



**Escuela de Administración**

**Gobierno de Datos para la Optimización Operativa en FAM TEAM**

**Trabajo de Grado**

**Andrés Felipe Abella Tunjano, David Santiago Castellanos Carrillo, María Paula  
Gómez Zapata**

**Bogotá, D.C.**

**2025**



**Escuela de Administración**

**Gobierno de Datos para la Optimización Operativa en FAM TEAM**

**Trabajo de Grado**

**Andrés Felipe Abella Tunjano, David Santiago Castellanos Carrillo, María Paula  
Gómez Zapata**

**Tutor: Erika Johana Salazar Bethancourt**

**Maestría en Business Analytics**

**Escuela de Administración**

**23 de mayo del 2025**

**Bogotá, Colombia**

**2025**

## Índice de Contenido

<i>Declaración de autonomía:</i> .....	7
<i>Declaración de exoneración de responsabilidad:</i> .....	8
Resumen Ejecutivo .....	9
Glosario.....	10
Palabras Clave.....	11
1. Introducción .....	11
1.1. Importancia del Gobierno de Datos en el Entorno Empresarial Actual.....	11
1.2. Gobierno de Datos para la Optimización Operativa en FAM TEAM .....	11
2. Objetivos .....	13
2.1. Objetivo General.....	13
2.2. Objetivos Específicos: .....	14
3. Alcance del Proyecto Aplicado.....	14
4. Cronograma.....	16
5. Límites del Proyecto .....	18
6. Desarrollo del Proyecto.....	18
6.1. Gobierno de Datos .....	18
<i>6.1.1. Diagnóstico Actual (AS-IS) .....</i>	<i>19</i>
<i>6.1.2. Madurez del Gobierno de Datos .....</i>	<i>20</i>
<i>6.1.3. Madurez de Riesgos Asociados al Estado Actual del Gobierno de Datos ...</i>	<i>23</i>

6.1.4.	<i>Modelo Objetivo del Gobierno de Datos</i> .....	26
6.1.5.	<i>Modelo Objetivo (TO-BE) Del Gobierno de Datos</i> .....	33
6.2.	Arquitectura de Datos Propuesta para FAM TEAM.....	38
6.3.	Estándares, Políticas y Controles Críticos para el Despliegue de la Arquitectura TO-BE	58
7.	Calidad de Datos .....	85
7.1.	Diagnóstico de Calidad de Datos – Archivo Cartera .....	86
8.	Modelo .....	93
9.	Validación de Estrategia de Proyecto Empresarial – Metodología TDSP+Lego .....	99
10.	Descripción de las Fuentes de Información .....	102
11.	Referencias Bibliográficas .....	104

### Índice de Tablas

<b>Tabla 1:</b>	<i>Cronograma de actividades</i> .....	16
<b>Tabla 2:</b>	<i>Niveles de Madurez del Gobierno de Datos</i> .....	21
<b>Tabla 3:</b>	<i>Subcategorías para evaluar el Gobierno de Datos</i> .....	21
<b>Tabla 4:</b>	<i>Heatmap – Evaluación de madurez del Gobierno de Datos de FAM TEAM</i> ....	22
<b>Tabla 5:</b>	<i>Calificación DAMA</i> .....	24
<b>Tabla 6:</b>	<i>Definición de roles</i> .....	31
<b>Tabla 7:</b>	<i>Propuesta distribución de roles</i> .....	32
<b>Tabla 8:</b>	<i>Matriz RACI</i> .....	32

<b>Tabla 9:</b> <i>Propuestas de actividades para el Gobierno de Datos 1</i> .....	47
<b>Tabla 10:</b> <i>Propuestas de actividades para el Gobierno de Datos 2</i> .....	49
<b>Tabla 11:</b> <i>Propuestas de actividades para el Gobierno de Datos 3</i> .....	50
<b>Tabla 12:</b> <i>Propuestas de actividades para el Gobierno de Datos 4</i> .....	51
<b>Tabla 13:</b> <i>Propuestas de actividades para el Gobierno de Datos 5</i> .....	54
<b>Tabla 14:</b> <i>Propuestas de actividades para el Gobierno de Datos 6</i> .....	55
<b>Tabla 15:</b> <i>Propuestas de actividades para el Gobierno de Datos 7</i> .....	56
<b>Tabla 16:</b> <i>Tabla diagnóstico de Calidad de Datos</i> .....	86
<b>Tabla 17:</b> <i>Indicadores de Calidad de Datos</i> .....	87
<b>Tabla 18:</b> <i>Tabla de acciones propuestas para mejorar la Calidad de Datos</i> .....	90

### Índice de Ilustraciones

<b>Ilustración 1:</b> <i>Interacción Gobierno de Datos</i> .....	30
<b>Ilustración 2:</b> <i>Macroproceso de Registro, Manipulación y Uso de Datos</i> .....	39
<b>Ilustración 3:</b> <i>Arquitectura de datos FAMTEAM</i> .....	41
<b>Ilustración 4:</b> <i>Arquitectura de datos propuesta</i> .....	46
<b>Ilustración 5:</b> <i>Roadmap de transición AS IS a TO BE</i> .....	77
<b>Ilustración 6:</b> <i>Implementación proceso de captura de datos en Gol</i> .....	83
<b>Ilustración 7:</b> <i>Implementación del proceso ETL automatizado</i> .....	83
<b>Ilustración 8:</b> <i>Repositorio estructurado en Azure</i> .....	84
<b>Ilustración 9:</b> <i>Estado actual de consulta de datos</i> .....	85
<b>Ilustración 10:</b> <i>Evaluación modelo de datos</i> .....	97
<b>Ilustración 11:</b> <i>Matriz de correlaciones</i> .....	97

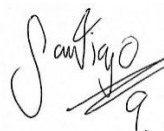
<b>Ilustración 12:</b> <i>Importancia de las variables en el modelo</i> .....	98
<b>Ilustración 13:</b> <i>Estudio y conclusiones del modelo de datos</i> .....	98

***Declaración de autonomía:***

“Declaramos bajo gravedad de juramento, que hemos escrito la presente tesis de maestría por nuestra propia cuenta, y que, por lo tanto, su contenido es original. Declaramos que hemos indicado clara y precisamente todas las fuentes directas e indirectas de información, y que esta tesis de maestría no ha sido entregada a ninguna otra institución con fines de calificación o publicación”.



Andrés Felipe Abella Tunjano, 29 de julio del 2025.



David Santiago Castellanos Carrillo, 29 de julio del 2025.



Maria Paula Gómez Zapata, 29 de julio del 2025.

***Declaración de exoneración de responsabilidad:***

“Declaramos que la responsabilidad intelectual del presente trabajo es exclusivamente de sus autores. La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él”.



Andrés Felipe Abella Tunjano, 29 de julio del 2025.



David Santiago Castellanos Carrillo, 29 de julio del 2025.



Maria Paula Gómez Zapata, 29 de julio del 2025.

## Resumen Ejecutivo

El proyecto “Gobierno de Datos para la Optimización Operativa en FAM TEAM” tiene como objetivo transformar la gestión de datos en FAM TEAM, una empresa líder en soluciones logísticas para el comercio exterior. La compañía enfrenta actualmente desafíos en la eficiencia operativa debido a la falta de un sistema de gobierno de datos unificado. Cada base de datos se maneja con formatos y convenciones variables, lo que dificulta la interoperabilidad y limita el potencial de análisis efectivo en toda la organización.

Este proyecto propone la implementación de un marco integral de gobierno de datos para estandarizar la recopilación, administración y uso de la información. A través de este marco, FAM TEAM podrá asegurar la calidad y coherencia de sus datos, facilitando su integración y optimización en áreas clave. Además, el proyecto incluye el desarrollo de visualizaciones de datos personalizadas y modelos analíticos para los departamentos comerciales y de cartera, habilitando la toma de decisiones informada y en tiempo real.

La implementación de un gobierno de datos permitirá a FAM TEAM mejorar significativamente su eficiencia operativa, reduciendo errores y redundancias en los procesos, y garantizando una mayor alineación de la información entre sus equipos. Las visualizaciones en tiempo real ayudarán a los equipos comerciales y de cartera a monitorear y responder rápidamente a los indicadores de rendimiento, identificando oportunidades de mejora y potencial de crecimiento.

En conjunto, este proyecto posicionará a FAM TEAM como una empresa de vanguardia en el uso estratégico de datos, optimizando sus operaciones y reforzando su capacidad para responder a las demandas del mercado con una base sólida de toma de decisiones.

## Glosario

**Gobierno de Datos:** Conjunto de prácticas, políticas y procedimientos que garantizan la gestión adecuada de los datos en una organización, asegurando su calidad, seguridad, y accesibilidad. Fuente: Data Governance Institute.

**Eficiencia Operativa:** Capacidad de una organización para optimizar sus procesos y recursos de manera que se logren los máximos resultados con los menores costos y esfuerzos posibles.

**Integración de Datos:** Proceso de combinar datos de diferentes fuentes en un sistema único y coherente, facilitando el acceso y análisis conjunto de información para mejorar la toma de decisiones. Fuente: IBM.

**Visualización de Datos:** Representación gráfica de datos mediante gráficos, tablas o dashboards que permite a los usuarios interpretar y analizar la información de manera eficiente. Fuente: Data.org.

**Bases de Datos Comerciales:** Conjunto de datos que contiene información relevante para el área comercial de FAM TEAM, incluyendo detalles de clientes, proveedores, y transacciones de ventas.

**Estándares de Datos:** Normas que establecen el formato, nombre y organización de los datos en una empresa, facilitando su comprensión y uso uniforme a través de diferentes áreas y sistemas.

**Dashboard Interactivo:** Plataforma visual que permite la consulta en tiempo real de datos relevantes, con gráficos y métricas clave que facilitan el análisis y la toma de decisiones informadas. Fuente: Tableau.

**Datos Transaccionales:** Información que registra cada operación comercial o logística de una empresa, como envíos, pedidos y entregas en el contexto de FAM TEAM.

**Calidad de Datos:** Grado de precisión, consistencia, y confiabilidad de los datos en una organización, esencial para el desarrollo de análisis y toma de decisiones basadas en información real y precisa. Fuente: ISO/IEC 25012.

### **Palabras Clave**

Gobierno de datos, eficiencia operativa, integración de datos, calidad de datos, visualización de datos, toma de decisiones, estandarización, análisis de datos, optimización comercial, logística y datos

## **1. Introducción**

### **1.1. Importancia del Gobierno de Datos en el Entorno Empresarial Actual**

En un entorno donde los datos se han convertido en uno de los activos más valiosos para las organizaciones, contar con un sistema sólido de gobierno de datos no es solo una ventaja competitiva, sino una necesidad estratégica. El gobierno de datos permite garantizar que la información sea precisa, accesible, coherente y segura a lo largo de toda la empresa, lo cual es esencial para operar con eficiencia, mitigar riesgos y tomar decisiones acertadas. Sin una adecuada gestión, los datos tienden a fragmentarse, perder calidad y generar inconsistencias que afectan directamente el desempeño organizacional.

### **1.2. Gobierno de Datos para la Optimización Operativa en FAM TEAM**

FAM TEAM es un grupo empresarial colombiano que opera de manera integrada en el sector logístico, conformado por tres compañías especializadas: Agentes Portuarios (AGP), Administración Marítima de Contenedores (AMC) y United Logistic Services (ULS). A través de estas empresas, el grupo ofrece una solución logística completa, articulando procesos desde la llegada de embarcaciones a puerto hasta la entrega final de mercancías al cliente, con un enfoque de eficiencia, trazabilidad y control.

AGP desempeña un papel fundamental en el inicio de la cadena logística, actuando como agencia marítima encargada del agenciamiento de motonaves, la gestión documental de importaciones y exportaciones, y la coordinación de operaciones portuarias. Su conocimiento del entorno marítimo y su capacidad operativa permiten garantizar el cumplimiento normativo y la fluidez en las operaciones de carga y descarga, lo cual representa una ventaja estratégica para los clientes y las líneas navieras que atiende.

Una vez los contenedores han sido desembarcados, AMC entra en operación, esta empresa se especializa en la gestión administrativa y física de contenedores, incluyendo procesos como la entrega, devolución, pagos de servicios, control de demoras y administración de patios.

Finalmente, ULS complementa el proceso logístico mediante servicios de transporte multimodal, conectando los puntos de origen y destino de la mercancía a nivel nacional e internacional. Como agente de carga, ULS coordina el movimiento de mercancías por vía marítima, aérea y terrestre. La complementariedad operativa entre AGP, AMC y ULS le permite a FAM TEAM consolidarse como un proveedor de servicios logísticos integrales, ofreciendo una experiencia de servicio que cubre de manera articulada todas las etapas del proceso logístico.

En la era digital, la correcta gestión y uso de datos se ha convertido en un factor crítico para el éxito de las empresas, especialmente en industrias de alta competencia como la logística. FAM TEAM, enfrenta actualmente desafíos significativos en términos de eficiencia operativa debido a la falta de un sistema de gobierno de datos estandarizado. Con bases de datos dispersas y formatos variables que se definen de manera independiente por diferentes áreas de la compañía, la integración y análisis de datos se han vuelto complejos, afectando la capacidad de la empresa para tomar decisiones informadas y rápidas.

El proyecto Gobierno de Datos para la Optimización Operativa en FAM TEAM tiene como objetivo establecer un marco integral de gobierno de datos que permita consolidar y estandarizar la información en toda la organización. A través de la creación de estándares de calidad y consistencia de datos, el proyecto busca eliminar redundancias y optimizar la interoperabilidad de las bases de datos. Además, se desarrollarán visualizaciones de datos específicas para los departamentos comerciales y de cartera, permitiendo a estos equipos acceder a insights relevantes en tiempo real y apoyar decisiones estratégicas basadas en información precisa y actualizada.

Este proyecto representa una oportunidad para que FAM TEAM no solo mejore su eficiencia operativa, sino también para consolidar una cultura de toma de decisiones basada en datos. Con una gestión de datos robusta y herramientas visuales, FAM TEAM podrá mejorar su capacidad de respuesta ante cambios del mercado y posicionarse como una empresa de referencia en el uso estratégico de datos dentro del sector logístico.

## **2. Objetivos**

### **2.1. Objetivo General**

Implementar un marco de gobernanza de datos en FAM TEAM que estandarice y optimice la gestión de información, acompañado de la implementación de modelos analíticos y tableros de indicadores clave de desempeño (KPIs) para las áreas comercial y de cartera, con el fin de mejorar la eficiencia operativa y facilitar una toma de decisiones informada y estratégica.

## **2.2. Objetivos Específicos:**

- Establecer un esquema de gobierno de datos que defina estándares de calidad, seguridad y acceso para garantizar la consistencia y confiabilidad de los datos en toda la organización.
- Diseñar e implementar 1 modelo analítico para segmentar clientes potenciales del área comercial, identificando al 80% de ellos en 3 categorías de valor (alto, medio y bajo), en 6 meses.
- Aumentar la tasa de asertividad de contacto de cartera en 30% en 6 meses, mediante un modelo predictivo que identifique a los clientes con mayor probabilidad de pago (80% de precisión)
- Diseñar e implementar dos tableros de visualización de indicadores clave de desempeño (KPIs): uno para el área comercial y otro para el área de cartera. Estos tableros permitirán el análisis en tiempo real de métricas esenciales y respaldaron la toma de decisiones estratégicas basadas en datos.

## **3. Alcance del Proyecto Aplicado**

El proyecto “Gobierno de Datos para la Optimización Operativa en FAM TEAM” abarcará las siguientes áreas clave de intervención dentro de la organización para asegurar una

mejora sustancial en la eficiencia operativa y en la calidad de la toma de decisiones en los departamentos de interés:

**Diseño y Establecimiento del Marco de Gobierno de Datos:** Esta fase incluye la creación de políticas, procedimientos y estándares de calidad que regulen la gestión de los datos en FAM TEAM. Se implementarán normas de acceso, integridad, y seguridad de la información que aseguren una base de datos confiable y coherente en toda la compañía.

**Estandarización e Integración de Bases de Datos:** Se realizará un análisis exhaustivo de las bases de datos existentes en todas las áreas para identificar inconsistencias y redundancias. A partir de ello, se definirán formatos uniformes y nomenclaturas comunes que faciliten la interoperabilidad entre sistemas. Esta integración permitirá una visión unificada de la información, clave para la toma de decisiones ágil y estratégica.

**Desarrollo de Visualizaciones de Datos y modelos para Áreas Comerciales y de Cartera:** A partir del marco de datos estandarizados, se crearán dashboards, reportes interactivos y 3 modelos analíticos que proporcionen a los equipos comerciales y de cartera acceso a métricas y KPIs en tiempo real y mejoren su toma de decisiones basadas en datos. Estas visualizaciones facilitarán la interpretación rápida de datos relevantes, promoviendo decisiones informadas y la detección oportuna de oportunidades y riesgos en las operaciones.

**Capacitación del Personal y Alineación Organizacional:** Para garantizar una correcta adopción del nuevo sistema de gobierno de datos, se impartirá capacitación a los empleados en el uso de las herramientas y en el entendimiento de los nuevos estándares. Esta formación busca promover una cultura organizacional basada en el análisis de datos y en la optimización operativa.

#### 4. Cronograma

**Tabla 1:** *Cronograma de actividades.*

Fase	Actividad	Duración	Fecha de Inicio	Fecha de Finalización
Fase 1: Análisis y Diagnóstico	Revisión y diagnóstico de las bases de datos actuales	8 semanas	06/08/2024	06/10/2024
Fase 1: Análisis y Diagnóstico	Identificación de necesidades y requerimientos	4 semanas	21/10/2024	28/11/2024
Fase 1: Análisis y Diagnóstico	Presentación del diagnóstico y ajustes necesarios	1 semana	29/11/2024	06/12/2024
Fase 2: Diseño del Marco de Gobierno de Datos	Desarrollo de políticas y estándares de calidad	2 semanas	09/12/2024	23/12/2024
Fase 2: Diseño del Marco de Gobierno de Datos	Definición de roles y accesos	1 semana	02/01/2025	08/01/2025
Fase 2: Diseño del Marco de Gobierno de Datos	Aprobación del marco de gobierno de datos	1 semana	09/01/2025	15/01/2025
Fase 3: Estandarización e Integración de Datos	Revisión y estandarización de variables y formatos	3 semanas	16/01/2025	06/02/2025

Fase 3: Estandarización e Integración de Datos	Centralización e integración de bases de datos	3 semanas	07/02/2025	27/02/2025
Fase 3: Estandarización e Integración de Datos	Pruebas de integración	1 semana	28/02/2025	06/03/2025
Fase 4: Desarrollo de Visualizaciones de Datos	Diseño de dashboards para el área comercial	2 semanas	07/03/2025	21/03/2025
Fase 4: Desarrollo de Visualizaciones de Datos	Diseño de dashboards para el área de cartera	2 semanas	22/03/2025	04/04/2025
Fase 4: Desarrollo de Visualizaciones de Datos	Validación y ajustes de visualizaciones	1 semana	07/04/2025	14/04/2025
Fase 5: Capacitación y Alineación	Entrenamiento de personal en uso de nuevas herramientas	2 semanas	15/04/2025	29/04/2025
Fase 5: Capacitación y Alineación	Evaluación de adopción y ajustes finales	1 semana	30/04/2025	07/05/2025
Fase 6: Cierre del Proyecto	Revisión final y documentación del proyecto	1 semana	08/05/2025	15/05/2025
Fase 6: Cierre del Proyecto	Presentación de resultados y lecciones aprendidas	1 semana	16/05/2025	23/05/2025

Fuente: Elaboración Propia

## **5. Límites del Proyecto**

El proyecto se centrará exclusivamente en la estandarización y visualización de datos para las áreas comerciales y de cartera. Las mejoras derivadas de este gobierno de datos serán aplicadas progresivamente a otros departamentos en el futuro, en función de los resultados obtenidos en las áreas iniciales de enfoque.

Este proyecto se enfoca en transformar la forma en que FAM TEAM gestiona y utiliza su información, estableciendo una base sólida para futuras iniciativas de analítica avanzada y optimización en toda la organización.

## **6. Desarrollo del Proyecto**

### **6.1. Gobierno de Datos**

El presente capítulo tiene como objetivo describir y analizar el estado actual del gobierno de datos en el grupo empresarial FAM TEAM, compuesto por tres compañías que prestan servicios complementarios dentro de la cadena logística: ULS (United Logistics Services), AGP (Agentes Portuarios) y AMC (Agencia Marítima de Contenedores). ULS se enfoca en ofrecer soluciones integrales de transporte internacional, AGP en la gestión operativa portuaria, y AMC en la limpieza y reparación de contenedores.

A partir de un ejercicio de consultoría que incluyó entrevistas, observación directa, revisión de plataformas tecnológicas y análisis documental, se realizó un levantamiento exhaustivo de los procesos y herramientas actuales para el manejo de la información dentro de estas compañías. Este diagnóstico aborda la infraestructura tecnológica, los flujos de trabajo

operativos, las capacidades organizacionales relacionadas con los datos, y los principales riesgos identificados en el ciclo de vida de la información. Todo este análisis se enmarca en el modelo DAMA-DMBOK v2, adoptando sus dimensiones funcionales para establecer una línea base robusta que sirva como punto de partida para el diseño del modelo TO-BE de gobernanza.

### ***6.1.1. Diagnóstico Actual (AS-IS)***

En el estado actual, FAM TEAM cuenta con múltiples áreas administrativas que dependen de los datos para su operación: comercial, tesorería, cartera, contabilidad, finanzas y logística. No obstante, no existe una estructura formal de gobierno de datos, lo que genera altos niveles de dependencia, fragmentación y riesgo operativo.

#### **Infraestructura y acceso**

Actualmente, FAM TEAM opera sobre una infraestructura tecnológica híbrida que combina desarrollos propios, sistemas fragmentados y herramientas de ofimática. Las principales tecnologías identificadas son:

**ERP Interno (GOL):** Plataforma que centraliza parte de la información operativa, incluyendo módulos como TARIF, utilizado por el equipo de Pricing para calcular las tarifas asociadas al transporte de carga. Aunque es el sistema central, GOL no está completamente integrado con otras herramientas clave, lo que limita su funcionalidad.

**Microsoft Azure:** Se utiliza como repositorio para almacenar archivos Excel que contienen reportes, cotizaciones, comprobantes, y documentos de gestión financiera. Sin embargo, la estructura de carpetas no responde a una lógica estandarizada ni jerárquica, lo que dificulta la trazabilidad y control de versiones. Además, el acceso está restringido a un número reducido de personas, lo que incrementa la dependencia operativa.

**Excel:** A pesar de contar con un ERP, Excel continúa siendo la herramienta predominante para registrar, consolidar, procesar y consultar información. Desde el registro de leads hasta la conciliación de facturas, la mayoría de las áreas trabaja con versiones locales de Excel que no están sincronizadas ni centralizadas.

Aunque en teoría todas las áreas tienen acceso a reportes desde GOL, la realidad es que el conocimiento profundo y la capacidad de manipulación de los datos se concentran exclusivamente en dos personas: Edgar y Henry, ambos subordinados del director de tecnología, Sergio Pombo. Estos dos colaboradores no solo controlan el acceso a Azure, sino que también son los únicos con capacidad para editar, depurar y estructurar datos.

### ***6.1.2. Madurez del Gobierno de Datos***

Para evaluar la madurez del Gobierno de Datos en FAM TEAM, se adopta un modelo híbrido basado en los lineamientos de DAMA-DMBOK v2, complementado con criterios operativos del Data Governance Institute (DGI) y el marco CMMI.

Se propone este modelo híbrido con el fin de integrar diferentes enfoques complementarios que permiten una visión más completa y contextualizada. Los lineamientos de DAMA-DMBOK v2 aportan una base conceptual sólida y ampliamente reconocida sobre las funciones y dimensiones del gobierno de datos. Por otra parte, los criterios operativos del DGI permiten traducir esas funciones en acciones prácticas y medibles. Finalmente, el modelo de madurez CMMI introduce una lógica de mejora continua y evaluación progresiva que facilita el diagnóstico y la planificación de la evolución organizacional.

La combinación de estos tres marcos permite adaptar la evaluación al estado real de la organización, evitando rigideces metodológicas y brindando una herramienta que es tanto

granular como accionable, lo cual es clave para organizaciones que se encuentran en etapas iniciales de madurez, como es el caso de FAM TEAM.

**Tabla 2:** *Niveles de Madurez del Gobierno de Datos*

Nivel	Estado	Descripción
0	Inexistente	No hay conciencia de la necesidad del gobierno de datos
1	Inicial / Ad-hoc	Actividades informales, sin roles, ni procesos definidos
2	Reactivo	Algunas prácticas emergen, pero sin consistencia ni formalización
3	Definido	Roles, políticas y procesos básicos formalizados y conocidos
4	Medio	Gobierno transversal, medio, con métricas y auditorías periódicas
5	Integrado	Mejora continua, automatización, parte de la cultura organizacional

Fuente: Elaboración Propia

**Tabla 3:** *Subcategorías para evaluar el Gobierno de Datos*

Subcategoría	Nivel Actual	Justificación
Conciencia Organizacional	1	No hay entendimiento formal sobre la necesidad de gobernar datos
Roles y responsabilidad	0	No existen roles definidos como el Data Owner o Data Steward
Políticas de datos	0	No existen políticas sobre el acceso, calidad o modificación de datos
Procesos de toma de decisiones	1	Las decisiones se toman caso a caso, de forma reactiva, por personas clave
Estructura organizacional	1	La gestión de datos depende de dos personas no formalmente asignadas

Monitoreo y métricas	0	No hay métricas sobre calidad, uso o gobierno de datos
Gestión del cambio	0	No hay estrategias para cambiar la cultura ni capacitar a los usuarios

Fuente: Elaboración Propia

Resultado global de FAM TEAM: Nivel 1 (Ad-hoc).

El gobierno de datos está en una etapa inicial, sin roles formales, procesos definidos ni políticas. Existe alta dependencia de personas específicas, lo que representa un riesgo significativo para la continuidad operativa y la confiabilidad de la información.

**Tabla 4:** Heatmap – Evaluación de madurez del Gobierno de Datos de FAM TEAM

Conciencia Organizacional	1
Roles y Responsabilidades	0
Políticas de Datos	0
Procesos de toma de decisiones	1
Estructura organizacional	1
Monitoreo y métricas	0
Gestión del cambio	0

Fuente: Elaboración Propia

### ***6.1.3. Madurez de Riesgos Asociados al Estado Actual del Gobierno de Datos***

Como parte del diagnóstico de madurez del Gobierno de Datos, se identificaron y evaluaron los principales riesgos inherentes al estado actual de las dimensiones propuestas por el marco DAMA-DMBOK. Esta evaluación tuvo como propósito establecer una línea base que permita priorizar intervenciones en el modelo TO-BE, asegurando una gobernanza efectiva y sostenible.

Para la identificación de riesgos, se realizó un levantamiento de hallazgos a partir de entrevistas, revisión documental y observación directa de procesos en las áreas involucradas con la gestión y uso de datos. Cada riesgo fue clasificado dentro de su respectiva dimensión DAMA (por ejemplo, Arquitectura de Datos, Calidad de los Datos, Seguridad, etc.), y posteriormente evaluado bajo dos criterios clave:

**Probabilidad:** Grado de ocurrencia del riesgo en el contexto actual, en una escala de 1 (muy baja) a 5 (muy alta).

**Impacto:** Consecuencias que tendría la materialización del riesgo sobre la operación, toma de decisiones, cumplimiento normativo u otros aspectos críticos, también en una escala de 1 (muy bajo) a 5 (muy alto).

El Nivel de Riesgo se calculó como el producto de ambas variables:

Nivel de Riesgo = Probabilidad × Impacto

Este puntaje permite cuantificar el riesgo y clasificarlo de forma sencilla en tres rangos:

**Bajo (1-12):** Riesgos que deben ser monitoreados, pero no requieren intervención inmediata.

**Medio (13-20):** Riesgos que requieren acciones correctivas en el corto plazo para evitar su escalamiento.

**Alto (21-25):** Riesgos críticos que deben ser atendidos de forma prioritaria, ya que comprometen significativamente la gestión de datos.

La matriz consolidada permitió visualizar no solo la severidad de los riesgos, sino también su concentración por dimensión, destacándose áreas sensibles como Calidad de Datos, Seguridad de los Datos y Arquitectura, donde predominan riesgos con niveles altos y medios. Esta información sirvió como insumo clave para el diseño del modelo TO-BE de gobernanza, orientado a mitigar dichos riesgos mediante políticas, estándares, roles y herramientas específicas.

**Tabla 5:** *Calificación DAMA*

Dimensión DAMA	Riesgo Identificado	Probabilidad	Impacto	Nivel de Riesgo
Calidad de los Datos	Errores en los datos	5	5	25
Arquitectura de Datos	Falta de estandarización en estructuras y campos críticos	5	5	25
Arquitectura de Datos	Uso de Excel como repositorio sin trazabilidad ni control	4	5	20

Calidad de los Datos	Duplicidad de registros críticos	5	4	20
Calidad de los Datos	Campos sin validación: ciudad, NIT, duplicidad de códigos	5	4	20
Seguridad y Privacidad de Datos	Accesos no formalizados ni trazables	4	5	20
Seguridad y Privacidad de Datos	Sin políticas de seguridad/modificación	4	4	20
Datos Maestros y de Referencia	Duplicidad e inconsistencias en datos maestros de clientes	5	4	20
Documentos y Contenido	No hay centralización ni control de versiones en documentación crítica	5	4	20
Modelado de Datos	Ausencia de modelo de datos formal y estandarizado	4	4	16
Business Intelligence	Reportes limitados en columnas, sin capacidad de análisis	4	4	16
Calidad de los Datos	Sin métricas de calidad	3	5	15
Seguridad y Privacidad de Datos	Acceso desigual a la información	5	3	15
Almacenamiento y Operaciones de Datos	Falta de respaldo estructurado de datos	3	5	15
Integración de Datos	Fragmentación del acceso entre áreas	4	3	12
Gestión de Metadatos	Falta de trazabilidad de reportes y definiciones	3	3	9

Fuente: Elaboración Propia

A partir de la evaluación de riesgos alineada con las 11 dimensiones funcionales del marco DAMA-DMBOK2, se concluye que FAM TEAM enfrenta amenazas críticas principalmente en las dimensiones de Arquitectura de Datos y Calidad de Datos, las cuales han sido priorizadas para el diseño del modelo objetivo (TO-BE). En términos de arquitectura, se identifican problemas relacionados con el uso fragmentado de Excel como repositorio final, la ausencia de un repositorio estructurado centralizado y la falta de estandarización técnica en el diseño de estructuras y campos. En cuanto a calidad, los errores de digitación, la duplicidad de registros críticos, la inexistencia de políticas de validación y la ausencia de métricas formales representan riesgos severos para la confiabilidad y consistencia de la información. Si bien se identificaron vulnerabilidades adicionales en dimensiones como Seguridad, Datos Maestros, Analítica o Metadatos, estas se consideran síntomas de una ausencia generalizada de Gobierno de Datos formal, que impide una gestión transversal, trazable y sostenible. Por tanto, el diagnóstico refuerza la urgencia de intervenir las dimensiones seleccionadas con una estrategia de gobierno estructurada que actúe como habilitador organizacional, técnico y cultural para toda la compañía.

#### ***6.1.4. Modelo Objetivo del Gobierno de Datos***

Con base en el diagnóstico presentado en el capítulo anterior, se propone un modelo objetivo de gobierno de datos para FAM TEAM que permita resolver las brechas actuales y establecer un sistema de gestión de la información sólido, sostenible y alineado con los objetivos estratégicos del negocio. Este modelo se fundamenta en los principios del DAMA-DMBOK v2 y está estructurado en torno a diez principios rectores, una nueva arquitectura organizacional de roles, un conjunto de políticas clave, herramientas de soporte, y una serie de procesos de gobernanza que garantizan la trazabilidad, calidad, seguridad y uso estratégico de los datos.

## **Principios rectores**

El Gobierno de Datos propuesto para FAM TEAM se fundamenta en un conjunto de principios rectores que permiten establecer una cultura sostenible de gestión de la información, alineada con los objetivos estratégicos de la organización y con las mejores prácticas internacionales promovidas por el marco DAMA-DMBOK v2.

### **Reconocimiento del dato como activo estratégico**

Los datos serán considerados recursos de valor crítico para el negocio. Esto implica:

- Asignar responsables formales para su gestión y calidad.
- Incluir su disponibilidad y confiabilidad como parte de la evaluación del desempeño organizacional.
- Incorporar su uso en la planificación estratégica.

### **Centralización de políticas y descentralización operativa**

El gobierno de datos establecerá políticas, estándares y controles de forma centralizada, pero permitirá su implementación y uso a través de equipos distribuidos que actúan con autonomía guiada. Esto fomenta:

- Escalabilidad.
- Flexibilidad operativa.
- Alineación global sin perder eficiencia local.

### **Transparencia, trazabilidad y auditabilidad**

Todos los procesos relacionados con la captura, modificación, acceso y uso de datos deben ser trazables y auditables. Esto implica:

Control de versiones.

Registro de accesos y cambios.

Documentación visible y mantenida.

Capacidad de reconstruir el camino de decisiones basadas en datos.

### **Definición clara de responsabilidades y rendición de cuentas**

Cada dato tendrá un Data Owner y posiblemente un Data Steward, quienes velarán por su integridad, disponibilidad y uso correcto. La gestión de datos dejará de ser una tarea informal y transversal para convertirse en una función estructurada, con indicadores, seguimiento y rendición de cuentas periódica.

El acceso a los datos no será un privilegio, sino una condición necesaria para el desempeño. Por lo tanto:

- Se implementará una política de acceso basada en roles (RBAC).
- Cada área podrá acceder a los datos que necesita para operar y tomar decisiones, sin depender de intermediarios.
- Se mantendrá trazabilidad y control sobre cada consulta y exportación.

### **Estandarización y calidad desde el origen**

Los datos deben cumplir reglas de integridad, validación y formato desde su punto de captura, para evitar reprocesos y errores posteriores. Esto se logrará mediante:

- Validaciones automáticas en formularios.
- Catálogos y listas controladas (ciudades, códigos, unidades, etc.).

- Restricciones de entrada (solo letras, sin caracteres especiales en campos críticos).

### **Uso de metadatos para facilitar el entendimiento**

Se construirá un diccionario y catálogo de datos accesible para toda la organización, que permita:

- Entender qué significan los datos.
- Saber quién los genera, quién los consume y para qué se usan.
- Facilitar el autoservicio y reducir la dependencia del área de tecnología.

### **Gestión activa del ciclo de vida de los datos**

Los datos tendrán un tratamiento estructurado desde su creación hasta su eliminación, incluyendo:

- Políticas de retención.
- Versiones de los documentos, histórico.
- Archivado seguro y eliminación controlada.

### **Fomento de la alfabetización en datos**

El modelo contempla una estrategia de capacitación progresiva en conceptos clave como:

- Modelo relacional.
- Consultas en SQL.
- Principios de calidad.
- Visualización de datos y toma de decisiones basada en evidencia.

### **Medición y mejora continua del gobierno de datos**

Se definirá un sistema de indicadores (KPIs) y mecanismos de retroalimentación para evaluar la efectividad del modelo. Esto incluirá:

Indicadores de calidad de datos.

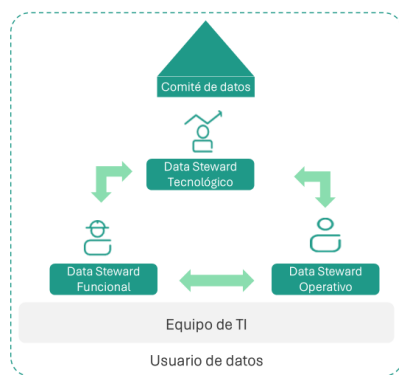
- Tiempos de respuesta en solicitudes.
- Niveles de satisfacción de los usuarios de datos.
- Grado de cumplimiento de políticas y estándares.

### **Roles y estructura organizacional**

En el modelo propuesto para FAM TEAM, se establece una estructura de gobierno de datos distribuida, en la que participan representantes de diferentes áreas de la organización. Cada actor tendrá funciones claras y medibles, con el objetivo de garantizar que los datos sean confiables, accesibles, seguros y gestionados con responsabilidad.

Esta estructura busca reemplazar la dependencia operativa de dos personas por una red de roles definidos que trabajen de forma colaborativa, bajo principios y políticas corporativas comunes.

### **Ilustración 1: *Interacción Gobierno de Datos***



Rol	Responsabilidades Principales	Competencias	Cargo sugerido en la organización
Comité de Gobierno de Datos	Órgano decisor. Define políticas, valida cambios estructurales, gestiona conflictos entre áreas, aprueba métricas.	Toma de decisiones, conocimiento del negocio, conocimiento de herramientas y visión estratégica	Director de TI: Sergio Pombo
Data Owner	Responsable estratégico de los datos de su área. Define reglas de negocio, autoriza accesos, representa a su área en el comité.	Conocimiento del negocio, visión estratégica, toma de decisiones	Líder de planificación financiera: Sergio Aragón
Data Steward	Responsable operativo. Garantiza calidad, estandarización, documentación y cumplimiento de políticas en su dominio.	Dominio técnico de los datos, habilidades en calidad y metadatos	Coordinador comercial: Karla Barreto
Equipo de TI	Gestiona la infraestructura, seguridad técnica, integración entre sistemas, acceso a bases y soporte a los demás roles.	Seguridad, bases de datos, automatización, gestión de plataformas	Ingenieros de desarrollo: Edgar León, Henry León
Usuario de datos	Usan los datos para tomar decisiones. Deben seguir estándares, reportar errores y participar en iniciativas de mejora.	Lectura e interpretación de datos, uso de reportes, comunicación	Analista de cartera: Andrea Cespedes Comercial: Ruby Salgado

Fuente: Elaboración Propia

**Tabla 6:** Definición de roles

Rol	Responsabilidades Principales	Competencias Requeridas
Comité de Gobierno de Datos	Órgano decisor. Define políticas, válida cambios estructurales, gestiona conflictos entre áreas, aprueba métricas.	Define políticas, válida cambios, gestiona conflictos, aprueba métricas.
Data Owner	Responsable estratégico de los datos de su área. Define reglas de negocio, autoriza accesos, representa a su área en el comité.	Conocimiento del negocio, visión estratégica, toma de decisiones
Data Steward	Responsable operativo. Garantiza calidad, estandarización, documentación y cumplimiento de políticas en su dominio.	Dominio técnico de los datos, habilidades en calidad y metadatos
Equipo de TI	Gestiona la infraestructura, seguridad técnica, integración entre sistemas, acceso a bases y soporte a los demás roles.	Seguridad, bases de datos, automatización, gestión de plataformas
Usuario de datos	Usan los datos para tomar decisiones. Deben seguir estándares, reportar errores y participar en iniciativas de mejora.	Lectura e interpretación de datos, uso de reportes, comunicación

Fuente: Elaboración Propia

**Tabla 7:** *Propuesta distribución de roles*

Rol	Área	Persona Sugerida	Rol Actual
Coordinador del Comité	Tecnología	Sergio Pombo	Director Tecnología
Equipo de TI	Tecnología	Edgar, Henry	Ingenieros de Tecnología
Data Owner	Finanzas	Líder de planificación financiera	Pendiente por confirmar personas
Data Steward	Comercial	Coordinador comercial	Pendiente por confirmar personas
Usuario de datos	Cartera	Analista de cartera	Pendiente por confirmar personas
Usuario de datos	Contabilidad	Analista de contabilidad	Pendiente por confirmar personas

Fuente: Elaboración Propia

### Herramientas de soporte

Para operar esta estructura se recomienda implementar:

Una matriz RACI para aclarar responsabilidades por procesos.

Un canal formal de comunicación y solicitudes (correo)

**Tabla 8:** *Matriz RACI*

Actividad / Proceso	Comité de Datos	Data Owner	Data Steward	Equipo TI	Usuario de Datos
Definir política de acceso	A	R	C	I	I
Validar la calidad de un conjunto de datos	I	A	R	C	I
Crear reportes de Azure	I	A	R	C	I

Gestionar permisos de acceso	I	A	C	R	I
Detectar y reportar errores de datos	I	C	R	C	R

Fuente: Elaboración Propia

La matriz RACI demuestra una asignación clara y coherente de responsabilidades dentro del modelo de gobernanza de datos. El Data Owner lidera estratégicamente la mayoría de los procesos, siendo responsable de decisiones clave como políticas de acceso y validación de calidad. El Data Steward cumple un rol operativo fundamental, ejecutando tareas relacionadas con la calidad y generación de reportes. El Comité de Gobierno de Datos actúa como aprobador de decisiones estructurales, asegurando la alineación con las políticas organizacionales. Por su parte, el Equipo de TI respalda técnicamente la gestión de accesos y plataformas, mientras que el Usuario de Datos participa activamente en la detección y reporte de errores, consolidando su papel como actor clave en la mejora continua de los datos.

#### ***6.1.5. Modelo Objetivo (TO-BE) Del Gobierno de Datos***

##### **Políticas**

La implementación del Gobierno de Datos en FAM TEAM requiere la formalización de políticas y procesos que garanticen la correcta gestión, uso, acceso, calidad y seguridad de la información. Estas políticas deben estar aprobadas por el Comité de Gobierno de Datos y comunicadas a todos los usuarios relevantes.

A continuación, se presentan las políticas fundamentales que se propone adoptar en esta fase:

Política de acceso a los datos

- **Objetivo:**

Asegurar que los datos estén disponibles para quienes los necesitan, con los permisos adecuados, en el momento oportuno y bajo principios de mínimo privilegio. Esta tarea en principio va a ser ejecutada por el coordinador del comité de del gobierno de datos (Sergio Pombo).

**Puntos clave:**

- Acceso definido por rol.
- Toda solicitud debe pasar por el Data Steward y ser aprobada por el Data Owner.
- Accesos deben estar registrados y auditados (logs en Azure).
- Revisión de accesos cada 3 meses.

**Beneficio:**

Mayor autonomía con control, reducción de cuellos de botella y mejora en la trazabilidad.

Política de calidad de datos

- **Objetivo**

Establecer estándares para asegurar que los datos sean correctos, completos, consistentes, oportunos y validados desde el origen.

**Puntos clave:**

- Implementación de reglas de validación en formularios del ERP (ej. ciudad sin caracteres especiales).
- Uso de catálogos controlados para campos críticos (ciudades, tipos de documento, unidades de negocio).

- Indicadores de calidad de datos por área (porcentaje de errores, duplicidades, etc.).
- Proceso formal de corrección documentado (ver interacción B.5).

**Beneficio:**

Reducción de errores en reportes, mejora en la confiabilidad y estandarización.

**Política de modificación y mantenimiento de datos**

- **Objetivo:**

Regular el proceso de modificación de datos críticos para asegurar trazabilidad, control y gobernabilidad.

**Puntos clave:**

- Solo los Data Stewards y el equipo de TI están autorizados para modificar campos maestros.
- Todas las modificaciones deben ser registradas con fecha, motivo y responsable en el SQL server management.
- No se permite la edición directa de archivos Excel que representen "la verdad" de los datos.

**Beneficio:**

Mayor control y reducción de errores no trazables.

**Política de documentación y metadatos**

- **Objetivo:**

Asegurar que cada activo de datos esté debidamente documentado en SQL server management para facilitar su uso, entendimiento y reutilización.

**Puntos clave:**

- Implementación de un catálogo de datos con definiciones, responsables, periodicidad y origen.
- Diccionario de datos obligatorio para nuevas bases o reportes.
- Revisión semestral del catálogo por parte del Data Steward.

**Beneficio:**

Mayor transparencia y reducción de dependencia de conocimiento informal.

Política de incorporación de nuevas fuentes de datos

- **Objetivo:**

Controlar y estandarizar la integración de nuevas fuentes o conjuntos de datos al ecosistema corporativo.

**Puntos clave:**

Toda nueva fuente debe ser evaluada y aprobada por el Comité.

Validación previa de calidad, origen, estructura y responsables.

Documentación en el catálogo antes de su uso.

**Beneficio:**

Evita silos nuevos, asegura control desde el origen y previene integraciones informales.

Política de capacitación y alfabetización

- **Objetivo:**

Desarrollar las competencias básicas en manejo y uso de datos en los distintos niveles organizacionales.

**Puntos clave:**

Plan de formación semestral en temas como SQL básico, lectura de reportes, Power BI, principios de calidad de datos.

Talleres liderados por los Data Stewards o equipos externos.

Sesiones cortas, prácticas y orientadas al caso FAM TEAM.

**Beneficio:**

Mayor autonomía de las áreas, reducción de dependencia y mejor uso de los activos de datos.

**Política de Respaldo y Recuperación**

- **Objetivo:** Asegurar la continuidad operativa mediante la disponibilidad y recuperación efectiva de la información crítica.

**Puntos clave:**

- Respaldo automático semanal de datos críticos.
- Procedimientos documentados de recuperación ante fallos.
- Almacenamiento redundante en medios externos (cloud/local).

**Beneficio:** Minimiza el impacto de pérdidas de información y garantiza continuidad del negocio ante incidentes.

Estos estándares serán documentados en una guía unificada y aprobados por el Comité de Gobierno de Datos. Su cumplimiento será obligatorio en todos los procesos que involucren el uso o creación de información.

### **Herramientas**

Matriz RACI por proceso (quién responde, aprueba, consulta o informa).

Sistema de solicitudes de datos (plataforma interna o canal formal).

Logs de acceso en Azure.

### **Procesos de soporte**

Registro y actualización de accesos.

Flujo de aprobación de nuevos reportes o fuentes.

Auditoría interna semestral de calidad y seguridad.

Evaluación continua de los KPIs del modelo de gobernanza.

En el siguiente capítulo se desarrollará la propuesta de arquitectura de datos que acompañará y habilitará la implementación de este modelo de gobernanza en FAM TEAM.

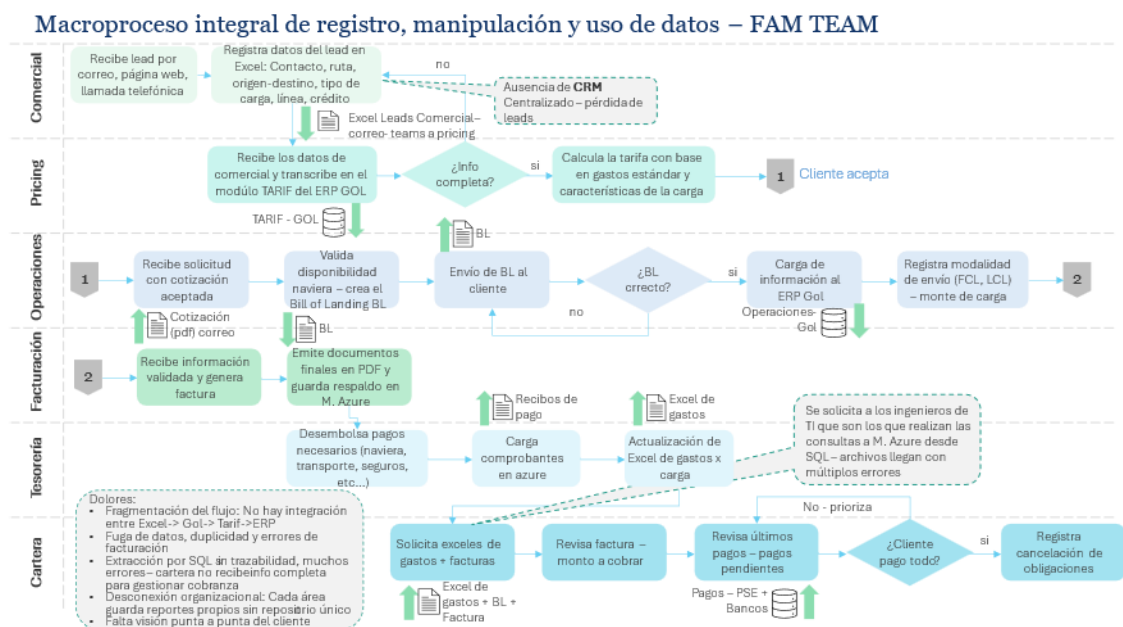
## **6.2. Arquitectura de Datos Propuesta para FAM TEAM**

### **Introducción**

La arquitectura de datos es el componente estructural que permite implementar, sostener y escalar un modelo de gobierno como el propuesto en el capítulo anterior. En el caso de FAM TEAM, el diseño de una arquitectura sólida busca resolver los problemas de fragmentación, baja

trazabilidad y duplicación de información identificados en el diagnóstico AS-IS, habilitando un entorno donde los datos sean accesibles, seguros, estandarizados y auditables.

**Ilustración 2: Macroproceso de Registro, Manipulación y Uso de Datos**



Fuente: Elaboración Propia

A través del análisis de las operaciones en la empresa ULS (United Logistics Services), se identificó un macroproceso transversal que refleja el flujo de datos desde la entrada de un cliente potencial hasta la aplicación de modelos analíticos para gestión de cobranza. Este proceso está compuesto por múltiples pasos y actores, distribuidos en seis áreas funcionales:

#### Comercial:

- Recibe leads por correo, página web o teléfono.
- Registra manualmente en Excel (antes) o en el CRM GOL (actualmente).
- Completa los datos básicos: cliente, origen, destino, volumen, crédito.

- Transfiere la información manualmente por correo electrónico o Teams al área de Pricing.

**Pricing:**

- Transcribe los datos recibidos en el módulo TARIF del ERP GOL.
- Verifica si la información está completa; si no, solicita correcciones.
- Calcula la tarifa con base en origen, destino, volumen, línea naviera y servicios requeridos.
- Exporta cotización en PDF y la remite nuevamente al equipo comercial.

**Cliente:**

- Recibe y acepta la cotización por correo o firma electrónica.

**Operaciones:**

- Recibe la solicitud confirmada.
- Verifica disponibilidad con la naviera y crea el documento emitido por el transportista o la naviera que certifica el recibo de la mercancía para su transporte: Bill of Lading (BL).
- Envía el BL al cliente para verificación.
- Si hay errores, se devuelve y se repite el proceso; si está correcto, se carga en GOL.
- Se registra la modalidad de envío (FCL o LCL) y el monto estimado de carga.

**Facturación:**

- Recibe la información operativa validada.

- Genera la factura en PDF con los valores definitivos.
- Respalda los documentos en Azure.

**Tesorería:**

- Realiza los desembolsos correspondientes a naviera, transporte, seguros.
- Carga los comprobantes en carpetas compartidas en Azure.
- Mantiene un Excel actualizado por cada carga.

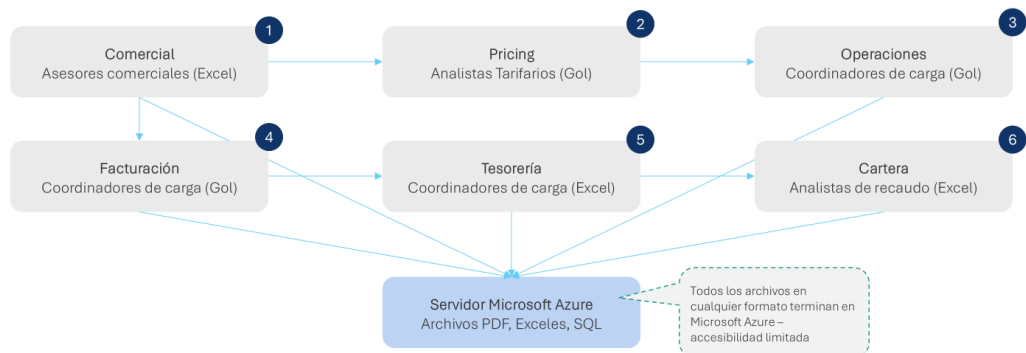
**Cartera:**

- Solicita documentos a otras áreas o busca en Azure mediante consultas SQL.
- Consolida información de Excel, BL y facturas.
- Verifica montos a cobrar y pagos realizados.
- Si detecta inconsistencias, las comunica a Facturación u Operaciones.
- Aplica un modelo predictivo de riesgo de impago.
- Prioriza clientes con base en la probabilidad de incumplimiento.

**Ilustración 3:** *Arquitectura de datos FAMTEAM*

Actualmente, FAM TEAM opera sobre una infraestructura tecnológica híbrida que combina desarrollos propios, sistemas fragmentados y Excel

Arquitectura previa



Problemas Identificados	Descripción
	<ul style="list-style-type: none"> <li>• <b>Duplicidad de registros:</b> Comercial y Operaciones trabajan con bases de datos diferentes, generando redundancias y discrepancias.</li> <li>• <b>Falta de integración:</b> Los datos no fluyen automáticamente entre las áreas, lo que dificulta la sincronización y la eficiencia.</li> <li>• <b>Procesos manuales:</b> Cartera utiliza Excel para su gestión, lo que limita la visibilidad y la trazabilidad en el ERP.</li> <li>• <b>Ausencia de políticas de llenado:</b> La falta de estándares ocasiona datos inconsistentes y en diferentes formatos.</li> </ul>

Fuente: Elaboración Propia

Este proceso ilustra la ausencia de automatización, la falta de conexión entre plataformas, y la dependencia de procedimientos manuales altamente expuestos al error humano.

## Principales Brechas Identificadas

### Fragmentación del flujo

- No existe integración entre las herramientas: Excel → GOL → TARIF → ERP.
- El traspaso de datos entre áreas se hace manualmente, generando duplicidades.
- Los errores en campos clave como tarifas, cliente o código de carga se replican en cascada.

### Dependencia de consultas manuales

- Solo dos personas (del área de tecnología) tienen capacidad para realizar extracciones.
- Las áreas funcionales dependen de ellos para cualquier requerimiento.
- Las consultas por SQL no cuentan con trazabilidad ni documentación.
- Cartera no siempre recibe la información necesaria para cobrar.

### **Desconexión organizacional**

- Cada área guarda sus propios reportes sin una lógica común.
- No existe un repositorio maestro ni un catálogo de datos.
- No se dispone de una "fuente única de la verdad".
- No hay visibilidad de punta a punta del proceso ni del cliente.

**Fragmentación del acceso:** El acceso a los datos es altamente segmentado y desigual. Cada área dispone únicamente de ciertos reportes, con base en su rol y nivel jerárquico, sin posibilidad de acceso a información transversal. De manera reciente, se ha otorgado acceso a Azure a dos personas del área financiera, quienes han logrado realizar cruces más efectivos gracias a su conocimiento técnico, pero esto no es la norma.

**Limitaciones de alfabetización y autonomía:** Existe una fuerte brecha de alfabetización en datos. La mayoría de los usuarios desconoce conceptos fundamentales como SQL, almacenamiento en la nube, modelos de datos o repositorios. Esto genera una fuerte dependencia de Edgar y Henry, quienes, además de sus múltiples responsabilidades, deben atender solicitudes manuales de extracción o transformación de datos, lo cual suele generar cuellos de botella e impactos en la productividad.

**Problemas de calidad y estandarización:** No existen políticas de calidad de datos ni mecanismos de validación en los formularios del ERP. Por ejemplo, campos como "ciudad" no cuentan con validaciones, permitiendo el ingreso de números, símbolos o errores ortográficos. Esto genera registros inconsistentes, duplicidades (como clientes escritos de forma diferente) y errores en los reportes finales. Además, no hay controles para garantizar unicidad en campos clave como el NIT, lo que compromete la confiabilidad de los análisis.

**Uso de Excel como repositorio final:** Los reportes generados desde GOL o Azure son exportados a Excel, donde cada área genera sus propios reportes sin trazabilidad. No existe un repositorio único ni seguimiento consolidado, lo que impide mantener una “fuente única de la verdad” y genera múltiples versiones de un mismo indicador según el área que lo analice.

### **Principios de Arquitectura de Datos**

**Centralización del repositorio:** Consolidar todos los datos operativos, financieros y comerciales en una única fuente estructurada, utilizando Azure como bodega unificada.

**Integración de sistemas:** Conectar GOL, TARIF y las hojas de cálculo a través de pipelines automáticos para eliminar tareas manuales.

**Gobernanza técnica:** Definir estándares técnicos de estructuras, catálogos, esquemas de datos y nomenclaturas para cada dominio.

**Modularidad y escalabilidad:** La arquitectura debe permitir incorporar nuevas fuentes o áreas de negocio de forma progresiva.

**Capacidad de autoservicio:** A través de dashboards y reportes estructurados que permitan a las áreas operar con autonomía.

### **Componentes Arquitectónicos Propuestos:**

#### **Repositorio Centralizado en Azure**

- Estructurado en carpetas por dominio de datos (Clientes, Operaciones, Cargas, Finanzas).
- Cada carpeta incluye subcarpetas para: fuentes originales, archivos depurados, reportes validados y respaldos históricos.
- Accesos gestionados por grupo y rol con autenticación.

#### **Pipeline de Integración (ETL)**

- Extracción automática desde GOL, TARIF y otras fuentes externas.
- Transformación de campos, validación de formatos, consolidación de registros.
- Carga programada en repositorio limpio y accesible por los Data Stewards.

#### **Catálogo de datos Corporativo**

- Plataforma digital que documenta cada conjunto de datos disponible: descripción, responsable, periodicidad, calidad esperada.
- Incluye diccionario de datos con nombres de campos, tipos y reglas de negocio.

#### **Dashboards y Herramientas de Consulta**

- Construcción de paneles de Power BI para cada área operativa y de control.
- Acceso controlado por perfiles definidos por el Comité de Gobierno.
- KPI disponibles en tiempo real con datos provenientes del repositorio limpio.

## Flujos de Datos (Data Flow)

A continuación, se resume el flujo ideal propuesto:

### Ilustración 4: Arquitectura de datos propuesta

#### ARQUITECTURA PROPUESTA



Fuente: Elaboración Propia

Captura de datos en GOL (Comercial, Tarifas, Operaciones)

Objetivo: Registrar información crítica de forma estructurada, validada y oportuna en el sistema ERP GOL.

Actividades clave:

- Registro de leads y cotizaciones (área comercial).
- Cálculo de tarifas en el módulo TARIF (pricing).
- Registro de operaciones logísticas, modalidad de carga, y BL (operaciones).

- Validación inicial de campos obligatorios y consistencia de datos.

Responsables: Usuarios de datos, Data Stewards operativos, líderes de cada área.

**Tabla 9:** *Propuestas de actividades para el Gobierno de Datos 1*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Diseñar formularios estandarizados	A	R	C	C	I
Definir campos obligatorios y validaciones	A	R	R	C	I
Parametrizar formularios en GOL	I	C	C	R	I
Registrar leads y solicitudes en GOL	I	I	I	I	R
Validar consistencia de los datos registrados	I	C	R	I	R
Reportar errores o inconsistencias	I	I	R	I	R
Capacitar usuarios finales en el uso de GOL	I	C	R	C	A
Actualizar manuales y plantillas de ingreso	I	R	R	C	I
Generar alertas de datos incompletos o inválidos	I	C	R	R	I

Fuente: Elaboración Propia

**Descripción de actividades clave:**

- Diseño de formularios: Establecer qué información debe capturarse desde comercial, operaciones y pricing en formatos estructurados.

- Validación técnica: Parametrización en GOL de campos obligatorios, restricciones de formato y lógica de negocio.
- Ingreso de datos: Registro directo por parte de usuarios finales desde las áreas operativas.
- Revisión y retroalimentación: Verificación continua de la calidad del dato ingresado y activación de mecanismos de corrección.
- Documentación: Actualización de instructivos y manuales, junto con entrenamiento recurrente para nuevas personas usuarias.

### **Validación y transformación (ETL automatizado)**

Objetivo: Estandarizar, depurar y transformar los datos capturados antes de su uso analítico.

Actividades clave:

- Extracción automática desde GOL y fuentes auxiliares.
- Aplicación de reglas de validación (formatos, duplicados, valores requeridos).
- Unificación de nomenclaturas y estructuras.
- Transformación de campos (fechas, nombres, códigos, etc.).

Responsables: Equipo de Tecnología, Data Stewards técnicos.

Para garantizar una arquitectura robusta, se define un proceso ETL automatizado con las siguientes fases y componentes:

#### **Fase 1: Extracción (Extract) – Gol**

Objetivo: Obtener datos crudos desde las fuentes originales sin alteración.

Orígenes:

- ERP GOL (módulos de comercial, tarifas, operaciones).
- Archivos estructurados del servidor Azure (facturación, tesorería, cartera).
- Formularios digitales y registros complementarios en Excel.

Mecanismos:

- Conexiones programadas mediante APIs o conectores ODBC.
- Extracción incremental para evitar sobrecarga.
- Almacenamiento temporal en zona de staging.

**Tabla 10:** *Propuestas de actividades para el Gobierno de Datos 2*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Definir fuentes autorizadas	A	R	C	C	I
Configurar conectores/API	I	C	I	R	I
Programar extracciones automáticas	I	C	C	R	I
Monitorear estado de extracción	I	I	C	R	I
Registrar incidentes o fallos	I	I	R	R	I

Fuente: Elaboración Propia

## **Fase 2: Transformación (Transform) – Azure Data Factory + SQL**

Objetivo: Limpiar, enriquecer y estandarizar los datos para su análisis y gobierno.

Actividades clave:

- Limpieza de datos: eliminación de duplicados, corrección de errores tipográficos, estandarización de campos.
- Validación de reglas de negocio: verificación de que los datos cumplan con condiciones específicas (ej. tarifas mayores a cero, códigos válidos de ciudad, etc.).
- Normalización: unificación de formatos (fechas YYYY-MM-DD, mayúsculas con tildes, campos numéricos sin símbolos).
- Enriquecimiento: incorporación de datos adicionales como códigos estándar, clasificación de clientes, etiquetas geográficas.
- Cálculo de indicadores preliminares: generación de flags de alerta, antigüedad de saldos, riesgo proyectado, etc.
- Auditoría y trazabilidad: registro de cada transformación aplicada por lote, fecha y usuario.

**Tabla 11:** *Propuestas de actividades para el Gobierno de Datos 3*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Definir reglas de validación y negocio	A	R	R	C	I
Estandarizar campos y formatos	I	C	R	R	I
Eliminar duplicados y limpiar datos	I	C	R	R	I
Aplicar enriquecimiento y derivaciones	I	R	R	C	I

Validar resultados de transformación	I	A	R	C	I
Documentar cada transformación aplicada	I	C	R	R	I

Fuente: Elaboración Propia

### Fase 3: Carga (Load) – Microsoft Azure

Objetivo: Ingresar los datos ya estructurados en el repositorio definitivo de Azure.

Estrategia de carga:

- Carga incremental con control de versiones.
- Carga por dominio (clientes, cargas, facturación, cartera).
- Verificación de consistencia posterior a la carga.
- Registro en logs con timestamp y hash de validación.

**Tabla 12:** *Propuestas de actividades para el Gobierno de Datos 4*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Configurar zona de carga en Azure	I	C	C	R	I
Definir carpetas y nomenclatura por dominio	I	R	R	C	I
Realizar carga incremental controlada	I	I	C	R	I
Verificar consistencia posterior a carga	I	A	R	R	I

Registrar logs de carga con hash	I	I	R	R	I
Validar accesibilidad y seguridad	C	C	R	R	I

Fuente: Elaboración Propia

### **Herramientas sugeridas para el ETL**

- Azure Data Factory o Power Automate para automatización de flujos.
- SQL Server Integration Services (SSIS) para transformaciones complejas.
- Python con Pandas para manipulación avanzada de datos.
- PowerShell o scripts Bash para tareas de administración programada.

### **Participación organizacional**

- Equipo de Tecnología: responsable de la configuración, programación y monitoreo del proceso ETL.
- Data Stewards: validan reglas de transformación y revisan resultados de calidad.
- Data Owners: definen los requerimientos de transformación y validación por dominio de datos.

Este proceso ETL permitirá a FAM TEAM mantener sus datos consistentes, auditables, actualizados y preparados para su uso operativo y estratégico.

El ecosistema actual de FAM TEAM está compuesto por múltiples fuentes heterogéneas, con datos operativos dispersos en Excel, GOL (CRM/ERP), módulos como TARIF, y bases administrativas cargadas manualmente. Estas fuentes presentan:

- Inconsistencias frecuentes en nomenclaturas.

- Datos duplicados o mal categorizados.
- Registros incompletos que requieren validación con reglas de negocio.

Ante este contexto, un enfoque ETL resulta más adecuado porque permite aplicar una capa intermedia de validación y limpieza antes de almacenar los datos en el repositorio estructurado de Azure. A diferencia del modelo ELT, que deposita los datos en bruto y luego los transforma, el ETL garantiza que solo datos estandarizados, validados y de calidad ingresen al entorno productivo, lo que es crítico para evitar contaminación de dashboards, reportes y modelos analíticos.

Las áreas que dependen de la información transformada —cartera, tesorería, operaciones y comercial— requieren información precisa y confiable en tiempo real para evitar errores de facturación, cobros duplicados o inconsistencias en la trazabilidad de servicios. Bajo un esquema ELT, las transformaciones se hacen después de cargar los datos, lo que implica:

- Aumento de la complejidad de monitoreo.
- Mayor riesgo de trabajar con datos sin limpiar.
- Potencial propagación de errores antes de ser corregidos.

El modelo ETL propuesto interrumpe ese riesgo en el paso de transformación previa, permitiendo:

- Aplicar reglas de negocio antes del almacenamiento.
- Documentar y auditar errores en zona de staging.
- Generar logs por excepción para rápida corrección y trazabilidad.

El equipo de TI y los Data Stewards de FAM TEAM están en proceso de profesionalización en herramientas como Azure Data Factory, Power BI, y catálogos documentales. El modelo ETL permite diseñar pipelines con estructuras claras y controladas, que:

- Centralizan las transformaciones en un entorno específico.
- Separan la lógica de negocio del almacenamiento final.
- Son más fáciles de auditar, mantener y escalar en fases tempranas.

### **Repositorio estructurado en Azure**

Objetivo: Centralizar la información en un entorno seguro, trazable y accesible por roles.

Actividades clave:

- Organización de carpetas por dominio (Clientes, Cargas, Finanzas, etc.).
- Subcarpetas para fuentes originales, archivos limpios, reportes y backups.
- Asignación de permisos basados en roles (RBAC).
- Control de versiones y respaldos periódicos.

Responsables: Equipo de Tecnología, Data Owner del dominio.

**Tabla 13:** *Propuestas de actividades para el Gobierno de Datos 5*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final

Diseñar estructura de carpetas por dominio	A	R	C	R	I
Establecer nomenclatura estándar	I	R	R	C	I
Asignar permisos de acceso por rol	C	C	R	R	I
Configurar políticas de backup y versiones	C	I	C	R	I
Monitorear uso y espacio disponible	I	I	C	R	I

Fuente: Elaboración Propia

### **Documentación en catálogo de datos**

Objetivo: Asegurar que todos los datos estén correctamente descritos y gobernados.

Actividades clave:

- Registro en el catálogo corporativo: nombre del conjunto, campos, responsable, periodicidad.
- Actualización del diccionario de datos.
- Enlace entre el repositorio y los documentos de referencia.

Responsables: Data Stewards, Comité de Gobierno de Datos.

**Tabla 14:** *Propuestas de actividades para el Gobierno de Datos 6*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Crear ficha técnica por conjunto de datos	I	R	R	C	I

Mantener actualizado el diccionario de datos	I	C	R	C	I
Validar integridad y consistencia documental	C	A	R	C	I
Publicar acceso al catálogo por área	I	C	R	C	I
Sensibilizar sobre uso del catálogo	C	C	R	I	A

Fuente: Elaboración Propia

### **Dashboards por área (Power BI)**

Objetivo: Facilitar el acceso a reportes visuales y confiables para la toma de decisiones.

Actividades clave:

- Diseño de visualizaciones por área (comercial, operaciones, cartera, tesorería).
- Conexión directa con Azure (no con archivos locales).
- Control de accesos por perfil.

Indicadores clave: tiempo de respuesta, calidad, desempeño de clientes, flujo de caja, etc.

Responsables: Equipo de analítica, Data Stewards, líderes de área.

**Tabla 15:** *Propuestas de actividades para el Gobierno de Datos 7*

Actividad	Comité de Gobierno	Data Owner	Data Steward	TI	Usuario Final
Diseñar tableros por área de negocio	C	R	R	C	C

Configurar permisos de visualización	C	C	R	R	I
Capacitar a usuarios en Power BI	I	C	R	C	A
Recoger retroalimentación de usuarios	I	C	R	I	R
Incorporar mejoras en reportes	I	R	R	C	C

Fuente: Elaboración Propia

### **Consulta y análisis por áreas (Comercial, Operaciones, Cartera)**

Objetivo: Promover la autonomía operativa en el uso y análisis de datos.

#### **Actividades clave:**

- Acceso a reportes dinámicos desde el dashboard.
- Generación de consultas simples sin necesidad de SQL (self-service BI).
- Identificación de errores, oportunidades y tendencias.
- Retroalimentación a Data Stewards sobre datos críticos.

Responsables: Usuarios de datos, líderes de procesos, Data Stewards.

#### Modelos analíticos y decisiones basadas en datos

Objetivo: Utilizar modelos predictivos y analíticos como insumo estructural de las decisiones.

Actividades clave:

Aplicación del modelo de riesgo de cartera para priorización de cobranza. (Se describirá más adelante)

- Análisis de rentabilidad por cliente o ruta.
- Identificación de cuellos de botella operativos.
- Simulación de escenarios de tarifas o tiempos de entrega.

Responsables: Equipo de analítica, Comité de Gobierno, Data Owners.

### **6.3. Estándares, Políticas y Controles Críticos para el Despliegue de la Arquitectura TO-BE**

Este subcapítulo describe los estándares técnicos, políticas organizacionales y controles operativos que deben implementarse para garantizar el despliegue exitoso de cada uno de los siete procesos clave definidos en la arquitectura de datos TO-BE. Cada conjunto se justifica con base en los riesgos mitigados y su importancia para la sostenibilidad del modelo de gobernanza.

#### **Captura de Datos en GOL**

##### **Estándares**

- Implementación de formularios inteligentes con lógica embebida que garantice la obligatoriedad de campos clave, validaciones automáticas de formato (fechas, valores numéricos, listas desplegadas) y condicionalidad entre campos.
- Definición de una nomenclatura estandarizada y centralizada para la codificación de entidades críticas (clientes, rutas, servicios), permitiendo interoperabilidad y trazabilidad. Ejemplos: CLIENTE\_<NIT>, RUTA\_<ORIGEN-DESTINO>, SERVICIO\_<FCL/LCL>.
- Inclusión automática de metadatos operativos y técnicos como usuario, fecha/hora de captura, versión del formulario, y canal de ingreso, habilitando auditoría y seguimiento continuo del flujo de datos.

## **Políticas**

**Política de integridad en la fuente:** Todo dato registrado en GOL debe cumplir obligatoriamente con los criterios mínimos de validación, completitud y consistencia definidos por los Data Stewards, sin excepciones operativas.

**Política de unicidad e identificación:** Toda entidad capturada (cliente, contacto, embarque, lead) debe estar asociada a un identificador único persistente en el ecosistema de datos de la organización, evitando ambigüedades y duplicidad de registros.

**Política de gestión de versiones:** Cualquier modificación estructural a los formularios de captura debe ser registrada formalmente con su versión correspondiente, aprobada por el Comité de Datos y comunicada a los usuarios impactados.

## **Controles**

- Configuración de reglas de validación en el sistema GOL con mecanismos de retroalimentación inmediata y mensajes de error orientados al usuario, reduciendo errores operativos y tiempo de capacitación.
- Generación de tableros de control mensuales que identifiquen patrones de registros fallidos, campos omitidos, reincidencias por unidad de negocio y causas comunes de rechazo.
- Ejecución de auditorías aleatorias bimensuales lideradas por los Data Stewards para verificar la aplicación de los estándares de captura, identificar posibles atajos o inconsistencias, y retroalimentar a los responsables de procesos.

## **Justificación**

La captura de datos representa el primer eslabón en la cadena de valor de la información y constituye un punto crítico en la prevención de errores sistemáticos. Una arquitectura de datos puede colapsar si la entrada es errónea, inconsistente o incompleta. En FAM TEAM, donde los datos son reutilizados transversalmente por las áreas comercial, operativa, financiera y analítica, la estandarización desde el punto de origen asegura fluidez, confiabilidad y agilidad en todos los procesos subsiguientes, desde la cotización hasta la analítica predictiva. Además, incorporar nomenclaturas y validaciones robustas habilita la automatización, reduce reprocesos, facilita auditorías y disminuye riesgos contractuales derivados de errores en la información.

Una captura de datos bien gobernada no solo mejora la calidad de los datos, sino que disminuye costos operativos, fortalece la experiencia del cliente y aumenta la eficiencia de los equipos. Este conjunto de estándares, políticas y controles representa la base sobre la cual se construye un modelo de datos confiable, escalable y alineado con los principios de calidad, trazabilidad y responsabilidad definidos por el marco DAMA.

## **Validación y Transformación (ETL)**

### **Estándares**

- Desarrollo modular y documentado de todos los procesos ETL, empleando prácticas de codificación limpias, reutilizables y mantenibles. Cada módulo (extracción, transformación, carga) debe contar con su propia lógica desacoplada y trazable.
- Uso obligatorio de control de versiones para los scripts y flujos ETL, con historial de cambios accesible y gestionado mediante repositorios (ej. Git), asegurando

gobernanza del código y trazabilidad de modificaciones.

- Implementación de validaciones automáticas en los pipelines: tipos de dato, detección de duplicados, verificación de rangos permitidos, control de nulos, consistencia entre campos relacionados, y estandarización de formatos (fechas, monedas, nomenclaturas).
- Segmentación estructural del proceso ETL en tres capas: zona de staging (almacenamiento temporal de datos brutos), zona de procesamiento/transformación, y zona de carga en destino, garantizando control y auditoría por etapa.
- Inclusión de metadatos de ejecución y calidad en cada flujo: tiempo de procesamiento, porcentaje de registros válidos, errores detectados y responsable de ejecución.

## **Políticas**

**Política de trazabilidad y linaje de datos:** Cada dato transformado debe poder ser rastreado hasta su fuente original a través de registros técnicos que documenten su flujo de transformación, permitiendo auditoría posterior y aplicabilidad de indicadores derivados.

**Política de documentación obligatoria:** Todo proceso ETL debe estar soportado por documentación accesible que describa su lógica de transformación, las reglas de negocio aplicadas por campo y sus responsables funcionales, facilitando la interoperabilidad y el

mantenimiento del flujo.

**Política de reversibilidad operativa:** Todo dato transformado debe conservar su versión original accesible (zona histórica o backup) para permitir auditoría forense, comparación retroactiva o reconstrucción de escenarios en caso de errores críticos.

### **Controles**

- Ejecuciones ETL deben generar logs detallados con: total procesado, errores, tiempo y Generación automática de logs de ejecución en cada corrida ETL, que incluyan cantidad de registros procesados, errores detectados, duración de cada etapa y responsable técnico. Estos logs deben almacenarse como evidencia para auditorías.
- Revisión mensual de las transformaciones críticas (aquellas que impactan KPIs estratégicos o procesos contables) a cargo de los Data Stewards y Owners responsables, incluyendo análisis de consistencia e impacto.
- Implementación de alertas automáticas configurables ante fallos estructurales o validaciones fallidas que superen umbrales definidos (por ejemplo: más de 5% de errores de tipo o más de 100 registros duplicados detectados).
- Registro de incidentes de transformación con mecanismos de retroalimentación y resolución, como parte del sistema de mejora continua de la arquitectura de datos.

### **Justificación:**

El proceso de Validación y Transformación (ETL) constituye el corazón técnico de la cadena de procesamiento de datos. Aquí se depura, estructura y estandariza la información bruta, convirtiéndola en un activo de alto valor analítico y operativo. En el contexto de FAM TEAM, donde los datos provienen de múltiples fuentes con formatos dispares y niveles de calidad variables, una arquitectura ETL robusta permite asegurar que los datos que alimentan dashboards, reportes y modelos analíticos sean confiables, consistentes y trazables.

Evitar que los procesos ETL se conviertan en “cajas negras” es esencial para garantizar la gobernabilidad del ecosistema de datos. El modularidad del código, la existencia de documentación clara, la trazabilidad de cada transformación y la generación de evidencia técnica (logs) no solo permiten el monitoreo en tiempo real, sino que facilitan el mantenimiento, la resolución de incidentes y la explicación de resultados ante instancias regulatorias o comités directivos.

Además, disponer de entornos separados para desarrollo, pruebas y producción evita errores en cadena y permite aplicar principios de calidad de software al gobierno de datos. En conjunto, estos lineamientos permiten que la transformación de datos deje de ser una tarea técnica aislada y se consolide como un proceso auditable, escalable y alineado con los principios de integridad, responsabilidad y transparencia del marco DAMA.

Repositorio estructurado en Azure

### **Estándares**

Organización jerárquica de carpetas en Azure Data Lake, estructurada por dominio funcional (ej. Comercial, Cartera, Operaciones) y subcarpetas por período (ej. Cartera/2025\_01), permitiendo segmentación, escalabilidad y control granular.

Convención de nomenclatura estandarizada para archivos, con estructura lógica y machine-readable: <DOMINIO>\_<FECHA>\_<TIPO>\_<VERSION>.csv, favoreciendo búsquedas automatizadas y control de versiones.

Implementación de backups automáticos con periodicidad diaria, almacenados en zonas de alta disponibilidad con redundancia geográfica (geo-replicación), cumpliendo estándares de continuidad del negocio.

Etiquetado automatizado de archivos mediante metadatos definidos por política (sensibilidad, tipo de contenido, fuente), habilitando clasificación, encriptación y control automatizado del ciclo de vida del dato.

### **Políticas**

Política de acceso seguro (RBAC): El acceso a carpetas y archivos se define bajo un modelo de control por roles (Role-Based Access Control), alineado con perfiles funcionales y revisado trimestralmente por el área de TI y el Comité de Datos.

Política de control de versiones: Todo archivo que sea insumo o resultado de procesos críticos (e.g. reportes financieros, bases para modelado) debe mantener un historial de versiones sin sobreescritura directa, habilitando auditoría y reversión.

Política de integridad de almacenamiento: Toda carpeta y archivo debe contar con un propietario funcional asignado (Data Owner), responsable de su actualización, documentación y ciclo de vida.

### **Controles**

Registro automatizado de accesos a carpetas y archivos críticos mediante logs de trazabilidad (incluye fecha, hora, usuario, tipo de acción), auditables desde Azure Monitor o herramientas integradas de seguridad.

Alarmas configuradas para detectar cargas fuera del estándar de nomenclatura o cambios no autorizados, activando alertas a los Data Stewards.

Validación mensual del cumplimiento del backup diario y pruebas programadas de restauración parcial, como mecanismo de contingencia y verificación operativa.

### **Justificación**

como el centro de gravedad donde convergen los esfuerzos de captura, transformación, consulta y modelado. Una arquitectura de almacenamiento sin estructura, sin control de acceso ni trazabilidad, genera caos, pone en riesgo la integridad de los datos y vulnera la confianza de los usuarios finales.

En organizaciones como FAM TEAM, donde el dinamismo operacional requiere disponibilidad de datos confiables y seguros, contar con una arquitectura en la nube gobernada permite escalar operaciones, cumplir regulaciones (como Habeas Data) y garantizar continuidad

operativa ante incidentes. Además, los estándares aquí definidos permiten habilitar automatizaciones posteriores como flujos de aprobación, retención de datos, y clasificación de archivos por sensibilidad, todo bajo una infraestructura alineada con los principios de arquitectura y seguridad del marco DAMA.

## **Documentación en Catálogo de Datos**

### **Estándares**

Todo conjunto de datos en uso activo o en etapa de análisis debe estar documentado mediante una ficha técnica estructurada que incluya: nombre del dataset, descripción funcional, unidad de medida, periodicidad de actualización, sistema fuente, dominio funcional, responsable de negocio (Data Owner), responsable técnico (Data Steward), y fecha de última actualización.

Cada variable o campo dentro del dataset debe contar con atributos detallados: tipo de dato (string, int, decimal, fecha), unidad de medida (si aplica), valores permitidos o rangos válidos, definición operacional y al menos un ejemplo práctico de uso.

El catálogo debe estar centralizado en una plataforma digital de consulta institucional (por ejemplo, Power BI Data Catalog, Azure Purview, o SharePoint estructurado), accesible para todos los perfiles autorizados, e integrado con los flujos de metadatos de los sistemas operativos y analíticos.

### **Políticas**

**Política de obligatoriedad documental:** Ningún dataset podrá ser promovido a producción o consumido por áreas funcionales si no cuenta con su respectiva ficha técnica validada y publicada en el catálogo institucional. Esto aplica tanto a fuentes crudas como a

datasets derivados o transformados.

**Política de actualización y vigencia:** Todo dataset activo debe ser revisado documentalmente de forma semestral o cada vez que experimente una modificación estructural, lógica de negocio o cambio de responsable. Las revisiones deben quedar registradas en el historial de versiones del catálogo.

**Política de accesibilidad controlada:** El catálogo debe permitir la consulta transversal bajo modalidad de solo lectura, garantizando la democratización del conocimiento sin comprometer la integridad documental. Se promoverá activamente su uso como canal oficial de interpretación de datos.

### **Controles**

Ejecución periódica de auditorías automáticas y manuales para detectar datasets activos que no cuenten con ficha técnica vigente o con documentación incompleta. Estas auditorías deben generar alertas a los Data Stewards responsables.

Revisión cruzada entre los metadatos operacionales de los archivos almacenados en Azure y la documentación registrada en el catálogo. Esto incluye campos, estructuras, y periodicidades.

Validación semestral de actualizaciones y consistencia documental por parte de los Data Owners y Stewards, con soporte en tableros de seguimiento y planes de remediación en caso de hallazgos.

### **Justificación**

El catálogo de datos es la columna vertebral del gobierno del conocimiento en una organización orientada a datos. Representa la formalización de la semántica, estructura y propósito de los datos que circulan en los sistemas, permitiendo que cualquier usuario —desde un analista de BI hasta un nuevo ingreso— pueda interpretar correctamente la información, sin depender de conocimientos tácitos o memoria organizacional informal.

En el caso de FAM TEAM, donde múltiples áreas operan sobre una misma base informacional con distintos objetivos, un catálogo bien documentado y accesible evita ambigüedades, reduce reprocesos, mejora el tiempo de respuesta analítica y facilita el control de calidad en el uso de datos. Además, se convierte en una herramienta clave para el onboarding de nuevos colaboradores, la auditoría interna, y el cumplimiento regulatorio en contextos donde la trazabilidad del dato y su interpretación coherente son requisitos indispensables.

Un entorno sin catálogo es un entorno expuesto al error, la dependencia de individuos clave y la pérdida de valor analítico. Un entorno con catálogo es un ecosistema donde la información se transforma en conocimiento institucional, reutilizable y sostenible.

### **Dashboards por Área (Power BI)**

## **Estándares**

Todos los dashboards deben conectarse de forma directa y automática a datasets certificados, almacenados y actualizados en Azure Data Lake o Synapse Analytics. Las conexiones deben realizarse mediante gateways seguros y controlados, garantizando la frescura de la información sin intervenciones manuales.

Las visualizaciones deben seguir una guía de diseño institucional que incluya: paleta de colores corporativa, jerarquía visual coherente (títulos, subtítulos, etiquetas), estandarización de tipos de gráficos según el tipo de variable (líneas para series temporales, barras para comparativos, tarjetas para KPIs, etc.), y uso de espacios blancos para mejorar la legibilidad.

Todos los dashboards deben incorporar filtros predefinidos y configurables por el usuario para dimensiones clave como periodo, cliente, producto, unidad de negocio y segmento, promoviendo exploración guiada sin alterar los cálculos base.

La capa semántica (modelo de datos en Power BI) debe ser estandarizada: nombres amigables, medidas documentadas y categorización clara de dimensiones e indicadores.

## **Políticas**

**Política de publicación controlada:** Todo dashboard debe pasar por una revisión formal de aseguramiento de calidad (QA) antes de ser publicado. Esta revisión incluye validación de consistencia numérica, revisión visual, performance de carga y validación de seguridad.

**Política de acceso diferenciado:** Los dashboards deben configurarse bajo un modelo de acceso por roles y perfiles. Información estratégica o sensible (por ejemplo: cartera vencida, rentabilidad, margen neto) debe estar restringida exclusivamente a usuarios con autorización expresa.

**Política de trazabilidad de indicadores:** Cada KPI publicado debe estar vinculado a una ficha técnica accesible desde el mismo dashboard, que contenga: definición del indicador, fórmula de cálculo, frecuencia de actualización, fuente de datos y responsable del dato.

### **Controles**

Validación mensual de consistencia entre los valores mostrados en el dashboard y el dataset fuente, especialmente en indicadores clave (ventas, cartera, operaciones) mediante validaciones cruzadas por parte de Data Stewards.

Seguimiento automatizado del uso de dashboards: usuarios activos, frecuencia de acceso, páginas más vistas, abandono temprano, zonas de menor interacción. Esta analítica de uso permite detectar dashboards obsoletos o mejoras potenciales.

Configuración de alertas automáticas por fallos de actualización, desconexión de fuente, errores de gateway o KPIs que no se estén calculando correctamente. Estas alertas deben ser monitoreadas por el equipo de soporte de BI y derivadas a los responsables técnicos.

### **Justificación**

Los dashboards constituyen el punto de contacto más frecuente entre los usuarios del negocio y el ecosistema de datos. Son la interfaz donde los datos se transforman en decisiones, y, por tanto, su integridad visual, funcional y conceptual es crítica para la credibilidad de todo el modelo de gobernanza.

Un dashboard con errores, desactualizado o mal diseñado puede inducir a decisiones erradas, pérdida de confianza institucional o rechazo del sistema por parte de las áreas funcionales. En FAM TEAM, donde los dashboards soportan procesos comerciales, financieros y operativos en tiempo casi real, la estandarización visual y la validación numérica garantizan una experiencia homogénea, confiable y alineada con los principios de “fuente única de verdad” y “dato confiable por diseño”.

Además, monitorear el uso de los dashboards permite entender patrones de consumo de información, priorizar mejoras, optimizar el desarrollo futuro y eliminar reportes que no agregan valor. La trazabilidad de indicadores, junto con el control de versiones y el acceso seguro, convierten a los dashboards no en simples herramientas visuales, sino en plataformas institucionales de apoyo estratégico para toda la organización.

### **Consulta y Análisis por Áreas**

#### **Estándares**

Todos los colaboradores deben tener acceso controlado y perfilado al catálogo de datos y a los dashboards predefinidos relevantes según su rol, área y nivel de decisión. Este acceso debe estar alineado con el modelo de datos institucional y soportado por autenticación corporativa.

Las solicitudes de nuevos reportes, visualizaciones o consultas específicas deben canalizarse mediante un procedimiento estandarizado que incluya: formato digital común (formulario), responsable del requerimiento, objetivo de negocio, prioridad, y tiempo estimado de respuesta (SLA). Esto promueve eficiencia, trazabilidad y gestión de la demanda.

Las herramientas de análisis institucional (ej. Power BI, Excel conectado a Azure, herramientas SQL) deben estar homologadas, actualizadas y mantenidas por el área de TI, asegurando coherencia tecnológica, integridad de resultados y soporte técnico en caso de incidentes.

### **Políticas**

**Política de autoservicio analítico:** Se debe capacitar y habilitar a los usuarios funcionales para la exploración y análisis autónomo de la información, sin depender de equipos de datos o tecnología para obtener métricas básicas o responder preguntas recurrentes.

**Política de protección de datos sensibles:** Toda consulta, reporte o visualización que contenga datos personales, financieros o estratégicos debe aplicar mecanismos de enmascaramiento por defecto (ocultamiento, o anonimización), conforme a normativas como Habeas Data y la política interna de confidencialidad.

**Política de trazabilidad y gobernanza de consultas:** Toda extracción, análisis o explotación de datos debe quedar registrada, ya sea de forma automática (en dashboards y

herramientas conectadas) o manual mediante bitácoras de solicitudes. Esto permite auditar el uso, rastrear decisiones y garantizar cumplimiento ético y normativo.

### **Controles**

Registro automático del uso de recursos analíticos (dashboards, catálogo, queries en Azure o SQL), incluyendo usuario, fecha/hora, propósito de consulta y tiempo de sesión. Esta data es analizada periódicamente para detectar necesidades de capacitación, brechas de adopción o uso ineficiente.

Medición continua del cumplimiento del SLA de solicitudes ad-hoc (desde la recepción hasta la entrega del reporte o consulta), con alertas para retrasos críticos y reporte mensual por unidad funcional.

Aplicación de encuestas semestrales de satisfacción y efectividad analítica por área, con foco en percepción de accesibilidad, utilidad, confiabilidad del dato y oportunidades de mejora. Los resultados deben alimentar un plan de mejora continua en analítica de negocio.

### **Justificación**

Una estrategia de gobierno de datos efectiva no se limita a definir estructuras técnicas: su verdadero impacto ocurre cuando empodera a los usuarios finales. En FAM TEAM, donde las decisiones operativas y comerciales deben tomarse con rapidez y fundamento, garantizar un

acceso oportuno, seguro y comprensible a la información es vital para mejorar la competitividad y la agilidad organizacional.

El modelo de consulta y análisis debe evolucionar desde una dependencia de analistas centrales hacia un esquema descentralizado y autosuficiente, donde cada área pueda resolver de forma autónoma sus preguntas clave, explorar tendencias y anticiparse a riesgos. Esto no solo reduce la sobrecarga de los equipos de BI, sino que acelera la reacción frente a cambios del mercado, reclamos de clientes o ajustes regulatorios.

Además, mediante el registro y análisis del uso de los recursos analíticos, la organización puede ajustar los entrenamientos, redefinir los dashboards, actualizar los catálogos y diseñar estrategias de adopción específicas por perfil. En resumen, esta dimensión convierte al gobierno de datos en un habilitador del desempeño organizacional, fortaleciendo la inteligencia distribuida y construyendo una cultura de decisiones basadas en evidencia en todos los niveles de FAM TEAM.

## **Modelaje Analítico en Azure**

### **Estándares**

Todo modelo predictivo o clasificatorio debe estar respaldado por una documentación técnica completa que incluya: justificación del problema de negocio, dataset utilizado, lógica de entrenamiento, ficha detallada de variables (tipo, fuente, tratamiento, correlaciones), y evaluación de su rendimiento con métricas específicas.

Los entornos de desarrollo, pruebas (UAT) y producción deben estar claramente segregados para evitar despliegues prematuros o no autorizados. Los modelos deben moverse

entre entornos mediante pipelines automatizados con validaciones técnicas.

Se deben definir umbrales mínimos de performance según el tipo de modelo y su impacto en el negocio. Indicadores obligatorios: precisión, recall, F1-score y AUC. Modelos con desempeño inferior al umbral establecido (ej.  $F1 < 0.75$ ) no podrán ser liberados sin justificación técnica y validación del Comité de Datos.

El código y los artefactos del modelo (scripts, métricas, logs, configuraciones de entorno) deben estar versionados en repositorios como Git y documentados bajo estándares reproducibles (ej. notebooks en MLFlow o Azure Machine Learning Studio).

## **Políticas**

**Política de aplicabilidad y comprensión funcional:** Todo modelo en uso debe contar con un resumen ejecutivo que explique su lógica, variables clave, criterios de decisión y restricciones, redactado en lenguaje no técnico. Este resumen debe ser accesible para los responsables de negocio que utilizan el modelo.

**Política de versionamiento controlado:** No se permitirá sobrescribir ni eliminar versiones anteriores de modelos productivos sin dejar trazabilidad completa. Toda actualización debe registrar cambios en variables, hiperparámetros, resultados y responsable técnico.

**Política de revisión ética y de sesgo algorítmico:** Todo modelo debe pasar por una revisión de impacto ético antes de su uso en producción, evaluando si introduce sesgos

injustificados por género, edad, raza, ubicación geográfica o cualquier otra categoría protegida. Se aplicarán métricas específicas de equidad y fairness para mitigar estos riesgos.

### **Controles**

Implementación de paneles de monitoreo activo en Azure (o herramienta equivalente) donde se observen en tiempo real: accuracy, tasa de error, distribución de predicciones, tasa de rechazo/aprobación, y desbalance de clases. Estos paneles deben estar disponibles para el equipo de Data Science y los Data Owners.

Validación de resultados por parte del Data Owner funcional antes del despliegue de cada versión del modelo. Esta validación debe incluir pruebas con casos reales, revisión de segmentaciones, y coherencia con el conocimiento del negocio.

Auditorías semestrales lideradas por el Comité de Datos, que incluyan revisión de inputs, documentación, métricas, reproducibilidad, gestión de versiones y uso real del modelo. Las auditorías deben levantar alertas si el modelo pierde validez, deja de ser explicable o su rendimiento se degrada.

### **Justificación**

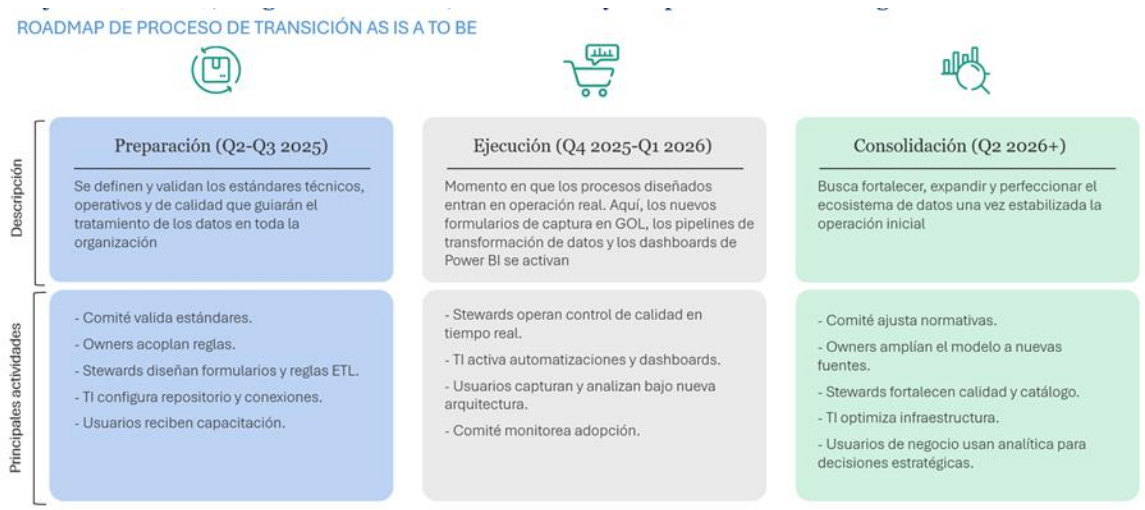
El modelaje analítico representa el punto más sofisticado del ciclo de vida de los datos y la fase en la que las decisiones dejan de ser reactivas y se transforman en proactivas. Sin embargo, el valor que genera un modelo predictivo no radica únicamente en su precisión

matemática, sino en su capacidad de ser entendido, auditado y gestionado como un activo estratégico de la organización.

En el contexto de FAM TEAM, donde la información es utilizada para anticipar riesgos de impago, detectar oportunidades comerciales y optimizar procesos logísticos, contar con modelos robustos, éticos y operativamente sostenibles es una condición necesaria para avanzar hacia una cultura data-driven madura.

Un modelo mal gobernado puede erosionar la confianza del usuario, generar decisiones erradas o incluso incumplir normativas si introduce sesgos o falta de explicabilidad. Por ello, esta dimensión exige un balance entre rigor técnico, transparencia operativa y alineación estratégica. Documentar, auditar, monitorear y comunicar de forma efectiva los modelos no es opcional: es la única forma de asegurar que el conocimiento predictivo se convierta en ventaja competitiva sostenida para FAM TEAM.

### Ilustración 5: Roadmap de transición AS IS a TO BE



Fuente: Elaboración Propia

Estructuramos un roadmap de 3 etapas para guiar de manera ordenada, progresiva y sostenible la transición del modelo de gestión de datos desde el estado actual (AS-IS) hacia la arquitectura objetivo (TO-BE), asegurando calidad, trazabilidad y adopción en toda la organización.

### **Objetivo del Plan de Transición**

Asegurar la adopción efectiva, ordenada y sostenible de los procesos definidos en la arquitectura TO-BE, mediante la activación gradual de los nuevos roles de gobierno de datos, la ejecución de actividades específicas por fase, y la conexión operacional entre los procesos críticos de captura, validación, consulta y modelaje de datos.

### **Fases del Plan de Transición**

El plan se divide en tres grandes fases secuenciales: Preparación, Ejecución y Consolidación. Cada fase está diseñada para cumplir objetivos específicos que aseguran la transición gradual desde el modelo fragmentado actual hacia un ecosistema de datos estructurado, gobernado y sostenible.

### **Fases de Transición Propuestas:**

#### **Fase de Preparación**

La fase de preparación constituye el cimiento estratégico de la transición hacia el modelo TO-BE. Durante este período, se definen y validan los estándares técnicos, operativos y de calidad que guiarán el tratamiento de los datos en toda la organización. El Comité de Gobierno establece las políticas centrales, mientras que los Data Owners traducen el conocimiento del

negocio en reglas concretas. Paralelamente, los Data Stewards diseñan formularios, parametrizan reglas de validación y documentan activos de información esenciales. El equipo de TI configura la infraestructura necesaria en Azure y habilita las conexiones automatizadas de datos, mientras los usuarios finales son capacitados en los nuevos procedimientos. Esta fase asegura que, antes de cualquier despliegue operativo, exista un lenguaje común, una plataforma adecuada y una cultura de datos mínima.

Las actividades clave que deben ocurrir en esta fase son:

**Comité valida estándares:** Aprobar políticas de captura de datos, estructura de almacenamiento, reglas de calidad y criterios de acceso. Esto define el marco común de operación para todas las áreas y asegura coherencia en el tratamiento de los datos. No hacerlo perpetuaría la fragmentación actual y afectaría la confiabilidad del ecosistema de datos.

**Owners definen reglas:** Establecer reglas de negocio específicas: campos obligatorios, valores válidos, formatos estándar. Permite que la captura y transformación de datos respondan a requerimientos reales del negocio. No definir reglas claras llevaría a errores sistemáticos e inconsistencias en la información crítica.

**Stewards diseñan formularios y reglas ETL:** Crear formularios de captura normalizados en GOL y parametrizar reglas de limpieza y validación en ETL. Esto estandariza el flujo de datos desde su origen. Si no se realiza, los errores manuales e incoherencias se trasladarán a todo el flujo analítico.

TI configura repositorio y conexiones: Construir la estructura de almacenamiento en Azure y configurar integraciones automáticas de datos. Permite una gestión centralizada, segura y escalable de los activos de información. No configurarlo mantendría la dispersión y los riesgos de pérdidas de información.

Usuarios reciben capacitación: Formar a usuarios en captura de datos, uso de dashboards y consulta del catálogo. Esto garantiza la adopción operativa del modelo de gobernanza. Sin capacitación, la resistencia al cambio comprometería el éxito de la transición.

### **Fase de Ejecución**

La fase de ejecución representa el momento en que los procesos diseñados entran en operación real. Aquí, los nuevos formularios de captura en GOL, los pipelines de transformación de datos y los dashboards de Power BI se activan y comienzan a ser utilizados. Los Data Stewards monitorean la calidad de los datos en tiempo real, registrando incidentes y ajustes necesarios. El equipo de TI supervisa el funcionamiento de las automatizaciones y asegura la disponibilidad de la infraestructura. Los usuarios finales incorporan el uso de dashboards y modelos predictivos en su gestión diaria. El Comité de Gobierno realiza el seguimiento de la adopción mediante KPIs, permitiendo identificar áreas de mejora inmediata. Esta fase es crítica para validar en campo la robustez de los procesos.

Las actividades por desarrollar en esta fase son:

Stewards operan control de calidad en tiempo real: Monitorear que los datos capturados cumplan estándares definidos, detectando errores al instante. Esto evita la propagación de datos defectuosos. Ignorarlo comprometería la confianza en dashboards y modelos analíticos.

TI activa automatizaciones y dashboards: Desplegar pipelines ETL, actualizar repositorios y publicar dashboards conectados a fuentes actualizadas. Automatiza la disponibilidad de información crítica. No hacerlo prolongaría la dependencia de procesos manuales, inseguros y lentos.

Usuarios capturan y analizan bajo nueva arquitectura: Ingresar datos usando formularios estandarizados y consultar dashboards en la toma de decisiones. Esto institucionaliza la cultura del dato como base del negocio. No capturar bajo el nuevo estándar arruinaría la integridad del sistema analítico.

Comité monitorea adopción: Medir KPIs de captura, calidad y uso de dashboards, realizando ajustes de política cuando sea necesario. Garantiza correcciones tempranas y continuidad del modelo. No monitorear expondría el proyecto a un retroceso hacia prácticas desordenadas.

## **Fase de Consolidación**

La fase de consolidación busca fortalecer, expandir y perfeccionar el ecosistema de datos una vez estabilizada la operación inicial. El Comité de Gobierno ajusta las políticas, estándares y procesos en función de los aprendizajes. Los Data Owners impulsan la incorporación de nuevas fuentes de datos, los Stewards refuerzan las auditorías de calidad, y el equipo de TI optimiza la performance de la infraestructura. Los usuarios finales consolidan su rol como analistas estratégicos, utilizando información confiable para decisiones de alto impacto. Esta fase garantiza que el modelo de gobernanza de datos no solo sea sostenible, sino que evolucione junto con las necesidades del negocio. En esta fase es crucial que:

**Comité ajusta normativas:** Revisar y actualizar cada seis meses las políticas y reglas de datos basadas en la experiencia operacional. Mantiene la vigencia y efectividad del modelo de gobernanza. Si no se actualizan, el modelo se volvería obsoleto frente a cambios de negocio.

**Owners amplían el modelo a nuevas fuentes:** Incorporar nuevas fuentes internas y externas, expandiendo la cobertura analítica del ecosistema. Esto refuerza el valor estratégico de los datos. No incorporar nuevas fuentes limitaría la capacidad competitiva de la organización.

**Stewards fortalecen calidad y catálogo:** Auditar la calidad de los datos regularmente y actualizar el catálogo corporativo de información. Sostiene la confianza de usuarios finales y facilita nuevos análisis. Sin auditorías, el deterioro progresivo de la calidad de datos socavaría el proyecto.

**TI optimiza infraestructura:** Mejorar performance de Azure, integrar nuevas capacidades y escalar almacenamiento conforme a crecimiento. Asegura la estabilidad y eficiencia de toda la arquitectura de datos. No optimizar podría colapsar la operación de sistemas analíticos críticos.

Usuarios usan analítica para decisiones estratégicas: Aplicar insights de dashboards y modelos predictivos en procesos comerciales, financieros y operativos. Cierra el ciclo de madurez de la organización basada en datos. No utilizar analítica condenaría el esfuerzo a una herramienta inactiva sin impacto tangible.

## Plan de Implementación

A continuación, se presenta la hoja de ruta para implementar cada uno de los siete procesos definidos en la arquitectura TO-BE, mediante fichas de iniciativa que detallan: objetivo, alcance, habilitadores, hitos y cronograma estimado.

### Ilustración 6: Implementación proceso de captura de datos en Gol

<b>Iniciativa</b>	1. Implementación del proceso de captura de datos en Gol
<b>Líder</b>	• Coordinador de Transformación Digital / Área Comercial
<b>Objetivo general</b>	• Formalizar y estandarizar la captura de datos desde origen, garantizando calidad desde el inicio del flujo
<b>Alcance / Objetivos específicos</b>	• Diseñar formularios con validaciones. • Parametrizar en GOL los campos obligatorios. • Capacitar usuarios y crear manuales operativos.
<b>Habilitadores</b>	Plataforma GOL, equipo TI, Data Stewards
<b>Presupuesto requerido</b>	Por definir

Hitos clave		Fecha fin
• Levantamiento de requisitos con áreas operativas		Q2 2025
• Diseño de formularios estandarizados		Q2 2025
• Validación de campos obligatorios y reglas		Q3 2025
• Parametrización técnica en GOL		Q3 2025
• Desarrollo de manuales de usuario		Q4 2025
• Capacitación por área funcional		Q4 2025
• Ejecución de pruebas piloto		Q4 2025
• Ajustes a partir del piloto		Q4 2025
• Inicio operativo general		Q1 2026

Cronograma																			
2025				2026				2027				2028				2029			
Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	

Fuente: Elaboración Propia

### Ilustración 7: Implementación del proceso ETL automatizado

<b>Iniciativa</b>	2. Implementación del proceso de ETL
<b>Líder</b>	• Jefe de Tecnología / Líder de Data Engineering
<b>Objetivo general</b>	• Garantizar la calidad, estandarización y trazabilidad de los datos mediante automatización
<b>Alcance / Objetivos específicos</b>	• Configurar pipelines de extracción y staging. • Aplicar reglas de transformación y validación. • Establecer logs y trazabilidad del proceso.
<b>Habilitadores</b>	• Azure Data Factory, Python, SSIS, catálogos de reglas
<b>Presupuesto requerido</b>	• Por definir

Hitos clave	Fecha fin
• Identificación de orígenes de datos	Q2 2025
• Configuración de arquitectura ETL	Q2 2025
• Configuración de zona de staging	Q3 2025
• Desarrollo del pipeline ETL	Q4 2025
• Implementación de reglas de negocio	Q1 2026
• Registro de logs y errores	Q1 2026
• Validación con DataStewards	Q1 2026
• Pruebas de calidad de datos	Q1 2026
• Puesta en producción	Q4 2026

2025												2026				2027				2028				2029			
Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4					

46

Fuente: Elaboración Propia

### Ilustración 8: Repositorio estructurado en Azure

<b>Iniciativa</b>	3. Repositorio estructurado en Azure
<b>Líder</b>	• Coordinador de Infraestructura / Líder de TI
<b>Objetivo general</b>	• Consolidar los datos empresariales en una bodega estructurada y segura
<b>Alcance / Objetivos específicos</b>	• Crear carpetas por dominio de datos. • Definir accesos RBAC y nomenclatura estándar. • Habilitar backups automáticos y control de versiones.
<b>Habilitadores</b>	• Microsoft Azure, políticas de acceso, scripts automatizados.
<b>Presupuesto requerido</b>	• Por definir

Hitos clave	Fecha fin
• Revisión de necesidades de almacenamiento por dominio	Q4 2025
• Diseño lógico de carpetas y estructura jerárquica	Q1 2026
• Validación de modelo por Comité de Gobierno	Q1 2026
• Creación de espacios físicos y configuración en Azure	Q2 2026
• Asignación de permisos RBAC y controles de acceso	Q2 2026
• Migración piloto de archivos depurados	Q3 2026
• Ejecución de limpieza y normalización	Q3 2026
• Integración con procesos ETL activos	Q4 2026
• Habilitación general del repositorio	Q4 2026

2025												2026				2027				2028				2029			
Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4					

47

Fuente: Elaboración Propia

### Recomendaciones de Change Management

Nombrar embajadores del cambio en cada unidad funcional.

Establecer una comunidad de práctica de datos por área.

Crear un boletín mensual con avances, casos de éxito y errores comunes.

Ejecutar encuestas trimestrales de percepción y adopción.

Establecer espacios de retroalimentación abiertos para ajustes en la implementación.

Este plan de implementación permite priorizar y calendarizar de manera realista el despliegue progresivo del modelo TO-BE y asegurar una adopción efectiva en todos los niveles organizacionales.

## **7. Calidad de Datos**

En el contexto actual de FAM TEAM, la dimensión de Calidad de Datos se encuentra caracterizada por la ausencia de políticas, controles automatizados y métricas formales. Los principales hallazgos relacionados con esta dimensión pueden reflejarse en el uso del ERP Gol, cuando el equipo de cartera requiere hacer una consulta del estado de cartera de un cliente:

**Ilustración 9:** *Estado actual de consulta de datos*

**CONSULTAR E IMPRIMIR DOCUMENTOS DE PAGOS Y RECAUDOS**

Movimiento de Obligaciones por Tercero  
 Obligaciones por Cuenta y Línea  
 CXC Todos los conceptos  
 Reporte general CXC

Tercero:  Fecha de Corte:

Línea:

Cuenta Inicial:  Cuenta Final:

**Generar Reporte**

**CARTERA DE CLIENTES**

Empresa: ULS Generado: Apr-21-2025

DOCUMENTO	FECHA	CRED	FILE	VCTO	USD	< 30	< 60	> 60
<b>860021615-AGUNSA COLOMBIA SAS</b>								
AP-012551	03-20-2025	1	250M012551	26	0.00	900,000.00	0.00	0.00
AP-012561	03-20-2025	1	250M012561	26	0.00	900,000.00	0.00	0.00
AP-012562	03-20-2025	1	250M012562	26	0.00	900,000.00	0.00	0.00
AP-012588	03-20-2025	1	250M012588	26	0.00	900,000.00	0.00	0.00
AP-012592	03-09-2025	1	250M012592	47	0.00	0.00	1,800,000.00	0.00
AP-012593	03-09-2025	1	250M012593	47	0.00	0.00	900,000.00	0.00
AP-012594	03-09-2025	1	250M012594	47	0.00	0.00	900,000.00	0.00
AP-012596	03-09-2025	1	250M012596	47	0.00	0.00	900,000.00	0.00
AP-012604	03-09-2025	1	250M012604	47	0.00	0.00	900,000.00	0.00
AP-012611	03-20-2025	1	250M012611	26	0.00	900,000.00	0.00	0.00
AP-012615	03-09-2025	1	250M012615	47	0.00	0.00	900,000.00	0.00
AP-012633	03-12-2025	1	250M012633	48	0.00	0.00	1,994,754.00	0.00
AP-022817	02-25-2025	1	250M022817	55	0.00	0.00	1,789,088.00	0.00
AP-022892	04-02-2025	1	250M022892	19	0.00	900,000.00	0.00	0.00
AP-022897	03-20-2025	1	250M022897	26	0.00	2,500,000.00	0.00	0.00
AP-101965	12-03-2024	1	240M101965	139	0.00	0.00	0.00	903,766.00

Fuente: ERP FAM TEAM

## 7.1. Diagnóstico de Calidad de Datos – Archivo Cartera

**Tabla 16:** Tabla diagnóstico de Calidad de Datos

Dimensión de Calidad DAMA	Descripción del Problema	Ejemplos observados	Impacto en la operación
Precisión	Variaciones en la escritura de nombres de clientes y ciudades.	"ABC S.A." vs "ABC SA" vs "abc s.a."; "BOGOTA" vs "Bogotá D.C."	Dificulta conciliaciones y genera falsos positivos en reportes.
Compleitud	Campos esenciales vacíos o nulos.	Campos "VALOR_FACTURA", "SALDO", "CONCEPTO" sin información.	Impide análisis de cartera vencida y cálculo de indicadores.
Consistencia	Formatos distintos para fechas y nombres.	"01/04/2025", "2025-04-01", "abril 1, 2025".	Genera errores en filtros, cálculos y visualización de datos.

Unicidad	Registros duplicados con diferencias mínimas.	Misma factura con distinta razón social, o duplicada sin motivo técnico.	Sobreestima saldos, induce errores en decisiones de cobranza.
Validez	Ingreso de caracteres no válidos en campos estructurados.	Nombres con símbolos especiales o espacios innecesarios: "CALI*" o " SANTA MARTA "	Reduce la calidad analítica y requiere depuración manual.
Trazabilidad	Ausencia de columnas con fecha de corte, versión, responsable o fuente.	Archivo sin encabezados técnicos que indiquen quién, cuándo y cómo fue generado.	Dificulta la validación cruzada y auditoría del dato.
Control de versiones	No existe historial de versiones ni lógica de nombramiento.	Archivo se comparte por correo o Teams sin control de versiones.	Riesgo de usar versiones desactualizadas, errores en decisiones operativas.

Fuente: Elaboración Propia

**Tabla 17:** *Indicadores de Calidad de Datos*

Dimensión DAMA	Indicador Calculado	Resultado
Precisión	% de facturas con datos de cliente estandarizados	85%
Compleitud	% de registros con todos los campos críticos llenos	82%
Consistencia	% de registros con formato de fecha correcto	88%

Unicidad	% de duplicados detectados por Cliente + N° Factura	5%
Validez	% de registros con valores válidos en campos tipo lista	90%
Trazabilidad	% de registros con fecha de vencimiento válida	85%
Control de versiones	No aplica (manejo manual de archivos Excel)	0%

Fuente: Elaboración Propia

En el estado actual (AS-IS), el área de Cartera de FAM TEAM opera sobre una base de datos exportada desde sistemas operativos a Excel, que presenta múltiples deficiencias de calidad estructural y semántica. La falta de validaciones en los formularios de origen permite la entrada de datos con errores tipográficos, formatos inconsistentes y omisiones críticas en campos como "valor de factura", "saldo", "fecha de vencimiento" o "concepto", lo que impide realizar análisis financieros confiables y oportunos. Además, se detectan registros duplicados, entradas con estructuras disímiles y ausencia de un control formal de versiones, lo que compromete la unicidad del dato y genera múltiples interpretaciones según el área que accede a la información. Estas condiciones exponen a la organización a sobrecostos en conciliaciones manuales, errores en la priorización de cobros y baja trazabilidad operativa.

Por otro lado, la descarga periódica de información no incluye metadatos técnicos esenciales como la fecha de corte, la fuente original o el responsable de la extracción, impidiendo la trazabilidad y auditabilidad del dato. La manipulación directa en Excel, sin un control estructurado ni medidas de seguridad adecuadas, refuerza un modelo de gestión informal y dependiente del conocimiento tácito de usuarios individuales. Esta práctica no sólo incrementa el riesgo de errores, sino que dificulta la interoperabilidad de los datos entre sistemas, afectando

directamente la calidad de los reportes gerenciales y la implementación de modelos analíticos robustos. La ausencia de una “fuente única de la verdad” y de reglas formales de calidad limita gravemente la capacidad de FAM TEAM para tomar decisiones basadas en datos y proyectar escenarios financieros confiables.

### **Estándares de tratamiento de datos:**

Para garantizar una gestión uniforme y eficiente de la información en todas las áreas de FAM TEAM, se definen los siguientes estándares para el tratamiento de datos:

**Formato uniforme:** Todo dato capturado debe respetar un formato previamente definido (ej. fechas en formato YYYY-MM-DD, nombres en mayúsculas con acentos, uso de separadores estándar para montos).

**Campos obligatorios:** Se establecerán campos obligatorios en formularios críticos (ej. NIT, ciudad, monto, tipo de cliente) con validaciones automáticas.

**Control de versiones:** Todo archivo o base de datos debe incluir control de versión (v1.0, v1.1, etc.), fecha de actualización y responsable.

**Etiquetado y nomenclatura:** Los archivos deberán seguir una convención estándar:

[Área]/[NombreBase][AAAA-MM-DD]\_[Versión]. Ejemplo: Comercial\_LeadsGOL\_2025-04-15\_v1.0.xlsx

**Separación de datos sensibles:** Se deberá identificar y clasificar los datos personales, financieros o confidenciales, para aplicar mecanismos de protección y acceso diferenciado.

**Temporalidad:** Los datos deberán contar con una fecha de creación y/o vigencia. Se establecerán ciclos de retención y eliminación según la categoría del dato.

**Tabla 18:** *Tabla de acciones propuestas para mejorar la Calidad de Datos*

Dimensión DAMA	Acción Correctiva Propuesta	Responsable Principal	Herramienta / Soporte	Beneficio Esperado
Precisión	Estandarización de campos como nombres y ciudades con listas controladas.	Data Steward – Cartera	Validaciones en formularios de ERP / Catálogos	Reducción de errores tipográficos y duplicidad de entidades.
Compleitud	Configurar campos obligatorios y alertas de valores nulos en formularios de ingreso.	Data Steward – Cartera	ERP GOL / Power Automate	Mejora del análisis financiero y confiabilidad del reporte.
Consistencia	Unificación de formatos de fechas, textos y códigos mediante reglas de transformación (ETL).	Equipo de TI + Data Steward	Azure Data Factory / Power BI	Uniformidad entre reportes y disminución de reprocesos.
Unicidad	Implementación de reglas de unicidad en campos clave (NIT, Factura) y de duplicación periódica.	Comité de Gobierno de Datos	SQL + Power BI	Evitar sobreestimaciones y errores contables.

Validez	Validaciones automáticas al ingreso (ej. longitud, caracteres válidos, listas desplegables).	Data Steward	Formularios ERP / Regla de negocio GOL	Mejora la calidad desde el origen, reduce correcciones.
Trazabilidad	Inclusión de columnas técnicas: fecha de corte, origen, responsable de extracción.	Equipo de TI	Power BI	Aumenta la auditabilidad y confianza en los datos.
Control de versiones	Implementación de control de versiones con nomenclatura estandarizada y registro de cambios.	Comité de Gobierno de Datos	Azure repositorio / Política documental	Mejora el seguimiento y evita el uso de archivos obsoletos.

Fuente: Elaboración Propia

La estrategia TO-BE para mejorar la calidad de datos en FAM TEAM se fundamenta en el establecimiento de estándares claros, procesos automatizados y responsabilidades distribuidas en torno al ciclo de vida del dato. Este modelo propone la implementación de controles desde el origen, mediante formularios con validaciones automáticas en el ERP GOL, listas desplegables, catálogos maestros y reglas de negocio que garanticen precisión, consistencia y validez. Estas medidas evitarán que se ingresen datos con errores tipográficos, formatos incorrectos o estructuras inadecuadas, promoviendo una cultura de calidad desde la captura de la información. Asimismo, se establecerán campos obligatorios para asegurar la completitud, evitando vacíos críticos que afecten la trazabilidad y el análisis financiero.

Adicionalmente, se formalizaron políticas, estándares y controles operativos que permitirán mantener la calidad de los datos en el tiempo. Entre ellos se destacan la política de calidad alineada con DAMA, la implementación de catálogos controlados, y las validaciones automáticas desde el origen mediante formularios ERP. Se incluyeron también controles de unicidad, revisión de KPIs mensuales, reportes automáticos en Power BI y estructuras mínimas de trazabilidad. Estos elementos brindan un marco robusto que sustenta no solo la integridad y consistencia de los datos, sino también la capacidad de auditar y tomar decisiones basadas en información verificada.

Además, el modelo considera la creación de indicadores específicos de calidad por dominio de datos, tales como porcentaje de errores, duplicados, valores nulos y tiempo promedio de corrección. Estos indicadores serán monitoreados por los Data Stewards y reportados periódicamente al Comité de Gobierno de Datos, permitiendo una gestión proactiva de la calidad y el fortalecimiento de la toma de decisiones. Complementariamente, se incluirán columnas técnicas en las bases operativas para registrar el origen del dato, la fecha de corte, y el responsable de su extracción, fortaleciendo la trazabilidad, la auditabilidad y la gobernabilidad de los datos utilizados por las áreas funcionales y los modelos analíticos.

Finalmente, el modelo TO-BE establece un marco de control de versiones y nomenclatura estándar que permite identificar fácilmente los datos más recientes y evitar el uso de archivos obsoletos. Para ello, se recomienda el uso de repositorios estructurados en Azure con control de accesos por rol (RBAC), respaldos automáticos y trazabilidad de cambios. Estas mejoras serán acompañadas por una estrategia de alfabetización en datos, que incluirá formación en principios de calidad, manejo de herramientas analíticas y uso responsable de la información.

Con estas acciones, FAM TEAM podrá consolidar una “fuente única de la verdad” y asegurar que sus decisiones estratégicas y operativas se basen en datos confiables, oportunos y bien gobernados.

## 8. Modelo

En este proyecto construimos un modelo analítico completo para predecir la probabilidad de que una empresa pague o no a tiempo, utilizando datos financieros, de comportamiento de mora, y condiciones contractuales. El proceso fue desarrollado en cinco grandes fases, cada una con objetivos específicos y decisiones técnicas fundamentadas en buenas prácticas de ciencia de datos y análisis de riesgo.

### **Fase 1: Exploración y Análisis Inicial de los Datos**

El primer paso esencial en la construcción del modelo predictivo fue la exploración de los datos a través de análisis estadísticos y visuales. Esto incluyó:

**Histogramas:** para visualizar la distribución de variables como ingresos, activos, pasivos, ganancia neta, etc. Esta herramienta permitió detectar asimetrías, valores extremos o distribuciones sesgadas.

**Boxplots:** ayudaron a identificar outliers (valores atípicos) y diferencias entre las clases objetivo (empresas que pagan vs. no pagan a tiempo).

**Correlaciones:** se calcularon para detectar multicolinealidad entre variables, lo cual podría afectar la estabilidad de algunos modelos.

Estas visualizaciones y estadísticas descriptivas fueron fundamentales para comprender la estructura del conjunto de datos, justificar transformaciones posteriores y detectar la necesidad de preprocesamiento.

## **Fase 2: Preparación y Preprocesamiento de los Datos**

La base de datos inicial contenía información detallada de empresas, incluyendo activos, pasivos, ingresos, ganancias, número de empleados, impagos recientes y condiciones de pago. Para construir un modelo confiable, fue necesario limpiar y preparar estos datos:

**Selección de variables:** Escogimos 16 variables predictoras numéricas que guardaban relación con la salud financiera de la empresa y su comportamiento de pago. Eliminamos variables tipo fecha para evitar errores de tipo y porque no aportaban valor directo al modelo sin transformaciones adicionales.

**Manejo de valores faltantes:** Aplicamos un imputador basado en la media para rellenar datos faltantes. Esto es esencial porque muchos modelos no pueden manejar valores nulos y eliminar registros reduciría el volumen de información disponible.

**Estandarización:** Escalamos todas las variables para que tuvieran media cero y desviación estándar uno. Esto mejora la estabilidad numérica del modelo y asegura que variables con escalas distintas no dominen la predicción.

**División del dataset:** Separamos el 25% de los datos como conjunto de prueba lo que nos permitió evaluar el modelo con datos que nunca ha visto, asegurando una medición realista de su capacidad predictiva.

### **Fase 3: Entrenamiento del Modelo Random Forest**

Elegimos el algoritmo de Random Forest (tras haber comparado con una regresión logística) por su robustez, capacidad para manejar relaciones no lineales y variables heterogéneas, y por su interpretabilidad mediante la importancia de variables.

#### **Rebalanceo con SMOTE:**

En el conjunto de entrenamiento original, existía un desbalance de clases significativo: muchas más empresas que pagaban a tiempo que las que no. Para evitar que el modelo aprendiera a predecir siempre la clase mayoritaria, se utilizó SMOTE (Synthetic Minority Over-sampling Technique), una técnica que genera datos sintéticos para balancear las clases. Esto permite:

Mejorar el recall y la f1-score de la clase minoritaria.

Construir un modelo justo y sensible a ambos resultados.

En lugar de simplemente duplicar ejemplos minoritarios, SMOTE genera nuevos ejemplos sintéticos a partir de los ya existentes, interpolando entre ellos. Esto permite aumentar el número de observaciones de la clase minoritaria de forma más realista y útil para el modelo (Microsoft Learn).

#### **Entrenamiento del modelo:**

Se desarrollaron y evaluaron dos modelos:

Regresión logística: modelo lineal clásico, fácil de interpretar, pero con capacidad limitada para capturar relaciones no lineales entre las variables.

Random Forest: un ensamble de múltiples árboles de decisión, que permite capturar interacciones complejas y patrones no lineales, con excelente rendimiento en tareas de clasificación.

Ambos modelos fueron evaluados sobre el conjunto de validación, y Random Forest superó a la regresión logística en todas las métricas

Para el desarrollo del random forest modelo con el que nos quedamos, lo entrenamos con los datos balanceados. Este modelo construye múltiples árboles de decisión y vota por la clase final. Es resistente al sobreajuste y captura interacciones complejas sin necesidad de transformaciones manuales.

**Evaluación:** Probamos el modelo con el 25% restante. Obtuvimos un AUC de 0.9465, un índice de Gini de 0.8931 y una precisión equilibrada de 0.9236, lo cual demuestra un excelente desempeño para distinguir entre buenos y malos pagadores.

#### **Fase 4: Guardado y Persistencia del Modelo**

Una vez entrenado, el modelo fue guardado como archivo .pkl junto con los objetos de escalado (scaler.pkl), imputación (imputer.pkl) y la lista de columnas (features.pkl). Esto permite reutilizar el modelo sin reentrenarlo y garantiza que futuros datos sean transformados de forma coherente. Este paso es crítico para la escalabilidad del modelo y su uso en ambientes productivos como APIs, dashboards o integraciones con otros sistemas de decisión.

## Ilustración 10: Evaluación modelo de datos

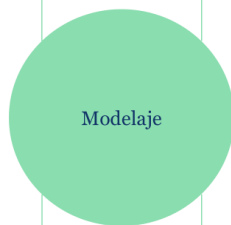
### MODELAJE

#### 1. Preparación y procesamiento de los datos

- Importamos la base de datos con información financiera y de comportamiento de clientes.
- Seleccionamos las variables predictoras relevantes, eliminando columnas tipo fecha
- Tratamos los valores faltantes imputándolos con la media, revisamos correlaciones
- Estandarizamos los datos (StandardScaler) para que todas las variables tuvieran la misma escala y fueran comparables.
- Dividimos el dataset en un conjunto de entrenamiento (75%) y uno de prueba (25%) para evaluar el desempeño real del modelo.

#### 2. Entrenamiento del modelo Random Forest

- Aplicamos rebalanceo (SMOTE) sobre el conjunto de entrenamiento para corregir el desbalance entre empresas que pagaban y no pagaban.
- Entrenamos un modelo Random Forest después de haber comparado con regresiones logísticas y diferentes sets de variables
- Evaluamos el modelo con el conjunto de prueba, obteniendo métricas de desempeño como matriz de confusión, reporte de clasificación, AUC, índice de Gini, y precisión equilibrada.



#### 3. Guardado del modelo y sus objetos auxiliares

- Guardamos el modelo entrenado en un archivo .pkl usando joblib, para poder reutilizarlo sin necesidad de reentrenarlo cada vez.
- Guardamos también el imputador y el escalador (imputer.pkl y scaler.pkl), asegurando que nuevos datos puedan ser procesados de la misma manera que los datos originales de entrenamiento.
- Guardamos la lista de variables usadas

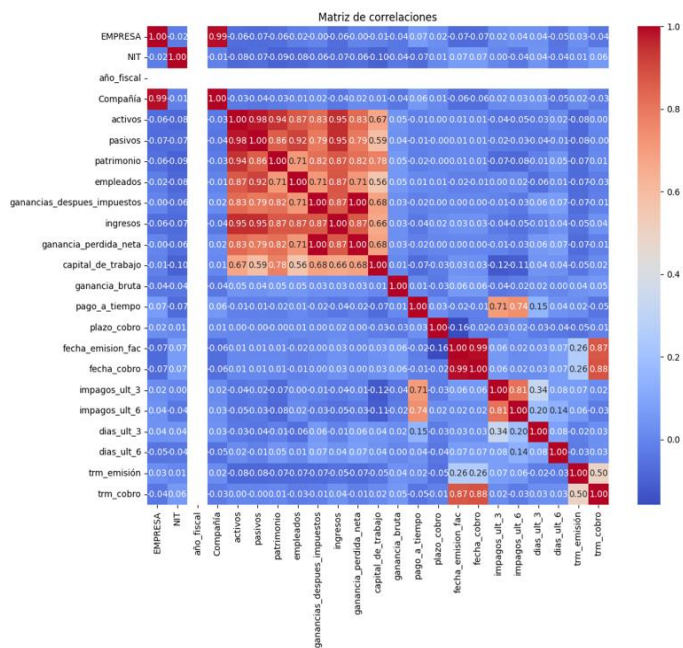
#### 4. Predicción de nuevas empresas

- Creamos una función personalizada que permite recibir los datos de una nueva empresa como un diccionario.
- Cargamos el modelo y los objetos .pkl, asegurando el orden correcto de las variables.
- Imputamos, escalamos y predecimos sobre nuevos registros de empresas.
- Calculamos y mostramos la probabilidad de que la nueva empresa pague o no pague a tiempo, usando el modelo ya entrenado.

Fuente: Elaboración Propia

## Ilustración 11: Matriz de correlaciones

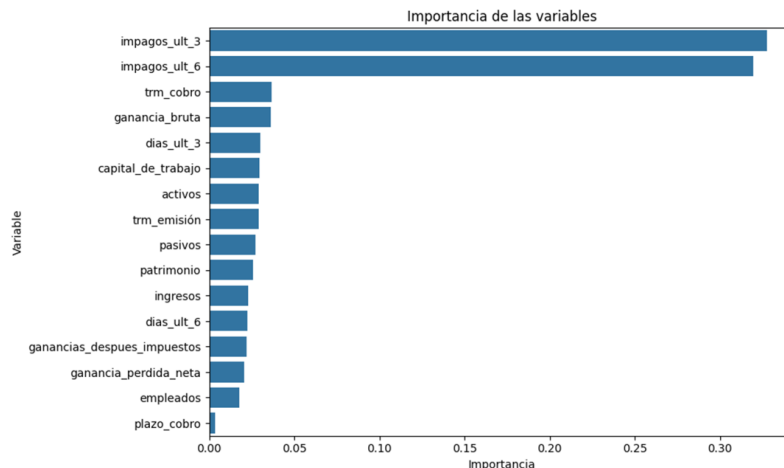
### VARIABLES



Fuente: Elaboración Propia

### Ilustración 12: Importancia de las variables en el modelo

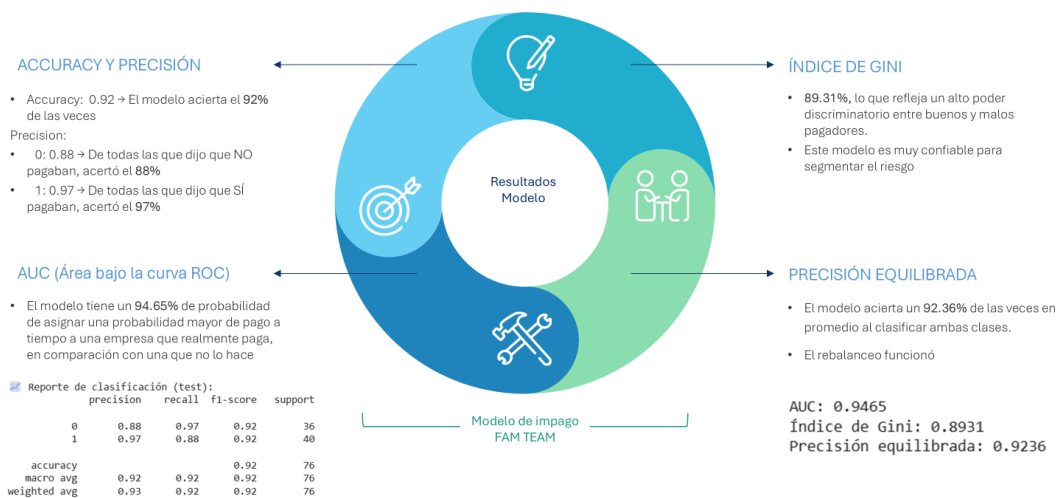
VARIABLES



Fuente: Elaboración Propia

### Ilustración 13: Estudio y conclusiones del modelo de datos

Después de varias iteraciones el random forest arroja un AUC de 0,95 y GINI de 0,89 – reflejando la confiabilidad del modelo para la toma de decisiones operativas



Fuente: Elaboración Propia

## **9. Validación de Estrategia de Proyecto Empresarial – Metodología TDSP+Lego**

A continuación, describiremos como cada una de las 5 etapas de la metodología TDSP ha sido y será aplicada en nuestro proyecto empresarial.

### **Business Understanding**

En esta fase identificamos los objetivos de negocio y los retos que el proyecto de gobierno de datos busca resolver. En el caso de FAM TEAM, realizamos un análisis profundo de los procesos operativos con énfasis en los de cartera y comerciales para comprender cómo las ineficiencias actuales afectan la eficiencia operativa y la toma de decisiones. Las actividades principales incluyeron:

Identificación de los objetivos estratégicos: Alineación del marco de gobierno de datos con los pilares corporativos de FAM TEAM, como liderazgo en logística y enfoque en el cliente.

Definición de indicadores clave (KPIs): Diseño de métricas para medir el éxito del proyecto, como la precisión en la segmentación de clientes o la mejora en la tasa de contacto efectivo de cartera.

Recolección de requisitos: Entrevistas y encuestas con stakeholders para determinar las necesidades específicas de información, deficiencias en gobierno y calidad de datos, modelaje y visualización de datos.

### **Data Acquisition and Understanding**

En esta fase recopilamos y analizamos los datos disponibles para identificar su calidad, relevancia y potencial de uso. Esto implicó un diagnóstico exhaustivo de las bases de datos existentes y las fuentes externas. En esta fase realizamos:

Revisión de las fuentes de datos: Evaluación de bases de datos comerciales, y de cartera para identificar inconsistencias y redundancias.

Análisis de calidad de datos: Uso de dimensiones como completitud, consistencia y precisión para diagnosticar problemas.

Integración inicial: Creación de un entorno de prueba para consolidar y unificar las fuentes de datos bajo un formato estandarizado.

### **Modeling**

En esta etapa, planeamos desarrollar los modelos analíticos y las herramientas de visualización que transformarán los datos de FAM TEAM en información útil. Esto incluye la implementación de dashboards y modelos predictivos. Los pasos principales que contemplaremos son:

Diseño de modelos predictivos: Desarrollo de algoritmos para segmentar clientes y predecir comportamientos, como el pago de deudas.

Prototipado de visualizaciones: Creación de dashboards interactivos para las áreas comercial y de cartera, enfocados en la interpretación rápida de KPIs.

Validación de modelos: Pruebas iniciales con datos históricos para ajustar la precisión y la utilidad de los modelos desarrollados.

### **Deployment**

Esta fase consiste en implementar las soluciones desarrolladas en el entorno operativo de FAM TEAM. En esta fase nuestro objetivo será garantizar que los modelos y visualizaciones sean accesibles y funcionales para los usuarios finales. Actividades principales incluyen:

Configuración de la infraestructura: Integración de los modelos y dashboards con las herramientas y sistemas de la organización.

Capacitación del personal: Entrenamiento en el uso de las herramientas y las mejores prácticas para interpretar los resultados.

Monitoreo inicial: Seguimiento del desempeño de las soluciones para identificar oportunidades de mejora.

### **Customer Acceptance**

Partiendo de lo que aprendimos en clase, y de que el éxito del sistema LEGO se valida mediante la aceptación de los usuarios y la mejora en los procesos de negocio. En esta fase, mediremos el impacto de las soluciones implementadas y recopilaremos feedback para realizar ajustes. Las actividades que realizaremos en esta fase son:

Evaluación de resultados: Comparación de los KPIs iniciales con los obtenidos tras la implementación.

Encuestas de satisfacción: Recolección de opiniones de los usuarios sobre la facilidad de uso y la utilidad de los dashboards y modelos.

Ajustes finales: Refinamiento de las herramientas y procesos con base en los comentarios recibidos.

## **10. Descripción de las Fuentes de Información**

### **Datos Internos de FAM TEAM:**

Bases de Datos Operativas: Registros detallados de las operaciones logísticas, incluyendo tiempos de entrega, rutas, y volúmenes de carga. Esta información será clave para estandarizar y unificar variables de referencia en las operaciones.

Datos Comerciales: Información sobre las transacciones comerciales, como ventas, clientes, pedidos y contratos. Estos datos serán utilizados para construir visualizaciones personalizadas para el área comercial, ayudando en el análisis de patrones de ventas y preferencias de los clientes.

Información de Cartera: Datos relacionados con el área de cartera, incluyendo el historial de pagos, deudas y niveles de morosidad. Estos datos serán esenciales para el diseño de dashboards que faciliten el seguimiento de cuentas por cobrar y la toma de decisiones en el área financiera.

### **Datos de Recursos Humanos:**

Roles y Accesos a Datos: Información sobre los roles y permisos actuales dentro de la compañía, proporcionada por el departamento de Recursos Humanos. Esta fuente permitirá definir y ajustar accesos en el marco de gobierno de datos para asegurar la seguridad y el cumplimiento en la gestión de la información.

### **Datos Externos:**

Tendencias de la Industria Logística: Informes de la industria y estudios de mercado sobre tendencias y mejores prácticas en la gestión de datos y optimización operativa en logística.

Esta información externa permitirá a FAM TEAM alinear sus prácticas de gobierno de datos con los estándares de la industria.

**Datos Económicos y del Mercado:** Información macroeconómica que impacta el sector logístico, como indicadores de inflación, tasas de cambio y políticas comerciales. Estos datos ayudarán a comprender el contexto en el que opera la empresa y a realizar análisis más precisos en las áreas de cartera y comercial.

**Datos de Benchmarking:**

**Mejores Prácticas de Gobierno de Datos:** Referencias de marcos de gobierno de datos exitosos en otras empresas del sector logístico, que servirán como base para el diseño de un modelo efectivo para FAM TEAM.

**Casos de Éxito en Visualización de Datos:** Ejemplos de visualizaciones y dashboards utilizados por líderes de la industria para el análisis de datos comerciales y financieros, los cuales proporcionarán ideas y estándares visuales que puedan aplicarse en las visualizaciones desarrolladas para FAM TEAM.

**Encuestas y Feedback del Personal:**

**Percepción y Necesidades de los Usuarios:** Encuestas y entrevistas con los equipos de los departamentos comercial y de cartera para identificar las necesidades específicas de visualización y el tipo de datos más relevantes para su toma de decisiones diaria. Este feedback permitirá adaptar el diseño de los dashboards a los requerimientos reales de los usuarios.

## 11. Referencias Bibliográficas

- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.
- Data Governance Institute. (n.d.). Definitions of Data Governance. Recuperado de <https://datagovernance.com/>
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. 2nd Edition. Burlingame, CA: Analytics Press.
- IBM Analytics. (n.d.). What is Data Governance? Recuperado de <https://www.ibm.com/analytics/data-governance>
- International Organization for Standardization (ISO). (2008). *ISO/IEC 38500:2008 – Corporate governance of information technology*. ISO.
- ISO/IEC 25012:2008. (2008). *Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model*. International Organization for Standardization.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.
- Microsoft. (2024, septiembre 1). *SMOTE: Referencia del componente*.
- Redman, T. C. (2018). *Getting in front on data: Who does what*. Harvard Business Review Press.

Scopesi, M., & Butelli, P. (2022). Data Governance y Data Analytics en el sector logístico: Buenas prácticas y casos de éxito. *Journal of Logistics*, 12(3), 345-367.