



## ESTRATEGIAS DE GOBIERNO Y CALIDAD DEL DATO

**Álvaro Muñoz Caballero**

**Tutora: Marta Cabedo Fabrés**

Trabajo Fin de Grado presentado en la Escuela Técnica Superior de Ingeniería de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Graduado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Curso 2020-21

Valencia, 30 de septiembre de 2021



## Resumen

El gobierno de datos consiste en la capacidad de una organización para gestionar el conocimiento que tiene sobre su información. El gobierno de datos proporciona un enfoque holístico para administrar, mejorar y aprovechar la información de forma que la empresa pueda ganar percepción y generar confianza en decisiones y operaciones empresariales.

Para que los datos e información generada por una empresa permita alcanzar sus objetivos y tener una visión global acertada, estos datos tienen que ser fiables y de alta calidad.

En este trabajo se mostrarán las distintas partes que constituyen el Gobierno de Datos, y se analizarán diferentes estrategias orientadas a verificar la Calidad de Datos. Se realizará un estudio de un caso real de gobierno de datos y se establecerá un procedimiento para llevar a cabo de forma exitosa el proceso.

## Resum

El govern de dades consisteix en la capacitat d'una organització per a gestionar el coneixement que té sobre la seua informació. El govern de dades proporciona un enfocament holístic per a administrar, millorar i aprofitar la informació de manera que l'empresa pugua guanyar percepció i generar confiança en decisions i operacions empresarials.

Per a que les dades i informació generada per una empresa permeta aconseguir els seus objectius i tenir una visió global encertada, aquestes dades han de ser fiables i d'alta qualitat.

En aquest treball es mostraran les diferents parts que constitueixen el Govern de Dades, i s'analitzaran diferents estratègies orientades a verificar la Qualitat de Dades. Es realitzarà un estudi d'un cas real de govern de dades i s'establirà un procediment per a dur a terme de forma exitosa el procés.

## Abstract

Data governance is the ability of an organization to manage the knowledge it has about its information. Data governance provides a holistic approach to managing, improving, and leveraging information so that the business can gain insight and build confidence in business decisions and operations.

For the data and information generated by a company to achieve its objectives and have an accurate global vision, this data must be reliable and of high quality.

In this work the different parts that constitute the Data Government will be shown, and different strategies aimed at verifying Data Quality will be analyzed. A real case study of data governance will be carried out and a procedure will be established to carry out the process successfully.



## Índice

Capítulo 1. Motivación, objetivos y organización del proyecto .....	3
1.1 Motivación .....	3
1.2 Objetivos .....	3
1.3 Organización .....	3
Capítulo 2. Introducción .....	4
2.1 Información y conocimiento .....	4
2.2 Activo empresarial .....	4
2.3 Ciclo de vida de los datos.....	4
2.4 Gestión de datos .....	5
2.5 Responsabilidad compartida .....	5
Capítulo 3. Gestión de Datos .....	6
3.1 Gobierno de datos.....	7
3.2 Arquitectura de datos .....	7
3.2.1 Modelo de datos empresarial.....	8
3.3 Desarrollo de datos.....	9
3.4 Operaciones de datos.....	9
3.5 Seguridad de datos .....	10
3.5.1 Reglamento General de Protección de Datos (RGPD).....	10
3.6 Datos Maestros y de Referencia.....	11
3.7 Almacén de datos e inteligencia de negocios.....	11
3.8 Contenidos y documentación .....	12
3.9 Metadatos .....	13
3.10 Calidad de datos .....	13
Capítulo 4. Ejemplo práctico .....	15
4.1 Gobierno de datos.....	15
4.2 Arquitectura del dato.....	15
4.2.1 Modelo de datos conceptual .....	15
4.2.2 Modelo de datos lógico .....	17
4.2.3 Modelo de datos físico .....	17
4.2.4 Generación de DDL .....	18
4.3 Operaciones de datos.....	19
4.3.1 Área de Administración .....	19
4.3.2 Área de Recepción .....	23
4.3.3 Área de Taller.....	23
4.3.4 Área Económica .....	28



4.4	Seguridad de datos .....	32
4.5	Datos maestros y de referencia.....	32
4.6	Almacén de datos e inteligencia de negocio .....	32
4.7	Calidad de datos .....	34
4.7.1	Perfilado de datos .....	34
4.7.2	Reglas de calidad.....	36
4.7.3	Resultados de las reglas de calidad .....	41
4.7.4	Resultados de las reglas de calidad .....	42
Capítulo 5.	Conclusión.....	44
Capítulo 6.	Bibliografía.....	45



## Capítulo 1. Motivación, objetivos y organización del proyecto

### 1.1 Motivación

La idea de este proyecto surge a raíz del auge del teletrabajo en las empresas debido a las restricciones puestas en marcha por la pandemia de COVID-19 que estamos viviendo. Esto, sumado a la necesidad de digitalización que tienen la PYMEs de nuestro país hace necesario que se empiece a trabajar con modelos de datos organizados y accesibles desde cualquier parte en cualquier momento.

### 1.2 Objetivos

El objetivo planteado en este documento es la explicación de que son los datos empresariales y las diversas funciones relacionadas con el gobierno del dato.

Todo esto se aplicará a un ejemplo práctico de una empresa en la que trabaja con datos locales y se le aplicarán los distintos pasos para que tenga una mayor disponibilidad, integridad, calidad y exactitud de estos.

### 1.3 Organización

El proyecto está organizado en primer punto de introducción, con la información relevante a los datos como activo empresarial. En el segundo punto tendremos la gestión del dato donde se explica la función de cada una de las actividades presentes. En el tercer punto, tendremos el ejemplo práctico donde aplicaremos lo desarrollado en los puntos anteriores y para finalizar, en el cuarto punto, tendremos las conclusiones del proyecto.

## Capítulo 2. Introducción

### 2.1 Información y conocimiento

Según la RAE un dato es la información dispuesta de manera adecuada para el tratamiento por una computadora.

En este contexto, los datos son la representación de hechos ya sea por texto, números, gráficos, imágenes u otros formatos que son capturados o almacenados.

La información son datos en un contexto (como podría ser un formato o una representación del tiempo), sin él no tienen significado para negocio ni se podría determinar su importancia.

Cuando utilizamos esta información y establecemos el contexto adecuado, contribuye a la comprensión y a su vez al conocimiento.

Una vez tenemos el conocimiento junto con la experiencia nos permitirá una toma adecuada de decisiones, es decir, tener una sabiduría para el manejo y uso correcto de los datos.

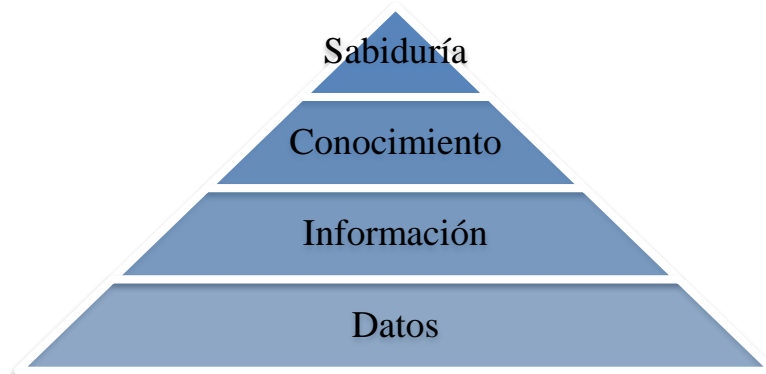


Figura 1. Relación entre datos, información, conocimiento y sabiduría.

[12] Mario G. Piattini Velthuis, Ismael Caballero Muñoz-Reja, Ana Isabel Gómez Carretero, Fernando Gualo Cejudo, Jorge Merino García, Bibiano Rivas García, “Calidad de Datos” Editorial Ra-Ma, 2018

### 2.2 Activo empresarial

Los datos y la información generada a partir de ellos son reconocidos como activos de la empresa y a partir de ellos se pueden alcanzar los objetivos establecidos.

Las organizaciones dependen de sus datos para tomar decisiones, pero también para ofrecer mejores servicios, tener un mayor conocimiento de sus clientes y para guiar sus operaciones tácticas o estratégicas. Para todo ello se necesitan datos fiables y de alta calidad. La gestión de datos puede proporcionar y controlar con eficacia los activos de datos.

[1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 25-26

### 2.3 Ciclo de vida de los datos

Como cualquier activo los datos tienen un ciclo de vida.

Estos son creados (o adquiridos), almacenados, mantenidos, utilizados y en algunos casos, destruidos. Durante este ciclo también pueden ser extraídos, exportados o importados, migrados, validados, editados, actualizados, limpiados, transformados o guardados.

Estos datos se pueden almacenar de forma estructurada como en bases de datos, archivos, documentos. También se pueden almacenar de manera desestructurada como correos electrónicos, documentos en papel, hojas de cálculo...



estén almacenados de una manera o de otra, los datos solo aportan valor al negocio cuando estos son utilizados.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 28-29

**Figura 2. Ciclo de vida de los datos.**

## 2.4 Gestión de datos

La gestión del dato es la función de planificación, control y entrega de activos de la información. Habitualmente se utiliza el termino en inglés Data Management.

Esta función incluye:

- Disciplinas de desarrollo, ejecución y supervisión
- Planes, políticas, proyectos, prácticas y procedimientos
- Controlar, proteger, entregar y mejorar
- Valor de los datos y activos de la información

El alcance y la escala de su aplicación varían con el tamaño, los medios disponibles y la experiencia que tenga la organización. Todo ello hace que la implementación de la gestión de los datos sea diferente, aunque su naturaleza sigue siendo la misma en todos los casos.

También se asegura de esta manera un mejor gestión e integridad de los datos.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 28-29

## 2.5 Responsabilidad compartida

La Gestión de Datos es una responsabilidad compartida entre los profesionales que se encargan de trabajar con los datos y aquellas personas que custodian los datos representando los intereses de la empresa. Estos son expertos en los datos a tratar y se encargaran de asegurar de que su calidad y el uso de los datos sea el adecuado.

Respecto a los profesionales de la gestión de los datos realizarán las funciones técnicas necesarias para salvaguardar y facilitar el uso de los datos donde sean necesarios.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 29-30

## Capítulo 3. Gestión de Datos

Según propone la principal organización internacional para gestores de datos (DAMA) en su libro DMBOK, la gestión de datos, Data Management en inglés, consiste en la planificación y ejecución de planes o proyectos que controlan y mejoran el valor de los datos y la información.

Su misión es satisfacer la disponibilidad de datos, la calidad, integridad y exactitud de estos para tener un mejor valor efectivo de la información que contienen. También asegurar el uso no adecuado garantizando su privacidad y confidencialidad.

Los objetivos de la gestión de datos deben ser específicos, medibles, ejecutables, realistas y con un plazo de tiempo especificado.

Para ello se establecen unas funciones que ayudaran a la gestión de los datos:

- 1- **Gobierno de Datos:** Establece control a través de la gestión de los activos.
- 2- **Arquitectura de Datos:** Define la organización y diseño de planos maestros.
- 3- **Desarrollo de Datos:** Diseñar, implementar y mantener soluciones para satisfacer necesidades de datos.
- 4- **Operaciones de Datos:** Planificación, control y soporte de los activos de datos durante toda su vida útil.
- 5- **Seguridad de Datos:** Establecimiento de políticas y procedimientos de seguridad.
- 6- **Datos Maestros y de referencia:** Actividades para asegurar la coherencia de los datos.
- 7- **Data Warehouse e Inteligencia de Negocio:** Tareas para proporcionar toma de decisiones, informes o análisis.
- 8- **Contenidos y Documentación:** Actividades para almacenar datos que se encuentran fuera de una base de datos.
- 9- **Metadatos:** Permitir el fácil acceso a los metadatos integrados.
- 10- **Calidad de Datos:** Tareas que aplican técnicas de calidad para evaluar y mejorar los datos.



Figura 3. Funciones de gestión de datos

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 30-31



### 3.1 Gobierno de datos

El gobierno de datos ejerce la autoridad y control, con la planificación y aplicación, de los datos activos. Para ello se incluyen organismos y normas para los métodos de toma de decisiones.

Para cumplir con esa finalidad, se definen y aprueba una estrategia para la gestión, procedimientos y arquitectura de datos, asegurándose de su cumplimiento. También resolver problemas con los datos y promover los valores de los datos activos.

Las actividades de un gobierno de datos se pueden dividir entre las actividades de planeación y las actividades de control.

Para la planeación de la Gestión de Datos hay que comprender las necesidades estratégicas de la empresa para poder desarrollar una estrategia de datos correcta. También hay que establecer una organización de roles y el nombramiento de unos administradores para estos roles. Además, de definir y desarrollar unas políticas de datos y estándares, así como crear una arquitectura de datos que permita la coherencia entre las distintas aplicaciones.

Entre las actividades de control de Gestión de Datos están la de supervisar la organización de datos, coordinar las distintas actividades, gestionar y resolver problemas relacionados a los datos, asegurar la conformidad a las regulaciones y políticas existentes, además de indicar el método de comunicación y promoción de estos datos.

El responsable de liderar el Gobierno de Datos es el director de datos, o CDO por sus siglas en inglés, este reportará directamente al CIO, o jefe de información, que es el responsable de la coordinación de la gestión de datos y el Gobierno de datos.

[1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 63-92

### 3.2 Arquitectura de datos

La gestión de la Arquitectura de Datos son aquellos procedimientos que se llevan a cabo para definir y mantener aquellas tareas que proporcionan un vocabulario común de estándares de negocios, requisitos de datos estratégicos. También se trazan diseños integrados en un alto nivel para que se alineen con la estrategia empresarial y sus requerimientos.

Para llegar a este fin, se debe planear la arquitectura con visión y previsión para proveer datos de alta calidad e identificar los requerimientos comunes para todas las aplicaciones. Al mismo tiempo, diseñar estructuras conceptuales y planes de requerimientos para los datos presentes y para largo plazo de la empresa.

Es por ello por lo que la Arquitectura de Datos se puede dividir en 3 grandes categorías de especificaciones:

- 1- **Modelos de datos de la empresa:** Componente central de la arquitectura, donde se describe las entidades de negocio, estructura de datos, el tipo de datos y sus relaciones.
- 2- **Análisis de la cadena de valor de la información:** Donde se alinean los datos con los procesos de negocios.
- 3- **Arquitectura de entrega:** Que incluye las arquitecturas de base de datos, su integración y almacenamiento, así como documentos, metadatos y programas asociados.

Por tanto, las actividades para desarrollar la Gestión de Arquitectura de Datos hay que entender las necesidades de la empresa, desarrollar un modelo de datos empresariales que se tendrán que alinear con los otros modelos existentes. Definir y mantener la integración entre todos los sistemas, las estructuras jerárquicas y las distintas tecnologías.

La realización de estas tareas recae en los arquitectos y modeladores de datos junto a administradores de bases de datos y siempre bajo la supervisión del CDO.

### 3.2.1 Modelo de datos empresarial

Para crear un modelo de datos empresarial se tienen que definir y documentar vocabulario, reglas de negocios y conocimiento, a su vez este se tiene que mantener para enriquecer cada vez más el modelo y ser más preciso para mostrar la arquitectura de la empresa.

Un modelo de datos se construye por capas de información organizadas jerárquicamente.

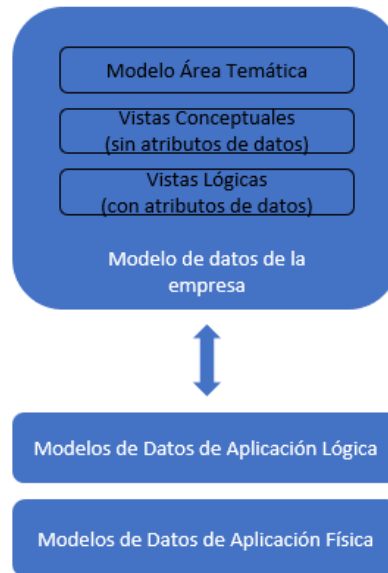


Figura 4. Capas de Modelo de Datos Empresariales

Estas capas están construidas de arriba hacia abajo, siendo el contenido más alto el nivel de la jerarquía fundamental y los más bajos definen los detalles y dependencias entre los datos.

- **Modelo de Área temática:** Son las entidades de negocio presentes en la empresa.
- **Vistas Conceptuales:** Se encarga de las áreas temáticas y relaciones entre entidades, es decir, los conceptos utilizados en la empresa y sus relaciones entre ellos.
- **Vistas Lógicas:** Se encarga de los atributos de datos esenciales para cada instancia de la entidad de negocios. Estos atributos son los datos comunes y estandarizados necesarios para la empresa.
- **Modelo de Datos de Aplicación Lógica:** Un modelo lógico de datos es un modelo que no es específico de una base de datos que describe aspectos relacionados con las necesidades de una organización para recopilar datos y las relaciones entre estos aspectos. Contiene representaciones de entidades y atributos, relaciones, identificadores y restricciones entre relaciones.
- **Modelo de Datos de Aplicación Física:** Un modelo de datos físicos es un modelo específico de bases de datos que representa objetos de datos relacionales (por ejemplo, tablas, columnas, claves principales y claves externas) y sus relaciones.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 93-116



### 3.3 Desarrollo de datos

El desarrollo de datos es el análisis, diseño, implementación y aplicación de las soluciones de datos para dar un mayor valor a los recursos de los datos de la empresa. Los componentes primarios de estas soluciones son las bases de datos y otras estructuras junto con componentes de soluciones de datos como informes o interfaces.

Para ello, se debe identificar y definir los requerimientos de datos para diseñar la estructura adecuada que mantenga la conformidad de los estándares, la integridad, seguridad y usabilidad de los datos activos. Una vez este todo definido implementar y mantener dicha solución.

La primera actividad a realizar sería saber las necesidades de información de negocio, así como los procesos de entrada y salida de datos.

A continuación, el análisis y construcción del modelo de datos. Empezando por un modelo conceptual que determina los objetos y sus relaciones seguido de un modelo de datos lógico donde se añaden los atributos de datos para cada entidad, siguiendo unos estándares de nomenclatura. Para pasar al modelo físico de datos donde se adaptaría el modelo lógico al lenguaje técnico y específico del sistema de gestión de base de datos que se vaya a emplear (Microsoft SQL Server, Oracle, Acces...).

Una vez modelado, se construiría e implementaría en el sistema, manteniendo tareas de verificación y testeo.

Los miembros del equipo participantes en el desarrollo de datos serían los expertos de negocio que proporcionarían los requerimientos. Arquitectos de datos y administradores de bases de datos para definir el modelo acorde con la aplicación de bases de datos a utilizar además de especialistas en servicios de acceso a datos para que se conecten los diversos programas sin problemas.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 117-162

### 3.4 Operaciones de datos

En la gestión de Operaciones de Datos se desarrolla, mantiene y se da soporte a los datos estructurados o base de datos para dar el mayor valor a los recursos de la empresa.

Los objetivos incluyen proteger y garantizar la integridad de los datos estructurados activos, administrar la disponibilidad de estos datos y optimizar el rendimiento de las transacciones de bases de datos.

Para ello se agrupan en 2 grandes grupos de actividades, Apoyo de Base de Datos y Gestión tecnológica de datos.

En las actividades de soporte de base de datos, hay que implementar y controlar las bases de datos, así como también garantizar la obtención de datos desde fuentes externas. Trazar planes de recuperaciones de datos y copias de seguridad de los datos almacenados. Establecer niveles y control en disponibilidad como en rendimiento de la base de datos. Además de diseñar planes para archivar, retener y purgar datos cuando estos ya no sean necesarios.

Para la gestión tecnológica se debe de entender los requerimientos de la empresa para poder definir y evaluar la arquitectura más idónea. Una vez elegida, se administrará, instalará en todos los usuarios necesarios haciendo un seguimiento en su inventario para evaluar necesidades y costes.

Los administradores de bases de datos o DBA serán los encargados de ejecutar estas actividades.



### 3.5 Seguridad de datos

La gestión de seguridad de datos es la planificación, desarrollo y ejecución de las políticas para proporcionar la autenticación, autorización y acceso a los datos activos.

Estos procedimientos aseguran que aquellas personas autorizadas puedan utilizar los datos evitando accesos inapropiados o a información restringida. Para llevar a cabo estas prácticas hay que cumplir con los intereses de la empresa, pero también con aquellas regulaciones en privacidad y confidencialidad.

Para llevar a cabo estas actividades hay que entender los requisitos del negocio en cuanto necesidades de seguridad y protección de datos sensibles, por tanto, se definirá una política, controles, procedimientos y estándares de seguridad. En estos controles de seguridad es donde se llevará a cabo el control de usuarios, permisos de acceso y contraseñas, siendo estas últimas las que tendrán unas normas extra para crear contraseñas complejas y fuertes.

También se deberá clasificar la confidencialidad de los datos y tener estos preparados para una auditoría de seguridad.

[1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 191-208

#### 3.5.1 Reglamento General de Protección de Datos (RGPD)

El Parlamento Europeo y el Consejo han aprobado el Reglamento General de Protección de Datos (RGPD) con la aspiración de unificar los regímenes de todos los Estados Miembros sobre la materia, en vigor desde el día 25 de mayo de 2016.

Se aplicarán 3 principios:

- **PRINCIPIO DE RESPONSABILIDAD.** Habrá que implementar mecanismos que permitan acreditar que se han adoptado todas las medidas necesarias para tratar los datos personales como exige la norma. Las organizaciones deben ser capaces de demostrar que cumplen dichas exigencias, lo cual obligará a desarrollar políticas, procedimientos, controles, etc.
- **PRINCIPIOS DE PROTECCIÓN DE DATOS POR DEFECTO Y DESDE EL DISEÑO.** Se deberán adoptar medidas que garanticen el cumplimiento de la norma desde el mismo momento en que se diseñe una empresa, producto, servicio o actividad que implique tratamiento de dato, como regla y desde el origen.
- **PRINCIPIO DE TRANSPARENCIA.** Los avisos legales y políticas de privacidad deberán ser más simples e inteligibles, facilitando su comprensión, además de más completos.

Obligaciones para empresas, administraciones y otras entidades:

- En ocasiones, será obligatorio designar un **delegado de Protección de Datos (DPO)**, interno o externo, que asista a las organizaciones en el proceso de cumplimiento normativo.
- **EVALUACIONES DE IMPACTO SOBRE LA PRIVACIDAD**, que determinen los riesgos específicos que supone tratar ciertos datos de carácter personal y prevean medidas para mitigar o eliminar dichos riesgos.
- Las **BRECHAS DE SEGURIDAD** deberán ser comunicadas a las autoridades de control y, en casos graves, a los afectados, con un máximo de 72 horas.
- **DATOS SENSIBLES:** Se amplían los datos especialmente protegidos, incluyendo ahora los datos genéticos y biométricos. Se incluyen también en esta categoría las infracciones y condenas penales, aunque no las administrativas.
- **GARANTÍAS ADICIONALES PARA LAS TRANSFERENCIAS INTERNACIONALES DE DATOS** fuera de la Unión Europea.



- **SANCIONES:** Las cuantías de las sanciones por incumplimiento de la norma crecen, pudiendo llegar a los 20 millones de euros o el 4% de la facturación global anual.

Nuevos derechos para los ciudadanos:

- **TRANSPARENCIA e INFORMACIÓN.** Las organizaciones, al tratar datos personales, deben proporcionar mayor información para favorecer la toma de decisiones por el ciudadano. Se tiene una especial consideración con los menores de edad en este punto.
- **CONSENTIMIENTO.** El consentimiento para poder tratar datos de carácter personal ha de ser inequívoco, libre y revocable y deberá darse mediante un acto afirmativo claro. No se admite consentimiento tácito.
- **DERECHO AL OLVIDO.** Se podrá revocar el consentimiento prestado para el tratamiento de datos personales en cualquier momento, pudiendo exigir la supresión y eliminación de los datos en redes sociales o buscadores de internet.
- **DERECHO A LA LIMITACIÓN DEL TRATAMIENTO.** Permite al ciudadano solicitar el bloqueo temporal del tratamiento de sus datos cuando existan controversias sobre su licitud.
- **PORTABILIDAD DE LOS DATOS.** Se permitirá al ciudadano solicitar la transferencia de los datos personales de un proveedor de servicios en Internet a otro.
- **DENUNCIAS.** Se podrán presentar denuncias a través de asociaciones de usuarios.
- **INDEMNIZACIONES.** Se reconoce la posibilidad de exigir indemnización de daños y perjuicios derivados del tratamiento ilícito de los datos personales.

[8] Reglamento General de Protección de Datos, <https://rgpd.es/> [Online]

### 3.6 Datos Maestros y de Referencia

Un dato maestro es un registro único que sirve de referencia para toda la empresa. Para ello hay que planificar, implementar y controlar las actividades para asegurar la consistencia de este dato maestro en todos los entornos y aplicaciones.

La administración de datos referenciales es el control del dominio de los valores definidos o vocabulario, incluyendo los términos estándares y otros identificadores únicos. La administración de datos maestros es el control de los valores de los datos maestros, que aseguran un uso compartido y consistente de estos.

Es decir, que el propósito de la administración de datos de datos maestros y de referencia es la de proporcionar una fuente autorizada a bajos costos y complejidad, a través de reutilización de datos y aplicación de estándares.

Las actividades que se llevan a cabo son la de entender las necesidades de integración de los datos maestros, identificando fuentes de origen e incluyéndolas en la arquitectura. Implementar soluciones que administren estos datos, así como reglas de coincidencia para verificar que en todos los sistemas se cumple que se tenga un dato único.

Estas tareas son aplicadas entre los arquitectos de datos como los administradores de bases de datos.

[1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 213-243

### 3.7 Almacén de datos e inteligencia de negocios

Un Almacén de Datos (Data Warehouse en inglés) es una combinación ente una base de datos y los programas utilizados para recolectar, limpiar, transformar y almacenar los datos de diversas fuentes. Esta combinación ayuda a Inteligencia de Negocio (Business Intelligence) para la

construcción de reportes para monitorizar, entender y tomar decisiones empresariales sobre la organización.

Los objetivos es proporcionar un almacenamiento de los datos, actuales e históricos, de alta calidad para asegurar un entorno confiable y de fácil acceso para la adquisición, consulta y entrega de los datos. Esto provee de un punto empresarial para la entrega de datos que apoyan las decisiones de gobierno de dato y donde se construye la infraestructura necesaria para la creación de análisis y reportes.

Para conseguirlo, el primer paso es entender las necesidades de información de Inteligencia de negocio. Definir, implementar y mantener una arquitectura de Almacén de Datos aprovechando la integración de los diversos procesos de negocio (como pueden ser servidores, bases de datos, seguridad, herramientas ETL y de calidad, calendarios, etc.). Implementar herramientas de Inteligencia de negocios para que se puedan procesar los datos y procesos. Con ello generar paneles de gestión, cuadros de mandos o análisis predictivos que ayuden en la toma de decisiones.

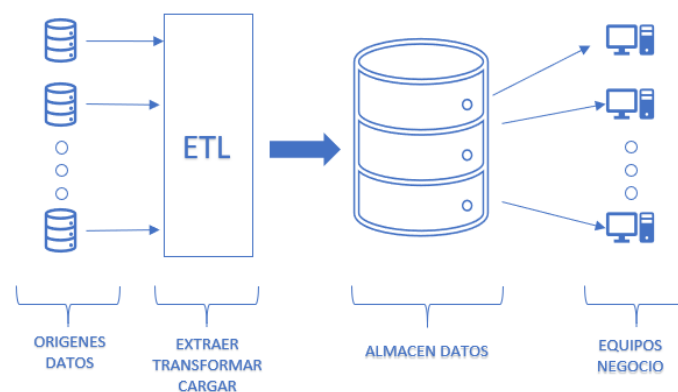


Figura 5. Esquema Almacén Datos

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 243-289

### 3.8 Contenidos y documentación

La gestión de documentación y contenidos se encarga de la captura, almacenamiento, acceso y uso de los datos almacenada fuera de las bases de datos, esta información puede ser archivos electrónicos o físicos como correos electrónicos, textos, graficas, imágenes, audios o videos.

Por tanto, hay que asegurar la disponibilidad de los datos almacenados en formatos no estructurados, para así permitir su recuperación y uso a través de todo el negocio. También esta información de se ser protegida legalmente como se indica en el apartado de seguridad.

Las actividades para ejecutar esta tarea se pueden dividir en 2 grades grupos, el de Gestión de Documentos y el de Gestión de Contenido.

- **Gestión de Documentos:** Creación de un plan para desarrollar estos documentos durante su ciclo de vida, para ello hay que implementar sistemas de gestión documental que permitan el acceso, almacenamiento y adquisición con sus correspondientes controles de seguridad. Toda la información integrada en los documentos debe tener una copia de seguridad e incluirse en las actividades de recuperación de datos, además de establecer cuanto tiempo deben permanecer guardados.
- **Gestión de Contenido:** Es la gestión de la información que contienen los documentos, esta debe ser almacenada y publicada para que sea accesible. Se define y mantiene la estructura de la documentación ya sea por formatos o por índices y se le proporciona acceso para toda la compañía. También debe de ser recuperable y estar actualizada.

Estas tareas se llevan a cabo por los administradores de sistema de documentos y registros.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 291-313

### 3.9 Metadatos

El termino metadato se usa para definir un dato y su contexto, que puede ser agrupados en tipos, finalidad, organización, entre otros.

Estos procesos proporcionan la integración entre metadatos de varias fuentes con lo que su acceso es más sencillo y se asegura una mayor calidad, así como también una mayor seguridad.

Para ello, una de las primeras actividades sería la de establecer una estrategia para la creación de la arquitectura en que se asegure unas metas, objetivos y una futura extensibilidad.

Crear normas de los metadatos para garantizar que toda la empresa opera de la misma forma (estas normas podrían ser normas ISO para procesamiento de terminología), esto asegurará una adquisición eficaz. También monitorizar estos procesos para garantizar su correcto funcionamiento.

[1]DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 313-346

### 3.10 Calidad de datos

La gestión en calidad de datos (Data Quality Management) es un proceso de apoyo la gestión del cambio, ya que una buena calidad de datos reporta una información exacta y de buen rendimiento para el negocio.

Para ello podemos aplicar el ciclo de gestión de calidad del dato

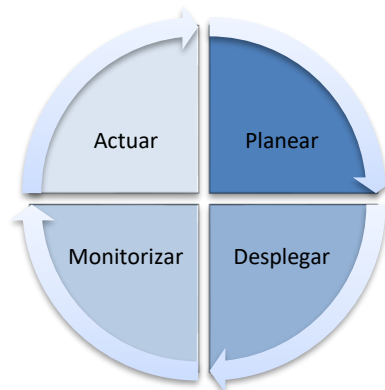


Figura 6. Ciclo de Gestión de calidad de datos

- **Planear:** Se idea un diagnóstico de la situación y se identifica los indicadores clave para medir la calidad.
- **Desplegar:** Implementación de procesos para medir y mejorar la calidad.
- **Monitorizar:** Seguimiento y medición en relación con las expectativas definidas.
- **Actuar:** Proceder a la resolución de problemas identificados que empeoran la calidad.

Las actividades para desarrollar la calidad del dato empezaran por definir los requisitos de calidad del dato, identificando componentes, evaluando como se clasifican los errores y especificando las reglas para corregirlos.

Por ello, es fundamental, realizar una evaluación de los datos para ello se implementará un perfilado de datos que indica que tiene un conjunto de datos y sus anomalías.



Estas anomalías nos ayudaran a identificar las reglas de calidad y como agruparlas por dimensiones. Un ejemplo de las dimensiones de las reglas de calidad de datos en:

- **Exactitud:** Indica la medida en que los datos representan correctamente las entidades que modelan.
- **Compleitud:** Ciertos atributos tengan valores asignados en un conjunto de datos.
- **Consistencia:** Los valores en un conjunto de datos son coherentes con los valores de otro conjunto de datos
- **Vigencia:** La información está al día.
- **Precisión:** Nivel de detalle del elemento, por ejemplo, precisión en datos numéricos.
- **Unicidad:** Ningún registro existe más de una vez en un conjunto de datos.

Una vez tenemos las dimensiones a aplicar, definiremos las reglas de calidad del dato y unos niveles de servicio, que nos servirán para probar y validar los requerimientos aplicados. Estas reglas deberán ser monitorizadas para ver la evolución de la calidad y resolver posibles problemas detectados y proceder a su limpieza.

Estas actividades serán llevadas por los analistas de calidad de datos.

[1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” pp 346-372

[12] Mario G. Piattini Velthuis, Ismael Caballero Muñoz-Reja, Ana Isabel Gómez Carretero, Fernando Gualo Cejudo, Jorge Merino García, Bibiano Rivas García, “Calidad de Datos” Editorial Ra-Ma, 2018



## Capítulo 4. Ejemplo práctico

Vamos a mostrar todo lo indicado anteriormente, con un ejemplo práctico basado en una empresa de alquiler de coches.

En esta empresa, los datos se registran en documentos Excel de forma local por cada área de trabajo. Este método de trabajo hace que los datos no sean ni consistentes ni exactos, además de que no se pueden acceder a ellos de forma remota. Por otra parte, dejan datos personales o sensibles sin protección y con una forma muy fácil de acceso.

Se pretende planificar y crear un modelo de gestión de datos adaptado, identificando las funciones y pasos necesarios para ello.

### 4.1 Gobierno de datos

Mediante el análisis de la compañía, se identifican necesidades como un modelo de gestión del dato. Para ello, se deberán estudiar las distintas áreas de negocio para poder crear una arquitectura del dato adecuada. También se aplicarán mejoras en la administración de datos maestros, la gestión de la seguridad o la calidad de los datos almacenados.

Para ello, se utilizarán herramientas como:

- Para la creación de modelos de datos: SAP PowerDesigner:
- Para la creación de ETLs: SAP Data Services Designer
- Para analizar la calidad de los datos: SAP Information Steward
- Para las bases de datos: Microsoft SQL Server
- Para la visualización de datos e informes: Microsoft Power BI

### 4.2 Arquitectura del dato

Por la tipología del negocio, se identifican 5 grandes áreas temáticas:

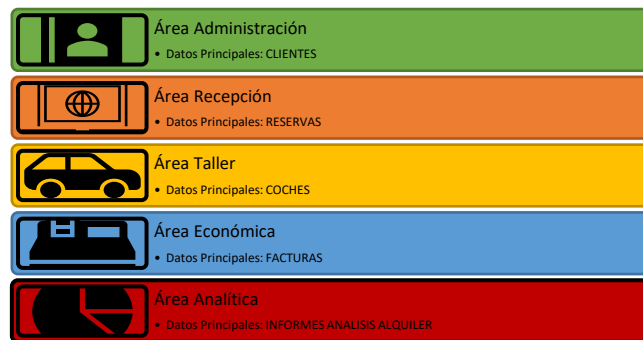


Figura 7. Áreas temáticas de la empresa

Analizando estas áreas y los documentos que utilizan actualmente para almacenar los datos, se creará un modelo de dato adecuado para esta compañía.

#### 4.2.1 Modelo de datos conceptual

En este modelo de datos vamos a identificar las relaciones entre las entidades más importantes. No se especifica ningún atributo ni ninguna clave que relacione las distintas tablas.

En el apartado del área analítica, se crearán sus propias vistas a partir de las tablas creadas.

Estas vistas están identificadas como el apartado “Informes” para poder ver la relación con las otras tablas (en líneas discontinuas).

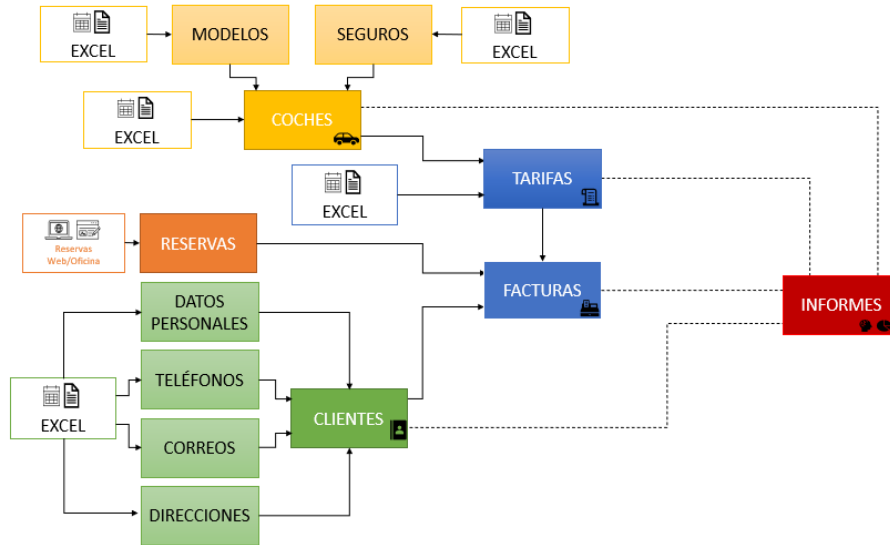


Figura 8. Modelo de datos conceptual esquemático

Llevando este esquema al programa utilizado para la creación de modelos de datos, quedaría de la siguiente forma:

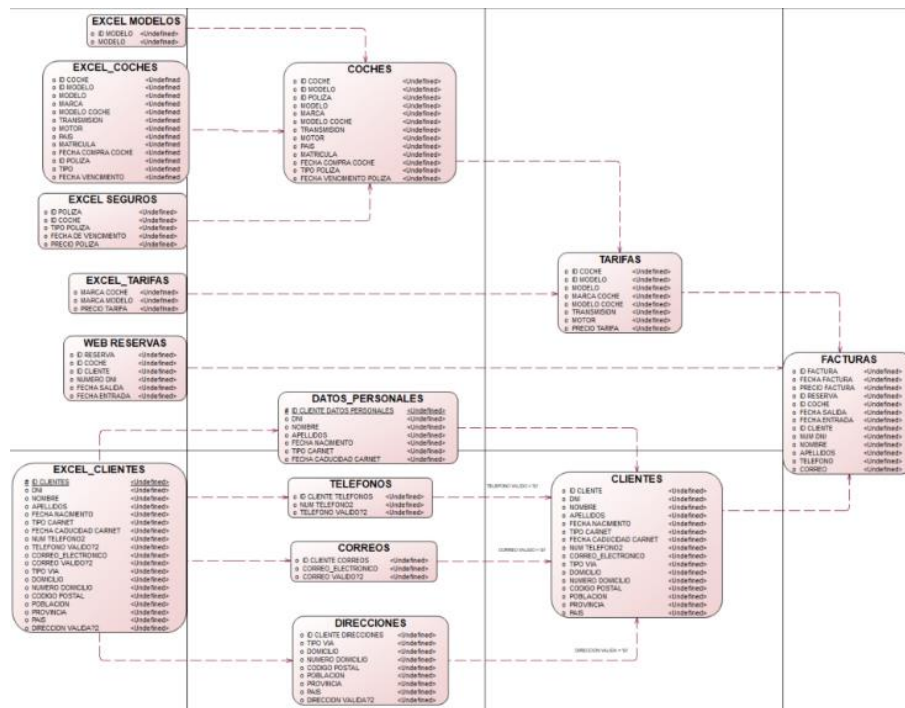


Figura 9. Modelo de datos conceptual

Como se ve en el modelo conceptual, se ha decidido dividir los datos pertenecientes a los clientes en las tablas Datos Personales, Teléfonos, Correos y Direcciones para nuevo crear una nueva tabla llamada Clientes. Se ha tomado esta decisión para tener mejor estructurados los datos y poder aplicarle medidas de seguridad de datos adecuadas a cada caso.

En el área de recepción ya se realiza en una tabla de bases de datos, así que aprovecharemos directamente esta tabla llamada Reservas.

Como se ha indicado anteriormente, el área de analítica creara en base de datos vistas según sus necesidades, así que no es necesario crear ninguna tabla en el modelo.



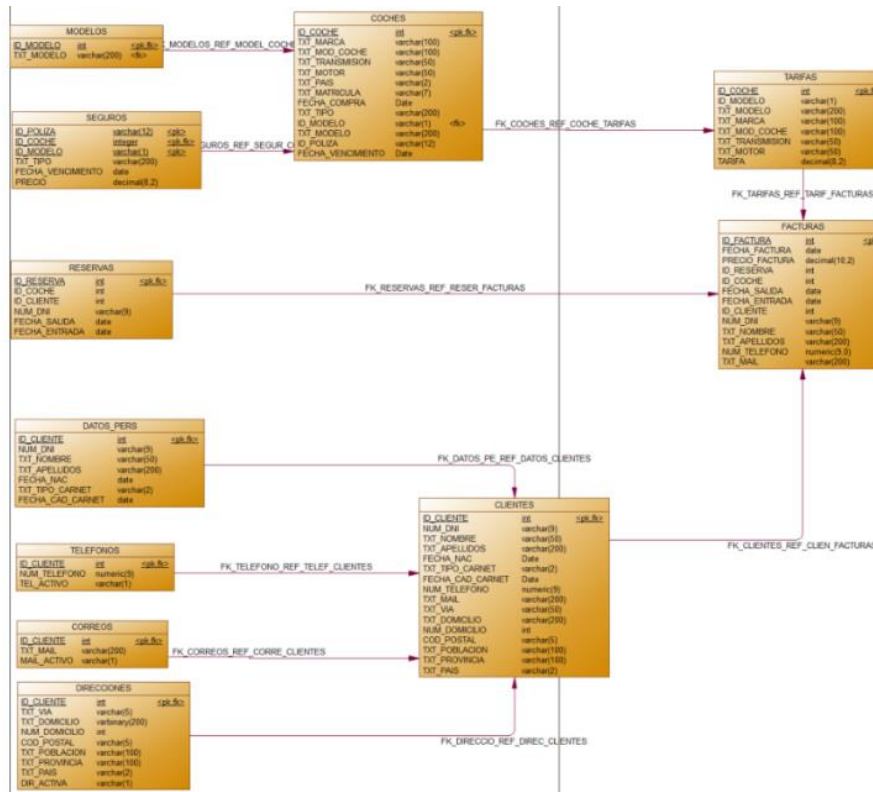


Figura 11. Modelo de datos físico

#### 4.2.4 Generación de DDL

Una vez tenemos completado el modelo de datos físico, generaremos los archivos DDL (Data Definition Language, por sus siglas en inglés). Estos nos permiten la creación, modificación o borrado de las tablas que componen el modelo en el lenguaje de Base de Datos indicado, en nuestro caso SQL.

Como ejemplo, tenemos la definición para la creación de la tabla Datos Personales y la tabla Seguros.

```

/*-----*/
/* Table: DATOS_PERS */
/*-----*/
create table DATOS_PERS (
  ID_CLIENTE          int           not null,
  NUM_DNI             varchar(9)    not null,
  TXT_NOMBRE         varchar(50)    null,
  TXT_APELLIDOS     varchar(200)   null,
  FECHA_NAC         date           null,
  TXT_TIPO_CARNET   varchar(2)     null,
  FECHA_CAD_CARNET date           null,
  constraint PK_DATOS_PERS primary key (ID_CLIENTE)
)
go

```

Figura 12. DDL creación tabla Datos Personales

```

/*-----*/
/* Table: DATOS_PERS */
/*-----*/
create table DATOS_PERS (
  ID_CLIENTE          int           not null,
  NUM_DNI             varchar(9)    not null,
  TXT_NOMBRE         varchar(50)    null,
  TXT_APELLIDOS     varchar(200)   null,
  FECHA_NAC         date           null,
  TXT_TIPO_CARNET   varchar(2)     null,
  FECHA_CAD_CARNET date           null,
  constraint PK_DATOS_PERS primary key (ID_CLIENTE)
)
go

```

Figura 13. DDL creación tabla Clientes

Ejecutando estos códigos en la ventana de consultas del programa utilizado para la gestión de la base de datos, crearemos todas las estructuras de las tablas necesarias de nuestro modelo. Con ello, podremos operar con ellas para la carga de los datos.

### 4.3 Operaciones de datos

Para poder operar con los datos, realizaremos una primera carga desde los documentos Excel que tienen los datos de origen, hacia las tablas de destino. Para ello, utilizaremos la herramienta de ETLs SAP Data Services Designer.

Una vez, los datos estén cargados en las tablas cualquier registro que se tenga que añadir o modificar se realizara directamente sobre las tablas.

En la forma de trabajar SAP Data Service Designer, cada ETL va dentro de un Dataflow (tendremos tantos como ETLs necesarias). Estos Dataflows (DF) se agrupan en Workflows (WF), se han creado un workflow por área de trabajo. Estos se crearán dentro de un Job de trabajo que será el que se ejecute para poner en funcionamiento las ETLs.



Figura 14. Orden jerárquico objetos programa SAP DSD

Por tanto, se creará un job de trabajo. Dentro de este tendremos los objetos WF, uno por área de trabajo.

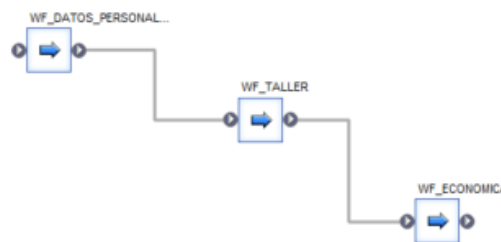


Figura 15. Esquema de construcción de Workflows

La ejecución de estos WF será secuencial, primero se ejecutará lo que contenga Datos\_Personales, a continuación, lo hará Taller y para finalizar Económica.

#### 4.3.1 Área de Administración

En el área de administración se guarda toda la información de clientes en un documento Excel.

Se elige este documento Excel como origen de los datos y mediante una ETL se cargan las diversas tablas con los campos necesarios en cada una de ellas, para así tener un mayor control sobre los datos personales.

A continuación, utilizando estas tablas y con otra ETL crearemos la tabla Clientes, donde tendremos los mismos datos que había en el documento Excel, pero de una forma más centralizada y de fácil acceso.

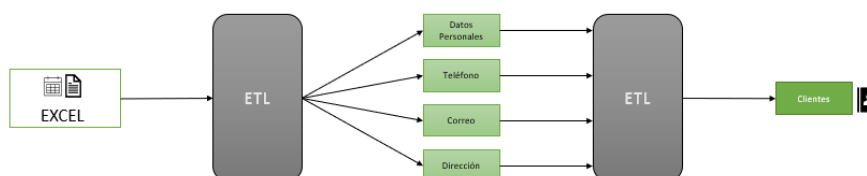


Figura 16. Esquema ETL Área Administración

Para hacer la carga en la tabla Clientes, como muestra la imagen anterior, hay que dividir en 2 workflows las distintas ETLs. Una donde el Excel carga las tablas Datos Personales, Teléfono, Correo y Dirección. Otra para desde estas tablas cargar la tabla Clientes.



Figura 17. Esquema construcción de Dataflows dentro del área de Administración

Para cargar las 4 tablas mencionadas anteriormente, en primer lugar, se tiene un documento Excel como origen. Para hacer la carga necesitaremos crear, en el programa, un archivo donde se indican el nombre de las columnas, su tipo y la ruta de donde se encuentra el documento Excel. Este objeto en el programa se llama Excel Workbook y lo nombraremos DATOS\_PERSONALES.

Las tablas de destino de estas ETLs son las tablas de la base de datos que se muestran en el modelo de datos.

Field Name	Data Type	Field Size	Precision	Scale	Content Type	Description
ID_CLIENTE	int					
NUM_DNI	varchar	9				
TXT_NOMBRE	varchar	50				
TXT_APELLIDOS	varchar	200				
FECHA_NAC	date					
TXT_TIPO_CARNET	varchar	2				
FECHA_CAD_CARNET	date					
TXT_VIA	varchar	50				
TXT_DOMICILIO	varchar	200				
NUM_DOMICILIO	int					
COD_POSTAL	varchar	5				
TXT_POBLACION	varchar	100				
TXT_PROVINCIA	varchar	100				
TXT_PAIS	varchar	2				
DIR_ACTIVIA	varchar	1				
NUM_TELEFONO	numeric		9	0		
TEL_ACTIVIVO	varchar	1				
TXT_MAIL	varchar	200				
MAIL_ACTIVIVO	varchar	1				

Format: Data Access

File Location: {none}  Delete file after transfer

Directory: P:\TFG

File name: EXCEL\_DATOS\_PERSONALES.xlsx

Access method:

Named range

Worksheet  Number

Range:

All fields

Custom range

Extend range

Code page: <default>

Use first row values as column names

Figura 18. Plantilla creación Excel Wordbook DATOS\_PERSONALES

Una vez donde tenemos este archivo que utilizaremos como origen, cargaremos los datos necesarios en cada una de las tablas. Para ello es necesario utilizar el objeto “Query” del programa, donde indicaremos la relación entre los campos de origen y los de destino.

Mostraremos la construcción de las querys Datos Personales y Correos

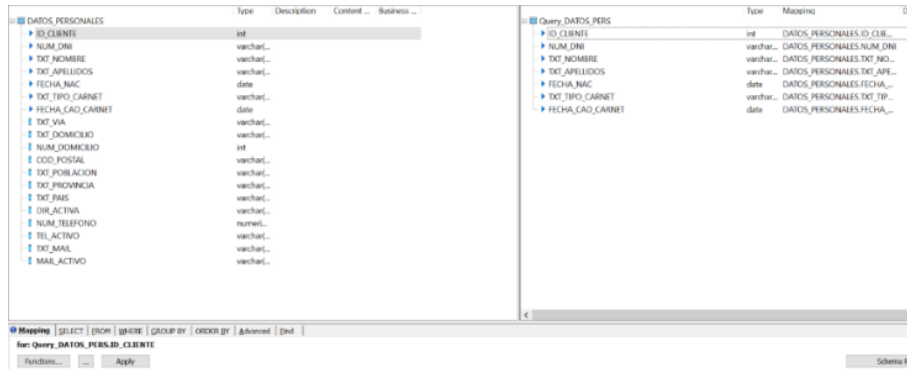


Figura 19. Construcción query Datos Personales

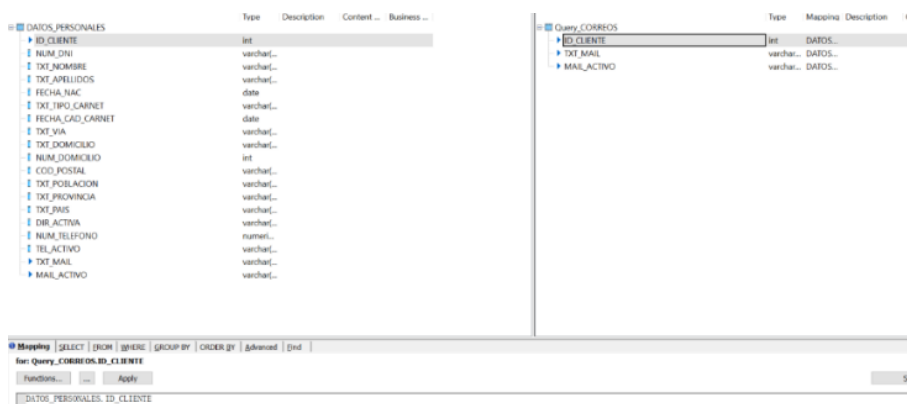


Figura 20. Construcción query Correos

Una vez seleccionados los campos de origen que necesitamos en cada tabla, quedaría la ETL de la siguiente manera:

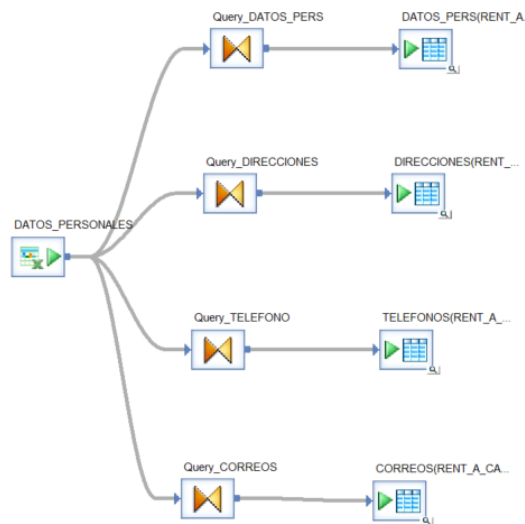


Figura 21. Construcción ETLs, con origen documento Excel DATOS\_PERSONALES

Para hacer la segunda ETL, tendríamos como origen las tablas que acabamos de cargar. Para ello necesitaríamos unir en un objeto “Query” las 4 tablas, indicando que columnas son necesarias en la tabla destino Clientes.

Column	Type	Mapping	Description
ID_CLIENTE	int	DATOS_PERS.ID_CLIENTE	
NUM_DNI	varchar(9)	DATOS_PERS.NUM_DNI	
TXE_NOMBRE	varchar(50)	DATOS_PERS.TXE_NOMBRE	
TXE_APELLIDOS	varchar(200)	DATOS_PERS.TXE_APELLIDOS	
FECHA_NAC	date	DATOS_PERS.FECHA_NAC	
TXE_TPO_CARNET	varchar(2)	DATOS_PERS.TXE_TPO_CARNET	
FECHA_CAD_CARNET	date	DATOS_PERS.FECHA_CAD_CARNET	
TXE_VIA	varchar(50)	DIRECCIONES.TXE_VIA	
TXE_DOMICILIO	varchar(200)	DIRECCIONES.TXE_DOMICILIO	
NUM_DOMICILIO	int	DIRECCIONES.NUM_DOMICILIO	
TXE_POBACION	varchar(100)	DIRECCIONES.TXE_POBACION	
TXE_PROVINCIA	varchar(100)	DIRECCIONES.TXE_PROVINCIA	
TXE_PAS	varchar(2)	DIRECCIONES.TXE_PAS	
NUM_TELEFONO	numeric(9,0)	TELEFONOS.NUM_TELEFONO	
TXE_MAIL	varchar(200)	CORREOS.TXE_MAIL	

Figura 22. Relación columnas tabla Clientes

También le indicaremos, en el apartado FROM como se unen las tablas, en este caso la relación entre todas es el campo ID\_CLIENTE.

Input schema(s)	From	Join rank	Cache
CORREOS	<input checked="" type="checkbox"/>	0	Automatic
DIRECCIONES	<input checked="" type="checkbox"/>	0	Automatic
TELEFONOS	<input checked="" type="checkbox"/>	0	Automatic
DATOS_PERS	<input checked="" type="checkbox"/>	0	Automatic

Left	Join Type	Right	Join Condition
DATOS_PERS	Inner join	DIRECCIONES	DATOS_PERS.ID_CLIENTE = DIRECCIONES.ID_CLIENTE
→	Inner join	TELEFONOS	DATOS_PERS.ID_CLIENTE = TELEFONOS.ID_CLIENTE
→	Inner join	CORREOS	DATOS_PERS.ID_CLIENTE = TELEFONOS.ID_CLIENTE

Figura 23. Relaciones entre tablas origen para carga tabla Clientes

Como en la tabla solo queremos cargar aquellos datos que están en uso actualmente, se aplicara un filtro en las tablas. Esto se hace en la pestaña inferior WHERE del objeto query.

```

DIRECCIONES.DIR_ACTIVADA = 'S'
and
TELEFONOS.TEL_ACTIVADO = 'S'
and
CORREOS.MAIL_ACTIVADO = 'S'

```

Figura 24. Filtrado registros para carga en tabla Clientes

La ETL de esta segunda parte, almacenada en el DF DF\_DATOS\_PERSONALES\_CLIENTES quedaría de la siguiente manera:

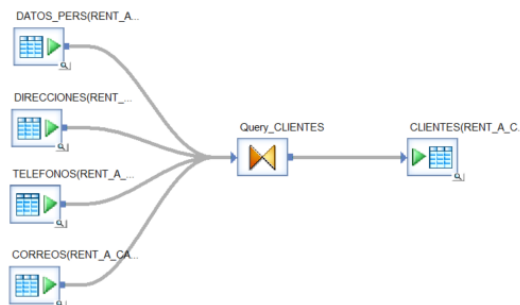


Figura 25. ETL carga tabla Clientes



### 4.3.2 Área de Recepción

En el área de recepción, se encargan de recibir las reservas ya sea físicamente o mediante la página web. Estos datos son almacenados por el programa de reservas en una tabla de nuestra base de datos, por tanto, ya tenemos creada la tabla con los datos de reservas y que se incluye en el modelo de datos.



Figura 26. Esquema carga tabla Reservas

### 4.3.3 Área de Taller

En el área de taller se almacena la información de las pólizas de seguros de los coches en un documento Excel, así como una pequeña lista con el código y el nombre que le correspondía a cada modelo.

La información de los coches que hay disponibles también se guardaba en un documento Excel. Esta información se extrae con una ETL y se le añade información que se encuentra en las tablas Modelos y Seguros para todo ello guardarlo en la tabla Coches.

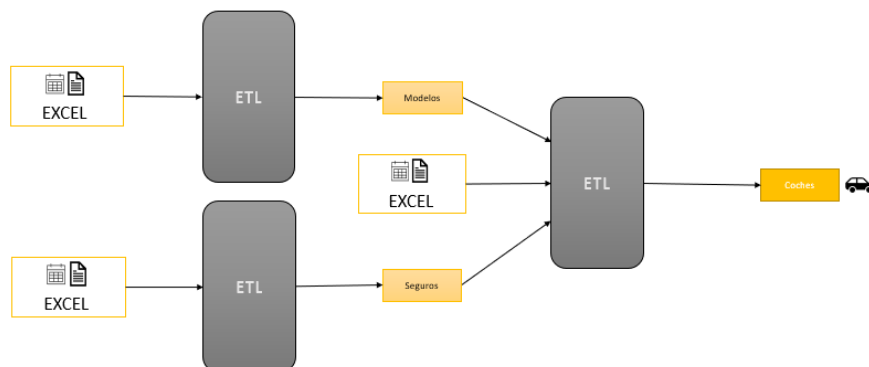


Figura 27. Esquema ETL Área Taller

Para ello, se va a dividir en 3 cargas distintas. Primero se cargará la tabla de Modelos, luego la tabla de Seguros y posteriormente la tabla de Coches. Las tablas de destino son las creadas en la base de datos y representadas en el modelo de datos.



Figura 28. Esquema construcción de Dataflows dentro del área de Taller

Para hacer la carga necesitaremos, como hemos visto anteriormente, crear un objeto Excel Workbook y los nombraremos MODELOS, SEGUROS y COCHES.

Format name:

Field Name	Data Type	Field Size	Precision	Scale	Content Type	Description
ID_MODELO	int					
TXT_MODELO	varchar	200				
*						

Format: **Data Access**

File Location:   Delete file after transfer

Directory:

File name:

Access method:

Named range

Worksheet   Number

Range:

All fields

Custom range

Extend range

Code page:

Use first row values as column names

Figura 29. Plantilla creación Excel Wordbook MODELOS

Format name: SEGUROS

Field Name	Data Type	Field Size	Precision	Scale	Content Type	Description
ID_POLIZA	varchar	12				
ID_COCHE	int					
TXT_TIPO	varchar	200				
FECHA_VENCIMIENTO	date					
PRECIO	decimal		10	2		

Format: Data Access

File Location: {none}  Delete file after transfer

Directory: P:\TFG

File name: SEGUROS.xlsx

Access method:

Named range

Worksheet  Number

Range:

All fields

Custom range

SEGUROS  Extend range

Code page: <default>

Use first row values as column names

Figura 30. Plantilla creación Excel Wordbook SEGUROS

Format name: COCHES

Field Name	Data Type	Field Size	Precision	Scale	Content Type	Description
ID_COCHE	int					
ID_MODELO	varchar	1				
TXT_MODELO	varchar	200				
TXT_MARCA	varchar	100				
TXT_MOD_COCHE	varchar	100				
TXT_TRANSMISION	varchar	50				
TXT_MOTOR	varchar	50				
TXT_PAIS	varchar	2				
TXT_MATRICULA	varchar	7				
FECHA_COMPRA	date					
ID_POLIZA	varchar	12				
TXT_TIPO	varchar	200				
FECHA_VENCIMIENTO	date					

Format: Data Access

File Location: {none}  Delete file after transfer

Directory: P:\TFG

File name: COCHES.xlsx

Access method:

Named range

Worksheet

Range: COCHES  Number

All fields

Custom range

Extend range

Code page: <default>

Use first row values as column names

**Figura 31. Plantilla creación Excel Workbook COCHES**

Una vez se han creado estos objetos, empezaremos con la ETL de Modelos, que se creara en el Dataflow DF\_MODELOS.



**Figura 32. ETL Modelos**

Con el objeto query seleccionaremos las columnas existentes en el documento Excel, para llevarlas a la tabla de destino. En este caso no aplicaremos ningún tipo de filtro.

Schema In: MODELOS

Type	Description	Content ...	Business ...
ID_MODELO	int		
TXT_MODELO	varchar(200)		

Schema Out: Query\_TALLER

Type	Mapping
ID_MODELO	MODELOS.ID_MODELO
TXT_MODELO	MODELOS.TXT_MODELO

Mapping: SELECT FROM WHERE GROUP BY ORDER BY Advanced End

For: Query\_TALLER.ID\_MODELO

Functions... Apply

MODELOS.ID\_MODELO

**Figura 33. Query Modelos**

Seguiremos con la ETL de Seguros, que se creara en el Dataflow DF\_SEGUROS.

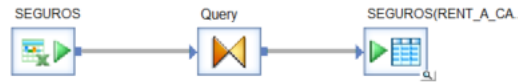


Figura 34. ETL Seguros

Como con la anterior extracción, utilizaremos el objeto query para seleccionar las columnas existentes en el documento Excel, para llevarlas a la tabla de destino. Tampoco aplicaremos ningún filtro.

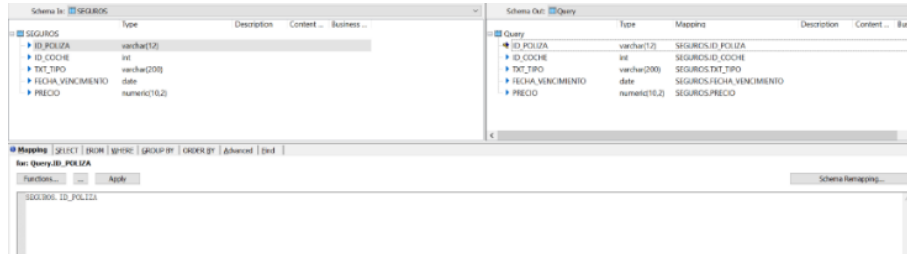


Figura 35. Query Seguros

Para hacer la carga en la tabla Coches, además de los datos que provienen del documento Excel, necesitamos los datos de las 2 tablas que acabamos de cargar, por eso se hace en tercer lugar en el dataflow DF\_COCHES.

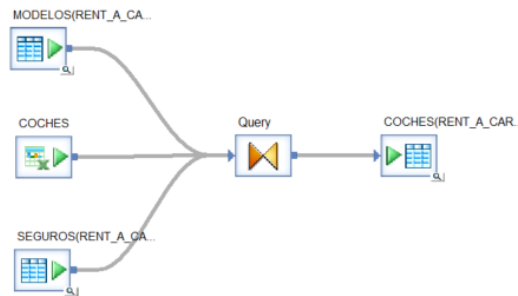


Figura 36. ETL Coches

Seleccionaremos las columnas de los elementos de origen mediante la query y utilizaremos el campo ID\_MODELO para unir las tablas COCHES con MODELOS y el campo ID\_POLIZA para unir los COCHES con la tabla de SEGUROS.

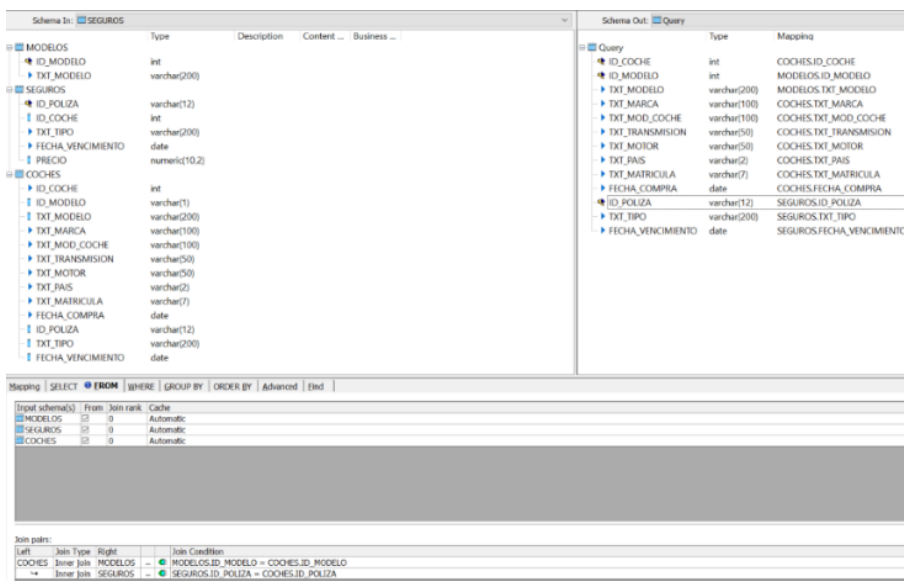


Figura 37. Query Coches

Con ello tendremos con registros todas las tablas que pertenecen al área de taller.

#### 4.3.4 Área Económica

Para el área económica necesitaremos 2 tablas, la tabla de Tarifas y la de Facturas.

Los datos de las tarifas se guardan en otro documento Excel con datos de coches, modelos y tarifa aplicada, a ese documento Excel se le añadirá la tabla Coches para rellenar la tabla con más datos y ser más específica.

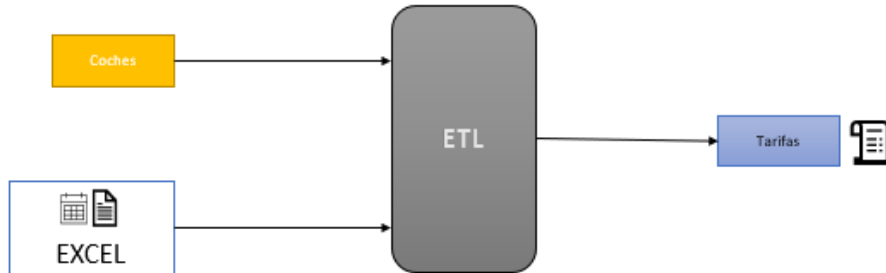


Figura 38. Esquema ETL tabla Tarifas

Para generar las facturas, se cargan los campos necesarios desde las tablas de Clientes, Tarifas y desde la Base de Datos de las reservas. Por ello, este paso se producirá después de la carga de la tabla Tarifas.

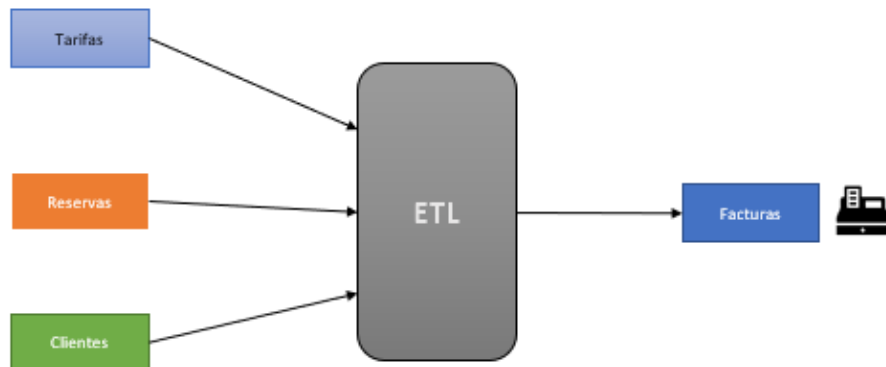


Figura 39. Esquema ETL tabla Facturas

Por tanto, la división de las ETLs en los Dataflows quedaría de la siguiente forma

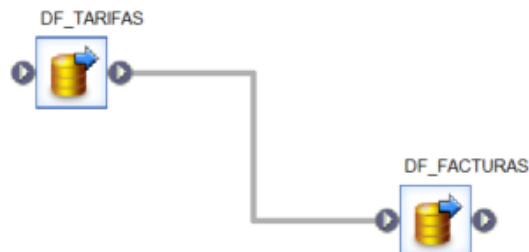


Figura 40. Esquema de construcción de Dataflows dentro del Área Económica

En primer lugar, crearemos un objeto Excel Workbook y los nombraremos como TARIFAS.

Field Name	Data Type	Field Size	Precision	Scale	Content Type	Description
TXT_MARCA	varchar	100				
TXT_MOD_COCHE	varchar	100				
TARIFA	decimal		10	2		

Figura 41. Plantilla creación Excel Wordbook TARIFAS

Este será el origen de los datos de la ETL junto a la tabla coches

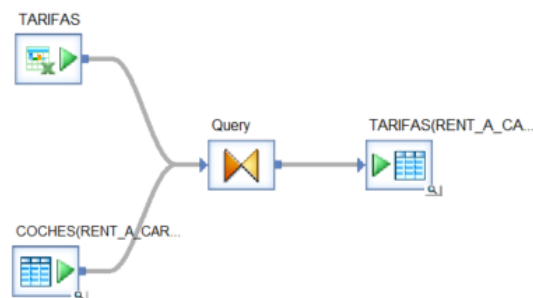


Figura 42. ETL Tarifas

Mediante una query seleccionaremos las columnas a cargar y la unión entre la tabla Coches y el documento Excel.

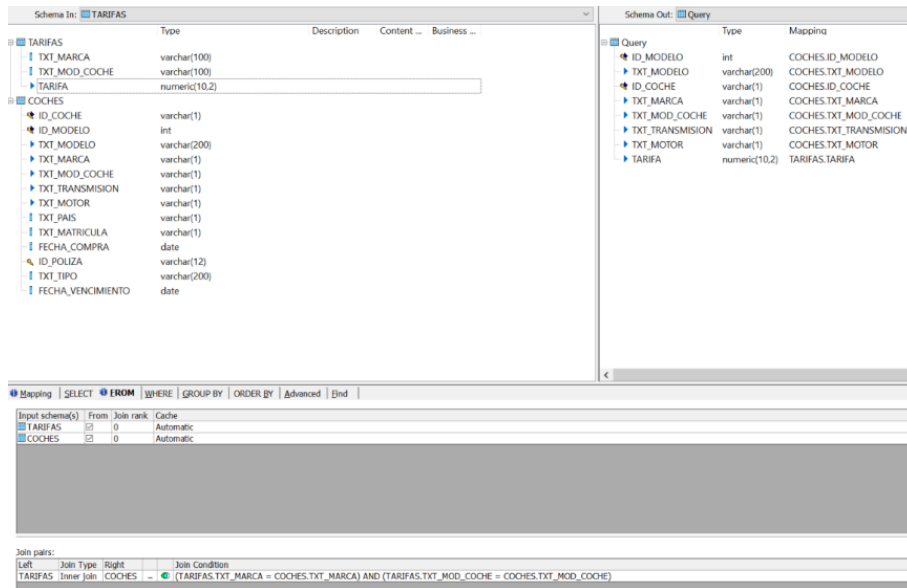


Figura 43. Query Tarifas

Como se ve en la imagen, se unirán por los campos TXT\_MARCA y TXT\_MOD\_COCHE y no se le aplicará ningún tipo de filtro.

Para la generación de la tabla Facturas contaremos con la siguiente ETL

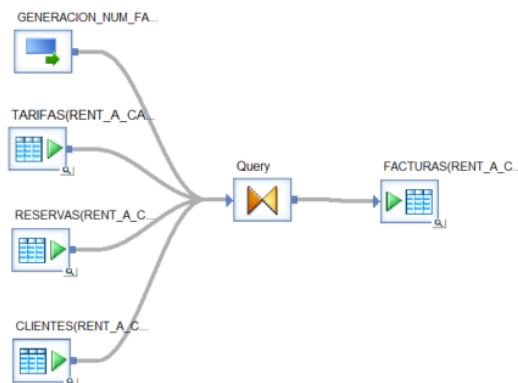


Figura 44. ETL Facturas

Como vemos en la imagen, además de las tablas Tarifas, Reservas y Clientes, tenemos un nuevo elemento en el programa llamado Row Generation, aunque lo renombramos como GENERACION\_NUM\_FACTURA. Este elemento crea una columna donde asigna un número secuencial a cada registro que se inserte, en nuestro caso del 1 al 10000.

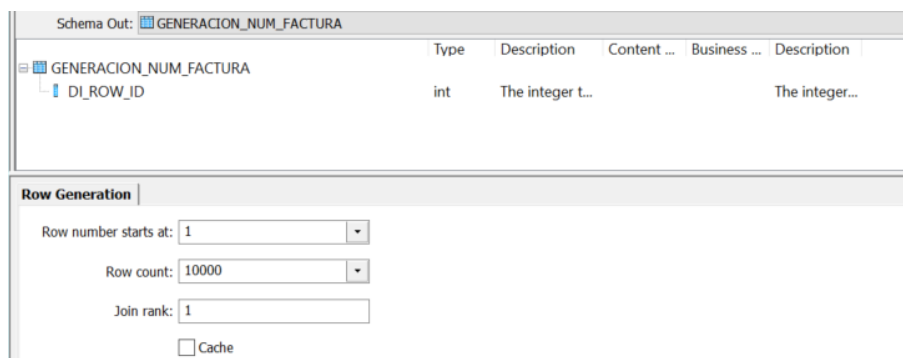


Figura 45. Elemento Row Generation



Para unir todos los elementos de origen tenemos la siguiente query

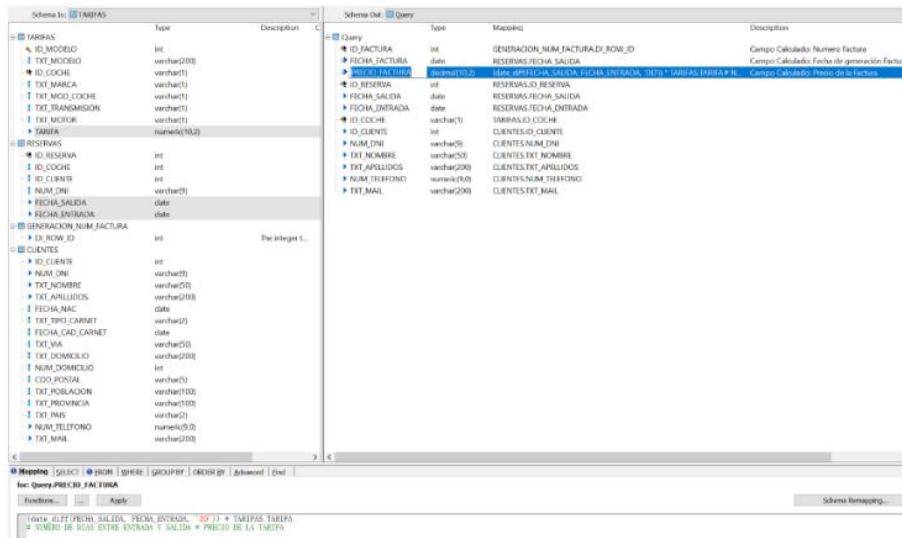


Figura 46. Query Facturas

Como se ve en la columna ID\_FACTURA, se le asigna la columna del objeto GENERACION\_NUM\_FACTURA. Para la fecha de la factura se le asignará la fecha salida de la reserva y para el campo que se le asigna el precio de la factura, realizaremos un cálculo donde se multiplicará la tarifa por los días que durará el alquiler (para ello utilizaremos la función date\_diff como se indica en la parte inferior).

Para unir todas las tablas utilizaremos el campo ID\_CLIENTE entre CLIENTES y RESERVAS y el campo ID\_COCHE entre RESERVAS y TARIFAS.

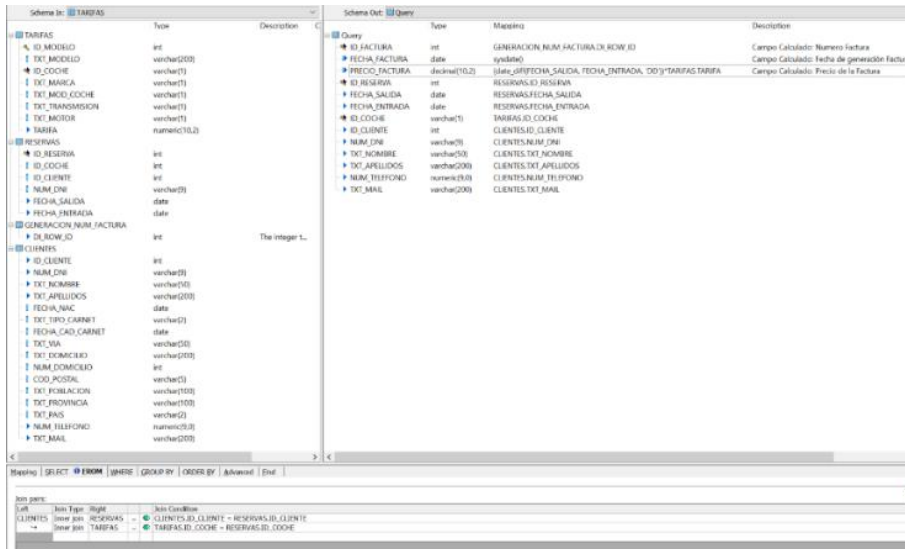


Figura 47. Query Facturas, con detalle en la unión entre tablas de origen

Con esta última carga ya tendremos los datos en todas las tablas necesarias identificadas en el modelo de datos.

Para añadir nuevos registros o modificar los existentes, ya se podrá hacer modificando la tabla en base de datos. En caso, de que se quiera cargar nuevos datos desde documentos Excel, se ejecutaría este proceso y estos datos serían insertados en las tablas.

#### 4.4 Seguridad de datos

Para gestionar la seguridad de los datos, en todos los programas relacionados (creación de modelos, creación de ETLs, calidad de datos, acceso y modificación de bases de datos y análisis de datos) se crearán los usuarios y se le asignarán los permisos correspondientes ya sean usuarios de consulta (con permisos solo de lectura) o usuarios administradores (con permisos de lectura, escritura y modificación)

Por ejemplo, un usuario del área de administración podrá tener permisos para modificar la tabla en base de datos, pero solo de consulta en las otras herramientas.

Un usuario del área analítica podrá tener acceso de administración las herramientas de análisis, bases de datos y calidad de los datos para crear sus cuadros de mandos informativos.

Por otra parte, al tener bien identificadas las columnas en los modelos de datos se pueden indicar si son datos personales, económicos o sensibles para aplicar las medidas de protección adecuadas.

#### 4.5 Datos maestros y de referencia

Mediante la creación del modelo de datos, se ha creado una tabla de referencia para cada área:

- Tabla CLIENTES para el área de Administración
- Tabla RESERVAS para el área de Recepción
- Tabla COCHES para el área de Taller
- Tabla FACTURAS para el área Económica
- Para el área analítica, se podrá crear en BBDD las vistas necesarias partiendo de estas tablas para la realización de cualquier análisis.

Se ha creado una tabla en base de datos que contiene los datos del estándar internacional de normalización ISO 3166-1 que proporciona códigos para los nombres de países y otras dependencias administrativas. Esta tabla la utilizaremos en las reglas de calidad.

También podemos decir que es un modelo estable y robusto ya que no se esperan nuevas fuentes de entradas de datos.

#### 4.6 Almacén de datos e inteligencia de negocio

Con las tablas creadas en base de datos según lo indicado en el modelo de datos, podemos crear, en esta misma base de datos, varias vistas que relacionen estos datos para crear análisis del negocio.

Se han creado 2 vistas de nombre COCHES\_RESERVA y FACTURAS\_CLIENTES y que contienen información de las tablas Coches, Facturas, Reservas y Clientes, Facturas, Coches respectivamente.

```
SELECT
COC.ID_COCHE, COC.TXT_MODELO, COC.TXT_MARCA, COC.TXT_MOD_COCHE, COC.TXT_TRANSMISION, COC.TXT_MOTOR,
RES.ID_RESERVA, RES.FECHA_SALIDA, RES.FECHA_ENTRADA,
DATEDIFF(DAY, FAC.FECHA_SALIDA, FAC.FECHA_ENTRADA) AS DIAS_ALQUILADOS, FAC.ID_FACTURA, FAC.PRECIO_FACT
FROM COCHES AS COC
INNER JOIN FACTURAS AS FAC ON FAC.ID_COCHE = COC.ID_COCHE
INNER JOIN RESERVAS AS RES ON RES.ID_RESERVA = FAC.ID_RESERVA
```

Figura 48. Código de la vista COCHES\_RESERVA

ID_COCHE	TXT_MODELO	TXT_MARCA	TXT_MOD_COCHE	TXT_TRANSMISION	TXT_MOTOR	ID_RESERVA	FECHA_SALIDA	FECHA_ENTRADA	DIAS_ALQUILADOS	ID_FACTURA	PRECIO_FACT
1	COMPACTO	FIAT	PANDA	MANUAL	GASOLINA	1	2021-01-01	2021-01-07	6	1	450.00
12	FURGONETAS	PEUGEOT	EXPERT	MANUAL	DIESEL	2	2021-01-05	2021-01-10	5	2	1250.00
8	SUV/CROSSOVER	RENAULT	CAPTUR	MANUAL	DIESEL	3	2021-02-01	2021-02-15	14	3	1750.00
5	MEDIANO	SEAT	LEON	MANUAL	DIESEL	4	2021-02-15	2021-03-01	14	4	1750.00
13	ELECTRICO	RENAULT	ZOE	AUTOMATICO	ELECTRICO	5	2021-03-01	2021-03-31	30	5	6000.00
7	SUV/CROSSOVER	PEUGEOT	2008	MANUAL	DIESEL	6	2021-03-01	2021-03-15	14	6	1750.00
15	TRABAJO	MERCEDES-BENZ	VITO	MANUAL	DIESEL	7	2021-04-01	2021-05-05	34	7	11900.00
11	FURGONETAS	OPEL	ZAFIRA	MANUAL	DIESEL	8	2021-04-15	2021-05-15	30	8	7500.00
10	MONOVOLUMEN	BMW	SERIE 2	MANUAL	GASOLINA	9	2021-05-02	2021-06-01	30	9	5250.00
1	COMPACTO	FIAT	PANDA	MANUAL	GASOLINA	10	2021-05-15	2021-08-15	92	10	6900.00

Figura 49. Registros ejecución vista COCHES\_RESERVA

```

SELECT
FAC.ID_RESERVA, FAC.FECHA_FACTURA, FAC.PRECIO_FACT, FAC.FECHA_SALIDA, FAC.FECHA_ENTRADA,
CLI.ID_CLIENTE, CLI.NUM_DNI, CLI.TXT_NOMBRE, CLI.TXT_APELLIDOS, CLI.COD_POSTAL, CLI.TXT_POBLACION, CLI.NUM_TELEFONO, CLI.TXT_MAIL,
COC.ID_COCHE, COC.TXT_MATRICULA, COC.TXT_MARCA, COC.TXT_MOD_COCHE

FROM CLIENTES AS CLI
INNER JOIN FACTURAS AS FAC ON CLI.ID_CLIENTE = FAC.ID_CLIENTE
INNER JOIN COCHES AS COC ON FAC.ID_COCHE = COC.ID_COCHE
    
```

Figura 50. Código de la vista FACTURAS\_CLIENTES

ID_RESERVA	FECHA_FACTURA	PRECIO_FACT	FECHA_SALIDA	FECHA_ENTRADA	ID_CLIENTE	NUM_DNI	TXNOMBRE	TXN_APELLIDOS	COD_POSTAL	TXN_POBLACION	NUM_TELEFONO	TXN_MAIL	ID_COCHE	TXN_MATRICULA	TXN_MARCA	TXN_MOD_COCHE
1	2021-01-01	400.00	2021-01-01	2021-01-07	10	60000000	NOMBRE10	APELLIDO10 APELLIDO30	46000	VALENCIA	123456789	NOMBRE10@CORP.ES	1	987654321	PAT	PANDA
2	2021-01-05	1200.00	2021-01-05	2021-01-10	1	012345678	NOMBRE1	APELLIDO1 APELLIDO21	46002	VALENCIA	111111111	NOMBRE1@CORP.ES	12	0007654	PEUGEOT	EXPERT
3	2021-02-01	1700.00	2021-02-01	2021-02-15	7	20202020	NOMBRE7	APELLIDO7 APELLIDO07	50005	VALENCIA	777777777	NUL	8	0003456	RENAULT	CAPTUR
4	2021-02-15	1700.00	2021-02-15	2021-03-01	20	412345678	NOMBRE20	APELLIDOS APELLIDO40	46610	GUARDASUARIA	500000000	NOMBRE20@CORP.ES	5	9876543	SEAT	LEON
5	2021-03-01	6000.00	2021-03-01	2021-03-31	8	33333333	NOMBRE8	APELLIDOS APELLIDO08	NUL	BARCELONA	888888888	NOMBRE8@CORP.ES	13	0000000	RENAULT	ZOE
6	2021-03-01	1700.00	2021-03-01	2021-03-15	11	88888888	NOMBRE11	APELLIDO11 APELLIDO11	46002	VALENCIA	123123123	NOMBRE11@CORP.ES	7	0002345	PEUGEOT	2008
7	2021-04-01	11900.00	2021-04-01	2021-05-05	13	88888888	NOMBRE13	APELLIDO13 APELLIDO13	46000	ALZIRA	78978978	NOMBRE13@CORP.ES	15	9101010	MERCEDES BENZ	VITO
8	2021-04-15	7000.00	2021-04-15	2021-05-15	15	01010101	NOMBRE15	APELLIDO15 APELLIDO15	46000	POLINA DE KOOPER	90104021	NOMBRE15@CORP.ES	11	0000000	ORIEL	ZAFIRA
9	2021-05-02	5200.00	2021-05-02	2021-05-01	6	11111111	NOMBRE6	APELLIDOS APELLIDO06	03003	VALENCIA	NUL	NOMBRE6@CORP.ES	10	0000000	BMW	SERIE 2
10	2021-05-15	6000.00	2021-05-15	2021-05-15	12	77777777	NOMBRE12	APELLIDO12 APELLIDO12	46000	BENICULL DE XUGUER	40404040	NOMBRE12@CORP.ES	1	9898322	PAT	PANDA

Figura 51. Registros ejecución vista FACTURAS\_CLIENTES

El resultado de estas vistas se ha utilizado en el programa Microsoft Power BI para crear varios cuadros de mando con informes de análisis.

Para la vista COCHES\_RESERVA se ha creado un cuadro de mandos donde se analiza por los distintos Modelos, Marcas, Transmisiones de los coches y tipos de motores el uso por días y las ganancias que generan.



Figura 52. Cuadro de mandos coches días y ganancia

En la parte superior obtenemos las relaciones por días, en la parte inferior las ganancias según cada tipo. En la parte media, tenemos los distintos valores de cada registro que se pueden utilizar para filtrar el resultado de las gráficas. En la siguiente imagen tenemos la misma gráfica pero con el filtro Motor Diesel activado.



Figura 53. Cuadro de mandos coches días y ganancia filtrado

Para la vista FACTURAS-CLIENTES se ha realizado el siguiente cuadro de mandos

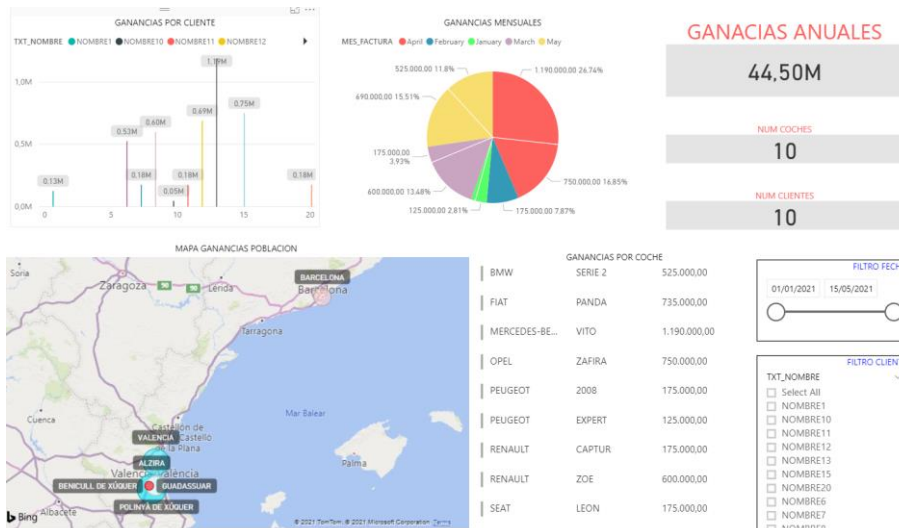


Figura 54. Cuadro de mandos ganancias por clientes

Donde se muestran las ganancias por cliente, las ganancias mensuales y anuales. El total de coches alquilados este año también el total de clientes. Las ganancias generadas por población del cliente y por coche. También disponemos de filtros de fecha y de cliente para tener más información.

Se podrán crear en base de datos las vistas necesarias para realizar tantos informes como se requieran.

## 4.7 Calidad de datos

Como hemos indicado anteriormente, se han definido unas tablas de referencia para cada área de negocio, por tanto, se aplicará el análisis de la calidad de los datos a estas.

### 4.7.1 Perfilado de datos

El primer paso, una vez tenemos cargadas las tablas en el programa, será realizar el perfilado de datos de estas tablas. Este perfilado nos indicara las distintas propiedades, valores, longitudes, la completitud y la distribución de estos datos. Lo que hará que tengamos una vista resumida del contenido total de registros de la tabla y podamos identificar posibles reglas de calidad a aplicar en las tablas.

El perfilado de la tabla CLIENTES del área de Administración sería el siguiente

Figura 55. Perfilado datos tabla Clientes

Este perfilado nos puede mostrar resultados como, por ejemplo, que existen distintos campos con un porcentaje de campos nulos (5% del total en los campos FECHA\_CAD\_CARNET, FECHA\_NAC, NUM\_DOMICILIO, NUM\_TELEFONO, TXT\_APELLIDOS, TXT\_NOMBRE y TXT\_TIPO\_CARNET). Estos campos nulos serían representativos para crear reglas de calidad para identificar que campo es el que falta.

Además, con el campo Pattern podemos identificar patrones como podrían ser los de la columna NUM\_DNI

Figura 56. Perfilado datos tabla Clientes, columna NUM\_DNI

Con la información del patrón podemos comprobar que hay el 90% que siguen la fórmula de 8 dígitos seguido de una letra y un 10% que siguen el formato letra, seguida de 7 dígitos seguida de otra letra (donde los 9 significan cifras y las X letras).

Para la tabla COCHES del área de Taller tendríamos

Figura 57. Perfilado datos tabla Coches

Para la tabla FACTURAS del área Económica

Figura 58. Perfilado datos tabla Facturas

También del área económica tenemos la tabla de TARIFAS



Category	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9
BB	1	10	6.76						
BB_MODEL0	vertr001	4	41						
BB_MODEL1	vertr002	75.00	100.00	272.44					
BB_MODEL2	vertr003	8969	9541						
BB_MODEL3	vertr004	COMERC.	TRABAJ0						
BB_MODEL4	vertr005	2968	707						
BB_MODEL5	vertr006	07896	046026						
BB_MODEL6	vertr007	4	10	6.90					

Figura 59. Perfilado datos tablas Tarifas

#### 4.7.2 Reglas de calidad

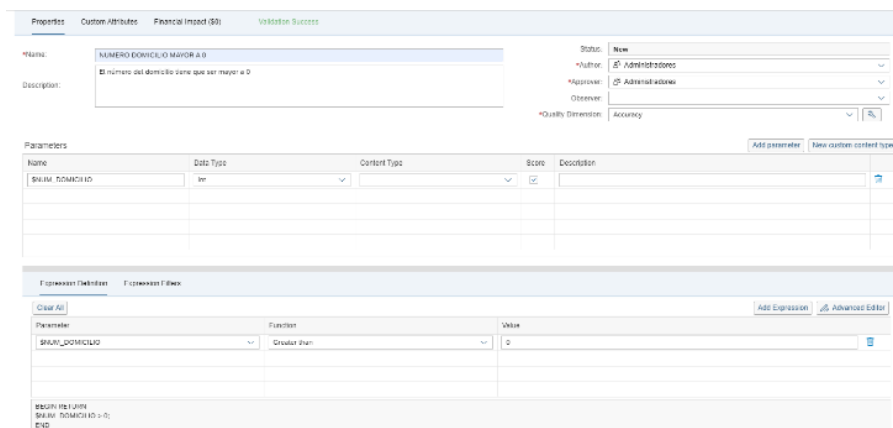
Analizando el resultado de estos perfilados, se ha decidido crear reglas de calidad.

Estas reglas de calidad obtienen una puntuación respecto a los registros totales y registros erróneos, se ha marcado un umbral de calidad en el que estar por debajo de 7 se marcará el resultado en rojo y será necesaria la corrección inmediata. Entre 7 y 9 aparecerá el resultado en amarillo indicando que el número de errores requiere una atención más especial y superior a 9 aparecerá en color verde, donde el número de errores será mínimo y habrá que corregirlos, pero no será tan urgente.

Se han creado reglas para las siguientes dimensiones:

- Accuracy (Precisión): Se comprobará que el número de los domicilios es superior a 0

Para crear esta regla se asignará un parámetro llamado NUM\_DOMICILIO y en el apartado de Expression Definition se le indicara que este parámetro tiene que ser superior a 0. En el apartado Quality Dimension indicamos Accuracy.



The screenshot shows the 'Properties' tab of a rule definition. The name is 'NUMERO DOMICILIO MAYOR A 0'. The description is 'El numero del domicilio tiene que ser mayor a 0'. The quality dimension is set to 'Accuracy'. The parameters table is as follows:

Nombre	Data Type	Content Type	Score	Description
NUM_DOMICILIO	int			

The expression definition is: `NUM_DOMICILIO > 0`.

Figura 60. Definición regla Accuracy

Una vez tengamos creada la regla, se le asignaran al parámetro NUM\_DOMICILIO que columnas queremos que se comprueben con la regla. En este caso, de la tabla Clientes el campo NUM\_DOMICILIO.



The screenshot shows the 'Rule String Editor' with the following configuration:

- Rule: NUMERO DOMICILIO MAYOR A 0
- Source: BB00\_TEST.de.CLIENTES
- Parameter: NUM\_DOMICILIO
- Field: NUM\_DOMICILIO

A color-coded progress bar at the bottom indicates the rule's status, showing it is currently in a yellow state.

Figura 61. Asignación de columnas para analizar

- **Completeness (Compleitud):** Se comprobarán si las columnas tienen algún registro vacío y también si las tarifas están rellenas.

Para comprobar que las columnas están rellenas, se ha creado un parámetro llamado CAMPO que analizará que los registros de las columnas indicadas no sean nulos o sean distintos a vacío. En el campo Quality Dimensión se elegirá la dimensión Completeness.

**Figura 62. Definición regla Completeness Campos Rellenos**

Al parámetro campo se asignarán columnas de las tablas Clientes y Coches

Rule Name:	CAMPOS RELLENOS			
Description:	Se comprueba que los campos están rellenos			
Author:	Administradores	Approver:	Administradores	Observer:
<b>Rule Bindings</b>				
<input type="checkbox"/>	Connection	Schema	Table	Columns
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_NOMBRE
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_APELLIDOS
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	COD_POSTAL
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_POBLACION
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_TIPO_CARNET
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	NUM_DNI
<input type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_DOMICILIO
<input checked="" type="checkbox"/>	BBDD_TEST	dbo	CLIENTES	TXT_PROVINCIA
<input type="checkbox"/>	BBDD_TEST	dbo	COCHES	TXT_MATRICULA
<input type="checkbox"/>	BBDD_TEST	dbo	COCHES	TXT_MODELO
<input type="checkbox"/>	BBDD_TEST	dbo	COCHES	ID_POLIZA
<input type="checkbox"/>	BBDD_TEST	dbo	COCHES	TXT_MOTOR
<input type="checkbox"/>	BBDD_TEST	dbo	COCHES	TXT_TIPO

**Figura 63. Asignación de columnas a analizar en la regla Campos Rellenos**

Para la regla de precio relleno, se hará la misma comprobación, pero en este caso el parámetro es de tipo decimal

Figura 64. Definición regla Completeness Precios Rellenos

Se le asignaran columnas de las tablas Facturas y Tarifas

Connection	Schema	Table	Columns
BBDD_TEST	dbo	FACTURAS	PRECIO_FACT
BBDD_TEST	dbo	TARIFAS	TARIFA

Figura 65. Asignación columnas a analizar precios

- Conformity (Conformidad): Se comprobará que el campo DNI de los clientes cumplen los formatos de DNI o NIE

Se creará un campo llamado DNI y este se comprobará con el patrón que tienen los DNI y NIE, corresponden en la búsqueda del patrón 9 a números y X para las letras.

Figura 66. Definición regla Conformity





Figura 70. Definición regla Timeliness

Se le asigna a la columna FECHA\_VENCIMIENTO de la tabla Coches que tiene almacenadas las fechas de los seguros de cada coche.

Connection	Schema	Table	Columns	Score
<input checked="" type="checkbox"/>	dbo	COCHES	FECHA_VENCIMIENTO	9.37

ID_COCHE	ID_MODELO	TXT_MODELO	TXT_MARCA	TXT_MOD_COC...	TXT_TRANSMIS...	TXT_MOTOR	TXT_PAIS
11	F	FURGONETAS	OPEL	ZAFIRA	MANUAL	DIESEL	ES

Figura 71. Asignación columnas a analizar en la regla Timeliness

- Uniqueness (Unicidad): Se comprueba que un número de DNI no esté repetido en varios clientes.

Se comprueba que el campo DNI es único

Figura 72. Definición regla Uniqueness

### 4.7.3 Resultados de las reglas de calidad

Para obtener los resultados, se tiene que crear una tarea para las reglas. En ella, añadimos las tablas que estén presentes en las reglas de calidad (si no añadimos una tabla, las columnas de esta no obtendrían resultados). Cuando tenemos ejecutada esta tarea, nos mostraría el resultado de cada columna.

- Regla Numero Domicilio mayor a 0 tenemos el siguiente resultado

Rule Name: NUMERO DOMICILIO MAYOR A 0										
Description: <input type="checkbox"/> número del domicilio tiene que ser superior a 0										
Author: Administradores			Approver: Administradores			Observer:				
Rule Bindings										
Connection	Schema	Table	Columns	Score	From Last	Failed	Total	Low	High	
<input type="checkbox"/>	db0	CLIENTES	NUM_DOMICILIO	9.50		1	20	7	9	

Figura 73. Resultado regla Accuracy

Como se ve en el resultado, solo hay 1 de los 20 registros que el domicilio es menor a 0, la razón que tiene el campo vacío.

- Regla de Campos Rellenos tenemos:

Rule Name: CAMPOS RELLENOS										
Description: <input type="checkbox"/> Se comprueba que los campos están rellenos										
Author: Administradores			Approver: Administradores			Observer:				
Rule Bindings										
Connection	Schema	Table	Columns	Score	From Last	Failed	Total	Low	High	
<input type="checkbox"/>	db0	CLIENTES	TXT_NOMBRE	9.50		1	20			
<input type="checkbox"/>	db0	CLIENTES	TXT_APELLIDOS	9.50		1	20			
<input type="checkbox"/>	db0	CLIENTES	COD_POSTAL	10.00		0	20			
<input type="checkbox"/>	db0	CLIENTES	TXT_FOBOLACION	10.00		0	20			
<input type="checkbox"/>	db0	CLIENTES	TXT_TIPO_CARNET	9.50		1	20			
<input type="checkbox"/>	db0	CLIENTES	NUM_DNI	10.00		0	20			
<input type="checkbox"/>	db0	CLIENTES	TXT_DOMICILIO	10.00		0	20			
<input type="checkbox"/>	db0	CLIENTES	TXT_PROVINCIA	10.00		0	20			
<input type="checkbox"/>	db0	COCHES	TXT_MATRICULA	10.00		0	16			
<input type="checkbox"/>	db0	COCHES	TXT_MODELO	10.00		0	16			
<input type="checkbox"/>	db0	COCHES	ID_POLIZA	10.00		0	16			
<input type="checkbox"/>	db0	COCHES	TXT_MOTOR	10.00		0	16			
<input type="checkbox"/>	db0	COCHES	TXT_TIPO	10.00		0	16			

Figura 74. Resultado regla Completeness (Campos Relleno)

Donde se observa que en la tabla Clientes, tiene columnas con registros vacíos o nulos.

- Regla Precios Rellenos

Rule Name: PRECIOS RELLENOS										
Description: <input type="checkbox"/> SE COMPRUEBA QUE LOS PRECIOS DE LA TARIFA Y DE LA FACTURA ESTAN RELLENOS										
Author: Administradores			Approver: Administradores			Observer:				
Rule Bindings										
Connection	Schema	Table	Columns	Score	From Last	Failed	Total	Low	High	
<input type="checkbox"/>	db0	FACTURAS	PRECIO_FACT	10.00		0	10			
<input type="checkbox"/>	db0	TARIFAS	TARIFA	10.00		0	16			

Figura 75. Resultado regla Completeness (Precios Relleno)

- La regla Formato DNI Correcto

Rule Name: FORMATO DNI CORRECTO										
Description: <input type="checkbox"/> Se comprueba que el campo DNI tiene el patrón correcto (ya sea para DNI o para NIE)										
Author: Administradores			Approver: Administradores			Observer:				
Rule Bindings										
Connection	Schema	Table	Columns	Score	From Last	Failed	Total	Low	High	
<input type="checkbox"/>	db0	CLIENTES	NUM_DNI	10.00		0	20			
<input type="checkbox"/>	db0	FACTURAS	NUM_DNI	10.00		0	10			

Figura 76. Resultado regla Conformity

- La regla País Matricula Correcto

Rule Name: PAIS MATRICULA CORRECTO							
Description: SE COMPRUEBA QUE EL PAIS DE LA MATRICULA SEA VALIDO							
Author: Administradores	Observer:						
Rule Bindings							
Connection	Schema	Table	Columns	Score	From Last	Failed	Total
<input checked="" type="checkbox"/> BBDD_TEST	dbo	COCHE	TXT_PAIS	10.00		0	16

Figura 77. Resultado regla Consistency

- La regla Seguro en Validez

Rule Name: SEGURO EN VALIDEZ									
Description: Se comprueba que la fecha del seguro es superior o igual a hoy									
Author: Administradores	Observer:								
Rule Bindings									
Connection	Schema	Table	Columns	Score	From Last	Failed	Total	Low	High
<input checked="" type="checkbox"/> BBDD_TEST	dbo	COCHE	FECHA_VENCIMIENTO	0.37		1	16	7	9

ID_COCHE	ID_MODELO	TXT_MODELO	TXT_MARCA	TXT_AYUD_COCH.	TXT_TRANSMIS.	TXT_MOTOR	TXT_PAIS	TXT_MATRICULA	FECHA_COMPRA	ID_POLITA	TXT_TIPO	FECHA_VENCL
11	F	FUSIONETAS	OPEL	ZAFIRA	MANUAL	DIESEL	ES	000000*	2021.01.02	benfjgym1251	TODOS RIESGO	2021.01.02

Figura 78. Resultado regla Timeliness

Como vemos en el resultado de la regla, el coche con ID\_COCHE = 11, tiene el seguro caducado.

#### 4.7.4 Resultados de las reglas de calidad

Los cuadros de mandos agrupan las reglas para poder ofrecer una visualización más sencilla. Se pueden crear tantos cuadros de mandos como sean necesarios.

La vista de la ventana de visualización muestra los resultados agrupados por cuadro de mando y dimensión.

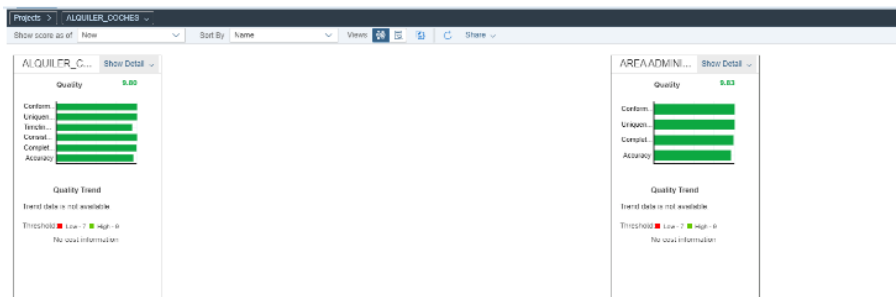


Figura 79. Cuadro de mandos, vista agrupada

Si queremos mostrar los resultados con más detalle:

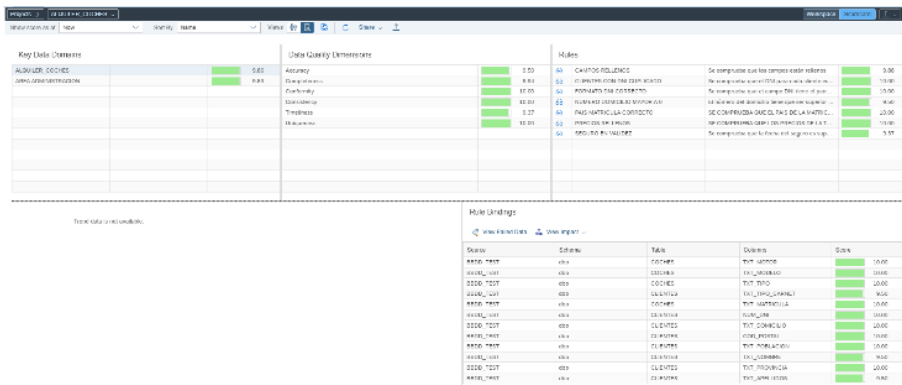


Figura 80. Cuadro de mandos, vista detalle

En nuestro caso hemos creado un cuadro de mando llamado ALQUILER\_COCHES que agrupa todas las reglas descritas anteriormente y otro llamado AREA ADMINISTRACION que agrupa solamente los resultados de las reglas que corresponden únicamente a las tablas del área de administración.

Para el cuadro de mando ALQUILER\_COCHES tenemos el siguiente cuadro de mando

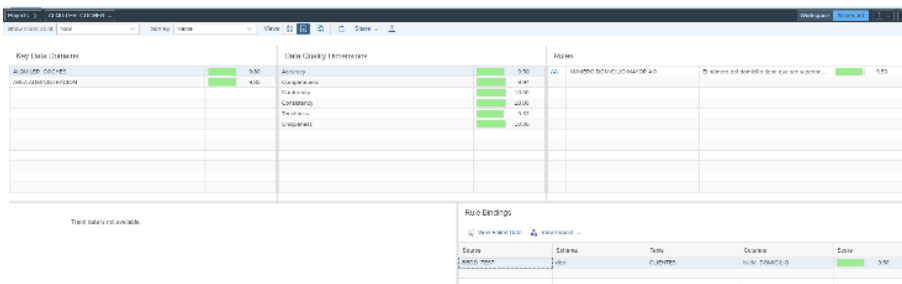


Figura 81. Cuadro de mandos ALQUILER\_COCHES

Para el área de administración el siguiente.

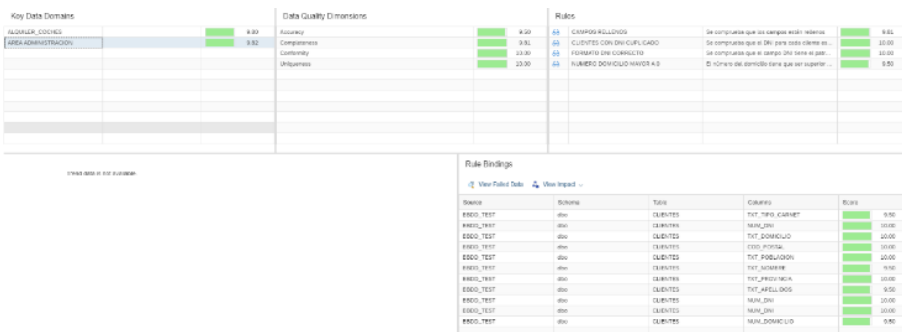


Figura 82. Cuadro de mandos AREA ADMINISTRACION

Los errores mostrados en los cuadros de mando se pueden descargar en documento Excel, donde se descargaría una hoja por regla que contenga un registro erróneo.

Todas estas reglas y formas de mostrar los registros que hemos identificados como no validos tienen como fin la corrección de estos para que el conjunto de datos almacenado sea lo más fiable posible.



## Capítulo 5. Conclusión

Mediante la aplicación de los distintos elementos de gobierno de dato, se ha pasado de mantener los datos en archivos Excel locales a tener estos datos identificados en tablas, organizadas con un modelo de datos. Estos datos están guardados en una base de datos accesible para toda la organización aplicando unas medidas de seguridad en el acceso y permisos. También se han detectado datos que no tienen registro o estos son erróneos, con las herramientas de calidad del dato, que de otra forma no se hubiera podido detectar.

Con la aplicación de estas medidas se ha conseguido una información más fiable, accesible y robusta. Haciendo que aumente el valor de la empresa y se tenga una visión más clara de ella. Con todo ello, se cumplen los objetivos marcados para este proyecto.



## Capítulo 6. Bibliografía

- [1] DAMA Internacional, “DAMA-DMBOK: Guía de fundamentos para la gestión de datos” Edición Impresa 2010
- [2] DAMA España, <https://www.damaspain.org/> [Online]
- [3] SAP Help Portal, “PowerDesigner Documentation”,  
<https://help.sap.com/viewer/abd3434b4987485c92057ab9392aadbe/16.6.6/en-US/c7da64d26e1b1014babfeec490561c31.html> [Online]
- [4] SAP Help Portal SAP PowerDesigner, “Building Data Models”  
<https://help.sap.com/viewer/856348b84a7c479489d5172a630f014d/16.6.6/en-US/c80aba926e1b1014a37293e03aeed040.html> [Online]
- [5] SAP Data Services, “Designer Guide”  
[https://help.sap.com/doc/PRODUCTION/5b775a8b23ce4bc49182298ddcb3b6e9/4.2.8/en-US/ds\\_42\\_designer\\_en.pdf](https://help.sap.com/doc/PRODUCTION/5b775a8b23ce4bc49182298ddcb3b6e9/4.2.8/en-US/ds_42_designer_en.pdf) [Online]
- [6] SAP Information Steward, “User Guide”  
[https://help.sap.com/doc/18ec99b2d06449c5b8b79c784d5a3af9/4.2.8/en-US/is\\_42\\_user\\_en.pdf](https://help.sap.com/doc/18ec99b2d06449c5b8b79c784d5a3af9/4.2.8/en-US/is_42_user_en.pdf) [Online]
- [7] Microsoft Power BI, “Documentación de Power BI”  
<https://docs.microsoft.com/es-es/power-bi/> [Online]
- [8] Reglamento General de Protección de Datos, <https://rgpd.es/> [Online]
- [9] ISO - International Organization for Standardization, “Norma ISO 3166-1”  
<https://www.iso.org/obp/ui/#iso:std:iso:3166:-1:ed-4:v1:en> [Online]
- [10] Informatica The Data Integration Company, “Gestión de datos maestros (MDM)”  
[https://www.informatica.com/content/dam/informatica-com/es/collateral/brochure/mdm\\_brochure.pdf](https://www.informatica.com/content/dam/informatica-com/es/collateral/brochure/mdm_brochure.pdf) [Online]
- [11] IBM, “Modelos lógicos de datos”  
<https://www.ibm.com/docs/es/ida/9.1.2?topic=modeling-logical-data-models> [Online]
- [12] Mario G. Piattini Velthuis, Ismael Caballero Muñoz-Reja, Ana Isabel Gómez Carretero, Fernando Gualo Cejudo, Jorge Merino García, Bibiano Rivas García, “Calidad de Datos” Editorial Ra-Ma, 2018