



**Producto Cloud B2B: Unificación y optimización del proceso de gestión de la información
del producto Cloud de Tigo Business**

Stefany Carolina De Sousa Lora

Informe de práctica empresarial para optar al título de Ingeniera de Sistemas

Asesor

Luis Hernando Silva Florez, Magister (MSc) en Ingeniería de Sistemas

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería de Sistemas

Medellín

2024

Cita	De Sousa Lora [1]
Referencia	[1] S. C. De Sousa Lora, “Producto Cloud B2B: Unificación y optimización del proceso de gestión de la información del producto Cloud de Tigo Business”, Trabajo de grado profesional, Ingeniería de Sistemas, Universidad de Antioquia, Medellín, 2024.
Estilo IEEE (2020)	



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio César Saldarriaga.

Jefe departamento: Danny Alejandro Múnera Ramírez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

TABLA DE CONTENIDO

RESUMEN	6
ABSTRACT	7
I. INTRODUCCIÓN	8
II. OBJETIVOS	9
A. Objetivo general	9
B. Objetivos específicos	9
III. MARCO TEÓRICO	10
IV. METODOLOGÍA	12
V. RESULTADOS	16
VI. ANÁLISIS	25
VII. CONCLUSIONES	27
REFERENCIAS	28

LISTA DE FIGURAS

Fig. 1. Proceso ETL	12
Fig. 2. Estandarización de los datos	17
Fig. 3. Ejemplo de nuevas columnas en la tabla de facturación	17
Fig. 4. Tabla inicial de consulta Cloud Server	18
Fig. 5. Tabla resultante tras la transformación	18
Fig. 6. Limpieza y conversión de campos numéricos	19
Fig. 7. Cambio de formato en campos de fecha	19
Fig. 8. Proceso de completar registros vacíos	20
Fig. 9. Ejemplo de nuevas columnas en tabla de altas y bajas de Cloud Server	21
Fig. 10. Tablas de facturación mes actual e histórico	22
Fig. 11. Tablas con el detalle del cliente mes actual e histórico	23
Fig. 12. Tabla final de altas y bajas de Cloud Server	23
Fig. 13. Tablero de facturación en Power BI	24
Fig. 14. Variaciones en la facturación por selección de filtros	24

SIGLAS, ACRÓNIMOS Y ABREVIATURAS

SQL	Structured Query Language
ETL	Extract, transform and load
B2B	Business-to-business

RESUMEN

Este proyecto tiene como objetivo mejorar la calidad de la información del producto Cloud 360 de Tigo Business y contribuir en la optimización de su gestión. Para lograrlo, se implementó un proceso de ETL haciendo uso de herramientas como Python y SQL con la finalidad de extraer, limpiar, clasificar y unificar todos los datos del producto que se encontraban dispersos en diversas fuentes. Con la data resultante, se construyeron tablas para almacenar la información sobre los clientes y servicios, así como de su respectiva facturación mensual. Adicionalmente, se diseñó un tablero en Power BI que se conecta a las tablas para presentar los datos más importantes y permitir que se tenga un mejor conocimiento sobre el producto y su estado financiero. El desarrollo de este proceso fue clave para lograr una clasificación e integración completa de la información, facilitando de esta manera su acceso y manejo. Esta nueva data contribuyó a su vez a la realización de informes detallados y confiables para un análisis y toma de decisiones más acertada.

***Palabras clave:* Cloud, ETL, Python, SQL, información, análisis, Power BI.**

ABSTRACT

The objective of this project is to improve the quality of the information related to Tigo's Business Cloud 360 product and contribute to the optimization of its management. To achieve this, an ETL process was implemented using tools such as Python and SQL in order to extract, clean, classify and unify all the available product data that was scattered in various sources. With the resulting data, tables were built to store information related to the customers and products, as well as the service's respective monthly billing. Additionally, a Power BI dashboard was designed that connects to the tables to present the most significant data and allow a better understanding of the product and its financial situation. The development of this process was key to achieve a complete classification and integration of the information, thus facilitating its access and management. This new data contributed in turn to the creation of detailed and reliable reports for more accurate analysis and decision-making.

***Keywords:* Cloud, ETL, Python, SQL, information, analysis, Power BI.**

I. INTRODUCCIÓN

Tigo es una empresa de telecomunicaciones radicada en Colombia que ofrece una amplia variedad de servicios tanto a clientes como empresas. Entre los servicios ofrecidos se encuentra la telefonía fija y móvil, internet fijo, televisión por cable o satélite, entre otras soluciones digitales de índole corporativa. Debido al gran portafolio de productos y servicios ofrecidos por la compañía, el sector encargado de la relación comercial con otras empresas, también conocido como B2B, actualmente enfrenta un desafío muy grande con el manejo de la información, específicamente de los productos Cloud ofrecidos. Esta dificultad en la gestión de la información se debe a que los datos se encuentran generalizados o mal clasificados y dispersos en distintas bases, dificultando su distinción, comprensión y análisis para la toma de decisiones.

Por este motivo, surge la necesidad de aplicar un proceso de ETL (extracción, transformación y carga), realizando inicialmente una búsqueda exhaustiva de la información necesaria en diversas fuentes. Posteriormente, se hace uso de Python y scripts SQL con el fin de extraer, limpiar, clasificar e integrar toda la data disponible sobre estos productos. El propósito de esto es convertir los datos extraídos en información confiable que permita distinguir los diferentes servicios y conocer su estado de facturación real. Adicionalmente, para facilitar un análisis más detallado y una toma de decisiones acertada, se presenta un tablero construido en Power BI con información clave sobre las altas y bajas de productos específicos, la cantidad de clientes y servicios, y la respectiva facturación mensual asociada a los mismos.

Es importante mencionar que, debido a la numerosa cantidad de datos y variedad de servicios Cloud existentes, fue necesario reducir el alcance del proyecto para enfocarse únicamente en un solo producto conocido como Tigo Cloud o Cloud 360. Sin embargo, se espera que en el futuro, se pueda continuar ampliando esta solución para cubrir los productos Cloud restantes, logrando así una optimización aún más completa de la gestión de la información.

II. OBJETIVOS

A. Objetivo general

Unificar la información del producto Cloud 360 de Tigo Business y optimizar su gestión.

B. Objetivos específicos

- Identificar las múltiples fuentes de datos a utilizar, como bases de datos, documentos o archivos.
- Buscar y extraer de dichas fuentes todos los identificadores únicos que correspondan a servicios Tigo Cloud y clasificarlos.
- Realizar cruces entre los identificadores encontrados y las bases correspondientes para obtener el detalle de los clientes y la facturación clasificada por tipo de servicio.
- Transformar los datos para garantizar que se ajusten a la necesidad de la solución. Esto incluye: limpieza de datos, estandarización, conversión de formatos, eliminación y creación de columnas, entre otros procesos.
- Cargar los datos transformados en nuevas tablas destino.
- Construir un tablero que presente información necesaria y relevante para el análisis y la toma de decisiones.

III. MARCO TEÓRICO

Las siglas B2B corresponden al término “business to business”, traducido al español como “de negocio a negocio”. Este concepto hace referencia al intercambio de servicios o productos entre empresas [1]. Se diferencia de las siglas B2C que significa “de empresa a consumidor” y representa la venta de productos y servicios a personas naturales o consumidores [2].

ETL se refiere al proceso que permite extraer, transformar y cargar datos. Durante este procedimiento se “recopilan datos de distintas fuentes para luego unificarlos a fin de facilitar el descubrimiento, la generación de informes, el análisis y la toma de decisiones” [3]. Los datos pueden ser obtenidos de diferentes tipos de fuente, cada una con una estructura única, por lo que resulta necesario aplicar procesos de transformación para asegurar que los datos conserven un mismo formato que permita su integración. Los datos procesados finalmente son almacenados en bases de datos, lagos de datos u otros almacenes de datos dependiendo de la necesidad de la implementación [3].

Los servicios Cloud son “infraestructuras, plataformas o sistemas de software que los proveedores externos alojan y ponen a disposición de los usuarios a través de Internet. Estos servicios facilitan el flujo de datos de los usuarios a través de Internet, desde los clientes frontend hasta los sistemas de los proveedores, y viceversa” [4].

Python es un lenguaje de programación fácil de aprender y usar, que permite desarrollar software, aplicaciones web e incluso es ampliamente utilizado en el área de machine learning y ciencias de datos. Este lenguaje es gratis y de código abierto, permitiendo que se pueda ejecutar en muchas plataformas y que se integre correctamente a todos los tipos de sistemas operativos [5].

Las siglas SQL, en español se refieren a “lenguaje de consulta estructurada” y corresponde a un lenguaje de programación que permite procesar información en una base de datos de tipo relacional. Una base de datos relacional es aquella que organiza la información en

tablas con sus respectivas filas y columnas y que permite establecer relaciones con otras tablas. Se puede utilizar el lenguaje SQL para construir consultas que permitan buscar, extraer, actualizar, eliminar o insertar información en las bases de datos [6].

Power BI es un software que permite construir tableros e informes interactivos y atractivos visualmente a partir de la importación de datos de diferentes fuentes, como pueden ser hojas de cálculo de Excel, bases de datos locales o en la nube, entre otras. Esta herramienta facilita la interpretación de la información y permite descubrir datos relevantes que optimizan la toma de decisiones en las empresas [7].

IV. METODOLOGÍA

Durante el desarrollo del proyecto se emplearon diversas herramientas que permitieron el cumplimiento de los objetivos generales y específicos planteados. Particularmente, se trabajó con los lenguajes de programación Python y SQL, y el gestor de bases de datos DBeaver. Además, se usaron módulos y librerías específicas de Python como Pandas, pyodbc, cx_Oracle, re, numpy y unicodedata para establecer las conexiones necesarias a las bases de datos, extraer y posteriormente transformar la información. Se utilizó la herramienta Power BI para construir el tablero que permitiría la visualización final de los datos, y finalmente, para facilitar la comunicación durante el proyecto, se utilizó la herramienta Teams. Con esta aplicación se realizaron reuniones periódicas para evaluar los avances de la implementación y aclarar bloqueos que pudieran surgir en el proceso.

En cuanto a la aplicación del proceso ETL y la construcción del tablero, a continuación, se describen detalladamente los aspectos metodológicos que se llevaron a cabo para su implementación [8, fig.1].



Fig 1. Proceso ETL.

- **Comprensión del problema y entorno de negocio**
 - Se comprendió el contexto del proyecto en el entorno de negocio.
 - Se estudió la necesidad específica de la solución, se definieron los objetivos y se acotó el alcance del proyecto para cubrir únicamente el producto Tigo Cloud.
 - Se identificaron los requisitos necesarios para llevar a cabo el proceso ETL y la construcción del tablero.

- **Identificación y exploración de las fuentes de datos**
 - Se identificaron las bases de datos que contienen información variada sobre los servicios Tigo Cloud.
 - Se inició la exploración de cada tabla para entender su estructura y contenido.
 - Se definieron las condiciones y los filtros necesarios que se debían tener en cuenta durante el proceso de rastreo de la información de los servicios en las bases de datos.

- **Extracción de la información**
 - Se emplearon módulos de Python como pyodbc y cx_Oracle para entablar la conexión a las distintas bases de datos fuente, sean SQL Server u Oracle. Posteriormente, se ejecutaron consultas SQL que permitieron extraer la información general de los clientes, servicios y facturación, así como también, los identificadores únicos asociados a los servicios Tigo Cloud.
 - Con la ayuda de la librería Pandas, se agruparon los datos obtenidos en distintos DataFrames y se clasificaron los identificadores hallados según el tipo de servicio al que pertenecen.
 - Se utilizó la librería Regular Expression (re) para eliminar algunos patrones presentes en los IDs que afectarían las demás operaciones a realizar. Adicionalmente, se removieron otros errores indeseados en los datos, como espacios, guiones, puntos, asteriscos y barras diagonales.
 - Haciendo uso nuevamente del módulo Pandas, se efectuaron los cruces necesarios entre las tablas que contienen los IDs únicos de los servicios y aquellos que tienen los datos de los clientes y la facturación. A raíz de esto, se obtuvo el detalle

específico de los clientes y la facturación mensual del producto Tigo Cloud con su respectiva clasificación por tipo de servicio.

- Por último, se realizó una consulta SQL adicional para obtener el detalle de las fechas de inicio y finalización de los servicios asociados a Cloud Server, el cual es un producto dentro del portafolio de Tigo Cloud. Esto con la finalidad de poder identificar cuántos de estos servicios se encuentran activos o cancelados.

- Transformación de los datos
 - Después de extraer la información necesaria, se continuó con el proceso de limpieza y conversión de formatos en Python. Se excluyeron valores duplicados en los datos y se eliminaron espacios en blanco. Además, se empleó la librería unicodedata para normalizar y eliminar acentos en las cadenas de texto de algunos campos ya que generaban variaciones innecesarias de un mismo dato.
 - Se eliminaron columnas sobrantes y se añadieron otras nuevas de acuerdo a la necesidad de cada DataFrame.
 - Respecto a la consulta sobre los servicios Cloud Server, fue necesario pivotar la tabla resultante debido a su estructura vertical compleja. Así mismo, se renombraron las columnas y se aplicaron procesos de limpieza y conversión de formatos en los datos de tipo fecha y numéricos. Para esto, se aplicaron métodos específicos de la librería Pandas como `pivot_table()`, `rename()` y `astype()`.
 - En la medida de lo posible, también se llenaron campos con registros vacíos para evitar datos incompletos que pudieran afectar la calidad de la información.

- Carga de la información
 - Con el uso nuevamente del módulo pyodbc, se crearon en la base SQL Server destino, las tablas de facturación y detalle del cliente para el mes actual, así como las tablas correspondientes que agrupan los meses anteriores. Adicionalmente, se creó la tabla que almacena la información sobre las altas y bajas del servicio Cloud Server de Tigo Cloud.

- Diseño de tablero/dashboard
 - Finalmente, se utilizó Power BI para diseñar el tablero y conectarlo a las tablas creadas, con el objetivo de presentar toda la información relevante sobre la facturación, los tipos de servicios y la cantidad de clientes. Además, se añadieron filtros como período, gerencia o región, que permiten una visualización detallada y personalizada de la información.

V. RESULTADOS

Esta sección contiene los resultados obtenidos tras la implementación de la metodología propuesta. Es importante mencionar que, algunas de las figuras presentes contienen campos con valores ocultos puesto que representan información confidencial de la compañía.

- Identificación y exploración de las fuentes de datos

Como resultado de la búsqueda de las fuentes de datos, se identificaron cuatro tablas en distintas bases que contenían información variada sobre los servicios Tigo Cloud. Algunas de las tablas tenían desde 800.000 hasta 3.000.000 registros cada una, por lo que el proceso de exploración fue extenso y demorado.

- Extracción de la información

La extracción de identificadores únicos de los servicios Tigo Cloud, realizada a partir de las bases de datos fuente, generó un listado de aproximadamente 23,000 registros. Estos identificadores fueron cruzados con las bases de datos correspondientes para obtener tanto la facturación como el detalle de los clientes asociados a estos servicios. Adicionalmente, se extrajo la información necesaria que permitió identificar las altas y bajas del servicio Cloud Server, que forma parte de Tigo Cloud.

- Transformación de los datos

Trás obtener las tablas de facturación y detalle del cliente en la fase de extracción de la información, fue necesario aplicar procesos de limpieza en algunos campos de estas tablas. Específicamente, en la Figura 2 se puede observar el antes y después de eliminar acentos y estandarizar los datos para evitar variaciones innecesarias a causa de la escritura.

ABC Departamento	ABC Departamento
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTA	BOGOTA D.C.
BOGOTÁ D.C.	BOGOTA D.C.
BOGOTÁ D.C.	BOGOTA D.C.
BOGOTÁ D.C.	BOGOTA D.C.
BOGOTÁ D.C.	BOGOTA D.C.

Fig 2. Estandarización de los datos.

Adicionalmente, fue necesario eliminar columnas sobrantes y añadir otras nuevas como “Período”, “Fuente”, “Nivel_1”, “Nivel_2”, “Nivel_3” y “Nivel_4”, evidenciadas en la Figura 3. La columna “Período” indica el año y mes al que corresponde la facturación o el detalle del cliente. Por otro lado, “Fuente” sólo está presente en el DataFrame de facturación y señala la base de datos de la cual fueron obtenidos los identificadores en la etapa anterior; esto con la finalidad de poder rastrear su origen en caso de ser necesario. Finalmente, las columnas “Nivel” presentes en ambas tablas corresponden a la clasificación de los servicios desde un nivel general hasta uno más detallado del producto. En el caso de la tabla de facturación, también se incorporó una columna llamada “Fact_Total” que contiene la suma de los montos de facturación agrupados por identificador.

ABC PERIODO	ABC FUENTE	ABC NEW_LEVEL1	ABC NEW_LEVEL2	ABC NEW_LEVEL3	ABC NEW_LEVEL4	123 FACT_TOTAL
202406				Tigo Cloud	Cloud Server	
202406				Tigo Cloud	Cloud Server	
202406				Tigo Cloud	Cloud Server	
202406				Tigo Cloud	Cloud Server	
202406				Tigo Cloud	Cloud Server	
202406				Tigo Cloud	Cloud Server	

Fig 3. Ejemplo de nuevas columnas en la tabla de facturación

Por otra parte, al extraer la data asociada a las fechas de inicio y finalización de los servicios Cloud Server, se obtuvo el DataFrame que se muestra en la Figura 4, el cual presenta una distribución en formato vertical que dificulta la interpretación de cada campo. Como consecuencia, fue necesario pivotar la tabla resultante, renombrar las

columnas y aplicar procedimientos de limpieza, cambios de formato, y adición de columnas, permitiendo así una mejor visualización de la información, como se evidencia en la Figura 5.

Index	PEDIDO_ID	CARACTERISTICA_ID	VALOR
0		1	
1		39	
2		2907	5/30/2024 11:43:40 AM
3		4388	None
4		1	
5		170	12
6		2682	20240531
7		4024	Cloud Server
8		4026	1
9		4027	180330
10		4028	180330

Fig 4. Tabla inicial de consulta Cloud Server

Index	PERIODO	PEDIDO_ID	IDENTIFICADOR	NIT	F_CREACION	F_INICIO	DESCRIPCION	CANTIDAD	FACTURADOR	VIGENCIA
0	202401				2024-01-30	2024-03-...	Cloud 360 -	1	None	12
1	202402				2024-02-10	2024-06-...	Cloud 360 -	1	None	12
2	202402				2024-02-12	2024-03-...	Cloud 360 -	1	None	36
3	202402				2024-02-13	2024-04-...	Cloud 360 -	1	None	12
4	202402				2024-02-14	2024-07-...	Cloud 360 -	1	None	24
5	202402				2024-02-22	2024-08-...	Cloud 360 -	1	None	12
6	202402				2024-02-22	2024-03-...	Cloud 360 -	1	None	12
7	202402				2024-02-23	2024-06-...	Cloud 360 -	1		4
8	202402				2024-02-26	2024-08-...	Cloud Server	1		22
9	202403				2024-03-05	2024-08-...	Cloud 360 -	1		1
10	202403				2024-03-08	2024-07-...	Cloud 360 -	1	None	12

Fig 5. Tabla resultante tras la transformación

Así mismo, como parte de la limpieza realizada en las tablas, las Figuras 6 y 7 detallan el proceso de depuración que fue necesario aplicar en los campos con valores

numéricos y de fecha. En la primera, se evidencia el estado inicial de los datos numéricos que se encontraban en formato String y contenían comas, puntos y espacios en blanco. Sin embargo, estos errores fueron corregidos, y el formato fue convertido a tipo Float.

VALOR_UNITARIO	VALOR_UNITARIO
608.112	608112
902.008	902008
2258734	2.25873e+06
9552334	9.55233e+06
1258916	1.25892e+06
531.164	531164
540.080	540080
11981326	1.19813e+07
2494853,	2.49485e+06

Fig 6. Limpieza y conversión de campos numéricos

Por su parte, la figura siguiente muestra el proceso de conversión de los campos con fechas. Originalmente, los datos contaban con un formato datetime, pero los registros de tiempo no aportaban información relevante. Por tanto, para simplificar la estructura y facilitar su manejo, se decidió convertir el campo a formato date únicamente.

F_INICIO	F_INICIO
2024-03-13 00:00:00	2024-03-13
2024-06-11 00:00:00	2024-06-11
2024-03-27 00:00:00	2024-03-27
2024-04-16 00:00:00	2024-04-16
2024-07-30 00:00:00	2024-07-30
2024-08-15 00:00:00	2024-08-15

Fig 7. Cambio de formato en campos de fecha

Continuando con la tabla de los servicios Cloud Server, dado que el campo que indicaba la fecha de finalización del servicio contenía registros vacíos, fue necesario completar la información calculando una fecha final estimada a partir de la fecha de

creación y la vigencia del mismo. Esto permitió asegurar que no existieran registros vacíos o incompletos, como se muestra en la Figura 8. Además, con esta nueva información se llevó a cabo el proceso de identificación de los servicios activos y cancelados, realizando una comparación de los campos de fecha de inicio y fecha final con respecto a la fecha actual. Como resultado se obtuvo la Figura 9, en donde se observa que se añadieron columnas nuevas en el DataFrame como "Estado" y "Periodo" que indican justamente si el servicio está activo o inactivo, así como el año y mes en que el servicio fue creado o anulado.

F_FIN	TIPO	F_ESTIMADA_FIN ▲
NaT	CLOUD360	2024-02-25
NaT	CLOUD360	2024-09-16
NaT	CLOUD360	2024-11-13
NaT	CLOUD360	2025-02-07
NaT	CLOUD360	2025-02-08
NaT	CLOUD360	2025-02-13
NaT	CLOUD360	2025-02-14
NaT	CLOUD360	2025-03-01
NaT	CLOUD360	2025-03-08
NaT	CLOUD360	2025-04-14

Fig 8. Proceso de completar registros vacíos

Index	PERIODO	F_CREACION	F_FIN	ESTADO
0	202401	2024-01-30	2025-03-13	Activo
1	202402	2024-02-10	2025-06-11	Activo
2	202402	2024-02-12	2027-03-27	Activo
3	202402	2024-02-13	2025-04-16	Activo
4	202402	2024-02-14	2026-07-30	Activo
5	202402	2024-02-22	2025-08-15	Activo
6	202402	2024-02-22	2025-03-08	Activo
7	202402	2024-02-23	2024-10-04	Activo
8	202402	2024-02-26	2026-06-15	Activo

Fig 9. Ejemplo de nuevas columnas en tabla de altas y bajas de Cloud Server

- Carga de la información

En la Figura 10 se presenta la estructura de las tablas destinadas a almacenar la información de facturación, tanto actual como histórica, ambas con un total de 35 columnas. De igual manera, la Figura 11 muestra la estructura de las tablas diseñadas para guardar el detalle del cliente, cada una con 30 columnas. La diferencia principal entre las tablas actuales e históricas radica en que las primeras contienen los datos del mes más reciente, mientras que las segundas almacenan la información correspondiente a meses anteriores.

MES ACTUAL	HISTORICO
RDF PERIODO	RDF PERIODO
RDF CLIENTE_ID	RDF CLIENTE_ID
RDF Nombre	RDF Nombre
RDF Suscripcion	RDF Suscripcion
RDF strServicioSuscrito	RDF strServicioSuscrito
RDFCodigo_Servicio	RDFCodigo_Servicio
RDFDescripcion_Servicio	RDFDescripcion_Servicio
RDFCodigo_Despartamento	RDFCodigo_Despartamento
RDFDepartamento	RDFDepartamento
RDFCodigo_Municipio	RDFCodigo_Municipio
RDFMunicipio	RDFMunicipio
RDFCodigo_Concepto	RDFCodigo_Concepto
RDFConcepto	RDFConcepto
RDFCodigo_PlanFacturacion	RDFCodigo_PlanFacturacion
RDFPla_Facturacion	RDFPla_Facturacion
RDFCodElementoMedicion	RDFCodElementoMedicion
RDFciclo	RDFciclo
RDFstrIdentificador	RDFstrIdentificador
RDFInterfaz	RDFInterfaz
RDF LINEA	RDF LINEA
RDF SUB_LINEA	RDF SUB_LINEA
RDF REGIONAL	RDF REGIONAL
RDF PLAZA	RDF PLAZA
RDF PRODUCTO_H	RDF PRODUCTO_H
RDF CONCEPTO_AGRUPADO	RDF CONCEPTO_AGRUPADO
RDF GERENCIA	RDF GERENCIA
RDF SEGB2B	RDF SEGB2B
RDF EMPRESA	RDF EMPRESA
RDF ESTADO_RGU	RDF ESTADO_RGU
I25 FACT_TOTAL	I25 FACT_TOTAL
RDF NEW_LEVEL1	RDF NEW_LEVEL1
RDF NEW_LEVEL2	RDF NEW_LEVEL2
RDF NEW_LEVEL3	RDF NEW_LEVEL3
RDF NEW_LEVEL4	RDF NEW_LEVEL4
RDF FUENTE	RDF FUENTE

Fig 10. Tablas de facturación mes actual e histórico

ACTUAL	HISTORICO
PERIODO	PERIODO
CLIENTE	CLIENTE
CLIENTE_ID	CLIENTE_ID
GERENCIA	GERENCIA
DEPARTAMENTO	DEPARTAMENTO
MUNICIPIO	MUNICIPIO
DIRECCION	DIRECCION
ESTRATO	ESTRATO
REGIONAL	REGIONAL
PLAZA	PLAZA
LINEA	LINEA
SUB_LINEA	SUB_LINEA
PRODUCTO_ID	PRODUCTO_ID
PRODUCTO	PRODUCTO
SUBPRODUCTO	SUBPRODUCTO
IDENTIFICADOR_ID	IDENTIFICADOR_ID
AGRUPADOR_ID	AGRUPADOR_ID
ESTADO_SERVICIO	ESTADO_SERVICIO
PLAN_FACTURACION	PLAN_FACTURACION
INTERFAZ	INTERFAZ
TIPO_SERVICIO	TIPO_SERVICIO
PRODUCTO_H	PRODUCTO_H
EJECUTIVO	EJECUTIVO
SUSCRIPCION	SUSCRIPCION
CODIGO_HOGAR	CODIGO_HOGAR
EMPRESA	EMPRESA
NEW_LEVEL1	NEW_LEVEL1
NEW_LEVEL2	NEW_LEVEL2
NEW_LEVEL3	NEW_LEVEL3
NEW_LEVEL4	NEW_LEVEL4

Fig 11. Tablas con el detalle del cliente mes actual e histórico

Similarmente, la Figura 12 muestra la estructura de la última tabla diseñada para almacenar la información sobre las altas y bajas de los servicios Cloud Server.

CLOUD_SERVER
PERIODO
PEDIDO_ID
IDENTIFICADOR
NIT
F_CREACION
F_INICIO
DESCRIPCION
CANTIDAD
FACTURADOR
VIGENCIA
123 VALOR_UNITARIO
123 VALOR_TOTAL
F_FIN
TIPO
ESTADO

Fig 12. Tabla final de altas y bajas de Cloud Server

- Diseño de tablero/dashboard

Para finalizar, se construyó un tablero de facturación en Power BI, vinculado a las tablas creadas. Este tablero representado en la Figura 13, permite visualizar la cantidad de clientes y servicios, así como la facturación total, clasificada por tipo de servicio. Como se evidencia en la Figura 14, se añadieron filtros que permiten seleccionar el período para comparar la facturación de meses anteriores e identificar variaciones, además de filtrar por gerencia, región o plaza. También se incluyó el número de altas y bajas del producto Cloud Server, que varía según el período seleccionado.

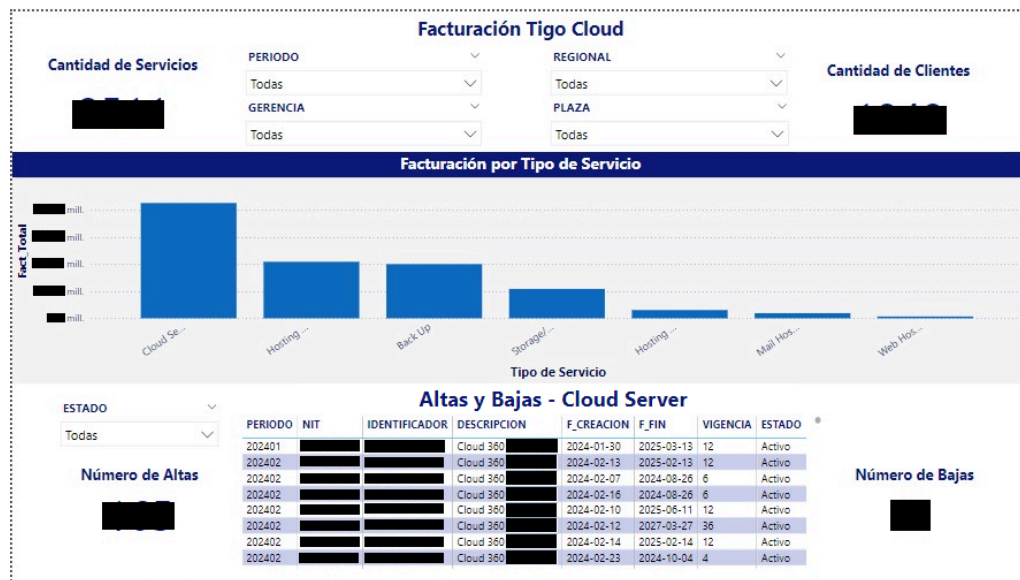


Fig 13. Tablero de facturación en Power BI

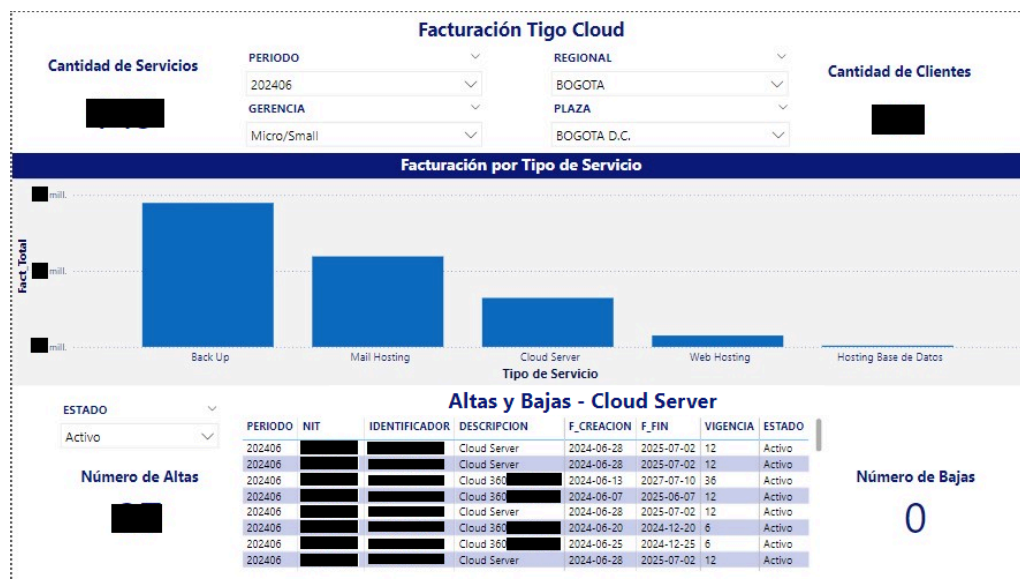


Fig 14. Variaciones en la facturación por selección de filtros

VI. ANÁLISIS

El procedimiento de extracción de la información que hace parte del proceso ETL fue bastante complejo de realizar, puesto que, se identificaron cuatro fuentes de datos distintas que contenían información variada sobre los servicios Tigo Cloud. Por este motivo, fue necesario realizar una exploración exhaustiva de las bases para poder comprenderlas e identificar los datos relevantes para la solución del problema. Sin embargo, el proceso aunque extenso resultó ser muy beneficioso para encontrar la información necesaria. Además, el uso de herramientas como Python y SQL simplificó en gran medida esta búsqueda y extracción de los datos, permitiendo un manejo más eficiente del alto volumen de registros.

De igual manera, la fase de transformación y limpieza de los datos también fue de gran importancia ya que como se observa en las Figura 2, la eliminación de acentos y la estandarización de los datos permitió evitar errores y variaciones en los mismos, garantizando que la información sea más clara y consistente. Además, la agregación de nuevas columnas en las tablas, evidenciada en la Figura 3, ayudó a brindar mayor detalle sobre los tipos de servicios, garantizando una mejor calidad en la información.

En las Figuras 4 y 5, se muestra la transformación realizada sobre la tabla con estructura vertical resultante de la consulta SQL para obtener las altas y bajas de Cloud Server, en donde fue necesario pivotar la tabla, renombrar las columnas y realizar limpieza de los datos. Como resultado se obtuvo una distribución más ordenada y entendible que mejora notablemente la visualización y comprensión de los datos.

El proceso de transformación de valores tipo String a Float en campos numéricos, como se muestra en la Figura 6, también fue muy importante ya que permitió conservar la coherencia y veracidad de los datos al mantener el formato correcto. Además, la conversión de campos datetime a date, observada en la Figura 7, ayudó a simplificar la información eliminando datos innecesarios, logrando así, que esta sea mucho más simple y fácil de manejar. En cuanto a los campos de fecha con registros vacíos, fue necesario corregirlos, como se evidencia en la Figura 8,

para garantizar completitud de la información y evitar interpretaciones incorrectas al momento de determinar el estado de los servicios.

Finalmente, como se observa en las Figuras 10, 11 y 12, las tablas que fueron creadas en la base de datos destino contienen una estructura que garantiza un correcto almacenamiento de los datos, facilitando a su vez un acceso rápido y sencillo a la información. Así mismo, la construcción del tablero en Power BI y su conexión a las tablas, ofreció una visualización completa de los productos, como se muestra en las Figuras 13 y 14, brindando detalles sobre la cantidad de clientes y servicios. Los filtros añadidos mejoraron significativamente la capacidad de análisis mediante la obtención de información específica según las necesidades y permitieron realizar comparaciones de la facturación a lo largo del tiempo.

VII. CONCLUSIONES

La implementación del proceso ETL en el proyecto fue sumamente necesaria y útil para solucionar la problemática existente con la dispersión y generalización de la información de los productos Cloud. Específicamente, este proceso permitió limpiar, estandarizar, categorizar y unificar los datos de los servicios Cloud 360 que provenían de distintas fuentes, con la finalidad de mejorar su calidad y permitir una distinción clara entre los tipos de servicios.

Así mismo, la creación de las nuevas tablas permitió consolidar y almacenar toda la información resultante del proceso en una base de datos destino de fácil acceso y gestión. Además, la estructura y organización de estas tablas facilitó la construcción del tablero en Power BI, con el cual se logró una visión clara de los datos más importantes sobre los servicios, su facturación y los clientes. La capacidad de filtrar los datos por período, gerencia o región, favoreció la realización de comparaciones para un análisis más detallado y oportuno de la evolución de la facturación a lo largo del tiempo. Todo esto con la finalidad de tener un conocimiento auténtico de los distintos tipos de servicios que forman parte de Tigo Cloud para una toma de decisiones mucho más informada.

Finalmente, es importante tener en cuenta que, la solución desarrollada tiene el potencial de seguirse ampliando para lograr unificar y clasificar los servicios Cloud restantes que quedaron fuera del alcance del proyecto. Para ello, se recomienda continuar aplicando procesos ETL que permitan una unificación completa de la información, así como también, no restringirse a una única fuente de datos. Esto es clave para recuperar la mayor cantidad posible de información, lo que a su vez permitirá realizar un análisis más integrado y detallado en el futuro.

REFERENCIAS

- [1] “¿B2B o B2C? ¿Cuál es la diferencia?”. LinkedIn Marketing. [En línea]. Disponible: <https://business.linkedin.com/es-es/marketing-solutions/success/marketing-terms/b2b-vs-b2c>
- [2] “¿Qué es ETL?”. Oracle | Cloud Applications and Cloud Platform. [En línea]. Disponible: <https://www.oracle.com/co/integration/what-is-etl/>
- [4] “¿Qué son los servicios de nube?”. Red Hat. [En línea]. Disponible: <https://www.redhat.com/es/topics/cloud-computing/what-are-cloud-services>
- [5] “¿Qué es Python?”. Amazon Web Services, Inc. [En línea]. Disponible: <https://aws.amazon.com/es/what-is/python/>
- [6] “¿Qué es SQL?”. Amazon Web Services, Inc. [En línea]. Disponible: [https://aws.amazon.com/es/what-is/sql/#:~:text=es%20importante%20SQL?,El%20lenguaje%20de%20consulta%20estructurada%20\(SQL\)%20es%20un%20len%20guaje%20de,los%20diferentes%20lenguajes%20de%20programación](https://aws.amazon.com/es/what-is/sql/#:~:text=es%20importante%20SQL?,El%20lenguaje%20de%20consulta%20estructurada%20(SQL)%20es%20un%20len%20guaje%20de,los%20diferentes%20lenguajes%20de%20programación)
- [7] “¿Qué es Power BI?”. Microsoft Learn. [En línea]. Disponible: <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>
- [8] “¿Qué es ETL (Extraer, Transformar, Cargar)? Una guía completa”. Astera. [En línea]. Disponible: <https://www.astera.com/es/type/blog/etl/>