 100$	VC: ventas cerradas; CE: cotizaciones emitidas.

Nota. Hoy la Gerencia solo conoce el margen bruto global ($\approx 30\%$); calcular el margen por línea es una tarea previa a la medición de este KPI.

TABLA 9*KPI DE LOGÍSTICA*

KPI	Fórmula	Variables
Tasa de entrega a tiempo	$TET = PET / TPE \times 100$	PET: pedidos entregados a tiempo; TPE: total de pedidos entregados.

KPI	Fórmula	Variables
Costo de transporte por pedido	$CTP = CTT / NP$	CTT: costo total de transporte; NP: número total de pedidos.

2.5 Informes actuales, necesidades y soluciones propuestas

Para cada área se contrastó el informe disponible con la necesidad detectada y con la brecha que aún no está cubierta. Siguiendo la recomendación del docente de continuar analizando cómo cubrir las necesidades, la última columna propone la solución con la que el proyecto responde a cada brecha y el capítulo donde se desarrolla.

TABLA 10

INFORMES ACTUALES, NECESIDADES DETECTADAS Y SOLUCIÓN PROPUESTA

Área	Informe actual	Necesidad detectada	Brecha no cubierta	Solución propuesta
Compras	Registro de compras locales e importadas en SitacPlus	Reabastecimiento basado en el histórico y no en el criterio manual	Modelo predictivo de punto de reorden	Modelo predictivo de reabastecimiento y semáforo de alertas (cap. 9)
Bodega	Registro de recepción y despacho	Visibilidad en tiempo real de la rotación por línea	Indicador automatizado de exactitud de inventario (ERI)	Tablero operativo de rotación y ERI alimentado por el datamart de Bodega (caps. 3 y 4)
Ventas	Registro de ventas por transacción	Rentabilidad por canal y segmento	Segmentación avanzada de clientes	Segmentación con datos internos y externos; tablero de margen por línea (caps. 4 y 6)
Logística	Registro de pedidos despachados	Optimización de rutas y tiempos	Indicador consolidado de costo por ruta y transportista	Sistematización de tarifas de transporte y KPI de costo por pedido (cap. 3)

2.6 Carencias identificadas en la carga y gestión de datos

TABLA 11

CARENCIAS DEL PROCESO DE INFORMACIÓN Y SU IMPLICACIÓN

Carencia	Implicación para el negocio
Alertas de stock basadas en criterio manual	Las alertas dependen del Líder de Bodega y no de un modelo con el histórico de rotación, lo que produce reabastecimientos tardíos o excesivos.
Ausencia de recurrencia de carga definida	No existe una rutina documentada de actualización (frecuencia, hora de corte, consolidación), por lo que la información disponible puede estar desfasada.
Repositorio único sin segmentación	Toda la información reside en un solo sistema, sin datamarts por proceso, lo que dificulta el análisis independiente de cada área.
Sin gobierno de datos ni dirección responsable	Nadie define políticas de calidad, estructura de archivos o criterios de ingestión; en organizaciones maduras esta función la asume una dirección de capacidades analíticas.
Fuentes externas sin proceso sistematizado	La consulta a la Superintendencia de Compañías y a los sitios de la competencia es manual y sin periodicidad definida.
Dependencia de hojas de cálculo	Sin capa semántica ni proceso ETL, la consolidación recae en planillas, con mayor riesgo de error humano y menor trazabilidad.
Falta de historial y trazabilidad	Se consulta el estado presente, pero no hay un mecanismo de conservación de históricos para análisis de tendencia o modelos predictivos.
Ausencia de identificador único transversal	Sin una llave común de cliente o de producto no es posible relacionar de forma consistente ventas, bodega, compras y logística.
Ausencia de márgenes por línea de producto	La Gerencia conoce solo el margen bruto global ($\approx 30\%$); sin margen por línea no puede medirse la rentabilidad de cada una de las once líneas.

2.7 Componentes de Business Intelligence aplicados a PIGE

Un sistema de BI combina datos, tecnología y procesos organizacionales para convertir la información en decisiones (Wixom y Watson, 2010). La tabla traduce cada componente a la realidad de PIGE.

TABLA 12*COMPONENTES DE BI EN PIGE*

Componente	Aplicación en PIGE
Problemática empresarial	Falta de información amigable, confiable y gráfica; solo se dispone de datos duros del sistema operacional (SitacPlus) y del servicio contable externo.
Equipo de análisis	Equipo 3 del Programa MBA, responsable de levantar el proyecto de BI. Los consumidores de la información son el docente y la Gerencia de la empresa.
Información de pedidos y expediciones	Registros de ventas, despacho y logística en SitacPlus y en las hojas de ruta del departamento logístico.
Información externa de tarifas de transporte	Hoy no está sistematizada. Debe incorporarse (flota propia y servicio contratado) para calcular con fiabilidad el costo de transporte por pedido; es una fuente externa adicional a las de la sección 2.2.
Data Warehouse	No existe aún; se construirá a partir de los datamarts lógicos de la sección 2.3.
Aplicación de BI	Power BI, junto con hojas de cálculo, como herramienta de generación, análisis y visualización de informes.

2.8 Protocolo de implantación

Todo proyecto de BI requiere un protocolo ordenado que pueda repetirse en iniciativas futuras. El proyecto de PIGE lo adopta en seis pasos, que además estructuran este documento:

- **Diagnosticar** la situación, los procesos y las carencias (capítulos 1 y 2).
- **Definir** los KPI y las fuentes de datos que los alimentan (capítulo 2).
- **Diseñar** el flujo ETL, la calidad de datos y la infraestructura (capítulo 3).
- **Visualizar:** definir qué se muestra, cómo y a quién (capítulo 4).
- **Organizar** el área responsable, sus roles y su presupuesto (capítulo 5).
- **Ampliar y explotar:** incorporar fuentes externas, arquitectura y analítica avanzada (capítulos 6 a 9).

Capítulo 3. Proceso ETL e infraestructura de datos (Semana 3)

3.1 Objetivo y alcance

Este capítulo diseña el proceso de extracción, transformación y carga (ETL) que consolidará la información dispersa de PIGE, hoy concentrada en SitacPlus, las hojas de ruta logísticas y fuentes externas, en los datamarts de Compras, Bodega, Ventas y Logística, como paso previo a la construcción de tableros. Cubre la definición conceptual del proceso, la calidad de la información, los horarios de carga, las variables y el acceso al futuro Data Warehouse, los sistemas implicados y los recursos, con la cuantificación de la infraestructura. El desarrollo técnico de scripts y flujos automatizados queda fuera de este alcance.

3.2 Proceso ETL

- **Extracción.** Al cierre de operaciones se extrae de SitacPlus la información transaccional (ventas, compras y movimientos de bodega) mediante exportación a archivo plano o consulta directa, según lo permita el sistema. En paralelo se recopilan las hojas de ruta de Logística y, con menor frecuencia, las tarifas de transporte y las fuentes externas.
- **Transformación.** Los archivos se limpian y estandarizan (formatos de fecha, unidades, códigos de artículo y de cliente), se valida su exactitud y se calculan los KPI definidos en el capítulo 2, utilizando hojas de cálculo con Power Query como capa intermedia.
- **Carga.** La información transformada se carga en los datamarts lógicos, que en una primera etapa residirán en archivos estructurados de referencia y, posteriormente, en un repositorio centralizado (data lake o Data Warehouse) que alimente directamente a Power BI.

PIGE no cuenta con una zona de aterrizaje (Landing Zone) ni con una interfaz automatizada de recolección; el proceso propuesto es el punto de partida para migrar de un manejo manual y disperso a un flujo estructurado y repetible. Conceptualmente, las fuentes internas y externas alimentan una capa de transformación que habilita las capacidades de reportería, indicadores y toma de decisiones (Kimball y Ross, 2013).

3.3 Calidad de la información disponible

TABLA 13

CALIDAD ACTUAL DE LAS FUENTES DE DATOS

Fuente	Calidad actual	Observación
SitacPlus (transacciones)	Media-alta	Datos completos y actualizados, pero sin identificador único transversal entre módulos.
Hojas de ruta de Logística	Media	Se registran en formato físico o en Excel no estandarizado; requieren digitalización consistente.
Tarifas de transporte	Baja	No están sistematizadas; dependen de cotizaciones puntuales de cada transportista, sin histórico consolidado.
Superintendencia de Compañías y competencia	Baja-media	Información pública con actualización manual y sin periodicidad garantizada.

La información interna transaccional tiene buena calidad de origen, pero exige estandarización antes de integrarse a un ETL confiable. Las fuentes externas y logísticas son las que demandarán mayor esfuerzo de depuración.

3.4 Horarios de carga y accesibilidad

TABLA 14

HORARIOS DE CARGA PROPUESTOS

Fuente	Horario	Frecuencia	Accesibilidad
SitacPlus (transacciones)	Cierre de operaciones, 18:00	Diaria	Jefatura Administrativa y Financiera
Hojas de ruta de Logística	Cierre de despacho, 19:00	Diaria	Líder de Bodega y despachadores
Tarifas de transporte	Primer día hábil del mes	Mensual	Jefatura Administrativa
Fuentes externas	Fin de trimestre	Trimestral	Gerencia General

La consolidación hacia los datamarts se realizaría el día siguiente a cada corte, de forma manual en esta etapa, con la meta de automatizar a mediano plazo la carga diaria de SitacPlus y de Logística.

3.5 Variables disponibles y gobierno de acceso

TABLA 15

VARIABLES POR DATAMART Y RESPONSABLE DE AUTORIZAR EL ACCESO

Datamart	Variables disponibles	Autoriza el acceso
Compras	Monto y volumen de compras locales e importadas, tiempos de entrega, pedidos no satisfechos, proveedor	Jefatura Administrativa
Bodega	Unidades por artículo, conteos físicos, tiempos de preparación y despacho, código de artículo	Líder de Bodega
Ventas	Unidades y precio por producto, vendedor y sucursal; cotizaciones; costos de producción	Gerencia de Ventas
Logística	Pedidos despachados, rutas, medios de transporte, tarifas, tiempos de entrega	Jefatura Administrativa

Mientras no exista un rol dedicado a la gobernanza de datos, se propone que la Gerencia General autorice el acceso consolidado a los cuatro datamarts y que cada jefatura administre el acceso a su propia área.

3.6 Sistemas y recursos

- **Sistemas:** SitacPlus como origen transaccional (extracción); hojas de cálculo con Power Query para limpieza y cálculo de KPI (transformación); Power BI para carga final y visualización; y correo electrónico y WhatsApp, canales actuales de intercambio de tarifas que deberán formalizarse.
- **Recursos internos:** la asesoría informática, para la extracción desde SitacPlus y la configuración de conexiones; y el equipo del proyecto, para el diseño del ETL, los KPI y los tableros.
- **Recursos externos:** el proveedor de SitacPlus, para habilitar exportaciones automatizadas o accesos directos a la base de datos; los transportistas, como fuente de tarifas; y el servicio contable externo, para el plan de cuentas y los estados financieros.

3.7 Dimensionamiento de la infraestructura y su costo

Una infraestructura de datos bien dimensionada debe indicar dónde se depositan, custodian y circulan los datos, y cuánto cuesta sostenerla. Dado el volumen de PIGE (un histórico de ventas de tres años y un catálogo de artículos, en el orden de megabytes), la primera etapa no requiere servidores nuevos: los datos se depositan en un repositorio en la nube (OneDrive o SharePoint), se procesan con herramientas ya instaladas y se publican en Power BI Service. La tabla cuantifica lo que puede estimarse hoy.

TABLA 16*DIMENSIONAMIENTO Y COSTO REFERENCIAL DE LA INFRAESTRUCTURA, PRIMERA ETAPA*

Componente	Dimensionamiento	Costo mensual (USD)	Costo anual (USD)
Depósito de datos (Zona de Aterrizaje)	Repositorio en la nube con pocos gigabytes; se asume suscripción Microsoft 365 vigente	0 incrementa 1	0
Procesamiento y modelado	Equipos de cómputo existentes con Power Query y Power BI Desktop (gratuito)	0	0
Visualización y distribución	Tres licencias Power BI Pro (Gerencia, Compras y analítica), a USD 14 por usuario al mes	42	504
Conexión con SitacPlus	Exportaciones automatizadas o puerta de enlace; depende del proveedor	Por cotizar	Por cotizar
Total cuantificado		42	504

Nota. Supuestos del equipo, por validar con la asesoría informática y el proveedor de SitacPlus. El costo de conexión se incorpora en el capítulo 8 como horas de asesoría.

Capítulo 4. Plan de visualización (Semana 4)

Antes de construir cualquier tablero es indispensable un plan de visualización que responda cuatro preguntas: qué información mostrar, cómo hacerlo, a quién va dirigida y con qué objetivo. Un plan explícito evita reportes genéricos y asegura que la inversión se traduzca en acciones concretas (Few, 2013; Knaflitz, 2015).

4.1 ¿Qué información se mostrará?

PIGE dispone de información dura de SitacPlus (maestro de artículos, proveedores, clientes y plan de cuentas) y de su servicio contable externo. El plan prioriza los datos que hoy generan valor operativo y estratégico en sus cuatro procesos, con el fin de explicar las relaciones de causa y dependencia entre compras, inventario, ventas y logística, y no de exhibir todo lo disponible:

- **Compras:** volúmenes de adquisiciones locales e importaciones, niveles de stock mínimo y alertas de reabastecimiento.
- **Bodega:** rotación de inventario por línea de producto.
- **Ventas:** volúmenes, márgenes, segmentación de clientes, ventas por sucursal y por vendedor en los ejercicios 2021 a 2023.
- **Logística:** pedidos despachados, ubicaciones de entrega y medios de transporte, con flota propia o servicio externo.

4.2 ¿Cómo se mostrará?

Cada tipo de dato admite una representación más intuitiva que otras. Power BI permitiría construir un ecosistema de cuadros de mando diferenciados, en la línea del tablero de comando integral (Kaplan y Norton, 1996):

- **Operacional, Bodega y Logística:** gráficos de barras para la rotación de inventario y mapas para visualizar rutas y zonas de despacho nacionales.
- **Táctico, Ventas:** gráficos de tendencia (evolución 2021-2023), dispersión de volumen frente a margen (cuando exista margen por línea) y segmentación por sucursal y vendedor.
- **Compras:** indicadores tipo semáforo para los niveles de stock mínimo, que señalen visualmente la necesidad de reabastecer.
- **Gerencia:** tarjetas de resultados con crecimiento de ventas, margen consolidado y cumplimiento de despachos.

4.3 ¿A quién va dirigida?

PIGE es una empresa familiar de 16 colaboradores con comunicación lineal, lo que facilita definir audiencias claras. El nivel de detalle, el lenguaje visual y la complejidad de cada tablero deben ajustarse al perfil del receptor: la jefatura de bodega no necesita la vista analítica de la gerencia, ni a la inversa. El docente planteó si todos los grupos verían los mismos KPI; la respuesta del equipo es que no, y la tabla siguiente precisa qué indicadores y con qué periodicidad ve cada audiencia.

TABLA 17

MATRIZ DE AUDIENCIAS, KPI Y PERIODICIDAD

Audiencia	KPI que consulta	Periodicidad y comparativo
Socios y Gerencia General	Crecimiento de ventas, margen bruto consolidado (y por línea, cuando se calcule), tasa de conversión, cumplimiento de despachos, índice de rotura de stock	Mensual y acumulado del año, frente al presupuesto y al mismo período del año anterior
Jefatura de Bodega	Rotación por línea, exactitud de inventario (ERI), tiempo de ciclo de preparación, alertas de stock mínimo	Diaria y semanal

Audiencia	KPI que consulta	Periodicidad y comparativo
Compras e Importaciones	% compras locales e importadas, tiempo promedio de reabastecimiento, índice de rotura, ranking de productos de mayor rotación	Semanal y mensual; acumulado frente al presupuesto
Equipo de Ventas	Venta por vendedor y sucursal, tasa de conversión, ranking de productos, pedido mínimo y pedido promedio	Mensual y acumulado frente al presupuesto y al año anterior
Logística y flota	Entrega a tiempo, costo de transporte por pedido, variación del número de envíos	Semanal y mensual
Personal en general	Tarjeta resumen: ventas acumuladas, cumplimiento de despachos y nivel de servicio	Mensual

4.4 ¿Cuál es el objetivo de la información?

El objetivo general es pasar de datos duros dispersos en el sistema contable a información accionable, confiable y visual. Específicamente, los tableros buscan:

- mejorar las decisiones de reabastecimiento, reduciendo quiebres y excesos de inventario;
- explicar el comportamiento comercial 2021-2023 para sostener o acelerar el crecimiento;
- optimizar los tiempos de entrega y priorizar las cuentas clave, en línea con la visión de ser un socio estratégico; y
- reemplazar criterios manuales y subjetivos, como la definición actual del stock mínimo, por alertas basadas en datos.

La información no es un fin en sí misma: debe traducirse en acciones de compras, bodega, ventas y logística, entregada a la persona adecuada, en el formato adecuado y en el momento oportuno.

4.5 Ampliación del catálogo de KPI y período de análisis

A partir de la retroalimentación recibida, el catálogo del capítulo 2 se complementa con tres indicadores de gestión comercial y logística y con una regla explícita sobre el período de análisis.

TABLA 18

KPI COMPLEMENTARIOS

KPI	Definición	Uso
Ranking de productos de mayor rotación	Top 10 de artículos por unidades vendidas, por línea de producto (y por margen cuando se disponga de él)	Priorizar el stock crítico y orientar la negociación con proveedores
Variación del número de envíos	$(\text{Envíos del período} - \text{envíos del período comparado}) / \text{envíos del período comparado} \times 100$	Detectar crecimiento o contracción de la operación logística
Pedido mínimo y pedido promedio	Menor valor facturado por pedido y valor promedio por pedido	Evaluar la rentabilidad de pedidos pequeños y definir condiciones comerciales

Período de análisis. En Compras y Ventas, los indicadores se calcularán desde el inicio del año (acumulado) y desde el inicio del mes, y se compararán contra el presupuesto y contra el mismo período del año anterior. Esto permite distinguir un desvío puntual de una tendencia.

4.6 Resumen del plan de visualización

TABLA 19*PLAN DE VISUALIZACIÓN DE PIGE*

Pregunta	Respuesta
¿Qué información?	KPI de Compras, Ventas, Bodega y Logística, centralizados en el Data Warehouse, con los indicadores específicos de cada audiencia (véase la matriz de audiencias).
¿Cómo mostrarla?	Dashboards con gráficos, mapas, alarmas y tarjetas en Power BI; al estar en línea, permiten presentar escenarios actuales y proyecciones.
¿A quién?	Socios, Gerencia, jefaturas y personal en general.
¿Con qué objetivo?	Socios: decisiones más rápidas y basadas en datos. Gerencia: formular estrategias para alcanzar los objetivos. Jefaturas: ejecutar los planes que garanticen la misión y la visión. Personal: conocer la situación de la empresa para aportar soluciones proactivas o sostener una gestión efectiva si los KPI son adecuados.

Capítulo 5. Área de analítica y caso de negocio inicial (Semana 5)

5.1 Posicionamiento y objetivo del área

Con 16 colaboradores, PIGE no puede sostener una gerencia de analítica de dedicación completa. Se propone que la función se posicione como un rol transversal que reporta directamente a la Gerencia General, junto a las demás gerencias, apoyado en la Asesoría Informática existente para la administración tecnológica y en un consultor externo para la construcción inicial de modelos y tableros. Así la analítica tiene visibilidad sobre Compras, Bodega, Ventas y Logística sin crear una gerencia intermedia.

Su objetivo es diseñar, construir y administrar los modelos analíticos e indicadores de gestión de PIGE, utilizando las fuentes, los KPI y el ETL ya definidos, para comprender el comportamiento de clientes, productos, proveedores y canales, y tomar decisiones que mejoren la rentabilidad y la eficiencia y reduzcan el riesgo de quiebres de stock.

5.2 Estructura y responsabilidades

TABLA 20

ESTRUCTURA DEL ÁREA DE ANALÍTICA

Rol	Función	Perfil
Responsable de Analítica (función ampliada)	Liderar el diseño, construcción y administración de indicadores y tableros a partir de los datamarts; comunicar hallazgos a la Gerencia General	Administrativo o financiero con manejo avanzado de Excel y Power BI
Asesoría Informática (existente)	Administrar la conexión entre SitacPlus y las herramientas de BI, garantizando seguridad y estabilidad	Ingeniero de sistemas (asesoría externa)
Consultor externo de Power BI (bajo demanda)	Construir la versión inicial de los tableros por datamart y transferir el conocimiento para que el Responsable los mantenga de forma autónoma	Consultor independiente o practicante de BI

La función de gerencia de analítica se cubre así con un rol interno de dedicación parcial y apoyo externo puntual, sin incrementar la planta fija.

5.3 Impacto esperado y presupuesto

Tomando como referencia el quiebre de stock (pedidos no satisfechos por falta de inventario), se estima que PIGE opera hoy con un 6 % aproximado, al depender de un criterio manual de reabastecimiento. Con un modelo basado en el histórico de rotación se proyecta reducirlo a 2,5 % en un año. Sobre las ventas de 2023 (USD 1,63 millones), cada punto porcentual equivale a unos USD 16.300 anuales en ventas recuperadas; los 3,5 puntos de mejora representarían USD 57.050 al año que hoy se pierden por falta de inventario. Con el margen bruto global de la empresa, de alrededor de 30 % (la Gerencia no dispone de márgenes por línea), eso equivale a unos USD 17.115 de margen anual.

TABLA 21

PRESUPUESTO ANUAL DE CREACIÓN DEL ÁREA DE ANALÍTICA

Categoría	Recurso	Costo mensual (USD)	Cantidad	Presupuesto anual (USD)
Labor (dedicación parcial)	Responsable de Analítica, 30 % de dedicación	300	1	3.600
Labor externa	Consultor externo de Power BI (bolsa de horas)	250	1	3.000
Materiales	Licencias Power BI Pro (USD 14 por usuario)	42	3 usuarios	504
Total				7.104

Frente a un costo anual recurrente de USD 7.104, las ventas recuperadas estimadas ascienden a USD 57.050, equivalentes a unos USD 17.115 de margen bruto al 30 %. Los

capítulos 9 y 10 valoran este resultado sobre margen y no sobre ventas, y lo someten a un análisis de sensibilidad.

5.4 Impacto en las iniciativas propuestas

Iniciativa 1: modelo predictivo de reabastecimiento. Hoy el Líder de Bodega decide el reabastecimiento con criterio manual y una efectividad estimada de 75 % de pedidos generados a tiempo. Con un modelo basado en la rotación histórica por artículo se proyecta elevarla a 90 %.

Iniciativa 2: consolidación de informes de gestión. La consolidación mensual de reportes se realiza hoy de forma manual en hojas de cálculo independientes. Con tableros automatizados se estima el siguiente ahorro:

TABLA 22

AHORRO EN LA CONSOLIDACIÓN DE INFORMES

Concepto	Hoy	Con el área de analítica
Costo por hora (asistente administrativo)	USD 4,85	USD 4,85
Horas mensuales de consolidación	20	6
Costo mensual	USD 97,00	USD 29,10
Costo anual	USD 1.164,00	USD 349,20
Ahorro anual estimado		USD 814,80

5.5 Gestión del cambio y adopción

Un proyecto bien diseñado solo tiene éxito si las personas que usan las herramientas están instruidas y motivadas. Buena parte de esa disposición depende de que cada usuario vea las ventajas del cambio para su propio trabajo y de que las herramientas sean intuitivas y fáciles de usar, algo que cualquiera advierte al enfrentarse a una aplicación nueva en su

oficina o a un cajero automático. La teoría de aceptación tecnológica sostiene que la utilidad y la facilidad de uso percibidas son los principales determinantes de la adopción (Davis, 1989), y la gestión del cambio exige un patrocinio visible y logros tempranos (Kotter, 1996). PIGE aplicará cuatro medidas:

- **Beneficio para cada rol.** Mostrar a cada usuario qué problema de su trabajo resuelve el tablero: alertas de reabastecimiento para Bodega y Compras, rankings y conversión para Ventas, costos por pedido para Logística.
- **Diseño sencillo.** Tableros con pocos indicadores por pantalla, semáforos y lenguaje del negocio, validados con los usuarios antes de su lanzamiento.
- **Formación práctica.** Capacitación breve y por rol, con acompañamiento del Responsable de Analítica durante las primeras semanas (su costo se estima en el capítulo 8).
- **Patrocinio y escucha.** La Gerencia General comunica el propósito del cambio y recoge las observaciones de los usuarios para mejorar los tableros.

5.6 Estructura del trabajo y valoración

Implantar un área de analítica no exige una gerencia dedicada, sino una función liderada por una persona con dedicación parcial, con apoyo de la Asesoría Informática y de un consultor puntual. El caso de negocio muestra una relación costo-beneficio favorable, pero depende de que las decisiones se tomen sobre datos bien interpretados; de lo contrario se reproducirían los problemas actuales: reabastecimientos tardíos, exceso de inventario y clientes insatisfechos. Por ello, el siguiente paso es validar las proyecciones con datos reales antes de formalizar la inversión.

PARTE II: DATOS EXTERNOS, ARQUITECTURA TECNOLÓGICA E INTELIGENCIA ARTIFICIAL

La segunda parte lleva el modelo de BI más allá de los datos que PIGE ya genera. Incorpora fuentes externas y no estructuradas bajo el marco legal ecuatoriano, define dónde y cómo se centralizarán los datos, selecciona las herramientas de código abierto proporcionales a su volumen y, finalmente, aplica ciencia de datos al problema de negocio de mayor valor, el reabastecimiento, y la complementa con un detector en tiempo casi real y con la valoración financiera integral del proyecto.

Capítulo 6. Obtención de datos externos y no estructurados (Semana 6)

6.1 Objetivo y alcance

Además de los datos de SitacPlus y de la información contable externa, el modelo de BI necesita fuentes adicionales que mejoren la perfilación de clientes, la gestión de cobranza, la planificación de compras y la eficiencia logística. Se identifican dos grandes orígenes: la compra de datos a entidades públicas y privadas del Ecuador y los datos que PIGE ya genera en sus propios canales, pero de forma desorganizada.

6.2 Marco legal: protección de datos personales en Ecuador

PIGE debe registrarse por la Ley Orgánica de Protección de Datos Personales (LOPDP), aprobada por la Asamblea Nacional el 10 de mayo de 2021 y publicada en el Registro Oficial el 26 de mayo de 2021. La norma alcanza a toda persona natural o jurídica, pública o privada, que trate datos personales en el territorio ecuatoriano, y concedió dos años de adecuación. Seis principios condicionan el proyecto:

- **Juridicidad, lealtad y transparencia:** todo tratamiento de datos de clientes, proveedores y prospectos debe ajustarse a la ley y comunicarse con claridad al titular.

- **Consentimiento previo, expreso e informado:** salvo las excepciones legales, solo se tratarán datos con autorización del titular.
- **Finalidad y minimización:** los datos comprados o recopilados deben servir a un propósito legítimo y explícito, evitando recolectar más de lo necesario.
- **Calidad y conservación:** la información usada en los modelos debe ser veraz, completa, exacta y actualizada.
- **Responsabilidad proactiva y demostrada:** PIGE debe poder acreditar en todo momento que ha implementado mecanismos efectivos de protección.
- **Aplicación favorable al titular:** ante la duda, la norma se interpreta en el sentido más favorable al dueño del dato.

En términos prácticos, antes de incorporar cualquier dato externo a los datamarts, el equipo verificará que la fuente cumpla la LOPDP y que PIGE cuente con una política de tratamiento de datos personales que informe a clientes y proveedores sobre la finalidad y sobre sus derechos de acceso, rectificación y eliminación (Asamblea Nacional del Ecuador, 2021).

6.3 Fuentes externas

PIGE vende principalmente a empresas (constructoras, industrias y distribuidores) y, en menor medida, a consumidores, y lo hace a crédito, por lo que el riesgo de cartera es relevante. Comprar información externa persigue dos fines: mejorar la perfilación y el scoring de riesgo de los clientes empresariales, para definir condiciones de crédito y anticipar la mora, y validar y enriquecer los datos de actividad económica y contacto de prospectos y proveedores.

TABLA 23*ENTIDADES ECUATORIANAS QUE PODRÍAN PROVEER INFORMACIÓN*

Entidad	Naturaleza	Datos relevantes para PIGE
Equifax Ecuador	Buró de crédito privado autorizado	Score crediticio (1-999) de clientes empresariales, historial de mora, operaciones de crédito vigentes y comportamiento de pago.
Aval Buró	Buró de crédito privado alternativo	Reporte complementario para contrastar el score y evitar castigar a un cliente por datos incompletos de una sola fuente.
Superintendencia de Compañías, Valores y Seguros	Entidad pública de control societario	Estado legal de la empresa (activa, disuelta, en liquidación), representante legal, accionistas y estados financieros de clientes de mayor volumen.
Servicio de Rentas Internas (SRI)	Entidad pública tributaria	Validación del RUC, estado tributario y tipo de contribuyente, útiles para depurar la base de clientes y proveedores.
Cámaras de Comercio (Quito y Guayaquil)	Entidades gremiales privadas	Actividad económica y sector de las empresas afiliadas, para segmentar prospectos de construcción e industria.

En Ecuador operan solo dos burós de crédito autorizados, Equifax Ecuador y Aval Buró, que cumplen para PIGE el rol de consolidar el comportamiento de pago de los clientes empresariales frente al sistema financiero y comercial. La información societaria y tributaria puede consultarse en los portales oficiales de la Superintendencia de Compañías, Valores y Seguros (s. f.) y del Servicio de Rentas Internas (s. f.).

6.3.1 Plataformas digitales: Meta y LinkedIn

Meta y LinkedIn no comercializan bases de datos de contactos; hacerlo vulneraría sus términos de servicio y el principio de consentimiento de la LOPDP. Ofrecen, en cambio,

acceso pagado a herramientas de segmentación y prospección sobre su propia audiencia, sin que PIGE llegue a poseer los datos personales en bruto.

TABLA 24

HERRAMIENTAS DE PLATAFORMAS DIGITALES

Herramienta	Qué ofrece realmente	Utilidad para PIGE
Meta Business Suite y Ads Manager	Audiencias segmentadas por ubicación, cargo, intereses y comportamiento; Meta conserva los datos y PIGE paga por el alcance	Campañas dirigidas a responsables de seguridad y salud ocupacional y de compras en constructoras e industrias de Quito y Guayaquil
LinkedIn Sales Navigator y Marketing Solutions	Filtros avanzados sobre perfiles públicos (cargo, sector, tamaño de empresa); no permite descargar bases masivas	Prospección activa de gerentes de compras e industriales para nutrir el embudo de Ventas

Debe evitarse cualquier práctica de extracción masiva y no autorizada de perfiles (scraping): infringe los términos de ambas plataformas y el consentimiento exigido por la LOPDP, por lo que no se considera fuente de datos para el proyecto.

6.3.2 Inteligencia de comercio exterior: Veritrade

Veritrade agrega y estructura los registros aduaneros públicos de importación y exportación, con consulta por producto, empresa (RUC), país de origen, valor CIF, puerto de ingreso y período (Veritrade, s. f.). Para PIGE, cuyo portafolio tiene un componente importado relevante, permite observar el comportamiento de importación de la propia competencia.

TABLA 25*DATOS DE VERITRADE Y SU UTILIDAD*

Dato disponible	Utilidad para PIGE
Empresas ecuatorianas que importan EPP (guantes, cascos, arneses, mascarillas, calzado), identificadas por RUC	Identificar competidores directos e indirectos hoy no mapeados y estimar su volumen por valor CIF acumulado.
País de origen y puerto o frontera de ingreso	Comparar rutas y países de abastecimiento frente a los proveedores de PIGE y evaluar la diversificación para reducir el riesgo de desabastecimiento.
Valor CIF y número de operaciones por período	Estimar precios de referencia de importación y detectar la estacionalidad del sector para mejorar la planificación de Compras.
Perfil histórico de un importador o exportador (hasta cinco años)	Evaluar a un proveedor extranjero potencial antes de negociar, verificando su trayectoria real.

Veritrade cierra una brecha que ninguna otra fuente cubre: mientras los burós, la Superintendencia y el SRI aportan visibilidad sobre el riesgo y la formalidad de los clientes, y Meta y LinkedIn aportan alcance comercial, solo Veritrade entrega inteligencia sobre el proceso de Compras y la estrategia de abastecimiento de la competencia. Al tratarse de registros aduaneros entre personas jurídicas, su riesgo legal bajo la LOPDP es considerablemente menor, lo que la convierte en una fuente prioritaria.

6.4 Datos internos desorganizados

PIGE genera internamente información valiosa que no se centraliza ni se estructura:

- **Llamadas y consultas de asesoría comercial:** consultas telefónicas y por WhatsApp sobre disponibilidad de EPP que no siempre quedan registradas en SitacPlus, con pérdida de contactos y de demanda no atendida.
- **Gestión de cobranza:** llamadas y acuerdos de pago documentados de forma dispersa y sin una causal de mora estandarizada que permita cruzarla con el score externo.
- **Horarios y patrones de pedido:** el momento en que los clientes empresariales piden (antes del cierre de obra o del inicio de turno) no se analiza y podría anticipar quiebres de stock.
- **Incidencias de despacho y transporte:** reclamos y novedades de entrega que se manejan por correo o verbalmente, sin una taxonomía común de causas.

6.5 Solución para estructurar la información

Text Analytics. Para los datos desorganizados se propone usar herramientas de análisis de texto que identifiquen palabras clave (producto solicitado, causal de mora, tipo de incidencia) y perfilen al cliente o al pedido, de modo que la información se consolide de forma estructurada y bajo los criterios del datamart correspondiente.

Captura de datos y ETL. Como estos canales no están conectados al futuro Data Warehouse, se requiere un proceso ETL que garantice su migración adecuada al modelo de BI, respetando en cada etapa los principios de minimización y finalidad de la LOPDP.

La incorporación seguirá cuatro pasos: seleccionar las fuentes externas e internas; relacionarlas con los procesos de Compras, Bodega, Ventas y Logística; definir casos de uso concretos (scoring de riesgo, segmentación de prospectos por sector y predicción de quiebres según patrones de pedido); y justificar el impacto esperado, principalmente en Ventas y Compras.

6.6 Avance técnico en Power BI

Durante la semana 6 el equipo construyó la versión inicial del archivo .pbix de PIGE, que documenta el proceso ETL aplicado a los datos de SitacPlus (Ferrari y Russo, 2016; Microsoft, s. f.):

- **Instalación y creación del informe.** Se instaló Power BI Desktop y se abrió un informe en blanco como punto de conexión de las fuentes.
- **Importación.** Se cargaron desde Excel las dos bases extraídas de SitacPlus: el Maestro de Artículos (catálogo de EPP) y el histórico de Ventas 2021-2023.
- **Transformación en Power Query.** Sobre el Maestro de Artículos se promovieron encabezados, se corrigieron tipos, se eliminaron columnas innecesarias y se filtraron filas. El histórico de ventas exigió una limpieza considerablemente mayor, con más de 35 pasos de reemplazo de valores, cambios de tipo y renombrado de columnas, por la inconsistencia de los datos de origen.
- **Columna condicional de origen.** Se creó una columna que clasifica cada artículo como Importado o Local según su marca (Pro Safety, EPI, DPL, Paredes, Kadesh, Coverguard, HAUK, Tecmater, Forte y Skyfall se consideran importadas; el resto, locales). Esta clasificación conecta el modelo con Veritrade y, más adelante, con el modelo predictivo.
- **Modelo de datos.** Se relacionaron el Maestro de Artículos (tabla de dimensión: marca, categoría, bodega, costo y existencia) y las Ventas 2021-2023 (tabla de hechos: cantidad, cliente, bodega y artículo) con cardinalidad uno a varios, lo que permite analizar ventas por marca, categoría y origen.

Los siguientes pasos son construir las medidas DAX y los primeros visuales (ventas por marca, participación de importados frente a locales y evolución 2021-2023) y, en fases

posteriores, integrar las fuentes externas. El primer paso concreto en materia de datos es definir con cuál buró (Equifax o Aval Buró) se validará a los clientes empresariales de mayor volumen, contratar una prueba con Veritrade y estructurar el registro de llamadas y reclamos internos.

Capítulo 7. Arquitectura de la información: Zona de Aterrizaje (Semana 7)

7.1 Objetivo y situación actual

El objetivo es adaptar el concepto de Zona de Aterrizaje (Landing Zone, LZ) a la realidad de PIGE, definiendo una arquitectura de información de bajo costo y fácil implementación que centralice datos internos y externos. PIGE no busca replicar una arquitectura corporativa de Big Data, sino una versión reducida que sirva de puente entre los datos existentes y las fuentes externas del capítulo 6.

Hoy cada área (Compras, Bodega, Ventas, Cobranza, Logística y Gerencia) accede de forma independiente a distintas fuentes, sin un repositorio único. SitacPlus concentra las ventas y el inventario, pero la información de asesoría comercial, cobranza y logística circula de forma manual, por correo o WhatsApp, sin registro estructurado, y las fuentes externas aún no se conectan a ningún proceso interno.

7.2 Modelo conceptual: Zona de Aterrizaje para PIGE

Se propone una LZ adaptada a la escala de la empresa: un componente centralizado y de bajo costo, alojado en la nube, que reciba tempranamente los datos de SitacPlus, de los archivos de Excel, de los canales no estructurados (llamadas, WhatsApp, correos de cobranza y reclamos de logística) y, en fases posteriores, de las fuentes externas. Su valor es ser el punto único desde el cual se alimenta el modelo de Power BI, evitando que cada área mantenga su propia copia y reduciendo el riesgo de trabajar con información desactualizada o inconsistente. La LZ consolida datos estructurados (SitacPlus, Excel), semiestructurados (archivos de Veritrade, reportes de los burós) y no estructurados (llamadas, WhatsApp, correos), y es la fuente única de los datamarts y del modelo de Power BI, tanto para el uso operativo como para el gerencial (Inmon, 2005).

7.3 Capacidades aplicables

Se omitieron las capacidades propias de arquitecturas de muy alto volumen y velocidad (acelerador SAS, motor NoSQL, procesamiento de flujos en tiempo real, librerías distribuidas y serialización avanzada), que no corresponden al volumen transaccional de PIGE. Las capacidades relevantes son:

TABLA 26

CAPACIDADES DE LA ZONA DE ATERRIZAJE APLICABLES A PIGE

Capacidad	Descripción para PIGE
Almacenamiento de bajo costo	Repositorio en la nube (OneDrive, SharePoint o Azure Blob) que centraliza los archivos de SitacPlus, los libros de Excel y los futuros archivos de Veritrade, sin infraestructura propia.
Integración de archivos planos	Carga de archivos .xlsx y .csv exportados de SitacPlus y Veritrade, como ya se hace en Power Query con el Maestro de Artículos y las Ventas 2021-2023.
Integración de bases de datos	Conexión directa, cuando esté disponible, con la base transaccional de SitacPlus, para evitar la exportación manual y reducir errores.
Motor de consultas tipo SQL	Power Query y DAX sobre los datos almacenados, para responder preguntas como ventas por marca, participación de importados y evolución 2021-2023.
Autenticación y autorización	Control de acceso por roles (Compras, Bodega, Ventas, Cobranza, Logística y Gerencia) en Power BI Service.
Gestión de metadatos	Diccionario de datos del Maestro de Artículos y de Ventas, con nombre, tipo y origen de cada campo.
Gestión de flujos de trabajo (ETL)	Programación del refresco de Power BI y del ETL que migrará llamadas, WhatsApp y correos de cobranza al datamart correspondiente.
Búsqueda y Text Analytics	Exploración de texto libre de llamadas y reclamos para extraer producto solicitado, causal de mora y tipo de incidencia.

Capacidad	Descripción para PIGE
Auditoría y cumplimiento (LOPD)	Trazabilidad de qué dato se compró o capturó, con qué finalidad y bajo qué autorización.
Machine learning (fase futura)	Uso posterior de los datos consolidados para modelos de scoring de riesgo y de predicción de quiebres de stock.

7.4 Estructura y codificación de los datos

No se imponen restricciones estrictas de codificación: los archivos Excel y CSV de SitacPlus, Veritrade o los burós pueden cargarse en formatos estándar (UTF-8, separados por comas o punto y coma), con el criterio ya aplicado en la limpieza del Maestro de Artículos y las Ventas. Se recomienda que todo archivo tabular incluya encabezados claros, para evitar reprocesos como los más de 35 pasos de limpieza que exigió el histórico de ventas.

7.5 Estructura propuesta del proyecto

- **Analizar proveedores de nube de bajo costo** (Microsoft 365 con OneDrive y SharePoint, Google Workspace o Azure) que PIGE ya podría tener o a los que pueda acceder con una inversión mínima.
- **Seleccionar el proveedor** más ajustado al presupuesto y a la madurez tecnológica, priorizando la integración con Power BI.
- **Definir el modelo de cómputo en la nube como SaaS** (Power BI Service y Microsoft 365), evitando por ahora los modelos PaaS o IaaS, que exigirían administrar infraestructura propia.
- **Establecer el flujo de integración** entre la LZ y el modelo de Power BI, de modo que los datos internos y, después, las fuentes externas alimenten un único datamart consolidado.

La computación en la nube pone al alcance de PIGE capacidades que antes eran exclusivas de grandes organizaciones, y el modelo SaaS es la opción más razonable en esta etapa. El paso siguiente es habilitar el repositorio, conectar el Maestro de Artículos y las Ventas ya trabajados y definir el proceso de captura y Text Analytics para llamadas y reclamos.

Capítulo 8. Arquitectura de código abierto (Semana 8)

8.1 Objetivo y criterio de selección

Este capítulo selecciona las herramientas de código abierto y los lenguajes que explotarán la Zona de Aterrizaje, con una evaluación crítica de su pertinencia frente al volumen real de datos. A diferencia de una gran corporación, PIGE es una empresa mediana cuyo volumen transaccional (ventas 2021-2023, catálogo de artículos y, a futuro, reportes de Veritrade y de los burós) no justifica aún un clúster propio. El criterio es, por tanto, de proporcionalidad: resolver las necesidades actuales sin sobredimensionar la inversión y dejar prevista una ruta de escalamiento.

8.2 Hadoop frente a Spark

Ambos son frameworks de la Fundación Apache, escalables y tolerantes a fallos (White, 2015; Zaharia et al., 2016). Se evaluaron frente a tres necesidades: un futuro modelo de scoring de riesgo, el análisis de texto de llamadas y reclamos, y la agilidad para iterar sobre los datos de ventas ya trabajados en Power Query.

TABLA 27

COMPARACIÓN DE HADOOP Y APACHE SPARK APLICADA A PIGE

Criterio	Hadoop (HDFS y MapReduce)	Apache Spark
Velocidad	Procesamiento por lotes en disco; más lento para iterar sobre un mismo conjunto de datos.	Procesamiento en memoria; mucho más rápido en cálculos iterativos como el scoring de riesgo.
Facilidad de uso	Requiere programar en Java sobre MapReduce; curva de aprendizaje alta para un equipo pequeño.	Soporta SQL, Python y R de forma nativa, lenguajes que el equipo ya usa o planea usar.

Criterio	Hadoop (HDFS y MapReduce)	Apache Spark
Machine learning y analítica avanzada	No incluye librerías propias; exige herramientas externas.	Incluye Spark MLlib, útil a futuro para el scoring crediticio y la predicción de quiebres.
Almacenamiento distribuido	Propio (HDFS), pensado para volúmenes muy grandes en varios nodos.	Sin sistema propio; se apoya en HDFS o, más pertinente para PIGE, en un repositorio en la nube de bajo costo.
Pertinencia para el volumen actual	Sobredimensionado: exige un clúster propio que no se justifica.	Preferible por flexibilidad, pero conviene contratarlo como servicio gestionado y solo cuando el volumen lo requiera.

Se selecciona Apache Spark como framework de referencia por su compatibilidad con Python y R y por su API de aprendizaje automático. Sin embargo, dado el volumen actual, se recomienda no desplegar un clúster propio de Spark ni de HDFS, sino contratarlo como servicio gestionado en la nube cuando el crecimiento de los datos lo justifique.

8.3 Lenguajes y herramientas complementarias

- **Python:** lenguaje principal de procesamiento y limpieza de datos (complementa el trabajo en Power Query) y herramienta de Text Analytics para extraer información de llamadas, WhatsApp y reclamos (McKinney, 2022).
- **R:** reservado para el análisis estadístico riguroso, en particular el futuro modelo de scoring de riesgo de clientes empresariales con datos de Equifax o Aval Buró y el historial de cobranza.
- **HDFS:** no se considera necesario en esta etapa; el repositorio en la nube propuesto como Zona de Aterrizaje resuelve el almacenamiento.

8.4 Ecosistema propuesto

TABLA 28

ECOSISTEMA DE CÓDIGO ABIERTO DE PIGE, DE LAS FUENTES A POWER BI

Componente	Rol en el proyecto
Python y Pandas	Motor principal de procesamiento hoy: limpieza y transformación de archivos de SitacPlus y Excel, y base del ETL hacia la Zona de Aterrizaje.
Apache Spark (gestionado)	Reservado para una fase futura, si el volumen crece de forma significativa (por ejemplo, con la incorporación masiva de Veritrade o de los burós); se contrataría como servicio en la nube y no como clúster propio.
R	Lenguaje estadístico para el futuro modelo de scoring de riesgo de clientes empresariales.
Python (Text Analytics)	Extracción de palabras clave en llamadas, WhatsApp y reclamos, y modelado de patrones de pedido para anticipar quiebres de stock.
Bases de datos	Se mantiene la base relacional de SitacPlus; por ahora no se requiere una base NoSQL, pues los datos son mayoritariamente tabulares.
Power BI Desktop y Service	Capa de visualización y entrega de resultados, que consume los datamarts generados desde Python, R y, eventualmente, Spark.

8.5 Costos de implementación y operación

La retroalimentación recibida señaló la conveniencia de cuantificar los costos del proyecto: horas del equipo, formación y equipamiento. La tabla los estima para el primer año. Las herramientas de código abierto no tienen costo de licencia y el volumen de datos permite usar equipos ya existentes, por lo que el esfuerzo se concentra en horas de trabajo y de formación.

TABLA 29*COSTOS ESTIMADOS DE IMPLEMENTACIÓN Y OPERACIÓN, PRIMER AÑO*

Rubro	Detalle y supuesto	Costo (USD)
Horas de proyecto: Responsable de Analítica	Aproximadamente 53 horas mensuales (30 % de la jornada); incluido en el presupuesto del capítulo 5	3.600,00
Horas de proyecto: consultor externo	Construcción inicial y transferencia de conocimiento; incluido en el presupuesto del capítulo 5	3.000,00
Licencias de Power BI Pro	Tres usuarios a USD 14 mensuales; incluido en el presupuesto del capítulo 5	504,00
Horas de Asesoría Informática	40 horas únicas para conexión con SitacPlus y ETL en Python, a USD 15 por hora (supuesto del equipo)	600,00
Formación de usuarios	Cuatro usuarios de Compras, Bodega y Ventas, 8 horas cada uno, a la tarifa de referencia de USD 4,85 por hora	155,20
Capacitación técnica en Python y R	Alrededor de 60 horas con material abierto, dentro de la dedicación del Responsable de Analítica	0,00
Máquinas e infraestructura	Equipos existentes; repositorio dentro de Microsoft 365; Spark gestionado no se contrata en esta etapa	0,00
Total primer año	USD 7.104 son costos recurrentes del área y USD 755,20 son costos únicos de implementación	7.859,20

Nota. Los costos de Asesoría Informática y de formación son supuestos del equipo, por validar. El efecto de estos costos sobre el caso de negocio se analiza en los capítulos 9 y 10.

8.6 Estructura y valoración

La combinación de Python para el procesamiento y el Text Analytics, R para el futuro scoring de riesgo y Spark como opción de escalamiento gestionado ofrece un balance adecuado entre capacidad técnica y proporcionalidad. El siguiente paso es implementar el procesamiento en Python sobre los datos ya cargados (SitacPlus, Maestro de Artículos y Ventas 2021-2023), dejando definida la ruta de incorporación de R y de Spark gestionado.

Capítulo 9. Implementación de inteligencia artificial (Semana 9)

9.1 *Objetivo de aplicación del modelo*

Resueltos el diagnóstico, la arquitectura y las herramientas, los datos consolidados deben servir para algo más que mostrar lo ocurrido. Este capítulo define cómo se aplicarían la ciencia de datos y el aprendizaje automático al proyecto de BI de PIGE, siguiendo los cinco puntos de estructura del material de la semana (Pérez Galán, s. f.), cuyo ejemplo es un modelo de cobranza bancaria que aquí se adapta a una empresa mediana.

El problema elegido es el reabastecimiento. Hoy la alerta de compra depende del criterio del Líder de Bodega, que advierte cuándo un artículo cae bajo el stock mínimo: una decisión experta pero sin apoyo del histórico de rotación. Se estima que solo el 75 % de los pedidos de reposición se genera a tiempo y que cerca del 6 % de los pedidos no se satisface por falta de inventario. Es el punto donde un modelo predictivo aporta más valor con menos complejidad, y la gestión de inventarios es un componente central del desempeño de la cadena de suministro (Chopra y Meindl, 2019).

El modelo predice, para cada artículo activo del Maestro de Artículos, la demanda esperada de las próximas semanas y la probabilidad de que el inventario no alcance a cubrirla. Con ello, Compras recibe con anticipación una alerta priorizada, con cantidad sugerida y punto de reorden calculados a partir del comportamiento histórico (Hyndman y Athanasopoulos, 2021; Silver et al., 2017).

TABLA 30*DEFINICIÓN DEL MODELO PREDICTIVO DE REABASTECIMIENTO Y RIESGO DE QUIEBRE*

Elemento	Definición para PIGE
Tipo de problema	Pronóstico de demanda por artículo (series de tiempo) y clasificación del riesgo de quiebre, entendido como la probabilidad de agotarse antes del siguiente ingreso de mercadería.
Unidad de análisis	Artículo del Maestro de Artículos, agregable por línea de producto y por origen (importado o local).
Variable a predecir	Unidades demandadas por período y ocurrencia de quiebre (pedido no satisfecho por falta de inventario).
Frecuencia	Semanal, tras el corte de SitacPlus; diaria cuando el ETL se automatice.
Resultado entregado	Demanda proyectada, probabilidad de quiebre, punto de reorden y cantidad sugerida, con semaforización en Power BI.
Herramientas	Python (Pandas, scikit-learn, statsmodels), R como apoyo estadístico y Power BI para la entrega; Spark queda reservado para una fase futura (Pedregosa et al., 2011; Seabold y Perktold, 2010).

Aplicación. El modelo se ejecuta en cada corte semanal sobre todos los artículos activos. Los de riesgo alto generan una alerta a Compras e Importaciones; los demás siguen en seguimiento normal hasta el siguiente corte. El punto de reorden no es igual para todos: en los importados, con tiempos de reposición más largos, la alerta debe activarse con más anticipación, y por eso la columna Importado o Local del modelo de Power BI es una variable de entrada.

Construcción. Exige reconstruir al menos dos años de historia, requisito que PIGE cumple con las ventas 2021-2023 ya relacionadas con el Maestro de Artículos. Se propone entrenar con los años más antiguos, reservar el último para validar y comparar el resultado contra un método base, como un promedio móvil o un suavizamiento exponencial. Si el

aprendizaje automático no supera con claridad al método base, se conserva el más simple, en coherencia con el criterio de proporcionalidad del capítulo 8 y con la práctica de validar siempre frente a un modelo de referencia (Provost y Fawcett, 2013).

Extensión prevista: propensión de pago. Con las fuentes externas integradas y el registro de cobranza ordenado, el mismo enfoque permitiría estimar la probabilidad de que un cliente empresarial en mora acepte una alternativa de negociación y priorizar la gestión de cobranza. Depende de datos que PIGE aún no tiene, por lo que queda fuera del caso de negocio de este capítulo.

9.2 Fuentes de datos

TABLA 31

FUENTES DE DATOS DEL MODELO

Fuente	Tipo	Datos que aporta	Uso en el modelo	Situación
SitacPlus: ventas 2021-2023	Interna, estructurada	Cantidad vendida, cliente, bodega y artículo	Variable a predecir: demanda por artículo y período	Disponible; cargada en Power BI
SitacPlus: Maestro de Artículos	Interna, estructurada	Marca, categoría, bodega, costo, existencia y origen	Atributos del artículo y stock actual	Disponible; cargado en Power BI
SitacPlus: compras e inventario	Interna, estructurada	Compras locales e importadas, tiempos de entrega, movimientos y pedidos no satisfechos	Tiempo de reposición e historial de quiebres	En el sistema; falta llevarla al datamart
Llamadas y consultas comerciales	Interna, no estructurada	Productos consultados que no quedan registrados	Señal de demanda no atendida (Text Analytics)	Fase posterior
Veritrade	Externa, semiestructurada	Importaciones del sector por producto, país y período	Estacionalidad y tendencia de los importados	Fase posterior

Variables. Los indicadores del capítulo 2 pasan a ser las variables de entrada: inventario promedio, rotación trimestral, tiempo promedio de reabastecimiento e índice de rotura de stock, a los que se suman la estacionalidad (mes y trimestre), la línea de producto, la marca, el origen y el costo. Así el trabajo hecho para los tableros se reutiliza.

Calidad. La información de SitacPlus es de calidad media-alta, pero carece de un identificador único y su histórico requirió más de 35 pasos de limpieza. Para un modelo predictivo estos defectos pesan más que en un tablero: un artículo registrado con dos códigos se lee como dos productos con menos demanda cada uno. Definir un código único de artículo es, por tanto, condición previa al entrenamiento, y conviene incorporar los años posteriores a 2023 para que el modelo aprenda de la demanda actual.

9.3 Áreas de aplicación y departamentos implicados

TABLA 32

PAPEL DE CADA DEPARTAMENTO FRENTE AL MODELO

Departamento	Papel	Qué recibe o aporta
Compras e Importaciones	Usuario principal	Recibe la alerta priorizada con cantidad sugerida y punto de reorden; decide y emite el pedido local o de importación.
Bodega (Líder de Bodega)	Validador y fuente de datos	Confirma existencias y conteos que alimentan el ERI; su criterio deja de ser el único disparador y contrasta lo que el modelo señala.
Ventas	Beneficiario y fuente de datos	Registra la demanda no atendida y las cotizaciones; gana certeza para ofrecer plazos de entrega, lo que favorece la conversión.
Logística	Beneficiario indirecto	Planifica despachos con menos pedidos pendientes por falta de mercadería.
Gerencia General	Patrocinador	Autoriza el acceso a los datamarts y consulta los indicadores de quiebre y de efectividad del reabastecimiento.
Equipo de analítica	Construcción y mantenimiento	El Responsable de Analítica construye, monitorea y reentrena; la Asesoría Informática asegura la conexión con SitacPlus; el consultor apoya la primera versión y transfiere conocimiento.

El cambio es de fondo y no de estructura. Antes, Bodega detectaba la caída del stock mínimo y avisaba a Compras; con el modelo, el datamart alimenta la predicción, Compras recibe una alerta ordenada por probabilidad de quiebre y Bodega confirma lo que hay físicamente. Las personas siguen decidiendo: el modelo sugiere y la experiencia del Líder de Bodega sirve para detectar cuándo se equivoca. Las alertas se mostrarán como semáforos en Power BI, con acceso por roles. Al tratarse de datos de artículos, compras y ventas por producto, el caso toca poco los datos personales; al llevar el análisis al nivel de cliente, aplicarán los principios de finalidad y minimización de la LOPDP.

9.4 Identificación de la mejora esperada

TABLA 33

SITUACIÓN ACTUAL Y PROYECTADA DE LOS INDICADORES

Indicador	Hoy	Con el modelo	Origen de la mejora
Quiebre de stock (pedidos no satisfechos)	≈ 6 %	≈ 2,5 % en un año	Modelo predictivo
Efectividad del reabastecimiento (pedidos a tiempo)	75 %	90 %	Modelo predictivo
Horas mensuales de consolidación de reportes	20 h	6 h	Tableros automatizados
Costo mensual de consolidación	USD 97,00	USD 29,10	Tableros automatizados (USD 4,85 por hora)

Conviene separar los efectos: los dos primeros indicadores dependen del modelo predictivo; los dos últimos provienen de automatizar los tableros y se logran aunque el modelo tarde en madurar. Incluir los cuatro en el caso de negocio es válido porque comparten la inversión, pero es útil saber de dónde viene cada beneficio.

Mejoras cualitativas. Compras más oportunas y con menos excesos de capital inmovilizado; mejor planificación de importaciones, al considerar el tiempo de reposición de cada artículo y la estacionalidad que muestra Veritrade; servicio más consistente a los clientes B2B de construcción, minería e industria, con menos pedidos incompletos; priorización basada en datos, pues las alertas se ordenan por probabilidad de quiebre y no por percepción; y una base para escalar hacia el scoring de riesgo y la propensión de pago.

Medición. El índice de rotura de stock, la efectividad del reabastecimiento y el error de pronóstico se seguirán cada mes en Power BI. El 6 % y el 75 % son estimaciones de la gerencia: en los primeros meses se medirán con datos de SitacPlus para fijar una línea base verificable.

9.5 Caso de negocio

Supuestos. Ventas de referencia de USD 1,63 millones (2023); cada punto de quiebre evitado equivale a unos USD 16.300 anuales en ventas recuperadas; el quiebre baja de 6 % a 2,5 % en un año de operación del modelo (3,5 puntos); y el margen bruto global es de alrededor de 30 %, dato confirmado por la Gerencia de PIGE, que no dispone de márgenes por línea. El modelo se construye con Python y R sin licencias adicionales y con la bolsa de horas del consultor ya presupuestada. Como PIGE vende a crédito, los ingresos recuperados se cobran con plazo, aspecto que se considera entre los riesgos.

Evaluación sobre margen y no sobre ventas. Las ventas recuperadas son ingresos y no utilidad; por eso el beneficio se calcula aplicando el margen bruto. Al no existir márgenes por línea, se usa el margen global de 30 %, y el cálculo podrá afinarse cuando la fase de ordenamiento de datos permita medir la rentabilidad de cada línea.

TABLA 34*BENEFICIOS Y COSTOS ANUALES CON EL MODELO EN RÉGIMEN*

Concepto	Valor anual (USD)
Ventas recuperadas por reducción del quiebre (3,5 puntos × 16.300)	57.050,00
Margen bruto sobre esas ventas (≈ 30 %)	17.115,00
Ahorro en consolidación de reportes (14 h × 4,85 × 12 meses)	814,80
Beneficio anual en régimen	17.929,80
Costo anual recurrente del área de analítica	7.104,00
Beneficio neto anual en régimen	10.825,80
Inversión del primer año (incluye USD 755,20 de costos únicos)	7.859,20

TABLA 35*ANÁLISIS DE ESCENARIOS, CON EL MODELO EN RÉGIMEN*

Escenario	Quiebre final	Puntos evitados	Ventas recuperadas (USD)	Margen bruto 30 % (USD)	Beneficio anual (USD)	Beneficio neto anual (USD)
Conservador	4,5 %	1,5	24.450	7.335,00	8.149,80	1.045,80
Intermedio	3,5 %	2,5	40.750	12.225,00	13.039,80	5.935,80
Esperado	2,5 %	3,5	57.050	17.115,00	17.929,80	10.825,80

Nota. Beneficio anual = margen bruto + ahorro de consolidación (USD 814,80). Beneficio

neto anual = beneficio anual menos el costo recurrente de USD 7.104. El capítulo 10

convierte estos flujos en un valor actual neto a cinco años.

En el escenario conservador el beneficio neto en régimen es apenas positivo, y al descontar la inversión inicial el valor actual neto a cinco años resulta ligeramente negativo (capítulo 10). Es la razón por la que el proyecto se implanta por etapas, con puntos de control antes de comprometer toda la inversión (Brealey et al., 2020).

Riesgos y supuestos críticos.

- **Calidad de los datos:** sin código único de artículo y con históricos que requirieron más de 35 pasos de limpieza, el modelo aprendería de información inconsistente.
- **Línea base estimada:** el 6 % de quiebre y el 75 % de efectividad deben verificarse en SitacPlus antes de comprometer metas.
- **Historia desactualizada:** los datos llegan hasta 2023; deben incorporarse los años posteriores antes de entrenar.
- **Adopción:** si Compras y Bodega no confían en las alertas, el modelo no cambia decisiones; se propone un piloto con las líneas de mayor rotación, comparando las alertas con lo que habría decidido el Líder de Bodega, y aplicar las medidas de gestión del cambio del capítulo 5.
- **Volumen moderado de datos:** en una empresa mediana un modelo sencillo puede rendir igual que uno complejo, por lo que la comparación con un método base es parte del diseño.
- **Venta a crédito:** el beneficio en caja llega después del beneficio en ventas, pues los ingresos recuperados se cobran con plazo; conviene acompañar el modelo con el scoring de riesgo de clientes del capítulo 6.

Ubicación en el cronograma. El modelo se desarrollaría sobre los datamarts y tableros de las fases 2 y 3 de la propuesta de inversión y entraría en operación en la fase de mejora continua, desde el mes 6; por eso el cálculo corresponde al primer año de operación del modelo y no al primer año calendario del proyecto.

Capítulo 10. Caso de uso en tiempo casi real y valoración integral (Semana 10)

10.1 Objetivo y alcance

El cliente industrial quiere su pedido en el mismo instante, y esa exigencia se traslada al análisis de los datos: no basta con saber qué ocurrió la semana pasada, hay que saber qué ocurre con el inventario mientras se vende, se despacha y se cotiza. Este capítulo aplica a PIGE la teoría de analítica en tiempo real revisada en la materia (Pérez Galán, s. f.) y cierra el trabajo con la valoración financiera integral del proyecto.

El modelo predictivo del capítulo 9 se ejecuta en un corte semanal. Entre un corte y otro PIGE sigue sin visibilidad: un respirador o un arnés puede agotarse un martes y no advertirse hasta que el Líder de Bodega lo note o llegue el siguiente corte. El caso de uso cierra esa brecha con una capa de detección de eventos casi en tiempo real sobre el mismo inventario. Se plantean dos diseños: el Diseño A, proporcional al tamaño de PIGE, con Python y Power BI; y el Diseño B, escalable, con Spark Streaming gestionado en la nube. Se recomienda el A y se deja el B como ruta de crecimiento, en coherencia con el criterio de proporcionalidad del capítulo 8.

El alcance cubre el diseño funcional, de datos y de reglas; no incluye el desarrollo del código ni la conexión técnica con SitacPlus, que depende de lo que permita el proveedor. Para probar el diseño sin tocar la operación, se propone reproducir (replay) el histórico de ventas 2021-2023 como si fuera un flujo de eventos en vivo, de modo que los umbrales se calibren con datos reales.

10.2 Marco teórico aplicado: qué significa tiempo real para PIGE

La materia distingue dos técnicas: el Data Stream, que analiza los datos mientras están en movimiento, y el CEP (Complex Event Processing), que busca eventos predefinidos

dentro de ese flujo (Kleppmann, 2017). En PIGE el flujo de transacciones de SitacPlus es el stream, y los eventos de interés (riesgo de quiebre, pico de demanda, demanda no atendida) son los que el CEP debe reconocer.

TABLA 36

TÉCNICAS DE ANÁLISIS EN TIEMPO REAL Y SU APLICACIÓN EN PIGE

Técnica	Descripción	Aplicación en PIGE
Data Stream	Análisis mientras los datos están en movimiento; prioriza la velocidad.	Cada venta, ingreso, despacho o cotización registrada en SitacPlus actualiza el estado del inventario apenas se genera.
CEP	Búsqueda de eventos predefinidos en el flujo de datos.	Reglas que reconocen cobertura menor al tiempo de reposición, picos de demanda o cotizaciones sobre artículos sin stock.
On-demand	Consulta periódica de nueva información mediante queries.	Es el modo actual: el Líder de Bodega consulta el stock cuando lo revisa. Se mantiene como respaldo.
Streaming	Flujo continuo de alto volumen y velocidad mediante un push.	Modo objetivo, adaptado al volumen de PIGE mediante ingestas incrementales cada pocos minutos.

Lo que realmente se busca reducir es la latencia, es decir, el tiempo entre que un evento ocurre y que puede usarse. La teoría distingue tres tipos: de la data, del análisis y de la acción. El modelo tradicional de data warehouse solo reduce la primera; el análisis en tiempo real busca reducir las tres.

TABLA 37

LATENCIAS DE PIGE: SITUACIÓN ACTUAL Y OBJETIVO DE DISEÑO

Latencia	Hoy	Objetivo de diseño
De la data	Carga diaria al cierre de operaciones (18:00) y consolidación manual al día siguiente.	Ingesta incremental cada 5 minutos en horario laboral.
Del análisis	Consolidación manual en hojas de cálculo (unas 20 horas al mes) y modelo semanal.	Segundos: reglas y modelo se evalúan en cada ciclo de ingesta.
De la acción	Depende de que el Líder de Bodega advierta la caída bajo el stock mínimo y avise a Compras.	Alerta priorizada al responsable, con meta de decisión de compra o revisión en menos de una hora.

Nota. Los valores del objetivo de diseño son metas por validar en el piloto; los de la columna Hoy provienen del diagnóstico de los capítulos 2, 3 y 9.

La teoría advierte que tiempo real no significa lo mismo en todas las industrias: milisegundos para un vehículo autónomo, segundos para sensores, minutos para un análisis de mercado. Decidir una compra de EPP no exige milisegundos, exige enterarse a tiempo, mientras todavía se puede reponer. Por eso el caso se clasifica como **near real time**: se procesan periódicamente los datos que llegan en un intervalo corto de minutos. Además, el data warehouse y el tiempo real no se excluyen: los datamarts siguen siendo la base de tendencias, márgenes y del modelo semanal, y la capa de tiempo real solo atiende los eventos que no pueden esperar al siguiente corte.

10.3 Fases del detector de quiebres

TABLA 38

FLUJO DEL DETECTOR, DE LA TRANSACCIÓN A LA ALERTA

N.º	Fase	Proceso	Salida
1	Captura de eventos	Extracción incremental cada 5 minutos de las transacciones de SitacPlus y de las hojas de ruta; si el proveedor solo permite exportación, un archivo automático cada 5 minutos hacia la Zona de Aterrizaje	Mensajes de evento
2	Filtro de mensajes	Validación de código, duplicados, fechas y unidades; los descartes se registran para corregirlos en origen	Eventos válidos
3	Consulta al repositorio	Búsqueda del estado e historial del artículo; si es nuevo o sin historia suficiente, el evento solo se almacena	Evento enriquecido
4	Detección de eventos	Reglas CEP y modelo estadístico	Evento clasificado
5	Alerta y acción	Semáforo en Power BI y notificación por correo al responsable, con cantidad sugerida de reposición	Alerta priorizada
6	Retroalimentación	Registro de la decisión tomada y de si la alerta fue acertada; reajuste de umbrales	Reglas recalibradas

La fase 2 pesa más en PIGE que en un caso ideal, porque SitacPlus carece de identificador único transversal y el histórico requirió más de 35 pasos de limpieza. La fase 6 reduce las falsas alarmas, un riesgo real en un equipo pequeño que dejaría de atender alertas que no resultan útiles.

10.4 Diseño del repositorio de datos

Un diseño ingenuo guardaría todos los movimientos en una sola tabla y la recorrería entera con cada evento, lo que se vuelve lento al crecer el histórico. Siguiendo el capítulo 8, que concluyó que por ahora no se requiere una base NoSQL, el diseño se mantiene en tablas relacionales o archivos estructurados dentro de la Zona de Aterrizaje.

TABLA 39*ESTRUCTURAS DE DATOS DEL DETECTOR*

Estructura	Contenido	Razón del diseño
Estado_articulo	Una fila por artículo: stock, demanda promedio y desviación recientes, tiempo de reposición, origen, línea y semáforo	Responde en una sola búsqueda sin recorrer el histórico
Movimientos	Eventos válidos, ordenados por fecha y partidos por mes	Conserva la trazabilidad y alimenta el modelo semanal
Alertas	Evento, regla, severidad, responsable, hora de emisión y decisión tomada	Permite medir la latencia de acción y la tasa de aciertos
Descartes	Eventos rechazados en la fase 2, con el motivo	Convierte los errores de datos en correcciones para el gobierno de datos

10.5 Reglas y algoritmo de detección

Cada evento tiene atributos definidos: artículo (código único del Maestro), línea de producto (una de las once), origen (importado o local, que determina el tiempo de reposición), tipo de evento (venta, ingreso, despacho, cotización o pedido no satisfecho), cantidad, canal (mostrador, WhatsApp, teléfono o web) y segmento de cliente (construcción, minería, industria o distribuidor).

La cobertura de un artículo se calcula como **CD = stock disponible / demanda diaria proyectada**. Un artículo está en riesgo cuando **CD < TPR + SS**, donde TPR es el tiempo promedio de reabastecimiento, KPI definido en el capítulo 2, y SS es un margen de seguridad en días, mayor para los importados. Los umbrales iniciales se calibran con el replay del histórico.

TABLA 40*EVENTOS QUE RECONOCE EL DETECTOR*

N.º	Evento	Regla (parámetros iniciales por calibrar)	Nivel	Responsable
E1	Riesgo de quiebre	Cobertura menor al tiempo de reposición más margen de seguridad	Rojo	Compras e Importaciones
E2	Stock crítico	Artículos de líneas críticas (respiratoria, anticaídas) bajo su umbral propio	Rojo	Compras e Importaciones
E3	Demanda no atendida	Cotización o pedido sobre un artículo con stock insuficiente	Rojo	Compras y Ventas
E4	Pico de demanda	Ventas de la ventana reciente a más de 3 desviaciones estándar de su comportamiento normal	Amarillo	Compras y Ventas
E5	Despacho retrasado	Pedido sin despachar pasado el plazo comprometido con el cliente	Amarillo	Logística
E6	Inconsistencia de datos	Movimiento con código inválido, duplicado o stock negativo	Amarillo	Bodega y analítica

Para el evento E4 se adopta un modelo estadístico basado en la distribución normal de Gauss: se calculan la media (μ) y la desviación estándar (σ) de la demanda en condiciones normales y se marca como anómalo un valor cuyo puntaje $z = (x - \mu) / \sigma$ supere el umbral definido. Es un sistema semi-supervisado, porque solo se necesita conocer las instancias normales, y su ventaja es la sencillez (Chandola et al., 2009).

Limitación. Con 16 colaboradores y una operación de decenas o centenas de transacciones diarias, muchos artículos se venden de forma esporádica y su desviación estándar es inestable. Por eso el modelo gaussiano se aplica solo a los artículos y líneas de mayor rotación, o a la demanda agregada por línea, y los de baja rotación se vigilan únicamente con las reglas E1 a E3, lo que coincide con el piloto propuesto en el capítulo 9.

10.6 Herramientas evaluadas y diseños propuestos

TABLA 41

HERRAMIENTAS DE BIG DATA Y DECISIÓN PARA PIGE

Herramienta	Pertinencia para PIGE	Decisión
Apache Hadoop	Procesa por lotes y exige un clúster propio; no sirve para detectar eventos al minuto	Descartada
Apache Spark	Spark Streaming es la opción natural si el volumen crece; se contrataría como servicio gestionado	Diseño B, fase futura
Python	Su menor velocidad es irrelevante con el volumen de PIGE; ya es la herramienta principal del ETL	Seleccionada (Diseño A)
Elasticsearch	No se requiere hoy; podría evaluarse para buscar texto en llamadas y WhatsApp	No requerida
Apache Storm	Procesa millones de datos por segundo, volumen muy superior al de PIGE	Descartada
Lenguaje R	Apoyo estadístico y futuro scoring de riesgo de clientes	Apoyo

Diseño A (recomendado). Un proceso en Python programado cada cinco minutos en horario laboral extrae los movimientos nuevos, los valida, actualiza las estructuras de la Zona de Aterrizaje, evalúa las reglas y publica el resultado para Power BI. Las alertas rojas se envían además por correo, de modo que no dependan de que alguien tenga abierto el tablero. Funciona con equipos existentes, el repositorio en Microsoft 365 y las tres licencias de Power BI Pro.

Diseño B (escalable). Spark Structured Streaming gestionado en la nube lee los archivos que llegan a la Zona de Aterrizaje en microlotes y aplica las mismas reglas. Solo se justificaría si el volumen crece de forma significativa, por ejemplo con la incorporación masiva de Veritrade, el análisis de texto de WhatsApp o un canal de comercio electrónico. Las reglas y estructuras se reutilizan, de modo que migrar de A a B no obliga a rediseñar el negocio.

Nota técnica por validar: la frecuencia mínima de actualización automática de los tableros de Power BI depende del tipo de licencia y de la forma de conexión. Si con las licencias Pro no es posible refrescar cada cinco minutos, el tablero se actualiza con la frecuencia permitida y las alertas por correo siguen saliendo al ritmo del proceso.

10.7 Fuentes de datos y departamentos implicados

El detector se alimenta de las ventas, las compras e inventario, el Maestro de Artículos, las cotizaciones y las hojas de ruta de Logística, con una frecuencia de cinco minutos para las transacciones y diaria para el maestro; la web de PIGE aporta solicitudes de cotización, y Veritrade, como fuente de estacionalidad, entra en una fase posterior porque no es un flujo continuo. Los departamentos tienen los papeles ya descritos en el capítulo 9, con dos matices:

- **Compras e Importaciones** recibe las alertas E1 a E4, ordenadas por criticidad y con cantidad sugerida; **Logística** recibe las E5; **Bodega** atiende las E6 y contrasta lo que el sistema señala.
- **Gerencia General** consulta desde el móvil los indicadores de quiebre y de efectividad y recibe las alertas rojas. Las personas siguen decidiendo: el detector avisa antes y con orden, y la experiencia del equipo sirve para reconocer una falsa alarma.

10.8 Mejora esperada y costo incremental

TABLA 42

SITUACIÓN ACTUAL Y OBJETIVO DE LOS INDICADORES

Indicador	Hoy	Objetivo	Origen de la mejora
Latencia de la data	Hasta más de 1 día	Máximo 5 minutos	Ingesta incremental

Indicador	Hoy	Objetivo	Origen de la mejora
Quiebre de stock	≈ 6 %	≈ 2,5 % en un año	Modelo del capítulo 9 y alertas E1 a E3
Efectividad del reabastecimiento	75 %	90 %	Modelo del capítulo 9 y alertas E1 a E3
Horas mensuales de consolidación	20 h	6 h	Tableros automatizados
Costo mensual de consolidación	USD 97,00	USD 29,10	Tableros automatizados

Nota. La capa de tiempo real no suma beneficios nuevos: es el mecanismo que permite sostener las metas del capítulo 9 entre un corte semanal y otro. El 6 % y el 75 % son estimaciones de la gerencia que deben verificarse en SitacPlus.

TABLA 43

COSTO INCREMENTAL DE LA CAPA DE TIEMPO REAL, PRIMER AÑO

Rubro	Detalle y supuesto	Costo (USD)
Asesoría informática adicional	30 horas para la ingesta incremental y el envío de alertas, a USD 15 por hora	450,00
Formación de usuarios	Cuatro usuarios, 2 horas cada uno, a USD 4,85 por hora	38,80
Equipo para ejecutar el proceso	Servidor existente o máquina virtual pequeña en la nube, USD 12 al mes (referencial)	144,00
Licencias y canal de alertas	Se usan las tres licencias de Power BI Pro ya presupuestadas y el correo electrónico	0,00
Total incremental		632,80

Nota. Supuestos del equipo, por validar con la asesoría informática y el proveedor. Con esta capa, la inversión del primer año sería de USD 8.492,00 (USD 7.859,20 más USD 632,80).

Para dimensionar qué aporta la capa por sí misma se presenta una sensibilidad sobre margen: cuánto cubre su costo si evita puntos de quiebre adicionales a los del caso base. Cada punto equivale a USD 16.300 en ventas, es decir, USD 4.890 de margen al 30 %.

TABLA 44*SENSIBILIDAD: QUIEBRE ADICIONAL EVITADO POR LA CAPA DE TIEMPO REAL*

Puntos adicionales evitados	Ventas recuperadas (USD)	Margen bruto 30 % (USD)	Margen / costo incremental
0,25	4.075	1.222,50	1,9 veces
0,50	8.150	2.445,00	3,9 veces
1,00	16.300	4.890,00	7,7 veces

Nota. Es un ejercicio de sensibilidad, no una proyección. El costo incremental del primer año se cubre con unos 0,13 puntos de quiebre evitado.

10.9 Valoración integral del proyecto

La valoración final se hace sobre flujos de cinco años y sobre margen, con tres parámetros confirmados por la Gerencia de PIGE: la empresa vende a crédito, el margen bruto global ronda el 30 % y se puede trabajar con una tasa de descuento de 12 %, similar al costo del dinero de un crédito. La inversión es de USD 7.859,20 el primer año y de USD 7.104 anuales a partir del segundo. El modelo entra en operación hacia el mes 6, de modo que los beneficios se materializan de forma gradual (Brealey et al., 2020).

TABLA 45*RESULTADO FINANCIERO DEL PROYECTO A CINCO AÑOS*

Indicador	Valor
Inversión del primer año	USD 7.859,20
Costo anual recurrente desde el segundo año	USD 7.104,00
Beneficio anual en régimen (margen + ahorro)	USD 17.929,80
Valor actual neto a cinco años (tasa de 12 %)	USD 26.990
Retorno sobre la inversión del primer año (VAN / inversión)	343 %
Recuperación de la inversión descontada	Alrededor del mes 15
Margen bruto mínimo para un VAN positivo	14,1 % (frente a un margen real de $\approx$ 30 %)

Nota. Cifras del modelo de flujos del equipo con los parámetros indicados. La capa de tiempo real reduciría el VAN en un monto del orden de su costo del primer año (USD 632,80), sin cambiar la conclusión.

TABLA 46

SENSIBILIDAD DEL VALOR ACTUAL NETO

Escenario	Supuesto	VAN a cinco años (USD)
Esperado	El quiebre baja 3,5 puntos (de 6 % a 2,5 %)	26.990
Conservador	El quiebre baja solo 1,5 puntos, con costos constantes	-1.715
Conservador ajustado	Igual, con la bolsa del consultor reducida al 50 % desde el segundo año	2.353

El resultado es sólido cuando se mide sobre margen: el proyecto sigue siendo rentable mientras el margen bruto supere el 14,1 %, muy por debajo del 30 % real. Pero no es inmune al desempeño: si el quiebre solo bajara 1,5 puntos y los costos se mantuvieran, el VAN sería ligeramente negativo. Por eso se propone una implantación por etapas, con dos puntos de control que acotan el riesgo máximo a la etapa en curso.

TABLA 47

CRONOGRAMA DE IMPLANTACIÓN EN CINCO FASES

Fase	Contenido	Período
F1	Línea base y gobierno de datos: medir quiebre y efectividad en SitacPlus, definir el código único y el margen bruto por línea	Meses 1 a 2
F2	Zona de Aterrizaje, ETL y datamarts	Meses 2 a 3
F3	Tableros y adopción por los usuarios, con formación por rol	Meses 3 a 5
F4	Prototipo y piloto del modelo en las líneas de mayor rotación, validado contra un método base	Meses 4 a 6
F5	Operación y mejora continua	Desde el mes 6

- **Control 1, mes 2.** Si la línea base real difiere del 6 % de quiebre y del 75 % de efectividad, se recalcula el caso de negocio antes de continuar.
- **Control 2, mes 6.** El modelo pasa a operación solo si supera al método base en el piloto; si no, se mantienen los tableros y el método simple.

TABLA 48

DOS TRAYECTORIAS PARA PIGE

Sin el proyecto	Con el proyecto
≈ 6 % de quiebre: unos USD 97.800 al año en ventas no atendidas	≈ 2,5 % de quiebre, con alertas priorizadas
Reposición por criterio de una persona; 75 % a tiempo	90 % de reposición a tiempo
Decisiones reactivas, en hojas de cálculo	KPI comunes frente al presupuesto
Crédito sin score ni registro de cobranza	Scoring de riesgo y propensión de pago
Proveedor reactivo que compite por precio	Socio estratégico de seguridad para cuentas clave

10.10 Riesgos y supuestos críticos

- **Acceso a SitacPlus.** Si el sistema solo permite exportaciones manuales, la latencia de la data no bajaría de las horas; es la validación más urgente con el proveedor, y el replay del histórico permite avanzar mientras tanto.
- **Calidad de los datos.** Sin código único, un mismo producto puede leerse como dos y distorsionar la cobertura.
- **Volumen bajo por artículo.** El modelo gaussiano es poco fiable en artículos de venta esporádica; se limita a los de mayor rotación.
- **Fatiga de alertas.** Demasiadas alertas amarillas llevarían al equipo a ignorarlas; se comienza con pocas reglas y se ajustan con la retroalimentación.

- **Línea base y margen.** El 6 % de quiebre y el 75 % de efectividad deben medirse en SitacPlus, y el margen por línea debe calcularse para afinar el caso.
- **Adopción.** El detector solo cambia decisiones si Compras y Bodega confían en él; se propone un piloto contrastado con lo que habría decidido el Líder de Bodega.
- **Necesidad real de tiempo real.** Para varias decisiones una carga diaria bien automatizada podría bastar; el piloto debe medir cuántas alertas habrían llegado tarde con el ciclo diario, y ese número justifica o no pasar a ciclos de cinco minutos.
- **Seguridad de la información.** Con más usuarios y alertas por correo, el acceso por roles, el cifrado y los acuerdos de confidencialidad son indispensables, como establece la LOPDP.

10.11 Conclusión del capítulo

Los datos en stream son los que fluyen mientras transcurre el tiempo. En PIGE están formados por las transacciones de SitacPlus y las hojas de ruta de Logística, y pueden procesarse en tiempo real, cuando cada dato se procesa de inmediato, o en casi tiempo real, cuando se procesa periódicamente lo que llega en un intervalo corto. Un monitor con ciclos de cinco minutos se ajusta a esta segunda modalidad y es proporcional al tamaño de PIGE: reduce las tres latencias, usa las herramientas ya definidas en la arquitectura del proyecto y mantiene su costo incremental por debajo de USD 650 en el primer año. No reemplaza al modelo semanal ni a los datamarts: los complementa. El modelo anticipa lo que probablemente ocurra; el detector avisa de lo que está ocurriendo.

CONCLUSIONES Y APLICACIONES

PIGE llegó a este proyecto con un activo valioso y subutilizado: años de información transaccional en SitacPlus, usada para registrar lo ocurrido y no para anticipar lo que viene. El recorrido de diez semanas muestra que la brecha no es de tecnología sino de método y de gobierno de datos: faltaban indicadores definidos, un flujo de carga repetible, un identificador único de artículos y clientes, márgenes por línea y una función responsable de la analítica.

La propuesta responde con proporcionalidad. En la Parte I se definieron los KPI por proceso, el ETL y los datamarts, un plan de visualización diferenciado por audiencia y un área de analítica de dedicación parcial. En la Parte II se incorporaron fuentes externas bajo el marco de la LOPDP, una Zona de Aterrizaje en la nube bajo modelo SaaS, un ecosistema de código abierto con Spark como opción gestionada de escalamiento, un modelo predictivo de reabastecimiento y un detector de quiebres en tiempo casi real.

La valoración sustenta la decisión. Con un margen bruto de 30 %, una tasa de descuento de 12 % y una inversión de USD 7.859,20 el primer año, reducir el quiebre de stock de 6 % a 2,5 % arroja un valor actual neto de unos USD 26.990 a cinco años, un retorno de 343 % sobre la inversión inicial y una recuperación hacia el mes 15; el proyecto es rentable mientras el margen bruto supere el 14,1 %. La salvedad es honesta: en el escenario conservador, con una mejora de solo 1,5 puntos, el VAN sería ligeramente negativo, y las líneas base de 6 % y 75 % son estimaciones de la gerencia que deben verificarse.

Como aplicación inmediata, se recomienda a la Gerencia aprobar la inversión liberada por etapas: medir la línea base en SitacPlus y definir el código único y el margen por línea (meses 1 a 3); construir un prototipo en Python validado contra un método base (meses 4 a 6); y probarlo en un piloto con las líneas de mayor rotación, acompañado de formación por rol. Cada etapa avanza solo si supera su punto de control. Ninguna herramienta genera valor

si quienes deben usarla no la adoptan, y el valor final depende de los datos y de las personas. Con las fuentes externas integradas, el mismo esquema permitirá abordar el scoring de riesgo y la propensión de pago, tan relevantes para una empresa que vende a crédito, y consolidar a PIGE como el socio estratégico de seguridad al que aspira su visión.

REFERENCIAS

- Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales* (Registro Oficial Suplemento 459, 26 de mayo de 2021).
- Brealey, R. A., Myers, S. C., & Allen, F. (2020). *Principles of corporate finance* (13.^a ed.). McGraw-Hill Education.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58. <https://doi.org/10.1145/1541880.1541882>
- 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>
- Chopra, S., & Meindl, P. (2019). *Supply chain management: Strategy, planning, and operation* (7.^a ed.). Pearson.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Ferrari, A., & Russo, M. (2016). *Introducing Microsoft Power BI*. Microsoft Press.
- Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (2.^a ed.). Analytics Press.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3.^a ed.).

OTexts. <https://otexts.com/fpp3/>

Inmon, W. H. (2005). *Building the data warehouse* (4.^a ed.). Wiley.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business School Press.

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3.^a ed.). Wiley.

Kleppmann, M. (2017). *Designing data-intensive applications*. O'Reilly Media.

Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.

Kotter, J. P. (1996). *Leading change*. Harvard Business School Press.

McKinney, W. (2022). *Python for data analysis* (3.^a ed.). O'Reilly Media.

Microsoft. (s. f.). *Power BI documentation*. <https://learn.microsoft.com/power-bi/>

Parmenter, D. (2015). *Key performance indicators: Developing, implementing, and using winning KPIs* (3.^a ed.). Wiley.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
<https://jmlr.org/papers/v12/pedregosa11a.html>

Pérez Galán. (s. f.). *Material didáctico de la asignatura* [Material de curso]. Universidad Internacional del Ecuador y Westfield Business School.

Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, 92–96.
<https://doi.org/10.25080/Majora-92bf1922-011>
- Servicio de Rentas Internas. (s. f.). *Portal institucional*. <https://www.sri.gob.ec>
- Silver, E. A., Pyke, D. F., & Thomas, D. J. (2017). *Inventory and production management in supply chains* (4.^a ed.). CRC Press.
- Superintendencia de Compañías, Valores y Seguros. (s. f.). *Portal institucional*.
<https://www.supercias.gob.ec>
- Veritrade. (s. f.). *Veritrade: Inteligencia comercial*. <https://www.veritradecorp.com>
- White, T. (2015). *Hadoop: The definitive guide* (4.^a ed.). O'Reilly Media.
- Wixom, B. H., & Watson, H. J. (2010). The BI-based organization. *International Journal of Business Intelligence Research*, 1(1), 13–28. <https://doi.org/10.4018/jbir.2010071702>
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, 59(11), 56–65. <https://doi.org/10.1145/2934664>