Maestría en

ADMINISTRACIÓN DE EMPRESAS

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

AUTORES:

Lic. Leslie Tamara Soria Muquis
Ing. Pablo Bolívar Mazón Pachacama
Ing. Mateo Fernando Ruiz Mosquera
Ing. Stefanie Andrea Aldás Romero
Econ. Jefferson Rodrigo Catota Sisalima

TUTOR: Mgs. José Luis Pérez Galán

Diseño e implementación de una solución de inteligencia de negocios y analítica predictiva para la gestión de costos, rentabilidad y toma de decisiones en Directech Cía. Ltda.

REINVENTEMOS

PROYECTO DE TITULACIÓN – FIN DE MÁSTER

MAESTRÍA EN ADMINISTRACIÓN DE EMPRESA ENL

Diseño e implementación de una solución de inteligencia de negocios y analítica predictiva para la gestión de costos, rentabilidad y toma de decisiones en Directech Cía. Ltda.

Por

Lic. Leslie Tamara Soria Muquis

Ing. Pablo Bolívar Mazón Pachacama

Ing. Mateo Fernando Ruiz Mosquera

Ing. Stefanie Andrea Aldás Romero

Econ. Jefferson Rodrigo Catota Sisalima

Octubre 2026

Aprobado



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

Yo, Cristian Javier Melo González e Ignacio Maroto, declaramos que, personalmente conocemos que los graduandos: Lic. Leslie Tamara Soria Muquis, Ing. Pablo Bolívar Mazón Pachacama, Ing. Mateo Fernando Ruiz Mosquera, Ing. Stefanie Andrea Aldás Romero y Econ. Jefferson Rodrigo Catota Sisalima, son los autores exclusivos de la presente investigación y que esta es original, auténtica y personal de ellos.



Cristian Melo
Coordinador MBA UIDE



Ignacio Maroto
Provost WBS

Autoría del Trabajo de Titulación

Nosotros, Lic. Leslie Tamara Soria Muquis, Ing. Pablo Bolívar Mazón Pachacama, Ing. Mateo Fernando Ruiz Mosquera, Ing. Stefanie Andrea Aldás Romero y Econ. Jefferson Rodrigo Catota Sisalima, declaramos bajo juramento que el trabajo de titulación titulado «Diseño e implementación de una solución de inteligencia de negocios y analítica predictiva para la gestión de costos, rentabilidad y toma de decisiones en Directech Cía. Ltda.» 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.



Leslie Tamara Soria Muquis
Correo electrónico:
lesoriamu@uide.edu.ec



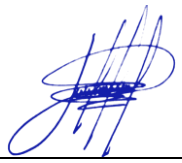
Pablo Bolívar Mazón Pachacama
Correo electrónico:
pamazonpa@uide.edu.ec



Mateo Fernando Ruiz Mosquera
Correo electrónico:
maruizmo@uide.edu.ec



Stefanie Andrea Aldás Romero
Correo electrónico: staldasro@uide.edu.ec



Jefferson Rodrigo Catota Sisalima
Correo electrónico:
jecatotasi@uide.edu.ec

Autorización de Derechos de Propiedad Intelectual

Nosotros, Lic. Leslie Tamara Soria Muquis, Ing. Pablo Bolívar Mazón Pachacama, Ing. Mateo Fernando Ruiz Mosquera, Ing. Stefanie Andrea Aldás Romero y Econ. Jefferson Rodrigo Catota Sisalima, en calidad de autores del trabajo de investigación titulado «Diseño e implementación de una solución de inteligencia de negocios y analítica predictiva para la gestión de costos, rentabilidad y toma de decisiones en Directech Cía. Ltda.», autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden se rigen por lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, octubre de 2026

Leslie Tamara Soria Muquis
Correo electrónico:
lesoriamu@uide.edu.ec

Pablo Bolívar Mazón Pachacama
Correo electrónico:
pamazonpa@uide.edu.ec

Mateo Fernando Ruiz Mosquera
Correo electrónico:
maruizmo@uide.edu.ec

Stefanie Andrea Aldás Romero
Correo electrónico: staldasro@uide.edu.ec

Jefferson Rodrigo Catota Sisalima
Correo electrónico:
jecatotasi@uide.edu.ec

Dedicatorias y Agradecimientos

Agradezco a Dios por estar presente en mi vida y permitirme llegar hasta este momento tan importante de mi formación profesional. A mis padres, por su amor incondicional, apoyo constante y motivación a lo largo de este camino.

A mi novio, Andrés, y a mi querida Dakota, por su amor, compañía y apoyo, que hicieron más llevadero este proceso y llenaron mis días de alegría. Asimismo, agradezco a mis compañeros de equipo por su compromiso, colaboración y esfuerzo conjunto en la realización de este trabajo. A todos ustedes, gracias por formar parte de este logro tan significativo en mi vida.

Leslie Soria

A Dios Padre, que nunca me ha dejado solo en cada paso dado; a mi familia, que me ha dado el aliento necesario para lograr mis objetivos, y a mis amigos, integrantes del grupo, cuyo esfuerzo y compromiso fueron clave para la culminación y el éxito del presente proyecto.

Pablo Mazón

Agradezco a mi familia por el apoyo, la empatía y, sobre todo, por la paciencia que mostraron a lo largo de este camino; sin duda, este logro también es de ellos. Asimismo, agradezco a mis compañeros por su inmensa colaboración y dedicación. Gracias a todos los que me apoyaron para poder cumplir esta meta.

Mateo Ruiz

Agradezco a Dios por guiarme y darme la fortaleza, sabiduría y perseverancia para alcanzar esta importante meta. A mis padres, Glen y Victoria, por ser mi inspiración de crecimiento, superación y éxito, y por enseñarme que el valor más grande de la vida es la familia. A mis cuatro hermanos, por su amor y apoyo incondicional en cada etapa de mi vida, brindándome siempre la certeza de que puedo contar con ellos.

Finalmente, dedico este logro a mi amada hija, Arianna, mi mayor motivación y motor de mejora continua, quien me inspira a ser mejor cada día y a demostrarle, a través del ejemplo, el amor y la valentía, que con esfuerzo, perseverancia y determinación los sueños se pueden alcanzar.

Stefanie Aldás

Agradezco profundamente a Dios, quien ha sido mi guía en medio de la incertidumbre y me ha dado fuerzas para continuar. A mis padres, porque sin ellos este sueño no habría sido posible. Todo lo que he alcanzado se lo debo a ellos.

A mis compañeros, con quienes compartí noches de desvelo, sacrificios e incluso diferencias; superamos cada dificultad hasta cumplir esta meta. A mi novia, Mónica: gracias por abrazarme en los momentos difíciles, creer en mí y recordarme que con tu amor lo puedo todo.

Y, de manera especial, a mis dos ángeles, Manuel y Consuelo. Su ejemplo de esfuerzo, sacrificio y devoción a Jesús sigue guiando mis pasos. Sé que, cuando miro al cielo, me sonríen y me bendicen.

Jefferson Catota

ÍNDICE DEL DOCUMENTO

ÍNDICE DEL DOCUMENTO 8
ÍNDICE DE FIGURAS Y TABLAS 12
INTRODUCCIÓN Y OBJETO DEL PROYECTO 15
PARTE I: DIAGNÓSTICO Y DISEÑO DEL SISTEMA DE BUSINESS INTELLIGENCE PARA DIRECTECH 17
CAPÍTULO 1: Diagnóstico inicial, necesidades de información y KPI de Directech 18
1.1. Presentación de la empresa 18
1.2. Identificación de los objetivos iniciales del proyecto: definición inicial del problema. 18
1.3. Revisión de las áreas y personas que demandan la información y el uso que hacen actualmente de la misma 19
1.4. Objetivos de cada una de las áreas con la información actual y necesidades detectadas 20
1.5. Definición de los KPI relevantes 22
1.6. Conclusiones del capítulo 24
CAPÍTULO 2: Fuentes de información, sistemas, almacenamiento y reportería actual 26
2.1. Origen de la información y recurrencia de carga: interna y externa 26
2.2. Sistemas y destino de la carga 27
2.3. Almacenamiento en data warehouse y datamarts 28
2.4. Carencias y problemas en la carga de datos 29
2.5. Informes que se generan actualmente, necesidades detectadas y no cubiertas 30
2.6. Conclusiones del capítulo 31
CAPÍTULO 3: Estructura de la información y procesos de carga de datos 33
3.1. Procesos de carga de datos (ETL, data lake) 33
3.2. Información y calidad de la información disponible 34
3.3. Horarios de carga y accesibilidad 35
3.4. Variables disponibles y acceso al Data Warehouse 36
3.5. Sistemas implicados en los procesos 37
3.6. Recursos de IT o externos empleados 37
3.7. Conclusiones del capítulo 38
CAPÍTULO 4: Arquitectura de Business Intelligence y perfiles de usuario 39

4.1. Elección de la arquitectura de Business Intelligence 39
4.2. Perfil de los usuarios de la información 43
4.3. Conclusiones del capítulo 45
CAPÍTULO 5: Estructura del departamento de Business Intelligence 46
5.1. Identificación de la situación actual del departamento 46
5.2. Definición de las carencias en formación y recursos 47
5.3. Coordinación de los flujos entre otras áreas 47
5.4. Organigrama, dependencias de otras áreas y solución de problemas 48
5.5. Recursos externos destinados: subcontratación 50
5.6. Business case de costes y beneficios 51
5.7. Conclusiones del capítulo 53
PARTE II: CONSTRUCCIÓN, ARQUITECTURA CLOUD E IMPLEMENTACIÓN DEL MODELO DE BI 55
CAPÍTULO 1: Modelo de Business Intelligence integrado EDUTECH – DIRECTECH e impacto potencial de Big Data 56
1.1. Transición del Bloque 1 al Bloque 2 56
1.2. Objetivos del modelo integrado 56
1.3. Selección e identificación de fuentes de datos 57
1.4. Relación de los datos 58
1.5. Casos de uso del modelo de Business Intelligence 59
1.6. Impacto de Big Data en el proyecto 60
1.7. Rol de PostgreSQL dentro de la solución 61
1.8. Evidencias técnicas de implementación 62
1.9. Conclusiones del capítulo 71
CAPÍTULO 2: Arquitectura Cloud y Landing Zone para el modelo de BI 72
2.1. Objetivo y alcance 72
2.2. Análisis de alternativas de proveedores Cloud 72
2.3. Selección de la alternativa de proveedor 73
2.4. Definición del modelo de Cloud Computing 74
2.5. Flujo de integración de la arquitectura Big Data con el Data Warehouse 75
2.6. Conclusiones del capítulo 77

CAPÍTULO 3: Implementación práctica, trazabilidad y gobierno operativo de la arquitectura BI 78
3.1. Objetivo del capítulo 78
3.2. Ajustes derivados de la retroalimentación del Capítulo 2 78
3.3. Arquitectura consolidada de la solución 79
3.4. Modelo operativo de ingesta y ETL 80
3.5. Implementación práctica realizada en Microsoft Fabric 82
3.6. Diseño de la siguiente etapa: Dataflow Gen2 y Power Query 85
3.7. Condiciones para una futura evolución hacia Big Data 86
3.8. Conclusiones del capítulo 86
CAPÍTULO 4: Inteligencia artificial integrada al Data Warehouse: riesgo de margen y sobrecostos 88
4.1. Objetivo y alcance 88
4.2. Objetivo de aplicación del modelo de data science y machine learning 89
4.3. Fuentes de datos utilizadas 91
4.4. Área de aplicación y departamentos de negocio implicados 95
4.5. Identificación de la mejora esperada 99
4.6. Business case de la implementación de IA 99
4.7. Riesgos y uso responsable de la IA 102
4.8. Plan de implementación 103
4.9. Conclusiones del capítulo 103
CAPÍTULO 5: Caso de uso Near Real Time para el control de costos y rentabilidad de proyectos 105
5.1. Identificación y definición del objetivo de aplicación del modelo Near Real Time 105
5.2. Fuentes de datos utilizadas 107
5.3. Área de aplicación y departamentos de negocio implicados 108
5.4. Identificación de la mejora esperada 109
5.5. Cuantificación de la mejora mediante business case 110
5.6. Alcance y condiciones de implementación 112
5.7. Conclusiones del capítulo 113
CONCLUSIONES Y APLICACIONES 114

REFERENCIAS 116
ANEXOS 119
Anexo A1. Código del notebook de entrenamiento y scoring en Microsoft Fabric 119
Anexo A2. Medidas DAX del tablero «Riesgo de margen» 121
Anexo A3. Evidencias técnicas previstas para la fase de implementación 121
Anexo A4. Esquema propuesto de almacenamiento de predicciones de IA en PostgreSQL 122

ÍNDICE DE FIGURAS Y TABLAS

Índice de figuras

Figura 1 *Logotipo de Directech Cía. Ltda.* 18
Figura 2 *Organigrama empresarial de Directech Cía. Ltda.* 26
Figura 3 *Modelo conceptual propuesto para BI DIRECTECH (fuentes → integración → capacidades)* 34
Figura 4 *Esquema estrella propuesto para el Data Warehouse de BI DIRECTECH* 41
Figura 5 *Estructura organizativa propuesta para la función de BI en Directech* 49
Figura 6 *E1. Creación de la base de datos en PostgreSQL / DBeaver* 63
Figura 7 *E2. Esquema BI y tablas creadas en PostgreSQL* 64
Figura 8 *E3. Diagrama entidad-relación del modelo* 65
Figura 9 *E4. Validación de la carga de datos* 66
Figura 10 *E5. Consulta de validación en SQL* 67
Figura 11 *E6. Conexión de Power BI con PostgreSQL* 68
Figura 12 *E7. Modelo relacional en Power BI* 69
Figura 13 *E8. Tablero «Resumen Ejecutivo»* 70
Figura 14 *E9. Tablero «Finanzas – Estado de Resultados»* 70
Figura 15 *Arquitectura propuesta de integración Cloud + BI* 76
Figura 16 *Arquitectura práctica consolidada para la prueba de concepto* 80
Figura 17 *Lakehouse LH_EDUTECH_DIRECTECH creado dentro del workspace del proyecto* 82
Figura 18 *Estructura inicial de Landing y RAW en el Lakehouse* 83
Figura 19 *Ejecución correcta del pipeline PL_01_LANDING_A_RAW* 84
Figura 20 *Archivos CSV preservados en la capa RAW después de la copia binaria* 85
Figura 21 *Diagrama de aplicación del modelo MRM en las etapas de una operación* 90
Figura 22 *Arquitectura de implementación de IA en Microsoft Fabric y Power BI* 94
Figura 23 *Matriz de priorización de operaciones según riesgo estimado y monto* 97
Figura 24 *Prototipo del tablero «Riesgo de margen» en Power BI* 98
Figura 25 *Beneficio anual en régimen frente al costo recurrente, y flujo neto acumulado a 36 meses por escenario* 102
Figura 26 *Arquitectura conceptual del caso de uso Near Real Time de BI DIRECTECH* 109

Índice de tablas

Tabla 1 *Información que demandan las áreas y uso actual* 19
Tabla 2 *Indicadores clave de desempeño (KPI)* 23
Tabla 3 *Identificación de los GAP por KPI* 24
Tabla 4 *Fuentes de datos internas de Directech* 26
Tabla 5 *Sistemas de información de Directech* 27
Tabla 6 *Necesidades de transformación de la información en Directech* 29

Tabla 7 *Diagnóstico de carencias y problemas en la gestión de datos* 29
Tabla 8 *Estado actual de la reportería en Directech* 30
Tabla 9 *Dimensiones de calidad de la información en Directech* 34
Tabla 10 *Tipos de datos disponibles en Directech* 35
Tabla 11 *Frecuencia de carga y accesibilidad por área* 35
Tabla 12 *Opciones de arquitectura de BI evaluadas* 39
Tabla 13 *Criterios de selección de arquitectura para Directech* 40
Tabla 14 *Componentes de la arquitectura seleccionada* 40
Tabla 15 *Herramientas de análisis y visualización evaluadas* 42
Tabla 16 *Perfiles de usuario identificados en Directech* 43
Tabla 17 *Necesidades de información por perfil de usuario* 44
Tabla 18 *Nivel de acceso y herramienta de consumo por perfil* 44
Tabla 19 *Roles y responsabilidades actuales relacionados con la información* 46
Tabla 20 *Carencias en formación y recursos identificadas* 47
Tabla 21 *Matriz de flujos y coordinación entre áreas* 47
Tabla 22 *Roles propuestos para la función de BI en Directech* 49
Tabla 23 *Mecanismo de escalamiento y solución de problemas* 50
Tabla 24 *Evaluación make vs. buy de los servicios requeridos* 51
Tabla 25 *Presupuesto estimado de implementación (año 1)* 51
Tabla 26 *Ahorro estimado en consolidación manual de reportes* 52
Tabla 27 *Beneficio estimado por reducción de sobrecostos* 53
Tabla 28 *Fuentes de datos del modelo integrado EDUTECH – DIRECTECH* 57
Tabla 29 *Capas de la arquitectura de integración* 58
Tabla 30 *Tablas de hechos del modelo integrado* 58
Tabla 31 *Preguntas gerenciales y relaciones del modelo* 59
Tabla 32 *Casos de uso del modelo de BI* 59
Tabla 33 *Impacto potencial de Big Data por área* 61
Tabla 34 *Niveles de madurez analítica aplicados al modelo* 61
Tabla 35 *Evidencias técnicas de la implementación* 62
Tabla 36 *Conteos esperados después de la carga* 66
Tabla 37 *Comparación de proveedores de nube pública* 72
Tabla 38 *Criterios de selección del proveedor Cloud* 74
Tabla 39 *Arquitectura tecnológica propuesta en Azure* 74
Tabla 40 *Modelos de servicio Cloud y su aplicación al proyecto* 74
Tabla 41 *Matriz ponderada de selección tecnológica para la prueba de concepto* 78
Tabla 42 *Resultado ponderado de la selección tecnológica* 79
Tabla 43 *Separación de responsabilidades tecnológicas en la arquitectura implementada* 80
Tabla 44 *Política de carga y frecuencia de actualización* 81
Tabla 45 *Responsabilidades operativas del flujo de información* 81
Tabla 46 *Secuencia de transformación prevista en Dataflow Gen2* 85
Tabla 47 *Condiciones que justificarían evaluar tecnologías de Big Data* 86

Tabla 48 *Definición del modelo predictivo de riesgo de margen* 89
Tabla 49 *Criterios de aceptación del modelo* 91
Tabla 50 *Variables del modelo disponibles en las tablas del proyecto* 92
Tabla 51 *Estructura del nuevo bloque de información hist_proyectos* 92
Tabla 52 *Pasos de implementación en Microsoft Fabric y Power BI* 94
Tabla 53 *Departamentos implicados y decisiones que habilita el modelo* 96
Tabla 54 *Situación actual frente a la situación con el modelo* 99
Tabla 55 *Impacto esperado sobre los KPI del proyecto* 99
Tabla 56 *Supuestos del business case* 100
Tabla 57 *Costos incrementales de la implementación de IA* 100
Tabla 58 *Beneficios anuales estimados por escenario* 101
Tabla 59 *Resultados del business case por escenario* 101
Tabla 60 *Riesgos de la implementación y medidas de mitigación* 102
Tabla 61 *Plan de implementación del modelo MRM* 103
Tabla 62 *Indicadores recalculados en cada ciclo Near Real Time* 106
Tabla 63 *Reglas de alerta del caso de uso Near Real Time* 107
Tabla 64 *Fuentes de datos y mecanismo de captura del caso de uso* 107
Tabla 65 *Departamentos implicados, alertas y tiempos de respuesta* 108
Tabla 66 *Mejora esperada del caso de uso Near Real Time* 109
Tabla 67 *Supuestos del business case Near Real Time* 110
Tabla 68 *Costos incrementales del caso de uso Near Real Time* 111
Tabla 69 *Beneficios anuales estimados del caso Near Real Time* 111
Tabla 70 *Resultados del business case Near Real Time por escenario* 112
Tabla 71 *Evidencias técnicas previstas para la implementación del modelo* 121

INTRODUCCIÓN Y OBJETO DEL PROYECTO

En un entorno empresarial cada vez más dinámico y competitivo, la capacidad de tomar decisiones estratégicas basadas en datos constituye una ventaja diferencial determinante (Mucci, 2026). El Business Intelligence (BI) emerge como la disciplina que permite transformar datos operativos y financieros dispersos en conocimiento accionable, proporcionando a las organizaciones una visión integral, oportuna y confiable de su desempeño (IBM, 2026b; Laudon & Laudon, 2020).

Directech Cía. Ltda. es una empresa ecuatoriana con más de veinte años de trayectoria (opera desde noviembre de 2005) en la provisión, desarrollo y gestión de soluciones tecnológicas integrales. Con presencia en once ciudades del país, una cartera de 15 clientes activos en los sectores público y privado y un volumen de ingresos aproximado de USD 900.000 anuales, la empresa ejecuta entre 5 y 10 proyectos tecnológicos por año en líneas de negocio que abarcan infraestructura y equipamiento tecnológico, redes y telecomunicaciones, seguridad informática, soporte técnico especializado y desarrollo e integración de soluciones (Directech Cía. Ltda., 2026).

A pesar de su consolidada operación, Directech enfrenta un desafío crítico: la ausencia de una herramienta analítica integrada que permita conocer con precisión el margen de ganancia real de cada proyecto, producto o servicio. La información financiera y operativa se encuentra distribuida en fuentes heterogéneas —sistema contable, reportes de proyectos, órdenes de trabajo, registros de ventas— sin un mecanismo de consolidación automatizado que permita visualizarla de manera unificada y en tiempo real. Esta brecha impide identificar con certeza qué proyectos son verdaderamente rentables, dónde se generan sobrecostos y cómo evoluciona la rentabilidad global de la empresa ante la variabilidad de sus costos.

El presente proyecto, titulado «Diseño e implementación de una solución de inteligencia de negocios y analítica predictiva para la gestión de costos, rentabilidad y toma de decisiones en Directech Cía. Ltda.» (en adelante, BI DIRECTECH), plantea un sistema de Business Intelligence con un Data Warehouse dimensional como repositorio analítico central, procesos ETL para consolidar información, tableros Power BI y una extensión de inteligencia artificial orientada al riesgo de sobrecostos y erosión de margen. La propuesta permite un análisis histórico y, en fases posteriores, alertas Near Real Time para apoyar decisiones preventivas. Se distingue expresamente entre las pruebas de concepto académicas realizadas con datos sintéticos y la implementación operativa futura en Directech Cía. Ltda.

Objetivo general: diseñar y validar un modelo de Business Intelligence para Directech Cía. Ltda. que centralice información comercial, financiera y operativa en un Data Warehouse, visualice los márgenes de ganancia y las desviaciones de costos e incorpore un componente de inteligencia artificial que permita estimar el riesgo de sobrecosto y de baja rentabilidad, para fortalecer decisiones basadas en datos.

Los objetivos específicos del proyecto son:

- Diagnosticar las necesidades de información de cada área de la empresa y definir los KPI de rentabilidad y costos, junto con sus brechas actuales.

- Identificar las fuentes de datos, los sistemas, los procesos de carga y la calidad de la información disponible.
- Seleccionar la arquitectura de BI, el modelo dimensional y la herramienta de visualización, y caracterizar los perfiles de usuario.
- Proponer la estructura organizativa de la función de BI y sustentar su viabilidad mediante un business case.
- Construir y validar un modelo BI y una prueba de concepto de ingesta en Microsoft Fabric; definir PostgreSQL como Data Warehouse para hechos, dimensiones y predicciones de IA; y diseñar un modelo de riesgo de margen/sobrecostos con visualización en Power BI y alertas Near Real Time.

El trabajo se organiza en dos partes. La Parte I presenta el diagnóstico y diseño de BI para Directech: necesidades y KPI (Capítulo 1), fuentes y reportería (Capítulo 2), procesos y calidad (Capítulo 3), arquitectura y usuarios (Capítulo 4), y organización y business case (Capítulo 5). La Parte I se centra en Directech porque en ella se originan las cotizaciones, los costos y la ejecución de los proyectos; la Parte II incorpora a EDUTECH, empresa vinculada a través de la cual se presentan las ofertas y se factura, para medir el margen completo de cada proyecto. La Parte II desarrolla el modelo integrado EDUTECH–DIRECTECH con datos académicos sintéticos (Capítulo 1), la arquitectura Cloud y la Landing Zone (Capítulo 2), la prueba de concepto en Fabric y PostgreSQL (Capítulo 3), la extensión con IA para riesgo de margen y sobrecostos (Capítulo 4), y el modelo de vigilancia Near Real Time (Capítulo 5). Fabric/OneLake soporta ingesta y procesamiento; PostgreSQL es el Data Warehouse analítico elegido y Power BI es la capa de consumo. Se distingue entre componentes implementados, prototipos verificados con pruebas y capacidades propuestas para la operación futura.

Metodológicamente, el proyecto sigue la guía del programa y el caso de referencia de Grupo Bancolombia (Pérez Galán, 2026a, 2026b, 2026c, 2026d, 2026e), adaptado a la escala de una PYME, y se apoya en los principios del modelado dimensional (Kimball & Ross, 2013) y del gobierno de datos (DAMA International, 2017).

El equipo de trabajo cuenta con vinculación directa con Directech Cía. Ltda. a través de una de sus integrantes, quien se desempeña como Coordinadora de Proyectos en la empresa, lo que garantiza un nivel de conocimiento organizacional estimado en 9/10 y el acceso a información operativa, financiera y comercial relevante, dentro de los lineamientos de confidencialidad establecidos por la organización.

PARTE I: DIAGNÓSTICO Y DISEÑO DEL SISTEMA DE BUSINESS INTELLIGENCE PARA DIRECTECH

CAPÍTULO 1: Diagnóstico inicial, necesidades de información y KPI de Directech

1.1. Presentación de la empresa

Nombre de la empresa: Directech Cía. Ltda.

Figura 1

Logotipo de Directech Cía. Ltda.



Nota. Tomado de Directech Cía. Ltda. (2026).

DIRECTECH es una empresa especializada en desarrollo de proyectos tecnológicos y la creación de ecosistemas digitales innovadores, orientados a satisfacer las necesidades específicas de cada cliente. La propuesta de valor se basa en la integración de soluciones tecnológicas avanzadas que combinan hardware, software y servicios de consultoría, con el objetivo de impulsar la transformación digital de empresas e instituciones en distintos sectores.

DIRECTECH se ha posicionado como un aliado estratégico para organizaciones que buscan modernizar sus procesos, optimizar recursos y mantenerse competitivas en un entorno cada vez más digitalizado. DIRECTECH no solo vende tecnología, sino que diseña e implementa soluciones a medida, asegurando una implementación eficiente y acompañamiento continuo en todas las etapas del proyecto: desde la planificación, pasando por la ejecución, hasta el soporte postventa.

Misión: Proveer, desarrollar y gestionar soluciones tecnológicas integrales a nivel nacional, ofreciendo servicios eficientes, confiables y adaptados a las necesidades de nuestros clientes. Directech Cía. Ltda. orienta sus operaciones hacia la innovación, la calidad del servicio y el uso estratégico de la información para optimizar sus procesos y generar valor.

1.2. Identificación de los objetivos iniciales del proyecto: definición inicial del problema

El objetivo inicial del proyecto consiste en diseñar e implementar un sistema de Business Intelligence para Directech Cía. Ltda. que permita visualizar de manera clara y en tiempo real el margen de ganancia real de sus proyectos, productos o servicios, considerando la variabilidad de los costos. Esto facilitará la toma de decisiones estratégicas basadas en datos, optimizando la rentabilidad y la eficiencia operativa.

Actualmente, Directech Cía. Ltda. no cuenta con una herramienta analítica integrada que permita consolidar de forma automatizada la información relacionada con costos variables, ventas y márgenes de ganancia. Esta situación limita la visibilidad sobre el margen real de los proyectos, productos o servicios, dificultando el ajuste oportuno de precios, la identificación de gastos innecesarios, la optimización de procesos y la toma de decisiones estratégicas con base en información confiable y actualizada.

1.3. Revisión de las áreas y personas que demandan la información y el uso que hacen actualmente de la misma

En Directech Cía. Ltda., la información relacionada con ventas, costos, proyectos, recursos técnicos, proveedores y márgenes de ganancia es utilizada por diferentes áreas de la organización. Cada área demanda información específica de acuerdo con sus responsabilidades, ya sea para la toma de decisiones estratégicas, la elaboración de propuestas comerciales, el control de proyectos, la ejecución técnica o el análisis financiero. En este sentido, el Business Intelligence permite recopilar, gestionar y analizar datos organizacionales para generar información útil que apoye la toma de decisiones empresariales (IBM, 2026b).

Actualmente, esta información se encuentra distribuida en diferentes fuentes, tales como sistema contable, informes financieros, cotizaciones, presupuestos, órdenes de trabajo, reportes de proyectos, registros de ventas y documentos internos de seguimiento. Esta situación genera la necesidad de consolidar los datos en una herramienta de Business Intelligence que permita visualizar de forma clara y oportuna la rentabilidad real de los proyectos, productos o servicios.

Tabla 1

Información que demandan las áreas y uso actual

Área / Persona	Información que demanda	Uso actual de la información
Gerencia General	Información consolidada sobre ingresos, costos, utilidad, margen bruto, margen neto, rentabilidad por proyecto y rentabilidad por línea de negocio.	Utiliza la información para tomar decisiones estratégicas, aprobar proyectos, evaluar resultados económicos, definir prioridades comerciales, controlar la rentabilidad y orientar el crecimiento de la empresa.
Área de Ventas	Información sobre precios, cotizaciones, historial de clientes, condiciones comerciales, descuentos aplicados y rentabilidad histórica por tipo de proyecto o cliente.	Utiliza la información para elaborar propuestas comerciales, negociar con clientes, definir precios de venta, estimar márgenes y evaluar el ámbito económico de nuevos negocios.
Área de Proyectos	Información sobre presupuestos aprobados, costos reales, avance de proyectos, recursos utilizados, tiempos de ejecución, proveedores y desviaciones presupuestarias.	Utiliza la información para planificar, ejecutar, controlar y cerrar proyectos tecnológicos, verificando el cumplimiento de plazos, recursos y costos definidos inicialmente.
Área Técnica	Información sobre requerimientos técnicos, equipos, licencias, materiales, horas de trabajo, soporte, instalaciones y mantenimientos.	Utiliza la información para ejecutar las soluciones tecnológicas, estimar recursos, reportar novedades, controlar tiempos de implementación y validar los insumos técnicos requeridos.
Área Administrativa y Financiera	Información sobre facturación, pagos a proveedores, costos directos, gastos operativos, compras, flujo de caja, registros contables y resultados financieros.	Utiliza la información para registrar ingresos y egresos, controlar costos, elaborar reportes financieros, gestionar pagos y apoyar el análisis económico de los proyectos.

Nota. Elaboración propia.

1.3.1. Uso actual de la información

Actualmente, la información es utilizada principalmente para fines de control operativo, seguimiento financiero y toma de decisiones gerenciales. Sin embargo, el uso de la información presenta limitaciones debido a que los datos no se encuentran completamente integrados en una sola herramienta de análisis. La toma de decisiones basada en datos se fundamenta en utilizar información y análisis, en lugar de depender únicamente de la intuición, para orientar las decisiones empresariales (Mucci, 2026).

La Gerencia requiere información consolidada para conocer la rentabilidad real de la empresa y de sus proyectos; sin embargo, depende de reportes generados desde distintas fuentes. El Área de Ventas utiliza información histórica y presupuestaria para preparar cotizaciones y negociar con clientes, pero no siempre cuenta con una visualización clara del margen real obtenido en proyectos anteriores. El Área de Proyectos realiza seguimiento a costos, tiempos y recursos, pero la comparación entre lo presupuestado y lo ejecutado puede requerir revisiones manuales. El Área Técnica registra o comunica información relacionada con recursos utilizados, horas de trabajo, materiales y soporte, pero dicha información no siempre se vincula directamente con el análisis financiero del proyecto. Finalmente, el Área Administrativa y Financiera dispone de registros contables y financieros, pero requiere integrar dichos datos con la información operativa y comercial para calcular de manera más precisa los márgenes reales.

Por lo tanto, se identifica que la información existe dentro de la organización, pero se encuentra dispersa entre diferentes áreas y fuentes. Esto limita la capacidad de análisis oportuno y dificulta que la empresa pueda conocer de manera rápida qué proyectos son más rentables, cuáles presentan sobrecostos, qué clientes generan mayor margen y qué líneas de negocio requieren ajustes.

1.3.2. Brecha identificada en el uso actual de la información

A partir de la revisión de las áreas que demandan información, se identifica la siguiente brecha principal:

- **Directech Cía.** Ltda. cuenta con información financiera, comercial y operativa relevante; sin embargo, esta información no se encuentra integrada en un sistema de Business Intelligence que permita analizar de forma consolidada los ingresos, costos, márgenes y rentabilidad real de los proyectos, productos o servicios.
- Esta brecha afecta principalmente a la Gerencia, Ventas, Coordinación de Proyectos, Área Técnica y Área Administrativa-Financiera, debido a que cada área utiliza información parcial para cumplir sus objetivos, pero no dispone de una visión integral que facilite la toma de decisiones basada en datos.

1.4. Objetivos de cada una de las áreas con la información actual y necesidades detectadas

El análisis de las áreas funcionales de Directech Cía. Ltda. evidencia que cada una utiliza información específica para cumplir sus responsabilidades y aportar al logro de los objetivos estratégicos de la organización. Sin embargo, la información disponible actualmente se encuentra distribuida en diferentes sistemas, reportes y registros operativos, lo que limita la capacidad de

realizar un análisis integral y oportuno del desempeño de los proyectos y de la rentabilidad del negocio. En este contexto, resulta necesario identificar los objetivos de cada área, la información con la que actualmente trabaja y las necesidades detectadas, con el propósito de establecer los requerimientos de información que posteriormente serán atendidos mediante el sistema de Business Intelligence.

1.4.1. Gerencia General

La Gerencia General tiene como principal objetivo dirigir estratégicamente la organización, garantizando el crecimiento sostenible de la empresa y la rentabilidad de sus proyectos. Para ello, actualmente utiliza información proveniente de reportes financieros, estados de resultados, informes comerciales y reportes de ejecución de proyectos. No obstante, esta información suele encontrarse dispersa entre diferentes fuentes, por lo que el análisis requiere consolidaciones manuales que consumen tiempo y dificultan una visión oportuna del negocio. Como necesidad principal, la Gerencia requiere disponer de información integrada y actualizada que le permita visualizar la rentabilidad por proyecto, cliente y línea de negocio, identificar desviaciones en los costos y evaluar el desempeño general de la empresa para respaldar la toma de decisiones estratégicas.

1.4.2. Área de Ventas

El objetivo del Área de Ventas es incrementar la comercialización de soluciones tecnológicas mediante propuestas competitivas que generen rentabilidad para la empresa y fortalezcan la relación con los clientes. Actualmente, el área trabaja con información relacionada con cotizaciones, historial comercial, precios, condiciones de negociación y registros de clientes. Sin embargo, no siempre dispone de información precisa sobre los costos reales de proyectos similares ni sobre el margen obtenido en negocios anteriores. La principal necesidad detectada consiste en acceder a información consolidada que permita conocer la rentabilidad histórica por cliente, producto o servicio, facilitando la elaboración de cotizaciones más competitivas y sostenibles para la organización.

1.4.3. Área de Proyectos

El Área de Proyectos tiene como objetivo planificar, coordinar y supervisar la ejecución de los proyectos tecnológicos, asegurando el cumplimiento de los plazos, presupuestos y alcances establecidos con los clientes. Actualmente utiliza presupuestos aprobados, cronogramas, reportes de avance, asignación de recursos y registros de costos. Sin embargo, el seguimiento entre los valores presupuestados y los costos reales requiere procesos manuales de comparación que dificultan la detección temprana de desviaciones. En consecuencia, el área necesita contar con información integrada que permita monitorear permanentemente la ejecución de cada proyecto, controlar la utilización de recursos y conocer el impacto de las variaciones de costos sobre la rentabilidad final.

1.4.4. Área Técnica

El Área Técnica tiene como objetivo garantizar la correcta implementación de las soluciones tecnológicas, asegurando la calidad del servicio y el uso eficiente de los recursos

técnicos disponibles. Para desarrollar sus actividades utiliza información relacionada con equipos, licencias, materiales, órdenes de trabajo, soporte técnico y horas de trabajo del personal especializado. No obstante, gran parte de esta información permanece aislada de la información financiera y comercial. La necesidad principal consiste en disponer de herramientas que permitan relacionar el consumo de recursos técnicos con el desempeño económico de cada proyecto, facilitando la evaluación de la productividad del personal y la optimización de los recursos tecnológicos utilizados.

1.4.5. Área Administrativa y Financiera

El Área Administrativa y Financiera tiene como objetivo mantener un adecuado control de los recursos económicos de la empresa, garantizando una gestión financiera eficiente y apoyando la sostenibilidad del negocio. Actualmente administra información relacionada con facturación, pagos a proveedores, costos, compras, flujo de caja y registros contables. Sin embargo, la integración con la información operativa y comercial es limitada, dificultando la obtención de reportes consolidados sobre la rentabilidad de los proyectos. Por ello, el área requiere disponer de información integrada y automatizada que facilite el análisis financiero, reduzca los tiempos de elaboración de reportes y permita evaluar de forma precisa el comportamiento de los costos, ingresos y utilidades de la organización.

1.4.6. Análisis general de las necesidades detectadas

El análisis realizado demuestra que todas las áreas de Directech Cía. Ltda. comparten una necesidad común: contar con información confiable, integrada y disponible en tiempo real para apoyar la toma de decisiones. Aunque cada departamento cumple funciones distintas, todos dependen de datos relacionados con costos, ingresos, recursos, clientes y proyectos. Actualmente, la dispersión de la información obliga a realizar procesos manuales de consolidación y análisis, lo que incrementa el tiempo de respuesta, limita la capacidad de identificar oportunidades de mejora y dificulta conocer con precisión la rentabilidad de cada proyecto. En este escenario, la implementación de una solución de Business Intelligence permitirá centralizar la información proveniente de las diferentes áreas de la empresa, generando una visión integral del negocio que facilite el seguimiento de los objetivos estratégicos y operativos.

1.5. Definición de los KPI relevantes

Uno de los principales desafíos que enfrenta Directech Cía. Ltda. es la dificultad para conocer con precisión el margen de ganancia real de cada proyecto, producto o servicio, debido a la variabilidad constante de los costos asociados a su ejecución. Actualmente, la información se encuentra distribuida en diferentes fuentes, lo que dificulta obtener una visión integral y oportuna del desempeño financiero y operativo de la empresa.

En este contexto, resulta indispensable definir indicadores clave de desempeño (Key Performance Indicators, KPI) que permitan medir de forma objetiva la rentabilidad, controlar el comportamiento de los costos y evaluar el cumplimiento de los objetivos estratégicos de la organización (Parmenter, 2020). Estos indicadores constituirán la base del sistema de Business

Intelligence, proporcionando información consolidada, confiable y actualizada para apoyar la toma de decisiones gerenciales.

La implementación de estos KPI permitirá identificar proyectos con mayor rentabilidad, detectar desviaciones entre los costos presupuestados y los costos reales, evaluar el retorno de la inversión y monitorear el comportamiento financiero de la empresa en tiempo real. De esta manera, Directech Cía. Ltda. dispondrá de una herramienta que facilitará la optimización de recursos, el fortalecimiento de la planificación estratégica y la mejora continua de sus procesos.

1.5.1. Marco metodológico para la definición de KPI

Los KPI definidos para el proyecto se estructuran en indicadores de rentabilidad y costos, alineados al marco de identificación de brechas (GAP):

Tabla 2

Indicadores clave de desempeño (KPI)

Nombre del indicador	Fórmula de cálculo	Descripción	Fuente de información
Margen Bruto (%)	(Ingresos del proyecto − Costos directos) / Ingresos del proyecto × 100	Mide la rentabilidad operativa de cada proyecto antes de gastos administrativos. Permite identificar qué proyectos generan mayor cobertura de costos directos.	Sistema contable, reportes de proyectos, órdenes de trabajo.
Margen Neto (%)	Utilidad neta del proyecto / Ingresos del proyecto × 100	Determina la rentabilidad final del proyecto considerando todos los costos directos, indirectos, administrativos y financieros asignados.	Estados financieros por proyecto, sistema contable.
Rentabilidad por Proyecto (%)	Utilidad del proyecto / Ingresos del proyecto × 100	Permite identificar cuáles proyectos generan mayor margen de ganancia y cuáles presentan menor rentabilidad debido a la variación de sus costos.	Reportes de proyectos, ingresos por proyecto y costos reales de ejecución.
Variación de Costos (%)	(Costo real − Costo presupuestado) / Costo presupuestado × 100	Detecta la desviación porcentual entre el presupuesto inicial y el costo real de ejecución. Un valor positivo indica sobrecosto; negativo, ahorro.	Presupuestos aprobados, órdenes de compra, registros de costos reales.
ROI por Proyecto (%)	(Beneficio neto del proyecto / Inversión total) × 100	Evalúa el retorno sobre la inversión de cada proyecto, facilitando la priorización estratégica de los proyectos más rentables.	Reportes financieros de proyectos, registros de inversión.
Utilidad Total (USD)	Ventas totales − Costos totales	Muestra la utilidad generada por la empresa en un período determinado, permitiendo evaluar el desempeño financiero global.	Sistema contable, estados financieros y consolidado de ingresos y costos de la empresa.

Nota. Elaboración propia.

1.5.2. Identificación de GAP por KPI

La siguiente tabla compara el resultado actual estimado de Directech con la meta objetivo para cada indicador, identificando la brecha (GAP) que el sistema de BI deberá contribuir a cerrar.

Tabla 3
Identificación de los GAP por KPI

Nombre del indicador	Resultado actual	Meta empresa	GAP	Área responsable
Margen Bruto por Proyecto (%)	~15 % promedio	≥ 25 %	+10 pp	Gerencia / Finanzas
Margen Neto por Proyecto (%)	~10 % promedio	≥ 18 %	+8 pp	Gerencia / Finanzas
Variación de Costos (%)	> +20 % sobrecosto	≤ +5 %	−15 pp	Gerencia / Proyectos
ROI por Proyecto (%)	Dato no disponible	≥ 20 %	Por definir	Gerencia / Finanzas
Margen Neto Global (%)	~10 %	≥ 15 %	+5 pp	Gerencia General

Nota. Elaboración propia con base en estimaciones internas de Directech Cía. Ltda.

1.5.3. Aporte de los KPI al sistema de Business Intelligence

Los indicadores definidos permitirán construir tableros de control interactivos que integrarán información financiera, comercial y operativa en una única plataforma. La visualización en tiempo real del comportamiento de los costos, ingresos y márgenes facilitará el análisis del desempeño de cada proyecto, producto y servicio, permitiendo identificar tendencias, detectar oportunidades de mejora y anticipar posibles desviaciones que afecten la rentabilidad de la organización (Few, 2013).

Adicionalmente, la clasificación de los KPI en las tres categorías metodológicas (Conocimiento, Organización y Motivación) garantiza una visión integral del negocio: desde la rentabilidad financiera por proyecto, pasando por la eficiencia operativa del equipo técnico, hasta el comportamiento comercial con los clientes. Esta estructura facilita la priorización de acciones correctivas y el seguimiento del impacto de cada mejora implementada.

En consecuencia, el sistema de Business Intelligence proporcionará información estratégica para fortalecer la toma de decisiones basada en datos, contribuyendo a una gestión más eficiente de los recursos, al incremento de la rentabilidad y al cumplimiento de los objetivos corporativos de Directech Cía. Ltda.

1.6. Conclusiones del capítulo

Directech Cía. Ltda. cuenta con información importante sobre ventas, costos, proyectos y resultados financieros; sin embargo, esta se encuentra dispersa en diferentes áreas y fuentes. Por ello, se evidencia la necesidad de implementar un modelo de Business Intelligence que integre la información y permita visualizar de forma clara la rentabilidad real, facilitando la toma de decisiones estratégicas, el control de costos y la mejora de la eficiencia operativa.

La identificación de las necesidades de cada área permite establecer los requerimientos de información que deberá satisfacer el sistema de Business Intelligence, proporcionando una base sólida para la definición de KPI, el monitoreo de la rentabilidad de los proyectos y el fortalecimiento de la gestión estratégica de la empresa.

La definición de los KPI constituye el punto de partida para pasar de un modelo basado en información dispersa y procesamiento manual hacia una toma de decisiones sustentada en datos confiables y oportunos. Estos indicadores responden directamente a la problemática central de la empresa (la imposibilidad de conocer el margen real de cada proyecto ante la variabilidad de sus costos), y la identificación de brechas confirma que la mayoría no se miden aún de forma sistemática, lo que refuerza la pertinencia del sistema de BI propuesto.

CAPÍTULO 2: Fuentes de información, sistemas, almacenamiento y reportería actual

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

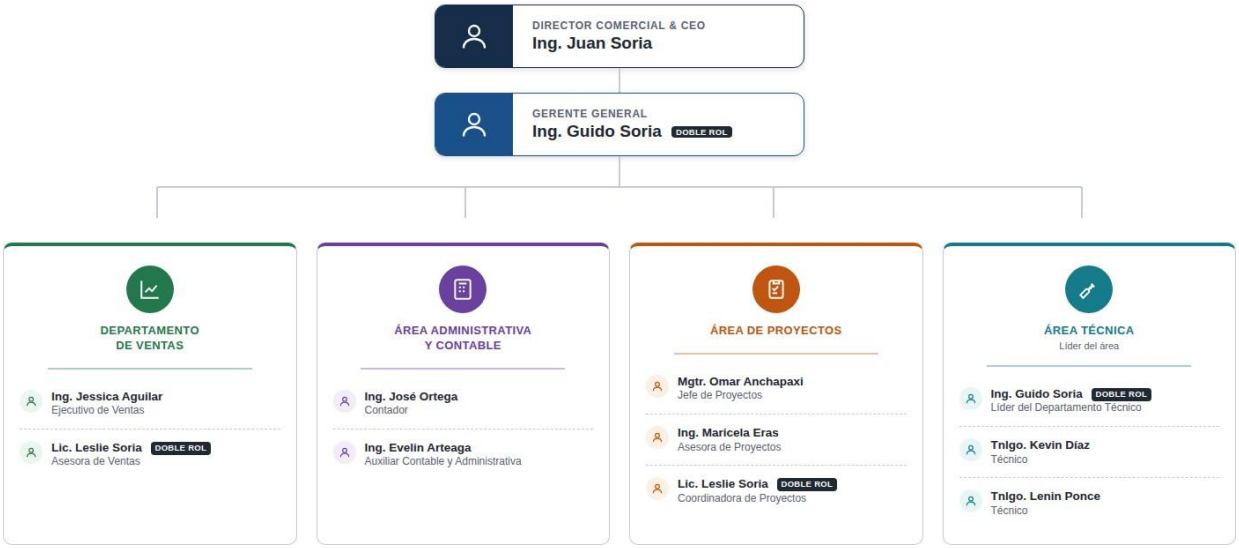
En el contexto del proyecto de Business Intelligence para Directech Cía. Ltda., se identificaron las fuentes de datos que alimentan la operación de la empresa. Directech es una compañía quiteña de soluciones tecnológicas integrales con aproximadamente 10 colaboradores, 15 clientes activos y entre 5 y 10 proyectos tecnológicos ejecutados por año. Sus fuentes de información se clasifican en internas y externas.

2.1.1. Fuentes internas

Las fuentes internas son aquellas generadas por los sistemas y procesos propios de la empresa durante su operación cotidiana. En Directech, la totalidad de la información operativa y financiera proviene de dos sistemas principales y de los registros manuales de cada área funcional.

Figura 2

Organigrama empresarial de Directech Cía. Ltda.



Nota. Elaboración propia con base en información de Directech Cía. Ltda.

Tabla 4

Fuentes de datos internas de Directech

Área	Datos generados	Frecuencia	Responsable
Dirección Comercial / Ventas	Cotizaciones, clientes, órdenes de venta, historial comercial	Diaria	Ing. Jessica Aguilar / Lic. Leslie Soria
Área Administrativa y Contable	Facturas, pagos, costos, compras, registros contables y estados financieros	Diaria	Ing. José Ortega / Ing. Evelin Arteaga

Área	Datos generados	Frecuencia	Responsable
Área de Proyectos	Presupuesto, cronograma, avance de proyectos	Semanal	Mgtr. Omar Anchapaxi / Ing. Maricela Eras / Lic. Leslie Soria
Área Técnica	Horas trabajadas, materiales utilizados, revisión técnica de equipos y apoyo en propuestas técnicas	Diaria	Ing. Guido Soria / Tnlgo. Kevin Díaz / Tnlgo. Lenin Ponce
Gerencia	Aprobación de proyectos, decisiones estratégicas y seguimiento de indicadores	Según necesidad	Ing. Juan Soria / Ing. Guido Soria

Nota. Elaboración propia.

2.1.2. Fuentes externas

Actualmente, Directech no cuenta con la integración de fuentes de datos externas de forma sistemática. Sin embargo, se identifican las siguientes fuentes potenciales que podrían enriquecer el análisis de BI en etapas posteriores del proyecto:

- **Información de proveedores:** cotizaciones, plazos de entrega y condiciones de compra de equipos y materiales tecnológicos.
- **Datos del sector tecnológico ecuatoriano:** indicadores de mercado, tendencias tecnológicas y benchmarking competitivo.
- **Información regulatoria:** normativa del SERCOP para proyectos con entidades del sector público.
- **Información financiera externa:** tasas de cambio e índices económicos relevantes para el costeo de proyectos con componente importado.

2.2. Sistemas y destino de la carga

Directech Cía. Ltda. opera con dos herramientas principales para la gestión de su información. La empresa presenta una arquitectura de información básica, lo que constituye tanto una limitación como una oportunidad para la implementación del proyecto de BI.

Tabla 5

Sistemas de información de Directech

Aspecto	Descripción
Sistemas utilizados	Directech utiliza el software contable Latinium para la gestión financiera y contable. El seguimiento de proyectos se realiza mediante Microsoft Excel. Actualmente la empresa no dispone de un ERP ni de un CRM, aunque se encuentra dentro de sus objetivos implementar un CRM para fortalecer la gestión comercial y el seguimiento de clientes.
Información almacenada por sistema	Latinium: clientes, ventas, facturación, compras, proveedores, cuentas por cobrar, cuentas por pagar, costos, registros contables y estados financieros. Excel: cronogramas, presupuestos, seguimiento de proyectos y control de avances.

Nota. Elaboración propia.

2.2.1. Proceso de negocio y flujo de datos

El flujo de información de Directech sigue el ciclo de vida de un proyecto tecnológico, desde la identificación de la oportunidad comercial hasta el cierre financiero. A continuación, se describe el proceso de negocio y cómo se genera la información en cada etapa:

1. El cliente solicita una solución tecnológica o el área comercial identifica una oportunidad de negocio.
2. El Departamento de Ventas prepara la propuesta comercial y coordina con el Área Técnica la revisión de equipos, materiales y requerimientos técnicos.
3. El Director Comercial & CEO (Ing. Juan Soria) analiza la propuesta y decide si el proyecto será ejecutado.
4. Una vez aprobado, el Jefe de Proyectos planifica el cronograma, asigna recursos y coordina la ejecución.
5. El Área Técnica ejecuta el proyecto registrando horas trabajadas y materiales utilizados.
6. El Área Administrativa y Contable registra compras, costos, facturación y pagos en Latinium.

Este proceso genera datos en cada etapa que actualmente quedan dispersos entre Latinium y hojas de cálculo de Excel, sin una integración que permita la trazabilidad completa del ciclo de vida de un proyecto.

2.3. Almacenamiento en data warehouse y datamarts

En la situación de partida de Directech Cía. Ltda. no existe un repositorio analítico centralizado ni una integración ETL formal. El Data Warehouse relacional se diseña e ilustra mediante la prueba de concepto académica posterior, sin atribuir a la empresa una plataforma operativa que todavía no se ha desplegado.

2.3.1. Situación actual del almacenamiento de datos

La información de Directech se encuentra distribuida en dos repositorios aislados:

- **Latinium (sistema contable):** almacena datos financieros, contables y de facturación. Genera reportes automáticos en formato PDF, pero su información no se exporta ni se integra con otras fuentes de manera sistemática.
- **Microsoft Excel (hojas de cálculo):** almacena cronogramas, presupuestos, seguimiento de proyectos y control de avances. La actualización es manual y depende de cada responsable de área.

No existe un proceso de consolidación nocturna ni una interfaz de carga automatizada, lo que genera que la información disponible para la toma de decisiones pueda estar desactualizada o ser inconsistente entre fuentes.

2.3.2. Propuesta de arquitectura de datos para Directech

Siguiendo los principios de un almacén de datos orientado al análisis (Inmon, 2005; Kimball & Ross, 2013), se propone para Directech una arquitectura simplificada que contemple:

- **Capa de fuentes:** Latinium (vía exportación de datos) y archivos Excel estandarizados como las dos fuentes primarias.
- **Capa de integración (ETL):** un proceso de extracción periódica que consolide la información de ambas fuentes en una base de datos única, eliminando duplicidades y estandarizando formatos.
- **Capa de almacenamiento:** una base de datos centralizada (Data Warehouse a escala PYME) donde se almacenen los datos históricos de proyectos, ventas, costos y clientes.
- **Capa de análisis y reportería:** dashboards e indicadores clave (KPI) accesibles para la Gerencia, permitiendo la toma de decisiones basada en datos consolidados.

Tabla 6

Necesidades de transformación de la información en Directech

Necesidad	Descripción de la necesidad
Reportería y análisis	Capacidad de generar reportes operativos, tácticos y gerenciales que consoliden información de Latinium y Excel. Incluye: estados financieros por proyecto, análisis de rentabilidad, comparación de costos presupuestados vs. reales, y exploración visual de datos mediante dashboards.
Analítica y minería	Capacidad de identificar patrones en el comportamiento de costos, rentabilidad por cliente, productividad del personal técnico y tendencias de ventas. Permitiría predecir desviaciones presupuestarias y segmentar clientes por nivel de rentabilidad.

Nota. Elaboración propia.

2.4. Carencias y problemas en la carga de datos

El diagnóstico realizado sobre la gestión de datos de Directech Cía. Ltda. revela múltiples carencias que limitan la capacidad analítica de la empresa y dificultan la toma de decisiones basada en información confiable y oportuna.

Tabla 7

Diagnóstico de carencias y problemas en la gestión de datos

Categoría	Descripción
Información difícil de obtener	Rentabilidad real de cada proyecto, margen de ganancia efectivo, productividad del personal técnico y comportamiento de los costos durante la ejecución. Aunque se dispone de información financiera y operativa, esta se encuentra distribuida en diferentes registros, lo que dificulta su análisis integral.
Procesos manuales	La consolidación de información para el análisis gerencial se realiza de forma manual. El cálculo de los costos reales por proyecto depende de registros y validaciones manuales, dificultando la obtención de indicadores oportunos y precisos.
Sistemas no integrados	Latinium y Microsoft Excel operan de forma aislada. No existe integración entre ambas herramientas ni con un sistema ERP o CRM, lo que ocasiona duplicidad de información y limita la disponibilidad de datos consolidados para la toma de decisiones.

Nota. Elaboración propia.

2.4.1. Análisis comparativo con el caso de referencia

En el caso de referencia utilizado como guía metodológica (Bancolombia), la Dirección de Capacidades Analíticas gobierna la estructura de archivos, certifica los campos y calendariza las actualizaciones incrementales. Directech presenta las siguientes brechas respecto de ese modelo de referencia:

- No existe un proceso de cierre ni generación automática de archivos de datos.
- No hay un repositorio centralizado que consolide información de todas las áreas.
- No se cuenta con un gobierno de datos que defina tipos de dato, frecuencia de actualización ni estándares de calidad (DAMA International, 2017).
- La extracción de información depende de consultas manuales individuales a cada sistema.

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

2.5.1. Reportes actuales

Tabla 8
Estado actual de la reportería en Directech

Aspecto	Descripción
Reportes generados	Reportes de ventas por cliente, estados financieros, reportes generales y detallados de ventas y compras, así como reportes de gastos. Permiten el seguimiento financiero y operativo de la empresa.
Usuarios de reportes	Contador (Ing. José Ortega), Gerente General (Ing. Guido Soria) y Director Comercial & CEO (Ing. Juan Soria). Los emplean para control financiero, evaluación del desempeño empresarial y toma de decisiones estratégicas.
Tipo de generación	Los reportes son generados automáticamente por Latinium en formato PDF. Sin embargo, parte de la información requerida para su elaboración debe ser ingresada manualmente al sistema, lo que incrementa el tiempo de procesamiento y puede afectar la oportunidad de la información.

Nota. Elaboración propia.

2.5.2. Necesidades detectadas: dashboard de BI

Con base en el levantamiento de información realizado con la Gerencia de Directech, se identificaron las siguientes necesidades de información que actualmente no están cubiertas por los reportes existentes. Estas necesidades constituyen los requerimientos para el diseño del dashboard de Business Intelligence:

- Ventas por cliente y por línea de negocio.
- Rentabilidad real de cada proyecto.
- Comparación entre costos presupuestados y costos reales de ejecución.
- Estado de los proyectos (planificados, en ejecución y finalizados).
- Gastos operativos asociados a cada proyecto.
- Compras realizadas por proyecto y su impacto en los costos.

- Margen bruto y margen neto por proyecto y de la empresa.
- Flujo de ingresos y egresos por período.
- Clientes con mayor volumen de ventas y mayor rentabilidad.
- **Indicadores financieros generales:** ingresos, costos, utilidad y margen de ganancia.

2.5.3. Indicadores clave (KPI) priorizados

La Gerencia de Directech considera como indicadores más relevantes para la toma de decisiones los siguientes:

- Margen bruto por proyecto.
- Margen neto por proyecto.
- Rentabilidad por proyecto.
- **Variación de costos (presupuestado vs.** real).
- Retorno sobre la inversión (ROI).
- Utilidad total.
- Ventas por cliente.
- Costos reales versus costos presupuestados.

2.5.4. Decisiones que no pueden tomarse por falta de información

El relevamiento con la Gerencia identificó las siguientes decisiones que actualmente no pueden fundamentarse en datos confiables y consolidados:

- Determinar con precisión la rentabilidad real de cada proyecto.
- Identificar oportunamente desviaciones entre el presupuesto inicial y los costos reales de ejecución.
- Medir objetivamente la productividad del personal técnico durante el desarrollo de los proyectos.
- Evaluar con indicadores financieros el desempeño de cada proyecto y cliente.
- Optimizar la asignación de recursos humanos y materiales con base en información confiable.
- Analizar tendencias de costos para establecer estrategias que incrementen la rentabilidad de la empresa.

2.6. Conclusiones del capítulo

El levantamiento de fuentes de datos realizado para Directech Cía. Ltda. evidencia que la empresa cuenta con información valiosa generada en sus operaciones diarias, pero que esta se encuentra fragmentada entre dos sistemas no integrados (Latinium y Excel), carece de estándares de gobierno de datos y no dispone de procesos automatizados de consolidación.

Se identificó la necesidad de implementar un proceso ETL que permita extraer, transformar y cargar los datos de ambas fuentes en un repositorio centralizado. Directech puede

implementar una solución de menor escala, acorde a su tamaño y complejidad operativa. Los principales hallazgos son:

- Directech genera datos operativos en cinco áreas funcionales con frecuencias que van desde diaria hasta semanal, pero sin un esquema de consolidación.
- Los dos sistemas actuales (Latinium y Excel) no están integrados, lo que genera duplicidad de información e impide obtener una visión unificada del negocio.
- Los reportes existentes se limitan a la información contable y no cubren las necesidades analíticas de la Gerencia, particularmente en lo relativo a rentabilidad por proyecto, productividad técnica y comparación de costos.
- La implementación de un Data Warehouse a escala PYME, complementado con dashboards de BI, permitiría a Directech pasar de un modelo de decisión basado en intuición a uno basado en datos consolidados y oportunos.

CAPÍTULO 3: Estructura de la información y procesos de carga de datos

3.1. Procesos de carga de datos (ETL, data lake)

Actualmente, Directech Cía. Ltda. no cuenta con un proceso ETL (extracción, transformación y carga) formalizado ni con un data lake. ETL es el proceso mediante el cual se combinan, limpian y organizan datos provenientes de múltiples fuentes en un único conjunto de datos consistente para su almacenamiento en un Data Warehouse (IBM, 2026a). La carga de información se realiza de forma manual: cada responsable de área ingresa o actualiza los datos directamente en Latinium (módulo contable/financiero) o en hojas de cálculo de Excel (proyectos, cronogramas, presupuestos). No existe un proceso de extracción automatizada desde estas fuentes, ni una rutina programada de consolidación (por ejemplo, cargas nocturnas o batch).

Como se planteó en el Capítulo 2 (sección 2.3.2), la propuesta para Directech consiste en implementar un proceso ETL de escala PYME que:

- Extraiga periódicamente los datos de Latinium (vía exportación) y de los archivos Excel estandarizados.
- Transforme la información, homologando formatos, eliminando duplicidades y validando consistencia entre fuentes.
- Cargue los datos consolidados en un repositorio único (Data Warehouse a escala PYME), evitando así los procesos manuales actuales de comparación y consolidación.

No se contempla, en esta fase, la implementación de un data lake, dado que el volumen y variedad de datos de Directech (dos fuentes estructuradas) no justifica dicha complejidad; un Data Warehouse relacional, estructurado bajo un modelo dimensional (Kimball & Ross, 2013), resulta más adecuado para sus necesidades actuales.

A modo de referencia, el siguiente modelo conceptual, adaptado del caso de Bancolombia utilizado como guía metodológica, ilustra cómo se relacionarían las fuentes de información, la capa de integración/almacenamiento y las capacidades de análisis en el proyecto BI DIRECTECH:

Figura 3

Modelo conceptual propuesto para BI DIRECTECH (fuentes → integración → capacidades)



Nota. Adaptado del caso de referencia Bancolombia (Pérez Galán, 2026a).

3.2. Información y calidad de la información disponible

La información disponible es relevante y cubre las áreas clave del negocio (ventas, costos, proyectos, finanzas), pero presenta limitaciones de calidad derivadas de su naturaleza manual y no integrada. La calidad de los datos suele evaluarse en función de su exactitud, completitud, consistencia y oportunidad (Redman, 2001), dimensiones que se analizan a continuación para el caso de Directech:

Tabla 9

Dimensiones de calidad de la información en Directech

Dimensión de calidad	Situación actual
Consistencia	Datos duplicados entre Latinium y Excel al no existir integración entre sistemas.
Actualización	Depende de cada responsable de área; no hay frecuencia estandarizada de refresco.
Confiabilidad	Los cálculos de costos reales y márgenes dependen de validaciones manuales, con riesgo de error humano.

Dimensión de calidad	Situación actual
Completitud	Existen datos no capturados sistemáticamente (p. ej., ROI por proyecto, actualmente "dato no disponible" según la identificación de GAP del Capítulo 1).
Trazabilidad	No es posible seguir el ciclo completo de un proyecto (comercial → técnico → financiero) en un solo sistema.

Nota. Elaboración propia con base en las dimensiones de calidad de Redman (2001).

En consecuencia, aunque la información existe, su calidad actual no es suficiente para sustentar un sistema de BI sin un proceso previo de limpieza, estandarización y gobierno de datos.

Adicionalmente, y siguiendo el mismo esquema de categorización de variables utilizado en el caso de referencia, se identifican los siguientes tipos de datos disponibles en Directech, en función de las necesidades de análisis de rentabilidad y costos. La variable en común que permite relacionar los datos entre las distintas fuentes es el identificador de cada proyecto:

Tabla 10

Tipos de datos disponibles en Directech

Tipo de dato	Variables disponibles
Ventas y comercial	Cotizaciones, historial de clientes, condiciones comerciales, descuentos aplicados, rentabilidad histórica por cliente o tipo de proyecto.
Costos y compras	Costos directos e indirectos, compras de equipos y materiales, órdenes de compra, proveedores.
Financiero-contable	Facturación, cuentas por cobrar/pagar, asientos contables, balances, flujo de caja.
Proyectos	Presupuesto aprobado, cronograma, avance de ejecución, recursos asignados, desviaciones presupuestarias.
Técnica	Horas trabajadas, materiales utilizados, equipos, licencias, soporte técnico.
Indicadores	Margen bruto, margen neto, rentabilidad por proyecto, variación de costos, ROI, utilidad total.

Nota. Elaboración propia.

3.3. Horarios de carga y accesibilidad

De acuerdo con el levantamiento de fuentes internas presentado en el Capítulo 2, la frecuencia de generación de datos varía según el área:

Tabla 11

Frecuencia de carga y accesibilidad por área

Área	Frecuencia de carga	Accesibilidad actual
Dirección Comercial / Ventas	Diaria	Un asesor de ventas tiene acceso directo a Latinium.
Área Administrativa y Contable	Diaria (carga) / Semanal (validación)	Contador y auxiliar contable, con acceso directo.

Área	Frecuencia de carga	Accesibilidad actual
Área de Proyectos	Semanal	Archivos Excel de uso local, sin acceso compartido formal.
Área Técnica	Diaria	Registros internos, no visibles para otras áreas.
Gerencia	Según necesidad	No tiene acceso directo a Latinium; depende de reportes generados por el contador.

Nota. Elaboración propia.

El acceso directo a Latinium está limitado a tres perfiles: el contador, la auxiliar contable y un asesor de ventas. En el Área Administrativa y Contable, la carga de información la realiza la auxiliar contable de forma diaria, mientras que el contador (Ing. José Ortega) efectúa una revisión semanal para verificar que dicha carga se haya realizado correctamente. Este control manual de calidad, si bien mitiga errores, introduce un desfase de hasta una semana entre el registro del dato y su validación, lo que retrasa la disponibilidad de información contable confiable para la Gerencia.

La Gerencia General y el Director Comercial & CEO no cuentan con acceso directo al sistema, por lo que dependen de que el contador les genere y entregue reportes puntuales para la toma de decisiones. No existe, además, un horario de carga estandarizado a nivel de toda la empresa, ni una ventana de consolidación definida entre áreas (por ejemplo, un cierre diario único). Esta limitación refuerza la necesidad identificada en el proyecto: la Gerencia requiere visibilidad autónoma y en tiempo real, sin depender de solicitudes manuales a terceros dentro de la organización. La propuesta de BI plantea automatizar la extracción desde Latinium y Excel con una frecuencia diaria, e incorporar validaciones automáticas de calidad que reduzcan la dependencia de la revisión manual semanal actualmente realizada por el contador.

3.4. Variables disponibles y acceso al Data Warehouse

Un hallazgo relevante para la viabilidad técnica del proyecto es que Latinium permite exportar a Excel los principales reportes del sistema, incluyendo compras, ventas, asientos contables y balances. Esta capacidad de exportación es clave para el diseño del proceso ETL propuesto, ya que reduce significativamente la complejidad de la fase de extracción: en lugar de requerir una integración directa o una API con Latinium, el proceso puede apoyarse en exportaciones periódicas a Excel como mecanismo de extracción, las cuales luego serían transformadas y cargadas al Data Warehouse. Esto valida que la arquitectura propuesta (capa de fuentes → ETL → almacenamiento → análisis) es técnicamente alcanzable con las herramientas que Directech ya posee, sin necesidad de inversión inmediata en conectores especializados.

Las variables disponibles para alimentar el Data Warehouse, considerando esta capacidad de exportación, son:

- **Desde Latinium (vía exportación a Excel):** compras, ventas, asientos contables, balances, facturación, cuentas por cobrar/pagar, costos.
- **Desde Excel (Área de Proyectos):** presupuesto, cronograma, avance de proyectos, recursos asignados.
- Desde el Área Técnica (registros internos): horas trabajadas, materiales utilizados.

El acceso al futuro Data Warehouse se propone diferenciado por perfil: Gerencia con visión consolidada de KPI (por primera vez, sin depender del contador); el asesor de ventas y el contador manteniendo su acceso operativo; y el Área de Proyectos con visibilidad sobre el cruce entre presupuesto y ejecución real.

A diferencia de organizaciones de mayor tamaño como el caso de referencia, donde una Dirección de Capacidades Analíticas centraliza el gobierno de datos y autoriza cada solicitud de acceso según la vocación del usuario, Directech no cuenta actualmente con una función formal de gobierno de datos, entendida como el conjunto de políticas y roles que garantizan la disponibilidad, integridad y seguridad de la información (DAMA International, 2017). Se recomienda que, en una primera etapa, este rol sea asumido por la Gerencia General en conjunto con la Coordinación de Proyectos, definiendo qué perfiles pueden consultar qué variables del Data Warehouse y por cuánto tiempo, evitando así que el crecimiento de la información derive en los mismos problemas de dispersión y falta de control que hoy afectan a la organización.

3.5. Sistemas implicados en los procesos

Los sistemas actualmente implicados son:

- **Latinium:** software contable que gestiona clientes, ventas, facturación, compras, proveedores, cuentas por cobrar/pagar y estados financieros.
- **Microsoft Excel:** herramienta utilizada para cronogramas, presupuestos, seguimiento de proyectos y control de avances.

Directech no dispone actualmente de un ERP ni de un CRM, aunque la implementación de un CRM está contemplada dentro de sus objetivos futuros para fortalecer la gestión comercial. Ningún sistema de Business Intelligence, Data Warehouse o herramienta de visualización (dashboards) está implementado hasta el momento; estos constituyen precisamente el objeto del proyecto.

3.6. Recursos de IT o externos empleados

Directech no cuenta con un departamento de TI dedicado propio. El soporte y mantenimiento de los equipos y proyectos de clientes recae en el Área Técnica interna. Sin embargo, sí existe un proveedor externo identificado: la empresa de software Latinium, con quien Directech mantiene un contrato de mantenimiento anual que incluye soporte técnico sobre el sistema contable. Este es actualmente el único recurso externo de TI contratado de forma recurrente por la organización.

Para la implementación del proyecto de BI, se estima necesario recurrir a:

- **Recursos internos:** la Coordinadora de Proyectos (vínculo directo con el equipo del proyecto) y el Área Técnica, que puede apoyar en la configuración de herramientas de extracción y visualización.
- **Recursos externos potenciales:** dado que el contrato actual con Latinium cubre soporte del sistema contable pero no contempla servicios de Business Intelligence, integración de datos o dashboards, se recomienda evaluar si Latinium ofrece algún módulo o API de exportación de datos aprovechable para el ETL, o si se requerirá contratar un proveedor

adicional especializado en BI/Data Warehouse a escala PYME, ya que Directech no dispone de personal interno especializado en modelado de datos o administración de bases de datos analíticas.

3.7. Conclusiones del capítulo

Extraer, transformar y cargar (ETL, extract, transform, load) es el proceso mediante el cual se compilan datos provenientes de distintas fuentes para su posterior organización y centralización en un único repositorio (IBM, 2026a). Este proceso constituye la base operativa de todo sistema de Business Intelligence, entendido como el conjunto de aplicaciones, infraestructura y prácticas que permiten acceder a la información y analizarla para mejorar y optimizar las decisiones empresariales (Gartner, s.f.).

En la mayoría de las empresas, la información potencialmente útil permanece dispersa o subutilizada: un estudio reciente sobre la brecha de valor de los datos encontró que solo un 32 % de las organizaciones logra obtener beneficios tangibles y medibles de su información, lo que implica que las dos terceras partes restantes obtienen pocos resultados concretos o ninguno (The Blue Owls, 2025). Otros estudios coinciden en que buena parte de las organizaciones enfrenta este mismo reto de calidad y aprovechamiento de datos (Nagle et al., 2017).

El caso de Directech Cía. Ltda. refleja precisamente esta situación: la empresa cuenta con información valiosa en Latinium y en archivos Excel, pero al no existir un proceso ETL formal, dicha información permanece fragmentada, limitando la capacidad de conocer con precisión la rentabilidad real de cada proyecto. A diferencia de organizaciones de mayor escala como el caso de referencia analizado, que cuenta con una Landing Zone, cargas automatizadas y un gobierno de datos centralizado, Directech requiere una solución ETL de menor escala, pero igualmente rigurosa en sus principios: extracción periódica, transformación estandarizada y carga hacia un repositorio único, modelado bajo un esquema dimensional.

En este proyecto, el reto consiste en transformar los datos recopilados de Latinium y Excel mediante un proceso ETL adaptado a la realidad y tamaño de Directech, que permita alimentar de forma confiable el Data Warehouse propuesto y, con ello, sustentar los KPI de rentabilidad y costos definidos previamente para la organización.

CAPÍTULO 4: Arquitectura de Business Intelligence y perfiles de usuario

En los capítulos anteriores se identificaron las fuentes de información disponibles (Latinium y Excel), se diagnosticaron las carencias en la gestión de datos y se definieron los indicadores clave de desempeño (KPI) de rentabilidad y costos que la organización requiere monitorear. Corresponde ahora tomar dos decisiones estructurales para el diseño de la solución: qué arquitectura de datos soportará el sistema de BI, y para qué perfiles de usuario se construirá la información, de manera que el diseño posterior de tableros y visualizaciones responda a necesidades reales y no a supuestos genéricos.

4.1. Elección de la arquitectura de Business Intelligence

La arquitectura de BI define cómo se integran, almacenan y estructuran los datos antes de llegar a la capa de análisis y visualización, siguiendo la guía metodológica utilizada como referencia para este proyecto (Pérez Galán, 2026a). Elegir correctamente esta arquitectura es determinante para Directech, ya que de ella depende la viabilidad técnica y económica de todo el proyecto, dado que la empresa no cuenta con un departamento de TI propio ni con presupuesto para infraestructura compleja.

4.1.1. Opciones de arquitectura evaluadas

Se evaluaron cuatro enfoques de arquitectura de datos aplicables a un proyecto de BI, considerando el volumen y variedad de información que actualmente maneja Directech (dos fuentes estructuradas: Latinium y Excel):

Tabla 12

Opciones de arquitectura de BI evaluadas

Arquitectura	Descripción	Aplicabilidad para Directech
Data Warehouse relacional (modelo dimensional)	Repositorio centralizado que integra datos estructurados y homogéneos bajo un modelo dimensional (hechos y dimensiones), orientado a consultas analíticas (Kimball & Ross, 2013).	Alta. Adecuado para un volumen bajo de fuentes estructuradas, como el caso de Directech (Latinium y Excel).
Data Lake	Repositorio que almacena datos en su formato original (estructurados, semiestructurados y no estructurados), sin un esquema previo definido.	Baja. Está pensado para grandes volúmenes y variedad de datos que Directech no genera actualmente.
Data Lakehouse	Arquitectura híbrida que combina la flexibilidad del data lake con las capacidades analíticas de un data warehouse (Armbrust et al., 2021).	Baja. Añade complejidad técnica e inversión que no se justifica para una PYME sin personal especializado en datos.
Solución basada solo en Excel/Power Query (sin repositorio centralizado)	Consolidación manual o semiautomática de archivos Excel, sin una base de datos central.	Insuficiente. Es la situación actual de Directech y no resuelve la duplicidad ni la falta de trazabilidad identificadas.

Nota. Elaboración propia.

4.1.2. Criterios de selección

La elección de la arquitectura se basó en los siguientes criterios, priorizados de acuerdo con la realidad operativa de Directech:

Tabla 13
Criterios de selección de arquitectura para Directech

Criterio	Consideración para Directech
Volumen y variedad de datos	Dos fuentes estructuradas (Latinium y Excel), sin datos no estructurados relevantes en el corto plazo.
Presupuesto disponible	Empresa con ingresos aproximados de USD 900.000 anuales; no es viable una inversión significativa en infraestructura de datos.
Personal disponible	No existe un departamento de TI propio; el soporte recae en el Área Técnica, sin especialización en administración de bases de datos analíticas.
Tiempo de implementación	Se requiere una solución que pueda implementarse y mostrar resultados dentro del alcance del proyecto de especialización.
Escalabilidad futura	La arquitectura debe permitir incorporar en el futuro nuevas fuentes (por ejemplo, un CRM contemplado dentro de los objetivos de la empresa) sin rediseñar todo el modelo.

Nota. Elaboración propia.

4.1.3. Arquitectura seleccionada y justificación

Se selecciona un Data Warehouse relacional a escala PYME, estructurado bajo un modelo dimensional (Kimball & Ross, 2013), alimentado mediante un proceso ETL (IBM, 2026a) que extrae la información desde Latinium (vía exportación a Excel) y desde los archivos Excel del Área de Proyectos y del Área Técnica. Esta arquitectura fue introducida como propuesta en el Capítulo 2; en este capítulo se formaliza como la decisión definitiva de arquitectura, justificada frente a las demás alternativas evaluadas.

Tabla 14
Componentes de la arquitectura seleccionada

Componente	Descripción
Capa de fuentes	Latinium (vía exportación a Excel) y archivos Excel estandarizados del Área de Proyectos y del Área Técnica.
Capa de integración (ETL)	Extracción periódica, homologación de formatos, eliminación de duplicidades y validación de consistencia entre fuentes, apoyada en Power Query.
Capa de almacenamiento	Data Warehouse relacional a escala PYME, modelado bajo esquema dimensional (estrella), con una tabla de hechos y dimensiones descriptivas.
Capa de análisis y consumo	Modelo de datos consumido por la herramienta de BI para la construcción de indicadores y tableros de control por perfil de usuario.

Nota. Elaboración propia.

Esta arquitectura no contempla, en esta fase, la implementación de un data lake ni de un modelo lakehouse, dado que el volumen y variedad de datos de Directech no justifica dicha complejidad. La solución elegida es proporcional al tamaño de la empresa y resuelve

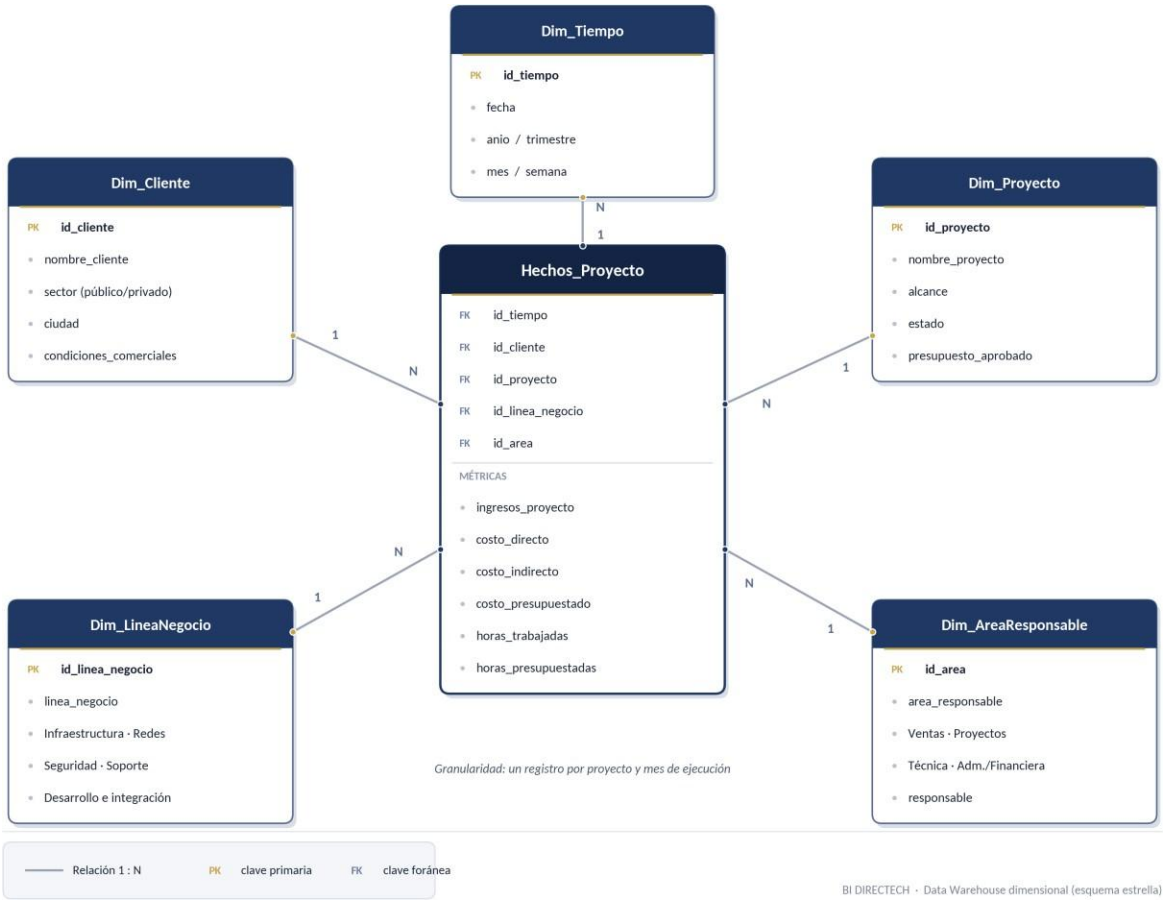
directamente la brecha diagnosticada: la ausencia de un repositorio único que consolide la información hoy dispersa entre Latinium y Excel.

4.1.4. Modelo dimensional propuesto (esquema estrella)

Para operacionalizar la arquitectura seleccionada, se propone un esquema estrella compuesto por una tabla de hechos central (Hechos_Proyecto), que concentra las métricas de rentabilidad y costos, y cinco dimensiones que permiten analizar dichas métricas desde distintas perspectivas: tiempo, cliente, proyecto, línea de negocio y área responsable.

Figura 4

Esquema estrella propuesto para el Data Warehouse de BI DIRECTECH



Nota. Elaboración propia con base en Kimball y Ross (2013).

Este modelo permitirá calcular directamente los KPI definidos en el Capítulo 1 (Margen Bruto, Margen Neto, Rentabilidad por Proyecto, Variación de Costos, ROI y Utilidad Total) mediante la combinación de las medidas de la tabla de hechos con los atributos de cada dimensión, sin necesidad de recalcular manualmente la información en Excel.

4.1.5. Herramienta de análisis y visualización seleccionada

Definida la arquitectura de datos, es necesario identificar qué herramienta consumirá el Data Warehouse propuesto para construir los indicadores y tableros de control que utilizarán los distintos perfiles de usuario. Se evaluaron cuatro alternativas —Power BI, Tableau, Google Looker Studio y Excel avanzado (Power Query/tablas dinámicas)— bajo los criterios de costo, integración con las fuentes actuales de Directech (Latinium y Excel), curva de aprendizaje y capacidad de generar tableros interactivos:

Tabla 15

Herramientas de análisis y visualización evaluadas

Herramienta	Descripción	Aplicabilidad para Directech
Power BI	Herramienta de Microsoft para modelado, análisis y visualización de datos, con versión de escritorio gratuita y publicación en la nube mediante licenciamiento Pro (Microsoft, 2026a).	Alta. Se integra de forma nativa con Excel y archivos planos, tiene una curva de aprendizaje accesible y permite publicar vistas diferenciadas por perfil sin personal especializado en programación.
Tableau	Plataforma de visualización reconocida por la flexibilidad y calidad de sus gráficos interactivos, con conectores a múltiples fuentes de datos.	Media. Ofrece capacidades de visualización robustas, pero su licenciamiento es más costoso que Power BI y su curva de aprendizaje es mayor para un equipo sin experiencia previa en BI.
Google Looker Studio	Herramienta gratuita de visualización de Google, orientada a la construcción de reportes conectados principalmente a fuentes de datos de Google y hojas de cálculo.	Baja. Su integración con fuentes locales como Latinium es limitada y depende en gran medida del ecosistema de Google.
Excel avanzado (Power Query/tablas dinámicas)	Uso de funciones avanzadas de Excel, tablas dinámicas y Power Query para consolidar y analizar datos sin una herramienta de BI dedicada.	Insuficiente como solución definitiva. Es la práctica actual de Directech; no permite construir tableros interactivos ni vistas diferenciadas por perfil.

Nota. Elaboración propia.

Se recomienda Microsoft Power BI como la herramienta de análisis y visualización para el proyecto BI DIRECTECH, por las siguientes razones:

- **Costo accesible:** Power BI Desktop es gratuito para construir el modelo y los tableros; solo se requiere licenciamiento Pro para publicar y compartir el acceso entre los perfiles de usuario identificados, lo que resulta viable para una PYME con ingresos aproximados de USD 900.000 anuales (Microsoft, 2026a).
- Integración directa con el esquema estrella propuesto: Power BI permite construir el modelo de hechos y dimensiones (Hechos_Proyecto, Dim_Cliente, Dim_Proyecto, Dim_Tiempo, Dim_LineaNegocio y Dim_AreaResponsable) mediante su motor de modelado, conectándose a los archivos exportados desde Latinium y a los archivos Excel del Área de Proyectos y del Área Técnica.
- **Curva de aprendizaje razonable:** su interfaz visual permite que el Área Técnica y la Coordinación de Proyectos mantengan y actualicen los modelos sin depender de un

desarrollador externo, dado que Directech no cuenta con personal especializado en programación.

- **Autoservicio por perfil de usuario:** mediante Power BI Service (nube) es posible publicar vistas diferenciadas para cada uno de los cinco perfiles definidos en la sección 4.2 (estratégico, táctico-proyectos, táctico-financiero, operativo-comercial y operativo-técnico), respetando el nivel de acceso que corresponde a cada uno.
- **Escalabilidad:** permite incorporar en el futuro nuevas fuentes de datos, como un eventual CRM contemplado dentro de los objetivos de la empresa, sin necesidad de migrar de herramienta ni de arquitectura.

En consecuencia, la arquitectura de Data Warehouse a escala PYME con modelo dimensional, sumada a Power BI como herramienta de análisis y visualización, conforma la propuesta tecnológica integral de BI DIRECTECH: desde la integración de los datos hasta su consumo por parte de cada perfil de usuario de la organización.

4.2. Perfil de los usuarios de la información

Definir para quién se construye la información es tan importante como definir cómo se almacena. Un mismo Data Warehouse puede alimentar necesidades muy distintas según el perfil que consulte la información: la Gerencia requiere una visión consolidada y estratégica, mientras que el Área Técnica requiere el detalle operativo de horas y materiales. Esta sección identifica y caracteriza los perfiles de usuario de BI DIRECTECH.

4.2.1. Metodología para la definición de perfiles

Se adopta un enfoque de perfiles (personas) que caracteriza a cada tipo de usuario a partir de cinco atributos: su objetivo principal dentro de la organización, la información que necesita consultar, la frecuencia con la que la requiere, el nivel de acceso que debe tener sobre el Data Warehouse y la herramienta a través de la cual consumirá la información. Esta caracterización se construye a partir del levantamiento de áreas y necesidades realizado en el Capítulo 1 (revisión de áreas que demandan información y objetivos por área).

4.2.2. Perfiles de usuario identificados en Directech

Tabla 16

Perfiles de usuario identificados en Directech

Perfil	Rol en la organización	Objetivo principal frente a la información
Perfil Estratégico	Gerencia General / Director Comercial & CEO	Dirigir estratégicamente la organización, evaluar la rentabilidad global y aprobar proyectos con base en indicadores consolidados.
Perfil Táctico – Proyectos	Coordinación de Proyectos	Planificar, controlar y cerrar proyectos, monitoreando el cumplimiento de presupuesto, plazos y recursos.
Perfil Táctico – Financiero	Contador / Área Administrativa y Contable	Controlar los recursos económicos de la empresa y garantizar la exactitud de los registros contables que alimentan el Data Warehouse.

Perfil	Rol en la organización	Objetivo principal frente a la información
Perfil Operativo – Comercial	Área de Ventas	Elaborar propuestas comerciales competitivas con base en la rentabilidad histórica por cliente o tipo de proyecto.
Perfil Operativo – Técnico	Área Técnica	Ejecutar las soluciones tecnológicas optimizando el uso de horas y materiales frente a lo presupuestado.

Nota. Elaboración propia.

4.2.3. Necesidades de información por perfil

Tabla 17

Necesidades de información por perfil de usuario

Perfil	Necesidad de información
Perfil Estratégico	Margen bruto y neto consolidados, utilidad total, ROI, rentabilidad comparada entre proyectos, clientes y líneas de negocio, y evolución de estos indicadores en el tiempo.
Perfil Táctico – Proyectos	Comparación entre presupuesto aprobado y costo real por proyecto, variación de costos, avance de ejecución y desviaciones frente a lo planificado.
Perfil Táctico – Financiero	Facturación, costos directos e indirectos, flujo de caja y estados financieros por proyecto, como insumo para el cálculo de márgenes.
Perfil Operativo – Comercial	Rentabilidad histórica por cliente y tipo de proyecto, condiciones comerciales y descuentos aplicados en cotizaciones anteriores.
Perfil Operativo – Técnico	Horas trabajadas y materiales utilizados por proyecto, frente a los recursos asignados originalmente.

Nota. Elaboración propia.

4.2.4. Nivel de acceso y herramienta de consumo por perfil

Dado que Directech no cuenta actualmente con una función formal de gobierno de datos (DAMA International, 2017), se propone que la Gerencia General, en conjunto con la Coordinación de Proyectos, defina y autorice los niveles de acceso descritos a continuación, evitando que el crecimiento de la información derive en los mismos problemas de dispersión que hoy afectan a la organización:

Tabla 18

Nivel de acceso y herramienta de consumo por perfil

Perfil	Nivel de acceso al Data Warehouse	Herramienta de consumo
Perfil Estratégico	Visión consolidada de todas las dimensiones y métricas, sin restricciones por proyecto o cliente.	Tablero de control publicado en la nube (autoservicio), consultable desde computador o dispositivo móvil.
Perfil Táctico – Proyectos	Acceso al detalle de presupuesto vs. ejecución de todos los proyectos bajo su coordinación.	Tablero de control interactivo con filtros por proyecto y período.
Perfil Táctico – Financiero	Acceso a las dimensiones financieras y de costos; mantiene además su acceso operativo directo a Latinium.	Reportes financieros existentes complementados con el tablero de control consolidado.

Perfil	Nivel de acceso al Data Warehouse	Herramienta de consumo
Perfil Operativo – Comercial	Acceso a la dimensión de clientes y a la rentabilidad histórica asociada a sus propias cotizaciones.	Tablero de control comercial con filtro por cliente y tipo de proyecto.
Perfil Operativo – Técnico	Acceso a la dimensión de horas y materiales por proyecto en el que participa.	Vista simplificada de seguimiento técnico, sin acceso a cifras financieras completas.

Nota. Elaboración propia.

Esta diferenciación de perfiles y accesos garantiza que cada usuario de Directech reciba información relevante para su rol, evitando la sobrecarga de datos y protegiendo la confidencialidad de la información financiera sensible frente a perfiles que no la requieren para su función.

4.3. Conclusiones del capítulo

- La arquitectura seleccionada —un Data Warehouse relacional a escala PYME con modelo dimensional— es la opción proporcional al volumen, variedad y presupuesto de Directech Cía. Ltda., frente a alternativas como el data lake o el lakehouse, que resultarían sobredimensionadas para su operación actual.
- El esquema estrella propuesto, con una tabla de hechos de proyectos y cinco dimensiones (tiempo, cliente, proyecto, línea de negocio y área responsable), permite calcular directamente los KPI de rentabilidad y costos definidos previamente, sin depender de cálculos manuales dispersos entre Latinium y Excel.
- La identificación de cinco perfiles de usuario (estratégico, táctico-proyectos, táctico-financiero, operativo-comercial y operativo-técnico) evidencia que Directech no requiere una única vista de información, sino un conjunto diferenciado de accesos y niveles de detalle, lo que sienta las bases para el diseño posterior de los tableros de control específicos por perfil.
- Microsoft Power BI se recomienda como la herramienta de análisis y visualización que consumirá el Data Warehouse propuesto, por su bajo costo, su integración directa con el esquema estrella diseñado y su capacidad de publicar vistas diferenciadas de autoservicio para cada uno de los perfiles de usuario identificados, sin requerir personal especializado en programación.

En conjunto, estas decisiones dan respuesta al objetivo inicial planteado para el proyecto BI DIRECTECH: disponer de una herramienta que permita visualizar de manera clara y en tiempo real el margen de ganancia real de los proyectos, considerando la variabilidad de los costos. La arquitectura de Data Warehouse dimensional resuelve la consolidación de las fuentes hoy dispersas entre Latinium y Excel; el esquema estrella habilita el cálculo directo de los KPI de rentabilidad; y la diferenciación de perfiles garantiza que dicha información llegue, en el nivel de detalle adecuado, tanto a la Gerencia como a cada área operativa.

CAPÍTULO 5: Estructura del departamento de Business Intelligence

Este capítulo desarrolla la estructura organizativa necesaria para sostener el sistema de BI DIRECTECH una vez implementado, siguiendo la misma secuencia utilizada en el caso de referencia (Pérez Galán, 2026b): situación actual, carencias, coordinación de flujos, organigrama y solución de problemas, recursos externos y business case.

5.1. Identificación de la situación actual del departamento

Directech Cía. Ltda. no cuenta actualmente con un departamento, gerencia o rol formalmente dedicado a la analítica de datos o al Business Intelligence. La gestión de la información se encuentra repartida de manera informal entre distintos colaboradores, como complemento a sus funciones principales, sin que exista una descripción de cargo, un presupuesto asignado ni un responsable único de la calidad y disponibilidad de los datos.

El levantamiento realizado en capítulos anteriores permite identificar quién asume hoy, de facto, cada una de las funciones que en una organización de mayor escala correspondería a un área de analítica (Davenport & Harris, 2017):

Tabla 19

Roles y responsabilidades actuales relacionados con la información

Función analítica	Quién la asume hoy	Situación actual
Consolidación de reportes financieros	Contador (Ing. José Ortega)	Genera reportes desde Latinium; la consolidación con datos de proyectos es manual.
Seguimiento de presupuesto vs. ejecución	Coordinación de Proyectos (Lic. Leslie Soria)	Comparación manual entre Excel de presupuesto y registros de costos reales.
Registro de datos técnicos (horas, materiales)	Área Técnica	Registro interno, no siempre vinculado al análisis financiero del proyecto.
Visión consolidada de rentabilidad	Gerencia General	Depende de reportes puntuales solicitados a otras áreas; no existe un tablero propio.
Administración de herramientas tecnológicas	Sin responsable definido	El soporte de Latinium se gestiona por contrato con el proveedor; no hay gestión de un Data Warehouse ni de Power BI.

Nota. Elaboración propia.

Esta situación es consistente con el diagnóstico realizado en capítulos previos: la información existe y es de buena calidad relativa, pero su gestión depende de personas específicas actuando de manera individual, sin coordinación formal ni continuidad garantizada si alguna de ellas se desvincula de la empresa. A diferencia del caso de referencia, donde ya existía una dirección de mayor jerarquía sobre la cual se construyó la gerencia analítica, en Directech la función de BI debe crearse desde cero, aprovechando al máximo los roles y el conocimiento organizacional ya disponibles, en particular el de la Coordinadora de Proyectos, quien impulsa el proyecto y cuenta con un nivel de conocimiento organizacional estimado en 9/10.

5.2. Definición de las carencias en formación y recursos

A partir de la situación actual descrita, se identifican las siguientes carencias que deben cerrarse para que la función de BI sea sostenible una vez concluida la fase de implementación del proyecto de especialización:

Tabla 20

Carencias en formación y recursos identificadas

Categoría	Carencia identificada	Efecto sobre el proyecto
Formación técnica	Ningún colaborador tiene experiencia previa en Power BI, modelado dimensional o lenguaje DAX.	Riesgo de dependencia total de un proveedor externo para el mantenimiento del sistema.
Formación en gobierno de datos	No existen políticas de calidad, actualización ni acceso a la información (DAMA International, 2017).	Riesgo de que la dispersión actual de los datos se replique dentro del propio Data Warehouse.
Recursos de personal	No existe un rol, ni siquiera a tiempo parcial, dedicado a la administración del sistema de BI.	El mantenimiento recaería de manera informal sobre la Coordinación de Proyectos, generando sobrecarga de funciones.
Recursos tecnológicos	No se dispone de licenciamiento de Power BI Pro ni de un repositorio central de datos.	Es necesario presupuestar el licenciamiento y la puesta en marcha del Data Warehouse.
Recursos financieros	No existe una partida presupuestal específica para analítica o Business Intelligence.	El proyecto requiere ser sustentado mediante un business case (ver sección 5.6).

Nota. Elaboración propia.

Estas carencias son de menor magnitud que las que enfrentaba la organización del caso de referencia antes de crear su gerencia analítica, dado que Directech no requiere modelos predictivos avanzados ni procesamiento de Big Data, sino la consolidación de dos fuentes estructuradas (Latinium y Excel). Sin embargo, la ausencia total de formación previa en herramientas de BI implica que la fase de implementación debe incluir necesariamente un componente de capacitación, y no únicamente de despliegue tecnológico.

5.3. Coordinación de los flujos entre otras áreas

El sistema de BI DIRECTECH depende de que cada área continúe generando su información con la calidad y periodicidad ya identificadas en el Capítulo 2 (fuentes internas), y de que exista un punto de coordinación que consolide dichos flujos hacia el Data Warehouse. Se propone la siguiente matriz de coordinación, que formaliza quién entrega qué información, con qué frecuencia y quién es responsable de validarla antes de su carga:

Tabla 21

Matriz de flujos y coordinación entre áreas

Área	Información que entrega	Frecuencia	Validador
Ventas	Cotizaciones, historial de clientes, condiciones comerciales	Diaria	Coordinación de BI

Área	Información que entrega	Frecuencia	Validador
Proyectos	Presupuesto, cronograma, avance, costos reales	Semanal	Coordinación de BI
Área Técnica	Horas trabajadas, materiales utilizados	Diaria	Coordinación de BI
Administrativa y Financiera	Facturación, costos, compras, flujo de caja (vía Latinium)	Diaria (carga) / Semanal (validación)	Contador
Gerencia General	Aprobación de proyectos, prioridades estratégicas	Según necesidad	Coordinación de BI

Nota. Elaboración propia.

A diferencia del modelo del caso de referencia, donde la coordinación entre gerencias se sostiene mediante secciones especializadas (Servicios de Inteligencia, Transformación de la Información, Automatización), en Directech esta coordinación se concentra en un único punto de contacto: la Coordinación de Business Intelligence. Esta figura es responsable de tres tareas frente a las demás áreas:

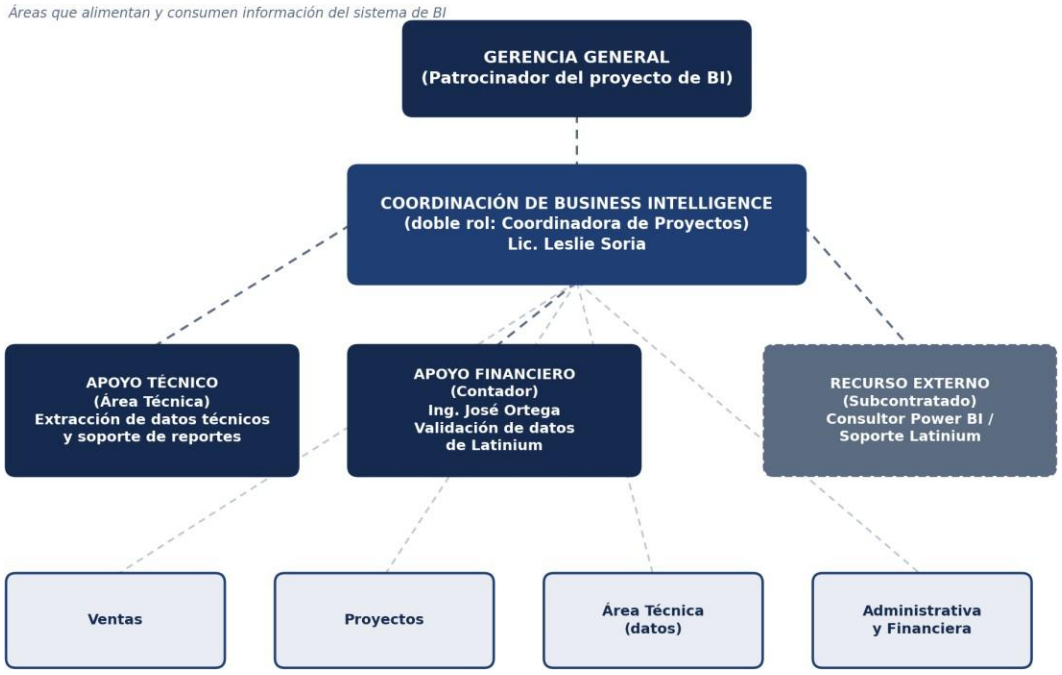
- Recordar y verificar que cada área entregue su información con la periodicidad acordada en la matriz de flujos.
- Resolver inconsistencias de primer nivel (por ejemplo, un proyecto sin código asignado o un costo registrado en un período incorrecto) directamente con el responsable del área.
- Comunicar a la Gerencia General cualquier brecha de información que no pueda resolverse a este nivel, siguiendo el mecanismo de escalamiento descrito en la sección 5.4.

5.4. Organigrama, dependencias de otras áreas y solución de problemas

Se propone una estructura ligera, proporcional al tamaño de Directech, que evita la creación de una gerencia independiente —inviable para una empresa de diez colaboradores— y en su lugar aprovecha roles ya existentes mediante la figura de doble rol, complementada con un recurso externo subcontratado durante la fase de implementación y mantenimiento especializado.

Figura 5

Estructura organizativa propuesta para la función de BI en Directech



Nota. Elaboración propia.

Tabla 22

Roles propuestos para la función de BI en Directech

Rol	Quién lo asume	Responsabilidad principal	Dedicación
Patrocinador (Gerencia General)	Ing. Guido Soria / Ing. Juan Soria	Aprobar el proyecto, priorizar recursos y usar el tablero consolidado para decisiones estratégicas.	Según necesidad
Coordinación de BI (doble rol)	Lic. Leslie Soria (Coordinadora de Proyectos)	Administrar el Data Warehouse y Power BI, coordinar la carga de datos y dar mantenimiento a los tableros.	Parcial (estimado 20 % de la jornada)
Apoyo técnico	Área Técnica	Verificar y entregar los datos de horas y materiales; apoyar en la extracción de reportes técnicos.	Puntual, dentro de sus funciones actuales
Apoyo financiero	Contador (Ing. José Ortega)	Validar la exportación de datos desde Latinium antes de su carga al Data Warehouse.	Puntual, dentro de sus funciones actuales
Recurso externo	Consultor / proveedor especializado	Implementar el modelo dimensional, configurar Power BI y capacitar a la Coordinación de BI.	Por contrato (fase de implementación y soporte anual)

Nota. Elaboración propia.

Esta estructura formaliza, sin generar un costo fijo de nómina adicional, un punto único de responsabilidad sobre el sistema de BI. El doble rol asignado a la Coordinación de Proyectos es consistente con el vínculo que ya existe entre el equipo del proyecto y Directech, y con el nivel de conocimiento organizacional (9/10) reportado en la introducción del proyecto.

5.4.1. Mecanismo de solución de problemas

Se define un esquema de escalamiento en tres niveles, análogo en su lógica al que sostiene la gerencia analítica del caso de referencia, pero adaptado a una estructura sin subgerencias intermedias:

Tabla 23

Mecanismo de escalamiento y solución de problemas

Nivel	Tipo de problema	Responsable de resolverlo	Tiempo esperado de respuesta
1	Dato faltante, incompleto o con formato incorrecto en la carga semanal.	Responsable del área que generó el dato, coordinado por la Coordinación de BI.	1 a 2 días hábiles
2	Inconsistencia entre fuentes (por ejemplo, un costo que no coincide entre Excel y Latinium).	Coordinación de BI, en conjunto con el Contador.	3 a 5 días hábiles
3	Falla técnica del modelo, del proceso ETL o de la herramienta Power BI.	Recurso externo (consultor / soporte especializado), convocado por la Coordinación de BI.	Según acuerdo de soporte contratado

Nota. Elaboración propia.

Cualquier problema que no pueda resolverse en el Nivel 2 debe ser comunicado a la Gerencia General, quien decide si amerita la intervención del recurso externo contratado. Este esquema evita que Directech dependa de una sola persona para la continuidad del sistema, distribuyendo la responsabilidad de acuerdo con la naturaleza del problema.

5.5. Recursos externos destinados: subcontratación

Directech no cuenta con un departamento de TI propio, y el soporte tecnológico recae actualmente en el Área Técnica y en el contrato de mantenimiento anual con Latinium. Dado que ningún colaborador tiene experiencia previa en Power BI ni en modelado dimensional, se recomienda subcontratar de manera puntual los servicios que requieren especialización, reservando para el personal interno las tareas de mantenimiento rutinario una vez completada la capacitación.

Tabla 24

Evaluación make vs. buy de los servicios requeridos

Servicio requerido	¿Interno o subcontratado?	Justificación
Diseño e implementación inicial del modelo dimensional (esquema estrella)	Subcontratado	Requiere experiencia específica en modelado de datos que hoy no existe en la organización.
Configuración inicial de Power BI y del proceso ETL	Subcontratado	Tarea de una sola vez, de alta especialización; no justifica contratar personal permanente.
Capacitación de la Coordinación de BI y del Área Técnica en Power BI	Subcontratado (una vez) + práctica interna	Es la vía para transferir capacidad a la organización y reducir la dependencia externa en el mediano plazo.
Carga y actualización periódica de datos	Interno (Coordinación de BI)	Es una tarea recurrente y de bajo riesgo técnico, adecuada para el personal ya vinculado.
Mantenimiento correctivo y soporte ante fallas técnicas mayores	Subcontratado (bajo demanda)	Directech no requiere un contrato de soporte permanente dado el bajo volumen de datos; conviene un esquema de bolsa de horas.
Soporte del sistema contable Latinium	Subcontratado (ya contratado)	Contrato de mantenimiento anual ya existente con el proveedor del software.

Nota. Elaboración propia.

Esta combinación de recursos —subcontratación puntual para la implementación y capacitación, y ejecución interna para la operación recurrente— es la opción proporcional al tamaño de Directech. Difiere del modelo del caso de referencia, donde la magnitud del volumen de datos y del presupuesto justifica un equipo interno permanente de más de diez personas; en Directech, un equipo de esa naturaleza no sería sostenible frente a un volumen de ingresos de aproximadamente USD 900.000 anuales.

5.6. Business case de costes y beneficios

Siguiendo la misma lógica utilizada en el caso de referencia —presentar el costo directo de constituir la función analítica frente a los beneficios esperados—, se estima a continuación el presupuesto de implementación y los beneficios cuantificables para Directech durante el primer año.

5.6.1. Presupuesto estimado de implementación (año 1)

Tabla 25

Presupuesto estimado de implementación (año 1)

Categoría	Concepto	Costo estimado (USD)
Costos directos – Servicios externos	Diseño e implementación del modelo dimensional y del proceso ETL (pago único)	1.800 – 2.500
Costos directos – Servicios externos	Capacitación inicial en Power BI (Coordinación de BI y Área Técnica)	600 – 900

Categoría	Concepto	Costo estimado (USD)
Costos directos – Servicios externos	Bolsa de horas de soporte y mantenimiento correctivo (anual)	500 – 800
Costos directos – Licenciamiento	Power BI Pro (5 licencias × USD 14/mes × 12 meses)	840
Costos directos – Personal interno	Tiempo dedicado por la Coordinación de BI (costo de oportunidad, 20 % de la jornada durante 3 meses de implementación)	900 – 1.200
Total estimado – año 1		4.640 – 6.240

Nota. Elaboración propia. Valores referenciales sujetos a cotización.

A partir del segundo año, el costo recurrente se reduce principalmente al licenciamiento de Power BI (aproximadamente USD 840 anuales, según el precio vigente de Power BI Pro de USD 14 por usuario al mes desde abril de 2025; Microsoft, 2026a) y a una bolsa de soporte externo de menor alcance, una vez que la Coordinación de BI haya asumido la operación cotidiana del sistema.

5.6.2. Beneficio 1: reducción de tiempo en la consolidación manual de reportes

Actualmente, el Contador y la Coordinación de Proyectos dedican tiempo de forma manual a comparar presupuestos con costos reales y a preparar la información que la Gerencia solicita de manera puntual. Se estima el siguiente ahorro, siguiendo el mismo formato de cálculo utilizado en el caso de referencia:

Tabla 26

Ahorro estimado en consolidación manual de reportes

Concepto	Sin sistema de BI (hoy)	Con sistema de BI
Costo hora hombre (promedio)	USD 12	USD 12
Horas invertidas al mes en consolidación manual	16	5
Costo mensual	USD 192	USD 60
Costo anual	USD 2.304	USD 720
Ahorro estimado anual	USD 1.584	

Nota. Elaboración propia.

5.6.3. Beneficio 2: reducción de sobrecostos en proyectos

La identificación de GAP del Capítulo 1 determinó una variación de costos actual superior al +20 % frente al presupuesto (sobrecosto), frente a una meta de ≤ 5 %. Aun si el sistema de BI únicamente contribuye a cerrar una parte de esta brecha —al permitir detectar desviaciones de manera oportuna en lugar de al cierre del proyecto—, el efecto sobre la rentabilidad es significativo dado el volumen de costos que maneja la empresa:

Tabla 27
Beneficio estimado por reducción de sobrecostos

Concepto	Valor estimado
Ingresos anuales de Directech	USD 900.000
Costos totales anuales estimados (margen neto actual ~10 %)	USD 810.000
Reducción conservadora de sobrecostos atribuible a la detección oportuna de desviaciones (3 % de los costos totales)	USD 24.300
Beneficio anual estimado (Beneficio 1 + Beneficio 2)	USD 25.884

Nota. Elaboración propia.

Comparando el beneficio anual estimado (aproximadamente USD 25.884) frente al costo de implementación en el primer año (entre USD 4.640 y USD 6.240), el proyecto presenta una relación beneficio-costo favorable ya desde su primer año de operación, incluso bajo el supuesto conservador de que el sistema de BI solo contribuye a cerrar una fracción de la brecha de sobrecostos identificada. Esta cifra no incluye beneficios adicionales de carácter cualitativo, como la mejora en la oportunidad de las decisiones comerciales o la reducción del riesgo de pérdida de información ante la rotación de personal.

Es importante señalar, en línea con la conclusión del caso de referencia, que estos beneficios solo se materializan si el sistema de BI se implementa y se utiliza correctamente: una interpretación errónea de los datos, o el abandono del mantenimiento una vez concluida la fase de implementación, puede llevar a decisiones equivocadas en lugar de a las mejoras aquí proyectadas.

5.7. Conclusiones del capítulo

- Directech no cuenta hoy con una función de analítica formalmente constituida; las tareas que en el caso de referencia asume una gerencia analítica completa están repartidas de manera informal entre la Contabilidad, la Coordinación de Proyectos y el Área Técnica, sin un responsable único ni continuidad garantizada.
- La estructura propuesta —una Coordinación de BI en doble rol, apoyada por el Área Técnica, el Contador y un recurso externo subcontratado— es proporcional al tamaño de Directech (aproximadamente diez colaboradores) y evita el costo de una gerencia independiente que no sería sostenible con el volumen de ingresos actual de la empresa.
- La matriz de coordinación de flujos y el mecanismo de escalamiento en tres niveles formalizan responsabilidades que hoy dependen de acuerdos informales entre colaboradores, reduciendo el riesgo de que el sistema de BI pierda continuidad.
- La combinación de subcontratación puntual (implementación, capacitación y soporte especializado) con operación interna recurrente (carga y mantenimiento diario) permite a Directech acceder a capacidades que hoy no posee, sin incurrir en el costo fijo de un equipo de TI propio.
- El business case muestra que el beneficio anual estimado del sistema de BI (cercano a USD 25.900, entre ahorro de tiempo y reducción conservadora de sobrecostos) supera ampliamente el costo de implementación en el primer año (entre USD 4.640 y USD

6.240), lo que sustenta financieramente la creación de esta función dentro de la organización.

- Al igual que en el caso de referencia, el éxito de esta estructura depende de que la información se interprete correctamente y de que el mantenimiento del sistema no se abandone una vez finalizada la fase de implementación.

PARTE II: CONSTRUCCIÓN, ARQUITECTURA CLOUD E IMPLEMENTACIÓN DEL MODELO DE BI

CAPÍTULO 1: Modelo de Business Intelligence integrado EDUTECH – DIRECTECH e impacto potencial de Big Data

1.1. Transición del Bloque 1 al Bloque 2

La Parte I de este trabajo diagnosticó la situación informacional de Directech Cía. Ltda. y diseñó, a nivel conceptual y organizativo, su sistema de Business Intelligence: KPI de rentabilidad, fuentes, arquitectura dimensional, perfiles de usuario y estructura del departamento. La Parte II avanza desde el diseño hacia la construcción y validación técnica del modelo. Para ello, el alcance se amplía a dos empresas del sector tecnológico, EDUTECH y DIRECTECH: empresas vinculadas, con los mismos propietarios, que operan de forma complementaria en el mismo ciclo comercial. Directech participa, tanto en el sector público como en el privado, en la etapa de proformas y cotizaciones; cuando corresponde presentar la oferta formal, en particular en los procesos de contratación pública del Ecuador, esta se realiza a través de EDUTECH, y una vez concluido el proceso de subasta, es también EDUTECH la que factura al cliente. En consecuencia, una misma oportunidad de negocio puede originarse en Directech y registrar sus ingresos en EDUTECH, de modo que el margen real de un proyecto solo puede medirse si se integra la información de ambas empresas en un único modelo. Por esta razón, el Bloque 2 integra a ambas empresas.

Con el fin de preservar la confidencialidad de la información real de Directech, cuyo acceso se rige por los lineamientos de la organización, en este bloque se utiliza una matriz de datos ficticios (sintéticos) que reproduce la lógica operativa de ambas empresas. De esta manera, el modelo puede construirse, cargarse y probarse de extremo a extremo sin exponer información sensible, y posteriormente puede alimentarse con los datos reales exportados desde Latinium y Excel descritos en la Parte I.

1.2. Objetivos del modelo integrado

El objetivo general de esta etapa es diseñar un modelo de Business Intelligence para EDUTECH y DIRECTECH que integre información comercial, financiera, operativa y geográfica mediante procesos de extracción, transformación y modelado de datos, con el propósito de generar indicadores estratégicos que faciliten la toma de decisiones gerenciales e identifiquen tendencias, oportunidades comerciales, niveles de rentabilidad y áreas de mejora organizacional.

Los objetivos específicos son:

- Identificar las principales fuentes de datos internas y externas que pueden ser utilizadas en el modelo de BI.
- Definir las relaciones entre ventas, compras, clientes, productos, proveedores, empresas, territorio y resultados financieros.
- Establecer casos de uso que permitan convertir los datos en información útil para la gestión gerencial.
- Analizar el impacto potencial de Big Data en las áreas, departamentos y procesos de negocio de las empresas.

- Plantear una arquitectura escalable que permita evolucionar desde análisis descriptivos hacia capacidades predictivas y prescriptivas.

1.3. Selección e identificación de fuentes de datos

Las fuentes de información se clasifican de acuerdo con su procedencia (interna o externa) y por su naturaleza (estructurada o semiestructurada). Las fuentes internas constituyen el núcleo transaccional del modelo, mientras que las externas permiten enriquecer el análisis y contextualizar el desempeño empresarial.

Tabla 28

Fuentes de datos del modelo integrado EDUTECH – DIRECTECH

Fuente	Procedencia	Tipo	Datos principales	Uso en BI
Ventas	Interna	Estructurada	Cliente, producto, cantidad, precio, fecha y empresa	Análisis comercial y tendencias
Compras	Interna	Estructurada	Proveedor, producto, cantidad, costo y fecha	Análisis de abastecimiento y costos
Devoluciones	Interna	Estructurada	Producto, motivo, cantidad, valor y documento asociado	Control de calidad y eficiencia comercial
Clientes	Interna	Estructurada	Tipo, ubicación, canal y estado	Segmentación y comportamiento
Productos	Interna	Estructurada	Categoría, marca, costo y precio	Rentabilidad, mix y portafolio
Proveedores	Interna	Estructurada	Categoría, ubicación y condición de pago	Gestión de proveedores
Estado de resultados	Interna	Estructurada	Ingresos, costos, gastos y utilidad	Desempeño financiero
Balance	Interna	Estructurada	Activos, pasivos y patrimonio	Liquidez, solvencia y endeudamiento
Información geográfica del Ecuador	Externa	Estructurada	Provincia, ciudad y región	Análisis territorial
Indicadores económicos	Externa	Estructurada	Inflación, actividad económica y otras variables	Contextualización de resultados
Datos abiertos	Externa	Estructurada / semiestructurada	Población, territorio y actividad económica	Análisis de potencial de mercado

Nota. Elaboración propia.

1.3.1. Arquitectura de integración

La solución se estructura en capas para separar las fuentes operativas del procesamiento analítico. Esta separación mejora la trazabilidad de las transformaciones y facilita futuras ampliaciones del modelo.

Tabla 29
Capas de la arquitectura de integración

Capa	Contenido	Función
1. Fuentes	Excel, PostgreSQL, datos geográficos y fuentes públicas	Capturar información interna y externa.
2. ETL / Power Query	Extracción, limpieza, estandarización y transformación	Corregir formatos, depurar registros y homologar claves.
3. Modelo dimensional	Tablas de hechos y dimensiones	Organizar la información bajo un esquema analítico.
4. Power BI / DAX	Medidas, KPI, segmentadores y visualizaciones	Transformar datos en información gerencial.
5. Gestión	Dashboards y análisis ejecutivo	Apoyar decisiones tácticas y estratégicas.

Nota. Elaboración propia.

El flujo conceptual resultante es: fuentes internas y externas → ETL → modelo dimensional → Power BI → indicadores y visualizaciones → toma de decisiones.

1.4. Relación de los datos

El modelo utiliza una arquitectura dimensional tipo estrella (Kimball & Ross, 2013). Las tablas de hechos almacenan los eventos o movimientos cuantificables, mientras que las dimensiones proporcionan el contexto necesario para analizar dichos eventos desde diferentes perspectivas.

1.4.1. Tablas de hechos

Tabla 30
Tablas de hechos del modelo integrado

Tabla	Contenido	Métricas principales
Fact_Ventas	Transacciones de ventas de EDUTECH y DIRECTECH	Cantidad, precio, subtotal, IVA, total y costo
Fact_Compras	Adquisiciones realizadas a proveedores	Cantidad, costo unitario, subtotal y total
Fact_Devoluciones	Devoluciones de ventas y compras	Cantidad, valor y motivo
Estado_Resultados	Resultados financieros mensuales	Ingresos, costos, gastos, utilidad operativa y utilidad neta
Balance_Mensual	Situación financiera mensual	Activos, pasivos y patrimonio

Nota. Elaboración propia.

1.4.2. Dimensiones principales

- **Dim_Fecha:** permite análisis por día, mes, trimestre y año.
- **Dim_Empresa:** distingue las operaciones de EDUTECH y DIRECTECH.
- **Dim_Producto:** clasifica productos por categoría, subcategoría, marca, costo y precio.

- **Dim_Cliente:** habilita análisis por tipo de cliente, canal, provincia, ciudad y región.
- **Dim_Proveedor:** permite evaluar compras, condiciones y concentración por proveedor.
- **Dim_Vendedor:** facilita medir el desempeño de ejecutivos y zonas comerciales.

1.4.3. Relaciones y preguntas de negocio

Las relaciones permiten pasar de un análisis aislado a una visión integrada. Por ejemplo, una venta puede asociarse con un cliente, producto, vendedor, empresa y fecha. El costo del mismo producto permite determinar el margen; la ubicación del cliente permite analizar el desempeño territorial; y las devoluciones ayudan a ajustar la interpretación de ventas y calidad del servicio.

Tabla 31

Preguntas gerenciales y relaciones del modelo

Pregunta gerencial	Relaciones utilizadas	Resultado esperado
¿Qué empresa genera mayor rentabilidad?	Ventas + productos + costos + devoluciones + resultados	Comparación de margen y utilidad por empresa
¿Qué productos aportan más utilidad?	Ventas + producto + costo	Ranking por margen y contribución
¿En qué provincias se concentra el negocio?	Clientes + provincia + ventas	Mapa y participación territorial
¿Qué proveedores tienen mayor peso?	Compras + proveedor + producto	Concentración y dependencia de abastecimiento
¿Qué clientes son estratégicos?	Ventas + cliente + fecha	Frecuencia, valor y recurrencia de compra

Nota. Elaboración propia.

1.5. Casos de uso del modelo de Business Intelligence

El modelo se orienta a resolver necesidades concretas de gestión. Cada caso de uso relaciona una necesidad empresarial con indicadores y decisiones potenciales.

Tabla 32

Casos de uso del modelo de BI

Caso de uso	Información	Indicadores	Decisión gerencial
Desempeño comercial	Ventas, clientes, productos y fechas	Ventas totales, crecimiento, ticket promedio, unidades	Definir estrategias comerciales y metas
Rentabilidad	Ventas, costos, devoluciones y productos	Margen bruto, utilidad, margen %	Priorizar productos y categorías rentables
Análisis geográfico	Clientes, ventas y provincias	Ventas por provincia, margen territorial, clientes	Detectar mercados y oportunidades de expansión
Gestión de compras	Compras, proveedores y productos	Compras, costo promedio, variación de costo	Negociar y diversificar proveedores
Gestión de clientes	Clientes, ventas y fechas	Frecuencia, valor, recurrencia, participación	Segmentar, fidelizar y recuperar clientes

Caso de uso	Información	Indicadores	Decisión gerencial
Gestión financiera	Balance y estado de resultados	Liquidez, endeudamiento, margen, ROA y ROE	Evaluar salud financiera y asignación de recursos

Nota. Elaboración propia.

1.5.1. Casos de uso prioritarios

Análisis comercial. La gerencia podrá observar la evolución mensual de ventas, comparar el desempeño entre EDUTECH y DIRECTECH, identificar productos y clientes de mayor contribución y analizar el comportamiento por provincia, canal y vendedor. Esto permite transformar una visión transaccional en una lectura de desempeño comercial.

Análisis de rentabilidad. El volumen de ventas por sí solo no determina el éxito de un producto. El modelo integra precio, cantidad, costo y devoluciones para estimar la contribución económica. Esta perspectiva permite identificar productos de alta rotación pero bajo margen, así como productos de menor volumen con alta rentabilidad. Este caso de uso conecta directamente con el objetivo central definido en la Parte I: conocer el margen real de cada proyecto, producto o servicio.

Análisis territorial. La dimensión geográfica permitirá visualizar el comportamiento por provincia y ciudad mediante mapas en Power BI. La concentración de ventas en Pichincha puede compararse con el crecimiento o margen de otras provincias, facilitando decisiones de expansión, cobertura y asignación comercial.

Segmentación de clientes. Como línea de evolución del modelo se plantea una clasificación RFM (recencia, frecuencia y valor monetario) para identificar clientes estratégicos, frecuentes, nuevos, en riesgo o inactivos. Este análisis conecta directamente el BI con estrategias de fidelización y gestión comercial.

1.6. Impacto de Big Data en el proyecto

El proyecto actual corresponde principalmente a una solución de Business Intelligence y no debe calificarse automáticamente como Big Data por el número de registros utilizados. El volumen disponible puede ser procesado eficientemente con tecnologías tradicionales como PostgreSQL, Power Query y Power BI. No obstante, la arquitectura propuesta puede evolucionar hacia escenarios de Big Data si las empresas incorporan nuevas fuentes con mayor volumen, velocidad y variedad (Marr, 2015; McAfee & Brynjolfsson, 2012).

1.6.1. Fuentes de datos para la evolución del modelo

- Registros de navegación y comportamiento en comercio electrónico.
- Interacciones en redes sociales y campañas digitales.
- Historial detallado de CRM, cotizaciones y oportunidades comerciales.
- Información de inventario y logística en tiempo real.
- Datos de servicio posventa, tickets y reclamos.
- Históricos externos de precios y disponibilidad de productos.

- Información de dispositivos o equipos conectados, cuando aplique.

1.6.2. Impacto por área y proceso

Tabla 33
Impacto potencial de Big Data por área

Área	Aplicación potencial	Beneficio esperado
Gerencia general	Integración de fuentes internas y externas	Decisiones basadas en información consolidada
Comercial	Analítica de comportamiento y propensión de compra	Mayor conversión y priorización de oportunidades
Marketing	Datos de campañas y canales digitales	Optimización de inversión y segmentación
Compras	Históricos de precios y desempeño de proveedores	Reducción de costos y menor dependencia
Finanzas	Series históricas y modelos predictivos	Mejor planificación y gestión de riesgo
Inventario	Pronóstico de demanda y rotación	Reducción de sobrestock y quiebres
Logística	Seguimiento de entregas y rutas	Mayor eficiencia operacional
Servicio al cliente	Análisis de tickets e interacciones	Mejora de satisfacción y tiempos de respuesta

Nota. Elaboración propia.

1.6.3. Evolución de la madurez analítica

Tabla 34
Niveles de madurez analítica aplicados al modelo

Nivel	Pregunta que responde	Ejemplo aplicado
Descriptivo	¿Qué ocurrió?	Ventas, compras, utilidad y clientes del periodo.
Diagnóstico	¿Por qué ocurrió?	Identificación de productos, territorios o costos que explican una variación.
Predictivo	¿Qué podría ocurrir?	Pronóstico de ventas, demanda o comportamiento del cliente.
Prescriptivo	¿Qué debería hacerse?	Recomendación de productos, mercados, inventario o acciones comerciales.

Nota. Elaboración propia con base en Sharda et al. (2020).

1.7. Rol de PostgreSQL dentro de la solución

PostgreSQL desempeña el papel de Data Warehouse relacional del proyecto: almacena tablas de hechos y dimensiones depuradas (esquema BI), mantiene el histórico necesario para los KPI y recibe las salidas de los modelos de inteligencia artificial (PostgreSQL Global Development Group, 2026). Microsoft Fabric/OneLake conserva los archivos de Landing y RAW y permite realizar transformaciones o entrenamiento; el Lakehouse es una zona de preparación y cálculo, no el destino definitivo de los resultados analíticos. Power BI consume preferentemente el modelo del Data Warehouse para que las cifras históricas y las predicciones tengan un origen gobernado y trazable.

- Centralización de datos de EDUTECH y DIRECTECH.
- Integridad mediante claves primarias y foráneas.
- Consultas SQL para validación y extracción.
- Conexión directa desde Power BI.
- Posibilidad de automatizar cargas futuras y ampliar el volumen de información.

1.8. Evidencias técnicas de implementación

Las evidencias técnicas demuestran la implementación del modelo, la estructura de la base de datos, el proceso de transformación y la conexión con Power BI.

Tabla 35

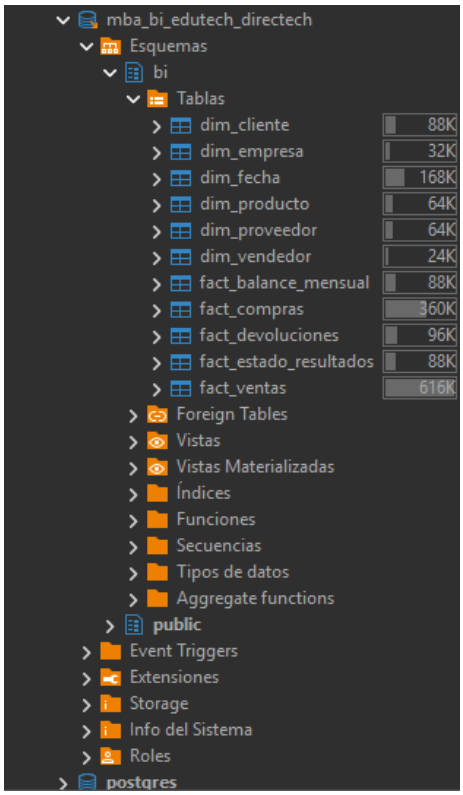
Evidencias técnicas de la implementación

Evidencia	Captura / evidencia	Qué demuestra
E1	PostgreSQL / DBeaver con las tablas cargadas	Persistencia y organización de la información
E2	Esquema de tablas	Nombre y columnas de las tablas
E3	Diagrama de relaciones	Modelo entidad-relación
E4	Validación de registros cargados	Conteo de datos cargados
E5	Consulta de datos	Sentencia SQL
E6	Conexión base de datos – Power BI	Extracción de datos
E7	Modelo relacional en Power BI	Modelo estrella y relaciones
E8	Tablero «Resumen Ejecutivo»	KPI consolidados para la Gerencia
E9	Tablero «Finanzas – Estado de Resultados»	Ingresos, costos, gastos y utilidad por empresa

Nota. Elaboración propia.

Figura 6

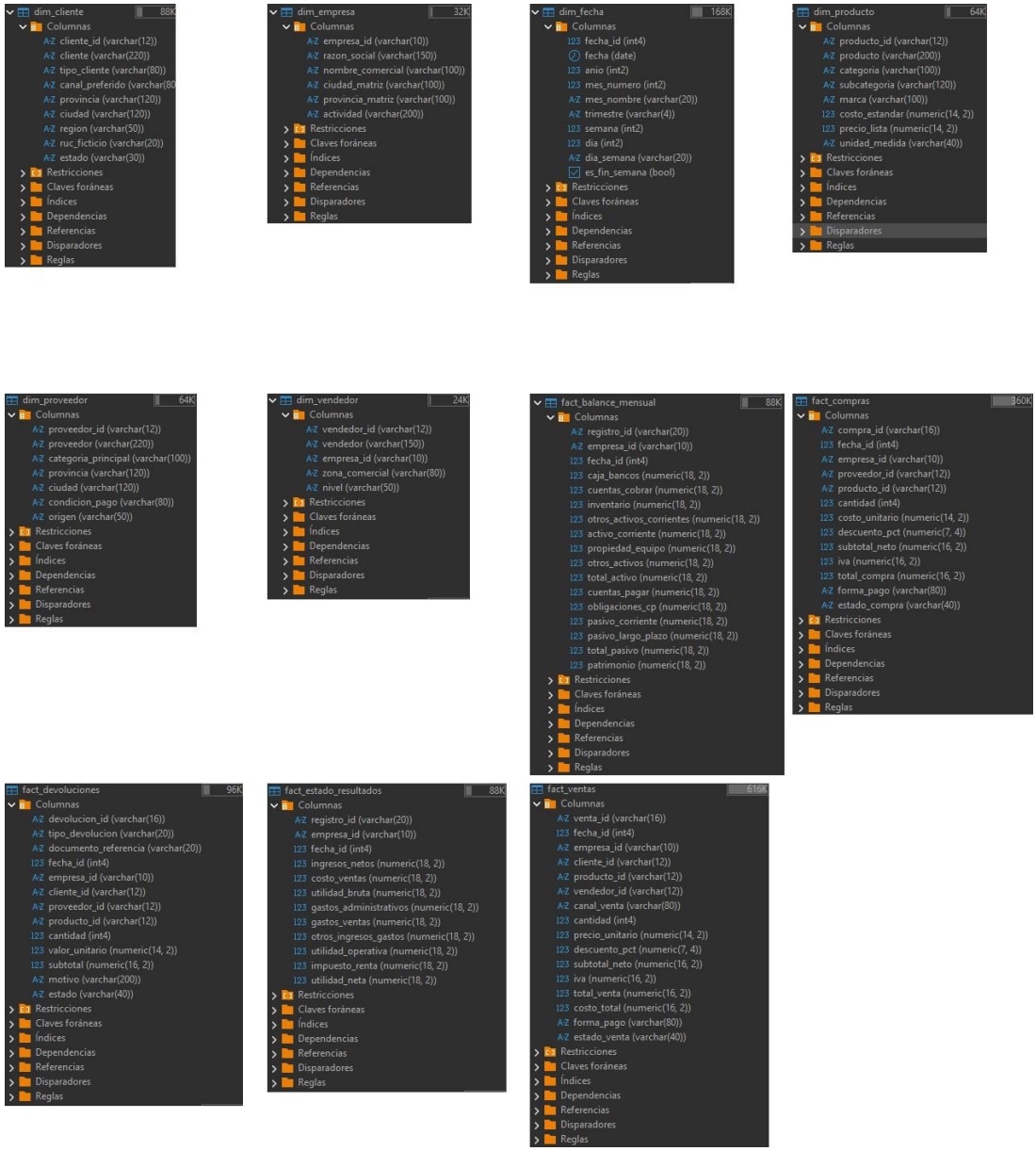
E1. Creación de la base de datos en PostgreSQL / DBeaver



Nota. Captura de la implementación del proyecto.

Figura 7

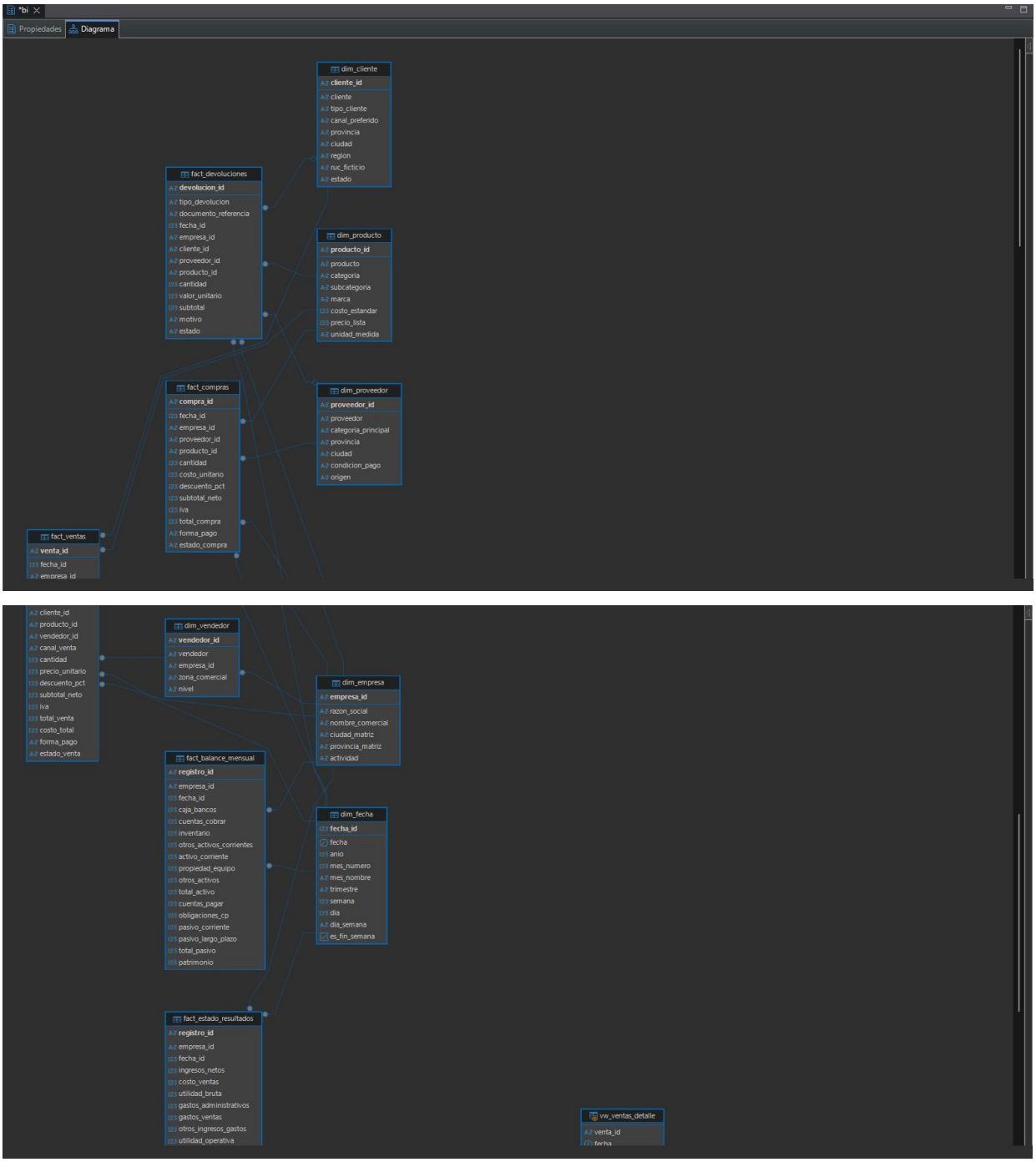
E2. Esquema BI y tablas creadas en PostgreSQL

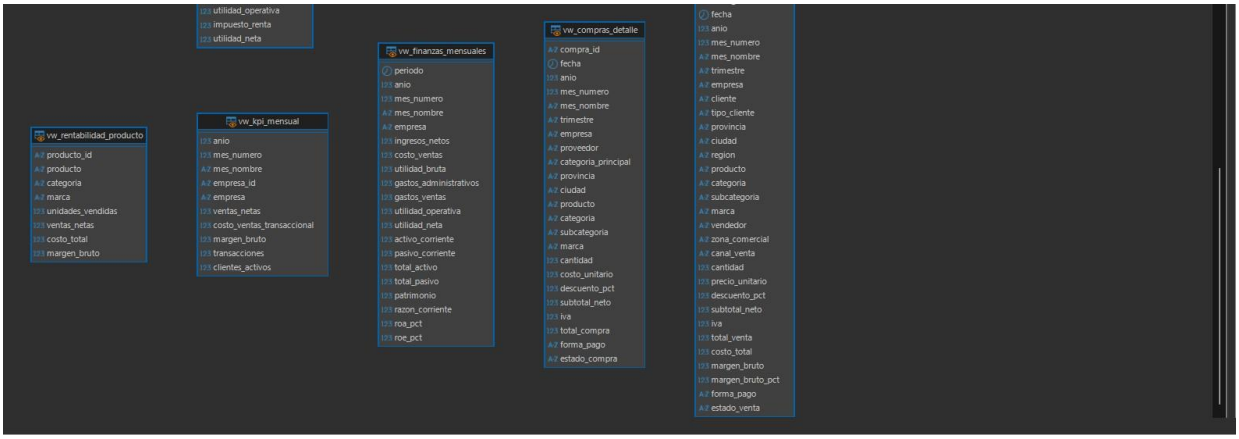


Nota. Captura de la implementación del proyecto.

Figura 8

E3. Diagrama entidad-relación del modelo

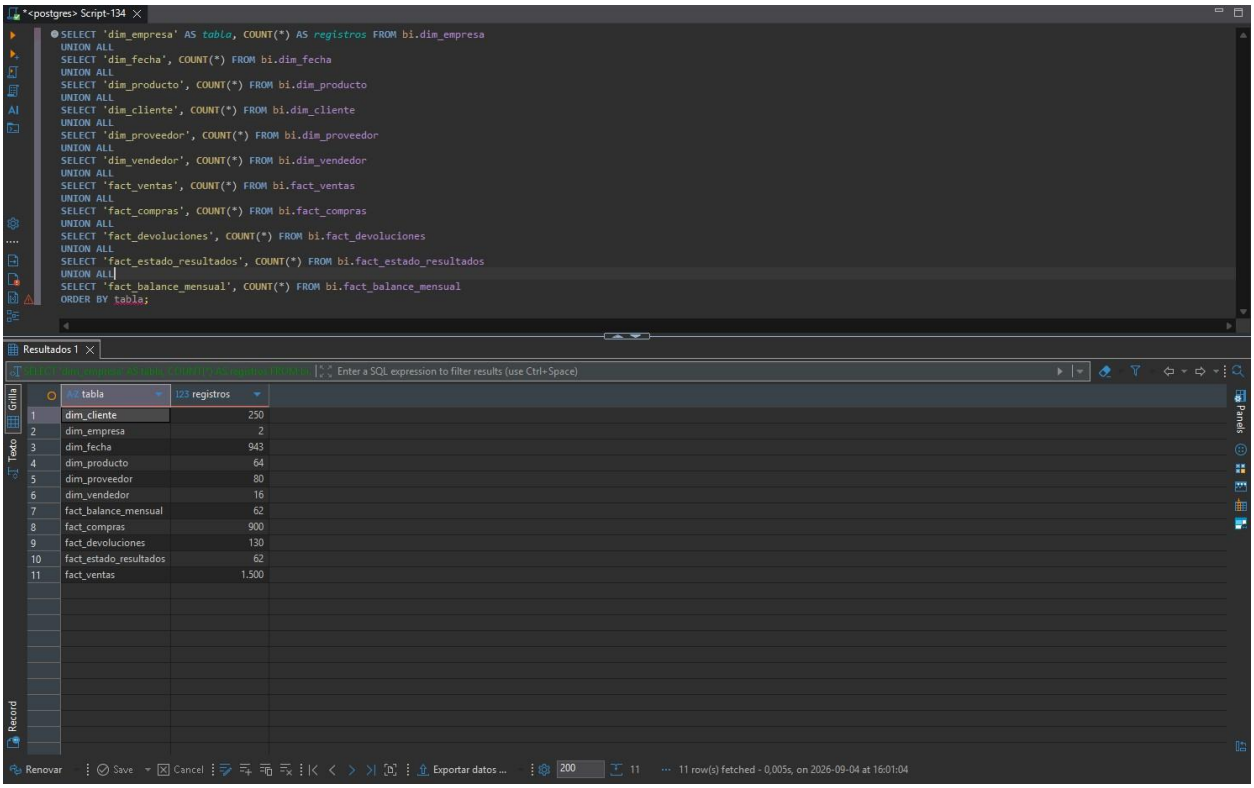




Nota. Captura de la implementación del proyecto.

Figura 9

E4. Validación de la carga de datos



Nota. Captura de la implementación del proyecto.

Tabla 36

Conteos esperados después de la carga

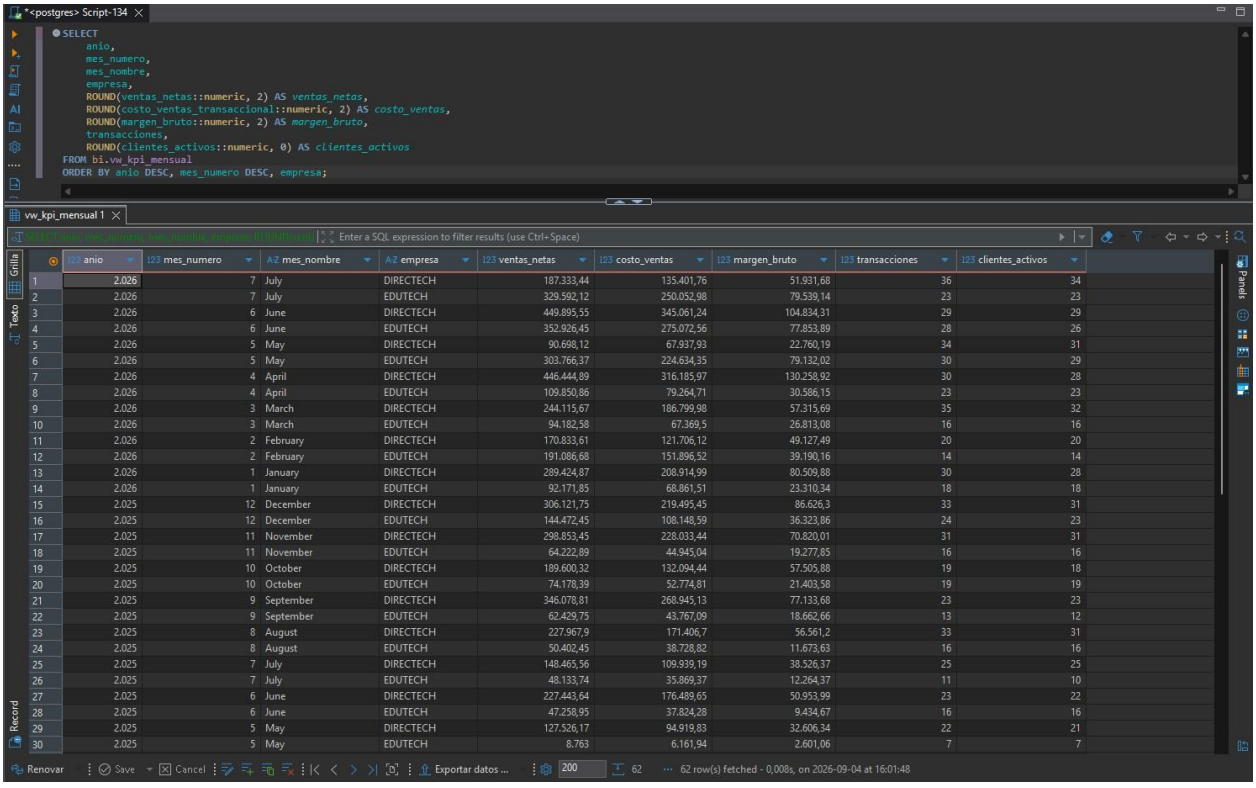
Tabla	Registros esperados
dim_empresa	2

Tabla	Registros esperados
dim_fecha	943
dim_producto	64
dim_cliente	250
dim_proveedor	80
dim_vendedor	16
fact_ventas	1.500
fact_compras	900
fact_devoluciones	130
fact_estado_resultados	62
fact_balance_mensual	62

Nota. Elaboración propia.

Figura 10

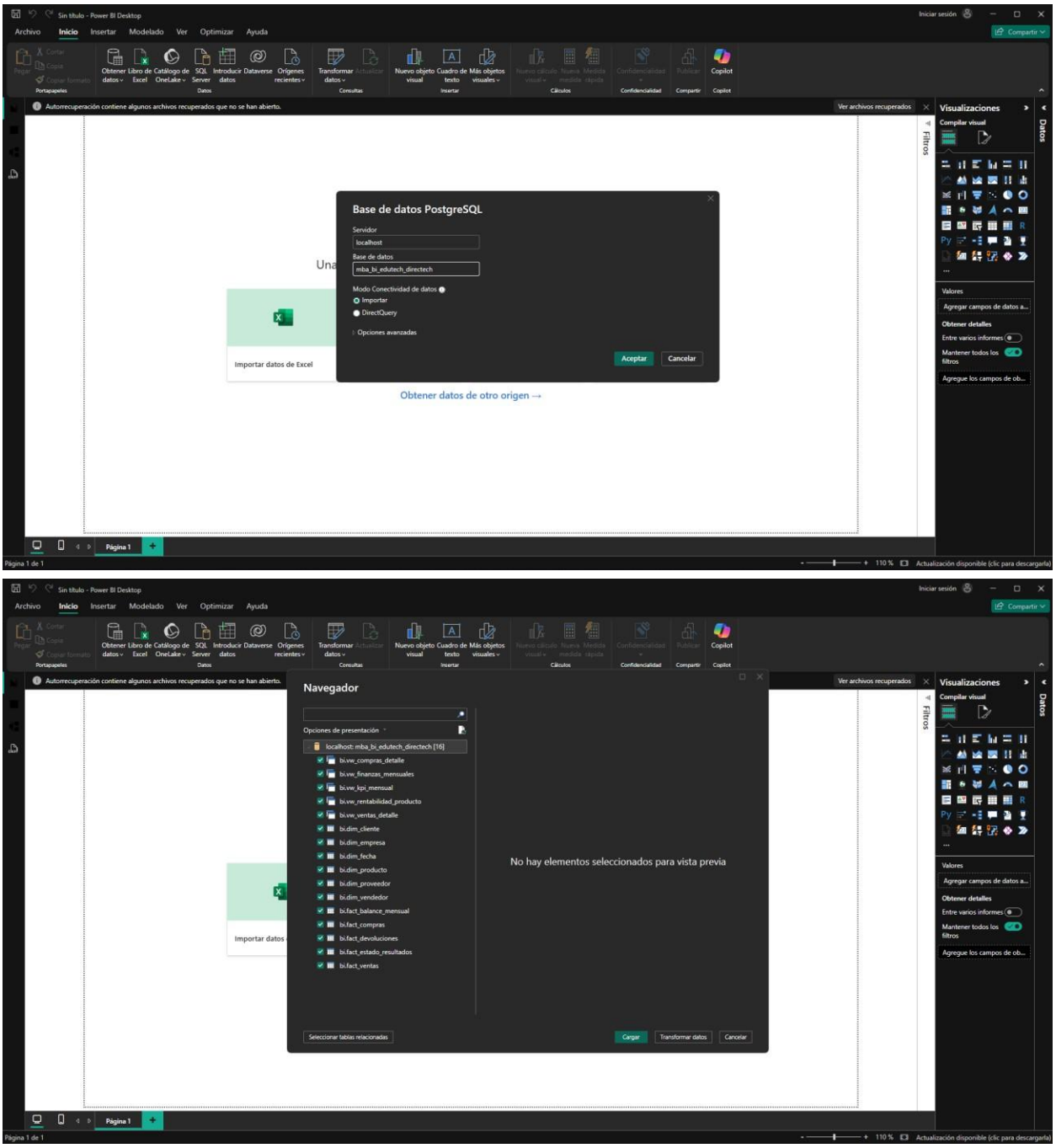
E5. Consulta de validación en SQL



Nota. Captura de la implementación del proyecto.

Figura 11

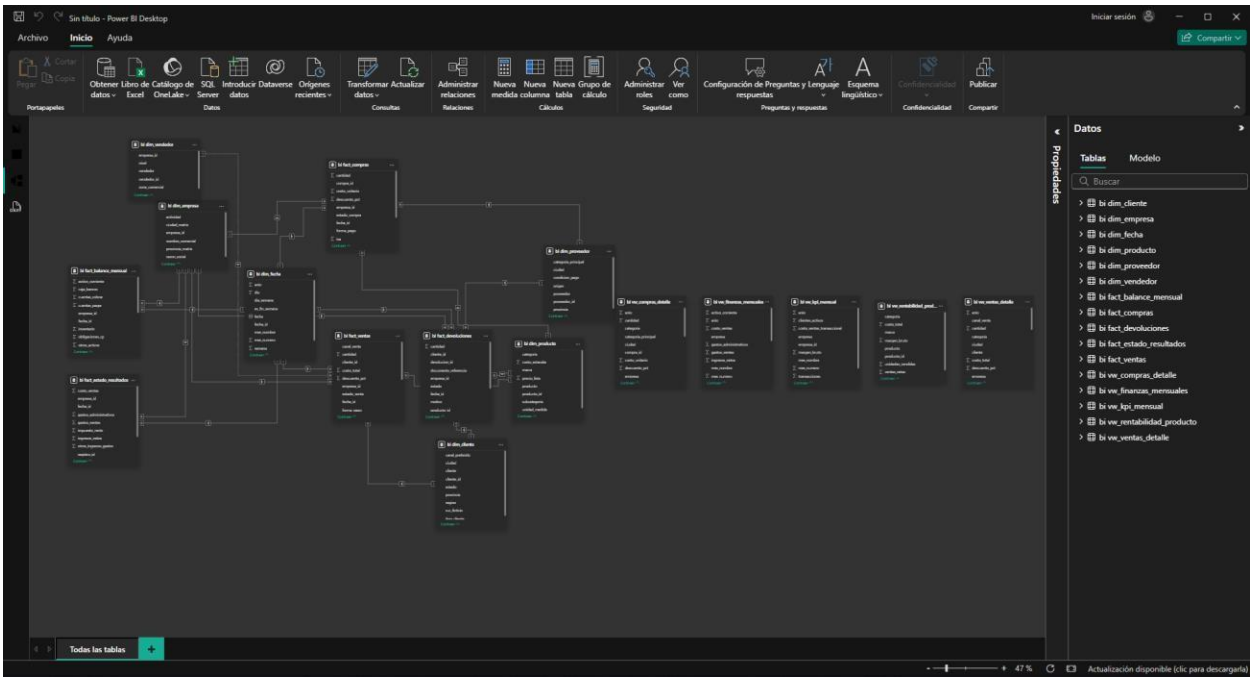
E6. Conexión de Power BI con PostgreSQL



Nota. Captura de la implementación del proyecto.

Figura 12

E7. Modelo relacional en Power BI

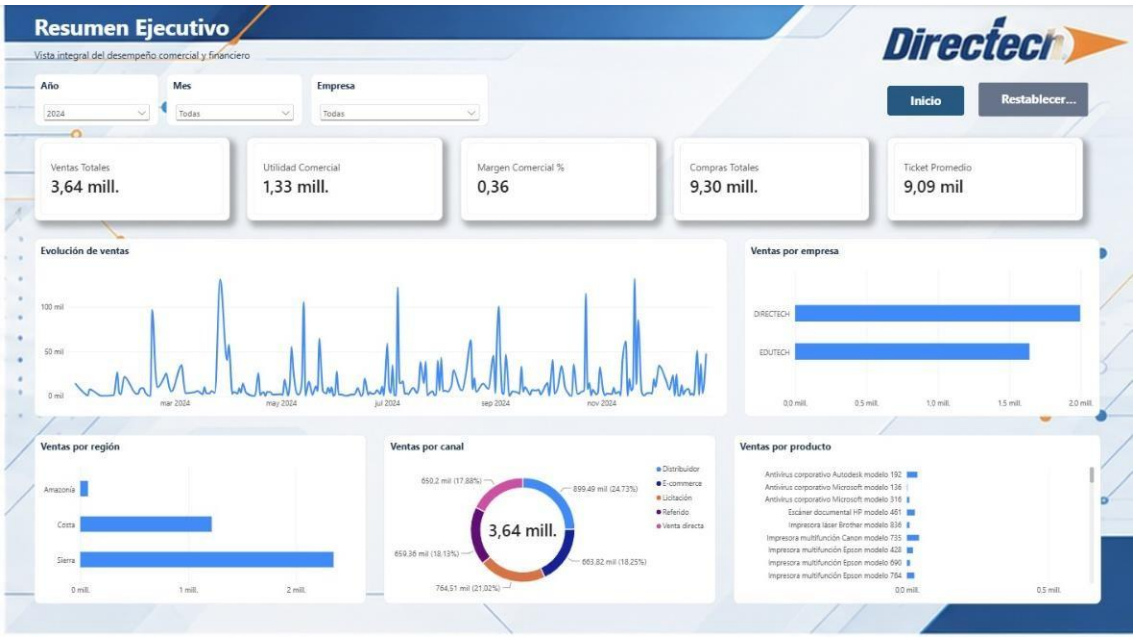


Nota. Captura de la implementación del proyecto.

Sobre este modelo se construyó el reporte gerencial en Power BI, organizado en páginas temáticas (Resumen Ejecutivo, Ventas, Compras y Proveedores, Productos y Devoluciones, Finanzas y Balance Financiero) a las que se accede desde una portada de navegación. A continuación se presentan los dos tableros más directamente vinculados con el objetivo del proyecto: la visión consolidada para la Gerencia y el estado de resultados. Cada página incorpora segmentadores por año, mes y empresa, y botones para volver al inicio o restablecer los filtros, de modo que cada perfil de usuario puede consultar la información sin depender de reportes preparados por Contabilidad.

Figura 13

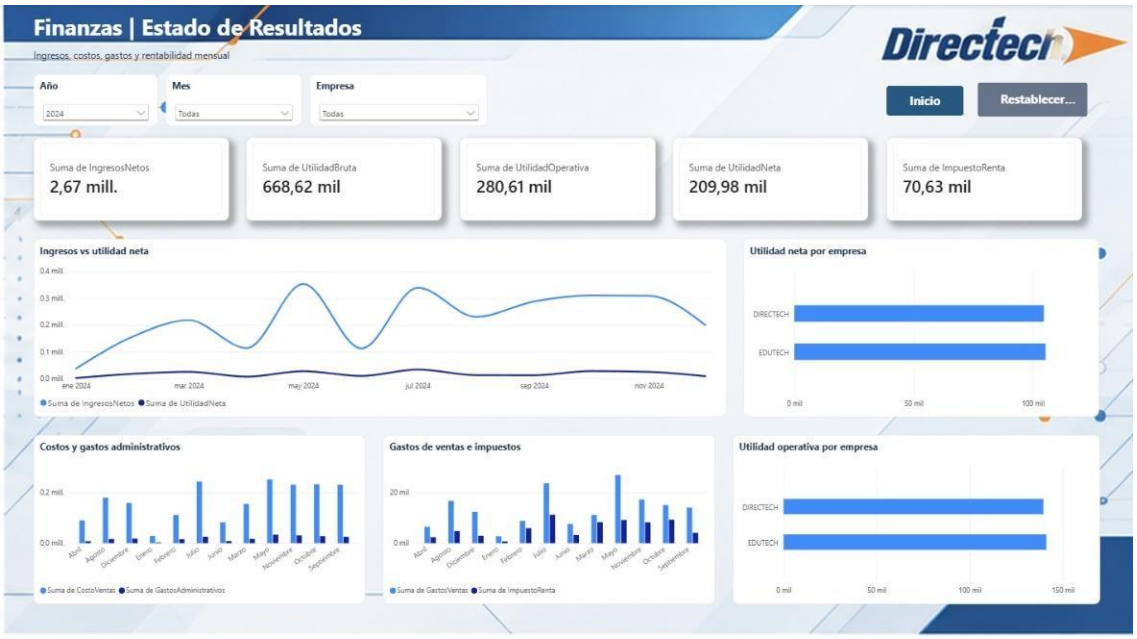
E8. Tablero «Resumen Ejecutivo»



Nota. Captura de la implementación del proyecto.

Figura 14

E9. Tablero «Finanzas – Estado de Resultados»



Nota. Captura de la implementación del proyecto.

1.9. Conclusiones del capítulo

La propuesta integra información comercial, operativa, financiera y geográfica de EDUTECH y DIRECTECH bajo una arquitectura de Business Intelligence orientada a la toma de decisiones. El modelo dimensional permite relacionar transacciones con clientes, productos, proveedores, empresas y periodos, facilitando el cálculo de indicadores que no serían visibles al analizar archivos independientes.

Los casos de uso muestran que el valor no se limita a las visualizaciones: el modelo permite identificar tendencias, concentraciones de riesgo y oportunidades de mejora. PostgreSQL centraliza los hechos y las dimensiones, y constituye también el destino propuesto de la tabla histórica de resultados de IA. Fabric provee la ingesta y el procesamiento analítico de apoyo, sin crear un segundo repositorio definitivo que duplique las métricas.

Finalmente, el uso de Big Data se plantea como una evolución potencial y no como una condición del modelo actual. A medida que las empresas incorporen información digital, logística, de CRM, servicio al cliente o fuentes externas de alta frecuencia, la arquitectura podrá avanzar hacia capacidades predictivas y prescriptivas, ampliando el impacto de los datos en la estrategia empresarial.

CAPÍTULO 2: Arquitectura Cloud y Landing Zone para el modelo de BI

La evolución del modelo de Business Intelligence de EDUTECH y DIRECTECH requiere una arquitectura tecnológica que permita integrar las fuentes estructuradas utilizadas actualmente con fuentes externas y semiestructuradas que puedan incorporarse progresivamente. La computación en la nube ofrece una alternativa escalable para almacenar, transformar y analizar información sin depender de una infraestructura física compleja administrada directamente por las empresas.

El caso de referencia de Bancolombia introduce el concepto de Landing Zone como una zona centralizada de exposición temprana de datos internos y externos (Pérez Galán, 2026c). Esta capa permite ejecutar procesos de calidad, transformación y enriquecimiento y convertirse posteriormente en fuente para bodegas de datos, datamarts y soluciones analíticas. Para EDUTECH y DIRECTECH se adopta esta lógica a una escala acorde con el proyecto académico y con la arquitectura de BI ya desarrollada.

2.1. Objetivo y alcance

El objetivo de este capítulo es definir una alternativa de arquitectura Cloud que complemente el modelo de Business Intelligence de EDUTECH y DIRECTECH. Para ello se comparan diferentes proveedores de nube pública, se selecciona la alternativa más adecuada para el proyecto, se determina el modelo de servicio Cloud y se establece el flujo de integración entre una Landing Zone, el Data Warehouse en PostgreSQL y Power BI.

La propuesta no reemplaza el modelo dimensional desarrollado previamente. Su propósito es incorporar una capa de almacenamiento e integración que facilite la recepción de datos con diferentes estructuras y permita una futura evolución hacia escenarios de mayor volumen, variedad y velocidad de información.

2.2. Análisis de alternativas de proveedores Cloud

Para seleccionar una plataforma se analizaron tres proveedores de nube pública con capacidad para soportar almacenamiento, integración de datos, bases PostgreSQL administradas y herramientas analíticas: Amazon Web Services (AWS), Microsoft Azure y Google Cloud Platform (GCP). La comparación se realiza considerando principalmente la compatibilidad con la arquitectura ya implementada, la facilidad de integración, la escalabilidad y la complejidad operativa para un proyecto de tamaño académico y empresarial medio.

Tabla 37

Comparación de proveedores de nube pública

Criterio	Amazon Web Services	Microsoft Azure	Google Cloud Platform
Almacenamiento / Data Lake	Amazon S3	Azure Data Lake Storage Gen2	Google Cloud Storage
Integración y ETL	AWS Glue	Azure Data Factory	Dataflow / servicios de integración
PostgreSQL administrado	Amazon RDS for PostgreSQL	Azure Database for PostgreSQL	Cloud SQL for PostgreSQL

Criterio	Amazon Web Services	Microsoft Azure	Google Cloud Platform
Analítica / BI	Amazon QuickSight	Power BI / Microsoft Fabric	Looker / BigQuery
Compatibilidad con PostgreSQL	Alta	Alta	Alta
Integración con Power BI	Disponible mediante conectores	Muy alta por ecosistema Microsoft	Disponible mediante conectores
Complejidad para el proyecto	Media	Baja–media	Media
Escalabilidad	Alta	Alta	Alta

Nota. Elaboración propia con base en Amazon Web Services (2026), Microsoft (2026b) y Google Cloud (2026).

2.2.1. Amazon Web Services (AWS)

AWS constituye una alternativa robusta para arquitecturas de datos. Amazon S3 puede funcionar como repositorio central de archivos, AWS Glue permite catalogar y transformar información y Amazon RDS ofrece PostgreSQL como servicio administrado. Su ecosistema es amplio y escalable, aunque para el alcance actual del proyecto implicaría incorporar herramientas adicionales diferentes de las ya utilizadas por el equipo.

2.2.2. Microsoft Azure

Azure dispone de Azure Data Lake Storage Gen2 para almacenamiento analítico, Azure Data Factory para procesos de integración y Azure Database for PostgreSQL como servicio administrado. Su principal ventaja para EDUTECH y DIRECTECH es la continuidad tecnológica con Power BI y con el entorno Microsoft ya utilizado para el procesamiento y visualización de información.

2.2.3. Google Cloud Platform (GCP)

Google Cloud ofrece Cloud Storage para almacenamiento, Cloud SQL para PostgreSQL y servicios como Dataflow y BigQuery para procesamiento y analítica. Es una alternativa técnicamente viable y con fuertes capacidades analíticas, aunque requeriría una mayor adaptación del flujo actual orientado a Power BI y herramientas Microsoft.

2.3. Selección de la alternativa de proveedor

Para el proyecto de Business Intelligence de EDUTECH y DIRECTECH se selecciona Microsoft Azure como proveedor Cloud de referencia. La selección se realiza por su adecuación al modelo existente y no porque los otros proveedores carezcan de capacidad técnica.

El proyecto utiliza PostgreSQL como repositorio dimensional y Power BI como herramienta de visualización. Azure constituye la plataforma de referencia para una eventual operación Cloud; la prueba de concepto de Landing y RAW se ejecutó en Microsoft Fabric/OneLake. Una migración futura del Data Warehouse a Azure Database for PostgreSQL o a un servicio equivalente deberá justificarse por sus costos y necesidades, sin asumir que dicho despliegue productivo ya está realizado.

Tabla 38

Criterios de selección del proveedor Cloud

Criterio de selección	Justificación para EDUTECH y DIRECTECH
Continuidad tecnológica	Conserva PostgreSQL y Power BI dentro de la solución.
Landing Zone	Separa la recepción de archivos del Data Warehouse estructurado.
Integración	Azure Data Factory permite automatizar pipelines ETL.
Escalabilidad	Los servicios pueden crecer conforme aumenten fuentes y volumen.
Administración	Reduce la operación manual de servidores y bases de datos.
Evolución futura	Permite ampliar fuentes y analítica sin reemplazar el modelo de BI.

Nota. Elaboración propia.

2.3.1. Arquitectura tecnológica propuesta

Tabla 39

Arquitectura tecnológica propuesta en Azure

Capa	Tecnología propuesta	Función principal
Fuentes	PostgreSQL, Excel/CSV, datos abiertos y fuentes externas	Generación y captura de información interna y externa.
Landing Zone	Azure Data Lake Storage Gen2	Recepción inicial y conservación de datos antes de su transformación.
Integración	Azure Data Factory + Power Query	Extracción, limpieza, homologación, validación y carga.
Data Warehouse	Azure Database for PostgreSQL	Almacenamiento estructurado bajo el modelo dimensional del proyecto.
Modelo analítico	Power BI	Relaciones, medidas, KPI y análisis.
Consumo	Power BI Service	Publicación de dashboards e información gerencial.

Nota. Elaboración propia.

2.4. Definición del modelo de Cloud Computing

Los servicios de computación en la nube suelen organizarse en tres modelos principales: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) y Software as a Service (SaaS) (Mell & Grance, 2011). La diferencia principal se encuentra en el nivel de administración que conserva la organización y el nivel de responsabilidad asumido por el proveedor.

Tabla 40

Modelos de servicio Cloud y su aplicación al proyecto

Modelo	Características	Aplicación al proyecto
IaaS	Proporciona infraestructura virtual. El cliente administra en mayor medida sistemas operativos, configuraciones y componentes instalados.	No se selecciona como modelo principal porque incrementaría las tareas técnicas de administración.

Modelo	Características	Aplicación al proyecto
PaaS	El proveedor administra la infraestructura base y la organización utiliza plataformas y servicios para desplegar datos y procesos.	Modelo principal propuesto para Data Lake, integración y PostgreSQL administrado.
SaaS	La aplicación completa es ofrecida como servicio y consumida por los usuarios finales.	Modelo complementario utilizado mediante Power BI Service para publicación y consumo de dashboards.

Nota. Elaboración propia con base en Mell y Grance (2011).

2.4.1. Modelo seleccionado: PaaS

Se define PaaS como el modelo principal de Cloud Computing para la arquitectura de datos. EDUTECH y DIRECTECH necesitan controlar la estructura de la información, las reglas de transformación, el modelo dimensional y los indicadores, pero no requieren administrar directamente servidores físicos o sistemas operativos para cada componente.

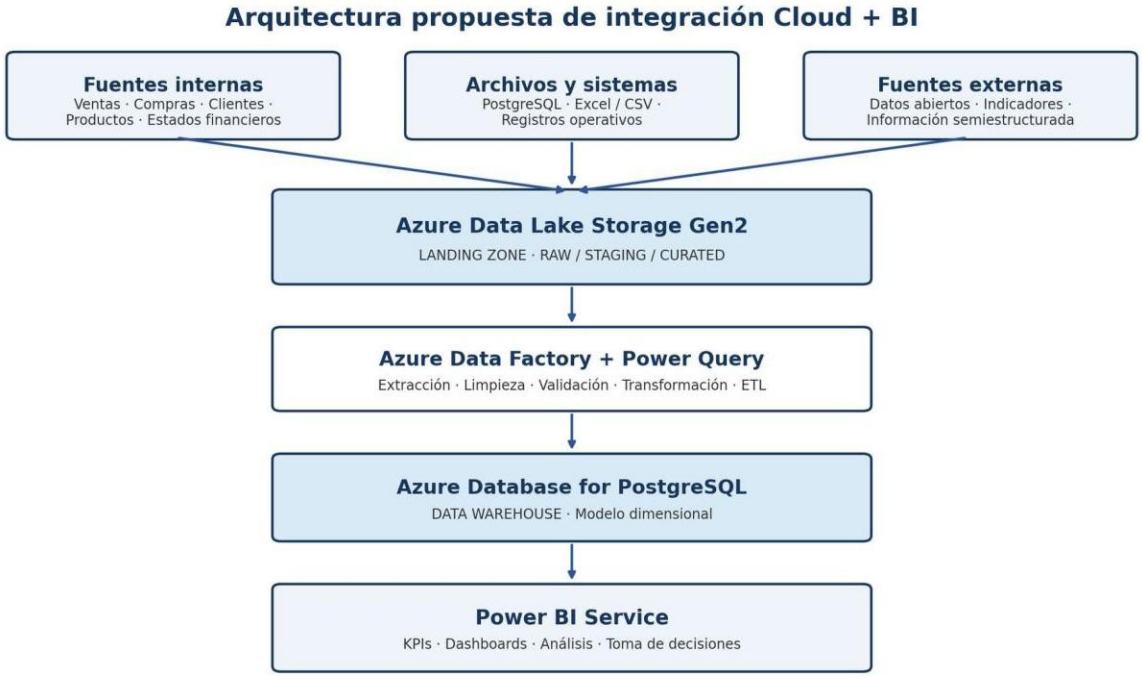
Bajo este enfoque, Azure Data Lake Storage Gen2, Azure Data Factory y Azure Database for PostgreSQL proporcionan servicios administrados que permiten concentrar los esfuerzos en la gestión y análisis de la información. De manera complementaria, Power BI Service funciona bajo un enfoque SaaS para distribuir los resultados analíticos a los usuarios.

2.5. Flujo de integración de la arquitectura Big Data con el Data Warehouse

La arquitectura propuesta mantiene el Data Warehouse en PostgreSQL como repositorio estructurado para el modelo dimensional e incorpora una capa previa de almacenamiento denominada Landing Zone. Esta capa recibe información procedente de diferentes fuentes sin exigir que todos los datos ingresen inicialmente con la misma estructura.

Figura 15

Arquitectura propuesta de integración Cloud + BI



Nota. Elaboración propia.

2.5.1. Etapa 1: fuentes de información

El flujo inicia con las fuentes internas y externas. En el entorno actual se incluyen las tablas y archivos relacionados con ventas, compras, clientes, productos, proveedores, devoluciones, estados financieros y otras fuentes utilizadas en el modelo académico. Conforme el proyecto evolucione podrán incorporarse datos abiertos, indicadores externos, registros semiestructurados y otras fuentes de mayor variedad.

2.5.2. Etapa 2: Landing Zone

Azure Data Lake Storage Gen2 funcionará como Landing Zone. Los datos se almacenarán inicialmente en una zona RAW conservando el formato de origen. Posteriormente podrán organizarse en zonas de preparación o STAGING y, una vez validados, en una zona CURATED lista para los procesos analíticos. Esta separación mejora la trazabilidad y evita modificar directamente el Data Warehouse cada vez que aparece una nueva fuente.

2.5.3. Etapa 3: integración y transformación

Azure Data Factory y Power Query se utilizarán para ejecutar procesos de extracción, transformación y carga. Durante esta fase se validarán formatos, fechas e identificadores; se

eliminarán duplicidades; se tratarán valores faltantes; y se homologarán estructuras antes de cargar la información en el modelo dimensional.

2.5.4. Etapa 4: Data Warehouse

Los datos estructurados y validados se cargarán en Azure Database for PostgreSQL. De esta manera se conserva la arquitectura ya desarrollada, formada por tablas de hechos y dimensiones relacionadas con empresa, fecha, cliente, producto, proveedor y vendedor. PostgreSQL continúa siendo la capa de persistencia principal para los datos que alimentan los indicadores de Business Intelligence.

2.5.5. Etapa 5: análisis y consumo

Power BI se conectará al Data Warehouse para construir y actualizar las medidas, indicadores y visualizaciones. Power BI Service permitirá publicar los dashboards para que los usuarios autorizados consulten información comercial, financiera y operativa sin acceder directamente a los repositorios de origen.

2.5.6. Relación con el enfoque Big Data

La propuesta actual continúa siendo principalmente una arquitectura de Business Intelligence. Sin embargo, la Landing Zone crea una base tecnológica para incorporar progresivamente datos estructurados, semiestructurados y no estructurados. Si en el futuro aumentan significativamente el volumen, la velocidad y la variedad de la información, esta capa facilitará la incorporación de procesamiento distribuido, analítica avanzada y modelos predictivos sin eliminar el Data Warehouse que soporta la reportería gerencial.

2.6. Conclusiones del capítulo

- El análisis de AWS, Microsoft Azure y Google Cloud demuestra que los tres proveedores cuentan con servicios capaces de soportar una arquitectura moderna de datos. La decisión debe considerar principalmente la compatibilidad con el entorno tecnológico y la complejidad operativa del proyecto.
- Microsoft Azure se selecciona como proveedor de referencia porque permite mantener PostgreSQL, incorporar una Landing Zone, automatizar procesos de integración y conservar Power BI como herramienta de análisis y visualización.
- El modelo Cloud principal propuesto es PaaS, ya que permite utilizar servicios administrados sin asumir la operación completa de servidores e infraestructura. Power BI Service complementa la solución mediante un modelo SaaS.
- La Landing Zone separa la recepción de datos de la estructura analítica final. Esto permite conservar información en su formato de origen, aplicar procesos de calidad y transformación y cargar al Data Warehouse únicamente la información preparada para análisis.
- La arquitectura propuesta amplía el modelo existente de EDUTECH y DIRECTECH y establece una base para incorporar nuevas fuentes y evolucionar hacia escenarios de mayor volumen, variedad y velocidad de datos.

CAPÍTULO 3: Implementación práctica, trazabilidad y gobierno operativo de la arquitectura BI

En el Capítulo 2 se definió una arquitectura Cloud de referencia para integrar fuentes estructuradas y semiestructuradas con un Data Warehouse en PostgreSQL y una capa de análisis en Power BI. La retroalimentación recibida confirmó la coherencia de la selección tecnológica, pero señaló la necesidad de fortalecer la justificación del proveedor, las decisiones operativas del proceso ETL y los criterios que determinarían una futura evolución hacia escenarios de Big Data.

En este capítulo el proyecto avanza desde el diseño conceptual hacia una prueba de concepto práctica. Para reducir complejidad técnica y mantener un enfoque acorde con un MBA, la implementación utiliza Microsoft Fabric como plataforma de ingesta y preparación visual de datos, PostgreSQL como repositorio estructurado final y Power BI como herramienta de análisis. El objetivo no es desarrollar software, sino demostrar cómo una arquitectura de datos bien gobernada mejora la calidad, trazabilidad y disponibilidad de la información para la toma de decisiones.

3.1. Objetivo del capítulo

Consolidar la arquitectura de Business Intelligence de EDUTECH y DIRECTECH incorporando criterios de selección tecnológica, gobierno operativo del ETL, trazabilidad, monitoreo y una primera implementación práctica en Microsoft Fabric, manteniendo PostgreSQL como Data Warehouse y Power BI como capa de consumo analítico.

3.2. Ajustes derivados de la retroalimentación del Capítulo 2

La retroalimentación del capítulo anterior se concentra en tres áreas: rigor en la selección del proveedor Cloud, definición operativa de las capas de datos y condiciones concretas para escalar hacia tecnologías de Big Data. Este capítulo incorpora estos tres ajustes dentro del diseño y de la implementación práctica.

3.2.1. Matriz ponderada de selección tecnológica

La comparación entre AWS, Microsoft Azure/Fabric y Google Cloud se replantea mediante una matriz ponderada. Las puntuaciones de 1 a 5 representan una valoración académica específica para este proyecto y no una clasificación universal de los proveedores.

Tabla 41

Matriz ponderada de selección tecnológica para la prueba de concepto

Criterio	Peso	AWS	Azure / Fabric	GCP
Coste para la prueba de concepto	15 %	4	5	4
Integración con Power BI y ecosistema Microsoft	20 %	3	5	3
Compatibilidad con PostgreSQL	10 %	5	5	5
Seguridad y gobierno	15 %	5	5	5
Opciones de región y residencia del dato	10 %	5	5	5
Capacidades y curva de aprendizaje del equipo	10 %	3	5	3

Criterio	Peso	AWS	Azure / Fabric	GCP
Portabilidad / dependencia del proveedor	10 %	4	3	4
Complejidad operativa	10 %	3	4	3

Nota. Elaboración propia.

Tabla 42

Resultado ponderado de la selección tecnológica

Proveedor	Puntuación ponderada	Interpretación para el proyecto
AWS	3,95 / 5	Alternativa robusta, pero con mayor cambio de herramientas para el equipo actual.
Azure / Fabric	4,70 / 5	Mayor continuidad con Power BI, experiencia visual y prueba de concepto ya habilitada.
GCP	3,95 / 5	Alternativa viable, con adaptación adicional frente al ecosistema ya utilizado.

Nota. Elaboración propia. Azure/Fabric obtiene la mayor adecuación para esta prueba de concepto.

Consideración de coste: la implementación académica actual utiliza una capacidad de prueba de Microsoft Fabric, PostgreSQL local, DBeaver Community y Power BI Desktop; por tanto, el coste incremental de software para la prueba de concepto es cero mientras la capacidad de prueba se encuentre activa. Este dato no representa el costo total de propiedad (TCO) de una implementación productiva, que deberá estimarse posteriormente mediante los calculadores oficiales y el volumen real de uso.

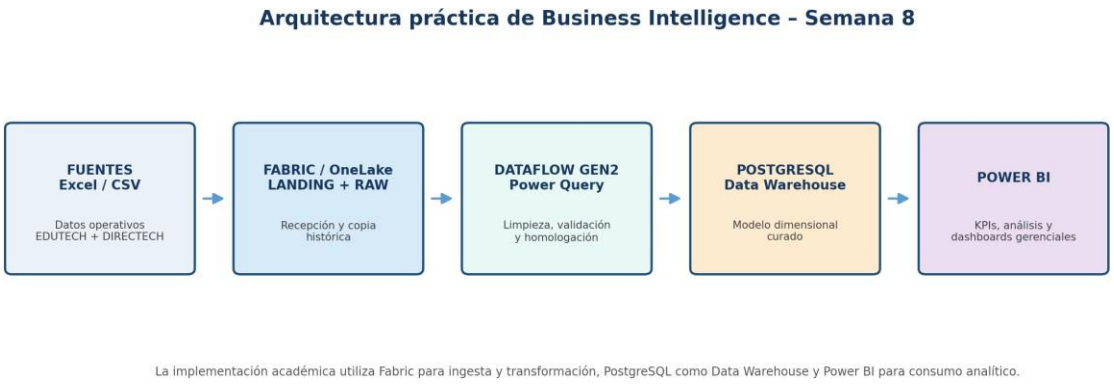
Residencia del dato: la capacidad utilizada en la prueba está asignada a la región South Central US. Esto resulta aceptable para datos sintéticos de carácter académico. En un escenario productivo con información real se deberá validar la región, los requisitos regulatorios —incluida la Ley Orgánica de Protección de Datos Personales (2021) del Ecuador— y las políticas corporativas antes de almacenar datos sensibles fuera de la jurisdicción definida por la organización.

3.3. Arquitectura consolidada de la solución

La arquitectura de este capítulo mantiene los principios planteados previamente, pero diferencia con mayor claridad las responsabilidades de cada componente. Fabric se utiliza para la ingesta, almacenamiento temporal y preparación visual; PostgreSQL permanece como Data Warehouse final; y Power BI se reserva para el modelo semántico, DAX y visualización gerencial (Microsoft, 2026c).

Figura 16

Arquitectura práctica consolidada para la prueba de concepto



Nota. Elaboración propia.

3.3.1. Responsabilidad de cada componente

Tabla 43

Separación de responsabilidades tecnológicas en la arquitectura implementada

Capa	Herramienta	Responsabilidad principal	Resultado
Fuentes	Excel / CSV	Generar y entregar información operativa	Archivos originales
Landing	Fabric Lakehouse / OneLake	Recepción de archivos sin transformación	Fuente recibida y trazable
RAW	Fabric Lakehouse / OneLake	Conservar copia histórica inalterada	Respaldo del dato original
Ingesta	Fabric Pipeline + Copy Job	Mover archivos entre Landing y RAW	Carga controlada y monitorizada
ETL	Dataflow Gen2 + Power Query	Limpiar, validar, homologar y consolidar	Datos preparados
Curated / DW	PostgreSQL	Persistir hechos y dimensiones depuradas	Modelo dimensional
Consumo	Power BI	Medidas, KPI y visualización	Dashboard gerencial

Nota. Elaboración propia.

3.4. Modelo operativo de ingesta y ETL

Uno de los principales ajustes frente al Capítulo 2 consiste en definir cómo funcionará la arquitectura en operación. Para ello se establecen reglas de carga, frecuencia, tratamiento de errores, monitoreo, trazabilidad y responsables.

3.4.1. Carga completa e incremental

La primera ejecución de cada conjunto histórico se realiza como carga completa (FULL). En las ejecuciones posteriores se utilizará una estrategia incremental para procesar únicamente registros nuevos o modificados, evitando reprocesar innecesariamente todo el histórico (Kimball & Caserta, 2004).

Tabla 44

Política de carga y frecuencia de actualización

Fuente	Carga inicial	Carga posterior	Frecuencia objetivo
Ventas	FULL	Incremental	Diaria
Compras	FULL	Incremental	Diaria
Devoluciones	FULL	Incremental	Diaria
Clientes	FULL	Incremental	Diaria
Productos	FULL	Incremental	Semanal
Proveedores	FULL	Incremental	Semanal
Estado de resultados	FULL	Incremental por periodo	Mensual
Balance	FULL	Incremental por periodo	Mensual

Nota. Elaboración propia.

3.4.2. Manejo de errores y calidad

Los registros que incumplan reglas esenciales —por ejemplo, identificadores vacíos, cantidades negativas, fechas inválidas o referencias inexistentes— no deben desaparecer silenciosamente ni mezclarse con los datos válidos. Se propone una salida de RECHAZADOS para revisión y un indicador de calidad de carga que compare registros recibidos, válidos y rechazados.

3.4.3. Monitoreo y trazabilidad

Fabric conserva el estado de ejecución de las canalizaciones y permite verificar si una actividad terminó correctamente, su duración y el momento de ejecución (Microsoft, 2026e). Adicionalmente, la nomenclatura de carpetas incorpora el tipo de carga y la fecha, por ejemplo FULL_2026_09_22, lo que facilita reconstruir el origen de una carga específica. En etapas posteriores se añadirá una tabla de control con fecha, archivo, filas recibidas, filas válidas, filas rechazadas y estado del proceso.

3.4.4. Responsabilidades

Tabla 45

Responsabilidades operativas del flujo de información

Actividad	Responsable principal	Valida	Resultado
Generación de archivos fuente	Área de negocio	Dueño del dato	Archivos recibidos
Ingesta Landing → RAW	Analista BI	Responsable del proyecto	Copia histórica
Depuración y homologación	Analista BI	Dueño del dato	Datos preparados

Actividad	Responsable principal	Valida	Resultado
Carga al Data Warehouse	Analista BI	Responsable técnico	PostgreSQL actualizado
Validación de KPI	Analista BI	Gerencia	Indicadores aprobados
Toma de decisiones	Gerencia	Áreas de negocio	Acciones de gestión

Nota. Elaboración propia.

En el caso de Directech, el rol de Analista BI corresponde a la Coordinación de BI en doble rol definida en el Capítulo 5 de la Parte I, y el dueño del dato a cada responsable de área identificado en la matriz de flujos.

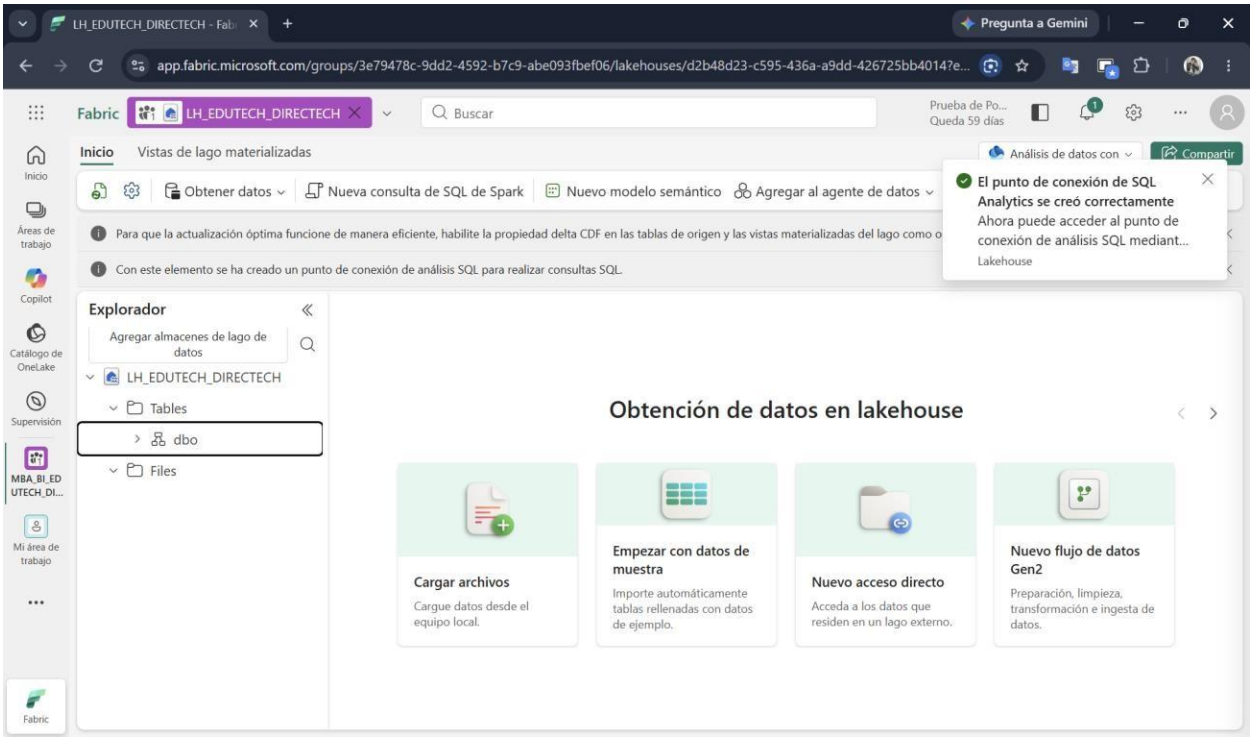
3.5. Implementación práctica realizada en Microsoft Fabric

La prueba de concepto ya cuenta con componentes funcionales. Las evidencias siguientes muestran el avance real desde la creación del Lakehouse hasta la ejecución de la primera carga completa desde Landing hacia RAW.

3.5.1. Creación del Lakehouse

Figura 17

Lakehouse LH_EDUTECH_DIRECTECH creado dentro del workspace del proyecto



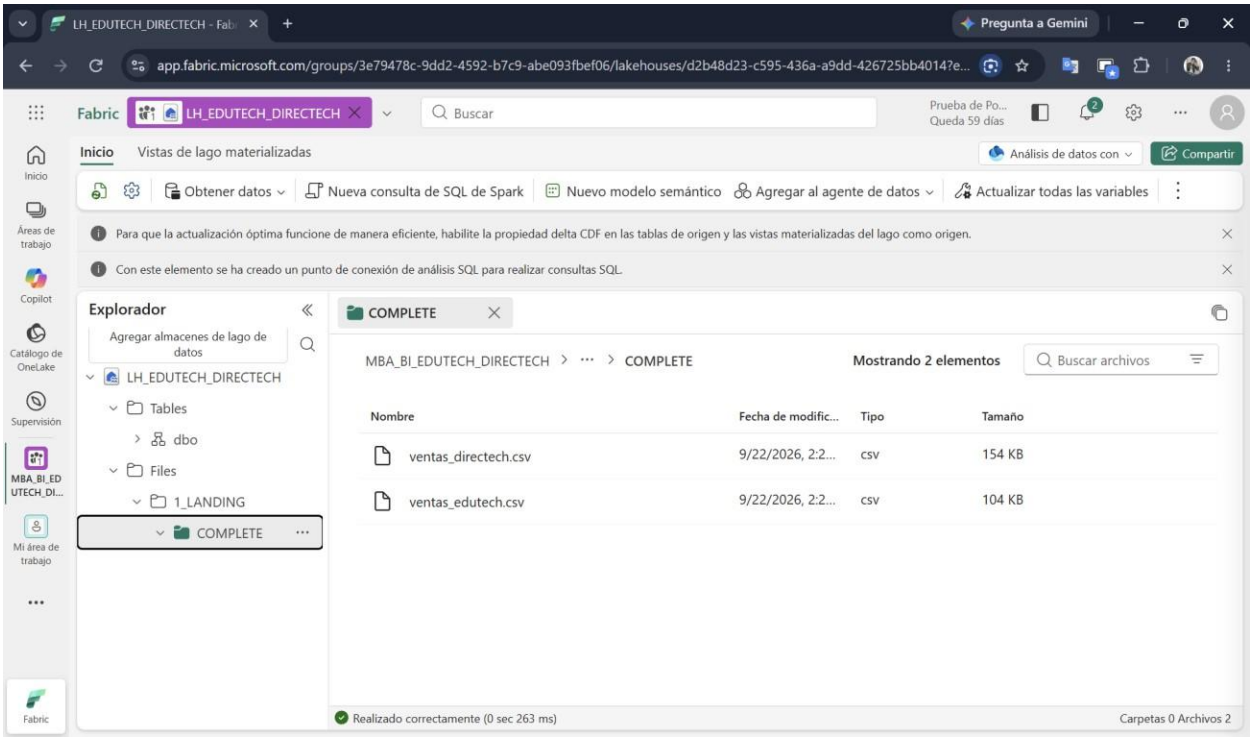
Nota. Evidencia de implementación propia.

El Lakehouse proporciona las áreas Files y Tables dentro de OneLake. Para la prueba se utiliza Files como espacio de recepción de archivos y se reserva Tables para etapas posteriores, cuando los datos ya se encuentren preparados para consumo estructurado.

3.5.2. Zona Landing con archivos originales

Figura 18

Estructura inicial de Landing y RAW en el Lakehouse



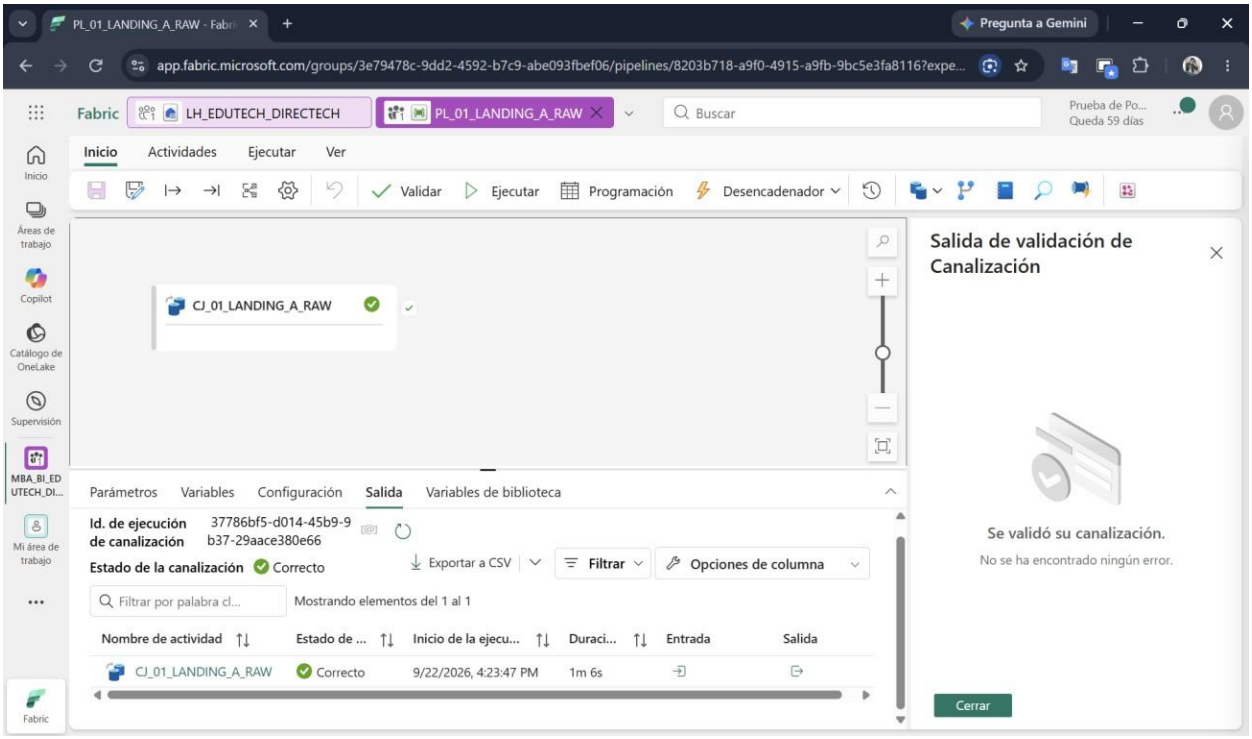
Nota. Evidencia de implementación propia.

Las ventas de EDUTECH y DIRECTECH se almacenan inicialmente en la carpeta 01_LANDING/FULL_2026_09_22. En esta zona no se aplican transformaciones: el propósito es conservar la información tal como fue recibida.

3.5.3. Pipeline y primera carga FULL

Figura 19

Ejecución correcta del pipeline PL_01_LANDING_A_RAW



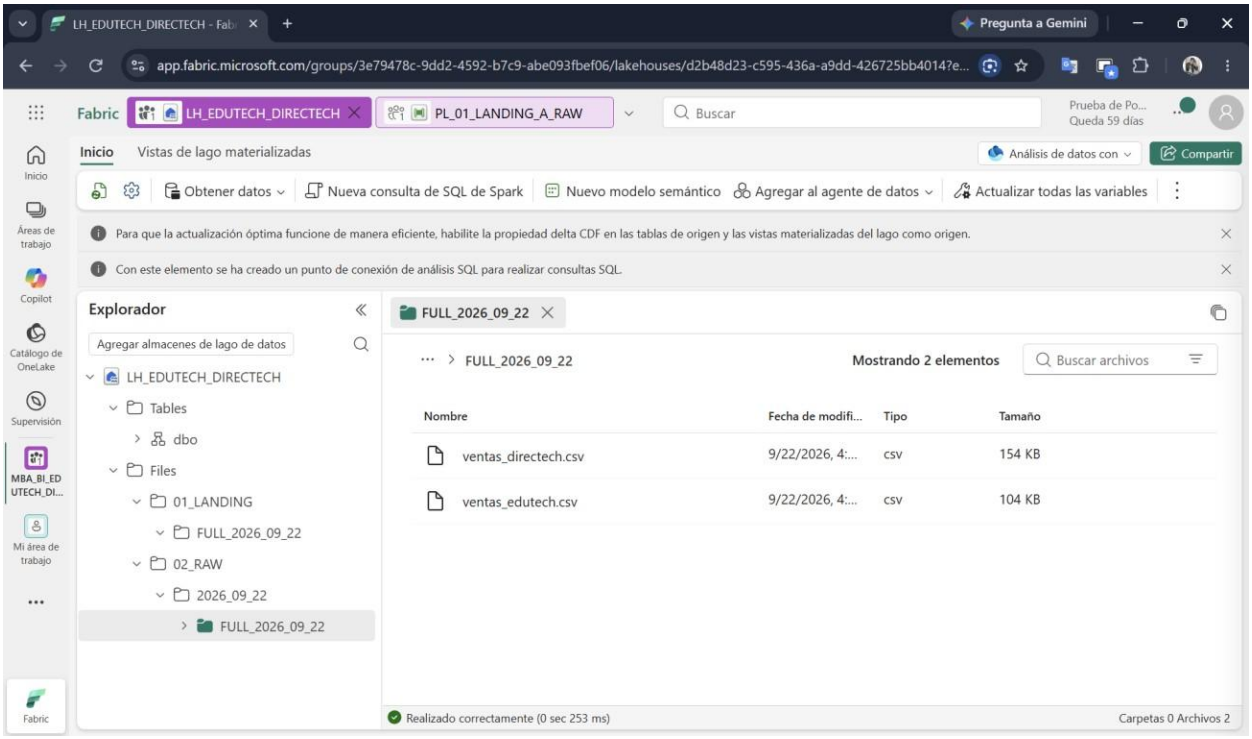
Nota. Evidencia de implementación propia.

El pipeline PL_01_LANDING_A_RAW orquesta un Copy Job que traslada los archivos desde Landing hacia RAW mediante copia binaria, evitando interpretar o alterar el contenido. La ejecución mostrada finalizó con estado Correcto y una duración aproximada de un minuto, lo que constituye la primera evidencia de monitoreo del proceso.

3.5.4. Verificación de la capa RAW

Figura 20

Archivos CSV preservados en la capa RAW después de la copia binaria



Nota. Evidencia de implementación propia.

La verificación final confirma que ventas_directech.csv y ventas_edutech.csv se mantienen con su extensión original dentro de 02_RAW/2026_09_22/FULL_2026_09_22. La conservación del formato confirma que la capa RAW actúa como respaldo histórico y no como zona de transformación.

3.6. Diseño de la siguiente etapa: Dataflow Gen2 y Power Query

La siguiente fase implementará el ETL visual mediante Dataflow Gen2 (Microsoft, 2026d). La elección responde al enfoque del MBA: permite trabajar con Power Query sin desarrollar código y hace visibles las decisiones de calidad de datos que posteriormente impactarán los indicadores gerenciales.

Tabla 46

Secuencia de transformación prevista en Dataflow Gen2

Etapa	Acción	Propósito
1	Lectura desde RAW	Tomar ventas_edutech.csv y ventas_directech.csv directamente desde OneLake.
2	Tipificación	Definir correctamente fechas, enteros, decimales y campos de texto.
3	Limpieza	Recortar espacios, limpiar caracteres y homologar mayúsculas/minúsculas.

Etapa	Acción	Propósito
4	Calidad	Detectar duplicados, nulos, cantidades no válidas y estados fuera de catálogo.
5	Reglas de negocio	Clasificar ventas efectivas, pendientes y canceladas.
6	Consolidación	Anexar EDUTECH y DIRECTECH en una única consulta de ventas.
7	Salida preparada	Publicar la información depurada para su carga a PostgreSQL.

Nota. Elaboración propia.

3.7. Condiciones para una futura evolución hacia Big Data

La arquitectura actual sigue siendo una solución de Business Intelligence tradicional: el volumen de datos es moderado, la actualización no exige procesamiento en tiempo real y las fuentes principales son estructuradas. Por tanto, no sería correcto presentar la implementación actual como Big Data.

La evolución hacia tecnologías de procesamiento distribuido deberá evaluarse cuando aparezcan condiciones concretas que excedan el nivel de servicio definido para la arquitectura actual. Los siguientes umbrales se plantean como criterios internos de decisión del proyecto, no como una definición universal de Big Data:

Tabla 47

Condiciones que justificarían evaluar tecnologías de Big Data

Condición	Criterio de escalamiento
Volumen	Crecimiento sostenido hacia decenas o cientos de millones de hechos, o históricos que deterioren los tiempos de refresco y consulta.
Velocidad	Necesidad de actualizar indicadores en minutos o segundos en lugar de cargas diarias.
Variedad	Incorporación masiva de JSON, logs, redes sociales, telemetría, documentos u otras fuentes semiestructuradas/no estructuradas.
Procesamiento	Incumplimiento recurrente del SLA de carga y transformación con PostgreSQL y Power Query.
Analítica avanzada	Necesidad de modelos predictivos a gran escala, procesamiento distribuido o streaming continuo.

Nota. Elaboración propia.

3.8. Conclusiones del capítulo

- Este capítulo transforma la arquitectura conceptual del capítulo anterior en una prueba de concepto funcional y observable.
- La comparación tecnológica incorpora criterios ponderados de coste, integración, seguridad, residencia del dato, capacidades del equipo, dependencia del proveedor y complejidad operativa.
- La secuencia Landing (Fabric) → RAW (Fabric) → ETL/curación → Data Warehouse dimensional (PostgreSQL) → Power BI establece trazabilidad y mantiene intacta la fuente original. La IA consulta los datos depurados y escribe sus predicciones en el mismo Data Warehouse.

- La primera carga FULL ya fue ejecutada correctamente en Microsoft Fabric y los archivos de ventas fueron preservados en formato CSV dentro de RAW.
- Dataflow Gen2/Power Query constituye la etapa de depuración propuesta; PostgreSQL funciona como Data Warehouse analítico y debe recibir, además de hechos y dimensiones, las predicciones históricas de IA. La carga FULL de Landing a RAW en Fabric es la evidencia técnica efectivamente ejecutada; los pasos posteriores deben distinguirse como implementación o propuesta según su verificación.
- La solución actual debe describirse como Business Intelligence; la adopción de tecnologías de Big Data se justificará únicamente si aumentan materialmente el volumen, velocidad, variedad o exigencias de procesamiento.

CAPÍTULO 4: Inteligencia artificial integrada al Data Warehouse: riesgo de margen y sobrecostos

Desde el Capítulo 1 de la Parte I se identificó el problema central de Directech Cía. Ltda.: la imposibilidad de conocer con oportunidad el margen real de cada proyecto, producto o servicio frente a la variabilidad de sus costos. Los indicadores levantados lo confirman: el margen bruto se ubica alrededor del 15 % frente a una meta de al menos 25 %, y la variación de costos supera el +20 % respecto de lo presupuestado, frente a una meta de hasta +5 %.

Los capítulos 1 a 3 de esta Parte sustentan la solución sobre un Data Warehouse dimensional en PostgreSQL, una arquitectura Cloud de referencia y una prueba de concepto en Microsoft Fabric para la recepción de archivos en Landing y RAW. El Data Warehouse constituye la fuente analítica única para los KPI históricos, mientras Fabric/OneLake proporciona datos preparados y un entorno de experimentación de IA. La integración posterior de Dataflow Gen2, predicciones y alertas corresponde a una evolución propuesta que debe verificarse con datos operativos antes de presentarse como implementada.

Se incorpora una capa de inteligencia artificial con dos usos complementarios: (1) el Modelo de Riesgo de Margen (MRM), que clasifica operaciones susceptibles de cerrar con margen bruto inferior a la meta comercial, y (2) una extensión progresiva de riesgo de sobrecostos para proyectos en ejecución, orientada a estimar la probabilidad de superar el presupuesto y el costo final esperado. Ambos resultados se diseñan para almacenarse históricamente en PostgreSQL, vinculados con proyectos, partidas, fechas y versiones del modelo, y consultarse desde Power BI. Se conserva la estructura metodológica de objetivos, fuentes, áreas, mejora esperada y business case.

4.1. Objetivo y alcance

El objetivo general es diseñar una capa de IA conectada al Data Warehouse de DIRECTECH, capaz de evaluar el riesgo de margen inferior al 25 % en operaciones y, una vez construido y validado un histórico suficiente, estimar el riesgo de sobrecostos y el costo final de proyectos en ejecución. Las predicciones deben persistir con fecha y versión en PostgreSQL y visualizarse en Power BI como apoyo a decisiones humanas, no como autorizaciones automáticas.

Los objetivos específicos son:

- Definir las variables objetivo de riesgo de margen y riesgo de sobrecosto, distinguiendo indicadores observados de resultados estimados por IA.
- Utilizar datos depurados de las capas Fabric/Power Query y un histórico por proyecto y partida, con el Data Warehouse PostgreSQL como origen/destino analítico gobernado.
- Establecer el área de aplicación, los departamentos implicados y las decisiones que el modelo habilita en cada perfil de usuario.
- Identificar la mejora esperada frente al proceso actual de cotización, compra y seguimiento de proyectos.

- Cuantificar la mejora mediante un business case incremental respecto del sistema de BI ya justificado en el Capítulo 5 de la Parte I.

El alcance comprende el diseño del MRM, el código de entrenamiento y scoring ilustrado en el Anexo A1, su arquitectura de integración y un diseño complementario para sobrecostos. El código se probó con datos de prueba de estructura equivalente, lo que demuestra viabilidad técnica y no precisión predictiva sobre DIRECTECH. El modelo para uso operativo requiere históricos reales, entrenamiento, validación temporal por proyecto y aprobación de sus métricas. El almacenamiento definitivo de resultados se define en bi.fact_predicciones_ia del Data Warehouse; la tabla Delta pred_riesgo_margen funciona únicamente como salida intermedia del notebook. La IA no sustituye PostgreSQL ni Power BI y no implica por sí misma una implementación Big Data.

4.2. Objetivo de aplicación del modelo de data science y machine learning

4.2.1. Problema de negocio que resuelve

En DIRECTECH, la propuesta comercial se elabora con costos estándar y con cotizaciones de proveedores que pueden cambiar entre la fecha de la oferta y la fecha de compra, en especial en equipos importados y en proyectos de mayor duración. A ello se suman los descuentos concedidos durante la negociación. El resultado es una erosión del margen que hoy solo se conoce al cierre, a través de los reportes de Latinium, cuando ya no es posible corregir el precio ni renegociar con el proveedor.

Hasta ahora, la revisión del riesgo de una cotización depende de la experiencia del vendedor y del monto de la operación. Con el modelo propuesto, esa revisión se apoya en un método predictivo que utiliza el comportamiento histórico de operaciones similares para anticipar su resultado, de la misma forma en que el caso de referencia pasó de segmentar por días de mora a segmentar por propensión estimada.

4.2.2. Modelo propuesto: Modelo predictivo de Riesgo de Margen (MRM)

Tabla 48

Definición del modelo predictivo de riesgo de margen

Elemento	Definición para DIRECTECH
Nombre	Modelo predictivo de Riesgo de Margen (MRM)
Tipo de problema	Clasificación binaria supervisada, con probabilidad como salida
Unidad de análisis	Operación o partida de proyecto para riesgo de margen; proyecto y corte temporal para la extensión de sobrecostos, sin mezclar proyectos entre entrenamiento y prueba
Variable objetivo	MRM: margen_bruto_real < 25 % (clasificación binaria). Extensión: costo_final_real > 1,05 × presupuesto_vigente_aprobado; requiere histórico de proyectos y validación independiente.
Momento de uso	Etapa 1: cotización · Etapa 2: aprobación y compra · Etapa 3: hitos de ejecución
Salida	MRM: probabilidad de margen bajo y nivel de riesgo. Extensión, tras validación: probabilidad de sobrecosto, costo final estimado y margen estimado. Las predicciones se versionan en PostgreSQL.

Elemento	Definición para DIRECTECH
Usuarios del resultado	Ventas, Coordinación de Proyectos, Área Administrativa y Contable, Área Técnica y Gerencia General

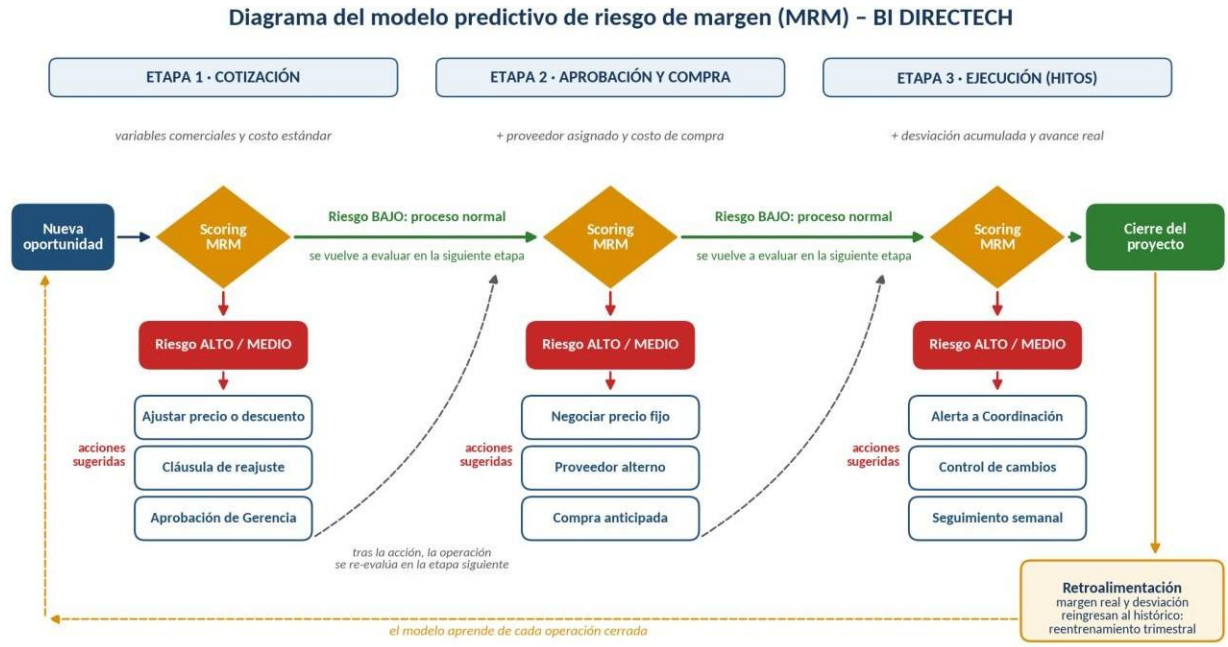
Nota. Elaboración propia.

La unidad de análisis evita entrenar únicamente con 5 a 10 proyectos anuales. Se exploran operaciones y partidas presupuestarias, que aportan más observaciones, pero múltiples partidas de un mismo proyecto no equivalen a proyectos independientes. Para evaluar sobrecostos se deberá separar el entrenamiento y prueba por proyecto y período, evitando que partidas del mismo proyecto aparezcan simultáneamente en ambos conjuntos. La matriz sintética académica demuestra el circuito de datos, pero no puede establecer por sí sola rendimiento predictivo ni ahorros reales.

4.2.3. Diagrama del modelo

Figura 21

Diagrama de aplicación del modelo MRM en las etapas de una operación



Nota. Elaboración propia.

El modelo ejecuta su rutina en cada etapa de la operación, de forma análoga a los buckets de mora del caso de referencia. Cuando la probabilidad es baja, la operación sigue el proceso normal y vuelve a evaluarse en la etapa siguiente, con más información disponible. Cuando el riesgo es medio o alto, el sistema sugiere acciones propias de la etapa: ajustar precio o descuento en la cotización, negociar precio fijo o buscar un proveedor alterno en la compra, y reforzar el seguimiento durante la ejecución. Al cierre, el margen real se reincorpora al histórico y el modelo se reentrena cada trimestre, de modo que aprende de cada operación concluida.

4.2.4. Enfoque técnico y criterios de aceptación

Se comparan dos algoritmos de la librería scikit-learn (Pedregosa et al., 2011). La regresión logística sirve como modelo de referencia interpretable, cuyos coeficientes pueden explicarse al Área de Ventas. El random forest permite capturar interacciones no lineales, por ejemplo entre el origen importado del producto y la volatilidad reciente de su costo. Las variables se construyen únicamente con información disponible antes de la operación: el costo real (costo_total) se utiliza para construir la variable objetivo y nunca como predictor, lo que evita la fuga de información hacia el modelo (Provost & Fawcett, 2013). La partición entre entrenamiento y prueba es temporal (80 % de las operaciones más antiguas para entrenar y 20 % de las más recientes para evaluar), de modo que el modelo se evalúe en las mismas condiciones en que se usará. El desbalance entre clases se corrige con ponderación de clases.

Tabla 49

Criterios de aceptación del modelo

Métrica	Qué mide	Umbral mínimo
AUC-ROC	Capacidad del modelo para ordenar las operaciones de mayor a menor riesgo	≥ 0,70
Recall (clase de riesgo)	Porcentaje de operaciones con margen bajo que el modelo logra alertar	≥ 70 %
Precisión (clase de riesgo)	Porcentaje de alertas que efectivamente terminan con margen bajo	≥ 50 %
Métrica de negocio	Margen real de las operaciones alertadas y corregidas frente a las no alertadas	Mejora sostenida por trimestre

Nota. Elaboración propia.

Si los modelos no alcanzan los criterios de aceptación con registros reales o no existe suficiente histórico, no se publicarán probabilidades como si fueran confiables. En ese caso funcionarán alertas determinísticas basadas en variación de costos, CPI y límites de margen, claramente identificadas como reglas de negocio. La evaluación se hará con separación temporal y agrupación por proyecto, comparando contra un modelo base y verificando también calibración y falsos positivos.

4.3. Fuentes de datos utilizadas

Las variables se extraerán de los hechos y dimensiones ya centralizados en PostgreSQL y, durante la preparación, de las tablas depuradas del Lakehouse de Fabric. Para la predicción de costos se requiere además un histórico que relacione el presupuesto aprobado, el avance real, las compras, los costos imputados, las horas técnicas y los cambios de alcance por proyecto y partida. No se consideran disponibles variables que todavía no hayan sido registradas y validadas. Dado que una misma operación puede originarse en Directech y facturarse a través de EDUTECH (Capítulo 1 de esta Parte), las variables de ingreso y margen se toman del modelo integrado de ambas empresas, de modo que el riesgo se evalúa sobre el margen completo de la operación.

4.3.1. Variables disponibles en el Lakehouse

Tabla 50

Variables del modelo disponibles en las tablas del proyecto

Tabla origen	Campos utilizados	Variable del modelo	Aporte al modelo
fact_ventas	cantidad, descuento_pct, precio_unitario, subtotal_neto, canal_venta, forma_pago, vendedor_id	cantidad, descuento, precio_vs_lista, margen_cotizado, canal, forma de pago, vendedor	Condiciones comerciales pactadas
fact_ventas	costo_total	Variable objetivo (margen real)	Solo para entrenar; nunca como predictor
dim_producto	categoria, marca, costo_estandar, precio_lista	Categoría (línea de negocio), marca, costo estándar	Exposición a la variación de costos según el tipo de bien
fact_compras + dim_fecha	costo_unitario, fecha, proveedor_id	var_costo_180d, cv_costo_180d, n_proveedores_180d, n_compras_180d	Tendencia y volatilidad del costo en los 180 días previos a la operación
dim_proveedor	origen	pct_importado_180d	Riesgo asociado a importaciones y plazos
dim_cliente	tipo_cliente, region	Sector (público o privado) y región	Condiciones contractuales y logísticas
dim_fecha	mes_numero	Mes de la operación	Estacionalidad
fact_devoluciones	motivo, cantidad	Tasa histórica de devoluciones (fase 2)	Costos de reproceso y garantías

Nota. Elaboración propia.

4.3.2. Nuevo bloque de información: historial de proyectos

Se propone reconstruir, inicialmente, al menos dos años de historial de proyectos en una tabla de trabajo hist_proyectos, cruzando presupuesto aprobado por partida (Excel de Proyectos), compras y costos ejecutados (Latinium), avance físico y horas técnicas (Área Técnica). Este conjunto se depurará en Fabric y se publicará como histórico gobernado en el Data Warehouse PostgreSQL. Su disponibilidad real, calidad y suficiencia estadística deberán comprobarse: contar con pocas decenas de proyectos puede impedir entrenar un predictor confiable aun cuando existan muchas partidas.

Tabla 51

Estructura del nuevo bloque de información hist_proyectos

Campo	Origen	Uso en el modelo
Código de proyecto, cliente y línea de negocio	Excel de Proyectos	Agrupación y variables de contexto
Partida y presupuesto aprobado por partida	Excel de Proyectos	Costo presupuestado (base de comparación)

Campo	Origen	Uso en el modelo
Costo real por partida	Latinium (compras y costos)	Etiqueta de sobrecosto al cierre (> 5 % del presupuesto aprobado); nunca predictor de un corte anterior
Fechas de cotización, compra y cierre	Excel y Latinium	Exposición temporal a cambios de precio
Cambios de alcance aprobados	Coordinación de Proyectos	Separar el sobrecosto controlable del contractual
Horas técnicas presupuestadas y reales	Área Técnica	Desviación de mano de obra por partida

Nota. Elaboración propia.

Este histórico permite definir una segunda etiqueta: riesgo_sobrecosto = 1 cuando el costo final real de un proyecto o partida supera el presupuesto vigente y aprobado en más del 5 %; de lo contrario, 0. Para evitar fuga temporal, cada predicción de ejecución solo puede utilizar información conocida hasta la fecha evaluada. El costo final real se reserva como etiqueta de entrenamiento y como criterio de evaluación, nunca como variable predictora.

4.3.3. Fuentes externas

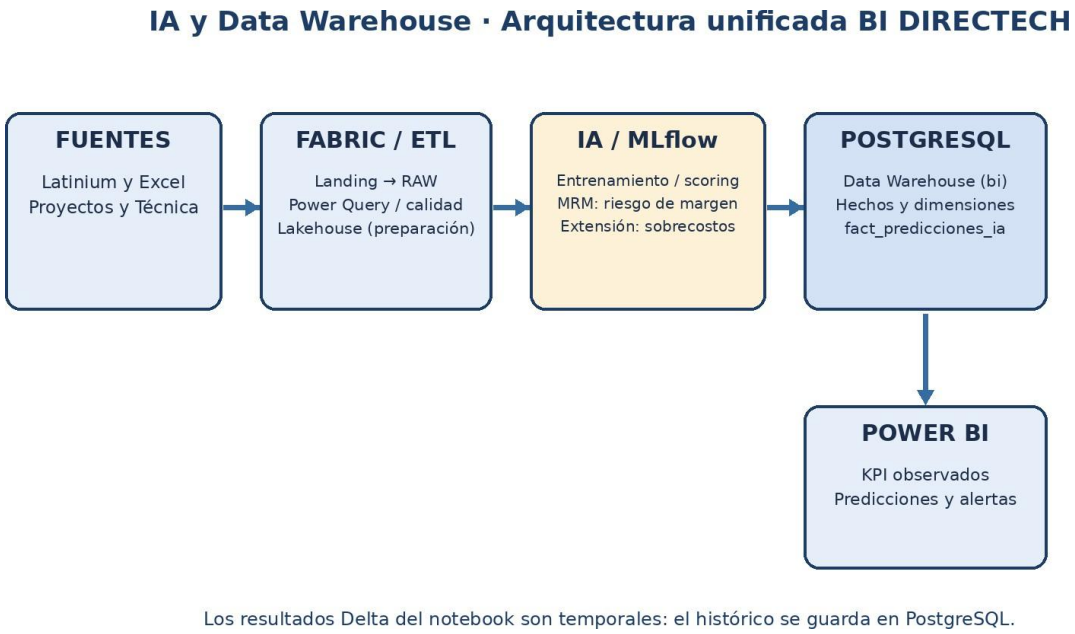
Como enriquecimiento posterior se consideran el Índice de Precios al Consumidor y los índices de precios de importación publicados por el INEC y el Banco Central del Ecuador, identificados como fuentes potenciales desde el Capítulo 2 de la Parte I. Se incorporarán únicamente después de la primera versión del modelo y si demuestran mejorar las métricas, para no añadir complejidad sin beneficio comprobado.

4.3.4. Implementación en Microsoft Fabric y Power BI

La arquitectura integra Microsoft Fabric como zona de ingesta, preparación y entrenamiento, con PostgreSQL como Data Warehouse de hechos, dimensiones y predicciones. El notebook puede producir una tabla Delta temporal de scoring, pero un proceso de validación y carga incremental debe persistir sus resultados en bi.fact_predicciones_ia antes de la actualización de Power BI. La siguiente figura refleja esta distribución de responsabilidades.

Figura 22

Arquitectura de implementación de IA en Microsoft Fabric y Power BI



Nota. Elaboración propia.

Tabla 52

Pasos de implementación en Microsoft Fabric y Power BI

Paso	Componente	Resultado
1. Publicar tablas depuradas	Dataflow Gen2 con destino adicional en Lakehouse (sección Tables, formato Delta)	Tablas de hechos y dimensiones disponibles para notebooks
2. Preparar variables y entrenar	Notebook NB_MRM_Entrenamiento (PySpark y scikit-learn) o interfaz AutoML low-code	Dos modelos candidatos comparados
3. Registrar experimentos	Experimento MLflow EXP_RIESGO_MARGEN	Métricas y parámetros por corrida
4. Versionar el modelo	Elemento ML Model MRM_DIRECTECH	Versión aprobada (AUC ≥ 0,70)
5. Calificar operaciones abiertas	Notebook NB_MRM_Scoring, programado semanalmente en un pipeline	Tabla Delta temporal pred_riesgo_margen, con fecha de evaluación y versión de modelo
6. Publicar en el Data Warehouse	ETL incremental con validación y UPSERT en PostgreSQL, esquema bi	bi.fact_predicciones_ia (histórico de predicciones por fecha, entidad y versión)
7. Consumir y compartir en Power BI	Power BI Desktop/Service con consulta al esquema bi de PostgreSQL; control de acceso por perfil	Dashboard con métricas observadas, predicciones IA y alertas

Nota. Elaboración propia.

Dado que el Capítulo 5 de la Parte I identificó que ningún colaborador tiene experiencia previa en programación analítica, Fabric ofrece también una interfaz AutoML low-code. Esta interfaz genera un notebook preconfigurado a partir de la tarea elegida (clasificación binaria), registra automáticamente las corridas en un experimento MLflow y guarda el mejor modelo como elemento ML Model (Microsoft, 2026g). La Coordinación de BI puede utilizarla para reentrenar el modelo sin modificar código, mientras el recurso externo mantiene el notebook del Anexo A1.

El seguimiento de experimentos y versiones se apoya en MLflow (Zaharia et al., 2018), integrado de forma nativa en Fabric (Microsoft, 2026j). Para el scoring se utiliza la probabilidad que entrega el modelo, necesaria para la matriz de priorización; cuando el volumen de datos crezca, Fabric ofrece la función PREDICT para calificar lotes de forma distribuida con el mismo modelo registrado (Microsoft, 2026f).

En Power BI, el visual Key influencers complementa el modelo porque muestra qué factores se asocian con una mayor probabilidad de riesgo, lo que facilita explicar las alertas a Ventas. Las funciones de Copilot no forman parte de la prueba de concepto: requieren una capacidad de pago F2 o superior y no están disponibles en la capacidad de prueba que utiliza actualmente el proyecto (Microsoft, 2026h).

4.3.5. Persistencia y gobierno de las predicciones en el Data Warehouse

Para evitar un segundo almacén analítico no gobernado, los resultados de los modelos se consolidarán en PostgreSQL, esquema bi, tabla fact_predicciones_ia. La granularidad propuesta será una predicción por entidad (proyecto, partida u operación), fecha de corte, horizonte y versión de modelo. Como mínimo se conservarán: prediccion_id, empresa_id, entidad_tipo, entidad_id, proyecto_id, partida_id, operacion_id, fecha_prediccion, horizonte, tipo_prediccion, probabilidad_margen_bajo, probabilidad_sobrecosto, costo_final_estimado, margen_estimado, nivel_riesgo, version_modelo, id_ejecucion, tipo_datos y fecha_carga. Las probabilidades no calculadas se almacenarán como NULL, nunca como cero.

El flujo previsto es Lakehouse/Delta (salida temporal del scoring) → validaciones de calidad y claves → carga incremental/UPSERT a bi.fact_predicciones_ia → modelo semántico de Power BI. Cada reejecución debe ser idempotente y conservar el histórico por corte y versión. Los datos sintéticos se etiquetarán para impedir que se mezclen con decisiones reales; las versiones de modelos no aprobadas no se mostrarán como predicciones operativas. La publicación física en PostgreSQL forma parte de la fase posterior al prototipo, y el Anexo A4 aporta el esquema SQL de referencia.

4.4. Área de aplicación y departamentos de negocio implicados

El área de aplicación principal es el ciclo comercial y de ejecución de proyectos, desde la cotización hasta el cierre. Es el punto en que se origina la brecha de margen y en el que confluyen Ventas, Proyectos, Compras y Gerencia. La tabla siguiente vincula cada departamento con el uso del modelo y con los perfiles de usuario definidos en el Capítulo 4 de la Parte I.

Tabla 53

Departamentos implicados y decisiones que habilita el modelo

Departamento	Responsable	Uso del modelo	Decisión que habilita	Perfil BI
Dirección Comercial / Ventas	Ing. Jessica Aguilar / Lic. Leslie Soria	Consulta la probabilidad de riesgo al preparar cada cotización	Ajustar precio o descuento; incluir cláusula de reajuste	Operativo – Comercial
Coordinación de Proyectos (Coordinación de BI)	Lic. Leslie Soria	Opera el modelo y da seguimiento al riesgo durante la ejecución	Seguimiento semanal y control de cambios de alcance	Táctico – Proyectos
Administrativa y Contable (compras)	Ing. José Ortega / Ing. Evelin Arteaga	Valida costos y proveedores en las operaciones de riesgo	Negociar precio fijo, proveedor alterno o compra anticipada	Táctico – Financiero
Área Técnica	Ing. Guido Soria / Tnlgo. Kevin Díaz / Tnlgo. Lenin Ponce	Aporta horas y materiales reales al histórico	Ajustar la estimación técnica de las partidas	Operativo – Técnico
Gerencia General	Ing. Juan Soria / Ing. Guido Soria	Revisa las operaciones de riesgo alto y monto mayor	Aprobar, renegociar o declinar la operación	Estratégico

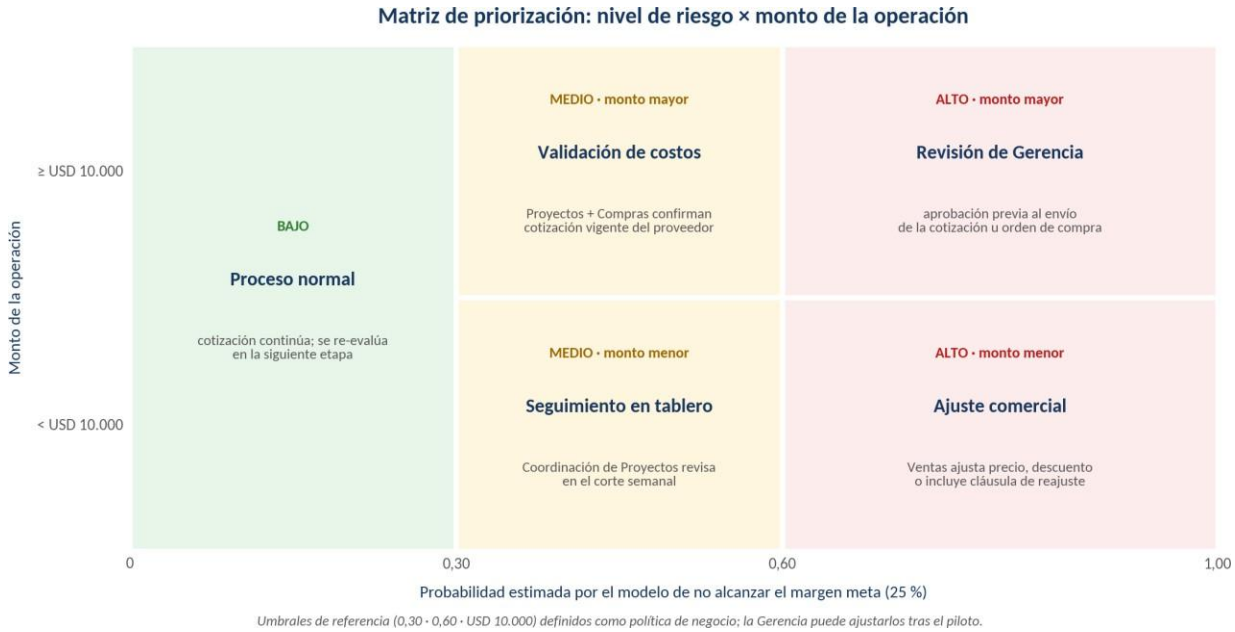
Nota. Elaboración propia.

4.4.1. Segmentación y priorización de las operaciones

Actualmente, las operaciones se revisan según su monto y la experiencia del responsable. Con el modelo, la priorización combina la probabilidad estimada con el monto de la operación, de modo que el esfuerzo de revisión se concentra donde el riesgo y el impacto económico son mayores.

Figura 23

Matriz de priorización de operaciones según riesgo estimado y monto



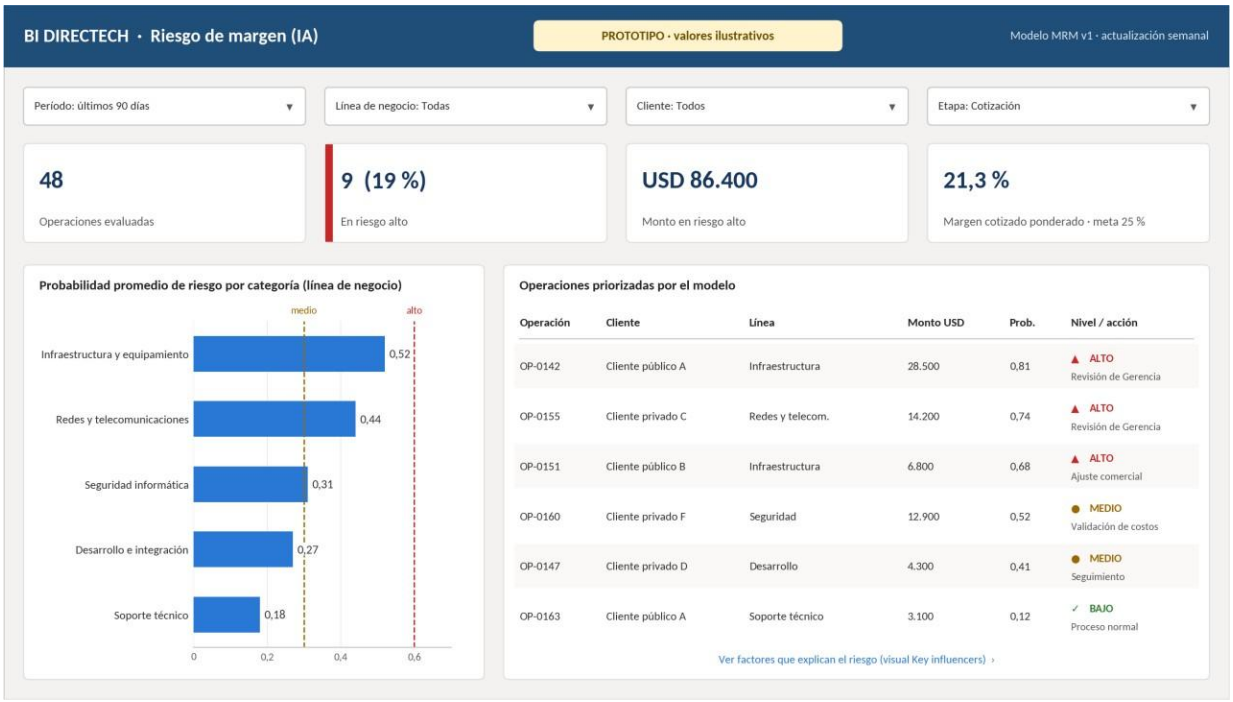
Nota. Elaboración propia. Umbrales de referencia definidos como política de negocio.

4.4.2. Consumo de los resultados en Power BI

Power BI consultará los resultados desde la tabla bi.fact_predicciones_ia del Data Warehouse PostgreSQL, relacionados con las dimensiones de proyecto, operación, cliente y tiempo. La tabla temporal pred_riesgo_margen del Lakehouse se considera salida de scoring, no fuente definitiva del tablero. La Figura 24 corresponde a un prototipo con datos ilustrativos, no a resultados de predicciones reales; los permisos se controlarán por rol y, cuando corresponda, mediante RLS. Las medidas DAX del Anexo A2 se adaptarán al nombre definitivo de la tabla cargada en Power BI.

Figura 24

Prototipo del tablero «Riesgo de margen» en Power BI



Nota. Elaboración propia. Valores ilustrativos; no corresponden a resultados del modelo.

4.4.3. Extensión de la IA al pronóstico de sobrecostos durante la ejecución

El MRM del Anexo A1 clasifica riesgo comercial de margen, pero no produce por sí mismo una probabilidad de sobrecosto ni una estimación de costo final de proyecto. Para incorporar ambos resultados se propone un segundo modelo supervisado, entrenado sobre cortes históricos de proyectos terminados, cuyo objetivo sea riesgo_sobrecosto = 1 cuando costo_final_real > 1,05 × presupuesto_aprobado_vigente. Entre los predictores admisibles al corte figuran porcentaje de avance, valor ganado, costo real acumulado, CPI, compromisos de compra, horas técnicas ejecutadas, plazo consumido, cambios de alcance aprobados y línea de negocio. La variable costo_final_real se reservará exclusivamente como etiqueta.

El costo_final_estimado podrá obtenerse en una primera etapa mediante una referencia transparente de valor ganado, por ejemplo EAC = presupuesto_aprobado_vigente / CPI cuando CPI sea positivo y el supuesto de estabilidad resulte razonable. Este cálculo será identificado como estimación por regla, no como IA. Un modelo de regresión podrá sustituirlo solamente después de comprobar menor error en proyectos externos al entrenamiento (MAE y error porcentual, con control por tamaño de proyecto). Las probabilidades requerirán métricas de discriminación y calibración, no solo AUC. Si los datos son insuficientes, Power BI mostrará reglas de alerta y ninguna probabilidad artificial.

4.5. Identificación de la mejora esperada

Tabla 54

Situación actual frente a la situación con el modelo

Aspecto	Situación actual	Con el modelo MRM
Momento en que se detecta el riesgo	Al cierre del proyecto, en los reportes de Latinium	Antes de enviar la cotización y en cada etapa
Criterio de priorización	Experiencia del vendedor y monto de la operación	Probabilidad estimada × monto (matriz de priorización)
Revisión de cotizaciones	Manual y homogénea (aprox. 8 h al mes)	Enfocada en riesgo medio y alto (aprox. 3 h al mes)
Aprendizaje organizacional	Depende de la memoria de las personas	Cada cierre alimenta el reentrenamiento del modelo
Trazabilidad de las decisiones	No se registra por qué se aceptó una operación	Versión del modelo, probabilidad y acción por operación

Nota. Elaboración propia.

Tabla 55

Impacto esperado sobre los KPI del proyecto

Indicador	Actual (Parte I)	Meta empresa	Aporte esperado del modelo (escenario base)
Margen bruto	~15 %	≥ 25 %	+0,85 pp, adicional al aporte del BI
Variación de costos	> +20 %	≤ +5 %	Evita ~7,4 % del sobrecosto residual
Operaciones de riesgo alto revisadas antes de cotizar	No medido	100 %	100 % desde el piloto
Tiempo de revisión de cotizaciones	~8 h al mes	—	~3 h al mes

Nota. Elaboración propia.

Conviene precisar el alcance de esta mejora: el modelo no cierra por sí solo la brecha de 10 puntos de margen. Esa brecha también depende de la política de precios, de las condiciones contractuales con clientes públicos y de la negociación con proveedores. El aporte del modelo consiste en dirigir la atención de la organización hacia las operaciones donde esas decisiones tienen mayor efecto, y hacerlo antes de que el margen se pierda.

4.6. Business case de la implementación de IA

El business case se construye como incremental respecto del presentado en el Capítulo 5 de la Parte I. Ese caso ya atribuyó al sistema de BI una reducción de sobrecostos de USD 24.300 anuales por detectar desviaciones durante la ejecución. El modelo actúa antes, en la cotización y la compra, y únicamente sobre el sobrecosto que el BI no alcanza a evitar. Por eso ese valor no se vuelve a contabilizar.

4.6.1. Supuestos

Tabla 56

Supuestos del business case

Supuesto	Valor	Fuente
Ingresos anuales	USD 900.000	Parte I
Costos directos (margen bruto ~15 %)	USD 765.000	Parte I
Sobrecosto anual implícito (variación de +20 % sobre lo presupuestado)	USD 127.500	Cálculo: 765.000 − 765.000 / 1,20
Reducción ya atribuida al sistema de BI	USD 24.300	Business case del Capítulo 5, Parte I
Sobrecosto residual sobre el que actúa el modelo	USD 103.200	Cálculo
Costo hora-hombre interno / consultor externo	USD 12 / USD 20	Capítulo 5, Parte I / supuesto
Capacidad Fabric F2 de pago por uso	USD 0,36 por hora; ~40 h al mes encendida	Microsoft (2026i)
Adopción en el año 1	50 % del beneficio (3 meses de implementación y adopción gradual)	Supuesto conservador
Tasa de descuento	12 % anual	Supuesto

Nota. Elaboración propia.

4.6.2. Costos de la implementación

Tabla 57

Costos incrementales de la implementación de IA

Concepto	Tipo	USD
Desarrollo del modelo y notebook (externo, 60 h × USD 20)	Único	1.200
Reconstrucción del histórico de 2 años (interno, 40 h × USD 12)	Único	480
Capacitación en Fabric Data Science (Coordinación de BI)	Único	300
Capacidad Fabric F2 pago por uso (~40 h/mes × USD 0,36)	Anual	173
Operación y monitoreo (Coordinación de BI, 4 h/mes × USD 12)	Anual	576
Reentrenamiento y soporte externo (10 h × USD 20)	Anual	200
Inversión inicial (pago único)	—	1.980
Costo recurrente anual	—	949
Total año 1	—	2.929

Nota. Elaboración propia. Las licencias de Power BI Pro ya están presupuestadas en el Capítulo 5 de la Parte I (actualizadas a USD 14 por usuario al mes), por lo que no se incluyen como costo incremental.

4.6.3. Beneficios por escenario

Tabla 58

Beneficios anuales estimados por escenario

Concepto	Conservador	Base	Optimista
Reducción de la erosión de margen (% de costos directos)	0,5 %	1,0 %	2,0 %
Margen protegido (USD por año)	3.825	7.650	15.300
Porcentaje del sobrecosto residual evitado	3,7 %	7,4 %	14,8 %
Ahorro en revisión de cotizaciones (5 h al mes × USD 12)	720	720	720
Total beneficio anual en régimen (USD)	4.545	8.370	16.020

Nota. Elaboración propia.

4.6.4. Resultados

Tabla 59

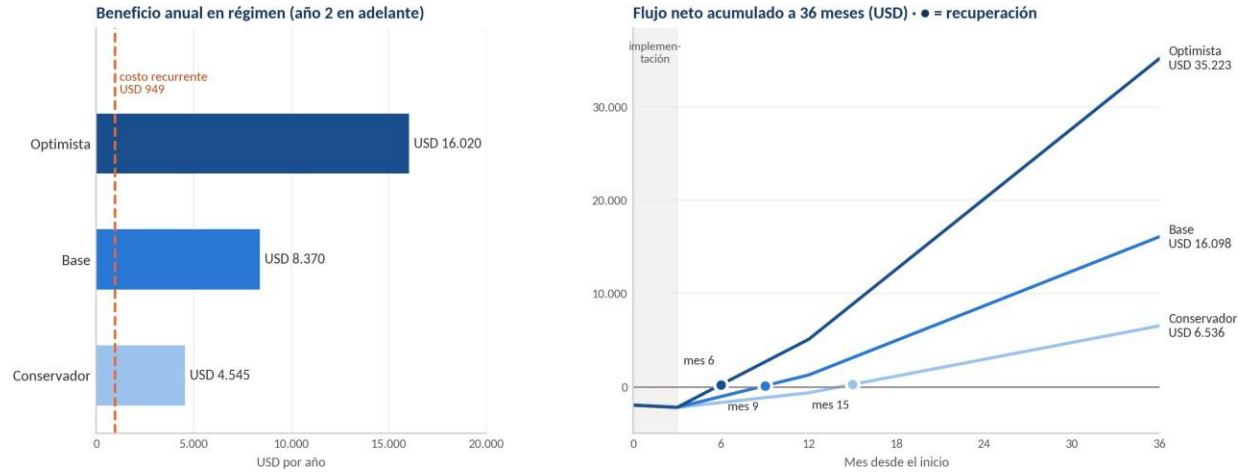
Resultados del business case por escenario

Indicador	Conservador	Base	Optimista
Flujo neto año 1 (USD)	−656	1.256	5.081
Flujo neto años 2 y 3 (USD, cada uno)	3.596	7.421	15.071
Flujo neto acumulado a 3 años (USD)	6.536	16.098	35.223
VAN a 12 %, 3 años (USD)	4.840	12.320	27.278
Relación beneficio/costo (3 años)	2,35	4,33	8,30
Periodo de recuperación	Mes 15	Mes 9	Mes 6
Aporte al margen bruto	+0,43 pp	+0,85 pp	+1,70 pp

Nota. Elaboración propia.

Figura 25

Beneficio anual en régimen frente al costo recurrente, y flujo neto acumulado a 36 meses por escenario



Nota. Elaboración propia.

En el escenario base, el modelo protege USD 7.650 de margen al año, genera un VAN de USD 12.320 a tres años y recupera la inversión en el mes 9. Incluso en el escenario conservador, que supone evitar apenas el 3,7 % del sobrecosto residual, el VAN es positivo (USD 4.840) y la inversión se recupera en el mes 15.

4.6.5. Punto de equilibrio

Para cubrir su costo recurrente anual (USD 949), el modelo necesita evitar USD 229 de erosión de margen al año, adicionales al ahorro de tiempo; esa cifra equivale al 0,03 % de los costos directos. Para recuperar todo el costo del primer año (USD 2.929), necesita evitar USD 2.209, es decir, el 0,29 % de los costos directos. Una sola cotización de monto mayor corregida a tiempo puede cubrir ese valor. Como en el business case del Capítulo 5 de la Parte I, estos beneficios solo se materializan si las alertas del modelo se traducen en acciones; un modelo cuyas recomendaciones no se aplican no genera valor.

4.7. Riesgos y uso responsable de la IA

Tabla 60

Riesgos de la implementación y medidas de mitigación

Riesgo	Mitigación
Histórico reducido (5 a 10 proyectos al año)	Unidad de análisis por operación y partida; algoritmos simples; umbral mínimo de AUC y regla de negocio de respaldo
Calidad de los datos (costos mal imputados, fechas inválidas)	Reglas de calidad y salida de RECHAZADOS en el Dataflow Gen2 definidas en el Capítulo 3
Deriva del modelo (cambios de proveedores o de precios)	Reentrenamiento trimestral; monitoreo del AUC y de la tasa real de riesgo

Riesgo	Mitigación
Decisiones automáticas sin criterio humano	El modelo recomienda y la persona responsable decide; la acción tomada queda registrada
Baja explicabilidad para los usuarios	Regresión logística como referencia; visual Key influencers; variables comprensibles para Ventas
Protección de datos	Datos principalmente empresariales (B2B); acceso por perfil mediante RLS; tratamiento de datos personales conforme a la Ley Orgánica de Protección de Datos Personales (2021)
Dependencia de la capacidad de prueba de Fabric	Planificar una capacidad F2 de pago por uso con pausa fuera de horario; Copilot queda fuera del alcance

Nota. Elaboración propia.

4.8. Plan de implementación

Tabla 61

Plan de implementación del modelo MRM

Fase	Semanas	Actividades	Entregable
0. Datos	1 – 3	Publicar las tablas en el Lakehouse; reconstruir hist_proyectos con dos años de historia	Histórico trazable y tablas preparadas; replicación al DW y controles de calidad
1. Modelo	4 – 6	Entrenar, comparar y registrar el MRM	Modelo v1 y experimento MLflow
2. Integración	7 – 8	Programar scoring; persistir versiones y resultados en bi.fact_predicciones_ia; conectar Power BI	Histórico de predicciones DW y tablero piloto
3. Piloto	9 – 12	Validar con proyectos reales no utilizados en entrenamiento; monitorear errores y registrar acciones	Métricas verificadas, aceptación o no del modelo y resultados del piloto
Operación	Continua	Scoring semanal, reentrenamiento trimestral y monitoreo de métricas	Nuevas versiones del modelo

Nota. Elaboración propia.

4.9. Conclusiones del capítulo

El MRM es una extensión de analítica predictiva para identificar anticipadamente operaciones con riesgo de margen bajo. Para materializar también el objetivo de la tesis sobre rentabilidad de proyectos, se diseña una segunda aplicación dirigida al riesgo de sobrecostos y costo final estimado durante la ejecución. No se afirmará que esta segunda aplicación ya se encuentra entrenada sobre datos reales: su despliegue está condicionado a un histórico suficiente, a la validación sin fuga de información y a la evaluación con proyectos no utilizados para entrenar.

Las variables de ambos casos proceden del modelo comercial-financiero y del historial de proyectos. Fabric concentra ingesta, curación y experimentos, mientras el Data Warehouse PostgreSQL conserva hechos, dimensiones y las predicciones versionadas en

bi.fact_predicciones_ia. Esta distinción elimina la duplicidad de repositorios de resultados y garantiza que Power BI pueda mostrar el indicador observado y su pronóstico con la misma trazabilidad.

La arquitectura admite notebooks o AutoML para entrenamiento, MLflow para versiones y un proceso programado que califique nuevas observaciones. La salida Delta pred_riesgo_margen es intermedia y se transfiere al Data Warehouse, donde cada registro preserva proyecto/partida, fecha de evaluación, horizonte, tipo de predicción, versión del modelo y fuente. Power BI consume el Data Warehouse; las alertas operativas pueden apoyarse en reglas calculadas antes de contar con un modelo predictivo aprobado.

El business case de IA corresponde a escenarios académicos incrementales respecto del BI básico; las cifras de protección de margen no representan ahorros comprobados. El escenario base presenta USD 7.650 anuales de margen potencialmente protegido y un VAN modelado de USD 12.320 a tres años, sujetos a calidad de datos, validación del modelo, adopción y respuesta humana. Los costos de operación y almacenamiento incremental deberán confirmarse antes del despliegue empresarial.

CAPÍTULO 5: Caso de uso Near Real Time para el control de costos y rentabilidad de proyectos

A lo largo del proyecto se ha identificado que uno de los principales problemas de Directech Cía. Ltda. es la dificultad para conocer oportunamente la rentabilidad real de cada proyecto tecnológico. El sistema de BI de la Parte I consolida la información y el modelo de IA del Capítulo 4 anticipa el riesgo en la cotización y la compra. Sin embargo, durante la ejecución los costos siguen llegando con retraso: la auxiliar contable registra a diario, el contador valida una vez por semana y el Área de Proyectos actualiza sus archivos semanalmente, de modo que un costo puede tardar hasta siete días en ser visible para quien debe actuar sobre él (Capítulo 3, Parte I).

El material de referencia distingue el procesamiento Real Time, en el que cada dato se procesa en el momento en que se genera, del Near Real Time, en el que la información se procesa en intervalos cortos y periódicos (Pérez Galán, 2026e). Para la realidad de Directech, el Near Real Time es la alternativa viable: las fuentes actuales no generan flujos continuos de eventos mediante API, y el ciclo de un proyecto se mide en semanas, por lo que reducir la latencia de días a horas captura la mayor parte del valor sin el costo de una arquitectura de streaming. Este capítulo desarrolla el caso de uso siguiendo la estructura de la guía: objetivo de aplicación, fuentes de datos, área de aplicación y departamentos implicados, mejora esperada y business case.

5.1. Identificación y definición del objetivo de aplicación del modelo Near Real Time

5.1.1. Caso de uso propuesto

El caso de uso se denomina «Sistema Near Real Time de monitoreo y alerta temprana de desviaciones de costos y rentabilidad de proyectos de Directech».

Su objetivo es reducir el tiempo entre el registro de un costo y su visibilidad para los responsables, de hasta siete días a un máximo de cuatro horas hábiles, recalculando en cada ciclo los indicadores de rentabilidad de cada proyecto en ejecución y generando alertas antes de que el sobrecosto se consolide como pérdida al cierre. El caso de uso actúa directamente sobre el KPI más crítico del proyecto: la variación de costos, que hoy supera el +20 % frente a una meta de hasta +5 % (Capítulo 1, Parte I).

5.1.2. Lógica de funcionamiento

1. Capturar, cada cuatro horas hábiles, los nuevos registros de compras, costos, facturación, materiales y horas de trabajo asociados a proyectos en ejecución (microcarga incremental).
2. Relacionar cada registro con el código del proyecto, su presupuesto aprobado, su porcentaje de avance y la línea de negocio correspondiente.
3. Recalcular en cada ciclo los indicadores de costo y rentabilidad del proyecto.
4. Clasificar el proyecto en un estado de seguimiento: normal, preventivo o crítico.

5. Actualizar el tablero en Power BI y enviar una alerta a los responsables cuando se supere un umbral.
6. Recalcular la clasificación MRM y, solo si la extensión de sobrecostos supera su validación, generar una nueva predicción de costo final y riesgo de sobrecosto, conservando ambas con fecha y versión en bi.fact_predicciones_ia.

5.1.3. Medición de la desviación con base en el avance

Comparar el costo acumulado con el presupuesto total no detecta el sobrecosto a mitad de un proyecto: un proyecto al 40 % de avance que ya consumió el 55 % de su presupuesto aparece todavía «por debajo» del total. Por ello, la desviación se mide con el método de valor ganado (Project Management Institute, 2021), que compara el costo real con el presupuesto correspondiente al avance logrado:

- Valor ganado (VG) = presupuesto aprobado del proyecto × porcentaje de avance real.
- Variación de costos sobre avance (%) = (costo real acumulado − VG) / VG × 100.
- Índice de desempeño de costos (CPI) = VG / costo real acumulado; un CPI menor a 1 indica que el proyecto gasta más de lo que avanza.

En el ejemplo anterior, con un presupuesto de USD 50.000, el valor ganado es USD 20.000 y el costo real USD 27.500: la variación sobre avance es +37,5 % y el CPI 0,73, lo que activa una alerta crítica que el método tradicional no habría mostrado hasta el cierre.

Tabla 62

Indicadores recalculados en cada ciclo Near Real Time

Indicador	Función dentro del modelo
Costo presupuestado y valor ganado	Línea base económica del proyecto, ajustada al avance real
Costo real acumulado	Cuánto se ha gastado hasta el último ciclo
Variación de costos sobre avance y CPI	Detectar desviaciones respecto de lo que se debía haber gastado a la fecha
Margen bruto proyectado	Rentabilidad operativa estimada al cierre si se mantiene el ritmo de gasto
Margen neto, rentabilidad por proyecto, ROI y utilidad	Seguimiento de los KPI definidos en el Capítulo 1 de la Parte I
Probabilidad de riesgo del MRM (etapa de ejecución)	MRM: probabilidad de margen bajo; extensión de IA, tras validación: probabilidad de sobrecosto y costo final estimado

Nota. Elaboración propia con base en Project Management Institute (2021).

5.1.4. Reglas de alerta propuestas para el prototipo

Los umbrales siguientes son una propuesta de diseño y deberán calibrarse con la información histórica de la tabla hist_proyectos definida en el Capítulo 4, para evitar tanto el exceso de alertas como las detecciones tardías.

Tabla 63

Reglas de alerta del caso de uso Near Real Time

Estado	Variación de costos sobre avance	Acción sugerida
Normal	≤ 3 % (CPI ≥ 0,97)	Continuar el seguimiento del proyecto.
Preventivo	> 3 % y ≤ 5 %	Revisar compras, recursos, horas y avance del proyecto.
Crítico	> 5 % (CPI < 0,95)	Alerta a Proyectos, Finanzas y Gerencia para acción inmediata.

Nota. Elaboración propia. El umbral crítico coincide con la meta del KPI de variación de costos (≤ +5 %).

5.2. Fuentes de datos utilizadas

Las fuentes son las ya identificadas en los capítulos 2 y 3 de la Parte I; lo que cambia en este caso de uso es el mecanismo y la frecuencia de captura. Un esquema Near Real Time no es posible si la información depende de exportaciones manuales, por lo que cada fuente requiere un mecanismo de captura definido. Esto incluye la facturación emitida por EDUTECH en los proyectos que se ofertan por su intermedio, sin la cual el margen recalculado en cada microcarga quedaría incompleto.

Tabla 64

Fuentes de datos y mecanismo de captura del caso de uso

Fuente	Evento de costo que aporta	Mecanismo de captura propuesto	Latencia objetivo
Latinium	Compra registrada, factura de proveedor, costo y facturación al cliente	Lectura programada de la base de datos o de una exportación automática, si el proveedor del software lo habilita; en su defecto, exportación diaria estandarizada a la carpeta compartida	≤ 4 h hábiles (con conexión) / ≤ 1 día (con exportación)
Excel – Proyectos	Avance real, presupuesto por partida y cambios de alcance	Plantilla única en repositorio compartido (OneDrive/SharePoint) leída por Dataflow Gen2 en cada ciclo	≤ 4 h hábiles
Registros del Área Técnica	Horas trabajadas y materiales utilizados	Formulario digital (Microsoft Forms o Power Apps) que registra cada parte de trabajo en una tabla	≤ 4 h hábiles
Maestros de clientes y proyectos	Código de proyecto, cliente, línea de negocio y responsable	Lectura en cada ciclo; cambios solo por altas o modificaciones	Según cambio
hist_proyectos (Capítulo 4)	Base histórica para calibrar umbrales	Actualización al cierre de cada proyecto	Trimestral

Nota. Elaboración propia.

La orquestación Near Real Time se propone como evolución de los componentes utilizados en la prueba de concepto del Capítulo 3: un pipeline programado trasladaría archivos desde Landing hacia RAW, Dataflow Gen2 validaría y depuraría cambios y el Data Warehouse PostgreSQL recibiría hechos y predicciones versionadas. La prueba acreditada actualmente corresponde a la carga FULL Landing → RAW; ni las microcargas automáticas cada cuatro

horas ni la integración directa con Latinium deben describirse como implementaciones ya verificadas. Cada ciclo dependerá de que el dato de origen exista y llegue a tiempo.

La frecuencia de cuatro horas hábiles (tres a cuatro ciclos diarios) es deliberada: permite mantener las licencias de Power BI Pro ya presupuestadas, cuyo número de actualizaciones programadas diarias es limitado (Microsoft, 2026a), mientras que una frecuencia horaria exigiría una capacidad superior sin un beneficio proporcional, dado que el ciclo de un proyecto se mide en semanas. El principal requisito de la fuente Latinium —confirmar con el proveedor si permite lectura directa o exportación automática— queda como condición previa de la implementación.

5.3. Área de aplicación y departamentos de negocio implicados

El área de aplicación principal es la gestión y control de proyectos en ejecución. El caso de uso es transversal, porque la rentabilidad de un proyecto depende de información comercial, financiera y operativa, y se apoya en los perfiles de usuario del Capítulo 4 y en el mecanismo de escalamiento del Capítulo 5 de la Parte I.

Tabla 65

Departamentos implicados, alertas y tiempos de respuesta

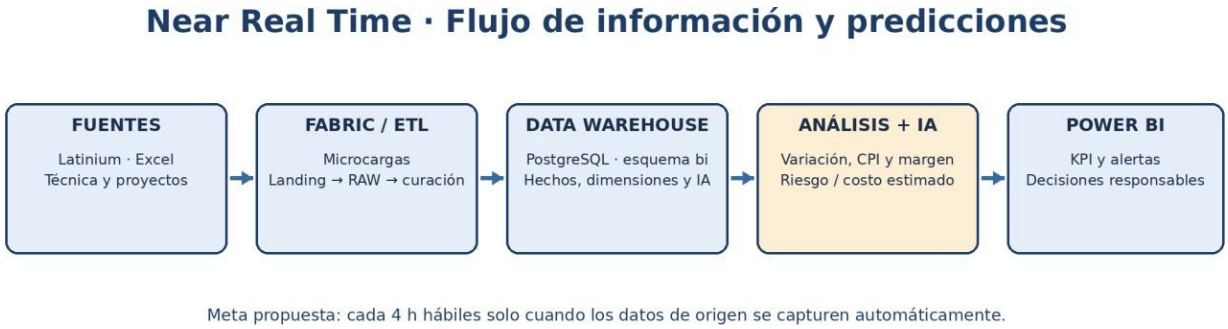
Área	Perfil BI	Participación en el modelo	Alerta que recibe	Tiempo de respuesta
Coordinación de Proyectos (Coordinación de BI)	Táctico – Proyectos	Área principal: supervisa desviaciones, coordina acciones y opera las microcargas	Preventiva y crítica	Preventiva: 48 h · Crítica: 24 h
Administrativa y Financiera	Táctico – Financiero	Valida el impacto económico y la calidad de los registros de Latinium	Crítica	24 h
Área Técnica	Operativo – Técnico	Registra horas y materiales en el formulario digital; propone ajustes técnicos	Preventiva de sus proyectos	48 h
Ventas / Comercial	Operativo – Comercial	Aporta condiciones comerciales; evalúa reajustes o cambios de alcance con el cliente	Crítica con impacto contractual	48 h
Gerencia General	Estratégico	Decide ante alertas críticas: renegociar, redistribuir recursos o aprobar el sobrecosto	Crítica (resumen diario)	24 h

Nota. Elaboración propia. Los tiempos de respuesta son una propuesta inicial que la Gerencia y la Coordinación de BI deberán aprobar.

Una alerta crítica que no se resuelva en el plazo se escala según los tres niveles definidos en el Capítulo 5 de la Parte I. Cada acción tomada se registra junto a la alerta, lo que permite medir posteriormente su efecto sobre el margen.

Figura 26

Arquitectura conceptual del caso de uso Near Real Time de BI DIRECTECH



Nota. Elaboración propia. El Lakehouse de Fabric recibe y prepara datos; PostgreSQL constituye el Data Warehouse analítico central, incluida la tabla bi.fact_predicciones_ia. La arquitectura Near Real Time es una propuesta de evolución, no un despliegue productivo confirmado.

5.4. Identificación de la mejora esperada

La mejora esperada del Near Real Time consiste en disminuir la latencia entre la generación de un costo y su disponibilidad para decidir. El objetivo de cuatro horas hábiles solo sería alcanzable con exportaciones automáticas o integración de origen y cargas programadas efectivamente disponibles; si Latinium mantiene exportaciones manuales diarias, su información tendrá al menos esa latencia. La integración del histórico, la actualización de Power BI y el scoring de IA deberán medirse por separado, sin equiparar frecuencia de consulta con frecuencia real de los datos.

Tabla 66

Mejora esperada del caso de uso Near Real Time

Indicador	Situación con el BI (cargas diarias y validación semanal)	Con Near Real Time
Tiempo entre el registro de un costo y su visibilidad	Hasta 7 días	Meta ≤ 4 h hábiles, solo con automatización efectiva de cada fuente; si no, latencia propia de la exportación
Momento de detección de una desviación	Corte semanal o cierre del proyecto	Siguiente microcarga
Base de comparación	Costo acumulado frente a presupuesto total	Costo real frente a valor ganado (CPI)
Tiempo de respuesta ante alerta crítica	No definido	≤ 24 horas
Proyectos con desviación > 5 % detectada antes del 50 % de avance	No se mide	Meta: 100 %

Indicador	Situación con el BI (cargas diarias y validación semanal)	Con Near Real Time
Riesgo de margen del MRM en ejecución	Recalculado semanalmente	MRM o extensión de sobrecostos recalculados en microcarga únicamente tras validación de los modelos
Variación de costos (KPI)	> +20 %	Contribuye a la meta ≤ +5 %

Nota. Elaboración propia.

Por ejemplo, si el precio de un equipo importado sube cuando el proyecto lleva un 30 % de avance, el costo aparece en el tablero en la siguiente microcarga y el CPI del proyecto cae por debajo de 0,95. Con esa alerta, Directech puede negociar el resto de las compras, sustituir proveedores, controlar las horas adicionales o tramitar un reajuste con el cliente, en lugar de descubrir la pérdida al cierre.

5.5. Cuantificación de la mejora mediante business case

El business case se plantea como incremental para no sumar dos veces el mismo ahorro. La Parte I atribuye al BI básico USD 24.300 anuales de reducción potencial de sobrecostos y el Capítulo 4 asigna al caso de IA, bajo el escenario base, USD 7.650 adicionales de margen potencialmente protegido. El capítulo Near Real Time modela un tercer beneficio exclusivamente sobre el remanente, sujeto a capturas más oportunas y respuesta a alertas. Estas cifras son supuestos de planificación, no resultados empresariales verificados.

5.5.1. Supuestos

Tabla 67

Supuestos del business case Near Real Time

Supuesto	Valor	Fuente
Ingresos anuales	USD 900.000	Parte I
Costos directos (margen bruto ~15 %)	USD 765.000	Parte I y Capítulo 4
Sobrecosto residual tras BI e IA (escenario base)	USD 95.550	103.200 − 7.650 (Capítulo 4)
Frecuencia de actualización	Cada 4 horas hábiles (3 a 4 ciclos diarios)	Supuesto de diseño
Latencia actual del dato de costo	Hasta 7 días	Capítulo 3, Parte I
Costo hora-hombre interno / consultor externo	USD 12 / USD 20	Parte I / Capítulo 4
Capacidad Fabric F2 de pago por uso	USD 0,36 por hora	Microsoft (2026i)
Adopción en el año 1	50 % del beneficio	Igual que el Capítulo 4
Tasa de descuento	12 % anual	Igual que el Capítulo 4

Nota. Elaboración propia.

5.5.2. Costos incrementales

Tabla 68

Costos incrementales del caso de uso Near Real Time

Concepto	Tipo	USD
Configuración de microcargas incrementales, tabla de control y reglas de alerta (externo, 30 h × USD 20)	Único	600
Plantilla estandarizada de Proyectos y formulario de registro del Área Técnica (interno, 20 h × USD 12)	Único	240
Cálculo del valor ganado y calibración de umbrales (interno, 16 h × USD 12)	Único	192
Protocolo de respuesta a alertas y capacitación a las áreas (interno, 10 h × USD 12)	Único	120
Capacidad Fabric F2 adicional en horario laboral (+120 h/mes × USD 0,36 × 12)	Anual	518
Monitoreo de cargas y gestión de alertas (Coordinación de BI, 4 h/mes × USD 12 × 12)	Anual	576
Soporte externo para ajustes (5 h × USD 20)	Anual	100
Inversión inicial (pago único)	—	1.152
Costo recurrente anual	—	1.194
Total año 1	—	2.346

Nota. Elaboración propia. Las licencias de Power BI Pro ya están presupuestadas en el Capítulo 5 de la Parte I; con tres a cuatro actualizaciones diarias no se requiere una licencia superior.

5.5.3. Beneficios por escenario

Tabla 69

Beneficios anuales estimados del caso Near Real Time

Concepto	Conservador	Base	Optimista
Erosión de margen evitada por detección anticipada (% de costos directos)	0,25 %	0,5 %	1,0 %
Margen protegido (USD por año)	1.913	3.825	7.650
Porcentaje del sobrecosto residual evitado	2,0 %	4,0 %	8,0 %
Ahorro en seguimiento manual del avance (3 h al mes × USD 12)	432	432	432
Total beneficio anual en régimen (USD)	2.345	4.257	8.082

Nota. Elaboración propia.

Como referencia de razonabilidad, el sobrecosto residual equivale a unos USD 1.840 por semana (95.550 / 52). Proteger USD 3.825 al año en el escenario base equivale a evitar alrededor de dos semanas de sobrecosto acumulado, coherente con detectar las desviaciones varios días antes que con el esquema actual.

5.5.4. Resultados

Tabla 70

Resultados del business case Near Real Time por escenario

Indicador	Conservador	Base	Optimista
Flujo neto año 1 (USD)	−1.174	−218	1.695
Flujo neto años 2 y 3 (USD, cada uno)	1.151	3.063	6.888
Flujo neto acumulado a 3 años (USD)	1.129	5.908	15.471
VAN a 12 %, 3 años (USD)	689	4.428	11.907
Relación beneficio/costo (3 años)	1,24	2,25	4,27
Periodo de recuperación	Mes 25	Mes 13	Mes 5
Aporte al margen bruto	+0,21 pp	+0,43 pp	+0,85 pp

Nota. Elaboración propia. Flujos mensuales con 50 % de adopción en el año 1; cifras redondeadas.

5.5.5. Punto de equilibrio e interpretación

Para cubrir su costo recurrente anual (USD 1.194), el caso de uso necesita evitar USD 762 de erosión de margen al año, adicionales al ahorro de horas, es decir, el 0,10 % de los costos directos. Para cubrir todo el costo del primer año (USD 2.346) con adopción completa, necesita evitar USD 1.914, el 0,25 % de los costos directos, que corresponde al escenario conservador.

El beneficio incremental del Near Real Time es menor que el del sistema de BI y el del modelo de IA, lo cual es esperable: la mayor parte del valor ya la capturan la consolidación de la información y la prevención en la cotización. Aun así, el caso es positivo en los tres escenarios. Se recomienda implementarlo después del sistema de BI, mediante un piloto de tres meses que mida la latencia real y el margen protegido antes de extenderlo a todos los proyectos. Como en los casos anteriores, el beneficio solo se materializa si las alertas se traducen en acciones dentro de los tiempos de respuesta definidos.

5.6. Alcance y condiciones de implementación

- El caso de uso es una evolución del modelo actual; no implica que Directech disponga hoy de streaming continuo ni de integración por API.
- Es condición previa confirmar con el proveedor de Latinium la posibilidad de lectura directa o exportación automática; sin ella, la latencia de esa fuente será de un día.
- Las áreas deben adoptar la plantilla única de Proyectos y el formulario digital del Área Técnica; sin registro oportuno de los eventos de costo, la mayor frecuencia de carga no aporta valor.
- Los umbrales de alerta y los tiempos de respuesta deben calibrarse con datos históricos y ser aprobados por la Gerencia y la Coordinación de BI.
- Una integración Real Time completa, mediante un CRM, API o eventos, se evaluará solo si el número de proyectos o la velocidad de los costos lo justifican, según los criterios de escalamiento del Capítulo 3.

5.7. Conclusiones del capítulo

- Near Real Time representa una etapa de diseño que busca reducir a cuatro horas hábiles la latencia del dato cuando la fuente permita extracción automática; no corresponde afirmar esa frecuencia para Latinium mientras dependa de exportaciones diarias manuales.
- La medición con valor ganado corrige una limitación del seguimiento actual: permite detectar el sobrecosto a mitad del proyecto, comparando el gasto con el avance y no con el presupuesto total.
- **La solución separa responsabilidades:** BI almacena y visualiza, IA estima riesgos tras su validación y Near Real Time propone actualizar hechos, alertas y predicciones con menor latencia. El destino de resultados es el Data Warehouse PostgreSQL, no un repositorio Delta definitivo independiente.
- La definición de alertas, responsables y tiempos de respuesta por área convierte la información en acciones, apoyada en los perfiles y el escalamiento ya definidos en la Parte I.
- El business case, construido como incremental para no contar dos veces los beneficios del BI y de la IA, es positivo en los tres escenarios: en el base protege USD 3.825 de margen al año, alcanza un VAN de USD 4.428 a tres años y recupera la inversión en el mes 13.

CONCLUSIONES Y APLICACIONES

El proyecto BI DIRECTECH partió de un problema concreto: Directech Cía. Ltda. no podía conocer con precisión el margen de ganancia real de cada proyecto, producto o servicio, porque su información financiera, comercial y operativa estaba dispersa entre Latinium, hojas de cálculo de Excel y registros manuales de cada área. A lo largo de los dos bloques del programa, esta problemática se abordó desde el diagnóstico hasta una prueba de concepto técnica. Los aspectos más importantes del trabajo son los siguientes:

- **Necesidades y KPI.** Todas las áreas de la empresa comparten una misma necesidad: información integrada, confiable y oportuna. Se definieron seis KPI de rentabilidad y costos (margen bruto, margen neto, rentabilidad por proyecto, variación de costos, ROI y utilidad total) y se cuantificaron sus brechas; la más crítica es un sobrecosto actual superior al 20 % frente a una meta de hasta el 5 %.
- **Datos y calidad.** La información existe, pero presenta problemas de consistencia, actualización, completitud y trazabilidad. La capacidad de Latinium para exportar a Excel hace viable un proceso ETL de escala PYME sin inversión inicial en conectores especializados.
- **Arquitectura y usuarios.** Se seleccionó un Data Warehouse relacional con esquema estrella y Power BI como herramienta de análisis, por ser la opción proporcional al volumen, presupuesto y capacidades de la empresa. Se definieron cinco perfiles de usuario con niveles de acceso diferenciados.
- **Organización y viabilidad.** La función de BI se sostiene mediante una Coordinación de BI en doble rol, apoyo interno de Contabilidad y del Área Técnica, y un recurso externo subcontratado. El business case estima un beneficio anual cercano a USD 25.900 frente a una inversión de USD 4.640 a USD 6.240 en el primer año.
- **Construcción y validación técnica.** El modelo integrado EDUTECH – DIRECTECH se implementó en PostgreSQL y se conectó con Power BI; se diseñó una arquitectura Cloud en Azure bajo el modelo PaaS con una Landing Zone, y se ejecutó con éxito la primera carga completa en Microsoft Fabric, con separación Landing → RAW → ETL → PostgreSQL → Power BI y criterios explícitos de trazabilidad y gobierno.
- Big Data como evolución, no como punto de partida. La solución actual es de Business Intelligence; tecnologías de Big Data solo se justificarán si crecen materialmente el volumen, la velocidad o la variedad de los datos, según los umbrales definidos.
- **De describir a anticipar.** El MRM se diseñó para clasificar operaciones con riesgo de margen bruto inferior al 25 %; se añadió una extensión para riesgo de sobrecostos y costo final estimado de proyectos. Fabric permite preparar y entrenar, y PostgreSQL conservará las salidas de IA en bi.fact_predicciones_ia para Power BI. Las métricas y ahorros corresponden a escenarios o pruebas académicas: el uso operativo requiere datos reales, evaluación por proyecto y aceptación formal.
- **Vigilancia Near Real Time.** Se propone medir desviaciones sobre valor ganado, capturar nuevos costos en microcargas y recalcular alertas y resultados de IA. La meta de cuatro horas depende de automatizar las fuentes y no está implementada con las exportaciones

manuales actuales. El business case incremental del caso de uso estima USD 3.825 de margen anual protegido en el escenario base, aún no realizado.

Para aplicar la propuesta en Directech Cía. Ltda. se recomienda: (1) formalizar responsables y permisos de los datos; (2) estandarizar Excel y confirmar la automatización de extracción desde Latinium; (3) consolidar hechos y dimensiones de proyectos y operaciones en el Data Warehouse PostgreSQL; (4) construir un histórico real de proyectos por partida y fecha; (5) comparar reglas de negocio con modelos IA, validando temporalmente por proyecto; (6) publicar cada resultado aprobado en bi.fact_predicciones_ia junto con su versión y fecha; (7) consumir el Data Warehouse desde Power BI; y (8) medir durante un piloto la latencia, la precisión y los ahorros realizados, antes de escalar a Near Real Time o adoptar decisiones comerciales apoyadas en predicciones.

Más allá de Directech, el enfoque desarrollado —diagnóstico de necesidades, KPI con brechas cuantificadas, arquitectura proporcional al tamaño de la empresa y gobierno de datos ligero— es replicable en otras PYME de servicios tecnológicos que enfrentan el mismo problema de información dispersa. Como advierte la literatura, el valor de los datos no está en la tecnología en sí, sino en su uso sistemático para decidir (Davenport & Harris, 2017; Provost & Fawcett, 2013): el éxito del sistema dependerá de que la organización lo mantenga, confíe en sus datos y lo incorpore a sus decisiones cotidianas.

REFERENCIAS

Amazon Web Services. (2026). *AWS documentation: Amazon S3, AWS Glue y Amazon RDS for PostgreSQL*. https://docs.aws.amazon.com/

Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. En *Proceedings of the 11th Conference on Innovative Data Systems Research (CIDR 2021)*. https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf

DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge* (2.ª ed.). Technics Publications.

Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning* (ed. actualizada). Harvard Business Review Press.

Directech Cía. Ltda. (2026). *Directech: Tecnología para ecosistemas educativos y empresariales* [Sitio web]. https://directech.com.ec/

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (2.ª ed.). Analytics Press.

Gartner. (s.f.). Business intelligence (BI). En *Gartner information technology glossary*. https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi

Google Cloud. (2026). *Google Cloud documentation: Cloud Storage, Cloud SQL for PostgreSQL, Dataflow y BigQuery*. https://cloud.google.com/docs

IBM. (2026a). *What is ETL (extract, transform, load)?* https://www.ibm.com/think/topics/etl

IBM. (2026b). *Business intelligence*. https://www.ibm.com/think/topics/business-intelligence

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

Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley.

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

Laudon, K. C., & Laudon, J. P. (2020). *Management information systems: Managing the digital firm* (16.ª ed.). Pearson.

Ley Orgánica de Protección de Datos Personales. (2021, 26 de mayo). Registro Oficial Quinto Suplemento n.º 459 (Ecuador).

Marr, B. (2015). *Big data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. Wiley.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68.

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing* (NIST Special Publication 800-145). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-145

Microsoft. (2026a). *Documentación de Power BI*. Microsoft Learn. https://learn.microsoft.com/power-bi/

Microsoft. (2026b). *Microsoft Learn: Azure Data Lake Storage, Azure Data Factory, Azure Database for PostgreSQL y Power BI Service*. https://learn.microsoft.com/azure/

Microsoft. (2026c). *Microsoft Fabric: OneLake and Lakehouse documentation*. Microsoft Learn. https://learn.microsoft.com/en-us/fabric/onelake/create-lakehouse-onelake

Microsoft. (2026d). *Microsoft Fabric Data Factory: Dataflow Gen2 documentation*. Microsoft Learn. https://learn.microsoft.com/en-us/fabric/data-factory/create-first-dataflow-gen2

Microsoft. (2026e). *Microsoft Fabric Data Factory: Copy job and pipeline documentation*. Microsoft Learn. https://learn.microsoft.com/en-us/fabric/data-factory/copy-job-activity

Microsoft. (2026f). *Model scoring with PREDICT in Microsoft Fabric*. Microsoft Learn. https://learn.microsoft.com/fabric/data-science/model-scoring-predict

Microsoft. (2026g). *Use the low-code AutoML interface in Fabric*. Microsoft Learn. https://learn.microsoft.com/fabric/data-science/low-code-automl

Microsoft. (2026h). *Copilot for Microsoft Fabric and Power BI: FAQ*. Microsoft Learn. https://learn.microsoft.com/fabric/fundamentals/copilot-faq-fabric

Microsoft. (2026i). *Microsoft Fabric pricing*. Microsoft Azure. https://azure.microsoft.com/pricing/details/microsoft-fabric/

Microsoft. (2026j). *Autologging in Synapse Data Science*. Microsoft Learn. https://learn.microsoft.com/fabric/data-science/mlflow-autologging

Microsoft. (2026k). *Real-Time Intelligence in Microsoft Fabric documentation*. Microsoft Learn. https://learn.microsoft.com/fabric/real-time-intelligence/

Mucci, T. (2026). *Data-driven decision making*. IBM. https://www.ibm.com/think/topics/data-driven-decision-making

Nagle, T., Redman, T. C., & Sammon, D. (2017, 11 de septiembre). Only 3% of companies' data meets basic quality standards. *Harvard Business Review*. https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards

Parmenter, D. (2020). *Key performance indicators: Developing, implementing, and using winning KPIs* (4.ª ed.). Wiley.

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

Pérez Galán, J. L. (2026a). *Visualización de información* [Material de guía metodológica, caso Grupo Bancolombia]. Westfield Business School.

Pérez Galán, J. L. (2026b). *Business case y creación del departamento* [Material de guía metodológica]. Westfield Business School.

Pérez Galán, J. L. (2026c). *Arquitectura de la información* [Material académico]. Westfield Business School.

Pérez Galán, J. L. (2026d). *Implementación de inteligencia artificial (IA)* [Material de guía metodológica]. Westfield Business School.

Pérez Galán, J. L. (2026e). *Caso de uso en Real Time* [Material académico]. Westfield Business School.

PostgreSQL Global Development Group. (2026). *PostgreSQL documentation*. https://www.postgresql.org/docs/

Project Management Institute. (2021). *A guide to the project management body of knowledge (PMBOK guide)* (7.ª ed.). Project Management Institute.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Redman, T. C. (2001). *Data quality: The field guide*. Digital Press.

Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, & artificial intelligence: Systems for decision support* (11.ª ed.). Pearson.

The Blue Owls. (2025). *The data value gap: Why only 32% of companies realise tangible benefits from data*. https://theblueowls.com/blog/the-data-value-gap-why-only-32-of-companies-realise-tangible-benefits-from-data/

Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F., & Zumar, C. (2018). Accelerating the machine learning lifecycle with MLflow. *IEEE Data Engineering Bulletin, 41*(4), 39–45.

ANEXOS

Anexo A1. Código del notebook de entrenamiento y scoring en Microsoft Fabric

El código se ejecuta en un notebook de Fabric con el Lakehouse del proyecto como origen predeterminado. Fue verificado de extremo a extremo con datos de prueba que replican los nombres y tipos de las tablas del Capítulo 1 de la Parte II: lectura, construcción de variables, entrenamiento, registro en MLflow, carga del modelo registrado, scoring y escritura de la tabla de resultados. Si el Dataflow Gen2 cambia algún nombre de columna, basta con ajustarlo en la celda 1. En Fabric, cada bloque corresponde a una celda del notebook.

```
# ===== CELDA 1 · Parámetros =====
from pyspark.sql import functions as F
META_MARGEN = 0.25     # meta de margen bruto (KPI, entrega Parte I)
VENTANA_DIAS = 180     # historia de compras previa a cada operación
EXPERIMENTO = "EXP_RIESGO_MARGEN"
MODELO      = "MRM_DIRECTECH"
NUM = ["cantidad", "descuento_pct", "precio_vs_lista", "margen_cotizado", "mes_numero",
       "var_costo_180d", "cv_costo_180d", "n_proveedores_180d", "pct_importado_180d",
       "n_compras_180d"]
CAT = ["categoria", "marca", "tipo_cliente", "region", "canal_venta", "forma_pago", "vendedor_id"]


# ===== CELDA 2 · Lectura de las tablas del Lakehouse =====
ventas   = spark.read.table("fact_ventas")
compras  = spark.read.table("fact_compras")
producto = spark.read.table("dim_producto")
cliente  = spark.read.table("dim_cliente")
provee   = spark.read.table("dim_proveedor")
dfecha   = spark.read.table("dim_fecha")
empresa  = (spark.read.table("dim_empresa")
            .filter(F.upper("nombre_comercial").contains("DIRECTECH"))
            .select("empresa_id"))


# ===== CELDA 3 · Base de operaciones y variable objetivo =====
base = (ventas.join(empresa, "empresa_id")
        .filter(F.col("subtotal_neto") > 0)
        .join(dfecha.select("fecha_id", "fecha", "mes_numero"), "fecha_id")
        .join(producto.select("producto_id", "categoria", "marca",
                              "costo_estandar", "precio_lista"), "producto_id")
        .join(cliente.select("cliente_id", "tipo_cliente", "region"), "cliente_id")
        .withColumn("estado", F.lower("estado_venta"))
        .withColumn("margen_cotizado",
                    1 - F.col("cantidad") * F.col("costo_estandar") / F.col("subtotal_neto"))
        .withColumn("precio_vs_lista", F.col("precio_unitario") / F.col("precio_lista"))
        .withColumn("margen_real", 1 - F.col("costo_total") / F.col("subtotal_neto"))
        .withColumn("riesgo_margen", (F.col("margen_real") < META_MARGEN).cast("int")))


# ===== CELDA 4 · Historia de costos de compra (solo datos anteriores a la operación) =====
cmp = (compras
       .join(dfecha.select("fecha_id",  F.col("fecha").alias("fecha_compra")),  "fecha_id")
       .join(provee.select("proveedor_id", "origen"), "proveedor_id")
       .select("producto_id", "fecha_compra", "proveedor_id",
               F.col("costo_unitario").cast("double").alias("costo_unitario"),
               F.lower("origen").rlike("import|internac|extranj|exterior")
                .cast("int").alias("importado")))
hist = (base.select("venta_id", "producto_id", "fecha",
                    F.col("costo_estandar").cast("double").alias("costo_std"))
        .join(cmp, "producto_id")
        .filter((F.col("fecha_compra") < F.col("fecha")) &
                (F.col("fecha_compra") >= F.date_sub(F.col("fecha"), VENTANA_DIAS)))
        .groupBy("venta_id")
        .agg((F.avg("costo_unitario") / F.first("costo_std") - 1).alias("var_costo_180d"),
             (F.stddev("costo_unitario") / F.avg("costo_unitario")).alias("cv_costo_180d"),
             F.countDistinct("proveedor_id").alias("n_proveedores_180d"),
             F.avg("importado").alias("pct_importado_180d"),
             F.count("*").alias("n_compras_180d")))
datos = (base.join(hist, "venta_id", "left")
         .select("venta_id", "fecha", "cliente_id", "producto_id", "estado", "riesgo_margen",
```

```
                F.col("subtotal_neto").cast("double").alias("monto"),
                *[F.col(c).cast("double").alias(c) for c in NUM],
                *[F.col(c).cast("string").alias(c) for c in CAT])
         .fillna("SIN_DATO", subset=CAT))


# ===== CELDA 5 · Entrenamiento, comparación y registro del modelo =====
import mlflow, mlflow.sklearn
import pandas as pd
from mlflow.models import infer_signature
from sklearn.base import clone
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score, recall_score, precision_score


hist_df = (datos.filter(~F.col("estado").rlike("cancel|pend"))     # solo operaciones cerradas
               .toPandas().sort_values("fecha"))
corte = int(len(hist_df) * 0.8)                                   # partición temporal 80/20
train, test = hist_df.iloc[:corte], hist_df.iloc[corte:]
X_tr, y_tr = train[NUM + CAT], train["riesgo_margen"]
X_te, y_te = test[NUM + CAT], test["riesgo_margen"]


prep = ColumnTransformer([
    ("num", Pipeline([("imp", SimpleImputer(strategy="median")),
                      ("esc", StandardScaler())]), NUM),
    ("cat",  Pipeline([("imp",  SimpleImputer(strategy="most_frequent")),
                      ("ohe", OneHotEncoder(handle_unknown="ignore"))]), CAT)])
candidatos = {
    "regresion_logistica": LogisticRegression(max_iter=1000, class_weight="balanced"),
    "random_forest": RandomForestClassifier(n_estimators=300, min_samples_leaf=5,
                                            class_weight="balanced", random_state=42)}


mlflow.set_experiment(EXPERIMENTO)
mlflow.autolog(disable=True)           # se registra de forma explícita
resultados = {}
for nombre, algoritmo in candidatos.items():
    with mlflow.start_run(run_name=nombre) as run:
        modelo = Pipeline([("prep", clone(prep)), ("clf", algoritmo)]).fit(X_tr, y_tr)
        prob = modelo.predict_proba(X_te)[:, 1]
        pred = (prob >= 0.5).astype(int)
        metricas = {"auc": roc_auc_score(y_te, prob),
                    "recall_riesgo": recall_score(y_te, pred),
                    "precision_riesgo": precision_score(y_te, pred, zero_division=0),
                    "tasa_riesgo_test": float(y_te.mean())}
        mlflow.log_params({"algoritmo": nombre, "n_train": len(train), "n_test": len(test)})
        mlflow.log_metrics(metricas)
        mlflow.sklearn.log_model(modelo,  artifact_path="modelo",
                                 signature=infer_signature(X_te, pred),
                                 input_example=X_te.head(3))
        resultados[nombre] = (metricas["auc"], run.info.run_id)


mejor = max(resultados, key=lambda k: resultados[k][0])
auc, run_id = resultados[mejor]
print({k: round(v[0], 3) for k, v in resultados.items()}, "-> mejor:", mejor)
if auc >= 0.70:                        # criterio mínimo de aceptación
    version = mlflow.register_model(f"runs:/{run_id}/modelo", MODELO).version
    print(f"Modelo {MODELO} registrado, versión {version}")
else:
    print("AUC < 0,70: no se registra; revisar variables o ampliar el histórico")


# ===== CELDA 6 · Scoring de operaciones abiertas y publicación para Power BI =====
# En producción esta celda vive en NB_MRM_Scoring (reutiliza las celdas 1 a 4 con %run)
VERSION = version                      # versión vigente del modelo registrado
modelo = mlflow.sklearn.load_model(f"models:/{MODELO}/{VERSION}")
nuevas = datos.filter(F.col("estado").contains("pend")).toPandas()   # cotizaciones abiertas
nuevas["prob_riesgo"] = modelo.predict_proba(nuevas[NUM + CAT])[:, 1]
nuevas["nivel_riesgo"] = pd.cut(nuevas["prob_riesgo"], [0, 0.30, 0.60, 1.0],
                                labels=["Bajo", "Medio", "Alto"],
                                include_lowest=True).astype(str)
```

```
def accion(r):                          # matriz de priorización (matriz de priorización)
    if r.nivel_riesgo == "Alto":
        return "Revisión de Gerencia" if r.monto >= 10000 else "Ajuste comercial"
    if r.nivel_riesgo == "Medio":
        return "Validación de costos" if r.monto >= 10000 else "Seguimiento en tablero"
    return "Proceso normal"


nuevas["accion"] = nuevas.apply(accion, axis=1)
nuevas["etapa"] = "Cotización"
nuevas["version_modelo"] = str(VERSION)
nuevas["fecha_scoring"] = pd.Timestamp.today().normalize()
columnas = ["venta_id", "fecha", "cliente_id", "producto_id", "categoria", "monto",
            "margen_cotizado", "prob_riesgo", "nivel_riesgo", "accion", "etapa",
            "version_modelo", "fecha_scoring"]
(spark.createDataFrame(nuevas[columnas])
      .write.mode("overwrite").format("delta").saveAsTable("pred_riesgo_margen"))
```

Anexo A2. Medidas DAX del tablero «Riesgo de margen»

```
Operaciones evaluadas = COUNTROWS ( pred_riesgo_margen )


Operaciones riesgo alto =
    CALCULATE ( [Operaciones evaluadas], pred_riesgo_margen[nivel_riesgo] = "Alto" )


% riesgo alto = DIVIDE ( [Operaciones riesgo alto], [Operaciones evaluadas] )


Monto en riesgo alto =
    CALCULATE ( SUM ( pred_riesgo_margen[monto] ), pred_riesgo_margen[nivel_riesgo] = "Alto" )


Probabilidad promedio = AVERAGE ( pred_riesgo_margen[prob_riesgo] )


Margen cotizado ponderado =
    DIVIDE (
        SUMX ( pred_riesgo_margen, pred_riesgo_margen[monto] * pred_riesgo_margen[margen_cotizado] ),
        SUM ( pred_riesgo_margen[monto] )
    )
```

Anexo A3. Evidencias técnicas previstas para la fase de implementación

El código del Anexo A1 fue verificado con datos de prueba; la ejecución del modelo en el workspace de Microsoft Fabric con los datos reales de Directech corresponde a la fase siguiente del proyecto. La tabla define las evidencias que deberán generarse en esa fase y que servirán como criterio de aceptación de la implementación: el modelo se considerará implementado cuando las seis evidencias estén disponibles y revisadas por la Coordinación de BI.

Tabla 71

Evidencias técnicas previstas para la implementación del modelo

Evidencia	Captura prevista	Qué demostrará
E1	Notebook NB_MRM_Entrenamiento ejecutado en el workspace	Preparación de variables y entrenamiento
E2	Experimento EXP_RIESGO_MARGEN con sus corridas	Comparación de algoritmos y métricas (AUC, recall, precisión)

Evidencia	Captura prevista	Qué demostrará
E3	Elemento ML Model MRM_DIRECTECH con su versión	Modelo aprobado y versionado
E4	Carga histórica de predicciones en bi.fact_predicciones_ia (PostgreSQL) y salida intermedia Delta	Persistencia definitiva en Data Warehouse; fecha, entidad, versión y trazabilidad
E5	Consulta desde Power BI al esquema bi de PostgreSQL	Integración Data Warehouse → Power BI con hechos y predicciones
E6	Tablero «Riesgo de margen» y visual Key influencers	Consumo por perfil y explicación de factores

Nota. Elaboración propia.

Anexo A4. Esquema propuesto de almacenamiento de predicciones de IA en PostgreSQL

El siguiente DDL constituye una especificación de implementación, no evidencia de que la tabla se encuentre creada en producción. Las claves de empresa, proyecto y operación se validarán contra las dimensiones reales del Data Warehouse antes del despliegue.

```
CREATE SCHEMA IF NOT EXISTS bi;
CREATE TABLE IF NOT EXISTS bi.fact_predicciones_ia (
    prediccion_id BIGINT GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    empresa_id BIGINT NOT NULL,
    entidad_tipo VARCHAR(20) NOT NULL,
    entidad_id VARCHAR(100) NOT NULL,
    proyecto_id BIGINT,
    partida_id VARCHAR(80),
    operacion_id BIGINT,
    fecha_prediccion TIMESTAMP NOT NULL,
    horizonte VARCHAR(40) NOT NULL,
    tipo_prediccion VARCHAR(40) NOT NULL,
    probabilidad_margen_bajo NUMERIC(7,6),
    probabilidad_sobrecosto NUMERIC(7,6),
    costo_final_estimado NUMERIC(18,2),
    margen_estimado NUMERIC(9,6),
    nivel_riesgo VARCHAR(12),
    version_modelo VARCHAR(60) NOT NULL,
    id_ejecucion VARCHAR(100) NOT NULL,
    tipo_datos VARCHAR(15) NOT NULL CHECK (tipo_datos IN ('sintetico','real')),
    fecha_carga TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP,
    CHECK (probabilidad_margen_bajo BETWEEN 0 AND 1 OR probabilidad_margen_bajo IS NULL),
    CHECK (probabilidad_sobrecosto BETWEEN 0 AND 1 OR probabilidad_sobrecosto IS NULL),
    CHECK (costo_final_estimado >= 0 OR costo_final_estimado IS NULL),
    UNIQUE (empresa_id, entidad_tipo, entidad_id,
        fecha_prediccion, horizonte, tipo_prediccion, version_modelo)
);
-- Claves opcionales de proyecto/partida/operación se vinculan por entidad_tipo
-- + entidad_id; UNIQUE preserva idempotencia por entidad y versión.
```

Criterios mínimos de carga y auditoría: validar código de proyecto y versión del modelo, almacenar fecha y horizonte de cada predicción, utilizar claves naturales normalizadas para evitar duplicados en reejecuciones, registrar filas rechazadas, restringir el acceso a cifras financieras y diferenciar explícitamente datos sintéticos de datos empresariales. Para la implementación real debe completarse el mapeo de claves foráneas y la política de retención del histórico.