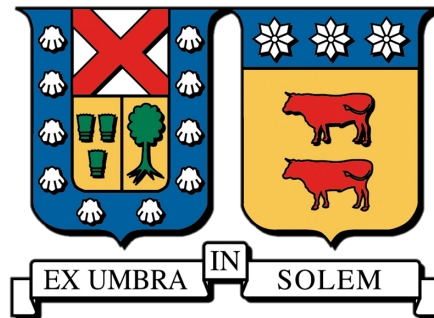


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



**“Plataforma Digital de Inteligencia de Negocio para el
Control de Información de Despacho con Foco en la
Interfaz Usuaría del Reporte Ejecutivo”**

NICHOLAS ANDREAS BERNAL ÁLVAREZ

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO
CIVIL TELEMÁTICO**

PROFESOR GUÍA:

NICOLÁS JARA

PROFESOR CORREFERENTE:

NICOLÁS GÁLVEZ

NOVIEMBRE 2022

Agradecimientos

A mis padres, Myriam Álvarez e Isaac Bernal, nada de esto no habría sido posible sin su infinito apoyo y paciencia. Los amo.

A Alfredo Osses, Constaza Burgos que desde tiempos inmemoriales me han aguantando, hoy son un gran pilar para mi salud mental.

A Alexis Alva y Catalina Zelaya que estuvieron desde el principio de esta etapa, siempre dando su apoyo. Y las risas no faltaron.

A Jesenia *Francisca* Salazar, llegaste en un momento difícil y nunca te fuiste. Te quiero mucho, viejis.

A Paula Maldonado, bueno ya tu sabes, fuiste indispensable en los últimos 100 metros.

Y a todos lo que con su presencia ayudaron a hacer más ameno transitar por este sendero tan escabroso y aportaron su granito de arena para formar a quien soy hoy.

Imagination is more important than knowledge. Knowledge is limited.

Imagination encircle the world.

Albert Einstein

Resumen

En esta memoria se desarrolla una plataforma de *Business Intelligence* para resolver el problema de control y análisis para el área de despacho de una empresa. Dado que está enmarcado dentro del Proceso de Memorias Multidisciplinarias (PMM), la empresa que propone el problema corresponde al banco Itau.

El objetivo principal de la solución desarrollada es la automatización de reportes para datos provenientes de múltiples fuentes. Para ello se planifica e implementa un proceso de recolección de registros de todos los proveedores de datos relacionados a la fabricación y despacho del producto. Estos registros luego son formateados y procesados para ser almacenados en un *data warehouse* de forma automática y actualizados diariamente, ya que esta es la periodicidad con la que se reciben los datos.

Lo anterior busca construir una estructura sólida para la creación de reportes, los que son creados usando la plataforma Tableau. El dashboard resultante presenta varios filtros que están relacionados a los productos, proveedores de despacho y sucursales, entre otros, y su principal atractivo es la automatización del reporte, esto debido que no necesita de intervención del usuario para mostrar la última información registrada.

Este trabajo, redujo el tiempo de actualización de la data almacenada, de dos semanas a un día. Por otro lado, el tiempo invertido en la creación de los reportes se redujo al mínimo debido a que, después de la implementación del trabajo realizado, solo se necesita modificar el reporte cuando se quiera agregar o quitar información del mismo.

Abstract

In this report, a Business Intelligence platform is developed to solve the control and analysis problem for the dispatch area of a company. Since it is framed within the Multidisciplinary thesis process (PMM for its acronym in Spanish), the company that proposes the problem corresponds to the bank Itau.

The main objective of the developed solution is the automation of reports for data coming from multiple sources. For this purpose, a process is planned and implemented to collect records from all data providers related to the manufacturing and dispatching of the product. These records are then formatted and processed to be stored in a data warehouse automatically and updated daily, since this is the periodicity with which the data is received.

The above seeks to build a solid structure for the creation of reports, which are created using the Tableau platform. The resulting dashboard presents several filters that are related to products, dispatch suppliers and branches, among others, and its main attraction is the automation of the report, because it does not require user intervention to display the latest information recorded.

This work reduced the time to update the stored data from two weeks to one day. On the other hand, the time invested in the creation of the reports was reduced to a minimum because, after the implementation of the work done, it is only necessary to modify the report when you want to add or remove information from it.

Glosario

Key Performance Indicator: Son métricas que se utilizan para sintetizar la información sobre la eficacia y productividad de las actividades que se lleven a cabo en un negocio.

Service Level Agreement: Se trata de un contrato firmado entre las partes involucradas en una negociación que determina cuáles son las responsabilidades de cada uno en relación a los servicios contratados.

Acrónimos

KPI: Key Performance Indicator

SLA: Service Level Agreement

ETL: Extract, Transform and Load

BI: Business Intelligence

OLAP: OnLine Analytical Processing

TC: Tarjeta de Crédito

TD: Tarjeta de Débito

KIT: Kit de cuenta corriente

Índice de figuras

2.1. Cuadrante Mágico de Gartner	12
3.1. Arquitectura Inteligencia de Negocio	16
3.2. Diagrama ETL	17
3.3. Diagrama de operación del banco	22
4.1. Modelo de datos	30
4.2. Esquema general de sistema	36
4.3. Vista inicial del dashboard	38
4.4. Vista del dashboard para tarjetas	39
4.5. Vista del dashboard para kits	41
4.6. Diagrama de ETL	42
5.1. Vista Inicial para todos los casos de uso	46
5.2. Diagrama lógico de interacción para el caso de uso 1	46
5.3. Emisiones de tarjetas de débito por mes	47
5.4. Diagrama lógico de interacción para el caso de uso 2	48
5.5. Lead Time para proveedor Envía Express	49
5.6. Gráfico de diferencia porcentual en las vistas Tarjetas y Kit	50
5.7. Diagrama de interacción para la ETL	51
5.8. Información de salida con archivos de entregas	52

Índice general

1. Introducción	5
1.1. Contexto	5
1.2. Objetivos	8
1.2.1. Objetivo General	8
1.2.2. Objetivos Específicos	8
1.3. Estructura	8
2. Estado del Arte	11
2.1. Software para la Inteligencia de Negocio	11
3. Caso de Estudio	15
3.1. Herramientas Utilizadas	16
3.1.1. Business Intelligence	16
3.1.2. Extracción, Transformación y Carga	17
3.1.3. Data Warehouse	19
3.1.4. Minería de Datos	20
3.2. Descripción del Caso de Estudio	22
3.2.1. Nuevos Registros	23
3.2.2. Registros durante el despacho	23
3.2.3. Análisis de datos	24
3.3. Problema	25
3.4. Oportunidad	25

4. Propuesta de Solución	27
4.1. Requisitos	27
4.1.1. Requisitos Funcionales	27
4.1.2. Requisitos No Funcionales	28
4.2. Diseño de la Arquitectura del Sistema	28
4.2.1. Tecnologías disponibles	28
4.2.2. Modelo de datos	30
4.3. Prototipo del sistema	36
4.3.1. Esquema general del sistema	36
4.3.2. Interfaz de usuario	37
4.3.3. ETLs	42
5. Validación del Prototipo	45
5.1. Casos de Uso: <i>Dashboard</i>	45
5.1.1. Cantidad de Emisiones de Tarjetas de Débito por Mes	46
5.1.2. Tiempo de Entrega Promedio de las Tarjetas de Débito por <i>un proveedor</i> por Mes	48
5.1.3. Diferencia del porcentaje de entrega del mes actual con el mes anterior durante el mismo periodo de tiempo	50
5.2. Casos de Uso: <i>Datawarehouse</i>	51
6. Conclusión	53

Capítulo 1

Introducción

1.1. Contexto

En la actualidad, los datos son la materia prima de un recurso vital para las empresas como lo es la información. Según un estudio hecho por DOMO, se generaron 79 zettabytes de *data* tan solo en el año 2021 y se prevé que esta cantidad llegue a 180 zettabytes para el 2025 [1]. Con estos antecedentes, es insostenible manipular, procesar y analizar esta cantidad exorbitante de datos de forma manual, independiente de si el encargado es una persona o un equipo de trabajo. Para solucionar el problema descrito, en los años noventa fue acuñado el termino *business intelligence* [2]. El *BI* es un método que utiliza herramientas tecnológicas avanzadas para acumular y analizar datos e información sobre una organización [3]. Este método permite a sus usuarios generar reportes, hacer predicciones y mejorar el rendimiento de su negocio, a través de diferentes herramientas como las ETL, los data warehouse y OLAP [4].

Una problemática que puede solucionar el *BI* es la inconsistencia de datos y procesos ineficientes. En este caso entendiendo al *Business Intelligence* como un paraguas que cubre una cantidad enorme de herramientas de análisis, bases de datos, entre otros [5]. Un claro ejemplo es Veolia, una empresa multinacional francesa enfocada en servicios medioambientales. Esta compañía, entregaba a sus clientes y ejecutivos, ho-

jas de cálculos con los datos necesarios para tomar decisiones, obteniendo un problema muy recurrente en estos entornos desordenados: múltiples fuentes de la verdad [6]. Esta inconsistencia generaba que los ejecutivos e inversionistas pasaran más tiempo discutiendo sobre números erróneos, que tomando decisiones importantes basadas en los datos recolectados. Hoy por hoy, implementando soluciones enfocadas en centralizar la información y presentarla de manera precisa a sus clientes y ejecutivos, Veolia implementó una plataforma estandarizada que procesa y analiza los datos provenientes de múltiples fuentes, para luego presentar la información obtenida en su portal de análisis. Lo anterior les permite enfocarse únicamente en la decisión que tomarán con esta información, ya que pueden confiar plenamente en su integridad.

Otra solución que se implementa usando *Business intelligence*, es el monitorio de stock o alarmas que permitan avisar a la empresa sobre algunos SLAs o KPIs que estén muy bajos a través de *dashboards* generados de manera automática, reduciendo así el tiempo y costos de generar reportes para visualizar estas métricas. Un ejemplo de esto es Inka Crop, una empresa de alimentos, la cual ha reducido su tiempo de generación de reporte en un 50 % [7]. Según comenta Gonzalo Campos, Key Account Manager, esta implementación le ayudó directamente con las métricas de stock en sus tiendas, avisando de forma oportuna cual debía ser reabastecida. Monitorizar esta métrica, les permite ahora avisar a los encargados de mover la mercadería del almacén a la tienda y así evitar perder ventas por falta de stock.

Resalta, de los ejemplos anteriores, que las empresas no tenían un proceso para obtener información a partir de los datos recolectados, impidiendo maximizar los beneficios que entrega este recurso. Es en este contexto que se desarrolla el presente trabajo de título, donde se busca resolver la problemática subyacente a la obtención de datos de múltiples fuentes, procesos no estandarizados para la obtención de estos y el despliegue centralizado y automatizado de los reportes ejecutivos para una empresa. El contexto particular de esta memoria es el despacho de producto y control de KPIs relacionados a su fabricación y entrega. Se desarrolla una plataforma integral de *business intelligence* que obtiene sus datos desde múltiples fuentes, bajo el supuesto de que no

están estandarizados, con el fin de procesarlos y almacenarlos en un data warehouse centralizado para la empresa. Esta acción busca que la información sea accesible desde cualquier área en caso de ser necesario. Por último, la plataforma presenta una interfaz, que muestra los datos almacenados en un *dashboard* dinámico separando por tipo de proveedor, sucursales, entre otros. Esta interfaz se actualiza de forma automática dependiendo de la periodicidad con la que el data warehouse reciba los datos.

La principal dificultad que presenta este trabajo es la heterogeneidad de las fuentes de datos. En primer lugar, aunque se explica con mayor detalle en los capítulos posteriores, la empresa genera datos de forma interna para crear los productos. Luego, estos datos son enviados al fabricante, el cual reporta diariamente el estado de todos los productos que ya fueron solicitados. En tercer lugar, una vez listos los productos y dependiendo de su tipo, puede que el mismo fabricante lo entregue o puede ser entregado por otro proveedor de despacho, generando así otra fuente de reporte de datos relacionados al estado de la entrega. Por último, si ninguna de las dos empresas despachadoras ha sido capaz de entregar el producto, se redirige a la sucursal, la cual debe generar registros que son otra fuente de datos para el sistema.

Para cuando se obtuvieron los requerimientos del caso, la empresa realizaba la recolección de los datos cada dos semanalmente, por lo que los reportes o *dashboards* que mostraban el progreso del despacho no podían ser actualizados hasta que los datos estuvieran ingresados en las bases de datos pese a que eran recibidos diariamente.

Para finalizar, la presente memoria esta enmarcada en el Programa de Memoria Multidisciplinaria y es el desafío entregado por el departamento de despacho del Banco Itaú.

1.2. Objetivos

1.2.1. Objetivo General

El objetivo de este trabajo es diseñar e implementar un sistema de gestión y manejo de la información proveniente de diferentes fuentes, principalmente enfocado en el área de reparto de productos, como también generar una solución para el despliegue de los principales indicadores del área, mediante alguna plataforma especializada.

1.2.2. Objetivos Específicos

- Obtener la data de todas las fuentes que alimentan de información al área de *Delivery*.
- Estandarizar y clasificar los datos obtenidos.
- Crear un modelo que almacene toda esta información.
- Integrar los datos al modelo.
- Investigar KPI's relevantes para esta unidad de negocio.
- Generar paneles para la visualización y análisis de los KPI's

1.3. Estructura

Este documento se estructura de la siguiente forma: En el capítulo 2 se presentan las plataformas punteras en el área de la inteligencia de negocio. Luego en el capítulo 3 se revisan los principales conceptos que se utilizan a lo largo de este trabajo y además se presenta el proceso actual de la empresa al momento de manejar y desplegar sus datos almacenados, se identificará la problemática y las oportunidades de mejora, las que posteriormente, en el capítulo 4, se usaran como propuesta a la solución describiendo la arquitectura del sistema propuesto y mostrando el prototipo alcanzado al final del

trabajo. En el capítulo 5 se presentan las validaciones y comparaciones del prototipo con el antiguo proceso mostrando algunos casos de uso del sistema implementado. Por último, en el capítulo 6, se entregaran las conclusiones del trabajo.

Capítulo 2

Estado del Arte

2.1. Software para la Inteligencia de Negocio

Las plataformas de inteligencia de negocio, ayuda a los usuario no técnicos, en especial empresarios, a desarrollar actividades como análisis, modelamiento, exploración, entre otros [8]. Existen muchas formas de conseguir esto, códigos dedicados, frameworks para la minería de datos, servicios web, y programas dedicados, todas ellas enmarcadas como herramientas de BI. Es tal la importancia de estas aplicaciones en las organizaciones, y tan crítico que las herramientas utilizadas sean completas, que el mercado en 2020 se estimaba sobre los USD 22 billones con una proyección para 2028 de USD 43 billones, lo que implica una tasa de crecimiento anual compuesto (CAGR) de 8,7 % [9]. Dentro del mercado de las herramientas BI existen fuertes competidores, con herramientas de alto nivel.

El cuadrante de Gartner de la figura 2.1, muestra de forma cualitativa las plataformas de *Business Intelligence* que hay en el mercado, analizando 12 capacidades críticas que muestran ventajas competitivas de los proveedores [8]



Figura 2.1: Cuadrante Mágico de Gartner

Se hace un repaso de las plataformas líderes según este reporte.

Microsoft Power BI

Microsoft Excel es una herramienta altamente utilizada por diferentes organizaciones para analizar tablas de datos y generar reportes. Sin embargo, Microsoft cuenta con una herramienta especializada para estos fines, Power BI. Del reporte de Gartner [8] se destaca su integración con Office 365 y Microsoft Teams, software altamente apreciado por las organizaciones en el actual contexto de teletrabajo. Esto significa que el acceso a esta herramienta podría ser útil y fácil de integrar en el flujo de trabajo. Sin embargo, se advierte también, en el reporte, que existe la limitante de sólo operar en un servidor Microsoft Azure. Ante esto, queda la interrogante si la potencial organización usuaria

de la solución estaría dispuesta a integrarse aún más con el ecosistema de Microsoft, o si quisiera mantener sus opciones abiertas. Cabe mencionar que dado el alcance de la base de usuarios, es posible que facilite converger con otras áreas que también usen esta herramienta.

Qlik

También dentro de la categoría de líderes de la industria, Qlik es otra opción para evaluar el estado del arte. El reporte de Gartner [8] informa que Qlik tiene una fuerte visión en machine learning e inteligencia artificial, ofreciendo sugerencias de forma automatizada y análisis complejos. Se destaca su flexibilidad en el despliegue tanto en la nube, y de distintos proveedores, como en servidor dedicado. Se advierte, sin embargo, la existencia de un portafolio complejo, con distintos tipos de integraciones entre sí, complicando la curva de aprendizaje. Corresponde evaluar si las herramientas destacadas de machine learning y sugerencias automatizadas son de interés para la potencial organización usuaria.

Tableau

El reporte de Gartner [8] destaca su lineamiento por el manejo visual de los datos, siendo una parte importante de su interfaz de funcionalidad “drag-and-drop”, eliminando los requerimientos de programación para obtener utilidades visuales y generación de reportes de gran calidad. Se advierte en el reporte que si bien existe una alternativa en la nube, Tableau está nativamente desarrollado para funcionar en un servidor dedicado dentro de la organización.

Capítulo 3

Caso de Estudio

En las empresas, aquellos que las lideran, siempre buscan tomar las mejores decisiones posibles para maximizar las ganancias. Esta acción es realmente compleja cuando se tiene en consideración la gran cantidad de factores que se manejan y que pueden afectar a aquella decisión, en este contexto cualquier paso en falso puede ser fatal. Por este motivo, las empresas buscaron disminuir esta incertidumbre enfocando su esfuerzo en algo que tenían pero no le daban la importancia necesaria: los datos. Esta información almacenada en sus servidores, ahora podía servir para tomar decisiones con un enfoque que la competencia quizás no tenía. Al darse cuenta de la gran ventaja que esto suponía, empezaron a invertir en una nueva área de la ciencia de la computación, naciendo así la inteligencia de negocio.

Este capítulo se divide en dos secciones. En primer lugar, se presentan las herramientas que se emplearon para llevar a cabo este trabajo. Y en segundo lugar, se presenta el caso de estudio describiendo el proceso actual de la empresa para recolectar datos y generar reportes, reconociendo sus falencias y presentando las oportunidades de mejoras que serán desarrolladas en la propuesta de solución.

3.1. Herramientas Utilizadas

3.1.1. Business Intelligence

Business intelligence (BI), o inteligencia de negocios, es la toma de decisiones en base al conocimiento extraído de datos recopilados por una organización. El propósito de esto es que las decisiones tengan una base fundamentada en el análisis. Es especialmente necesario en las unidades operativas de las organizaciones, con el fin de optimizar sus procesos. A su vez, es un área enfocada en los altos cargos de gerencia, cuyo conocimiento de las herramientas informáticas es escaso o simplemente nulo [5], debido a esto, es el área de la ciencia de datos que se vio obligada a bajar los estándares de sofisticación en aquellos programas enfocados en estos usuarios, los gerentes, para así darle la necesaria independencia que tenían con el departamento de TI de la empresa. Un ejemplo de la estructura de un sistema utilizado para la inteligencia de negocio se puede ver en la figura 3.1. Los datos proveniente de múltiples fuentes se recolectan para estandarizarlos y procesarlos, izquierda de la figura 3.1, antes de ser almacenados en un data warehouse, una vez ahí pueden ser consultados por diferentes aplicaciones que analizan la data procesada, derecha de la figura 3.1, obteniendo información relevante para tomar decisiones.

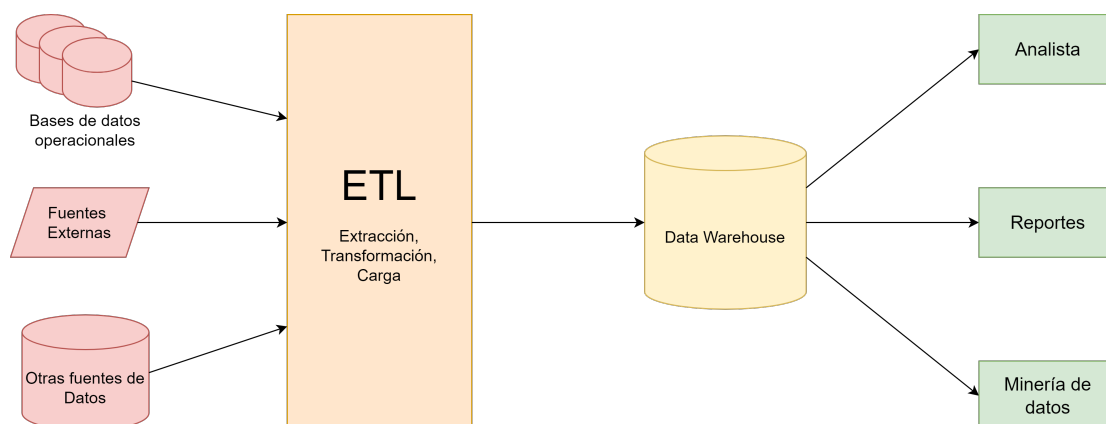


Figura 3.1: Arquitectura Inteligencia de Negocio

Teniendo en cuenta el enfoque del negocio, toda organización que lleva un registro de sus acciones acumula una increíble cantidad de datos a lo largo de su existencia, y con la globalización estas empresas pueden recopilar y almacenar una cantidad enorme de información [10]. Esta creciente fuente de registros trae consigo un gran problema al momento de manipularlos y analizarlos, que hace muy complicado el proceso de manipulación y posterior análisis de la data, para ello herramientas como el aprendizaje de maquinas, la inteligencia artificial o la minería de datos, son imprescindibles al momento de procesar la información y de manipular aquellos indicadores claves que son requeridos para tomar decisiones importantes dentro del negocio.

3.1.2. Extracción, Transformación y Carga

Estos tres procesos corresponden a la primera fase del BI, debido que sin esta se tendría un desorden que hace inoperable los datos obtenido por la empresa y por tanto inútiles. Estos procesos, en su conjunto, buscan unificar y transformar la data en información.

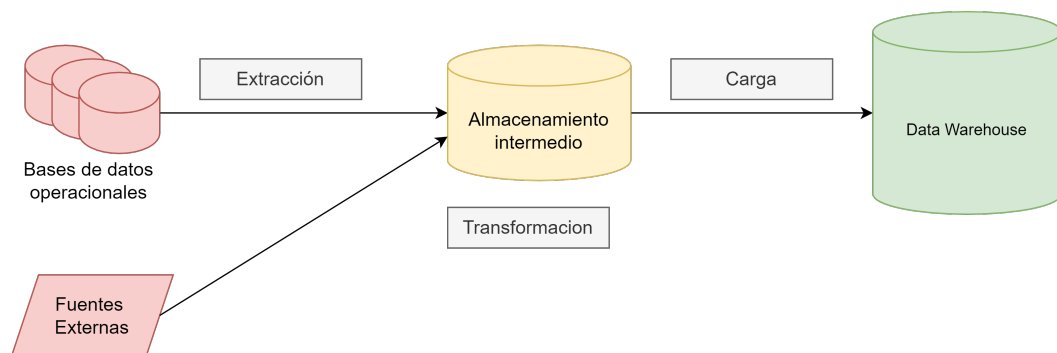


Figura 3.2: Diagrama ETL

Extracción

El primer proceso del manejo de los datos es la extracción, ubicado a la izquierda de la figura 3.2. En este punto, los datos son recopilados de diferentes fuentes como bases

de datos operacionales, archivos, entre otros. La extracción busca consolidar y filtrar la data para facilitar el rendimiento de las operaciones que se ejecutan sobre los datos [11]. Se debe considerar como el punto de partida del manejo de datos. Este proceso suele ser complicado debido a que la información que se reúne no necesariamente tiene formato claro, y se intenta interpretar y corregir lo esencial únicamente para tener uniformidad en la data que se entregará al siguiente proceso.

Transformación

Una vez terminada la extracción, se pasa a una etapa donde son limpiados, sacando o tratando aquellos datos que son anómalos para que no afecten al conjunto total de registros, que sean sacados o tratados depende netamente del especialista que conoce cuales son los riesgos de ambas acciones y como afecta cada uno al negocio. Después de limpiarlos, se formatean todos los registros que quedan [11].

Un ejemplo de lo anterior puede ser el atributo RUN de un usuario, que dependiendo del origen del registro el formato puede venir escrito con guión, sin guión o con puntos y con guión. El analista debe decidir que formato utilizará y aplicarlo a todas las fuentes.

Además, en esta etapa se pueden calcular atributos compuestos, llamados agregaciones. Esto se debe a que manejar una gran cantidad de información es costoso en términos computacionales y hacerlo a la hora de analizar los datos en un gráfico, por ejemplo, el tiempo se carga seria extenso y poco atractivo. Esta acción no es obligatoria, pero muy recomendable.

Por ultimo esta información es almacenada en un repositorio [11], como se muestra en la figura 4.6.

Carga

Es la etapa final en el manejo de datos. Luego de transformar la data en un recurso utilizable, esta es enviada al *data warehouse* en donde se encuentra al alcance de todo aquel que quiera ocuparla para el análisis o generación de reportes, con cualquier tipo

de software del mercado o incluso uno personalizado. La razón detrás de cambiar el lugar de almacenamiento, recae en la necesidad de guardar la información en una base de datos especialmente diseñada para ampliar la escalabilidad y reducir los tiempos de respuesta [11]. Ahora la gerencia puede conectarse y hacer sus reportes completamente independizados del departamento de informática de la empresa.

3.1.3. Data Warehouse

El Data warehouse corresponde a una base de datos especializada en la carga masiva de datos, usualmente de registro histórico, es decir, son de solo lectura. La implementación de esta base de dato se enfoca principalmente en el acceso eficaz y eficiente a la información almacenada en ellos [11]. Por ello es que se pueden distinguir, en el proceso de creación, dos tipos de Data warehouse, el lógico y el físico. El primero, es la presentación estructural que toman los datos una vez cargados, con esto facilitamos la compresión de la estructura de los datos como por ejemplo creando tablas que representen el origen de los datos para saber, en caso de una gran cantidad de datos corruptos, cual es el origen. Esta presentación también ayuda al posterior mantenimiento de la base de datos. El segundo tipo, el físico, corresponde a la implementación del modelo lógico en la base de datos.

Ambos tipos de Data Warehouse son importantes al momento de planear donde y como se guarda la información. Sin embargo no son necesariamente iguales, ya que vienen a solucionar problemas diferentes. El lógico busca ordenar la información para una mejor compresión y planeación mientras que el físico busca que la información sea de rápido acceso. Un ejemplo de lo mencionado, sería crear varias tablas de datos relacionadas por algún tipo de índice llegando tal vez a una tercera o cuarta forma normal, sin embargo es muy engorroso por lo que carece completamente de eficiencia a la hora solicitar la información a la base de datos y es ahí donde entra actuar el modelo físico del data warehouse, que habitualmente solo llega a la primera forma normal

manteniendo la mayoría de los datos en una sola tabla para aumentar considerablemente los tiempos de respuestas de cualquier tipo de solicitud que se le haga.

3.1.4. Minería de Datos

Al momento que los datos van subiendo en volumen su capacidad y el tiempo que se necesita para poder realizar análisis crece con ellos. Para solucionar este problema, la minería de datos, emplea diferentes herramientas para realizar este proceso tales como la inteligencia artificial, aprendizaje automático, estadísticas, entre otros.

Esta área de la ciencia de datos se caracteriza por 3 procesos. Extraer la información de la fuente, para luego procesarla y transformarla en datos coherentes entre sí, y luego analizarla. Este último proceso crea valor a la fuente, ya que permite crear atributos que dependen de otros, cuyos cálculos son complejos y demorarían mucho si se hiciesen en tiempo real.

Durante sus procesos, la minería de datos realiza tareas que permitirán distinguir diferentes patrones en la información que luego servirán como datos complementarios a los recolectados, estas tareas se distinguen en dos tipos, tareas descriptivas y tareas predictivas y cada una de estas usa métodos diferentes para obtener sus resultados.

Tareas Descriptivas

Son aquellas tareas que buscan generar asociaciones entre diferentes atributos de la fuente de datos. De estas tareas podemos distinguir cuatro tipos. Correlación, Detección de datos anómalos, asociación y agrupación.

- **Correlación:** Esta tarea se enfoca principalmente en atributos numéricos y usando la covarianza u otro indicador de correlación. Busca saber si dos o más datos están relacionados de manera lineal o de algún otro modo.
- **Detección de datos anómalos:** El objetivo es encontrar datos que escapen a la tendencia general, llamados también *outliers*. Esta tarea es útil para detectar si-

tuaciones fuera de lo común que dependiendo del contexto podrían ser fraudes, fallas en un sistema, entre otros.

- **Asociación:** Su finalidad es encontrar alguna relación entre los atributos del conjunto de datos proporcionado, y a diferencia de las tareas correlación, esta no se enfoca en los parámetros numéricos sino que es independiente de la naturaleza del dato.
- **Agrupación:** Es una tarea que consiste en obtener grupos de tal forma que los elementos de estos grupos sean similares entre sí, este criterio puede ser usado para buscar grupos poblacionales en el marketing, para así determinar en qué tipo de consumidor hay que enfocar los recursos.

Tareas Predictivas

Este tipo de tareas busca predecir, a partir de una muestra, si un nuevo dato pertenece o no al modelo. En este tipo de tareas encontramos 3 tipos. Clasificación, Categorización y Regresión.

- **Clasificación:** Busca generar una función sobreyectiva, llamada clasificador, con los datos de la muestra. Esta tarea puede ser de dos tipos Binaria o suave. La primera tiene dos grupos de clasificación y todos los datos solo pueden ir a parar a uno de ellos y la segunda se le agrega un parámetro al resultado que identifica el grado de precisión a la predicción hecha por la función principal.
- **Categorización:** Al igual que la tarea de clasificación esta tiene una función que predice en qué grupo se encuentra el nuevo dato, sin embargo lo que diferencia a ambas tareas es el clasificador, ya que este puede asignar a un mismo dato diferentes grupos. El fin de esta tarea es buscar una relación entre los datos.
- **Regresión:** Se enfoca en datos de naturaleza numérica, y busca encontrar una función que represente lo más fielmente posible los datos. Esta tarea entrega un

valor llamado error cuadrático de la regresión que muestra la relación entre la función y los datos, cuyos valores van de 0 a 1, siendo 1 que la función describe en su totalidad al conjunto de datos.

3.2. Descripción del Caso de Estudio

Como se comentó al inicio del capítulo, la gran cantidad de información que una empresa almacena es necesaria para tomar decisiones certeras. Sin embargo, cuando se delega las tareas a otras empresas, el origen y el formato de los datos suelen variar. Es por ello que la primera problemática que se busca solucionar, es la unificación y estandarización las fuentes de información para toda aquella empresa que necesite delegar el despacho de sus productos. Y como segunda problemática, es necesario automatizar la creación de reportes que concentren los KPIs del departamento de despacho de la empresa.

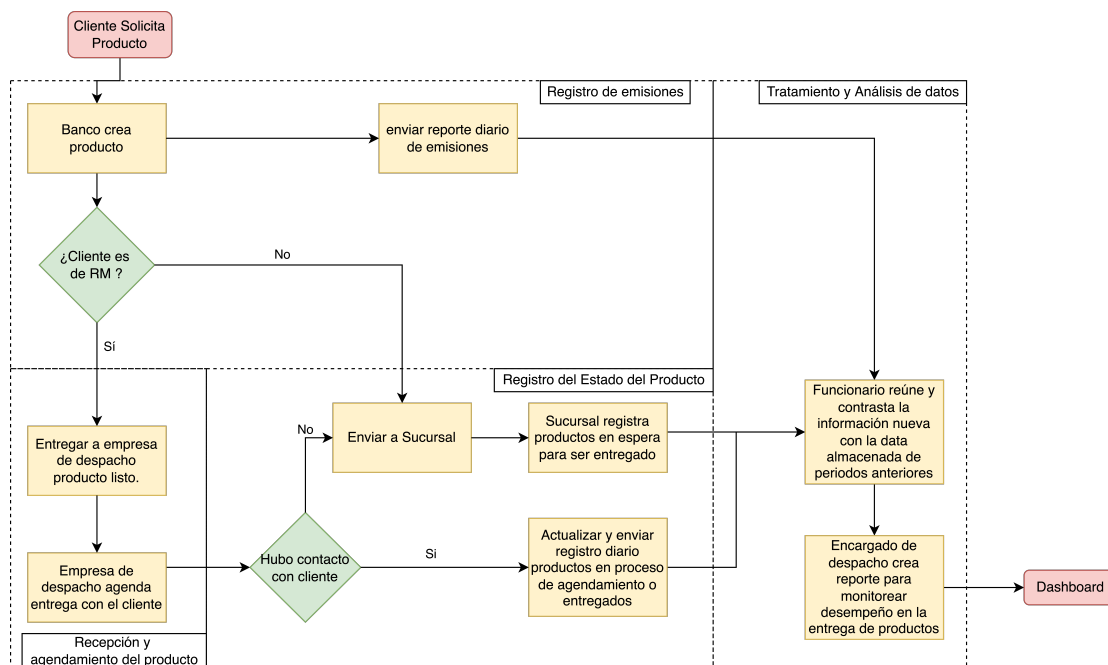


Figura 3.3: Diagrama de operación del banco

Particularmente, como caso de estudio, se trabaja bajo el contexto de una entidad

bancaria genérica, desde ahora *El Banco*, con un departamento de despacho que delega el proceso de entrega a una o varias empresas externas, la que quedan a cargo de la mantención y gestión del proceso. Este departamento debe gestionar la entrega de los productos físicos, como tarjetas de crédito, débito o kit de cuenta corriente, a los clientes que lo soliciten. El flujo de operaciones mencionado, se describe en el diagrama de la figura 3.3.

3.2.1. Nuevos Registros

En el contexto explicado, *El banco* crea dos productos diferentes dependiendo de si el cliente es nuevo o no. Esto es importante ya que el producto se crea y se registra de manera diferente para cada caso. Por otro lado, cabe destacar, que la creación de cualquier producto no es una tarea del departamento de despacho, ya que este solo se dedica a gestionar el envío de productos, es por esto que en el diagrama de la figura 3.3 se encuentra oculto en el bloque “*Banco crea producto*”, sin embargo, los encargados deben facilitar un archivo diario con los nuevos productos emitidos. Con estos archivos, a un funcionario se le asigna la tarea de recoger toda la información generada sobre las emisiones y contrastarla con los datos de periodos anteriores.

3.2.2. Registros durante el despacho

El banco obtiene información del estado de sus productos, de forma diaria, por dos entes diferentes, sus sucursales y las empresas externas.

Las empresas externas tienen un contrato con *el banco* que describe como debe entregarle la información de sus productos, pero cada empresa puede tener diferentes métodos de entrega al cliente, que en este contexto pueden ser agendamiento y entrega simple. El agendamiento permite hacer un rastreo más prolijo del proceso de despacho, permitiendo saber si el producto esta pronto a ser entregado o no. Mientras que la entrega simple permite a la empresa entregar el producto a cualquier persona mayor

de edad que se encuentre en el domicilio sin previa coordinación. Los encargados del despacho entregan un informe en formato Excel *al banco* con el estado actual de los productos, solo si este ha cambiado, como se puede apreciar en la sección *Registro del estado del producto* de la figura 3.3. El formato de este informe queda a discreción de cada empresa ya que cada una tiene su propio proceso de entrega.

En el caso de las sucursales, estas reciben productos por dos posibles razones. La primera vinculada al cliente, cuando este desea ir a buscar su nuevo producto a alguna sucursal o si la empresa de despacho no logra contactarlo para coordinar la entrega. Y la segunda razón, se relaciona con el origen de la solicitud de un nuevo producto. Si la creación de un producto es solicitada por una región ese producto no pasará por las empresas de despacho y el cliente deberá ir a buscarlo a su sucursal más cercana, como se puede ver en la figura 3.3. Gracias a que las sucursales son parte de la entidad bancaria, tienen acceso directo a la bases de datos *del banco*, por lo que solo deben registrar diariamente cuales son los productos que fueron entregados.

3.2.3. Análisis de datos

El banco asigna a una persona para que reúna toda la información de los últimos 15 días y la junte con los registros históricos almacenados en un Excel. Esto se debe a que la información es creada diariamente y su origen es diverso. Esta complejidad impide que el personal del banco tenga como tarea diaria reunir todos los datos en un solo documento para crear el informe de KPIs que se requiere. Una vez que la persona encargada logra juntar los registros antiguos con los nuevos, el personal del departamento de despacho encargado de realizar los reportes de desempeño, diseña las vistas o gráficos que resuman los indicadores que el área necesita. Estas vistas luego son exportadas a un archivo PDF que es utilizado en las reuniones de control de despacho.

3.3. Problema

Con el proceso descrito, se distinguen varios problemas. Para comenzar, las diferentes fuentes de datos en donde se guarda la información, generan un retraso al momento de necesitar los registros para el análisis de los indicadores de desempeño. Tener toda esta información centralizada permitiría que los reportes, al menos en el área de despacho, fueran más rápidos, y otros departamentos podrían acceder a esta información con facilidad en caso de necesitarla. Por otro lado, el cruce de los registros de las emisiones y entregas hechas por sucursal o por las empresas externas, es manual. Esto evidentemente puede generar errores al momento de actualizar la información en el Excel donde se extrae el informe de KPIs. Por último, la periodicidad con la que se integran los nuevos datos a los registros históricos no permite generar reportes de forma diaria ni dinámica.

3.4. Oportunidad

Habiendo entendido el proceso y descubierto los problemas del mismo, a continuación se describen las oportunidades observadas.

Por el lado de los datos de cada paquete, a simple vista se puede detectar que el proceso de consolidación los datos debe ser automatizado, pero esto es complicado si se considera que las fuentes son diferentes. por lo que la oportunidad se encuentra en poder leer los archivos que llegan desde las empresas externas, darle un formato a esos registros y luego guardarlos en una base de datos interna.

Solucionando el problema de múltiples fuente de información, el cruce entre las bases de datos es más sencillo, ya que estas se encuentra dentro de la institución y con un formato estandarizado. La unión mencionada alimenta una nueva tabla en una base de datos interna, así cualquier departamento que necesite esta información no debe procesarla sino solo usarla.

Para finalizar, el análisis de los registros, o sea los reportes de KPI's, puede hacerse

de forma automática y periódica. Esto gracias a que toda la información esta centralizada y se actualiza tantas veces como la incorporación de nuevos datos lo permita. Dando lugar a que las reuniones no dependan del proceso de actualización quincenal que ahora se tiene. Por otro lado, siguiendo con el análisis de los datos, si se quita Excel como generador de reportes y se usa algún *software* como los comentados en el capítulo 2, que son especializados para el manejo de grandes volúmenes de datos y además muy intuitivos en su uso, permitiría al personal encargado de los reportes optimizar el tiempo dedicado a esta tarea.

Capítulo 4

Propuesta de Solución

En este capítulo se detalla la propuesta de solución que busca abordar los problemas y oportunidades descritos en el capítulo anterior. En resumen, el problema se reduce a la complejidad de acceso a los datos de despacho, lo que se traduce en una sobrecarga de trabajo rutinario a los funcionarios, aumentando el periodo entre reportes. La principal oportunidad es automatizar el proceso utilizando herramientas tecnológicas. Como solución se propone un sistema integrado por dos partes. Por un lado reunir, centralizar y estandarizar los datos del departamento de despacho. Y por el otro, construir un *dashboard* que facilite la visualización de los principales KPIs asociados al proceso.

4.1. Requisitos

Como punto inicial de una propuesta, se deben considerar los requisitos del proyecto. Estos pueden ser funcionales o no funcionales, y en particular para el caso de estudio son los siguientes.

4.1.1. Requisitos Funcionales

- Desplegar los KPI's comúnmente utilizados en una interfaz o *dashboard*.

- La información mostrada en el *dashboard* debe ser dinámica, esto quiere decir que los gráficos desplegados deben responder a filtros.
- Cargar automáticamente los datos provenientes de los proveedores.
- Capturar información desde bases de datos internas.
- Procesar los datos recibidos y alertar en caso de errores o inconsistencias.

4.1.2. Requisitos No Funcionales

- Procesar y cargar los datos debe ser fuera del horario de oficina.
- Restringir los permisos de edición del *dashboard*.
- La información desplegada debe ser consistente con los informes que ya maneja el área.
- El usuario que use el *dashboard* no debe necesitar conocimientos específicos de la plataforma en donde se implemente el *dashboard* para poder usarla, es decir, debe ser intuitiva.

4.2. Diseño de la Arquitectura del Sistema

Con los objetivos claros y los requerimientos extraídos de las entrevistas con los usuarios, se puede iniciar con el proceso de diseño e implementación del sistema. En esta sección se detalla cuales fueron las decisiones que se tomaron al momento de escoger la tecnología o diseño de una interfaz además de describir el sistema implementado.

4.2.1. Tecnologías disponibles

En contexto bancario explicado en el capítulo anterior, es fácil inferir que no debe ser necesario ser personal calificado en el área de las tecnologías para interactuar, tanto

con las bases de datos como con plataformas que permitan un análisis de los datos almacenado en ellas. Por esta razón, estas plataformas tecnológicas deben ser intuitivas y es con este enfoque que se escogen las diferentes tecnologías en este proyecto.

Base de datos

En general cuando la información es almacenada internamente, esta tiene un formato claro, más aun si esta información corresponde a un departamento en concreto dentro la empresa, como lo es en el caso de estudio planteado en el capítulo anterior. Es por ello que se puede categorizar naturalmente en tablas toda la información sin crear estructuras complejas. También cabe destacar que, para el modelo de datos de despacho, la cantidad de entidades es baja facilitando aun más estandarizar y ordenar la información.

Lo anterior lleva a seleccionar una base de datos relacional, y en particular, para el caso de estudio se hará uso de Microsoft SQL server, por ser una plataforma ampliamente disponible en el mercado, aunque al utilizar funciones del estándar SQL, la base de datos podría ser trasladada a cualquier otro motor de base de datos relacional.

Dashboard

Para ofrecer una solución amigable a los usuarios del sistema, es necesario cambiar Excel, como programa para crear y visualizar gráficos, por una plataforma que esté enfocada a la inteligencia de negocio. De los *software* mencionados en el capítulo 2 el *dashboard* es implementado en Tableau, que es considerado uno de los líderes en esta área. Esto se debe a múltiples factores como poseer una amplia base de usuarios activos, una amplia documentación y cursos entregados por la mismo Tableau que permite una curva de aprendizaje mucho más rápida, entre otros. Sin embargo, la principal razón de escoger esta plataforma es que tiene un esquema de trabajo intuitivo, lo que permite crear una solución que una organización puede utilizar sin necesitar que sus usuarios

tenga muchos conocimientos del *software*.

4.2.2. Modelo de datos

En la figura 4.1 se observa el modelo de datos para la solución propuesta, el que presenta 3 entidades principales: *Paquete*, *Etap*a y *Encargado*. Con estas se analiza la información de despacho de los productos. Adicionalmente, el modelo se complementa con otras 2 entidades menos relevantes, pero que igualmente logran guardar información útil para futuras iteraciones del proyecto.

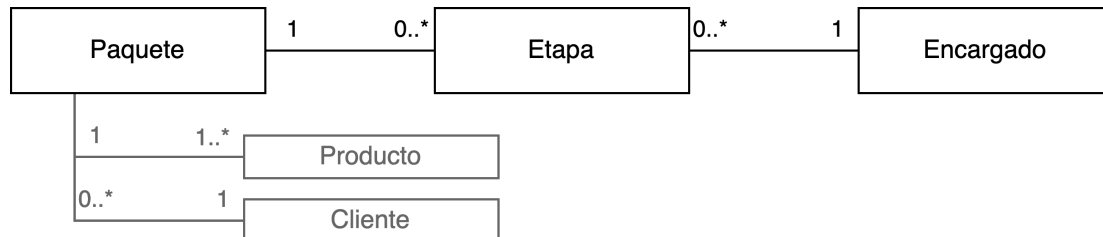


Figura 4.1: Modelo de datos

Entidades Principales

A continuación se describen las tres principales entidades donde se almacena la información de los paquetes y su estado.

1. Paquete

Esta entidad contiene el registro de todos los despachos, es decir, en términos generales, corresponde a la orden de despacho. Y en particular, para el caso de estudio, esta entidad se encarga de separar las emisiones de tarjetas que pertenecen a un Kit de las que no, ya que las primeras deben enviarse en conjunto.

- **ID**: Se crea un identificador de uso interno para individualizar los paquetes. Si bien existe la clave natural con las columnas Número + RutTitular + RutTh + Emisión, el *ID* permite facilitar las relaciones entre las entidades,

en particular entre Paquete y Etapa. Cuando se informa un nuevo registro en la vida de un paquete, se debe comprobar que este exista, para lo cual se utiliza parte de la clave natural (Número + RutTitular), que es lo que se informa junto al estado, por lo tanto al rescatar el paquete correspondiente, resulta conveniente enlazar con este identificador.

- **Número:** Corresponde al número de cuenta, en el caso de kit de cuenta corriente, o al número de tarjeta, en el caso de tarjetas de crédito o débito singulares. Se podría hacer un símil con el número de orden, pero con la diferencia que este número puede “eventualmente” duplicarse, lo cual lo descalifica como identificador único.
- **RutTitular:** Corresponde al RUT del titular de la cuenta o tarjeta, que en el caso general correspondería al dueño del paquete.
- **RutTh:** Corresponde al RUT del tarjeta habiente de la tarjeta, que en el caso general correspondería a quien recibe el paquete.
- **Emisión:** Corresponde a la fecha de emisión del primer producto. Esta fecha es una desnormalización, pues se duplica de la etapa inicial del paquete, pero dada la frecuencia con que se utilizará este atributo, es conveniente destacarla en esta entidad.

- **Tipo:** Corresponde al tipo de paquete. En el caso de estudio puede ser KIT, TC, TD, o CONDELL, dependiendo si es un kit de cuenta corriente, una tarjeta de crédito/débito o un paquete de otra unidad del banco. Se utilizará como filtro de información.

- **Causal:** Corresponde a la razón de la emisión del paquete. Para el caso de estudio, puede ser NUEVO, EMISIONES, REEMISIONES, RENOVACIONES, entre otros. Se utilizará también como filtro de información.

- **Segmento:** Es una clasificación interna, correspondiente al tipo de cliente titular. Tiene implicación en el flujo de determinación del encargado del despacho.

- **Encargado:** Corresponde al ID de la entidad responsable, inicialmente, de la entrega, pudiendo ser alguno de los proveedores o alguna sucursal en particular.

- **isRM:** Indica si el paquete es entregado en la región metropolitana. Este campo es una implementación particular del caso de estudio.

- **Registro:** Es una marca de tiempo que corresponde la fecha y hora de escritura del paquete.

2. Etapa

Esta entidad se encarga de guardar el historial del proceso de entrega de los productos. La mayoría de los estudios se realizarán sobre esta entidad, para medir progreso, hacer seguimiento y verificar entrega.

- **ID:** Número de identificación del registro.
- **Fecha:** Fecha de la etapa.
- **Estado:** Nombre de la etapa.
- **Motivo:** Información extra sobre la razón de estar en esa etapa.
- **Mesmovil:** Atributo particular de la implementación del caso estudio. En este se calcula la diferencia entre meses nominales desde la emisión a la fecha de la etapa, por lo tanto, no tiene relación directa con la cantidad de días pasados.
- **LT:** Atributo particular de la implementación del caso estudio. En este se calcula la diferencia, en días hábiles, entre la emisión a la fecha de la etapa, por lo tanto, días feriados y fines de semana no son contados.
- **Paquete:** Llave foránea del paquete correspondiente.
- **Encargado:** Llave foránea del encargado correspondiente.

3. Encargado

Esta entidad guarda la información sobre el responsable asociado a la etapa de un paquete. En el caso de estudio, corresponde tanto a proveedores externos, como a sucursales.

- **ID:** Identificador del encargado. En la implementación del caso estudio, se utilizó el código de sucursal interno y se añadieron códigos para los proveedores en un rango reservado.
- **Nombre:** Nombre del encargado
- **Tipo:** Separa los tipos de encargado, según las particularidades de la implementación. En el caso estudio se utilizan las clasificaciones Proveedor, Sucursal regiones, Sucursal RM.
- **Región:** Corresponde, si aplica, a la región que pertenece el encargado.
- **Comuna:** Corresponde, si aplica, a la comuna que pertenece el encargado.
- **Dirección:** Dirección del encargado.
- **Segmento:** Corresponde, si aplica, al segmento que atiende el encargado.

Entidades Complementarias

Este tipo de entidades, como se mencionó anteriormente, no son relevantes para el diseño del *dashboard*, esto quiere decir, que no son usadas al momento de desarrollar gráficos, mas si son relevantes para la implementación de las ETL.

1. Cliente

Esta entidad se encarga de guardar los identificadores R.U.T. y Nombre de los clientes. Se actualiza cada vez que ingresa una emisión de productos.

2. Producto

Esta entidad se encarga de guardar la información de los productos singulares emitidos, lo cual para el caso de estudio significa tarjetas de crédito y débito, pertenezcan o no a un Kit de cuenta corriente. Para el caso genérico, esta entidad sería el detalle de todas las órdenes de despacho.

Para finalizar esta sección, destacar que todos aquellos atributos que pueden estar duplicados, sirven para evitar operaciones innecesarias que demorarían el despliegue de gráficos, haciendo poco fluida la experiencia de usuario al momento de analizar la información. Un ejemplo de esto es el campo *Emisión* en Paquete y *Fecha* en Etapa, cuyos valores son idénticos cuando el producto esta en el estado de emisión.

Por otro lado, también hay atributos que se almacenan para evitar cálculos adicionales, como son *Mesmovil* y *LT*. Esto tiene como propósito evitar cálculos costosos, en términos de procesamiento, para aquellos valores que varían solo al momento de actualizar la información en la base de datos.

4.3. Prototipo del sistema

4.3.1. Esquema general del sistema

Una forma fácil de entender un sistema es plasmarlo en un diagrama de alto nivel, mostrando los diferentes módulos que lo componen e indicando como interactúan las diferentes partes de este sistema.

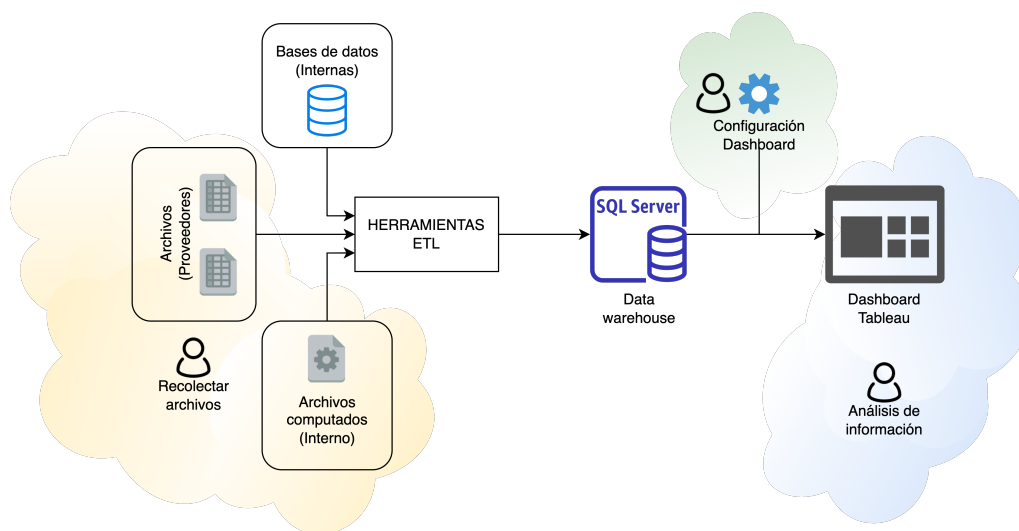


Figura 4.2: Esquema general de sistema

En la figura 4.2 se observa un diagrama que muestra las diferentes partes del sistema implementado. A la izquierda se pueden ver las fuentes de información que abastecen, una vez procesados, al *Data Warehouse*. Estas fuentes pueden ser bases de datos internas, archivos internos preprocesados o archivos de los proveedores de despacho. Los dos últimos, solo para el caso de estudio, son recolectados de forma manual por un funcionario interno. Estos archivos son entregados, junto con los demás registros, como entrada a la ETL para ser procesados, dándoles un formato e identificando data corrupta o inconsistente. Una vez el procesamiento de los registros termina, se carga la información al *data warehouse* en las entidades que se describieron en la sección anterior.

Por último, al lado derecho de la figura 4.2 se ubican los módulos de análisis técnico de la información almacenada, desplegando un resumen ejecutivo relevante para el departamento de despacho. Este resumen se presenta en un formato de *dashboard* utilizando herramientas especializadas en inteligencia de negocio, que para el caso de estudio es el *software* Tableau. En este segmento se observan dos módulos, el primero se encarga de generar el *dashboard* con la información requerida, con un enfoque intuitivo para la interfaz. el segundo corresponde al usuario final del *dashboard*, aquel que lee el reporte analizando los gráficos para tomar decisiones de negocio. Cabe destacar que en esta sección del diagrama, ambos módulos corresponden a uno solo y, como se ha mencionado antes, es completamente independiente del área tecnológica de la empresa.

4.3.2. Interfaz de usuario

A continuación se presentan las tres vistas principales del dashboard, aplicadas según lo esperado para el caso estudio.

Vista Inicial

En la figura 4.3, se presenta la vista inicial del sistema, que se caracteriza por hacer un resumen de toda la información almacenada de los últimos 6 meses. Como se observa, la interfaz se divide en tres secciones.

En la sección A, se encuentra una columna con dos indicadores. El primero compara las entregas a la fecha del mes en curso con el mismo periodo del mes anterior. El segundo, muestra tanto el porcentaje de entregas del mes actual como del mes anterior en el mismo periodo de tiempo. Ambos indicadores se muestran primero en forma general para luego ser desglosada por tipo de producto. Adicionalmente, el contenedor de los indicadores usados para los productos como un botón que direcciona al usuario a otros *dashboard*.

Los gráficos de *Emisiones Mensuales* y *Entregados*, se ubican en la sección *B*. A la izquierda se resumen las emisiones de los últimos 6 meses, detallada por tipo y cantidad de producto, marcando proporciones y cantidades absolutas de cada uno. Por otro lado, el gráfico a la derecha muestra cuanto se demoran en entregar cada producto, mostrando el porcentaje acumulado del mes móvil en el que se entrego el producto, entendiendo como mes móvil lo indicado en el atributo de la entidad etapa (capítulo 5.2.2).

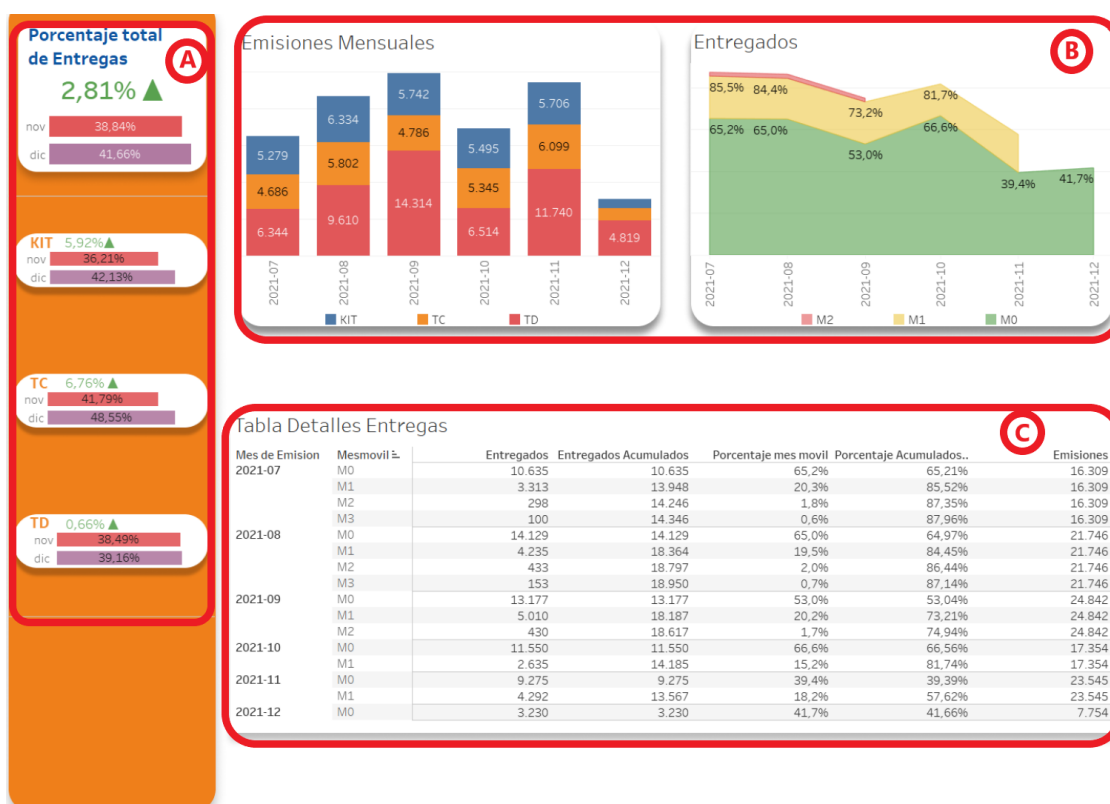


Figura 4.3: Vista inicial del dashboard

Para finalizar, la sección *C* contiene una tabla con el detalle de los dos gráficos anteriores, agregando el parámetro de entregas, mostrando esta información de forma absoluta y acumulada por mes móvil y mes de emisión.

Vista Tarjetas

El propósito de esta vista es agrupar la información de las tarjetas que no pertenecen a un Kit de cuenta corriente. Al igual que la vista anterior, para mayor claridad en la explicación, el *dashboard* de la figura 4.4 se divide en secciones.



Figura 4.4: Vista del dashboard para tarjetas

Al igual que la vista anterior, la sección A muestra el indicador de diferencia porcentual y los gráficos de porcentajes de entrega del mes actual y el mes anterior manteniendo el mismo periodo de tiempo. No obstante, el primero esta presente solo en la parte superior de la columna, ya que es un resumen global de las entregas y debe estar presente en todas las vistas. Por otro lado, los botones de esta columna no direccionan al usuario, sino que su propósito es filtrar por tipo de producto al que se refieren las métricas de todo el *dashboard*, destacando en blanco el producto que se está filtrando. Por último, en esta vista, el contenedor del *Porcentaje total de entregas* funciona como botón que direcciona al usuario a la **Vista Inicial**.

La sección *B* muestra el mismo tipo de información que el descrito para la figura 4.3. El principal cambio es que ambos gráficos varían dependiendo de los filtros, de las secciones *A* y *D*, activos sobre el *dashboard*. Cabe destacar que, si bien se muestra el mismo tipo de información en esta sección (emisiones y entregas), las emisiones no se dividen por tipo de producto sino que ahora la división se hace por encargado de la entrega, es decir, proveedores de despacho (PROVEEDOR), sucursales de regiones (REGIONES), y sucursales de la región metropolitana (RM).

En lo referente a la sección *C*, esta consta de dos partes, y cada una tiene filtros que afectan los gráficos de *A* y los ubicados en esta sección. La parte superior muestra detalles importantes de los proveedores, teniendo un filtro desplegable que contiene todos los proveedores de despacho contratados por la empresa. Mientras que la parte inferior el filtro agrupa las sucursales en regiones y región metropolitana. Los graficos que se ubican acá, corresponden a al atributo *LT*, explicado en la capítulo 5.2.2, para los estados entregado, en ambos casos (sucursal y proveedores), agendado y contactado. Para cerrar esta sección, destacar que se debe mantener concordancia entre los filtros aplicados en *C* con los aplicados en *D*, de lo contrario los gráficos de *A* no tendrán valores que mostrar.

Finalizando con esta vista, la sección *D*, ubicada en la parte superior del *dashboard*, presenta dos filtros: Proveedores y Sucursal. Esto afectan a las secciones *A* y *B* modificando la información presentada en los gráficos de emisión y entregas por mes móvil, interceptando el filtro aplicado en *A* con los aplicado por esta sección. El funcionamiento de estos filtro esta relacionado a los *tickets* en negro que se puedan ver. Cuando ambos están presente no hay filtros aplicados, por lo que la información mostrada en las demás secciones corresponde a el total de la información. Al hacer seleccionar uno de los recuadros desaparece el *ticket* no seleccionado filtrando los registros del producto por ese parámetro. Para para desactivar el filtro aplicado se debe hacer *click* sobre el parámetro seleccionado, con esto aparece el *ticket* que falta y la vista vuelve a su estado "sin filtro".

Vista Kit

La última vista implementada, agrupa toda la información relacionada a los kit de cuenta corriente. Como en la vista anterior, la figura 4.5 se divide en cuatro parte para facilitar la explicación.



Figura 4.5: Vista del dashboard para kits

El *dashboard* de las figura 4.5, es idéntico al de la figura 4.4 en cuanto a estructura, pero, debido a la información que muestra, algunas funciones no se encuentran en esta vista.

Las secciones A y B se mantienen en funcionalidad pero muestran solo la información de kits de cuenta corriente.

Con respecto a la sección C, esta presenta un cambio que se debe únicamente al caso de estudio. La parte superior de esta sección ya no cuenta con la lista de proveedores de despacho, debido a que para el producto kit la entidad bancaria solo trabaja con una empresa. A pesar de esto, las parte de sucursales sigue funcionando igual que en la

Vista Tarjetas.

Por último, los filtros de la sección *D*, funcionan igual de que la vista anterior, alterando las secciones *A* y *B* al mismo tiempo.

Cabe destacar que la razón de replicar la **Vista Tarjeta** en la **Vista Kit** fueron dos, evitar la sobre carga de información al momento de calcular los filtros, lo que haría a la interfaz muy lenta, impactando directamente en la experiencia del usuario. Y el segundo motivo está enfocado en el usuario final y busca no abrumarlo con muchos datos en una única vista.

4.3.3. ETLs

Las herramientas ETL, como se mencionó en el capítulo 3, son las encargadas de recibir los archivos fuente, procesar sus datos para ajustarlos al modelo de datos propuesto y almacenarlos en la base de datos definitiva, como se muestra en diagrama de la figura 4.6.

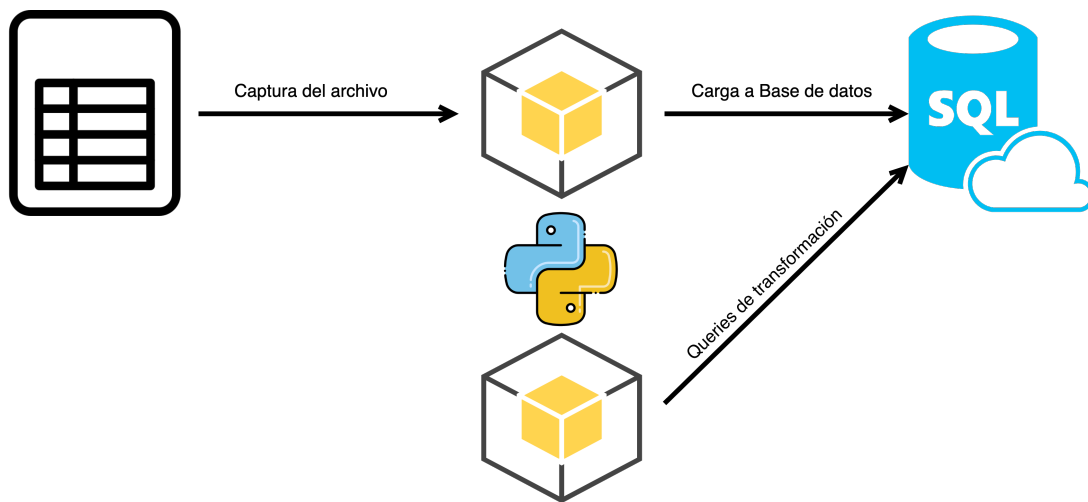


Figura 4.6: Diagrama de ETL

En el diagrama, se muestra que existe un script en python encargado de capturar los archivos entregados. Este script tiene dos frentes de acción, primeramente, la extracción en sí, donde se seleccionan las columnas de interés, se rellena con una columna

con la fecha correspondiente, y se renombran las columnas, para luego cargar las filas extraídas en una serie de tablas, cuyo rol es ser el backlog de estados por incorporar.

Una vez se extraen todos los datos informados, viene el segundo frente, correspondiente a la ejecución de procedimientos almacenados que tomen lo informado en las bases backlog, contrastarlo con lo existente en la base de datos final y aprobar su incorporación. Este paso nos permite que la plataforma sea flexible ante descoordinaciones de datos en la tabla de etapas, pues algo informado se mantendrá en el backlog hasta que le corresponda ser incorporado.

Capítulo 5

Validación del Prototipo

A lo largo de esta memoria, se presentó el proceso utilizado para dar solución al problema que existe cuando un sistema dedicado al análisis de datos no está actualizado. Durante este proceso se reunió información del procedimiento actual (*Capítulo 3*), se identificaron requerimientos, se diseñó una propuesta de solución y se presentó la primera iteración de la propuesta (*Capítulo 4*).

Dentro del contexto de la primera iteración, a continuación se describen algunos casos de uso tanto del *dashboard* como del proceso que actualiza la información en las bases de datos.

5.1. Casos de Uso: *Dashboard*

Con el prototipo de interfaz propuesto, a continuación se presentan 3 KPIs que se pueden obtener de él. En este contexto, cada KPI se relaciona directamente a uno de los *casos de uso* que se presentan, y estos comienzan siempre en la *Vista Inicial* de la figura 5.1, ya que es la primera vista a la que accede el usuario al momento de utilizar el *dashboard*. Además, como se mencionó en el capítulo anterior, todos los elementos que se encuentran en la sección A en esta interfaz, que representan productos, corresponden a botones para cambiar de vista. Para el caso de estudio solo hay tres tipos de producto: TC, TD y KIT.

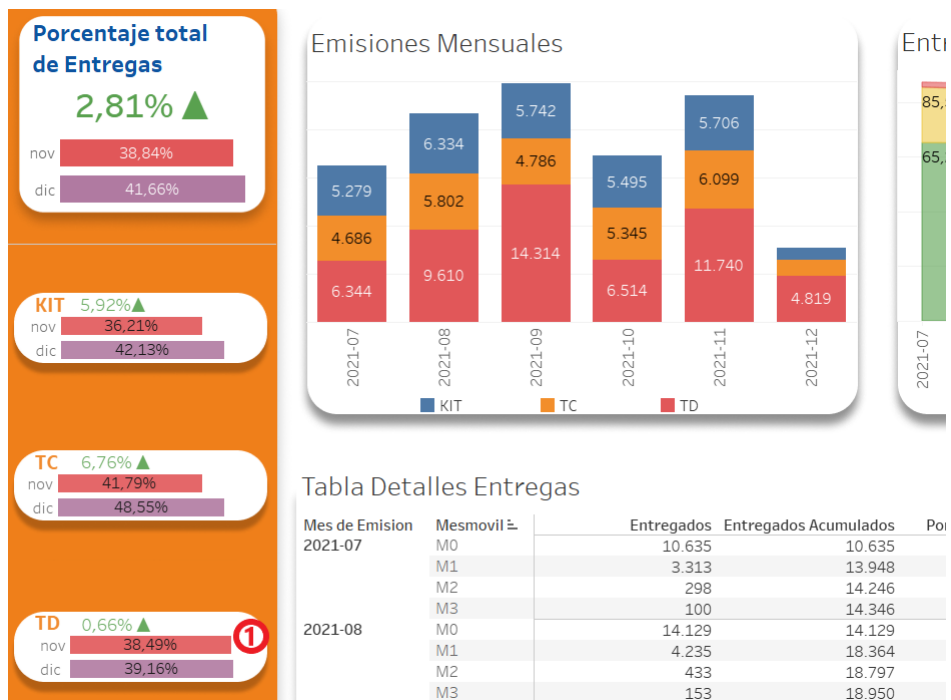


Figura 5.1: Vista Inicial para todos los casos de uso

5.1.1. Cantidad de Emisiones de Tarjetas de Débito por Mes

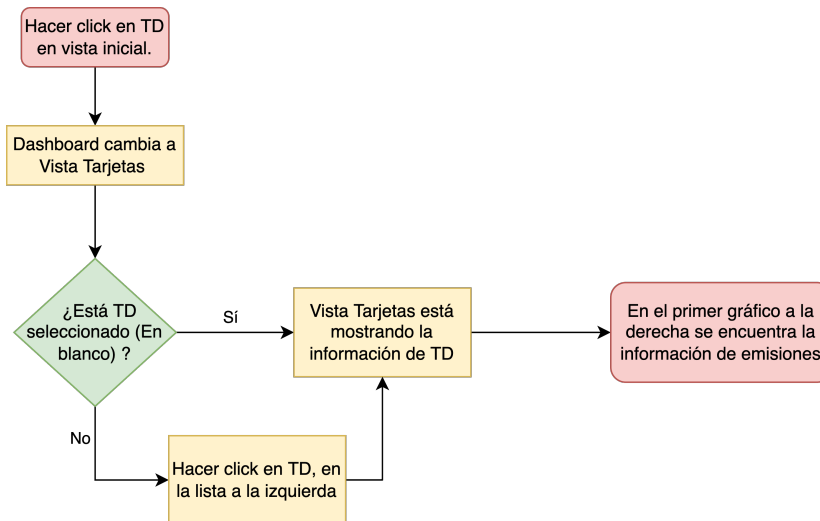


Figura 5.2: Diagrama lógico de interacción para el caso de uso 1

En el diagrama de la figura 5.2 se puede distinguir el flujo que debe seguir el usuario para obtener la información relacionada a las emisiones de las tarjetas de débito. Estos

pasos se detallan a continuación.

1. Hacer clic en el producto TD (figura 5.1) para cambiar a la *Vista Tarjetas*.
2. Una vez en la *Vista Tarjetas*, a diferencia de la vista inicial, solo un producto aparece en blanco, como se ve en la figura 5.3, esto significa que dicho producto esta seleccionado y la información presentada en el *dashboard* corresponde a él.
 - a) Si no se encuentra seleccionado TD en la *Vista Tarjetas*, clicar en TD genera que toda la información presentada en el *dashboard* solo pertenezca a las tarjetas de débito (figura 5.3).
3. Las emisiones se encuentran en el gráfico *Emisiones Mensuales* de la sección B, separadas por PROVEEDOR, RM y REGIONES.

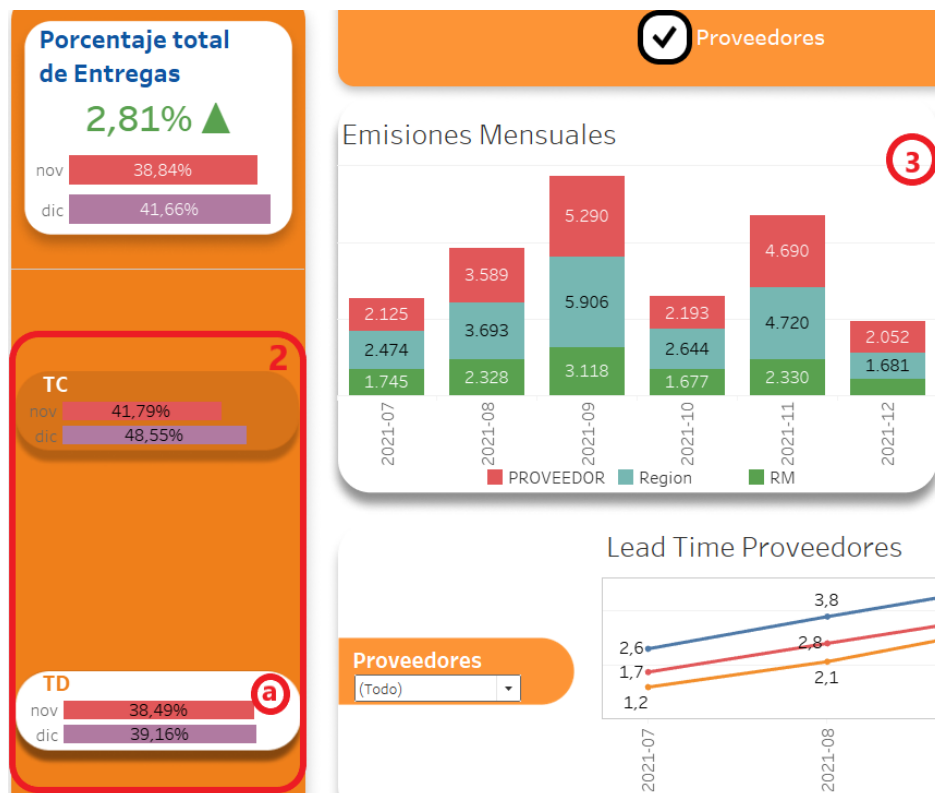


Figura 5.3: Emisiones de tarjetas de débito por mes

Para visualizar este mismo caso de uso para otros productos, comenzando en *Vista Inicial*, se debe clicar en el producto deseado y una vez la vista cambie, este debe estar de color blanco para que el *dashboard* muestre la información relacionada a este producto.

5.1.2. Tiempo de Entrega Promedio de las Tarjetas de Débito por un proveedor por Mes

En el capítulo 5.2.2, se presentaron las entidades del modelo de datos, donde se describió la entidad *Etapa* que se encarga de guardar la parte del proceso de entrega en la que se encuentra un paquete, y uno de sus atributos es *Lead Time* (LT), que corresponde al tiempo de demora. Se usa este atributo derivado para medir el tiempo de espera. Cada vista tiene su propio gráfico que muestra esta información.

Al igual que en el caso de uso anterior, se presenta un diagrama de flujo, ver figura 5.4, para describir de manera lógica los pasos que debe seguir el usuario para obtener la información relacionada al tiempo de demora.

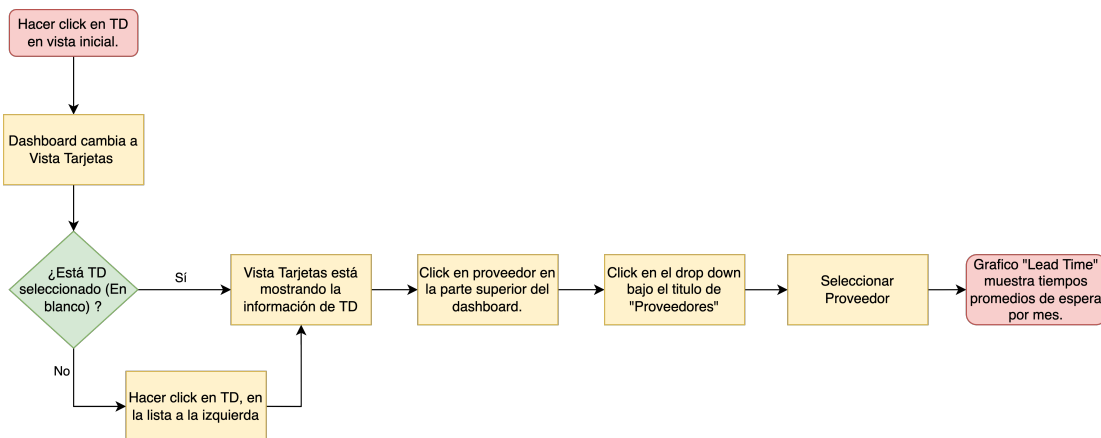


Figura 5.4: Diagrama lógico de interacción para el caso de uso 2

A continuación, una descripción detallada de la interacción del usuario con el *dashboard*:

1. Hacer clic en el producto TD (figura 5.1), para cambiar a la *Vista Tarjetas*
2. Si no se encuentra seleccionado TD en la *Vista Tarjetas*, clicar en TD genera que toda la información presentada en el *dashboard* solo pertenezca a las tarjetas de débito (figura 5.5).
3. Seleccionar el filtro de proveedores de la sección *D* para que toda la información presente en la interfaz sea relacionada a los proveedores (figura 5.5).
4. En la sección *C* de la vista se muestran los gráficos asociados a la variable *Lead Time* divididos por proveedor y por sucursales (figura 5.5).
 - a) A la derecha de esta sección se encuentra el filtro que permite mostrar información del gráfico *Lead Time* por proveedor de servicio de despacho.
 - b) Clicando en el filtro y seleccionando un elemento de la lista que despliega, el gráfico de *Lead Time* cambia, mostrando únicamente información de del proveedor seleccionado.

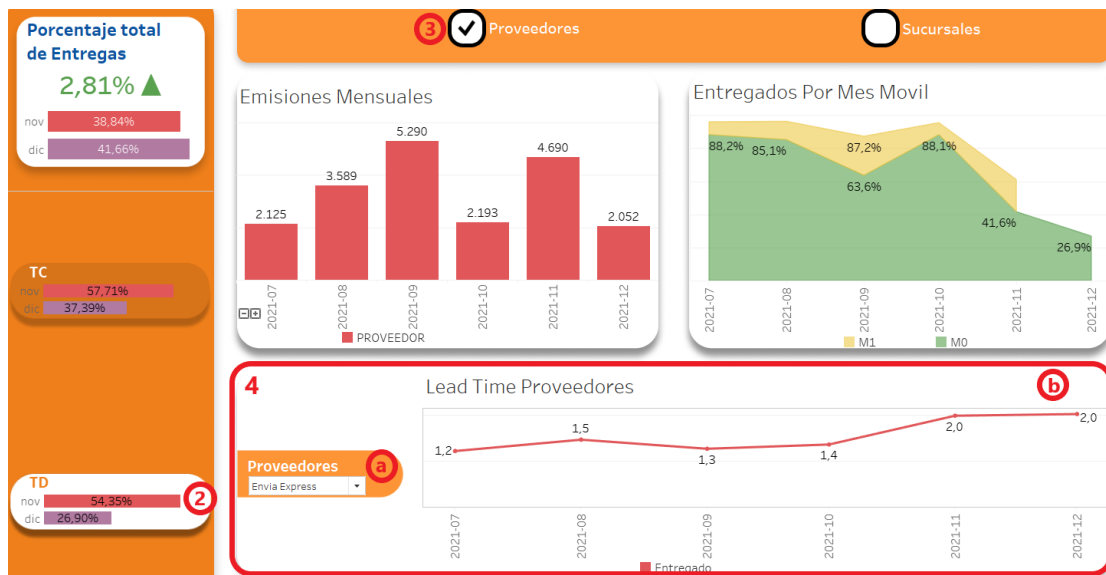


Figura 5.5: Lead Time para proveedor Envía Express

5.1.3. Diferencia del porcentaje de entrega del mes actual con el mes anterior durante el mismo periodo de tiempo

Para este caso de uso, el usuario necesita saber si, para el día en que se esta usando el *dashboard*, el porcentaje de entregas de un producto en particular es mayor o menor que el porcentaje de entregas del mismo producto al mismo día pero del mes anterior. Por ejemplo, si el usuario está viendo el *dashboard* el día 9 de diciembre, se compararían las entregas hasta esta fecha con las realizadas hasta el 9 de noviembre.

1. Para esta métrica, solo basta estar en la *Vista inicial*, ya que la sección A muestra esta información en cada uno de sus elementos como gráficos de barra horizontal con el porcentaje sobre las barras, indicando el mes de cada una. Esto se puede ver en la figura 5.1.

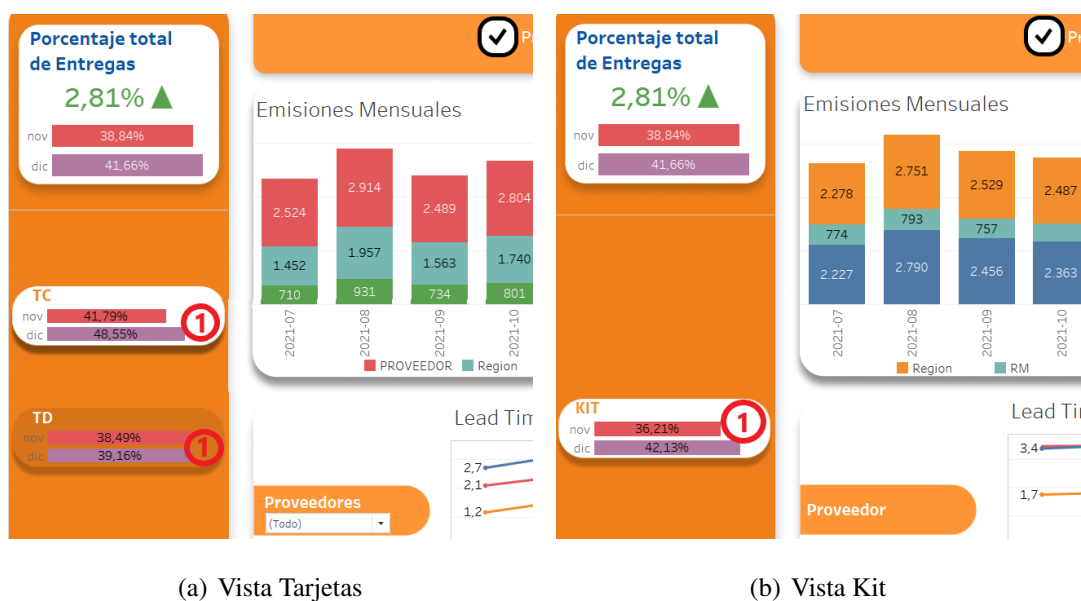


Figura 5.6: Gráfico de diferencia porcentual en las vistas Tarjetas y Kit

Por otro lado, cada vista tiene este indicador, y si el usuario lo desea puede aplicar filtros de proveedor o sucursal sobre el. En la figura 5.6 se puede ver este KPI en las vistas tarjetas y kit.

5.2. Casos de Uso: *Datawarehouse*

El único caso de uso que se presenta relacionado con la base de datos, es capturar los reportes que envían los proveedores para actualizar los registros almacenados en las fuentes de datos internas de la empresa.

Dado que el caso de estudio no permite que las empresas proveedoras accedan a la infraestructura internas de la empresa por temas de seguridad, existen dos opciones: Crear una infraestructura dedicada que supla las necesidades de seguridad de la empresa, o mantener el proceso que se tiene actualmente que reúne los reportes entregados por los proveedores en una carpeta, para luego incorporarlos a la base de datos. Considerando que crear una infraestructura escapa del alcance de este proyecto, se decide dejar los reportes en una carpeta, y es en este contexto, que el diagrama de interacción del usuario para ejecutar la ETL, ver figura 5.7, y su respectiva explicación se detallan a continuación.



Figura 5.7: Diagrama de interacción para la ETL

1. El empleado asignado debe reunir los reportes entregados por los proveedores para dejarlos en el lugar definido al momento de implementar la solución.
2. Ejecutar el programa desarrollado en python que corresponde a la ETL. Una vez la ejecución termine el programa informa la cantidad de filas agregadas a la base de datos, tal y como se muestra en la figura 5.8.

Informe de productos entregados 10-12-2021.xlsx
Dpro_Entregas 2021-12-10
Se agregaron 209 filas correspondientes a tarjetas
Se agregaron 67 filas correspondientes a KIT
Se agregaron 38 filas correspondientes a Rechazos
removido EXCELS/folder_new/Informe de productos entregados 10-12-2021.xlsx
Informe de productos entregados 27-11-2021.xlsx
Dpro_Entregas 2021-11-27
Se agregaron 131 filas correspondientes a tarjetas
Se agregaron 192 filas correspondientes a KIT
Se agregaron 8 filas correspondientes a Rechazos
removido EXCELS/folder_new/Informe de productos entregados 27-11-2021.xlsx

Figura 5.8: Información de salida con archivos de entregas

Estas son todas las intervenciones que el usuario tiene en relación con las fuentes de datos. Terminado el paso 2, la información ya se encuentra disponible para su uso por el *dashboard*.

Capítulo 6

Conclusión

A lo largo de este trabajo de título, se desarrolló una solución que automatiza la extracción de información desde diferentes fuentes y la creación de informes para el control de operaciones. Se abordó la problemática a través de un caso de estudio, sin embargo, esta solución se puede utilizar en cualquier empresa o departamento que tenga la misma necesidad.

Con el contexto del problema, se estudió la necesidad del usuario con el objetivo de determinar sus requerimientos, para luego presentar una solución que se dividió en dos partes: el *almacenamiento y procesamiento de los datos* y la *interfaz*. La primera considera la gestión de los datos que provienen de diferentes fuentes con el propósito de centralizarlos. Mientras que la segunda parte, corresponde al análisis de la información almacenada, creando un *dashboard*, con varias vistas, capaz de agrupar los datos y entregar los índices claves para monitorear el rendimiento de los procesos asociados a estas métricas, que en el caso de estudio corresponden a las empresas de despacho y las entregas realizadas por las sucursales.

Separar el desarrollo en dos, permitió trabajar una solución a la medida para ambas partes de la problemática. Por el lado de la base de datos, debido a que el proceso era completamente manual, se encontraron varias dificultades propias del caso de estudio, como la recepción de informes por parte de las empresas externas, que carecían de un

formato compatible con los registros almacenados en las fuentes de datos internas de la entidad bancaria. También, y gracias a la manera de recibir los datos de las empresas de despacho, se debía reunir todos los reportes para juntarlos con los registros antiguos, almacenados sin ningún sistema tecnológico que ayudara a evitar errores o hiciera el trabajo de forma automática. Ahora bien, la parte que se relacionaba con los reportes de la información recolectada también presentaba oportunidades importantes de automatización, ya que estos se confeccionaban manualmente cada vez que se requerían, limitando la información que se podía presentar por dos principales factores: los datos nuevos estuvieran incorporados a los antiguos y que el personal designado pudiera desarrollar el informe en un tiempo prudente.

Por estas razones, la solución del problema se encontraba principalmente en la automatización de los procesos que agregan nueva información y que generan reportes con los registros guardados. Se implementó un sistema para evitar que el usuario manipule los datos al momento de actualizar la información que llega diariamente, y este mismo sistema procura que los reportes sean automáticos sin la necesidad de ser actualizados mes a mes para alguna reunión. Sin embargo, debido al contexto del caso de estudio, no se pudo separar completamente al usuario del proceso, ya que este debe juntar los informes de las empresas externas y guardarlos en una carpeta creada exclusivamente para este propósito.

En conclusión, con este trabajo se logró mejorar la eficiencia en los recursos del equipo, que para el caso de estudio fue el departamento de despacho de la entidad bancaria. Esta mejora consiste específicamente en la reducción de *horas hombre* utilizadas en tareas rutinarias, lo que permite disponerlas para la labor propia del área. También se simplifica el flujo de trabajo, eliminando los errores asociados a la manualidad del proceso, y deja a disposición la información en un periodo de espera que solo depende de la periodicidad con la que se reciban los reportes, pudiendo así ser utilizada por el *dashboard*, actualizándose automáticamente junto con la base de datos, mejorando ampliamente las operaciones del equipo.

Bibliografía

- [1] DOMO. Data Never Sleeps 9.0, 2022.
- [2] Hsinchun Chen, Roger H L Chiang, Veda C Storey, Carl H Lindner, and J Mack Robinson. Business intelligence and analytics: From big data to big impact, 2012.
- [3] Ronda Harrison, Angelique Parker, Gabrielle Brosas, Raymond Chiong, and Xue-mei Tian. The role of technology in the management and exploitation of internal business intelligence. *Journal of Systems and Information Technology*, 17:247–262, 8 2015.
- [4] Papa Senghane Diouf, Aliou Boly, and Samba Ndiaye. *Variety of data in the ETL processes in the cloud: state of the art*. 2018.
- [5] Yu Wei Chang, Ping Yu Hsu, and Zeng Yuan Wu. Exploring managers’ intention to use business intelligence: The role of motivations. *Behaviour and Information Technology*, 34:273–285, 3 2015.
- [6] Microsoft. Environmental services company builds sustainable, data-driven solutions with Power BI and Azure, 2022.
- [7] InStoreView. Inka Crops: Decisiones rápidas en el Canal Moderno, 2022.
- [8] Austin Kronz, Kurt Schlegel, Julian Sun, David Pidsley, and Anirudh Ganeshan. Magic quadrant for analytics and business intelligence platforms. 3 2022.
- [9] Fortune Business Insights. Market Search Report, 2022.

- [10] David Reinsel, John Gantz, and John Rydning. The digitization of the world from edge to core, 2018.
- [11] Gautam Srivastava, S. Muneeswari, Revathi Venkataraman, V. Kavitha, and N. Parthiban. A review of the state of the art in business intelligence software. *Enterprise Information Systems*, 16:1–28, 2022.