

UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

FACULTAD DE INGENIERÍA Y GESTIÓN

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



**“OPTIMIZACIÓN DE QUERYS PARA MEJORAR EL RENDIMIENTO DE LOS
PROCESOS ETL EN EL ÁREA DE INTELIGENCIA COMERCIAL DEL BANCO
PICHINCHA”**

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de

INGENIERO DE SISTEMAS

PRESENTADO POR EL BACHILLER

HINOSTROZA NUÑEZ, BRYAN JOSUE

Villa El Salvador

2019

DEDICATORIA

El presente trabajo va dedicado a mis padres, quienes, con su esfuerzo y apoyo desde temprana edad, me inculcaron el valor del estudio y el esfuerzo para lograr las metas que me trace.

AGRADECIMIENTOS

- A mis padres, porque siempre se sacrificaron por darme lo mejor y enseñarme el valor de la humildad y constancia para conseguir mis objetivos
- Al Banco Pichincha por brindarme las facilidades para poder llevar a cabo este proyecto.
- A mi señorita enamorada, quien me brindó su apoyo en todo momento dentro y fuera de las aulas para alcanzar mis metas.
- A mi familia por ser el soporte que me permitió seguir adelante y ser la razón de mi esfuerzo.
- A mi asesor, por su consejo y asesoría para la correcta elaboración de este trabajo.

ÍNDICE

DEDICATORIA.....	ii
AGRADECIMIENTOS	iii
LISTA DE FIGURAS	vii
LISTA DE TABLAS.....	ix
INTRODUCCIÓN.....	1
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	2
1.1 Descripción de la realidad problemática	2
1.2 Justificación del Problema	5
1.2.1 Justificación Teórica.....	5
1.2.2 Justificación Práctica.....	5
1.2.3 Justificación Tecnológica	6
1.2.4 Justificación Medio Ambiental	6
1.3 Delimitación del Proyecto	6
1.3.1 Teórica	6
1.3.2 Temporal.....	7
1.3.3 Espacial.....	7
1.4 Formulación del Problema	8
1.4.1 Problema General	8
1.4.2 Problemas Específicos.....	8
1.5 Objetivos.....	8
1.5.1 Objetivo General	8
1.5.2 Objetivos Específicos	8
CAPÍTULO II: MARCO TEÓRICO.....	10
2.1 Antecedentes.....	10

2.1.1	Antecedentes Internacionales	10
2.1.2	Antecedente Nacional	14
2.2	Bases Teóricas	15
2.2.1	Lenguaje de Consulta Estructurado	15
2.2.2	Inteligencia de Negocios	17
2.2.3	Proceso de Extracción, Transformación y Carga	18
2.2.4	Servicio de Integración de SQL Server	20
2.2.5	Inteligencia Comercial	23
2.2.6	Banco Pichincha Perú	23
2.3	Definición de Términos Básicos.....	25
CAPÍTULO III: DESARROLLO DEL TRABAJO DE SUFICIENCIA PROFESIONAL		30
3.1	Análisis Exploratorio Situacional.....	31
3.1.1	Mapeo de Fuentes	31
3.1.2	Descripción del proceso de ETL actual.....	33
3.1.3	Medición de los Tiempos ETL Actual	41
3.2	Construcción de la Solución	48
3.2.1	Elaboración de los diccionarios de datos.	48
3.2.2	Descripción del Proceso ETL Propuesto.....	49
3.2.3	Medición de los Tiempos ETL Propuesta.....	69
3.3	Validación del Modelo Propuesto	74
3.3.1	Validación a Nivel Contenedor.....	74
3.3.2	Validación a Nivel Tarea	75
CONCLUSIONES.....		76
RECOMENDACIONES.....		77
BIBLIOGRAFÍA.....		78

ANEXOS.....	80
Tiempos de Ejecución ETL anterior	80
Tiempos de Ejecución ETL propuesta.....	84
Formato de Diccionario de Datos	87

LISTA DE FIGURAS

Figura 1 Tipos de declaraciones de sentencias SQL.....	15
Figura 2 Transformación de los datos para toma de decisiones.....	18
Figura 3 Proceso ETL.....	19
Figura 4 Arquitectura SSIS.....	22
Figura 5 Etapas del trabajo de investigación.....	30
Figura 6 Fuentes Servidor SBIBFPR03.....	31
Figura 7 Fuentes Servidor SBDBFPR15.....	32
Figura 8 Estructura ETL actual.....	33
Figura 9 Validar Activos (Sequence Container).....	34
Figura 10 Carga Temporales (Sequence Container).....	35
Figura 11 Actualizar Correcciones (Sequence Container).....	36
Figura 12 Stock Actual (Sequence Container).....	37
Figura 13 Saldos Retenidos (Sequence Container).....	38
Figura 14 Desembolsos Hamber (Sequence Container).....	39
Figura 15 Data Adicional (Sequence Container).....	40
Figura 16 Reporte PLD (Sequence Container).....	40
Figura 17 Tiempo de Ejecución de los Contenedores por Día (min).....	42
Figura 18 Tareas contenedor actualizar correcciones.....	44
Figura 19 Tareas del contenedor carga temporales.....	46
Figura 20 Estructura ETL propuesta.....	49
Figura 21 Estructura propuesta de validar información.....	50
Figura 22 Consulta Validar TipoCambio.....	51
Figura 23 Estructura propuesta de carga intermedia.....	52
Figura 24 Eliminación de índice en la consulta.....	53
Figura 25 Reemplazo de sentencia IN por INNER JOIN.....	53
Figura 26 Estructura propuesta de actualizar correcciones.....	55
Figura 27 Modificación de la tarea Carga PLD.....	56
Figura 28 Modificación de la tarea Actualización.....	57
Figura 29 Modificaciones de la tarea Update PLD.....	58

Figura 30 Modificaciones de la tarea Correcciones PLD	59
Figura 31 Modificaciones de la tarea Actualizar Canal	60
Figura 32 Modificación de la tarea Reenganche Fuente: Elaboración propia	61
Figura 33 Estructura propuesta de stock PLD	62
Figura 34 Consulta de la tarea Cálculo del Stock	63
Figura 35 Estructura propuesta de data adicional	64
Figura 36 Consulta de la tarea Info Leo Canales.....	65
Figura 37 Consulta de la tarea Indicadores Consolidado	66
Figura 38 Consulta de la tarea T_TC_Agencias	66
Figura 39 Estructura propuesta de reporte PLD	67
Figura 40 Estructura propuesta de desembolsos Hamber	68
Figura 41 Tiempo de ejecución de los contenedores propuesta (min)	70
Figura 42 Tareas propuesta contenedor actualizar correcciones	72

LISTA DE TABLAS

Tabla 1 Tiempo de ejecución por sentencia en tablas sin índices	13
Tabla 2 Tiempo de ejecución por sentencia en tablas con índices	13
Tabla 3 Tiempo de ejecución diario por contenedor	42
Tabla 4 Análisis de tiempos de ejecución por contenedor	43
Tabla 5 Tiempos de ejecución diarios del contenedor actualizar correcciones	44
Tabla 6 Análisis de los tiempos del contenedor actualizar correcciones	45
Tabla 7 Tiempos de ejecución diarios del contenedor carga temporales	46
Tabla 8 Análisis de los tiempos del contenedor carga temporales	47
Tabla 9 Tiempos de ejecución diarios por contenedor - ETL propuesto	70
Tabla 10 Análisis de los tiempos por contenedor - ETL propuesta	71
Tabla 11 Tareas del contenedor propuesto, actualizar correcciones.....	72
Tabla 12 Análisis de los tiempos de las tareas del contenedor actualizar correcciones propuesto	73
Tabla 13 Comparativo de tiempos entre proceso actual y propuesta por contenedor	74
Tabla 14 Comparativo de tiempos entre proceso actual y propuesta por tarea.....	75
Tabla 15 Tiempos de ejecución ETL anterior	80
Tabla 16 Tiempos de ejecución ETL propuesta.....	84
Tabla 17 Diccionario de datos, primera parte	87
Tabla 18 Diccionario de datos, segunda parte.....	87

INTRODUCCIÓN

La optimización de consultas es un concepto clave a la hora de construir una consulta SQL, puesto que su incorrecta estructuración o hacer uso inadecuado de las sentencias puede desembocar en una ejecución que tarde más de lo debido, consuma mayor cantidad de recursos y en el peor de los casos, devuelva datos inconsistentes, por ello es necesario mantenerse actualizado con las mejores prácticas y tener siempre el apoyo de un diccionario de datos que nos facilite la comprensión de los datos con los que vamos a trabajar.

En el presente trabajo de investigación se evalúa el rendimiento de un proceso de extracción, transformación y carga en el área Inteligencia Comercial del Banco Pichincha, donde nuestro indicador principal será el tiempo de ejecución que tardan los contenedores y sus tareas, con el fin de encontrar el componente que más carga de trabajo causa al servidor. Se trabajó con la siguiente estructura.

El Capítulo I hace referencia al planteamiento del problema en general, que estará compuesto por la descripción de la realidad problemática, la justificación del problema, la delimitación del proyecto, la formulación del problema y los objetivos que se buscan cumplir.

El Capítulo II engloba el marco teórico, compuesto por antecedentes de investigaciones similares que nos aportan un enfoque de trabajo aplicable al presente, contiene las bases teóricas sobre las cuales fue planteado el trabajo de investigación y nos aporta un diccionario de términos básicos para una mejor comprensión.

Por último el Capítulo III, referente al trabajo de investigación se trabajará en base a tres partes principales; en la primera se realiza el análisis exploratorio de la situación actual del proceso; en la segunda se llevará a cabo la construcción del modelo de solución propuesta, tomando en consideración lo rescatado de la parte primera; por último se realizará la validación del modelo propuesto, contrastando los indicadores obtenidos entre de la primera y segunda parte.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

Castro (Castro, 2015) menciona que para las organizaciones, la información es un recurso esencial, puesto que es necesaria para su operatividad diaria y la toma de decisiones estratégicas que le permitan mantener un alto nivel competitivo y constante crecimiento.

Para apoyar en la toma de decisiones, en un inicio se buscaba recopilar información sobre una porción del mundo con apoyo de los sistemas de información y estaba basado en la información de libros contables, recuentos, censos (civiles y militares), etc. Hoy en día, con la informatización de las organizaciones y el desarrollo de aplicaciones de software, la finalidad principal de los sistemas de información ha pasado a ser la de dar soporte a los procesos básicos de la organización (ventas, producción, personal, etc.). (Marqués, 2015).

Hoy en día la combinación de buenas prácticas, tecnologías y capacidades utilizadas por distintas compañías para la recopilación e integración de información, sumada a la aplicación de las reglas del negocio para brindar una visión de los datos de forma más comprensible y que mejore el desempeño de la organización, es conocida bajo el nombre de Inteligencia de Negocios o Business Intelligence (BI). (Rodriguez Padilla, 2014)

Sumado al concepto de Inteligencia de Negocios nació también el de Data Mart, el cual según Sinnexus (Sinergia e Inteligencia de Negocio S.L., s.f.) es un almacén de datos orientado a un proceso de negocios en específico, además que posee la estructura adecuada para el análisis a detalle de la información, desde distintos puntos de vista.

El proceso de extracción, transformación y carga (ETL) es la esencia de un almacén de datos, ya que su correcto diseño y construcción permite extraer los datos y asegurar los estándares de calidad y veracidad de los datos, además permite el ajuste de estos para que fuentes separadas puedan ser

usadas juntas y finalmente se entreguen los datos en un formato útil para el consumo de los usuarios en la toma de decisiones. (Kimball & Caserta, 2004)

Para las empresas del sector bancario es indispensable contar con procesos que se encarguen de asegurar la calidad de los datos y su disponibilidad para la toma de decisiones. Para alcanzar este cometido el área de Inteligencia Comercial del Banco Pichincha cuenta con distintos procesos ETL, los cuales permiten recolectar datos desde distintas fuentes (bases de datos SQL Server, db2, archivos planos, etc.); integrarlas y depositarlas en otro repositorio tras haber realizado las transformaciones y limpieza de datos debida.

Las bases generadas con las ETL son consumidas por reportes que permiten hacer el seguimiento de distintos indicadores para medir la evolución de los productos ofertados por el banco. Para efectos de la presente investigación centraremos especial atención en los Préstamos de Libre Disponibilidad (PLD) y el proceso que nos apoya en la recolección, tratamiento y depósito de su información en las bases para su explotación; ya que los reportes generados son utilizados desde por asesores en agencia, para validar el avance de sus colocaciones y obtener mayores comisiones; hasta por gerentes para medir el comportamiento del producto y tomar decisiones en base a la información presentada, es necesario mitigar los problemas que han podido ser identificados y que son mencionados a continuación:

- **Dificultad para realizar modificaciones**, causadas en su mayoría por el desconocimiento de los orígenes de datos y las fórmulas para calcular algunas variables.
- **Tiempos de ejecución elevados**, debido a consultas mal estructuradas o que no llevan un plan de ejecución adecuado y al uso incorrecto de algunas herramientas del Business Intelligence Development Studio 2008 (BIDS).

- **Errores en la carga de información**, causada en su mayoría al volver a ejecutar el paquete tras haberse ocurrido un error en alguna de las fases de extracción, limpieza y transformación o carga de datos.

1.2 Justificación del Problema

1.2.1 Justificación teórica

Hoy en día, el nivel de importancia que ha alcanzado el Gobierno de Datos es mayor al que años atrás podíamos apreciar en las organizaciones sin importar su tamaño o sector, esto se debe a que es más notorio que la falta de gestión de información influye drásticamente en la toma de decisiones, al no existir procesos y/o políticas adecuados que permitan asegurar la calidad de los datos. (Martinez, 2012).

Una de las labores más esenciales en la administración y el mantenimiento de una base de datos siempre va a ser la optimización de consultas. Cuando se trabaja con bases de datos de gran volumen hasta la tarea más sencilla puede volverse compleja, ya que en vez de retornar una pequeña cantidad de registros podemos hacer que esta crezca exponencialmente por el uso de una consulta inadecuada. (Esteban, 2010).

La inteligencia de negocios se puede aplicar en cualquier campo de estudio donde exista la necesidad de resolver un problema referente a datos y desarrollar soluciones para estas; siendo el área de Marketing donde este conjunto de buenas prácticas ha tenido mayor impacto. (Matute, 2013).

1.2.2 Justificación práctica

El presente trabajo de investigación nos permitirá tener una visión general de las características de los datos que tenemos a disposición en los distintos repositorios habilitados para el área de Inteligencia Comercial del Banco Pichincha, además de tener la documentación necesaria para comprender el flujo de información, lo cual facilitará las tareas de modificaciones y auditoría de datos, así como también optimizar el proceso de ejecución de ETL actualmente implementado en el área de

Inteligencia Comercial, reduciendo así el tiempo y consumo de recursos del servidor.

1.2.3 Justificación tecnológica

Es necesaria la automatización de los procesos ETL dentro de las instituciones bancarias para poder dejar el trabajo manual de ejecución, seguimiento y validación y poder centrar el esfuerzo de los colaboradores en la explotación y análisis de estos datos, para ello la comprensión de los datos y fuentes de información existentes son necesarias para asegurar la integridad de las mismas, junto a las buenas prácticas de optimización de consultas y construcción de ETL robustas capaces de responder a posibles fallos durante sus procesos de ejecución.

1.2.4 Justificación medio ambiental

Como uno de los objetivos del presente trabajo de investigación se busca la reducción de tiempos de ejecución, lo cual afectará también en el tiempo de funcionamiento de los equipos inmersos en este proceso, desembocando en la reducción del uso de energía eléctrica, además, tanto los diccionarios de datos como el resto de la documentación serán entregados de forma digital y no se harán uso de materiales y/o equipos que puedan ser nocivos para la salud o el medio ambiente.

1.3 Delimitación del Proyecto

1.3.1 Teórica

La investigación se enfocará en la aplicación de técnicas de optimización de consultas para mejorar los tiempos de ejecución de los procesos de ETL y reducir el consumo de recursos usados por el servidor.

Según Bobrowski (Bobrowski, Chin-Lee, Closkey, Frazzini & Sokolsky, 1996), la optimización es el proceso durante el cual se decide y

elige la forma más eficiente de llevar a cabo una consulta a una base de datos. Es un importante paso en el procesamiento de cualquier sentencia de manipulación de datos.

1.3.2 Temporal

La presente investigación se llevará a cabo en el periodo de Febrero – Abril del 2019; para efectos de la validación de los resultados obtenidos se utilizará información recuperada de la primera fase de exploración.

La primera fase del proyecto, la Exploración de los datos tendrá una duración de 2 semanas (del 4 al 17 de febrero), mientras que la Construcción del modelo propuesto se llevará en el periodo del 18 de febrero al 10 de marzo, durante 3 semanas, la etapa de Implementación será entre el 11 y 16 de marzo, y la fase final de Validación se llevará a cabo del 17 al 31 de marzo.

1.3.3 Espacial

El Banco Pichincha es una entidad privada con presencia en Ecuador, Colombia, España, Panamá y Perú. Cuenta con una red de 79 oficinas en el Perú, de las cuales, 40 de ellas se encuentran ubicadas en la capital, incluyendo su sede principal.

El lugar de desarrollo del presente trabajo de investigación será la provincia de Lima, departamento de Lima, Perú, en las instalaciones de la oficina principal del Banco Pichincha, la cual se encuentra ubicada en la Av. Ricardo Palma 278, Miraflores.

1.4 Formulación del problema

1.4.1 Problema general

¿De qué manera el mapeo de datos y la optimización de queries permitirán mejorar el rendimiento de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha?

1.4.2 Problemas específicos

- ¿De qué manera el mapeo de datos y la optimización de queries disminuirán la dificultad para realizar modificaciones de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha?
- ¿De qué manera el mapeo de datos y la optimización de queries mejorarán los tiempos de ejecución elevados de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha?
- ¿De qué manera el mapeo de datos y la optimización de queries mitigarán los errores en la carga de información de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha?

1.5 Objetivos

1.5.1 Objetivo general

Realizar el mapeo de datos y la optimización de queries para mejorar el rendimiento de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha.

1.5.2 Objetivos específicos

- Realizar el mapeo de datos y la optimización de queries para disminuir la dificultad al realizar modificaciones en los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha.

- Realizar el mapeo de datos y la optimización de queries para mejorar los tiempos de ejecución elevados de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha.
- Realizar el mapeo de datos y la optimización de queries para mitigar los errores en la carga de información de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha.

CAPÍTULO II: MARCO TEÓRICO

2.1 Antecedentes

En esta sección, se ha revisado la literatura de las últimas investigaciones referente a este tema de investigación y destacan las siguientes que se mencionan a continuación.

2.1.1 Antecedentes internacionales

Según Galvis (Galvis, Gómez, & Bustamante, 2013), su trabajo de investigación tuvo como finalidad la evaluación y selección de un marco de trabajo para el desarrollo del modelo lógico proceso de extracción, transformación y carga.

En el cual, el marco de trabajo seleccionado sería puesto a prueba en la construcción de una solución de BI que apoye en la toma de decisiones y gestión del Proyecto de Círculos de Aprendizaje ejecutado por la Universidad de Magdalena en Colombia.

Como conclusiones del trabajo de investigación obtuvo que:

- El diseño de los procesos ETL es fundamental en la implantación de una solución de Inteligencia de Negocios para asegurar su mantenibilidad, comprensión y extensibilidad.
- Si bien en el trabajo anteriormente mencionado se utilizó el enfoque de modelado basado en objetos y se logró adaptar a las necesidades del proyecto. El autor hace mención que la técnica utilizada está lejos de ser la mejor para su aplicación en todo tipo de proyectos.
- Debido a la naturaleza de los procesos de ETL (entradas, procesos y salidas) una notación útil que combinaría sería el diagrama de clases UML.

La investigación de Frutos (Frutos, 2016), tuvo como finalidad comparar la utilidad y aplicabilidad de dos herramientas para la optimización de consultas, ya que al conocer el impacto de las

operaciones permitiría mejorar las definiciones de las consultas y reducir sus costes.

Como conclusiones de este trabajo de investigación se obtuvo que:

- No hay una solución perfecta, sino que la decisión de las herramientas a usar para optimizar consultas está ligada a un amplio conjunto de factores como el tipo de base de datos, el enfoque de rapidez y disponibilidad, el nivel de conocimiento del usuario y la curva de aprendizaje que esté dispuesto a asumir.
- Algo que influye en la labor de optimizar consultas en las bases de datos es la formación y capacitación constante. Mientras más se deja de lado la actualización en temas de tecnología es mayor el coste que asumirá la organización, ya sea de recursos o el ocasionado por vulnerabilidades.
- El conocimiento de las tablas y datos, el entorno de consulta y otra información que los sistemas no son capaces de recopilar al ejecutar las consultas son un factor importante al momento de elegir los planes de ejecución.

Por otro lado, Méndez (Duque, Hernández, Pérez, Arroyave, & Espinosa, 2016) en su artículo, muestra un modelo con la finalidad de mejorar el rendimiento de procesos de ETL y lograr mayor eficiencia en la tarea de poblar un Data Mart.

Como conclusiones de este trabajo se obtuvo que:

- Sobresale la complejidad en las tareas de ETL durante el proceso de construcción del Data Warehouse (DW), ya sea en costos, tiempo de desarrollo/ejecución y consumo de recursos. Sin embargo, al someter el modelo de ETL planteado a un gran volumen de datos hidrometeorológicos se obtuvieron buenos resultados de rendimiento.

- El modelo fue aplicado en un caso de investigación específico, pero puede ser aplicado a otros dominios sin mayor complejidad tan solo siguiendo los pasos planteados en el trabajo.

Según Guo (Guo, Sun, Yuan, & Yue, 2015), los autores proponen un nuevo enfoque para los procesos de ETL, donde se aplicarían virtual tables (tablas virtuales) para llevar a cabo la etapa de transformación antes de la extracción y carga. Este nuevo enfoque ha sido denominado TEL, ya que cambiaría el orden de las siglas en inglés por Transform – Extract – Load.

De las conclusiones obtenidas en este trabajo de investigación se rescata lo siguiente:

- Reducción de la carga efectuada en el proceso de transmisión de datos.
- Aumenta el rendimiento de las consultas gracias al uso de capas de acceso.
- Es denominado por los autores como un enfoque práctico y factible.

Graneros (Graneros, 2017) busca analizar y comparar los resultados de la ejecución de distintas consultas con diferente complejidad en dos motores de base de datos distintos (MSSQL Server y MySQL), para efectos de su investigación hizo uso de un ambiente dedicado donde comparó cuál de estos motores se desempeña mejor.

Las pruebas realizadas sin indexación de las tablas arrojaron que las cláusulas in, distinct y exists, tardan menos tiempo y generan menos costo computacional al ser ejecutadas en la base de datos MySQL, mientras que para ejecutar consultas como left join, inner join y producto cartesiano, la mejor opción es MSSQL Server; tal como se observa en la Tabla 1 a continuación.

Tabla 1 Tiempo de ejecución por sentencia en tablas sin índices

Cláusula	Tiempo (s)	
	MSSQL	MySQL
In	247.03	66.48
In con Distinct	241.26	66.81
Exists	235.98	77.92
Left Join	75.2	3873.97
Producto Cartesiano	111.01	3939.59
Inner Join	109.23	3901.5

Fuente: Elaboración propia

Tras haber realizado la indexación de las tablas se realizaron las comparaciones de las consultas nuevamente, los resultados se muestran en la Tabla 2, ubicada a continuación.

Tabla 2 Tiempo de ejecución por sentencia en tablas con índices

Cláusula	Tiempo (s)	
	MSSQL	MySQL
In	18.4166	4.75
In con Distinct	18.4377	4.54
Exists	18.3757	9.13
Left Join	16.3063	4.46
Producto Cartesiano	15.9667	39.85
Inner Join	14.513	39.86

Fuente: Elaboración propia

Cabe resaltar que, como conclusión el autor menciona que la indexación de las tablas mejora el rendimiento y tiempo de ejecución de las consultas; sin embargo, ralentiza las operaciones de insert, update y delete, ya que por cada una de estas se debe ir actualizando el índice.

2.1.2 Antecedente nacional

En caso de las investigaciones nacionales, se destacó el trabajo de investigación de Salazar (Salazar, 2017); tuvo por finalidad la implementación de un Data Mart que apoye al área comercial de la empresa Azaleia para mejorar el proceso de extracción de información de los usuarios, reduciendo la carga operativa y mejorando la gestión de los indicadores.

Como conclusiones de este trabajo de investigación se obtuvo que:

- Es necesario tomar en consideración las características y capacidades del motor de base de datos con el fin de asegurar el rendimiento en el tiempo de respuesta para los usuarios.
- El volumen de información existente en las bases es mandatorio para el diseño de la arquitectura utilizada en el Data Mart, por tanto, también en los procesos de ETL.
- El tiempo de integración de datos y generación de reportes se vio disminuido drásticamente, lo que permite brindar más tiempo para el análisis de los indicadores.

2.2 Bases teóricas

2.2.1 Lenguaje de consulta estructurado

El lenguaje de consulta estructurado o Structured Query Language (SQL) es el que permite la comunicación con el sistema gestor de base de datos como SQL Server, Oracle, etc., estas se dividen en lenguaje de definición de datos o Data Definition Language (DDL), lenguaje de manipulación de datos o Data Manipulation Language (DML) y lenguaje de control de datos o Data Control Language (DCL). (Rodríguez, 2008).

En la Figura 1 se muestran las sentencias agrupadas por los tipos de consultas de datos mencionados en el párrafo anterior.

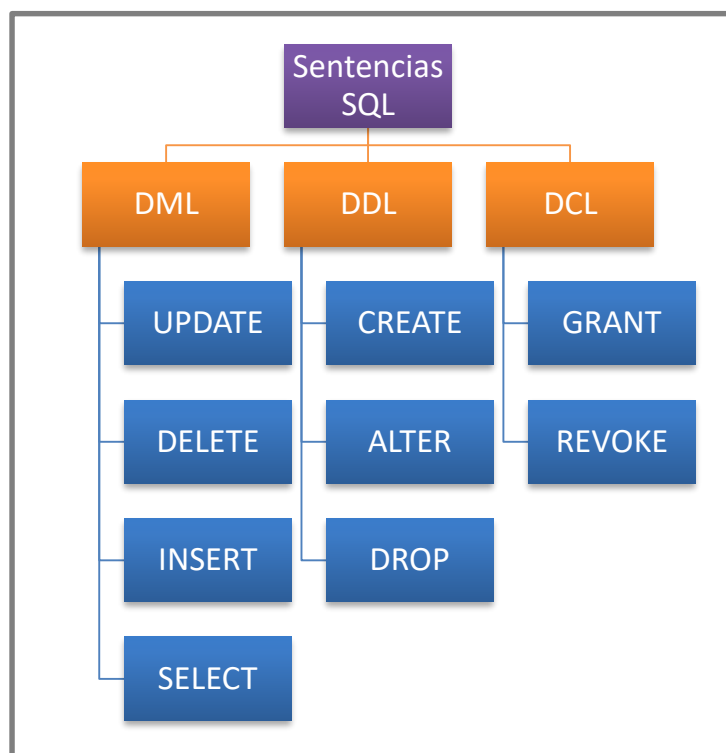


Figura 1 Tipos de declaraciones de sentencias SQL

Fuente: Elaboración propia

2.2.1.1 Lenguaje de Definición de Datos

El lenguaje de definición de datos o Data Definition Language (DDL), incluye aquellos comandos que son utilizados

para la creación de una base de datos y sus distintos componentes como: índices, tablas, procedimientos almacenados, triggers, etc. (Reyes, 2002).

Algunas de las sentencias que comprende son:

- Sentencia Create, permite la creación de un objeto SQL dentro de un sistema gestor de bases de datos relacionales (RDBMS).
- Sentencia Drop, es utilizada para eliminar de forma rápida los objetos de un sistema gestor de bases de datos (SGBD); por ejemplo, índices, tablas y bases de datos.
- Sentencia Alter, esta sentencia es utilizada para modificar las definiciones de objetos del SGBD anteriormente especificados.

2.2.1.2 Lenguaje de Manipulación de Datos

El lenguaje de manipulación de datos o Data Manipulation Language (DML), incluyen aquellas sentencias usadas para modificar, borrar, insertar y consultar los datos almacenados en una base de datos, tabla u otro objeto. (Reyes, 2002).

Algunas de las sentencias que comprende son:

- Sentencia Select, permite consultar los datos existentes en una tabla.
- Sentencia Insert, permite la inserción de nuevos registros en una tabla.
- Sentencia Delete, es utilizada para eliminar los registros de una tabla, permite especificar los datos a eliminar.
- Sentencia Update, es utilizada para modificar todos los datos de la tabla o de forma específica acorde a las necesidades.

2.2.1.3 Lenguaje de Control de Datos

El lenguaje de control de datos o Data Control Language (DCL), es proporcionado por el SGBD para administrar el acceso a los datos de una base de datos.

Algunas de las sentencias que comprenden son:

- Sentencia Grant, tiene por finalidad el brindar accesos a uno varios usuarios o roles sobre tareas específicas como son select, insert, update, delete, etc.
- Sentencia Revoke, es utilizada para eliminar o quitar los permisos otorgados por la sentencia grant.

2.2.2 Inteligencia de Negocios

El término inteligencia de negocios o Business Intelligence (BI) hace referencia al almacenamiento y procesamiento de grandes volúmenes de datos con el fin de transformarlos en información y conocimiento que permita apoyar en la toma de decisiones en tiempo real. (Bernabeu, 2010).

Para Vitt (Vitt, Luckevich, & Mismar, 2002) es utilizado por desarrolladores de software y expertos para hacer referencia a un conjunto de tecnologías, programas, aplicaciones y procesos.

BI tiene como objetivo principal ser un soporte en la toma de decisiones para mejorar el rendimiento de las compañías y ser una ventaja competitiva en el mercado, brindando la facilidad de responder a los cambios con mayor rapidez.

En la Figura 2 mostrada a continuación se presentan las etapas de transformación de los datos para llegar a la toma de decisiones.



Figura 2 Transformación de los datos para toma de decisiones

Fuente: (Estevas, 2012)

2.2.3 Proceso de Extracción, Transformación y Carga

Los procesos de extracción, transformación y carga o ETL (Extract, Transform and Load), son aquellos que permiten la lectura de datos desde distintas fuentes, realizar transformaciones para asegurar su calidad e integrarlos para luego ser depositadas en un destino. Estos procesos constan de tres fases como se aprecia en la Figura 3:

- Fase de Extracción.
- Fase de Transformación.
- Fase de Carga

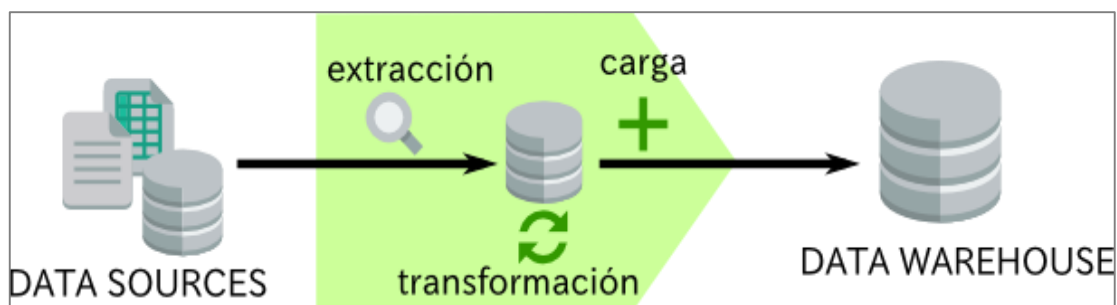


Figura 3 Proceso ETL

Fuente: (Bernabeu, 2010)

2.2.3.1 Fase de Extracción

Durante esta etapa se busca obtener los datos desde los sistemas o bases origen, los cuales tras pasar por una revisión serán preparados para iniciar el proceso de transformación. Una consideración importante tomada en esta fase es evitar el bloqueo de la base o sistema de donde se originan los datos, sin afectar su uso cotidiano. (POWERDATA SOLUTION S.L., 2017).

Para Bernabeu (Bernabeu, 2010) el hacer uso de un almacenamiento intermedio brinda grandes ventajas como evitar la dependencia de los procesos transaccionales en línea (OLTP), aumenta la facilidad de integrar distintas fuentes y permitir gestionar los metadatos generados por el ETL.

2.2.3.2 Fase de Transformación

Durante esta etapa se aplican las reglas del negocio o funciones que permitan asegurar la calidad de los datos y convertirlos en datos limpios que aporten valor al momento de ser cargados. Las reglas aplicadas en esta etapa deben ser claras, independientes, inteligibles, declarativas y sobre todo deben de ser útiles para el negocio. (POWERDATA SOLUTION S.L., 2017).

2.2.3.3 Fase de Carga

Durante esta última parte del proceso ETL, los datos provenientes de la fase de transformación son almacenados en la base de datos destino. Dos de las formas más básicas para realizar este proceso son:

- Acumulación Simple, incluye la carga de resúmenes de datos, considerada la forma más sencilla y común de realizar esta fase.
- Rolling, se refiere a la carga de datos en distintos niveles de granularidad, es decir, distintos niveles de agrupación. (POWERDATA SOLUTION S.L., 2017)

Bernabeu (Bernabeu, 2010) hace mención de dos tareas en las que se divide este proceso; por un lado se tiene la carga inicial y por el otro, la actualización o carga periódica, donde esta última se encarga de refrescar la base con los últimos datos agregados.

2.2.4 Servicio de Integración de SQL Server

El servicio de integración de SQL Server o SQL Server Integration Service (SSIS) es una herramienta de integración y transformación de datos, permite construir soluciones empresariales sin la necesidad de tener conocimiento de lenguajes de programación.

Acorde a la documentación de Microsoft (Microsoft, 2017), SSIS permite a los usuarios extraer datos desde distintos orígenes y facilitar su descarga o almacén en uno o muchos destinos, además de facilitar las tareas de limpieza, transformaciones y minería.

La arquitectura del servicio de integración de SQL Server está ideada para separar la administración de paquetes y el flujo de control de la transformación y el movimiento de los datos. Esta arquitectura es definida por dos motores diferentes, automatizables y extensibles al configurar SSIS. (Microsoft, 2017).

2.2.4.1 Motor en tiempo de ejecución de SSIS

Es el encargado de administrar y ejecutar los paquetes, esto incluye el control de eventos, variables, registros y orden de ejecución. Permite automatizar las tareas de crear, configurar y ejecutar paquetes, así mismo el desarrollo de tareas personalizadas. (Microsoft, 2017).

2.2.4.2 Motor de flujo de datos de SSIS

Administra la tarea de flujo de datos (Data Flow Task), la cual se especializa en la recolección, transformación y depósito de datos de diferentes orígenes. Está compuesta por componentes adicionales que pueden ser de origen, para transformación o carga en uno o varios destinos. La programación del motor de flujo de datos permite a los desarrolladores la construcción de componentes especializados. (Microsoft, 2017).

En la Figura 4 se puede apreciar la arquitectura del servicio de integración de Microsoft, mostrando a detalle los componentes de cada motor mencionados en los puntos 2.2.4.1 y 2.2.4.2.

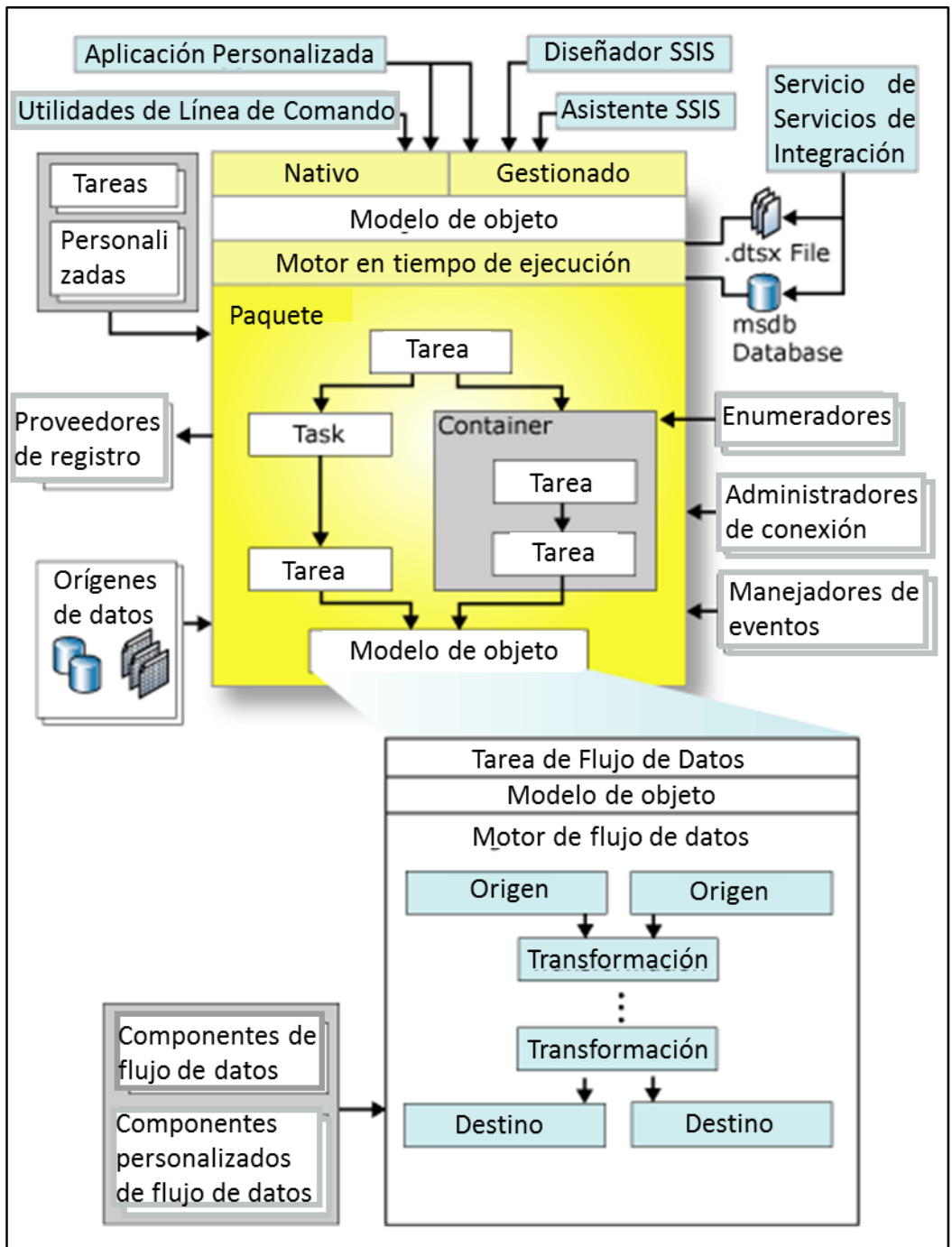


Figura 4 Arquitectura SSIS

Fuente: (Microsoft, 2017)

2.2.5 Inteligencia Comercial

Según Gómez (Gómez, 2018) la inteligencia comercial es el conjunto de aptitudes y habilidades intelectuales que permite a una persona comprender y comunicar la visión comercial de una idea. Sin embargo, al referirnos al lado comercial de una idea no debe interpretarse solo como una medida económica, sino también en la influencia que tendrá sobre las personas que la consumirán.

Investigación de mercado no es lo mismo que Inteligencia Comercial, ya que esta última es capaz de integrar todas las fuentes de información que existen dentro y fuera de la empresa, siendo la investigación de mercado una herramienta para obtener los datos necesarios y que no se encuentran en otras fuentes. (Deniz Miguez, 2003).

2.2.6 Banco Pichincha Perú

El Banco Pichincha Perú es una empresa privada del sector bancario, multiproducto y orientado a los segmentos de mediana empresa, microfinanzas y banca persona; cuenta con 79 agencias a nivel nacional. Pertenece al grupo Pichincha y fue fundado en Ecuador en 1906. Siendo la mayor institución bancaria en el país vecino, recibe su nombre por el volcán Pichincha, el cual a sus pies tiene a la ciudad de Quito. (Banco Pichincha, 2019).

Dentro de los productos que ofrece el Banco Pichincha encontramos ahorros y depósitos, seguros, tarjetas, créditos y otros servicios a la medida. Siendo los préstamos de libre disponibilidad quienes destacan dentro de los productos crediticios, ya que están orientados a cubrir necesidades de consumo exclusivamente, además que el cliente puede garantizar su crédito mediante un depósito a plazo, dando lugar a lo que se conoce como PLD con garantía líquida; en caso el cliente desee garantizar su crédito con una propiedad física a través de una

hipoteca, el producto pasaría a ser un PLD con garantía hipotecaria.
(Banco Pichincha, 2019).

2.3 Definición de Términos Básicos

Archivos Planos. Son aquellos archivos que están compuestos solo de texto, sin formato, por ejemplo, archivos con extensión .csv, .txt y .xlsx.

Auditoría de base de datos. Proceso orientado a la mediación y monitoreo de los niveles de acceso de los usuarios a las bases de datos corporativas.

Base de Datos. Es un repositorio que permite guardar grandes volúmenes de datos de forma organizada y mantenerlos disponibles para su posterior uso.

Business Intelligence Development Studio 2008. Herramienta de la suite de SQL Server 2008 R2, permite el desarrollo y construcción de los procesos ETL.

Calidad de Datos. Procesos, técnicas y operaciones que se encargan de asegurar la fiabilidad de los datos en las empresas y/o instituciones.

Data Mart (DM). Hace referencia a una base de datos departamental, especializada en un área de negocio específica.

Data Warehouse (DW). Es una base de datos a nivel corporativo, ya que está compuesto por un conjunto de Data Mart. Contiene información centralizada de todas las áreas de negocio y procesos de la empresa.

Dato. Representación de una variable cualitativa o cuantitativa.

DB2. Es un sistema gestor de bases de datos relacionales diseñado por IBM, es considerado un estándar orientado para plataformas de sistemas mandatorios.

Diagrama de Clases UML. Es un diagrama que describe la estructura de un sistema y muestra sus clases, características y operaciones, además de las relaciones entre objetos.

Eventos. Sucesos que ocurren en tiempo de ejecución de paquetes del SSIS, desencadenados por tareas o contenedores.

Indicadores. Datos que permiten medir de forma cuantitativa un hecho y valorar sus características evolutivas.

Índices. Es un objeto de base de datos que mejora el tiempo de consulta al crear un identificador a cada fila de una tabla, permitiendo un rápido acceso a los registros.

Información. Conjunto organizado de datos procesados pertenecientes a un mismo contexto que constituyen un mensaje.

Inteligencia de Negocios. Del inglés Business Intelligence (BI); es un conjunto de habilidades que permiten generar información a partir de los datos y luego transformarla en conocimiento, para mejorar el proceso de toma de decisiones.

Framework. Es un marco de trabajo conceptual, desarrollado bajo un estándar para enfocar un tipo de problemática.

Mapeo De Datos. Es el proceso de establecer relaciones entre los modelos de datos separados

Marketing. Es el conjunto de técnicas y estudios con el objeto de entender el comportamiento de un producto y su comercialización.

Metadatos. Son información que caracteriza a los datos producidos; como su calidad, historia, disponibilidad y descripciones del contenido.

Modelo Lógico. Es una representación gráfica de entidades que contiene relaciones, atributos y restricciones entre estos.

MySQL. Es un sistema gestor de base de datos relacionales de código libre desarrollado por Oracle y basado en el lenguaje SQL.

OLTP. Procesamiento de Transacciones en Línea, es un tipo de procesamiento que facilita aplicaciones que usualmente trabajan con ingreso y recuperación de información.

Optimización. Proceso por el cual se busca reducir el tiempo de ejecución de una consulta y mejorar el rendimiento de la misma.

Oracle database. Es un sistema gestor de base de datos objeto – relacional, desarrollado por la corporación Oracle.

Paquete. Es la unidad de trabajo del servicio de integración de SQL, contiene los elementos de conexión, flujos de datos, eventos, variables, etc., ensamblados con la plataforma gráfica SSIS o mediante programación.

Planes de ejecución. Es una estimación del comportamiento de una consulta si esta fuese ejecutada para probar su desempeño.

PLD. Los préstamos de Libre Disponibilidad son créditos orientados al consumo personal, no para inversiones.

Procedimientos almacenados. Es un conjunto de instrucciones T-SQL que pueden ser ejecutadas cuando sean necesarias.

Proceso. Conjunto de tareas ordenadas e interconectadas que tienen un inicio y fin.

Querys. Consultas que se realizan en un lenguaje de programación estructurado con la finalidad de manipular los datos u objetos en una base de datos.

Registro. Es un conjunto de campos que contienen datos de una entidad, también es conocido como fila o tupla.

Servidor. Es un ordenador, que cumple el rol de suministrar la información a una serie de usuarios dentro de la organización, o a clientes fuera de esta.

SGBD. Son un conjunto de programas que permiten la manipulación de la información contenida en una base de datos.

Sistemas de información. Conjunto de elementos orientados al tratamiento de datos, que tienen una entrada, proceso, salida y retroalimentación.

Software. Es la parte lógica de una computadora, el conjunto de programas que le permiten ejecutar ciertas tareas.

SQL Server. Es un sistema gestor de bases de datos relacionales diseñado por Microsoft, brinda un entorno visual para la administración y consulta de bases de datos.

Tablas. Objetos de base de datos que guardan toda su información, los datos son organizados en filas (registros) y columnas (campos).

Tablas temporales. Son objetos de una base de datos que guardan información durante un tiempo limitado, luego de su uso o su ejecución pasar a ser eliminadas de forma automática por el SGBD.

TEL. Variante del proceso de extracción, carga y transformación en el que el proceso de carga es llevado a cabo antes que el de extracción para aminorar su tiempo de carga en los depósitos finales.

Triggers. Es una clase especial de procedimiento almacenado que se ejecuta al producirse un evento específico en la base de datos.

UML. El Lenguaje de Modelado Unificado es un lenguaje de modelado visual para la arquitectura, diseño e implementación de softwares.

Variable. Son objetos susceptibles de ser modificados, varían en función de una determinada condición o cálculo.

Virtual tables. Son vistas de consultas entre tablas que no se almacenan de forma física en la base de datos.

CAPÍTULO III: DESARROLLO DEL TRABAJO DE SUFICIENCIA PROFESIONAL

El presente trabajo de investigación utiliza un marco de desarrollo similar al de un proyecto de Inteligencia de Negocios, por lo cual estará comprendido por etapas similares, así como se muestra en la Figura 5.



Figura 5 Etapas del trabajo de investigación

Fuente: Elaboración propia

3.1 Análisis Exploratorio Situacional

3.1.1 Mapeo de Fuentes

El proceso de extracción, transformación y carga de los PLD es alimentado por fuentes internas del banco, provenientes de dos servidores de base de datos SQL Server; de la revisión del proceso ETL se pudieron rescatar las tablas mostradas en la Figura 6 y Figura 7.

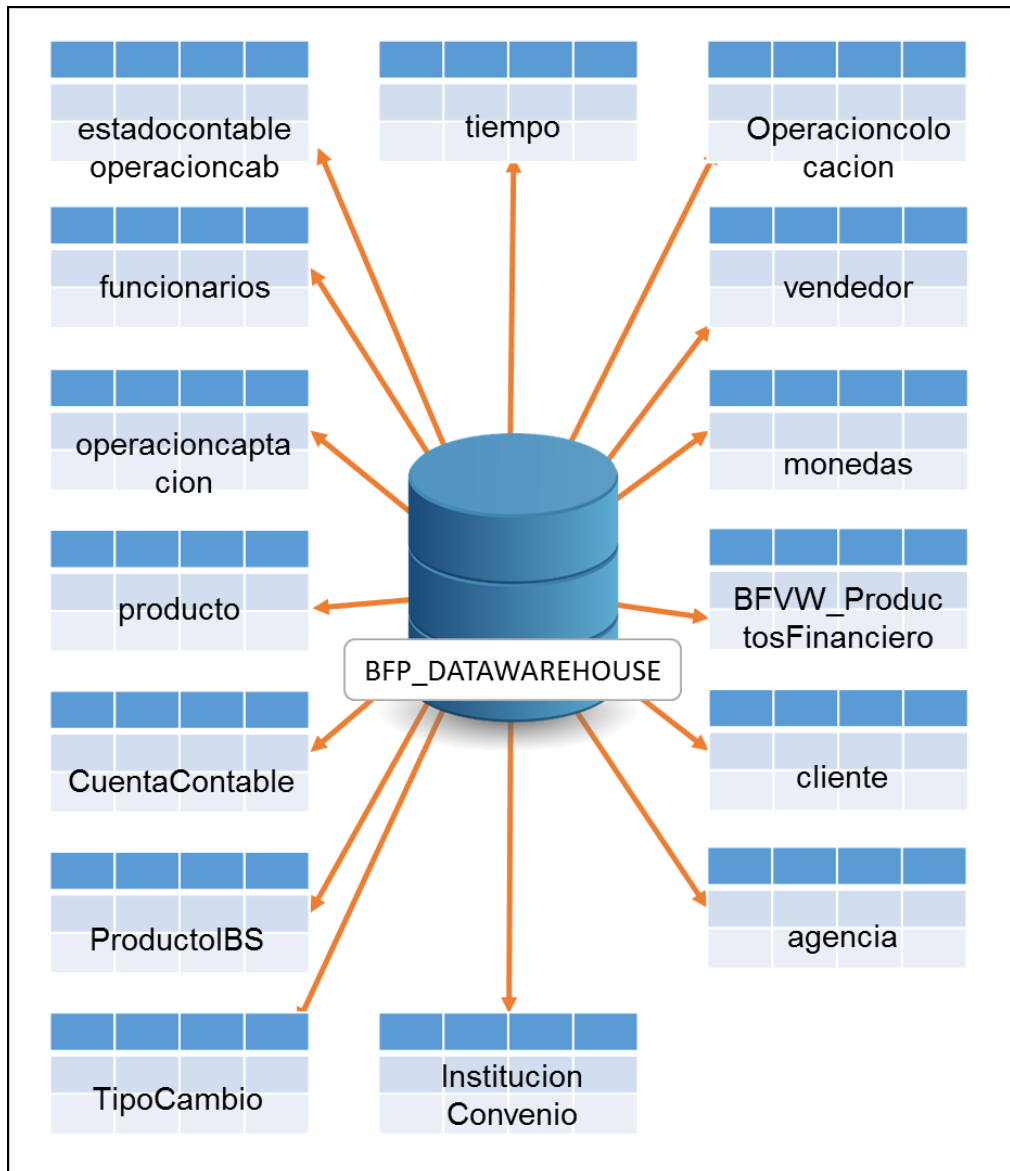


Figura 6 Fuentes Servidor SBIBFPR03

Fuente: Elaboración propia

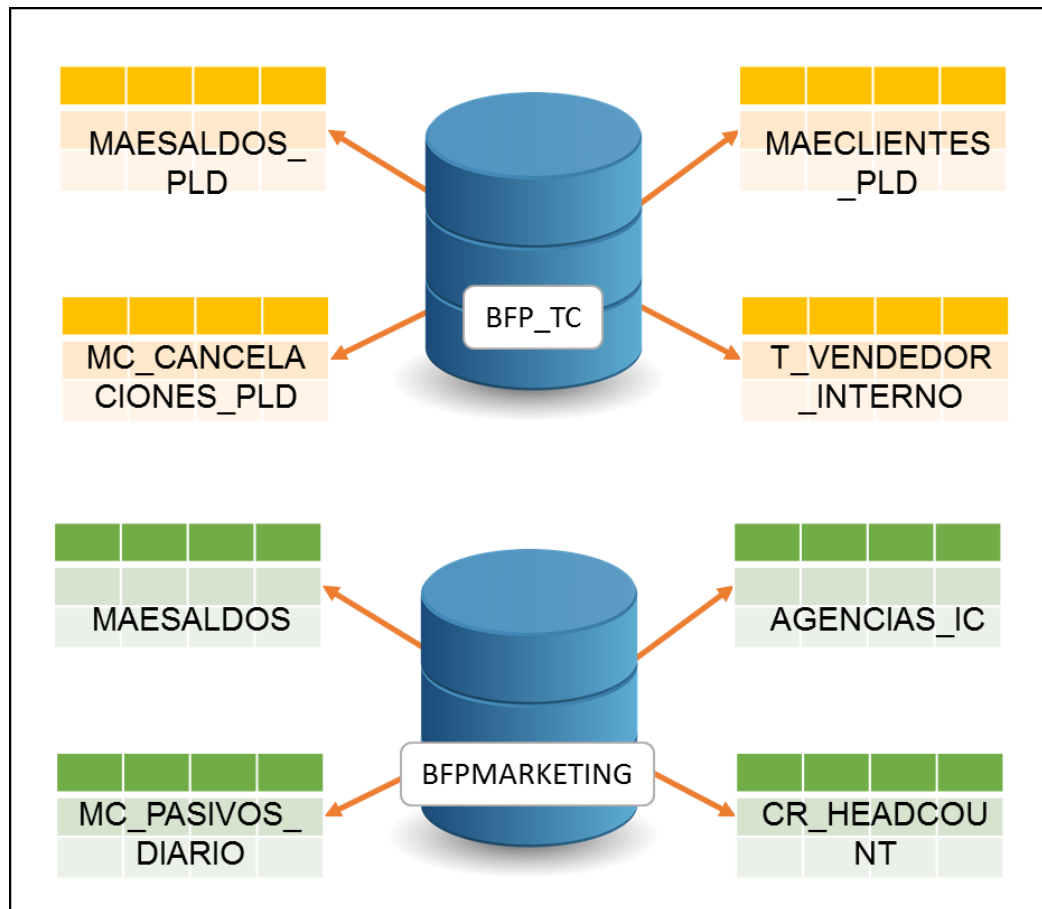


Figura 7 Fuentes Servidor SBDBFPR15

Fuente: Elaboración propia

Cabe mencionar que las tablas pertenecientes al servidor SBIBFPR03 que se muestran en la Figura 6 son externas al área de Inteligencia Comercial, por lo que su administración y actualización se encuentran fuera de nuestras capacidades y por tanto tenemos una dependencia directa que en algunos casos puede ser perjudicial por la hora de inicio de ejecución de nuestros procesos.

3.1.2 Descripción del proceso de ETL actual

El proceso de ETL utilizado para la generación de la base con información referente a los Préstamos de Libre Disponibilidad está comprendido por sub procesos de actualización de información como son: validación de activos, carga temporales, actualizar correcciones, stock actual, saldos retenidos y data adicional; y dos sub procesos que apoyan en la generación de los reportes: reporte PLD y desembolsos Hamber, estos últimos descritos mas no evaluados ya que sus rendimientos están ligados netamente a la capacidad de procesamiento del equipo en donde se ejecute la ETL, el cual puede variar drásticamente al encontrarse en un servidor de producción a diferencia de una computadora personal de uso de oficina.

En la Figura 8 se puede apreciar la estructura y el diseño de la ETL antes mencionada.

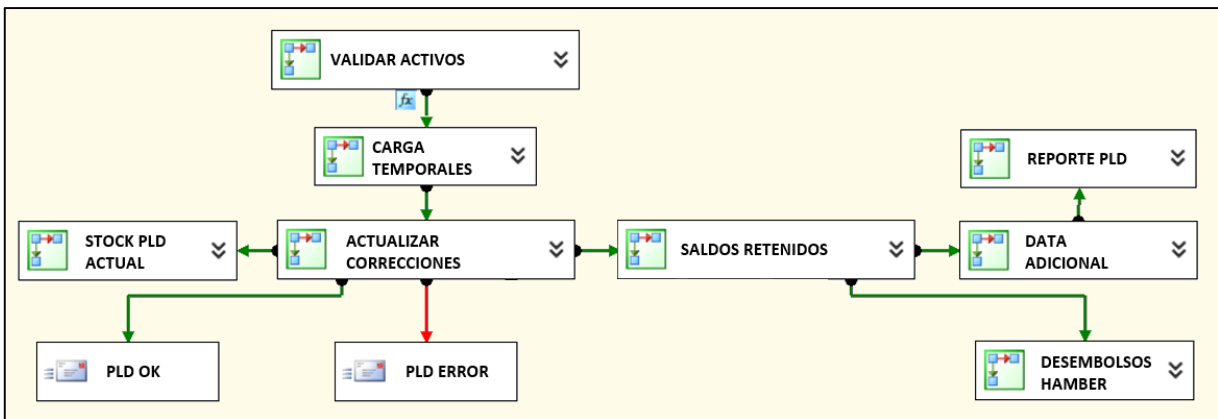


Figura 8 Estructura ETL actual

Fuente: Elaboración Propia

Los subprocesos de la ETL se encuentran agrupados por Sequence Container, los cuales son subconjuntos del flujo de control principal, estos pueden incluir distintas tareas y pueden estar ordenados de forma secuencial o paralela.

A continuación, se describirá la funcionalidad de cada uno de los contenedores, indicando también su estructura interna y las tareas que lo componen.

3.1.2.1 Validar Activos

Este contenedor inicia con el envío de un correo electrónico indicando el inicio del proceso ETL, para luego continuar con el cálculo de la fecha a la que se va a actualizar la información y la recuperación de la cantidad de registros existentes en las fuentes, este último es usado para condicionar la ejecución de la siguiente tarea y contenedor, ya que de no existir registros el proceso envía un mensaje de alerta mencionando que no hay información disponible en las fuentes y además detiene la ejecución del proceso; la estructura se detalla a continuación en la Figura 9.

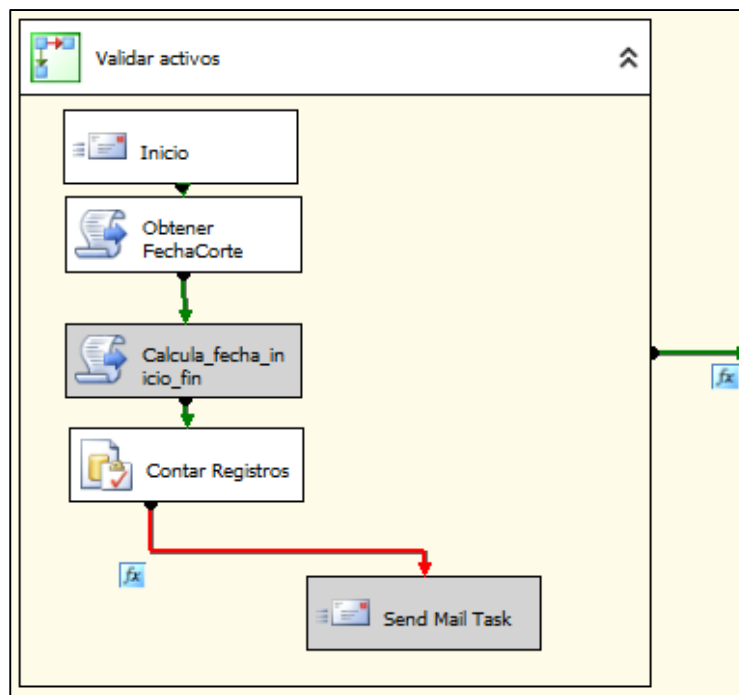


Figura 9 Validar Activos (Sequence Container)

Fuente: Elaboración propia

3.1.2.2 Carga Temporales

Durante este subproceso se realiza la extracción y transformación de los datos en las fuentes del servidor SBIBFPR03 para generar las tablas intermedias que más adelante serán consolidadas en la tabla final. Las tablas intermedias son cargadas en el servidor SBDBFPR15 y contienen información referente a los saldos de los créditos vigentes, los clientes y los créditos cancelados de forma histórica.

La estructura interna de este contenedor está compuesta por tareas de ejecución de consultas T-SQL y por tareas de flujo de datos, son detalladas en la Figura 10.

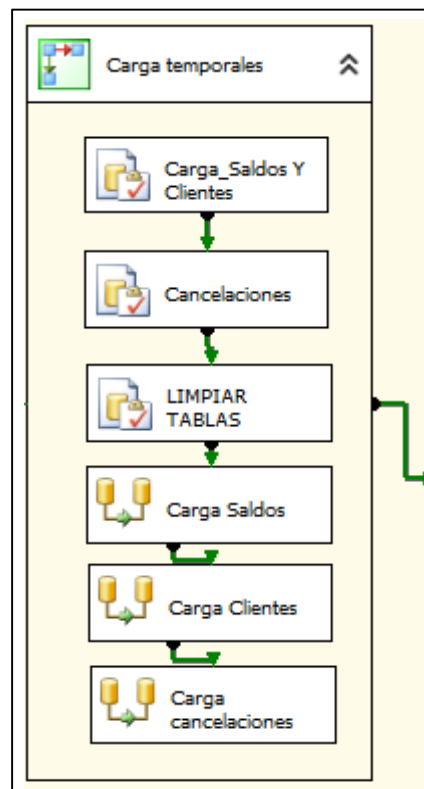


Figura 10 Carga Temporales (Sequence Container)

Fuente: Elaboración propia

3.1.2.3 Actualizar Correcciones (Sequence Container)

Durante esta etapa del proceso se realiza el volcado de la información existente en las tablas intermedias generadas en el contenedor anterior, aplicando las lógicas del negocio acordadas con otros equipos del banco. La correcta ejecución de este subproceso permite el paso a la ejecución de otros procesos y el envío de un correo informativo a usuarios que consumen esta tabla indicando que ya se encuentra disponible con la información actualizada, en caso hubiese ocurrido algún error durante su ejecución enviará una alerta para proceder con su revisión y ejecución desde el punto donde ocurrió el error.

En la Figura 11 se puede apreciar la estructura interna de las tareas inmersas en el contenedor.

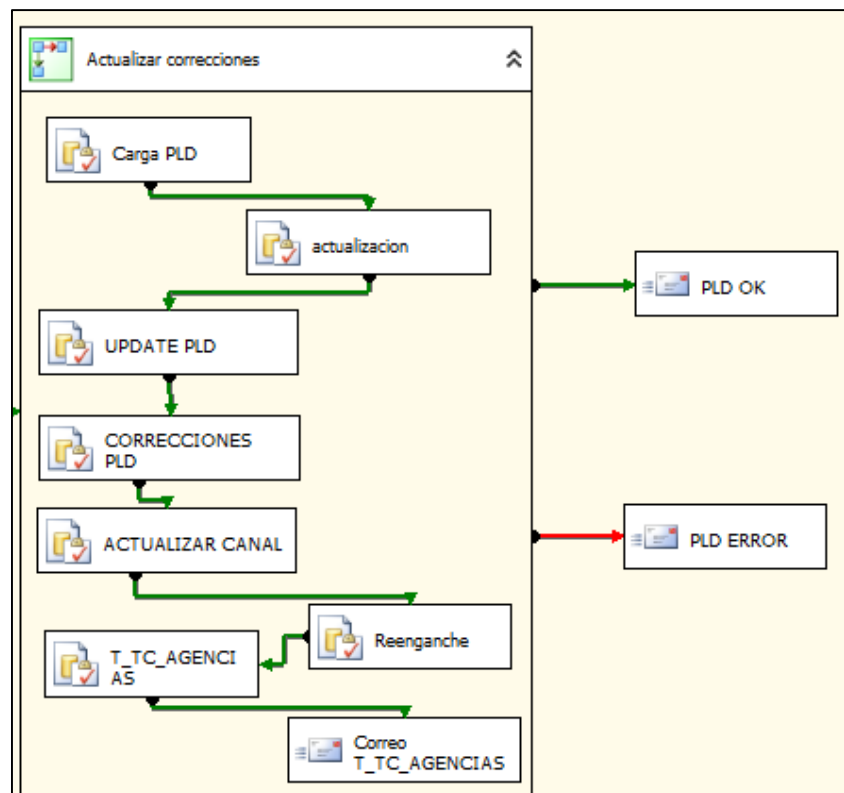


Figura 11 Actualizar Correcciones (Sequence Container)

Fuente: Elaboración propia

3.1.2.4 Stock Actual

La función de este contenedor es realizar el cálculo del stock actual de Préstamos de Libre Disponibilidad vigentes y su envío al equipo de Banca Empresa, quienes utilizarán esta información como más les convenga.

A diferencia de otros contenedores en los que se utiliza la tarea de envío de correos, en este se optó por la tarea de script utilizando Visual Basic para poder enviar un mensaje personalizado y que tenga un mejor performance.

Este subproceso está compuesto por dos tareas secuenciales: el cálculo del stock y el envío de la información, las cuales se muestran a continuación en la Figura 12.

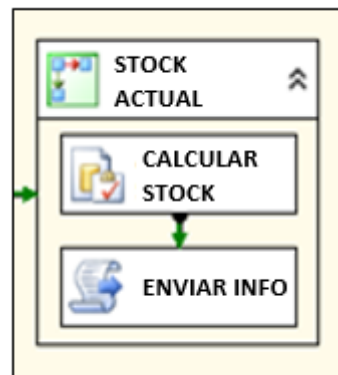


Figura 12 Stock Actual (Sequence Container)

Fuente: Elaboración propia

3.1.2.5 Saldos Retenidos

Durante este contenedor se realiza la limpieza de las tablas donde se almacenará la información de aquellos créditos que han sido aprobados y aceptados por cliente pero que aún no han sido desembolsados por falta de firma del cronograma de pagos.

El proceso es llevado a cabo para ver los créditos en esta condición en el mes actual como en el mes anterior, no se realiza la búsqueda de créditos retenidos en un lapso de tiempo mayor ya que por lo general el tiempo de vida de estos no es mayor a dos semanas.

La Figura 13 muestra la arquitectura actual de este sub proceso, así como el orden secuencial de las tareas que lo conforman.

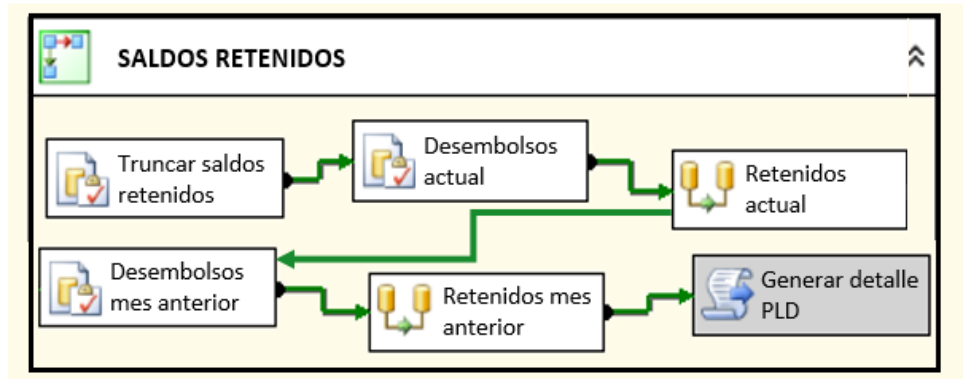


Figura 13 Saldos Retenidos (Sequence Container)

Fuente: Elaboración propia

3.1.2.6 Desembolsos Hamber

Las tareas inmersas en este contenedor ayudan a actualizar los reportes con información de los desembolsos del mes, que es compartido a una lista de usuarios finales.

La estructura de las tareas de este contenedor se puede visualizar en la Figura 14.

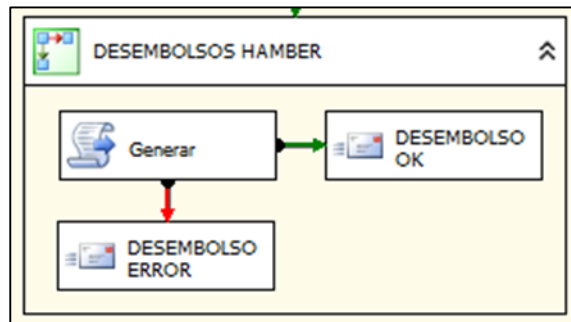


Figura 14 Desembolsos Hamber (Sequence Container)

Fuente: Elaboración Propia

3.1.2.7 Data Adicional (Sequence Container)

Esta etapa de la ETL se utiliza para calcular información puntual solicitada por otros equipos de trabajo. Las tareas en este contenedor trabajan de forma paralela ya que no son dependientes.

Las tareas que lo componen se aprecia en la Figura 15.



Figura 15 Data Adicional (Sequence Container)

Fuente: Elaboración propia

3.1.2.8 Reporte PLD (Sequence Container)

Durante esta fase se accede al reporte de avance de PLD y haciendo uso de macros se refresca la información y se generan los reportes diarios, que son almacenados de forma local como también en un servidor de archivos. Por último, se envía un correo que indica el estado de la actualización del reporte, que puede ser correctamente ejecutado o error durante su actualización.

La estructura de las tareas de este contenedor se puede ver en la Figura 16.

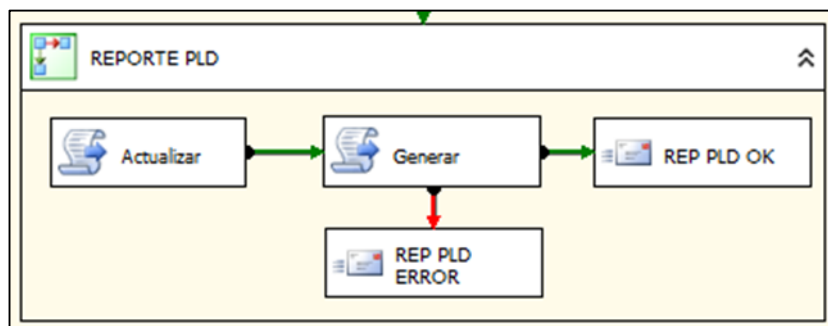


Figura 16 Reporte PLD (Sequence Container)

Fuente: Elaboración propia

3.1.3 Medición de los Tiempos ETL Actual

Para determinar los cuellos de botella dentro del proceso ETL antes descrito se procedió a realizar la medición de tiempos de ejecución entre cada tarea para luego poder identificar el contenedor y tarea que puede ser optimizada, recordando la ley de Pareto se espera que el 20% de las tareas consuman el 80% del total de tiempo de ejecución.

El análisis de los tiempos de ejecución será llevado a cabo en dos partes, en la primera se realizará la búsqueda a nivel de contenedor para identificar en cuál de estos encontramos las tareas con mayor necesidad de mejora, en la segunda nos enfocaremos en las tareas que forman parte del contenedor, buscando encontrar aquella o aquellas que generen los picos en tiempo de ejecución; cabe indicar que como se mencionó en un inicio no se evaluarán los contenedores con los procesos de actualización de los reportes ya que su ejecución está ligada directamente al equipo de cómputo donde sea ejecutada la ETL.

3.1.3.1 Análisis a nivel contenedor.

Se procedió a evaluar el tiempo de ejecución de cada contenedor con el fin de encontrar el máximo y proceder en el siguiente paso con su análisis a nivel de tarea.

Esta primera búsqueda arrojó los resultados (en minutos) expuestos en la Tabla 3, Figura 17 y Tabla 4.

Tabla 3 Tiempo de ejecución diario por contenedor

DURACION (min)	Contenedores						
Fecha de Ejecución	ACTUALIZAR CORRECCIONES	CARGA TEMPORALES	DATA ADICIONAL	SALDOS RETENIDOS	VALIDAR ACTIVOS	Total general	
04/02/2019	50.8	12.5	1.5	3.4	0.4	68.6	
05/02/2019	88.1	15.6	4.8	3.7	0.4	112.6	
06/02/2019	49.6	14.6	4.5	7.0	0.4	76.1	
07/02/2019	41.3	13.2	6.0	12.3	2.1	74.9	
08/02/2019	37.6	20.3	2.0	5.3	2.1	67.3	
09/02/2019	41.6	20.9	2.1	5.4	0.9	70.9	
10/02/2019	54.3	15.5	2.5	5.9	0.9	79.1	
11/02/2019	92.6	18.6	6.2	7.2	0.9	125.5	
12/02/2019	54.5	18.8	2.5	7.1	1.1	84.0	
13/02/2019	55.7	16.7	3.9	7.4	1.1	84.8	
14/02/2019	96.5	22.8	7.2	9.7	1.6	137.8	
15/02/2019	58.0	21.8	3.5	9.6	1.6	94.5	
16/02/2019	59.2	19.7	3.9	9.4	1.6	93.8	
17/02/2019	94.1	22.8	5.6	5.7	1.6	129.8	

Fuente: Elaboración propia

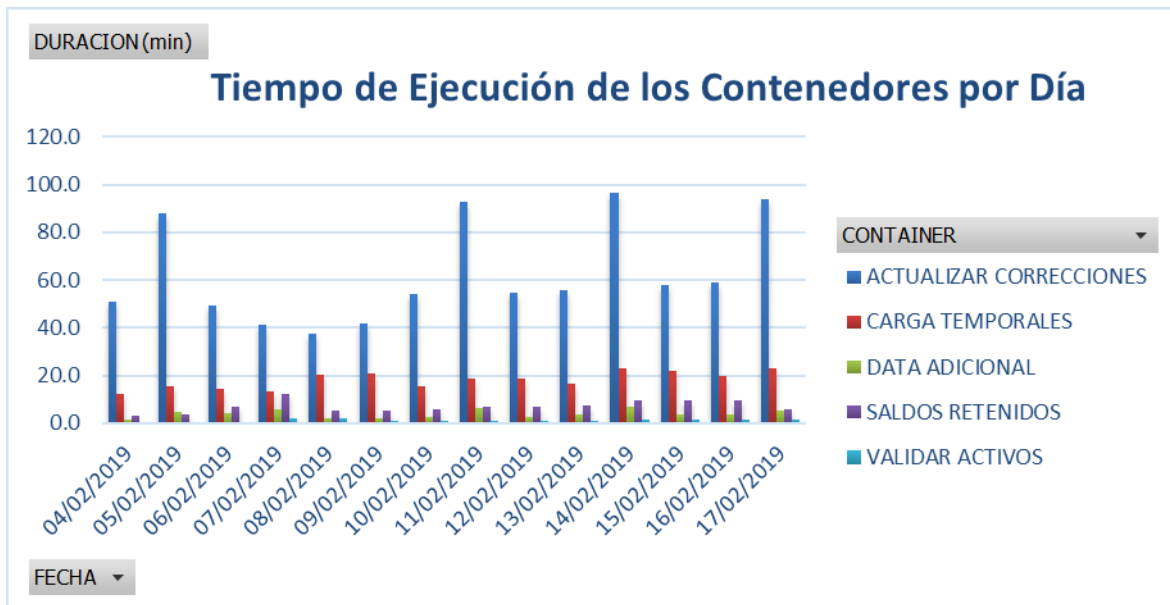


Figura 17 Tiempo de Ejecución de los Contenedores por Día (min)

Fuente: Elaboración propia

Tabla 4 Análisis de tiempos de ejecución por contenedor

	MIN	MAX	PROM
ACTUALIZAR CORRECCIONES	37.6	96.5	62.4
CARGA TEMPORALES	12.5	22.8	18.1
DATA ADICIONAL	1.5	7.2	4.0
SALDOS RETENIDOS	3.4	12.3	7.1
VALIDAR ACTIVOS	0.4	2.1	1.2

Fuente: Elaboración propia

De estos resultados se pudo inferir que los picos de tiempo de procesamiento se dan en primer lugar en el contenedor Actualizar Correcciones, con un tiempo promedio de ejecución de 62.4 minutos y el contenedor Carga Temporales, con un tiempo promedio de ejecución de 18.1 minutos.

3.1.3.2 Análisis a nivel de tarea

Se procedió a continuación a analizar por separado los contenedores antes mencionados con la intención de encontrar la tarea que tarda más tiempo en ejecutarse o si todas las tareas tienen un tiempo de ejecución similar.

Para el contenedor Actualizar Correcciones se obtuvieron los resultados expuestos en la Tabla 5, Figura 18 y Tabla 6.

Tabla 5 Tiempos de ejecución diarios del contenedor actualizar correcciones

DURACION	Tareas								
Fecha	ACTUALIZACION	ACTUALIZAR CANAL	CARGA PLD	CORRECCIONES PLD	REENGANCHE	T_TC_AGENCIAS	UPDATE PLD	TOTAL	
04/02/2019	0.5	0.6	33.3	2.6	0.4	0.4	13	50.8	
05/02/2019	0.8	3.8	41.3	10.8	2.1	0.4	28.9	88.1	
06/02/2019	4.8	0.7	33.3	5.9	0.4	0.4	4.1	49.6	
07/02/2019	2.2	2.4	24.6	3.2	2.6	2.1	4.2	41.3	
08/02/2019	0.9	1	23.5	3.3	1.1	0.9	6.9	37.6	
09/02/2019	1	1.1	25.8	2.4	1.2	0.9	9.2	41.6	
10/02/2019	1	1.1	33.8	3.1	0.9	0.9	13.5	54.3	
11/02/2019	1.3	4.5	41.8	11.5	2.8	1.1	29.6	92.6	
12/02/2019	5.5	1.4	34	6.6	1.1	1.1	4.8	54.5	
13/02/2019	1.2	1.3	34	3.3	1.1	1.1	13.7	55.7	
14/02/2019	2	5	42.5	12	3.3	1.6	30.1	96.5	
15/02/2019	6	1.9	34.5	7.1	1.6	1.6	5.3	58	
16/02/2019	1.7	1.8	34.5	3.8	1.6	1.6	14.2	59.2	
17/02/2019	2	4.2	42.5	12	2.5	0.8	30.1	94.1	

Fuente: Elaboración propia

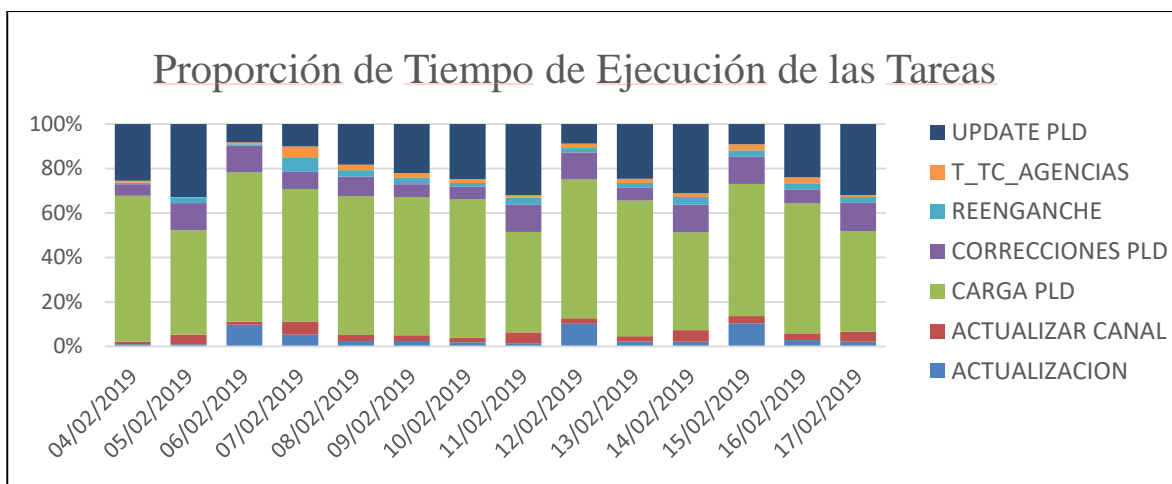


Figura 18 Tareas contenedor actualizar correcciones

Fuente: Elaboración propia

Tabla 6 Análisis de los tiempos del contenedor actualizar correcciones

	MIN	MAX	PROM	MEDIAN
ACTUALIZACION	0.5	6	2.2	1.5
ACTUALIZAR CANAL	0.6	5	2.2	1.6
CARGA PLD	23.5	42.5	34.2	34
CORRECCIONES PLD	2.4	12	6.3	4.85
REENGANCHE	0.4	3.3	1.6	1.4
T_TC_AGENCIAS	0.4	2.1	1.1	1
UPDATE PLD	4.1	30.1	14.8	13.25

Fuente: Elaboración propia

El resultado del análisis del primer contenedor muestra que la tarea Carga PLD es la que tarda más tiempo en ejecutarse, consumiendo aproximadamente el 60% del total del tiempo de ejecución.

De igual forma se realiza el análisis del segundo contenedor (Carga Temporales), el cual, si bien tiene un tiempo de ejecución mucho menor en comparación con el primero, se revisará para validar que el pico de tiempo no recaiga en una tarea específica.

Para este análisis se obtuvieron los resultados expuestos en la Tabla 7, Figura 19 y Tabla 8.

Tabla 7 Tiempos de ejecución diarios del contenedor carga temporales

DURACION (min)	Tareas							
Fecha	CANCELACIONES	CARGA CANCELACIONES	CARGA CLIENTES	CARGA SALDOS	CARGA SALDOS Y CLIENTES	LIMPIAR TABLAS	TOTAL	
04/02/2019	3	0.4	0.4	0.5	7.7	0.5	12.5	
05/02/2019	4.7	0.4	0.5	0.6	8.8	0.6	15.6	
06/02/2019	4.9	0.4	0.5	0.5	7.8	0.5	14.6	
07/02/2019	4.6	2.1	2.2	2.2	7.8	2.1	21	
08/02/2019	5.3	0.9	2.1	2.2	7.7	2.1	20.3	
09/02/2019	3.9	0.9	0.9	1	13.3	0.9	20.9	
10/02/2019	3.5	0.9	0.9	1	8.2	1	15.5	
11/02/2019	5.2	0.9	1	1.1	9.3	1.1	18.6	
12/02/2019	5.6	1.1	1.2	1.2	8.5	1.2	18.8	
13/02/2019	3.7	1.1	1.1	1.2	8.4	1.2	16.7	
14/02/2019	5.9	1.6	1.7	1.8	10	1.8	22.8	
15/02/2019	6.1	1.6	1.7	1.7	9	1.7	21.8	
16/02/2019	4.2	1.6	1.6	1.7	8.9	1.7	19.7	
17/02/2019	5.9	1.6	1.7	1.8	10	1.8	22.8	

Fuente: Elaboración propia

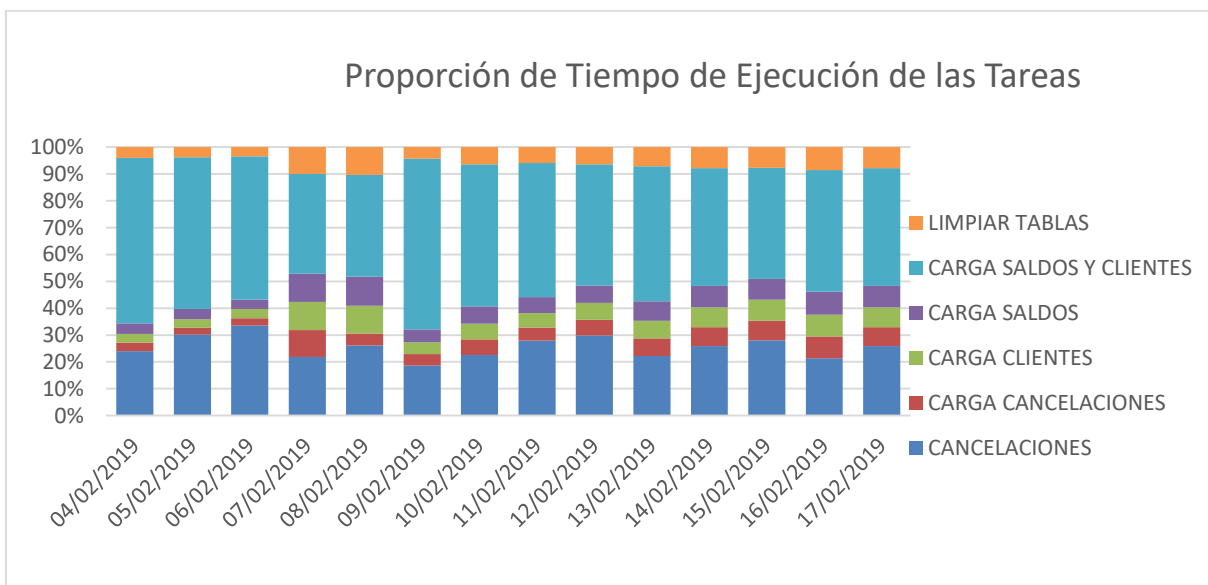


Figura 19 Tareas del contenedor carga temporales

Fuentes: Elaboración propia

Tabla 8 Análisis de los tiempos del contenedor carga temporales

	MIN	MAX	PROM	MEDIAN
CANCELACIONES	3	6.1	4.8	4.8
CARGA CANCELACIONES	0.4	2.1	1.1	1
CARGA CLIENTES	0.4	2.2	1.3	1.15
CARGA SALDOS	0.5	2.2	1.3	1.2
CARGA SALDOS Y CLIENTES	7.7	13.3	9.0	8.65
LIMPIAR TABLAS	0.5	2.1	1.3	1.2

Fuentes: Elaboración propia

Las pruebas indican que las tareas que generan el pico de tiempo para el contenedor Carga Temporales, son aquellas donde se crean las tablas temporales, como lo son Carga Saldos y Clientes y la tarea Cancelaciones.

En ambos contenedores expuestos anteriormente se puede notar que hay una tarea predominante, la cual ocupa el mayor tiempo de ejecución respecto al total, por lo que en el desarrollo de la solución propuesta se hará especial enfoque en solucionar y optimizar las consultas inmersas en estas, con el fin de aminorar el tiempo de ejecución final del proceso ETL.

3.2 Construcción de la Solución

3.2.1 Elaboración de los diccionarios de datos

Una de las desventajas en el actual proceso ETL es la inexistencia de diccionarios de datos de las fuentes, lo cual no permite conocer las tablas que tenemos a nuestra disposición y los tipos de datos que se almacenan en estas, por esto se definió un estándar para la construcción de los diccionarios de datos aprobado por el área de Inteligencia Comercial, la cual nos apoyará a la hora de construir las consultas T-SQL.

El formato del diccionario de datos tiene la siguiente estructura:

- Fuente origen. Hace referencia al tipo de origen de la fuente.
- Servidor. Nombre del servidor en que se encuentra alojada la fuente.
- Base. Nombre de la base de datos en que se encuentra alojada la fuente.
- Tabla. Nombre de la tabla fuente.
- Campos. Nombre del campo de la tabla.
- Tipo Dato. Tipo de dato admitido por el campo.
- Cálculo. Fórmula o función para calcular el campo.
- Descripción. Descripción general de los datos almacenados en el campo.
- Archivo origen. Nombre del archivo en caso la fuente de origen sea un archivo plano.
- Campo origen. Nombre del campo original antes de ser transformado.
- Tipo de dato origen. Tipo de dato del campo original antes de ser transformado.
- Transformación. Fórmula o función por la que fue transformado el campo.
- Valores. Valores admitidos por el campo.

3.2.2 Descripción del Proceso ETL Propuesto

Luego de realizar el análisis del ETL anterior se pudo identificar que el contenedor que toma más tiempo de ejecución es actualizar correcciones y la tarea asociada a esta es carga pld, por lo que si bien se realizó una modificación en la mayoría de los componentes de la ETL se tuvo especial atención en los puntos antes mencionados.

Tras la revisión de las fuentes que se realizó anteriormente también se pudo detectar que el subproceso de saldos retenidos estaba generando información repetida, ya que estos datos eran generados por un proceso aparte, por lo que se decidió quitar de la solución propuesta, quedando la estructura del proceso de extracción, carga y transformación de la siguiente manera, como se muestra en la Figura 20.

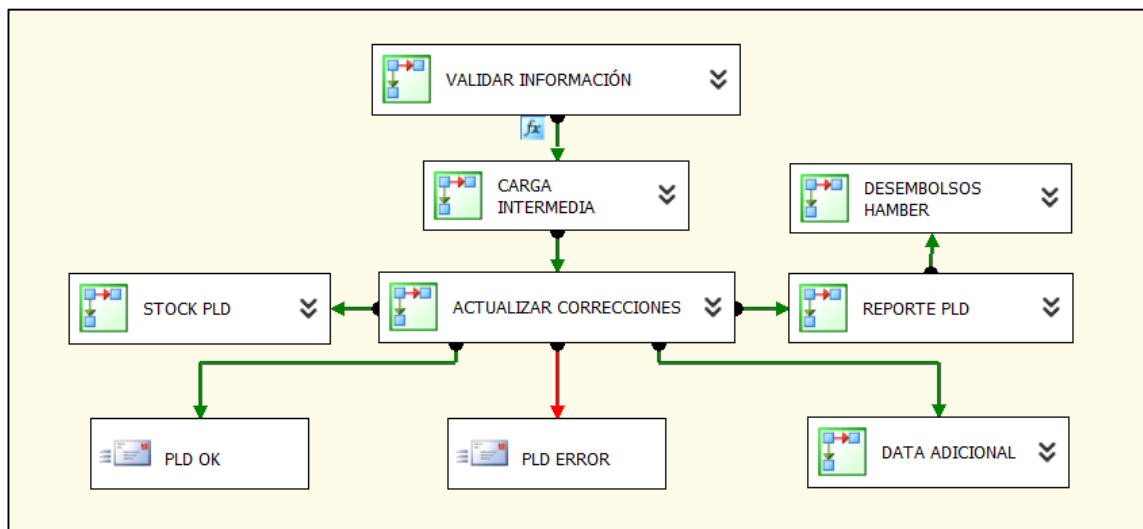


Figura 20 Estructura ETL propuesta

Fuente: Elaboración propia

El objetivo de cada sub proceso o contenedor sigue siendo el mismo, pero ha sido reordenado de forma que permita su mejor comprensión y tenga una correcta ejecución. A continuación, se detalla la función de cada tarea y las modificaciones en caso hayan sido realizadas.

3.2.2.1 Validar Información

Durante este paso se busca obtener los parámetros generales para ejecutar la ETL y llevar a cabo la primera validación de existencia de información, tal como se puede apreciar en la Figura 21.

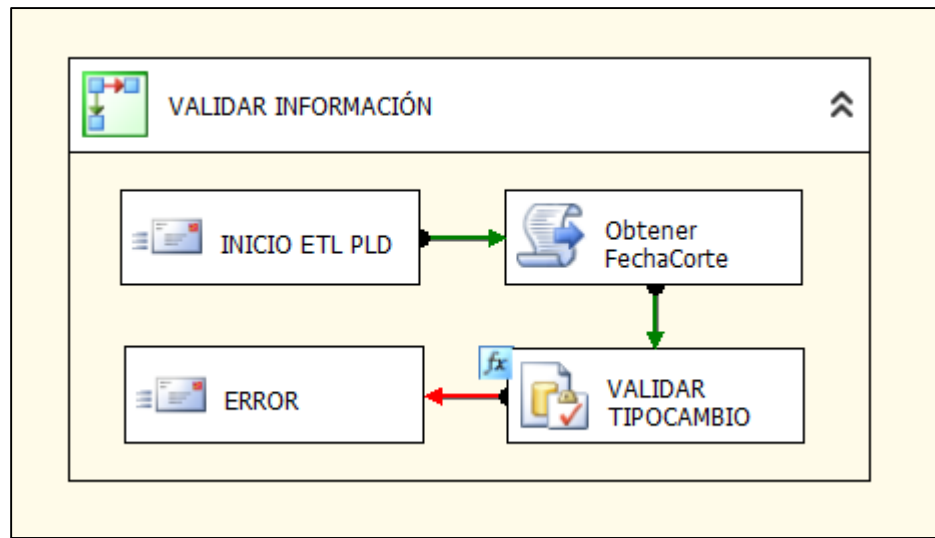


Figura 21 Estructura propuesta de validar información

Fuente: Elaboración propia

- Inicio ETL PLD. Al igual que en la ETL anterior este proceso realizar el envío de un correo indicando la hora de inicio del proceso.
- Obtener FechaCorte. Permite obtener la fecha a la que se realizará la carga de información, adicionalmente ahora también calcula el día de la semana de la ejecución, ambos datos son almacenados en variables independientes ya que serán utilizados por los distintos subprocesos siguientes.
- Validar tipocambio. Devuelve la cantidad de registros para el tipo de cambio a la fecha que se desea cargar la

información. En la Figura 22 se aprecia la consulta utilizada en esta tarea.

```
--CONSULTA ANTIGUA
select count(*) conteo
from dbo.TipoCambio tc left outer join dboTiempo t
on tc.ctiempo = t.ctiempo
where Fecha = ?

--CONSULTA NUEVA
select count(1) conteo
from dbo.TipoCambio tc inner join dbo.Tiempo t
on tc.ctiempo = t.ctiempo
where Fecha = ?
```

Figura 22 Consulta Validar TipoCambio

Fuente: Elaboración propia

- Error. En caso no se hayan encontrado registros del tipo de cambio a la fecha de ejecución, se dispara esta alerta indicando que no hay información disponible en las fuentes, por tanto, el proceso general será detenido.

3.2.2.2 Carga Intermedia

Durante esta etapa del proceso se realiza la construcción de tablas intermedias que contienen información a detalle de los créditos y clientes activos durante el mes, así como también las cancelaciones que han ocurrido hasta la fecha requerida, la arquitectura de este proceso de detalla en la Figura 23.

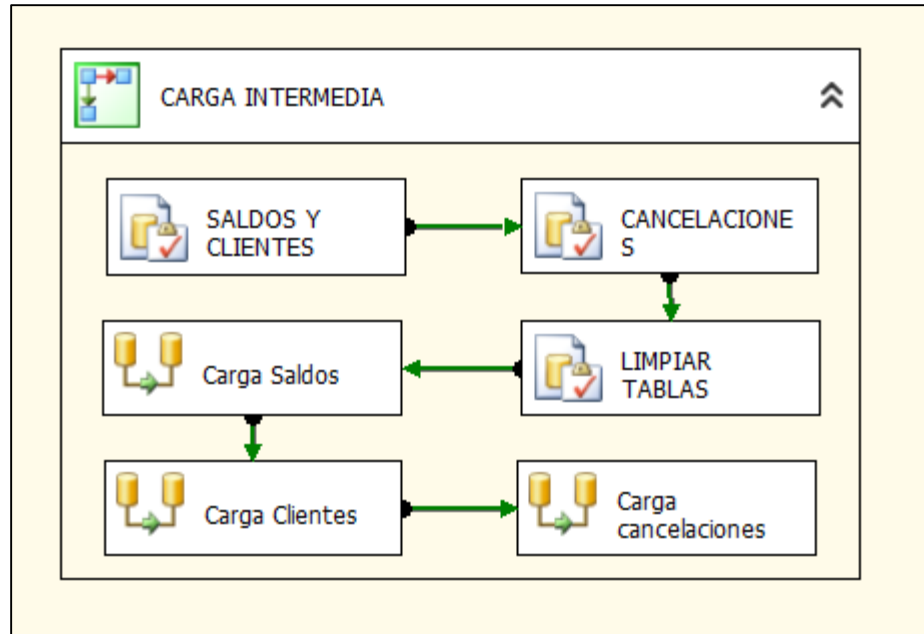


Figura 23 Estructura propuesta de carga intermedia

Fuente: Elaboración propia

- Saldos y clientes. Extrae y transforma la información de las fuentes del servidor SBISBDR03 para depositarlas en tablas temporales, una con información referente a los saldos de los créditos y otra con la información detallada de los clientes vigentes. Esta tarea fue el segundo punto de interés para optimizar la consulta, por tanto, se revisó a detalle y se encontró existencia de índices que no eran utilizados y generaban el aumento en tiempos de ejecución. En la Figura 24 aprecia la indexación de dos columnas, de las cuales solo la columna *numerooperacion* es utilizada durante el resto del proceso para ejecutar los join, mientras que en la Figura 25 se ve el cambio de la sentencia IN por la sentencia LEFT JOIN, lo cual permitió mejorar el tiempo de ejecución de la consulta.

```

if(object_id('tempdb..#operaciones') is not null ) drop table #operaciones
select *
into #operaciones
FROM #temp3
union all
-SELECT * FROM #temp4

create index idx_01 on #operaciones(cOperacion); -- Este índice fue eliminado
create clustered index idx_02 on #operaciones(NumeroOperacion)

```

Figura 24 Eliminación de índice en la consulta

Fuente: Elaboración propia

```

--ANTIGUO
if(object_id('tempdb..#operacioncolocacion') is not null )
-drop table #operacioncolocacion

SELECT *
INTO #operacioncolocacion
FROM operacioncolocacion
WHERE cOperacion > 0
AND NumeroOperacion IN (SELECT NumeroOperacion FROM #tmp_operacioncolocacion)
-AND fechafinvigencia < @fecha_corte+2

--NUEVO
if(object_id('tempdb..#operacioncolocacion') is not null )
-drop table #operacioncolocacion
|
SELECT a.*
INTO #operacioncolocacion
FROM operacioncolocacion a inner join #tmp_operacioncolocacion b
on a.NumeroOperacion = b.NumeroOperacion
WHERE cOperacion > 0
AND fechafinvigencia < @fecha_corte+2

```

Figura 25 Reemplazo de sentencia IN por INNER JOIN

Fuente: Elaboración propia

- Cancelaciones. Esta tarea se encarga de generar una tercera tabla temporal con información de los créditos que han sido cancelados durante el periodo actual.
- Limpiar tablas. Realiza la limpieza de las tablas intermedias donde serán depositados los datos de las

tablas temporales, en caso de la tabla de cancelaciones solo elimina los registros del periodo.

- Carga saldos. Esta tarea permite el depósito de la tabla temporal con información de los saldos en una tabla física que será usada como intermedia.
- Carga clientes. Esta tarea permite el depósito de la tabla temporal con información de los clientes en una tabla física que será usada como intermedia.
- Carga cancelaciones. Esta tarea permite el depósito de la tabla temporal con información de las cancelaciones en una tabla física que será usada como intermedia y que además guarda un histórico.

3.2.2.3 Actualizar Correcciones

Durante esta etapa del proceso se realiza el volcado de la información existente en las tablas intermedias generadas en el contenedor anterior, aplicando las lógicas del negocio acordadas con otros equipos del banco, su estructura se aprecia en la Figura 26.

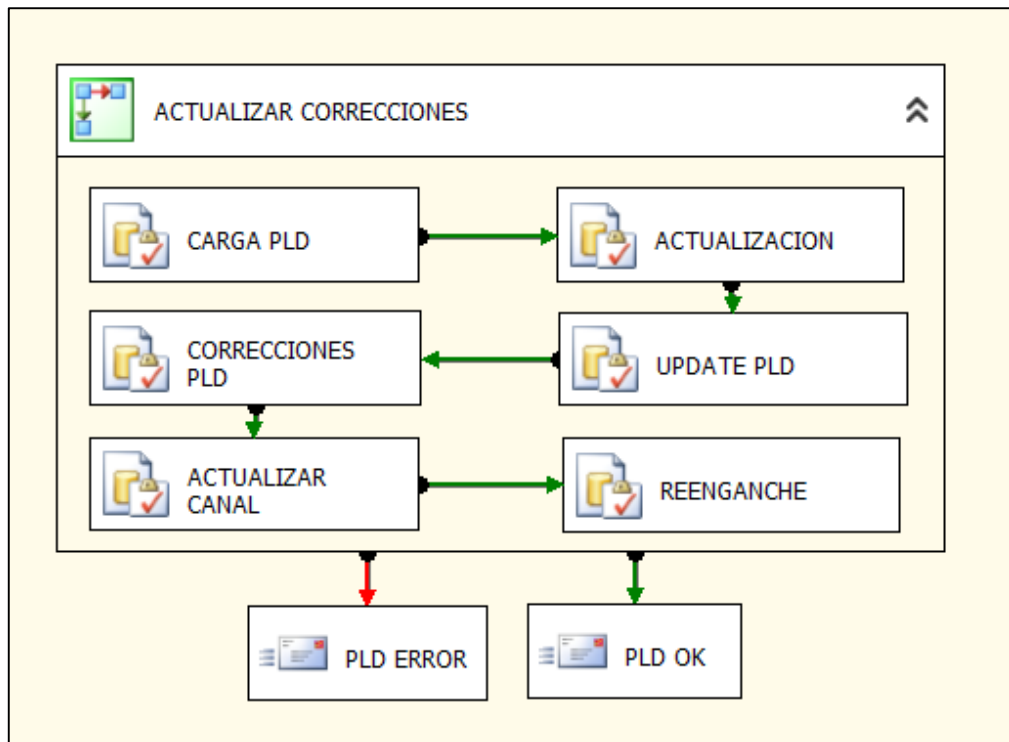


Figura 26 Estructura propuesta de actualizar correcciones

Fuente: Elaboración propia

- Carga PLD. Esta tarea tuvo especial atención, ya que se observó tras el análisis que es la que mayor tiempo de duración tarda y consume la mayor cantidad de recursos del servidor. Tras revisar la consulta T-SQL inmersa en esta tarea se realizó la eliminación de índices a tablas temporales que no contenían muchos registros y causaban mayor lentitud en las sentencias update, así también se pudo identificar que para tablas que contenían demasiados registros el aplicar la consulta update e insert teniendo dos columnas indexadas tardaban demasiado tiempo en ejecutarse, en primera instancia se optó por dejar solo uno de los índices pero luego se pudo encontrar que había posibilidad de agregar los datos mediante un

left join en vez de la actualización de los campos, lo cual permitió disminuir enormemente los tiempos.

En la Figura 27 se puede ver el caso de la creación de una tabla temporal donde luego se le agregaba una columna adicional que sería actualizada por una consulta update más adelante, estos tres pasos fueron reemplazados por una consulta left join a la primera tabla para agregar la columna de forma directa.

```
--ANTIGUO
if object_id('tempdb..#PLD_TEMP_NUEVO') is not null drop table #PLD_TEMP_NUEVO
SELECT P.*,
    case when p.fechaaperturaoriginal>='20150101' and v.COD_EMP is not null then v.CANAL
        else 'AGENCIAS' end as CANAL,
    case when p.fechaaperturaoriginal>='20150101' and v.COD_EMP is not null then v.NOM_SUP else 'AGENCIAS' end as NOM_SUP
INTO #PLD_TEMP_NUEVO
FROM #PLD_TEMP P
left join (select * from bfpmarketing..cr_headcount where estado='A') c on p.CodigoVendedor=c.cod_vend
left join T_VENDEDOR_INTERNO v on c.codigo =v.COD_EMP and est_act='Activo'

if object_id('tempdb..#PLD_TEMP') is not null drop table #PLD_TEMP
alter table #PLD_TEMP_NUEVO
add agencia_maesaldos varchar(100)

update #PLD_TEMP_NUEVO
set agencia_maesaldos= m.agencia
from #PLD_TEMP_NUEVO pld left join #MAESALDOS m
on (pld.numerooperacion=m.numerooperacion)

--NUEVO

if object_id('tempdb..#PLD_TEMP_NUEVO') is not null drop table #PLD_TEMP_NUEVO
SELECT P.*,
    case when p.fechaaperturaoriginal>='20150101' and v.COD_EMP is not null then v.CANAL else 'AGENCIAS'
        end as CANAL,
    case when p.fechaaperturaoriginal>='20150101' and v.COD_EMP is not null then v.NOM_SUP else 'AGENCIAS'
        end as NOM_SUP,
    CAST(m.agencia as varchar(100)) agencia_maesaldos
INTO #PLD_TEMP_NUEVO
FROM #PLD_TEMP P
left join (select * from bfpmarketing..cr_headcount where estado='A') c on p.CodigoVendedor=c.cod_vend
left join T_VENDEDOR_INTERNO v on c.codigo =v.COD_EMP and est_act='Activo'
left join #MAESALDOS m on P.numerooperacion=m.numerooperacion
```

Figura 27 Modificación de la tarea Carga PLD

Fuente: Elaboración propia

- Actualización. Actualiza el importe desembolsado y datos del colaborador que realizó la venta del crédito que actualmente no se encuentre vigente.

En la Figura 28 se puede apreciar la modificación realizada en la consulta.

```
--ANTIGUO
UPDATE PLD
SET IMPORTE_DESEMBOLSADO_SOLES = MP.IMPORTEDESEMBOLSADO *
(CASE
    WHEN MP.MONEDA='NUEVO SOL'
    THEN 1
    ELSE ( SELECT TIPOCAMBIO
           FROM BFFMARKETING..BI_TIPOCAMBIO
           where CONVERT(CHAR(8),FECHA,112)=@fecha_corte
                AND MONEDA ='US DOLLAR')
    END),
    CODIGOVENDEDOR = MP.CODIGOVENDEDOR,
    NOMBREVENDEDOR = MP.NOMBREVENDEDOR
FROM PLD P INNER JOIN MAESALDOS_PLD MP ON (P.NUMEROOPERACION=MP.NUMEROOPERACION)
WHERE P.PERIODO=@PERIODO
AND P.FLAGSTK=0
--AND P.IMPORTE_DESEMBOLSADO_SOLES IS NULL

--NUEVO
DECLARE @tipocambio float
select @tipocambio = tipocambio
from BFFMARKETING..BI_TIPOCAMBIO where fecha=@fecha_corte and moneda = 'US DOLLAR'

UPDATE PLD
SET IMPORTE_DESEMBOLSADO_SOLES= MP.IMPORTEDESEMBOLSADO * (CASE WHEN MP.MONEDA='NUEVO SOL' THEN 1 ELSE @tipocambio END ),
CODIGOVENDEDOR = MP.CODIGOVENDEDOR,
NOMBREVENDEDOR = MP.NOMBREVENDEDOR
FROM PLD P INNER JOIN MAESALDOS_PLD MP ON (P.NUMEROOPERACION=MP.NUMEROOPERACION)
WHERE P.PERIODO=@PERIODO
AND P.FLAGSTK=0
--AND P.IMPORTE_DESEMBOLSADO_SOLES IS NULL
```

Figura 28 Modificación de la tarea Actualización

Fuente: Elaboración propia

- Update PLD. Incluye actualizaciones puntuales como cambio del nombre de una agencia, en general se usa esta tarea para parchar errores en la carga de información causados por inconsistencia de algunas lógicas. Se eliminaron las consultas que se encontraban desfasadas y no generaban ninguna modificación en las bases. En la Figura 29 se muestran algunas de las consultas utilizadas durante esta tarea.

```

UPDATE PLD
SET PLD.TIPO_PLD=P2.TIPO_PLD
FROM #PLD_MES_ANT_2 P2
WHERE PLD.NUMEROOPERACION=P2.NUMEROOPERACION
AND PLD.FECHACANCELACION>=@fecha_inicio
AND PLD.TIPO_PLD<>P2.TIPO_PLD
AND PLD.PERIODO=@PERIODO

UPDATE PLD
SET pld.CodigoVendedor =p2.CodigoVendedor,
pld.NombreVendedor =p2.NombreVendedor,
pld.importe_desembolsado_soles =p2.importe_desembolsado_soles
FROM #PLD_MES_ANT_2 P2
WHERE PLD.NUMEROOPERACION=P2.NUMEROOPERACION
AND PLD.FECHACANCELACION>=@fecha_inicio
AND PLD.PERIODO=@PERIODO

---actualiza cancelaciones totales
update pld set
fechacancelacion=a.fechacancelacion,
importe_desembolsado_soles=a.importe_desembolsado_soles,
saldo_cancelaciones_soles=a.saldo_cancelaciones_soles
from
(select * from pld where periodo=@PERIODO-1
and fechacancelacion is not null)a
where pld.numerooperacion=a.numerooperacion
and pld.periodo=@PERIODO

```

Figura 29 Modificaciones de la tarea Update PLD

Fuente: Elaboración propia

- Correcciones PLD. Aquellos registros que viajaron desde las fuentes con inconsistencia en sus datos son corregidos durante este proceso. Su modificación es de forma reactiva, es decir al encontrar nuevas casuísticas se van agregando actualizaciones. Se eliminaron las consultas que se encontraban desfasadas y no generaban ninguna modificación en las bases. En la Figura 30 se muestran algunas de las consultas utilizadas durante esta tarea.

```

if object_id('tempdb..#PLD_MES_ANT') is not null drop table #PLD_MES_ANT
SELECT *
INTO #PLD_MES_ANT
FROM PLD
WHERE PERIODO=@PERIODO-1
AND FECHAAPERTURAORIGINAL>=@mes_anterior

UPDATE PLD
SET AGENCIA_MAESALDOS=PLD_ANT.AGENCIA_MAESALDOS
FROM #PLD_MES_ANT PLD_ANT
WHERE PLD.PERIODO=@PERIODO AND PLD.FECHAAPERTURAORIGINAL>=@fecha_inicio
AND PLD.CANAL='AGENCIAS'
AND PLD.AGENCIA_MAESALDOS='OF. PRINCIPAL'
AND PLD.NOMBREVENDEDOR=PLD_ANT.NOMBREVENDEDOR

update pld set tipo_pld='COLABORADORES'
where periodo=@PERIODO-1
and canal='COLABORADORES'
and tipo_pld<>'COLABORADORES'

UPDATE PLD
SET AGENCIA = 'AG. CHULUCANAS'
WHERE PERIODO = @PERIODO
AND AGENCIA = 'AG. CARSA CHULUCANAS'

UPDATE PLD
SET AGENCIA_MAESALDOS = 'AG. CHULUCANAS'
WHERE PERIODO = @PERIODO
AND AGENCIA_MAESALDOS = 'AG. CARSA CHULUCANAS'

```

Figura 30 Modificaciones de la tarea Correcciones PLD

Fuente: Elaboración propia

- Actualizar Canal. Se realiza la actualización manual del canal de venta de cada crédito en base al colaborador que realizó la venta, estas modificaciones son indicadas por un equipo especializado que comparte con el equipo de Inteligencia Comercial una lista actualizada mes a mes. En la Figura 31 se muestran algunas de las consultas utilizadas durante esta tarea.

```

update PLD set canal='ESPECIALISTAS PLD'
WHERE codigovendedor IN ('8929','8410','7015','9588','4038','3995','10783','8159','2646',
                          '2781','7625','3913','2A1F','11783','4825','D005','2FDC')

update PLD set canal='FFV EXTERNA 2'
WHERE codigovendedor IN ('D996','E014','D986','E008','D998','E016','E134',
                          'E135','E132','E133','E407','E405','E404','E403','E402')

--DENNIS ACUÑA
update PLD set canal='FFV EXTERNA 1'
WHERE codigovendedor IN ('D994','E009','E250','D995','E010','E130','E131','E248',
                          'E249','E277','E278','E409','E408','E406','E414')

update pld set canal='COLABORADORES'
where codigovendedor in ('5503','1712','7629','4133','3002')

update PLD set canal='AGENCIAS'
WHERE codigovendedor IN ('8430','10274','8466','2B2C','2D1C',
                          '27F5','D779','12069','2F25','2F5C','2D42')

update PLD set canal='BANCA TELEFONICA'
WHERE codigovendedor IN ('E275','E276','E401')
AND periodo=@PERIODO
and fechaaperturaoriginal >= @fecha_inicio

update PLD set canal='CALL CENTER'
WHERE codigovendedor IN ('E376','E299')
AND periodo=@PERIODO
and fechaaperturaoriginal >= @fecha_inicio

```

Figura 31 Modificaciones de la tarea Actualizar Canal

Fuente: Elaboración propia

- Reenganche. Durante esta tarea se aplica la lógica establecida por el equipo de expertos en los Préstamos de Libre Disponibilidad, la cual permitirá hacer el cálculo del monto neto del crédito. Se modificó la consulta para trabajar con variables tipo tabla para almacenar la información temporal, ya que no se realizaban operaciones sobre esta una vez creada, tal como se puede apreciar en la Figura 32.

```

|--ANTIGUO
if(object_id('tempdb..#temp_reenganche') is not null ) drop table #temp_reenganche
select a.numerooperacion,
(case when year(fechaaperturaoriginal)*100+month(fechaaperturaoriginal)=periodo
and venta_reenganche is not null then venta_reenganche else
(case when year(fechaaperturaoriginal)*100+month(fechaaperturaoriginal)=periodo
and flagstk=1 and cproductoult>0 then isnull(importe_desembolsado_soles,0)
else 0 end)end)importe_neto
into #temp_reenganche
from pld a
left join (--importe desembolsado
select a.codigoclientebanco,
sum(a.importe_desembolsado_soles-b.saldo_corte_soles) as venta_reenganche
from (
select codigoclientebanco,
sum(importe_desembolsado_soles)importe_desembolsado_soles
from pld
where periodo=@PERIODO
)
)

```

```

--NUEVO
declare @temp_reenganche as table(numerooperacion bigint, importe_neto int)
INSERT INTO @temp_reenganche
SELECT P.NUMEROOPERACION,
CASE
WHEN P.TIPO_PLD = 'GARANTIA LIQUIDA'
THEN( CASE
WHEN C.SALDO_CANCELACION IS NULL
THEN P.IMPORTE_DESEMBOLSADO_SOLES
ELSE A.IMPORTE_DESEMBOLSADO_SOLES - IIF(C.SALDO_CANCELACION<0,0,C.SALDO_CANCELACION)
END)
ELSE( CASE
WHEN B.SALDO_CANCELACION IS NOT NULL AND B.FECHA_CANCELACION >= P.FECHAAPERTURAORIGINAL
THEN A.IMPORTE_DESEMBOLSADO_SOLES - IIF(B.SALDO_CANCELACION<0,0,B.SALDO_CANCELACION)
ELSE P.IMPORTE_DESEMBOLSADO_SOLES
END)
END IMPORTE_NETO
FROM PLD P LEFT JOIN (--IMPORTE DESEMBOLSADO TOTAL POR CLIENTE
SELECT CODIGOCLIENTEBANCO, FECHAAPERTURAORIGINAL,
SUM(IMPORTE_DESEMBOLSADO_SOLES) IMPORTE_DESEMBOLSADO_SOLES
FROM PLD
WHERE PERIODO = @PERIODO
)

```

Figura 32 Modificación de la tarea Reenganche

Fuente: Elaboración propia

- PLD Ok. Envía un correo informando que la tabla ya está disponible con la información actualizada; es enviado a una lista de usuarios finales que la consumen.
- PLD Error. Envía un correo de alerta indicando que ocurrió un problema durante la ejecución de alguna tarea del contenedor, por lo que se debe revisar lo sucedido de forma manual y volver a ejecutar desde el último punto guardado.

3.2.2.4 Stock PLD

La función de este contenedor es realizar el cálculo del stock de Préstamos de Libre Disponibilidad vigentes y su envío al equipo de Banca Empresa, quienes utilizarán esta información como más les convenga, las tareas que comprende este contenedor se detallan en la Figura 33.

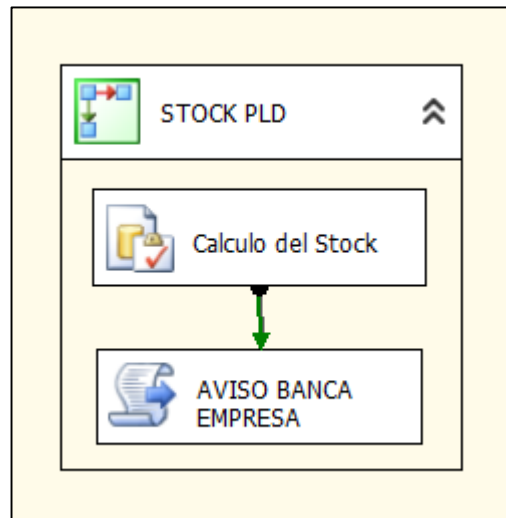


Figura 33 Estructura propuesta de stock PLD

Fuente: Elaboración propia

- Calculo del Stock. Ejecuta una consulta que permite realizar el cálculo del stock vigente de los Créditos de Libre Disponibilidad, expresando en saldo (soles) y número de clientes. La consulta utilizada se visualiza en la Figura 34.

```
DECLARE @fechaCorte DATE
DECLARE @periodo INT
SET @fechaCorte = ?
SET @periodo = year(@fechaCorte)*100 + month(@fechaCorte)

SELECT SUM(SALDO_CORTE_SOLES) Monto, COUNT(DISTINCT CodigoClienteBanco) Clientes
FROM PLD
WHERE PERIODO = @periodo
AND flagstk>0 AND cproductoult>0
AND codigocuentacontable not like '8103%' AND codigocuentacontable not like '8109%'
```

Figura 34 Consulta de la tarea Cálculo del Stock

Fuente: Elaboración propia

- Aviso banca empresa. Realiza el envío de un correo personalizado al equipo de Banca Empresa con los datos antes calculados.

3.2.2.5 Data Adicional

Esta etapa de la ETL se utiliza para calcular información puntual solicitada por otros equipos de trabajo, serán almacenados en otras fuentes que sirven para dar mantenimiento a otros procesos y/o reportes, tal como se muestra en la Figura 35.

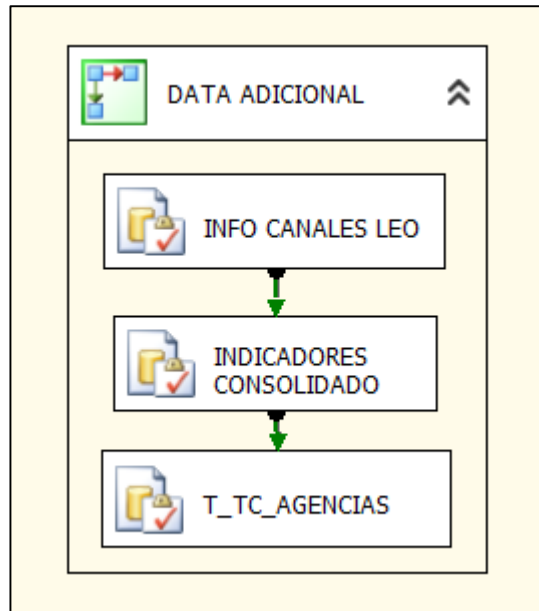


Figura 35 Estructura propuesta de data adicional

Fuente: Elaboración propia

- Info Leo Canales. La consulta de esta tarea ingresa los desembolsos del mes a una fuente externa que consolida información de distintos productos, la cual luego es consumida por una macro que permite la creación de archivos diarios organizados por productos y agencias. La consulta utilizada se visualiza en la Figura 36.

```

DECLARE @fecha_corte as datetime
DECLARE @fecha as int
DECLARE @periodo as int

set @fecha_corte = ?
SET @fecha = year(@fecha_corte)*10000+ month(@fecha_corte)*100 +day(@fecha_corte)
SET @periodo = year(@fecha_corte)*100+ month(@fecha_corte)

DELETE FROM bfpmarketing..lb_info_Canales
WHERE producto = 'PLD' and periodo = @periodo and Fecha_desembolso = @fecha

insert into bfpmarketing..lb_info_Canales
select  p.Periodo,
        'PLD' Producto,
        a.codigoagenciaCodigo_Oficina,
        p.Oficina,
        convert(varchar(10),p.fechaaperturaoriginal,112) Fecha_desembolso,
        p.codigoendedorCodigo_usuario,
        p.nombreendedorColaborador,
        p.nombreclienteNombre_cliente,
        p.numerooperacionNumero_credito,
        p.codigoclientebancoCodigo_cliente_banco,
        'PEN' Moneda,
        p.importe_desembolsado_solesImporte_bruto,
        p.venta_neta_solarizadoimporte_neto
from    (SELECT *, IIF(Agencia_maesaldos LIKE '%HELL', 'AG. SCHELL', Agencia_maesaldos) Oficina FROM bfp_tc..pld) p
left join bfpmarketing..agencias_ic a
on p.oficina collate SQL_Latin1_General_CP1_CI_AS= a.agencia collate SQL_Latin1_General_CP1_CI_AS
where p.periodo = @periodo and p.fechaaperturaoriginal = convert(varchar(10),@fecha_corte,112)

```

Figura 36 Consulta de la tarea Info Leo Canales

Fuente: Elaboración propia

- Indicadores Consolidado. Esta tarea realiza el cálculo de los principales indicadores de medición del producto, tales como el crecimiento, las ventas netas, etc.; la información consolidada junto con la de otros productos es presentada por otro reporte cuya función es consolidar diferentes KPI's para el seguimiento de los productos del banco. La consulta utilizada se visualiza en la Figura 37.

```

--LIMPIAR DIA
DELETE FROM BFPMARKETING..INDICADORES_CONSOLIDADO WHERE PRODUCTO = 'PLD' AND FECHA = @FECHACORTE

--VENTAS NETAS
SELECT @VENTAS = SUM(VENTA_NETA_SOLARIZADO) FROM PLD
WHERE PERIODO = @PERIODO AND FECHAAPERTURAORIGINAL >= @FECHAAPERTURAORIGINAL
AND FLAGSTK>0 AND CPRODUCTOULT>0
AND TIPO_PLD <> 'COLABORADORES'
AND codigocuentacontable not like '8103%' AND codigocuentacontable not like '8109%'

--CRECIMIENTO
SELECT @CRECIMIENTO =
SUM(CASE WHEN periodo = @PERIODO then SALDO_CORTE_SOLES else 0 end) -
SUM(case when periodo = @PERIODO_ANT then SALDO_CORTE_SOLES else 0 end)
FROM PLD
WHERE flagstk=1 and cproductoult>0 AND codigocuentacontable not like '8103%' AND codigocuentacontable not like '8109%'

```

Figura 37 Consulta de la tarea Indicadores Consolidado

Fuente: Elaboración propia

- T_TC_Agencias. Resume la información a un nivel con menor detalle, la información es agrupada por agencia y colaborador, en vez de desglosarla por cada crédito como en la tabla original. Guarda este resumen en una tabla histórica que contiene la información al cierre de cada mes. La consulta utilizada se visualiza en la Figura 38.

```

SELECT CASE WHEN codigovendedor in ('E033','E034','E375','E400','E397') THEN agencia_maesaldos COLLATE SQL_Latin1_General_CP1_CI_AS
ELSE b.agencia END AS AGENCIA_FINAL,
CASE WHEN codigovendedor in ('E033','E034','E375','E400','E397') THEN codigovendedor
ELSE codigo END AS COD_VENDEDOR,
nombrevendedor AS NOM_VEND_WF,
'BANCO' EMPRESA,
puesto,
NULL AS META,
sum(CASE WHEN flagstk='1' THEN 1 else 0 end) AS CANTIDAD,
'PLD' AS PRODUCTO,
sum (case when venta_neta_solarizado is not null then venta_neta_solarizado else isnull(importe_desembolsado_soles,0)end)AS MONTO,
PERIODO,
(CONVERT(CHAR(08),getdate()-1,112)) AS FECHA_CORTE,
CANAL
INTO #AGENCIAS_PLD
FROM PLD P left join (select * from bfpmarketing..cr_headcount where estado='A' or (estado <> 'A' and[fecha sal]>='20190601')) b
on p.codigovendedor=b.cod_vend
WHERE PERIODO='201906'
AND FECHAAPERTURAORIGINAL>='20190601'
and flagstk='1'
and cproductoult>0
GROUP BY CASE WHEN codigovendedor in ('E033','E034','E375','E400','E397') THEN agencia_maesaldos COLLATE SQL_Latin1_General_CP1_CI_AS
ELSE b.agencia END,
CASE WHEN codigovendedor in ('E033','E034','E375','E400','E397') THEN codigovendedor ELSE codigo END,
nombrevendedor,P.PERIODO,P.CANAL, B.puesto

```

Figura 38 Consulta de la tarea T_TC_Agencias

Fuente: Elaboración propia

3.2.2.6 Reporte PLD

Durante esta fase se accede al reporte de avance de PLD y haciendo uso de macros se refresca la información y se generan los reportes diarios, que son almacenados de forma local como también en un servidor de archivos, su estructura se detalla en la Figura 39.

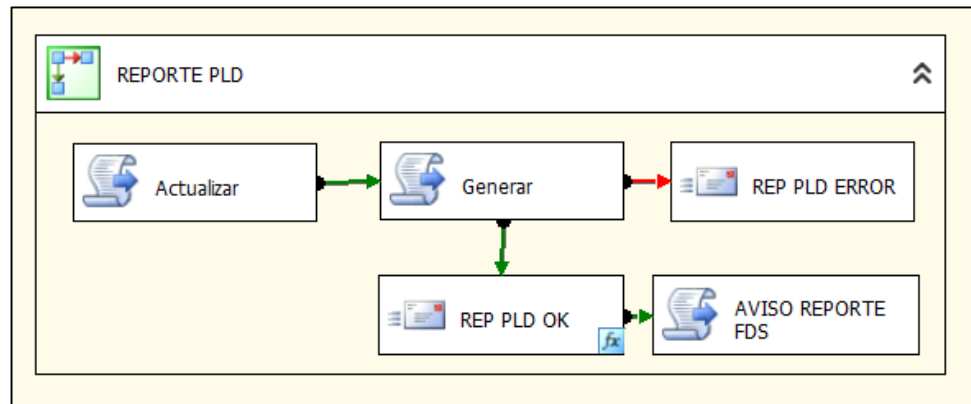


Figura 39 Estructura propuesta de reporte PLD

Fuente: Elaboración propia

- Actualizar. Accede al reporte de PLD y ejecuta la función Actualizar, la cual refresca todas las tablas conectadas que conforman el reporte.
- Generar. Accede al reporte de PLD que se encuentra activo y guarda una copia local con la fecha de la información procesada y una copia adicional en un servidor de archivos que será consumido por un portal, donde los usuarios finales podrán descargarlo y visualizarlo.
- REP PLD Ok. En caso se hayan generado los reportes de forma correcta, se enviará un correo indicándolo al equipo de Inteligencia Comercial, quienes con previa revisión

informarán al resto de usuarios finales que el archivo ya se encuentra disponible.

- Aviso reporte FDS. Esta tarea agregada permite el envío del correo corporativo a todas las agencias y grupos del banco que requieran acceso al avance de PLD los días sábado.
- REP PLD Error. En caso hubiese fallado la actualización o la generación de los archivos diarios (en la máquina local o en el servidor) se enviará un correo al equipo para su revisión y en caso sea necesario su reproceso manual.

3.2.2.7 Desembolsos Hamber

Las tareas inmersas en este contenedor ayudan a actualizar los reportes con información de los desembolsos del mes, que es compartido a una lista de usuarios finales, tal como se muestra en la Figura 40.

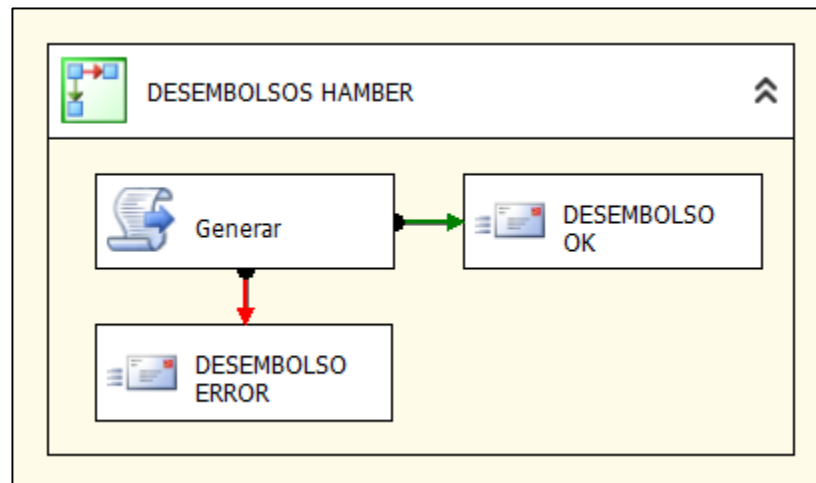


Figura 40 Estructura propuesta de desembolsos Hamber

Fuente: Elaboración propia

- Generar. Este script permite acceder a un archivo Excel (reporte de desembolsos) y ejecutar la macro programada para actualizar la información y crear una copia en una ruta específica.
- Desembolso OK. En caso no haya ocurrido algún error durante la actualización del reporte o la creación de la copia, se realizará el envío de un correo al usuario que hace uso del reporte.
- Desembolso Error. En caso haya ocurrido algún error durante la actualización del reporte o la creación de la copia, se realizará el envío de un correo al usuario encargado de dar mantenimiento a dicho reporte.

3.2.3 Medición de los Tiempos ETL Propuesta

El análisis de los tiempos de ejecución será llevado a cabo con la misma metodología utilizada para medir los tiempos del proceso ETL anterior, es decir yendo de lo general (contenedor) a lo particular (tarea), con el fin de conseguir los mismos indicadores y poder contrastarlos más adelante. Cabe resaltar que la unidad de medida utilizada es el minuto.

3.2.3.1 Análisis a nivel contenedor.

Se procedió a evaluar el tiempo de ejecución de cada contenedor con el fin de encontrar los nuevos valores de tiempo de ejecución, arrojando los siguientes resultados en la Tabla 9, Figura 41 y Tabla 10.

Tabla 9 Tiempos de ejecución diarios por contenedor - ETL propuesto

DURACION (min)	Contenedores					
Fecha de Ejecución	ACTUALIZAR CORRECCIONES	DATA ADICIONAL	VALIDAR INFORMACION	CARGA INTERMEDIA	Total general	
11/03/2019	48.2	0.8	0.0	10.3	59.4	
12/03/2019	59.7	4.1	0.0	13.4	77.2	
13/03/2019	47.1	0.5	0.0	12.4	59.9	
14/03/2019	26.9	2.0	0.0	10.2	39.0	
15/03/2019	31.6	0.3	0.0	9.1	41.0	
16/03/2019	35.6	0.5	0.0	15.7	51.8	
17/03/2019	48.2	0.8	0.0	10.3	59.4	
18/03/2019	67.1	4.1	0.0	13.4	84.7	
19/03/2019	47.1	0.5	0.0	12.4	59.9	
20/03/2019	48.2	0.8	0.0	10.3	59.4	
21/03/2019	73.2	4.1	0.0	13.4	90.7	
22/03/2019	47.1	0.5	0.0	12.4	59.9	
23/03/2019	48.2	0.8	0.0	10.3	59.4	
24/03/2019	65.5	4.1	0.0	13.4	83.1	

Fuente: Elaboración propia

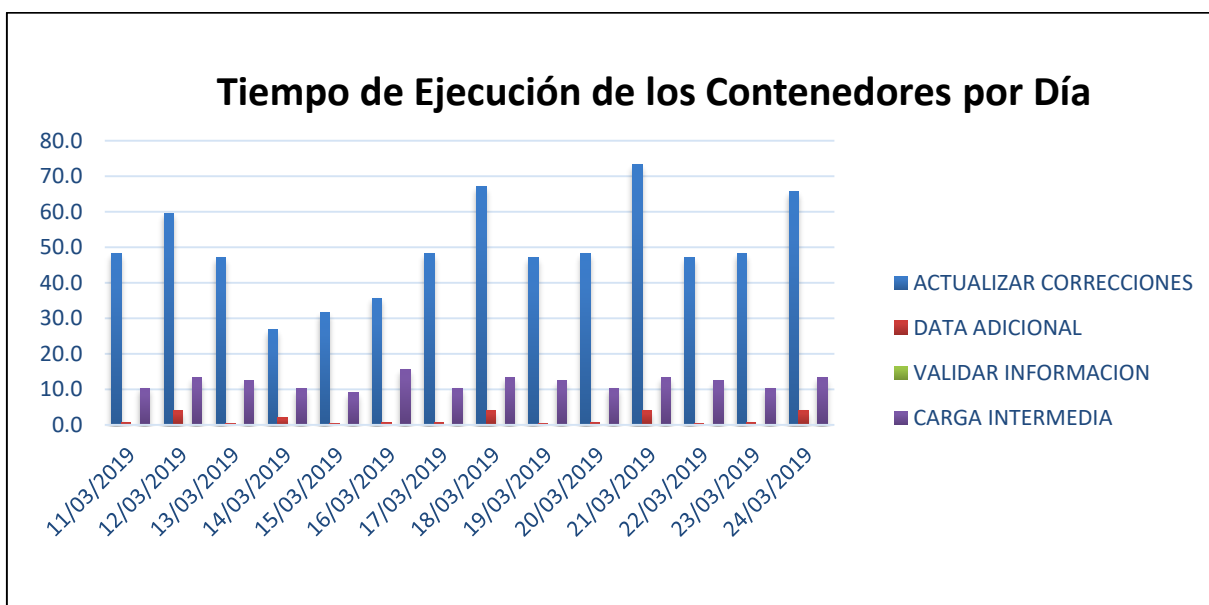


Figura 41 Tiempo de ejecución de los contenedores propuesta (min)

Fuente: Elaboración propia

Tabla 10 Análisis de los tiempos por contenedor - ETL propuesta

	MIN	MAX	PROM
ACTUALIZAR CORRECCIONES	26.9	73.2	49.5
DATA ADICIONAL	0.3	4.1	1.7
VALIDAR ACTIVOS	0.0	0.0	0.0
CARGA INTERMEDIA	9.1	15.7	11.9

Fuente: Elaboración propia

De estos resultados se pudo inferir que los picos de tiempo de procesamiento siguen estando en el contenedor Actualizar Correcciones, con un tiempo promedio de ejecución de 49.5 minutos y el contenedor Carga Intermedia, con un tiempo promedio de ejecución de 11.9 minutos.

3.2.3.2 Análisis a nivel de tarea

Se procedió a continuación a analizar el contenedor actualizar correcciones, para detallar el tiempo de ejecución por tarea, solo se analizará este contenedor a fondo ya que es el que nos proporciona mayor interés de solucionar.

Los resultados arrojados se muestran en las Tablas 11 y 12 y la Figura 42.

Tabla 11 Tareas del contenedor propuesto, actualizar correcciones

DURACION (min)	Tareas							
Fecha	ACTUALIZACION	ACTUALIZAR CANAL	CARGA PLD	CORRECCIONES PLD	REENGANCHE	UPDATE PLD	TOTAL	
11/03/2019	0.15	0.21	32.91	2.26	0.07	12.62	48.22	
12/03/2019	0.44	3.43	31.00	10.47	1.79	12.52	59.66	
13/03/2019	4.41	0.39	32.92	5.59	0.03	3.72	47.05	
14/03/2019	0.17	0.35	22.55	1.12	0.58	2.11	26.88	
15/03/2019	0.05	0.13	22.61	2.43	0.27	6.09	31.59	
16/03/2019	0.17	0.29	25.00	1.54	0.32	8.30	35.62	
17/03/2019	0.15	0.21	32.91	2.26	0.07	12.62	48.22	
18/03/2019	0.44	3.43	33.71	10.47	1.79	17.29	67.14	
19/03/2019	4.41	0.39	32.92	5.59	0.03	3.72	47.05	
20/03/2019	0.15	0.21	32.91	2.26	0.07	12.62	48.22	
21/03/2019	0.44	3.43	28.53	10.47	1.79	28.53	73.20	
22/03/2019	4.41	0.39	32.92	5.59	0.03	3.72	47.05	
23/03/2019	0.15	0.21	32.91	2.26	0.07	12.62	48.22	
24/03/2019	0.44	3.43	32.86	10.47	1.79	16.52	65.52	

Fuente: Elaboración propia

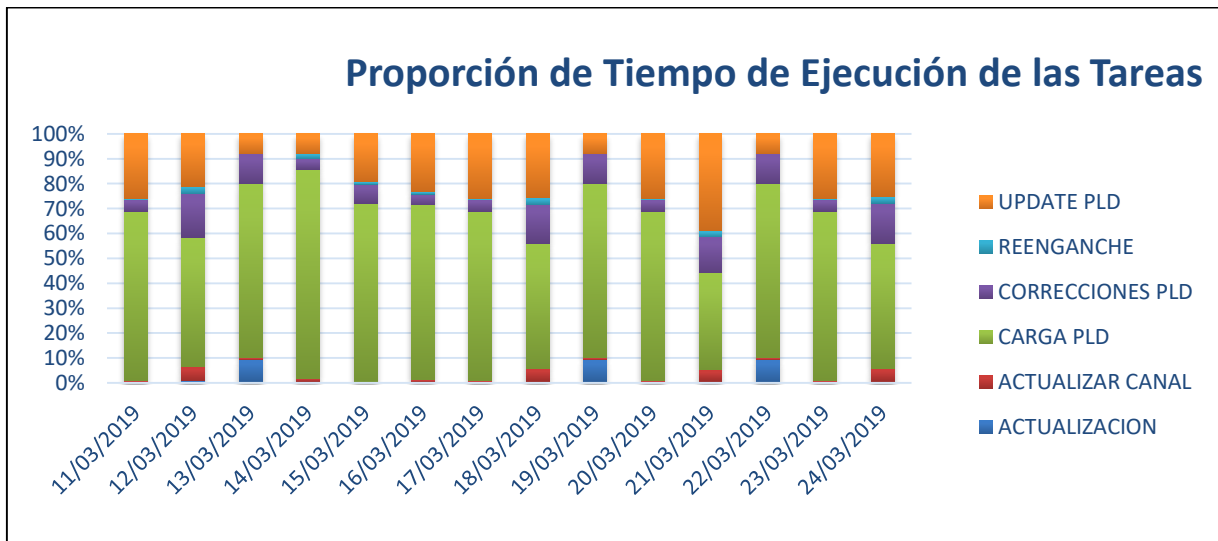


Figura 42 Tareas propuesta contenedor actualizar correcciones

Fuente: Elaboración propia

Tabla 12 Análisis de los tiempos de las tareas del contenedor actualizar correcciones propuesto

	MIN	MAX	PROM	MEDIAN
ACTUALIZACION	0.05	4.41	1.14	0.31
ACTUALIZAR CANAL	0.13	3.43	1.18	0.37
CARGA PLD	22.55	33.71	30.48	32.91
CORRECCIONES PLD	1.12	10.47	5.20	4.01
REENGANCHE	0.03	1.79	0.62	0.17
UPDATE PLD	2.11	28.53	10.93	12.57

Fuente: Elaboración propia

La tarea carga PLD tiene el mayor tiempo de duración promedia con 30.48 minutos, si bien sigue siendo alto en el punto 3.3 validaremos si se pudo disminuir el tiempo en comparación de la ETL anterior.

3.3 Validación del Modelo Propuesto

En esta sección del trabajo de investigación se procede a comparar los resultados obtenidos entre la ETL anterior y la ETL propuesta.

Se tomará en consideración la forma de trabajo y análisis que se ha venido trabajando desde los puntos anteriores, comparando los tiempos a nivel contenedor y luego a nivel tarea, centrándose en el contenedor actualizar correcciones.

3.3.1 Validación a Nivel Contenedor

Como se puede observar en la Tabla 13, el performance del ETL propuesto es mayor, permitiendo disminuir considerablemente los tiempos de ejecución, aun así, se puede notar que el contenedor que mantiene los niveles más altos en tiempo de ejecución sigue siendo actualizar correcciones.

Tabla 13 Comparativo de tiempos entre proceso actual y propuesta por contenedor

	MIN		MAX		PROM	
	ANTERIOR	PROPUESTA	ANTERIOR	PROPUESTA	ANTERIOR	PROPUESTA
ACTUALIZAR CORRECCIONES	37.60	26.88	96.50	73.20	62.42	49.55
CARGA TEMPORALES	12.50	9.09	22.80	15.69	18.69	11.92
DATA ADICIONAL	1.50	0.34	7.20	4.14	4.01	1.71
SALDOS RETENIDOS	3.40	-	12.30	-	7.08	-
VALIDAR ACTIVOS	0.40	0.00	2.10	0.01	1.19	0.00

Fuente: Elaboración propia

Adicionalmente se puede observar que el tiempo antes utilizado para ejecutar el subproceso de saldos retenidos ya no es utilizado, puesto

que, gracias al mapeo de datos anteriormente realizado se pudo detectar la existencia de una base con la misma información.

3.3.2 Validación a Nivel Tarea

El mayor esfuerzo en la optimización de las consultas fue dirigido al contenedor actualizar correcciones y en específico a la tarea carga PLD, por lo que en esta sección revisaremos la variación de los tiempos de ejecución en el subproceso antes mencionado, en base a los resultados mostrados en la Tabla 14.

Tabla 14 Comparativo de tiempos entre proceso actual y propuesta por tarea

	MIN		MAX		PROM		MEDIAN	
	ANTERIOR	PROPUESTA	ANTERIOR	PROPUESTA	ANTERIOR	PROPUESTA	ANTERIOR	PROPUESTA
ACTUALIZACION	0.50	0.05	6.00	4.41	2.21	1.14	1.50	0.31
ACTUALIZAR CANAL	0.60	0.13	5.00	3.43	2.20	1.18	1.60	0.37
CARGA PLD	23.50	22.55	42.50	33.71	34.24	30.48	34.00	32.91
CORRECCIONES PLD	2.40	1.12	12.00	10.47	6.26	5.20	4.85	4.01
REENGANCHE	0.40	0.03	3.30	1.79	1.62	0.62	1.40	0.17
T_TC_AGENCIAS	0.40	-	2.10	-	1.06	-	1.00	-
UPDATE PLD	4.10	2.11	30.10	28.53	14.83	10.93	13.25	12.57

Fuente: Elaboración propia

Se observa que la solución propuesta disminuye el tiempo de ejecución de todas las tareas, no solo la de carga PLD que era en la cual se centra, en esta última se logró disminuir en promedio 4 minutos del tiempo de ejecución aplicando los conceptos de optimización consultas mencionados en la sección 3.2.2.

CONCLUSIONES

Del presente trabajo de investigación se concluye que:

La aplicación del mapeo de datos como la optimización de queries permitieron mejorar el rendimiento de los procesos ETL en el área de Inteligencia Comercial del Banco Pichincha, al disminuir los tiempos de ejecución y la complejidad de las consultas.

La dificultad para realizar modificaciones en los procesos ETL se vio disminuida al tener un diccionario de datos actualizado de nuestras fuentes disponibles, permitiendo también reemplazar las consultas complejas por otras más simples.

Los tiempos de ejecución del proceso ETL propuesto son menores a los del proceso actual, ya que con el mapeo de datos y la optimización de queries se logró identificar las sentencias que brindan un mayor performance y eliminar procesos que generaban información similar en distintas fuentes.

La optimización de las consultas permitió eliminar los errores de información duplicada o con valores nulos que se generaban con el proceso ETL anterior y con el mapeo de datos realizado se pudo identificar nuevas fuentes con información de mayor disponibilidad.

RECOMENDACIONES

Respecto a las conclusiones antes mencionadas, se proponen las siguientes recomendaciones:

La complejidad de las modificaciones se ve fuertemente disminuida al mantener actualizado el diccionario de las fuentes, ya que facilitan en gran medida el proceso de extracción y transformación de la información, además de ser una herramienta útil de consulta.

Para lograr optimizar el tiempo de ejecución de los procesos, es necesario mantener una capacitación constante acerca de las buenas prácticas al momento de armar una consulta, para así evitar caer en redundancias y minimizar el uso de tiempo y recursos de los servidores.

La disminución de errores en la carga de información, asegurando la calidad de los datos, es una de las principales tareas en el proceso ETL, por lo que siempre es recomendable realizar diagramas que permitan entender el flujo de la información y las transformaciones por las que han pasado.

Adicionalmente tomando en cuenta el estado actual de la infraestructura de datos del área de Inteligencia Comercial se recomienda:

Tener acceso a las fuentes principales de los OLTP para alimentar una data warehouse propiedad del área de Inteligencia Comercial permitiría disponibilizar la información y la generación de reportes a primera hora, ya que al consumir de la data warehouse externo al área, nos encontramos limitados a su disponibilidad.

La plataforma utilizada actualmente en el área de Inteligencia Comercial para la integración de datos es el Business Intelligence Development Studio 2008, la cual es una versión desfasada, la versión actual (SQL Server Data Tools) contiene grandes mejoras y facilita enormemente la construcción de las ETL, además de tener un motor de optimización de consultas mejorado.

BIBLIOGRAFÍA

- Banco Pichincha. (15 de 02 de 2019). *Banco Pichincha del Perú*. Obtenido de <https://www.pichincha.pe/>
- Bernabeu, R. D. (2010). *HEFESTO*. Córdoba, Argentina.
- Castro, J. (12 de Junio de 2015). *Blog Corponet*. Obtenido de <https://blog.corponet.com.mx/la-importancia-de-la-informacion-para-la-toma-de-decisiones-en-la-empresa>
- Deniz Miguez, C. (2003). *tatum*. Obtenido de http://www.tatum.es/intranet/tatum2003/fotos/pub_fichero355.pdf
- Duque, N., Hernández, E., Pérez, Á., Arroyave, A., & Espinosa, A. (2016). Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales. *Ciencia e Ingeniería Neogranadina*, 95-109.
- Esteban. (4 de Octubre de 2010). *Guíadev*. Obtenido de <https://guiadev.com/optimizando-consultas-lentas-sql/>
- Estevas, R. (2012). *Social Web Intelligence - SWI. Mánager Responsable de desarrollo*. Obtenido de <http://socialwebintelligence.com/>
- Frutos, R. G. (2016). Optimización de consultas en bases de datos relacionales.
- Galvis, E., Gómez, L., & Bustamante, A. (2013). ETL Processes modeling techniques: an alternatives review and its application in a BI solution development project. *Scientia et Technica*, 185-191.
- Gómez, J. (27 de Noviembre de 2018). *CEREM*. Obtenido de <https://www.cerem.pe/blog/habilidades-de-inteligencia-comercial>
- Graneros, N. (2017). *Evaluación de performance en bases de datos*. La Plata.
- Guo, S.-S., Sun, A.-B., Yuan, Z.-M., & Yue, Q. (2015). A New ETL Approach Based on Data Virtualization. *Journal of Computer Science and Technology*, 91-104.
- Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Canada: Wiley Publishing, Inc.

- Kimball, R., & Ross, M. (2002). *The data warehouse toolkit : the complete guide to dimensional modeling Second Edition*. Canadá: Wiley Computer Publishing.
- Marqués, M. P. (2015). *BUSINESS INTELLIGENCE Técnicas, herramientas y aplicaciones*. Madrid: RC Libros.
- Martinez, J. (23 de Julio de 2012). *IBM*. Obtenido de developerWorks: <https://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/gobierno-datos-pdf.pdf>
- Matute, G. (12 de Abril de 2013). *ESAN*. Obtenido de <https://www.esan.edu.pe/conexion/actualidad/2013/04/12/inteligencia-negocios-empresa/>
- Microsoft. (05 de 03 de 2017). *Documentación de SQL*. Obtenido de <https://docs.microsoft.com>
- POWERDATA SOLUTION S.L. (6 de Junio de 2017). *BLOG POWERDATA*. Obtenido de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl>
- Reyes, P. (2002). *Introducción al SQL de InterBase: DDL y DML*. Madrid: Danysoft Internacional.
- Rodriguez Padilla, J. M. (2014). *Cómo hacer inteligente su negocio: business intelligence a su alcance*. México: Grupo Editorial Patria.
- Rodríguez, V. J. (2008). *Administración de base de datos con SQL Server 2008*. Excellence Innova.
- Salazar, J. (2017). IMPLEMENTACIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA COMERCIAL DE LA EMPRESA AZALEIA - BASADO EN METODOLOGÍA ÁGIL SCRUM.
- Sinergia e Inteligencia de Negocio S.L. (s.f.). *Sinnexus.com*. Obtenido de https://www.sinnexus.com/business_intelligence/datamart.aspx
- Vassiliadis, P., Simitsis, A., Georgantas, P., Terrovitis, M., & Skiadopoulos, S. (2005). A generic and customizable framework for the design of ETL scenarios. *Information Systems*, 492-525.
- Vitt, E., Luckevich, M., & Mismar, S. (2002). *Business Intelligence. Técnicas De Análisis Para La Toma De Decisiones Estratégicas*. España: McGraw-Hill.

ANEXOS

Tiempos de Ejecución ETL anterior

A continuación, se muestran en la Tabla 15 los siete primeros días de los tiempos de ejecución recuperados del proceso ETL anterior, comprenden desde el 4 de febrero hasta el 10 de febrero inclusive.

Tabla 15 Tiempos de ejecución ETL anterior

FECHA	CONTAINER	TASK	HORA INICIO	HORA FIN
04/02/2019	VALIDAR ACTIVOS	VALIDAR ACTIVOS	07:30:04 a.m.	07:30:04 a.m.
04/02/2019	CARGA TEMPORALES	CARGA SALDOS Y CLIENTES	07:30:04 a.m.	07:37:25 a.m.
04/02/2019	CARGA TEMPORALES	CANCELACIONES	07:37:25 a.m.	07:40:02 a.m.
04/02/2019	CARGA TEMPORALES	LIMPIAR TABLAS	07:40:02 a.m.	07:40:12 a.m.
04/02/2019	CARGA TEMPORALES	CARGA SALDOS	07:40:12 a.m.	07:40:20 a.m.
04/02/2019	CARGA TEMPORALES	CARGA CLIENTES	07:40:20 a.m.	07:40:24 a.m.
04/02/2019	CARGA TEMPORALES	CARGA CANCELACIONES	07:40:24 a.m.	07:40:24 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:40:24 a.m.	08:13:19 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:13:19 a.m.	08:13:28 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:13:28 a.m.	08:26:05 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:26:05 a.m.	08:28:21 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:28:21 a.m.	08:28:33 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:28:33 a.m.	08:28:37 a.m.
04/02/2019	ACTUALIZAR CORRECCIONES	T_TC_AGENCIAS	08:28:37 a.m.	08:28:38 a.m.
04/02/2019	SALDOS RETENIDOS	TRUNCAR SALDOS RETENIDOS	08:28:38 a.m.	08:28:38 a.m.
04/02/2019	SALDOS RETENIDOS	DESEMBOLSOS ACTUAL	08:28:38 a.m.	08:29:26 a.m.
04/02/2019	SALDOS RETENIDOS	RETENIDOS ACTUAL	08:29:26 a.m.	08:29:28 a.m.
04/02/2019	SALDOS RETENIDOS	DESEMBOLSOS DEL MES ANTERIOR	08:29:28 a.m.	08:30:15 a.m.
04/02/2019	SALDOS RETENIDOS	RETENIDOS MES ANTERIOR	08:30:15 a.m.	08:30:16 a.m.
04/02/2019	DATA ADICIONAL	LB_INFO_CANALES	08:30:16 a.m.	08:30:36 a.m.

04/02/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:30:36 a.m.	08:31:03 a.m.
05/02/2019	VALIDAR ACTIVOS	VALIDAR ACTIVOS	07:30:08 a.m.	07:30:08 a.m.
05/02/2019	CARGA TEMPORALES	CARGA SALDOS Y CLIENTES	07:30:08 a.m.	07:38:34 a.m.
05/02/2019	CARGA TEMPORALES	CANCELACIONES	07:38:34 a.m.	07:42:53 a.m.
05/02/2019	CARGA TEMPORALES	LIMPIAR TABLAS	07:42:53 a.m.	07:43:08 a.m.
05/02/2019	CARGA TEMPORALES	CARGA SALDOS	07:43:08 a.m.	07:43:22 a.m.
05/02/2019	CARGA TEMPORALES	CARGA CLIENTES	07:43:22 a.m.	07:43:30 a.m.
05/02/2019	CARGA TEMPORALES	CARGA CANCELACIONES	07:43:30 a.m.	07:43:31 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:43:31 a.m.	08:24:31 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:24:31 a.m.	08:24:58 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:24:58 a.m.	08:53:29 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:53:29 a.m.	09:03:58 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	09:03:58 a.m.	09:07:24 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	09:07:24 a.m.	09:09:11 a.m.
05/02/2019	ACTUALIZAR CORRECCIONES	T_TC_AGENCIAS	09:09:11 a.m.	09:09:12 a.m.
05/02/2019	SALDOS RETENIDOS	TRUNCAR SALDOS RETENIDOS	09:09:13 a.m.	09:09:13 a.m.
05/02/2019	SALDOS RETENIDOS	DESEMBOLSOS ACTUAL	09:09:13 a.m.	09:09:54 a.m.
05/02/2019	SALDOS RETENIDOS	RETENIDOS ACTUAL	09:09:54 a.m.	09:09:56 a.m.
05/02/2019	SALDOS RETENIDOS	DESEMBOLSOS DEL MES ANTERIOR	09:09:56 a.m.	09:11:05 a.m.
05/02/2019	SALDOS RETENIDOS	RETENIDOS MES ANTERIOR	09:11:05 a.m.	09:11:07 a.m.
05/02/2019	DATA ADICIONAL	LB_INFO_CANALES	09:11:08 a.m.	09:11:47 a.m.
05/02/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	09:11:47 a.m.	09:15:15 a.m.
06/02/2019	VALIDAR ACTIVOS	VALIDAR ACTIVOS	07:30:05 a.m.	07:30:06 a.m.
06/02/2019	CARGA TEMPORALES	CARGA SALDOS Y CLIENTES	07:30:06 a.m.	07:37:30 a.m.
06/02/2019	CARGA TEMPORALES	CANCELACIONES	07:37:30 a.m.	07:42:01 a.m.
06/02/2019	CARGA TEMPORALES	LIMPIAR TABLAS	07:42:01 a.m.	07:42:13 a.m.
06/02/2019	CARGA TEMPORALES	CARGA SALDOS	07:42:13 a.m.	07:42:21 a.m.
06/02/2019	CARGA TEMPORALES	CARGA CLIENTES	07:42:21 a.m.	07:42:28 a.m.
06/02/2019	CARGA TEMPORALES	CARGA CANCELACIONES	07:42:28 a.m.	07:42:28 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:42:28 a.m.	08:15:23 a.m.

06/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:15:23 a.m.	08:19:48 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:19:48 a.m.	08:23:31 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:23:31 a.m.	08:29:06 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:29:06 a.m.	08:29:29 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:29:29 a.m.	08:29:31 a.m.
06/02/2019	ACTUALIZAR CORRECCIONES	T_TC_AGENCIAS	08:29:31 a.m.	08:29:33 a.m.
06/02/2019	SALDOS RETENIDOS	TRUNCAR SALDOS RETENIDOS	08:29:33 a.m.	08:29:33 a.m.
06/02/2019	SALDOS RETENIDOS	DESEMBOLSOS ACTUAL	08:29:33 a.m.	08:30:48 a.m.
06/02/2019	SALDOS RETENIDOS	RETENIDOS ACTUAL	08:30:48 a.m.	08:30:49 a.m.
06/02/2019	SALDOS RETENIDOS	DESEMBOLSOS DEL MES ANTERIOR	08:30:49 a.m.	08:31:16 a.m.
06/02/2019	SALDOS RETENIDOS	RETENIDOS MES ANTERIOR	08:31:16 a.m.	08:31:17 a.m.
06/02/2019	DATA ADICIONAL	LB_INFO_CANALES	08:31:17 a.m.	08:31:20 a.m.
06/02/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:31:20 a.m.	08:31:43 a.m.
07/02/2019	VALIDAR ACTIVOS	VALIDAR ACTIVOS	07:30:04 a.m.	07:30:04 a.m.
07/02/2019	CARGA TEMPORALES	CARGA SALDOS Y CLIENTES	07:30:06 a.m.	07:37:30 a.m.
07/02/2019	CARGA TEMPORALES	CANCELACIONES	07:29:20 p.m.	07:31:50 p.m.
07/02/2019	CARGA TEMPORALES	LIMPIAR TABLAS	07:31:50 p.m.	07:31:52 p.m.
07/02/2019	CARGA TEMPORALES	CARGA SALDOS	07:31:52 p.m.	07:32:00 p.m.
07/02/2019	CARGA TEMPORALES	CARGA CLIENTES	07:32:00 p.m.	07:32:07 p.m.
07/02/2019	CARGA TEMPORALES	CARGA CANCELACIONES	07:32:07 p.m.	07:32:07 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:32:07 p.m.	07:54:40 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	07:54:40 p.m.	07:54:50 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	07:54:50 p.m.	07:56:57 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	07:56:57 p.m.	07:58:04 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	07:58:04 p.m.	07:58:25 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	07:58:25 p.m.	07:59:00 p.m.
07/02/2019	ACTUALIZAR CORRECCIONES	T_TC_AGENCIAS	07:59:00 p.m.	07:59:00 p.m.
07/02/2019	SALDOS RETENIDOS	TRUNCAR SALDOS RETENIDOS	07:59:00 p.m.	07:59:01 p.m.

07/02/2019	SALDOS RETENIDOS	DESEMBOLSOS ACTUAL	07:59:01 p.m.	07:59:37 p.m.
07/02/2019	SALDOS RETENIDOS	RETENIDOS ACTUAL	07:59:37 p.m.	07:59:38 p.m.
07/02/2019	SALDOS RETENIDOS	DESEMBOLSOS DEL MES ANTERIOR	07:59:38 p.m.	08:00:52 p.m.
07/02/2019	SALDOS RETENIDOS	RETENIDOS MES ANTERIOR	08:00:52 p.m.	08:00:52 p.m.
07/02/2019	DATA ADICIONAL	LB_INFO_CANALES	08:00:52 p.m.	08:01:07 p.m.
07/02/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:01:07 p.m.	08:02:49 p.m.
08/02/2019	VALIDAR ACTIVOS	VALIDAR ACTIVOS	08:17:42 p.m.	08:17:43 p.m.
08/02/2019	CARGA TEMPORALES	CARGA SALDOS Y CLIENTES	08:17:43 p.m.	08:23:21 p.m.
08/02/2019	CARGA TEMPORALES	CANCELACIONES	08:23:21 p.m.	08:26:37 p.m.
08/02/2019	CARGA TEMPORALES	LIMPIAR TABLAS	08:26:37 p.m.	08:26:38 p.m.
08/02/2019	CARGA TEMPORALES	CARGA SALDOS	08:26:38 p.m.	08:26:44 p.m.
08/02/2019	CARGA TEMPORALES	CARGA CLIENTES	08:26:44 p.m.	08:26:48 p.m.
08/02/2019	CARGA TEMPORALES	CARGA CANCELACIONES	08:26:48 p.m.	08:26:48 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	08:26:48 p.m.	08:49:25 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:49:25 p.m.	08:49:28 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:49:28 p.m.	08:55:33 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:55:33 p.m.	08:57:59 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:57:59 p.m.	08:58:07 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:58:07 p.m.	08:58:23 p.m.
08/02/2019	ACTUALIZAR CORRECCIONES	T_TC_AGENCIAS	08:58:23 p.m.	08:58:24 p.m.
08/02/2019	SALDOS RETENIDOS	TRUNCAR SALDOS RETENIDOS	08:58:24 p.m.	08:58:24 p.m.
08/02/2019	SALDOS RETENIDOS	DESEMBOLSOS ACTUAL	08:58:24 p.m.	08:58:51 p.m.
08/02/2019	SALDOS RETENIDOS	RETENIDOS ACTUAL	08:58:51 p.m.	08:58:52 p.m.
08/02/2019	SALDOS RETENIDOS	DESEMBOLSOS DEL MES ANTERIOR	08:58:52 p.m.	08:59:19 p.m.
08/02/2019	SALDOS RETENIDOS	RETENIDOS MES ANTERIOR	08:59:19 p.m.	08:59:19 p.m.
08/02/2019	DATA ADICIONAL	LB_INFO_CANALES	08:59:19 p.m.	08:59:31 p.m.

Fuente: Elaboración propia

Tiempos de Ejecución ETL propuesta

Al igual que para los tiempos de ejecución de la ETL anterior, se muestra una parte de los datos recuperados, comprenden desde el 20 al 24 de marzo inclusive en la Tabla 16.

Tabla 16 Tiempos de ejecución ETL propuesta

FECHA	CONTAINER	TASK	HORA INICIO	HORA FIN
20/03/2019	VALIDAR INFORMACION	VALIDAR INFORMACION	07:30:04 a.m.	07:30:04 a.m.
20/03/2019	CARGA INTERMEDIA	CARGA SALDOS Y CLIENTES	07:30:04 a.m.	07:37:25 a.m.
20/03/2019	CARGA INTERMEDIA	CANCELACIONES	07:37:25 a.m.	07:40:02 a.m.
20/03/2019	CARGA INTERMEDIA	LIMPIAR TABLAS	07:40:02 a.m.	07:40:12 a.m.
20/03/2019	CARGA INTERMEDIA	CARGA SALDOS	07:40:12 a.m.	07:40:20 a.m.
20/03/2019	CARGA INTERMEDIA	CARGA CLIENTES	07:40:20 a.m.	07:40:24 a.m.
20/03/2019	CARGA INTERMEDIA	CARGA CANCELACIONES	07:40:24 a.m.	07:40:24 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:40:24 a.m.	08:13:19 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:13:19 a.m.	08:13:28 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:13:28 a.m.	08:26:05 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:26:05 a.m.	08:28:21 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:28:21 a.m.	08:28:33 a.m.
20/03/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:28:33 a.m.	08:28:37 a.m.
20/03/2019	DATA ADICIONAL	T_TC_AGENCIAS	08:28:37 a.m.	08:28:38 a.m.
20/03/2019	DATA ADICIONAL	LB_INFO_CANALES	08:30:16 a.m.	08:30:36 a.m.
20/03/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:30:36 a.m.	08:31:03 a.m.
21/03/2019	VALIDAR INFORMACION	VALIDAR INFORMACION	07:30:08 a.m.	07:30:08 a.m.
21/03/2019	CARGA INTERMEDIA	CARGA SALDOS Y CLIENTES	07:30:08 a.m.	07:38:34 a.m.
21/03/2019	CARGA INTERMEDIA	CANCELACIONES	07:38:34 a.m.	07:42:53 a.m.
21/03/2019	CARGA INTERMEDIA	LIMPIAR TABLAS	07:42:53 a.m.	07:43:08 a.m.
21/03/2019	CARGA INTERMEDIA	CARGA SALDOS	07:43:08 a.m.	07:43:22 a.m.
21/03/2019	CARGA INTERMEDIA	CARGA CLIENTES	07:43:22 a.m.	07:43:30 a.m.
21/03/2019	CARGA INTERMEDIA	CARGA CANCELACIONES	07:43:30 a.m.	07:43:31 a.m.

21/03/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:43:31 a.m.	08:12:03 a.m.
21/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:24:31 a.m.	08:24:58 a.m.
21/03/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:24:58 a.m.	08:53:29 a.m.
21/03/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:53:29 a.m.	09:03:58 a.m.
21/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	09:03:58 a.m.	09:07:24 a.m.
21/03/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	09:07:24 a.m.	09:09:11 a.m.
21/03/2019	DATA ADICIONAL	T_TC_AGENCIAS	09:09:11 a.m.	09:09:12 a.m.
21/03/2019	DATA ADICIONAL	LB_INFO_CANALES	09:11:08 a.m.	09:11:47 a.m.
21/03/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	09:11:47 a.m.	09:15:15 a.m.
22/03/2019	VALIDAR INFORMACION	VALIDAR INFORMACION	07:30:05 a.m.	07:30:06 a.m.
22/03/2019	CARGA INTERMEDIA	CARGA SALDOS Y CLIENTES	07:30:06 a.m.	07:37:30 a.m.
22/03/2019	CARGA INTERMEDIA	CANCELACIONES	07:37:30 a.m.	07:42:01 a.m.
22/03/2019	CARGA INTERMEDIA	LIMPIAR TABLAS	07:42:01 a.m.	07:42:13 a.m.
22/03/2019	CARGA INTERMEDIA	CARGA SALDOS	07:42:13 a.m.	07:42:21 a.m.
22/03/2019	CARGA INTERMEDIA	CARGA CLIENTES	07:42:21 a.m.	07:42:28 a.m.
22/03/2019	CARGA INTERMEDIA	CARGA CANCELACIONES	07:42:28 a.m.	07:42:28 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:42:28 a.m.	08:15:23 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:15:23 a.m.	08:19:48 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:19:48 a.m.	08:23:31 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:23:31 a.m.	08:29:06 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:29:06 a.m.	08:29:29 a.m.
22/03/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:29:29 a.m.	08:29:31 a.m.
22/03/2019	DATA ADICIONAL	T_TC_AGENCIAS	08:29:31 a.m.	08:29:33 a.m.
22/03/2019	DATA ADICIONAL	LB_INFO_CANALES	08:31:17 a.m.	08:31:20 a.m.
22/03/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:31:20 a.m.	08:31:43 a.m.
23/03/2019	VALIDAR INFORMACION	VALIDAR INFORMACION	07:30:04 a.m.	07:30:04 a.m.
23/03/2019	CARGA INTERMEDIA	CARGA SALDOS Y CLIENTES	07:30:04 a.m.	07:37:25 a.m.
23/03/2019	CARGA INTERMEDIA	CANCELACIONES	07:37:25 a.m.	07:40:02 a.m.
23/03/2019	CARGA INTERMEDIA	LIMPIAR TABLAS	07:40:02 a.m.	07:40:12 a.m.

23/03/2019	CARGA INTERMEDIA	CARGA SALDOS	07:40:12 a.m.	07:40:20 a.m.
23/03/2019	CARGA INTERMEDIA	CARGA CLIENTES	07:40:20 a.m.	07:40:24 a.m.
23/03/2019	CARGA INTERMEDIA	CARGA CANCELACIONES	07:40:24 a.m.	07:40:24 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:40:24 a.m.	08:13:19 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:13:19 a.m.	08:13:28 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:13:28 a.m.	08:26:05 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:26:05 a.m.	08:28:21 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	08:28:21 a.m.	08:28:33 a.m.
23/03/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	08:28:33 a.m.	08:28:37 a.m.
23/03/2019	DATA ADICIONAL	T_TC_AGENCIAS	08:28:37 a.m.	08:28:38 a.m.
23/03/2019	DATA ADICIONAL	LB_INFO_CANALES	08:30:16 a.m.	08:30:36 a.m.
23/03/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	08:30:36 a.m.	08:31:03 a.m.
24/03/2019	VALIDAR INFORMACION	VALIDAR INFORMACION	07:30:08 a.m.	07:30:08 a.m.
24/03/2019	CARGA INTERMEDIA	CARGA SALDOS Y CLIENTES	07:30:08 a.m.	07:38:34 a.m.
24/03/2019	CARGA INTERMEDIA	CANCELACIONES	07:38:34 a.m.	07:42:53 a.m.
24/03/2019	CARGA INTERMEDIA	LIMPIAR TABLAS	07:42:53 a.m.	07:43:08 a.m.
24/03/2019	CARGA INTERMEDIA	CARGA SALDOS	07:43:08 a.m.	07:43:22 a.m.
24/03/2019	CARGA INTERMEDIA	CARGA CLIENTES	07:43:22 a.m.	07:43:30 a.m.
24/03/2019	CARGA INTERMEDIA	CARGA CANCELACIONES	07:43:30 a.m.	07:43:31 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	CARGA PLD	07:43:31 a.m.	08:16:23 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZACION	08:24:31 a.m.	08:24:58 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	UPDATE PLD	08:24:58 a.m.	08:41:29 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	CORRECCIONES PLD	08:53:29 a.m.	09:03:58 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	ACTUALIZAR CANAL	09:03:58 a.m.	09:07:24 a.m.
24/03/2019	ACTUALIZAR CORRECCIONES	REENGANCHE	09:07:24 a.m.	09:09:11 a.m.
24/03/2019	DATA ADICIONAL	T_TC_AGENCIAS	09:09:11 a.m.	09:09:12 a.m.
24/03/2019	DATA ADICIONAL	LB_INFO_CANALES	09:11:08 a.m.	09:11:47 a.m.
24/03/2019	DATA ADICIONAL	INDICADORES CONSOLIDADO	09:11:47 a.m.	09:15:15 a.m.

Fuente: Elaboración propia

Formato de Diccionario de Datos

Los diccionarios de datos se encuentran divididos en dos partes, la primera, mostrada en la Tabla 17, contiene los detalles de los campos de la fuente a tratar.

Tabla 17 Diccionario de datos, primera parte

fuelle origen	servido	base	tabla	campos	Tipo Datc	Calculo	Descripci
propia	SBIBFPR15	BFPMARKETING	agencias_ic	segmento	vchar		segmento de
propia	SBIBFPR15	BFPMARKETING	agencias_ic	ZONAL	vchar		zonal definic
propia	SBIBFPR15	BFPMARKETING	agencias_ic	CLUSTER	vchar		cluster defin
propia	SBIBFPR15	BFPMARKETING	agencias_ic	codigoagencia	float		codigo agenc
propia	SBIBFPR15	BFPMARKETING	agencias_ic	AGENCIA	nvarchar		agencia
propia	SBIBFPR15	BFPMARKETING	agencias_ic	ZONA	nvarchar		
propia	SBIBFPR15	BFPMARKETING	agencias_ic	RED	nvarchar		
propia	SBIBFPR15	BFPMARKETING	agencias_ic	TIPO	nvarchar		Banco o Cars
propia	SBIBFPR15	BFPMARKETING	agencias_ic	ESTADO	nvarchar		Lima / Provir
propia	SBIBFPR15	BFPMARKETING	agencias_ic	DPTO#	nvarchar		Departamen
propia	SBIBFPR15	BFPMARKETING	agencias_ic	PROV#	nvarchar		Provincia do
propia	SBIBFPR15	BFPMARKETING	agencias_ic	DISTRITO	nvarchar		Distrito don
propia	SBIBFPR15	BFPMARKETING	agencias_ic	BANCARIA	nvarchar		
propia	SBIBFPR15	BFPMARKETING	agencias_ic	SBS	nvarchar		
propia	SBIBFPR15	BFPMARKETING	agencias_ic	CAT#	nvarchar		
propia	SBIBFPR15	BFPMARKETING	agencias_ic	DIRECCION	nvarchar		Dirección
propia	SBIBFPR15	BFPMARKETING	agencias_ic	TELEFONO	nvarchar		telefono

Fuente: Elaboración propia

En la segunda parte, mostrada en la Tabla 18, se pueden observar el detalle de los datos de las fuentes de origen de la tabla.

Tabla 18 Diccionario de datos, segunda parte

ARCHIVO ORIGEN	CAMPO ORIGEN	TIPO DE DATO	TRANSFORMACIÓN	VALORES

Fuente: Elaboración propia