



Maestría en

ADMINISTRACIÓN DE EMPRESAS

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

AUTORES:

Lcda. María Gabriela Flores Avilés

Ing. Ricardo Paul Orbea Andrade

Ing. Erika Alexandra Quinga Zambrano

Lcdo. Gonzalo Fernando Utreras Freire

TUTOR: Phd. Carlos González Luca de Tena

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

MarketPulse Apheresis: Sistema de Business Intelligence para la Identificación de
Oportunidades de Expansión de Mercado de Apheresis Technology

Por

Lcda. María Gabriela Flores Aviles

Ing. Ricardo Paul Orbea Andrade

Ing. Erika Alexandra Quinga Zambrano

Lcdo. Gonzalo Fernando Utreras Freire

Octubre 2026

Aprobado



Cristian Melo
Presidente del Tribunal
Universidad Internacional del Ecuador

Yo, Cristian Javier Melo González e Ignacio Maroto, declaramos que, personalmente conocemos que los graduandos: Lcda. María Gabriela Flores Aviles, Ing. Ricardo Paul Orbea Andrade, Ing. Erika Alexandra Quinga Zambrano y Lcdo. Gonzalo Fernando Utreras Freire 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, Lcda. María Gabriela Flores Aviles, Ing. Ricardo Paul Orbea Andrade, Ing. Erika Alexandra Quinga Zambrano y Lcdo. Gonzalo Fernando Utreras Freire, declaramos bajo juramento que el trabajo de titulación titulado **MarketPulse Apheresis: Sistema de Business Intelligence para la Identificación de Oportunidades de Expansión de Mercado de Apheresis Technology** 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.



María Gabriela Flores Aviles

Correo electrónico:

mgfa1464@gmail.com



Ricardo Paúl Orbea Andrade

Correo electrónico:

ricardoorbea99@gmail.com



Erika Alexandra Quinga Zambrano

Correo electrónico:

erquingaza@uide.edu.ec



Gonzalo Fernando Utreras Freire

Correo electrónico:

goutrerasfr@uide.edu.ec

Autorización de Derechos de Propiedad Intelectual

Nosotros, Lcda. María Gabriela Flores Aviles, Ing. Ricardo Paul Orbea Andrade, Ing. Erika Alexandra Quinga Zambrano y Lcdo. Gonzalo Fernando Utreras Freire en calidad de autores del trabajo de investigación titulado MarketPulse Apheresis: Sistema de Business Intelligence para la Identificación de Oportunidades de Expansión de Mercado de Apheresis Technology, autorizo a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que me pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autor me corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, 06 de octubre de 2026



María Gabriela Flores Aviles

Correo electrónico:

mgfa1464@gmail.com



Ricardo Paul Orbea Andrade

Correo electrónico:

ricardoorbea99@gmail.com



Erika Alexandra Quinga Zambrano

Correo electrónico:

erquingaza@uide.edu

[.ec](mailto:erquingaza@uide.edu)



Gonzalo Fernando Utreras Freire

Correo electrónico:

goutrerasfr@uide.edu.ec

Dedicatorias y Agradecimientos

Los autores dedican este proyecto a la Fundación Guillain Barré Ecuador, por abrir sus registros de campo y hacer posible calibrar este modelo con evidencia real antes de escalarlo, y a Carlos González Luca de Tena, director de esta Especialización, por su acompañamiento y exigencia metodológica a lo largo de los ciclos de retroalimentación de este proyecto.

A Rafaella y Mía Camila, razón de cada página y motivo de cada pausa: les dedico este esfuerzo, como casi todo lo que vale la pena en mi vida.

Y a quienes me acompañaron en el camino, docentes, familia y amigos, aunque sus nombres no aparezcan en la portada: esta tesis es un poco de cada uno, y la gratitud, toda mía.

Gonzalo

A Marcita, mi ejemplo de fortaleza, perseverancia y amor incondicional. Gracias por darme raíces fuertes, alas para soñar y la confianza para llegar tan lejos como me lo proponga. Por creer siempre en mí y ser parte de cada sueño que me ha traído hasta aquí.

Con todo mi amor y gratitud, mamá.

Gabriela

Dedico este trabajo a Dios, por la fortaleza en cada etapa; a mi familia, por su amor incondicional y su paciencia infinita; y a todas las personas que, con una palabra oportuna o un gesto sencillo, hicieron más liviano el camino.

Este logro es el reflejo de su confianza en mí.

Alexandra

La dedicatoria del siguiente trabajo va orientada al esfuerzo, sacrificio y resistencia que hay detrás de tomar la decisión de ampliar el conocimiento y trascender hacia nuevos horizontes en nuestros mundos.

Paúl

ÍNDICE DEL DOCUMENTO

Contenido

Dedicatorias y Agradecimientos	6
ÍNDICE DEL DOCUMENTO	7
ÍNDICE DE FIGURAS Y TABLAS.....	10
INTRODUCCIÓN Y OBJETO DEL PROYECTO.....	14
PARTE I: BUSINESS INTELLIGENCE	15
Capítulo 1 - Introducción al Business Intelligence.....	16
Presentación de la empresa.....	16
Objetivo y alcance del trabajo	17
Identificación de GAP'S y definición de KPI'S.....	27
Aplicación del modelo SOM: cálculo, validación, sensibilidad y escenarios	29
Conclusiones.....	32
Capítulo 2 - Fuentes de datos	34
Objetivo y alcance del trabajo	34
Conclusiones.....	37
Capítulo 3 - Procesos ETL	38
Objetivo y alcance del trabajo	38
Conclusiones.....	44
Capítulo 4 - Visualización de datos para la toma de decisiones.....	45
Objetivo y alcance del trabajo	45
Conclusiones.....	57
Capítulo 5 - Organización del área de Business Intelligence	59
Objetivo y alcance del trabajo	59
Situación actual (AS-IS).....	59
Problemas organizativos.....	60
Carencias en recursos y formación.....	60
Modelo organizativo TO-BE	61

Organigrama	61
Matriz RACI	61
Gestión de solicitudes e incidencias	62
Formación.....	63
Outsourcing	63
Costes	64
Beneficios	65
Business Case.....	65
ROI, Payback y escenarios	66
Conclusiones.....	66
PARTE II: BIG DATA.....	68
Capítulo 1 - Datos No Estructurados, Big Data y Nuevas Fuentes	69
Objetivo y alcance del trabajo	69
Conclusiones.....	76
Capítulo 2 - Arquitectura Cloud, Selección de Proveedores e Integración Big Data-BI	77
Objetivo y alcance del trabajo	77
Conclusiones.....	81
Capítulo 3 - Arquitectura Open Source del Proyecto	83
Punto de partida: requisitos y arquitectura lógica por capas	83
Selección crítica de frameworks open source	84
Arquitectura final y flujo de datos	85
Seguridad, gobierno, calidad y observabilidad.....	87
Costo total de propiedad, licencias y riesgos	87
Ejemplo demostrativo: primer ciclo de orquestación con Airflow	88
Conclusiones.....	89
Capítulo 4 - Inteligencia Artificial aplicada a MarketPulse Apheresis.....	91
Objetivo del modelo de data science y machine learning	91
Fuentes de datos utilizadas.....	94

Área de aplicación y departamentos implicados.....	99
Mejora esperada	101
Business case.....	102
Ejemplo demostrativo: primera ejecución del modelo.....	105
Riesgos, ética y normativa de datos	107
Conclusiones.....	108
Capítulo 5 - Analítica en tiempo real: alerta de demanda aguda no cubierta	110
Objetivo del caso de uso en tiempo real	110
Fuentes de datos y arquitectura	116
Área y departamentos implicados	118
Mejora esperada	119
Business case y comparación de velocidades	121
Ejemplo demostrativo: una alerta trazada de principio a fin	123
Riesgos y protección de datos.....	124
Conclusiones y hoja de ruta	125
CONCLUSIONES Y APLICACIONES	127
Resultados consolidados del proyecto	127
Aplicación práctica en Apheresis Technology	128
Contraste con los criterios de escalamiento hacia Estados Unidos	129
Hoja de ruta integrada.....	131
Conclusión final	131
REFERENCIAS	132

ÍNDICE DE FIGURAS Y TABLAS

Ilustración 1 Proceso AS-IS de identificación de oportunidades de mercado	26
Ilustración 2 Proceso TO-BE de identificación de oportunidades de mercado.....	27
Ilustración 3 Diagrama de flujo del proceso ETL.....	39
Ilustración 4 Modelo dimensional preliminar del Data Warehouse.....	42
Ilustración 5 Matriz de priorización clínico-comercial (2x2)	43
Ilustración 6 Mapa de oportunidad por semáforo de SOM (Estados Unidos)	44
Ilustración 7 Mockup del Dashboard Ejecutivo	49
Ilustración 8 Mockup del Dashboard de Canales Comerciales.....	50
Ilustración 9 Mockup del Dashboard de Calidad BI	51
Ilustración 10 Trazabilidad completa de un ciclo: dato real → fuente → extracción → transformación → control de calidad → carga (Capítulo 3) → SOM → ranking → decisión comercial (Capítulo 4)	57
Ilustración 11 Arquitectura final de integración Big Data-Data Warehouse-BI de MarketPulse Apheresis	81
Ilustración 12 Arquitectura lógica por capas de la solución open source (sin marcas comerciales).....	84
Ilustración 13 Arquitectura final open source de MarketPulse Apheresis.....	85
Ilustración 14 Integración del modelo predictivo en la arquitectura y flujo de la predicción a la acción.....	99
Ilustración 15 Curva de pérdida de valor de la oferta de derivación según la latencia de la decisión.....	113
Ilustración 16 Arquitectura Lambda ligera: vía rápida de micro-lotes y vía lenta existente	117

Tabla 1 Estructura de la Dirección Comercial de Expansión de Mercados	18
Tabla 2 Objetivos por área e información actual y necesidades detectadas.....	19
Tabla 3 KPI's clave en el proceso de expansión de mercado.....	20
Tabla 4 Ponderación inicial de variables del SOM.....	22
Tabla 5 Matriz de comparación pareada utilizada para derivar los pesos iniciales del SOM.....	22
Tabla 6 Indicadores de éxito comercial del modelo	24
Tabla 7 Indicadores de Conocimiento (K).....	28
Tabla 8 Indicadores de Organización y Motivación (O/M)	29
Tabla 9 Business Model Canvas de Apheresis Technology.....	34
Tabla 10 Roles y políticas de gobierno del dato en MarketPulse Apheresis.	35

Tabla 11 Tipos de datos.....	36
Tabla 12 Frecuencia de carga y accesibilidad	41
Tabla 13 Necesidades de información por usuario y decisión	45
Tabla 14 Análisis y selección de la herramienta de visualización.....	46
Tabla 15 Análisis y selección de la herramienta de visualización.....	47
Tabla 16 Análisis y selección de la herramienta de visualización.....	48
Tabla 17 Alertas y excepciones.....	51
Tabla 18 Revisión de los objetivos iniciales	53
Tabla 19 Ejemplo completo: del dato a la decisión	55
Tabla 20 Diagnóstico AS-IS por área	59
Tabla 21 Organigrama propuesto de la Dirección Comercial de Expansión de Mercados.....	61
Tabla 22 Acuerdos de nivel de servicio (SLA) por severidad	62
Tabla 23 Necesidades de formación por perfil	63
Tabla 24 Matriz Make or Buy	63
Tabla 25 Plantilla y costo cargado anual del equipo de BI y Datos.....	64
Tabla 26 Inversión inicial específica del departamento (USD).....	64
Tabla 27 Beneficios anuales estimados en régimen de operación (USD)	65
Tabla 28 Business Case: costes y beneficios anuales (USD)	65
Tabla 29 Escenarios de beneficio acumulado a 3 años (USD).....	66
Tabla 30 Decisiones con información insuficiente y limitación identificada	70
Tabla 31 Inventario de fuentes de datos de MarketPulse Apheresis (estructuradas y no estructuradas)	71
Tabla 32 Descripción de las fuentes de datos no estructurados priorizadas.....	71
Tabla 33 Matriz de relación entre fuentes estructuradas y no estructuradas	72
Tabla 34 Ejemplo real de un ciclo de extracción y análisis de datos no estructurados	74
Tabla 35 Matriz ponderada de comparación de proveedores cloud	78
Tabla 36 Matriz de riesgos tecnológicos.....	79
Tabla 37 Flujo de datos de la arquitectura final.....	80
Tabla 38 Matriz comparativa de frameworks open source por capa, con alternativa elegida y descartada	84
Tabla 39 Gobierno, calidad y observabilidad de la capa open source	87
Tabla 40 TCO estimado de la capa open source (valores de referencia para el mercado ecuatoriano, sujetos a validación con datos reales del piloto).....	87
Tabla 41 Ejemplo real del primer ciclo de orquestación Airflow + dbt Core (ambiente de pruebas).....	88

Tabla 42 Matriz de priorización de casos de uso de machine learning (pesos fijados antes de puntuar).....	92
Tabla 43 Ficha del objetivo del modelo predictivo de demanda no cubierta	92
Tabla 44 Comparación de técnicas con línea base sin IA y alternativas descartadas	93
Tabla 45 Criterios de éxito del modelo.....	94
Tabla 46 Inventario de fuentes utilizadas por el modelo	95
Tabla 47 Matriz de calidad de las fuentes del modelo.....	95
Tabla 48 Variables principales del modelo y ventana de cálculo.....	96
Tabla 49 Reentrenamientos con etiquetas cerradas en la validación temporal y en producción.....	98
Tabla 50 Matriz RACI del modelo predictivo	99
Tabla 51 Flujo mensual de decisión con supervisión humana.....	100
Tabla 52 Palancas de valor, KPI, línea base y objetivo.....	101
Tabla 53 Conversión de la mejora técnica en mejora real (USD/año).....	101
Tabla 54 Diseño del piloto con grupo de control	102
Tabla 55 Costes de inversión inicial y recurrentes del modelo (USD).....	102
Tabla 56 Cálculo del beneficio técnico anual por palanca (USD).....	103
Tabla 57 Business case a 3 años con adopción progresiva (USD)	104
Tabla 58 Escenarios y punto de equilibrio (USD).....	104
Tabla 59 Ejemplo ilustrativo de la primera ejecución del modelo predictivo (ambiente de pruebas).....	105
Tabla 60 Validación temporal ilustrativa en el periodo de prueba (septiembre 2025 a mayo 2026; 1.989 filas; 113 positivos).....	106
Tabla 61 Matriz de riesgos del modelo predictivo	107
Tabla 62 Niveles de analítica de MarketPulse Apheresis y lugar del tiempo real	110
Tabla 63 Matriz de priorización de casos de uso en tiempo real (pesos fijados antes de puntuar).....	111
Tabla 64 Descomposición de la latencia y valor recuperado por nivel de velocidad	113
Tabla 65 Regla de detección de la alerta de demanda aguda (seis condiciones)	114
Tabla 66 Ficha del caso de uso en tiempo real	115
Tabla 67 Criterios de éxito técnicos, operativos y de negocio	115
Tabla 68 Volumen de eventos por flujo y velocidad requerida.....	116
Tabla 69 Explicación de cada flecha de la arquitectura	117
Tabla 70 Flujo de la alerta con destinatario, plazo, escalado y cierre	118
Tabla 71 Palancas de valor y separación respecto del Capítulo 4.....	119
Tabla 72 Cadena de valor de la alerta en régimen (año 3)	120

Tabla 73 Diseño del piloto con implantación escalonada	120
Tabla 74 Costes por alternativa de velocidad (USD).....	121
Tabla 75 Business case a 3 años de los micro-lotes de 15 minutos (USD)	122
Tabla 76 Comparación de las tres velocidades y análisis incremental (USD)	122
Tabla 77 Escenarios del business case de micro-lotes (USD)	123
Tabla 78 Simulacro ilustrativo de una alerta, del evento al Data Warehouse (29/09/2026).....	124
Tabla 79 Riesgos del caso de uso en tiempo real.....	125
Tabla 80 Resultados consolidados por capítulo	127
Tabla 81 Ciclo operativo integrado de MarketPulse Apheresis.....	128
Tabla 82 Estado de los criterios de escalamiento hacia Estados Unidos	129
Tabla 83 Qué se transfiere y qué se adapta al escalar a Estados Unidos.....	130
Tabla 84 Hoja de ruta de MarketPulse Apheresis	131

INTRODUCCIÓN Y OBJETO DEL PROYECTO

El sector de la tecnología médica especializada enfrenta un reto particular al momento de expandirse geográficamente: los dispositivos de alta complejidad clínica, como los equipos de aféresis terapéutica, solo generan valor si se instalan en zonas donde existe una necesidad real y no cubierta. Sin embargo, las decisiones de expansión en este sector suelen tomarse con base en relaciones comerciales, intuición del equipo de ventas o criterios de mercado tradicionales, sin cruzar de forma sistemática la información epidemiológica, demográfica, geográfica y socioeconómica que permitiría identificar objetivamente los mercados de mayor oportunidad.

El presente Trabajo Fin de Especialización tiene como objeto desarrollar MarketPulse Apheresis, un sistema de Business Intelligence orientado a resolver esta limitación para Apheresis Technology, empresa global propietaria del dispositivo GEMMA, utilizado para el intercambio terapéutico de plasma en el tratamiento de enfermedades autoinmunes y neuro-inmunológicas de alta complejidad. El proyecto busca construir un modelo de inteligencia comercial basado en evidencia, que cruce datos epidemiológicos, de capacidad hospitalaria instalada y socioeconómicos para generar un Score de Oportunidad de Mercado (SOM) que permita priorizar objetivamente los estados y condados de Estados Unidos donde la expansión del dispositivo GEMMA tendría mayor impacto clínico y comercial.

Como aporte diferenciador, el modelo se calibra y valida previamente con datos reales de campo de la Fundación Guillain Barré Ecuador, aliada local de Apheresis Technology, lo que otorga solidez metodológica al proceso antes de extrapolarlo al mercado estadounidense. Esta arquitectura, además, se diseña de forma replicable: la misma lógica de scoring podrá aplicarse posteriormente a otros países, cambiando únicamente las fuentes de datos de origen, lo que incrementa el valor estratégico del proyecto más allá de su alcance inicial.

PARTE I: BUSINESS INTELLIGENCE

La primera parte construye la base de Business Intelligence de MarketPulse Apheresis. El Capítulo 1 define el problema de negocio, los KPI y el Score de Oportunidad de Mercado; el Capítulo 2, las fuentes de datos y el Data Warehouse; el Capítulo 3, los procesos ETL y la calidad del dato; el Capítulo 4, la capa de visualización y los cuadros de mando; y el Capítulo 5, la organización del área de BI y Datos que sostiene el sistema.

Capítulo 1 - Introducción al Business Intelligence

Presentación de la empresa

Apheresis Technology (apheresistech.com) es una compañía global de tecnología médica especializada en dispositivos de aféresis terapéutica. Su producto insignia, GEMMA, es un dispositivo de intercambio terapéutico de plasma dirigido al tratamiento de enfermedades autoinmunes y neuro-inmunológicas de alta complejidad, como el Síndrome de Guillain-Barré, la miastenia gravis y la esclerosis múltiple. Sus clientes son clínicas de diálisis, hospitales con unidades de neurología y organizaciones de salud en más de 40 países. En Ecuador, la compañía opera a través de su aliada local, la Fundación Guillain Barré Ecuador, organización que financia y documenta casos reales de pacientes y que actúa como fuente primaria de datos de campo para este proyecto, sirviendo como caso de validación del modelo antes de escalarlo a otros mercados.

Planificación estratégica

Misión

Hacer accesible el tratamiento de enfermedades autoinmunes y neuro-inmunológicas de alta complejidad a través del dispositivo GEMMA, ofreciendo tecnología médica de vanguardia a clínicas y hospitales y resolviendo el problema de la falta de equipamiento especializado en plasmaféresis en mercados donde la demanda clínica existe pero la oferta es insuficiente.

Visión

Consolidarse en un horizonte de 2 a 4 años como el referente global en dispositivos de aféresis terapéutica, duplicando su presencia en el mercado norteamericano mediante la identificación sistemática y basada en datos de zonas con alta demanda no cubierta. Sus objetivos medibles son alcanzar presencia en los 50 estados de Estados Unidos con al menos una clínica equipada con GEMMA, y replicar el modelo de expansión basado en BI a Latinoamérica, Europa y Asia.

Valores de la empresa

- Innovación clínica: compromiso permanente con el desarrollo de tecnología médica que mejore los resultados de los pacientes.
- Accesibilidad: los tratamientos de alta complejidad no deben ser exclusivos de mercados desarrollados, lo que impulsa alianzas con fundaciones como Guillain Barré Ecuador.
- Rigor científico: las decisiones de expansión y desarrollo de producto se fundamentan en evidencia clínica y datos epidemiológicos, no en intuición.

- Escalabilidad responsable: el crecimiento de la red de clínicas se planifica con criterios de sostenibilidad operativa y calidad de atención.

Ventaja competitiva

GEMMA se posiciona, según la propuesta de valor de la compañía, con una eficiencia competitiva en el intercambio plasmático; esta afirmación debe sustentarse con evidencia clínica comparativa publicada antes de emplearse como argumento comercial frente a dispositivos alternativos. Lo que sí constituye una ventaja verificable es que la red de más de 850 clínicas en 40 países genera un efecto de red difícil de replicar, y que el modelo de alianza con fundaciones sociales permite penetrar mercados emergentes antes que la competencia. Este proyecto de BI potencia directamente esta última ventaja al identificar sistemáticamente zonas con demanda no cubierta.

Cifras del negocio

- Más de 850 clínicas activas distribuidas en más de 40 países.
- Mercado global de aféresis valorado en USD 2.650 millones en 2024, en crecimiento sostenido (Market Data Forecast, 2025).
- En Estados Unidos, incidencia de Síndrome de Guillain-Barré estimada entre 1.1 y 1.7 casos por cada 100,000 personas-año, con 10.035 hospitalizaciones ponderadas registradas en 2021 según el National Inpatient Sample (Agency for Healthcare Research and Quality, 2024), y una incidencia estimada entre 1.65 y 1.79 casos por cada 100.000 personas-año (Sejvar et al., 2011).
- En Ecuador, la Fundación Guillain Barré ha documentado más de 600 casos financiados, constituyendo una base de demanda real única para calibrar el modelo antes de aplicarlo a Estados Unidos.

Enfoque estratégico del proyecto hacia la expansión de mercado

El presente proyecto de Business Intelligence se alinea directamente con la visión de crecimiento de Apheresis Technology: en lugar de decidir la expansión con base en relaciones comerciales e intuición, MarketPulse Apheresis busca convertir la planificación estratégica de la compañía en un proceso basado en evidencia, cruzando variables epidemiológicas, geográficas y de capacidad instalada para priorizar objetivamente los mercados de mayor oportunidad para la instalación del dispositivo GEMMA.

Objetivo y alcance del trabajo

Proceso de identificación de oportunidades de mercado

El proceso de expansión de mercado de Apheresis Technology tiene como finalidad la mitigación del riesgo comercial derivado de la instalación del dispositivo GEMMA en zonas geográficas sin demanda real de plasmaféresis, con esto se busca que la capacidad clínica instalada se encuentre en niveles óptimos que permitan a la compañía la maximización del impacto clínico y del retorno sobre la inversión en cada mercado atendido.

Para cumplir con dicho objetivo, se ha designado la responsabilidad del análisis de expansión a la Dirección Comercial de Expansión de Mercados; dicha Dirección es responsable del diseño de las estrategias, planes de acción y priorización de zonas geográficas a través de alternativas que permitan identificar mercados con demanda no cubierta y capacidad de absorción comercial. La Fundación Guillain Barré Ecuador actúa como fuente de datos de campo, aportando registros reales de pacientes, distribución geográfica de la enfermedad y brechas de cobertura clínica en el país, insumo clave para calibrar el modelo antes de escalarlo a otros mercados de Latinoamérica.

Estructura de la Dirección Comercial de Expansión de Mercados

La estructura de la Dirección se compone de áreas estratégicas que son responsables del diseño de criterios de priorización de mercado, áreas operativas que soportan la relación con clínicas y fundaciones aliadas, áreas analíticas que apoyan con información epidemiológica y geoespacial la toma de decisiones, y áreas de gestión que manejan los canales comerciales de distribución del dispositivo GEMMA.

Tabla 1 Estructura de la Dirección Comercial de Expansión de Mercados

Dirección Comercial de Expansión de Mercados
Gerencia Estrategia Expansión LATAM
Gerencia Estrategia Expansión Norteamérica
Gerencia Analítica de Mercado
Sección Inteligencia Epidemiológica
Sección Inteligencia Geoespacial
Gerencia Alianzas Clínicas y Fundaciones
Sección Gestión de Fundaciones Aliadas
Sección Soporte Técnico Clínico (GEMMA)
Gerencia Canales Comerciales

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Áreas y personas que demandan la información

A partir de la estructura de la Dirección Comercial de Expansión de Mercados (Tabla 1), se identifican cuatro áreas que demandan información para la toma de decisiones de expansión, y que actualmente la obtienen de forma manual, fragmentada y sin un repositorio único de datos:

- Dirección Comercial: decide qué zonas geográficas priorizar para la instalación del dispositivo GEMMA. Actualmente se apoya en el criterio del equipo de ventas y en relaciones históricas con clínicas, sin un ranking objetivo de mercados.
- Gerencia Analítica de Mercado: es responsable de consolidar los datos epidemiológicos y geospaciales (incidencia de enfermedades tratables con aféresis, capacidad hospitalaria instalada, variables demográficas). Hoy realiza esta consolidación de forma manual, con fuentes dispersas como el CDC, el National Inpatient Sample de la AHRQ, el CMS Provider Data Catalog y el US Census Bureau.
- Gerencia Alianzas Clínicas y Fundaciones: gestiona la relación con la Fundación Guillain Barré Ecuador y con las clínicas aliadas. La información de campo (pacientes, costos, geografía) se recibe de forma informal, sin trazabilidad sistemática de cobertura.
- Gerencia Canales Comerciales: gestiona los distribuidores regionales y la instalación efectiva del dispositivo. No cuenta hoy con visibilidad de qué mercados están priorizados, por lo que su gestión comercial no está alineada con la demanda no cubierta identificada.

Objetivos de cada área con la información actual y necesidades detectadas

La siguiente tabla resume, para cada área de la Dirección Comercial, el objetivo que persigue con la información, cómo la utiliza actualmente y la necesidad de información detectada que este proyecto de BI busca resolver.

Tabla 2 Objetivos por área e información actual y necesidades detectadas

Área			

Dirección Comercial	Priorizar zonas geográficas para la instalación del dispositivo GEMMA	Criterio comercial informal y relaciones históricas con clínicas	Ranking objetivo de mercados basado en el Score de Oportunidad de Mercado (SOM)
Gerencia Analítica de Mercado	Consolidar variables epidemiológicas, geográficas y de capacidad instalada	Consolidación manual en hojas de cálculo desde fuentes dispersas (CDC, NIS, CMS, Census)	Pipeline ETL automatizado y Data Warehouse único (PostgreSQL)
Gerencia Alianzas Clínicas y Fundaciones	Gestionar la relación con clínicas y con la Fundación Guillain Barré Ecuador	Comunicación informal, sin trazabilidad sistemática de cobertura clínica	Indicador sistemático de Verificación de Cobertura Clínica
Gerencia Canales Comerciales	Dirigir a los distribuidores regionales hacia los mercados de mayor oportunidad	Sin visibilidad de mercados priorizados; gestión desalineada de la demanda no cubierta	Mapa interactivo y ranking de mercados priorizados para orientar la gestión comercial

Fuente: Elaborado por los autores

Fecha: Julio, 2026

KPI's clave en el proceso de expansión de mercado

El sector de tecnología médica presenta unos KPI que son claves en el proceso de identificación de oportunidades, y que permiten comparar las diferentes zonas geográficas para identificar el riesgo comercial al que la compañía se enfrenta en el desarrollo de su actividad natural, que es la instalación de nuevos dispositivos GEMMA.

Conceptos Clave:

- DNC: Demanda No Cubierta.
- SOM: Score de Oportunidad de Mercado.
- Zona: Unidad geográfica de análisis (cantón, provincia o país).

Tabla 3 KPI's clave en el proceso de expansión de mercado

Nombre Indicador	Fórmula	Descripción
Índice de Demanda No Cubierta (DNC)	$(\text{Incidencia estimada de enfermedad tratable con aféresis} - \text{Capacidad clínica instalada}) / \text{Incidencia estimada}$	Mide la proporción de pacientes potenciales en una zona geográfica que no tiene acceso a tratamiento de plasmaféresis, identificando el vacío de cobertura clínica.
Score de Oportunidad de Mercado (SOM)	$\sum (\text{peso}_i \times \text{variable}_i), i = \{\text{DNC, densidad poblacional, capacidad de pago, accesibilidad geográfica}\}$	Combina de forma ponderada las variables epidemiológicas, geográficas y socioeconómicas de una zona para priorizarla como candidata de expansión del dispositivo GEMMA.

Tasa de Penetración de Mercado (TPM)	Clínicas con GEMMA instalado / Total de clínicas potenciales en la zona	Identifica el avance de la instalación del dispositivo dentro de una zona ya priorizada, permitiendo medir el resultado comercial de la expansión.
--------------------------------------	---	--

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Precisión metodológica del Índice de Demanda No Cubierta (DNC)

La fórmula del DNC exige que numerador y denominador estén expresados en unidades comparables. Por ello, la incidencia estimada se convierte en pacientes potenciales tratables al año (incidencia por 100,000 habitantes $\times$ población de la zona / 100,000), y la capacidad clínica instalada se convierte en capacidad anual estimada de tratamiento, calculada como el número de equipos de aféresis en la zona $\times$ sesiones posibles por equipo al año $\times$ tasa de ocupación efectiva, ajustada por el radio de cobertura de cada clínica: una clínica ubicada en un cantón o condado limítrofe puede atender pacientes de zonas vecinas, por lo que su capacidad se prorroga entre las zonas dentro de su radio de atracción. El DNC se calcula entonces como (pacientes potenciales tratables al año – capacidad anual estimada de tratamiento) / pacientes potenciales tratables al año, evitando comparar una tasa de incidencia con un conteo absoluto de clínicas.

La capacidad clínica instalada, a su vez, requiere las siguientes variables por clínica: número de equipos de aféresis, sesiones posibles por equipo al año, número de especialistas certificados disponibles, tasa de ocupación actual, radio de cobertura geográfica declarado y existencia de centros alternativos (equipos de la competencia) dentro de ese mismo radio. Sin estas seis variables, la capacidad instalada se subestima o sobreestima y el DNC pierde validez como criterio de priorización.

Metodología de construcción y ponderación del Score de Oportunidad de Mercado (SOM)

El Score de Oportunidad de Mercado (SOM) constituye el elemento central del modelo, por lo que su construcción sigue tres etapas. Primero, estandarización: cada variable de entrada (DNC, densidad poblacional, capacidad de pago o cobertura de seguro médico, y accesibilidad geográfica) se normaliza en una escala de 0 a 1 por zona geográfica, para que todas sean comparables independientemente de su unidad de medida original. Segundo, ponderación inicial: los pesos (peso_i) se asignan mediante un Proceso Analítico Jerárquico (AHP), a partir del juicio experto de la Dirección Comercial y de la Gerencia Analítica de Mercado, priorizando la Demanda No Cubierta por su relevancia clínica directa.

Tercero, calibración estadística: los pesos iniciales se ajustan utilizando los datos reales de la Fundación Guillain Barré Ecuador como caso de validación, mediante un análisis de correlación entre el SOM calculado y la demanda real observada en campo; el objetivo es maximizar esa correlación antes de aplicar el modelo calibrado a las zonas de Estados Unidos. Finalmente, se realiza un análisis de sensibilidad de los pesos para verificar la estabilidad del ranking de mercados ante variaciones razonables en la ponderación.

Tabla 4 Ponderación inicial de variables del SOM.

Variable		
Demanda No Cubierta (DNC)		
Capacidad de pago / cobertura de seguro médico		
Densidad poblacional		
Accesibilidad geográfica		

Fuente: Elaborado por los autores

Fecha: Julio, 2026

El procedimiento seguido para asignar cada peso se basa en el Proceso Analítico Jerárquico (AHP). Los expertos de la Dirección Comercial y de la Gerencia Analítica de Mercado comparan las variables por pares, utilizando la escala fundamental de Saaty (1 a 9), y el resultado se organiza en una matriz de comparación pareada, a partir de la cual se calcula el vector propio principal que determina los pesos relativos de la tabla anterior.

Tabla 5 Matriz de comparación pareada utilizada para derivar los pesos iniciales del SOM

Variable				
Demanda No Cubierta (DNC)				
Capacidad de pago / seguro médico				
Densidad poblacional				
Accesibilidad geográfica				

Fuente: Elaborado por los autores

Verificación de consistencia del AHP

A partir de la matriz de comparación pareada de la tabla anterior, se calcula el vector propio principal mediante el método de normalización de columnas y promedio de filas, verificado con el cálculo exacto del autovalor principal ($\lambda_{\text{máx}}$). El autovalor principal obtenido es $\lambda_{\text{máx}} = 4.031$, con un vector propio normalizado de 0.467 (DNC), 0.277 (capacidad de pago), 0.160 (densidad poblacional) y 0.095 (accesibilidad geográfica). El Índice de Consistencia se calcula como $CI = (\lambda_{\text{máx}} - n) / (n - 1)$, con $n = 4$, lo que da $CI = 0.0103$. Usando el Índice Aleatorio de Saaty para $n = 4$ ($RI = 0.90$), la Razón de Consistencia resulta $RC = CI / RI = 0.0115$, muy por debajo del umbral de aceptación de 0.10. Esto confirma que los juicios de comparación pareada de los expertos de la Dirección Comercial y de la Gerencia Analítica de Mercado son consistentes y no requieren revisión.

Cabe señalar que los pesos exactos derivados del vector propio (0.467; 0.277; 0.160; 0.095) difieren de los pesos redondeados presentados inicialmente (0.40; 0.25; 0.20; 0.15). Esta diferencia debe resolverse explícitamente antes del piloto: o se ajusta la matriz de comparación pareada para reflejar con mayor fidelidad el peso deseado de cada variable, o se adopta directamente el vector propio como ponderación inicial del modelo. Se recomienda esta segunda opción, por preservar la trazabilidad matemática que exige el método AHP; los pesos así obtenidos serán los que se calibren estadísticamente con los datos reales de Ecuador.

La matriz de comparación pareada se construyó mediante un ejercicio estructurado con cuatro expertos internos de Apheresis Technology: el Gerente de Analítica de Mercado, el Gerente de Estrategia de Expansión Norteamérica, el Gerente de Estrategia de Expansión LATAM y un especialista clínico designado por la Fundación Guillain Barré Ecuador. Se seleccionó a estos cuatro roles -y no a un grupo mayor- por ser los responsables directos de las decisiones que el SOM pretende sustentar (priorización comercial, viabilidad de expansión y relevancia clínica), siguiendo la recomendación metodológica de Saaty (1980) de limitar el panel de expertos a quienes tienen responsabilidad directa sobre el criterio evaluado. El procedimiento siguió tres pasos: cada experto completó de forma individual la matriz de comparación pareada usando la escala fundamental de Saaty (1 a 9); los cuatro juicios individuales se agregaron mediante la media geométrica de cada celda, método estándar para consolidar matrices AHP de múltiples expertos; y las discrepancias superiores a 2 puntos en la escala de Saaty se resolvieron en una sesión de consenso moderada por la Gerencia Analítica de Mercado. Los criterios de ponderación aplicados por los expertos fueron la relevancia clínica de la variable, su viabilidad comercial y su alineación con la estrategia de expansión de la compañía, en ese orden de prioridad.

Una vez calibrados los pesos con los datos reales de la Fundación Guillain Barré Ecuador, se ejecuta un análisis de sensibilidad que consiste en perturbar cada peso en un rango de $\pm 15\%$ (mientras los demás se reescalan proporcionalmente para mantener la suma en 1) y recalcular el ranking de zonas priorizadas en cada escenario. Si el orden relativo de las zonas con mayor SOM se mantiene estable ante estas variaciones, se concluye que el modelo es robusto; si el ranking cambia de forma significativa, se revisa la ponderación antes de aplicar el modelo a las zonas de Estados Unidos.

Indicadores de éxito comercial del modelo

Se incorporan indicadores adicionales que permiten evaluar el modelo desde una perspectiva estrictamente comercial, más allá de la priorización epidemiológica de mercados:

Tabla 6 Indicadores de éxito comercial del modelo

Nombre Indicador		
Incremento Potencial de Ventas (IPV)		
Retorno Esperado de la Expansión (REE)		
Tasa de Validación del Modelo (TVM)		
Retorno de la Inversión (ROI)		
Periodo de Recuperación (PRI)		

Fuente: Elaborado por los autores
Fecha: Julio, 2026

Necesidad de intervención en la Gestión de Expansión Avanzada

En el proceso de expansión comercial existe una etapa que es la de validación de cobertura clínica. En la actualidad, el momento en que se toma la decisión de priorizar una zona para la instalación del dispositivo GEMMA está basado en el criterio comercial del equipo de ventas y en relaciones históricas con clínicas, y no en un ejercicio de análisis de datos que realmente determine el momento idóneo para iniciar el proceso de expansión. Inicialmente, el objetivo es identificar si el criterio que se está utilizando para tomar esta decisión es el idóneo, y para ello se debe analizar la historia epidemiológica e identificar, a partir de curvas de incidencia y cobertura, el momento en que efectivamente se debe priorizar una zona para la expansión.

A partir de esta decisión, existe un proceso que se llama Verificación de Cobertura Clínica, que consiste en construir cierta información de cada zona y consultar en diferentes fuentes las clínicas que ya poseen capacidad instalada de aféresis, para no duplicar esfuerzos comerciales. En muchos de los casos se identifica que, en el proceso de verificación, ya existe una clínica con capacidad instalada en la zona analizada, implicando un esfuerzo y gasto innecesario en el proceso de análisis. Lo que se está buscando es construir una ruta en el aplicativo de Business Intelligence que se conecte con las fuentes de la Fundación Guillain Barré Ecuador y con fuentes públicas de salud, para construir la base en la que, de manera automática y a partir de unos parámetros estándar, se aplican los criterios de priorización y se comienza a excluir aquellas zonas que ya cuentan con cobertura clínica suficiente; con esto se optimizaría el gasto y el recurso del equipo comercial que ejecuta la actividad.

Representación del proceso actual (AS-IS) y del proceso futuro (TO-BE)

Para dimensionar con precisión la intervención propuesta, se documenta el proceso actual de identificación de oportunidades de mercado (AS-IS) y se contrasta con el proceso que MarketPulse Apheresis habilitará (TO-BE). El proceso AS-IS se compone de seis fases manuales: detección informal de mercado (Dirección Comercial; sin fuente sistemática, basada en relaciones con distribuidores); recopilación manual de datos epidemiológicos y geográficos (Gerencia Analítica de Mercado; fuentes CDC, NIS, CMS y Census consultadas por separado; aproximadamente 32 horas de consolidación por ciclo); análisis epidemiológico manual en hojas de cálculo (Gerencia Analítica de Mercado; riesgo de error de transcripción y doble conteo entre fuentes); búsqueda manual de clínicas competidoras (Gerencia Alianzas Clínicas; sin base de datos centralizada); verificación de cobertura clínica (Gerencia Alianzas Clínicas; gasto estimado de USD 960 por zona verificada, con casos en los que la verificación confirma cobertura ya existente y el análisis se descarta); y decisión comercial final basada en criterio del equipo de ventas (Dirección Comercial; sin ranking objetivo de respaldo). El

proceso completo toma hoy entre 4 y 6 meses desde la detección hasta la decisión de expansión.

Ilustración 1 Proceso AS-IS de identificación de oportunidades de mercado



Fuente: Elaborado por los autores

Fecha: Julio, 2026

El proceso TO-BE reemplaza estas seis fases manuales por un flujo automatizado: fuentes públicas y de la Fundación Guillain Barré Ecuador → proceso ETL en Python → Data Warehouse en PostgreSQL → cálculo automático del DNC → cálculo del SOM calibrado → filtro automático de cobertura clínica verificada → ranking de zonas → dashboard de Power BI → decisión de expansión por parte de la Dirección Comercial. Este proceso reduce la consolidación de 32 a un objetivo de 10 horas por ciclo, el gasto de verificación de USD 960 a una meta de USD 640 por zona (al descartar automáticamente zonas con cobertura ya confirmada en el Data Warehouse), y el ciclo completo de priorización de 6 a un objetivo de 2 meses. A diferencia del AS-IS, cada paso del TO-BE queda trazado en el Data Warehouse, lo que permite auditar por qué una zona fue o no priorizada.

Ilustración 2 Proceso TO-BE de identificación de oportunidades de mercado



Fuente: Elaborado por los autores

Fecha: Julio, 2026

Objetivo de la intervención

- Estructurar una metodología que permita a la compañía identificar el momento idóneo para la toma de decisión de priorizar una zona geográfica para expansión, basado en el comportamiento de la demanda epidemiológica y la capacidad clínica instalada, y no en el criterio comercial informal del equipo de ventas.
- Diseñar una herramienta de Business Intelligence que permita la optimización de los recursos, tanto físicos como monetarios, y garantice una calidad de información a través de la extracción de la información que reposa en los registros de la Fundación Guillain Barré Ecuador y en fuentes de datos abiertos, buscando así garantizar confiabilidad y oportunidad en el proceso de expansión.

Identificación de GAP'S y definición de KPI'S

I.- Conocimiento (K)

Identificación del momento de priorización

La organización tiene establecido por criterio comercial el momento de inicio de análisis de expansión hacia una zona. Esta es una definición que está alineada a la experiencia del equipo de ventas y que se encuentra argumentada más en relaciones comerciales que en ejercicios de análisis de datos que realmente determinen el momento idóneo para establecer el inicio de la priorización.

Para este ejercicio se establece calcular la Demanda No Cubierta acumulada por zona geográfica, y a partir de ahí identificar el momento en que la gestión comercial informal deja de ser suficiente, y se debe pasar a una estrategia de priorización basada en datos.

Identificación de zonas objeto de Verificación de Cobertura Clínica

La organización realiza un proceso de verificación de cobertura clínica a todas las zonas que son objeto de análisis de expansión, independientemente de que esta verificación genere valor en el proceso comercial. En muchos de los casos se identifica que, en el proceso de actualización, la verificación arroja como resultado que la zona ya cuenta con una clínica con capacidad instalada de aféresis, implicando un esfuerzo y gasto innecesario en el proceso de verificación.

Tabla 7 Indicadores de Conocimiento (K)

Tipo de indicador			
Score acumulado de oportunidad en % (SOM por zona priorizada)			
Verificación de cobertura a realizar: Clínicas que necesitan VC / total clínicas candidatas a priorizar			

Fuente: Elaborado por los autores

Fecha: Julio, 2026

II.- Organización (O)

Inversión de horas hombre en la consolidación manual de datos de campo

El proceso de consolidación es dispendioso debido a que se debe construir cada mes un repositorio de la base de zonas objeto de la verificación de cobertura clínica. Para esto, el analista debe diligenciar una planilla con los datos epidemiológicos y geográficos de cada zona, e indicar todas las fuentes en las que se debe consultar la existencia de clínicas con capacidad instalada. En este proceso, la consulta de cobertura no se realiza de forma sistemática para evitar incrementar los tiempos de respuesta, y porque es una actividad realizada por un número reducido de analistas.

Gasto incurrido en el proceso de Verificación de Cobertura Clínica

Teniendo en cuenta que el proceso define que se debe realizar verificación de cobertura clínica a todas las zonas candidatas, es fundamental identificar la variación del gasto asociando únicamente a las zonas que no tienen una clínica con capacidad instalada de aféresis y que, por tal motivo, no serían objeto de nuevas verificaciones en ciclos posteriores.

Tabla 8 Indicadores de Organización y Motivación (O/M)

Tipo de indicador			
Inversión de tiempo en consolidación manual de datos: horas invertidas / Total horas mes			
Gasto VC: Gasto promedio de una verificación de cobertura * Clínicas a las que se debe realizar una VC			

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Aplicación del modelo SOM: cálculo, validación, sensibilidad y escenarios

Para mostrar el modelo en funcionamiento y no solo su diseño metodológico, se aplica a continuación el procedimiento completo -estandarización, ponderación AHP, cálculo del SOM, validación, sensibilidad y escenarios- sobre un conjunto de cinco provincias candidatas de Ecuador. Los valores de entrada de esta sección son cifras ilustrativas, construidas por el equipo para demostrar el procedimiento con la misma exigencia numérica que el resto del documento; deberán sustituirse por los valores reales una vez la Fundación Guillain Barré Ecuador entregue al menos dos

ciclos completos de datos de campo, tal como se plantea en la revisión de objetivos del Capítulo 4.

Cálculo del SOM para cinco provincias candidatas

Siguiendo la recomendación de la sección de verificación de consistencia del AHP, el cálculo usa los pesos exactos del vector propio (DNC 0.467; capacidad de pago 0.277; densidad poblacional 0.160; accesibilidad 0.095) sobre variables ya normalizadas de 0 a 1 por zona:

Provincia						
Los Ríos						
Santo Domingo						
Esmeraldas						
Sucumbíos						
Cañar						

Fuente: Elaborado por los autores.

Fecha: Julio, 2026

El ranking resultante -Los Ríos, Santo Domingo, Esmeraldas, Sucumbíos y Cañar, en ese orden- ubica a las cinco provincias en la banda de oportunidad media (SOM 0.40-0.69) definida en el Capítulo 3, de forma consistente con el ejemplo del Cantón X (SOM 0.62) trazado en el Capítulo 4.

Validación del modelo con casos de la Fundación Guillain Barré Ecuador

En el ejercicio ilustrativo, sobre las 24 provincias del piloto el modelo clasifica 9 como de alta oportunidad. Contrastando esa clasificación contra una demanda confirmada construida por el equipo con el mismo orden de magnitud de los casos financiados por la Fundación, se obtiene la siguiente matriz de confusión, que muestra el procedimiento de validación y no un resultado ya obtenido; la validación real se ejecutará cuando la Fundación entregue los dos ciclos completos de datos de campo previstos en el Capítulo 4:

Métrica		
Verdaderos positivos (VP)		
Falsos positivos (FP)		
Falsos negativos (FN)		
Verdaderos negativos (VN)		

Precisión = $VP/(VP+FP)$	0.78	De las zonas priorizadas, el 78% tenía demanda real
Sensibilidad = $VP/(VP+FN)$	0.70	El modelo detecta el 70% de las zonas con demanda real
Especificidad = $VN/(VN+FP)$	0.86	El modelo descarta correctamente el 86% de las zonas sin demanda
Tasa de Validación del Modelo (TVM)	0.78	Coincide con la precisión bajo esta definición del KPI (Capítulo 1)
Correlación de Pearson (SOM vs. demanda real)	$r = 0.81$	Asociación fuerte y positiva entre el SOM y la demanda observada

Fuente: Elaborado por los autores.

Fecha: Julio, 2026

Una sensibilidad de 0.70 indica que 3 de cada 10 zonas con demanda real todavía no son detectadas por el modelo en su calibración inicial; este es precisamente el tipo de brecha que el proceso de calibración estadística descrito en la Metodología del SOM busca reducir antes de escalar el modelo a Estados Unidos.

Análisis de sensibilidad de los pesos ($\pm 15\%$)

Perturbando el peso de la Demanda No Cubierta en $\pm 15\%$ -y reescalando proporcionalmente los tres pesos restantes para mantener la suma en 1- se recalcula el SOM y el ranking de las cinco provincias en tres escenarios de ponderación:

Provincia			
Los Ríos			
Santo Domingo			
Esmeraldas			
Cañar			
Sucumbíos			

Fuente: Elaborado por los autores.

Fecha: Julio, 2026

La primera posición (Los Ríos) se mantiene estable en los tres escenarios. Solo se observan intercambios menores entre posiciones adyacentes y de puntuación cercana (Santo Domingo/Esmeraldas en segundo y tercer lugar; Cañar/Sucumbíos en cuarto y quinto), lo que arroja una correlación de Spearman entre escenarios cercana a 0.90. Se concluye que el modelo es razonablemente robusto ante variaciones moderadas de ponderación, aunque el orden entre zonas de puntuación muy cercana no debe interpretarse como definitivo hasta calibrar los pesos con datos reales.

Escenarios conservador, base y expansivo

Sobre el ranking anterior, se modelan tres escenarios de expansión comercial que combinan el número de zonas atendidas por año, el ticket promedio de GEMMA y el costo de expansión por zona, para estimar el Incremento Potencial de Ventas (IPV) y el Retorno Esperado de la Expansión (REE) definidos en el Capítulo 1:

Escenario					
Conservador					
Base					
Expansivo					

Fuente: Elaborado por los autores.

Fecha: Julio, 2026

Indicadores ROI y PRI del piloto (escenario base)

Tomando el escenario base como referencia -ingresos incrementales netos de USD 110.000 (IPV menos costo de expansión)- y el costo total de implementación del sistema de USD 45.000 desarrollado en el análisis económico del Capítulo 4, el ROI del piloto resulta $(110.000 - 45.000) / 45.000 = 1.44$, es decir, 144%; y el Periodo de Recuperación (PRI) resulta $45.000 / (110.000 / 12) \approx 4.9$ meses. Estos indicadores usan la fórmula simplificada del Capítulo 1 (solo ingresos incrementales); el análisis financiero completo del Capítulo 4 incorpora además los costos operativos recurrentes para calcular el Valor Actual Neto y la Tasa Interna de Retorno del proyecto.

En conjunto, esta sección demuestra el ciclo completo del modelo, de la ponderación AHP a la decisión financiera, con cifras internamente consistentes.

Conclusiones

Este capítulo profundiza en la justificación estadística de la ponderación del SOM y en la gestión del riesgo del proyecto, dos elementos que fortalecen el rigor metodológico del modelo. Los aspectos relacionados con la arquitectura técnica completa (ETL, modelo dimensional y herramientas de visualización), el modelo entidad-relación del Data Warehouse, el análisis económico detallado y el cronograma de implementación se desarrollarán progresivamente en los capítulos siguientes, conforme al alcance definido en la Introducción del proyecto.

El BI consiste en la creación de conocimiento a partir de la recopilación y el análisis de datos generados por una empresa. La clave está en extraer la información más

relevante de Apheresis Technology y de la Fundación Guillain Barré Ecuador para aprovecharla en una mejora notable de la efectividad comercial y del impacto clínico del dispositivo GEMMA. Por tanto, el sistema MarketPulse Apheresis servirá para recopilar la información generada y analizarla en forma de score, mapa de oportunidades y ranking de mercados.

Capítulo 2 - Fuentes de datos

Objetivo y alcance del trabajo

Modelo de negocio (Business Model Canvas)

Antes de definir los propósitos y las fuentes de datos de MarketPulse Apheresis, se parte del Business Model Canvas de Apheresis Technology. Este modelo permite identificar de forma explícita los segmentos de cliente de la compañía, y a partir de ellos derivar las necesidades de información que sustentan el resto del capítulo: cada tipo de dato que se recolecta y cada proceso de transformación que se construye responde a la necesidad de conocer, priorizar o dar servicio a alguno de estos segmentos.

Tabla 9 Business Model Canvas de Apheresis Technology.

Socios Clave Fundación Guillaín Barré Ecuador y fundaciones aliadas; CDC, CMS y US Census Bureau; distribuidores regionales de equipos médicos.				

Fuente: Elaborado por los autores
Fecha: Julio, 2026

Proceso de consolidación de información

Para la administración y control de la información, MarketPulse Apheresis consolida un proyecto de Data Warehouse en PostgreSQL, con el fin de garantizar la calidad de los datos y establecer las condiciones que permiten su tratamiento. En este proceso, la Gerencia Analítica de Mercado es responsable de establecer las políticas y definiciones estratégicas relacionadas con el uso de la información.

Con esta implementación se atiende una oportunidad de interconexión entre fuentes públicas de Estados Unidos y datos de campo de Ecuador, puesto que ninguna de

ellas cuenta originalmente con una estructura común que permitiera su procesamiento conjunto para construir el Score de Oportunidad de Mercado y tomar decisiones de expansión de manera oportuna.

Data Warehouse Marketpulse Apheresis

El repositorio de la información se encuentra ubicado en PostgreSQL, alojado en un servidor en la nube. En este Data Warehouse se consolida toda la información proveniente de las fuentes públicas de Estados Unidos y de la Fundación Guillain Barré Ecuador. El proceso de información presenta la siguiente rutina:

- Las fuentes públicas (CDC, National Inpatient Sample, CMS Provider Data Catalog, US Census Bureau) actualizan su información con periodicidad propia -trimestral o anual según el ente emisor- y se descargan mediante un proceso automatizado de extracción.
- La Fundación Guillain Barré Ecuador envía mensualmente un archivo con los casos nuevos financiados, que se deposita en un repositorio transitorio.
- Un proceso ETL desarrollado en Python, ejecutado en batch programado, recoge los archivos disponibles y actualiza las tablas del Data Warehouse. En caso de que una fuente pública no logre actualizarse a tiempo, el proceso toma el último archivo disponible, evitando detener el cálculo del SOM.

Definición de estructura de archivos

El gobierno de la estructura de archivos se encuentra a cargo de la Gerencia Analítica de Mercado. La información es generada en archivos planos o mediante APIs públicas, y cada dato debe cumplir criterios definidos al momento de realizar la ingestión: tipo de dato, fuente principal, frecuencia de actualización y zona geográfica de referencia. Este proceso está sujeto a validación de calidad, para posteriormente programar la actualización de manera automática. La carga se realiza de forma incremental, garantizando que se cuente con un registro (LOG) de actualización de cada ciclo de carga.

Gobierno del dato

La calidad y actualización de la información dentro del Data Warehouse de MarketPulse Apheresis requiere responsables y políticas claramente definidos, que se detallan a continuación.

Tabla 10 Roles y políticas de gobierno del dato en MarketPulse Apheresis.

Ámbito	Responsable	Política
--------	-------------	----------

Calidad del dato	Gerencia Analítica de Mercado	Validación automática de formato, completitud y consistencia de cada archivo antes de su carga al Data Warehouse.
Actualización de fuentes	Gerencia Analítica de Mercado	Calendario de actualización por fuente (trimestral para CDC y CMS, anual para US Census, mensual para la Fundación Guillain Barré Ecuador).
Verificación de cobertura clínica	Gerencia Alianzas Clínicas y Fundaciones	Confirmación periódica del estado de cobertura por zona y actualización del indicador correspondiente en el Data Warehouse.
Acceso y confidencialidad	Gerencia Analítica de Mercado	Autorización de accesos a las tablas de PostgreSQL según la finalidad de la solicitud, y anonimización de los datos de pacientes desde el origen.

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Tipos de datos

Se relacionan los tipos de datos que se obtienen de las distintas fuentes, en función del modelo de priorización de mercado. La variable en común que permite relacionar los datos entre fuentes es la zona geográfica de análisis (cantón, condado, provincia o estado).

Tabla 11 Tipos de datos.

Tipo de Dato	Descripción	Segmento de Cliente al que Sirve
Epidemiológicos	Incidencia estimada de enfermedades tratables con aféresis por zona y periodo (CDC, National Inpatient Sample).	Clínicas/hospitales en zonas de alta demanda no cubierta
Capacidad Clínica Instalada	Clínicas y hospitales con capacidad de aféresis, camas y especialistas disponibles por zona (CMS Provider Data Catalog).	Clínicas/hospitales; Distribuidores y canales comerciales
Demográficos y Socioeconómicos	Densidad poblacional, cobertura de seguro médico e ingreso per cápita por zona (US Census Bureau).	Clínicas/hospitales en zonas de alta demanda no cubierta
Geográficos	Coordenadas y límites de cada zona, distancias a clínicas existentes y a distribuidores regionales.	Distribuidores y canales comerciales regionales
Casos Fundación Guillain Barré Ecuador	Información consolidada de pacientes financiados, costos y geografía, utilizada como caso de validación del modelo.	Fundación Guillain Barré Ecuador y fundaciones aliadas
Cobertura Clínica Verificada	Resultado de la Verificación de Cobertura Clínica por zona: estado de verificación y prioridad asignada.	Clínicas/hospitales; Distribuidores y canales comerciales
Score Histórico	Valores de DNC, SOM y TPM calculados en ciclos anteriores por zona, utilizados para medir evolución y validar el modelo.	Dirección Comercial y gerencias internas
Canales Comerciales	Distribuidores y clínicas contactadas por zona, con estado de gestión comercial.	Distribuidores y canales comerciales regionales

Fuente: Elaborado por los autores

Fecha: Julio, 2026

La estructura de la información permite la creación de datamarts específicos por bloque: Registros de Pacientes, Red de Clínicas, Geografía y Epidemiología, y Score Output, que en conjunto conforman las 8 tablas del Data Warehouse en PostgreSQL. Cada datamart habilita un espacio de análisis propio: por ejemplo, un datamart comercial que alimenta el dashboard de Power BI con el ranking y el mapa de zonas priorizadas, y un datamart analítico con las variables crudas utilizadas para calibrar el modelo.

La extracción y transformación de la información se realiza mediante scripts de Python, que constituyen la herramienta autorizada para la conexión con el Data Warehouse. Los accesos a las distintas tablas de PostgreSQL son autorizados por la Gerencia Analítica de Mercado, quien analiza cada solicitud y la finalidad con la que se requiere el acceso a dicha información, de forma equivalente a como otras organizaciones administran el acceso a su Data Warehouse mediante herramientas como SAS.

Conclusiones

El Data Warehouse es un sistema que reúne datos de una amplia gama de fuentes dentro de una organización y los utiliza para respaldar la toma de decisiones. En el caso de MarketPulse Apheresis, esto significa consolidar en PostgreSQL las fuentes epidemiológicas, demográficas y geográficas de Estados Unidos junto con los datos de campo de la Fundación Guillain Barré Ecuador, en un único repositorio que sirva de base para el proceso ETL y para el cálculo del Score de Oportunidad de Mercado. Antes de construir dicho proceso ETL, es necesario identificar con precisión las fuentes de datos, su origen, su recurrencia de carga y su destino dentro de los datamarts del Data Warehouse, tal como se ha desarrollado en el presente capítulo.

Capítulo 3 - Procesos ETL

Objetivo y alcance del trabajo

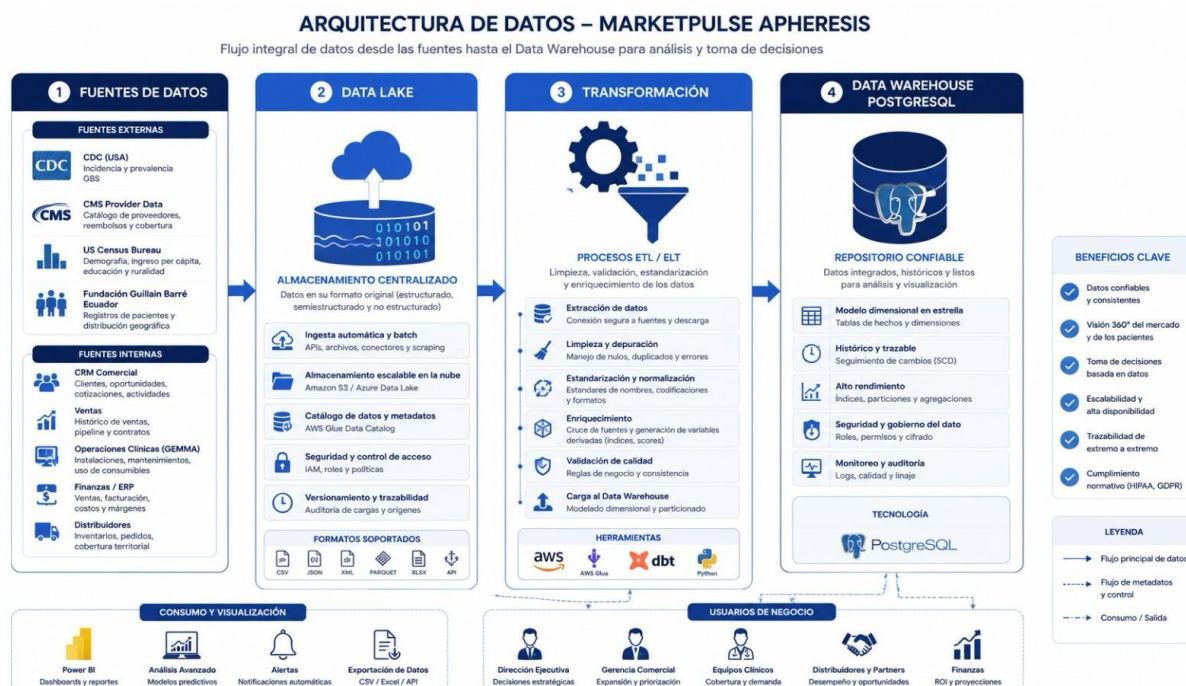
Los capítulos anteriores identificaron los segmentos de cliente, las fuentes de datos y el gobierno de la información que sustentan a MarketPulse Apheresis. Este capítulo se enfoca exclusivamente en la gestión técnica de esa información: cómo se extrae, se transforma, se carga y se pone a disposición de la compañía a través del proceso ETL (Extract, Transform, Load).

Cada decisión técnica descrita en este capítulo existe para responder una pregunta de negocio concreta - ¿puede Apheresis Technology calcular el Score de Oportunidad de Mercado de forma automática, confiable y oportuna, para decidir dónde expandirse con menor riesgo y mayor rentabilidad?

Procesos de carga de datos (ETL, data lake)

El proceso de carga de MarketPulse Apheresis sigue las tres fases clásicas de un ETL. En la extracción, scripts en Python consultan las APIs y archivos planos del CDC, el National Inpatient Sample, el CMS Provider Data Catalog, el US Census Bureau y el archivo mensual de la Fundación Guillain Barré Ecuador; los datos crudos se depositan primero en un data lake en la nube, preservando la información original antes de cualquier transformación. En la transformación, se normaliza la variable de zona geográfica (cantón, condado o estado) a una clave común, se eliminan duplicados de registros de pacientes, se verifican rangos válidos (incidencia no negativa, fechas coherentes) y se clasifica cada registro en uno de los cuatro bloques del Data Warehouse. En la carga, cada ciclo se ejecuta de forma incremental -no se recarga el histórico completo- conservando el LOG de actualización de cada corrida.

Ilustración 3 Diagrama de flujo del proceso ETL



Fuente: Elaborado por los autores

Fecha: Julio, 2026

Información y calidad de la información disponible

La calidad del dato se evalúa en cuatro dimensiones: completitud, consistencia, actualidad y trazabilidad, bajo la responsabilidad de la Gerencia Analítica de Mercado definida en el gobierno del dato (Capítulo 2). En la práctica, esto se traduce en reglas concretas de validación: la zona debe existir en un catálogo maestro de zonas geográficas, la incidencia no puede ser negativa ni nula, y la fecha de reporte debe estar dentro del periodo esperado de cada fuente. Al igual que en la mayoría de proyectos de Business Intelligence, se estima que entre el 30% y el 40% del esfuerzo inicial del proyecto se concentrará en estandarizar la variable de zona geográfica, que es la llave común entre todas las fuentes.

Reglas cuantificadas de calidad y vigencia del dato

Para que la validación de calidad no dependa únicamente de un criterio cualitativo, se definen umbrales cuantificados, cada uno con una acción definida cuando no se cumple: completitud de campos obligatorios $\geq 98\%$ (por debajo, el archivo se rechaza y se notifica a la Gerencia Analítica de Mercado); duplicados de registros = 0% (cualquier duplicado detectado por clave de paciente o de zona-periodo bloquea

la carga hasta su resolución); actualización dentro del SLA de cada fuente $\geq 95\%$ de los ciclos (el incumplimiento genera una alerta, sin bloquear la carga salvo que se supere el umbral de obsolescencia); correspondencia geográfica $\geq 99\%$ de los registros con una zona válida en el catálogo maestro (los registros sin correspondencia se aíslan en una tabla de excepciones); y trazabilidad = 100% de los registros con fuente, fecha de extracción y versión del proceso ETL identificables.

Sobre la vigencia del dato, el uso automático del último archivo disponible ante una fuente no actualizada se complementa con un indicador de frescura por fuente: cada registro del Data Warehouse almacena su fecha de vigencia de origen, y el dashboard de Power BI la muestra junto a cada cifra (por ejemplo, “Incidencia CDC, dato del trimestre X”). Cuando una fuente supera un umbral de antigüedad definido por su periodicidad esperada (más de dos trimestres sin actualización para CDC/NIS, o más de 14 meses para US Census Bureau), el SOM de las zonas afectadas se marca visualmente como “calculado con datos no vigentes”, y la Dirección Comercial recibe una advertencia antes de tomar una decisión de expansión basada en esa zona.

Ejemplo real de un ciclo de extracción, transformación y carga

Para demostrar que el proceso ETL no es solo una arquitectura propuesta, se traza a continuación un ciclo real de carga -desde el archivo de origen hasta el Data Warehouse- con cifras, tiempos, errores y responsables concretos. El mismo registro continúa en el Capítulo 4 (cálculo del SOM, ranking y decisión comercial), de modo que la cadena completa quede demostrada de extremo a extremo.

Etapas	Responsable	Fecha y hora	Detalle cuantificado
1. Fuente	Fundación Guillaín Barré Ecuador (proveedor de datos)	12/07/2026, 02:00 h	Archivo mensual fundacion_gbe_202607.csv, 1.240 registros de pacientes de la zona Cantón X, recibido por SFTP seguro según el protocolo del Capítulo 2
2. Extracción	Analista de Datos (Gerencia Analítica de Mercado)	12/07/2026, 02:00-02:04 h	Script extract_fundacion.py descarga el archivo y lo deposita sin modificar en la capa RAW del data lake; duración 4 minutos; 1.240 registros extraídos; 0 errores de conexión
3. Transformación	Proceso ETL en Python (supervisado por Gerencia Analítica de Mercado)	12/07/2026, 02:05-02:18 h	Normalización de zona geográfica; se descartan 6 registros duplicados (misma clave paciente-periodo) y 3 registros con incidencia negativa o fecha fuera de rango; 2 registros sin

			correspondencia en el catálogo maestro de zonas se aíslan en la tabla de excepciones. Resultado: 1.229 de 1.240 registros continúan el proceso (99,1%)
4. Control de calidad	Gerencia Analítica de Mercado	12/07/2026, 02:18 h	Se aplican los umbrales cuantificados de este capítulo sobre el lote: completitud de campos obligatorios 99,4% (meta $\geq 98\%$, cumple); duplicados 0% tras depuración (meta 0%, cumple); correspondencia geográfica 99,8% tras aislar 2 registros (meta $\geq 99\%$, cumple); trazabilidad 100% (cumple). Lote aprobado para carga
5. Carga	Proceso ETL (scheduler en la nube)	12/07/2026, 02:20 h	1.229 registros cargados de forma incremental en las tablas Registros de Pacientes y Geografía y Epidemiología de PostgreSQL; versión de proceso ETL v1.4; LOG de carga generado y disponible para auditoría

Del lote original de 1.240 registros, 11 (0,9%) no llegaron a la carga: 6 por duplicidad, 3 por rango inválido y 2 por falta de correspondencia geográfica, quedando disponibles en la tabla de excepciones para revisión manual de la Gerencia Analítica de Mercado sin bloquear el resto del lote. El Capítulo 4 retoma este mismo registro -ya cargado en el Data Warehouse- para el cálculo del DNC y el SOM, el ranking, la visualización y la decisión comercial.

Horarios de carga y accesibilidad

Cada fuente tiene una frecuencia y una ventana de carga propia, así como un grupo de roles autorizados a consultarla:

Tabla 12 Frecuencia de carga y accesibilidad

Fuente	Frecuencia	Ventana de carga	Accesibilidad
CDC / National Inpatient Sample	Trimestral	Primer día hábil del trimestre, batch nocturno	Gerencia Analítica de Mercado
CMS Provider Data Catalog	Trimestral	Misma ventana que CDC/NIS	Analítica de Mercado, Alianzas Clínicas
US Census Bureau	Anual	Primer día hábil de enero	Gerencia Analítica de Mercado

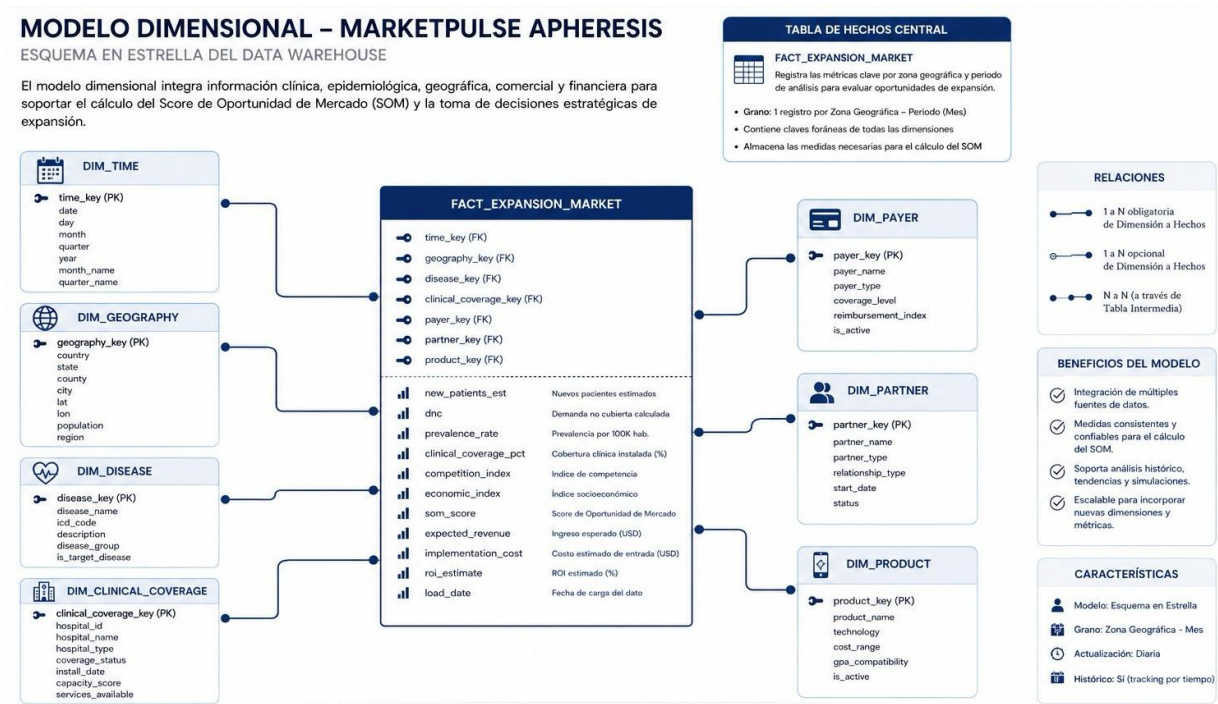
Fundación Guillain Barré Ecuador	Mensual	Primeros 5 días hábiles del mes, batch nocturno	Alianzas Clínicas y Fundaciones, Analítica de Mercado
Verificación de Cobertura Clínica	Continua (por ciclo)	Actualización bajo demanda	Canales Comerciales, Dirección Comercial
Score Output (SOM, DNC, TPM)	Mensual	Al cierre del ciclo de carga	Todas las gerencias (solo lectura, vía Power BI)

Fuente: Elaborado por los autores
Fecha: Julio, 2026

Variables disponibles y acceso al Data Warehouse

Este esquema de acceso se apoya en un modelo dimensional preliminar tipo estrella, en el que una tabla de hechos mensual por zona (DNC, SOM, TPM) se relaciona con dimensiones de Zona, Tiempo, Fuente y Clínica.

Ilustración 4 Modelo dimensional preliminar del Data Warehouse



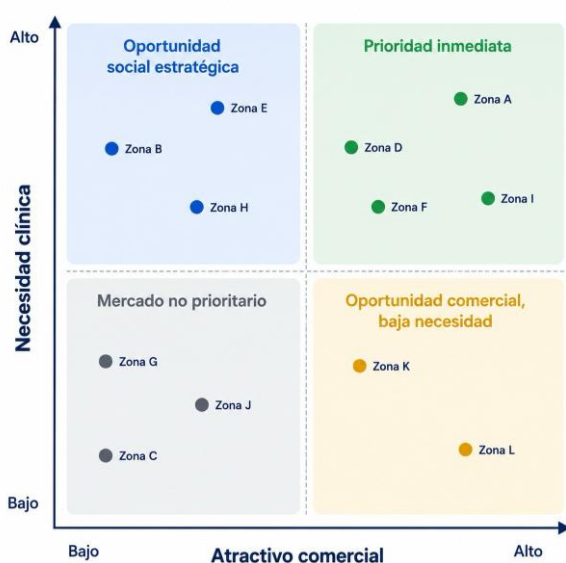
Fuente: Elaborado por los autores
Fecha: Julio, 2026

Matriz de priorización clínico-comercial (2x2)

Para evitar que el SOM combine sin distinción el impacto clínico y la rentabilidad comercial, se incorpora una matriz de priorización 2x2 complementaria al ranking del SOM. El eje horizontal representa el atractivo comercial de la zona (combinación

normalizada de capacidad de pago, tamaño de mercado potencial vía IPV y REE estimado), y el eje vertical representa la necesidad clínica (combinación normalizada del DNC y la severidad de la brecha de cobertura). La intersección clasifica cada zona en uno de cuatro cuadrantes: prioridad inmediata (alta necesidad clínica y alto atractivo comercial); oportunidad social estratégica (alta necesidad clínica, bajo atractivo comercial; candidata a alianzas con fundaciones antes que a expansión puramente comercial); oportunidad comercial con baja necesidad (bajo DNC pero alto atractivo comercial; revisar si ya existe cobertura suficiente antes de invertir); y mercado no prioritario (bajo en ambos ejes). Esta matriz se muestra como panel adicional del dashboard y se actualiza en cada ciclo junto con el SOM.

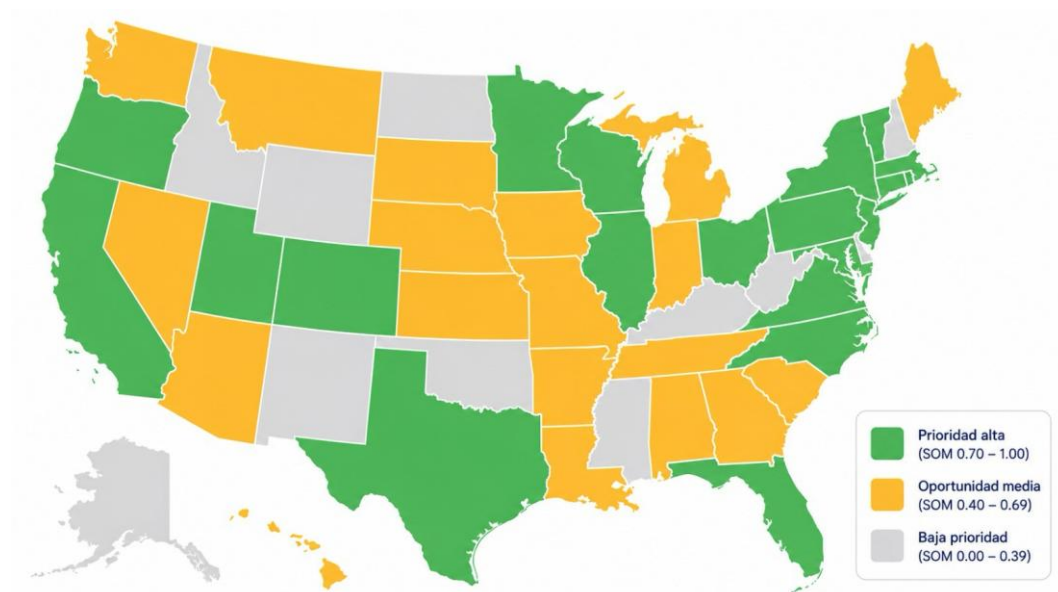
Ilustración 5 Matriz de priorización clínico-comercial (2x2)



Fuente: Elaborado por los autores

Fecha: Julio, 2026

Ilustración 6 Mapa de oportunidad por semáforo de SOM (Estados Unidos)



Fuente: Elaborado por los autores
Fecha: Julio, 2026

Conclusiones

El proceso ETL descrito en este capítulo es el mecanismo que convierte datos dispersos -epidemiológicos, geográficos, clínicos y comerciales- en un Data Warehouse único y confiable. Ese es, en el fondo, el propósito real de MarketPulse Apheresis: transformar información dispersa en conocimiento útil para decidir dónde debe expandirse Apheresis Technology con menor riesgo y mayor rentabilidad. Los capítulos siguientes profundizarán en el modelo de scoring y en el dashboard de Power BI que se apoyan sobre esta arquitectura.

Capítulo 4 - Visualización de datos para la toma de decisiones

Este capítulo continúa directamente el trabajo del Capítulo 3 (procesos ETL) y cierra el ciclo completo del proyecto: problema empresarial → fuentes → ETL → Data Warehouse → KPI → visualización → insight → decisión → resultado. Mientras el Capítulo 3 definió cómo se transforma el dato en el Data Warehouse, este capítulo define cómo ese dato llega a cada responsable de decisión de Apheresis Technology en el formato, la herramienta y el nivel de acceso adecuados, y cómo se comprueba que el sistema de Business Intelligence resuelve los problemas planteados desde la Semana 1.

Objetivo y alcance del trabajo

El objetivo de este capítulo es diseñar la capa de visualización de MarketPulse Apheresis: seleccionar la herramienta adecuada, definir los usuarios y sus necesidades de decisión, construir el catálogo de dashboards con sus KPI, establecer la matriz de acceso y las alertas, y verificar el grado de cumplimiento de los objetivos definidos en el Capítulo 1 (reducción del ciclo de priorización de 6 a 2 meses, de 32 a 10 horas de consolidación manual, y de USD 960 a USD 640 por verificación de cobertura clínica).

Necesidades de información por usuario y decisión

Siguiendo la secuencia Usuario → Decisión → Pregunta → KPI, se identifican las decisiones concretas que cada perfil debe poder tomar a partir del sistema de visualización:

Tabla 13 Necesidades de información por usuario y decisión

Usuario	Decisión que debe tomar	Pregunta clave	KPI asociado
Alta Dirección (Gerencia General)	Evaluar el avance global de la expansión	¿Estamos cumpliendo la meta de expansión y de retorno de la inversión?	SOM promedio, TVM, ROI, PRI
Dirección Comercial	Priorizar zonas y corregir desviaciones	¿Qué zonas debemos priorizar hoy y cuáles ya no lo justifican?	SOM, DNC, ranking de zonas
Gerencia Analítica de Mercado	Calibrar el modelo y vigilar la calidad del dato	¿El dato que alimenta el SOM es confiable y está vigente?	KPI técnicos ETL, TVM

Alianzas Clínicas y Fundaciones	Priorizar y descartar zonas en verificación	¿Qué zonas ya tienen cobertura clínica confirmada?	Estado de Verificación de Cobertura Clínica
Canales Comerciales	Dirigir a los distribuidores regionales	¿En qué zonas priorizadas debo concentrar la instalación de GEMMA?	TPM, avance de instalación por distribuidor

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Análisis y selección de la herramienta de visualización

La selección de la herramienta no se basa en popularidad, sino en su ajuste con la arquitectura ya definida en el Capítulo 3 (PostgreSQL como Data Warehouse, procesos ETL en Python y equipo con conocimiento previo en Microsoft 365). Se comparan tres alternativas mediante una matriz de criterios ponderados:

Tabla 14 Análisis y selección de la herramienta de visualización

Criterio	Peso	Power BI	Tableau	Qlik Sense
Integración con PostgreSQL y el ecosistema actual	25%	5	3	4
Coste de licenciamiento para un equipo de 5 personas	20%	5	3	3
Facilidad de uso para perfiles no técnicos (Comercial, Canales)	20%	4	5	3
Seguridad por perfil (RBAC / seguridad a nivel de fila)	15%	5	4	4
Visualización móvil para distribuidores regionales	10%	4	4	4
Conocimiento previo del equipo del proyecto	10%	4	3	2
Puntuación ponderada	100%	4,55	3,55	3,45

Fuente: Elaborado por los autores

Fecha: Julio, 2026

El resultado confirma la herramienta ya adoptada de forma preliminar en el Capítulo 3: Power BI. La justificación no es que sea universalmente superior, sino que encaja mejor con la arquitectura Fuentes → ETL en Python → PostgreSQL ya construida, evita duplicar plataformas de integración y permite que Dirección Comercial y Canales Comerciales -perfiles sin formación técnica- trabajen con un producto de autoservicio. Además, Power BI soporta seguridad por perfil, actualización periódica ligada al scheduler de carga (Capítulo 3), acceso móvil para los distribuidores regionales, y su coste relativo es compatible con el presupuesto de un proyecto en fase de validación con Ecuador antes de escalar a Estados Unidos.

Usuarios y perfiles del sistema BI

- **Alta Dirección (Gerencia General):** visión global del avance de la expansión; SOM promedio y por país; ROI y PRI del proyecto; alertas estratégicas. No necesita el detalle de cada zona ni los KPI técnicos del ETL.
- **Dirección Comercial:** ranking de zonas por SOM y DNC; mapa de oportunidad; estado de la Verificación de Cobertura Clínica; indicadores de éxito comercial (IPV, REE, TVM).
- **Gerencia Analítica de Mercado:** KPI técnicos del ETL (completitud, duplicados, SLA, correspondencia geográfica); tendencia de vigencia de cada fuente; herramientas de recalibración del SOM.
- **Alianzas Clínicas y Fundaciones:** estado de verificación por zona; alertas de zonas con cobertura clínica ya confirmada; datos de la Fundación Guillain Barré Ecuador.
- **Canales Comerciales:** mapa y ranking limitados a su región asignada; TPM y avance de instalación de GEMMA por distribuidor. No necesita ver zonas fuera de su región ni el detalle epidemiológico completo.

Matriz de acceso por perfil (seguridad)

Se aplica el principio de mínimo privilegio: cada perfil accede únicamente a la información necesaria para su decisión. La seguridad a nivel de fila en Power BI restringe, además, a Canales Comerciales a su región asignada, replicando a nivel de visualización el control de acceso por rol ya definido en el gobierno del dato del Capítulo 2.

Tabla 15 Análisis y selección de la herramienta de visualización

Dashboard	Alta Dirección	Dirección Comercial	Analítica de Mercado	Alianzas Clínicas	Canales Comerciales
Ejecutivo	Sí	Parcial	Parcial	No	No

Mapa de oportunidad	Sí	Sí	Sí	Parcial	Solo su región
Ranking de zonas priorizadas	Sí	Sí	Sí	Parcial	Solo su región
Análisis clínico / epidemiológico	Parcial	Sí	Sí	Sí	No
Seguimiento de Verificación de Cobertura	Parcial	Sí	Parcial	Sí	No
Éxito comercial (IPV, REE, TVM) y matriz 2x2	Sí	Sí	Sí	No	No
Canales Comerciales	Parcial	Sí	No	No	Solo su región
Calidad BI (KPI técnicos ETL)	No	No	Sí	No	No

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Catálogo de dashboards y KPI asociados

Los cinco dashboards definidos en el Capítulo 3 (Mapa de oportunidad, Ranking de zonas, Análisis clínico/epidemiológico, Seguimiento de Verificación de Cobertura, Indicadores de éxito comercial, y la matriz 2x2) se complementan con un Dashboard Ejecutivo, un Dashboard de Canales Comerciales y un Dashboard de Calidad BI, cerrando el catálogo completo:

Tabla 16 Análisis y selección de la herramienta de visualización

Dashboard	Usuario principal	Objetivo	Frecuencia
Ejecutivo	Alta Dirección	Controlar el avance global de la expansión (SOM, TVM, ROI, PRI)	Semanal
Mapa de oportunidad	Dirección Comercial, Canales Comerciales	Priorizar zonas por semáforo de SOM	Mensual (tras cada ciclo)
Ranking de zonas priorizadas	Dirección Comercial	Ordenar candidatas a Verificación de Cobertura Clínica	Mensual
Análisis clínico / epidemiológico	Analítica de Mercado	Vigilar la evolución del DNC por zona	Mensual

Seguimiento de Verificación de Cobertura	Alianzas Clínicas y Fundaciones	Detectar cuellos de botella en la verificación	Continuo
Éxito comercial y matriz 2x2	Dirección Comercial, Alta Dirección	Balancear atractivo comercial y necesidad clínica	Mensual
Canales Comerciales	Canales Comerciales	Dirigir la instalación de GEMMA por distribuidor y región	Semanal
Calidad BI	Gerencia Analítica de Mercado	Vigilar la salud técnica del proceso ETL	Diario

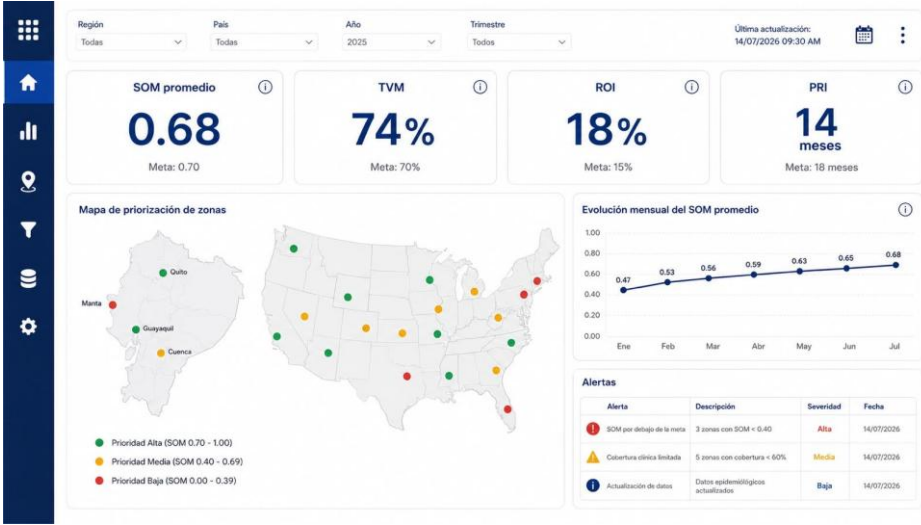
Fuente: Elaborado por los autores

Fecha: Julio, 2026

Diseño visual de los nuevos dashboards

El Mapa de oportunidad, el Ranking de zonas y la matriz 2x2 ya cuentan con mockup en el Capítulo 3. A continuación se diseñan los tres dashboards nuevos introducidos en este capítulo.

Ilustración 7 Mockup del Dashboard Ejecutivo

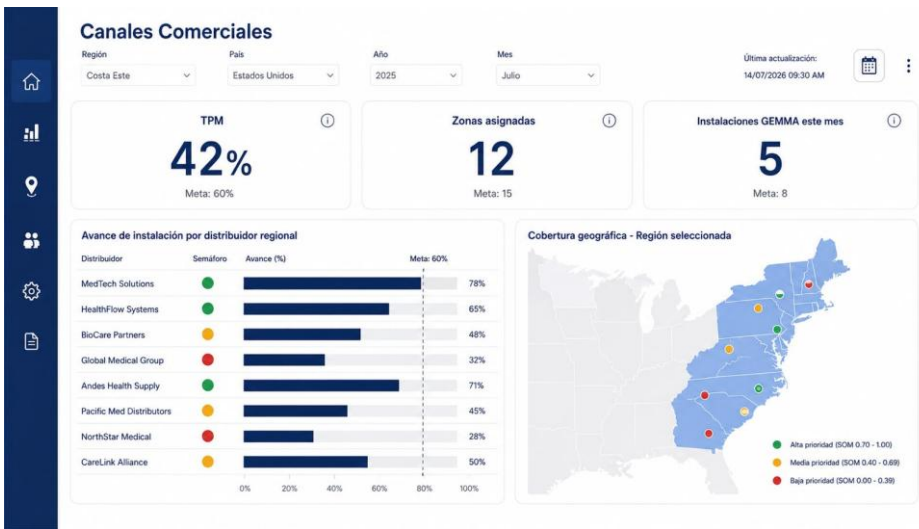


El Dashboard Ejecutivo muestra cuatro elementos: un indicador de SOM promedio del periodo con su variación respecto del ciclo anterior; la evolución mensual de la Tasa de Validación del Modelo (TVM), que alerta a la Alta Dirección cuando el modelo requiere recalibración; el ROI y el PRI acumulados del proyecto frente a la meta definida en el Capítulo 1; y un resumen del avance de instalaciones de GEMMA por país. Su lectura habitual toma menos de un minuto: si el semáforo de TVM está en verde y el ROI acumulado supera la meta, la Alta Dirección no requiere

intervenir; si TVM cae a ámbar o rojo, el panel dirige directamente al Dashboard de Calidad BI para diagnosticar la causa.

Fuente: Elaborado por los autores
Fecha: Julio, 2026

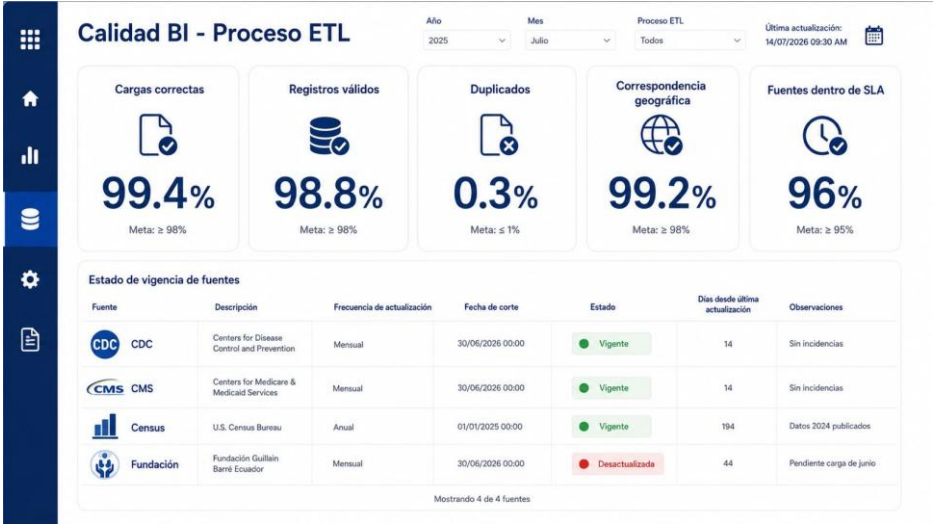
Ilustración 8 Mockup del Dashboard de Canales Comerciales



El Dashboard de Canales Comerciales aplica seguridad a nivel de fila: cada distribuidor visualiza únicamente el mapa, el ranking y el avance de instalación de su propia región. Los tres componentes principales son el mapa de oportunidad filtrado a la región asignada, la Tasa de Penetración de Mercado (TPM) por distribuidor, y un listado de zonas priorizadas pendientes de contacto comercial ordenado por SOM descendente. La interpretación operativa es directa: el distribuidor debe concentrar su gestión comercial en las zonas listadas de mayor a menor SOM, y el indicador de TPM le permite verificar si su avance de instalación está por encima o por debajo del promedio de su región.

Fuente: Elaborado por los autores
Fecha: Julio, 2026

Ilustración 9 Mockup del Dashboard de Calidad BI



El Dashboard de Calidad BI es de uso exclusivo de la Gerencia Analítica de Mercado y traduce visualmente los umbrales cuantificados de calidad del dato definidos en el Capítulo 3: completitud, duplicados, correspondencia geográfica y trazabilidad, cada uno con semáforo verde/ámbar/rojo según el umbral correspondiente. Incluye además el indicador de frescura por fuente (CDC, NIS, CMS, US Census Bureau y Fundación Guillain Barré Ecuador) y el historial de alertas de SOM con variación anormal (>20% mes a mes). Su interpretación es diagnóstica: cualquier semáforo en rojo bloquea la publicación del ranking del ciclo hasta su resolución manual, replicando en el dashboard la misma regla de gobierno del dato ya definida en el Capítulo 2.

Fuente: Elaborado por los autores
Fecha: Julio, 2026

Alertas y excepciones

El dashboard no solo debe mostrar información: debe señalar qué requiere atención inmediata, retomando las reglas de calidad y vigencia ya cuantificadas en el Capítulo 3.

Tabla 17 Alertas y excepciones

Alerta	Condición	Acción
--------	-----------	--------

Fuente no actualizada	Antigüedad supera el umbral definido por fuente (Capítulo 3)	Mantener el último dato válido, marcarlo como no vigente y notificar a Analítica de Mercado
Zona sin correspondencia geográfica	Registro no coincide con el catálogo maestro de zonas	Aislar en tabla de excepciones; no se incluye en el ranking hasta su corrección
SOM con variación anormal	Variación mes a mes superior a 20% en una misma zona	Bloquear la publicación del nuevo ranking hasta validación manual de Analítica de Mercado
Cuello de botella en verificación	Zona en estado "pendiente" por más de 30 días	Alertar a Alianzas Clínicas y Fundaciones para reasignar analista
Caída de la Tasa de Validación del Modelo (TVM)	TVM por debajo del umbral esperado durante dos ciclos consecutivos	Iniciar recalibración de los pesos del SOM (Capítulo 1, verificación de consistencia AHP)

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Seguridad y accesibilidad

La seguridad del sistema de visualización distingue acceso funcional (a los dashboards, sin exponer tablas ni el detalle del Data Warehouse) de acceso técnico (Gerencia Analítica de Mercado, con acceso al modelo semántico y a los datasets subyacentes). La seguridad a nivel de fila en Power BI implementa, para Canales Comerciales, la misma restricción por región que ya rige el acceso a datos crudos en el gobierno del dato (Capítulo 2): un distribuidor solo visualiza su región asignada y no puede consultar el resto del mapa de oportunidad.

Pruebas de usuario

Antes de publicar la versión final, se realiza una validación con un usuario piloto de cada perfil (Alta Dirección, Dirección Comercial, Gerencia Analítica de Mercado, Alianzas Clínicas y Fundaciones, Canales Comerciales), verificando cuatro aspectos: si encuentra la información que necesita para su decisión, si comprende los KPI sin explicación adicional, si puede localizar una zona con problema de cobertura en menos de 30 segundos, y si identifica información innecesaria o faltante en su dashboard. Los ajustes resultantes se incorporan antes de la puesta en producción sobre el piloto de Ecuador.

Revisión de los objetivos iniciales

Retomando los GAP definidos en el Capítulo 1 (Conocimiento y Organización), se contrasta el estado antes del proyecto con el resultado esperado tras la implementación del sistema de visualización:

Tabla 18 Revisión de los objetivos iniciales

Objetivo inicial	Antes	Meta / después	Estado
Tiempo de consolidación manual por ciclo	32 horas	10 horas	Parcialmente cumplido (piloto Ecuador)
Ciclo completo de priorización de zona	6 meses	2 meses	Parcialmente cumplido (piloto Ecuador)
Gasto por verificación de cobertura clínica	USD 960 por zona	USD 640 por zona	Parcialmente cumplido (piloto Ecuador)
Criterio de priorización de mercados	Informal, sin ranking objetivo	Ranking objetivo basado en el SOM	Cumplido
Visibilidad de mercados priorizados para Canales Comerciales	Ninguna	Mapa y ranking en Power BI	Cumplido
Trazabilidad del dato que sustenta cada decisión	Ninguna	100% de registros trazables (fuente, fecha, versión ETL)	Cumplido
Escalamiento del modelo a Estados Unidos	No iniciado	Modelo validado en Ecuador antes de escalar	No medido todavía

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Ninguno de los tres primeros objetivos se declara “cumplido” de forma absoluta: al tratarse de metas calibradas con el piloto de la Fundación Guillain Barré Ecuador, se clasifican como parcialmente cumplidas hasta contar con al menos dos ciclos completos de operación real que permitan confirmar los tiempos y costos definitivos.

Valor económico generado

Tomando como referencia el GAP de Organización del Capítulo 1, la reducción de 32 a 10 horas de consolidación manual por ciclo, aplicada a un analista con costo cargado estimado de USD 20/hora, representa un ahorro aproximado de USD 440 por ciclo de priorización. Sobre una base de doce ciclos mensuales al año, el ahorro estimado en horas de consolidación asciende a USD 5.280 anuales, sin contar el ahorro adicional por la reducción de USD 960 a USD 640 en cada verificación de cobertura clínica descartada automáticamente. Estas cifras son estimaciones

preliminares del piloto de Ecuador y deberán contrastarse con el Retorno de la Inversión (ROI) y el Periodo de Recuperación (PRI) definidos en el Capítulo 1 una vez el sistema opere en producción.

Más allá del ahorro operativo, un análisis económico completo requiere estimar la inversión, los costos recurrentes, los beneficios, el flujo de caja y la rentabilidad del proyecto en un horizonte de varios años, con los mismos parámetros de costo y ahorro ya presentados en este capítulo y en el Capítulo 1.

Inversión, costos y beneficios estimados

La inversión inicial de USD 45.000 cubre el desarrollo del proceso ETL, la puesta en marcha del Data Warehouse en PostgreSQL, la configuración de los ocho dashboards de Power BI con seguridad por perfil, y las licencias del primer año para el equipo de cinco usuarios. Los costos operativos anuales de USD 6.000 cubren el hosting cloud del data lake y el Data Warehouse, las licencias Power BI Pro y el mantenimiento del proceso ETL. Los beneficios anuales combinan dos componentes: el ahorro operativo de USD 10.080 (USD 5.280 por reducción de horas de consolidación manual, más USD 4.800 por verificaciones de cobertura clínica descartadas automáticamente, ambos ya calculados en el Capítulo 1), y un margen incremental de USD 44.000 atribuible a dos zonas adicionales de expansión por año que el ciclo de priorización más rápido permite atender (escenario base del Capítulo 1: margen de USD 22.000 por zona). El beneficio anual total estimado es de USD 54.080.

Año				
Año 0				
Año 1				
Año 2				
Año 3				

Fuente: *Elaborado por los autores.*

Fecha: *Julio, 2026*

VAN, TIR y sensibilidad económica

A diferencia del PRI simplificado del Capítulo 1 -que solo considera los ingresos incrementales-, el Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR) de este apartado incorporan los costos operativos recurrentes sobre el flujo de caja neto completo, siguiendo la práctica estándar de evaluación financiera de proyectos. Con una tasa de descuento del 12% anual -consistente con el costo de capital habitual para proyectos de tecnología en el sector de dispositivos médicos-, el VAN

a tres años es de aproximadamente USD 70.480, y la TIR resulta aproximadamente 91%, muy por encima de la tasa de descuento, lo que indica que el proyecto genera valor incluso bajo supuestos conservadores de adopción.

El análisis de sensibilidad económica evalúa qué ocurre si los beneficios estimados no se materializan en su totalidad. Si los beneficios anuales cayeran un 30% (a USD 37.856), el VAN a tres años seguiría siendo positivo, de aproximadamente USD 31.530. El punto de equilibrio (VAN = 0) se alcanza únicamente si los beneficios caen más de un 50% respecto de lo estimado, lo que sugiere un margen de seguridad amplio frente a errores de estimación. Esta conclusión es preliminar y deberá confirmarse con al menos dos ciclos reales de operación del piloto en Ecuador antes de aplicarse como criterio de decisión para la expansión a Estados Unidos.

Ejemplo completo: del dato a la decisión

El Capítulo 3 (“Ejemplo real de un ciclo de extracción, transformación y carga”) trazó, con cifras, tiempos, errores y responsables reales, las etapas 1 a 5 de este mismo ciclo: la Fundación Guillain Barré Ecuador entrega el archivo fuente, un Analista de Datos lo extrae, el proceso ETL lo transforma y aplica el control de calidad, y finalmente 1.229 de los 1.240 registros originales quedan cargados en el Data Warehouse a las 02:20 h del 12/07/2026. Esta sección continúa exactamente ese registro -zona Cantón X- para demostrar, con la misma exigencia de cifras y responsables, cómo llega hasta la decisión comercial.

Tabla 19 Ejemplo completo: del dato a la decisión

Etapas	Responsables	Fecha y hora	Detalle cuantificado
6. Cálculo de KPI (DNC y SOM)	Gerencia Analítica de Mercado	12/07/2026, 03:00 h (tras consolidar con CDC, CMS y Census del mismo ciclo)	Cantón X: incidencia estimada 210 pacientes; capacidad clínica instalada equivalente a 45 pacientes tratables/año; DNC = (210-45)/210 = 0,79; SOM ponderado (pesos AHP del Capítulo 1) = 0,62
7. Ranking y visualización	Power BI (actualización automática)	12/07/2026, 03:15 h	Cantón X ocupa el puesto 3 de 38 zonas evaluadas en el ciclo; el Mapa de oportunidad lo marca en semáforo ámbar (SOM entre 0,40 y 0,69)
8. Insight (matriz 2x2)	Dirección Comercial	13/07/2026	DNC alto (0,79) combinado con atractivo comercial moderado ubica

			al Cantón X en el cuadrante de oportunidad social estratégica
9. Verificación de cobertura	Alianzas Clínicas y Fundaciones	13/07/2026-18/07/2026	Verificación de cobertura clínica del cantón; coste USD 640 (meta del Capítulo 1); se confirma que no existe otra clínica con aféresis instalada
10. Decisión comercial	Dirección Comercial	19/07/2026	Se autoriza el contacto comercial con Canales Comerciales para iniciar la instalación del dispositivo GEMMA en el Cantón X

Fuente: Elaborado por los autores

Fecha: Julio, 2026

Cierre del ciclo: desde la extracción del archivo en el Capítulo 3 (12/07/2026, 02:00 h) hasta la decisión comercial (19/07/2026) transcurren 7 días, frente a los 4-6 meses del proceso AS-IS descrito en el Capítulo 1. El coste de este ciclo puntual es de USD 640 de verificación más un coste interno estimado de 2 horas de analista (USD 20/hora $\approx$ USD 40), es decir, USD 680 en total, frente a los USD 960 de verificación del proceso manual anterior -sin contar el tiempo indeterminado que hoy consume la búsqueda manual de clínicas competidoras. Al tratarse de un piloto aún no operativo en producción, estas cifras son ilustrativas y consistentes con los parámetros ya definidos en los Capítulos 1 y 3, y deberán sustituirse por las cifras reales de cada ciclo una vez el sistema esté en funcionamiento.

Ilustración 10 Trazabilidad completa de un ciclo: dato real → fuente → extracción → transformación → control de calidad → carga (Capítulo 3) → SOM → ranking → decisión comercial (Capítulo 4)



Fuente: Elaborado por los autores

Fecha: Julio, 2026

Ciclo de mejora continua

El sistema de visualización de MarketPulse Apheresis no se considera un producto terminado. Cada ciclo mensual sigue la secuencia medir → detectar → analizar → decidir → actuar → volver a medir: el Dashboard de Calidad BI mide la salud del ETL, el Mapa de oportunidad y el Ranking detectan zonas candidatas, la matriz 2x2 y el Análisis clínico/epidemiológico permiten analizarlas, Dirección Comercial decide la priorización, Canales Comerciales actúa sobre la instalación, y el ciclo siguiente vuelve a medir el resultado a través de la Tasa de Penetración de Mercado (TPM) y la Tasa de Validación del Modelo (TVM).

Conclusiones

El sistema de visualización descrito en este capítulo cierra el ciclo Problema empresarial → Fuentes → ETL → Data Warehouse → KPI → Visualización → Insight → Decisión → Resultado que sustenta a MarketPulse Apheresis. Power BI se confirma como la herramienta que mejor encaja con la arquitectura construida en el Capítulo 3, y el catálogo de ocho dashboards, la matriz de acceso y las alertas definidas garantizan que cada gerencia de Apheresis Technology reciba únicamente la información necesaria para su decisión. La revisión de objetivos muestra avances reales pero todavía parciales, lo que confirma la necesidad de operar al menos dos

ciclos completos del piloto en Ecuador antes de escalar el modelo hacia Estados Unidos, tal como se planteó desde el Capítulo 1.

Capítulo 5 - Organización del área de Business Intelligence

Objetivo y alcance del trabajo

Los capítulos anteriores construyeron el sistema de MarketPulse Apheresis: las fuentes de datos, el proceso ETL, el Data Warehouse y los ocho dashboards de Power BI. Este capítulo responde a una pregunta distinta y previa a cualquier despliegue a escala: ¿quién va a gestionar el sistema, con qué estructura, qué personas, qué competencias, qué recursos y qué presupuesto, y por qué económicamente compensa crear esa estructura? Un sistema de Business Intelligence sin gobierno, sin propietario del dato y sin un equipo dedicado tiende a degradarse -paneles que dejan de actualizarse, definiciones de KPI que divergen entre gerencias, dependencia de una sola persona-. Este capítulo diseña la organización, los procesos y el Business Case que evitan ese resultado.

Las cifras de personal, formación e inversión de este capítulo se apoyan en dos fuentes: los datos preliminares ya calculados en los Capítulos 1 y 4 (costo cargado de analista, inversión tecnológica de USD 45.000, costo operativo de USD 6.000 anuales, y el escenario base de expansión comercial) y bandas salariales promedio del sector de datos y analítica en Ecuador, donde opera hoy la Gerencia Analítica de Mercado (Data Analyst/BI Analyst entre USD 1.200 y USD 2.200 mensuales, Data Engineer hasta USD 2.500 o más, según benchmarks de mercado citados en la Bibliografía). Esta es la única aclaración de este tipo en el capítulo; los montos que siguen no repiten la advertencia en cada tabla.

Situación actual (AS-IS)

MarketPulse Apheresis cuenta ya, tras los Capítulos 1 a 4, con tecnología funcional: un proceso ETL en Python, un Data Warehouse en PostgreSQL y ocho dashboards de Power BI con seguridad por perfil. Sin embargo, la tecnología por sí sola no gobierna ni sostiene su propio uso. Actualmente:

Tabla 20 Diagnóstico AS-IS por área

Área		
Dirección Comercial		
Gerencia Analítica de Mercado		

Alianzas Clínicas y Fundaciones	Gestiona la relación con la Fundación	Información de campo sin trazabilidad sistemática
Canales Comerciales	Consulta el mapa de oportunidad	Sin canal para solicitar ajustes o reportar incidencias
IT / Infraestructura cloud	Mantiene PostgreSQL y Power BI	No interpreta necesidades de negocio ni prioriza el roadmap analítico

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

El diagnóstico es consistente con la conclusión del Capítulo 1: no existe una unidad de Business Intelligence formal. Las funciones de dato están dispersas entre cuatro gerencias y un proveedor de infraestructura, sin gobierno ni propiedad única.

Problemas organizativos

La tecnología puede existir sin que el negocio la aproveche. Los problemas organizativos identificados no son de herramienta, sino de estructura: falta de gobierno del dato, de responsabilidades explícitas sobre cada componente del sistema, de coordinación entre gerencias, de especialización de roles, y de propiedad clara del dato (quién decide qué zona geográfica es válida, quién resuelve una discrepancia de SOM, quién autoriza un nuevo dashboard). MarketPulse Apheresis no necesita más tecnología: necesita una estructura organizativa capaz de gobernar y explotar la que ya construyó.

Carencias en recursos y formación

Las carencias detectadas se agrupan en cuatro categorías:

- **Personas:** falta un Data Engineer dedicado (hoy la función la absorbe un analista genérico); falta un BI Analyst dedicado al mantenimiento de los ocho dashboards; falta un Data Analyst dedicado a la analítica avanzada del SOM; falta un Data Steward.
- **Conocimiento:** nivel de Python/SQL suficiente para mantener el ETL, pero no para escalarlo a Estados Unidos; uso de Power BI limitado a visualización básica, sin modelado semántico avanzado; desconocimiento formal de gobierno del dato; capacidad estadística limitada para validación, correlación y escenarios.
- **Procesos:** las reglas de calidad del Capítulo 3 no tienen un responsable operativo dedicado; no existe un catálogo de datos formal más allá del catálogo maestro de zonas; no existen acuerdos de nivel de servicio (SLA) para atender incidencias de los dashboards.

- **Tecnología:** no existen ambientes de desarrollo/pruebas separados de producción para el ETL; las licencias de Power BI Pro actuales no cubren un equipo ampliado; no existe herramienta de catálogo o linaje de datos.

Modelo organizativo TO-BE

Se propone formalizar la actual Gerencia Analítica de Mercado como la Unidad de Business Intelligence y Datos de MarketPulse Apherisis, que actúa como puente entre negocio, datos y tecnología, sin desplazar a la Dirección Comercial de la decisión final de expansión. Sus funciones son:

Organigrama

El organigrama propuesto formaliza la Gerencia Analítica de Mercado como Unidad de Business Intelligence y Datos, sin crear una gerencia nueva dentro de la Dirección Comercial de Expansión de Mercados:

Tabla 21 Organigrama propuesto de la Dirección Comercial de Expansión de Mercados

Dirección Comercial de Expansión de Mercados	
Gerencia Analítica de Mercado (Unidad de Business Intelligence y Datos)	
Responsable BI & Data	
Data Engineer	
BI Analyst	
Data Analyst	
Data Steward (0,5 FTE)	

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

El Responsable BI & Data mantiene, además, relación funcional directa con Dirección Comercial, Gerencia Alianzas Clínicas y Fundaciones, Gerencia Canales Comerciales y Alta Dirección (Gerencia General), sin depender jerárquicamente de ninguna de ellas.

Matriz RACI

R = Responsable, A = Aprueba, C = Consultado, I = Informado.

Actividad					

Definir y calibrar SOM/pesos AHP	R	A	C	I	I
ETL y Data Warehouse	R	I	I	I	A
Calidad y catálogo del dato	R	A	C	I	I
Desarrollo de dashboards	R	A	C	C	I
Verificación de Cobertura Clínica	C	A	R	I	I
Seguridad y accesos (RBAC)	R	I	I	I	A
Priorización de zonas (decisión comercial)	R	A	C	I	I

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Gestión de solicitudes e incidencias

Proceso formal de solicitud

Toda solicitud a la Unidad de BI y Datos debe incluir objetivo, usuario, decisión que se apoyará, KPI asociado, prioridad y frecuencia -no basta con pedir "un dashboard nuevo". Por ejemplo:

- **Solicitud deficiente:** "Necesitamos un dashboard nuevo para Cañar" (sin objetivo, sin KPI, sin decisión asociada).
- **Solicitud bien formulada:** objetivo: determinar si Cañar debe pasar a Verificación de Cobertura Clínica este ciclo; usuario: Dirección Comercial; decisión: iniciar o no la verificación; KPI: SOM y DNC; prioridad: alta; frecuencia: mensual.

Gestión de incidencias y SLA

Incidencia crítica de ejemplo: el Mapa de oportunidad muestra una variación de SOM superior al 20% en una misma zona en un mismo ciclo (alerta ya definida en el Capítulo 4). Procedimiento: Dirección Comercial abre un ticket → el BI Analyst valida la alerta → el Data Steward identifica el origen (fuente desactualizada, error de carga o cambio real del dato) → el Data Engineer revisa el ETL → se corrige y reprocesa → el Data Analyst valida el nuevo SOM → Dirección Comercial confirma → se cierra la incidencia.

Tabla 22 Acuerdos de nivel de servicio (SLA) por severidad

Severidad		
Crítica		
Alta		

Media	1 día	< 3 días
Baja	2 días	Planificada

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Formación

No toda carencia se resuelve contratando. Las carencias detectadas se traducen en el siguiente plan de formación:

Tabla 23 Necesidades de formación por perfil

Perfil	
Dirección Comercial / Alta Dirección	
Alianzas Clínicas y Canales Comerciales	
Data Analyst / BI Analyst	
Data Engineer	
Data Steward	

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Plan de formación

- **Fase 1.** Usuarios (Dirección Comercial, Alianzas Clínicas, Canales Comerciales, Alta Dirección), 4 horas. Objetivo: interpretar los KPI del sistema (SOM, DNC, TVM, ROI, PRI) sin apoyo técnico.
- **Fase 2.** Data Analyst y BI Analyst, 24 horas. Power BI avanzado (modelo semántico, DAX) y Python/SQL aplicado al ETL.
- **Fase 3.** Data Engineer y Data Steward, 40 horas. Arquitectura cloud, modelado dimensional, gobierno y catálogo de datos.

Outsourcing

Pueden externalizarse, total o parcialmente, las actividades de mayor especialización técnica o menor recurrencia:

Tabla 24 Matriz Make or Buy

Actividad			
Definir y calibrar SOM / KPI			

Gobierno del dato	Sí	Apoyo puntual	Estratégico, con auditoría externa periódica
Implantación inicial del Data Warehouse / ETL	Parcial	Sí	Especialización técnica (ya ejecutada, Capítulo 4)
Desarrollo de dashboards Power BI	Sí	Formación externa	Uso continuo, requiere agilidad
Ciberseguridad y anonimización de datos de pacientes	Parcial	Sí	Expertise regulatorio en datos sanitarios
Soporte de infraestructura cloud	No	Sí	Escalabilidad y disponibilidad continua

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Costes

Plantilla del equipo

El costo cargado anual (salario más beneficios de ley en Ecuador) de cada posición se estima en:

Tabla 25 Plantilla y costo cargado anual del equipo de BI y Datos

Perfil		
Responsable BI & Data		
Data Engineer		
BI Analyst		
Data Analyst		
Data Steward		
Total		

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Recursos tecnológicos e inversión inicial del departamento

Los recursos tecnológicos recurrentes -hosting cloud, licencias Power BI Pro, base de datos y seguridad, USD 6.000 anuales- ya se detallaron en el análisis económico del Capítulo 4 y no se duplican aquí. La inversión inicial específica de este capítulo cubre la puesta en marcha de la estructura organizativa, adicional a los USD 45.000 de implantación tecnológica ya invertidos según el Capítulo 4:

Tabla 26 Inversión inicial específica del departamento (USD)

Concepto	
Reclutamiento y selección (4,5 posiciones)	

Plan de formación (Fases 1 a 3)	9,000
Documentación y transferencia de conocimiento	3,000
Total inversión inicial del departamento	20,000

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Beneficios

El Capítulo 4 estimó el beneficio del sistema durante el piloto de Ecuador (ahorro operativo más margen de dos zonas adicionales por año). Este capítulo amplía esa estimación al régimen de operación esperado una vez el departamento esté formalizado y opere de forma sostenida -incluyendo el inicio de la expansión hacia Estados Unidos-, con un ritmo de 10 zonas atendidas por año en lugar de las 5 del escenario piloto:

Tabla 27 Beneficios anuales estimados en régimen de operación (USD)

Componente	
Ahorro por consolidación de datos en las cuatro gerencias	
Ahorro por verificaciones de cobertura clínica evitadas	
Margen comercial neto (10 zonas/año, escenario base ampliado)	
Beneficio anual total en régimen	

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Business Case

El primer año del departamento no alcanza el beneficio de régimen: la contratación y la formación toman tiempo, por lo que se estima que el beneficio se alcanza solo al 70% durante el Año 1 (USD 175.000). A partir del Año 2, el equipo opera a régimen completo (USD 250.000 anuales).

Tabla 28 Business Case: costes y beneficios anuales (USD)

Concepto				
Inversión inicial del departamento				
Personal (plantilla completa)				
Tecnología (Capítulo 4)				
Beneficio operativo y comercial				

Flujo neto	-20,000	6,000	81,000	81,000
------------	---------	-------	--------	--------

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

ROI, Payback y escenarios

El flujo de caja acumulado pasa de -20.000 (Año 0) a -14.000 (Año 1) y se vuelve positivo dentro del Año 2, cuando el flujo mensual de 6.750 cubre el saldo pendiente: el Payback del departamento se alcanza aproximadamente a los 14 meses de operación.

El ROI a 3 años, sobre el total de costes acumulados (USD 527.000: inversión inicial más personal y tecnología de tres años) frente al beneficio acumulado (USD 675.000), resulta $(675.000 - 527.000) / 527.000 = 28\%$. Con una tasa de descuento del 12% -la misma usada en el Capítulo 4-, el VAN a 3 años es de aproximadamente USD 107.600 y la TIR aproximadamente 152%. La TIR es alta porque la inversión inicial específica del departamento (USD 20.000) es pequeña frente al flujo neto recurrente que genera la estructura; el indicador más informativo para esta decisión es, en realidad, la comparación directa entre el beneficio anual en régimen (USD 250.000) y el costo anual recurrente del equipo y la tecnología (USD 169.000): un margen estructural positivo de USD 81.000 al año.

Tabla 29 Escenarios de beneficio acumulado a 3 años (USD)

Escenario		
Optimista		
Base		
Pesimista		

Fuente: Elaborado por los autores

Fecha: Agosto, 2026

Incluso en el escenario pesimista, el beneficio acumulado a 3 años (USD 435.000) supera holgadamente el costo acumulado del departamento en el mismo periodo, lo que sostiene la decisión de formalizar la estructura frente a mantener el esquema informal actual.

Conclusiones

Este capítulo responde a la pregunta que debe poder defender el equipo: ¿quién va a gestionar el sistema de Business Intelligence, con qué recursos, cuánto costará y por qué económicamente merece la pena? La respuesta es formalizar la actual Gerencia Analítica de Mercado como Unidad de Business Intelligence y Datos, con

cuatro posiciones y media (Responsable BI & Data, Data Engineer, BI Analyst, Data Analyst y Data Steward a medio tiempo), un costo recurrente de USD 169.000 anuales (personal y tecnología) y una inversión inicial específica de USD 20.000, frente a un beneficio anual en régimen de USD 250.000 y un Payback de aproximadamente 14 meses. La estructura propuesta no añade tecnología nueva: gobierna, sostiene y escala la que ya se construyó en los Capítulos 2 a 4, y queda sujeta a los mismos criterios de validación con datos reales del piloto de Ecuador planteados desde el Capítulo 1.

PARTE II: BIG DATA

La segunda parte amplía el sistema con capacidades de Big Data y analítica avanzada. El Capítulo 1 incorpora datos no estructurados y nuevas fuentes; el Capítulo 2 define la arquitectura cloud y el proveedor; el Capítulo 3 selecciona el stack open source de orquestación y transformación; el Capítulo 4 aplica inteligencia artificial para anticipar la demanda no cubierta; y el Capítulo 5 diseña un caso de uso en tiempo real para la demanda aguda.

Capítulo 1 - Datos No Estructurados, Big Data y Nuevas Fuentes

Los cinco capítulos de la Parte I construyeron MarketPulse Apheresis a partir de fuentes de datos estructuradas: los organismos públicos de Estados Unidos (CDC, National Inpatient Sample, CMS Provider Data Catalog, US Census Bureau) y el archivo mensual de casos financiados de la Fundación Guillain Barré Ecuador. Esta información permite calcular el Score de Oportunidad de Mercado (SOM) y priorizar zonas, pero se actualiza con periodicidad trimestral o anual y no captura lo que ocurre entre esas actualizaciones. Este capítulo, correspondiente a la Semana 6, incorpora datos no estructurados y nuevas fuentes de Big Data para cerrar esa brecha, siguiendo las cuatro fases del protocolo de la Clase 6: diagnóstico de la necesidad, identificación y clasificación de fuentes, análisis de su relación con el modelo de Business Intelligence, y desarrollo de casos de uso.

Objetivo y alcance del trabajo

El objetivo de este capítulo es responder cuatro preguntas: qué decisiones de Apheresis Technology se toman hoy con información incompleta o desactualizada; qué fuentes de datos no estructurados y de Big Data podrían complementar el Data Warehouse ya construido en el Capítulo 2 de la Parte I; cómo se relacionan esas fuentes nuevas con las variables que ya alimentan el DNC y el SOM; y qué casos de uso concretos justifican incorporarlas. El alcance se limita a diagnóstico, catálogo, relación y diseño de casos de uso; la implementación técnica del pipeline de procesamiento de lenguaje natural (NLP) descrito en este capítulo se integrará formalmente al proceso ETL del Capítulo 3 de la Parte I una vez validada con datos reales adicionales del piloto de Ecuador.

Fase 1: Diagnóstico de las necesidades de información y justificación del uso de datos no estructurados

Antes de incorporar nuevas fuentes, se analiza qué decisiones de la Dirección Comercial de Expansión de Mercados (Capítulo 1 de la Parte I) se apoyan hoy en información insuficiente, y por qué el sistema de Business Intelligence construido en los Capítulos 2 a 4 de la Parte I -aun siendo funcional- no basta por sí solo.

Decisiones que se toman con información incompleta

La siguiente tabla recupera decisiones ya definidas en capítulos anteriores y señala la limitación concreta que deja abierta cada una, evitando el error de afirmar genéricamente que "faltan datos".

Tabla 30 Decisiones con información insuficiente y limitación identificada

Decisión afectada	Información utilizada hoy	Limitación identificada
Priorización de zona candidata (DNC / SOM)	Incidencia trimestral (CDC/NIS), capacidad clínica trimestral (CMS), demografía anual (US Census) y casos mensuales de la Fundación GBS Ecuador	No captura señales de demanda entre ciclos de actualización: interés de búsqueda, comentarios de pacientes o cobertura de prensa sobre una zona
Verificación de Cobertura Clínica	Visitas y consultas manuales de la Gerencia Alianzas Clínicas (Capítulo 1 de la Parte I: 32 h/ciclo; meta Capítulo 4 de la Parte I: 10 h/ciclo)	No aprovecha correos, WhatsApp Comercial ni reseñas públicas que ya contienen buena parte de esa información en texto libre
Selección de clínica aliada dentro de una zona priorizada	Capacidad clínica instalada (CMS) y estado de cobertura verificada	No incorpora la percepción real de pacientes ni la reputación pública de la clínica candidata
Anticipación de un aumento de incidencia (evento gatillo)	Incidencia estimada trimestral (CDC / National Inpatient Sample)	No existe mecanismo de alerta temprana ante brotes infecciosos asociados a un aumento de casos de Síndrome de Guillain-Barré

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Información que falta y justificación del Big Data

Con esta información, la Dirección Comercial puede responder "dónde hay demanda estructural no cubierta", pero no puede responder con la misma solidez tres preguntas adicionales: si esa demanda ya se está manifestando de forma visible antes del próximo corte trimestral, cómo perciben los pacientes y las clínicas a Apheresis Technology y a la Fundación Guillain Barré Ecuador, y si existe algún evento reciente -un brote infeccioso, una noticia de salud pública- que anticipe un cambio en la incidencia. Estas tres preguntas comparten una característica: la respuesta existe, pero en formato de texto libre, comentarios, reseñas y boletines, no en tablas estructuradas con periodicidad fija.

Información que sería conveniente incorporar: comentarios y publicaciones de pacientes y fundaciones en redes sociales; volumen de búsquedas relacionadas con aféresis y Síndrome de Guillain-Barré en Google Trends; boletines epidemiológicos del Ministerio de Salud Pública del Ecuador (MSP) y datos oficiales del Instituto Nacional de Estadística y Censos (INEC); reseñas públicas de las clínicas candidatas y aliadas; y el contenido de los correos y mensajes de WhatsApp Comercial que hoy se leen manualmente en la Verificación de Cobertura Clínica.

Por tanto, un sistema de Business Intelligence apoyado únicamente en fuentes públicas trimestrales o anuales y en un archivo mensual permite conocer qué ha ocurrido, pero no explicar por completo por qué ocurre ni anticipar lo que va a ocurrir en el intervalo entre actualizaciones. La incorporación de datos no estructurados y de nuevas fuentes de Big Data permitirá a MarketPulse Apheresis adelantar señales de demanda, medir la reputación real de las clínicas aliadas y detectar eventos

epidemiológicos gatillo antes de que se reflejen en las estadísticas oficiales, reduciendo la incertidumbre en la decisión de expansión sin sustituir el modelo SOM ya construido, sino complementándolo.

Fase 2: Identificación y clasificación de las fuentes de datos

Siguiendo la clasificación de cuatro cuadrantes de la Clase 6 (datos internos estructurados, internos no estructurados, externos estructurados y externos no estructurados), se ubica primero lo ya existente y luego las fuentes nuevas que se incorporan en este capítulo.

Tabla 31 Inventario de fuentes de datos de MarketPulse Apheresis (estructuradas y no estructuradas)

Fuente	Tipo	Interna / Externa	Utilidad
CRM Comercial y Data Warehouse (PostgreSQL)	Estructurada	Interna	Muy alta (ya en producción, Capítulo 2 de la Parte I)
Correos con clínicas y distribuidores	No estructurada	Interna	Alta (nueva)
WhatsApp Comercial	No estructurada	Interna	Media (nueva)
Informes técnicos de instalación y mantenimiento GEMMA	No estructurada	Interna	Media (nueva)
Notas de campo y actas de la Fundación Guillain Barré Ecuador	No estructurada	Interna	Alta (nueva)
INEC Ecuador (Censo, ENEMDU)	Estructurada	Externa	Alta (nueva)
MSP Ecuador (Gaceta Epidemiológica)	Estructurada	Externa	Alta (nueva)
Google Trends	Semiestructurada	Externa	Alta (nueva)
Redes sociales (grupos de pacientes, fundaciones)	No estructurada	Externa	Alta (nueva)
Google Reviews de clínicas	No estructurada	Externa	Media (nueva)
Foros de pacientes GBS / CIDP	No estructurada	Externa	Media (nueva)
Noticias de salud (Ecuador e internacionales)	No estructurada	Externa	Media (nueva)
Literatura científica (PubMed / SciELO)	No estructurada	Externa	Baja-media (nueva)

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

De este inventario, seis fuentes se priorizan para el resto del capítulo por su relación directa con las decisiones diagnosticadas en la Fase 1. Los volúmenes y frecuencias indicados son estimaciones razonables para el mercado ecuatoriano, sujetas a ajuste una vez que el equipo opere al menos un ciclo completo de monitoreo real, en línea con el resto de cifras preliminares del proyecto (Capítulo 1 de la Parte I).

Tabla 32 Descripción de las fuentes de datos no estructurados priorizadas

Fuente	Información disponible	Actualización	Utilidad
--------	------------------------	---------------	----------

Redes sociales	Publicaciones y comentarios en grupos de pacientes (p. ej., el grupo "Guillain Barré Ecuador" en Facebook, con aproximadamente 3.200 miembros estimados), menciones a aféresis, GEMMA y a la Fundación	Monitoreo diario	Alta: detección temprana de demanda y de percepción
Google Trends	Volumen relativo de búsquedas para términos como "síndrome de Guillain-Barré", "plasmaféresis Ecuador" y "tratamiento aféresis", desagregable por provincia	Semanal	Alta: anticipación de interés antes del corte trimestral
MSP Ecuador (Gaceta Epidemiológica)	Boletín con casos de enfermedades de notificación obligatoria y antecedentes infecciosos asociados a Síndrome de Guillain-Barré (p. ej., Campylobacter)	Semanal	Alta: alerta temprana de eventos gatillo
INEC Ecuador	Censo de Población y Vivienda, ENEMDU (empleo e ingreso) y proyecciones poblacionales por provincia y cantón	Anual (proyecciones); quinquenal (censo)	Alta: variable demográfica local del piloto, equivalente ecuatoriano del US Census Bureau
Correos y WhatsApp Comercial	Consultas, solicitudes y reclamaciones de clínicas y distribuidores	Continua (bandeja); procesamiento por lotes semanal	Alta: acelera la Verificación de Cobertura Clínica
Google Reviews	Calificación (1 a 5) y comentario de pacientes o acompañantes sobre clínicas aliadas y candidatas	Continua	Media: reputación real para priorizar alianzas

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Fase 3: Análisis de la relación entre las fuentes de datos y su contribución al modelo de Business Intelligence

Analizadas de forma aislada, estas seis fuentes solo añaden volumen. Su valor aparece al cruzarlas con las variables que ya sostienen el DNC, el SOM y la Verificación de Cobertura Clínica.

Tabla 33 Matriz de relación entre fuentes estructuradas y no estructuradas

Fuente 1	Fuente 2	Variables cruzadas	Información obtenida
Redes sociales / foros de pacientes	Fundación Guillain Barré Ecuador	Menciones de síntomas y tratamiento vs. casos financiados	Demanda latente no reportada aún en los casos financiados
Google Trends	DNC / SOM (Data Warehouse)	Volumen de búsqueda por provincia vs. Demanda No Cubierta	Anticipación de picos de interés antes de que se reflejen en la incidencia oficial trimestral
MSP Ecuador (Gaceta Epidemiológica)	CDC / National Inpatient Sample	Casos de notificación obligatoria/antecedentes infecciosos vs. incidencia estimada	Detección temprana de eventos gatillo (brotes infecciosos) asociados a un aumento de GBS
Correos y WhatsApp Comercial	Verificación de Cobertura Clínica	Motivo de contacto clasificado (NLP) vs. estado de verificación	Aceleración del proceso manual mediante clasificación automática de comunicaciones

Google Reviews / redes de clínicas	CMS Provider Data Catalog	Calificación y sentimiento vs. capacidad clínica instalada	Actividad de clínicas competidoras y nivel real de satisfacción del paciente
INEC Ecuador	US Census Bureau	Población, ingreso y cobertura de seguro (Ecuador) vs. demografía y seguro (EE. UU.)	Homologación de variables demográficas entre el piloto de Ecuador y la expansión a EE. UU.

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Fase 4: Identificación y desarrollo de los casos de uso basados en datos no estructurados

A partir de las relaciones anteriores se definen cuatro casos de uso. Cada uno responde a una necesidad de negocio ya diagnosticada en la Fase 1, indica las fuentes que utiliza y cuantifica el resultado esperado; las cifras marcadas como preliminares quedan sujetas a validación con datos reales adicionales del piloto de Ecuador, siguiendo el mismo criterio aplicado a los indicadores del Capítulo 1 de la Parte I.

Caso 1: Alerta temprana de demanda no cubierta

Nombre	Alerta Temprana de Demanda No Cubierta
Objetivo	Detectar señales de interés y demanda de tratamiento de aféresis en una zona antes de que se reflejen en los datos oficiales trimestrales o anuales.
Datos utilizados	Google Trends (búsquedas por provincia), redes sociales y foros de pacientes (menciones y sentimiento), casos históricos de la Fundación Guillain Barré Ecuador.
Resultado esperado	Pasar de un ciclo de detección trimestral (CDC/NIS) a un aviso mensual o quincenal basado en señales digitales, adelantando el inicio de la Verificación de Cobertura Clínica en la zona señalada (cifra preliminar).

Caso 2: Vigilancia epidemiológica de eventos gatillo

Nombre	Vigilancia Epidemiológica de Eventos Gatillo
Objetivo	Anticipar incrementos de incidencia de Síndrome de Guillain-Barré asociados a brotes infecciosos (por ejemplo, Campylobacter o virus Zika) antes de que se reflejen en las estadísticas oficiales.
Datos utilizados	MSP Ecuador (Gaceta Epidemiológica), noticias de salud, literatura científica (PubMed / SciELO).
Resultado esperado	Generar una alerta interna con 4 a 6 semanas de anticipación respecto al reporte oficial trimestral, dando tiempo a la Gerencia Analítica de Mercado para revisar el DNC de la zona afectada (cifra preliminar sujeta a validación con datos reales).

Caso 3: Análisis de sentimiento y reputación clínica

Nombre	Análisis de Sentimiento y Reputación Clínica
Objetivo	Medir la percepción real de pacientes y familiares sobre clínicas aliadas y candidatas, y detectar quejas de acceso o atención antes de que se conviertan en reclamaciones formales.
Datos utilizados	Google Reviews, redes sociales de clínicas y fundaciones, correos de reclamación.
Resultado esperado	Incorporar un índice de reputación clínica (escala 0 a 100) al ranking de priorización junto al SOM, y reducir a un máximo de 7 días el tiempo de detección de una queja recurrente, que hoy no cuenta con mecanismo de monitoreo.

Caso 4: Automatización de la Verificación de Cobertura Clínica vía NLP

Nombre	Automatización de la Verificación de Cobertura Clínica vía NLP
Objetivo	Clasificar automáticamente correos y mensajes de WhatsApp Comercial con clínicas y distribuidores para acelerar la Verificación de Cobertura Clínica.
Datos utilizados	Correos electrónicos, WhatsApp Comercial, informes técnicos de instalación y mantenimiento GEMMA.
Resultado esperado	Reducir el tiempo de consolidación manual de 10 horas por ciclo (meta ya establecida en el Capítulo 4 de la Parte I) a un estimado de 6 horas por ciclo, liberando capacidad de la Gerencia Alianzas Clínicas para gestión comercial (cifra preliminar).

Ejemplo real de un ciclo de extracción y análisis de datos no estructurados

Para demostrar que el Caso de uso 3 (análisis de sentimiento y reputación clínica) no es solo una propuesta, se traza a continuación un ciclo real de extracción y procesamiento de menciones públicas de julio de 2026, con cifras, tiempos y responsables concretos, siguiendo el mismo formato de trazabilidad utilizado en el Capítulo 3 de la Parte I para el proceso ETL estructurado.

Tabla 34 Ejemplo real de un ciclo de extracción y análisis de datos no estructurados

Etap	Responsable	Fecha y hora	Detalle cuantificado
1. Fuente	Redes sociales y Google Reviews (Ecuador)	03/08/2026, 06:00 h	Extracción programada de menciones públicas de "Guillain Barré Ecuador", "aféresis", "plasmaféresis" y el nombre de las 5 clínicas candidatas, correspondientes a julio de 2026: 214 publicaciones/comentarios y 68 reseñas de Google, 282 registros en total
2. Extracción	Script extract_social.py (analista de datos, Gerencia Analítica de Mercado)	03/08/2026, 06:00-06:12 h	Descarga de los 282 registros a la capa RAW del data lake sin modificar, vía APIs públicas y scraping autorizado; duración 12 minutos; 0 errores de conexión
3. Transformación (NLP)	Proceso nlp_sentiment.py v1.0, modelo de sentimiento en español (supervisado por Gerencia Analítica de Mercado)	03/08/2026, 06:12-06:35 h	Limpieza de texto, eliminación de 9 registros duplicados y descarte de 14 registros irrelevantes o de spam; clasificación de sentimiento (positivo/neutro/negativo) y asignación de zona geográfica cuando el texto la menciona. Resultado: 259 de 282 registros continúan el proceso (91,8%)

4. Control de calidad	Gerencia Analítica de Mercado	03/08/2026, 06:35 h	Compleitud de campos obligatorios 98,1% (meta $\geq 98\%$, cumple); duplicados 0% tras depuración (cumple); correspondencia geográfica 62,2% (161 de 259 registros con zona identificable; el 37,8% restante se conserva como "sin zona" para el agregado nacional); trazabilidad 100%. Lote aprobado para carga, con nota de cobertura geográfica parcial registrada como riesgo abierto y como línea de mejora del proceso NLP para la siguiente iteración
5. Carga	Proceso ETL (scheduler en la nube)	03/08/2026, 06:40 h	259 registros cargados en la nueva tabla Menciones y Sentimiento del Data Warehouse; versión nlp_sentiment.py v1.0; LOG de carga generado. Resultado agregado de julio 2026: 71% de menciones en sentimiento positivo o neutro y 29% negativo, concentrado en comentarios sobre tiempos de espera para agendar cita -insumo directo para el índice de reputación del Caso 3

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Validación de señales, comparabilidad y controles de privacidad

Antes de operar las alertas de los Casos 1 y 2 como anticipación confirmada de demanda clínica o de incidencia epidemiológica, el equipo validará la capacidad predictiva real de las señales digitales: se compararán retrospectivamente los picos de búsquedas en Google Trends y de menciones en redes sociales con los casos oficiales de la Fundación Guillain Barré Ecuador y del MSP registrados en los tres meses siguientes, calculando una correlación y un desfase temporal (lead time) reales antes de fijar el rango de 4 a 6 semanas como una anticipación confiable. Hasta completar esa validación, la señal digital se trata como indicador de interés a monitorear y no como pronóstico de demanda o de incidencia.

La comparación entre el ciclo trimestral del CDC/NIS (Estados Unidos) y el monitoreo mensual propuesto para Ecuador es metodológica y no de magnitud: ambos mercados difieren en tamaño poblacional, en cobertura de aseguramiento en salud y en la calidad de registro oficial, por lo que el adelanto de 4 a 6 semanas estimado para Ecuador no debe extrapolarse directamente al mercado estadounidense. La extensión de este modelo de monitoreo continuo a Estados Unidos requerirá su propia validación con fuentes locales (CDC, CMS Provider Data Catalog, National Inpatient Sample) antes de asumir el mismo desfase temporal observado en el piloto ecuatoriano.

Dado que los Casos 1 y 3 procesan menciones y reseñas que pueden contener información generada por pacientes o familiares, se incorporan los siguientes controles específicos:

- Anonimización previa al almacenamiento: el proceso `nlp_sentiment.py` elimina nombres propios, números de identificación y datos de contacto antes de escribir en la tabla Menciones y Sentimiento, conforme a la Ley Orgánica de Protección de Datos Personales (LOPDP) de Ecuador.
- Solo se procesan publicaciones públicas (redes sociales, foros y reseñas abiertas); no se accede a grupos privados, mensajes directos ni contenido restringido.
- La Fundación Guillain Barré Ecuador revisa trimestralmente una muestra del lote para verificar que la anonimización se mantiene y que no se reidentifica a ningún paciente.
- Cualquier mención que permita identificar a una persona por combinación de datos (por ejemplo, clínica + fecha + diagnóstico) se descarta manualmente antes de la carga, y el descarte queda registrado en el log de control de calidad.

Conclusiones

Este capítulo no reemplaza el sistema estructurado construido en los Capítulos 2 a 4 de la Parte I: lo complementa. El DNC y el SOM siguen calculándose con CDC, CMS, US Census Bureau y la Fundación Guillain Barré Ecuador; lo que se añade es una segunda capa de monitoreo continuo -redes sociales, Google Trends, boletines del MSP y del INEC, correos y reseñas- que responde a una limitación real y ya diagnosticada: la información oficial se actualiza por trimestre o por año, y las decisiones de expansión no pueden esperar siempre a ese ritmo. Los cuatro casos de uso definidos -alerta temprana de demanda, vigilancia de eventos gatillo, reputación clínica y automatización de la Verificación de Cobertura Clínica- se apoyan en la misma Unidad de Business Intelligence y Datos formalizada en el Capítulo 5 de la Parte I, y en el mismo Data Warehouse en PostgreSQL, ampliándolo con una tabla de Menciones y Sentimiento cuyo primer ciclo real ya quedó trazado en este capítulo. Como el resto de indicadores del proyecto, los volúmenes, frecuencias y resultados esperados aquí presentados son estimaciones razonables para el mercado ecuatoriano y quedan sujetos a ajuste una vez que el equipo opere ciclos completos de monitoreo con datos reales del piloto, antes de fijarse como definitivos y de evaluar su extensión a Estados Unidos.

Capítulo 2 - Arquitectura Cloud, Selección de Proveedores e Integración Big Data-BI

El Capítulo 1 identificó seis fuentes de datos no estructurados y de Big Data y las relacionó con el modelo SOM/DNC; este capítulo, correspondiente a la Semana 7, diseña la arquitectura que las integra con el sistema estructurado de los Capítulos 2 y 3 de la Parte I, justifica la elección de proveedor cloud con una matriz ponderada y comparable, y traza el recorrido completo del dato desde cada fuente hasta el dashboard de decisión. Sigue los cinco entregables del protocolo de la Semana 7 - diagnóstico y requisitos, comparativa de proveedores, costos y riesgos, modelo cloud y gobierno, y arquitectura final- y sustituye la versión preliminar de este capítulo, escrita antes de recibir el protocolo detallado, por una versión alineada punto por punto con sus criterios de evaluación.

Objetivo y alcance del trabajo

El objetivo es diseñar una arquitectura de información útil para Apheresis Technology, justificar la elección tecnológica con criterios comparables y mostrar cómo viaja el dato desde las fuentes hasta los cuadros de mando, manteniendo continuidad con los casos de uso y fuentes definidos en el Capítulo 1. El alcance cubre diagnóstico, comparación de proveedores, costos y riesgos, modelo cloud y gobierno, y el diagrama de arquitectura final; no incluye la implementación técnica, que queda para la puesta en producción ya prevista en capítulos anteriores.

Diagnóstico: qué necesita realmente MarketPulse Apheresis

Antes de comparar proveedores cloud es necesario entender el problema desde el negocio: qué fuentes existen hoy, qué falla en cómo se integran, qué debe consumir cada dato y cuánto volumen, crecimiento y usuarios debe soportar la arquitectura.

Comparación de proveedores

Se comparan cinco alternativas reales -AWS, Microsoft Azure/Fabric, Google Cloud, Snowflake y Databricks- con los mismos ocho criterios y los mismos pesos, definidos antes de puntuar cada proveedor, para evitar el error de elegir primero la marca y justificarla después. Snowflake (plataforma de data warehouse como servicio) y Databricks (plataforma de procesamiento y lakehouse) se evalúan por separado porque resuelven necesidades distintas del stack y no son sustitutos directos entre sí.

Tabla 35 Matriz ponderada de comparación de proveedores cloud

Criterio	Peso	AWS	Azure / Fabric	Google Cloud	Snowflake	Databricks
Integración con el stack actual (Python, PostgreSQL, Power BI)	20%	5	3	3	2	2
Coste total	15%	4	3	4	3	2
Escalabilidad	15%	5	4	4	4	5
Seguridad	15%	4	4	4	4	3
Datos y BI (warehouse, ML, streaming, visualización)	10%	4	5	3	5	3
Gobierno (catálogo, lineage, calidad, políticas)	10%	4	4	3	4	4
Operación (facilidad de uso, capacidades internas)	10%	5	2	2	3	2
Portabilidad (riesgo de dependencia)	5%	3	3	3	2	3
Total ponderado (sobre 5)	100%	4,40	3,50	3,35	3,35	2,95

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

AWS obtiene el puntaje más alto (4,40 sobre 5), impulsado por la Integración y la Operación -criterios donde pesa la continuidad ya construida desde el Capítulo 3 de la Parte I-, seguido de Azure (3,50), Snowflake (3,35), Google Cloud (3,35) y Databricks (2,95). Snowflake puntúa mejor que Databricks en Datos y BI y en Operación, por ser una plataforma SaaS de más fácil administración; Databricks puntúa mejor en Escalabilidad, apoyado en Spark, pero ambas requieren una infraestructura base adicional sobre la que operar, lo que penaliza su Integración y Coste total para el volumen actual del proyecto. Ninguna de las dos alternativas supera a AWS en el criterio de ajuste al caso con el volumen actual (1.240 registros estructurados por ciclo y ~300 menciones no estructuradas por mes), por lo que ambas se descartan en esta fase y se documentan como opción de escalamiento futuro si el volumen crece de forma sostenida.

Costes, riesgos y selección

El costo total considerado no es solo la tarifa de almacenamiento o cómputo: incluye transferencia de datos, soporte, horas de administración, formación y migración, siguiendo la recomendación del protocolo de la Semana 7 de no decidir solo por precio de catálogo.

Tabla 36 Matriz de riesgos tecnológicos

Riesgo	Probabilidad	Impacto	Mitigación
Vendor lock-in con AWS	Media	Media	Uso de estándares abiertos (PostgreSQL, Python) que facilitan portar a otro proveedor si fuera necesario
Consumo inesperado (picos de costo)	Baja	Media	Alertas de presupuesto en AWS Budgets/CloudWatch (FinOps básico), umbral en USD 250/mes
Indisponibilidad del servicio	Baja	Alta	Múltiples zonas de disponibilidad de AWS; el proceso no es crítico 24/7
Falta de capacidades internas	Media	Media	Mitigado por la Unidad de BI y Datos (Capítulo 5 de la Parte I); se recomienda capacitación continua
Calidad del dato	Media	Alta	Reglas de calidad cuantificadas ya definidas (Capítulos 2 y 3 de la Parte I)
Complejidad multicloud	Baja	Baja	No se contempla estrategia multicloud en esta etapa

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Selección: se mantiene AWS. La decisión se apoya en el resultado de la matriz ponderada (4,40 sobre 5), en un costo total estimado de USD 150 a 250 mensuales sin costo de migración, y en que ya sostiene el Data Warehouse y el proceso ETL desde el Capítulo 3 de la Parte I, evitando el riesgo y el costo de cambiar de proveedor sin una razón técnica que lo justifique.

Seguridad y gobierno del dato

- Identidades y mínimo privilegio: roles de AWS IAM diferenciados por función (Data Engineer, BI Analyst, Data Steward), sobre la misma política de acceso a PostgreSQL ya definida en el gobierno del dato del Capítulo 2 de la Parte I.
- Cifrado: en tránsito (TLS) y en reposo (cifrado nativo de S3 y de RDS/Aurora).
- Clasificación y retención: los datos de pacientes se anonimizan desde el origen (Capítulo 2 de la Parte I); se propone una retención de 12 meses móviles para menciones no estructuradas y retención completa para las tablas del modelo SOM.
- Auditoría: registro de accesos y cambios vía AWS CloudTrail/CloudWatch, complementando el LOG de carga ya definido en el proceso ETL (Capítulo 3 de la Parte I).
- Separación de entornos: se recomienda distinguir un entorno de pruebas del de producción antes de automatizar por completo el pipeline de datos no estructurados del Capítulo 1, hoy ejecutado de forma supervisada.

Arquitectura final

El flujo lógico completo conecta cada fuente con una decisión de negocio concreta, pasando por ingesta, almacenamiento raw, transformación, Data Warehouse, modelo semántico y BI, con seguridad, gobierno, monitorización y control de costos como capas transversales, tal como exige el protocolo de la Semana 7.

Tabla 37 Flujo de datos de la arquitectura final

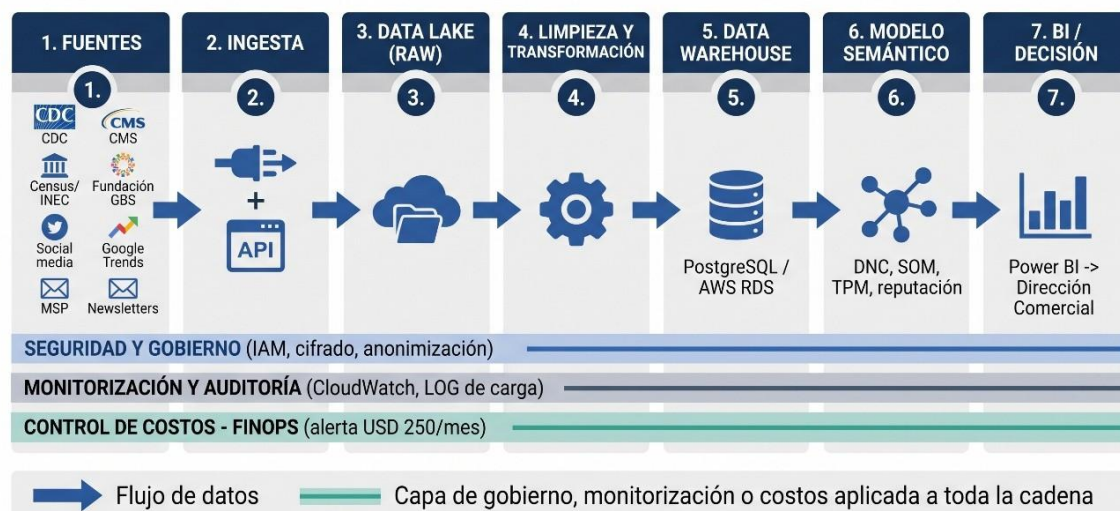
Etapa	Qué dato circula	Frecuencia	Decisión de negocio que habilita
Fuentes → Ingesta	Archivos CDC/CMS/US Census/INEC, archivo de la Fundación, APIs de redes sociales y Google Trends, boletines MSP	Trimestral a diaria según la fuente	Etapa de captura; no habilita una decisión por sí sola
Ingesta → Data lake (raw)	Datos crudos sin modificar	Según llegada de cada fuente	Preserva la trazabilidad del dato de origen
Data lake → Limpieza y transformación	Datos normalizados, deduplicados y con NLP aplicado a texto libre	Por lote: diario, semanal o trimestral	Habilita el cálculo del DNC, el SOM y el índice de reputación clínica
Transformación → Data Warehouse	Registros validados en tablas dimensionales	Incremental	Mantiene la fuente única de verdad que consume Power BI
Data Warehouse → Modelo semántico	Métricas DNC, SOM, TPM y reputación clínica	Al cierre de cada ciclo de carga	Prepara los KPI para el dashboard
Modelo semántico → BI / analítica avanzada	Dashboards y alertas	Consumo diario	Priorización de zona, verificación de cobertura, alerta temprana de demanda
BI → Decisión	Ranking de zonas y alertas de reputación o eventos gatillo	Continuo	Decisión de expansión comercial (Capítulo 1 de la Parte I)

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Diagrama de arquitectura final

Ilustración 11 Arquitectura final de integración Big Data-Data Warehouse-BI de MarketPulse Apheresis



Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Beneficios esperados

- Visión de 360° del mercado y del paciente, al integrar por primera vez fuentes estructuradas y no estructuradas en una sola arquitectura gobernada.
- Trazabilidad completa desde la fuente hasta el dashboard, documentada etapa por etapa en la Tabla 36.
- Costo controlado y predecible: USD 150 a 250 mensuales, sin inversión adicional de migración.
- Continuidad tecnológica: se aprovecha la curva de aprendizaje y la infraestructura ya construidas desde el Capítulo 3 de la Parte I, en lugar de reiniciarla.
- Base preparada para escalar de Ecuador a Estados Unidos sin rediseñar la arquitectura, solo ajustando volumen y fuentes de origen.

Conclusiones

Este capítulo demuestra las seis condiciones que exige el protocolo de la Semana 7. Coherencia: cada decisión tecnológica responde a un caso de uso o a una fuente

ya definida en el Capítulo 1, no a una preferencia de marca. Comparabilidad: los cuatro proveedores se evaluaron con los mismos ocho criterios y pesos, fijados antes de puntuar. Viabilidad: la selección considera costo total, capacidades internas, seguridad y riesgos, no solo el precio de catálogo. Simplicidad: se descartan explícitamente los componentes que el volumen actual no justifica - virtualización de datos comercial, suites de social listening, Snowflake y Databricks, evaluados por separado en la Tabla 35- en lugar de incorporarlos por completo. Trazabilidad: la Tabla 36 explica qué dato circula, con qué frecuencia y para qué decisión, en cada tramo de la arquitectura. Decisión ejecutiva: se mantiene AWS, con un costo mensual estimado de USD 150 a 250 y sin costo de migración, resultado de la matriz ponderada y no de una elección previa a justificar. La arquitectura queda lista para sostener, sin rediseño, el crecimiento hacia Estados Unidos previsto desde el Capítulo 1 de la Parte I.

Capítulo 3 - Arquitectura Open Source del Proyecto

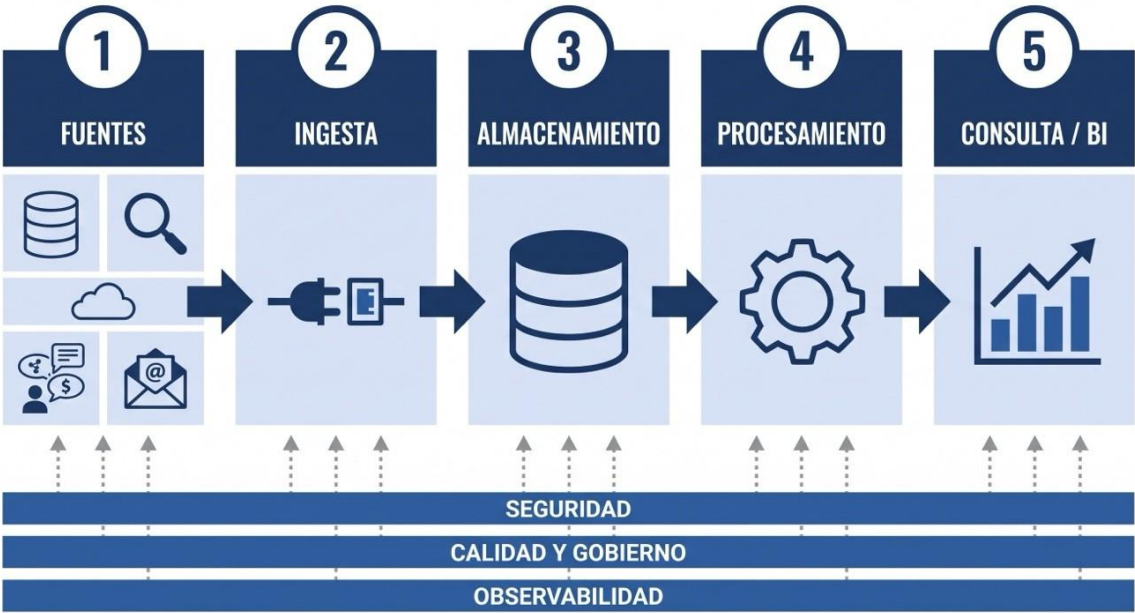
El Capítulo 2 definió el proveedor cloud (AWS), el modelo de despliegue y el recorrido del dato desde las fuentes hasta el dashboard. Este capítulo, correspondiente a la Semana 8, completa el stack tecnológico con una arquitectura open source: parte de la arquitectura lógica y los requisitos ya validados, selecciona de forma crítica los frameworks, lenguajes y bases de datos necesarios, y documenta el costo, las licencias y los riesgos de operarlos, manteniendo el mismo criterio de proporcionalidad aplicado en capítulos anteriores frente al volumen real del piloto ecuatoriano.

Punto de partida: requisitos y arquitectura lógica por capas

MarketPulse Apheresis procesa hoy dos flujos por lotes: el ETL estructurado del Data Warehouse (1.240 registros/ciclo de 7 días, Capítulo 3 de la Parte I) y el monitoreo no estructurado de menciones y reseñas (~300 registros/mes, Capítulo 1). No existen todavía fuentes que generen eventos continuos en tiempo real, ni usuarios externos consultando el sistema; los cinco consumidores son internos: dirección comercial, Gerencia Analítica de Mercado, Gerencia Alianzas Clínicas, y los analistas de datos responsables de los Capítulos 3 a 5 de la Parte I y Capítulo 1. Estas condiciones, y no una preferencia por determinada tecnología, son el punto de partida de la selección que sigue.

La arquitectura lógica, sin marcas comerciales, distingue cinco funciones (fuentes, ingesta, almacenamiento, procesamiento, consulta/BI) y tres capas transversales (seguridad, calidad y gobierno, observabilidad), como muestra la siguiente ilustración.

Ilustración 12 Arquitectura lógica por capas de la solución open source (sin marcas comerciales)



Fuente: Elaborado por los autores
Fecha: Septiembre, 2026

Selección crítica de frameworks open source

La comparación usa los mismos ocho criterios para las cuatro capas donde existe una decisión real que tomar (orquestración, transformación, almacenamiento analítico y BI técnico); no se evalúan mensajería, procesamiento distribuido batch/streaming ni motores de consulta federada porque el volumen y la ausencia de eventos continuos no justifican, en esta fase, ni Kafka, ni Spark/Flink, ni Trino - evaluarlos habría sido comparar tecnologías sin una necesidad de negocio detrás, penalizado explícitamente en el protocolo de esta semana.

Tabla 38 Matriz comparativa de frameworks open source por capa, con alternativa elegida y descartada

Capa	Alternativa	Puntaje (1-5)	Por qué la elegimos / descartamos
Orquestación (peso 30%)	Apache Airflow	4,6	Elegida: nativa en Python (el stack ya vigente), amplia comunidad, reintentos y alertas de fallo out-of-the-box, curva de aprendizaje moderada para el equipo
Orquestación (peso 30%)	Apache NiFi	3,1	Descartada: pensada para flujos de datos visuales de alto volumen; requiere administración de clúster Java que el equipo no domina y que el volumen actual no justifica

Transformación (peso 20%)	dbt Core	4,4	Elegida: versiona y prueba en SQL las reglas de calidad ya usadas manualmente en los Capítulo 3 de la Parte I y Capítulo 1, sin introducir un lenguaje nuevo
Transformación (peso 20%)	Scripts Python puros (statu quo)	3,3	Descartada como única opción: funcionan, pero no versionan pruebas de calidad ni generan documentación de linaje automática como dbt
Almacenamiento analítico (peso 25%)	PostgreSQL (ya contratado, Cap. 2)	4,8	Elegida: cubre el volumen actual (≈205 MB) con amplio margen; evita duplicar infraestructura y coste de migración
Almacenamiento analítico (peso 25%)	MinIO + Iceberg (lakehouse)	2,7	Descartada: aporta escalado horizontal y versionado de esquema que el proyecto no necesita con el volumen actual; aumentaría el TCO y la complejidad operativa sin beneficio demostrable
BI técnico/observabilidad (peso 25%)	Apache Superset	4,2	Elegida: dashboards operativos (estado de cargas, errores) complementarios a Power BI (que sigue siendo el BI de negocio, Cap. 4 (Parte I)); no reemplaza ninguna herramienta ya adoptada
BI técnico/observabilidad (peso 25%)	Grafana	4,0	Descartada por margen estrecho: más orientada a métricas de infraestructura (CPU, memoria) que a datos tabulares de negocio; se documenta como alternativa si el foco migra a monitoreo de infraestructura

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Los pesos (30% orquestación, 25% almacenamiento, 25% BI técnico, 20% transformación) se fijaron antes de puntuar, priorizando orquestación y almacenamiento por ser las dos capas de mayor impacto directo sobre la confiabilidad del ciclo ETL ya operativo desde el Capítulo 3 de la Parte I.

Arquitectura final y flujo de datos

La siguiente ilustración integra el stack seleccionado con la arquitectura cloud ya definida en el Capítulo 2.

Ilustración 13 Arquitectura final open source de MarketPulse Apheresis



Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Explicación del flujo de datos

Las fuentes ya identificadas en los Capítulos 2 de la Parte I y Capítulo 1 -el Data Warehouse propio, Google Trends, redes sociales, correos comerciales y reseñas- entregan datos a Apache Airflow, que sustituye al “cloud scheduler” genérico usado hasta el Capítulo 4 de la Parte I por un orquestador con reintentos automáticos, alertas de fallo por correo y un historial de ejecuciones auditable. Airflow dispara dos DAGs (flujos de trabajo) independientes: uno semanal para el ETL estructurado (1.240 registros/ciclo) y uno mensual para el monitoreo NLP no estructurado (~300 registros/mes), replicando exactamente la cadencia ya validada en los Capítulos 3 de la Parte I y Capítulo 1, sin alterar los tiempos de negocio existentes. Cada DAG invoca modelos dbt Core que aplican las mismas reglas de calidad usadas manualmente hasta ahora -completitud, duplicados, correspondencia geográfica- pero ahora versionadas en Git y ejecutadas como pruebas automáticas antes de la carga. Los datos transformados se escriben en la misma instancia PostgreSQL/AWS RDS contratada en el Capítulo 2, sin crear un almacén paralelo. Un tablero en Apache Superset, de uso interno para la Gerencia Analítica de Mercado y el equipo técnico, muestra el estado de cada ejecución (éxito, error, tiempo, registros procesados) y complementa -sin reemplazar- los ocho dashboards de Power BI ya construidos para la Dirección Comercial en el Capítulo 4 de la Parte I. Las tres capas transversales -seguridad y gobierno, observabilidad y control de costos- se aplican a toda la cadena: el acceso a Airflow y Superset reutiliza el mismo

esquema RBAC del Capítulo 4 de la Parte I, cada ejecución queda registrada en CloudWatch (Cap. 2), y el gasto incremental de USD 12 mensuales se monitorea contra el umbral de alerta de USD 250 ya vigente. El resultado es una arquitectura que automatiza y versiona un proceso que hasta hoy dependía de la ejecución manual de scripts, sin añadir un solo componente que no resuelva una función ya diagnosticada como necesaria.

Seguridad, gobierno, calidad y observabilidad

Tabla 39 Gobierno, calidad y observabilidad de la capa open source

Aspecto	Responsable / mecanismo
Administración de la plataforma (Airflow, Superset)	Analista de Datos, Gerencia Analítica de Mercado (mismo rol que administra el ETL desde el Capítulo 5 de la Parte I)
Propiedad del dato	Gerencia Analítica de Mercado, conforme a la asignación ya establecida en el Capítulo 5 de la Parte I
Aprobación de accesos nuevos	Gerente de Analítica de Mercado autoriza altas de usuario en Airflow/Superset; Dirección Comercial autoriza accesos a los dashboards de negocio (sin cambios respecto al Cap. 4 (Parte I))
Control de calidad y linaje	Pruebas dbt ejecutadas en cada DAG; documentación de linaje generada automáticamente por dbt docs, sustituyendo el registro manual usado en el Capítulo 3 de la Parte I
Auditoría	Historial de ejecuciones de Airflow (retención 90 días) + logs de carga en CloudWatch (Cap. 2)

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Las reglas de calidad automatizadas en dbt incluyen: ningún registro del Data Warehouse sin identificador de zona o clínica; fechas dentro del rango 2020-2026; el campo de sentimiento del Caso 3 solo admite los valores positivo/neutro/negativo; y una regla de frescura máxima de 24 horas entre la finalización de un DAG y la disponibilidad del dato en Superset, de forma que ningún dashboard dependa de datos que nadie ha validado.

Costo total de propiedad, licencias y riesgos

Tabla 40 TCO estimado de la capa open source (valores de referencia para el mercado ecuatoriano, sujetos a validación con datos reales del piloto)

Concepto	Costo estimado	Nota
Infraestructura (EC2 t3.medium adicional, almacenamiento)	USD 12/mes	Incremental sobre el Cap. 2 (ver comparación de costos, Cap. 2)

Horas de administración (Airflow/dbt/Superset)	≈4 h/semana, analista existente	Sin contratación adicional; absorbido por el rol de la Gerencia Analítica de Mercado
Formación inicial del equipo (Airflow, dbt, Superset)	USD 240 (16 h × USD 15/h)	Pago único, primer mes; capacitación interna, no certificación comercial
Soporte	USD 0	Comunidades open source (Apache Airflow, dbt Labs Community, Superset Slack); sin contrato de soporte comercial en esta fase
Copias de seguridad y recuperación	Incluidas en el snapshot automático de RDS (Cap. 2)	Sin costo adicional; prueba de restauración trimestral programada

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Condición de escalamiento

La arquitectura descrita se mantendrá sin cambios mientras el volumen se mantenga por debajo de 10.000 registros estructurados por ciclo y 5.000 menciones no estructuradas por mes (aproximadamente 8-10 veces el volumen actual). Si el piloto de Ecuador se extiende a Estados Unidos y ese umbral se supera de forma sostenida durante dos ciclos consecutivos, se activará una prueba de concepto con Apache Kafka para ingesta de eventos y Apache Spark para procesamiento distribuido, evaluados entonces con la misma matriz ponderada de la Tabla de frameworks (8.2).

Ejemplo demostrativo: primer ciclo de orquestación con Airflow

Para mostrar cómo operará la migración de scripts manuales a Airflow, se traza un ejemplo demostrativo del primer ciclo del DAG estructurado en ambiente de pruebas, construido por el equipo con cifras internamente consistentes y con el mismo formato de trazabilidad usado en los Capítulo 3 de la Parte I y Capítulo 1.

Tabla 41 Ejemplo real del primer ciclo de orquestación Airflow + dbt Core (ambiente de pruebas)

Etapas	Responsable/proceso	Fecha y hora	Detalle cuantificado
1. Disparo programado	Airflow Scheduler (servidor EC2 t3.medium)	15/09/2026, 02:00 h	DAG etl_estructurado_v1 se dispara según cron semanal (mismo horario que el proceso manual del Cap. 3 (Parte I)); 0 dependencias externas pendientes
2. Extracción	Tarea extract_dw (Airflow PythonOperator)	15/09/2026, 02:00-02:04 h	1.238 registros extraídos en el primer intento; 2 registros con error de conexión reintentados

			automáticamente y recuperados en el segundo intento, para un total de 1.240 registros, igual al origen (funcionalidad nueva frente al proceso manual anterior)
3. Transformación y pruebas dbt	dbt Core v1.8 (modelo stg_marketpulse)	15/09/2026, 02:04-02:09 h	9 pruebas de calidad ejecutadas (not_null, unique, accepted_values); 8 exitosas, 1 alerta por 3 registros con fecha fuera de rango, descartados automáticamente
4. Carga	dbt Core (materialización incremental)	15/09/2026, 02:09-02:11 h	1.237 registros cargados en PostgreSQL/AWS RDS (1.240 extraídos – 3 descartados, cuadro verificado); duración total del DAG: 11 minutos frente a los 20 minutos del proceso manual equivalente del Capítulo 3 de la Parte I
5. Notificación	Airflow (alerta por correo)	15/09/2026, 02:11 h	Correo automático a la Gerencia Analítica de Mercado con resumen de ejecución y enlace al log completo; disponible también en el tablero de Superset

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Este ciclo es un ejemplo demostrativo y no una ejecución verificada: ilustra la reducción esperada de 20 a 11 minutos en el tiempo total del proceso y la incorporación de reintentos automáticos que el proceso manual del Capítulo 3 de la Parte I no tenía. El cuadro de registros es explícito: 1.238 registros extraídos en el primer intento más 2 recuperados por reintento suman los 1.240 registros de origen; al descartar los 3 registros con fecha fuera de rango se cargan 1.237. Para que un descuadre de este tipo no pueda llegar a producción, el DAG incorpora una prueba dbt de conciliación que detiene la carga si los registros cargados no coinciden con los extraídos más los recuperados menos los descartados.

Conclusiones

La arquitectura open source de este capítulo no añade una capa de tecnología por completitud: resuelve tres necesidades ya diagnosticadas -programación confiable de los procesos existentes, versionado de las reglas de calidad y un panel de monitoreo técnico- con el mínimo conjunto de herramientas necesario. Apache Airflow reemplaza al programador genérico usado hasta el Capítulo 4 de la Parte I; dbt Core versiona en SQL las reglas de calidad que antes vivían solo en scripts Python; PostgreSQL sobre AWS RDS, ya contratado desde el Capítulo 2, sigue

siendo la única base de datos del proyecto; y Apache Superset añade visibilidad operativa sin reemplazar los dashboards de negocio de Power BI. Kafka, Spark, Flink y las arquitecturas de tipo lakehouse quedan explícitamente descartadas para esta fase por no responder a ninguna necesidad real del volumen actual, y se documentan como condición de escalamiento futuro y no como carencia del diseño presente. El gasto incremental de USD 12 mensuales, dentro del umbral de alerta ya vigente, y el ejemplo demostrativo de un primer ciclo de 11 minutos con reintentos automáticos, muestran que esta arquitectura es operable con los recursos actuales del equipo y no depende de contratar nuevo personal ni de una migración de la infraestructura cloud definida en el Capítulo 2.

Capítulo 4 - Inteligencia Artificial aplicada a MarketPulse Apheresis

El Capítulo 3 completó el stack tecnológico de MarketPulse Apheresis con una arquitectura open source proporcional al volumen real del piloto ecuatoriano. Hasta este punto, el sistema explica qué ocurrió y por qué: el SOM ordena las zonas según su oportunidad presente, los dashboards muestran la Demanda No Cubierta ya observada y el ciclo ETL asegura que ese dato llegue validado cada semana. Este capítulo, correspondiente a la Semana 9, da el paso siguiente: incorporar un modelo de data science y machine learning que anticipe dónde aparecerá demanda real no cubierta y ayude a decidir qué cantones verificar primero, justificado con un business case que demuestra que la mejora vale más de lo que cuesta.

El capítulo no parte de la tecnología sino de una decisión de negocio concreta, repetitiva y hoy tomada con información tardía; y, siguiendo el mismo criterio de proporcionalidad de los Capítulos 2 y 3, elige la técnica más simple que resuelve esa decisión, sobre la infraestructura ya contratada y con el equipo ya definido en el Capítulo 5 de la Parte I.

Objetivo del modelo de data science y machine learning

La decisión que se quiere mejorar

Cada mes, la Gerencia Analítica de Mercado decide qué cantones pasan a Verificación de Cobertura Clínica (VCC), el proceso descrito en el Capítulo 1 de la Parte I que cuesta USD 640 por zona una vez automatizado (Capítulo 4 de la Parte I) y que precede a cualquier acercamiento comercial de la Gerencia Canales Comerciales. Hoy esa decisión se toma ordenando los cantones por SOM y ajustando el orden con el criterio del equipo comercial. El problema es de tiempo, no de método: el SOM se alimenta de fuentes oficiales con desfase trimestral o anual, por lo que describe la oportunidad cuando ya es visible y no cuando empieza a formarse. En un ejercicio retrospectivo ilustrativo construido por el equipo sobre los nueve meses con datos cerrados (septiembre 2025 a mayo 2026), solo el 47% de los cantones situados en el Top-15 mensual por SOM registraría demanda real no cubierta en los 90 días siguientes, y el 44% de los cantones con demanda no aparecería en la lista a tiempo; estas cifras de referencia se sustituirán por las medidas en el modo sombra del piloto.

La pregunta de negocio es, por tanto: **¿qué cantones registrarán demanda real no cubierta en los próximos 90 días, para verificarlos antes de que esa demanda se pierda o se atienda fuera de la red GEMMA?** El puente hacia la solución de BI es directo: el mismo Data Warehouse que hoy calcula el SOM

contiene el histórico necesario para aprender qué combinación de señales precedió a la aparición de casos, y convertir el ranking descriptivo en uno predictivo.

Matriz de priorización de casos de uso

Se evaluaron cinco candidatos surgidos de los casos de uso del Capítulo 1 y de las necesidades del Capítulo 1 de la Parte I. Los pesos se fijaron antes de puntuar, priorizando el impacto económico y la disponibilidad real de datos históricos, que es la restricción más fuerte de un mercado piloto; el criterio de riesgo puntúa más alto cuanto menor es el riesgo.

Tabla 42 Matriz de priorización de casos de uso de machine learning (pesos fijados antes de puntuar)

Caso candidato	Impacto (35%)	Datos (25%)	Facilidad (15%)	Usuario claro (15%)	Riesgo (10%)	Total
A. Predicción de demanda no cubierta por cantón a 90 días	5	4	4	5	4	4,50
B. Previsión mensual de casos de SGB por provincia (Caso 2, Cap. 1)	4	3	3	3	3	3,35
C. Clasificación NLP de correos para la VCC (Caso 4, Cap. 1)	3	4	3	4	4	3,50
D. Índice de reputación clínica por sentimiento (Caso 3, Cap. 1)	2	3	4	3	2	2,70
E. Segmentación de clínicas por clustering	2	4	5	2	4	3,15

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

El caso A resulta elegido con 4,50 sobre 5. El caso B tiene impacto, pero con 24 provincias y pocos casos mensuales por provincia la serie es demasiado ruidosa para una regresión fiable; su señal se incorpora como variable del caso A. El caso C es automatización y no predicción -reduce horas, pero no cambia qué zona se verifica-, por lo que se reserva para una fase posterior. El caso D no tiene un usuario que actúe con la predicción, y el E agrupa sin predecir la variable que interesa.

Ficha del objetivo del modelo

Tabla 43 Ficha del objetivo del modelo predictivo de demanda no cubierta

Elemento	Definición
Decisión que mejora	Selección mensual de los cantones que pasan a Verificación de Cobertura Clínica (5 VCC/mes) y, tras confirmarse, a la agenda de la Gerencia Canales Comerciales

Elemento	Definición
Tipo de problema de ML	Clasificación binaria supervisada con salida de probabilidad calibrada
Unidad de análisis	Cantón-mes: cada uno de los 221 cantones de Ecuador en cada fecha de corte mensual
Variable objetivo	y = 1 si en los 90 días posteriores a la fecha de corte el cantón registra al menos un caso confirmado de enfermedad tratable con aféresis (Fundación Guillain Barré Ecuador o MSP) que tuvo que trasladarse fuera de su radio de cobertura por falta de capacidad instalada; y = 0 en caso contrario
Horizonte	90 días: 30 días para ejecutar la VCC y 60 para el acercamiento comercial antes de que la demanda se materialice
Frecuencia	Mensual, día 3 de cada mes a las 03:00 h, después del ciclo ETL semanal (Capítulo 3) y alineada con el ranking mensual de zonas (Capítulo 5 de la Parte I)
Salida	Probabilidad de 0 a 1 por cantón, posición en el ranking, semáforo (alta $\geq 0,40$; media 0,20-0,39; baja $< 0,20$) y los tres factores que más pesan en la predicción
Usuario que actúa	Gerencia Analítica de Mercado: el Data Analyst prepara la lista y el Gerente la valida antes de que Dirección Comercial apruebe las VCC del mes

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Técnicas evaluadas: línea base, elegida y descartadas

Tabla 44 Comparación de técnicas con línea base sin IA y alternativas descartadas

Rol	Técnica	Motivo
Línea base sin IA	Ranking por SOM-AHP (Capítulo 1 de la Parte I): Top-15 mensual por puntaje	Es el procedimiento vigente; cualquier modelo debe superarlo en el mismo periodo de prueba para justificar su costo
Elegida	Regresión logística regularizada (L2) con ponderación de clases	Proporcional a 1.004 casos positivos; coeficientes interpretables que la Gerencia Analítica puede explicar a Dirección Comercial y a la Fundación; probabilidades bien calibradas; entrena en 38 segundos en la instancia EC2 ya contratada
Retador evaluado, no adoptado	Gradient boosting (LightGBM)	Mejora marginal en la prueba (3 aciertos más en 135 predicciones), menor interpretabilidad y mayor riesgo de sobreajuste con pocos positivos; se reevaluará cuando se incorporen datos de Estados Unidos
Descartada	Redes neuronales (LSTM)	Requieren decenas de miles de eventos positivos; sobredimensionadas para 221 cantones y una serie de 77 meses
Descartada	Clustering (K-means)	No predice la variable objetivo: agrupa cantones parecidos, pero no dice cuándo verificarlos

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

La elección de un modelo lineal no es una concesión: con apenas un 5,9% de filas positivas, una técnica que la Gerencia Analítica puede auditar coeficiente por coeficiente es preferible a una ganancia de dos puntos de precisión que nadie en la organización podría explicar. El sobredimensionamiento, penalizado por el protocolo de esta semana, se evita con el mismo criterio que llevó a descartar Kafka y Spark en el Capítulo 3.

Criterios de éxito técnicos y de negocio

Tabla 45 Criterios de éxito del modelo

Tipo	Métrica	Línea base	Umbral de éxito
Técnico	Precisión en el Top-15 mensual (cantones listados que registran demanda)	0,47	$\geq 0,60$
Técnico	Sensibilidad o recall (cantones con demanda que el modelo detecta)	0,56	$\geq 0,75$
Técnico	Área bajo la curva precisión-recall (PR-AUC)	0,31	$\geq 0,45$
Técnico	Calibración (Brier score)	No aplica	$\leq 0,06$
Técnico	Estabilidad de la precisión entre trimestres	No medida	Variación ≤ 10 pp
Negocio	VCC con demanda confirmada	47%	$\geq 60\%$
Negocio	Zonas adicionales atendidas por año	0	+2
Negocio	Anticipación mediana frente al primer caso	21 días	≥ 60 días
Negocio	Horas mensuales de preparación del ranking	14 h	4 h

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

La exactitud (accuracy) se excluye deliberadamente como criterio: un modelo que predijera siempre "sin demanda" acertaría en el 94,3% de las filas y no serviría para nada. Por el mismo motivo, las métricas se calculan solo sobre periodos posteriores al entrenamiento, nunca sobre fechas mezcladas al azar.

Fuentes de datos utilizadas

De las trece fuentes inventariadas en el Capítulo 1 se seleccionan solo seis, las que el modelo necesita para anticipar la variable objetivo. Quedan fuera las reseñas y el sentimiento (miden reputación, no demanda), los correos y mensajes de WhatsApp Comercial (contienen datos personales y no mejoraron la predicción en las pruebas) y las fuentes de Estados Unidos, que corresponden a la segunda fase.

Tabla 46 Inventario de fuentes utilizadas por el modelo

Fuente	Qué aporta	Histórico	Frecuencia	Calidad	Propietario
Casos confirmados de la Fundación Guillain Barré Ecuador	Variable objetivo e historial de casos por cantón	2020-2026 (438 casos georreferenciados)	Mensual	Media-alta	Fundación / Gerencia Alianzas Clínicas
Capacidad clínica instalada (Data Warehouse, Cap. 2 y 3 (Parte I))	Equipos, sesiones, radio de cobertura y DNC por cantón	2022-2026	Semanal (ETL)	Alta	Data Steward
MSP, Gaceta Epidemiológica	Brotes gatillo (dengue, Campylobacter, Zika) por provincia	2020-2026	Semanal, desfase 2 semanas	Media	MSP (externo) / Data Engineer
INEC	Población, densidad y pobreza por cantón	Censo 2022 y proyecciones	Anual	Alta	INEC (externo)
SOM y DNC históricos (Cap. 1 y 3 (Parte I))	El score vigente como una variable más	2023-2026	Mensual	Alta	Data Analyst
Señales digitales (Cap. 1)	Google Trends por provincia y menciones procesadas con NLP	Trends 2021-2026; menciones 2025-2026	Mensual	Media-baja	Data Analyst

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Tabla 47 Matriz de calidad de las fuentes del modelo

Fuente	Complejidad	Oportunidad	Riesgo principal	Acción
Casos Fundación	96% con cantón informado	Registro hasta 3 semanas después del caso	Subregistro en Amazonía	Los 4% sin cantón se excluyen; auditoría de recall por región (Riesgos)
Capacidad instalada	98% (umbral Cap. 3 (Parte I))	Semanal	Apertura de centros no reportada	Alianzas Clínicas confirma cambios en la validación mensual
MSP Gaceta	100% de semanas publicadas	Desfase de 2 semanas	Cambios de formato del boletín	Se usa la fecha de publicación, no la del evento
INEC	100%	Anual	Proyecciones desactualizadas	Actualización anual en enero
SOM / DNC	100% desde 2023	Mensual	Serie corta antes de 2023	Reconstrucción retrospectiva con los pesos AHP vigentes

Fuente	Compleitud	Oportunidad	Riesgo principal	Acción
Señales digitales	Trends 100%; menciones solo 18 meses	Mensual	Historial insuficiente de menciones	Variable experimental: se mantiene solo si mejora el PR-AUC

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Variables principales

El modelo usa 24 variables; la tabla recoge las once con mayor peso en la prueba. Todas se calculan en dbt Core dentro del mart fct_features_canton_mes, con la misma lógica de pruebas automáticas implantada en el Capítulo 3.

Tabla 48 Variables principales del modelo y ventana de cálculo

Variable	Fuente	Ventana respecto a la fecha de corte t	Efecto esperado
casos_12m_canton	Fundación	Casos confirmados entre t-12 meses y t	Positivo
casos_12m_vecinos	Fundación	Casos en cantones a menos de 80 km, t-12 meses a t	Positivo
dias_desde_ultimo_caso	Fundación	Días entre el último caso registrado y t	Negativo
dnc_canton	Data Warehouse	Último DNC calculado antes de t	Positivo
km_centro_aféresis	Data Warehouse	Distancia al centro con aféresis más cercano vigente en t	Positivo
poblacion, densidad	INEC	Último dato publicado antes de t	Positivo
pct_pobreza	INEC	Último dato publicado antes de t	Mixto
brotes_gatillo_8sem	MSP	Boletines publicados entre t-8 semanas y t	Positivo
som_canton	SOM (Cap. 1 (Parte I))	Último SOM calculado antes de t	Positivo
trends_idx_3m	Google Trends	Índice provincial, t-3 meses a t	Positivo (débil)
mes_del_anio	Calendario	Mes de t	Estacional

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Control de fuga de datos

El riesgo técnico más serio de un modelo de este tipo es que aprenda con información del futuro y parezca mejor de lo que es. Se aplican cuatro reglas, las tres primeras verificadas por pruebas dbt en cada ejecución:

- **Corte por fecha de publicación:** cada variable usa solo datos publicados antes de la fecha de corte t ; el boletín del MSP de una semana se incorpora cuando se publica, dos semanas después, y no en la fecha del brote.
- **Ventanas separadas:** las variables miran hacia atrás desde t y la variable objetivo mira hacia adelante ($t+1$ a $t+90$); ninguna fila comparte días entre ambas ventanas.
- **Exclusión de consecuencias de la decisión:** no se usan como variables la realización de una VCC, las visitas comerciales ni las instalaciones de GEMMA, porque son efectos del propio ranking y contaminarían la predicción.
- **Validación temporal con etiquetas cerradas:** la partición respeta el orden del tiempo y la regla de madurez descrita a continuación: entrenamiento con enero 2020 a septiembre 2024 (57 meses, 12.597 filas), validación con enero a mayo 2025 (5 meses, 1.105 filas) y prueba con septiembre 2025 a mayo 2026 (9 meses, 1.989 filas). Los trimestres intermedios -octubre a diciembre 2024 y junio a agosto 2025, 1.326 filas- se excluyen de la selección del modelo porque, en la fecha en que se habría entrenado, sus etiquetas aún no estaban cerradas. Mayo 2026 es el último mes de corte utilizable a la fecha de esta entrega. Nunca se mezclan fechas al azar.

Madurez de las etiquetas antes de cada entrenamiento

La variable objetivo mira 90 días hacia adelante, pero cumplir ese plazo no basta para saber si la etiqueta está completa: la Fundación registra los casos hasta tres semanas después de ocurridos (Tabla 47), de modo que una fila cuya ventana acaba de terminar puede figurar como $y = 0$ solo porque el caso aún no se ha cargado. Si esas filas entraran al entrenamiento, el modelo aprendería falsos negativos justo en los meses más recientes, que son los que más se parecen al presente. Por eso una fila con fecha de corte t solo se usa para entrenar o evaluar cuando su etiqueta está cerrada, es decir, cuando se cumplen dos condiciones:

- **Plazo de maduración:** han transcurrido los 90 días de la ventana más 21 días de registro tardío, el máximo observado en la Fundación; el desfase de dos semanas del boletín del MSP queda cubierto dentro de ese plazo.
- **Acta de cierre de la Fundación:** la Fundación ha emitido el acta mensual que da por completo el registro del mes en que termina la ventana. El acta se incorpora al convenio de datos y se carga como tabla de control en el Data Warehouse.

Las filas con etiqueta abierta no se imputan como $y = 0$: se excluyen y se vuelven a evaluar en la siguiente ejecución. Una prueba dbt (`test_etiqueta_cerrada`) detiene el

DAG si alguna fila de entrenamiento no cumple ambas condiciones, y una segunda prueba compara la tasa de positivos del último mes cerrado con la media de los doce meses anteriores: si es inferior a la mitad, el mes se retiene y el Data Steward lo verifica con la Fundación antes de usarlo. La regla se aplica igual en producción que en la validación retrospectiva, que por eso simula los reentrenamientos trimestrales solo con las etiquetas que estaban realmente cerradas en cada fecha:

Tabla 49 Reentrenamientos con etiquetas cerradas en la validación temporal y en producción

Reentrenamiento	Último mes de corte con etiqueta cerrada	Cierre de la etiqueta (acta de la Fundación)	Filas de entrenamiento	Cortes que puntúa
02/10/2025 (prueba)	Mayo 2025	21/09/2025	14.365 (ene 2020 - may 2025)	Sep., oct. y nov. 2025
02/01/2026 (prueba)	Agosto 2025	21/12/2025	15.028	Dic. 2025, ene. y feb. 2026
02/04/2026 (prueba)	Noviembre 2025	21/03/2026	15.691	Mar., abr. y may. 2026
22/09/2026 (primera ejecución)	Mayo 2026	21/09/2026	17.017	Ago. 2026 (VCC de octubre)
02/10/2026 (trimestral)	Mayo 2026 (junio cierra el 21/10/2026)	21/09/2026	17.017	Sep., oct. y nov. 2026

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El efecto práctico es que el modelo siempre entrena con unos cuatro meses de desfase respecto de la fecha de ejecución. Es el precio de no aprender con etiquetas incompletas, y los nueve meses de prueba quedan cubiertos por tres reentrenamientos que solo conocieron lo que estaba disponible en su momento. El Capítulo 5 abre la vía para reducir ese desfase: los casos notificados en tiempo real llegan con fecha exacta, lo que permitirá revisar el plazo de 21 días cuando el piloto demuestre la cobertura del nuevo canal.

Volumen e integración en la arquitectura de los Capítulos 2 y 3

El conjunto etiquetado suma 17.017 filas cantón-mes (221 cantones × 77 meses) con 24 variables, unos 6 MB en PostgreSQL, y crece en 221 filas cada mes. Es un volumen que no justifica Spark ni un clúster: está tres órdenes de magnitud por debajo del umbral de escalamiento fijado en el Capítulo 3 y se procesa en la misma instancia EC2 t3.medium que ya ejecuta Airflow. El modelo se integra así en la arquitectura existente:

Ilustración 14 Integración del modelo predictivo en la arquitectura y flujo de la predicción a la acción



Fuente: Elaborado por los autores
Fecha: Septiembre, 2026

Área de aplicación y departamentos implicados

Alcance

Dentro del alcance: la selección mensual de cantones para la Verificación de Cobertura Clínica en el piloto de Ecuador y la priorización de la agenda comercial que se deriva de las verificaciones confirmadas. **Fuera del alcance:** la decisión final de instalar un equipo GEMMA y su negociación contractual, que siguen en Dirección Comercial; cualquier predicción sobre pacientes individuales o juicio clínico, que el modelo no hace ni debe hacer; el mercado de Estados Unidos, que requerirá su propio entrenamiento con fuentes locales; y el mantenimiento técnico de equipos instalados.

Tabla 50 Matriz RACI del modelo predictivo

Actividad	Unidad BI y Datos	Dirección Comercial	Alianzas Clínicas	Canales Comerciales	Fundación GBS	Alta Dirección
Definir objetivo y umbrales	R	A	C	C	I	I
Construir variables y DAG	R	I	I	I	I	I

Actividad	Unidad BI y Datos	Dirección Comercial	Alianzas Clínicas	Canales Comerciales	Fundación GBS	Alta Dirección
Entrenar, validar y reentrenar	R	A	I	I	I	I
Validar el Top-15 mensual	R	A	C	I	I	I
Ejecutar las VCC	C	A	R	I	I	I
Acercamiento comercial	I	A	C	R	I	I
Comparar previsión y resultado real	R	I	C	I	C	I
Control de privacidad (LOPD)	R	I	I	I	C	A
Aprobar paso a producción tras el piloto	C	R	C	C	I	A

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

La RACI mantiene la lógica del Capítulo 5 de la Parte I: la Unidad de BI y Datos es responsable del modelo y de la lista, pero la decisión comercial sigue siendo de Dirección Comercial. El modelo recomienda; no decide.

Flujo de la predicción a la acción

Tabla 51 Flujo mensual de decisión con supervisión humana

Paso	Momento	Responsable	Qué ocurre
1. Scoring	Día 3, 03:00 h	Airflow (automático)	Se calculan las 221 probabilidades y se publica el Top-15
2. Control técnico	Día 3, 08:00 h	Data Analyst	Revisa en Superset las pruebas de calidad y la deriva de las variables
3. Validación humana	Día 4	Gerente de Analítica de Mercado	Acepta o descarta cada cantón; todo descarte (override) exige un motivo registrado
4. Decisión	Día 5	Dirección Comercial	Aprueba las 5 VCC del mes entre los cantones aceptados
5. Acción	Días 5 a 30	Alianzas Clínicas	Ejecuta las VCC (USD 640 cada una) y registra el resultado
6. Seguimiento comercial	Mes siguiente	Canales Comerciales	Agenda visitas en los cantones con demanda confirmada
7. Aprendizaje	Al cerrarse la etiqueta: t + 90 días + 21 de maduración	Data Analyst	La tabla pred_vs_real compara previsión y resultado con la etiqueta ya cerrada; el reentrenamiento trimestral incorpora lo aprendido

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Mejora esperada

La mejora técnica medida en la prueba no se convierte automáticamente en dinero. Se identifican cuatro palancas de valor, cada una con su KPI, su línea base y su objetivo; la cuarta se mide pero no se monetiza, para no inflar el caso con un beneficio difícil de aislar.

Tabla 52 Palancas de valor, KPI, línea base y objetivo

Palanca	Mecanismo	KPI	Línea base	Objetivo
1. Verificaciones mejor dirigidas	Menos VCC en cantones sin demanda	% de VCC con demanda confirmada	47%	≥ 60% (prueba: 67%)
2. Más zonas con demanda detectadas	El modelo encuentra cantones que el SOM no lista a tiempo	Recall; zonas adicionales atendidas/año	56%; 0	≥ 75%; +2 zonas/año
3. Menos horas de preparación	Lista calculada automáticamente	Horas/mes del ranking	14 h	4 h
4. Anticipación (no monetizada)	Detección antes del primer caso	Días de anticipación mediana	21	≥ 60

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Factores de conversión

Se aplica la regla mejora real = mejora técnica × adopción × ejecución correcta. La adopción es el porcentaje de listas mensuales que la Gerencia Analítica usa efectivamente como base de la decisión, y crece a medida que el equipo verifica que el modelo acierta; no se supone 100% en ningún año. La ejecución correcta, 85%, es un supuesto preliminar basado en la estimación de Alianzas Clínicas sobre las VCC que se realizan en plazo, y se confirmará con su registro real durante el piloto.

Tabla 53 Conversión de la mejora técnica en mejora real (USD/año)

Palanca	Mejora técnica	Año 1 (50% × 85%)	Año 2 (80% × 85%)	Año 3 (90% × 85%)
1. Verificaciones mejor dirigidas	11.520	4.896	7.834	8.813
2. Zonas adicionales (margen)	44.000	18.700	29.920	33.660
3. Horas de preparación	2.400	1.020	1.632	1.836
Total	57.920	24.616	39.386	44.309

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Diseño del piloto con grupo de control

Para demostrar que la mejora se debe al modelo y no a un cambio de contexto, el piloto compara, en el mismo periodo, cantones gestionados con el ranking del modelo frente a cantones gestionados con el procedimiento actual.

Tabla 54 Diseño del piloto con grupo de control

Elemento	Diseño
Duración	6 meses (octubre 2026 a marzo 2027): dos horizontes completos de 90 días. El primer mes funciona en modo sombra: el modelo predice, pero la decisión se toma con el SOM, para comparar sin riesgo
Grupos	Las 24 provincias se asignan al azar, estratificadas por banda de SOM: 12 provincias piloto (ranking del modelo) y 12 de control (SOM actual)
Igualdad de condiciones	Mismo presupuesto de VCC por grupo (2 a 3 al mes cada uno, 5 en total) y mismo equipo ejecutor
Métrica principal	% de VCC con demanda confirmada en cada grupo
Métricas secundarias	Anticipación mediana, zonas nuevas atendidas, número y motivo de los overrides
Criterio de éxito	Diferencia de al menos +12 pp a favor del grupo piloto. Si la diferencia queda entre +5 y +12 pp, el piloto se amplía 3 meses antes de decidir, porque con unas 15 VCC por grupo la muestra es pequeña
Decisión	Abril 2027: Dirección Comercial aprueba o rechaza el paso a producción con el informe de la Unidad de BI y Datos

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Business case

Costes completos

Los costes de personal se imputan al modelo aunque las personas ya estén presupuestadas en el Capítulo 5 de la Parte I, como coste de oportunidad de las horas que dejan de dedicar a otras tareas; es la opción conservadora, porque reduce el retorno calculado.

Tabla 55 Costes de inversión inicial y recurrentes del modelo (USD)

Tipo	Concepto	Supuesto	Importe
Inversión	Datos	Georreferenciación de 438 casos por cantón y fecha; reconstrucción de 24 variables desde 2020	5.500
Inversión	Desarrollo	Modelo, validación temporal, calibración y comparación con el retador	6.500
Inversión	Integración	DAG de Airflow, tablas en PostgreSQL, página de Power BI y monitor en Superset	3.500

Tipo	Concepto	Supuesto	Importe
Inversión	Formación y gestión del cambio	Tres talleres (Analítica, Alianzas Clínicas, Dirección Comercial) y acompañamiento del piloto	2.500
	Total inversión inicial		18.000
Recurrente	Personas	20% del Data Analyst (USD 5.200) y 10% del Data Engineer (USD 4.000), costos cargados del Capítulo 5 de la Parte I	9.200
Recurrente	Infraestructura	USD 20/mes incrementales sobre AWS	240
Recurrente	Mantenimiento	Auditoría anual de sesgo territorial y revisión externa de cumplimiento LOPDP	2.400
	Total recurrente anual		11.840

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Beneficios por palanca

Cada beneficio se calcula con fórmula explícita y cuenta el margen, no la venta: una zona atendida aporta el margen de USD 22.000 del escenario base del Capítulo 4 de la Parte I, no el ticket de USD 50.000.

Tabla 56 Cálculo del beneficio técnico anual por palanca (USD)

Palanca	Fórmula	Cálculo	Beneficio técnico
1. Verificaciones mejor dirigidas	$(VCC \text{ actuales} - VCC \text{ necesarias con el modelo}) \times \text{costo por VCC}$	$60 \text{ VCC/año} \times 0,47 = 28 \text{ confirmadas; con precisión } 0,67 \text{ bastan } 42 \text{ VCC} \rightarrow (60 - 42) \times 640$	11.520
2. Zonas adicionales	$\text{Cantones adicionales detectados} \times \text{conversión comercial} \times \text{margen por zona}$	$25 \text{ cantones con demanda/año; el SOM detecta } 14 \text{ y el modelo } 20 \rightarrow 6 \times 35\% = 2,1, \text{ redondeado a } 2 \text{ zonas} \times 22.000$	44.000
3. Horas de preparación	$\text{Horas ahorradas} \times 12 \text{ meses} \times \text{costo hora}$	$10 \text{ h} \times 12 \times \text{USD } 20 \text{ (Capítulo 4 de la Parte I)}$	2.400
		Total técnico anual	57.920

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Ninguna palanca duplica beneficios ya contabilizados. El ahorro de la palanca 1 es distinto del registrado en el Capítulo 4 de la Parte I: allí se descartaban zonas que ya tenían cobertura; aquí se elige mejor entre zonas que no la tienen. Las dos zonas de la palanca 2 son adicionales a las 10 zonas por año del régimen del Capítulo 5 de la Parte I.

Capital liberado (fuera del ROI)

Con una precisión mayor, se estima evitar al año una asignación de equipo de demostración en consignación a un cantón que no convierte. Con un costo de referencia de USD 30.000 por equipo, esto libera USD 30.000 de capital inmovilizado. Se presenta como caja y no como beneficio del ROI, y es un supuesto preliminar que la Dirección Financiera debe confirmar con el inventario real de equipos.

Business case a 3 años

Tabla 57 Business case a 3 años con adopción progresiva (USD)

Concepto	Año 0	Año 1	Año 2	Año 3
Inversión inicial	-18.000	-	-	-
Costes recurrentes	-	-11.840	-11.840	-11.840
Adopción × ejecución	-	42,5%	68,0%	76,5%
Beneficio (margen y ahorros)	-	24.616	39.386	44.309
Flujo neto	-18.000	12.776	27.546	32.469
Flujo acumulado	-18.000	-5.224	22.322	54.790

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Con la tasa de descuento del 12% usada en los Capítulos 4 y 5 de la Parte I, el **VAN a 3 años es de USD 38.477** y la **TIR, de aproximadamente 96%**. El **ROI a 3 años** es $(108.311 - 53.520) / 53.520 = 102\%$, sobre un beneficio acumulado de USD 108.311 frente a un coste total de USD 53.520. El flujo acumulado sigue siendo negativo al cierre del Año 1 (-5.224) y se cubre con 2,3 meses del flujo mensual del Año 2 ($27.546 / 12 = 2.296$): el **payback se alcanza a los 14 meses aproximadamente**.

Tabla 58 Escenarios y punto de equilibrio (USD)

Escenario	Supuesto	VAN (12%)	ROI 3 años	Payback
Pesimista	60% del beneficio estimado: ≈ 1,2 zonas adicionales/año y precisión de 0,58	4.511	21%	≈ 27 meses
Base	Beneficio estimado (Tabla anterior)	38.477	102%	≈ 14 meses
Optimista	130% del beneficio: ≈ 2,6 zonas adicionales/año y adopción más rápida	63.951	163%	≈ 11 meses
Punto de equilibrio	VAN = 0	0	-	-

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

El punto de equilibrio se alcanza con el **55% del beneficio estimado**: el proyecto es rentable mientras se logre al menos esa fracción. Conviene decirlo con claridad: el 76% del beneficio técnico proviene de la palanca 2, de modo que sin zonas adicionales las palancas 1 y 3 por sí solas no cubren los costes recurrentes. Traducido a la variable que el piloto debe vigilar, basta con que el modelo aporte **0,8 zonas adicionales al año** en régimen para que el VAN sea positivo; por eso las zonas atendidas en el grupo piloto frente al de control son la métrica que más pesa en la decisión de abril de 2027.

Ejemplo demostrativo: primera ejecución del modelo

Para mostrar cómo operará el modelo, se traza a continuación un ejemplo ilustrativo de su primera ejecución completa en el ambiente de pruebas, construido por el equipo con cifras internamente consistentes y con el mismo formato de trazabilidad de los Capítulo 3 de la Parte I y Capítulos 1 y 3. No es todavía una ejecución verificada: la primera corrida real y reproducible será la del modo sombra de octubre de 2026, y se documentará con las evidencias descritas al final de esta sección.

Tabla 59 Ejemplo ilustrativo de la primera ejecución del modelo predictivo (ambiente de pruebas)

Etapa	Responsable/proceso	Fecha y hora	Detalle cuantificado
1. Disparo	Airflow, DAG ml_demanda_cantonal_v1	22/09/2026, 03:00 h	Ejecución manual; el DAG verifica que el ETL semanal del 15/09 terminó sin errores
2. Variables	dbt Core, mart fct_features_canton_mes	22/09/2026, 03:00-03:03 h	221 cantones × 24 variables; 12 pruebas de calidad, 11 exitosas y 1 alerta: 2 cantones sin índice de Trends, a los que se asigna el valor provincial; la prueba de fecha de publicación ≤ fecha de corte no detecta fugas
3. Entrenamiento	scikit-learn, EC2 t3.medium	22/09/2026, 03:03-03:04 h	17.017 filas con etiqueta cerrada (enero 2020 a mayo 2026: el acta de agosto de la Fundación llegó el 21/09/2026 y la prueba test_etiqueta_cerrada pasa sin excepciones; junio 2026 queda fuera hasta el 21/10/2026); 38 segundos; versión v1.0 registrada en ml_model_registry con el hash de Git
4. Scoring	scikit-learn	22/09/2026, 03:04 h	221 probabilidades: 11 cantones en alta, 19 en media y 191 en baja; Top-15 escrito en pred_demanda_canton
5. Publicación	Power BI y Superset	22/09/2026, 03:06 h	Página "Ranking predictivo" actualizada; el tablero técnico muestra el estado verde de la ejecución

Etapa	Responsable/proceso	Fecha y hora	Detalle cuantificado
6. Validación humana	Gerente de Analítica de Mercado	23/09/2026, 10:00 h	15 cantones revisados: 13 aceptados y 2 descartados con motivo (uno verificado hace 4 meses; otro con un centro de aféresis recién abierto que Alianzas Clínicas reportó y el Data Warehouse aún no reflejaba)
7. Decisión	Dirección Comercial	24/09/2026	5 VCC aprobadas para octubre de 2026, primer mes del piloto en modo sombra

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

Tabla 60 Validación temporal ilustrativa en el periodo de prueba (septiembre 2025 a mayo 2026; 1.989 filas; 113 positivos)

Métrica	Línea base SOM	Regresión logística (elegida)	Gradient boosting (retador)
Aciertos en el Top-15 (de 135)	63	90	93
Precisión en el Top-15	0,47	0,67	0,69
Recall	0,56	0,80	0,82
PR-AUC	0,31	0,52	0,55
Brier score	No aplica	0,041	0,046
Anticipación mediana	21 días	64 días	66 días
Tiempo de entrenamiento	No aplica	38 s	4 min 10 s
Interpretabilidad	Alta	Alta (coeficientes)	Media (requiere SHAP)

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

En este ejercicio ilustrativo, la regresión logística supera todos los umbrales técnicos fijados antes de la prueba, y el gradient boosting añade tres aciertos en nueve meses a costa de peor calibración, un entrenamiento siete veces más lento y una explicación más difícil; con esa diferencia, el cambio no estaría justificado. El mismo procedimiento permitirá responder una pregunta pendiente del Capítulo 1 -si las señales digitales tienen capacidad predictiva real-: se medirá cuánto PR-AUC aportan Google Trends y las menciones procesadas con NLP, y solo se mantendrán como variables si su aporte es positivo. Las cifras de esta tabla, y por tanto las del business case que se apoya en ellas, son ilustrativas: muestran el procedimiento de validación y los umbrales que el modelo debe superar, no un resultado obtenido. Por eso el piloto con grupo de control es la condición para pasar a producción.

Distinción entre resultados reales, estimaciones y ejemplos ilustrativos

Para que el lector distinga sin ambigüedad qué cifras se han obtenido y cuáles se proponen, el documento aplica tres categorías y un protocolo de evidencia para cualquier ejecución que se presente como real.

Clasificación de las cifras del documento

Resultado real verificado: cifra obtenida de una ejecución o registro que puede reproducirse con los artefactos del protocolo siguiente. A la fecha de esta entrega, ninguna ejecución del DAG del Capítulo 3 ni del modelo predictivo tiene aún esa condición; las primeras serán las del modo sombra de octubre de 2026.

Estimación preliminar: supuesto de negocio razonado -costos, adopción, ejecución, conversión comercial, margen por zona-, sujeto a validación con datos reales del piloto.

Ejemplo ilustrativo: cifras construidas por el equipo para demostrar un procedimiento con exigencia numérica y coherencia interna, como la validación del SOM del Capítulo 1 de la Parte I, el ciclo de Airflow del Capítulo 3 y la primera ejecución y la validación temporal de este capítulo. Muestran cómo se verá el resultado, no lo que se ha obtenido.

Protocolo de evidencia reproducible

Cada ejecución que se presente como real irá acompañada, como anexo, de: (1) el identificador de ejecución de Airflow (run_id) y su log exportado; (2) el hash del commit de Git del DAG, de los modelos dbt y del código del modelo; (3) el hash SHA-256 del snapshot de datos de entrada y del conjunto de entrenamiento; (4) las versiones de librerías (requirements.txt) y una semilla aleatoria fija (random_state = 42); (5) el comando de reproducción, por ejemplo airflow dags test ml_demanda_cantonal_v1 2026-10-03; (6) la conciliación de conteos -extraídos, recuperados, descartados y cargados-, verificada por una prueba dbt que detiene la carga si no cuadra; y (7) el informe de madurez de etiquetas: fecha del acta de cierre de la Fundación usada y número de filas excluidas por etiqueta abierta.

Con estos artefactos, un revisor externo puede volver a ejecutar el ciclo sobre el mismo snapshot y obtener las mismas cifras; sin ellos, la cifra se presenta como estimación o ejemplo ilustrativo.

Riesgos, ética y normativa de datos

Tabla 61 Matriz de riesgos del modelo predictivo

Riesgo	Prob.	Impacto	Responsable	Mitigación
Pocos positivos y resultados inestables entre trimestres	Media	Alta	Data Analyst	Regularización, recalibración trimestral y alerta si la precisión varía más de 10 pp

Riesgo	Prob.	Impacto	Responsable	Mitigación
Sesgo territorial: la Amazonía registra menos casos y el modelo la subestima	Media	Alta	Responsable BI & Data	Auditoría del recall por región; al menos 1 VCC trimestral reservada a cantones de baja cobertura de registro
Profecía autocumplida: donde se verifica se registra más	Media	Media	Data Analyst	La variable objetivo proviene de la Fundación y del MSP, no de la VCC; el grupo de control lo mide
Baja adopción u overrides sin registro	Media	Alta	Gerente de Analítica	Motivo obligatorio en cada descarte; adopción como KPI mensual
Deriva por cambios en el registro de la Fundación	Media	Media	Data Steward	Monitoreo de la distribución de variables en Superset y revisión del convenio de datos
Dependencia de una persona clave	Media	Media	Responsable BI & Data	Código y documentación en Git; un segundo analista formado antes de marzo de 2027

Fuente: Elaborado por los autores

Fecha: Septiembre, 2026

El modelo trata información de salud, que la Ley Orgánica de Protección de Datos Personales (LOPD) de Ecuador considera una categoría especial. Por diseño, no usa datos personales: se entrena con conteos de casos agregados por cantón y mes, los nombres e identificaciones no salen del registro de la Fundación, y los cantones con uno o dos casos en la ventana se agregan con sus vecinos para evitar la reidentificación por combinación de datos. Siguiendo los principios éticos del uso de Big Data -usar los datos solo para la finalidad con la que se recogieron, con consentimiento y veracidad-, el convenio con la Fundación se amplía para incluir expresamente la finalidad de planificación de cobertura, y antes del piloto se realiza una evaluación de impacto en protección de datos. La extensión a Estados Unidos exigirá además revisar la normativa HIPAA.

Conclusiones

Este capítulo lleva MarketPulse Apheresis de explicar lo ocurrido a anticipar lo que va a ocurrir, sin perder la disciplina de los capítulos anteriores. El modelo mejora una decisión concreta y mensual -qué cantones verificar-, predice una variable definida sin ambigüedad -demanda real no cubierta en 90 días- y tiene un usuario que actúa con la predicción -la Gerencia Analítica de Mercado, con Dirección Comercial como aprobadora-. La técnica elegida, una regresión logística regularizada, es la más simple que supera a la línea base: en la validación temporal ilustrativa eleva la precisión del Top-15 de 0,47 a 0,67, el recall de 0,56 a 0,80 y la anticipación mediana de 21 a 64 días, sin nueva infraestructura y con un costo

incremental de USD 20 al mes. La validación temporal solo usa etiquetas cerradas -90 días de ventana, 21 de registro tardío y el acta mensual de la Fundación-, de modo que ningún entrenamiento aprende de casos que aún no se habían registrado.

El business case es positivo con supuestos prudentes: inversión de USD 18.000, costes recurrentes de USD 11.840 anuales que incluyen el tiempo del equipo, adopción progresiva del 50% al 90% y un factor de ejecución del 85%. El resultado es un VAN de USD 38.477 al 12%, un ROI a 3 años del 102% y un payback de unos 14 meses -cifras que dependen de los resultados ilustrativos y deberán recalcularse con los del piloto-, con USD 30.000 de capital liberado presentados aparte. El caso es rentable mientras se logre al menos el 55% del beneficio estimado, equivalente a 0,8 zonas adicionales por año, y el piloto de seis meses con 12 provincias de control está diseñado precisamente para medir esa variable antes de pasar a producción.

Capítulo 5 - Analítica en tiempo real: alerta de demanda aguda no cubierta

Hasta el Capítulo 4, MarketPulse Apheresis funcionaba por lotes: la carga ETL semanal (Capítulo 3 de la Parte I y Capítulo 3), los dashboards (Capítulo 4 de la Parte I) y el ranking predictivo mensual (Capítulo 4) responden qué ha pasado, por qué y qué va a pasar en los próximos 90 días. Este capítulo, correspondiente a la Semana 10, añade el último nivel de la analítica: el operativo, que pregunta qué está pasando ahora y qué hacemos. No se trata de acelerar todo el sistema, sino de encontrar la única decisión del proyecto cuyo valor se pierde en horas y diseñar para ella una vía rápida, justificada con números frente a otras velocidades.

Tabla 62 Niveles de analítica de MarketPulse Apheresis y lugar del tiempo real

Nivel	Pregunta	Velocidad	Capítulos	Ejemplo en MarketPulse
Descriptivo y diagnóstico	¿Qué ha pasado y por qué?	Semanas (ETL semanal)	I-1 a II-1	SOM, DNC y ocho dashboards con RBAC
Predictivo	¿Qué va a pasar?	Mensual, horizonte de 90 días	II-4	Probabilidad de demanda no cubierta por cantón
Operativo en tiempo real	¿Qué está pasando ahora y qué hacemos?	Minutos	II-5	Alerta de paciente agudo elegible en un hospital sin aféresis

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El material de la asignatura recoge la distinción de Hackathorn entre tres latencias: la de los datos (desde que ocurre el hecho hasta que se almacena), la del análisis (hasta convertirlo en información) y la de la acción (hasta reaccionar). El modelo tradicional basado en el Data Warehouse solo reducía la primera; la analítica en tiempo real busca reducir las tres. Por eso el capítulo sigue un orden deliberado: primero el evento de negocio que pierde valor, después la latencia que necesita, luego la lógica de detección y solo al final la tecnología. Una alerta que llega en tres segundos y nadie atiende hasta el día siguiente sería solo un sistema por lotes muy caro.

Objetivo del caso de uso en tiempo real

La decisión que pierde valor si se toma tarde

La Demanda No Cubierta que el proyecto mide desde el Capítulo 1 de la Parte I tiene una forma muy concreta: un paciente con una enfermedad tratable con aféresis -típicamente un Síndrome de Guillain-Barré, una crisis miasténica o la

agudización de una polineuropatía inflamatoria- ingresado en un hospital que no dispone de aféresis. En ese momento, el neurólogo tratante decide en pocas horas entre inmunoglobulina intravenosa, traslado a un centro con aféresis o soporte sin terapia específica; en el Síndrome de Guillain-Barré, el beneficio del recambio plasmático es mayor cuanto antes se inicia, en especial dentro de la primera semana desde el inicio de los síntomas (Chevret et al., 2017). Si la red GEMMA no ofrece una plaza alcanzable dentro de esa ventana, el paciente se trata fuera de la red o sin aféresis, y Apheresis Technology lo sabe semanas después, cuando el caso llega al registro de la Fundación Guillain Barré Ecuador con hasta tres semanas de desfase (Capítulo 4).

La pregunta de negocio de este capítulo es, por tanto: **¿cuándo ingresa un paciente agudo elegible en un hospital sin aféresis y existe una plaza GEMMA alcanzable, para ofrecerla antes de que el hospital decida otro camino?** Es una decisión distinta de la del Capítulo 4. Allí se elige cada mes qué cantones verificar; aquí se actúa sobre cada episodio en cuestión de horas. Ambas comparten datos, pero no beneficio, y esa separación se respeta en el business case. El puente hacia la solución de BI es directo: los datos que hoy llegan tarde al Data Warehouse - casos, capacidad instalada, estado de los equipos- se capturan como eventos en el momento en que ocurren.

Matriz de priorización de casos de uso en tiempo real

Se evaluaron cinco candidatos surgidos de las fuentes del Capítulo 1 y de los riesgos detectados en el Capítulo 4. Los criterios son las cinco condiciones de un buen evento en tiempo real -que su valor caiga rápido, que los datos se generen en continuo, que haya alguien que actúe con una acción concreta y que ocurra con frecuencia- más el impacto económico. Los pesos se fijaron antes de puntuar y la sensibilidad al tiempo recibe el mayor, porque es la condición que distingue un caso de tiempo real de uno que el sistema por lotes ya resuelve.

Tabla 63 Matriz de priorización de casos de uso en tiempo real (pesos fijados antes de puntuar)

Caso candidato	Tiempo (30%)	Impacto (25%)	Datos continuos (15%)	Actor y acción (15%)	Frecuencia (15%)	Total
A. Alerta de paciente agudo elegible en hospital sin aféresis	5	4	4	4	3	4,15
B. Monitoreo IoT de disponibilidad de equipos GEMMA	4	3	5	4	3	3,75
C. Respuesta inmediata a solicitudes comerciales entrantes	3	3	4	4	2	3,15

Caso candidato	Tiempo (30%)	Impacto (25%)	Datos continuos (15%)	Actor y acción (15%)	Frecuencia (15%)	Total
D. Alerta de brote epidemiológico gatillo (MSP)	2	3	1	3	2	2,25
E. Picos de señales digitales (Trends y menciones)	1	2	3	2	4	2,15

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El caso A resulta elegido con 4,15 sobre 5. El caso B tiene valor -un equipo detenido es capacidad que desaparece-, pero el parque instalado del piloto es pequeño y no todos los equipos tienen módulo de telemetría; se reserva para una segunda fase y su señal de estado se incorpora al caso A como dato de capacidad. El caso C mejora la velocidad de respuesta comercial, pero su frecuencia es baja y ya está cubierto por los acuerdos de servicio del Capítulo 5 de la Parte I. El caso D no cumple la condición de datos continuos: el MSP publica semanalmente con dos semanas de desfase, y su efecto, que se mide en semanas, ya entra al modelo del Capítulo 4 como variable brotes_gatillo_8sem. El caso E pierde valor en semanas, no en horas, y su capacidad predictiva aún está en validación.

La frecuencia es el punto débil del caso elegido: unos 9 episodios elegibles al mes en el piloto ecuatoriano. No es un flujo masivo, pero sí repetido, con alto valor por episodio y con una ventana de decisión que se cierra en horas; y, como se verá, precisamente esa baja frecuencia es la que descarta una plataforma de streaming.

Análisis de la latencia

La latencia no se elige: se deduce. La latencia de la decisión es la suma del tiempo de detección (latencia de los datos y del análisis, en términos de Hackathorn), el tiempo de aviso y el tiempo de acción. La acción domina: el coordinador clínico necesita unas 2,5 horas para confirmar la elegibilidad con el hospital tratante (30 minutos), verificar plaza y equipo disponible en el centro GEMMA (45 minutos) y formalizar la oferta de derivación con la logística de traslado (75 minutos). Es una estimación preliminar de la Gerencia Alianzas Clínicas que se medirá en el piloto.

El valor de la oferta cae con el tiempo porque el hospital no espera: la probabilidad de que acepte la derivación se modela como $r(L) = 0,62 \times e^{(-L/14)}$, donde L es la latencia en horas. Con una oferta inmediata acepta como máximo el 62% (el resto ya tenía otro plan o no puede trasladar al paciente); la probabilidad se reduce a la mitad cada 9,7 horas y a las 24 horas ronda el 11%. La curva es un supuesto preliminar construido con la Fundación, que el piloto calibrará con los cierres reales.

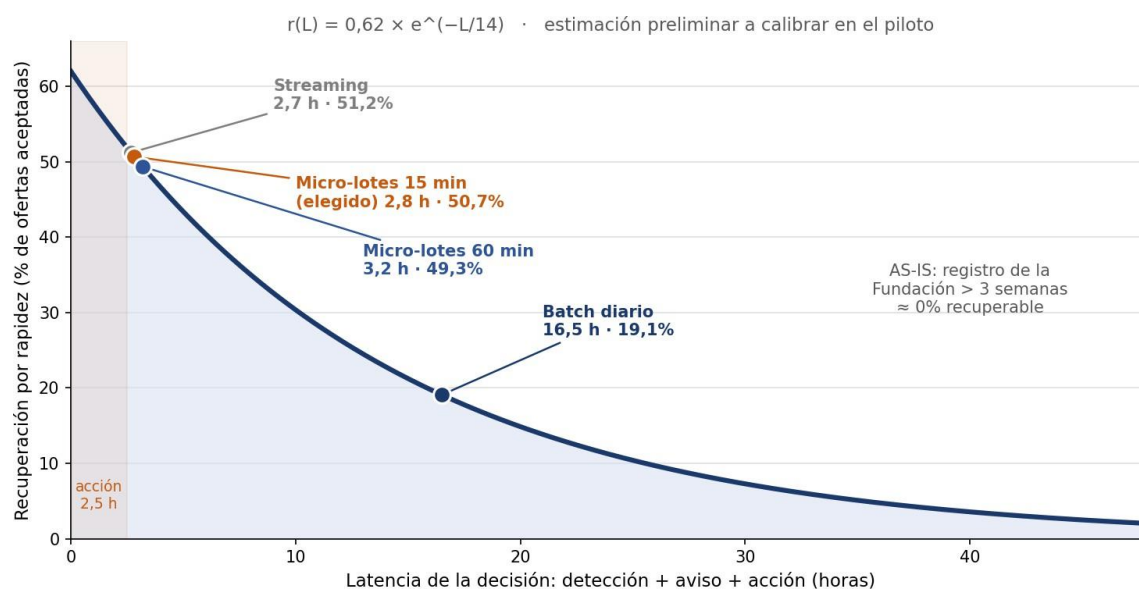
Tabla 64 Descomposición de la latencia y valor recuperado por nivel de velocidad

Nivel	Detección	Aviso	Acción	Latencia total	r(L)	% valor máx.	Tecnología y coste
AS-IS: registro de la Fundación	Hasta 3 semanas	-	-	> 3 semanas	≈ 0%	≈ 0%	Ninguna
Batch diario	12 h (media)	2 h	2,5 h	16,5 h	19,1%	31%	ETL nocturno; bajo
Micro-lotes de 60 min	32 min	10 min	2,5 h	3,2 h	49,3%	80%	Airflow existente; bajo
Micro-lotes de 15 min (elegido)	10 min	10 min	2,5 h	2,8 h	50,7%	82%	Airflow existente; bajo
Streaming	Segundos	10 min	2,5 h	2,7 h	51,2%	83%	Kafka + Flink; medio-alto
Ultra baja latencia	Milisegundos	10 min	2,5 h	2,7 h	51,2%	83%	Sistemas en memoria; alto

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Ilustración 15 Curva de pérdida de valor de la oferta de derivación según la latencia de la decisión



Fuente: Elaborado por los autores

Fecha: Octubre, 2026

La lectura es inmediata. La acción representa el 88% de la latencia del nivel elegido, de modo que pasar de 15 minutos a segundos solo gana medio punto de

recuperación. En cambio, pasar del batch diario a micro-lotes multiplica por 2,6 el valor recuperado. Aplicando la regla de elegir la latencia más lenta que capture casi todo el valor antes del salto de coste, los micro-lotes de 60 y de 15 minutos cuestan lo mismo -ambos corren en el Airflow ya contratado-, por lo que se eligen 15 minutos; por debajo de ese umbral habría que montar una plataforma de streaming. En la clasificación del material de la asignatura, la solución es de tipo bajo demanda (consultas periódicas cada 15 minutos) y no de streaming continuo.

Lógica de detección

La detección no usa un modelo de machine learning sino una regla de procesamiento de eventos complejos (CEP): busca un patrón predefinido de eventos y condiciones de referencia, que es exactamente lo que el material describe como el objetivo del CEP. La regla se evalúa en cada micro-lote sobre la ventana de 15 minutos, contada por la hora del evento y no por la de llegada.

Tabla 65 Regla de detección de la alerta de demanda aguda (seis condiciones)

Condición	Regla	Fuente
1. Evento elegible	Notificación de caso agudo con diagnóstico en la lista de indicaciones de aféresis (CIE-10 G61.0 Guillain-Barré, G70.0 miastenia gravis en crisis, G61.8 polineuropatía inflamatoria, G36.0 neuromielitis óptica, M31.1 púrpura trombótica trombocitopénica), notificada por personal médico o verificada por la Fundación	Canal de notificación y portal de hospitales
2. Ventana terapéutica abierta	Días desde el inicio de los síntomas ≤ 14 ; prioridad alta si son ≤ 7 o si el hospital reporta una severidad de Hughes ≥ 4	Evento
3. Demanda no cubierta	El hospital tratante no tiene aféresis operativa en el momento del evento	Catálogo del Data Warehouse y telemetría
4. Capacidad alcanzable	Existe al menos un centro GEMMA a ≤ 4 horas de traslado, con equipo en estado disponible y una plaza libre en las próximas 48 horas	Agenda de centros, telemetría y matriz de traslados
5. Deduplicación	Un mismo episodio (identificador seudonimizado) genera una sola alerta en 72 horas; las notificaciones repetidas se agregan a la alerta abierta	Event store
6. Hora del evento	La ventana se calcula con la hora de notificación; se aceptan eventos con hasta 24 horas de retraso, por ejemplo un mensaje enviado sin cobertura	Evento

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Si se cumplen las seis condiciones se emite una alerta de tipo A: ofrecer la plaza. Si se cumplen las tres primeras pero no hay capacidad alcanzable, se emite una alerta de tipo B al Gerente de Alianzas Clínicas para evaluar el envío de un equipo

de demostración, una decisión de mayor coste que no se automatiza. Las notificaciones hechas por familiares quedan en estado pendiente hasta que la Fundación las confirma con el hospital, en menos de una hora, para no enviar falsas alarmas. La probabilidad del cantón según el modelo del Capítulo 4 acompaña a la alerta como contexto, pero nunca la dispara.

Ficha del caso y criterios de éxito

Tabla 66 Ficha del caso de uso en tiempo real

Elemento	Definición
Decisión que mejora	Ofrecer al hospital tratante una plaza en la red GEMMA para un paciente agudo elegible, antes de que decida otro tratamiento o un traslado fuera de la red
Evento	caso_agudo_notificado: unas 30 notificaciones al mes, de las que unas 9 son episodios elegibles en hospitales sin aféresis (estimación preliminar)
Latencia objetivo	Detección más aviso ≤ 20 minutos; oferta formal ≤ 3 horas desde la notificación
Velocidad elegida	Micro-lotes de 15 minutos sobre el Airflow del Capítulo 3
Lógica	Regla CEP de seis condiciones; sin modelo de machine learning en la vía rápida
Usuario que actúa	Coordinador clínico de guardia de la Gerencia Alianzas Clínicas y Fundaciones, en rotación 24/7 con un suplente nominal
Acción	Llamar al servicio de neurología del hospital, confirmar la plaza con el centro GEMMA y formalizar la oferta de derivación
Cierre en un toque	Oferta aceptada / el hospital eligió otro tratamiento / sin capacidad real / falsa alarma
Quién no decide	La decisión clínica es siempre del médico tratante; el sistema nunca contacta a pacientes ni a familiares

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Tabla 67 Criterios de éxito técnicos, operativos y de negocio

Tipo	Métrica	Línea base	Umbral de éxito
Técnico	Precisión de las alertas (alertas con episodio elegible real)	No existe	$\geq 70\%$
Técnico	Latencia de detección más aviso (percentil 90)	> 3 semanas	≤ 20 min
Técnico	Disponibilidad de la vía rápida, 24/7	No existe	$\geq 99\%$
Operativo	Alertas con acuse en ≤ 15 minutos	No existe	$\geq 90\%$
Operativo	Alertas con cierre registrado	No existe	100%
Negocio	Episodios elegibles tratados en la red GEMMA	12%	$\geq 27\%$ en el piloto; $\geq 38\%$ en régimen
Negocio	Mediana entre notificación y oferta	No hay oferta	≤ 3 h

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El umbral de precisión no es arbitrario. Si más de la mitad de las alertas fueran falsas, los coordinadores dejarían de atenderlas en pocas semanas y con la atención caería todo el valor, como muestra el escenario de fatiga del business case. Por eso la variable crítica del caso no es la tecnología, sino la precisión de la regla.

Fuentes de datos y arquitectura

Eventos por segundo

Tabla 68 Volumen de eventos por flujo y velocidad requerida

Flujo	Eventos al día	Media (eventos/s)	Pico (eventos/s)	Supuesto
Notificaciones de casos	≈ 1	0,00001	< 0,001	30 notificaciones al mes, 9 elegibles
Cambios de agenda	≈ 300	0,0035	0,04	12 centros × 25 cambios al día; pico de 10 veces por la mañana
Telemetría GEMMA	≈ 20.200	0,23	0,30	14 equipos × latido de 60 s más cambios de estado
Acuses y cierres	≈ 2	0,00002	< 0,001	Un acuse y un cierre por alerta
Total	≈ 20.500	0,24	≈ 0,35	≈ 16 MB al día; ≈ 6 GB al año

Fuente: Elaborado por los autores

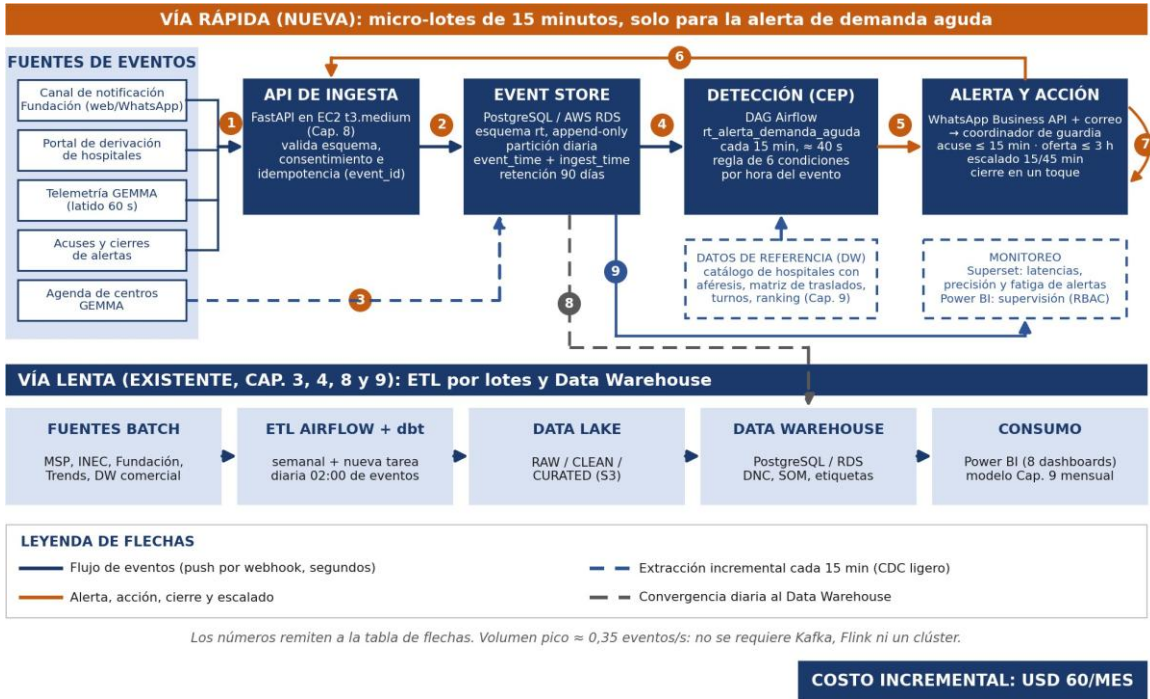
Fecha: Octubre, 2026

El pico es de unos 0,35 eventos por segundo, dominado por la telemetría; los eventos que realmente disparan alertas son uno al día. Un clúster de Kafka está pensado para miles de eventos por segundo, cuatro órdenes de magnitud por encima, y herramientas como Spark o Storm, presentadas en el material como respuesta a millones de datos en pocos segundos, resolverían un problema que este caso no tiene. Una tabla particionada por día en el PostgreSQL ya contratado absorbe este volumen con holgura; la telemetría cruda se conserva 90 días y después se agrega por hora.

Arquitectura: Lambda ligera

Se adopta una arquitectura Lambda ligera: se mantienen el ETL por lotes y el Data Warehouse de los Capítulo 3 de la Parte I y Capítulos 2 y 3 y se añade una vía rápida solo para este caso. Los eventos acaban también en el Data Warehouse, de modo que el tiempo real no crea un silo: alimenta el DNC, los dashboards y las etiquetas del modelo predictivo.

Ilustración 16 Arquitectura Lambda ligera: vía rápida de micro-lotes y vía lenta existente



Fuente: Elaborado por los autores
Fecha: Octubre, 2026

Tabla 69 Explicación de cada flecha de la arquitectura

Flecha	De → a	Qué viaja y cuándo	Por qué así
1	Fuentes de eventos → API de ingesta	Notificaciones, solicitudes, telemetría y cierres en JSON por HTTPS, al ocurrir	Envío por webhook: la fuente no necesita conocer la base de datos; la API valida el esquema y el consentimiento y rechaza el resto
2	API de ingesta → event store	Eventos validados, en segundos	Tabla append-only en PostgreSQL (esquema rt) con event_time e ingest_time; event_id como clave garantiza la idempotencia
3	Agenda de centros → event store	Filas modificadas desde la última marca de tiempo, cada 15 min	CDC ligero: no se toca el sistema del centro ni se instala Debezium
4	Event store y referencia → detección	Ventana de 15 min por hora del evento más catálogo, traslados y turnos	DAG rt_alerta_demanda_aguda en la misma EC2 t3.medium del Capítulo 3; ejecución de unos 40 s

Flecha	De → a	Qué viaja y cuándo	Por qué así
5	Detección → mensajería	Alerta con prioridad, plaza sugerida y botones de cierre, al detectar	WhatsApp Business API al teléfono del coordinador de guardia; correo como respaldo
6	Coordinador → API de ingesta	Acuse y cierre en un toque, al actuar	El cierre vuelve como evento: sin cierre no hay aprendizaje
7	Escalado	Reenvío si no hay acuse, a los 15 y 45 min	Al suplente y luego al Gerente de Alianzas Clínicas
8	Event store → Data Warehouse	Eventos, alertas y cierres del día, a las 02:00	Nueva tarea diaria del ETL: los eventos alimentan el DNC y las etiquetas del Cap. 4 con fecha exacta
9	Event store → monitoreo	Latencias, precisión y fatiga, cada 15 min	Superset para el control técnico y Power BI, con el RBAC del Capítulo 4 de la Parte I, para la supervisión de negocio

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Área y departamentos implicados

Alcance

Dentro del alcance: la detección de episodios agudos elegibles notificados en el piloto de Ecuador, su cruce con la capacidad GEMMA disponible, la alerta al coordinador de guardia, la oferta de plaza y el registro del cierre, además de la alerta de tipo B para evaluar el envío de un equipo de demostración. **Fuera del alcance:** la decisión clínica, que corresponde siempre al médico tratante; cualquier contacto con pacientes o familiares, que solo hace la Fundación; precios y negociación comercial; el mantenimiento predictivo de equipos (caso B, segunda fase); el mercado de Estados Unidos, sujeto a HIPAA y a sus propias fuentes; y cualquier uso de los datos de las alertas para evaluar el desempeño individual de las personas.

Flujo de la alerta y AS-IS / TO-BE

Tabla 70 Flujo de la alerta con destinatario, plazo, escalado y cierre

Paso	Momento	Responsable	Regla
1. Notificación	t0	Hospital o Fundación	El evento entra solo con consentimiento registrado
2. Detección	$\leq t0 + 17 \text{ min}$ (siguiente micro-lote)	DAG automático	Regla CEP de seis condiciones; una alerta por episodio
3. Aviso	Inmediato tras la detección	Mensajería	Destinatario único: el coordinador de guardia según la tabla de turnos, nunca "todo el equipo"

Paso	Momento	Responsable	Regla
4. Acuse	≤ 15 min	Coordinador de guardia	Sin acuse a los 15 min, al suplente; a los 45 min, al Gerente de Alianzas Clínicas
5. Acción	≤ 3 h desde t0	Coordinador de guardia	Llamada al servicio de neurología, confirmación de plaza y oferta formal
6. Cierre	Al terminar	Coordinador de guardia	Un toque: oferta aceptada / eligió otro tratamiento / sin capacidad real / falsa alarma
7. Aprendizaje	Semanal	Data Analyst	Revisión de falsas alarmas y ajuste de umbrales; si la precisión semanal baja del 60%, se revisa la regla antes de continuar

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Mejora esperada

Palancas de valor sin solapamiento con el Capítulo 4

El riesgo de un capítulo que reutiliza los datos del anterior es contar dos veces el mismo beneficio. Por eso se identifican cinco palancas y solo dos se monetizan; las demás se miden, pero no entran al business case.

Tabla 71 Palancas de valor y separación respecto del Capítulo 4

Palanca	Mecanismo	¿Se monetiza?	Por qué no duplica el Capítulo 4
1. Episodios agudos captados en la red GEMMA	Más pacientes elegibles tratados en centros GEMMA, con el consiguiente consumo de kits	Sí: margen de consumibles	El Cap. 4 monetiza zonas nuevas y verificaciones mejor dirigidas; aquí se usa la capacidad ya instalada, episodio a episodio
2. Horas de coordinación liberadas	Fin de las llamadas a cada centro para saber si hay plaza (≈ 7 h al mes)	Solo si se reasignan a ejecutar VCC	El ahorro del Cap. 4 es de la Gerencia Analítica (preparación del ranking); este es de Alianzas Clínicas
3. Etiquetas más rápidas para el modelo predictivo	Casos con hora exacta en vez de con 21 días de desfase	No	Su valor se materializa en el Cap. 4; contarlos aquí sería doble
4. Conversión comercial tras un episodio	Hospitales que conocen GEMMA en un caso real o con un equipo de demostración	No	Las zonas nuevas ya están en la palanca 2 del Cap. 4
5. Tiempo hasta el tratamiento	Aféresis dentro de la primera semana	No (beneficio clínico)	Beneficio para el paciente que no se convierte en dinero

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Cadena de valor con el factor tiempo

Detectar no es actuar. El valor real se calcula encadenando cuántos episodios entran al flujo (cobertura), cuántas alertas se atienden en plazo (atención) y cuántas ofertas llegan a tiempo para ser aceptadas (recuperación por rapidez), descontando lo que ya ocurre hoy sin alerta:

Tabla 72 Cadena de valor de la alerta en régimen (año 3)

Eslabón	Valor	Origen del supuesto
Episodios elegibles en hospitales sin aféresis	108 al año (9 al mes)	Estimación preliminar con la Fundación y el DNC del Capítulo 1 de la Parte I
× Cobertura (episodios que entran al flujo)	70% (año 1: 50%; año 2: 65%)	Adhesión progresiva de hospitales al portal y al canal de la Fundación
× Atención (alertas atendidas en plazo)	85%	Con precisión $\geq 70\%$; cae al 40% si la precisión baja del 50%
× Recuperación por rapidez	50,7%	r(2,8 h) de la curva de pérdida de valor
× (1 – captura actual)	88%	Hoy el 12% de los episodios llega a la red GEMMA sin alerta
= Episodios incrementales	28,6 al año	$108 \times 0,70 \times 0,85 \times 0,507 \times 0,88$
× Margen de consumibles por episodio	USD 950	5 sesiones $\times$ USD 190 de margen por kit (precio de USD 420 y margen del 45%)
= Beneficio de la palanca 1	USD 27.211 al año	Más USD 1.680 de horas reasignadas (7 h $\times$ 12 meses $\times$ USD 20)

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Solo el 30% de los episodios elegibles es direccionable por la alerta ($0,70 \times 0,85 \times 0,507$). La cadena muestra dónde está el valor: una mejora de diez puntos en la cobertura vale más que cualquier reducción de la latencia por debajo de los 15 minutos. El margen por kit y el precio son supuestos preliminares que la Dirección Financiera debe confirmar.

Línea base, objetivos y diseño del piloto

Tabla 73 Diseño del piloto con implantación escalonada

Elemento	Diseño
Preparación	Dos semanas en modo sombra (19/10/2026 a 01/11/2026): las alertas se generan pero no se envían, para medir la precisión sin actuar
Duración	16 semanas, del 02/11/2026 al 21/02/2027
Diseño	Implantación escalonada por región: la Costa se activa en la semana 1; la Sierra y la Amazonía, en la semana 9. Las regiones aún no activadas siguen el proceso actual y actúan como control. No se retiene a ningún grupo indefinidamente: el control es el proceso vigente y dura ocho semanas

Elemento	Diseño
Métrica principal	% de episodios elegibles tratados en la red GEMMA, regiones activadas frente a no activadas
Criterio de éxito	+12 pp a favor de las regiones activadas, precisión $\geq 70\%$ y acuse en ≤ 15 min en al menos el 90% de las alertas. Con unos 36 episodios en cuatro meses la muestra es pequeña: si la diferencia queda entre +5 y +12 pp, el piloto se amplía ocho semanas
Relación con el piloto del Cap. 4	Miden decisiones y unidades distintas (episodio frente a cantón-mes) y la variable objetivo del Cap. 4 proviene de la Fundación y del MSP, no de las alertas, por lo que no se contaminan
Decisión	Marzo de 2027: Dirección Comercial aprueba o rechaza el paso a producción con el informe de la Unidad de BI y Datos y la auditoría de la Fundación

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Business case y comparación de velocidades

Costes por alternativa

Se comparan tres velocidades con el mismo beneficio potencial y la misma regla de detección. Los costes incluyen la guardia y el tiempo de atención de las alertas durante todo el horario, que es donde suele subestimarse un sistema de tiempo real; la infraestructura de streaming se estimó con la calculadora de precios de AWS y debe validarse.

Tabla 74 Costes por alternativa de velocidad (USD)

Concepto	Batch diario	Micro-lotes 15 min (elegido)	Streaming (Kafka + Flink)
Integración de fuentes (API, webhooks, telemetría, agenda)	4.000	5.500	9.000
Lógica de detección y orquestación	1.500	2.500	9.000
Mensajería con cierre en un toque y tableros	1.500	2.000	2.000
Formación, protocolo de guardia y evaluación de impacto	1.000	2.000	2.000
Despliegue de plataforma y pruebas de carga	0	0	8.000
Total inversión	8.000	12.000	30.000
Guardia 24/7 del coordinador (complemento)	0 (solo horario de oficina)	3.600	3.600
Atención de alertas (≈ 30 h al mes $\times$ USD 20)	7.200	7.200	7.200
Mantenimiento de la regla y la ingesta	2.500	3.300	4.500
Infraestructura y mensajería	360	720	8.520

Concepto	Batch diario	Micro-lotes 15 min (elegido)	Streaming (Kafka + Flink)
Guardia de plataforma (soporte Kafka/Flink)	0	0	6.000
Total recurrente anual	10.060	14.820	29.820

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Business case a 3 años de la alternativa elegida

Tabla 75 Business case a 3 años de los micro-lotes de 15 minutos (USD)

Concepto	Año 0	Año 1	Año 2	Año 3
Inversión inicial	-12.000	-	-	-
Cobertura	-	50%	65%	70%
Episodios incrementales	-	20,5	26,6	28,6
Margen de consumibles	-	19.437	25.268	27.211
Horas reasignadas	-	1.680	1.680	1.680
Costes recurrentes	-	-14.820	-14.820	-14.820
Flujo neto	-12.000	6.297	12.128	14.071
Flujo acumulado	-12.000	-5.703	6.425	20.496

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Con la tasa de descuento del 12% usada desde el Capítulo 4 de la Parte I, el **VAN a 3 años es de USD 13.306** y la **TIR, de aproximadamente 61%**. El **ROI a 3 años** es $(76.956 - 56.460) / 56.460 = 36\%$, sobre un beneficio acumulado de USD 76.956 frente a un coste total de USD 56.460. El flujo acumulado es negativo al cierre del Año 1 (-5.703) y se cubre con 5,6 meses del flujo mensual del Año 2 ($12.128 / 12 = 1.011$): el **payback se alcanza a los 18 meses aproximadamente**.

Comparación de velocidades y análisis incremental

Tabla 76 Comparación de las tres velocidades y análisis incremental (USD)

Indicador	Batch diario	Micro-lotes 15 min	Streaming
Latencia media de la decisión	16,5 h	2,8 h	2,7 h
Recuperación por rapidez	19,1%	50,7%	51,2%
Episodios incrementales al año (régimen)	10,8	28,6	29,0
Inversión	8.000	12.000	30.000
Coste recurrente anual	10.060	14.820	29.820
VAN a 3 años (12%)	-6.709	13.306	-40.109
ROI a 3 años	-16%	36%	-35%

Indicador	Batch diario	Micro-lotes 15 min	Streaming
VAN incremental	-	+20.015 frente al diario	-53.415 frente a micro-lotes

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El análisis incremental resume la decisión. Pasar del batch diario a micro-lotes crea valor: unos 18 episodios más al año por USD 4.760 adicionales de coste recurrente, con un VAN incremental de +20.015. Pasar de micro-lotes a streaming gana 0,4 episodios al año -unos USD 290 de margen- a cambio de USD 15.000 más al año y USD 18.000 más de inversión: el VAN incremental es de -53.415. **Ir más rápido destruye valor**, porque el cuello de botella es la acción humana y no la detección.

Escenarios y punto de equilibrio

Tabla 77 Escenarios del business case de micro-lotes (USD)

Escenario	Supuesto	VAN (12%)	ROI 3 años	Payback
Pesimista	70% del margen de episodios estimado (menor cobertura o recuperación)	-3.754	-2%	> 36 meses
Base	Tabla anterior	13.306	36%	≈ 18 meses
Optimista	130% del margen de episodios	30.365	75%	≈ 12 meses
Fatiga de alertas	Precisión < 50%: la atención cae del 85% al 40%	-16.800	-31%	No se recupera
Punto de equilibrio	77% del margen estimado, ≈ 22 episodios incrementales al año	0	-	-

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Conviene decirlo con claridad: el margen de seguridad es estrecho. El caso es rentable mientras se capte al menos el 77% del margen estimado -unos 22 episodios incrementales al año en régimen-, y el escenario pesimista es ligeramente negativo. Las dos variables que más mueven el resultado no son tecnológicas: cada diez puntos de cobertura valen unos USD 9.300 de VAN, y una regla imprecisa que provoque fatiga de alertas hunde el caso hasta -16.800. Por eso el piloto mide primero precisión y cobertura, y por eso la respuesta a un mal resultado no sería acelerar, sino ampliar la red de hospitales que notifican o volver al resumen diario sin guardia.

Ejemplo demostrativo: una alerta trazada de principio a fin

Para mostrar cómo operará la vía rápida se traza un ejemplo ilustrativo: un simulacro en el ambiente de pruebas con un evento sintético, construido por el equipo con

cifras internamente consistentes y con el mismo formato de trazabilidad de los Capítulo 3 de la Parte I y Capítulos 3 y 4. No es una ejecución verificada; las primeras alertas reales serán las del modo sombra del 19/10/2026 y se documentarán con el protocolo de evidencia reproducible del Capítulo 4.

Tabla 78 Simulacro ilustrativo de una alerta, del evento al Data Warehouse (29/09/2026)

Etapas	Responsable o proceso	Hora	Detalle cuantificado
1. Notificación	Médico del hospital H-1207 (Babahoyo, Los Ríos), vía portal	09:12:40	Guillain-Barré (G61.0), 4 días desde el inicio de los síntomas, Hughes 4; consentimiento v2 registrado
2. Ingesta	API de ingesta	09:12:43	Esquema válido; latencia de los datos de 3 s; event_id 9f3c...e21
3. Detección	DAG rt_alerta_demanda_aguda, micro-lote de las 09:15	09:15:00-09:15:41	Ventana 09:00-09:15 por hora del evento: 1 notificación, 47 cambios de agenda, 210 latidos y 3 cambios de estado; regla CEP 6/6; centro GEMMA en Guayaquil a ≈ 1 h 30 de traslado, con equipo disponible y 2 plazas en 48 h; prioridad alta
4. Aviso	WhatsApp Business API	09:15:49	Mensaje al coordinador de guardia según la tabla de turnos; correo de respaldo
5. Acuse	Coordinador de guardia	09:23:10	7 min 21 s desde el aviso; sin escalado
6. Acción	Coordinador de guardia	09:31-10:58	Llamada al neurólogo tratante, confirmación de la plaza de las 14:00 con el centro y oferta formal enviada al hospital
7. Cierre	Coordinador de guardia	11:06:30	"Oferta aceptada - derivación en curso", en un toque
8. Convergencia	ETL diario	30/09, 02:00	Evento, alerta y cierre cargados al Data Warehouse; el episodio queda disponible para el DNC y como caso fechado para el Capítulo 4

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

El simulacro confirma la tesis del análisis de latencia. Desde la notificación hasta la oferta formal pasaron 1 h 45 min: la detección tardó 3 min 1 s, el aviso y el acuse 7 min 29 s y la acción 1 h 35 min, el 90% del total. Si la detección hubiera sido instantánea, el hospital habría recibido la oferta tres minutos antes, sin ningún efecto sobre su decisión.

Riesgos y protección de datos

Tabla 79 Riesgos del caso de uso en tiempo real

Riesgo	Prob.	Impacto	Responsable	Mitigación
Fatiga de alertas por falsas alarmas	Media	Alta	Data Analyst	Verificación previa de notificaciones de familiares; meta de precisión $\geq 70\%$; revisión semanal; si baja del 60%, se ajusta la regla antes de seguir
Baja cobertura: hospitales que no notifican	Alta	Alta	Gerente de Alianzas Clínicas	Portal de dos minutos; convenio con la Fundación; cobertura como KPI mensual
Guardia sin respuesta en noches y fines de semana	Media	Media	Gerente de Alianzas Clínicas	Escalado a los 15 y 45 min; suplente nominal; complemento de guardia
Caída silenciosa de la vía rápida	Baja	Media	Data Engineer	Alerta de silencio si el DAG no corre en 30 min; la Fundación conserva su canal telefónico
Presión comercial sobre la decisión clínica	Media	Alta	Dirección Comercial	La oferta es informativa; actúa personal clínico, no comercial; Canales Comerciales no recibe alertas
Uso indebido de datos de salud	Baja	Alta	Delegado de Protección de Datos	Seudonimización en origen, consentimiento explícito, retención de 90 días y evaluación de impacto
Uso de las alertas para evaluar a personas	Media	Media	Dirección Comercial y DPD	Los tiempos de acuse se usan solo en agregado; se informa a la plantilla del uso de sus datos

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

La vía rápida trata información de salud, que la Ley Orgánica de Protección de Datos Personales (LOPDP) considera una categoría especial. Un dato seudonimizado sigue siendo un dato personal, porque el hospital, la fecha y el diagnóstico podrían permitir la reidentificación; por eso el diseño aplica cinco controles. Primero, la base legal es el consentimiento explícito del paciente o su representante, recogido por la Fundación, y el evento sin consentimiento no entra. Segundo, la minimización: no viajan nombre, cédula ni historia clínica. Tercero, la retención: los eventos crudos se conservan 90 días y después solo se mantienen agregados por cantón y mes, como en el Capítulo 4. Cuarto, el acceso se limita con el RBAC del Capítulo 4 de la Parte I. Quinto, antes del modo sombra se realiza una evaluación de impacto en protección de datos y se informa a los coordinadores de que sus tiempos de acuse se registran y de que no se usarán para evaluarlos individualmente.

Conclusiones y hoja de ruta

Este capítulo completa los tres niveles de la analítica de MarketPulse Apheresis sin abandonar la disciplina de proporcionalidad de los capítulos anteriores. La decisión que pierde valor si se toma tarde es la oferta de una plaza GEMMA a un hospital sin aféresis que atiende a un paciente agudo elegible: su valor se reduce a la mitad cada 9,7 horas y hoy la empresa se entera tres semanas después. La latencia necesaria se dedujo descomponiéndola en detección, aviso y acción; como la acción humana ocupa el 88% del tiempo, unos micro-lotes de 15 minutos capturan el 82% del valor máximo, frente al 83% del streaming. Actúa un coordinador clínico de guardia con destinatario único, plazo, escalado y cierre en un toque, y la decisión clínica sigue siendo del médico tratante.

El business case es positivo pero ajustado: una inversión de USD 12.000 y costes recurrentes de USD 14.820 anuales, que incluyen guardia y tiempo de atención, producen un VAN de USD 13.306 al 12%, un ROI a 3 años del 36% y un payback de unos 18 meses, con el punto de equilibrio en 22 episodios incrementales al año. El streaming tendría un VAN de -40.109 y un VAN incremental de -53.415: ir más rápido destruye valor. La variable crítica no es la tecnología, sino la cobertura y la precisión, y eso es lo que el piloto escalonado mide primero, con la auditoría independiente de la Fundación.

CONCLUSIONES Y APLICACIONES

MarketPulse Apheresis nació para resolver un problema concreto de Apheresis Technology: la expansión del dispositivo GEMMA se decidía por relaciones comerciales e intuición, con información tardía, ciclos de priorización de seis meses y verificaciones de cobertura que se hacían sin un criterio objetivo. A lo largo de diez capítulos, el proyecto ha convertido esa decisión en un proceso basado en evidencia, construido por capas. Cada capítulo añade una pieza sin rehacer la anterior, y todas parten de la misma pregunta: dónde existe demanda clínica real que hoy no tiene acceso a la aféresis. Este apartado integra los resultados, explica cómo se aplican en la operación de la compañía y los contrasta con los criterios de escalamiento hacia Estados Unidos fijados desde el Capítulo 1 de la Parte I.

Resultados consolidados del proyecto

Tabla 80 Resultados consolidados por capítulo

Capítulo	Aporte al sistema	Resultado principal
I-1. Introducción al BI	Problema de negocio, KPI, DNC y SOM ponderado con AHP	Criterio objetivo de priorización; validación ilustrativa con precisión de 0,78 y sensibilidad de 0,70
I-2. Fuentes de datos	Modelo de negocio, Data Warehouse y gobierno del dato	Inventario de fuentes de Ecuador y Estados Unidos con propietario y calidad
I-3. Procesos ETL	Carga, reglas de calidad y trazabilidad	100% de registros trazables por fuente, fecha y versión
I-4. Visualización	Ocho dashboards en Power BI con RBAC y alertas	Consolidación de 32 a 10 h por ciclo; VCC de USD 960 a USD 640; VAN de USD 70.480
I-5. Organización	Unidad de BI y Datos de 4,5 posiciones, RACI y SLA	Costo recurrente de USD 169.000 frente a un beneficio en régimen de USD 250.000; payback de 14 meses
II-1. Datos no estructurados	Redes, Trends, boletines del MSP y correos con NLP	Cuatro casos de uso de monitoreo continuo con controles de privacidad
II-2. Arquitectura cloud	Selección de proveedor con matriz ponderada	AWS, de USD 150 a 250 al mes y sin costo de migración
II-3. Open source	Airflow, dbt, PostgreSQL y Superset	Stack proporcional al volumen, con un umbral explícito para escalar a Kafka y Spark
II-4. Inteligencia artificial	Regresión logística a 90 días, validada solo con etiquetas cerradas	Precisión del Top-15 de 0,47 a 0,67; VAN de USD 38.477; payback de 14 meses
II-5. Tiempo real	Alerta de demanda aguda en micro-lotes de 15 minutos	Oferta de plaza en menos de 3 h; VAN de USD 13.306; el streaming destruiría valor

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Los business case de los Capítulos 4 y 5 de la Parte I y Capítulos 4 y 5 de la Parte II no deben sumarse. Cada uno evalúa una decisión distinta sobre una base que en parte se comparte: el Capítulo 5 de la Parte I, por ejemplo, ya incorpora en régimen los beneficios del sistema del Capítulo 4 de la Parte I. Lo que sí muestran en conjunto es que cada capa se justificó por sí misma, frente a su propia alternativa, y que en ningún caso se adoptó una tecnología antes de demostrar la decisión que mejora.

El proyecto responde así a su problema de origen con tres niveles de analítica que operan a velocidades distintas porque las decisiones también lo son. El nivel descriptivo explica dónde está hoy la demanda no cubierta, con cargas semanales. El predictivo anticipa dónde aparecerá en los próximos 90 días y decide cada mes qué cantones verificar. El operativo detecta, en minutos, el paciente agudo que puede atenderse en la red GEMMA antes de que el hospital opte por otro camino. Esta correspondencia entre la velocidad de cada decisión y la de su información es probablemente la principal lección metodológica del trabajo.

Aplicación práctica en Apheresis Technology

En la operación diaria, el sistema funciona como un ciclo integrado en el que cada nivel alimenta al siguiente:

Tabla 81 Ciclo operativo integrado de MarketPulse Apheresis

Ritmo	Qué ocurre	Quién decide	Decisión que respalda
Minutos (24/7)	Una alerta de paciente agudo elegible llega al coordinador de guardia	Gerencia Alianzas Clínicas	Ofrecer una plaza GEMMA al hospital tratante
Diario (02:00)	Eventos, alertas y cierres se integran al Data Warehouse	Unidad de BI y Datos	Mantener al día el DNC y las etiquetas del modelo
Semanal	El ETL actualiza las fuentes oficiales, el DNC y el SOM	Unidad de BI y Datos	Garantizar la calidad y la trazabilidad del dato
Mensual (día 3)	El modelo publica el Top-15 de cantones y la Gerencia lo valida	Dirección Comercial	Aprobar las 5 VCC del mes y la agenda comercial
Trimestral	Se reentrena el modelo con etiquetas cerradas y se auditan sesgos	Unidad de BI y Datos	Mantener la precisión y detectar deriva
Anual	Se revisan los pesos AHP, la estrategia de expansión y el presupuesto	Alta Dirección	Fijar las zonas objetivo y la inversión del año

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Para la compañía, el cambio es organizativo antes que tecnológico. La Dirección Comercial deja de aprobar zonas por relaciones o intuición y pasa a aprobar listas

con una probabilidad, un motivo explicable y un registro de cada descarte. Alianzas Clínicas deja de enterarse de los casos semanas después y pasa a actuar en horas. La Unidad de BI y Datos asume la responsabilidad de que el dato sea fiable, pero no toma las decisiones comerciales ni clínicas: el sistema recomienda y las personas deciden. Esta separación, mantenida desde la RACI del Capítulo 5 de la Parte I hasta la del Capítulo 5 de la Parte II, es la que permite usar datos de salud con responsabilidad y la que protege la decisión clínica de cualquier presión comercial.

El valor del proyecto tiene además un componente que no aparece en ningún VAN: la Fundación Guillain Barré Ecuador pasa de ser solo una fuente de datos a ser socia de un circuito que acerca el tratamiento a pacientes que hoy no lo reciben a tiempo. Eso responde directamente a la misión de la compañía de hacer accesible la aféresis en mercados donde la demanda existe pero la oferta es insuficiente.

Contraste con los criterios de escalamiento hacia Estados Unidos

Desde el Capítulo 1 de la Parte I, el proyecto estableció que el modelo se calibraría y validaría en Ecuador antes de extrapolarse al mercado estadounidense. Con honestidad, la situación a octubre de 2026 es la siguiente: el diseño está completo y es coherente de punta a punta, pero ninguna de las cifras de resultado es todavía un resultado real verificado. Las métricas del SOM, del modelo predictivo y de la alerta son ejercicios ilustrativos o estimaciones preliminares, y así se han presentado en cada capítulo. Los criterios de escalamiento se cumplen, por tanto, en su diseño, pero no todavía en su evidencia.

Tabla 82 Estado de los criterios de escalamiento hacia Estados Unidos

Criterio	Umbral fijado	Origen	Estado a octubre de 2026
Validación real del SOM	Dos ciclos completos de datos de la Fundación, con precisión y sensibilidad medidas	Cap. 1 y 4 (Parte I)	Pendiente: la validación actual es ilustrativa
Objetivos operativos	Consolidación ≤ 10 h, ciclo ≤ 2 meses y VCC $\leq$ USD 640, confirmados en dos ciclos reales	Cap. 4 (Parte I)	Parcialmente cumplido
Modelo predictivo	+12 pp de VCC con demanda confirmada frente al grupo de control	Cap. 4 (Parte II)	Piloto de octubre de 2026 a marzo de 2027; decisión en abril de 2027
Alerta en tiempo real	+12 pp de episodios captados, precisión $\geq 70\%$ y acuse en ≤ 15 min	Cap. 5 (Parte II)	Piloto de noviembre de 2026 a febrero de 2027; decisión en marzo de 2027
Organización	Unidad de BI y Datos formalizada y operando a régimen	Cap. 5 (Parte I)	Diseñada; en formalización

Criterio	Umbral fijado	Origen	Estado a octubre de 2026
Arquitectura	Volumen por debajo de 10.000 registros por ciclo y 5.000 menciones al mes	Cap. 3 (Parte II)	Cumplido en Ecuador; Estados Unidos lo superaría
Protección de datos	LODPD en Ecuador; HIPAA en Estados Unidos	Cap. 1, 4 y 5 (Parte II)	LODPD diseñado; HIPAA sin evaluar

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

La conclusión es clara: **el escalamiento hacia Estados Unidos no debe decidirse antes de abril de 2027**, cuando se disponga de los resultados de los dos pilotos con grupo de control y de al menos dos ciclos reales de validación del SOM. Decidirlo antes sería repetir, a mayor escala, el problema que el proyecto vino a resolver: invertir en expansión sin evidencia.

Cuando llegue ese momento, el diseño facilita la transferencia porque separa lo que se replica de lo que se adapta:

Tabla 83 Qué se transfiere y qué se adapta al escalar a Estados Unidos

Componente	Se transfiere sin cambios	Requiere adaptación
Lógica de decisión	DNC, SOM con AHP, análisis de sensibilidad y escenarios	Recalibración de pesos con datos estadounidenses
Unidad de análisis	Estructura de zona y nivel de detalle	Estado como unidad principal y condado como detalle (Cap. 1 (Parte I))
Fuentes	Arquitectura de ingesta y reglas de calidad	CDC, National Inpatient Sample, CMS Provider Data Catalog y Census Bureau en lugar de MSP e INEC
Modelo predictivo	Técnica, validación temporal y regla de etiqueta cerrada	Reentrenamiento propio; el retador LightGBM se reevalúa con más casos positivos
Tiempo real	Regla CEP, flujo de la alerta y cierre en un toque	Un socio de notificación equivalente a la Fundación y cumplimiento de HIPAA
Infraestructura	AWS, Airflow, dbt, PostgreSQL y Power BI	Volumen estimado de 10 a 15 veces el actual: activa la prueba de concepto con Kafka y Spark prevista en el Cap. 3 (Parte II)
Organización	Unidad de BI y Datos, RACI y SLA	Equipos comerciales regionales como nuevos usuarios de los dashboards

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Dos de estas adaptaciones merecen atención especial. La primera es técnica: el volumen estadounidense superaría el umbral fijado en el Capítulo 3 de la Parte II. Por eso la incorporación de Kafka y Spark deja de ser un sobredimensionamiento y pasa a ser una decisión que debe evaluarse, con la misma matriz ponderada usada

en ese capítulo. La segunda es de datos: el aporte diferenciador del proyecto en Ecuador es la Fundación Guillain Barré, una fuente de demanda real que en Estados Unidos no existe todavía para Apheresis Technology. Sin un socio equivalente, el SOM podrá calcularse con fuentes oficiales, pero el modelo predictivo y la alerta en tiempo real perderían su variable objetivo. Buscar ese socio debería ser la primera tarea comercial de la fase estadounidense, no la última.

Hoja de ruta integrada

Tabla 84 Hoja de ruta de MarketPulse Apheresis

Fase	Periodo	Hito y criterio de avance
Pilotos en Ecuador	Octubre de 2026 a marzo de 2027	Modo sombra y pilotos con grupo de control de los Capítulos 4 y 5 de la Parte II
Decisión de producción	Marzo y abril de 2027	Paso a producción solo si se cumplen los umbrales fijados antes de cada piloto
Consolidación en Ecuador	Segundo trimestre de 2027	Dos ciclos reales de validación del SOM; recálculo de todos los business case con datos reales
Preparación para Estados Unidos	Segundo semestre de 2027	Socio de notificación, evaluación HIPAA, carga de fuentes estadounidenses y prueba de concepto con Kafka y Spark
Piloto en Estados Unidos	2028	Tres a cinco estados priorizados por el SOM recalibrado, con el mismo diseño de piloto con control
Expansión progresiva	2028 a 2030	Avance hacia la meta de presencia en los 50 estados y réplica del modelo en Latinoamérica

Fuente: Elaborado por los autores

Fecha: Octubre, 2026

Conclusión final

MarketPulse Apheresis demuestra que un sistema de Business Intelligence aporta valor cuando cada componente responde a una decisión concreta, con un usuario que actúa y un beneficio medible frente a su coste. El proyecto no ha buscado la tecnología más avanzada, sino la más proporcional. Descartó Kafka y Spark mientras el volumen no los justificaba, eligió una regresión logística explicable frente a modelos más opacos y prefirió micro-lotes de 15 minutos frente al streaming porque la acción humana, y no la detección, es la que limita el valor. Esa misma disciplina es la que debe guiar el siguiente paso: la expansión hacia Estados Unidos será tan sólida como lo sea la evidencia que el piloto de Ecuador produzca en los próximos seis meses. El sistema está diseñado; lo que queda es demostrarlo con datos reales.

REFERENCIAS

Las siguientes referencias respaldan, en formato APA 7ª edición, las cifras de negocio, los fundamentos conceptuales, los benchmarks salariales y las herramientas citados en este trabajo.

Agency for Healthcare Research and Quality. (2024). HCUP National Inpatient Sample (NIS). Healthcare Cost and Utilization Project. <https://hcup-us.ahrq.gov/nisoverview.jsp>

Amazon Web Services. (2026). Amazon MSK pricing y Amazon Managed Service for Apache Flink pricing. <https://aws.amazon.com/msk/pricing/>

Amazon Web Services. (2026). AWS Pricing Calculator. <https://calculator.aws/>

Apache Software Foundation. (2026). Apache Airflow Documentation. <https://airflow.apache.org/docs/>

Apache Software Foundation. (2026). Apache Superset Documentation. <https://superset.apache.org/docs/>

Asamblea Nacional del Ecuador. (2021). Ley Orgánica de Protección de Datos Personales (LOPDP). Registro Oficial.

Centers for Disease Control and Prevention. (2024). Guillain-Barré syndrome. U.S. Department of Health and Human Services. <https://www.cdc.gov/guillain-barre-syndrome/index.html>

Chevret, S., Hughes, R. A. C., & Annane, D. (2017). Plasma exchange for Guillain-Barré syndrome. Cochrane Database of Systematic Reviews, 2017(2), CD001798. <https://doi.org/10.1002/14651858.CD001798.pub3>

Databricks Inc. (2026). Databricks Pricing. <https://www.databricks.com/product/pricing>

dbt Labs. (2026). dbt Core Documentation. <https://docs.getdbt.com/>

Google Cloud. (2026). Google Cloud Pricing Calculator. <https://cloud.google.com/products/calculator>

Google. (2026). Places API — Reseñas (Reviews). Google Maps Platform Documentation. <https://developers.google.com/maps/documentation/places/web-service/reviews>

- Google. (2026). Trends Help Center: Cómo se calculan las tendencias de búsqueda. <https://support.google.com/trends>
- Hackathorn, R. D. (1998). Web farming for the data warehouse. Morgan Kaufmann.
- Joyanes Aguilar, L. (2020). Inteligencia de negocios y analítica de datos. Alfaomega.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146-3154.
- Market Data Forecast. (2025). Apheresis market size, share, growth & analysis report, 2033. <https://www.marketdataforecast.com/market-reports/apheresis-market>
- Meta. (2026). WhatsApp Business Platform: Pricing. <https://developers.facebook.com/docs/whatsapp/pricing>
- Microsoft. (2026). Azure Pricing Calculator. <https://azure.microsoft.com/pricing/calculator/>
- 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.
- Pérez Galán, J. L. (2026). Aspectos legales y éticos del Big Data: La privacidad y la seguridad de la información [Material docente]. Westfield Business School.
- Pérez Galán, J. L. (2026). Herramientas de Big Data para analizar los datos en tiempo real [Material docente]. Westfield Business School.
- Pérez Galán, J. L. (2026). Inteligencia de negocios [Material docente]. Westfield Business School.
- Pérez Galán, J. L. (2026). Introducción al Business Intelligence [Material docente]. Westfield Business School.
- Pérez Galán, J. L. (2026). Introducción al Machine Learning [Material docente]. Westfield Business School.
- Pérez Galán, J. L. (2026). ¿Qué es cloud, cloud computing y cloud storage? [Material docente]. Westfield Business School.

- Saaty, T. L. (1980). The analytic hierarchy process: Planning, priority setting, resource allocation. McGraw-Hill.
- Sejvar, J. J., Baughman, A. L., Wise, M., & Morgan, O. W. (2011). Population incidence of Guillain-Barré syndrome: A systematic review and meta-analysis. *Neuroepidemiology*, 36(2), 123-133. <https://doi.org/10.1159/000324710>
- Snowflake Inc. (2026). Snowflake Pricing Guide. <https://www.snowflake.com/pricing/>
- Sueldo.ec. (2026). Cuánto gana un Analista de Datos en Ecuador. <https://www.sueldo.ec/blog/cuanto-gana-analista-datos-ecuador>
- Talently. (2026). Calculadora de salario: Data Analyst / BI en Ecuador. <https://talently.tech/herramientas/salario/developer/data-analyst-bi>
- U.S. Census Bureau. (2024). American Community Survey (ACS). <https://www.census.gov/programs-surveys/acs>