



Universidad Cenfotec
Maestría en Tecnología de Bases de Datos

Documento final de Proyecto de Investigación Aplicada 2

Implementación de una infraestructura escalable del producto tarjetas

Brown Sirias Agustín
Chavarría Arroyo Miguel

Agosto, 2020

DEDICATORIA

A *Dios*, pues obramos en su nombre y todos nuestros logros son para su gloria.

Agustín Brown Sirias:

A mi esposa, Glorieta Marchena Rosales, por su comprensión y apoyo en todo momento, así como por impulsarme a terminar este proyecto. A mis abuelos (Pedro y Ángela) y madre (Reyna), que desde un inicio me motivaron para iniciar con mis estudios en la universidad, los llevo siempre en mi corazón.

Miguel Chavarría Arroyo:

A mi familia, por todo el apoyo durante mi formación académica y profesional. En especial a mi padre Jorge Chavarría Gorgona y a mi madre Ana Arroyo Montero, por su ejemplo de humildad, honestidad, tenacidad, trabajo y esfuerzo; además, por demostrarme que no hay obstáculo invencible y que en su gran mayoría todos radican en la mente. A mi esposa Yahaira Aguilar Jiménez, por su apoyo incondicional, por el sacrificio de días de descanso, por siempre mostrarse dispuesta a ser anfitriona de las largas jornadas de trabajo y por recordarme a diario la importancia de concluir este proyecto.

AGRADECIMIENTOS

En primer lugar a *Dios*, por permitirnos vivir en un país lleno de imperfecciones, pero con las condiciones necesarias para desarrollar nuestro potencial utilizando nuestro propio esfuerzo. A nuestra *familia*, por el apoyo y paciencia durante este tiempo de estudio. A nuestros *compañeros de maestría y profesores*, quienes muy amablemente siempre nos apoyaron brindando su opinión sincera, crítica constructiva y aporte de ideas.

Agradecemos especialmente a todo el personal de CENFOTEC, por su comprensión y apoyo para poder finalizar esta maestría, aunque fue un camino duro y difícil.

Gracias a nuestro tutor y profesor, José Cabezas, por toda la paciencia y motivación, por su disponibilidad y entrega, y por creer en este proyecto, el cual fue un motor para poder finalizar.

Gracias al compañero Miguel Chavarría, por su entrega, dedicación y por dar lo mejor, fue un gusto trabajar con usted.

Gracias al compañero Agustín Brown, por ser la persona pujante del equipo, por su empatía y disciplina, y por siempre buscar e impulsar la excelencia en este proyecto, fue un gusto trabajar con usted.

HOJA DE APROBACIÓN DEL PROYECTO



Universidad Cenfotec
Carrera de Postgrado
Maestría en Tecnología de Bases de Datos

TRIBUNAL EXAMINADOR

Este proyecto fue aprobado por el Tribunal Examinador de la carrera: **Maestría en Tecnologías de Bases de Datos**, requisito para optar por el título de grado de **Maestría**, para los estudiantes: **Agustín Brown Sirias y Miguel Chavarría Arroyo**

JOSE ALBERTO
CABEZAS JAIKEL
(FIRMA)

Digitally signed by JOSE
ALBERTO CABEZAS JAIKEL
(FIRMA)
Date: 2020.08.11 14:32:40
-06'00'

MBD José Cabezas Jaikel

Tutor

LUIS ALBERTO MONGE
FUENTES (FIRMA)

Firmado digitalmente por LUIS
ALBERTO MONGE FUENTES
(FIRMA)
Fecha: 2020.08.11 16:55:57 -06'00'

MBA Luis Monge Fuentes

Lector 1

IGNACIO TREJOS
ZELAYA (FIRMA)

Digitally signed by IGNACIO
TREJOS ZELAYA (FIRMA)
Date: 20.20.08.11 18:38:22
-06'00'

M. Sc. Ignacio Trejos Zelaya

Lector 2

San José, Costa Rica, 10 de agosto de 2020

Firmado digitalmente, de conformidad con la Ley de Certificados, Firmas Digitales y Documentos Electrónicos N° 8454, destacando el artículo 9°.

RESUMEN EJECUTIVO

Este proyecto expone los pasos necesarios para crear las estructuras de datos requeridas, que proporcionarán un alto nivel de detalle de información relacionada con el producto tarjetas (crédito y débito) de forma diaria.

La secuencia de pasos utilizada describe cómo se implementó la automatización de los procesos manuales para generar una plataforma robusta que permitió la elaboración de una estructura flexible de análisis e inteligencia de negocios.

Los procedimientos analizados y la implementación de los mismos, para la carga de la información en la base de datos, se realizó utilizando las mejores prácticas al buscar la optimización en cada nivel de los procesos. El sentido primordial será siempre agrupar la información de los archivos de manera congruente para que los datos obtenidos sigan una estructura conveniente.

La segmentación de la información en varios esquemas de trabajos permitió que los pasos de la carga de la información se fuesen llevando a cabo de manera natural y, por ende, la información quedó respaldada de modo congruente.

Con esta propuesta se buscó proporcionar una herramienta que permita generar un depósito de datos para su autoservicio, el cual sirva como punto inicial para posteriormente depurarlo según sus necesidades.

TABLA DE CONTENIDO

1. CAPÍTULO 1: INTRODUCCIÓN	1
1.1. ANTECEDENTES DEL PROBLEMA	1
1.2. JUSTIFICACIÓN	1
1.3. VIABILIDAD	2
1.3.1. Factibilidad técnica	2
1.3.2. Factibilidad económica	2
1.3.3. Factibilidad operacional	4
1.4. PLANTEAMIENTO DEL PROBLEMA	4
1.4.1. Problema general	4
1.4.2. Subproblemas del proyecto	4
1.5. OBJETIVOS DEL PROYECTO	5
1.5.1. Objetivo general	5
1.5.2. Objetivos específicos	5
1.6. CRONOGRAMA DEL DESARROLLO DE LA PROPUESTA	6
1.7. MARCO DE REFERENCIA ORGANIZACIONAL Y SOCIOECONÓMICO	8
1.7.1. Reseña histórica	8
1.7.2. Misión	9
1.7.3. Visión	9
1.7.4. Valores de la organización	9
1.8. ALCANCES Y LIMITACIONES DEL PROYECTO	11
1.8.1. Alcances	11
1.8.2. Limitaciones	11
1.9. ESTADO DE LA CUESTIÓN	12
1.9.1. Objetivo	12
1.9.2. Problema	12
1.9.3. Pregunta	12
1.9.4. Palabras clave	12
1.9.5. Intervención	13
1.9.6. Población	13

1.9.7. Aplicación.....	13
1.9.8. Selección de recursos.....	14
1.9.9. Verificación de referencias.....	14
1.9.10. Selección de estudios.....	14
1.9.11. Ejecución.....	15
1.9.12. Selección de los resultados iniciales.....	15
1.9.13. Evaluar la calidad de los estudios.....	15
1.9.14. Extraer los datos de los estudios primarios.....	16
1.9.15. Análisis de resultados.....	16
2. CAPÍTULO 2: MARCO TEÓRICO.....	17
2.1. FUENTES PRIMARIAS Y SECUNDARIAS.....	17
2.2. BASE DE DATOS.....	17
2.3. PROCESAMIENTO ANALÍTICO EN LÍNEA (OLAP).....	18
2.4. BASE DE DATOS TABULAR.....	19
2.5. DATA MART.....	19
2.6. ESQUEMA ESTRELLA.....	20
2.6.1. Tablas de hechos.....	20
2.6.2. Tablas de dimensiones.....	20
2.7. BASE DE DATOS COLUMNAR.....	21
2.8. SAP HANA.....	22
2.8.1. Almacenamiento por columnas.....	22
2.8.2. Eliminar índices.....	24
2.8.3. Procesar más datos con menos memoria.....	25
2.8.4. Vistas calculadas.....	25
3. CAPÍTULO 3: MARCO METODOLÓGICO.....	26
3.1. TIPO DE INVESTIGACIÓN.....	26
3.2. ALCANCE DE LA INVESTIGACIÓN.....	26
3.3. ENFOQUE DE LA INVESTIGACIÓN.....	26
3.4. SUJETOS Y FUENTES DE INFORMACIÓN.....	27
3.5. DESCRIPCIÓN DE INSTRUMENTOS.....	27
4. CAPÍTULO 4: ANÁLISIS Y DIAGNÓSTICO DE LA SITUACIÓN ACTUAL.....	29

4.1. PANORAMA INICIAL	29
4.2. ETAPA DE LA RECOPIACIÓN DE INFORMACIÓN.....	30
4.3. ANÁLISIS DE LA SITUACIÓN ACTUAL	30
4.3.1. Identificación de roles y manejo de la información.....	31
4.3.2. Análisis del software utilizado.....	31
4.3.3. Identificación de fuentes de datos.....	32
4.3.4. Identificación de indicadores	32
4.3.5. Expectativas y necesidades de las áreas involucradas	33
4.4. ANÁLISIS FODA DEL PROYECTO	34
4.5. DEFINICIÓN DEL ALCANCE Y LA ESTRATEGIA.....	34
5. CAPÍTULO 5: PROPUESTA DE LA SOLUCIÓN.....	37
5.1. PROPUESTA DE SOLUCIÓN	37
5.2. DISEÑO FÍSICO, ENTORNOS Y HERRAMIENTAS DE IMPLEMENTACIÓN.....	39
5.2.1. Entornos de implementación.....	40
5.2.2. Herramientas para el desarrollo y la implementación.....	43
5.3. INVENTARIO DE OBJETOS, ESTRUCTURAS Y HERRAMIENTAS UTILIZADOS EN LA IMPLEMENTACIÓN DE LA SOLUCIÓN	44
5.3.1. Fuentes de datos primarias.....	44
5.3.2. Fuentes de datos secundarias	46
5.3.3. Base de datos operativa.....	46
5.3.4. Bases de datos analíticas.....	47
5.3.5. Procesos de extracción, transformación y carga.....	49
5.3.3. Herramientas analíticas de BI.....	60
5.3.4 Herramientas de visualización	64
6. CAPÍTULO 6: CONCLUSIONES Y RECOMENDACIONES.....	68
6.1. CONCLUSIONES.....	68
6.2. RECOMENDACIONES	70
REFERENCIAS.....	72
APÉNDICES.....	75
APÉNDICE 1. COMPOSICIÓN DE LA TARJETA.....	75
APÉNDICE 2. DATOS DEL PROCESADOR.....	75

APÉNDICE 3. TRAMAS DE AUTORIZACIÓN	76
APÉNDICE 4. CONJUNTO DE DATOS	76
APÉNDICE 5. RECOPIACIÓN DE INFORMACIÓN	77
APÉNDICE 6. ESTRATEGIA DEFINIDA EN FASES.....	77
APÉNDICE 7. DISTRIBUCIÓN DE TABLAS.....	78
APÉNDICE 8. INVENTARIO DE OBJETOS.....	78
APÉNDICE 9. COMPARACIÓN DE INDICADORES	79
APÉNDICE 10. ACTA DE ENTREGA Y ACEPTACIÓN DE PRODUCTO	80

ÍNDICE DE TABLAS

Tabla 1. Costos estimados.....	3
Tabla 2. Cronograma de la propuesta	7

ÍNDICE DE FIGURAS

Figura 1. Estructura del proyecto.....	6
Figura 2. Arquitectura SAP HANA.....	22
Figura 3. Almacenamiento basado en filas y columnas.....	23
Figura 4. Arquitectura multicore.....	24
Figura 5. Infraestructura propuesta.....	38
Figura 6. Infraestructura propuesta detallada en sus respectivas fases.....	39
Figura 7. Diseño físico.....	40
Figura 8. Entorno de desarrollo.....	41
Figura 9. Entorno de pruebas.....	42
Figura 10. Entorno de producción.....	43
Figura 11. ETL principal.....	50
Figura 12. ETL cargas complementarias.....	51
Figura 13. ETL cargas catálogos.....	52
Figura 14. ETL cargas archivos del procesador.....	53
Figura 15. ETL cargas tablas históricas.....	54
Figura 16. ETL cargas tablas principales.....	55
Figura 17. ETL cargas tablas modelo estrella.....	57
Figura 18. ETL SAP Data Services.....	59
Figura 19. Vistas calculadas SAP HANA.....	60
Figura 20. Modelo de datos tabular.....	61
Figura 21. Modelo de datos multidimensional.....	63
Figura 22. Pivoteo en Excel de modelos tabulares/multidimensionales.....	65
Figura 23. Reporte implementado en Reporting Services.....	66
Figura 24. Tablero de control implementado en la herramienta SAC.....	67

1. CAPÍTULO 1: INTRODUCCIÓN

1.1. Antecedentes del problema

Dado que la mayoría de entidades financieras no cuenta con una plataforma de procesamiento de tarjetas de crédito y débito, esta labor es realizada por una entidad externa conocida como procesadora de tarjetas. Este tipo de relación entre empresas conlleva a la utilización de varias herramientas (internas y externas) que llevan a cabo los controles de las transacciones hechas por los tarjetahabientes.

Por lo tanto, el punto sensible de este proceso se centró en generar mecanismos que permitan la unificación de la información (tanto local como la proporcionada por el proveedor) en una plataforma centralizada, automatizada y actualizada, la cual proporcione eficientemente información que se encuentre disponible para su posterior análisis.

1.2. Justificación

Al fundamentarse en el concepto general de una transacción y sus cuentas, se generó una estructura base que recopilara los datos de los diferentes sistemas involucrados, de tal forma que se homologaran y sincronizaran los datos con el fin de que se integraran en tiempo, orden y forma.

Posteriormente, se consideró la creación de estructuras de datos derivadas que posibilitaran una exploración profunda de la información y, así, la identificación de los principales focos de análisis en temas de tarjetas de crédito y débito.

Aunado a esto, se tomó en cuenta una fase de automatización de procesos que independizará los datos de los procesos manuales, de modo que se logrará la eficiencia en el tiempo de los procesos operativos y la disminución de la posibilidad de un error humano derivado de los procesos manuales.

La herramienta resultante permitió ver y/o analizar información de tarjetas desde diferentes perspectivas requeridas en la organización, además de la posibilidad de integrarla con el depósito de datos corporativo de la entidad.

1.3. Viabilidad

1.3.1. Factibilidad técnica

La empresa financiera auspiciadora del proyecto cuenta con los requisitos en cuanto a *hardware* y *software*. Adicional, las licencias necesarias se encuentran disponibles, por lo cual se determinó que el proyecto era técnicamente factible.

Dado que el cliente cuenta con una herramienta en la nube, el mismo solicitó que los tableros fuesen confeccionados con dicha herramienta.

1.3.2. Factibilidad económica

Los costos para este proyecto se estimaron con base en las horas hombre requeridas para su elaboración, por lo que se realizó un costo estimado que se presenta en la tabla 1.

Tabla 1. *Costos estimados*

Costos estimados					
	RECURSOS DISPONIBLES	TAREAS DEL PROYECTO	HORAS	COSTO	TOTAL TAREA
Análisis	2	Diseño de instrumentos para recolectar los requerimientos	4	\$ 20	\$ 80
	2	Análisis de los requerimientos	20	\$ 20	\$ 400
	2	Documentar los requerimientos establecidos	20	\$ 20	\$ 400
	2	Reuniones con los usuarios expertos en el tema de tarjetas para establecer los requerimientos	20	\$ 20	\$ 400
			Subtotal	64	\$ 80
Diseño	2	Definición de la metodología	3	\$ 20,0	\$ 160,0
	2	Diseño del modelo	160	\$ 20,0	\$ 3 200
	2	Diseño de los procesos de extracción y transformación de datos	40	\$ 20,0	\$ 800
	2	Diseño de la visualización de los datos	40	\$ 20,0	\$ 800
			Subtotal	243	\$ 80
Desarrollo	2	Desarrollo del modelo	40	\$ 20,0	\$ 800
	2	Desarrollo de los procesos	40	\$ 20,0	\$ 800
	2	Desarrollo de la visualización	40	\$ 20,0	\$ 800
			Subtotal	120	\$ 60
Pruebas	2	Probar el modelo desarrollado	40	\$ 20	\$ 800
	2	Prueba de los procesos	40	\$ 20	\$ 800
	2	Revisión de la visualización	40	\$ 20	\$ 800
			Subtotal	120	\$ 60

Documentación	2	Documentación del modelo	40	\$ 20	\$ 800
	2	Documentación de los procesos	40	\$ 20	\$ 800
	2	Documentación de la visualización	40	\$ 20	\$ 800
		Subtotal	120	\$ 60	\$ 2 400
		Total	667	\$ 340	\$ 13 440

Fuente: Elaboración propia

1.3.3. Factibilidad operacional

La entidad financiera cuenta con las fuentes de datos que contienen la información del producto tarjetas, la cual involucra la importación de información estandarizada a nivel de archivos con información de cuentas y transacciones proporcionadas por el procesador, además de los procesos realizados a nivel de *core* bancario; por lo tanto, se contó con los insumos necesarios para la factible operacional.

1.4. Planteamiento del problema

1.4.1. Problema general

¿Cómo construir una infraestructura de datos robusta que permita la administración de la información histórica y actual del producto tarjetas, y que sirva de plataforma para el análisis y la toma de decisión?

1.4.2. Subproblemas del proyecto

- ¿Cómo identificar las fuentes de datos involucradas en el producto tarjetas?
- ¿Cómo definir el flujo correcto para la carga de los datos desde las fuentes iniciales?

- ¿Cómo modelar nuevas estructuras para agrupar, unificar e integrar los datos en los conceptos que se ajusten a la lógica del negocio?
- ¿Cómo desarrollar la secuencia lógica de carga de datos en las estructuras definidas?
- ¿Cómo diseñar esquemas para los repositorios de información?
- ¿Cómo elaborar los indicadores relevantes para la visualización y toma de decisiones?

1.5. Objetivos del proyecto

Existen varias taxonomías para la definición de los objetivos:

- Bloom, 1956.
- Bloom revisada, 2001.
- Marzano y Kendall, 2006.
- Churches, 2008.

En este proyecto, la definición de los objetivos se llevó a cabo basándose en la taxonomía de Bloom original, debido a que se trataron de ordenar los procesos desde el más simple hasta el más complejo.

1.5.1. Objetivo general

Construir una infraestructura de datos robusta que permita la administración de la información histórica y actual del producto tarjetas, y que sirva de plataforma para el análisis y la toma de decisión.

1.5.2. Objetivos específicos

- Identificar las fuentes de datos involucradas en el producto tarjetas.

- Definir el flujo correcto para la carga de los datos desde las fuentes iniciales.
- Modelar nuevas estructuras para agrupar, unificar e integrar los datos en los conceptos que se ajusten a la lógica del negocio.
- Desarrollar la secuencia lógica de carga de datos en las estructuras definidas.
- Diseñar esquemas para los repositorios de información.
- Elaborar los indicadores relevantes para la visualización y toma de decisiones.

1.6. Cronograma del desarrollo de la propuesta

En la figura 1 se expone la estructura del proyecto:

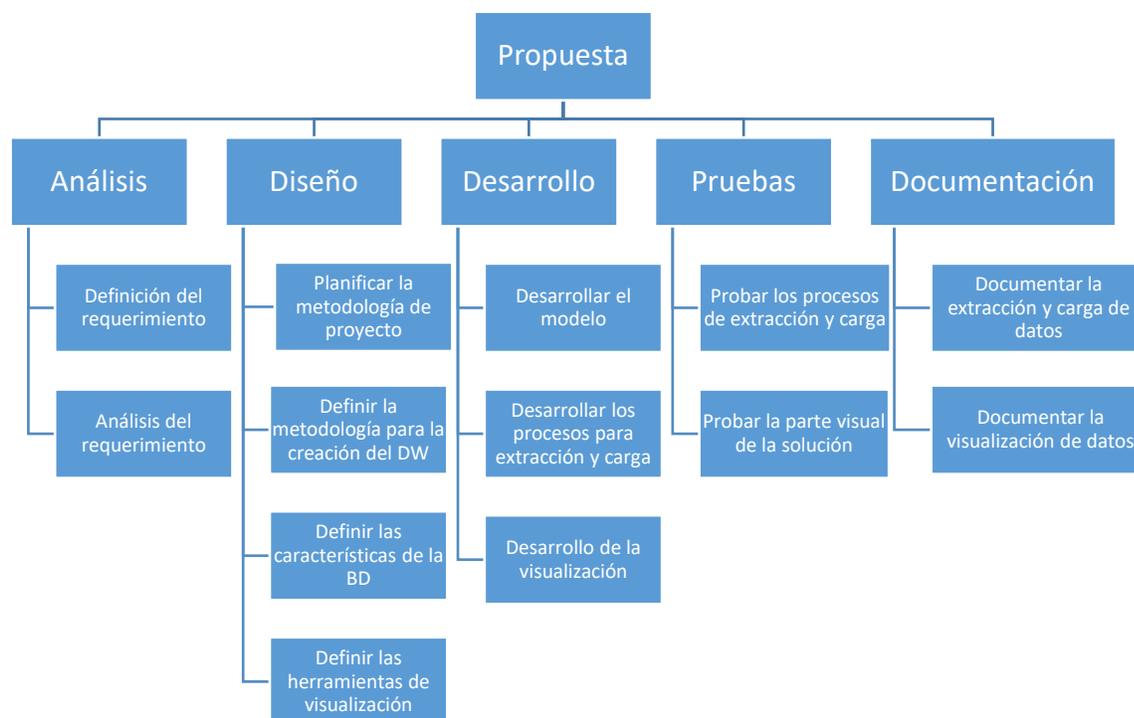


Figura 1. Estructura del proyecto

Fuente: Elaboración propia

Las fechas expuestas en el cronograma utilizan el formato día/mes/año como se aprecia en la tabla 2:

Tabla 2. *Cronograma de la propuesta*

	TAREAS DEL PROYECTO	FECHA INICIO	FECHA FIN
Análisis	Definición de los requerimientos	13/01/2020	15/01/2020
	Análisis de los requerimientos	16/01/2020	28/01/2020
Diseño	Definición de metodología	29/11/2020	17/02/2020
	Diseño del modelo	18/02/2020	06/03/2020
	Diseño de los procesos de extracción y transformación de datos	07/03/2020	27/03/2020
	Diseño de la visualización de los datos	29/03/2020	04/04/2020
Desarrollo	Desarrollo del modelo	05/04/2020	11/04/2020
	Desarrollo de los procesos	12/04/2020	18/04/2020
	Desarrollo de la visualización	19/04/2020	30/04/2020
Pruebas	Probar el modelo desarrollado	01/05/2020	05/05/2020
	Prueba de los procesos	06/05/2020	10/05/2020
Documentación	Revisión de la visualización	11/05/2020	17/05/2020
	Documentar los pasos necesarios para implementar el modelo	18/05/2020	25/05/2020

Fuente: Elaboración propia

1.7. Marco de referencia organizacional y socioeconómico

1.7.1. Reseña histórica

En la página de Coopenae (2019), se brinda la siguiente reseña histórica de la institución:

El 23 de julio de 1966 se fundó la Cooperativa Escolar de la Asociación Nacional de Educadores ‘COOPE-ESCOLAR R.L.’, nombre original de lo que hoy es Coopenae.

Al inicio de sus labores, Coopenae ofrecía diversos servicios para el desarrollo del ejercicio profesional de los educadores, entre estos: poligrafiado, textos, útiles escolares y ayudas visuales, entre otros. Estos servicios los brindaba en un pequeño expendio en el sexto piso de las instalaciones de ANDE; sin embargo, en 1969 se trasladó al Barrio La Dolorosa, a un local que perteneció a la Sociedad de Seguros del Magisterio Nacional.

En 1977, al cambiar su nombre por el de Coopenae, también varió su enfoque, pues ya no solo facilitaba insumos escolares a educadores, sino que abrió sus servicios de crédito a otro grupo importante del gremio de la educación: los educadores pensionados.

Después de este importante paso, en 1983 Coopenae inició un proceso de especialización en los servicios financieros y, luego de evaluar el trabajo realizado, decidió eliminar la actividad comercial que desarrolló desde el inicio. Un año después creó los Servicios de Mutualidad, donde destacó el Socorro Mutuo Solidario y en 1991 creó el FIC, un ahorro a largo plazo con características muy particulares.

Coopenae ingresó al sistema financiero formal en 1993 bajo la supervisión de la Auditoría General de Entidades Financieras (hoy SUGEF), lo cual contribuyó para alcanzar altos estándares de seguridad. Aunado a esto, en 1995 la Administración desarrolló de forma pionera y visionaria, el proyecto de regionalización (extender sus servicios a otras zonas del país).

En el año 2001, Coopenae llegó a consolidarse como la cooperativa de ahorro y crédito líder del país, título que ostenta a la fecha. En 2004, la Cooperativa tomó una de las decisiones más trascendentales pues decidió abrir todos sus servicios al sector público.

Desde ese momento, Coopenae ha ido mejorando y ampliando su portafolio de productos y servicios. En enero de 2008, inició operaciones Coopenae Seguros, como una agencia comercializadora afiliada al Instituto Nacional de Seguros; también, abrió la prestación de servicios a los trabajadores del sector privado que sean afines a la Cooperativa. Hoy, Coopenae es una empresa cooperativa que se ocupa de satisfacer las necesidades de sus asociados por medio de los servicios que ofrece.

1.7.2. Misión

“Somos la empresa cooperativa que representa la mejor opción del mercado para satisfacer necesidades a través de servicios financieros, elevando la calidad de vida de nuestros asociados-clientes” (Coopenae, 2019).

1.7.3. Visión

“Seremos una empresa de servicios financieros líder y modelo en el contexto global, donde la meta suprema es el deleite del asociado en su condición de dueño y cliente” (Coopenae, 2019).

1.7.4. Valores de la organización

Coopenae (2019) indica que sus valores son los siguientes:

- Sentido de propósito (somos-existimos): Cada labor que realizamos, por sencilla que parezca, conlleva un profundo sentido de aportar algo positivamente determinante en la vida de nuestros asociados, compañeros y comunidad. Tenemos claro el propósito

por el cual hacemos lo que hacemos. Ninguna conducta, decisión o acción debe realizarse sin que exista una consecuencia significativa en pro de nuestra misión organizacional. Esto explica por qué nuestros colaboradores se reconocen como parte integral de la Cooperativa.

- **Servicio:** Creemos que servir es la expresión misma de la vocación humana, y como tal se constituye en un deber moral de cada colaborador. Nuestra acción de servir está caracterizada por una penetrante empatía hacia los sentimientos, emociones y necesidades de todos con quienes interactuamos; este es nuestro principal distintivo organizacional.
- **Pasión:** Entendemos la pasión como la búsqueda permanente de la excelencia en cada acción que emprendemos o responsabilidad que asumimos. Aprovechamos las nuevas experiencias y conocimientos no solo para realizar nuestro oficio de manera óptima, sino también como una inspiración para perfeccionar las prácticas organizacionales.
- **Transparencia:** La transparencia es la actitud que se deriva de actuar con integridad y honestidad. Cada asunto de la gestión organizacional será tratado con claridad por los involucrados; de igual forma no habrá dificultad al exponer los resultados de nuestras decisiones y acciones. Creemos en la rendición de cuentas como una acción que evidencia la correcta actuación de Coopenae.
- **Confianza:** Creemos en la confianza como el factor clave para el éxito de toda relación. Queremos crear confianza en todas las relaciones que establezcamos. Entendemos que la confianza se construye desde una base sólida de competencias y carácter de cada colaborador, lo cual sustenta la creación de relaciones de confianza que soportadas en sistemas y procesos configuran la confianza organizacional. Creemos en la velocidad

de la confianza y en la confianza inteligente como impulsores económicos y de la contribución distintiva de la organización.

- Respeto mutuo: Para cumplir con los objetivos organizacionales las relaciones entre los miembros de la empresa serán en un ambiente en el que prime la comprensión y tolerancia a las diferencias e ideas individuales, así como el trato con dignidad a los demás sin distinciones jerárquicas.

1.8. Alcances y limitaciones del proyecto

1.8.1. Alcances

Construir los procesos de recolección de información mediante los cuales se puedan generar las estructuras necesarias para proveer una plataforma que permita el análisis avanzado de los datos de tarjetas de crédito.

1.8.2. Limitaciones

- Los archivos fuente deben poseer una longitud fija y un diccionario de datos donde se especifique la longitud y orden de los datos.
- La revisión de la estructura de cada archivo debe realizarse cada cierto tiempo para validar su integridad, dado que en ocasiones el proveedor de los archivos agrega o quita campos sin previa notificación, lo cual afecta el procesamiento adecuado de los mismos.
- Posterior al proceso de descarga de los archivos, se deben efectuar las revisiones pertinentes para certificar la validez de los mismos.

- Aun cuando en procesos posteriores a la carga de los archivos, los datos son normalizados y clasificados bajo un enfoque lógico y de negocio, el ámbito se limita a los datos obtenidos de la fuente.
- La fecha más actualizada de los datos importados de los archivos siempre corresponderá a los datos procesados el día anterior al día en curso (T-1).

1.9. Estado de la cuestión

1.9.1. Objetivo

Identificar estudios previos que permitan definir los pasos y procesos que se requieren para la implementación de una infraestructura escalable del producto tarjetas.

1.9.2. Problema

En la actualidad Coopenae no cuenta con la información actualizada de los saldos ni movimientos del producto tarjetas, debido a que depende del procesador para que envíe dicha información hasta final de mes. Este hecho provoca que los usuarios no puedan efectuar seguimientos y análisis de la información durante el mes en curso.

1.9.3. Pregunta

¿Existe alguna implementación que permita visualizar los saldos y movimientos de tarjetas de crédito y débito al día anterior (t-1)?

1.9.4. Palabras clave

Se utilizaron las siguientes palabras claves: “infraestructura escalable del producto tarjetas”, “escalable de tarjetas”, “tarjetas ETL”.

1.9.5. Intervención

Al identificar las fuentes de datos necesarias, seguir los pasos correctos e involucrar los procesos relacionados, fue posible crear una infraestructura que proporcione la mayor información actualizada del producto tarjetas. Dicha información puede ser ajustada y complementada para crear *data marts* que pueden integrarse al almacén de datos corporativo, al mismo tiempo que puede ser el insumo para análisis más complejos. El paso final involucró la creación de objetos de visualización que sirvan de herramienta para el análisis continuo y actualizado del producto tarjetas.

1.9.6. Población

La población observada fueron los documentos que explican el orden y la estructura en la que se componen los archivos proporcionados por el procesador, además de los manuales de sistema que explican cómo subsistemas del *core* bancario pueden ser relacionados con las estructuras previas para proveer mayor cantidad de información complementaria.

1.9.7. Aplicación

Las áreas de aplicación que se pueden ver beneficiadas de la investigación son las relacionadas con la administración, control y mejora continua de los servicios brindados por la organización, ya que se pueden determinar, entre otras cosas:

- Medir el porcentaje de ingresos diarios provenientes de la cartera del producto tarjetas (crédito y débito).
- Medir la rentabilidad y crecimiento mensual del producto tarjetas (crédito y débito).
- Evaluar el crecimiento mediante la comparativa interanual y su comportamiento.

- Identificar los comercios de mayor consumo y el comportamiento de compra de los tarjetahabientes
- Realizar alianzas estratégicas con los comercios que puedan elevar el consumo de los tarjetahabientes.

Estos son algunos de los usos que se le puede dar a la investigación si es aplicada.

1.9.8. Selección de recursos

- Fuentes de métodos de búsqueda: se utilizaron motores de búsqueda web.
- Cadenas de búsqueda: se emplearon las siguientes palabras claves: “infraestructura escalable del producto tarjetas”, “escalable de tarjetas”, “tarjetas ETL”.
- Lista de recursos:
 - Google Académico.
 - Documentación oficial de los proveedores de bases de datos Oracle, SQL server y SAP HANA.

1.9.9. Verificación de referencias

El proceso de identificación de las cadenas de búsquedas fue estratégico para optimizar el esfuerzo; por lo tanto, dedicar tiempo y esfuerzo puede significar ahorrar mucho tiempo.

1.9.10. Selección de estudios

Se utilizaron los siguientes criterios para incluir/excluir estudios:

- Incluir/excluir estudios por título.
- Incluir/excluir estudios por palabras clave.

- Incluir/excluir estudios en función de los resúmenes.

1.9.11. Ejecución

Tomando como base las palabras claves definidas anteriormente, se realizó una búsqueda de todas las investigaciones relevantes que consideraran el problema planteado de forma total o parcial y luego se argumentó el factor diferenciador del enfoque formulado para esta investigación.

De igual manera, se consideraron aquellas investigaciones que tuvieron alguna importancia complementaria que pudieran apoyar o no el enfoque planteado para esta investigación, con la finalidad de presentar un argumento sólido en exposición de la propuesta.

1.9.12. Selección de los resultados iniciales

Se realizó una primera iteración para verificar los resultados obtenidos con las cadenas de búsqueda. A nivel científico, no se encontraron resultados que concordaran con el tipo de investigación formulado; sin embargo, a nivel comercial se obtuvo información por parte de las marcas (Visa y MasterCard).

1.9.13. Evaluar la calidad de los estudios

Se validó la credibilidad, relevancia y rigor de los documentos obtenidos en los resultados de búsqueda, en especial la información obtenida de las marcas (Visa y MasterCard). No obstante, al evaluar con detalle los documentos, se observó información de trazabilidad y operatividad transaccional de las tramas de información (estado en que se reciben los datos del procesador de tarjetas).

1.9.14. Extraer los datos de los estudios primarios

Se extrajeron los documentos de definición de los datos iniciales y se complementaron con documentos académicos que justifican el manejo de dichos datos, para así sustentar que esta investigación sigue las mejores prácticas en la creación, diseño e implementación de almacenes de datos corporativos, aun cuando a nivel científico, académico o comercial no se ha publicado ninguna investigación similar.

1.9.15. Análisis de resultados

Utilizando como base los documentos de definición de datos y tramas por parte de las marcas (Visa y MasterCard) y la definición académica de un almacén de datos, basada en un sistema de facturación (modelo encabezado/detalle), se determinaron los siguientes resultados.

Dada la utilización e interpretación de los datos, las estructuras se han creado bajo la definición de un elemento base, denominado número de tarjeta, el cual es considerado único, y bajo su definición propia complementa una estructura jerárquica ascendente de cuatro niveles (número de tarjeta, cuenta, producto, BIN). Alrededor de este elemento base, se han integrado otros datos como saldos, fechas importantes y tarjetahabientes, que permiten una asociación estable y lógica de los datos, proveyendo una línea de información equivalente al modelo encabezado.

Por otro lado, las transacciones derivadas del elemento base (contenidas en una trama), posibilitan detalles propios equivalentes al modelo detalle (con la variante de que el flujo de los datos define dos conjuntos de transacciones diferentes explicados en apartados posteriores).

2. Capítulo 2: Marco teórico

En esta sección se describen las tecnologías que fueron utilizadas a lo largo de proyecto para satisfacer los diferentes requerimientos de almacenamiento, procesamiento y presentación de la información usada para el desarrollo de la infraestructura propuesta.

2.1. Fuentes primarias y secundarias

En cuanto a las fuentes primarias, Raffino (2020) indica: “Las fuentes primarias son aquellas más cercanas posible al evento que se investiga, es decir, con la menor cantidad posible de intermediaciones”.

Por su parte, respecto a las fuentes secundarias, Raffino (2020) señala: “Las fuentes secundarias, en cambio, se basan en las primarias y les dan algún tipo de tratamiento, ya sea sintético, analítico, interpretativo o evaluativo, para proponer a su vez nuevas formas de información”.

2.2. Base de datos

En relación con la base de datos, Silberschatz, Korth y Sudarshan (2002) se refieren al sistema gestor de bases de datos (SGBD), el cual es una colección de datos interrelacionados y programas que permiten acceder a los mismos. Esta colección de datos se conoce por lo general como base de datos y contiene información significativa para una empresa. Además, agregan lo siguiente en cuanto al SGBD:

El objetivo principal de un SGBD es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera que sea tanto práctica como eficiente.

Los sistemas de bases de datos se diseñan para gestionar grandes cantidades de información. La gestión de los datos implica tanto la definición de estructuras para

almacenar la información como la provisión de mecanismos para la manipulación de la información. Además, los sistemas de bases de datos deben proporcionar la fiabilidad de la información almacenada, a pesar de las caídas del sistema o los intentos de acceso sin autorización. Si los datos van a ser compartidos entre diversos usuarios, el sistema debe evitar posibles resultados anómalos (Silberschatz et al., 2002, p. 1).

2.3. Procesamiento analítico en línea (OLAP)

De acuerdo con Microsoft (2018), el OLAP es una tecnología que tiene como fin la organización de grandes bases de datos empresariales y brinda un análisis complejo sin afectar de manera negativa los sistemas transaccionales. Asimismo, explica lo siguiente en cuanto a las bases de procesamiento de transacciones en línea (OLTP) y su relación con el OLAP:

Las bases de datos que utiliza una empresa para almacenar todas sus transacciones y registros se llaman bases de datos de procesamiento de transacciones en línea (OLTP). Normalmente, estas bases de datos tienen registros que se introducen uno cada vez. A menudo contienen una gran cantidad de información de valor para la organización. Sin embargo, las bases de datos que se usan para OLTP no se diseñaron para el análisis. Por lo tanto, obtener respuestas de estas bases de datos es costoso en términos de tiempo y esfuerzo. Los sistemas OLAP se han diseñado para ayudar a extraer de los datos esta información de inteligencia empresarial con un alto rendimiento. Esto se debe a que las bases de datos OLAP se optimizan para cargas de trabajo grandes en lecturas y pequeñas en escrituras (Microsoft, 2018).

2.4. Base de datos tabular

Respecto a la base de datos tabular, Microsoft (2020) detalla lo siguiente en referencia a los modelos tabulares de Analysis Services y su relación con DirectQuery como modo de consulta alternativo cuando los modelos son muy grandes:

Los modelos tabulares de Analysis Services son bases de datos que se ejecutan en memoria o en el modo DirectQuery y se conectan a los datos directamente desde los orígenes de datos relacionales de *back-end*. Mediante el uso de algoritmos de compresión de última generación y procesador de consultas multiproceso, el motor de análisis de Vertipaq Analysis Services ofrece un acceso rápido a los objetos y los datos del modelo tabular mediante la creación de informes de aplicaciones cliente como Power BI y Excel.

Aunque los modelos en memoria son el valor predeterminado, DirectQuery es un modo de consulta alternativo para los modelos que son demasiado grandes para caber en la memoria o cuando la volatilidad de los datos impide una estrategia de procesamiento razonable. DirectQuery consigue paridad con los modelos en memoria gracias a la compatibilidad con una amplia gama de orígenes de datos, la capacidad de controlar las columnas y las tablas calculadas en un modelo DirectQuery, la seguridad de nivel de fila a través de expresiones DAX que llegan a la base de datos *back-end* y la consulta, optimizaciones que producen un rendimiento más rápido.

2.5. Data mart

Imhoff, Galemno y Geiger (2003) brindan las siguientes características del *data mart*:

(...) el *data mart*, es un subconjunto de datos de almacenamiento y es donde la mayoría de las actividades en el entorno de BI tienen lugar. Los datos de cada *data mart*, por lo general

se adapta para una capacidad o función particular, como el análisis de la rentabilidad del producto, análisis de KPI, análisis demográficos de los clientes, y así sucesivamente.

Todas las variedades de *data mart* tienen características universales y únicas. Esto quiere decir que contienen un subconjunto de datos de almacenamiento de datos, que puede ser colocado físicamente con el almacén de datos o en su propia plataforma independiente, y que varían en tamaño desde unos pocos megabytes a varios gigabytes a terabytes.

2.6. Esquema estrella

En cuanto al esquema estrella, Adamson (2010) indica que es un diseño tridimensional para una base de datos relacional, en el cual “Dimensiones relacionadas se agrupan como columnas en tablas de dimensiones, y los hechos se almacenan como columnas en una tabla de hechos”. Además, Adamson (2010) explica que su nombre se debe a que cuando se dibuja con la tabla de hechos, en el centro se forma una especie de estrella o asterisco.

2.6.1. Tablas de hechos

Estas tablas presentan los hechos, pero también incluyen llaves sustitutas que “hacen referencia a cada una de las tablas de dimensiones asociadas” (Adamson, 2010), por ejemplo, la cantidad de órdenes, costo en dólares y en colones, así como productos, vendedores, clientes y fechas de las órdenes de compra.

2.6.2. Tablas de dimensiones

Con relación a las tablas de dimensiones, Adamson (2010) describe las distintas funcionalidades de las mismas:

Las tablas de dimensiones sirven para proporcionar un rico contexto que es necesario para el estudio de los hechos. En consultas e informes, las dimensiones se utilizan para especificar cómo se rodaron hechos a su nivel de agregación. Valores de dimensiones se pueden utilizar para filtrar informes. Ellos se pueden utilizar para proporcionar un contexto para cada medición, por lo general en forma de etiquetas textuales que preceden hechos en cada fila de un informe. También pueden ser utilizados para impulsar las relaciones maestro detalle, totalizando, tabulación cruzada, o clases.

2.7. Base de datos columnar

A continuación, Amazon (2019) explica la base de datos columnar al compararla con una base de datos relacional y vincularla a bases de datos NoSQL:

Mientras una base de datos relacional está optimizada para almacenar filas de datos, normalmente para aplicaciones transaccionales, una base de datos en columnas está optimizada para lograr una recuperación rápida de columnas de datos, normalmente en aplicaciones analíticas. El almacenamiento basado en columnas para las tablas de bases de datos es un factor importante en el desempeño de las consultas analíticas, ya que reduce notablemente los requisitos globales de E/S del disco, y disminuye el volumen de datos que hay que cargar desde él.

De la misma forma que otras bases de datos NoSQL, las bases de datos columnares están diseñadas para reducir la escala utilizando clústeres distribuidos de *hardware* de bajo coste para aumentar el desempeño, de manera que resultan ideales para el almacenamiento de datos y el procesamiento de *Big Data*.

2.8. SAP HANA

De acuerdo con Schmitz (2015):

SAP HANA es una aplicación que utiliza tecnología de base de datos *in-memory* para el procesamiento de grandes cantidades de datos en tiempo real, a gran velocidad. Al no necesitar el acceso a disco para procesar los datos, y tomarlos directamente de la RAM, se acelera considerablemente la obtención de resultados y el análisis de los datos.

Esta única plataforma *in-memory* combina una base de datos que cumple con ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad) con servicios de aplicaciones, herramientas flexibles para la adquisición de datos y analíticas de alta velocidad. Así, SAP HANA almacena y recoge los datos utilizados por las aplicaciones y puede integrar datos de múltiples fuentes con datos transaccionales en tiempo real.

En la figura 2 se puede apreciar la arquitecta de SAP HANA:



Figura 2. Arquitectura SAP HANA
Fuente: SAP, 2017

2.8.1. Almacenamiento por columnas

SAP (2019a) explica el almacenamiento de datos columnar y relacional que lleva a cabo SAP HANA:

Como sistema de gestión de base de datos relacional de propósito general, la plataforma SAP HANA almacena datos en tablas columnares y usa velocidad de procesamiento *in-memory* y tecnología de procesadores *multicore* para acelerar el procesamiento tanto transaccional como analítico.

El almacenamiento columnar aumenta la eficiencia y brinda una variedad de otros beneficios. Entre ellos se incluyen la capacidad de procesar analíticas con mayor rapidez, maximizar la compresión de datos para reducir el tamaño de la base de datos y eliminar los índices.

La organización de los datos en formato columnar ofrece muchas ventajas para el procesamiento analítico, por sobre el enfoque tradicional de almacenar los datos en filas. Mientras que las bases de datos relacionales almacenan datos como tablas bidimensionales que constan de filas y columnas, los datos se organizan *in-memory* y en el disco como secuencias lineales. Tanto una base de datos orientada en filas como una orientada en columnas almacenan los datos en una secuencia de registros.

En la figura 3 se aprecia el almacenamiento en tablas basado en filas y en columnas

Tabla: Ventas

País	Producto	Ventas
EE. UU.	Alpha	3.000
EE. UU.	Beta	1.250
Japón	Alpha	700
RU	Alpha	450

Consulta de ventas totales

Seleccionar suma (ventas) de ventas;

Almacenamiento en filas

Fila 1		Fila 2		Fila 3		Fila 4					
EE. UU.	Alpha	3.000	EE. UU.	Beta	1.250	Japón	Alpha	700	RU	Alpha	450

Almacenamiento columnar

País				Producto				Ventas			
EE. UU.	EE. UU.	Japón	RU	Alpha	Beta	Alpha	Alpha	3.000	1.250	700	450

Figura 3. Almacenamiento basado en filas y columnas

Fuente: SAP, 2019a

Aunado a lo anterior, SAP (2019a) señala:

El almacenamiento columnar también le permite tomar ventaja de la arquitectura de procesadores *multicore* para el procesamiento de datos paralelo. Cada procesador puede explorar una columna o parte de una columna en paralelo, para un uso óptimo del *hardware* y para lograr velocidad

La figura 4 muestra la arquitectura de procesadores *multicore*:

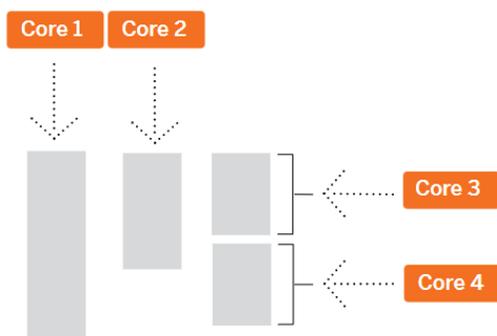


Figura 4. Arquitectura multicore
Fuente: SAP, 2019a

2.8.2. Eliminar índices

En relación con la eliminación de índices en SAP HANA, el texto de SAP (2019a) detalla:
Con SAP HANA, puede acceder a cualquier dato a alta velocidad sin necesidad de que los administradores de bases de datos creen índices.

En el almacenamiento columnar, cada columna funciona como un índice. Esto elimina la necesidad de crear índices, el espacio para almacenar los índices y el procesamiento adicional para actualizar los índices durante el procesamiento transaccional. Sus administradores ahora pueden pasar más tiempo dando soporte a proyectos de innovación y menos tiempo solucionando problemas de rendimiento de las aplicaciones e informes actuales.

2.8.3. Procesar más datos con menos memoria

SAP (2019a) establece en cuanto al procesamiento de más datos con menos memoria en SAP HANA, lo siguiente:

La codificación por diccionario con SAP HANA hace posible convertir valores de *string* en números enteros y procesar más datos usando cachés de CPU de alta velocidad –tales como L1, L2 y L3–. Además, SAP HANA usa múltiples técnicas de compresión tales como codificación prefija, *run-length*, de *cluster*, aislada e indirecta para que aun más datos puedan almacenarse y procesarse *in-memory* y en cachés de CPU de alta velocidad. La plataforma también selecciona automáticamente el algoritmo de compresión correcto en base al contenido de la columna, brindando así la máxima compresión.

2.8.4. Vistas calculadas

Por último, respecto a las vistas calculadas, SAP (2019b) menciona:

Se utiliza una vista calculada para definir sectores más avanzados en los datos en la base de datos SAP HANA. Las vistas calculadas pueden ser simples y reflejar la funcionalidad que se encuentra tanto en las vistas de atributos como en las vistas analíticas. Sin embargo, generalmente se usan cuando el caso de uso comercial requiere una lógica avanzada que no está cubierta en los tipos anteriores de vistas de información.

Por ejemplo, las vistas calculadas pueden tener capas de lógica de cálculo, pueden incluir medidas provenientes de múltiples tablas de origen, pueden incluir lógica SQL avanzada, entre otros. La base de datos de la vista calculada puede incluir cualquier combinación de tablas, vistas de columnas, vistas de atributos y vistas analíticas. Puede crear uniones, uniones, proyecciones y niveles de agregación en las fuentes.

3. CAPÍTULO 3: MARCO METODOLÓGICO

3.1. Tipo de investigación

Este proyecto fue una investigación exploratoria; al respecto, Hernández, Fernández y Baptista (2010) indican que los estudios exploratorios se llevan a cabo cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes. Es decir, cuando la revisión de la literatura reveló que tan solo hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio, o bien, si se desea indagar acerca de temas y áreas desde nuevas perspectivas.

3.2. Alcance de la investigación

Esta investigación tuvo un alcance exploratorio, dado que expone información general respecto a un fenómeno o problema poco conocido, incluyendo la identificación de posibles variables por estudiar en un futuro.

3.3. Enfoque de la investigación

Este proyecto de investigación utilizó un enfoque mixto, pues empleó componentes de los métodos cualitativo y cuantitativo. Hernández et al. (2010) señalan en cuanto a este enfoque:

Los métodos mixtos representan un conjunto de procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y el análisis de datos cuantitativos y cualitativos, así como su integración y discusión conjunta, para realizar inferencias producto de toda la información recabada (metainferencias) y lograr un mayor entendimiento del fenómeno bajo estudio.

3.4. Sujetos y fuentes de información

Las principales fuentes de información utilizadas para este proyecto fueron:

- Documentación oficial de los vendedores de bases de datos SAP, Oracle y Microsoft.
- Documentación brindada por el área de arquitectura de bases de datos.
- Documentos, libros y artículos que brinden apoyo referente a temas del diccionario de datos o almacenes de datos.
- Reuniones con el patrocinador.
- Reuniones con el personal del área de arquitectura de bases de datos.

Por su parte, los sujetos de información utilizados para este proyecto se citan a continuación:

- Involucrados en el proyecto.
- Los distintos motores de bases de datos relacionales que utiliza la Cooperativa (SQL Server, Oracle, SAP HANA).

3.5. Descripción de instrumentos

En esta investigación se utilizaron los siguientes instrumentos:

- Reuniones programadas con los involucrados: este instrumento ayudó a conocer el panorama inicial del proyecto, determinar las pautas para la recolección de la información y establecer su alcance.
- Grupos de discusión: este instrumento involucró a los analistas de nivel avanzado de cada unidad de negocio relacionada al proyecto, con la intención de definir los objetivos idóneos por cumplir con el proyecto.

- Documentación específica y de procesos: estos instrumentos sirvieron para comprender las estructuras de datos relacionadas al proyecto y ayudaron en el planteamiento de la estrategia para la creación de las nuevas estructuras requeridas para satisfacer el proyecto.
- Observación: mediante este instrumento se lograron complementar los conceptos vinculados a la lógica del negocio y su integración.

4. Capítulo 4: Análisis y diagnóstico de la situación actual

Este capítulo retoma el apartado: “Antecedentes del problema y planteamiento del problema”, con la intención de ampliarlo al explicar la manera en que se trató el problema experimentado por la empresa financiera auspiciadora y la forma en que se definieron los pasos para reunir la información suficiente con el fin de definir una ruta, así como un alcance y una estrategia por seguir.

4.1. Panorama inicial

El primer paso consistió en realizar un análisis contextual del problema, con la intención de comprender, evaluar y dimensionar la situación. De este modo, del contexto se conocía lo siguiente:

A nivel organizacional el producto de tarjetas es gestionado por dos unidades de negocio: la Unidad de Medios Electrónicos de Pago (MEP) y el Departamento Comercial. Ambas unidades trabajan bajo diferentes enfoques los datos transaccionales generados del producto. El primero trabaja muy de cerca con el procesador de tarjetas para garantizar el control de la operatividad diaria y el segundo se enfoca en las métricas del producto colocado y las estrategias anuales del crecimiento de la cartera.

La organización utiliza dos herramientas para la administración de los procesos del producto: un sistema de transaccional para gestiones conocido como SAT (Sistema de Administración de Tarjetas) y una herramienta suministrada por el procesador de tarjetas denominada Siscard, que permite administrar y controlar todas las características y funcionalidades del producto.

4.2. Etapa de la recopilación de información

El siguiente paso fue establecer una serie de reuniones con las unidades o departamentos involucrados con el proyecto, con el propósito de conocer sus necesidades y recabar información referente a los siguientes puntos:

- Identificar los roles participantes y conocer acerca del manejo de los datos que realizan.
- Explorar las herramientas utilizadas por los usuarios finales, de tal manera que sea posible identificar las fuentes de datos y cómo estas se relacionan.
- Identificar parámetros de tiempo y efectividad operacional que puedan ser utilizados como indicadores de rendimiento en la solución resultante del proyecto.

4.3. Análisis de la situación actual

Una vez recopilada la información de las reuniones con los involucrados del proyecto de ambas unidades de negocio, esta se clasificó para efectuar el siguiente análisis:

4.3.1. Identificación de roles y manejo de la información

Rol	Manejo de la información
Usuario	Trabaja directamente con los sistemas transaccionales y reportes disponibles. De manera casual utiliza los archivos proporcionados por el procesador de tarjetas.
Analista	Trabaja con ciertos reportes disponibles, pero en su mayoría utiliza los archivos proporcionados por el procesador de tarjetas.
Jefe	Solicita la mayoría de la información requerida a los analistas de área y en ocasiones debe pagar para acceder a reportes personalizados por el procesador de tarjetas. Maneja indicadores manuales para evaluar la productividad.
Gerente	Trabaja con reportes solicitados previamente al procesador de tarjetas, los cuales son de poca utilidad para la toma de decisiones.

4.3.2. Análisis del software utilizado

Como se expuso, se utilizan dos herramientas para la administración del producto de tarjetas, pero ambas no interaccionan entre sí y la herramienta SAT cumple la función de gestionar solicitudes (colocación, reemplazo -por robo o pérdida-, reposición por vencimiento, entre otras), además se debe considerar que la mayoría de las solicitudes terminan en procesos manuales que son efectuados en el sistema Siscard.

Por otra parte, Siscard es una herramienta B2B, en donde el procesador de tarjetas provee su plataforma para la administración de los datos, pero estos pertenecen al proveedor, por ende, no se tiene acceso a los mismos de forma directa; sin embargo, el proveedor facilita una unidad compartida tipo FTP, en la cual se colocan archivos planos con la información transaccional al día anterior (T-1), dichos archivos proporcionan un 78 % de la información transaccional hecha por el procesador.

4.3.3. Identificación de fuentes de datos

Como resultado del análisis del *software* utilizado y de la opinión de los usuarios y expertos de las áreas relacionadas, se identificó que los archivos compartidos por el procesador de tarjetas representan la fuente principal de los datos, mientras los datos del sistema SAT y algunos otros submódulos del sistema transaccional representan fuentes secundarias de datos, que después de algunas consideraciones técnicas podrían integrarse a las fuentes primarias.

4.3.4. Identificación de indicadores

Se identificó que en el caso de la Unidad de Medios Eléctricos de Pago, se cuenta con un sistema que gestiona las solicitudes de datos tanto internas como para otras áreas de la empresa, y registra el tiempo estimado de resolución de una solicitud en días, la cantidad promedio de solicitudes ingresadas por semana, la cantidad de solicitudes resueltas por semana y el tiempo promedio utilizado en resolver una solicitud en días. Estos parámetros conforman el conjunto de indicadores de rendimiento.

En tanto, el tiempo estimado de resolución de todas las solicitudes ingresadas en días proporciona el indicador de rendimiento operacional requerido. Teniendo en cuenta que el 90 % de la información de tarjetas sale de la Unidad de Medios Electrónicos de Pago, estos indicadores fueron la referencia ideal para medir la factibilidad y éxito del proyecto.

4.3.5. Expectativas y necesidades de las áreas involucradas

Rol	Expectativas	Necesidades
Usuario	Contar con reportes adicionales que le ayuden a cumplir con sus funciones diarias.	Automatización de procesos manuales.
Analista	Contar con una herramienta que le permita acceder a los datos integrados de forma centralizada y manejar grandes volúmenes de datos.	Integración de datos de sistemas independientes. Automatización de procesos manuales.
Jefe	Al igual que los analistas, acceder a una herramienta con datos integrados y centralizados.	Reportes de control y de alto nivel.
Gerente	Reportes de control de mando y de tomas de decisión.	Reportes de alto nivel y de control estratégico.

Se detectó que las necesidades a nivel de usuarios y de analistas suelen ser muy similares dado que ambos roles actualmente generan muchos reprocesos y procesos manuales, pero existe claridad acerca de la expectativa de lo que se debe cumplir en cada rol. En tanto, los niveles superiores sí presentan necesidades más acordes con sus expectativas.

4.4. Análisis FODA del proyecto

<p>Fortalezas</p> <p>El producto tarjetas a nivel operacional es muy bien administrado, los mandos medios y altos conocen bien los procesos.</p>	<p>Oportunidades</p> <p>Se requiere de herramientas de análisis y control que permitan trabajar con mayor eficiencia y explorar con mayor profundidad todas las características del producto.</p>
<p>Debilidades</p> <p>El manejo tecnológico no es centralizado, sino que se genera desde cada departamento o unidad de trabajo, adicionalmente el crecimiento a nivel de infraestructura tecnológica es limitado. El soporte es lento y en ocasiones no cuenta con el nivel adecuado para atender emergencias o situaciones de urgencia.</p>	<p>Amenazas</p> <p>Las áreas carecen de herramientas de control, que les permitan prevenir fraudes internos, errores de información o fuga de información.</p>

4.5. Definición del alcance y la estrategia

Alcance: Creación de un depósito de datos enfocado al producto tarjetas (crédito y débito) que permita el análisis óptimo y profundo de todas sus características y posibilite la evolución del mismo según las necesidades lo requieran. El depósito de datos cuenta con dos tipos de módulos (*data marts*): históricos y diarios (T-1), y se implementa bajo dos tecnologías, a saber, tabular y multidimensional.

Estrategia: El proyecto cuenta de 5 fases: Previa, Integración, Modelado, Deposito de Datos y visualización.

Fase 1 - Previa: tomando en cuenta que la fuente primaria corresponde a archivos maestros codificados, se comenzó con importar la información diaria de los archivos a una base de datos. Posteriormente, se crearon los catálogos necesarios para traducir cada uno de los códigos

integrados en la información importada. Luego, se realizaron las tablas para contener la información integrada, clara y limpia, y por último se hicieron copias de los datos en tablas históricas.

De forma paralela, se importarán las fuentes secundarias con el formato más detallado y con una estructura enfocada a extender el concepto de las fuentes primarias.

Fase 2 - Integración: en esta fase se crearon las tablas de uso diario que contemplan las tablas de las fuentes primarias y secundarias ya integradas. Esta fase fue de gran importancia porque es en donde se pueden efectuar las diferentes pruebas de validación de datos a partir de los sistemas transaccionales (Siscard y SAT).

Fase 3 - Modelado: en esta fase se creó una nueva base de datos llamada Staging Area, en la cual se hizo una copia de las tablas realizadas en la fase anterior. En esta nueva base de datos se establecieron los procedimientos necesarios para convertir las tablas ingresadas en un modelo estrella (copo de nieve) que fue el insumo de la siguiente fase. Se hicieron cinco modelos estrella: saldos, autorizaciones, movimientos, financiamientos y solicitudes.

Fase 4 – Depósito de datos: esta fase se compuso de dos partes, la primera consistió en una base de datos con modelo estrella que importó las tablas de la fase anterior y generó el depósito de datos de tarjetas en una base de datos multidimensional, en la cual se llevan los datos diarios e históricos. La segunda parte fue la creación soluciones tabulares para el análisis de segmentos específicos de información a los cuales se les quiera dar seguimiento. Estas soluciones pueden tomar como insumo tablas de la fase dos o tres según se requiera.

Fase 5 – Visualización: esta fase consistió en el uso de herramientas para la presentación dinámica de los datos (tableros de control, infográficos, reportes dinámicos o envío programado de informes), que permitan a los mandos altos conocer de forma general el comportamiento del producto tarjetas. También contempló la habilitación de accesos controlados de datos para el

análisis de los datos mediante herramientas de autoservicio (Excel, SAC o R), lo cual posibilite a los analistas del área evaluar los comportamientos de los datos.

Adicional, a partir del cierre de la fase 2, se realizaron sesiones de capacitación con las personas relacionadas con el proyecto, iniciando con los usuarios y analistas, y posteriormente con los jefes y gerentes. Al finalizar el desarrollo de la fase 4, se comenzó con un plan complementario con los jefes de área, para evaluar los indicadores de rendimiento y establecer el impacto del proyecto y grado de éxito del mismo.

5. Capítulo 5: Propuesta de la solución

5.1. Propuesta de solución

Tomando como punto de partida *el alcance y la estrategia* definidos al final del capítulo anterior, la propuesta consistió en la creación de una infraestructura de datos flexible y robusta que permita la administración óptima de los datos relacionados al producto tarjetas (crédito y débito), posibilitando a los diferentes actores o usuarios acceder de forma centralizada a una plataforma de datos, donde puedan consultar los procesos operativos o realizar análisis más elaborados y específicos, con información histórica o reciente (T-1).

Con la intención de emparejar las tecnologías y los roles de los usuarios, y con esto garantizar los niveles de seguridad, se optó por seguir un enfoque vertical sobre la infraestructura de datos y, de esta manera, visualizarla en dos partes: una parte operativa y una parte analítica.

La parte operativa es una base de datos relacional donde se reunieron todas las fuentes de datos (fuentes primarias y secundarias, identificadas en el capítulo anterior) y se proporcionó una estructura compuesta de todas las características operativas y funcionales del producto tarjetas. Para mejorar los tiempos de respuesta y posibilitar el cruce de datos entre características del producto, las tablas principales se desplegaron por medio de una base de datos tabular y las consultas identificadas como recurrentes fueron implementadas mediante reportes de rápido acceso y con la opción de utilizar filtros dinámicos.

Por otro lado, la parte analítica está compuesta de dos bases de datos relacionales y una base de datos multidimensional. Las bases de datos relacionales se encargaron de la transformación del modelo relacional a un modelo estrella funcional, utilizando la primera base de datos como ejecutor de modelado y habilitando la segunda base de datos como repositorio tanto de las estructuras requeridas como de los datos. Adicional, en esta base de datos se implementaron procedimientos para la creación de métricas que fueron incluidas como indicadores de eficiencia

y control. Por otra parte, la base de datos multidimensional utilizó la segunda base de datos como insumo para generar el *data mart* que se integró al *data warehouse* corporativo.

Finalmente, la herramienta SAC (SAP Analytics Cloud) consumió el *data mart* para la creación de tableros gerenciales y de control.

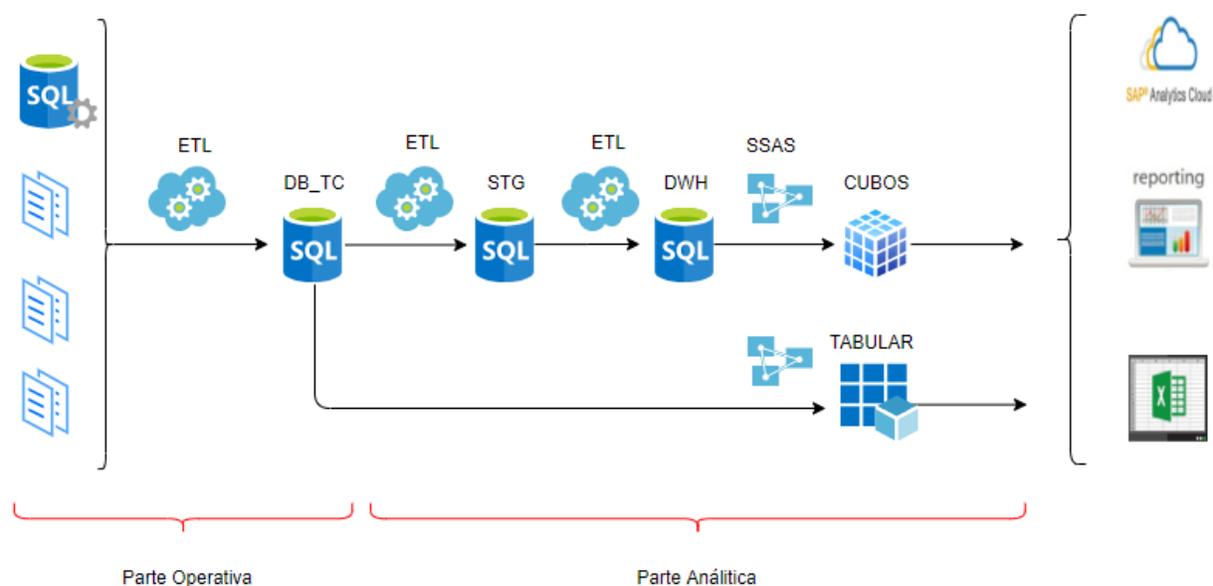


Figura 5. Infraestructura propuesta
Fuente: Elaboración propia

La figura 5 muestra el flujo de los datos por medio de los procesos establecidos, lo que resultó en la consolidación de repositorios de información útiles en las diferentes etapas del proyecto.

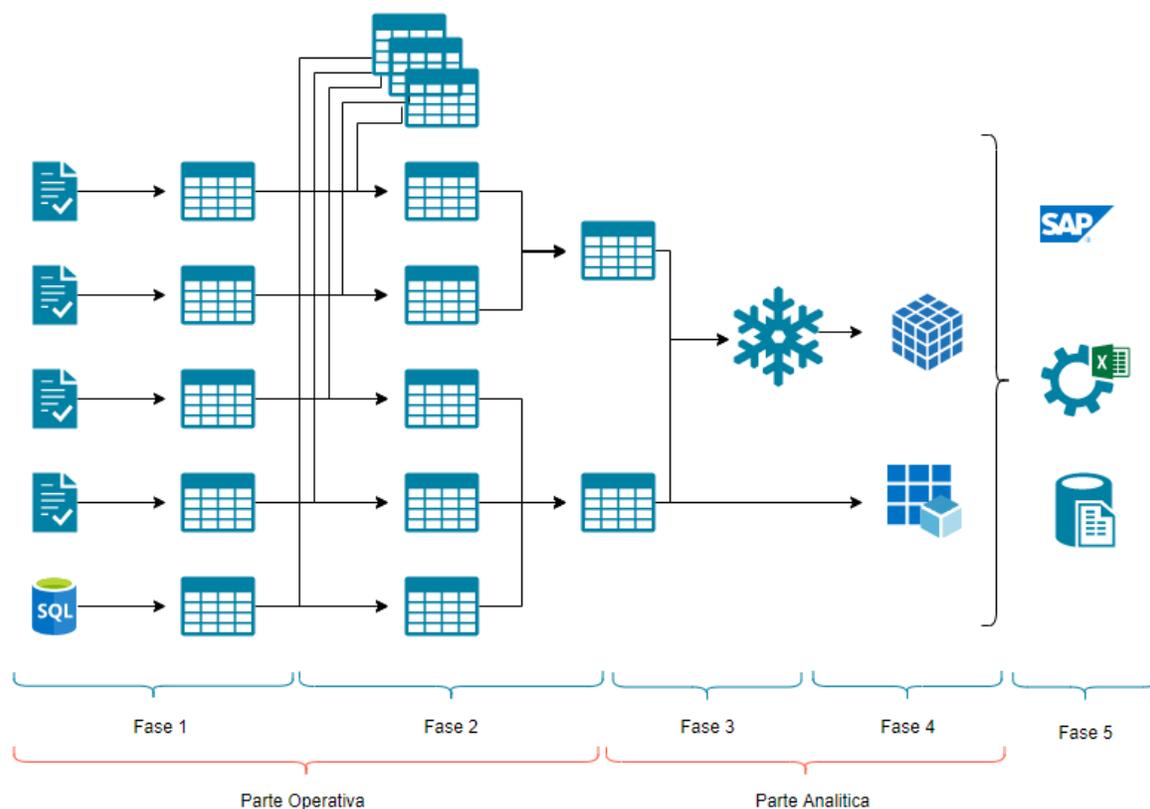


Figura 6. Infraestructura propuesta detallada en sus respectivas fases
Fuente: Elaboración propia

En la figura 6 se aprecia de manera más detallada la infraestructura propuesta en la figura 5, pero desplegada y ajustada a las fases expuestas en la estrategia. Además, establece la separación de la parte operativa y la parte analítica.

5.2. Diseño físico, entornos y herramientas de implementación

Para el adecuado aprovechamiento de los recursos, un óptimo funcionamiento de las herramientas de implementación y un crecimiento sostenido y controlado de la infraestructura, se propuso la utilización de cuatro servidores físicos que son distribuidos de la siguiente manera: un servidor de bases de datos, un servidor para bases de datos analíticas y dos servidores de visualización de datos, haciendo la aclaración de que el servidor para bases de datos analíticas

corresponde al conjunto de servidores dedicados para el *data warehouse* corporativo, el cual consta de dos instancias, una multidimensional y una tabular. También se cuenta con un servidor para reportes dinámicos y esto cubre uno de los servidores de visualización.

De esta manera se añadió un nuevo servidor físico de base de datos al entorno de servidores empresariales y un servicio en la nube para la utilización de un servidor de visualización (en la plataforma SAC de SAP).

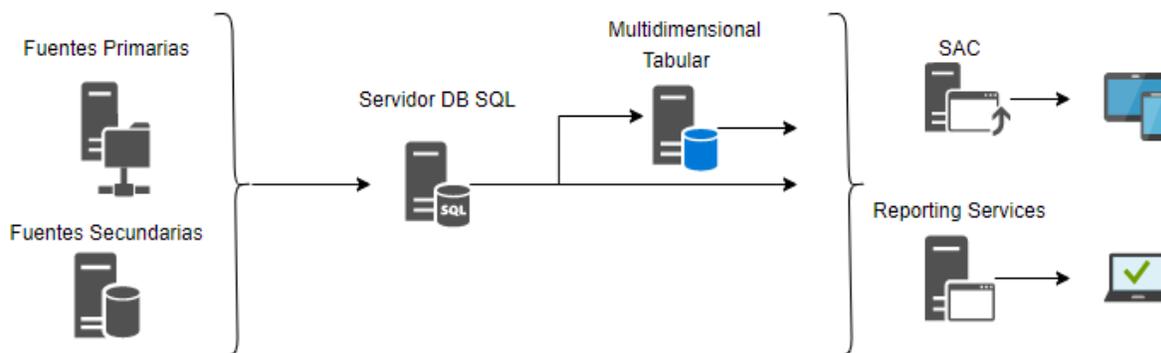


Figura 7. Diseño físico
Fuente: Elaboración propia

5.2.1. Entornos de implementación

Siguiendo los estándares de implementación establecidos por el Departamento de TIC (encargado del soporte de las aplicaciones y *hardware* físico), se definieron tres entornos de implementación de la siguiente forma: un entorno de desarrollo, un entorno de pruebas y un entorno de producción.

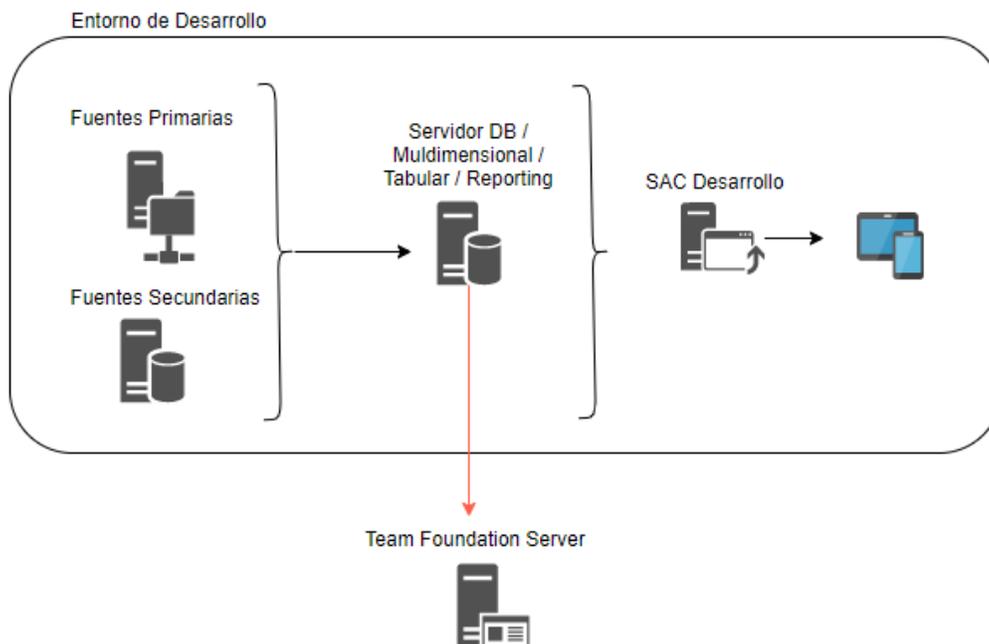


Figura 8. Entorno de desarrollo
Fuente: Elaboración propia

El entorno de desarrollo constó de un servidor que integra todas las herramientas de desarrollo requeridas, es de libre acceso para los desarrolladores del proyecto y está vinculado con el servidor de Team Foundation donde se guardaron las versiones de desarrollo listas para ser probadas o ser desplegadas en los servidores de producción.

A este entorno se le relacionó un servicio en la nube conectado a un repositorio que sirvió de entorno de desarrollo (en este caso, al estar en la nube la plataforma SAC tiene procesos muy sencillos para pasar la implementación entre entornos).

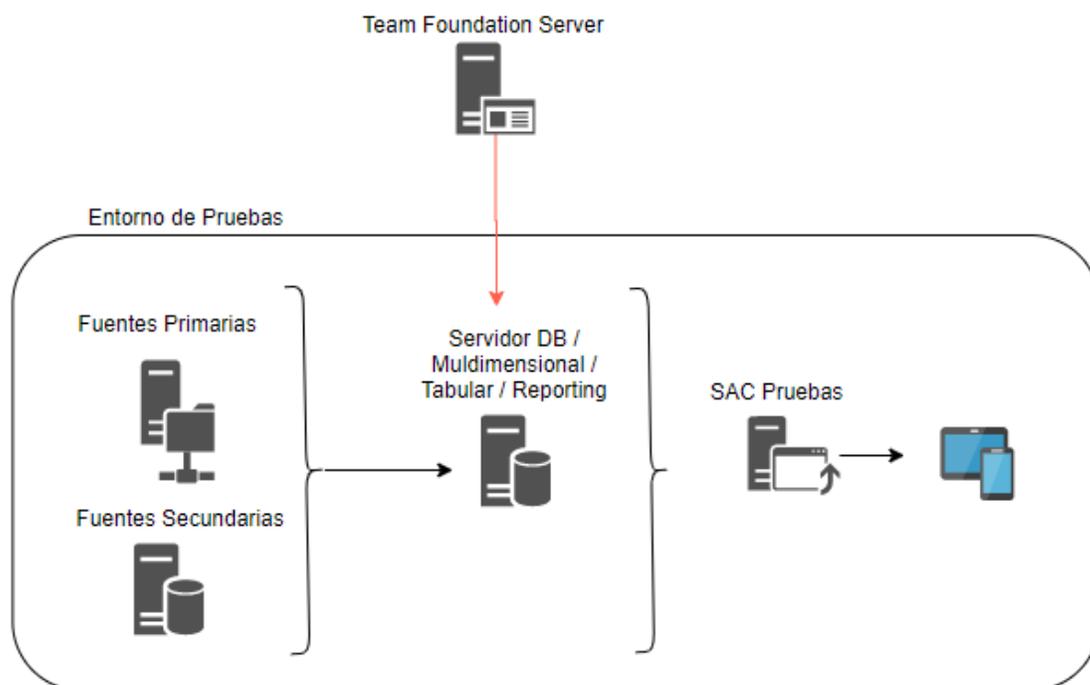


Figura 9. Entorno de pruebas

Fuente: Elaboración propia

El entorno de pruebas (al igual que el entorno de desarrollo) constó de un servidor con características muy similares a los servidores de producción, solamente que las herramientas se usaron de forma diferente, ya que cada desarrollo se probó de manera separada, para garantizar un comportamiento similar al entorno de producción. Los desarrollos se acceden por medio de Team Foundation, el cual es gestionado por el Departamento de TIC.

Ahora bien, existen procedimientos para la ejecución de las pruebas. En un inicio TIC lanza una serie de pruebas de rendimiento para aceptar que el desarrollo no afectará el funcionamiento de todos los elementos y herramientas colocados en los ambientes de producción; luego, se habilita a los desarrolladores el ambiente con un periodo de tiempo establecido para que se puedan ejecutar las pruebas de funcionalidad y respuesta.

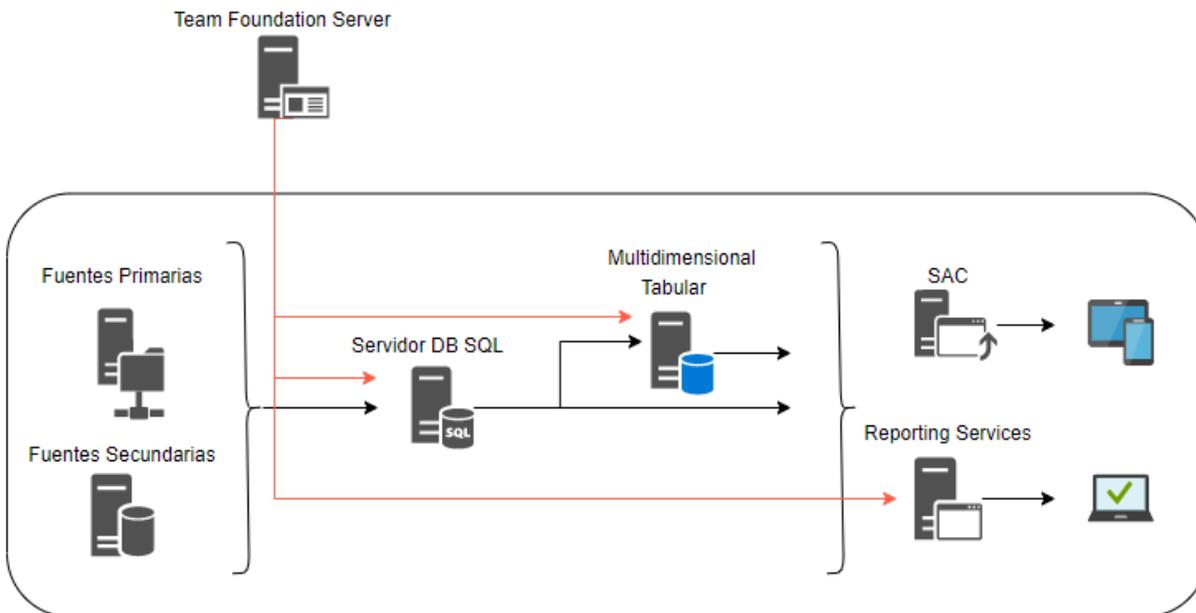


Figura 10. Entorno de producción
Fuente: Elaboración propia

El entorno de producción es únicamente accedido por el Departamento de TIC. Para aplicar los desarrollos, ajustes o cambios, se debe realizar una solicitud de pase a producción, por medio de un sistema de procesos, en donde se indique: el funcionamiento del desarrollo, el servidor donde debe aplicarse el despliegue, el número de paquete, las consideraciones adicionales, además de un *script* paso a paso de la ejecución del paquete. La aplicación de los pases a producción se efectúa luego de todas las aprobaciones y se programa el día de pase, notificando a todas las partes.

5.2.2. Herramientas para el desarrollo y la implementación

Para llevar a cabo este proyecto, se determinó el uso de herramientas disponibles en la organización y, de esta manera, se redujo en gastos excesivos que afectarían la factibilidad y viabilidad del mismo.

Dentro de las herramientas utilizadas en el desarrollo, se encuentran:

- Visual Studio 2015 (para vincular los desarrollos con el Team Foundation Server).

- DataTools 2016 (para la implementación de proyectos de BI, principalmente Integration Service, Analysis Services y Reporting Service).
- SQL Server 2017 Enterprise Edition (para la creación de bases de datos relacionales, además de tres instancias: una para bases de datos multidimensionales, otra para bases de datos tabulares y una instancia más para *reporting services*).

5.3. Inventario de objetos, estructuras y herramientas utilizados en la implementación de la solución

5.3.1. Fuentes de datos primarias

La fuente primaria está compuesta por un conjunto de 17 archivos proporcionados por el procesador de tarjetas mediante un FTP (de forma diaria) y administrado por una herramienta de transferencia de archivos. Los archivos contienen información sobre:

Nombre del archivo	Descripción
Información Básica de Tarjetas	Proporciona información relacionada a la tarjeta.
Información Básica Tarjetas Extendido	Proporciona información detallada y ampliada del archivo de información básica de tarjetas.
Maestro de Cuenta	Proporciona información relacionada a la cuenta de la tarjeta.
Detalle de Cuenta	Proporciona información adicional al archivo de cuentas.
Extensión Tarjeta Tipo	Proporciona información acerca de las características del tipo de tarjeta asignada.
Archivo de Autorizaciones	Proporciona información acerca del proceso de transacciones autorizadas.
Autorizaciones Denegada	Proporciona información acerca de las transacciones fallidas y detalla por qué el fallo.
Disponible para Financiamiento	Proporciona información acerca del crédito adicional disponible para el tarjetahabiente.
Datos Adicionales al Movimiento EF	Proporciona información acerca de los créditos adicionales colocados en los tarjetahabientes.
Archivo de Movimientos	Proporciona información acerca de las transacciones debitadas y contabilizadas.
Archivo de Movimientos Extendido	Proporciona información adicional al archivo de movimientos.
Maestro de Saldos de Premiación	Proporciona información acerca de sistemas de premiación (puntos o millas), según el tipo de tarjeta y su respectivo saldo.
Total de Puntos por Mes del Tarjetahabiente	Proporciona información del balance mensual de los puntos (millas) ganados por el tarjetahabiente.
Información Histórica de Transacciones	Proporciona información complementaria al archivo de autorizaciones.
Información de Localidades y Agencias	Proporciona información acerca de las sucursales o agencias del banco.

Información de Tarjetahabientes	Proporciona información acerca de los clientes que son usuarios de tarjetas.
---------------------------------	--

5.3.2. Fuentes de datos secundarias

Las fuentes de datos secundarias corresponden a tablas dentro del *core* de la Cooperativa que proporcionan información acerca de solicitudes, gestiones, servicios y procesos que son propios de la entidad bancaria y no se reflejan en el sistema facilitado por el procesador. Estas tablas fueron transferidas de forma íntegra a las bases de datos de proyecto en curso.

5.3.3. Base de datos operativa

En este repositorio de datos se almacena la información diaria proveniente de las fuentes primarias y secundarias. Está compuesto por cinco tipos de tablas que cumplen tareas diferentes y de procedimientos que transforman e integran los datos para conformar estructuras robustas de datos lógicamente equilibradas. Dentro de los tipos de tablas, se encuentran:

- **Tablas de migración:** dada la variedad de datos almacenados en los archivos, se procedió a realizar un proceso de extracción y carga, para luego transformar los datos de manera conveniente en cada caso. De este modo, estas tablas representan un repositorio temporal de los datos.
- **Tablas catálogo:** son tablas que proporcionaron definiciones y características a campos codificados en las *tablas de migración*. Así, las *tablas de migración* fueron relacionadas con las *tablas catálogos* y, con esto, se lograron definiciones más interpretativas de los procesos o propiedades de los campos, lo que mejoró el entendimiento de los datos.
- **Tablas de integración:** estas tablas integraron la información de varias *tablas de migración* comunes de forma normalizada, además de aplicar en el proceso la integración

de las *tablas catálogo*, lo que dio como resultado tablas integrales, robustas y completas. Un punto importante por destacar es que en este tipo de tablas se integraron las tablas importadas de las fuentes de datos secundarias y es donde los datos de ambas fuentes de datos (primaria y secundaria) se integraron.

- **Tablas históricas:** almacenaron la información histórica (diaria) generada en las tablas de integración y fueron de gran utilidad para consultas o comparaciones entre eventos de dos distintos periodos de tiempo.
- **Tablas principales:** estas tablas corresponden a las *tablas de integración* potencializadas con cálculos, indicadores y características adicionales (proporcionadas y sugeridas en su mayoría por los usuarios y analistas de los sistemas). Estas tablas fueron la base para la creación de bases de datos tabulares, asimismo fueron el insumo de las bases de datos analíticas para crear los modelos estrella/copo de nieve, que se utilizaron en la creación de las bases de datos multidimensionales.

5.3.4. Bases de datos analíticas

Se componen de dos repositorios de datos. *El primer repositorio* tomó como insumo las *tablas principales* de la *base de datos operativa* y para cada una se efectuaron dos etapas de procesos. En la primera etapa se aplicaron procedimientos para identificar y extraer tipos de datos que pudieran ser considerados característicos, se agruparon las características similares y a partir esto se generaron *tablas tipo dimensión*.

La segunda etapa comenzó con la creación de una copia de las *tablas principales* y fueron llamadas *tablas hechos*; después, mediante una serie de procedimientos para cada *tabla hechos*, se reemplazó la descripción de las características por los códigos de las *tablas tipo dimensión*.

Para el alcance de este proyecto, se consideraron cinco tipos de tablas principales con la finalidad de ser utilizadas en esta sección:

Información _Tarjetahabientes	Se compone de información detallada de los clientes que poseen tarjetas (crédito y/o débito) con la institución.
Información _Saldo _Tarjetas	Se compone de información detallada del balance de la tarjeta, además de integrar fechas de los últimos procesos, lo cual permite tener un conocimiento general del comportamiento del uso de la tarjeta.
Información _Autorizaciones	Se compone de la información detallada de la trama transaccional generada por el uso de la tarjeta, en cualquier tipo de establecimiento o negocio.
Información _Movimientos	Se compone de la información detallada de las transacciones validadas y contabilizadas por el procesador de tarjetas, además de integrar procesos internos de la administración de la cuenta relacionada a la tarjeta.
Información _Financiamiento	Se compone de información detallada de las líneas de crédito adicionales aprobadas a los tarjetahabientes.

Se identificaron cuatro tablas principales que se utilizaron para el análisis y la creación de las bases de datos tabulares (este tema es ampliado en la sección 5.3.6).

En el segundo repositorio disponible se tomaron como insumo las tablas hechos y las tablas tipo dimensión, con las que después de algunos ajustes (que contemplaron el incremento o modificación de identificadores o la creación de jerarquías), se procesó la relación de las tablas para la generación de cada modelo estrella/copo de nieve que dio paso a las estructuras que conforman cada data mart y, luego, fueron integradas al data warehouse corporativo.

5.3.5. Procesos de extracción, transformación y carga

Para lograr un proceso ordenado y eficiente que permitiera múltiples movimientos seguros entre repositorios, con transformaciones de datos controladas, con tiempos de respuesta aceptables y con facilidad de uso y control, se optó por la utilización de la herramienta Integration Services (proporcionada en el kit de soluciones de Data Tools de Microsoft).

De esta manera, se consiguió que la mayor parte de procesos siguiera un flujo definido y automático para la extracción, transformación y carga de los datos, desde las fuentes primarias y secundarias, pasando cada uno de los tres repositorios de bases de datos (expuestos con anterioridad) y quedando integrados en las estructuras estratégicamente creadas (*data warehouse* corporativo) para el análisis de alto nivel que proporcione la información más precisa y confiable y que dé el sustento argumentativo necesario para todas las tomas de decisión acerca de la operativa del negocio.

A continuación, se detalla cada uno de los siete paquetes de ETL creados. Al respecto, se explica el orden de su creación y su posterior reordenamiento con el objetivo de mostrar el flujo lógicamente secuencial que posibilita la independencia individual de cada ETL, aunque unidas todas las partes conforman un ciclo integral.

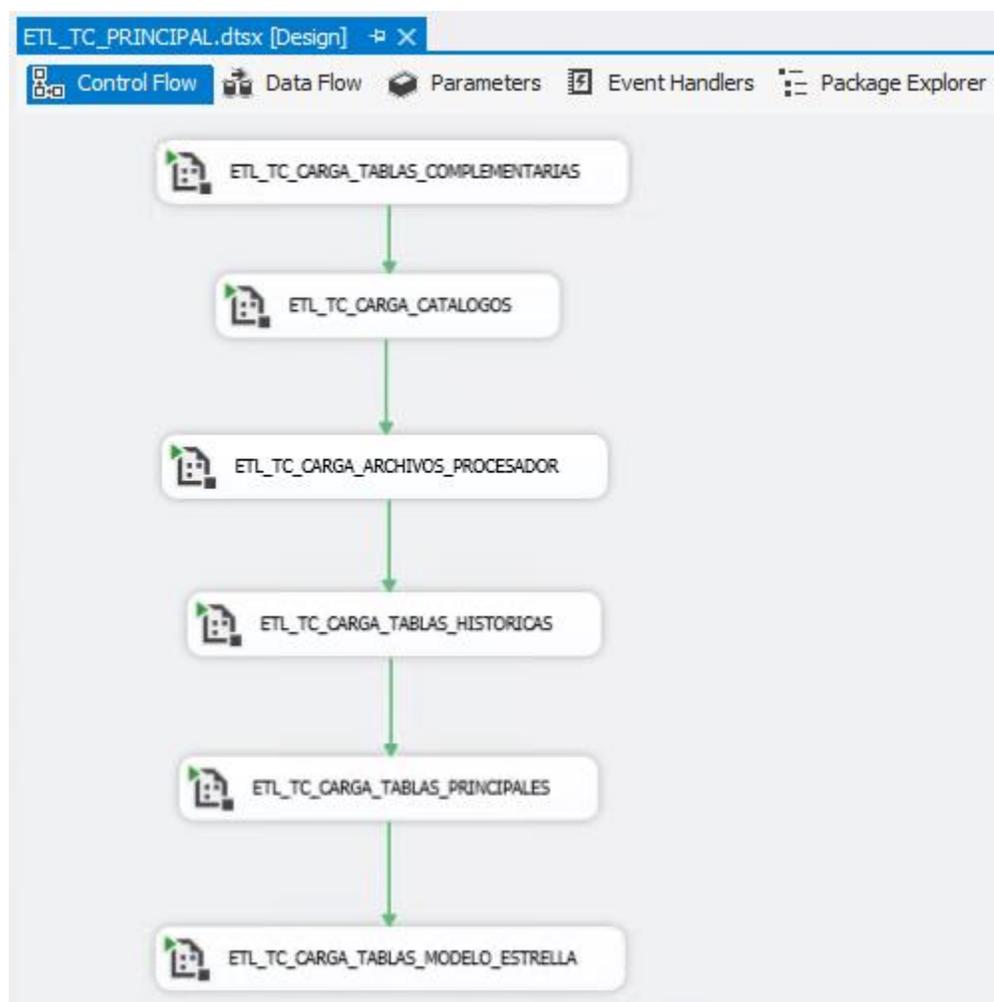


Figura 11. ETL principal
Fuente: Elaboración propia

Aunque es el primer paquete ETL en ser ejecutado, fue el último en ser creado, debido a que es un paquete de control que une a los demás paquetes y efectúa la ejecución secuencial de cada uno. La independencia de cada paquete permite que si alguno de los pasos falla, no sea necesario comenzar desde el principio, sino que se puede resolver el error y continuar desde el punto donde se falló.

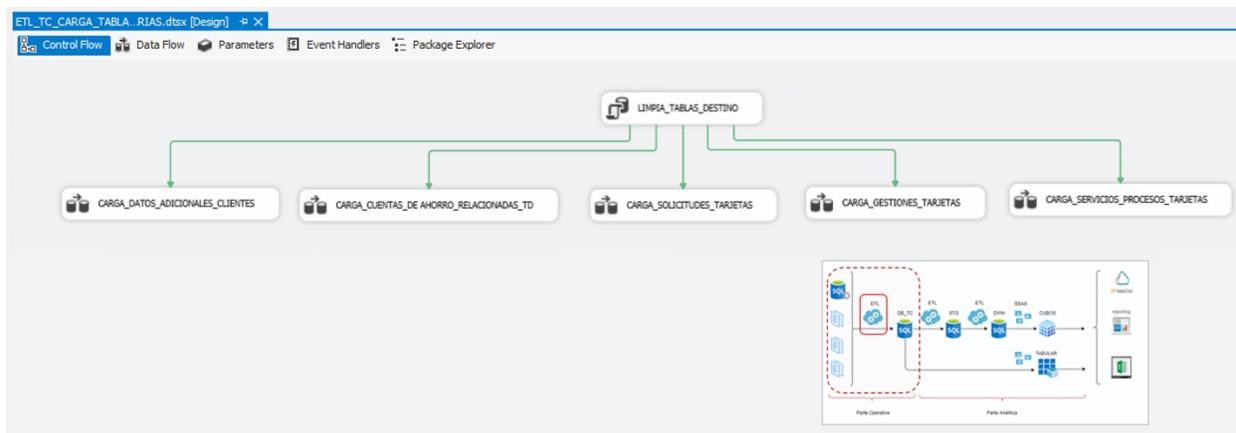


Figura 12. ETL cargas complementarias
Fuente: Elaboración propia

Este ETL fue el segundo en ser creado y se mantiene como el segundo proceso en ser ejecutado. Al respecto, la jerarquía de ejecución se basó en los siguientes argumentos: la importación de estos datos no afecta ningún proceso anterior, la complejidad del paquete es baja y se requiere que la importación se efectúe en momentos en que el *core* bancario esté liberado de carga.

El objetivo de este paquete se enfoca en importar los datos de las *fuentes secundarias* a *tablas de migración* y como se muestra en la imagen adjunta de la *infraestructura propuesta*, se encuentra en la primera etapa de la parte operativa.

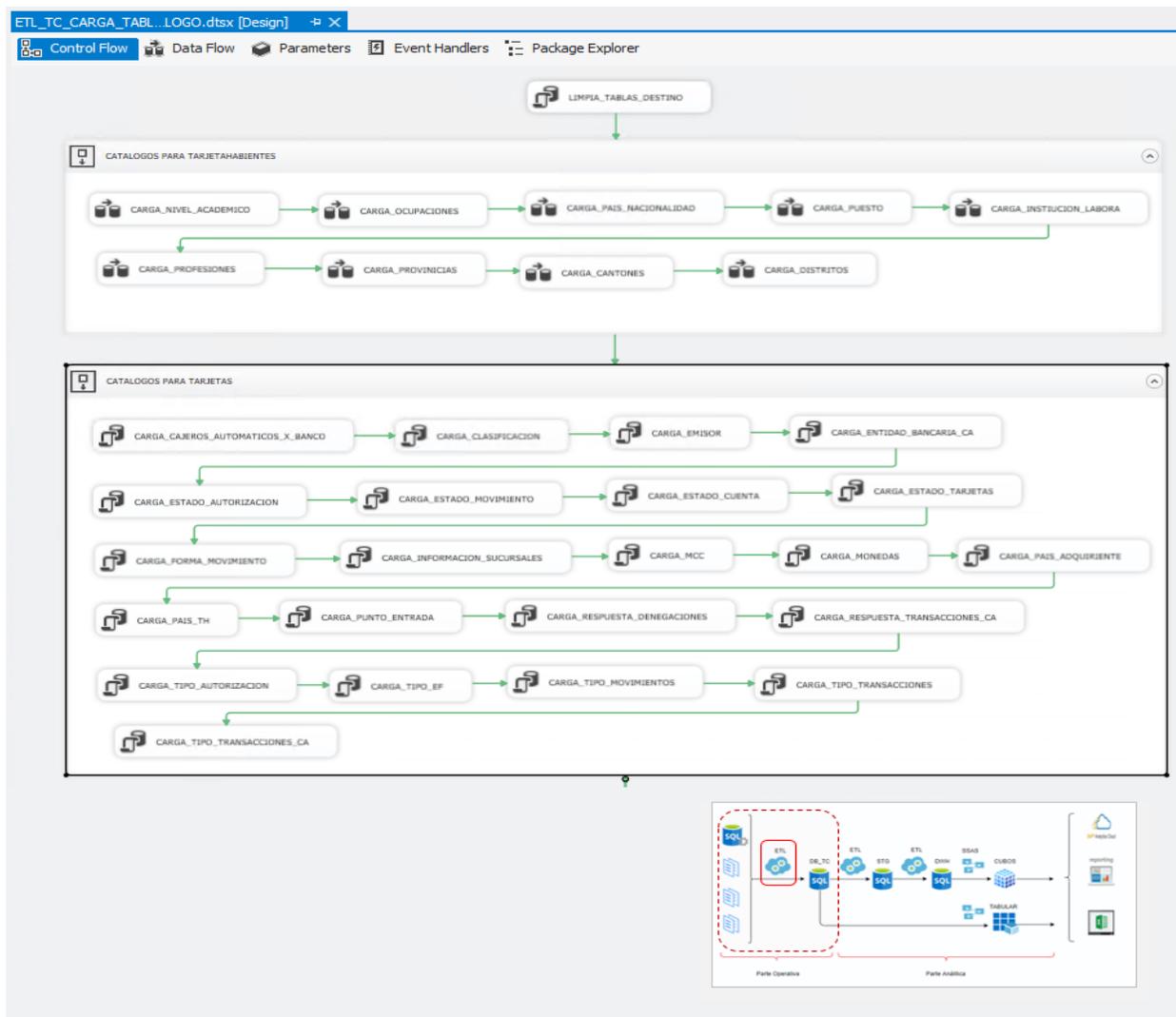


Figura 13. ETL cargas catálogos
Fuente: Elaboración propia

Este ETL sigue la misma secuencia del anterior, al haber sido el tercero en ser creado y ser el tercero en la línea de ejecución. En cuanto a esto, se tomaron los mismos argumentos para ubicarlo en la jerarquía de ejecución: no interfiere con ningún proceso anterior, complejidad baja y distintas fuentes de origen con conjuntos de datos pequeños.

El objetivo de este paquete se enfoca en importar datos descriptivos que complementen los datos de las fuentes de datos primarias y secundarias. Según la *infraestructura propuesta*, y como se muestra en la imagen adjunta, este proceso se encuentra en la primera etapa de la parte operativa.

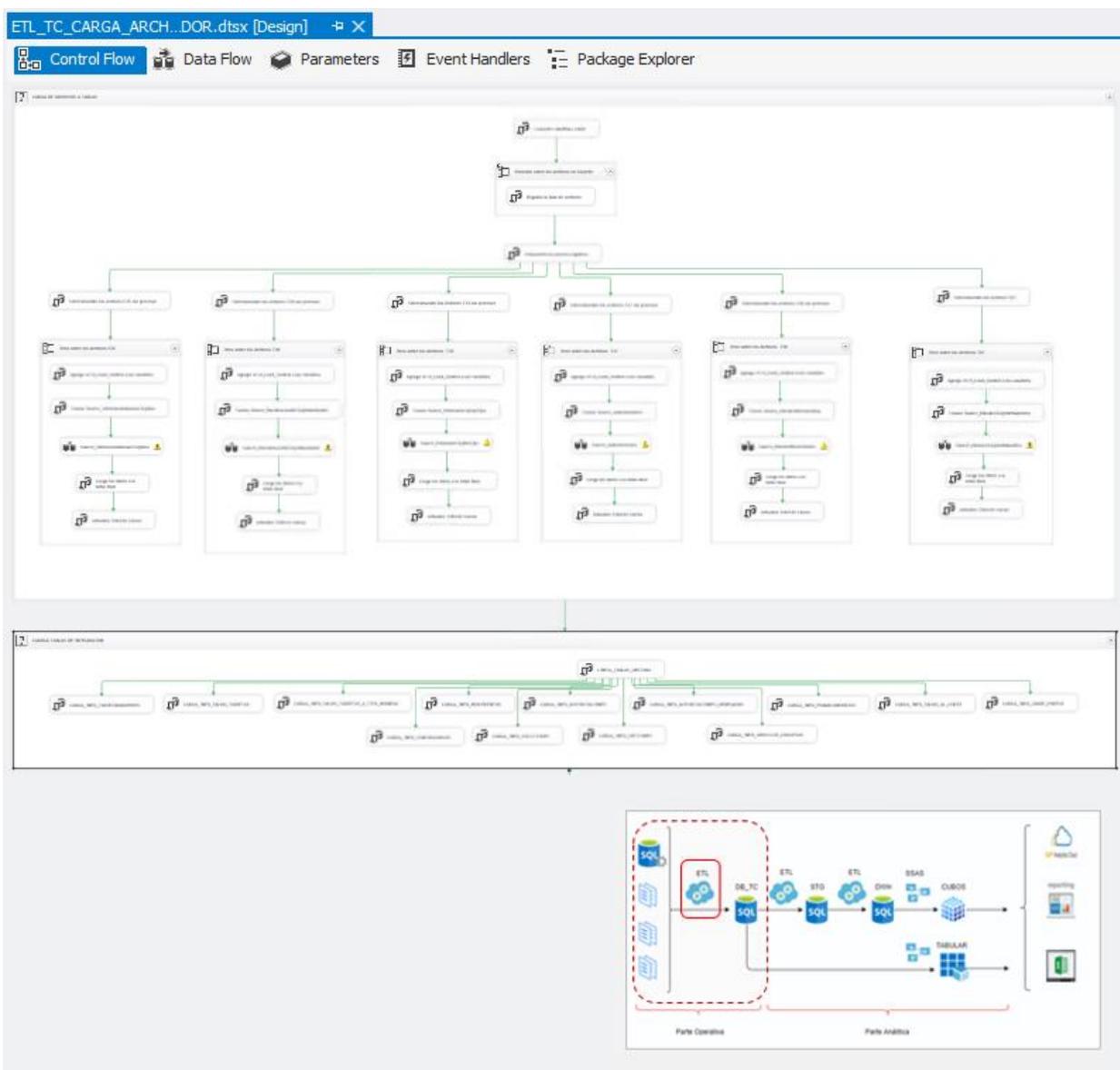


Figura 14. ETL cargas archivos del procesador
Fuente: Elaboración propia

Este ETL fue el primer proceso en ser creado, pero el cuarto en línea de ejecución. Siguiendo los mismos argumentos para ubicarlo en la jerarquía de ejecución, se concluyó que requería previamente de la carga de las *tablas catálogo*, la cantidad de archivos por cargar afectaba el nivel de complejidad y lo ubica en medio-alto y aunque no tenía ninguna otra dependencia, el tiempo de carga era significativamente mayor al de los dos ETL anteriores.

El objetivo de este paquete se enfoca en importar los datos de la *fuentes primaria* (archivos planos) y mediante una serie de procedimientos, cargar los datos en las *tablas de migración* de forma paralela. De modo secuencial, se efectúan procedimientos que realizan algunas transformaciones requeridas para completar algunos campos que almacenan datos calculados o datos derivados.

Finalmente, un último procedimiento secuencial relaciona las *tablas de migración* con las *tablas catálogo* para generar las *tablas de integración* que representan el resultado del proceso. Según la *infraestructura propuesta*, y como se muestra en la imagen adjunta, este proceso se encuentra en la primera etapa de la parte operativa.

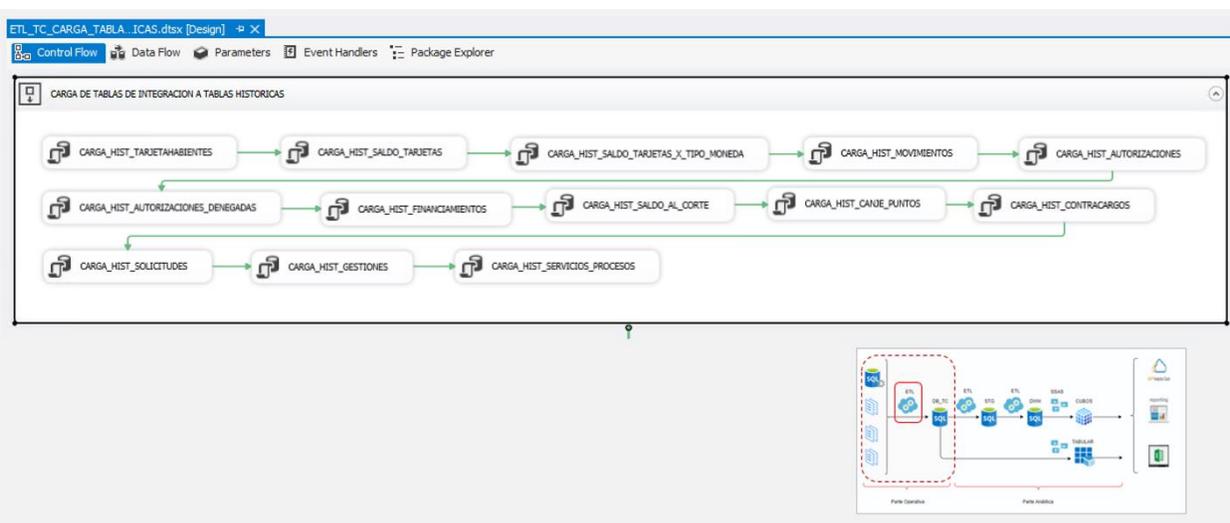


Figura 15. ETL cargas tablas históricas
Fuente: Elaboración propia

En un principio este ETL se ubicaba dentro del ETL anterior, sin embargo, implicaba mucho trabajo cuando se encontraban fallos de ejecución porque alguno de los archivos originales cambiaba su estructura o se realizaban actualizaciones en el *core* bancario y esto afectaba las tablas migradas. Lo anterior implicó hacer el ejercicio de evaluación de jerarquía de ejecución (requiere un proceso anterior, la complejidad es baja y el resultado son tablas importantes que son de alta

El objetivo de este paquete se enfoca en cohesionar *tablas de integración* para lograr tablas con la mayor cantidad de atributos comunes posible, además de permitir que las demás tablas generadas puedan relacionarse entre sí. Esto hace que las tablas agrupen conceptos, o sea, que puedan separarse con solo seleccionar algunos campos de la tabla, o manejarlos como un todo al tomar en cuenta todos los campos.

Esta particularidad permite que dichas tablas sean el insumo ideal para las bases de datos tabulares, al igual que la base óptima para generar modelos estrella/copo de nieve para ser usados en la parte analítica del proyecto. Como se observa en la imagen adjunta de la *infraestructura propuesta*, este proceso forma parte del segundo de los procesos de ETL (primero en la parte analítica).



Figura 17. ETL cargas tablas modelo estrella
Fuente: Elaboración propia

Este ETL fue el quinto en ser creado, pero el séptimo y último en orden de ejecución, de acuerdo con los argumentos para ubicarlo en la jerarquía de ejecución: requiere de las *tablas de principales* creadas en el ETL; el nivel de complejidad es alto, pues hace una separación de los campos entre medias y atributos; y el resultado es la base para los procesos analíticos.

El objetivo de este paquete se enfoca en crear la base de tablas que son utilizadas en la herramienta Analysis Services para crear los *data marts* correspondientes. Como se puede apreciar

en la imagen adjunta de la *infraestructura propuesta*, este proceso es el segundo conjunto de ETL (segundo en la parte analítica) y con esto se cierra el conjunto de procesos automáticos de ejecución diaria.

Según las imágenes de ETL, el kit de soluciones Data Tools presenta el entorno para la configuración de objetos y creación de procesos, pero al igual que toda herramienta de desarrollo requiere de un proceso de depuración (efectuado por el proceso Build) y un proceso de despliegue (realizado por el proceso Deploy), este último crea y ubica la solución ejecutable dentro del motor de SQL Server (previamente elegido) y desde el motor se procede con sus correspondientes configuraciones (determinación de parámetros, cronograma de ejecución, monitoreo y control de la solución). De esta forma, la solución queda automatizada y controlada por el mismo motor de bases de datos.

Existe adicionalmente otro proceso de ETL, pero que se ejecuta de manera independiente y bajo otra tecnología, utilizando la herramienta Data Integration de SAP. Este ETL se agregó en la última etapa del proyecto para extraer de la base de datos DWH, las *tablas del modelo estrella* que sirven de insumo para la fase de visualización (dado que la herramienta se encuentra en la nube, requiere importar los datos desde la ubicación de la herramienta en sitio, en este caso SQL Server). El proceso efectúa algunos ajustes de transformación para consumir y almacenar los datos en su base de datos columnar.

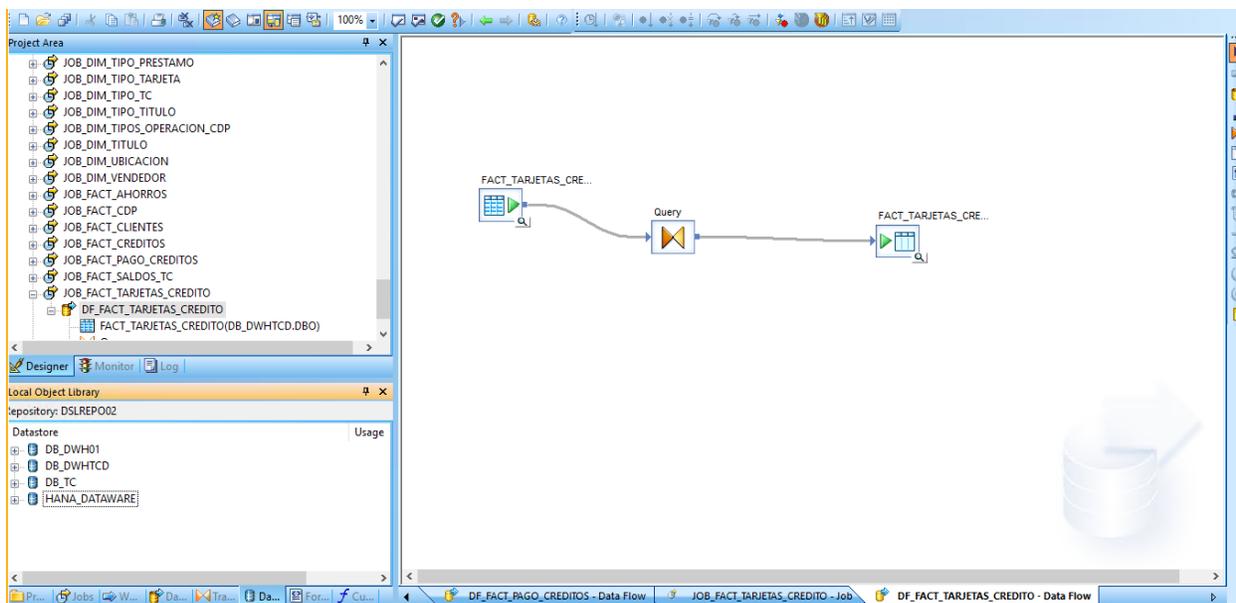


Figura 18. ETL SAP Data Services

Fuente: Elaboración propia

Como se aprecia en la imagen anterior, los datos son cargados a la base de datos SAP HANA, y al igual que los procesos de SSAS, se realizan cargas diarias de la información para ser consumidos en la herramienta de visualización SAC. De esta manera, se utiliza una herramienta que sacara el mayor provecho de los datos para proveer tableros de control e informes dinámicos de gran calidad y eficiencia.

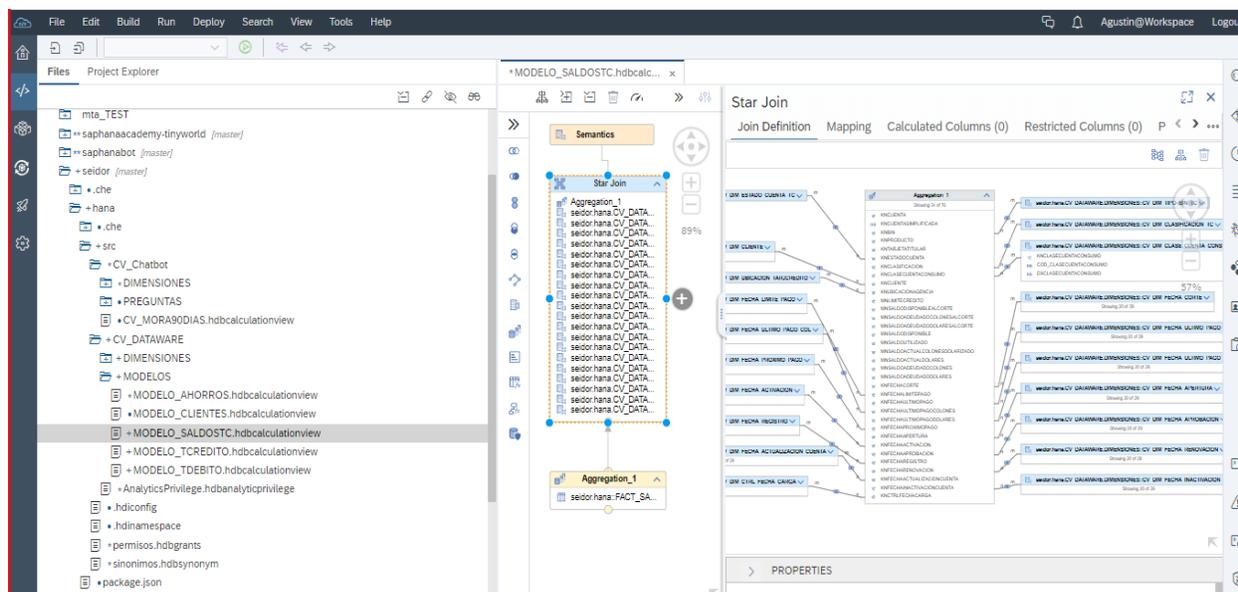


Figura 19. Vistas calculadas SAP HANA

Fuente: Elaboración propia

Al igual que SSIS, se hizo un paso intermedio para que SAC pueda visualizar la información, este proceso consistió en la creación de vistas calculadas para las dimensiones que fueron utilizadas en los modelos de saldos de tarjetas e información de tarjetas de crédito y de débito. Como se muestra en la figura anterior, se mantuvo el mismo concepto de diagrama estrella para el modelado de los datos, brindando las medidas y dimensiones respectivas.

5.3.3. Herramientas analíticas de BI

Además de la herramienta Integration Services, el kit de soluciones de Data Tools de Microsoft cuenta con una herramienta para el tratamiento analítico de los datos, tomando como insumo bases de datos almacenadas de forma relacional; dicha herramienta se llama Analysis Services.

De los tres tipos de soluciones disponibles por esta herramienta, se implementaron dos tipos para el presente proyecto: la primera fue una solución con un *modelo de datos tabular* y la

segunda un *modelo de datos multidimensional*. Ambas soluciones utilizan motores de datos distintos por lo que se requiere contar con dos instancias diferentes: una instancia para bases de datos tabulares (con un motor que administra las tablas en memoria) y una instancia para bases de datos multidimensionales (con un motor analítico que crea conjuntos de datos conocidos como *cubos* a partir de segmentos de la *base de datos multidimensional* denominados *data marts*).

Num_Tarjeta	Bin	Producto	Cuenta	Cuenta_Simplificada	Tipo_Tarjeta_Emisor	Emisor	Cod_Tipo_Tarjeta	Tipo_Tarjeta	Tarjeta_Titular	Estado
4345270014040005	434527	43452700	4345270...	14040	TD	VISA CLASICA WINK	T	TITULAR	4345270014040005	ACTIV
4345270014018001	434527	43452700	4345270...	14018	TD	VISA CLASICA WINK	T	TITULAR	4345270014018001	ACTIV
4345270014021005	434527	43452700	4345270...	14021	TD	VISA CLASICA WINK	T	TITULAR	4345270014021005	ACTIV
4345270014031004	434527	43452700	4345270...	14031	TD	VISA CLASICA WINK	T	TITULAR	4345270014031004	ACTIV
4345270005703009	434527	43452700	4345270...	05703	TD	VISA CLASICA WINK	T	TITULAR	4345270005703009	ACTIV
4345270005835009	434527	43452700	4345270...	05835	TD	VISA CLASICA WINK	T	TITULAR	4345270005835009	ACTIV
4345270006068006	434527	43452700	4345270...	06068	TD	VISA CLASICA WINK	T	TITULAR	4345270006068006	ACTIV
4345270006134006	434527	43452700	4345270...	06134	TD	VISA CLASICA WINK	T	TITULAR	4345270006134006	ACTIV
4345270006187004	434527	43452700	4345270...	06187	TD	VISA CLASICA WINK	T	TITULAR	4345270006187004	ACTIV
4345270005704007	434527	43452700	4345270...	05704	TD	VISA CLASICA WINK	T	TITULAR	4345270005704007	ACTIV
4345270006382001	434527	43452700	4345270...	06382	TD	VISA CLASICA WINK	T	TITULAR	4345270006382001	ACTIV
4345270005792002	434527	43452700	4345270...	05792	TD	VISA CLASICA WINK	T	TITULAR	4345270005792002	ACTIV
4345270006429000	434527	43452700	4345270...	06429	TD	VISA CLASICA WINK	T	TITULAR	4345270006429000	ACTIV
4345270006457001	434527	43452700	4345270...	06457	TD	VISA CLASICA WINK	T	TITULAR	4345270006457001	ACTIV
4345270005911008	434527	43452700	4345270...	05911	TD	VISA CLASICA WINK	T	TITULAR	4345270005911008	ACTIV

Count_TC: 248700
Count_TC_Unicos: 248577

Saldo_Tarjetas | Movimientos | Autorizaciones | Financiamientos | Autorizaciones_Denegadas | Distribucion_Transacciones_Mes_Anterior | Saldo_Tarjetas_Extendido

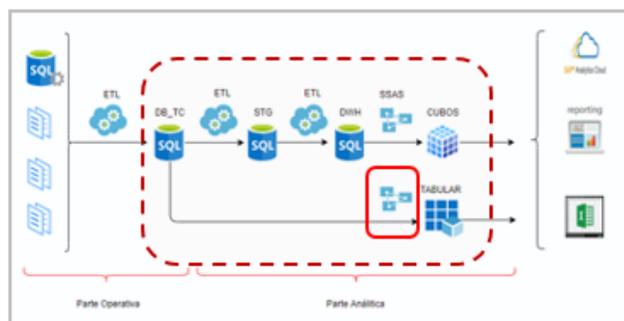


Figura 20. Modelo de datos tabular
Fuente: Elaboración propia

El *modelo tabular* presenta una gran ventaja para el análisis de datos de manera exploratoria, al posibilitar la creación de una estructura intermedia entre una base de datos relacional y una multidimensional; es decir, permite la interrelación entre las tablas del modelo y la construcción de métricas que hacen la labor de los componentes de una tabla de hechos, mientras que el resto de la tabla se comporta como una gran tabla tipo dimensión.

Estas características son de gran utilidad para involucrar y entrenar a los usuarios en el manejo de modelos analíticos, que con el paso del tiempo pueden ir evolucionando hasta convertirse en modelos más complejos como los de un modelo multidimensional.

En este caso, se implementó un modelo para el área de medios electrónicos de pagos, con la meta de proveer un modelo de datos que fuera familiar y limitado a sus necesidades. Como insumo se utilizaron *tablas de integración* y unas pocas *tablas históricas* e internamente se incorporaron diferentes tipos de métricas que complementaron el modelo. Su utilización se realiza por medio de la herramienta Excel, además de incorporarse en reportes dinámicos (más detalle en el apartado de herramientas de visualización).

Como se observa en la imagen adjunta de la *infraestructura propuesta*, este proceso forma parte del proceso de Analysis Service (SSAS), en la parte analítica. La implementación de la solución se efectuó mediante el despliegue en una instancia de *base de datos tabular*, desde donde se administra y controla desde entonces.

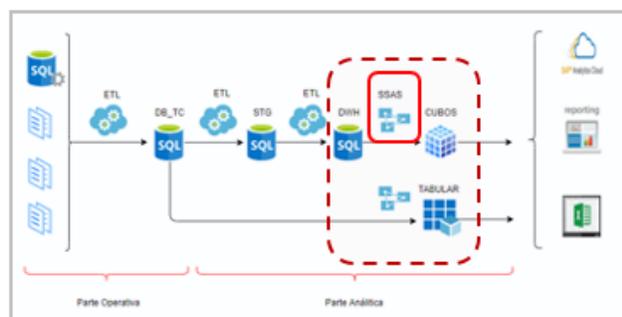
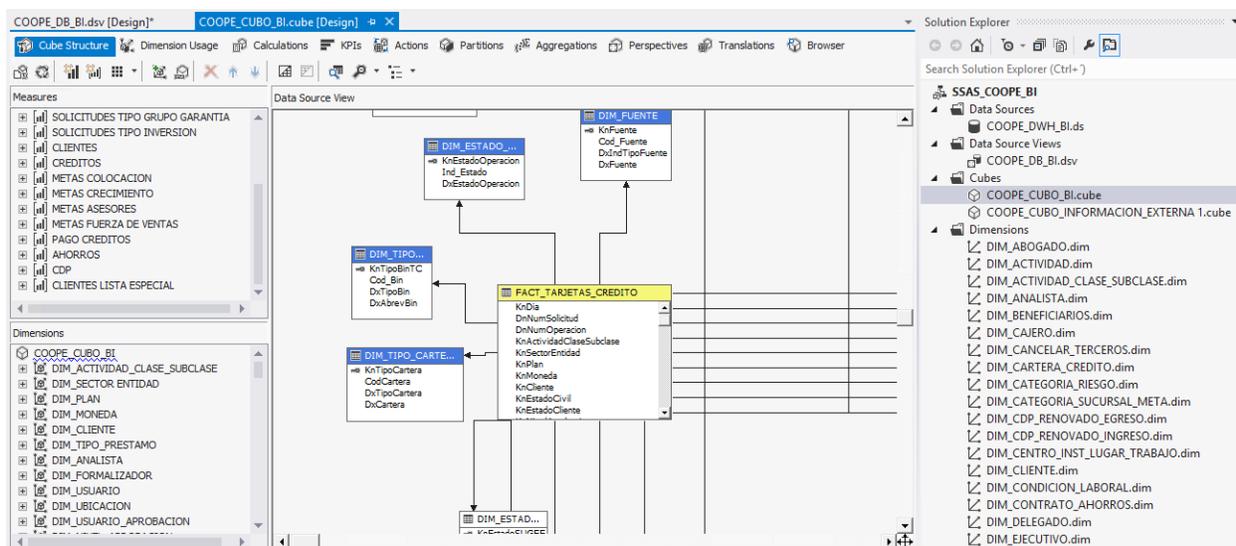


Figura 21. Modelo de datos multidimensional
Fuente: Elaboración propia

El *modelo de datos multidimensional* vino a llenar el vacío de datos que existía acerca del producto tarjetas dentro del *data warehouse* corporativo, permitiendo a los analistas avanzados incorporar información más detallada en cuanto a los distintos atributos del producto en sus trabajos y, de esta manera, generar modelos de decisión más precisos y confiables, además de ayudar en la creación de campañas comerciales mejor dirigidas y con mayor impacto.

Ahora bien, dado que la empresa ya contaba con un modelo multidimensional de grandes proporciones, los analistas usuarios de esta herramienta no requirieron de mucho entrenamiento para rápidamente comenzar con la exploración de la información.

Con respecto a la imagen adjunta de la *infraestructura propuesta*, se aprecia que este también forma parte de los procesos de Analysis Service (SSAS), en la parte analítica. La implementación se integró a la solución corporativa, la cual se despliega en una instancia de *bases de datos multidimensional*, desde donde se administra y controla.

5.3.4 Herramientas de visualización

Esta sección representó una parte estratégica del proyecto, pues es mediante las herramientas utilizadas en este apartado que los usuarios lograron tener contacto con los datos procesados a lo largo de todo el proyecto. De este modo, se buscó cumplir con tres características: satisfacer los datos requeridos por el usuario, proporcionar herramientas intuitivas de fácil manejo y permitir el autoservicio de los datos para los usuarios más avanzados (cumpliendo en todo momento con los roles de cada usuario). Asimismo, la liberación de los datos por medio de las herramientas se realizó igualmente de manera estratégica, para que los usuarios se familiarizaran con estas e incrementaran su uso.

Primero se liberaron los modelos tabulares y multidimensionales para ser accedidos mediante la herramienta Excel, de esta forma los usuarios intermedios y avanzados (en su mayoría analistas de datos de cada área involucrada) podían utilizar los modelos para verificar sus datos (esto generó confianza en los modelos) y posteriormente para producir análisis más complejos, pero importantes con el fin de mostrar la versatilidad y funcionabilidad de los modelos.

Etiquetas de fila	Fecha_Actualizacion	Cuenta_Num_Tarjeta	Count_Autorizaciones	Sum_Monto_Autorizado	Avg_Tipo_Cambio	Sum_Monto_Colones
#414452	01/01/2020	8	8	395,000.00	569.93	395,000.00
	02/01/2020	21	27	1,504,000.00	570.09	1,504,000.00
	03/01/2020	29	34	1,186,000.00	569.50	1,186,000.00
	04/01/2020	43	51	1,876,000.00	569.50	1,876,000.00
	05/01/2020	34	39	881,000.00	569.50	881,000.00
	06/01/2020	38	44	2,075,000.00	569.37	2,075,000.00
	07/01/2020	34	36	1,625,100.00	567.47	1,681,747.00
	08/01/2020	31	39	984,153.67	569.47	1,071,527.36
	09/01/2020	31	35	903,000.00	569.48	903,000.00
	10/01/2020	26	29	1,110,000.00	569.80	1,110,000.00
	11/01/2020	26	27	979,000.00	569.84	979,000.00
	12/01/2020	23	24	1,008,000.00	569.84	1,008,000.00
	13/01/2020	27	31	1,422,000.00	568.61	1,422,000.00
	14/01/2020	22	24	855,000.00	566.59	855,000.00
	15/01/2020	30	35	1,444,000.00	565.84	1,444,000.00
	16/01/2020	18	20	899,000.00	564.67	899,000.00
	17/01/2020	27	29	1,143,160.00	563.46	1,233,139.20
	18/01/2020	23	27	906,000.00	563.37	906,000.00
	19/01/2020	13	15	550,000.00	563.37	550,000.00
	20/01/2020	24	26	742,000.00	561.54	742,000.00
	21/01/2020	16	19	877,000.00	561.54	877,000.00
	22/01/2020	11	13	415,000.00	563.31	415,000.00
	23/01/2020	10	12	759,000.00	565.35	759,000.00
	24/01/2020	18	20	746,000.00	567.19	746,000.00
	25/01/2020	19	20	652,000.00	567.19	652,000.00
	26/01/2020	9	10	282,000.00	567.19	282,000.00
	27/01/2020	16	21	1,234,000.00	566.51	1,234,000.00
	28/01/2020	15	16	983,000.00	565.53	983,000.00
	29/01/2020	13	15	494,000.00	566.55	494,000.00
	30/01/2020	16	16	675,000.00	567.22	675,000.00
	31/01/2020	15	17	731,000.00	567.06	731,000.00

Figura 22. Pivoteo en Excel de modelos tabulares/multidimensionales
Fuente: Elaboración propia

Luego se desarrolló una serie de reportes previamente identificados (de utilización para cierres de mes, seguimiento de comportamientos, solicitud recurrente, entre otros) en Reporting Services (esta herramienta dispone de una plataforma web de fácil acceso y administración), donde se dispusieron carpetas con el nombre de cada unidad de negocio asociada al proyecto y dentro de dichas carpetas se colocaron los reportes implementados (a los cuales también se les podía administrar su acceso de ser necesario). Los reportes se crearon con un concepto dinámico en donde el usuario puede ir aplicando una serie de filtros para obtener un conjunto completo de datos o el subconjunto deseado.

Inicio > Medios_de_Pago > Detalle_Transacciones_X_Plataforma Inicio | Mis suscripciones | Configuración del sitio | Ayuda

Periodo: 2020-06 Marca: MASTERCARD; VISA Ver informe

Categoría_Tarjeta: BLACK Tipo_Plataforma: Plataforma de Compras; Plataform

Plataforma: AIRBNB; AMAZON; GLOVO; GOOC

1 de 1 Buscar | Siguiente

Detalle_Transacciones_X_Plataforma

Marca	Categoría Tarjeta	Tipo Tarjeta Emisor	Tipo Plataforma	Plataforma	DO				
					# Transacciones	Monto Total Transacciones	Monto Promedio Transaccion	Monto Total Comision	Monto Promedio Comision
MASTERCARD	BLACK	TC	Plataforma de Compras	AMAZON	21	7,136.58	339.84	78.47	3.74
				WISH	2	32.00	16.00	0.35	0.18
			Plataforma de Entrega de Comida	GLOVO	9	152.89	16.99	1.68	0.19
			Plataforma de Hospedaje	AIRBNB	1	1,167.24	1,167.24	12.84	12.84
			Plataforma de Musica Online	SPOTIFY	9	59.91	6.66	0.60	0.07
			Plataforma de Pago	GOOGLE	16	1,368.39	85.52	14.99	0.94
				PAYPAL	13	371.30	28.56	4.05	0.31
			Plataforma de Transporte	UBER	207	3,339.24	16.13	36.77	0.18
			Plataforma de Video Streaming	NETFLIX	16	239.95	15.00	2.68	0.17
				PLAYSTATION	3	50.17	16.72	0.50	0.17

Figura 23. Reporte implementado en Reporting Services

Fuente: Elaboración propia

Finalmente, mediante la utilización de la herramienta SAC (SAP Analytics Cloud), se implementó la creación de tres tipos de tableros de control diferentes y dos tipos de infográficos. Los tableros fueron creados para proveer a las jefaturas y gerencias de información generalizada de la operativa diaria del producto tarjetas, en tanto que los infográficos fueron incluidos en una revista virtual originada con el objetivo de informar a la Junta Directiva de la empresa. De esta manera, se dio visibilidad a los datos generados durante todo el proyecto a nivel de los mandos altos de la empresa.

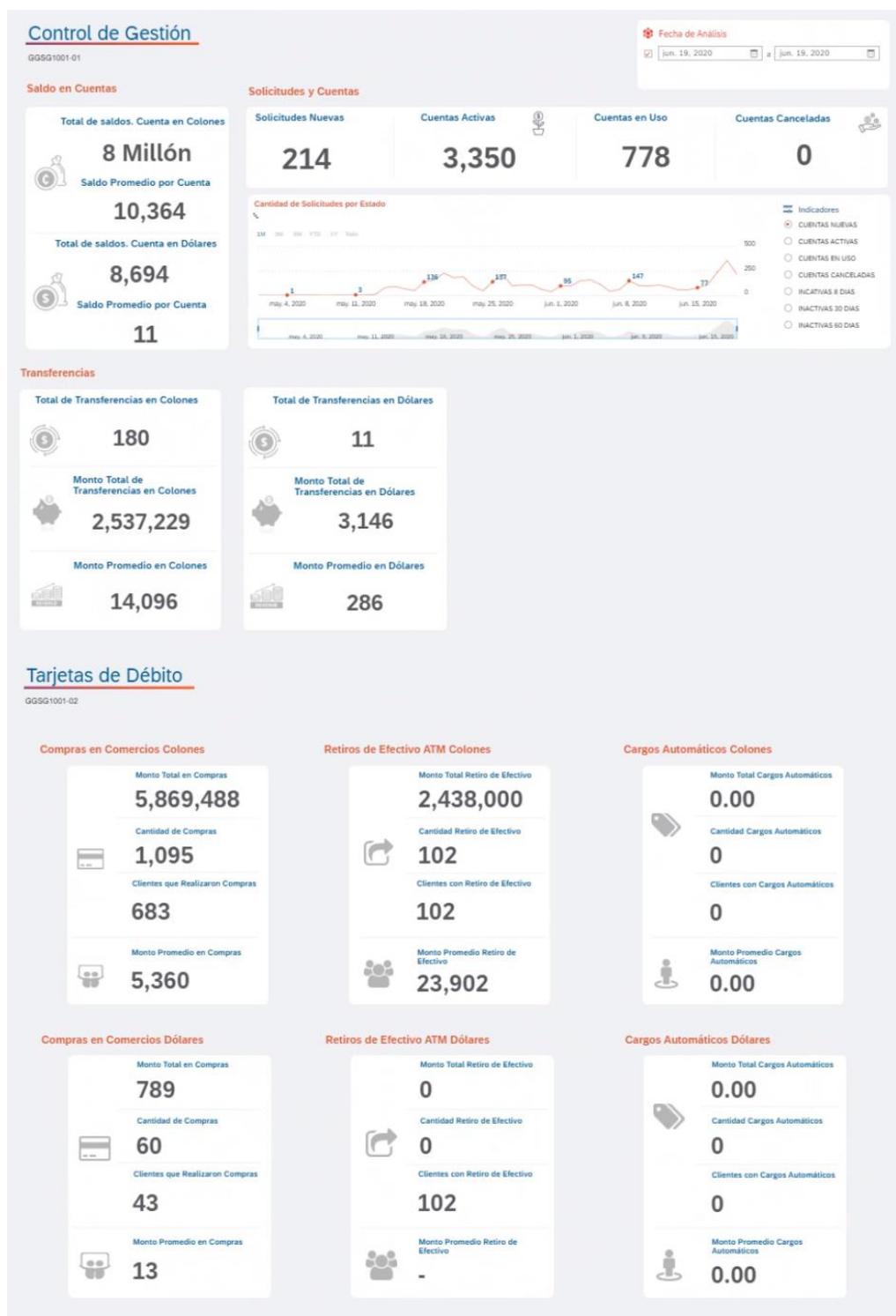


Figura 24. Tablero de control implementado en la herramienta SAC
Fuente: Elaboración propia

6. Capítulo 6: Conclusiones y recomendaciones

6.1. Conclusiones

En este proyecto se construyó una infraestructura de datos robusta, que permita la administración de la información histórica y actual del producto tarjetas, y que sirva de plataforma para el análisis y la toma de decisión. Lo más importante de la infraestructura creada es que dispone de información completa y actualizada (T-1), lo cual posibilita que las unidades de negocio accedan a datos que con anterioridad no se disponían. En relación con esto, el elemento que más ayudó a lo largo del proceso fue contar con usuarios expertos que conocían el ciclo completo del negocio, lo que posibilitó el avance sin contratiempos a lo largo de todo el proceso. Por su parte, lo más difícil a lo que se enfrentó en el proyecto en torno a este punto fue la resistencia al cambio tecnológico por parte de la mayoría de los usuarios, debido a que estos estaban acostumbrados a dedicar mayor parte de su tiempo en procesos manuales.

En segundo lugar, se identificaron las fuentes de datos involucradas en el producto tarjetas. Lo más importante de este punto fue disponer oportunamente de la información requerida para avanzar y finalizar este subproceso. El aspecto que más ayudó fue poder contar con la definición de las estructuras de la información mediante una guía de procesos, pues esto disminuyó el tiempo de consulta a los usuarios expertos. Por otro lado, lo más difícil en este proceso fue definir el alcance de las fuentes secundarias utilizadas, ya que no existe documentación del sistema transaccional para establecer un vínculo entre los datos.

Además, se definió el flujo correcto para la carga de los datos desde las fuentes iniciales. En cuanto a este paso, lo más relevante fue crear un proceso eficiente de carga. Al respecto, lo que más ayudó a lograr esta carga eficiente fue contar con el conocimiento de las herramientas utilizadas en el proceso. Sin embargo, lo más difícil fue establecer la secuencia de carga, de tal

manera que se pudiera contar con un objeto de control que evite la duplicación de eventos y garantice el cumplimiento de todos los pasos establecidos.

Asimismo, se creó un nuevo modelo de estructuras para agrupar, unificar e integrar los datos en los conceptos que se ajusten a la lógica del negocio. En cuanto a esto, lo más importante de este proceso fue que se crearon los cimientos de todo el flujo entre las diferentes bases de datos participantes. Por su parte, el aspecto que más ayudó a lograr el modelado fue la participación de los usuarios expertos en la composición y definición de los distintos conceptos del negocio. No obstante, lo más difícil fue contextualizar los aportes de los usuarios del negocio, debido a que ellos conocían ampliamente los conceptos, pero requirió interpretarlos y ajustarlos para el entorno de desarrollo.

También, se desarrolló la secuencia lógica de la carga de datos en las estructuras definidas. Con relación a esto, lo más relevante fue automatizar el proceso, de tal manera que se tomaran todas las condiciones y características requeridas para garantizar la calidad y disponibilidad de los datos. Lo que más ayudó fue la definición de modelado anterior, pues permitió identificar el orden y secuencia de carga de todos los conceptos requeridos. Por su parte, lo más difícil fue realizar las pruebas de comprobación de los datos cargados, ya que la organización no contaba con una herramienta que posibilitara hacer la paridad de los datos a nivel histórico.

Aunado a lo expuesto, se diseñaron nuevos esquemas para el almacenamiento en repositorios de información, empleando diversas tecnologías (bases de datos tabulares y bases de datos multidimensionales). En relación con esto, lo más importante de la creación de estos repositorios fue que cumplieran con diferentes funciones, porque el aporte de estos apoya el autoservicio de datos con usuarios de nivel intermedio y alto, lo que resultó en un cambio conceptual de las labores operativas, liberando a los usuarios de procesos manuales y enfocándolos en un análisis de información, lo cual incrementa la productividad de las unidades de negocio

involucradas y genera información de apoyo para la toma de decisión. Por otra parte, el elemento que más ayudó en la creación de los esquemas fue la utilización de herramientas interactivas y de fácil implementación que permitieran enfocar las funciones de los esquemas en modelos intuitivos y amigables para el usuario final. Sin embargo, lo más difícil de esta etapa fue convencer a los mandos altos de utilizar los repositorios recientemente creados, pues no confiaban en la información brindada en los nuevos esquemas de datos.

Por último, se establecieron los indicadores relevantes para la visualización y toma de decisiones. Lo más importante del establecimiento de indicadores fue clasificarlos en dos tipos: indicadores de control/rendimiento e indicadores de toma de decisión. Los primeros permiten medir la utilización de las estructuras y herramientas creadas, en tanto que los segundos posibilitan las mediciones propias de los datos para la toma de decisión. En cuanto a esto, lo que más ayudó en el establecimiento de indicadores fue la evaluación del sistema de las solicitudes de datos, pues esta facilitó identificar el tipo de consulta requerido por cada unidad de negocio. No obstante, lo más difícil de esta etapa fue lograr un consenso entre las unidades de negocio involucradas para definir indicadores únicos que satisficieran los requerimientos de las áreas participantes.

6.2. Recomendaciones

Dentro de un proyecto tan ambicioso como lo fue este, siempre se desea que haya una mejora continua del mismo; por lo tanto, se enumeran las siguientes recomendaciones:

- A nivel del producto tarjetas, se establece:
 - Continuar con la integración y unificación de datos provenientes de otras fuentes:
 - Transferencias a nivel de cuentas o número de tarjeta.
 - Pagos por ventanilla y por deducción automática.
 - Relacionar los datos de tarjetas según el canal digital utilizado por el cliente.

- Proveer visibilidad del producto tarjetas ante los diferentes estados que pueda asumir el asociado (en especial para renunciaciones y fallecimientos).
 - Proveer un seguimiento más detallado de los procesos de renovación y reposición de tarjetas, además de los vencimientos por mes en curso.
 - Automatización de procesos manuales identificados, que para este proyecto no fueron considerados.
 - Crear una dimensión que muestre las cuentas ligadas a una tarjeta de débito.
- A nivel de infraestructura de desarrollo y producción, se recomienda:
 - Crear un repositorio único de fuentes de datos (lago de datos).
 - Independizar el *hardware* y simplificar la administración (migración a bases de datos a la nube), lo que impactaría en tiempos de respuesta más efectivos.
 - Lograr la independencia del DWH corporativo (migración a la nube), lo que ayudaría en la disminución de los costes de tiempo de procesamiento.
 - A nivel de reportes, tableros de control e informes, se aconseja:
 - Continuar con la creación de nuevos reportes referenciando procesos automatizados.
 - Continuar con la creación de nuevos informes que incluyan solicitudes de datos identificadas recientemente como solicitudes recurrentes.
 - Crear un nuevo tablero de datos referenciando los diferentes canales digitales disponibles.
 - Crear un reporte que muestre la bitácora de cambios de las cuentas ligadas a una tarjeta de débito y sus atributos.

Referencias

- Adamson, C. (2010). *The complete reference star schema*. Nueva York, Estados Unidos: McGraw-Hill.
- Amazon. (2019). *¿Qué es una base de datos columnar?* Recuperado de: <https://aws.amazon.com/es/nosql/columnar/>
- Biolchini, J., Gomes, P., Cruz, A. y Hort, G. (2005). *Systematic Review in Software Engineering*, Recuperado de: <https://www.cos.ufrj.br/uploadfile/es67905.pdf>
- Coopenae. (2019). *Historia, visión, misión y valores de Coopenae*. Recuperado de: <https://www.coopenae.fi.cr/nuestra-historia/>
- Evertec. (2019). *Procesamiento emisor*. Recuperado de: <https://www.evertecinc.com/es-pr/services/processing/procesamiento-emisor/>
- Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la investigación*. México: McGraw-Hill.
- Hurtado de Barrera, J. (2005). *Cómo formular objetivos de investigación: un acercamiento desde la investigación holística*. España: Quirón Ediciones.
- Imhoff, C., Gallemmo, N. y Geiger, J. (2003). *Mastering data warehouse design relational and dimensional techniques*. Indianapolis, Estados Unidos: Wiley Pub.
- Inmon, W. (2005). *Building the data warehouse*. Indianapolis, Estados Unidos: Wiley Pub.
- Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*. Recuperado de: <http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>
- Martín, M. (2006). *Sistemas gestores de bases de datos: ciclos formativos, grado superior*. Madrid, España: McGraw-Hill.

- MasterCard. (2019). *Procesamiento de pagos*. Recuperado de: <https://latinamerica.mastercard.com/es-region-lac/acerca-de-mastercard/que-hacemos/procesamiento-de-pagos.html>
- Microsoft. (2018). *Procesamiento analítico en línea (OLAP)*. Recuperado de: <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/online-analytical-processing>
- Microsoft. (2020). *Información general sobre el modelado tabular*. Recuperado de: <https://docs.microsoft.com/es-es/analysis-services/tabular-models/tabular-models-ssas?view=asallproducts-allversions>
- Raffino, M. (2020). *Fuentes de información*. Recuperado de: <https://concepto.de/fuentes-de-informacion/>
- SAP. (2017). *¿Qué es SAP HANA?* Recuperado de: <https://www.ematzgroup.com/single-post/2017/08/09/%C2%BFQu%C3%A9-es-SAP-HANA>
- SAP. (2019a). *Aproveche la base de datos in-memory de SAP*. Recuperado de: <https://www.sap.com/latinamerica/documents/2017/12/64894d8c-e37c-0010-82c7-eda71af511fa.html>
- SAP. (2019b). *Sap Hana Database*. Recuperado de: <https://help.sap.com/viewer/04030263a0d041309a039fa3ea586720/7.3.20/en-US>
- Schmitz, A. (2015). *¿Qué es SAP HANA?* Recuperado de: <https://news.sap.com/latinamerica/2015/09/que-es-realmente-sap-hana/>
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for DataMining. *Journal of Data Warehousing*, 5(4), 13-23.
- Silberschatz, A., Korth, H. y Sudarshan, S. (2002). *Fundamentos de bases de datos*. Madrid, España: McGraw-Hill.

Zubizarreta, A. (1986). *La aventura del trabajo intelectual. Cómo estudiar e investigar*. México: Fondo Educativo Interamericano.

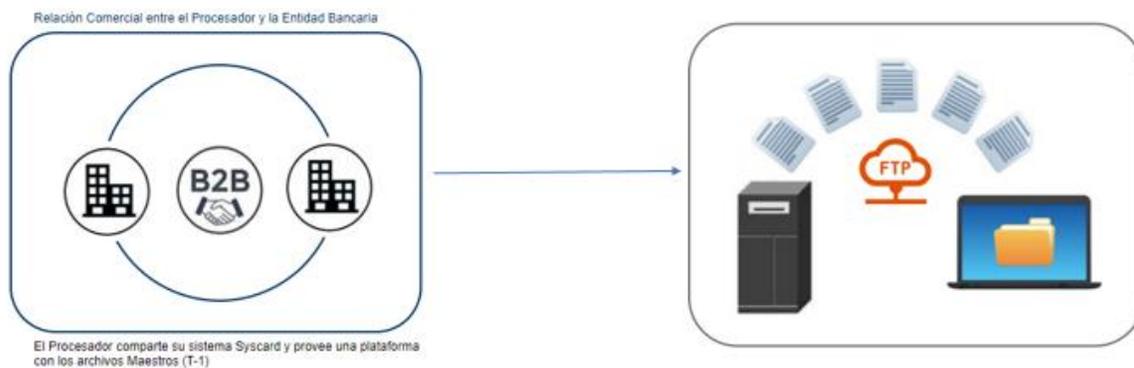
Apéndices

Apéndice 1. Composición de la tarjeta



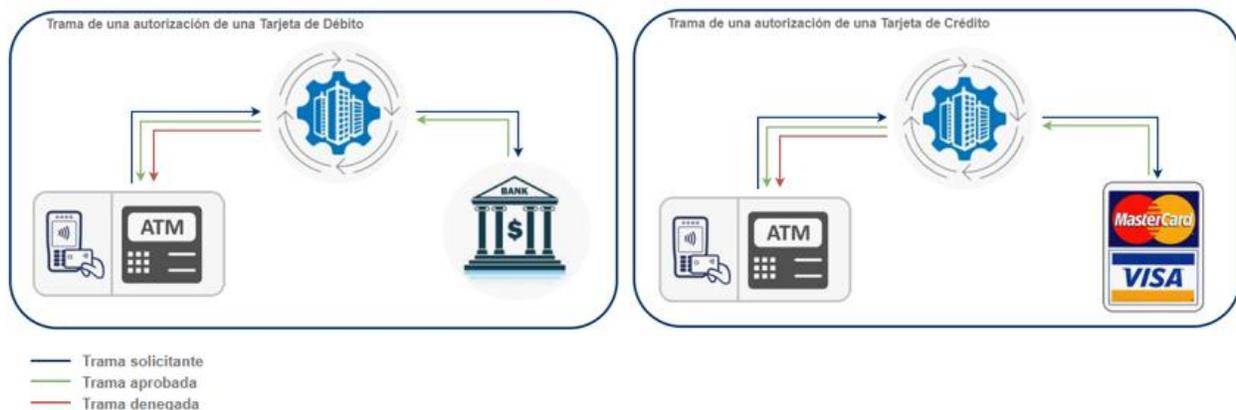
Fuente: Elaboración propia

Apéndice 2. Datos del procesador



Fuente: Elaboración propia

Apéndice 3. Tramas de autorización



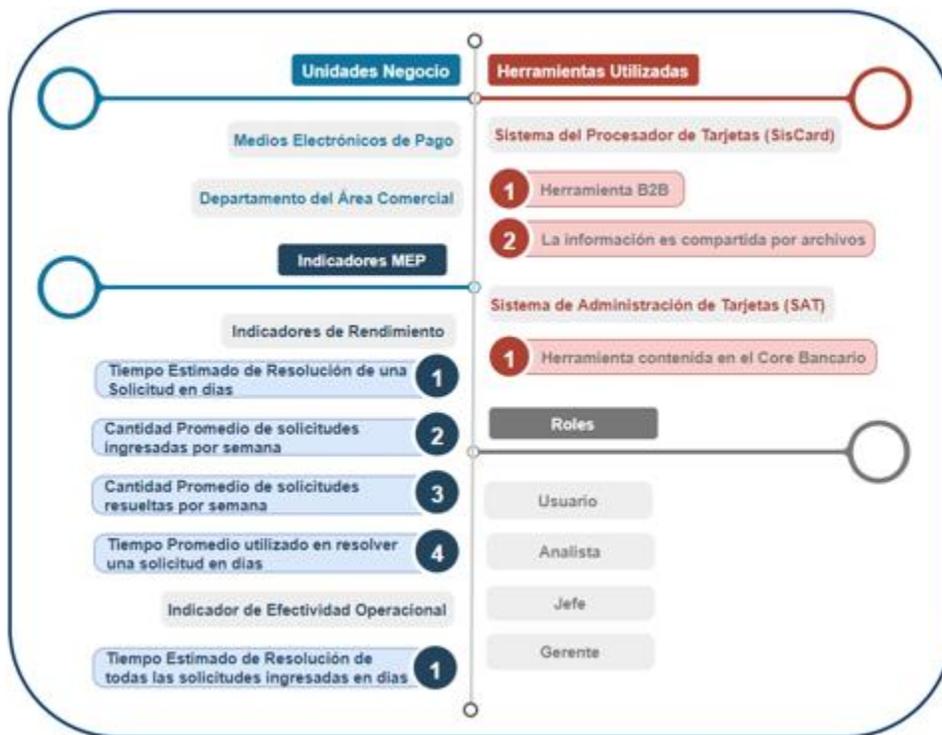
Fuente: Elaboración propia

Apéndice 4. Conjunto de datos



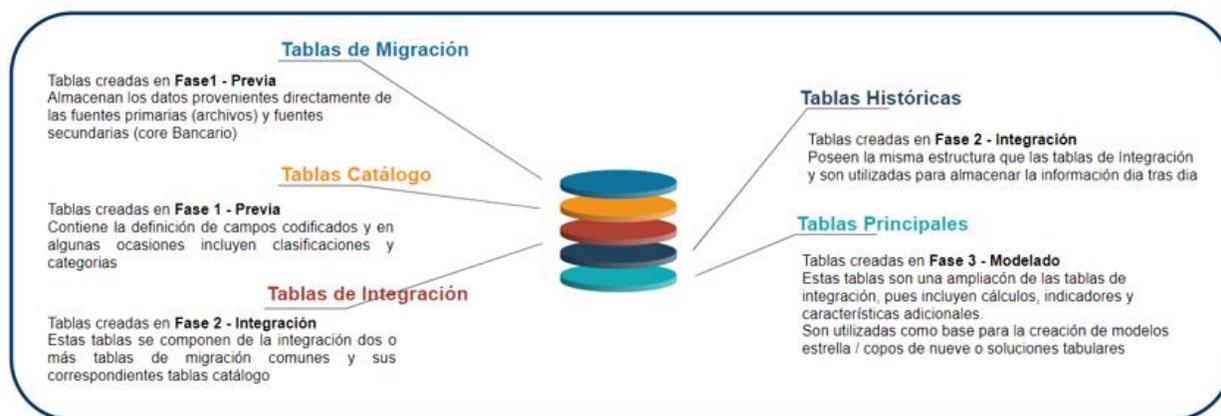
Fuente: Elaboración propia

Apéndice 5. Recopilación de información



Fuente: Elaboración propia

Apéndice 6. Estrategia definida en fases



Fuente: Elaboración propia

Apéndice 7. Distribución de tablas



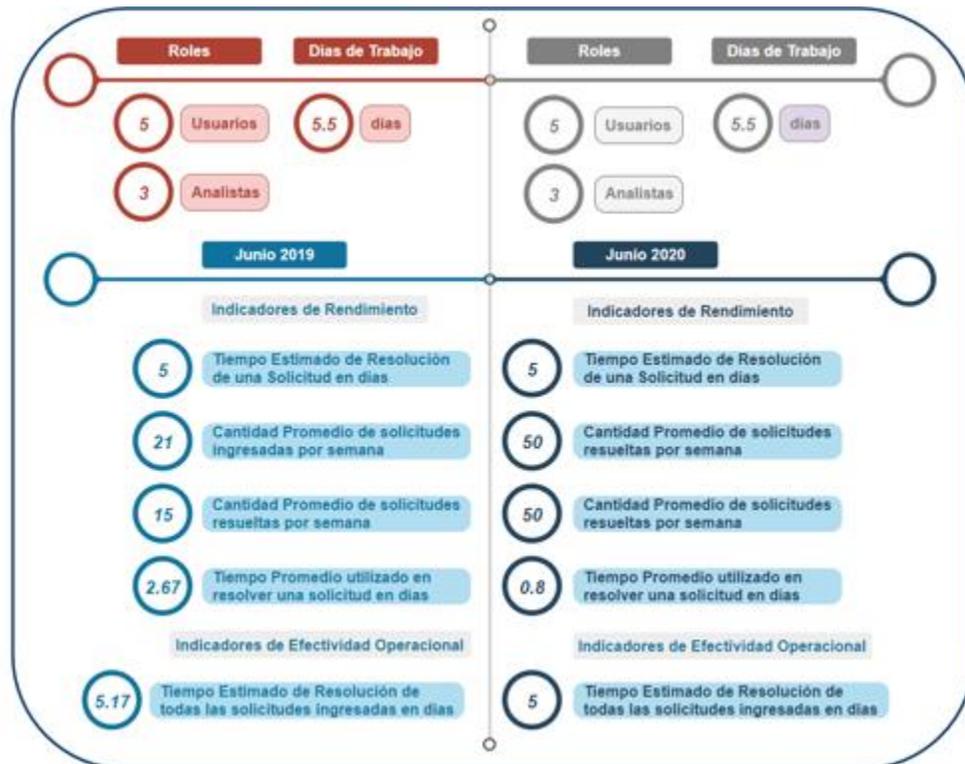
Fuente: Elaboración propia

Apéndice 8. Inventario de objetos



Fuente: Elaboración propia

Apéndice 9. Comparación de indicadores



Fuente: Elaboración propia

Apéndice 10. Acta de entrega y aceptación de producto

	COOPENAE, R. L. GERENCIA GESTIÓN ESTRATÉGICA	Código: FOR PMO 01-0918-06.v1
	ACTA DE ENTREGA Y ACEPTACIÓN DE PRODUCTO	Riesgo Asociado:
		Página 1 de 2

PRY 20190515001–Creación de una BD para administrar el producto tarjetas Acta de entrega y aceptación de producto 15-mayo-2020

1. Control de aprobaciones:

Los funcionarios suscritos que participaron en el proyecto dan fe de la recepción y aceptación del (los) entregable(s) indicado(s) en este documento, y que el contenido del presente documento refleja con total fidelidad los comentarios y las observaciones relevantes que fueron aportados. Los abajo firmantes confirman que esta recepción y aceptación implica la transferencia de la propiedad y responsabilidad en su utilización del (los) entregable(s) descritos.

Rol	Participante	Puesto	Firma
Sponsor	Cindy Domonkos Obando	Gerencia Medios de Pago	<i>Firma no requerida, aprobación válida por correo electrónico</i>
PM	Álvaro Duran Vargas	Gerencia de Gestión Estratégica	<i>Firma no requerida, aprobación válida por correo electrónico</i>
Líder funcional /QA	Agustín Brown Sirias	Gestión Estratégica	<i>Firma no requerida, aprobación válida por correo electrónico</i>
Líder técnico	Miguel Chavarría Arroyo	Gestión Estratégica	<i>Firma no requerida, aprobación válida por correo electrónico</i>

2. Control de distribución

Las siguientes personas serán informadas sobre el resultado de la reunión.

Interesado	Puesto	Interés
José Joaquín Vargas Zamora	Gerente de Canales	Notificar

* Interés: Notificar, archivar, informar, retroalimentar.

Copia de Archivo

 <p style="text-align: center;">COOPENAE, R. L. GERENCIA GESTIÓN ESTRATÉGICA</p> <p style="text-align: center;">ACTA DE ENTREGA Y ACEPTACIÓN DE PRODUCTO</p>	Código: FOR PMO 01-0918-06.v1
	Riesgo Asociado:
	Página 2 de 2

3. Datos generales

Número y nombre de proyecto:	<i>Creación de una BD para administrar el producto tarjetas</i>
Producto:	<i>Base de datos de análisis</i>
Fase del proyecto:	<i>Ejecución</i>
Fecha de entrega:	<i>15/05/2020 día en que se está realizando la entrega</i>
Fecha de aceptación:	<i>20/05/2020 día en que el grupo multidisciplinario o funcionarios acepta la entrega del producto</i>

4. Contenido del entregable

A continuación, se describe el contenido que debe tener el entregable por recibir:

- 3 bases de datos tabulares para el análisis de usuarios analíticos.
- 25 informes de Reporting Services.
- 1 tablero de control.

5. Criterios de aceptación

A continuación, se detallan los criterios de aceptación que se espera contenga el entregable por recibir:

- Funcionabilidad de las herramientas desarrolladas.
- Cumplimiento de los requerimientos en cada entregable.
- Actualización diaria de los entregables.
- Comprobación de la información según las fuentes de datos.

6. Observaciones: *No hay observaciones*

Copia de Archivo