UNIVERSIDAD DE CUENCA

Facultad de Ingeniería

Carrera de Ingeniería de Sistemas

Microservicios aplicados a la integración y análisis de datos orientados al cálculo de una tarifa diferencial para aparcaderos de vehículos privados. Caso de estudio: Campus Central de la Universidad de Cuenca.

Trabajo de Titulación previo a la obtención del Título de Ingeniero de Sistemas

Autores:

Freddy Leonardo Abad León

Cl: 0104496765

Correo electrónico: fre-leoabad@hotmail.com

Esteban Dario Vizhñay Enderica

Cl: 0104781422

Correo electrónico: stebanviz@gmail.com

Tutora:

Ing. Elina María Ávila Ordóñez PhD. (c)

Cl: 0917624868

Co-tutor:

Ing. Víctor Hugo Saquicela Galarza PhD.

Cl:0103599577

Cuenca-Ecuador

12 de octubre de 2020

Resumen:

La ciudad de Cuenca, Ecuador ha evidenciado en la última década un incremento en el uso de modos motorizados, lo que ha contribuido a la congestión vehicular en sus zonas céntricas. Para atender a esta nueva demanda, se observa el crecimiento en la oferta de aparcaderos que operan de forma independiente sin una regulación de las tarifas a cobrar. Así, en la ciudad resulta relativamente barato aparcar vehículos lo que ha facilitado la masificación de su uso. Para evitar las consecuencias negativas asociadas a una movilidad motorizada se debería definir un plan tarifario que incida en la selección del modo de transporte. El objetivo de este trabajo es implementar el modelo matemático propuesto por el Grupo de Investigación Movilidad Activa y Sostenible de la Universidad de Cuenca que define un plan tarifario diferencial del uso de aparcamiento con caso de estudio del Campus Central de la Universidad de Cuenca. Los métodos utilizados son Hefesto, Crisp-DM y Modelo Tarifario: el primero utilizado en la construcción de un Data Warehouse para consolidar los datos recolectados en los aparcaderos del campus y visualizar indicadores de interés a través de cubos de información; el segundo siguió un ciclo de procesos ordenado para extracción de conocimiento del Data Warehouse mediante técnicas de Data Mining; finalmente, se validó el modelo matemático comparado tarifas personalizadas para los servidores universitarios. Los materiales utilizados fueron: Modelo Tarifario, Datos Administrativos y de Movilidad; y, entre las tecnologías utilizadas se encuentran: Suite de Pentaho, Spring Boot, entre otros. Los resultados obtenidos incluyen un Data Warehouse para responder preguntas de interés, además de ser una fuente de datos para aplicar Minería de Datos con sus consecuentes modelos de Inteligencia Artificial, y la creación de un microservicio para calcular la tarifa personalizada mediante el modelo, que es utilizado por un aplicativo web al alcance del grupo de movilidad. Finalmente, se concluye analizando el uso del Data Warehouse, y observando su impacto en el modelo; y la tarifa que tendría cada usuario, con base a los objetivos centrales de la tarifa de cobro.

Palabras claves: Microservicios. Data warehouse. Data mining.

Freddy Leonardo Abad León
Esteban Dario Vizhñay Enderica

Abstract:

The use of motorized modes in Cuenca, Ecuador, has increased in the last decade, contributing to traffic congestion in its central area. To meet this growing demand, the supply of parking lots that operate independently without regularization on the rates to be charged is increased. Thus, in the city, parking a vehicle is cheap, which has facilitated its massification. To avoid the negative consequences associated with motorized mobility, it is convenient to define a rate plan to influence the mode choice. The aim of this study is to implement the mathematical model proposed by the Active and Sustainable Mobility Group of the University of Cuenca defining a differential rate plan for the use of parking with a case study of the Central Campus of the University of Cuenca. The methods used are Hefesto, Crisp-DM and Tariff Model: the first one used in the construction of a Data Warehouse to consolidate the data collected in the campus parking lots and visualize indicators of interest through information cubes; the second followed an orderly process cycle for the extraction of knowledge from the Data Warehouse through Data Mining techniques; finally, the mathematical model was validated by comparing personalized rates for university servers. The materials used were: Tariff Model, Administrative and Mobility Data; and, among the technologies used are: Pentaho Suite, Spring Boot, among others. The results obtained include a Data Warehouse to answer questions of interest, in addition to being a data source to apply Data Mining with its consequent Artificial Intelligence models, and the creation of a microservice to calculate the personalized rate through the model, which is used by a web application available to the mobility group. Finally, it is concluded by analyzing the use of the Datawarehouse, and observing its impact on the model; and the rate that each user would have, based on the central objectives of the collection rate.

Keywords: Microservices. Data warehouse. Data mining.

Índice

Índice de Formulas

Índice de Tablas Capitulo 3

Índice de Figuras

Índice de Anexos

Listado de Abreviaturas

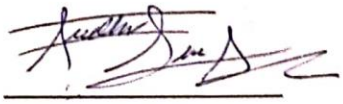
- ANT: Agencia Nacional de Tránsito
- API: Application Programming Interface
- Crisp-DM: Cross Industry Standard Process for Data Mining
- CSS: Cascading Style Sheets
- DB: Data Base
- DI: Data Integration
- DM: Data Mining
- DS: Data Sources
- DTICS: Dirección de tecnologías de la información y comunicación de la Universidad de Cuenca
- DW: Data Warehouse
- ETL: Extract, Transform and Load
- GAD Cuenca: Gobierno autónomo descentralizado municipal del cantón cuenca
- HAL: Hypertext Application Language
- HTML: HyperText Markup Language
- HTTP: HyperText Transfer Protocol Secure
- IA: Inteligencia Artificial
- IEEE: Institute of Electrical and Electronics Engineers
- JS: JavaScript
- JSON: JavaScript Object Notation
- KDD: Knowledge Database Discovery
- K-NN: K-Nearest Neighbors
- MAS: Grupo de Investigación Movilidad Activa y Sostenible de la Universidad de Cuenca
- ML: Machine Learning
- MSA: Microservice Architecture
- OAuth: Open Authorization
- OLAP: On-Line Analytical Processing
- OLTP: OnLine Transaction Processing
- PBI: Pentaho Business Intelligence
- PDI: Pentaho Data Integration
- PSW: Pentaho Schema Workbench
- REST: Representational State Transfer
- SGBD: Sistema de Gestión de Bases de Datos
- SPA: Single Page Application

- SQL: Structured Query Language
- SOA: Service Oriented Architecture
- TCP/IP: Protocolo de Control de Transmisión
- TS: TypeScript
- WWW: World Wide Web
- XML: Extensible Markup Language

Cláusula de Propiedad Intelectual

Freddy Leonardo Abad León, autor del trabajo de titulación Microservicios aplicados a la integración y análisis de datos orientados al cálculo de una tarifa diferencial para aparcaderos de vehículos privados. Caso de estudio: Campus Central de la Universidad de Cuenca.", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, 12 de octubre de 2020



Freddy Leonardo Abad León

C.I: 0104496765

Cláusula de Propiedad Intelectual

Esteban Dario Vizhñay Enderica, autor del trabajo de titulación "Microservicios aplicados a la integración y análisis de datos orientados al cálculo de una tarifa diferencial para aparcaderos de vehículos privados. Caso de estudio: Campus Central de la Universidad de Cuenca.", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, 12 de octubre de 2020

Esteban Dario Vizhñay Enderica

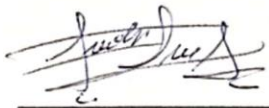
C.I: 0104781422

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Freddy Leonardo Abad León en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación "Microservicios aplicados a la integración y análisis de datos orientados al cálculo de una tarifa diferencial para aparcaderos de vehículos privados. Caso de estudio: Campus Central de la Universidad de Cuenca.", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 12 de octubre de 2020



Freddy Leonardo Abad León

C.I: 0104496765

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Esteban Dario Vizhñay Enderica en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación "Microservicios aplicados a la integración y análisis de datos orientados al cálculo de una tarifa diferencial para aparcaderos de vehículos privados. Caso de estudio: Campus Central de la Universidad de Cuenca.", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 12 de octubre de 2020

Esteban Dario Vizhñay Enderica

C.I: 0104781422

Agradecimiento

A Dios por guiar mis pasos, a mis padres, Martha y Luis, por siempre apoyar y orientar sabiamente mis decisiones, a mi hermana, Alexandra, por animarme a lograr todo objetivo que me proponga.

A todos los docentes, con quienes tuve la fortuna de recibir instrucción; por forjarme como un profesional responsable, de manera especial a la tutora de tesis, Ing. Elina Avila y al Ing. Victor Saquicela co-tutor, por guiarnos, brindándonos su tiempo y canalizando su conocimiento para la obtención de este trabajo de titulación.

A la Lic. Rosita Avila por aconsejarme sabiamente en los momentos más difíciles de la vida universitaria.

A mis colegas de carrera: Paola, Esteban, Fernando, Bryan y Edisson por hacer de la vida universitaria más cálida.

Al grupo de Movilidad Activa y Sostenible, por la colaboración y tiempo ofrecido oportunamente para la obtención de este trabajo de titulación.

Finalmente, agradezco a toda la comunidad universitaria: vinculación con la sociedad, asos escuelas, grupos académicos, etc. Todos los espacios donde colabore, en esta gran Universidad, donde se me recalco vivamente la importancia de ser un profesional con responsabilidad social.

Freddy Abad L.

Agradecimiento

A Dios, por brindarme padres que me fomentaron valores esenciales para tener una vida con metas y poderlas cumplir, a mis padres Mauro y Mariana “Viejos” por apoyarme y nunca dejarme solo en cada momento a lo largo de este gran camino, a pesar de las dificultades siempre estuvieron ahí para mí, agradezco nuevamente a Dios por gozar de ese privilegio.

A mi hermano, Pablo por sus consejos para convertirme en un buen profesional, mis hermanas, Nathaly y Paulina por estar siempre ahí a mi lado sacándome sonrisas y mostrarme su apoyo siempre.

A los docentes, que compartieron su tiempo y conocimiento conmigo, lo que ha permitido forjarme como un buen profesional, de manera especial a mi tutora de tesis, Ing. Elina Avila y al Ing. Victor Saquicela co-tutor, por el apoyo y su tiempo brindado para poder finalizar con éxito este trabajo de titulación.

A mis colegas de carrera, Christian, Freddy, Bryan, Edisson, William, Paola y Paul por estar en todo momento brindando su apoyo y permitir que cada ciclo sea lo más relajado posible.

Al grupo de Movilidad Activa y Sostenible, por la colaboración y tiempo ofrecido para la obtención de este trabajo de titulación.

Finalmente agradezco a Adriana Elizabeth Durán González por brindarme su apoyo y amor incondicional en todo momento durante este largo camino, agradezco cada palabra de aliento y ánimos que me sirvieron en momentos difíciles para continuar, por los momentos que me regaló durante cada ciclo. Por convertirse en un motivo para finalizar mi carrera y sobre todo que estuviera orgullosa del profesional que quería llegar ser. Agradezco a Dios por ponerte en mi camino.

Esteban Vizhñay

Dedicatoria

A mis padres, Martha y Luis, por ser mi motivación, mi fortaleza, mi guía. Éste es ***su logro***.

A mi hermana, Alexandra, por ser mi confidente, mi consejera, mi motivadora.

¡Lo Logramos!

Freddy Abad L.

Dedicatoria

Este trabajo de titulación se lo dedico a mis padres Mauro y Mariana, ya que sin ellos no lo hubiera logrado.

A mis hermanas y mis sobrinos para que puedan tenerme como ejemplo que se pueden lograr las cosas que uno se propone a pesar de todas las dificultades.

Esteban Vizhñay

Capítulo I

1. Introducción

Este capítulo ofrece una introducción que inicia presentando el contexto sobre las problemáticas globales y locales del tema central de este trabajo, para luego presentar la justificación y motivación de este trabajo de titulación. Además, se presenta el modelo matemático a implementar, algunos problemas identificados que pueden presentarse durante el desarrollo y se detallan las tareas globales a realizarse para solventarlos. De manera formal, se describe la pregunta de investigación, la cual está directamente basadas en las preguntas del proyecto principal desarrollado por el Grupo de Investigación Movilidad Activa y Sostenible de la Universidad de Cuenca (MAS).

1.1. Contexto

Como afirma Simićević, et al., (2013) la política de aparcaderos tiene un fuerte impacto no solo en la operación del subsistema de estacionamiento sino también en todo el sistema de transporte y la ciudad en general. Además, Klementschitz, et al., (2007) indica que las políticas de aparcamiento se pueden diseñar y emplear para influir en el comportamiento de movilidad en las zonas urbanas. Las respuestas del conductor a la política de aparcamiento como la tarifa y limitación de tiempo pueden variar según Scholefield, et al. (1997) como, por ejemplo: el tipo de estacionamiento, ubicación del estacionamiento, modo de transporte, etc.

A nivel global, se tienen casos prácticos de implementación de modelos matemáticos para el cálculo de tarifas de parqueadero, este es el caso de Atenas y Dublín. En la ciudad de Atenas, en Grecia, Tsamboulas, D. A. (2001) implementó un modelo de Regresión logística multinomial donde demostró que los conductores con el propósito de trabajar tienen más probabilidades de cambiar el modo de transporte o la hora del día, antes de cancelar el viaje o cambiar el destino. Permitiendo calcular la probabilidad de uso de automóviles y el comportamiento del conductor cuando se incrementa el precio del aparcamiento. En el caso de Dublín, Irlanda, Andrew Kelly, J., & Peter Clinch, J., (2006) establecieron el modelo PROBIT, donde se investigó las diferencias de las respuestas al precio del estacionamiento entre usuarios de propósito comercial y otros. Así, demostraron que a precios más bajos no hay diferencias entre categorías de usuarios. La diferencia aumenta a medida que se incrementa el precio. Recomendaron el establecimiento de un umbral de precios entre grupo de conductores y sus propósitos de movilización.

A nivel local, Moscoso (2012) explica que, dentro de la zona céntrica de la ciudad de Cuenca, la creación de aparcaderos privados ha aumentado el nivel de contaminación y congestión vehicular. Además, los datos del GAD Cuenca (2015) informan que alrededor del 46% de las plazas del Centro Histórico corresponden a usos ilegales o “no regulados”. Frente a este contexto, apoyado por la información presentada anteriormente, el cálculo de tarifas diferenciales a aparcaderos de vehículos privados toma relevancia como una forma de solucionar dichos problemas, como lo garantizan Albert, et al (2006). Estos explican en su trabajo, que la tarifa de estacionamiento es una medida que se toman para resolver los problemas de congestión vehicular presente en una zona. Por ejemplo, el aplicar un recargo adicional a la tarifa cancelada por los conductores que llegan durante las horas pico de la mañana. Esto afectará la predisposición del usuario a utilizar su vehículo privado y sobre todo el aparcadero, promoviendo el uso de medios alternativos de transporte.

1.2. Necesidad

La investigación de esta problemática y sobre todo al cálculo de tarifas para aparcaderos ha tomado relevancia a lo largo de los años, para Anastasiadou, et al (2009). Las medidas de política de aparcaderos no solo afectan el sistema operativo de este, también generan impactos en la movilidad y el sistema socioeconómico de una ciudad. Con lo anterior mencionado, se demuestra que la tarifa de aparcadero impacta más de lo que se espera en una ciudad, tanto a nivel económico como de movilidad humana. Así, Lam (2004), propone la creación de modelos de elección y comportamiento que imponga el uso de esquemas de cobro basados en parámetros de tiempo/demanda y los hábitos de los conductores.

Según Zoeter et al., (2014), el uso de un modelo matemático, en lugar de una estimación sin bases científicas provee un umbral de precios según los hábitos del conductor, manteniendo en equilibrio la ocupación de los aparcaderos. Así, se disminuye indirectamente la congestión vehicular de la zona colindante. La estimación de una tarifa debe ser supervisada, Zoeter, et al, (2014), confirman que el estacionamiento es un recurso escaso en centros urbanos, si el estacionamiento es gratuito, o a un precio bajo, se utilizan de manera ineficiente, ya que los conductores no están incentivados a evitar horas pico. En casos de tarifas muy altas, se limita la movilidad de la población. Estos son motivos para el uso de un modelo matemático que se ajuste a la disponibilidad del aparcadero, su ubicación, el criterio y comportamiento de los usuarios, entre otros.

Este trabajo lleva a cabo el modelo propuesto por Avila-Ordóñez, et. al. (2019), que asentará un plan tarifario del uso de aparcamiento en el interior de los predios del campus central de la Universidad de Cuenca. El modelo matemático actualmente se encuentra en fase teórica y en proceso de validación en casos reales, este trabajo de titulación validará dicho modelo, apoyado por datos reales facilitados por los proveedores de información relacionados (DTICS & MAS Universidad de Cuenca). Asumimos que durante el desarrollo de este trabajo abordaremos los siguientes problemas:

- **Heterogeneidad en los datos:** Las instituciones que proveen datos al modelo matemático, son autónomas, su dirección de tecnologías funciona por separado. Estos datos evidentemente presentarán diversidad de sintaxis y semántica.
- **Integración de distintas fuentes de datos:** Garantizar que los datos que use el modelo, y su retroalimentación con nuevos datos, no afecten a los datos históricos presentes en un determinado momento.
- **Garantizar el funcionamiento del modelo y sus componentes:** Incorporar las distintas etapas de desarrollo: integración, limpieza y consolidación de datos, además las implementaciones del modelo matemático representan desafíos para garantizar el funcionamiento total o parcial del sistema. El aislamiento de componentes es un tópico para solucionar.

- **Protección de los datos e información:** El entorno del sistema será online, se desplegará en Internet, así el sistema y sus datos deben asegurarse ante ataques informáticos que manipulen el correcto funcionamiento de este.

Estos problemas se resolverán mediante técnicas de limpieza, estandarización e integración de datos, además de su consolidación.

1.3. Tarea

El trabajo de titulación enfrenta los problemas a solventar, descritos en la sección 1.2. Así, se propone:

- **Abordar la heterogeneidad de los datos:** Debido a la diversa procedencia de datos, se desarrollará un proceso ETL que "limpie" y estandarice los datos a utilizar en las siguientes etapas. Adicionalmente, se crearán reportes de los cortes de datos realizados, que evidencian anomalías sintácticas y semánticas, y proveen acciones para que el modelo matemático tenga mayor precisión.
- **Integrar diversas fuentes de datos:** Al ser diversas instituciones públicas, las que provean de los datos, cada una manejando diversos formatos, se propone la implementación de un DW.
- **Aislar los componentes del sistema mediante Microservicios:** Un sistema funcional e integral garantiza el cumplimiento de propiedades emergentes, tales como la Durabilidad y el Aislamiento. Una arquitectura moderna, como MSA aborda estas propiedades correctamente, aislando cada componente por separado, así como su ejecución en conjunto. Así cada etapa del sistema se desarrolla como un Microservicio: análisis de datos y cálculo del modelo matemático. Además, funciona en conjunto con otros componentes del sistema, cumpliendo con los objetivos específicos de esta tesis.
- **Implementar estándares de seguridad que protejan los datos:** Ante la necesidad de anticipar los datos de ataques informáticos que alteren, eliminen o utilicen los datos e información para fines ajenos al tema de tesis, se pondrá en práctica el estándar OAuth2, este garantizará la comunicación segura y rápida entre componentes.

1.4. Pregunta de Investigación

¿Cómo lograr una correcta integración de los datos disponibles por el momento y habilitar la posibilidad de agregar más fuentes de datos para futuros trabajos?

1.5. Hipótesis

- El uso de una arquitectura de MSA disminuye las dependencias entre equipos de trabajo y sus etapas de desarrollo. Esto resultará, en un código de producción más rápido e independiente.
- Las técnicas de DM permitirán extraer información y conocimiento relevante de los datos que se proporcionen. Así el manejo de dichos datos será de manera eficiente y consecuentemente, la toma de decisión será más acertada.
- La implementación de las etapas de recolección y limpieza de datos facilitarán la integración de nuevas fuentes, sin afectar la consistencia histórica, retro - alimentando el modelo matemático.

1.6. Objetivos generales y específicos

Objetivo General

Diseñar e Implementar un Data Warehouse que permita integrar y analizar los datos provenientes de los ingresos y salidas del campus central, además de implementar un aplicativo web que consuma un microservicio que calcula la tarifa diferencial de acuerdo al modelo matemático propuesto por el grupo de Movilidad Activa y Sostenible de la Universidad de Cuenca.

Objetivos específicos

- Plantear e implementar una infraestructura de software para la integración y análisis de datos provenientes de las fuentes secundarias y de los aparcaderos.
- Aplicar un proceso de Extract, Transform and Load a los datos provenientes de las puertas de ingreso (12 de abril y Economía) y salida (Arquitectura y Filosofía) del campus central, correspondientes a 8 zonas de aparcaderos (Economía, Psicología, entre otros), y correlacionarse con los datos socioeconómicos del cuerpo docente, empleados y trabajadores.
- Aplicar una o varias técnicas de Machine Learning refiriendo la metodología de Data Mining, Crisp-DM, en el proceso de ejecución. Estas técnicas reconocerán variantes de las variables o descartarán las propuestas por el Grupo de Investigación sobre las dinámicas de Movilidad Humana.

- Validar el modelo matemático implementado en una infraestructura web, obteniendo un cálculo funcional en consideración de las variables propuestas por el grupo de investigación y/o las variantes producto del Data Mining.

Capítulo II

2. Marco Teórico

En este capítulo se dan a conocer los conceptos principales de los términos y técnicas fundamentales utilizados en el desarrollo de este documento. La distribución de este capítulo inicia presentando una visión general del plan tarifario para el uso del aparcamiento en el interior de los predios de la Universidad de Cuenca. Esta sección, explica el modelo matemático propuesto por Avila-Ordóñez, et al., (2019), sus factores y pesos, además de la definición de la fórmula y los casos de estudio tomados en cuenta, para el cálculo de la tarifa diferencial. Prosigue detallando sobre DI, mediante metodologías de DW, que permiten el diseño, desarrollo, despliegue y mantenimiento ordenado. Este proceso puede considerarse uno de los más extensos e importantes en el desarrollo de este trabajo de titulación, por lo cual, recomendamos cautela en su lectura. Así continúa, definiendo la Minería de Datos y Obtención de Conocimiento, presentando un mapeo de los procesos que debe pasar un problema planteado a resolver con DM. Esta sección abarca diversos sub temas, como: *Machine Learning*, pormenorizando este campo de las ciencias de la computación y como se plantea soluciones a los problemas que requieren que computadoras puedan aprender por sí solas. *Crisp-DM*, como modelo estándar abierto para los procesos de obtención de conocimiento, en base a los datos proporcionados a los modelos de ML. *Arquitecturas MSA* que detalla características de este tipo de arquitecturas modernas, que impactan directamente a los sistemas informáticos desarrollados en esta tesis. Además, detalla la importancia de su uso en la actualidad e incluye las ventajas y desventajas de su implementación, añadiendo los desafíos que presenta su desarrollo. Este subtema además analiza la Escalabilidad en una MSA, explicando las razones por las cuales en sistemas escalables se debe usar arquitecturas orientadas al escalamiento. La *Aplicación de APIs*, define las Application Programming Interface empleados para el desarrollo de sistemas web. *Diseño de Microservicios* orientado por mensajes, precisa los enfoques que puede tomar los Microservicios, las ventajas y desventajas que presenta cada enfoque, además de demostrar cómo el trabajo de abstracción en etapas de diseño afecta directamente a la reducción de tiempo de desarrollo de los sistemas. *Vulnerabilidad, pruebas y seguridad REST*, pormenoriza los servicios REST como interfaz para conectar sistemas basados en el protocolo HTTP, analizando las vulnerabilidades que enfrentan y las seguridades que se toman en su implementación. *Patrones de diseño*, define los patrones arquitectónicos aplicables en el desarrollo de software, explicando la factibilidad de su uso, además de los beneficios y desafíos que presenta integrarlos. *Desarrollo de las Aplicaciones Web*, detalla la evolución del desarrollo web hasta la actualidad, donde se mayoritariamente se desarrolla mediante componentes en SPA. Además de explicar las ventajas a nivel de usabilidad y mantenibilidad de este tipo de desarrollo en entornos web actuales que pueden desempeñarse de manera óptima para los requerimientos actuales.

2.1. Plan Tarifario para el uso del aparcamiento en el interior de los predios de la Universidad de Cuenca

El modelo matemático de Avila-Ordóñez, et. Al., (2019) proporciona un plan tarifario del uso de aparcamiento en el interior de los predios de la Universidad de Cuenca. Actualmente, la tarifa de uso es de 15$ por aparcadero para la planta docente y administrativa del campus central (V. Garate, personal communication, 2020). El modelo busca ajustar esta tarifa considerando 4 diferentes factores: *Distancia, Cautividad, Titularidad y Dedicación*. Los porcentajes para los factores definidos son una propuesta de Avila-Ordóñez, et. Al., (2019), que serán validados en la etapa final.

El **factor distancia** se refiere a la lejanía entre el lugar de residencia y el campus central de la Universidad de Cuenca (tomando como referencia los centroides de estos). El factor distancia usa la fórmula de Haversine (Shylaja, 2015) que permite determinar la distancia del arco terrestre entre dos coordenadas geográficas (longitud, latitud). La Fórmula 1 muestra la fórmula matemática de Haversine, para calcular la distancia entre dos puntos geográficos. Donde es φ es la latitud, λ es la longitud, R es el radio de la tierra (radio medio= 6371 km) y los ángulos deben estar en radianes para pasar a las funciones trigonométricas.

$$a = sin^2(\Delta\varphi/2) + cos(\varphi_1) * cos(\varphi_2) * sin(\Delta\lambda/2)$$

$$c = 2 * a * tan2(\sqrt{a}, \sqrt{(1-a)})$$

$$d = R * c$$

Fórmula 1 Haversine.

El factor de distancia permite el cálculo de la tarifa mínima dependiendo de la lejanía de la residencia del usuario con el campus donde aparca su automóvil. Según el factor de distancia, se puede calcular el factor de mayoración, el cual es equivalente al porcentaje de cercanía o lejanía entre los puntos medidos.

Según Avila-Ordóñez, et al., (2019), la tarifa mínima dependiente del factor de mayoración es equivalente a:

- 10 % si la distancias entre el lugar de residencia y el campus central son menores o iguales a 1 km.
- 5 % si la distancia entre el lugar de residencia y el campus central es mayor a 1 km, pero menor o igual a 5 km.
- 0% si la distancia entre el lugar de residencia y el campus central es mayor a 5 km.

El **factor de cautividad** se refiere a cuán "cautivo" está una persona a usar un medio de transporte dependiendo la ubicación de su residencia, considerando la oferta de transporte público a su alcance. El cálculo para el factor de cautividad, se realiza en base al trazo de radios de 400 m alrededor de las paradas de bus de toda la ciudad. Para el cálculo de la distancia entre las paradas de buses y los

lugares de residencia se aplica la fórmula de Haversine, proceso que se repite en las paradas cercanas al campus central. Si los usuarios se encuentran fuera de estos radios se consideran cautivos del vehículo privado. La función de cautividad, se ve afectada también, por el tiempo de viaje en bus y el número de transferencias realizadas en cada viaje. Para determinar los intervalos de mayoración se tomaron cinco percentiles del tiempo de viaje, el mismo que varía desde 0 hasta 55.26 minutos. El promedio del tiempo de viaje es 22.48 minutos con una desviación estándar de 17.98 minutos. Si el tiempo de viaje está en el primer percentil se mayora la tarifa mínima con un 10 %. Si el tiempo de viaje se encuentra en el segundo percentil el factor de mayoración es del 7.5 %. Si el tiempo de viaje se encuentra en el tercer percentil se afecta la tarifa mínima con un 5 %. Finalmente, si está en el cuarto percentil se afecta la tarifa mínima con un 2.5 %, en el resto de casos no grava la tarifa mínima.

El **factor dedicación** se refiere al tipo de jornada de trabajo que mantiene el docente o administrativo con la Universidad. Así, influye en la tarifa mínima dependiendo del tipo de contrato que tenga el usuario del aparcadero con la Universidad, referente a la dedicación en horas que tenga. Esta tarifa, proponen Avila-Ordóñez, et al., (2019), se grave en:

- 10 % si el docente o administrativo tiene un contrato a tiempo parcial con la Universidad de Cuenca.
- 5 % si el contrato es por medio tiempo.
- 0% si los contratos son a tiempo completo.

El **factor de titularidad** se refiere al tipo de contrato que mantiene el docente o administrativo con la universidad. Así, influye en la tarifa mínima dependiendo del tipo de contrato referente al tipo de servicio que presta. Esta tarifa, proponen Avila-Ordóñez, et. Al., (2019), se mayora en:

- 10 % para docentes o administrativos con un contrato por servicios profesionales.
- 5 % para docentes o administrativos con un contrato por servicios ocasionales.
- 0% para docentes o administrativos titulares.

Así, explicado los factores y la teoría que estos conllevan, (Avila-Ordóñez, et. al. ,2019) propone la siguiente fórmula de cálculo de la tarifa (Ver Fórmula 2):

$$Tarifa_j = TM(1 + (w_1)(p) + (w_2)(q) + (w_3)(r) + (w_4)(s))$$

Fórmula 2 Modelo de cálculo de tarifa diferencial

En donde:

- $Tarifa_j$ tarifa individual de una persona j
- w_i son los pesos dados para cada factor.
- **TM** es la tarifa mínima.
- **p, q, r, s** son, respectivamente los factores distancia, cautividad, dedicación y titularidad.

2.2. Integración de Datos

La integración de datos es el proceso que consolida datos de diversas fuentes en una/s única/s bases de datos aplicables al dominio del problema a resolver. Esto permite una visión unificada, reforzada por procedimientos como: acceso a las fuentes, extracción de los datos, limpieza, mapeo de datos, y finalmente la carga en una única fuente de datos. Esta sección es imprescindible para analizar los datos y convertirlo en información y posteriormente en conocimiento. Este proceso permite explotar los datos, en productos o servicios que mejoren la calidad de procesos de una empresa, aplicando soluciones para problemas de su dominio.

2.2.1. Data Warehouse y Data Mart

Kimball & Ross (2013), define al DW como "un almacén de datos que extrae, limpia, conforma y entrega una fuente de datos dimensional para la consulta y el análisis". Es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Permite realizar informes y análisis de datos y se considera un componente fundamental de la inteligencia empresarial. Conforma un expediente completo de una organización, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos (especialmente OLAP). Los almacenes de datos contienen a menudo grandes cantidades de información que se subdividen a veces en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario. El DW se puede definir, a su vez, como la unión de varios Data Marts, los cuales confirman Kimball & Ross (2013), son sistemas orientados a la consulta, en el que se producen procesos batch de carga de datos (altas) con una frecuencia baja y conocida. En otras palabras, un Data Mart es un DW orientado a una sola tarea o proceso del negocio. Kimball & Ross (2013), declaran, al ETL como el "proceso de mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y, finalmente cargarlos en otra base de datos, Data Mart, o DW". Esto con la finalidad de analizar estos datos, o como parte de apoyo de un proceso de negocio.

2.2.2. Metodología de Hefesto

Hefesto es descrita por Bernabeu and Mattío, (2017), como una metodología cuya propuesta está fundamentada en una extensa investigación, comparación de metodologías existentes y el aporte

de experiencias propias en procesos de diseño e implementación de DW. Es necesario mencionar que esta metodología se adapta muy bien al ciclo de vida de desarrollo de software de metodologías ágiles. Hefesto está compuesta por los siguientes pasos:

1. **Análisis de Requerimientos**
    a. Preguntas del Negocio
    b. Indicadores y Perspectivas
    c. Modelo Conceptual
2. **Análisis de Data Sources**
    a. Hechos e Indicadores
    b. Mapeo
    c. Granularidad
    d. Modelo Conceptual Ampliado
3. **Modelo lógico del Data Warehouse**
    a. Tipología
    b. Tablas de Dimensiones
    c. Tablas de Hechos
    d. Uniones
4. **Integración de datos**
    a. Carga Inicial
    b. Actualización

1. Análisis de requerimientos

En esta etapa se identifican los requerimientos de los usuarios utilizando preguntas que expliquen los objetivos de su organización. Se identifican indicadores y perspectivas que se tomarán en cuenta para construir el DW.

a. Preguntas de Negocio

Se obtiene y determina la información clave de alto nivel esencial para lograr las metas y ejecutar las estrategias de la empresa, además facilita la toma de decisiones. Debe tenerse en cuenta que la información recolectada, es la que proveerá el soporte para desarrollar los pasos sucesivos.

Un punto importante que debe tenerse en cuenta, es que la información debe estar soportada de alguna manera por la DS, ya que, de otra forma, no se podrá elaborar el DW.

b. Indicadores y Perspectivas

Posterior al planteamiento de las preguntas de negocio, se procede a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Así los indicadores, son valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, importes, promedios, etc.

En cambio, las perspectivas se refieren a las entidades mediante las cuales se quieren examinar los indicadores. Con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, etc. Cabe destacar, que el tiempo se suele considerar comúnmente como una perspectiva.

c. Modelo Conceptual

Un Modelo Conceptual es una descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de Objetos, Relaciones y Atributos. Un ejemplo es lo que se visualiza en la Figura 1. A través de este Modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos.

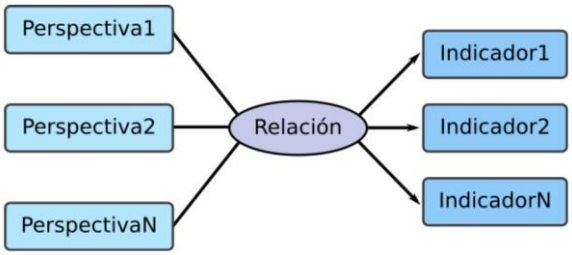


Figura 1 Modelo Conceptual. Tomado de Bernabeu and Mattío, (2017)

2. Análisis de Data Sources

Esta etapa tiene la finalidad de determinar cómo serán calculados los indicadores y establecer el mapeo entre el Modelo Conceptual creado en el paso anterior y los datos de la empresa.

a. Hechos e Indicadores

El cálculo de los indicadores, se define tomando en cuenta lo siguiente:

- Hechos que lo conforman, con su respectiva fórmula de cálculo, por ejemplo: Hechos 1 + Hechos 2.

- Función de agregación que se utilizará. Por ejemplo: Suma, Promedio, Conteo, etc.

b. Mapeo

En esta etapa se examinan los DS e identifican sus características propias, para asegurar que los DS disponibles contengan los datos requeridos. Luego, se establece cómo serán obtenidos los elementos que se definieron en el Modelo Conceptual, determinando de esta manera una correspondencia directa entre elementos del Modelo Conceptual y DS.

c. Granularidad

Teniendo como base el mapeo establecido en la etapa anterior, se debe presentar a los usuarios los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle qué significa cada campo y/o valor de los datos encontrados en los DS, por lo cual, conviene investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

Luego de exponer frente a los usuarios los datos existentes, explicando su significado, valores posibles y características, estos deben decidir cuáles son los que consideran relevantes para consultar los Indicadores y cuáles no.

d. Modelo Conceptual Ampliado

Esta etapa contempla la ampliación del Modelo Conceptual, colocando debajo de cada perspectiva los campos seleccionados y debajo de cada indicador su respectiva fórmula de cálculo. Ver Figura 2.

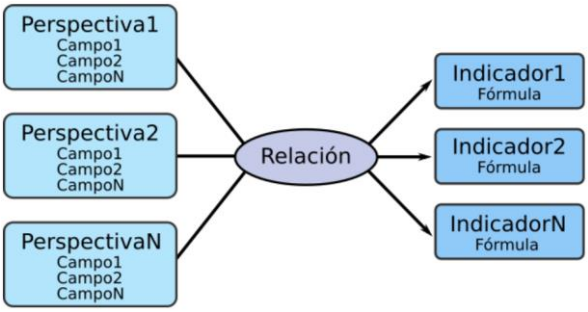


Figura 2 Modelo Conceptual Ampliado. Tomado de Bernabeu and Mattío, (2017)

3. Modelo lógico del DW

Un Modelo Lógico se define como la representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

a. Tipología

El esquema escogido para este trabajo de titulación corresponde al Modelo Estrella, sin embargo, existen otros dos esquemas que se pueden considerar:

1. Copo de Nieve
2. Constelación

La elección del tipo de esquema se realizará tomando en cuenta la que mejor se adapte a los requerimientos y necesidades de los usuarios, recordando que el modelo lógico seguirá este esquema.

b. Tablas de Dimensiones

Todas las perspectivas identificadas en los pasos anteriores y sus campos deben realizar el siguiente proceso:

1. Elegir un nombre que identifique la Tabla de Dimensión.
2. Añadir un campo que represente su clave principal.
3. Definir los nombres de los campos, si estos no son intuitivos.

Este proceso puede usarse en cualquiera de los tres tipos de esquemas mencionados. Un ejemplo del proceso de creación de dimensiones se representa en la Figura 3.

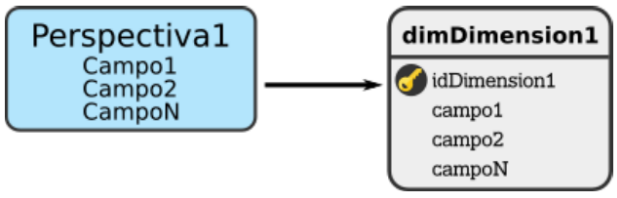


Figura 3 Ejemplo de creación de una tabla dimensión. Tomado de Bernabeu and Mattío, (2017)

c. Tablas de Hechos

El proceso de creación de Tabla de Hecho conlleva los siguientes pasos:

- Asignar un nombre a la Tabla de Hechos que represente la información que contiene, área de investigación, negocio enfocado, etc.
- Definir la clave primaria, que se compone de la combinación de las claves primarias de cada Tabla de Dimensión relacionada.
- Crear tantos campos de Hechos como Indicadores se hayan definido en el modelo conceptual y se les asignará un nombre.

Este proceso se representa en la Figura 4.

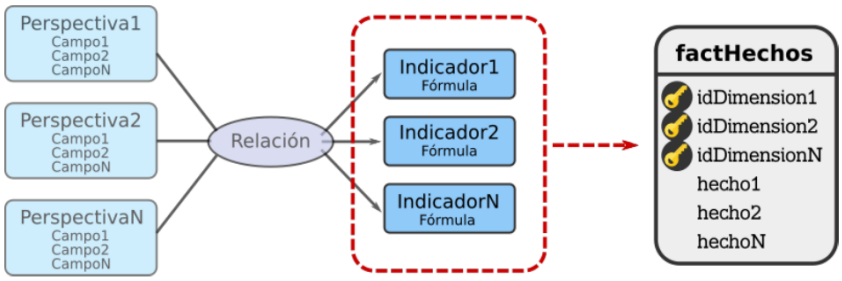


Figura 4 Creación de un hecho. Tomado de Bernabeu and Mattío, (2017)

d. Uniones

Este concepto se define como la relación correspondiente entre tablas de Dimensiones y tablas de Hechos. Ver Figura 5.

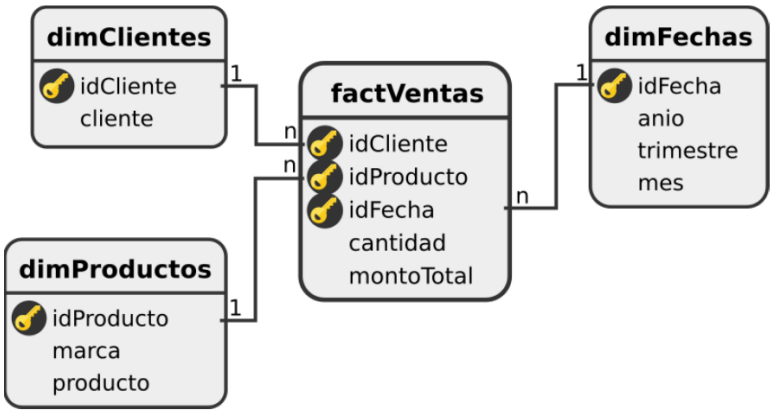


Figura 5 Ejemplo de una unión entre dimensiones y hecho. Tomado de Bernabeu and Mattío, (2017)

4. Integración de datos

El paso posterior a la creación del Modelo Lógico, es la carga de datos. Consecuentemente se definen las reglas y políticas de actualización, así como también los procesos que se llevarán a cabo.

a. Carga Inicial

En esta etapa se realiza la carga inicial del DW, en el modelo construido en los pasos anteriores. Este paso conlleva realizar ciertas tareas previas, tales como procesos ETL para asegurar la limpieza y calidad de los datos.

Uno de los principios básicos del DW es la carga correcta de datos, evitando los "Missing Values" (valores faltantes), "Outliers" (datos anómalos) o sin integridad. Estos tipos de datos conllevan a una degradación de la calidad del DW en las etapas de consultas a las bases de datos y creación de reportes. Para este fin se establecen condiciones y restricciones para asegurar que solo se utilicen los datos de interés.

Es importante fijar un orden de carga, donde las tablas dimensión serán las primeras en ser cargadas y posteriormente las tablas de hechos. Así se garantiza la integridad de datos en valores de claves foráneas y de esta manera se evita problemas de rechazo de datos por parte del SGBD.

b. Actualización

Esta etapa establece las políticas y estrategias de actualización periódica a los cubos de datos. Así se llevan a cabo las siguientes acciones:

- Determinar el proceso de limpieza de datos y calidad de datos, definir los procesos ETL, etc., que deberán realizarse para actualizar los datos del DW.
- Especificar de forma general y detallada las acciones que deberá realizar cada herramienta de software definidos en la Sección de Materiales Y Métodos.

2.3. Minería de Datos y Obtención de Conocimiento

Debido a la gran cantidad de datos recolectados a lo largo de los años en sistemas de información y los generados día a día, el humano carece de la capacidad de extraer conocimiento sin depender de herramientas computacionales. El resultado de esto, como indica Han (2011), es que en muchas ocasiones estos datos son rara vez visitados al momento de tomar decisiones de negocio. Haciendo que las decisiones importantes no se tomen en base a la historia que cuentan los repositorios de datos, sino más bien en la intuición, debido a la falta de herramientas para extraer el conocimiento incorporado en el gran volumen de datos existente (Shastri, Mansotra, 2019). Es por esto que Han

(2011) llama al desarrollo de herramientas que permitan convertir estas "tumbas" en "minas de oro" del conocimiento.

Según Witten (2015), la minería de datos es el proceso automático de descubrir patrones en los datos. Los patrones útiles permiten hacer predicciones no triviales sobre nuevos datos. Si bien los términos KDD y DM están entrelazados como se puede notar en el concepto descrito anteriormente. Para Bramer, (2016) en su libro, KDD es la "extracción no trivial de información implícita, previamente desconocida y potencialmente útil de los datos". Llevando al descubrimiento del conocimiento a un papel central, el cual está tan entrelazado que muchos autores tratan al DM como un sinónimo de KDD (Han, 2011). Debido a su cercanía en términos conceptuales, se puede tomar como referencia el proceso de KDD para la realización de procesos de minería de datos. El proceso de KDD como lo describe Han (2011) consta de 7 pasos:

1. Limpieza de datos
2. Integración de datos
3. Selección de datos
4. Transformación de datos
5. Minería de datos
6. Evaluación de patrones
7. Presentación del conocimiento

2.3.1. Machine Learning

El ML, como define Marques (2018), es un campo de las ciencias de la computación que se encarga de "aprender" dado un conjunto de datos. Este campo representa la estructura y comportamientos de los datos estudiados en un dominio de problema. Así, aprende sin "memorizar" sino creando modelos matemáticos a partir de los datos históricos que ayudan a concluir sobre ejemplares no entrenados. Esta disciplina forma parte de la IA, donde se trata de imitar el funcionamiento de redes neuronales cerebrales, mediante procesos computacionales.

Como determina Bramer (2016), se definen cuatro técnicas de DM: Clasificación, Clustering, Predicción, Reglas de Asociación. A continuación, se define las técnicas Clasificación y Clustering, implementadas en el capítulo 4 - Materiales y Métodos.

Clasificación

La técnica de clasificación, busca definir los valores de un determinado atributo, conocido en un conjunto de datos de prueba, en base a los valores de otros atributos. El objetivo de esta técnica, es

inducir un modelo para poder predecir una clase dados los valores de los atributos (Sucar, 2015). Saquicela (2015) define dos etapas en esta técnica:

- **Construcción del modelo:** Engloba las fases de entrenamiento y validación que nos permiten conocer la fiabilidad del mismo.
- **Explotación de los modelos extraídos:** Estos modelos son aplicados para la estimación de los valores para la etiqueta (o atributo a estimar) para los casos en los cuales no se conocen.

Uno de los algoritmos utilizados para las técnicas de clasificación es K-NN. El algoritmo K-NN es uno de los algoritmos de clasificación más conocidos y un ejemplo de aprendizaje supervisado. Este algoritmo utiliza un conjunto de ejemplos ya clasificados denominados conjuntos de entrenamiento o training para clasificar los nuevos ejemplos. No se crea un nuevo modelo, sino que el modelo es el propio conjunto de training (Berástegui Arbeloa, 2018).

El algoritmo clasifica cada nuevo ejemplo calculando la distancia de ese ejemplo con todos los conjuntos de training. El principal problema con este algoritmo es encontrar el valor de "k" con el cual obtener el mayor rendimiento al clasificar, es importante el uso de técnicas de validación para obtener el "k" apropiado como puede ser Cross Validation.

Clustering

La técnica de clustering, es un proceso de agrupación de datos de entrada en grupos de similares características. Esta técnica, no supervisada, no necesita ningún atributo que caracteriza la clase de equivalencia a la que pertenece cada una de las instancias entrantes, puesto que su salida es definir este atributo. Existen diversos métodos de clustering (Saquicela, 2015):

- **Métodos de particionado:** Consiste en la diferenciación de N grupos de objetos de forma que cada grupo contenga al menos un elemento y cada elemento pertenezca a uno de estos grupos.
- **Métodos Jerárquicos:** Implementa la definición de jerarquías sobre los objetos que representan las relaciones de semejanza.
- **Métodos basados en densidades:** Está basado en el crecimiento de los clusters siempre que satisfagan unos criterios o densidad.
- **Métodos basados en Celdas:** Comprende en la división de los valores en un número finito de intervalos por dimensión de forma que el espacio total queda dividido en parcelas.
- **Métodos de modelado:** Consiste en ajustar los datos que se disponen a determinados modelos o funciones matemáticas usando mecanismos estadísticos-numéricos para realizar dicho ajuste.

Uno de los algoritmos más usados para clustering es *K-Means*. Saquicela (2015) lo describe como: "Proceso interactivo de refinamiento de un número de N de clusters, definidos a priori". Mantiene el siguiente ciclo de ejecución:

1. El algoritmo comienza con la selección aleatoria de N posiciones aleatorias del espacio de datos.
2. Para cada uno de los elementos de los datos de entrada se calcula el punto más próximo a N elegidos, esta determina el cluster al que pertenece cada dato.
3. Se re-calcula para cada cluster del punto medio. Y se repite el paso 2.

Este algoritmo concluye en cuanto los clusters son estables y no se han modificado los puntos medios a lo largo de dos interacciones. Además, es necesario mencionar que su aplicación a datos de gran tamaño es muy limitada, además en los casos de datos categóricos la identificación del punto medio requiere una interacción extra de todos los datos al finalizar cada prueba.

2.3.2. Técnicas de validación de modelos de IA

Los modelos de IA, ya sea de Aprendizaje Supervisado o No Supervisado, deben ser validados, esto refiere a la necesidad de medir su efectividad. Las técnicas de validación, dependen de las técnicas de IA aplicadas.

2.3.2.1. Elbow Curve

La curva de codo o "Elbow Curve" es una heurística para determinar el número de conglomerados en un conjunto de datos ("Elbow method (clustering)," n.d.). Este método permite decidir en base a resultados gráficos que refieren a la variación explicada en función del número de conglomerados. Así, el número de conglomerados será aquel que forme un codo en este gráfico. El ejemplo de la Figura 6, se muestra como en n=4 clusters es óptimo a comparación de otras opciones de agrupamiento (Bonaros, 2020).

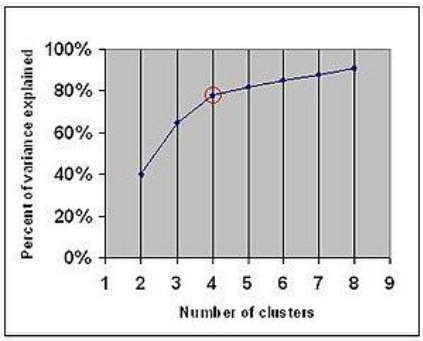


Figura 6 Elbow curve, para el cual en n=4 será el número de clusters óptimo. Tomado de (Bonaros, 2020)

2.3.2.2. Silhouette Curve

La curva de Silueta, refiere al resultado de la técnica de Silueta, el cual es un método de interpretación y validación de la coherencia dentro de grupos de datos (Ramirez, 2018). El valor de silueta es una medida de qué tan similar es un objeto a su propio grupo o cluster en comparación con otros grupos (separación). La silueta "varía de -1 a +1, donde un valor alto indica que el objeto se corresponde bien con su propio grupo y no con los grupos vecinos. Si la mayoría de los objetos tienen un valor alto, entonces la configuración de agrupamiento es apropiada. Si muchos puntos tienen un valor bajo o negativo, entonces la configuración de la agrupación en clústeres puede tener demasiados o muy pocos clústeres" ("Silhouette (clustering)," n.d.). La silueta se puede calcular con métricas de distancia, sea esta la distancia euclidiana o la distancia de Manhattan (scikit-learn, 2020).

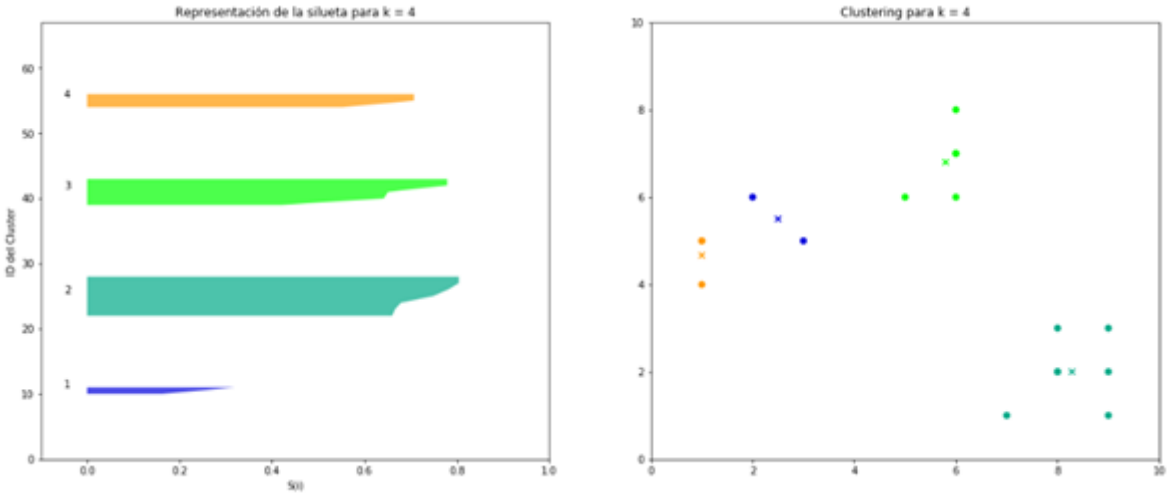


Figura 7 Gráfico de curva de silueta. Tomado de (Ramirez, 2018)

2.3.2.3. Matriz de Confusión

La matriz de confusión, de error o de coincidencia, es una tabla que permite la visualización del rendimiento de un algoritmo de aprendizaje supervisado o no supervisado. Las filas de la matriz representan las instancias en una clase predicha, mientras que cada columna representa las instancias en una clase real ("Confuxion matrix" n.d.). Así, esta tabla mide el rendimiento para un problema de clasificación de aprendizaje automático donde la salida puede ser de dos o más clases (Narkhede, 2019). La Figura 8.a y 8.b, visualiza un ejemplo de matriz de confusión, donde la diagonal serán los valores correctamente clasificados (TP = True Positive), mientras que la diagonal superior representa los Falsos Positivos (FP = False Positive, valores predichos falsamente como positivos). Mientras que la diagonal inferior representa los Falsos Negativos (FN = False Negative, o valores predichos falsamente como negativos).

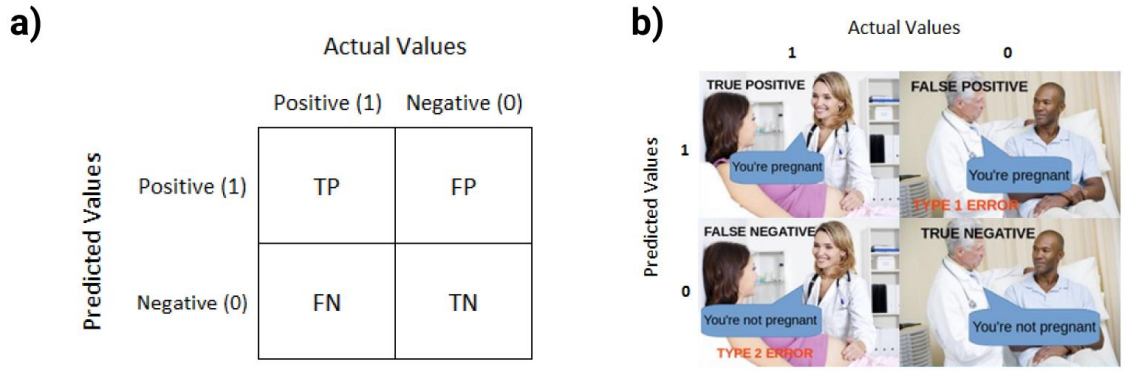


Figura 8 Matriz de Confusión o coincidencia. Tomado de (Narkhede, 2019)

2.3.3. Crisp-DM

CRISP-DM es una metodología o guía de referencia de minería de datos para desarrollo de proyectos analíticos. Esta se considera como un proceso jerárquico, que tiene cuatro niveles de abstracción: Fases, tareas generales, tareas específicas e instancias de proceso. (Enriquez, 2016) (Ver Figura 9). Abarcando desde lo más general, hasta los casos más específicos de un proyecto de DM. Teniendo una ventaja metodológica, al desarrollar de manera organizada un proyecto de esta naturaleza.

Esta secuencia de pasos, si bien es ordenada, no es estrictamente rígida, ajustándose a las necesidades del proyecto en proceso. Así, cada fase se estructura en varias tareas generales de segundo nivel. Estas tareas generales se proyectan a tareas específicas, las cuales describen las acciones que se desarrollan para situaciones particulares. El proceso indicado en la Figura 9, se estructura de la siguiente manera (Romero, 2020):

1. *Primer nivel - Fases:* Este nivel conceptualiza las fases y tareas que se desarrollaran total, parcialmente o no se desarrollaran en los siguientes niveles.
2. *Según nivel - Tareas generales:* Este nivel, conceptualiza las tareas generales que se detallarán en los siguientes niveles.
3. *Tercer nivel - Tareas específicas:* Este nivel detalla a las tareas especializadas describiéndolas como las acciones de las tareas generales
4. *Cuarto nivel - Instancias de procesos:* Este nivel integra un conjunto de acciones, decisiones y resultados sobre el proceso de DM en curso.

Estos niveles y sus correspondientes tareas, pueden realizarse en orden diferente, iterando entre sí, las veces que el problema a solventarlo exija.

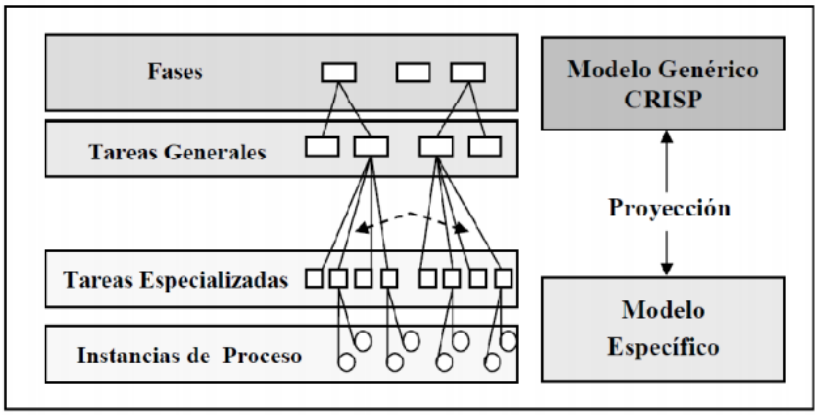


Figura 9 Esquema de niveles de CRISP-DM. Tomado de (CRISP-DM, 2000)

El ciclo de vida de un proyecto que implementa la metodología CRISP DM, se describe en la Figura 10. Este ciclo de vida sigue 6 etapas iterativas y consecutivas, permitiendo regresar, o adelantar entre fases (Enriquez, 2016).

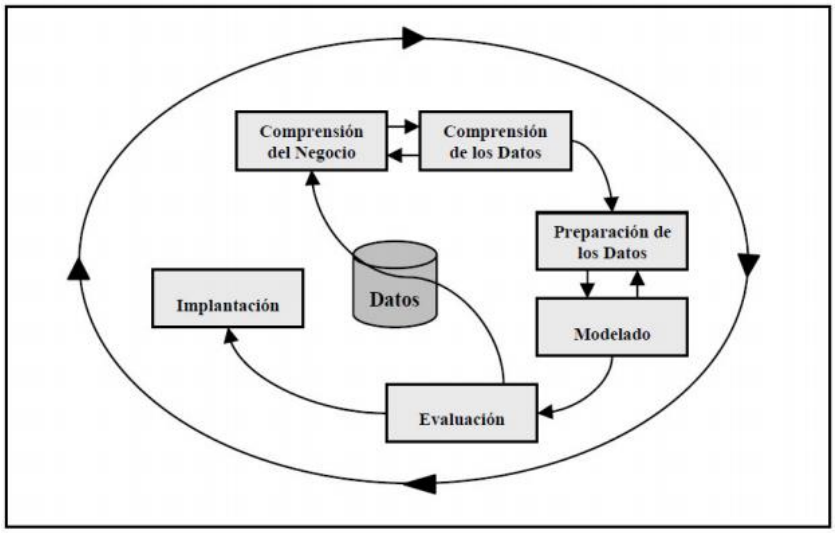


Figura 10 Ciclo de vida de CRISP–DM. Tomado de (CRISP-DM, 2000)

Entre las tareas generales de esta metodología, se encuentran las siguientes fases, cada una con distintas subfases (Enriquez, 2016).

1. ***Fase de comprensión del negocio o problema:*** Comprende el análisis de los objetivos y requisitos desde una perspectiva comercial o institucional. Es considerada como la fase más importante, ya que convierte los requisitos en objetivos técnicos y en un plan de proyecto. Es primordial el entendimiento más completo del problema que se desea resolver, así, se recolectarán los datos correctos e interpretarán correctamente los resultados (Romero, 2020). Esta fase comprende distintas subfases como: Determinar objetivos del negocio, Valorar la situación, Determinar objetivos de DM y Producir un plan de proyecto (Saquicela, 2015).
2. ***Fase de comprensión de los datos:*** Implica la comprensión, familiarización y obtención de los datos, así, identifica problemas de la calidad de estos, descubre relaciones evidentes que

permitan definir las primeras hipótesis. Esta fase y las dos siguientes demandan el mayor esfuerzo y tiempo en un proyecto de DM (Romero, 2020). Esta fase comprende distintas subfases como: Recoger los datos de inicio, Describir los datos, Explorar los datos, Verificar la calidad de los datos (Saquicela, 2015).

3. ***Fase de preparación de los datos:*** Cubre todas las actividades para construir el conjunto de datos final en el cual se aplicarán los modelos. Esta fase implica su preparación para adaptarlos a las técnicas de DM que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos. La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato (Romero, 2020). Esta fase comprende distintas subfases como: Seleccionar los datos, Limpiar los datos, Construir los datos, Integrar los datos, Formatear los datos (Saquicela, 2015).
4. ***Fase de modelado:*** Implica la selección de las técnicas de modelado apropiadas para el modelo y la calibración a valores óptimos de los parámetros de las técnicas seleccionadas. Las técnicas a utilizar en esta fase se eligen en función de ser apropiada al problema, disponer de datos adecuados, cumplir los requisitos del problema, tener el tiempo adecuado para obtener un modelo, dominio de la técnica (Romero, 2020). En esta etapa, se define si los datos están en condiciones óptimas o de ser necesario regresar a la fase preparación de los datos. Esta fase comprende distintas subfases como: Seleccionar técnica de modelado, Generar diseño del test, Construir el modelado, Valorar el modelo (Saquicela, 2015).
5. ***Fase de evaluación:*** Esta fase evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Esta evaluación considera el rendimiento del modelo/s y la integridad de todos los pasos, verificando que se han incluido todos los objetivos del negocio o investigación. Además de medir la fiabilidad que para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Esta etapa, considera múltiples herramientas para la interpretación de los resultados, por ejemplo, las matrices de confusión, empleadas concurrente en problemas de clasificación, y consisten en una tabla que indica cuántas clasificaciones se han hecho para cada tipo, la diagonal de la tabla representa las clasificaciones correctas, y los valores fuera de la diagonal, las clasificaciones incorrectas.

   Evaluado el modelo en relación a estas técnicas, se analiza las deficiencias del modelo, y si es factible pasar al siguiente paso, o caso contrario volver a la fase de modelamiento. Si el modelo generado es válido en función de los criterios de éxito establecidos en la fase anterior, se procede a la explotación del modelo en relación a los objetivos del negocio (Romero, 2020). Esta fase comprende distintas subfases como: Evaluar los resultados, Proceso de revisión y Determinar pasos siguientes (Saquicela, 2015).

6. ***Fase de implementación:*** Obtenido el modelo validado, se debe transformar el conocimiento obtenido en acciones dentro del proceso de negocio. Esto se realiza mediante recomendaciones que el analista emite basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso, por ejemplo, en análisis de riesgo crediticio. Adicional, se debe tener en cuenta que el último paso no es la implementación de un modelo, pues se deben documentar y presentar los resultados de manera comprensible para el usuario, para lograr un incremento del conocimiento. Además, de asegurar el mantenimiento de la aplicación y la posible difusión de los resultados. Esta fase comprende distintas subfases como: Implantación del plan, Monitorización del plan y mantenimiento, Producir informe final y Revisar el proyecto (Saquicela, 2015).

2.4. Microservicios

El desarrollo de sistemas informáticos ha tenido una constante evolución, partiendo de las arquitecturas monolíticas hasta las modernas MSA. La presente subsección analiza ampliamente las ventajas y desventajas de usar una arquitectura basada en microservicios, además de los desafíos que presenta desarrollar este tipo de arquitecturas. Posteriormente, analiza su integración con conceptos de APIs, para la construcción de sistemas heterogéneos y escalables. Finalmente se define los enfoques por los cuales se puede optar en la etapa de diseño de microservicios, dependiendo de los datos que se trataran y del dominio del proyecto a ejecutar.

2.4.1. Arquitectura de Microservicios

Según las normas IEEE 42010 (2011), la arquitectura de software se define como: "Conceptos o propiedades fundamentales de un sistema en su entorno encarnado en sus elementos, relaciones y en los principios de su diseño y evolución", de esta manera se constituye como una parte fundamental de un sistema informático. La importancia de estas tiene un profundo efecto en la calidad de lo que se construye sobre sí, promoviendo el desarrollo exitoso y eventual mantenimiento del sistema.

El enfoque del tema de tesis, que se plantea, permite utilizar arquitecturas tipo SOA. En el estudio de Newman, 2015 se define a SOA como un enfoque de diseño en el que múltiples servicios colaboran para proporcionar un conjunto final de capacidades, de esta manera promover la reutilización del software entre dos o más aplicaciones de usuario final, por ejemplo, usando los mismos servicios.

La arquitectura moderna de SOA utilizada es MSA. Un MSA es una interpretación moderna de la arquitectura orientada a servicios usada para construir sistemas distribuidos (Dragoni, N. et al., 2016). Para Nadareishvili, (2015) un microservicio es un componente de alcance limitado que se puede implementar de forma independiente y que admite la interoperabilidad mediante de la comunicación

basada en mensajes. Por otro lado, el estudio de Ingeno, (2018) propone a MSA como una arquitectura de aplicaciones de software que utilizan servicios pequeños, autónomos e independientes. Estos servicios utilizan interfaces bien definidas y se comunican entre sí a través de protocolos estándar y livianos. La interacción con un microservicio se realiza con una interfaz bien definida. Un sistema está conformado normalmente, con varios microservicios. Cada microservicio debe ser una caja negra para los consumidores, ocultando su implementación y complejidad. Cada microservicio se enfoca en hacer una tarea bien y puede trabajar junto con otros microservicios para realizar otras más complejas.

Algunos beneficios y desafíos presentes en una arquitectura de microservicios son (Bryant, 2015):

- **Beneficios**
    - Disminuye las dependencias entre equipos, lo que resulta en un código de producción más rápido.
    - Permite que muchas iniciativas se ejecuten en paralelo.
    - Soporta múltiples tecnologías/idiomas/frameworks.
    - Permite una degradación elegante del servicio.
    - Promueve la facilidad de innovación a través del "código desechable": es fácil fallar y seguir adelante.
- **Desafíos**
    - Mantener múltiples entornos de ensayo en varios equipos y servicios es difícil.
    - Definir la propiedad de los servicios es difícil.
    - La implementación debe automatizarse.
    - Las API ligeras deben definirse.
    - Es un reto cumplir las normas y brindar a los ingenieros plena autonomía en la producción.
    - La gestión de las entradas y salidas requiere esfuerzo.
    - Los informes a través de varias bases de datos de servicio son difíciles.

Así, los distintos servicios de una MSA se comunican con otros empleando una red para conseguir el objetivo final. Estos servicios pueden usar protocolos simples (Pisani, M. et al., 2016).

2.4.2. Escalabilidad en una MSA

Un enfoque de desarrollo, implementado mediante una arquitectura MSA, otorga múltiples ventajas en el desarrollo de sistemas grandes, Nadareishvili et al. (2016). Entre estas ventajas se encuentra la escalabilidad.

Una arquitectura MSA es óptima para sistemas escalables debido a:

- Aislamiento de fallas

- Independencia de almacenamiento de datos
- Enfoque a la granularidad de servicios, un servicio por cada proceso
- Interfaces bien definidas
- Comunicación liviana en los protocolos

2.4.3. Aplicación de APIs

Al considerar los límites de una arquitectura MSA, pasa a segundo plano el código fuente a escribir, considerándose importante, cómo se comunican los distintos componentes de un sistema, y más importante aún, qué tan rápido lo hacen. Una arquitectura MSA puede implementarse con arquitecturas sencillas y rudimentarias como SOAP con formatos XML, así como una arquitectura REST con JSON (Nadareishvili et al., 2016). Teniendo como objetivo el alto nivel de separación, independencia y modularidad, el acoplamiento de los componentes es importante. La flexibilidad que brinda la arquitectura MSA, permite realizar aplicaciones web y móviles con back-ends robustos, así como sencillos (Subramanian & Raj, 2019).

2.4.4. Diseño de Microservicios orientado por mensajes

El diseño de un MSA puede realizarse mediante dos enfoques:

- ***Orientado por mensajes:*** Este enfoque trabaja para escribir código de componentes que se pueda refactorizar de manera segura con el tiempo, refiriéndose a las interfaces compartidas entre componentes. “La noción de mensajería como una forma de compartir información entre componentes se remonta a las ideas iniciales sobre cómo funciona la programación orientada a objetos. Un caso de implementación es cuando los desarrolladores pueden exponer los puntos de entrada generales en un componente (por ejemplo, una dirección IP y un número de puerto) y recibir mensajes específicos de tareas al mismo tiempo. Esto permite cambios en el contenido del mensaje como una forma de refactorizar componentes de forma segura a lo largo del tiempo. El enfoque orientado por mensajes es ampliamente usado en la actualidad, incluyendo a estos las APIs y web services”. (Nadareishvili et al., 2016)
- ***Orientado por hipermedia:*** Este enfoque parte del enfoque orientado por mensajes, agregando complejidad y funcionalidad. Esencialmente, conceptualiza a los “mensajes pasados entre componentes, siendo estos más que solo datos, conteniendo descripciones de posibles acciones (por ejemplo, enlaces y formularios). Estos se acoplan libremente a datos y acciones. Por ejemplo, las API Gateway de API y App-Stream de Amazon admiten respuestas en el formato HAL. Las API de estilo

hipermedia adoptan la capacidad de evolución y el acoplamiento flexible como los valores centrales del estilo de diseño. También puede conocer este estilo como API con Hypermedia como el motor del estado de la aplicación (API HATEOAS)" (Nadareishvili et al., 2016).

El caso de estudio de este trabajo, se enfoca en el uso de datos planos. Dado esto, un enfoque orientado por mensajes, en comparativa con uno orientado por hipermedia es el correcto. Este, según el trabajo de Nadareishvili et al., (2016), opera con formatos de mensajes tales como Avro, Protobuf o Thrift a través de TCP/IP, así como JSON a través de HTTP. Al adoptar un enfoque orientado a mensajes, se pueden exponer los puntos de entrada generales en un componente y recibir mensajes específicos de tareas al mismo tiempo. Esto permite cambios en el contenido del mensaje como una forma de refactorizar componentes de manera segura a lo largo del tiempo.

2.5. Patrones de Diseño

Los patrones de diseño, Erich Gamma, et al., (2008) son "descripciones de clases y objetos relacionados que están particularizados para resolver un problema de diseño general en un determinado contexto". A continuación, una síntesis de algunos patrones representativos.

2.5.1. Patrones Estructurales

Los patrones estructurales se preocupan de cómo se combinan las clases y los objetos para formar estructuras más grandes (Erich Gamma, et al., 2008).

2.5.1.1. Patrón Facade

El Patrón Facade proporciona una interfaz unificada para un conjunto de interfaces de un subsistema. Define una interfaz de alto nivel que hace que el subsistema sea más fácil de usar.

Aplicabilidad

Se usará el patrón Facade cuando:

- Se pretende proporcionar una interfaz simple para un subsistema complejo. Debido a que los subsistemas suelen volverse más complicados a medida que van evolucionando. Para este caso, Facade proporciona una vista simple del subsistema que resulta adecuada para la mayoría de clientes. Si se busca más personalización se debería ir más allá de Facade.

- Cuando existen dependencias entre los clientes y las clases que implementan una abstracción. Facade permite desacoplar el subsistema de sus clientes y de otros subsistemas, promoviendo así la independencia entre subsistemas y la portabilidad.
- Se busca dividir en capas nuestros subsistemas. Donde Facade proporciona un punto de entrada en cada nivel del subsistema.

En la Figura 11 el patrón Facade en la práctica.

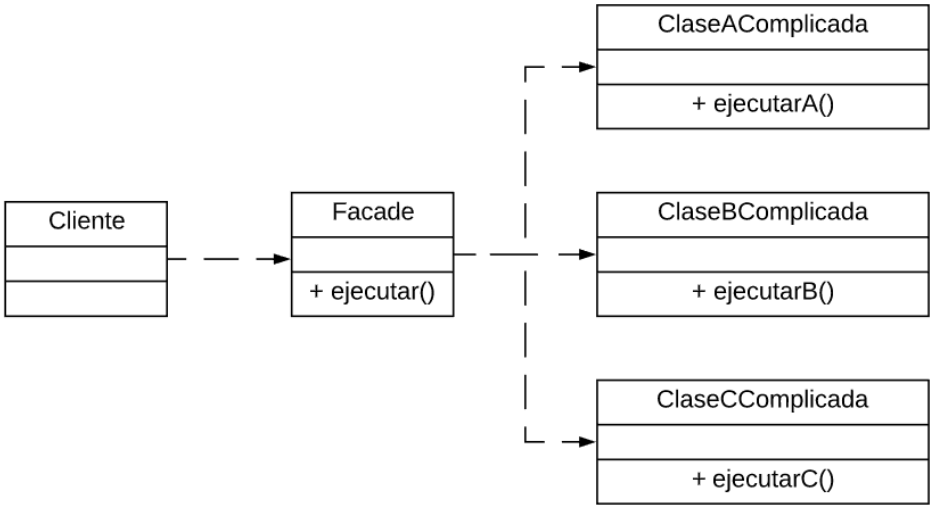


Figura 11 Patrón Facade. Tomado de (Erich Gamma et.al, 2008)

Ventajas de uso

Facade nos proporciona las siguientes ventajas:

- Ocultar a los clientes los componentes del subsistema, logrando reducir los objetos con los que tratan los clientes y permitiendo que el subsistema sea más fácil de usar.
- Promueve un débil acoplamiento entre el subsistema y sus clientes. Un acoplamiento débil nos permite modificar los componentes de un subsistema sin que sus clientes se vean afectados.
- No impide que las aplicaciones usen las clases del subsistema en caso de que sea necesario. Permitiendo facilidad de uso y generalidad.

2.6. Aplicaciones Web

Las aplicaciones web son sistemas informáticos que se almacenan en un servidor remoto y acceden a través de una red (usualmente en Internet) usando una interfaz de navegador. (Rouse, 2019) Las aplicaciones web pueden diseñarse para diversos usos y campos de acción, estos se encuentran disponibles en todo lugar que disponga de acceso a la red de datos del servidor que lo aloja.

El funcionamiento de una aplicación web implica diversos servidores: web, aplicación y base de datos. Así, los servidores web administran las solicitudes que provienen de un cliente, mientras que

el servidor de aplicaciones completa la tarea solicitada. Por lo general, el servidor de aplicaciones se complementa con un servidor de base de datos para almacenar cualquier información necesaria en el aplicativo web (Rouse, 2019).

Una de las grandes ventajas de las aplicaciones web, es su ciclo de desarrollo corto, desarrollados en dos frentes, FrontEnd y BackEnd.

El front-end, o desarrollo del lado del cliente, son "todas las tecnologías de diseño y desarrollo web que corren en el navegador" (¿"What is front-end development?," 2009) e implica la conversión de datos obtenidos desde el lado del servidor, o back-end, por lo tanto, es una interfaz gráfica para la interacción con el usuario (Nicole, 2017). Su desarrollo implica HTML5, CSS y JS, complementados con frameworks web como Angular, React, Vue, entre otros (Dhaduk, 2020).

El back-end, o desarrollo del lado del servidor, es la "capa de acceso a datos de un software, que no es directamente accesible por los usuarios, ésta contiene la lógica de la aplicación que gestiona dichos datos" (Nicole, 2017) accedidos mediante un servidor de base de datos, un gestor de archivos o a través de servicios web (Ver Figura 12). El desarrollo de esta capa utiliza diversos lenguajes de programación complementado con frameworks compatibles al lenguaje que se use. Algunos ejemplos de lenguajes usados para el back-end, son Django, Flask para Python, Spring Boot para Java o NodeJS para Javascript.

El back-end y front-end interactúan recurrentemente, como se define en la Figura 12, mediante un protocolo de comunicación, generalmente HTTP, ampliamente usado como protocolo de comunicación para las transferencias de información en la WWW (W3C, 2014).

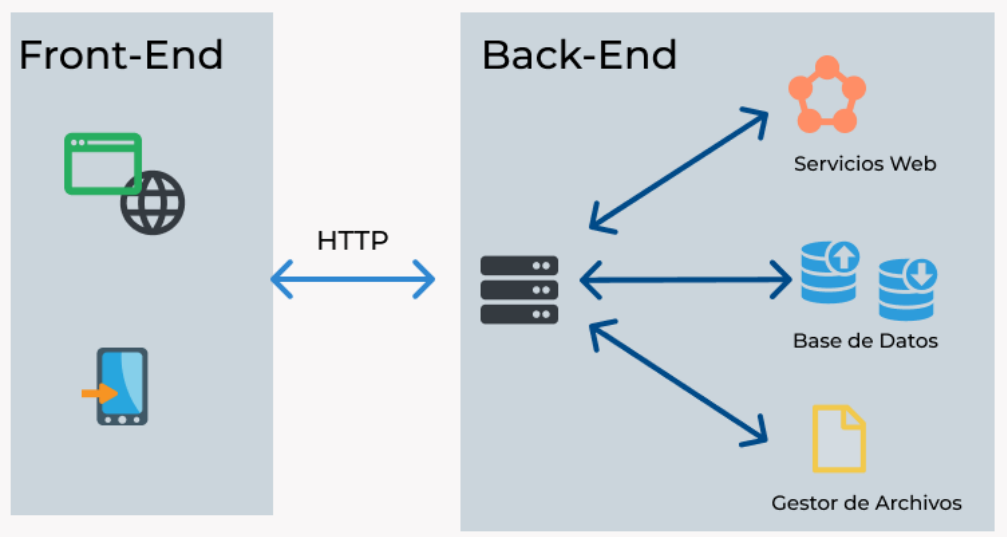


Figura 12 Arquitectura Back-End & FrontEnd

2.6.1. Single-page application (SPA)

Las SPA son un tipo de aplicación web moderna, con la característica de caber en una sola página con el propósito de dar una experiencia más fluida a los usuarios, creando un símil con una aplicación de escritorio (Arranz, 2015). La característica predominante de SPA, es su carga única de los códigos de HTML, JS, y CSS, permitiendo desplegar dinámicamente los elementos cuando se requiera en la aplicación web, como respuesta a las acciones del usuario, en tiempos reducidos (Flanagan, 2006). Este tipo de aplicación web, en conjunto con tecnologías permiten la navegabilidad en páginas lógicas dentro de la aplicación (Santamaria, 2015). La interacción con las aplicaciones de página única puede involucrar comunicaciones dinámicas con el servidor web que está detrás (Arranz, 2010). La Figura 13 define la arquitectura que una SPA usa, por lo general, para consumir servicios del lado del servidor (Arranz, 2015). Así, una SPA, se define con componentes, que interactúan directamente con el usuario. Estos componentes hacen peticiones de servicios, los cuales se definen mediante modelos. Finalmente, los servicios interactúan directamente con el backend, mediante algún protocolo de comunicación, usualmente HTTP (Arranz, 2010).

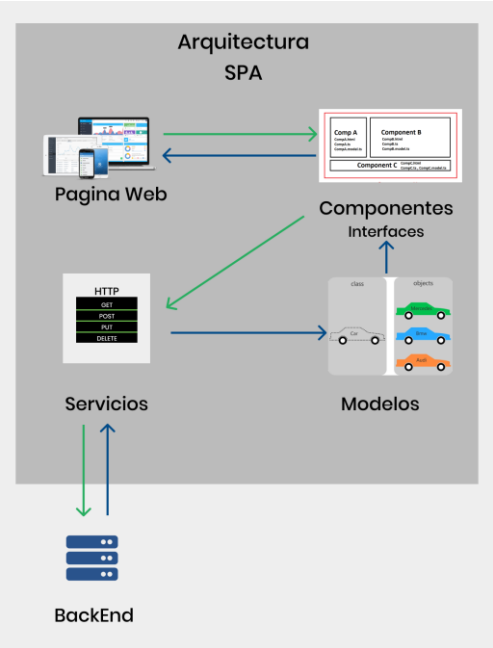


Figura 13 Single-Page Application

Capítulo III

3. Marco Histórico

Este capítulo detalla el marco histórico de los temas principales de este trabajo de titulación. Inicia explicando a breves rasgos la importancia de un modelo matemático que tome en cuenta múltiples factores de movilidad humana. La influencia del precio de estacionamiento en el comportamiento del usuario del garaje de estacionamiento. Además de la determinación de la tarifa de estacionamiento mediante metodologías de evaluación contingentes. Continúa analizando el impacto de los vehículos motorizados privados y su influencia en el problema de transporte público en el centro histórico: el caso de Cuenca-Ecuador. Además, examina un plan de movilidad y el uso de espacios públicos por los sistemas de aparcamiento. Prosigue pormenorizando el uso de nuevos algoritmos para la gestión de la demanda de estacionamiento y un despliegue a escala de ciudad. Finalmente, detalla el uso de KDD aplicada a la ingeniería de transporte.

Este marco histórico detalla investigaciones recientes, aplicables para problemas de la actualidad, ya sea del desarrollo de software, o de temas de movilidad humana. Las investigaciones consultadas fueron realizadas en un rango desde 1992 hasta la actualidad, para temas de movilidad; y para temas de desarrollo de software fueron consultadas aquellas publicadas en un rango desde 2012 hasta la actualidad.

3.1. Tarifas de aparcaderos

Esta sección resalta la importancia y efectos de utilizar una política de cobro para las tarifas de aparcaderos, además de su impacto e influencia sobre los usuarios y en algunos casos el área donde reside el aparcadero.

3.1.1. El efecto de los cargos de estacionamiento y el límite de tiempo para el uso del automóvil y el comportamiento de estacionamiento

La política de estacionamiento tiene un fuerte impacto no solo en la operación del subsistema de estacionamiento sino también en todo el sistema de transporte y la ciudad en general (Simićević, et al., 2013). Por ejemplo, los estudios han demostrado que el factor más importante para reducir el uso del automóvil es el precio del estacionamiento (Higgins, 1992). Además, la tarifa de estacionamiento se considera la segunda mejor medida para resolver la congestión del tráfico después de la carga de congestión (Albert, et al., 2006), pero se usa con mucha más frecuencia debido a su implementación relativamente simple (Marsden, 2006; Verhoef et al., 1995). Aunque una buena política de

estacionamiento tiene muchas implicaciones positivas para el transporte sostenible, una mala política de estacionamiento puede tener el efecto contrario (Simićević, et al., 2013). Recientemente, se puede observar una creciente preocupación respecto a la política de estacionamiento podría afectar negativamente la competitividad y la eficiencia comercial en un área (D'Acierno, et al., 2006). Para establecer adecuadamente la política de estacionamiento y definir las medidas apropiadas, es decir, garantizar que se cumplan los objetivos sin un impacto adverso en el sistema de transporte y otros sistemas de una ciudad, se deben predecir los efectos de dicha política (Simićević, Vukanović & Milosavljević, 2013).

3.1.2. Influencia del precio de estacionamiento en el comportamiento del usuario del garaje de estacionamiento

El nivel de movilidad y motorización registra un aumento continuo en casi todos los países del mundo, lo que ha llevado al estado en el que la demanda de tráfico excede la capacidad de la carretera (Simicevic, et al., 2012). Esto anteriormente se resolvía expandiendo la capacidad, pero en la actualidad por costos y problemas de desarrollo respecto a recursos limitados, tal solución no es factible o solo es posible en pequeña medida. Como nos indica Simićević, et al, (2012) la tarifa de estacionamiento se ha convertido en una de las políticas más poderosas de gestión de la demanda de tráfico. Esto apoyado por la idea de (Vuchic, V, 1999) que menciona que los usuarios son particularmente sensibles a los costos directos de cada viaje, debido a la inclusión del cargo por estacionamiento.

3.1.3. Determinación de la tarifa de estacionamiento utilizando la metodología de valoración contingente

Hoy en día, la propiedad y el estacionamiento de automóviles han recibido mucha atención, ya que puede influir seriamente en la vida de las personas en los pueblos y ciudades (Tam, Lam 2004). La disponibilidad de plazas de aparcamiento está relacionada directamente con la movilidad urbana. Por lo tanto, las medidas de política de estacionamiento no solo afectan el sistema operativo de estacionamiento, sino que también generan impactos en la movilidad y el sistema socioeconómico de una ciudad (Anastasiadou, et al., 2009). Sin embargo, la viabilidad de un proyecto relacionado con la implementación de nuevos servicios de estacionamiento depende principalmente del método y la precisión de la estimación de la tarifa de estacionamiento apropiada que se cobrará a los usuarios potenciales. Una tarifa irrazonablemente alta disminuirá la demanda y puede poner en peligro la recuperación del proyecto. Hay que considerar lo que nos indican (Anastasiadou, et al., 2009) que las estrategias de precios de estacionamiento también se usan ampliamente en función de modelos de

elección y comportamiento porque imponen el uso de esquemas de cobro basados en parámetros de tiempo / demanda y los hábitos de los conductores.

3.2. Movilidad en Cuenca y Estacionamientos Privados

Los últimos años en Cuenca se ha observado un reparto modal que pone en los primeros lugares al vehículo privado y al transporte público en bus, ambos modos motorizados y masificados, esto ha hecho que ciertas zonas de la ciudad se vean seriamente afectadas por los efectos negativos de esta situación. En esta sección se proveen algunos detalles sobre la situación actual:

3.2.1. Los vehículos motorizados privados y el problema de transporte público en los centros históricos: el caso de Cuenca-Ecuador

La creación de parqueaderos en la zona céntrica de la ciudad ha aumentado el nivel de contaminación y de congestión vehicular en el centro de la ciudad (Moscoso Cordero, 2012). Los corazones de manzana que en un principio fueron utilizados como huertos han desaparecido en gran parte para dejar espacio para los coches, lo que genera aún más contaminación, teniendo en cuenta que esos espacios verdes eran los responsables de la descontaminación del aire; es necesario señalar que no hay muchas zonas verdes en el Centro Histórico destinadas a esta función. La principal complicación es la cantidad de tiempo que las unidades de transporte público tienen que invertir para recorrer el centro de la ciudad. Debido a la alta congestión en horarios pico a lo largo del día, un autobús se tarda aproximadamente 15 minutos para moverse 300 metros, produciéndose una exposición más larga de la ciudad a los gases contaminantes. Este problema va de la mano con la contaminación acústica causada por los coches en el interior del Centro Histórico, afectando a los peatones y causando dificultades de audición (Moscoso Cordero, 2012).

3.2.2. Plan de Movilidad y espacios públicos

El estacionamiento en edificaciones cuenta con 6.588 plazas distribuidas entre inmuebles de explotación privada (6.268) y edificios de titularidad pública (320) (Ilustre Municipalidad de Cuenca, 2015). La mayor concentración se localiza en el Centro Histórico (54%), abarcando la mitad de las plazas existentes; le sigue en porcentaje El Ejido (35%) y la Feria Libre (10%); las demás zonas poseen pocas plazas o carecen de ellas. Un dato preocupante es que alrededor del 46% de las plazas del Centro Histórico, que pertenecen a los estacionamientos de corazón de manzana, corresponden a usos ilegales o “no regulados”. Se tiene una oferta de estacionamientos que genera excesivo tráfico de agitación y favorece al ingreso de vehículos privados hasta zonas urbanas excesivamente interiores y vulnerables a la presencia de transporte motorizado. Como una forma de frenar el uso del vehículo privado y para

integrarlo a un nuevo modelo de movilidad, es necesario implementar “estacionamientos de borde” en ciertos lugares estratégicos de la ciudad, de tal manera que los usuarios accedan al transporte público y completen su recorrido y objetivos gracias a este medio.

El sistema de aparcamiento de Cuenca sufre un desbalance al analizar la zona de mayor concentración o demanda de viajes, el centro urbano. Debido al crecimiento del parque automotor de la ciudad y la escasez de suelo no hacen viable la construcción de nuevos parqueaderos, por lo que optar por nuevas infraestructuras no es una opción. La presión que sufre el Centro Histórico se da por el excesivo número de parqueaderos existentes que han incentivado al uso del vehículo particular especialmente por motivo “laboral”, que son personas que tienen una plaza fija en algún parqueadero debido a que su lugar de trabajo se encuentra en el sector, provocando congestión vehicular. Se necesita una normativa que promueva un uso eficiente de estas áreas, diferenciándose según la tipología de vehículos (Municipalidad de Cuenca, 2015). Un incremento en las plazas de estacionamiento agravaría la situación, debido a que un mayor número de vehículos ingresarán con este fin, aumentando la agitación del tráfico.

3.3. Data Mining aplicado al transporte

A continuación, se describe la importancia de usar a los datos como medio para buscar soluciones a problemas que enfrenta una ciudad, por ejemplo, la congestión vehicular. Una de las soluciones que se presenta, es la regulación de las tarifas de aparcamiento.

3.3.1. Nuevos algoritmos para la gestión de la demanda de estacionamiento y un despliegue a escala de ciudad

El DM para el bien social, permite que los servicios públicos escasos se utilicen de manera más eficiente y ayuda a reducir las externalidades no deseadas, como la congestión vehicular y la contaminación. El estacionamiento en la calle es un recurso escaso en muchos centros urbanos. Si el estacionamiento en la calle es gratuito, o tiene un precio significativamente por debajo de las tarifas del mercado, se utilizará de manera ineficiente, ya que los conductores no están adecuadamente incentivados para evitar las horas pico y las ubicaciones pico (Zoeter, Dance, Clinchant & Andreoli, 2014). Varios especialistas en estacionamiento (Shoup, 1997) han sugerido un enfoque más cauteloso: revisar las tarifas de estacionamiento a intervalos regulares, digamos cada mes, con base en los datos obtenidos de los sensores de estacionamiento. De esa forma, los conductores pueden memorizar las tarifas alrededor de su oficina, restaurante favorito, etc. y ajustar su comportamiento en consecuencia.

3.3.2. El problema con las actualizaciones basadas en la ocupación promedio

Basar los cambios en las tarifas con respecto a la tasa de ocupación promedio en un período de revisión, no es cierto para la mayoría de los modelos de utilidad razonables (Zoeter et. al. 2014). Si las tarifas cambian en función de la ocupación promedio, un período de subutilización seguido de congestión puede conducir a una situación perfecta "ni demasiado vacía ni demasiado llena", en promedio, y por lo tanto enmascarar los problemas que realmente han ocurrido. Esto es de particular importancia si las ventanas de tiempo se determinan algorítmicamente en función de los datos: una mañana tranquila seguida de una tarde congestionada se agrupan, ya que conjuntamente tendrán una ocupación promedio perfecta (Zoeter et. al. 2014). Los métodos para obtener la tarifa deben ser simples de entender, fáciles de ver y conducir a políticas de precios que sean fáciles de recordar y de aplicar para los conductores.

3.4. Datawarehouse aplicado a problemas de transporte

3.4.1. Data Mining aplicada a la ingeniería de transporte

Los pasos básicos involucrados en el DM y el KDD (Usama M, et. al. 1996) se puede visualizar en la Figura 14, los cuales son:

1. Comprensión del dominio de la aplicación.
2. Colección de conjunto de datos objetivo
3. Limpieza de datos y preprocesamiento
4. Almacenamiento de datos
5. Selección de la tarea de selección de datos relevantes
6. Selección de la tarea de minería de datos
7. Selección de herramienta de minería de datos - redes neuronales, algoritmos genéticos, árboles de decisión, más cercano, etc.
8. Minería de datos - identificación de relaciones -Clases, agrupaciones, asociaciones, patrones secuenciales
9. Interpretación de los resultados.
10. Consolidación del conocimiento descubierto

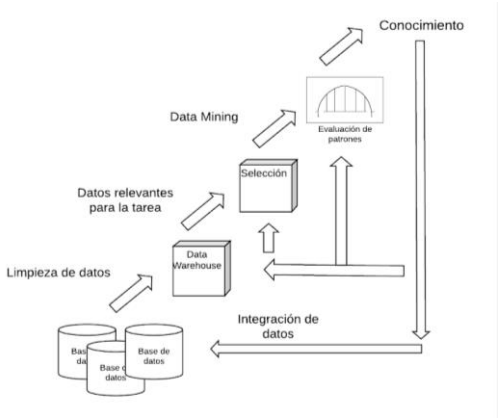


Figura 14 Data Mining y el proceso de descubrimiento de conocimiento. Tomado de (Usama M, et. al. 1996)

Capítulo IV

4. Materiales y Métodos

Este capítulo presenta los materiales y métodos usados a lo largo del desarrollo de este trabajo de titulación. Éste, al tratar una temática que aborda múltiples aristas del conocimiento humano, unificando temas de movilidad, con diversas áreas de desarrollo de software, tales como el Data Warehouse, Data Mining, Microservicios y Desarrollo de Aplicativos Web desarrolla los procedimientos definidos en cada sección del Capítulo II, Marco Teórico.

Así, este capítulo inicia analizando el tipo de alcance de investigación. Continúa precisando los diferentes materiales usados, estos fueron agrupados según su ámbito de uso: Modelos Teóricos, Datos, Herramientas de software existente y Lenguaje de Programación y Frameworks utilizados para el desarrollo de software. Seguidamente se detalla a los proveedores de los datos empleados y los propios datos según sus características. Prosigue especificando la metodología de Hefesto en el contexto del trabajo de titulación, pormenorizando todas sus etapas, como el análisis de requerimientos, el análisis OLTP, el modelo lógico del Data Warehouse, finalmente la integración de los datos y las reglas de actualización. En todas las etapas de la Metodología de Hefesto se detalla sus subetapas correspondientes, llevadas a cabo consecutiva y ordenadamente.

El capítulo continúa con la especificación sigilosa de los pasos seguidos para la Minería de Datos según el modelo estándar Crisp-DM, y sus etapas. Además del proceso para la construcción de los Dashboards que permiten analizar profundamente cada pregunta de la sección 4.4. Continuamente, se detalla el proceso de desarrollo del aplicativo web SPA con Angular, en conjunto con la arquitectura del sistema web.

Finalmente, se desarrolla el proceso para la construcción del microservicio con Spring Boot, el cual provee un REST –API consumido por la aplicación web SPA desarrollada.

4.1. Análisis del tipo de alcance

El tema propuesto se plantea como un proyecto técnico que desarrolla uno de los objetivos del proyecto de investigación central “Parámetros que influyen en la adhesión y permanencia a un sistema de viaje compartido en una comunidad universitaria. Estudio de caso: Programa de Viaje Compartido de la Universidad de Cuenca”, donde el alcance se establece como descriptivo y correlacional. A continuación, se justifica la elección:

- **Descriptivo:**
    - Este alcance, medirá la pertinencia de las variables referentes a los usuarios del aparcadero, con respecto al modelo matemático propuesto.
- **Correlacional:**

- Este alcance, permite desarrollar uno de los objetivos propuestos, al identificar valores de las variables sobre las dinámicas de movilidad que permitan afinar el cálculo de la tarifa diferencial.
- El modelo matemático planteado aún se encuentra en fase de pruebas, por lo que se busca relaciones que aumenten la efectividad de este modelo.

4.2. Materiales

Los materiales empleados se dividen en 3 grupos en los cuales se clasifica según su uso:

4.2.1. Modelo Teóricos

El Modelo Matemático a implementarse pertenece al "Plan tarifario para el uso del aparcamiento en el interior de los predios de la Universidad de Cuenca" creado por Avila-Ordóñez, et al., (2019). Dicho modelo cuenta con cuatro factores que influyen en el cálculo: distancia, cautividad, titularidad y dedicación. Estos factores se explicaron con anterioridad en
Marco Teórico-Sección 2.1. La fórmula de cálculo se detalló en la Fórmula 2.

4.2.2. Datos

Los datos utilizados en el estudio provienen de MAS y DTICS de la Universidad de Cuenca. Los datos se subdividieron en dos grupos: Administrativos y Movilidad. Los datos administrativos son provistos por DTICS y MAS, en donde se detallan las características socioeconómicas del personal administrativo que hace uso de los aparcaderos de la Universidad. En cuanto a movilidad se refieren a datos recolectados por DTICS y el grupo de investigación en los ingresos y salidas a los aparcaderos del Campus Central de la Universidad de Cuenca y los obtenidos en el estudio de auto compartido "Yo Te Llevo" (Avila & Cazorla, 2019). Las variables que contendrán los datos administrativos y de aparcadero se detallan en la Tabla 1 y 2.

Variables APARCADEROS
hora
fecha
puerta
tipo (tarjeta, placa)

id (# tarjeta, placa)

Tabla 1 Variables referentes al apartado de Aparcadero

Variables ADMINISTRATIVO
Cédula
Edad
Género
Provincia de Residencia
Cantón de Residencia
Dirección de Residencia Actual
Tipo de Servidor (Empleado, Profesor, Investigador)
Remuneración Salarial
Titularidad (contratado, titular, etc.)
Dedicación (medio tiempo, tiempo completo, etc.)
Campus en el cual labora de manera prioritaria
Núm. Vehículos que posee
Placa vehículo principal
Placa vehículo secundaria
Avalúo aproximado de los vehículos
Tiene Plaza Estacionamiento
ID Tarjeta Estacionamiento
Tiene Discapacidad
Tipo Discapacidad

Núm. de integrantes del grupo familiar
Núm. familiares (grupo familiar) que laboran en la Universidad
Tiene Bicicleta
Tiene acceso a estacionamiento de bicicletas
Tiene moto
Placa moto

Tabla 2 Variables referentes al apartado Administrativo.

4.2.3. Herramientas de software existente

La etapa del ETL hace uso de herramientas de software dependiendo la etapa que se realiza. La creación de un Data Warehouse tiene una etapa previa de limpieza y estandarización de estos, en los cuales se usará Pentaho Data Integration (Hitachi Vantara Editors, 2019) y la librería Pandas de Python (NumFOCUS, 2020). Posteriormente, se realiza la creación de los cubos de datos a través de PSW (Hitachi Vantara Editors, 2018) y finalmente se crean Dashboards Gráficos mediante Grafana 7.0 (Grafana Labs Editors, 2020). El sistema gestor de base de datos usado para todas las etapas de estudio, implementación y despliegue es PostgreSQL v12. Todos los procesos incluidos fueron parcial o totalmente complementados con el gestor de hojas de cálculo Microsoft Excel. Finalmente, se empleó el gestor de referencias Mendeley para la edición de este documento.

4.2.4. Lenguajes de Programación y Frameworks utilizados para el desarrollo de software

Para las etapas de creación de microservicios, la implementación del modelo y sus sub etapas de validación de datos se usará el lenguaje de programación Java 1.8 junto con el Framework SpringBoot (Apache Editors, 2020).

El desarrollo del aplicativo web, que consume los servicios del microservicio antes mencionado, requiere del lenguaje de programación Typescript y su framework para el lado del servidor Angular v7. El aplicativo web hace uso de Bootstrap v4, como biblioteca multiplataforma de código abierto para diseño de sitios y aplicaciones web ("Bootstrap (framework)", n.d). Finalmente, este

aplicativo web, al ser desarrollado por y para la Universidad de Cuenca, se emplea el Manual de Imagen Institucional para las etapas de diseño del aplicativo (UCuenca, 2018).

4.3. Proveedores de datos, y especificación de datos analizados

Los proveedores de los datos para este trabajo son:

- MAS:
    - Datos socioeconómicos del personal que emplea los aparcaderos universitarios
        - Archivo Excel con 1712 registros con información del personal docente, administrativos, empleados entre otros de la Universidad, además de una estimación inicial de los factores de cálculo del modelo matemático.
- DTICS:
    - Datos administrativos y de aparcaderos (datos del sistema de lectura por tarjetas y por detección de placa).
        - Administrativos con 26 columnas.
        - Aparcaderos con 5 columnas.
- ANT: datos vehiculares obtenidos mediante un API-Rest provisto por Ecuador Legal Online, un proveedor externo a la ANT que trata los datos vehiculares identificados por la placa del vehículo.

Para resguardar la privacidad de los usuarios de los aparcaderos y sus hábitos de movilidad, los datos se obtienen bajo un acta de confidencialidad de datos por parte de la DTICS y MAS. Los archivos contemplados en el análisis se detallan según su número de registros, esto permite evidenciar su manejo como Cubo de Datos y no solo por SGBD, se visualiza el número de registros por archivo con datos del año 2019-2020 en la Tabla 3.

Fuente de Datos	Núm. Registros
InfoAutos	1095
PlacasAcceso	17695
resultado_ddbb	797
registroentrada12deabril	19620
registroentradaeconomia	54367
registrosalidaArquitectura	38414
registrosalidaFilosofia	15915
datosAdministrativos	1712
datosAdminXFactor	1712

Tabla 3 Archivo base para los cubos multidimensionales

4.4. Metodología de Hefesto

La metodología de Hefesto consta de varias fases de desarrollo, las cuales son:

1. Análisis de Requerimientos
2. Análisis OLTP
3. Modelo Lógico del DW
4. Integración de Datos

Estas se explican de manera práctica para nuestro caso de estudio a continuación.

4.4.1. Análisis de Requerimientos

Esta fase consta en entrevistas iterativas con los involucrados en el modelo tarifario, buscando obtener preguntas a ser respondidas. A estas preguntas se les identifica perspectivas e indicadores, de esta manera se obtiene un primer diseño del modelo conceptual. Este proceso puede repetirse por cambios en las preguntas propuestas y su enfoque.

Preguntas del Negocio

Las preguntas identificadas tratarán de responder con estadística descriptiva, analizando el comportamiento de cada variable en la información obtenida. Las preguntas formuladas son:

1. Número de vehículos que ingresan y salen del campus central de acuerdo a una fecha, hora, puerta y modalidad de ingreso (tarjeta o reconocimiento de placa). Las puertas para analizar corresponden a: Puerta Principal, Puertas de las facultades de: Jurisprudencia, Economía y Arquitectura. Ver Figura 15 secciones A, B, C, E determinado.

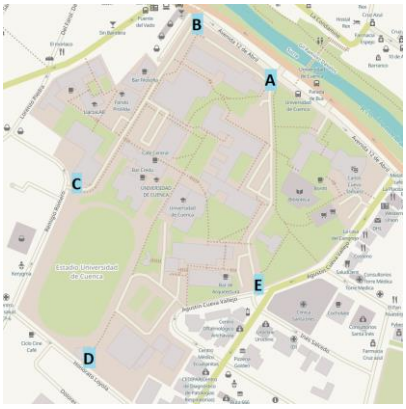


Figura 15 Zonas de aparcamiento en el campus central

2. Número de personas que ingresan al campus central por género, fecha y hora.
3. Número de personas según el tipo de servidor y titularidad que transitan por cada puerta del campus central.

4. Número de personas que ingresan al campus central, según la dependencia laboral (lugar donde laboran) y la dedicación (tiempo completo, medio tiempo, etc.).
5. Número de personas respecto a su remuneración salarial que usan los aparcaderos.
6. Número de personas que usan los aparcaderos, respecto a su edad (a mayor edad, se puede inferir mayor necesidad de transporte privado).
7. Número de vehículos que transitan por los aparcaderos según el cantón de la placa del vehículo.
8. Número de personas que ingresan al campus central de acuerdo con el tipo de discapacidad.
9. Número de vehículos que ingresan/salen de acuerdo con el cilindraje del vehículo, año, clase de vehículo por fecha y hora.
10. Número de personas que ingresan de acuerdo al factor de Cautividad.
11. Número de personas que ingresan de acuerdo al factor de Dedicación.
12. Número de personas que ingresan de acuerdo al factor de Distancia.
13. Número de personas que ingresan de acuerdo al factor de Titularidad.
14. Número de personas que ingresan/salen de acuerdo al tipo de servidor y distancia de origen.
15. Número de personas que ingresan/salen de acuerdo con la titularidad y distancia de origen.

Indicadores y Perspectivas

Los indicadores y perspectivas se generan a partir de preguntas bien formuladas, estos se identifican a continuación por código de color (Indicador en Naranja y Perspectiva en Celeste) o se puede observar de manera global en la Tabla 4.

1. Número de vehículos que ingresan y salen del campus central de acuerdo a una fecha, hora, puerta y modalidad de ingreso (tarjeta o reconocimiento de placa).
2. Número de personas que ingresan al campus central por género, fecha y hora.
3. Número de personas según el tipo de servidor y titularidad que transitan por cada puerta del campus central.
4. Número de personas que ingresan al campus central, según la dependencia laboral (lugar donde laboran) y la dedicación (tiempo completo, medio tiempo, etc.).
5. Número de personas respecto a su remuneración salarial que usan los aparcaderos.
6. Número de personas que usan los aparcaderos, respecto a su edad (a mayor edad, se puede inferir mayor necesidad de transporte privado).
7. Número de vehículos que transitan por los aparcaderos según el cantón de la placa del vehículo.
8. Número de personas que ingresan al campus central de acuerdo al tipo de discapacidad.
9. Número de vehículos que ingresan/salen de acuerdo al cilindraje del vehículo, año, clase de vehículo por fecha y hora.
10. Número de personas que ingresan de acuerdo al factor de Cautividad.
11. Número de personas que ingresan de acuerdo al factor de Dedicación.

12. Número de personas que ingresan de acuerdo al factor de Distancia.
13. Número de personas que ingresan de acuerdo al factor de Titularidad.
14. Número de personas que ingresan/salen de acuerdo al tipo de servidor y distancia de origen.
15. Número de personas que ingresan/salen de acuerdo con la titularidad y distancia de origen.

Indicador	**Perspectivas**
Número de vehículos	Fecha
Número de personas	Hora
	Puerta
	Modalidad de ingreso
	Clase de vehículo
	Cilindraje
	Cantón vehículo
	Año del vehículo
	Tipo de servidor
	Titularidad
	Dedicación
	Dependencia Laboral
	Remuneración Laboral
	Edad
	Tipo de discapacidad
	Género
	factor Distancia
	factor Cautividad
	factor Dedicación
	factor Titularidad

Tabla 4 Indicadores y Perspectivas

Modelo Conceptual

El modelo conceptual referente a los indicadores y perspectivas identificadas anteriormente se detallan en la Figura 16.

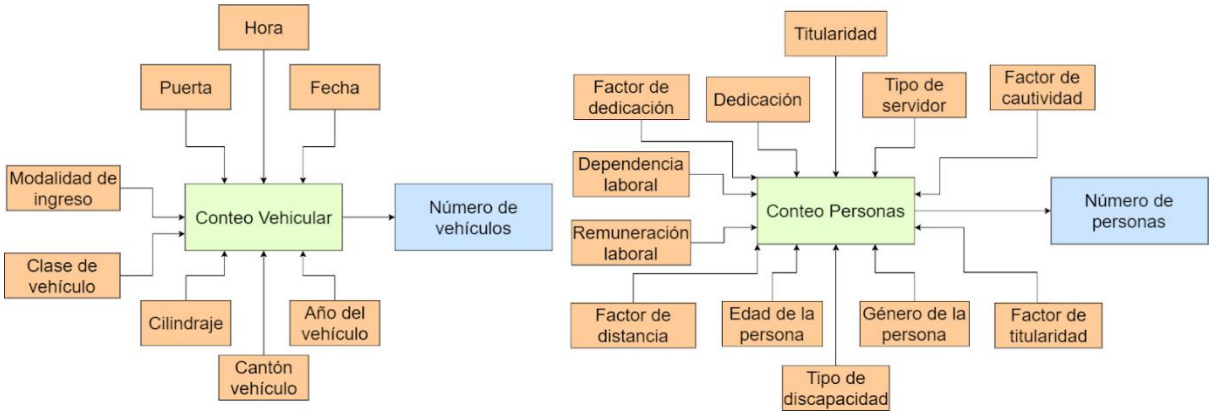


Figura 16 Modelo Conceptual

4.4.2. Análisis OLTP

La segunda fase implica el análisis OLTP, la cual consta de varios pasos como la conformación de indicadores, establecimiento de correspondencias, nivel de granularidad y finalmente el modelo conceptual ampliado. A continuación, se detalla cada uno de los pasos anteriormente mencionados.

Hechos e Indicadores

Los hechos, indicadores y función de agregación se analizan en los siguientes literales:

- Indicador: Número de Vehículos
  - **Hecho:** Número de Vehículos
  - **Función de agregación:** COUNT
  - **Aclaración:** representa el aforo vehicular de ingreso y salida.
- Indicador: Número de Personas
  - **Hecho:** Número de Personas
  - **Función de agregación:** COUNT
  - **Aclaración:** cada ingreso y salida vehicular es considerado como un ingreso o salida de una persona, debido a la falta de información acerca de cuántas personas ingresan por vehículos o salen.
- Indicador: Conteo
  - **Hecho:** Conteo
  - **Función de agregación:** COUNT
  - **Aclaración:** este Indicador y Hecho representará a los dos anteriormente descritos, debido a la decisión de tratar a cada ingreso vehicular también como un ingreso de

persona, entonces un ingreso o salida representa un vehículo y una persona al mismo tiempo.

Mapeo - Correspondencia y Granularidad

Los datos por extraerse fueron proporcionados por la DTICS y MAS detallados en la primera sección del documento. La Figura 17 indica los campos que conforman cada perspectiva. La correspondencia entre perspectivas e indicadores se observa en la Figura 18.

Fecha fechaCompleta	**Hora** horaCompleta	**Clase** descripcion	**Cilindraje** cilindraje	**Factor de Distancia** valor
Puerta descripcion	**Modalidad** descripcion	**Canton Matricula** descripcion	**Año del vehículo** año	**Factor de Dedicacion** valor
tipoServidor descripcion	**titularidad** descripcion	**remuneracion** valor	**edad** edadPersona	**Factor de Cautividad** valor
dedicacion descripcion	**dependencia** descripcion	**tipoDiscapacidad** tipo	**Genero** generoPersona	**Factor de Titularidad** valor

Figura 17 Granularidad de las perspectivas

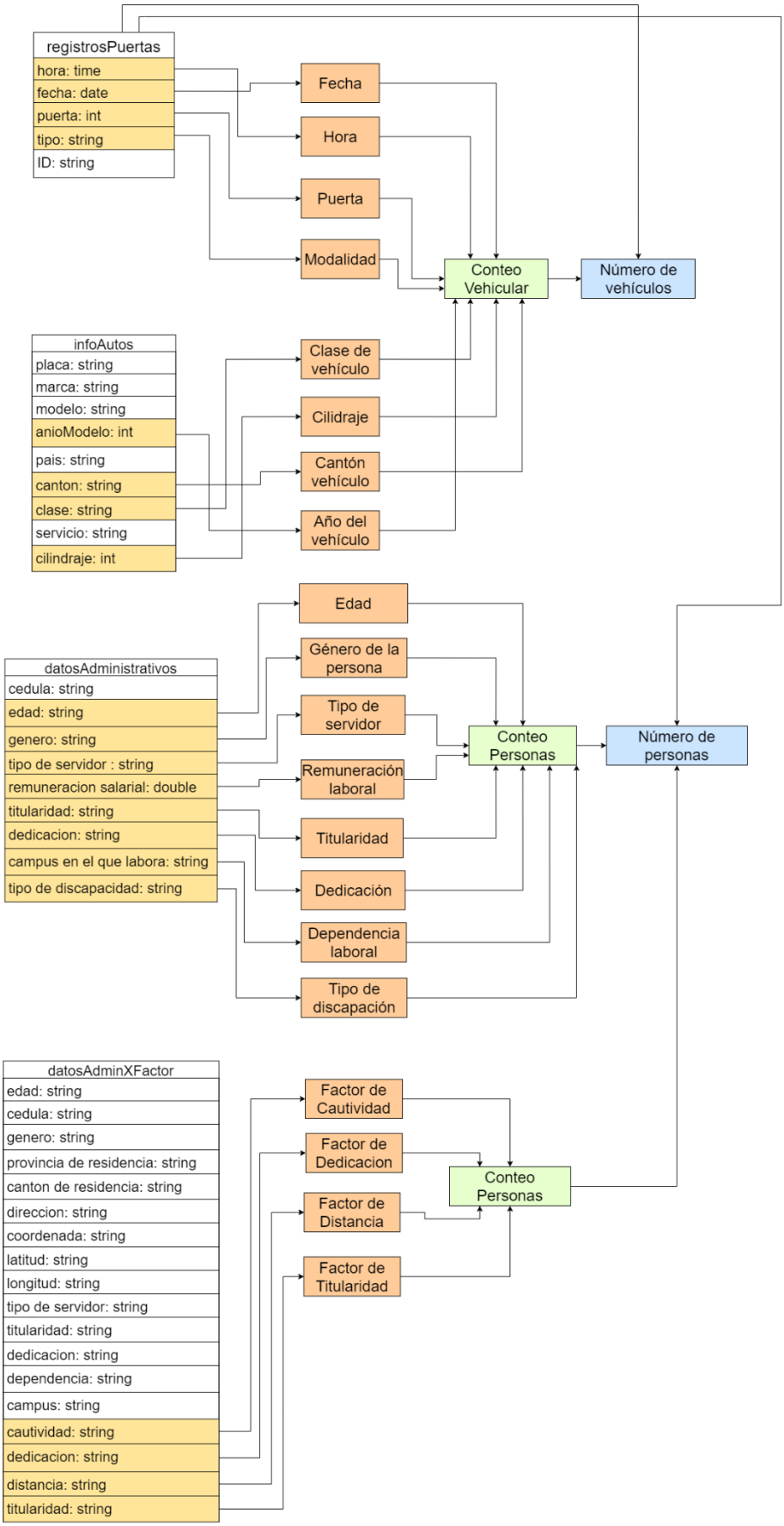


Figura 18 Correspondencia entre perspectivas y hechos con las fuentes de datos

Modelo Conceptual Ampliado

El modelo conceptual ampliado aplicable a nuestras perspectivas e indicadores se definen en la Figura 19.

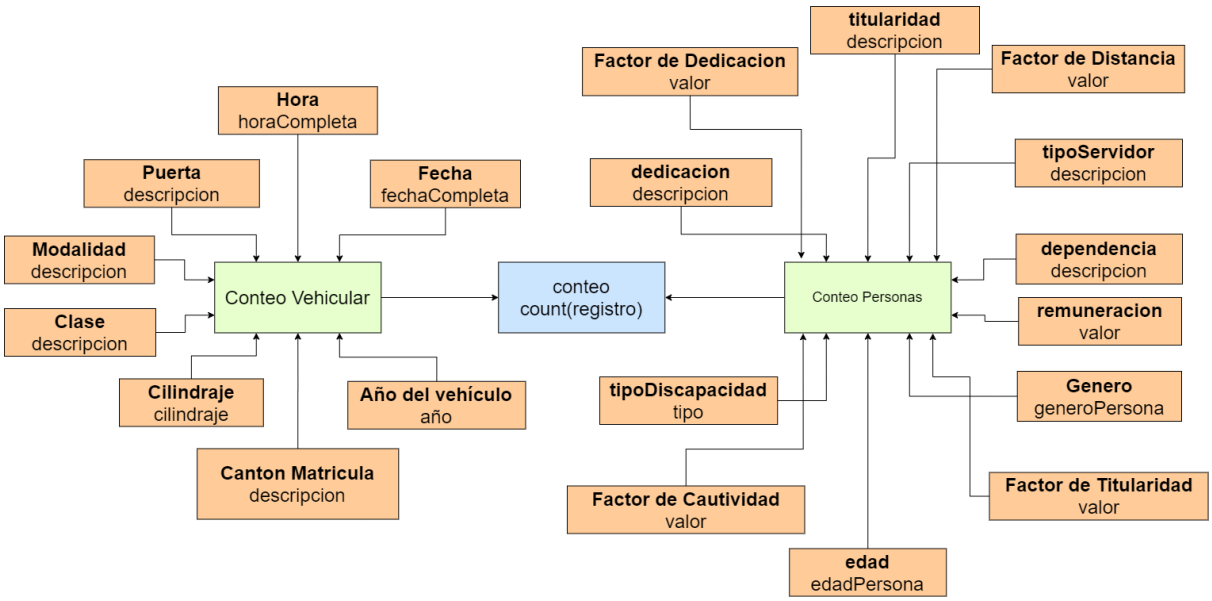


Figura 19 Modelo Conceptual ampliado

4.4.3. Modelo Lógico del DW

El modelo lógico de un DW es la representación de una estructura de datos que puede procesarse y almacenarse en algún SGBD.

Tipología

Analizando los datos existentes y las preguntas a responder se selecciona el Esquema en Estrella. Esto se debe a los datos y la forma en que se pueden organizar.

Tablas de Dimensiones

Cada perspectiva identificada anteriormente, debe ser representada como una tabla Dimensión, para lo cual se realizan los siguientes pasos:

- Elegir un nombre que identifique la tabla de Dimensión.
- Añadir un campo que represente su clave principal.
- Definir los nombres de los campos, si estos no son intuitivos.

En la Figura 20 se observa las ocho primeras Dimensiones descritas a continuación:

- **Perspectiva Fecha:** se escoge la fecha más antigua y la más actual, para obtener un rango de fechas que constituirán los datos de esta dimensión.
- **Perspectiva Hora:** se utiliza el formato 24h00 (00:00 - 23:59).

- **Perspectiva Puerta:** puertas de ingreso y salida consideradas.
- **Perspectiva Modalidad:** modalidad tarjeta magnética o detección de placa.
- **Perspectiva Clase:** las clases de vehículos que transitan por los aparcaderos, como, por ejemplo: automóvil, camión, camioneta, etc.
- **Perspectiva Cilindraje:** refiere a los cilindrajes de los vehículos que transitan por los aparcaderos.
- **Perspectiva Cantón:** indica el cantón y provincia al que pertenece determinado vehículo.
- **Perspectiva Año:** años de fabricación de los vehículos que transitan por los aparcaderos.

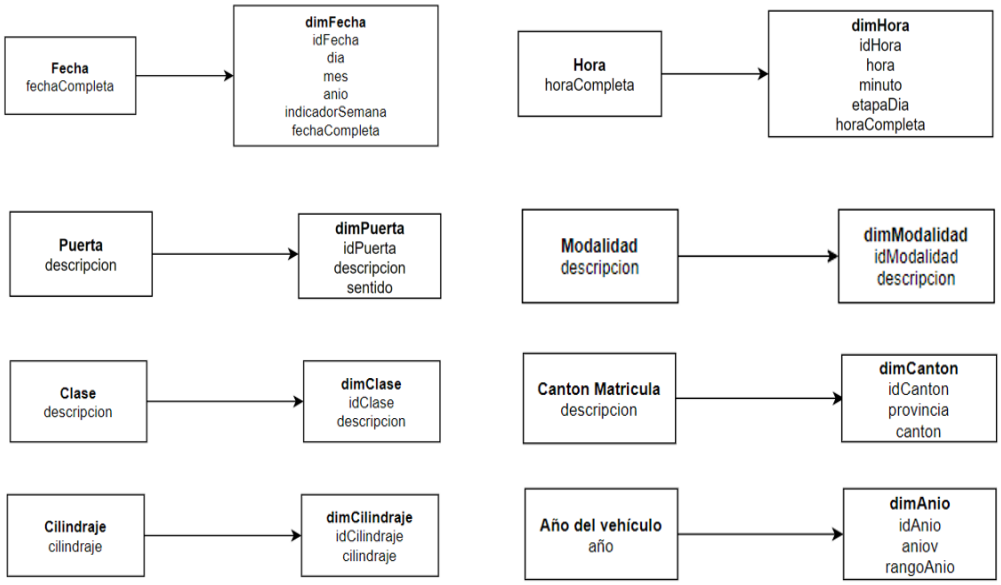


Figura 20 Relación de perspectivas a tablas dimensiones

En la Figura 21 se observa las siguientes ocho Dimensiones:

- **Perspectiva Servidor:** permite diferenciar entre docente, empleado y trabajador.
- **Perspectiva Titularidad:** personas contratadas o titulares usuarios de los aparcaderos.
- **Perspectiva Dedicación:** indica si la persona trabaja: tiempo completo, medio tiempo, tiempo parcial u horas clase.
- **Perspectiva Dependencia:** almacena los distintos destinos laborales de los usuarios de los aparcaderos, no es necesario que laboren en el campus central, pueden trabajar en otros campus.
- **Perspectiva Remuneración:** establece rangos de remuneración salarial que van entre 1 SBU a 2 SBU, de 2 SBU a 3 SBU, etc.
- **Perspectiva tipoDispacidad:** permite diferenciar a las personas de acuerdo al tipo de discapacidad que posean.
- **Perspectiva Edad:** almacena las edades de los usuarios de los aparcaderos, se asigna un rango de edad, por ejemplo: Entre 20 a 30, Entre 30 a 40, etc.
- **Perspectiva Género:** identifica el género de nacimiento de cada servidor.

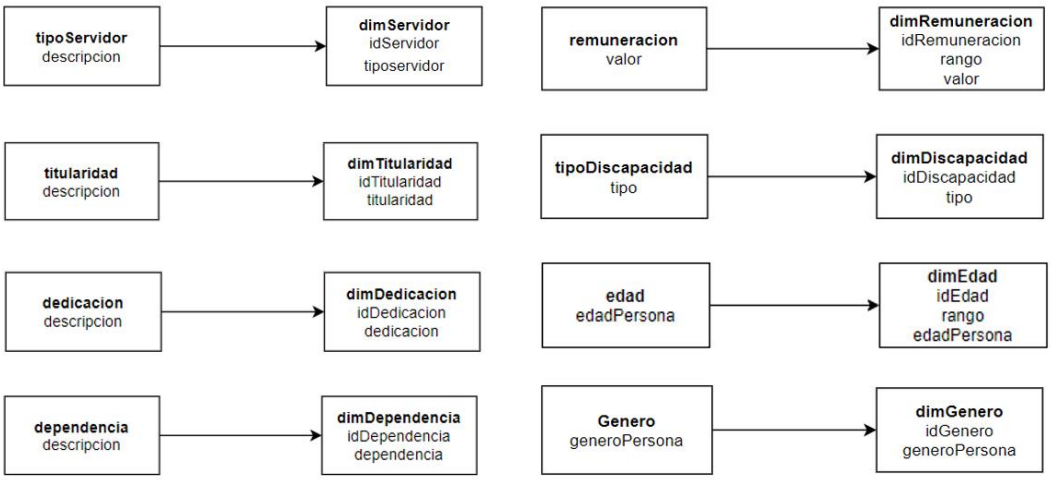


Figura 21 Perspectivas a tablas dimensiones

En La Figura 22 se observa las cuatro últimas Dimensiones:

- **Perspectiva Factor Distancia, Cautividad, Dedicación, y Titularidad:** se almacena el valor de los factores calculados por el grupo MAS.

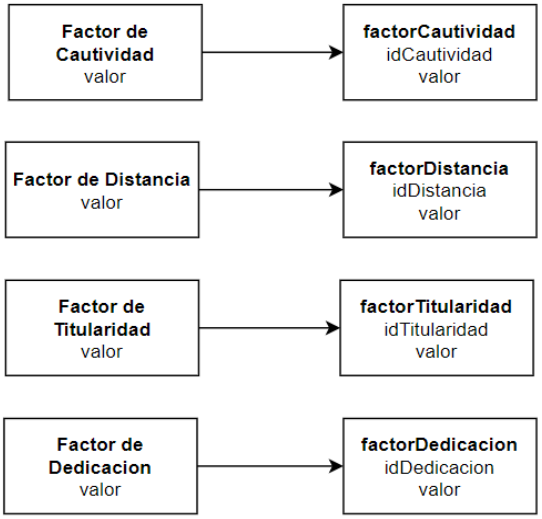


Figura 22 Perspectivas a tablas dimensiones

Tablas de Hechos

Los indicadores y sus operaciones deben ser representadas mediante una Tabla de Hechos, para lo cual se realizan los siguientes pasos:

- Asignar un nombre a la Tabla de Hechos que represente la información que contiene, área de investigación, negocio enfocado, etc.
- Definir la clave primaria, que se compone de la combinación de las claves primarias de cada tabla de Dimensión relacionada.

- Crear tantos campos de Hechos como indicadores se hayan definido en el modelo conceptual y se les asignará un nombre.

Como se ha explicado en la fase dos de la metodología de Hefesto, se posee un único indicador que representa número de vehículos como número de personas, para dicho caso (Bernabeu and Mattío 2017) mencionan lo siguiente: “Si en dos o más preguntas de negocio figuran los mismos Indicadores, pero con diferentes perspectivas de análisis, existirán tantas tablas de Hechos como preguntas que cumplan esta condición.”. A partir de lo mencionado anteriormente se decidió crear tres cubos con el mismo indicador, pero con diferentes perspectivas las cuales son las siguientes: Aparcadero, Vehículo y Administrativo. Las cuales se pueden visualizar en la Figura 23.

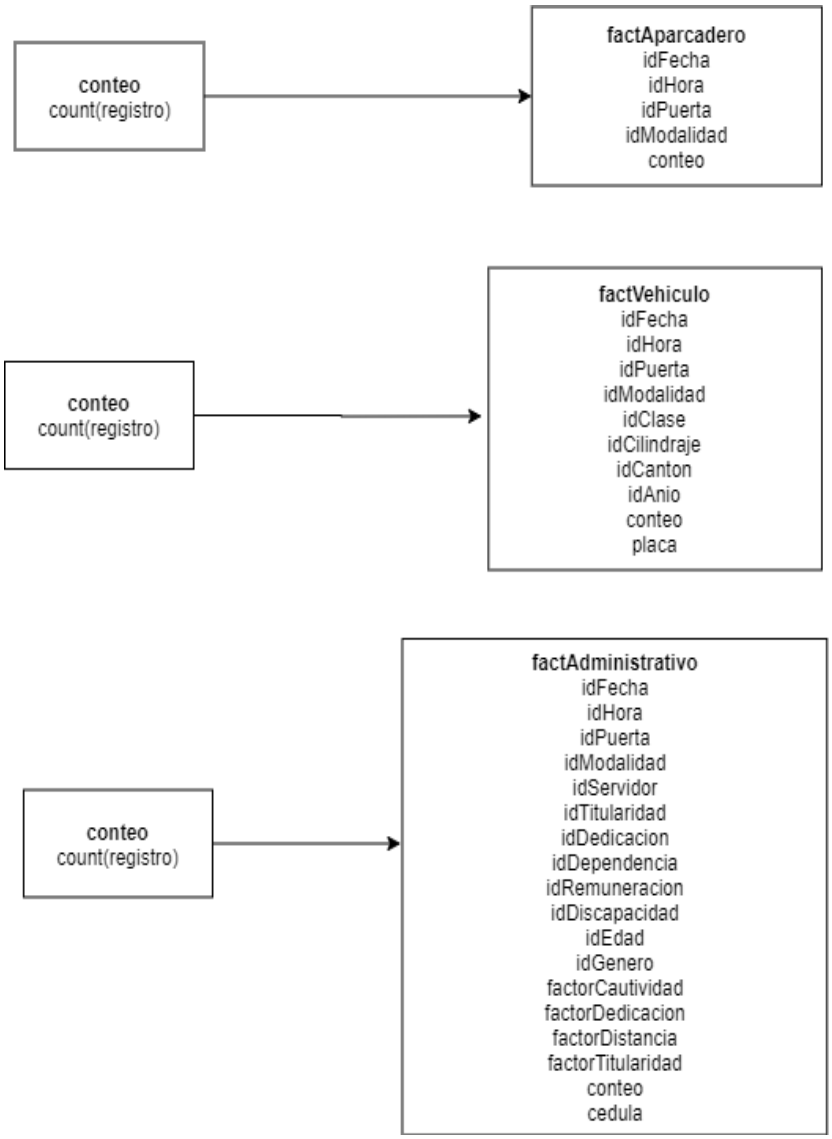


Figura 23 Indicadores a tablas hechos

Unión

Se define la relación correspondiente entre Tablas de Dimensiones y Tablas de Hechos. En las Figuras 24, 25 y 26 se visualizan dichas uniones.

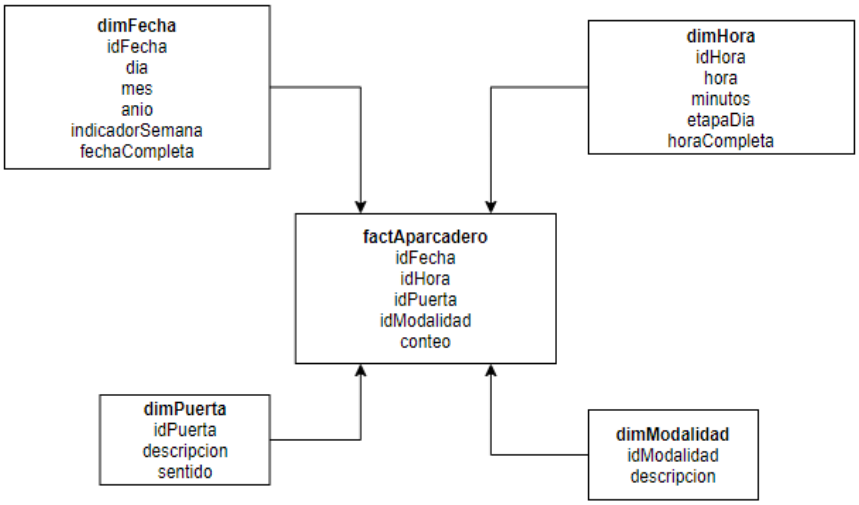


Figura 24 Uniones entre Dimensiones y Hecho con perspectiva aparcadero

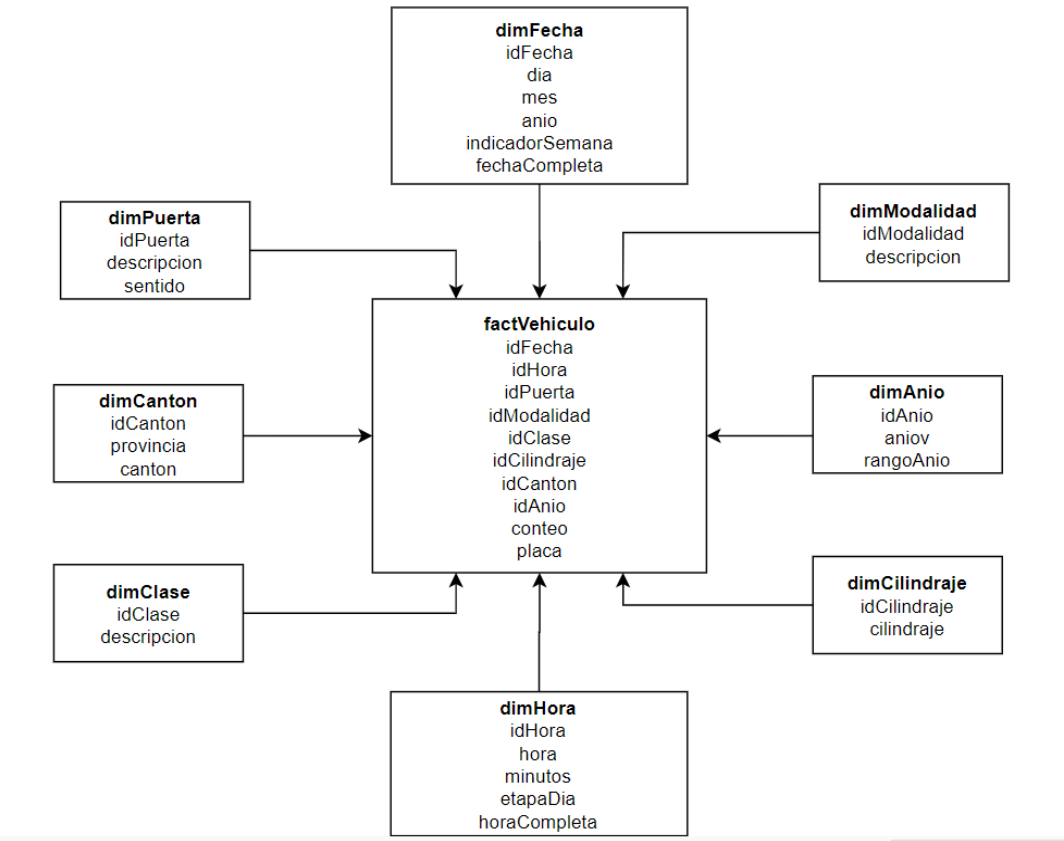


Figura 25 Uniones entre Dimensiones y Hecho con perspectiva vehículo

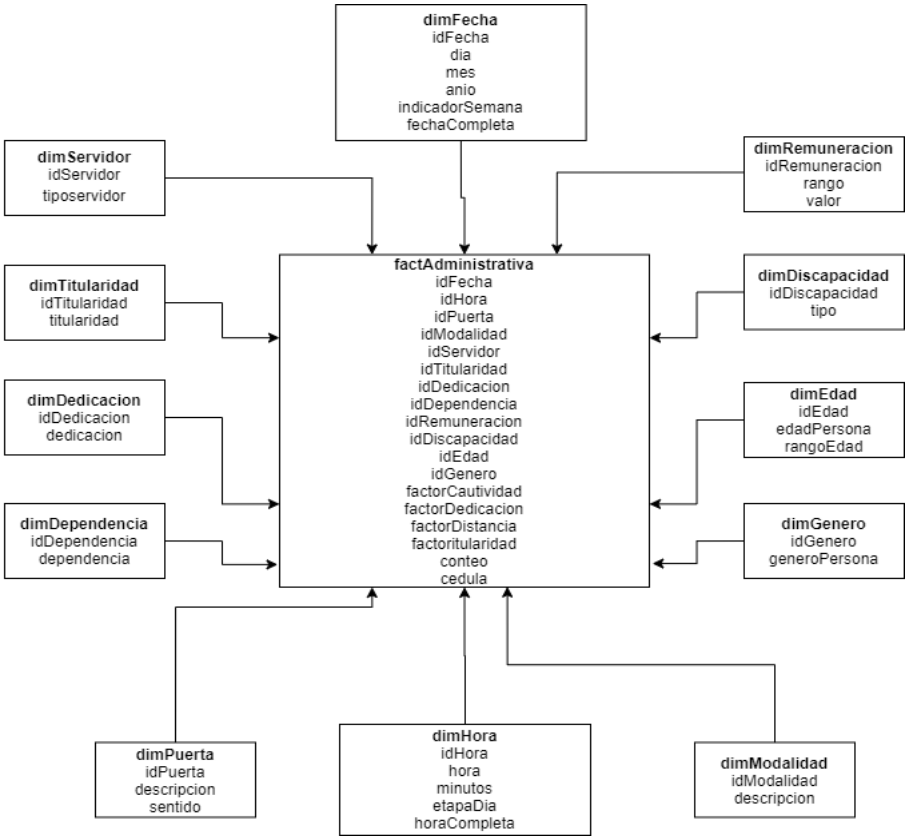


Figura 26 Uniones entre Dimensiones y Hecho con perspectiva administrativa factor

En la Figura 26 las Dimensiones factorCautividad, factorDedicacion, factorDistancia y factorTitularidad forman parte del Hecho administrativo. Esta particularidad se define como "Dimensiones degeneradas" descritos por Saquicela, (2015) como *"Atributos que no son hechos en sí mismo, pero tampoco claves de ninguna tabla dimensión*" además que los valores de estas Dimensiones serán utilizados como criterios de análisis coincidiendo con lo que Bernabéu & Mattío, (2017) menciona: "*Las dimensiones degeneradas hacen referencia a un campo que será utilizado como un criterio de análisis y que es almacenado en la Tabla de Hechos*". Finalmente, por temas del dominio del problema el almacenar valores cuantitativos como tablas Dimensiones, es contraproducente en etapas futuras de análisis.

4.4.4. Integración de Datos

En esta etapa se procede a cargar los datos para poblar las Dimensiones y Hechos que forman parte de nuestro DW. Consecuentemente se definirán reglas y políticas de actualización para el mantenimiento del DW.

Carga Inicial

Consiste en detallar los procesos seguidos para realizar la primera carga de datos al DW como son las Tablas de Dimensiones y Hechos, además de explicar acerca del tratamiento o preparación de los datos que se haya realizado.

Carga de Dimensión Fecha

El proceso de carga se observa en la Figura 27, cada etapa será explicada a continuación:

1. Lectura de archivo Excel que posee fechas entre la más antigua y la más actual.
2. Extracción del día de la semana, el mes de la fecha, el año de la fecha y el día del mes de la fecha.
3. Mapeo para obtener el día exacto, por ejemplo: lunes, martes, miércoles, etc.
4. Mapeo para obtener los meses del año, por ejemplo: enero, febrero, marzo, etc.
5. Creación de una columna que indica "día de la semana" o "fin de semana".
6. Unión del año, mes y día para formar un id de fecha.
7. Ordenación de las columnas.
8. Almacenamiento de la dimensión fecha en archivo csv.
9. Lectura del archivo del paso 8.
10. Carga de la Dimensión en la BD.

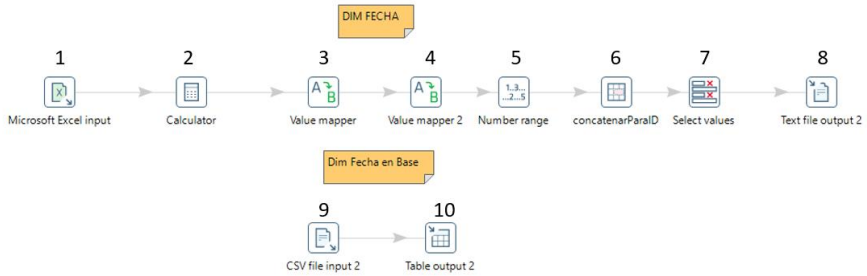


Figura 27 Proceso para la dimensión fecha

Carga de Dimensión Hora

La Figura 28 representa el proceso para esta Dimensión:

1. Lectura de archivo Excel con horas y minutos desde las 00:00 a 23:59
2. Creación de secuencia para el id de la Dimensión.
3. Establecimiento de rangos de valores para obtener una nueva columna que describa la etapa del día (mañana, tarde, noche y amanecer).
4. Ordenamiento las columnas
5. Almacenamiento de la dimensión en un archivo csv.
6. Lectura del archivo del paso 6.
7. Almacenamiento de la Dimensión en la BD.

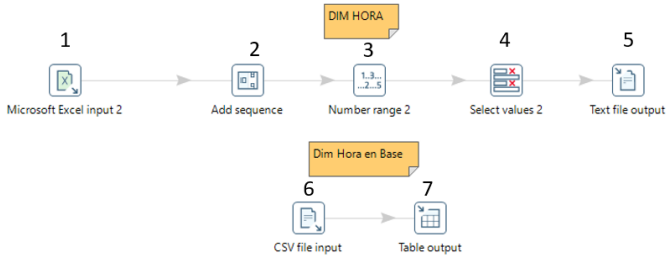


Figura 28 Proceso para la dimensión hora

Carga de Dimensiones Puerta, Modalidad, Clase, Cilindraje, Cantón, Servidor, Titularidad, Dedicación, Dependencia, Discapacidad, Género.

Estas Dimensiones ya contaban con la información preparada, su proceso de almacenamiento de CSV a la DB se visualiza en la Figura 29.

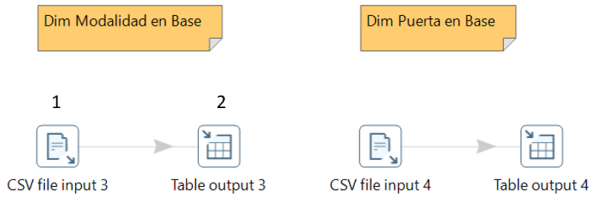


Figura 29 Proceso de carga para varias Dimensiones

Carga de Dimensión Año vehicular, Remuneración y Edad

Estas dimensiones tuvieron el proceso de la Figura 30 descrita a continuación:

1. Lectura de un archivo Excel con los datos para dichas Dimensiones.
2. Creación de una nueva columna de acuerdo a rangos definidos.
3. Almacenamiento en archivo csv de las Dimensiones.
4. Lectura del archivo del paso 3.
5. Almacenamiento en la BD.

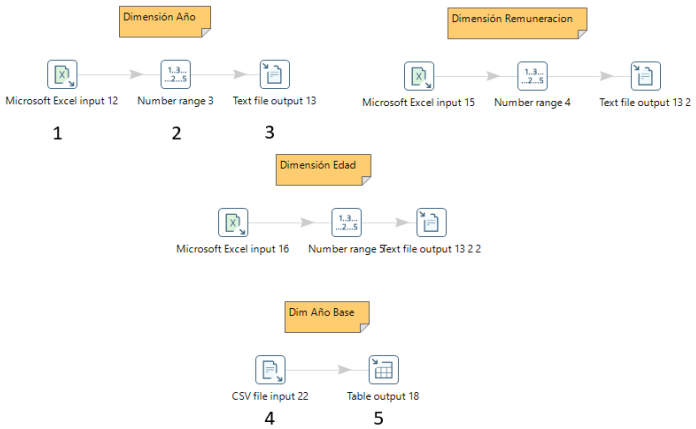


Figura 30 Dimensiones año, remuneración y edad

Antes de realizar la carga de los datos para los Hechos aparcadero, vehículo, administrativo y administrativo factor, se realizaron varios procesos sobre los datos de los registros de ingreso y salida del campus que observan en la Figura 31 y descritos a continuación:

1. Lectura de los registros de ingreso y salida de las diferentes puertas del campus central (Cuatro de estos registros utilizan la tarjeta y uno con detección de placas).
2. Agregación de dos constantes para el id de las dimensiones puerta y modalidad.
3. Reemplazamiento de los caracteres "-" por ":" para la hora y para la fecha "-" por "/".
4. Creación de un archivo CSV para la salida de estos procesos.

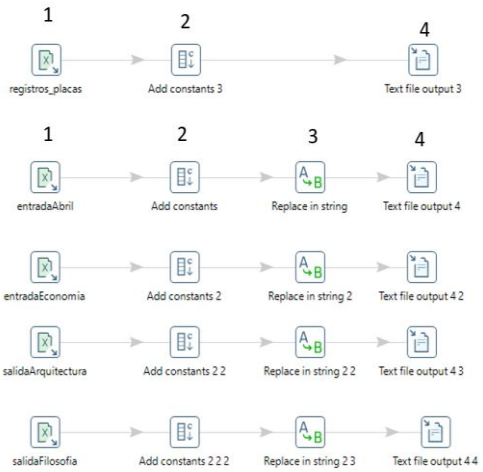
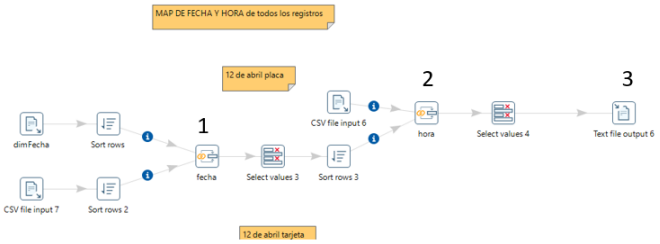


Figura 31 Preprocesamiento de registros de entrada y salida

En seguida se mapeo las fechas y horas con cada archivo csv correspondientes a las diferentes puertas. Este proceso se muestra en la Figura 32 con la estructura final de los archivos de ingresos y salidas. Los pasos más importantes son:

1. Join entre la dimensión Fecha y los registros por el campo Fecha.
2. Join entre la dimensión Hora y la salida del paso 1 por el campo Hora.
3. Creación de un archivo csv de salida para cada modalidad (cámara o tarjeta).



idFecha	idHora	idPuerta	idModalidad	ID
20200130	358	1	2	ABG9619
20200303	366	1	2	ABG9619
20200211	374	1	2	ABA9606
20200206	375	1	2	ABA9606

Figura 32 Proceso para obtener los archivos de ingresos y salidas de cada puerta de ingreso/salida.

Además, se realizó un preprocesamiento para los datos administrativos mediante la librería Pandas de Python, obteniendo un archivo csv con los datos administrativos tratados. Permitiendo reducir el esfuerzo en la carga de los Hechos aparcaderos y administrativos que hacen uso de este archivo para diferentes cruces. Este proceso se observa en la Figura 33, sus pasos son:

1. Lectura de archivo csv con los datos administrativos.
2. Extracción del valor de la edad, ya que la columna fue calculada y está compuesta de la edad con meses, días, horas, etc.
3. Mapeo sobre las columnas género, tipo de servidor, titularidad, dedicación, tipo de discapacidad.
4. Verificación de placas, deben tener 4 dígitos en su parte numérica.
5. Creación de un archivo csv de salida.

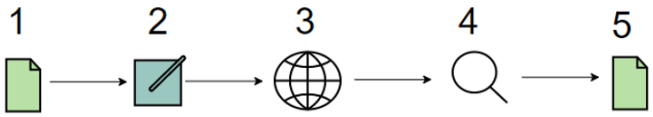
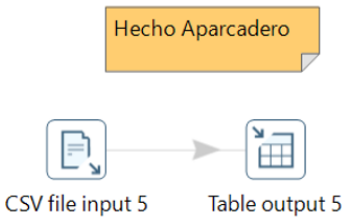


Figura 33 Secuencia de operaciones que se realizó a los datos administrativos

Carga de Hecho Aparcadero

Con el preprocesamiento realizado solo resta unir los dos archivos (registros tarjeta y placa), eliminar la columna “ID” y agregar una columna que se llame “conteo” y proceder a almacenar en la DB como se muestra en la Figura 34.



idfecha	idhora	idpuerta	idmodalidad	conteo
20191013	1	1	1	1
20191016	13	1	1	2
20191026	14	1	1	3
20190927	15	1	1	4
20190908	31	1	1	5
20191026	35	1	1	6
20191013	40	1	1	7
20190629	45	1	1	8
20191024	46	1	1	9
20190629	62	1	1	10
20191024	62	1	1	11
20191024	62	1	1	12

Figura 34 Hecho Aparcadero

Carga de Hecho Vehículos

Los dos archivos de registros, uno por cada modalidad (tarjeta y detección de placa), deben consolidarse en uno solo. Se crearon dos procesos que realizan los cruces necesarios para obtener los datos vehiculares y nos devuelva dos archivos uno para cada modalidad. Ver Figura 35. Posteriormente, mediante Python se realizaron los mapeos faltantes generando dos archivos csv para cada modalidad, después se realizó la unificación de estos en un solo archivo para así ingresar en la DB como tabla de Hecho vehículo. Ver Figura 36.

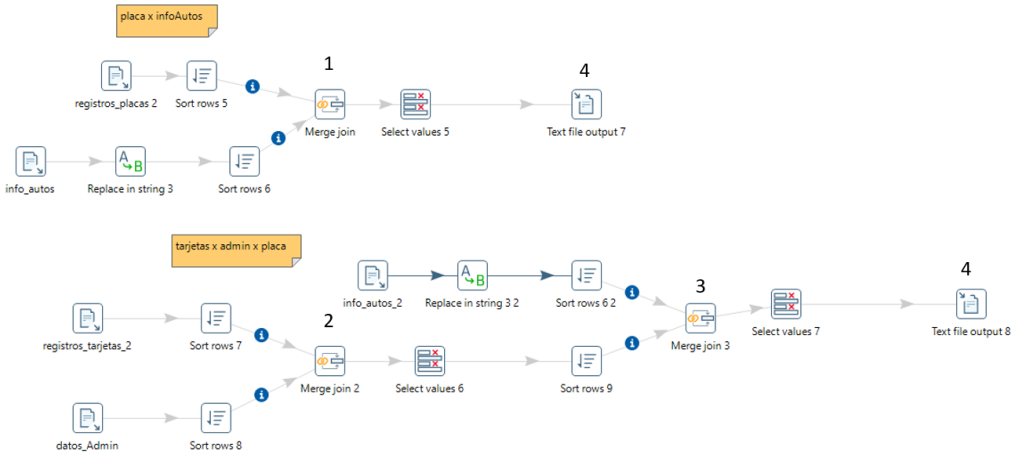
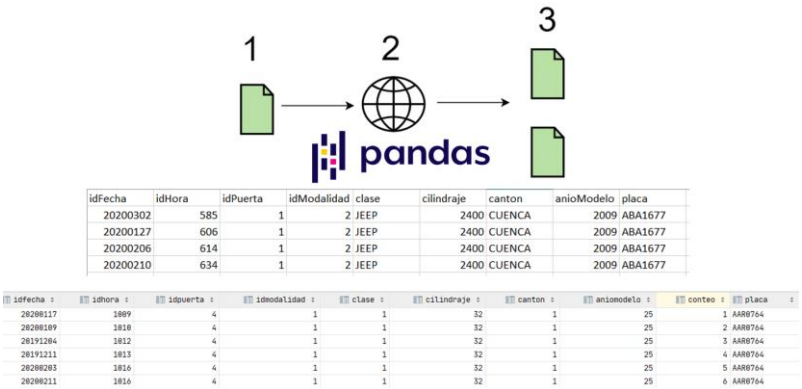


Figura 35 Cruce entre registros y la información vehicular



idFecha	idHora	idPuerta	idModalidad	clase	cilindraje	canton	anioModelo	placa
20200302	585	1	2	JEEP	2400	CUENCA	2009	ABA1677
20200127	606	1	2	JEEP	2400	CUENCA	2009	ABA1677
20200206	614	1	2	JEEP	2400	CUENCA	2009	ABA1677
20200210	634	1	2	JEEP	2400	CUENCA	2009	ABA1677

idfecha	idhora	idpuerta	idmodalidad	clase	cilindraje	canton	aniomodelo	conteo	placa
20200117	1009	4	1	1	32	1	25	1	AAR0764
20200109	1010	4	1	1	32	1	25	2	AAR0764
20191204	1012	4	1	1	32	1	25	3	AAR0764
20191211	1013	4	1	1	32	1	25	4	AAR0764
20200203	1016	4	1	1	32	1	25	5	AAR0764
20200211	1016	4	1	1	32	1	25	6	AAR0764

Figura 36 Hecho Vehículos

Carga de Hecho Administrativo

Los datos administrativos deben ser previamente tratados, para lo cual, a partir de este tratamiento se realiza los siguientes pasos:

1. En la Figura 37, se realiza el cruce mediante placas y tarjetas con los datos administrativos.

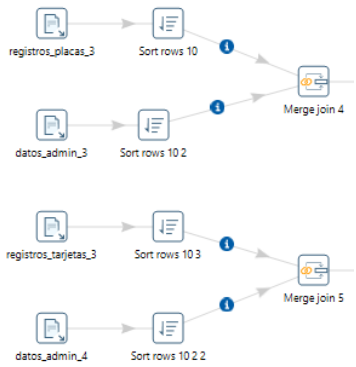


Figura 37 Cruce de los ingresos por placa y tarjeta con los datos administrativos

2. En la Figura 38, se realiza el cruce mediante la remuneración entre la salida del paso 1 previamente ordenada por dicho campo y la dimensión remuneración creada anteriormente.

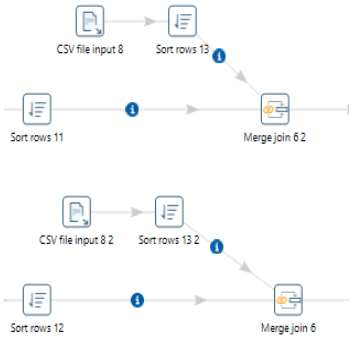


Figura 38 Cruce por medio de la remuneración

3. En la Figura 39, se realiza el cruce mediante la edad entre la salida del paso 2 previamente ordenada por dicho campo y la dimensión edad creada anteriormente.

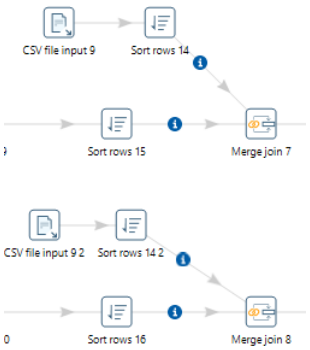


Figura 39 Cruce por medio de la edad

4. En la Figura 40, se realiza el cruce mediante la cédula entre la salida del paso 3 previamente ordenada por dicho campo y el archivo que contiene los factores. Al finalizar esta etapa se creó un archivo único para todos los ingresos y salidas.

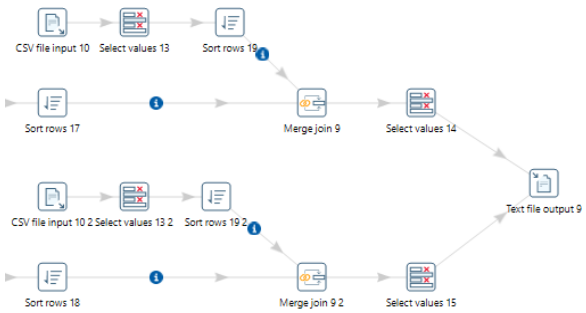


Figura 40 Cruce por medio de la cédula para los factores

5. Finalmente se utiliza Excel para el cruce entre la dependencia y la dimensión dependencia a través de la función buscarV y se procedió a guardar en la base el hecho. Ver Figura 41.

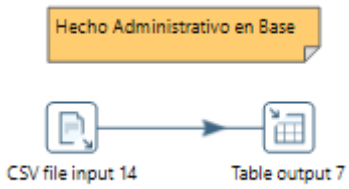


Figura 41 Carga del Hecho Administrativo

La implementación de los cubos multidimensionales jerárquicos se realizó en PSW. La Figura 42, los identifica con sus tablas y dimensiones respectivas. Mientras en la Figura 43 se visualiza un ejemplo de consulta en el cubo aparcadero con la herramienta Pentaho Server.

Figura 42 Esquema de Cubos realizado en Schema Workbench

Puerta - Puerta		Puerta 12 Abril	Puerta Arquitectura	Puerta Economia	Puerta Filosofia
Anio	Mes	registro	registro	registro	registro
2019	Agosto	101	101	101	-
	Diciembre	-	7.978	7.978	2.774
	Julio	1.506	1.506	1.506	-
	Junio	967	967	967	-
	Noviembre	1.286	9.009	9.009	3.138
	Octubre	8.595	8.597	8.597	724
	Septiembre	7.164	5.807	5.807	999
2020	Abril	-	-	3	-
	Enero	3.502	3.492	10.095	4.001
	Febrero	7.423	-	6.139	2.960
	Marzo	6.769	-	4.164	1.318

Figura 43 Ejemplo de una consulta en Pentaho Server

Actualización

Las políticas de actualización de datos, son las siguientes:

- La información debe ser refrescada cada semestre. Este tiempo se estima en base al proceso de recolección de datos que realiza el DTICS, además del tiempo promedio de cálculo del modelo tarifario. Esto refiere, por ejemplo, que los hábitos de movilidad de un usuario varían según su horario de trabajo. Siendo así, un profesor varía sus hábitos de movilidad cada ciclo, a diferencia de un administrativo que cambia sus hábitos de movilidad en un año calendario, sin embargo, se toma el tiempo menor de cambios del universo de usuarios.
- La información de las Dimensiones a excepción de fecha y hora debe ser cargada siempre en su totalidad.
- La información para los Hechos debe ser reemplazada en su totalidad, manteniendo un respaldo de los registros antiguos.
- Es necesario un periodo de prueba para optimizar las actualizaciones ya sea con la ayuda de procesos de ETL o de scripts de Python-Pandas.
- El archivo datosAdminXFactor.csv debe ser procesado previamente por el grupo MAS. Esto debido a que este archivo es producto de un proceso interno del grupo, ajeno a los objetivos y procesos de este trabajo de titulación.

La ejecución de las políticas de actualización de datos, debe tomar en cuenta:

- Que todos los registros nuevos pasen por el proceso ETL que los prepara y realiza el mapeo de fechas y horas respectivamente.
- La Dimensión Fecha necesita establecer siempre una fecha fin hasta de dos meses en adelante del último registro que se quiera ingresar.
- Los Hechos vehículo y administrativos necesitan realizar los cruces respectivos y se mantendrán los archivos resultantes de estos cruces para tener un control de la información almacenada.
- Cualquier actualización debe mantener un histórico de respaldos del DW con fecha y hora en que se ha realizado el respaldo.

Procedimiento para nuevos usuarios

Dados los tiempos de actualización de los datos del DW, o fuera de estos tiempos, se puede presentar la situación de un nuevo usuario que tramite la obtención de un lugar en los aparcaderos de la universidad. El nuevo usuario debe presentar los datos detallados en la Tabla 5 en su solicitud para obtener un lugar en el aparcadero.

Variables Nuevo Usuario
Cédula
Edad
Género
Provincia de Residencia
Cantón de Residencia
Dirección de Residencia Actual
Tipo de Servidor (Empleado, Profesor, Investigador, etc.)
Remuneración Salarial
Titularidad (contratado, titular, etc.)
Dedicación (medio tiempo, tiempo completo, etc.)
Campus en el cual labora de manera prioritaria
Núm. Vehículos que posee
Placa vehículo principal
Placa vehículo secundaria
Avalúo aproximado de los vehículos
ID Tarjeta Estacionamiento
Tiene Discapacidad
Tipo Discapacidad
Núm. de integrantes del grupo familiar
Núm. familiares (grupo familiar) que laboran en la Universidad
Tiene Bicicleta
Tiene acceso a estacionamiento de bicicletas
Tiene moto
Placa moto

Tabla 5 Datos necesarios para un nuevo usuario.

Según estos datos, se calcularán los valores de factores y pesos del modelo tarifario. Estos valores serán asignados por MAS. Paralelamente, los datos de la Tabla 5, serán procesados con las técnicas ETL, explicadas en la sección Materiales y Métodos-Sección 4.4.4 Posterior a su procesamiento, se ingresarán estos datos junto a los valores de factores y pesos del modelo tarifario a la Tabla de Hechos Administrativos. Dada la placa vehicular principal, se consumirá la API provista por EcuadorLegalOnline para obtener las características principales del vehículo. Estos datos serán procesados, así mismo, con las técnicas ETL de la sección Materiales y Métodos-Sección 4.4.4 e ingresadas en la Tabla de Hechos Vehículos. En el caso de la Tabla de Hechos Aparcadero, no se

ingresará ningún registro, puesto que esta tabla refiere a las veces que ingresa o sale un vehículo de los aparcaderos, datos que por su dominio se deben registrar en las puertas, según su modalidad de ingreso.

4.5. Crisp-DM

La aplicación de minería de datos se realizó mediante la metodología de CRISP-DM que permite aplicar Clasificación, Clusterización, Predicción de valores o Análisis de dependencias o asociaciones.

Comprensión del Negocio

El dominio del negocio global comprende la integración de fuentes de datos para implementar un DW, se propone aplicar minería de datos. Las técnicas y su objetivo se describen a continuación:

- **Clasificación:** el modelo propuesto, K-NN crea un mecanismo con los datos existentes de entrada y salida (aprendizaje supervisado) detallados en la sección 2 de la metodología de Hefesto, permitiendo comprender los datos y clasificar nuevos ingresos, según diez criterios de agrupamiento, propuesto inicialmente. Este modelo no establece una relación entre variables, sino busca patrones de agrupamiento directamente influidos por una etiqueta o dato de salida. A manera de ejemplo, la clasificación agrupa a los usuarios según los datos administrativos y de vehículo, detallados en la sección 2 de minería de datos. Un nuevo usuario se clasificará de acuerdo al modelo propuesto previamente entrenado. El dato de salida del DataSet, es la variable rango de ingreso, el cual establece como Rango 1, Rango 2 hasta Rango 10 los datos, esta distribución y asignación de variable corresponde a cada tipología de usuario.
- **Clusterización:** el modelo de K-Means valida los criterios de agrupamiento del modelo de clasificación, así se busca confirmar si el agrupamiento en 10 clústeres por rango de ingresos económicos es correcto. Este modelo de clusterización si visualiza la relación entre variables, estas relaciones están basadas en los datos existentes, es decir que, al ingresar nuevos datos, o modificar los datos existentes, los criterios de agrupamiento en “k” grupos y el modelo de clusterización deberán ser consultados nuevamente. Así, el problema a solventar se detalla a continuación
    - Una persona con características de tipo de servidor, titularidad, dedicación, remuneración, discapacidad, edad y género tiene un vehículo con características: clase de vehículo, cilindraje y año de lanzamiento. Se plantea identificar el número de clústeres apropiado para estas características, con la finalidad de proponer distintos valores de factores de cálculo según el número de clústeres que afinen el modelo de cálculo de tarifa diferencial. Ver Figura 44.

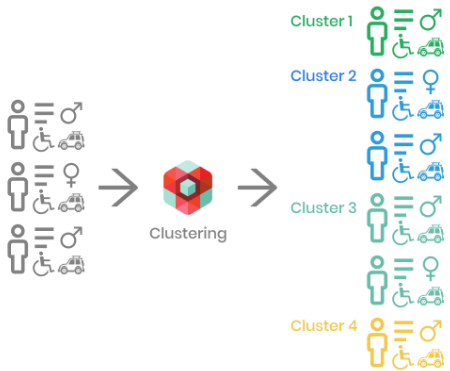


Figura 44 Diagrama del proceso de clustering de servidores según características propias y del vehículo que conduce.

Las fuentes de datos usadas para estos modelos son extraídas de los cubos de datos correspondientes a vehículos y administrativos del DW. Ver Figura 45.

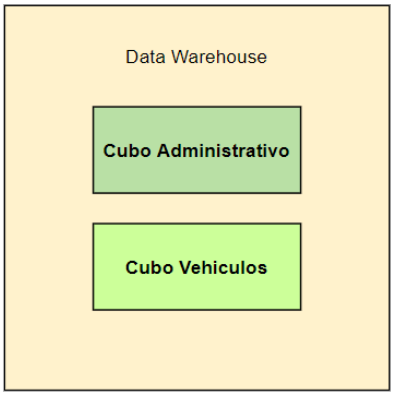


Figura 45 Fuentes de datos para el Data Mining

Comprensión de los Datos

En la implementación de la metodología de Hefesto se ha descrito las fuentes de datos, como se extrajo de dichas fuentes los datos de interés, así como cualquier tratamiento sobre estos para la construcción del DW, por lo tanto, los datos para el DM son tal cual como Hefesto. Estos datos tienen todos sus registros completos, no existen valores nulos en estos DataSet. Finalmente, de nuestras fuentes de datos tomaremos los siguientes valores, en un formato de archivo csv:

- Características de una persona como: tipo de servidor, titularidad, dedicación, remuneración, discapacidad, edad y género.
- Características vehiculares como: clase de vehículo, cilindraje y año de lanzamiento.

Preparación de los Datos

Los datos existentes están completos, es decir no existen valores nulos o vacíos entre estos. Sin embargo, para los modelos de clasificación y clusterización no deben ingresar datos categóricos, por tal se ha planteado las siguientes técnicas de preprocesamiento de datos categóricos.

- **Codificación de enteros o de etiqueta**: A cada valor de categoría único se le asigna un valor entero. Por ejemplo, para un salario: Bajo, Medio y Alto se asignan valores 1, 2 y 3 respectivamente. **Esta técnica requiere una relación ordinal entre registros.**
- **Codificación One-Hot**: Las variables categóricas que no contienen relación ordinal, requiere aplicar la técnica de One Hot Encoding, en la que se los representa binariamente por característica [5]. Por ejemplo, una variable tipo de servidor: administrativo, investigador; necesita de 2 variables binarias: categoria_administrativo, categoria_investigador. Estas variables tendrán valores 1 (True) y 0 (False).

La tabla 6 agrupa a las variables usadas en cada técnica de preprocesamiento de datos categóricos.

Técnicas de preprocesamiento de datos categóricos	
One Hot Encoding	**Codificación de Etiqueta**
Servidor	Remuneración
Titularidad	Año de Vehículo
Dedicación	Edad
Discapacidad	
Clase de vehículo	

Tabla 6 Variables según el uso de técnicas de preprocesamiento de datos categóricos

El modelo clasificatorio hace uso de todas las variables disponibles para su modelamiento, mientras para el clustering *se debe mantener variables correlacionadas esto incrementa la precisión de agrupamiento*. Debido a lo mencionado anteriormente para el modelo de clustering se realizó el siguiente análisis de variables:

- **Análisis de variables**: Los datos y variables del DataSet se representan mediante una gráfica pairplot, la cual establece la distribución en variables individuales y entre dos variables del DataSet. Ver Figura 46.

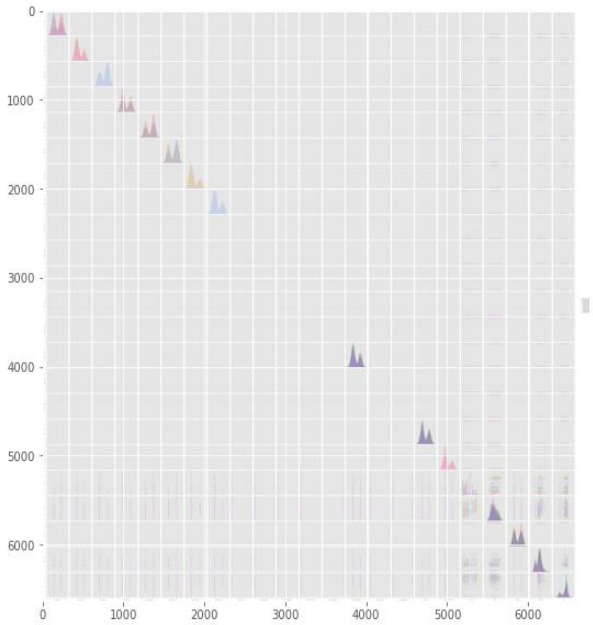


Figura 46 Gráfica Pairplot que establece relaciones entre variables del DataSet

La Figura 46 identifica que las variables con mayor distribución entre sí mismas y con las demás, corresponden a las siguientes características: sueldo, edad, género, cilindraje, año. Estas variables son seleccionadas para el modelo de clusterización como datos de entrada.

Modelado y Validación

Los modelos para la clasificación y clusterización se detallan a continuación. Además, su validación se encuentra en la parte final de los modelos, permitiendo entender los resultados de estos modelos.

Modelo de Clasificación -kNN

A partir del DataSet preprocesado, se divide en datos de entrenamiento y prueba, con la finalidad de medir la precisión del modelo clasificatorio y evitar el overtraining y el overfitting. La distribución de datos corresponde a 75% y 25% respectivamente. Posterior a la separación se normaliza los datos entre 0 y 1 sean categóricos o numéricos, para disminuir los rangos en variables tales como remuneración y cilindraje. Para el entrenamiento del modelo clasificatorio se debe identificar los “k” vecinos apropiados, este “k” se puede identificar elaborando un “Elbow Curve” que identifique la precisión del modelo según el “k” respectivo. Para la obtención de esta curva se ejecutó veinte veces el modelo con distintos “k”. La Figura 47 identifica los “k” entre [1, 9] los cuales obtienen un accuracy entre [65 ,75] %, los cuales, aunque no son óptimos, son aceptables para un modelo clasificatorio.

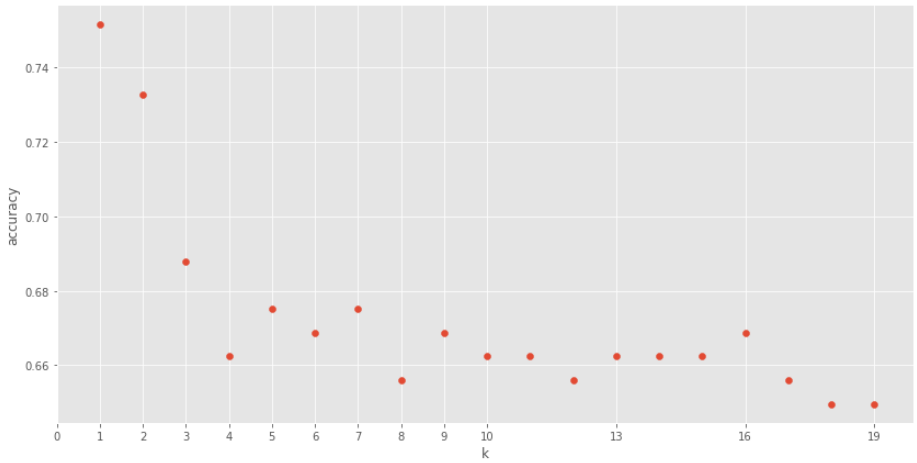


Figura 47 "Elbow Curve", para la identificación del k apropiado para kNN

Identificados los valores para "k" se ejecuta el modelo entre [1,9] para "k", obteniendo los resultados de la Tabla 7.

k	Accuracy Train	Accuracy Test
1	1	0.75
2	0.87	0.73
3	0.82	0.69
4	0.78	0.66
5	0.76	0.68
6	0.74	0.67
7	0.74	0.68
8	0.68	0.66
9	0.69	0.67

Tabla 7 Accuracy de los modelos clasificatorios según el k asignado.

Los valores "k" empleados son 2, 3, 5, debido a mayores accuracy estos valores permitirán entrenar los modelos con buena precisión. El modelo con "k" = 5, obtiene la matriz de confusión visualizada en la Figura 48, en donde se observa pocos falsos positivos.

```
Matriz de Confusion
[[ 9  2  0  0  0  0  0  0  0  0]
 [ 5  7  5  0  0  0  0  0  0  0]
 [ 0  4 20  2  3  0  0  0  0  0]
 [ 0  3  6  2  0  0  0  0  0  0]
 [ 0  0  1  0 27  0  0  1  0  0]
 [ 0  0  0  1  0  1  1  3  0  0]
 [ 0  0  0  0  0  0  0  4  0  0]
 [ 0  0  1  0  1  0  0 26  1  1]
 [ 0  0  0  0  0  0  0  0  5  0]
 [ 0  0  0  0  0  0  0  0  6  9]]
```

Figura 48 Matriz de Confusión para "k" =5

Los modelos clasificatorios se guardan para clasificaciones futuras de nuevos usuarios a los cuales se busque asignar un rango de ingreso económico. Un ejemplo de aplicación de los modelos clasificatorios es un docente titular, con dedicación de tiempo completo, con remuneración salarial de 1200$, de género masculino, con edad de 46 años con auto año 2013 de cilindraje 1400, es clasificado como **Rango 3**.

Modelo de Clusterización K-Means

El proceso de entrenamiento de K-Means, parte de la selección de un "k" óptimo. Esta selección puede realizarse en prueba y error, o usando la técnica de "Elbow Curve". Esta se emplea para encontrar el número de clústeres "k" óptimo mediante una métrica (WCSS, scores, inertia), siendo los más óptimos aquellos "k" que realizan el quiebre de una curva descendente. La métrica elegida es inertia, el cual refiere a la suma de las distancias al cuadrado de las muestras con respecto al centro de agrupación más cercano. Ver Figura 49.

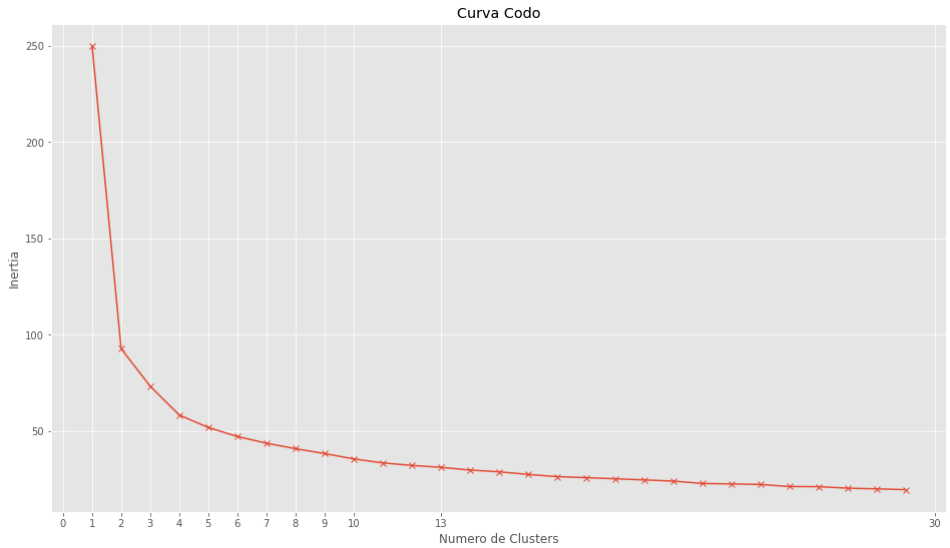


Figura 49 Elbow Curve del modelo de clusterización

La curva de codo indica que con un "k" = 2 y "k" = 4 obtienen clústeres más precisos para los datos existentes. Así, el modelo se entrena con estos k, obteniendo la Figura 50.a para "k" =2 y la Figura 50.b para "k" =4.

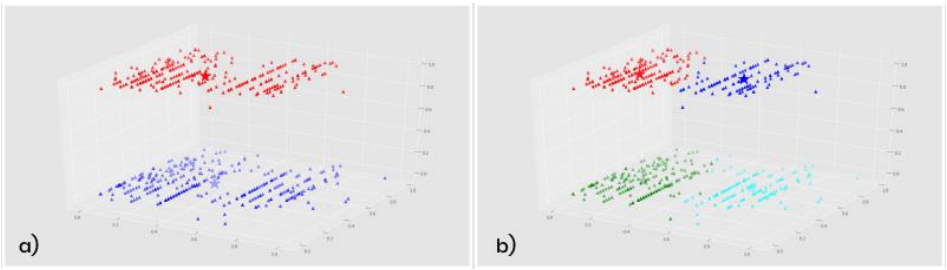


Figura 50 Clusterización de datos, a con k=2 y b con k=4

Validación del modelo KMeans

La Clusterización es un método no supervisado difícil de medir su precisión, sin embargo, se puede validar según el "k" óptimo de la curva de codo con el método de la gráfica de silueta. Esta gráfica permite identificar la agrupación de datos y su distribución, lo cual refiere a un mejor agrupamiento disminuyendo los errores por outliers.

- Los coeficientes de silueta tienen rango de -1 a 1, para este caso se mantiene un rango -0.1 a 1.
- Silhouette Score da el valor promedio para todas las muestras, esto da una perspectiva de la densidad y la separación de la forma agrupada.
- Las puntuaciones de silueta se realizan para cada muestra.

La Figura 51, identifica el análisis de silueta [6] para los distintos "k", esta gráfica de silueta identifica una distribución correcta entre los dos y cuatro clústeres, ya que no están muy cercanos del 0 o 1 indicando que los valores están a una distancia espaciada del centroide, además de que su distribución es correcta entre número de clústeres. Los valores de promedio de score de silueta (Ver Tabla 8) indican una mejor elección el k=2.

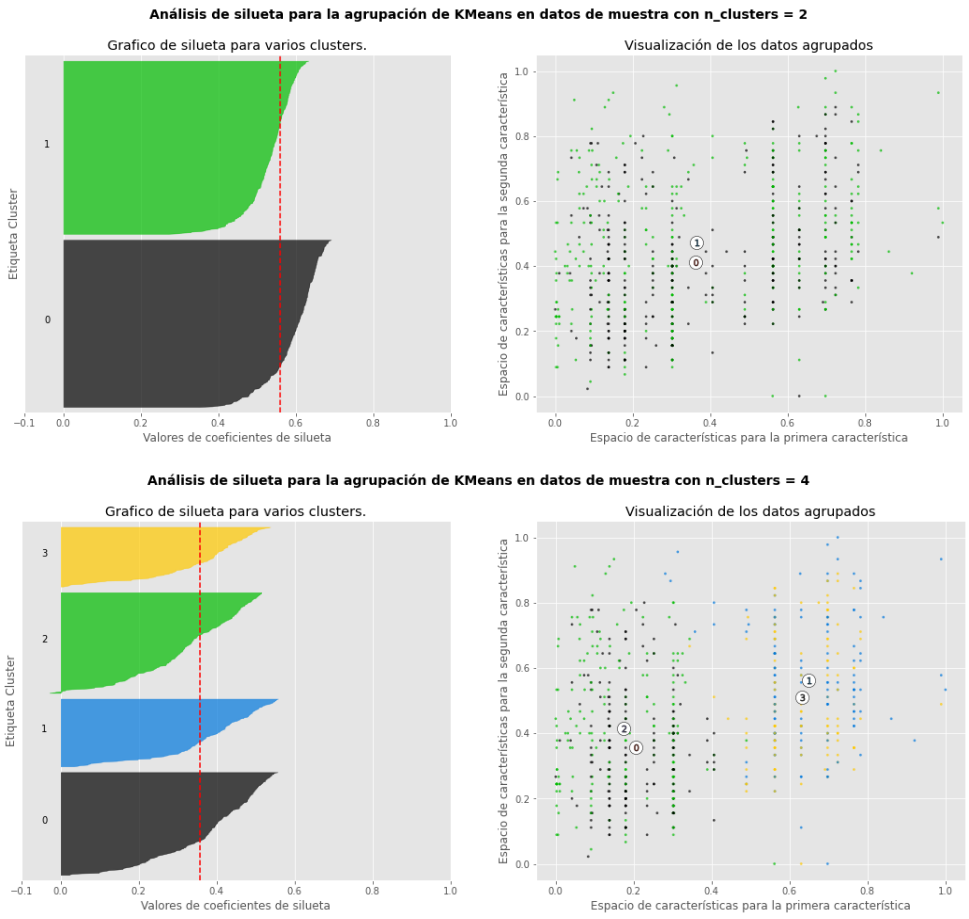


Figura 51Método de análisis de la silueta, para un k=2, y k=4

K	Promedio de silhouette_score
2	0.56048
4	0.35696

Tabla 8 Promedio de precisión de los análisis de silueta

Entre el análisis de score y las gráficas de silueta, se optó por un "k" =4 debido a la distribución de datos entre cada cluster, los valores de score distribuidos alrededor del valor promedio y con fines prácticos el dominio del problema de clusterización.

Así, el modelo entrenado con k=4, se prueba con los siguientes datos (normalizados) remuneración de 650$, edad de 44, masculino, cilindraje del vehículo de 1400 y año de modelo 1971. Estos valores se asignan al Cluster 0. El clustering, evidenció en su etapa de análisis de variables la relación existente entre las características: Remuneración, Edad y Discapacidad, en cuanto al modelo tarifario propuesto por Avila-Ordóñez, et al., (2019). Estas características fueron sugeridas oportunamente al grupo de investigación para su consideración en su fórmula de cálculo.

Análisis de la representatividad de los clusters

El análisis correspondiente al significado de cada clúster, es decir, a que representa cada clúster del modelo de KMeans se realizó mediante un árbol de decisiones analizando las etiquetas resultantes del mencionado modelo. Los resultados son:

Clúster 0

- Hombres, un sueldo promedio entre 641.50$ y 2143$. El grupo etario corresponde a edades menores a 49 años

Clúster 1

- Hombres con sueldo superior a 4653$.
- Mujeres con sueldo superior a 1556 $. El grupo etario corresponde a edades iguales o menores a 39 años y medio.

Clúster 2

- Mujeres con sueldo menor a 2700$. El grupo etario corresponde edades entre 39 años y medio a 61 años y medio

Clúster 3

- Hombres con sueldo menor o igual a 641.50$ y entre 2143$ y 4653$. El grupo etario corresponde a edades mayores a 49 años

4.6. Validación del Modelo Tarifario

Análisis de Pesos Tentativos

Los pesos apropiados para aplicar al modelo matemático implementado deben ser propuestos en base a un objetivo orientado al cobro de la tarifa. Esto refiere, por ejemplo, si el objetivo central del modelo será la recaudación mínima, es decir, que las tarifas cubran los gastos operativos de los aparcaderos, los pesos se ajustarán orientados a esta finalidad. Otro ejemplo, sería una recaudación elevada, es decir, que las tarifas a parte de cubrir los gastos operativos de los aparcaderos, produzcan una ganancia significativa, por tal los pesos se ajustaran a esta finalidad.

Justificado el ajuste de los pesos, se propone analizar las tarifas con diferentes valores, obedeciendo a las características (factores de dedicación, titularidad, cautividad y distancia) de los servidores universitarios. Este análisis debe realizarse cada semestre, al momento de dar mantenimiento al DW, sección Materiales y Métodos-Sección 4.4.4 y contempla los datos válidos de factores de cálculo, eliminando a los registros nulos.

Los pasos a seguir en este análisis de pesos son:

1. Cálculo y análisis de pesos “techo”
2. Análisis de los factores de cálculo a los usuarios de los aparcaderos
3. Propuesta de relaciones entre factores de cálculo
4. Análisis final de los pesos con las relaciones determinadas en el paso 3

La parte inicial de este análisis o paso 1, busca obtener los pesos “techo”, los cuales serán aquellos que, corriendo el modelo matemático, su tarifa máxima no exceda dos tarifas mínimas. La Tabla 9, indica que al tener id tarifa 1 y 2, se cumple el requisito, los demás se descartan.

	Pesos				Tarifa (USD)		
ID Tarifa	Factor Distancia	Factor Cautividad	Factor Dedicación	Factor Titularidad	Mínimo	Máximo	Promedio
1	1	1	1	1	15	21	18.38
2	2	2	2	2	15	27	21.77
3	3	3	3	3	15	33	25.14
4	4	4	4	4	15	39	28.52

5	5	5	5	5	15	45	31.89

Tabla 9 Análisis de tarifas cuando todos los pesos son iguales

Partiendo de la obtención de pesos "techo" se procede a analizar las variaciones entre pesos, relacionando con las características de los factores de los servidores universitarios (paso 2). Las características mencionadas, se extrajo de la Tabla de Hechos Administrativa de la base de datos obtenida mediante el DW Materiales y Métodos Sección 4.4.3. Los datos obtenidos se reflejan en la Figura 52, de estos datos, se puede concluir que un servidor universitario promedio que usa los aparcaderos del Campus Central cumple con las siguientes características: tiene un contrato por servicios profesionales (53%) a medio tiempo (80%), vive a una distancia promedio de 1 km a 5 km del campus central (68%), con un tiempo promedio de viaje de 44 minutos a 55 minutos (50%).

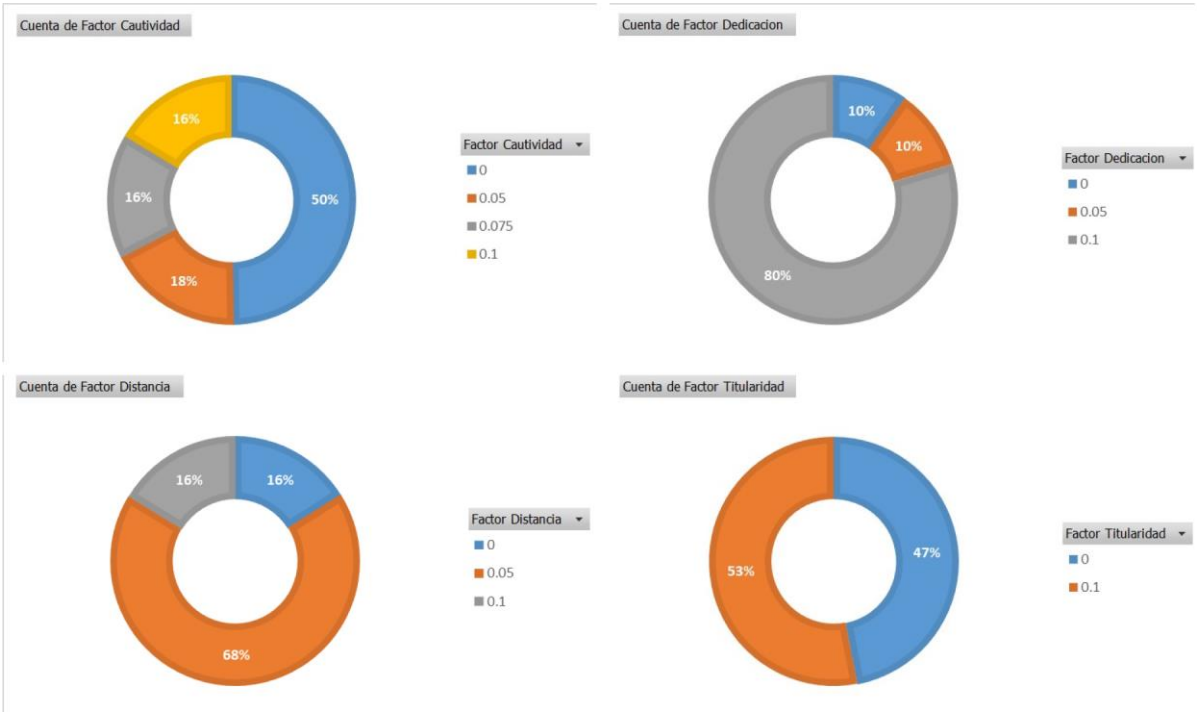


Figura 52 Factores Promedio de un servidor universitario que usa los aparcaderos del campus central.

El análisis planteado previamente, permite ejecutar el paso 3, el cual estima que los factores cautividad y distancia deben ser "penados" bajo los mismos pesos, por sus relaciones implícitas, ya que, a mayor distancia o tiempo de viaje, el servidor será más propenso a usar su vehículo privado. Los factores de titularidad y dedicación deben ser penados independientemente, ya que no mantienen una íntima relación entre sí. Dado este análisis, se procede a hacer una combinación donde los 4 factores, tengan pesos de 1 o 2. El paso 4, se refleja en la Tabla 10 muestra los resultados de correr el modelo

con todas las combinatorias con estos pesos. Estos pesos, en comparación de los casos de estudio de Avila-Ordóñez, et. al. (2019), muestran un incremento generalizado de ~1$ en cualquiera de las combinaciones de pesos. Esta particularidad se debe principalmente al número de servidores universitarios y la actualidad de sus datos, siendo así que los casos de estudio de Avila-Ordóñez, et. al. (2019), tomo en cuenta datos recolectados con corte de 2016, y el presente análisis tomo en cuenta datos con corte de 2020. Finalmente, se estima que los pesos recomendables para el modelo son los correspondientes al:

- ID Tarifa 3, donde el factor distancia, cautividad, dedicación y titularidad tienen pesos 1,1,2,1 respectivamente. Esto ya que cumple la condición de peso techo, relación entre variables y su tarifa promedio muestra un incremento mínimo al promedio de servidores universitarios.
- ID Tarifa 14, donde el factor distancia, cautividad, dedicación y titularidad tienen pesos 2,2,1,2 respectivamente. Estos pesos gravan la tarifa inicial, su rango de recaudación es mayor al anterior mencionado, y la gran mayoría de servidores universitarios no tienen un incremento mayor a 5$ de la tarifa inicial.

Pesos				Tarifa		
Factor Distancia	Factor Cautividad	Factor Dedicación	Factor Titularidad	Mínimo	Máximo	Promedio
1	1	1	1	15	21	18.38
1	1	1	2	15	22.5	18.41
1	1	2	1	15	22.5	18.69
1	2	1	1	15	24	18.92
1	2	1	2	15	24	18.99
1	1	2	2	15	24	19
1	2	2	1	15	24	19.11
2	1	1	1	15	25.5	19.24
2	1	1	2	15	25.5	19.27

1	2	2	2	15	25.5	19.29
2	1	2	1	15	25.5	19.35
2	2	1	1	15	25.5	19.47
2	1	2	2	15	25.5	19.48
2	2	1	2	15	25.5	19.52
2	2	2	1	15	25.5	19.6
2	2	2	2	15	27	19.71

Tabla 10 Análisis de tarifas con pesos diferentes en sus factores

4.7. DashBoard

La creación de los DashBoards dinámicos se realizó mediante Grafana Labs. Dadas las preguntas que se buscan responder se agrupó y separó en nueve DashBoards, que acceden a la DB y mediante Querys en SQL permite al usuario ver la distribución de datos a través de los distintos indicadores. Los tipos de gráficos utilizados son correspondientes a los que Grafana tiene por defecto y otros instalados a través de plugins. Los nueve Dashboards implementados tienen los tópicos:

- Dashboard 0 - Panel General. Este panel, globaliza los datos históricos de ingresos dependiendo de la modalidad de toma de datos: reconocimiento de placas o tarjetas magnéticas. Este dashboard es dinámico, respecto al año de lectura de los datos. Ver Anexo 1.
- Dashboard 1 - Panel con información referente a determinada información de los vehículos que transitan por los parqueaderos del campus central de la Universidad. Este Dashboard es dinámico, es decir varía según: Puertas, Mes, Año y Dia. Ver Anexo 2.
- Dashboard 2 - Panel con información referente al género de nacimiento de los servidores que hacen uso de los parqueaderos del campus central de la Universidad. Este Dashboard es dinámico, es decir varía según: Puerta, Mes, Año, Dia y Género. Ver Anexo 3.
- Dashboard 3 - Panel con información referente a los tipos de servidores que hacen uso de los parqueaderos del campus central de la Universidad. Este Dashboard es dinámico, es decir varía según: Puerta, Mes, Año, Dia y Tipo de Servidor. Ver Anexo 4.

- Dashboard 4 - Panel con información referente a las dependencias donde laboran los servidores universitarios. Este Dashboard es dinámico, es decir varía según: Puerta, Mes, Año, Dia y Dependencia. Ver Anexo 5.
- Dashboard 5 - Panel con información referente a los tipos de discapacidades que enfrentan los servidores universitarios. Este Dashboard es dinámico, es decir varía según: Puerta, Mes, Año, Dia y Discapacidad. Ver Anexo 6.
- Dashboard 6 - Panel con información referente a los vehículos que transitan por el campus central en referencia a su cantón de matrícula. Este Dashboard es dinámico, es decir varía según: Puerta, Mes, Año, Dia y Discapacidad. Ver Anexo 7.
- Dashboard 7 - Panel con información referente a los factores de cálculo. Ver Anexo 8.
- Dashboard 8 - Panel de despliegue de información general de los modelos de Inteligencia Artificial. Además de información específica del modelo de clasificación y de clusterización. Ver Anexo 9.

4.8. Single-page Applications con Angular

Como requerimiento de despliegue del aplicativo web, se definió la necesidad de desarrollar una SPA, mediante Angular v7. Este requerimiento de desarrollo permite la integración con los sistemas de la universidad, además de un constante mantenimiento organizado. En la toma de requerimientos, basados en entrevistas con el equipo de desarrollo de DTICS, se presentó la necesidad de que el aplicativo web sea construido bajo una arquitectura Componentes, Modelos y Servicios. Estos requerimientos fueron desarrollados, de tal manera que la Figura 53 los engloba. Las vistas home y calcular modelo del aplicativo web desarrollado se puede ver en los Anexos 10 y 11.

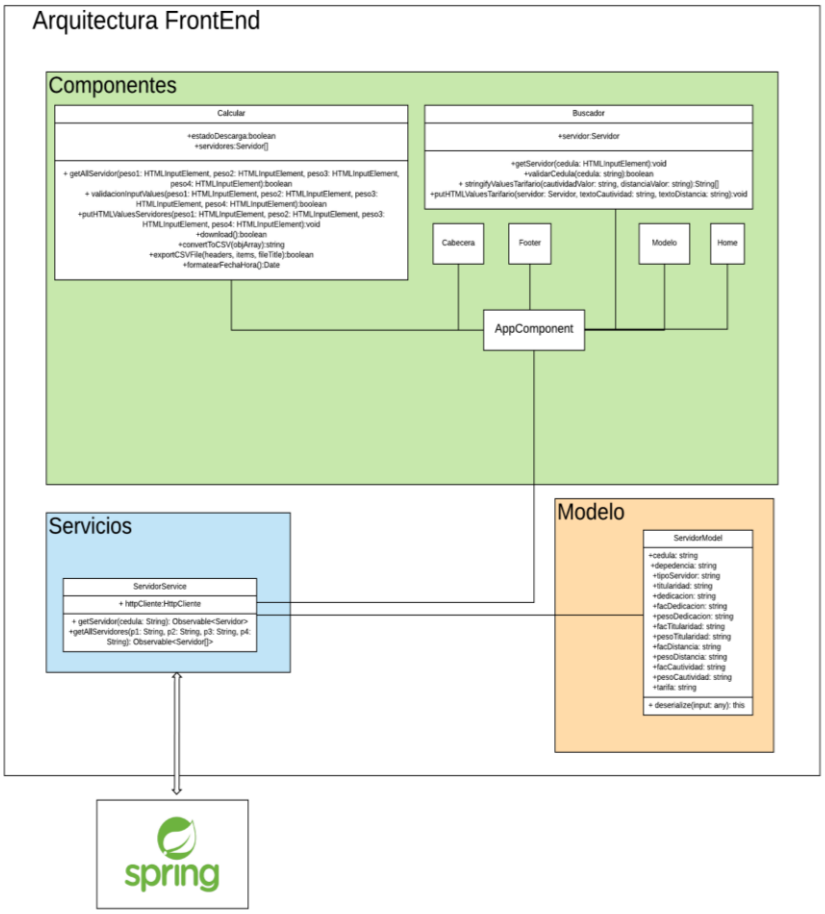


Figura 53 Arquitectura Front End para una SPA

El proyecto está agrupado en 3 paquetes:

Componentes

Los SPA, como se definieron en el marco teórico, sección Marco Teórico-Sección 2.6.1, son aplicaciones de una sola página, es decir, renderiza toda la aplicación al abrirla, y navega a través de secciones o componentes (Kevin, 2019). Los componentes a breves rasgos, son porciones de una página web, que contiene un archivo HTML, CSS Y TS. Su funcionamiento se aísla, y permite la reutilización de código. Estos componentes son orquestados por un componente máster (Kevin, 2017). Los componentes del proyecto se definen a continuación:

- App: Este componente, es el orquestador máster de los demás componentes, al cargarse, renderiza todos los componentes, y su navegabilidad permite su uso bajo determinados casos definidos previamente.
- Buscador: Este componente, refiere a la sección buscador de cédula, el cual contiene un input validado, para buscar la tarifa que debería pagar el titular de la cédula. Este componente hace la validación de captura de datos del usuario y despliegue de datos obtenidos del servidor.
- Cabecera: La componente cabecera, al igual que footer, son estáticos, ya que permite navegar a través de todos los componentes.

- Calcular: Este componente, permite, que dado 4 pesos (Dedicación, Titularidad, Distancia y Cautividad) se presente las tarifas para todos los servidores registrados en la base de datos.
- Contenido: Este componente, es el visualizado en el home, el cual tiene información del proyecto ejecutado.
- Footer: Este componente al igual que la cabecera, son estativos, y visualiza información de interés general, como contactos, página web, etc.
- Modelo: Este componente, detalla el modelo tarifario implementado, además de varios casos de estudio.
- NotFound: Este componente maneja errores ya sea por url erróneas y otros errores que no permiten navegar por los componentes definidos.

Servicios

El paquete servicios, detalla 2 archivos, los cuales hacen las peticiones al servidor mediante el protocolo HTTP, mediante un servicio REST. Estos servicios deben manejar un tipo de modelo definido en el paquete modelos (Kevin, 2019).

- Servidor: Este servicio, hace una petición tipo GET, al lado del servidor, para que devuelva información que se muestra en el componente Buscador. Entre otras cosas el JSON devuelto por el backend contiene nombre, cédula, factores de cálculo y tarifa diferencial.
- Servidores: Este servicio, hace una petición tipo GET, al lado del servidor, para que devuelva información que se muestra en el componente Calcular. Transporta los 4 pesos para calcular las tarifas correspondientes, y devuelve un JSON con la información requerida. Entre otras cosas el JSON devuelto por el backend contiene cédula, factores de cálculo y tarifa diferencial.

Modelos

El paquete de modelos, define 2 archivos, los cuales contienen clases que permiten abstraer correctamente las respuestas a las peticiones del paquete servicios (Kevin, 2017). Los modelos implementados son:

- Servidor: Este archivo define la clase Servidor, con los atributos: cedula, nombres, dependencia, tipoServidor, titularidad, dedicacion, facDedicacion, pesoDedicacion, facTitularidad, pesoTitularidad, facDistancia, pesoDistancia, facCautividad, pesoCautividad, tarifa. Además de una función que deserealiza el JSON devuelto por el backend, en un objeto de esta clase.
- Servidores: Este archivo define la clase Servidores, con los atributos: cedula, facDedicacion, facTitularidad, facDistacancia, facCautividad, tarifa. Además de una función que deserealiza el JSON devuelto por el backend, en un objeto de esta clase.

4.9. Microservicio con Spring Boot

Tomando como referencia la Figura 54 se explica el proceso de creación del microservicio para el cálculo de la tarifa diferencial a partir del uso del modelo matemático propuesto por Avila-Ordóñez, et al., (2019). Además de recalcar el uso del Patrón Facade requerido por DTICS en la creación de un microservicio.

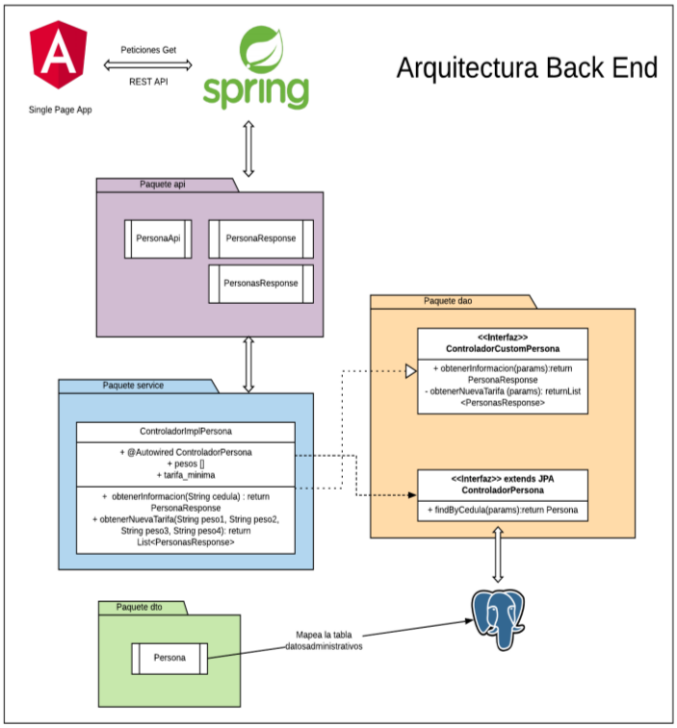


Figura 54 Arquitectura del microservicio

El proyecto se encuentra divido en cuatro paquetes:

Paquete API

Se puede encontrar las siguientes tres clases:

- PersonaApi: clase que representa nuestro API REST y pone a disposición las urls para acceder a sus métodos.
- PersonaResponse: objeto que se utiliza como respuesta ante las solicitudes HTTP.
- PersonasResponse: objeto que se utiliza como respuesta ante las solicitudes HTTP.

Paquete DAO

Se definen las siguientes interfaces para el acceso de datos:

- ControladorPersona: extiende de JpaRepository, permite el acceso a varios métodos genéricos y además el definir queries Jpa personalizadas.
- ControladorCustomPersona: define los métodos a ser implementados por nuestra clase de servicio.

Paquete DTO

Se define una clase que representa una tabla dentro de nuestra base de datos, por lo cual se mapea dicha tabla en una clase y sus columnas en atributos de esa clase.

Paquete SERVICE

La clase dentro de este paquete implementa ControladorCustomPersona, ésta incluye el código que representa la lógica de negocio del microservicio.

El patrón Facade dentro del microservicio se puede apreciar en la Figura 55. Donde la clase PersonaApi actúa como fachada al implementar sus propios métodos que hacen uso de los métodos de la clase ControladorImplPersona. Por lo cual si se involucran otros métodos y tablas en el microservicio es necesario crear una nueva clase ControladorImplObjeto que contenga la lógica y sus métodos propios. Consecuentemente, PersonaApi hará uso de estos nuevos métodos en sus métodos ya definidos.

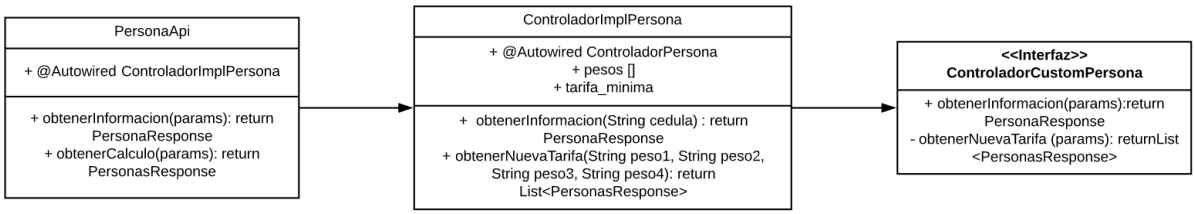


Figura 55 Patrón Facade dentro del microservicio

Capítulo V

Este último capítulo dará a conocer las conclusiones finales que se han obtenido en base a los objetivos planteados al inicio de este trabajo de titulación, las aportaciones que se han logrado con el trabajo realizado, las recomendaciones propuestas y finalmente los trabajos que se podrían efectuar a futuro.

5. Conclusiones

En el desarrollo de este trabajo se han alcanzado los objetivos inicialmente planteados en cuanto a:

Diseñar e Implementar un Data Warehouse que permita integrar y analizar los datos provenientes de los ingresos y salidas del campus central, además de implementar un aplicativo web que consuma un microservicio que calcula la tarifa diferencial de acuerdo al modelo matemático propuesto por el grupo de movilidad

La Metodología Hefesto permite la construcción de Data Warehouse o Data Mart de manera ordenada y secuencial, el desarrollo de esta metodología se presentó en en Materiales y Métodos-Sección 4.4. Los cubos creados y que forman parte de este Data Warehouse brindan a sus usuarios información de su interés y permite el uso de perspectivas para el análisis de datos. Con respecto al análisis de datos, el Data Warehouse presenta la información de manera clara y precisa para su visualización mediante los DashBoards implementados en Materiales y Métodos-Sección 4.5. Estas visualizaciones complementan las etapas de análisis, al consolidar los datos en información que permita tomar decisiones pertinentes en cuanto a ocupación de los parqueaderos y tarifas según las características de los servidores universitarios que hacen uso de estos. Además, el análisis de variables es apegado a las costumbres de movilidad de los servidores universitarios que constantemente cambian, y deben revisarse, para que los parqueaderos puedan gestionarse priorizando la realidad actual de los servidores. A manera de ejemplo, la realidad de 2019 permite crear una tarifa para el 2020, sin embargo, debido a la crisis sanitaria que enfrenta el país en el año corriente, 2020, evidenciará que para el primer semestre 2021, los hábitos de movilidad de la comunidad universitaria variarán, esto como consecuencia de los cambios de personal administrativo y docente, las cargas horarias hacia los docentes titulares, la desvinculación de servidores universitarios con contratos por servicios profesionales u ocasionales, la disminución de las remuneraciones salariales, entre otros factores que se evidenciará en los semestres del 2021.
Asimismo, la etapa de Minería de Datos, permitió extraer información y conocimiento relevante de los datos que se proporcionaron. El conocimiento generado es:

- La necesidad de las características Edad, Discapacidad y Remuneración en el modelo matemático, el análisis de la tarifa diferencial y de los hábitos de movilidad de los servidores universitarios debido a que:
    - El factor **edad** influye directamente en la preferencia de uso de tipo transporte. A mayor edad, los servidores optan por el uso del vehículo privado, esto se demostró en los grupos etarios identificados en base a los datos del Data Warehouse.
    - El factor **discapacidad** influye directamente en la preferencia de uso del vehículo privado, por sobre el transporte público, esto puede intuir a una mejor comodidad en la movilización entre los lugares de residencia y trabajo. Además, de intuir la falta de políticas en la ciudad que den preferencia a las personas con discapacidad en los sistemas de transporte público. Este tópico se detalla en Trabajos Futuros, sección 5.2.
    - El factor **remuneración** influye directamente en el uso de los parqueaderos, a mayor remuneración, el servidor opta por usar un parqueadero privado, en este caso el de la universidad.

    Estas características mostraron una correlación importante entre sí, por lo cual se recomendó oportunamente al grupo de investigación para su uso en la fórmula de cálculo.
- La relación íntima entre las características de movilidad y socioeconómicas, concluyendo, que un servidor universitario con dedicación de medio tiempo es más propenso a dejar su automóvil en los parqueaderos del campus central. Así también, un servidor que vive a una distancia de 1 a 5 km, con un tiempo de viaje de 44 a 55 minutos, prefiere el uso del vehículo, por sobre el transporte público, ya que la relación distancia-cautividad influye en su decisión.

Las técnicas de Inteligencia Artificial, evidenciaron la ventaja de su uso para la toma de decisión será más acertada, en su dominio. Finalmente, el aplicativo web y el microservicio se crearon de acuerdo a los requerimientos de DTICS para su correcta integración con la infraestructura de la Universidad de Cuenca, estos se encuentran desplegados en los servidores de la Universidad y disponibles para su uso inmediato.

Plantear e implementar una infraestructura de software para la integración y análisis de datos provenientes de las fuentes secundarias y de los aparcaderos.

Para lograr este objetivo se hizo uso de herramientas OpenSource del grupo Hitachi Vantara las cuales son utilizadas por la Universidad de Cuenca. Estas herramientas son Pentaho Data Integration (P.D.I.) para los procesos de Extract, Transform and Load. Dichos procesos fueron estandarizados para su apropiada futura aplicación por parte de DTICS, esto orientado a su mantenimiento. Además, mencionar el uso de la librería Pandas y Microsoft Excel como complementos al proceso ETL como tal. La

siguiente herramienta utilizada fue Pentaho Business Intelligence, utilizada para contener los cubos, responder las preguntas propuestas en la sección Materiales y Métodos-Sección 4.4.1 y permitir el acceso a estos mediante una interfaz intuitiva y accesible para usuarios no técnicos. Finalmente, para el análisis de datos se hizo uso del lenguaje de programación Python y las librerías Scikit-learn y Seaborn, para el tratamiento, análisis, modelado y validación de los algoritmos implementados en este trabajo de titulación. Estas herramientas permitieron una curva de aprendizaje mínima y una consecuente implementación exitosa.

Aplicar un proceso de Extract, Transform and Load (ETL) a los datos provenientes de las cuatro puertas de ingreso (12 de abril y Economía), salida (Arquitectura y Filosofía) del campus central, correspondientes a 8 zonas de aparcaderos (Economía, Psicología, entre otros), y correlacionarse con los datos socioeconómicos del cuerpo docente, empleados y trabajadores.

Los procesos ETL permiten compilar los datos a partir de varias fuentes de datos, posteriormente organizarlo y finalmente centralizarse esto se puede observar con detalle en Materiales y Métodos-Sección 4.4.4.1. Los procesos ETL nos permitieron generalizar los registros de ingresos y salidas para que puedan ser tratados como uno solo, además de facilitar los cruces entre la información administrativa y vehicular necesarias para poder construir los cubos pertinentes. Finalmente, los procesos ETL ofrecen solución a problemas de integración con los datos, pero es importante ir más allá del problema como los proveedores de información. Para nuestro caso pudimos identificar dos problemas muy relevantes que pueden afectar a los procesos ETL: *i)* Falta de datos en grandes periodos de tiempo, para nuestro caso cinco meses en el año 2019. *ii)* Falta de estandarización en los archivos entregados por los proveedores de información de los registros de ingresos y salidas, cada archivo entregado contenía diferentes columnas o valores dentro de estos.

Aplicar una o varias técnicas de Machine Learning refiriendo la metodología de Minería de Datos, Crisp-DM, en el proceso de ejecución. Estas técnicas reconocerán variantes de las variables o descartarán las propuestas por el Grupo de Investigación sobre las dinámicas de Movilidad Humana.

Para la Minería de Datos se hizo uso de la metodología CRISP DM, recomendado para proyectos de Inteligencia Artificial, el cual implementa un desarrollo ordenado y secuencial. El uso de esta metodología incrementa el porcentaje de éxito en los objetivos de la etapa de minería de datos, colaborando en la obtención de relaciones entre variables y su consecuente agrupamiento de servidores según las características socioeconómicas y de movilidad, priorizando las variables correlacionadas íntimamente. Las técnicas de Inteligencia Artificial, planteadas en las primeras etapas de CRISP DM,

fueron Clustering y Clasificación, las cuales se desarrollaron mediante los algoritmos de K-NN y K-Means, respectivamente. Estos algoritmos se entrenaron múltiples veces, variando valores específicos de cada algoritmo, validando finalmente su precisión, y entregando aquellos que, ajustados a los datos obtenidos en las etapas de Data Warehouse, presentaron resultados prometedores. El análisis inicial de variables de Clustering evidenció la importancia de las características: Remuneración, Edad y Discapacidad. Estas fueron sugeridas al grupo de investigación para su posible futura incorporación en el modelo matemático de la tarifa diferencial.

Validar el modelo matemático implementado en una infraestructura web, obteniendo un cálculo funcional en consideración de las variables propuestas por el grupo de investigación y/o las variantes producto del Data Mining.

El modelo matemático de Avila-Ordóñez, et. al. (2019), se validó oportunamente con los datos del Data Warehouse, esto, debido a una comparativa realizada en la sección Materiales y Métodos-Sección 4.6.1, donde se aplicó el modelo variando los pesos correspondientes a cada factor de los servidores universitarios que *hacen uso de los parqueaderos*, debido a que los factores son estáticos para cada servidor. Este análisis se realizó usando el aplicativo web generado, permitiendo identificar la necesidad de tener pesos techo, o máximos; estos pesos se estimaron según la necesidad a solventar en la tarifa a cobrar, es decir, si la finalidad es cubrir los gastos operativo o además de cubrirlos, generar ganancias que no afectan mayoritariamente a la economía de los usuarios del aparcadero. En los análisis, se plantearon los dos casos, aplicando además un análisis de variables que demostró la correlación entre distancia y cautividad, las cuales fueron definidas por igual. Los pesos presentaron tarifas máximas, menores al doble de la tarifa actual, mientras el promedio de tarifas calculadas presentó un incremento de menos de 5$ en la tarifa actual, esto evidencia una mínima afectación a la economía de los servidores universitarios, y una mayor recaudación por parte de la Universidad, si se aplicase este modelo.

5.1. Recomendaciones

Con los problemas que se identificaron durante la construcción del Data Warehouse se propone:

- **Inspeccionar el proceso de recolección de datos en las puertas del estudio.**
    - P.E.: Octubre del 2019 las puertas: 12 de abril, Economía y Arquitectura poseen registros en todos sus días. Todo lo contrario, a la puerta de Filosofía que cuenta únicamente con los tres últimos días del mes.
- **Completar los datos administrativos**
    - Uno de los problemas más grandes, fue no poder ejecutar el modelo a todos los servidores universitarios, debido a la falta de datos certeros de residencia en la ciudad (indispensable para el cálculo de los factores de cautividad y distancia).

- **Ampliar los casos de estudio, a otros campus de la UC u otra institución pública.**
    - Obtener un modelo tarifario robusto.
- **Las instituciones involucradas deben cooperar íntegramente en la obtención de los datos.**
    - Evitar el ralentamiento del ciclo de construcción del DW.
- **Exploración con otras técnicas de DM, ampliando casos de estudio.**
    - Validar las conclusiones obtenidas en este trabajo.
- **Integración de nuevas fuentes de datos**
    - Complementen el DW.

5.2. Trabajos Futuros

Como trabajos futuros con respecto a este trabajo de titulación, se debe considerar:

- **Uso de los resultados del modelo de Clusterización para fines de MAS.**
    - PE.: Campañas de Marketing dirigidas de concientización de auto compartido (del Grupo de Investigación) para la optimización de la gestión de parqueaderos de la Universidad de Cuenca.
- **Mantenimiento adecuado a futuro del DW**
    - Datos, Información y Conocimiento obtenido puede ser utilizado por las autoridades de la Universidad para establecer nuevas reglas en torno a los aparcaderos o a conocer patrones de movilidad de la población universitaria que puede conllevar a estudios complementarios.
- **Incremento de la robustez del Modelo matemático.**
    - Añadir más datos históricos.
    - Complementar el modelo en otros campus de la UC.
    - Implementar el modelo en otras instituciones públicas.

Bibliografía

Newman, S. (2015). Building Microservices (1st ed., p. 8). O'Reilly.

Nadareishvili, I. (2015). Microservice architecture (1st ed., p. 6). Sebastopol, CA: O'Reilly Media, Inc.

Tsamboulas, D. A. (2001). Parking fare thresholds: A policy tool. Transport Policy, 8(2), 115–124. https://doi.org/10.1016/S0967-070X(00)00040-8

Andrew Kelly, J., & Peter Clinch, J. (2006). Influence of varied parking tariffs on parking occupancy levels by trip purpose. Transport Policy, 13(6), 487–495. https://doi.org/10.1016/j.tranpol.2006.05.006

Klementschitz, R., Stark, J., & Sammer, G. (2007). Integrating Mobility Management in Land Development Planning with Off-Street Parking Regulations. Journal of Urban Planning and Development, 133(2), 107-113. https://doi.org/10.1061/(asce)0733-9488(2007)133:2(107)

Bryant, D. (2015). Scaling Microservices at Gilt with Scala, Docker and AWS. Retrieved 30 April 2020, from https://www.infoq.com/news/2015/04/scaling-microservices-gilt/

Bramer, M., (2016). Principles of Data Mining. 3rd ed. pp.2-3.

Simićević, J., Vukanović, S., & Milosavljević, N. (2013). The effect of parking charges and time limit to car usage and parking behaviour. Transport Policy, 30, 125-131.

Higgins, D., 1992. Parking taxes: effectiveness, legality and implementation, some general considerations. Transportation 19 (3), 221–230.

Albert, G., Mahalel, D., 2006. Congestion tolls and parking fees: a comparison of the potential effect on travel behaviour. Transport Policy 13, 496–502.

D'Acierno, L., Gallo, M., Montella, B., 2006. Optimisation models for the urban parking pricing problem. Transport Policy 13, 34–48.

Simicevic, Jelena & Milosavljevic, Nada. (2012). Influence of Parking Price on Parking Garage Users' Behaviour. PROMET-Traffic & Transportation. 24. 413-423.

Calthrop, E., Proost, S., Van Dender, K. (2000): Parking policies and road pricing, Urban Studies, Vol. 37, 63–76.

Vuchic, V (1999).: Transportation for livable cities, Centre for Urban Policy Research, New Jersey, United States of America.

Tam, M. L., and Lam, W. H. K. 2004. "Balance of car ownership under user demand and road network supply conditions—Case study in Hong Kong." J. Urban Plann. Dev., 1301, 24–36.

Anastasiadou, M., Dimitriou, D. J., Fredianakis, A., Lagoudakis, E., Traxanatzi, G., & Tsagarakis, K. P. (2009). Determining the parking fee using the contingent valuation methodology. Journal of Urban Planning and Development, 135(3), 116-124.

Moscoso Cordero, M. (2012). Private motor vehicles and the problem of public transport in the historic centres: the case of Cuenca-Ecuador. Estoa, 1(1), 79-93. doi: 10.18537/est.v001.n001.09

Municipalidad de Cuenca (2015), “Plan de movilidad y espacios públicos”, Revista Municipio de Cuenca, Ecuador

Sudhir Kumar Barai (2003) Data mining applications in transportation engineering, Transport, 18:5, 216-223

Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy (Eds.). (1996). Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, USA.

Zoeter, Onno & Dance, Christopher & Clinchant, Stéphane & Andreoli, Jean-Marc. (2014). New algorithms for parking demand management and a city-scale deployment. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 10.1145/2623330.2623359.

Shoup, D. (1997). The High Cost of Free Parking. Journal Of Planning Education And Research, 17(1), 3-20. doi: 10.1177/0739456x9701700102

Garate, V. (2020). Informacion Parqueadero Campus Central [Correo electrónico].

Han, J. (2011). Data mining: concepts and techniques. Burlington: Elsevier Science.

Witten, I. H. (2005). Data mining: practical machine learning tools and techniques. Ámsterdam Boston, MA: Morgan Kaufman.

Romero, C., y Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618.

ISO/IEC/IEEE 42010:2011. (s. f.). Retrieved 22 de abril de 2020, de https://www.iso.org/standard/50508.html

Ávila, E., Cazorla, P., Vanegas, P. (2019). Modelo de Gestión de la zona de aparcamiento en la Universidad de Cuenca (1.a ed., Vol. 1). Cuenca, Ecuador: Universidad de Cuenca.

Ingeno, J. (2018, agosto 1). Software Architect 's Handbook. Retrieved 22 de abril de 2020, de https://learning.oreilly.com/library/view/software-architects-handbook/9781788624060/

Nadareishvili, I., Mitra, R., McLarty, M., & Amundsen, M. (2016, agosto 1). Microservice Architecture. Retrieved 22 de abril de 2020, de https://learning.oreilly.com/library/view/microservice-architecture/9781491956328/

Subramanian, H., & Raj, P. (2019, enero 15). Hands-On RESTful API Design Patterns and Best Practices. Retrieved 22 de abril de 2020, de https://learning.oreilly.com/library/view/hands-on-restful-api/9781788992664/

Márquez, V. (2018, octubre 29). ¿Qué es exactamente Machine Learning? Retrieved 22 de abril de 2020, de https://medium.com/latinxinai/qué-es-exactamente-machine-learning-77441201a65b

Rençberoğlu, E. (2019, abril 1). Técnicas fundamentales de ingeniería de características para el aprendizaje automático. Retrieved 22 de abril de 2020, de https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114

Zheng, A., & Casari, A. (2018, abril 1). Feature Engineering for Machine Learning. Retrieved 22 de abril de 2020, de https://learning.oreilly.com/library/view/feature-engineering-for/9781491953235/

Goonewardana, H. (2019, febrero 28). PCA: Application in Machine Learning. Retrieved 22 de abril de 2020, de https://medium.com/apprentice-journal/pca-application-in-machine-learning-4827c07a61db

Kimball, R., & Ross, M. (2013, julio 1). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition. Retrieved 22 de abril de 2020, de https://learning.oreilly.com/library/view/the-data-warehouse/9781118530801/

Dragoni, N., Giallorenzo, S., Lluch Lafuente, A., Mazzara, M., Montesi, F., Mustafin, R., & Safina, L. (2016, June 16). Microservices: yesterday, today, and tomorrow. Retrieved May 1, 2020, from https://arxiv.org/pdf/1606.04036v1.pdf

Pisani, M., García, N., Miraballes, R., & Llambias, G. (2016, April 1). Aplicación de Microservicios sobre una arquitectura SOA con restricciones de calidad de servicio. Retrieved May 1, 2020, from https://www.researchgate.net/publication/324681050_Aplicacion_de_Microservicios_sobre_una_arquitectura_SOA_con_restricciones_de_calidad_de_servicio

Shastri, S., & Mansotra, V. (2019, May 1). KDD-Based Decision Making: A Conceptual Framework Model for Maternal Health and Child Immunization Databases. Retrieved May 1, 2020, from https://www.researchgate.net/publication/333276602_KDD-Based_Decision_Making_A_Conceptual_Framework_Model_for_Maternal_Health_and_Child_Immunization_Databases

Shylaja B S, (2015). From Navigation to Star Hopping: Forgotten Formulae. Retrieved Jul 23, 2020, from https://www.ias.ac.in/article/fulltext/reso/020/04/0352-0359

Hitachi Vantara Editors, (2019). Pentaho Data Integration. Retrieved Jul 23, 2020, from https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform/pentaho-data-integration.html

Bernabeu, D., & Mattío, G. (2017). Introducción DATA WAREHOUSING: Marco Conceptual HEFESTO: Metodología Data Warehouse. 182.

NumFOCUS, (2020). Pandas. Retrieved Jul 23, 2020, from https://pandas.pydata.org

Hitachi Vantara Editors, (2018). Pentaho Schema Workbench. Retrieved Jul 23, 2020, from https://help.pentaho.com/Documentation/8.2/Products/Schema_Workbench

Grafana Labs Editors, (2020). Grafana 7.0. Retrieved Jul 23, 2020, from https://grafana.com

Apache Editors, (2020). Microservices. Retrieved Jul 23, 2020, from https://spring.io/microservices

Sucar, L. E. (2015, August 23). Métodos de Inteligencia Artificial [Diapositivas]. Scribd. https://es.scribd.com/document/429503370/MetIA-09

Elbow method (clustering). (n.d.). In Wikipedia. Retrieved August 24, 2020, from https://en.wikipedia.org/wiki/Elbow_method_(clustering)

Bonaros, B. (2020, August 19). K-Means Elbow Method code for Python –. Retrieved August 24, 2020, from https://predictivehacks.com/k-means-elbow-method-code-for-python/#:%7E:text=The%20Elbow%20method%20is%20a,its%20assigned%20center(distortions)

Ramirez, J. (2018, December 25). K-means: Elbow Method and Silhouette - Jonathan Ramirez. Retrieved August 24, 2020, from https://medium.com/@jonathanrmzg/k-means-elbow-method-and-silhouette-e565d7ab87aa

scikit-learn. (2015, January 1). Selecting the number of clusters with silhouette analysis on KMeans clustering. Retrieved August 24, 2020, from https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

Silhouette (clustering). (n.d.). In Wikipedia. Retrieved August 24, 2020, from https://en.wikipedia.org/wiki/Silhouette_(clustering)#:%7E:text=The%20silhouette%20value%20is%20a,poorly%20matched%20to%20neighboring%20clusters

Narkhede, S. (2019, August 29). Understanding Confusion Matrix - Towards Data Science. Retrieved August 24, 2020, from https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62?gi=164be2eabf3f

Romero J. (2020, April 27). Metodologías de Minería de Datos. Retrieved August 24, 2020, from https://jorgeromero.net/metodologias-de-mineria-de-datos/

Saquicela, V. (2015, 9 octubre). Técnicas de Data Mining [Diapositivas]. https://www.ucuenca.edu.ec

Rouse, M. (2019, August 26). Web application (Web app). Retrieved August 24, 2020, from https://searchsoftwarequality.techtarget.com/definition/Web-application-Web-app

What is front-end development? (2009, September 28). Retrieved August 24, 2020, from https://www.theguardian.com/help/insideguardian/2009/sep/28/blogpost

Nicole. (2017, May 23). Qué es Frontend y Backend. Retrieved August 24, 2020, from https://platzi.com/blog/que-es-frontend-y-backend/

Dhaduk, H. (2020, June 10). Best Frontend Frameworks of 2020 for Web Development. Retrieved August 24, 2020, from https://www.simform.com/best-frontend-frameworks/

W3C. (2014, June 11). HTTP - Hypertext Transfer Protocol Overview. Retrieved August 24, 2020, from https://www.w3.org/Protocols/

Arranz, J. M. (2015, September 21). The Single Page Interface Manifesto. Retrieved August 24, 2020, from http://itsnat.sourceforge.net/php/spim/spi_manifesto_en.php

UCuenca, C. U. (2018, June 24). Manual de Imagen Institucional. Retrieved August 24, 2020, from https://issuu.com/comunicacionuniversidaddecuenca/docs/manual_de_imagen_u_cuenca

Kevin. (2019, April 18). Angular 7 models. Retrieved August 24, 2020, from https://medium.com/swlh/angular-7-models-cd0cd80f5e33

Kevin. (2017, September 7). Working with models in Angular. Retrieved August 24, 2020, from https://nehalist.io/working-with-models-in-angular/

Enriquez C. (2016, July 16). Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM. Retrieved August 24, 2020.

Flanagan D. (2006, May 27). JavaScript - The Definitive Guide. 5th ed., O'Reilly, Sebastopol. Retrieved August 24, 2020

Santamaria J. (2015). The Single Page Interface Manifesto. Retrieved August 24, 2020, from http://itsnat.sourceforge.net/php/spim/spi_manifesto_en.php

Arranz J. (2010, September 7). Tutorial: Single Page Interface Web Site With ItsNat. Retrieved August 24, 2020, from https://dzone.com/articles/tutorial-single-page-interface

Bootstrap (framework). (n.d.). In Wikipedia. Retrieved August 24, 2020, from https://es.wikipedia.org/wiki/Bootstrap_(framework)

Chapman P, et. al. (2020) CRISP-DM 1.0. (n.d.). Retrieved August 24, 2020, from https://the-modeling-agency.com/crisp-dm.pdf

Anexos

Anexo 1 Panel General

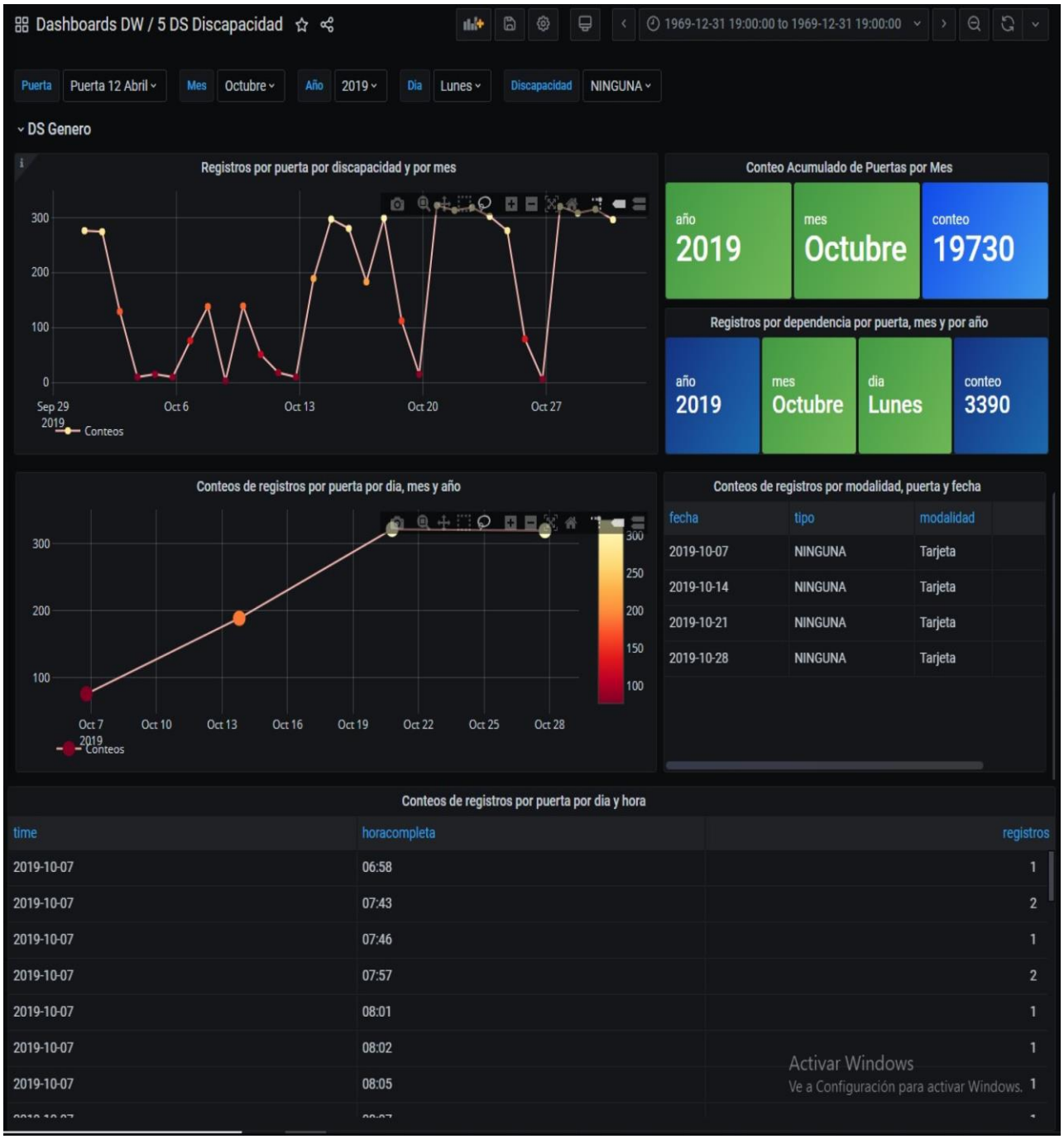
Anexo 2 Panel con información referente a los vehículos que transitan por los parqueaderos del campus central de la Universidad

Anexo 3 Panel con información referente al género de nacimiento de los servidores que hacen uso de los parqueaderos del campus central de la Universidad

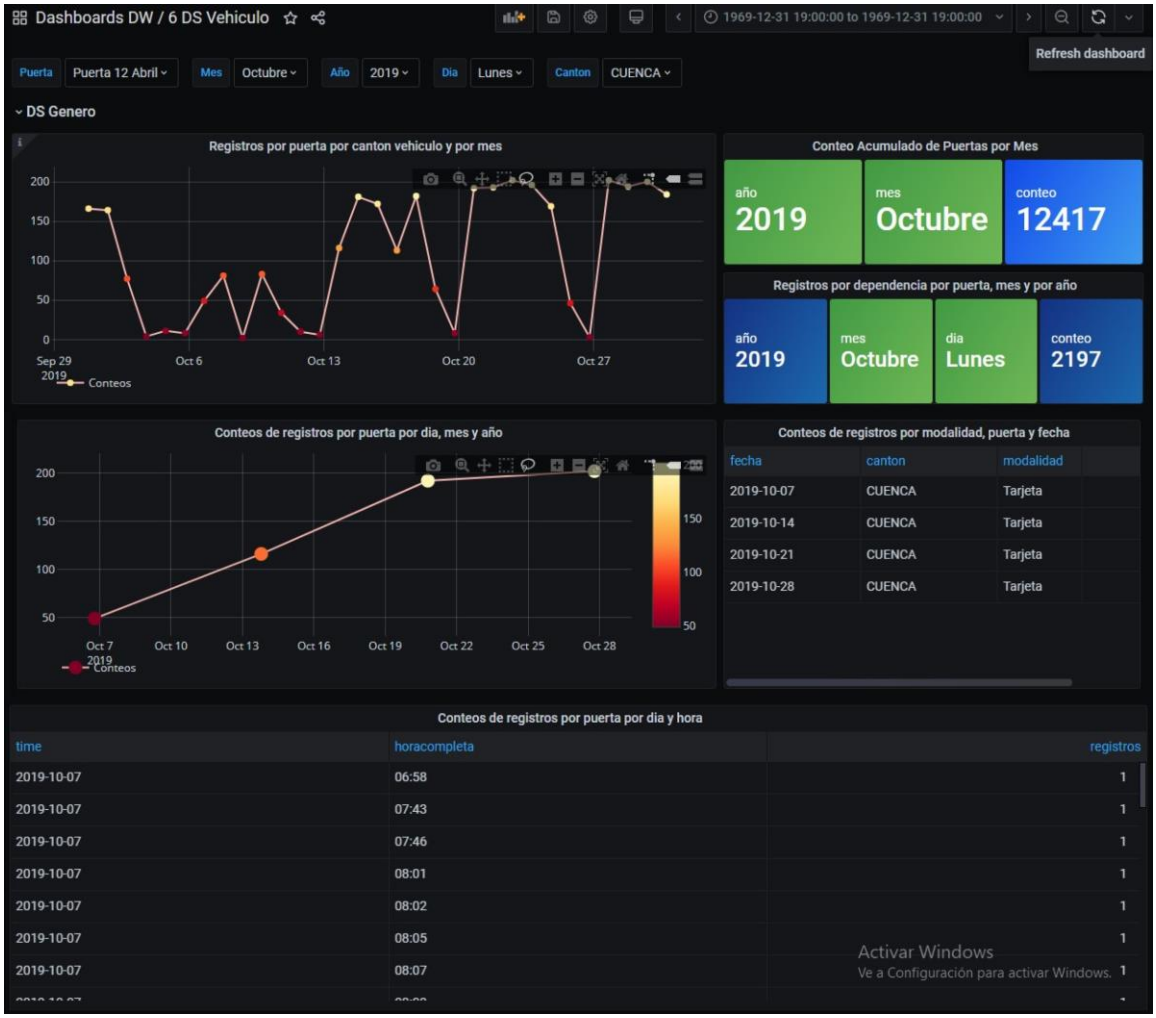
Anexo 4 Panel con información referente a los tipos de servidores que hacen uso de los parqueaderos del campus central de la Universidad

Anexo 5 Panel con información referente a las dependencias donde laboran los servidores universitarios

Anexo 6 Panel con información referente a los tipos de discapacidades que enfrentan los servidores universitarios

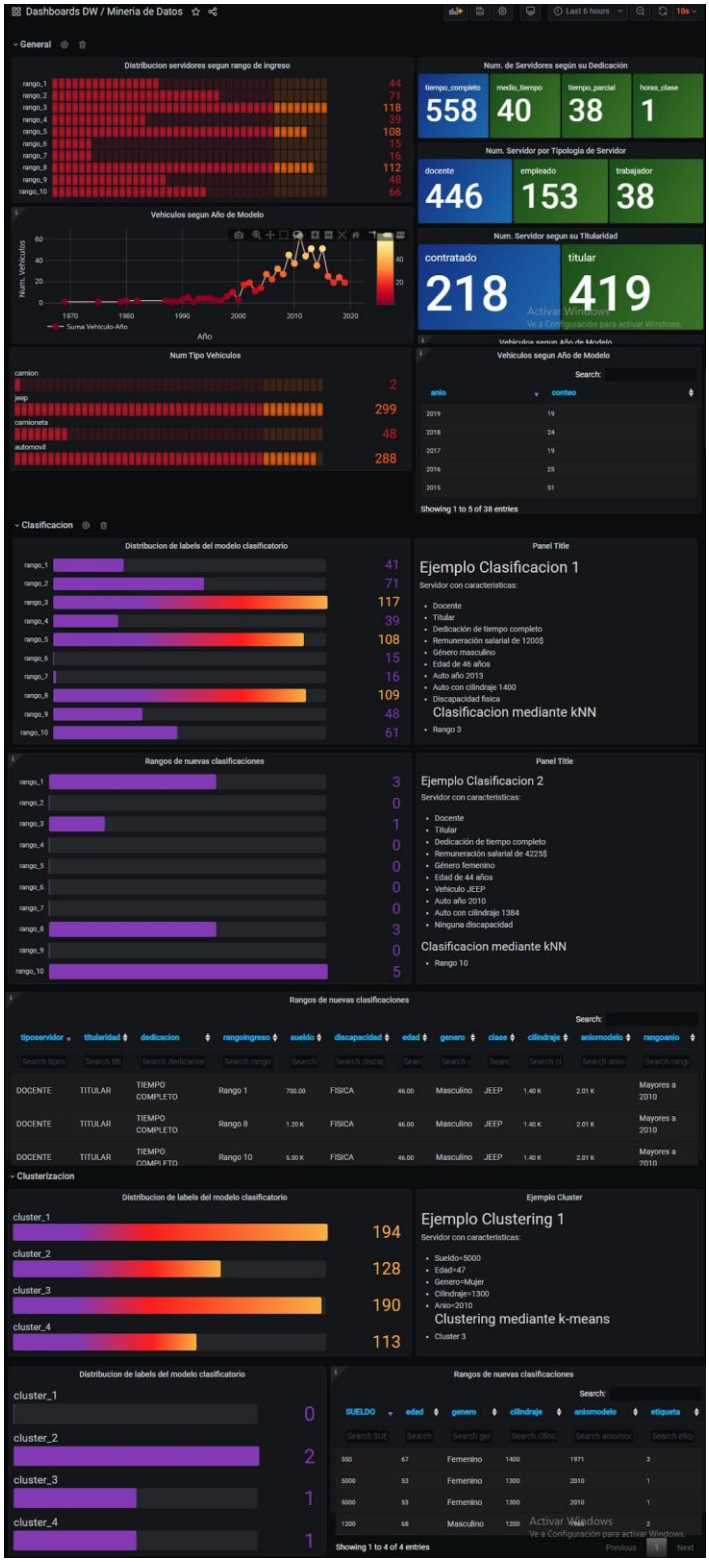


Anexo 7 Panel con información referente a los vehículos que transitan por el campus central en referencia a su cantón de matrícula.

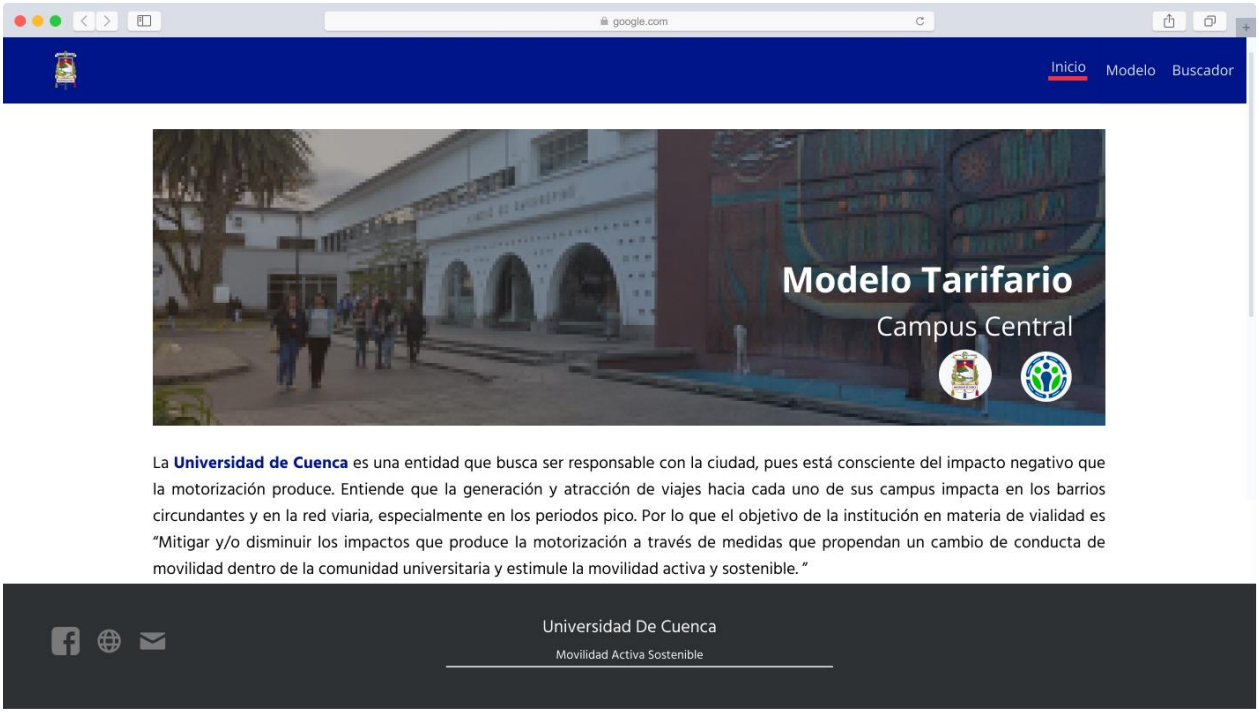


Anexo 8 Panel con información referente a los factores de cálculo

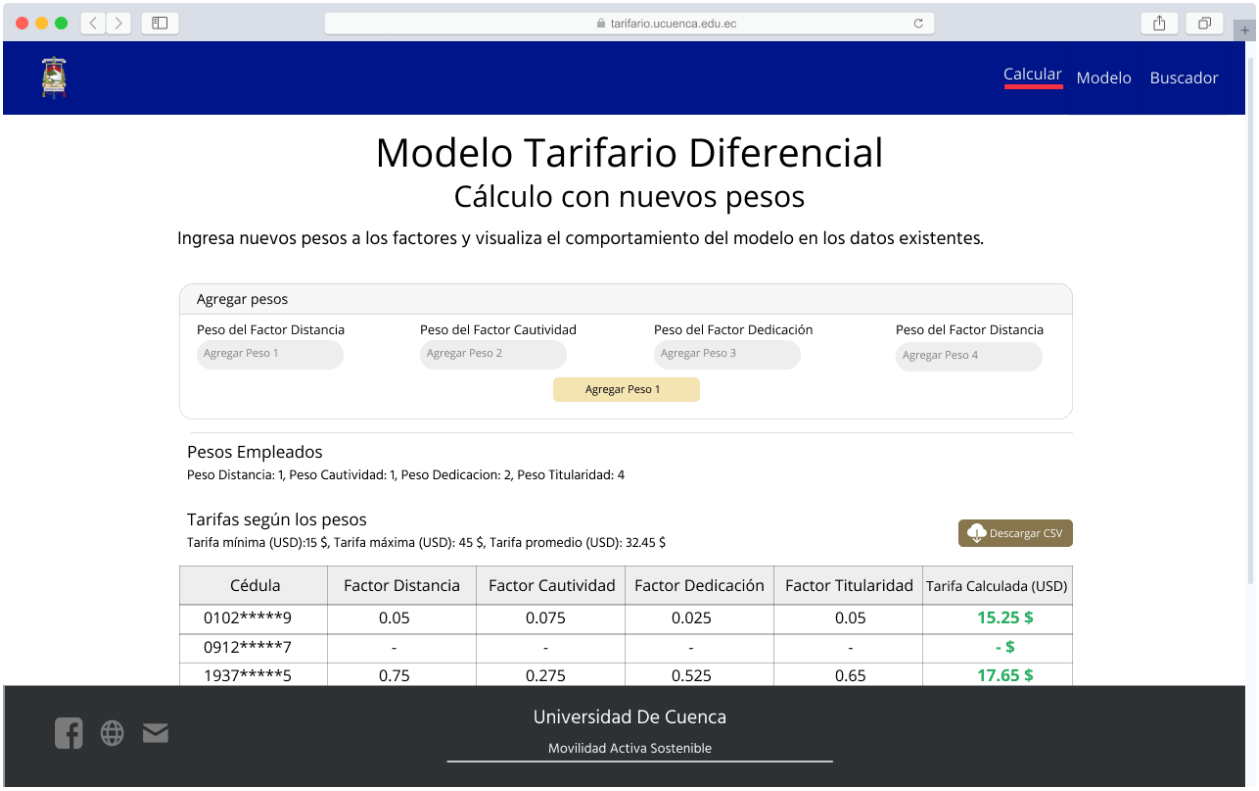
Anexo 9 Panel de despliegue de información de los modelos de Inteligencia Artificial



Anexo 10 Vista Home aplicativo web



Anexo 11 Vista Calcular Modelo aplicativo web



Anexo 12 Manual complementario para el Data Warehouse desarrollado

1. Acerca del Manual

El presente manual complementa al desarrollo de la metodología de Hefesto en el capítulo IV del trabajo de titulación. El principal objetivo de este documento, es la documentación correcta de los procesos de ETL y las consecuentes bases de datos. Esta documentación facilitará el mantenimiento propuesto en la sección Materiales y Métodos Sección 4.4.4, la cual corresponde a la actualización semestral de los datos.

2. Visión Global

Esta sección, permitirá al lector preparar las herramientas necesarias para desarrollar todo el proceso de mantenimiento del Data Warehouse.

a. Especificaciones

Es necesario el uso de:

- Pentaho Data Integration
- Microsoft Excel
- Python y la librería Pandas
- PostgreSQL v12 como SGBD

3. Procesos Requeridos

Los procesos definidos en las siguientes secciones y subsecciones corresponden a la integración de datos, desde las distintas entidades proveedoras de datos. Los pasos ordenados incluyen: La preparación y carga de datos, en base al modelo conceptual ampliado de la sección 4.2 – Análisis OLTP. Esta sección incluye: Carga de las Dimensiones Fecha, Hora, Puerta, Modalidad, Clase, Cilindraje, Cantón, Servidor, Titularidad, Dedicación, Dependencia, Discapacidad, Género, Año vehicular, Remuneración y Edad, además de la carga de tablas hecho vehículo. Posterior a esta carga de datos, se realiza el tratamiento sobre los archivos de ingresos y salidas y datos administrativos, para las tablas de hecho administrativo y estacionamiento.

a. Preparación y carga de datos

En esta sección se explica en secuencia los pasos de preparación de datos para el DW. Los pasos a continuación detallados serán complementados con los siguientes pasos. Por tal, se recomienda inicialmente la lectura de esta sección y a continuación la implementación de esta sección.

1. **Obtención de datos vehiculares**

La obtención de datos vehiculares se realizó mediante un script en Python que hace uso de la librería Request y consume un servicio de la página web EcuadorLegalOnline. El archivo de entrada, debe ser semejante al de la Figura 1, debe siempre cumplirse la regla de tener tres valores alfabéticos y 4 valores numéricos, sin ningún carácter especial. Esto, debido a las limitantes de consultas que provee el servicio web consumido para la obtención de los datos.

placa
ABD6496
ABD3276
PBA5196

Figura 1 Plantilla del archivo entrada para el Script de obtención de datos vehiculares

2. **Carga de Dimensión Fecha**

Los pasos necesarios para este proceso son:

1. Creación de archivo Fecha en formato xlsx

Este archivo se debe crear manualmente, contiene únicamente la columna "fecha", la cual se llena con todas las fechas a partir de la más antigua de los registros hasta la fecha más reciente al mantenimiento que se realice. El formato debe mantenerse como DD/MM/AAAA como se aprecia en la Figura 2.

fecha
24/6/2019
25/6/2019
26/6/2019
27/6/2019
28/6/2019

Figura 2 Formatos fecha para el archivo fecha.xlsx

2. Creación del archivo dimFecha.csv

Utilizar el proceso ETL que se puede visualizar el Figura 3, lo que produce como salida el archivo dimFecha.csv.

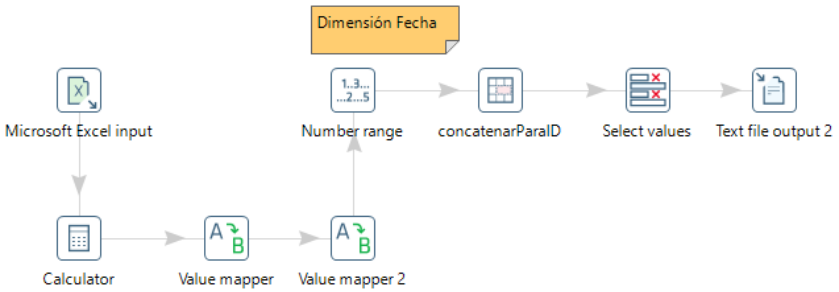


Figura 3 Proceso ETL de la Dimensión Fecha

3. Cargar la dimensión fecha en la base de datos.

Para cargar el archivo producto del paso anterior, en la base de datos del DW, se debe realizar el proceso de la Figura 4.

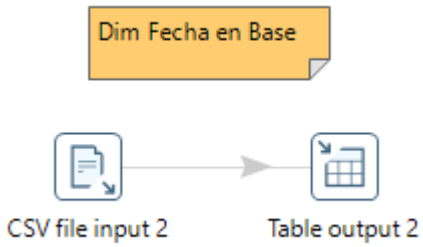


Figura 4 Proceso de almacenamiento en la BD de Dimensión Fecha

3. Carga de Dimensión Hora

Los pasos necesarios para la creación y carga de la tabla Dimensión hora son:

1. Creación de archivo Hora en formato xlsx

Esta sección crea un archivo Excel que contiene las columnas: hora, minuto y horacompleta en formato 24 horas. Las columnas hora y minuto deben respetar el formato "00" y para horacompleta el formato será general. Este archivo se puede apreciar en la Figura 5.

hora	minuto	horacompleta
00	00	00:00
00	01	00:01
00	02	00:02
00	03	00:03
00	04	00:04
00	05	00:05
00	06	00:06

Figura 5 Archivo hora.xlsx

2. Creación del archivo dimHora.csv.

Este paso realiza un proceso ETL, el cual se visualiza en la Figura 6. La salida de este proceso es el archivo dimHora.csv.

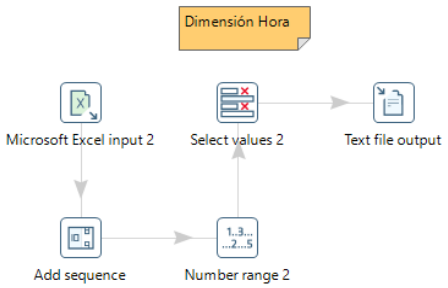


Figura 6 Proceso ETL para la Dimensión Hora

3. Almacenamiento en la Base de Datos.

   Carga el archivo producto del paso anterior, en la base de datos del DW, mediante el proceso de la Figura 7.

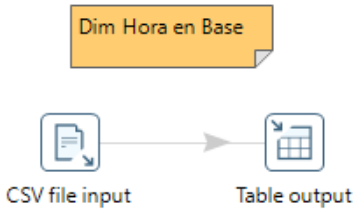


Figura 7 Proceso de almacenamiento en la BD de Dimensión Hora

4. **Carga de Dimensiones Puerta, Modalidad, Clase, Cilindraje, Cantón, Servidor, Titularidad, Dedicación, Dependencia, Discapacidad, Género.**

Previo a los procesos en adelante definidos, es importante puntualizar que los datos que se almacenan en cada archivo dependen de los datos suministrados por los entes recolectores de datos. Los pasos necesarios para este proceso son:

1. Creación de archivos csv y carga de cada dimensión
   a. Puerta

      El primer paso es la creación de un archivo csv con tres columnas: idPuerta, puerta y sentido. Estas puertas, ya sea de entrada o salida, corresponden a los puntos de recolección de datos, pueden variar según las necesidades futuras de la universidad. Ver Figura 8. Este archivo se almacena en la base de datos del DW, como se indica en la Figura 9.

idPuerta	puerta	sentido
1	Puerta 12 Abril	Entrada
2	Puerta Economia	Entrada
3	Puerta Arquitectura	Salida
4	Puerta Filosofia	Salida

Figura 8 Plantilla del archivo dimPuerta.csv

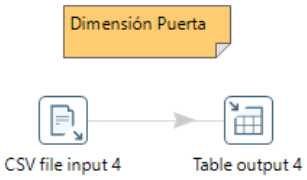


Figura 9 Proceso de carga de Dimensión Puerta

b. Modalidad

Esta subsección crea un archivo csv con dos columnas: idModalidad y modalidad. Estas difieren, en las técnicas que se utilicen en la recolección de datos. La plantilla de este archivo se puede visualizar en la Figura 10. Este archivo, debe ser almacenado en la BD, mediante el proceso de la Figura 11.

idModalidad	modalidad
1	Tarjeta
2	Camara

Figura 10 Plantilla del archivo dimModalidad.csv

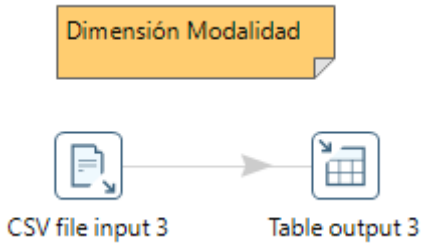


Figura 11 Proceso de carga de Dimensión Modalidad

c. Clase

Esta subsección crea un archivo csv con dos columnas: idClase y Descripcion. Este archivo dependerá de los tipos de vehículos que se permitan ingresar en los estacionamientos, además del tipo de vehículo, obtenido mediante el API REST de la ANT. Ver Figura 12. Este archivo se almacena en la base de datos, como se visualiza en la Figura 13.

idClase	Descripcion
1	AUTOMOVIL
2	CAMION
3	CAMIONETA
4	JEEP
5	MOTOCICLETA
6	OMNIBUS
7	VEHICULO UTILITARIO

Figura 12 Plantilla del archivo dimClase.csv

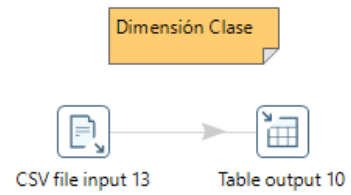


Figura 13 Proceso de carga de Dimensión Clase

d. Cilindraje

Esta subsección crea un archivo csv con dos columnas: idCilindraje y valor. Estos corresponden a los obtenidos mediante el API REST de la ANT. Ver Figura 14. El archivo anterior será la entrada del proceso de la Figura 14, el cual almacena en la BD.

idCilindraje	valor
1	0
2	100
3	115
4	125
5	150

Figura 14 Plantilla del archivo dimCilindraje.csv

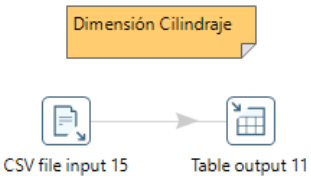


Figura 15 Proceso de carga de Dimensión Cilindraje

e. Cantón

Esta subsección crea un archivo csv con tres columnas: idCanton, PROVINCIA y CANTON. Al igual que las dos subsecciones anteriores, estos datos se obtienen mediante el API REST de la ANT. Ver Figura 16. El archivo obtenido se almacena en la BD, como se visualiza en la Figura17.

idCanton	PROVINCIA	CANTON
1	AZUAY	CUENCA
2	AZUAY	GIRON
3	AZUAY	GUALACEO
4	AZUAY	NABON
5	AZUAY	PAUTE
6	AZUAY	PUCARA
7	AZUAY	SAN FERNANDO
8	AZUAY	SANTA ISABEL
9	AZUAY	SIGSIG

Figura 16 Plantilla del archivo dimCanton.csv

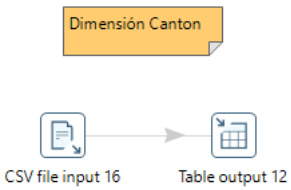


Figura 17 Proceso de carga de Dimensión Cantón

f. Servidor

Esta subsección crea un archivo csv con dos columnas: idServidor y tiposervidor. Ver Figura 18. Este archivo es la entrada del proceso de la Figura 19.

idServidor	tiposervidor
1	DOCENTE
2	EMPLEADO
3	TRABAJADOR

Figura 18 dimTipoServidor.csv

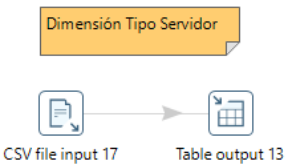


Figura 19 Proceso de carga de Dimensión Tipo Servidor

g. Titularidad

Esta subsección crea un archivo csv con dos columnas: idServidor y tiposervidor. Ver Figura 20. El archivo anterior será la entrada del proceso de la Figura 21.

idTitularidad	titularidad
1	CONTRATADO
2	TITULAR

Figura 20 Plantilla del archivo dimTitularidad.csv

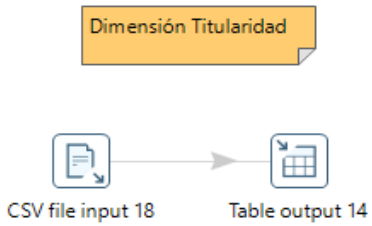


Figura 21 Proceso de carga de Dimensión Titularidad

h. Dependencia

Esta subsección crea un archivo csv con dos columnas: iddependencia y dependencia. Ver Figura 22. Este archivo es la entrada del proceso de la Figura 23.

iddependenc	dependencia
1	AUDITORÍA INTERNA
2	AULA DE DERECHOS HUMANOS
3	CENTRO DE DOCUMENTACION REGIONAL JUAN BAUTISTA VASQUEZ
4	COMISION DE EVALUACION INTERNA
5	COORDINACION ADMINISTRATIVA
6	COORDINACION FINANCIERA

Figura 22 Plantilla del archivo dimDependencia.csv

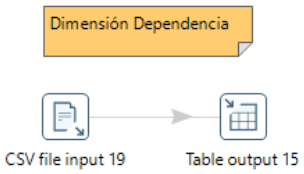


Figura 23 Proceso de carga de Dimensión Dependencia

i. Discapacidad

Esta subsección crea un archivo csv con dos columnas: iddiscapacidad y tipo. Ver Figura 24. Este archivo es la entrada del proceso de la Figura 25.

iddiscapacidad	tipo
1	AUDITIVA
2	FISICA
4	NINGUNA
3	VISUAL

Figura 24. Plantilla del archivo dimDiscapacidad.csv

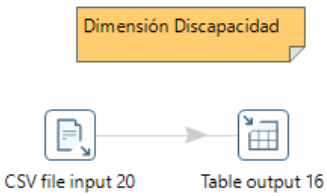


Figura 25. Proceso de carga de Dimensión Discapacidad

j. Género

Esta subsección crea un archivo csv con dos columnas: idgenero y género. Ver Figura 26. Este archivo es la entrada del proceso de la Figura 27.

idgenero	genero
0	Masculino
1	Femenino

Figura 26 Plantilla del archivo dimGenero.csv

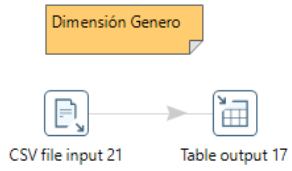


Figura 27 Proceso de carga de Dimensión Género

5. Carga de Dimensión Año vehicular, Remuneración y Edad

Previo a los procesos en adelante definidos, es importante puntualizar que los datos que se almacenan en cada archivo dependen de los datos suministrados por los entes recolectores de datos. Los pasos necesarios para este proceso son:

1. Crear archivos excel y carga de cada dimensión
    a. Año Vehicular

        Esta subsección crea un archivo xlsx con dos columnas: idAnio y anioModelo Ver Figura 28. Este archivo es la entrada del proceso de la Figura 29, el cual genera un archivo salida, listo para almacenar en la BD como se visualiza en la Figura 30.

idAnio	anioModelo
1	1969
2	1971
3	1973
4	1975

Figura 28 Plantilla del archivo AnioVehicular.xlsx

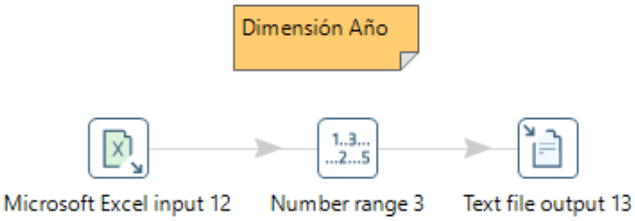


Figura 29 Proceso para generar la Dimensión Año

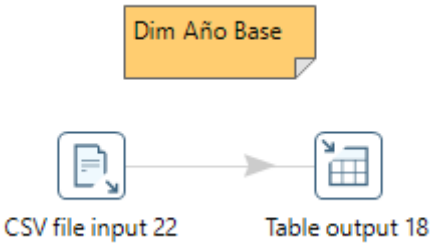


Figura 30 Proceso de carga de la Dimensión Año

    b. Remuneración

        Esta subsección crea un archivo xlsx con dos columnas: idRemuneracion y valor. Ver Figura 31. Este archivo es la entrada del proceso

de la Figura 32. El cual debe ser almacenado en la BD, mediante el proceso de la Figura 33.

idRemuneracion	valor
1	550
2	561
3	566
4	578
5	585
6	596
7	600

Figura 31 Plantilla del Archivo Remuneracion.xlsx

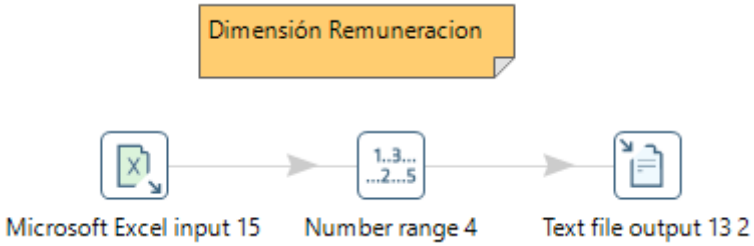


Figura 32 Proceso para generar la Dimensión Remuneración

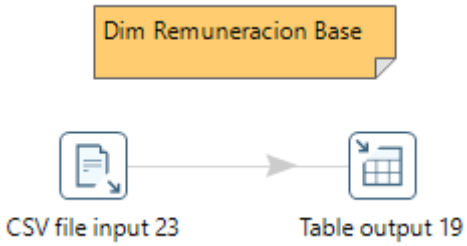


Figura 33 Proceso de carga de Dimensión Remuneración

c. Edad

Esta subsección crea un archivo xlsx con dos columnas: idEdad y Edad. Ver Figura 34. Este archivo es la entrada del proceso de la Figura 35, el cual será almacenado en la BD como se visualiza en el proceso de la Figura 36.

idEdad	Edad
1	0
2	11
3	25
4	27

Figura 34 Plantilla del archivo Edad.xlsx

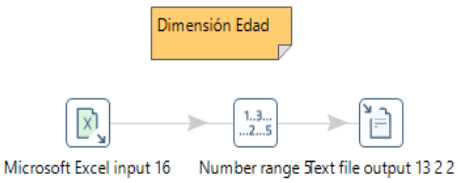


Figura 35 Proceso para generar la Dimensión Edad

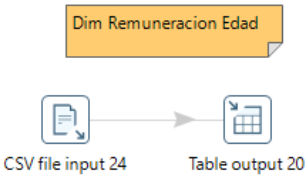


Figura 36 Proceso de carga de Dimensión Remuneración

6. Tratamiento sobre los archivos de ingresos y salidas

Esta sección vamos hace uso de los archivos facilitados por DTICS por lo que es **primordial** respetar las columnas y sus nombres, se deberá seguir el siguiente orden:

1. Archivos de ingresos y salidas

   Estos 5 archivos, de los cuales un archivo pertenece a ingresos por la Av. 12 de abril mediante la detección de la placa (cámara) y los restantes son dos archivos de ingreso (Av. 12 de abril y Economía) y dos de salida (Arquitectura y Filosofía) mediante el uso de la tarjeta. En la Figura 37 se aprecia sus columnas, nombres y formatos originales entregados.

Fecha	Hora	puerta	Tipo	ID	Funcion	Resultado
22/1/2020	16:17:59	1	placa	PSO0720	Llega	Acceso
22/1/2020	16:22:37	1	placa	ABF2269	Llega	Acceso
22/1/2020	16:23:05	1	placa	ABF2269	Llega	Acceso

Camara Av. 12 de Abril

ID de Tarjeta	Hora de Ingreso	Fecha de Ingreso	puerta
61626	18_25_43	24-06-19	1
22988	18_28_07	24-06-19	1
13095	13_22_27	27-06-19	1

Tarjeta Av. 12 de Abril

ID de Tarjeta	Hora de Ingreso	Fecha de Ingreso
61626	18_25_43	24-06-19
22988	18_28_07	24-06-19
13095	13_22_27	27-06-19

Tarjeta Economia

ID de Tarjeta	Hora de Ingreso	Fecha de Ingreso
61626	18_25_43	24-06-19
22988	18_28_07	24-06-19
13095	13_22_27	27-06-19

Tarjeta Arquitectura

ID de Tarjeta	Hora de Ingreso	Fecha de Ingreso
26784	16_36_01	06-09-19
17041	16_58_56	06-09-19
16999	17_11_42	06-09-19

Tarjeta Filosofía

Figura 37 Archivos de Ingreso y Salidas

2. Utilizar el proceso ETL detallado en la Figura 38. Cada proceso devolverá un archivo csv las cuales poseen las mismas columnas, con la diferencia lógica de la puerta a la cual representa.

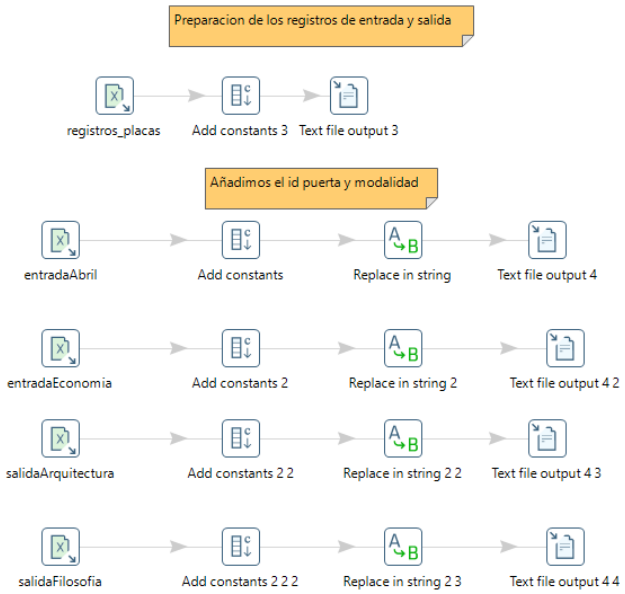


Figura 38 Procesos ETL de los archivos de ingresos y salidas

3. En cada archivo salida del paso 2, es necesario hacer un mapeo de las fechas y horas con la finalidad de ya tener listo las dimensiones fecha y hora. Además, se juntan los archivos de tarjeta en un solo archivo. Ver Figura 39.

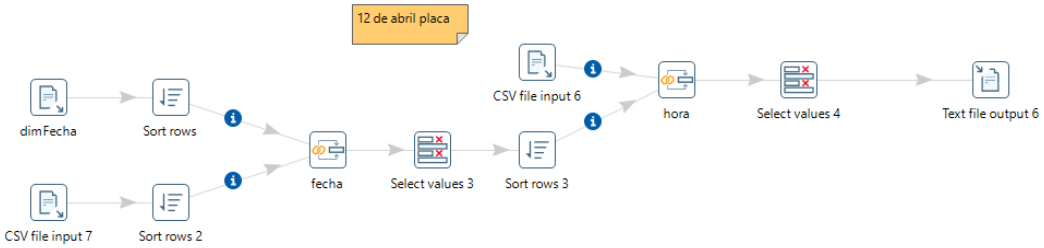


Figura 39. Ejemplo de mapeo de campos

7. Tratamiento sobre los datos administrativos

Todo este proceso se realiza mediante un script de Python. El archivo administrativo debe poseer los campos placa y tarjeta completos caso contrario se descartan. Esto se debe a su importancia para hacer el cruce entre los datos vehiculares y los datos administrativos para la creación de sus cubos respectivos.

8. Carga de Hecho Vehículo

El script que utiliza es el llamado “cruces.py” y la función llamada “hechoVehiculo”. En la Figura 40 se observa todos los procesos necesarios para este Hecho de orden de Izquierda a Derecha.

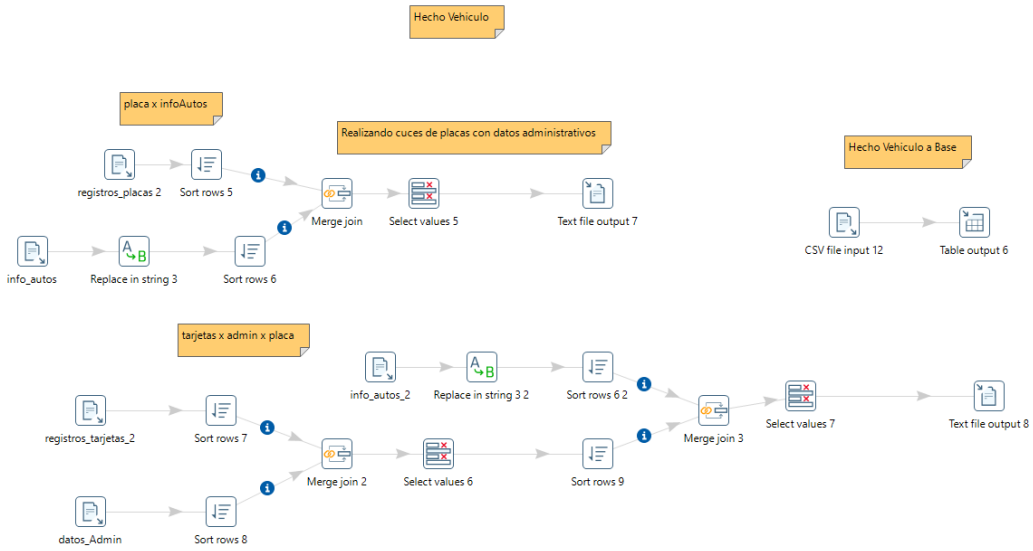


Figura 40 Proceso para la carga del Hecho Vehículo