



UNIVERSIDADE DA CORUÑA

Facultad de Economía y Empresa

Trabajo de
Fin de grado

**Bases de datos y
minería de datos
en entornos
empresariales: el
caso de la
empresa Carlin.**

Vanesa Décima Barravecchia

Tutores: Saavedra Places,
Ángeles; González Soto, Julio.

Grado en Ciencias Empresariales

Año 2014

Resumen

La tarea de acceder a la información, de manera eficiente y rápida, está cobrando cada vez mayor interés por parte de las empresas. Toda organización debería contar con un proceso analítico estructurado y formal, que facilite el proceso de toma de decisiones.

Para ello se diseñará un sistema de información que se ajuste a las necesidades de una empresa en concreto, mediante la aplicación del modelo Entidad-Relación.

Este modelo hará posible el estudio del funcionamiento interno de una empresa, proporcionando a través del análisis y explotación de los datos una nueva visión de negocio. Algunas de las técnicas que servirán de apoyo en el desarrollo de este trabajo serán la minería de datos (*data mining*) y dentro de los almacenes de datos (*data warehouses*), las técnicas OLAP (*On-Line Analytical Processing*) que permitirán clasificar los datos en diferentes dimensiones, con distintas combinaciones para obtener información no evidente y útil en la toma de decisiones.

Abstract

Companies are actually increasing their interest in accessing quickly and efficiently to information. Every organization should have analytical, structured and formal process in order to make it easier the decision making. To do so, this project will design an information system by applying the ER model. The ER model will make it possible studying the organization from the inside giving a whole new vision of the business process. Some of techniques that will be used in this project are data mining, and from the data warehouses the OLAP techniques. This techniques will make it possible to cluster and clasify the data in order to have valueble hidden information useful in the desicion making.

Palabras clave: minería de datos, bases de datos, toma de decisiones, descubrimiento de conocimiento, almacenes de datos.

Índice

1. INTRODUCCIÓN	7
1.1 Motivación.....	7
1.2 Objetivos del trabajo.....	8
1.3 Estructura de la Memoria.....	10
2. FUNDAMENTOS TEÓRICOS	11
2.1 Metodología de diseño de Bases de Datos.....	14
2.2 Diseño Conceptual: el modelo Entidad-Relación	15
2.2.1 Entidades.....	16
2.2.2 Atributos	17
2.2.3 Identificador	18
2.2.4 Relaciones y restricciones en los tipos de relación	19
2.3 El Modelo Relacional	23
2.3.1 Claves.....	25
2.3.2 Restricciones de integridad.....	26
2.3.3 Reglas de transformación	28
3. PLANIFICACIÓN	31
3.1 Definición de tareas.....	31
3.2 Esfuerzo	33
3.3 Recursos.....	34
3.4 Temporalización	35
3.5 Seguimiento y conclusiones.....	36
4. DISEÑO DE LA BASE DE DATOS	39

4.1 Selección del dominio/minimundo.	39
4.2 Diseño Conceptual	40
4.3 Diseño Lógico	47
4.4 Diseño Físico	50
4.5 Alimentación de la Base de datos	54
4.5.1 Información real	54
4.5.2 Generación de datos.....	60
5. EXPLOTACIÓN DE LA INFORMACIÓN	62
5.1 Almacenes de datos (data warehouses).....	62
5.2 Análisis de datos: Consultas SQL	65
5.2.1 Franquicias	65
5.2.2 Proveedores	71
5.2.3. Artículos	72
6. CONCLUSIONES	76
Bibliografía.....	78

Índice de figuras

Figura 1. Sistema de procesamiento de una Base de Datos.	11
Figura 2. Etapas de diseño de una BD.....	14
Figura 3. Tipo de entidad y conjunto de entidades.	17
Figura 4. Identificador único.	19
Figura 5. Ejemplo de Relación.	19
Figura 6. Participación y Cardinalidad	20
Figura 7. Correspondencia de cardinalidades. (a) Uno a uno. (b) Uno a muchos.	21
Figura 8. Correspondencia de cardinalidades. (a) Muchos a uno. (b) Muchos a muchos.	22
Figura 9. Ejemplo de una entidad débil.	23
Figura 10. Ejemplo de superclave.....	25
Figura 11. Transformación de una relación 1: N, al modelo relacional.	30
Figura 12. Diagrama de Gantt estimado.	35
Figura 13. Diagrama de Gantt final.....	38
Figura 14. Modelo Entidad-Relación de Carlin.....	41
Figura 15. Relaciones BD Carlin.....	51
Figura 16. Diseño de la tabla Artículos.	52
Figura 17. Diseño de la tabla Franquicias.....	52
Figura 18. Diseño de la tabla Proveedores.....	53
Figura 19. Diseño de la tabla líneas de facturas compras.....	53
Figura 20. Alimentación de la tabla proveedores.....	54
Figura 21. Alimentación de la tabla Categorías.....	55
Figura 22. Subcategorías.....	55
Figura 23. Código web Carlin.....	56
Figura 24. Alimentación de la tabla Subcategorías.....	57
Figura 25. Nombres de cada uno de los artículos.	58
Figura 26. Alimentación de la tabla Artículos.....	58
Figura 27. Alimentación de la tabla Franquicias.....	59
Figura 28. Alimentación de la tabla Códigos Postales.	59

Figura 29. Generación de precios.	60
Figura 30. Generación de datos de la tabla líneas compra proveedores.	61
Figura 31. Arquitectura de un almacén de datos.	63
Figura 32. Consulta y gráfico de las compras en verano de las franquicias de Galicia.	66
Figura 33. Consulta y gráfico de las ventas en verano de las franquicias de Galicia.	67
Figura 34. Consulta y gráfico del total de ventas por CC.AA.	68
Figura 35. Consulta y gráfico de la procedencia de los clientes de la franquicia 4.	69
Figura 36. Consulta y gráfico de la evolución de dos franquicias según sus ventas.	70
Figura 37. Consulta y gráfico del gasto en proveedores.	71
Figura 38. Consulta y gráfico de ventas por subcategoría.	72
Figura 39. Marcas que más se venden en la subcategoría de bolígrafos.	73
Figura 40. Regla de asociación entre los portaminas y las minas FABER CASTELL.	74

Índice de tablas

Tabla 1. Símbolos básicos para diagramas de entidad-relación.....	16
Tabla 2. Relación “Departamentos”	24
Tabla 3. Relación no válida por uso de valor múltiple.....	26
Tabla 4. Relación incorrecta.....	27
Tabla 5. Violación de restricción de clave.....	27
Tabla 6. Violación de la integridad referencial.....	28
Tabla 7. Asignación de horas por tarea.....	33

1. INTRODUCCIÓN

1.1 Motivación

En las organizaciones, públicas o privadas, las decisiones se toman de forma continua. Tales decisiones pueden ser más o menos complejas, tener efectos a largo o a corto plazo e involucran a personas, con distintos roles y niveles jerárquicos. Shackle (1966), define la decisión como un corte entre el pasado y el futuro. Como tomar una decisión supone escoger la mejor alternativa de entre las posibles, se necesita información sobre cada una de estas alternativas y sus consecuencias respecto a un objetivo. La información es la materia prima, el *input* de la decisión y una vez tratada adecuadamente dentro del proceso de la toma de decisión se obtiene como *output* la acción de ejecutar.

Para adoptar algunos tipos de decisiones se suelen utilizar modelos. Estos se pueden definir como una representación simplificada de una parte de la realidad. El principal objetivo de un modelo, es permitir una mejor comprensión y descripción de la parte de la realidad que representa. Esa mejor comprensión de la realidad permite tomar mejores decisiones. Por ello, las empresas gestionan grandes flujos de información, y la minería de datos es la herramienta que permite analizar y explotar los grandes conjuntos de datos, Molina López y García Herrero (2006), explican que los datos tal cual se almacenan en las bases de datos no suelen proporcionar beneficios directos; su valor real, reside en la información que podamos extraer de ellos, es decir, información que nos ayude a tomar decisiones o a mejorar la comprensión de los fenómenos que nos rodean. En cambio para Bramer (2013), la minería de datos es la etapa de análisis del proceso de descubrimiento del conocimiento, y se trata de la extracción de información implícita, desconocida y potencialmente útil para la toma de decisiones.

“data rich but knowledge poor”.

1.2 Objetivos del trabajo

El objetivo general del presente trabajo, es diseñar y construir una base de datos que responda a unos requisitos de almacenamiento de información y de explotación de la misma, en el marco de un área de empresa. El diseño de la base de datos operacional, se llevará a cabo a partir del minimundo o universo de discurso¹ de la compañía, con el fin de ofrecer una visión abstracta de la información. Se analizará la situación de la organización a partir de sus objetivos y estrategias futuras.

Se ha seleccionado una empresa que utiliza como modelo de negocio las *franquicias* y se ha demostrado que es la forma de crecimiento empresarial más exitosa en los últimos tiempos. Franquicias es un vocablo cada vez más usado en el país, y esto es debido a que vivimos una realidad económica y los efectos de una globalización que une personas, culturas y economías, exponiéndolas a nuevas formas de hacer negocios y por lo tanto a nuevos retos empresariales.

Mediante esta empresa se pretende extraer información de un conjunto de datos y transformarlo en una estructura comprensible para su uso posterior en la toma de decisiones, aplicando minería de datos.

Las herramientas de software que servirán de apoyo, serán principalmente, un gestor de bases de datos, un gestor de hojas de cálculo y un procesador de texto, que como sabemos hoy en día, son las aplicaciones más usuales en la gestión de las empresas.

En resumen, los objetivos son:

- Diseñar e implementar un sistema de información, que se ajuste lo máximo posible a la realidad.
- Aplicar la tecnología existente y ver su impacto en el ámbito de una empresa.
- Derivar de los datos, información relevante para la toma de decisiones.

¹ Minimundo o universo de discurso: es la representación en una base de datos de algún aspecto del mundo real. Este concepto jugará un papel importante a la hora de diseñar la base de datos del caso Carlin.

- Apoyarse en herramientas de software que permitan visualizar la información de manera clara y concisa.
- Valorar críticamente el conocimiento, la tecnología y la información disponible para resolver problemas en el día a día de una organización.
- Estudiar como el uso de las bases de datos, sirve de apoyo a nivel gerencial en el ámbito empresarial.

1.3 Estructura de la Memoria

Este trabajo está dividido en cinco grandes bloques. En el primer bloque se hablará de los *fundamentos teóricos*, centrándose en el concepto y en el sistema de procesamiento de una base de datos. A modo de ejemplo, se citarán casos reales.

El segundo bloque constará de tres capítulos bien definidos. El primer capítulo se centra en la *metodología* a seguir en el diseño de las bases de datos. En este punto se dará un repaso general de las distintas fases del diseño de las mismas. En el segundo capítulo se hablará del modelo *Entidad-Relación*, que es el modelo de datos seleccionado para representar el minimundo del caso real de la empresa *Carlin* y en el tercer capítulo, se trazarán las líneas generales del *Modelo Relacional* y su importancia en el diseño lógico de las bases de datos.

El tercer bloque estará dedicado a la *Planificación* del proyecto, en el que de manera detallada se expondrán las tareas, el esfuerzo, los recursos, la temporalización y las desviaciones en la realización del proyecto.

En el cuarto bloque se expondrá el porqué de la selección de la *empresa Carlin* como caso de estudio, sus necesidades de información y sus objetivos futuros. Por otro lado, se elaborará la hoja de especificación de requisitos, es decir, la descripción a alto nivel del funcionamiento y gestión operacional de la empresa, que conllevará la aplicación de las distintas fases de la metodología de las bases de datos: el *diseño conceptual*, el *diseño lógico* y el *diseño físico*. A continuación, se poblará la base de datos en la medida de lo posible con la información real de la compañía y mediante la generación de datos.

En el quinto bloque se hará una pequeña introducción a la importancia de los *almacenes de datos* y se realizará la *explotación de los datos*. Dicha explotación se apoyará en herramientas como el lenguaje SQL de consultas y el data mining, que harán posible la toma de decisiones a nivel gerencial.

El sexto bloque estará destinado a las *conclusiones*.

2. FUNDAMENTOS TEÓRICOS

Para un usuario final de una aplicación, no es necesario saber cómo están representados los datos internamente, solo le preocupa que su aplicación responda a sus necesidades de manera eficiente y correcta.

La mayoría de las aplicaciones que se desarrollan requieren del almacenamiento en memoria externa de información. Durante estos años se han utilizado diferentes formas de organización. Aquellos sistemas, cuyo objetivo principal es la gestión de datos que pueden ser creados, eliminados y consultados provocaron un interés especial por parte de los desarrolladores de software con vistas a aplicar mecanismos de organización eficaces. Tras la proliferación de los sistemas de gestión de datos, se desarrolló una forma de organización conocida como sistemas de procesamiento de base de datos o sistemas orientados a los datos. Los datos se organizan y mantienen en un conjunto estructurado que no está diseñado para una aplicación concreta, sino que por el contrario, tienden a satisfacer las necesidades de información de toda la organización, necesidades cuya diversidad se ve acentuada con el transcurso del tiempo.

Para procesar los datos almacenados se utiliza un conjunto de programas que se agrupan en un sistema gestor de base de datos (SGBD), que permite incluir, extraer y consultar información.

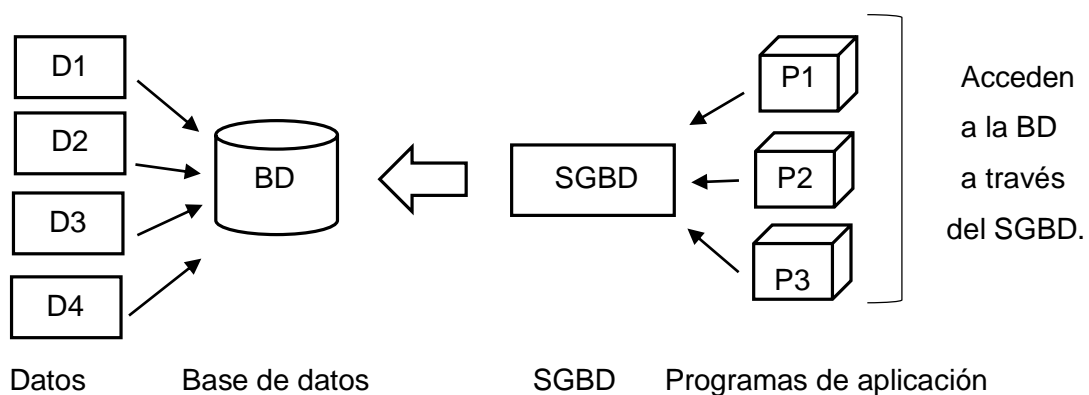


Figura 1. Sistema de procesamiento de una Base de Datos.

El SGBD es el software que interactúa con los programas de aplicación del usuario y con la base de datos.

El término Base de datos aparece en los años 70, y se ha definido como:

“Colección de datos interrelacionados” (Elmasri, 2007).

“Colección o depósito de datos integrados, con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real; los datos que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de estas, y su definición y descripción, únicas para cada tipo de datos, han de estar almacenados junto con los mismos. Los procedimientos de actualización y recuperación, comunes y bien determinados, habrán de ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de datos.” (De Miguel, 1993).

“Una colección compartida de datos lógicamente relacionados, junto con una descripción de estos datos, que están diseñados para satisfacer las necesidades de información de una organización.” (Connolly & Begg, 2005)

Se podría decir que una base de datos, es un depósito centralizado, posiblemente de gran tamaño, compuesto por datos que pueden ser utilizados simultáneamente por múltiples departamentos y usuarios. En lugar de disponer de una serie de archivos desconectados con datos redundantes, todos los datos están integrados, manteniéndose al mínimo las posibles duplicaciones. La base de datos deja de ser propiedad de un departamento y pasa a ser un recurso corporativo compartido. La base de datos almacena no sólo los datos operacionales de la organización, sino también una descripción de dichos datos.

Actualmente, podemos encontrar múltiples casos reales de aplicaciones de las bases de datos:

- Banca: Información de clientes, cuentas, transacciones, préstamos...
- Líneas aéreas: información de clientes, horarios, vuelos, destinos...
- Universidades: información de estudiantes, carreras, horarios, asignaturas...

- Transacciones de tarjetas de crédito: para compras con tarjetas de crédito y generación mensual de extractos.
- Finanzas: para almacenar información sobre grandes empresas, ventas y compras de documentos formales financieros, como bolsa y bonos.
- Producción: para la gestión de la cadena de producción y para el seguimiento de la producción de elementos en las factorías, inventarios de elementos en almacenes y pedidos de elementos.
- Recursos humanos: para información sobre los empleados, salarios, impuestos y beneficios, y para la generación de las nóminas.

2.1 Metodología de diseño de Bases de Datos

El diseño de una base de datos es un proceso complejo que abarca decisiones a muy distintos niveles. La complejidad se controla con más facilidad si se fracciona el problema en distintas fases y se resuelven independientemente, utilizando técnicas específicas. El diseño de un entorno, que responda a las necesidades de la empresa que se está modelando, exige prestar atención a un amplio conjunto de consideraciones y aspectos adicionales del uso esperado de la base de datos, puesto que influyen en gran variedad de opciones de diseño en los niveles físico, lógico y conceptual.



Figura 2. Etapas de diseño de una BD.

La fase inicial del diseño de la base de datos es la caracterización completa de las necesidades de datos de los posibles usuarios. El diseñador de la base de datos debe interactuar con los expertos y los usuarios del dominio para realizar esta tarea. El resultado de esta fase es una especificación de requisitos.

El **diseño conceptual** parte de las especificaciones de requisitos y su resultado es el esquema conceptual de la base de datos. Un esquema conceptual, es una descripción de alto nivel de la estructura de la base de datos, independientemente del SGBD (*Sistema gestor de base de datos*) que se vaya a utilizar para manipularla, que suele ser gráfico y proporciona una visión detallada de la empresa. Un modelo conceptual es un lenguaje que se utiliza para describir esquemas conceptuales. El objetivo del diseño conceptual, es describir el contenido de información de la base de datos y no las

estructuras de almacenamiento que se necesitarán para manejar esta información. En este caso, el modelo conceptual que se utilizará es el diagrama entidad-relación (E-R).

El **diseño lógico** parte del esquema conceptual y da como resultado un esquema lógico. Un esquema lógico es una descripción de la estructura de la base de datos en términos de las estructuras de datos que puede procesar un tipo de SGBD. Un modelo² lógico es un lenguaje usado para especificar esquemas lógicos. El modelo de implementación de los datos suele ser el modelo relacional. El diseño lógico depende del tipo de SGBD que se vaya a utilizar.

El **diseño físico** parte del esquema lógico y da como resultado un esquema físico. Un esquema físico es una descripción de la implementación de una base de datos en memoria secundaria: las estructuras de almacenamiento y los métodos utilizados para tener un acceso eficiente a los datos. Por ello, el diseño físico depende del SGBD concreto y el esquema físico se expresa mediante su lenguaje de definición de datos.

2.2 Diseño Conceptual: el modelo Entidad-Relación

El modelo de datos entidad-relación (E-R), se desarrolló para facilitar el diseño conceptual de bases de datos permitiendo la especificación de un esquema de la empresa que representa la estructura lógica global de la base de datos.

Fue desarrollado por Peter Chen en 1976, para permitir al diseñador expresar las propiedades conceptuales de la BD en un esquema. Chen (1976) describe, “El modelo E-R puede ser usado como una base para una vista unificada de los datos”, adoptando “el enfoque más natural del mundo real que consiste en entidades y relaciones”. Así es, que el modelo entidad-relación está formado por una serie de conceptos que permiten describir la realidad mediante un conjunto de representaciones gráficas y lingüísticas.

² Existen distintos modelos de administración de datos, con frecuencia son utilizados: el modelo de bases de datos jerárquicas, el modelo de bases de datos red y el modelo de bases de datos relacionales. Este último modelo será el que desarrollaremos con más profundidad en el presente trabajo.



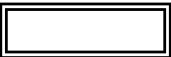
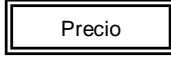



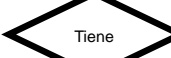




Símbolo	Significado	Ejemplo
	Entidad Fuerte	
	Entidad Débil	
	Atributo	
	Relación	
	Atributo multivaluado	
	Atributo Derivado	

Tabla 1. Símbolos básicos para diagramas de entidad-relación.
Elaboración propia.

2.2.1 Entidades

Una **Entidad**³ es cualquier tipo de objeto sobre el que se quiere guardar información: cosa, persona, concepto abstracto o suceso. Toda entidad tiene un conjunto de propiedades que la identifican, que se denominan atributos.

Toda entidad debe cumplir tres reglas:

- Tener existencia independiente (Elmasri, 2007),
- Debe poder distinguirse de las demás, no pudiendo haber duplicados,
- Tener propiedades que la describan.

Un **tipo de entidad** define una colección (o conjunto) de entidades que tienen los mismos atributos. Mientras que un **conjunto de entidades**⁴ es una colección de entidades del mismo tipo.

³ Ejemplos de entidades pueden ser: coches, casas, clientes, conciertos, excursiones, etc.

⁴ El conjunto de todas las personas que son clientes en un banco dado, por ejemplo, se puede definir como el conjunto de entidades *cliente*.

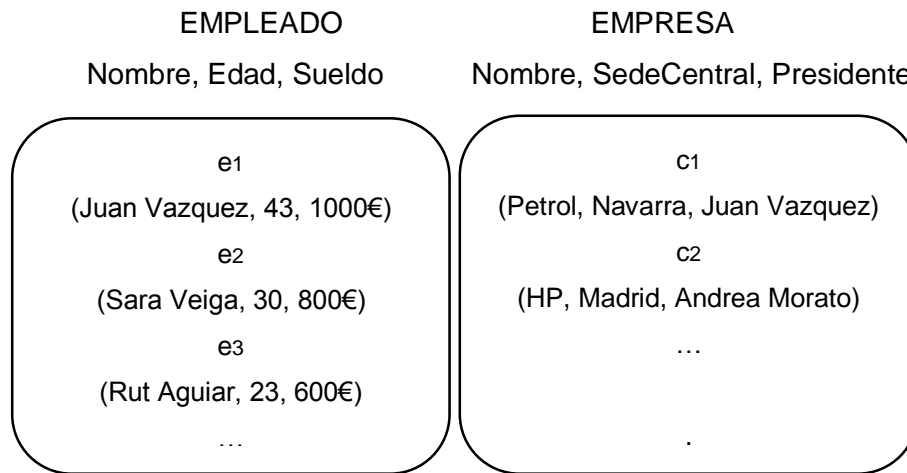


Figura 3. Tipo de entidad y conjunto de entidades.

Fuente: (Elmasri, 2007).

La *figura 3*, muestra dos tipos de entidades EMPLEADO Y EMPRESA, y una lista de atributos de cada una. También se ilustran unas cuantas entidades individuales de cada tipo, junto con los valores de sus atributos. La colección de todas las entidades de un tipo de entidad en particular de la base de datos en cualquier momento del tiempo se denomina conjunto de entidades.

2.2.2 Atributos

Un atributo, es cada una de las propiedades o características que tiene un tipo de entidad o un tipo de relación⁵. Cada atributo tiene un conjunto de valores asociados denominado *dominio*. El **dominio** define todos los valores posibles que puede tomar un atributo por cada entidad individual. Un ejemplo es, si el rango de edades permitido para la entidad empleados está entre 16 y 70 años, podemos especificar el conjunto de valores del atributo Edad de EMPLEADO, como un conjunto de números enteros entre 16 y 70. De forma parecida, podemos especificar el conjunto de valores para el atributo Nombre como un conjunto de cadenas de caracteres alfabéticos separados por espacios en blanco, etc.

Volviendo al concepto de atributo, se pueden clasificar en:

- **Simple o Compuestos.** Un atributo *simple* es un atributo que tiene un solo componente, que no se puede dividir en partes más pequeñas que tengan un

⁵ El concepto de relación se definirá en el punto 2.2.4.

significado propio. Un atributo *compuesto*, es aquel formado por varios componentes, cada uno con un significado per se.

*Por ejemplo, el atributo **nombre**, puede estar estructurado como un atributo compuesto, consistente en **nombre**, **primer_apellido** y **segundo_apellido**. (Silberschatz A, 2007).*

- **Monovaluados o Multivaluados.** Un atributo *monovaluado* es aquel que tiene un solo valor para cada ocurrencia de la entidad o relación a la que pertenece. Un atributo *multivaluado* es aquel que tiene varios valores para cada ocurrencia de la entidad o relación a la que pertenece.

*Para un ejemplo de atributo multivaluado, considérese un conjunto de entidades **empleado** con el atributo **número_teléfono**. Cada empleado puede tener cero, uno o varios números de teléfono, y empleados diferentes pueden tener diferente cantidad de teléfonos. (Silberschatz A, 2007).*

- **Almacenado o Derivado.** Un atributo *derivado* es aquel que representa un valor que se puede obtener a partir de otra información existente en el modelo, es decir del valor de uno o varios atributos, que no necesariamente deben pertenecer a la misma entidad o relación.

*Por ejemplo, supóngase que el conjunto de entidades **cliente**, tiene el atributo **edad**, que indica la edad del cliente. Si el conjunto de entidades **cliente** tiene también un atributo **fecha_de_nacimiento**, se puede calcular **edad** a partir de **fecha_de_nacimiento** y de la fecha actual. Por tanto **edad** es un atributo derivado. En este caso, **fecha_de_nacimiento** puede considerarse un atributo almacenado. (Silberschatz A, 2007)*

2.2.3 Identificador

Antes de explicar que es un identificador, es necesario definir que son los valores NULL (nulos).

En algunos casos, es posible que una entidad en particular no tenga un valor aplicable para un atributo. Por ejemplo, el atributo NumeroApartamento de una dirección sólo se aplica a las direcciones correspondientes a edificios de apartamentos, y no a otros tipos de residencias, como las casas unifamiliares. La dirección de una casa unifamiliar tendría el valor NULL para su atributo NumeroApartamento. Los valores NULL también se pueden utilizar, cuando no se conoce el valor de un atributo para una entidad en particular, por ejemplo si no conocemos el número de teléfono de un determinado empleado.

Todo tipo de entidad debe tener un *identificador*, siendo un atributo o conjunto de atributos que determina de modo único cada ocurrencia de esa entidad. Un identificador, no admite nulos y no admite duplicados. (Véase figura 4).

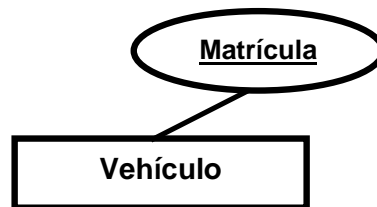


Figura 4. Identificador único.

La *Matrícula*, identifica unívocamente a un vehículo en concreto. El atributo *Matrícula* será **clave** o **identificador** de la entidad *vehículo*.

2.2.4 Relaciones y restricciones en los tipos de relación

Una **relación** expresa un vínculo, una asociación o correspondencia entre varias entidades. (Véase figura 5).

Un **tipo de relación** es el conjunto de relaciones de la misma naturaleza.

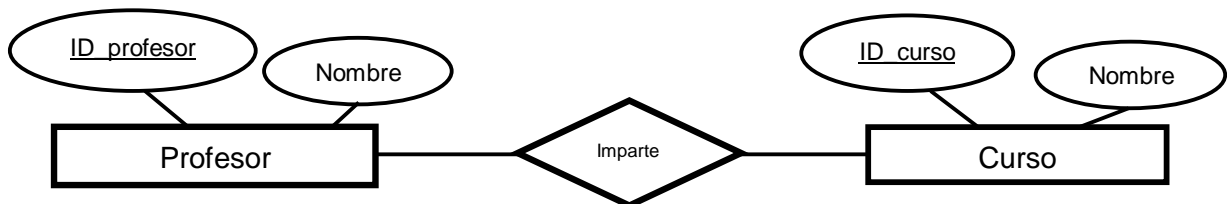


Figura 5. Ejemplo de Relación.

Imparte es una relación que vincula las entidades Profesor y Curso.

Los tipos de relaciones normalmente tienen ciertas **restricciones** que limitan las posibles combinaciones entre las entidades que pueden participar en el conjunto de relaciones correspondiente. Estas restricciones están determinadas por la situación del minimundo representado por las relaciones. Por ejemplo, si una empresa tiene por norma que cada empleado debe trabajar únicamente para un departamento, entonces tendríamos que describir esta restricción en el esquema.

Se pueden distinguir dos tipos de restricciones de relación: razón de cardinalidad y participación.

La razón de **cardinalidad** de una relación binaria⁶, especifica el número *máximo* de instancias de relación en las que una entidad puede participar.

La restricción de **participación** especifica si la existencia de una entidad depende de si está relacionada con otra entidad a través de un tipo de relación. Esta restricción especifica el número *mínimo* de instancias de relación en las que puede participar cada entidad.

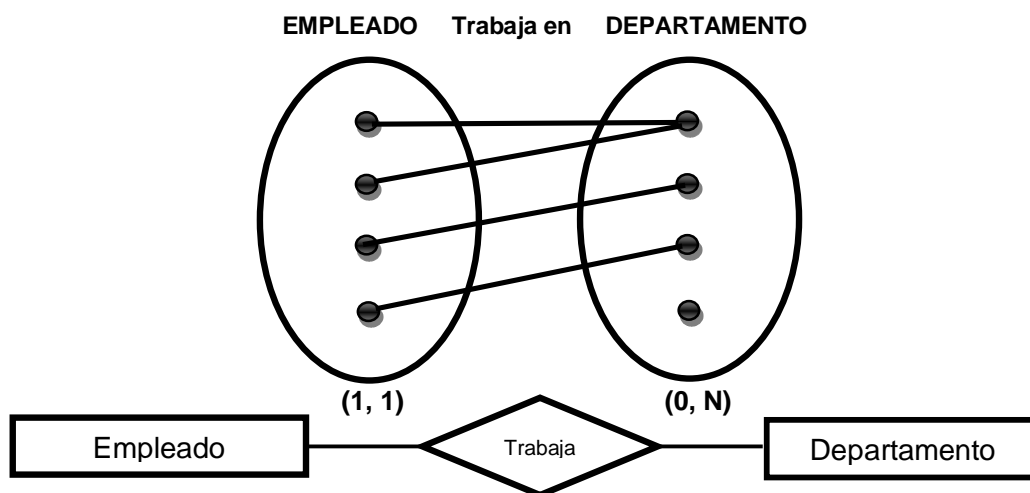


Figura 6. Participación y Cardinalidad

Fuente: (Elmasri, 2007).

⁶ Existen distintos tipos de relaciones, n-arias (ternarias, cuaternarias...), reflexivas (unarias), exclusivas, jerárquicas y binarias. Siendo esta última en la que nos centraremos para el diseño de la base de datos.

En el ejemplo de la *figura 6*, el tipo de relación binaria Trabaja en, DEPARTAMENTO: EMPLEADO tiene una razón de cardinalidad de (1, N), que significa que cada departamento puede estar relacionado con (emplear a) cualquier cantidad de empleados, pero un empleado puede estar relacionado con (trabajar para) un solo departamento.

En cuanto la participación, si la política de la empresa dice que cada empleado debe trabajar para un departamento, entonces la entidad empleado solo puede existir si participa en al menos una instancia de relación Trabaja en.

Correspondencia de cardinalidades

La correspondencia de cardinalidades, o como se ha visto anteriormente, razón de cardinalidad, expresa el número de entidades a las que otra entidad se puede asociar mediante un conjunto de relaciones.

Dado un conjunto de relaciones binarias R entre los conjuntos de entidades A y B, la correspondencia de cardinalidades puede ser:

- **Uno a Uno 1: 1:** Cada entidad de A se asocia, a lo sumo, con una entidad de B, y cada entidad de B se asocia, a lo sumo, con una entidad de A. (Véase la *figura 7.a*).
- **Uno a Muchos 1: N:** Cada entidad de A se asocia con cualquier número (cero o más) de entidades de B. Cada entidad de B, sin embargo, se puede asociar, a lo sumo, con una entidad de A. (Véase la *Figura 7.b*).

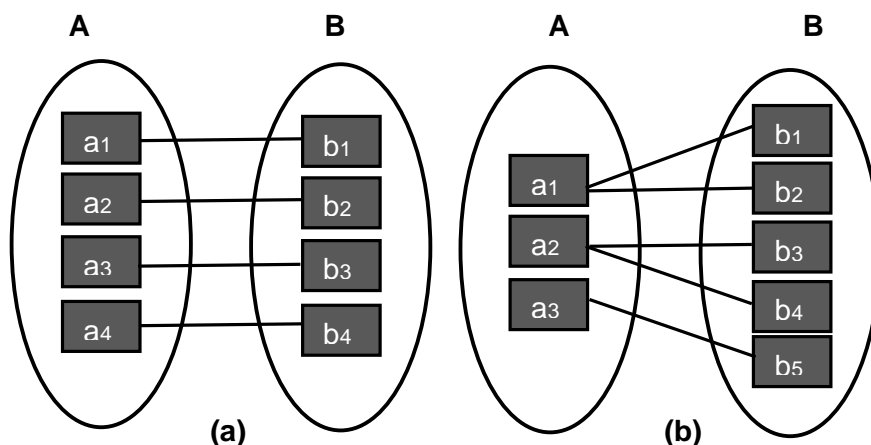


Figura 7. Correspondencia de cardinalidades. (a) Uno a uno. (b) Uno a muchos.

Fuente: (Silberschatz A, 2007).

- **Muchos a Uno N: 1:** Cada entidad de A, se asocia, a lo sumo, con una entidad de B. Cada entidad de B, sin embargo, se puede asociar con cualquier número (cero o más) de entidades de A. (Véase la Figura 8.a).
- **Muchos a Muchos M: N:** Cada entidad de A se asocia con cualquier número (cero o más) de entidades de B, y cada entidad de B se asocia con cualquier número (cero o más) de entidades de A. (Véase la Figura 8.b).

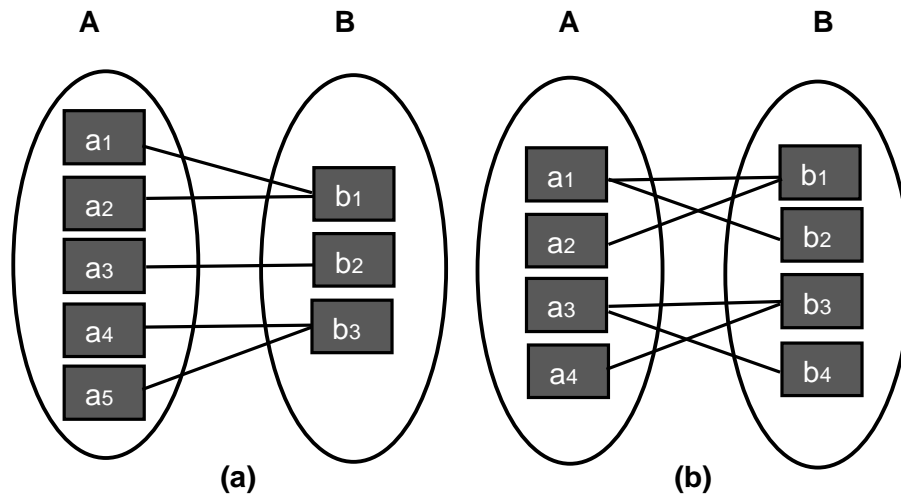


Figura 8. Correspondencia de cardinalidades. (a) Muchos a uno. (b) Muchos a muchos.

Fuente: (Silberschatz A, 2007)

La correspondencia de cardinalidades adecuada para un conjunto de relaciones dado, depende de la situación del mundo real que el conjunto de relaciones modele.

Una vez que se han visto las relaciones binarias, tiene sentido hablar de los tipos de entidades, como se ha mencionado anteriormente en el punto 2.2.1.

Hay dos tipos de entidades: fuertes y débiles. Una entidad **fuerte**, es aquella que se identifica por sí misma. Una entidad **débil** es aquella cuyos atributos no la identifican completamente, sino que sólo la identifican de forma parcial. Esta entidad debe participar en una relación que ayude a identificarla.

Como una entidad débil tiene dependencia de identificación, posee una clave parcial o discriminador, siendo su definición:

“El conjunto de atributos que pueden identificar sin lugar a dudas las entidades débiles que están relacionadas con la misma entidad fuerte”. (Véase figura 9).

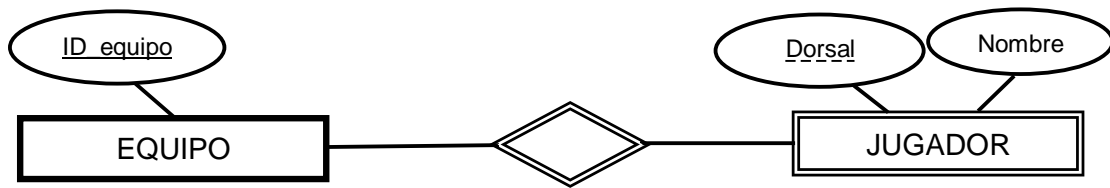


Figura 9. Ejemplo de una entidad débil.

Jugador tiene dependencia de identificación de la entidad fuerte *Equipo*. El atributo *ID_equipo* es la clave de la entidad *Equipo* y el atributo *Dorsal* es un discriminador para la entidad débil *Jugador*, puesto que junto con el atributo *ID_equipo* permite determinar que jugador es de un determinado equipo.

2.3 El Modelo Relacional

En 1970, Edgar F. Codd publicó un trabajo⁷ proponiendo un nuevo modelo de datos que perseguía como objetivo general la flexibilidad y sencillez; el modelo ofrece estructuras de datos simples y lenguajes sencillos para los usuarios, facilitando la comprensión y utilización por estos. Sin menoscabo de este objetivo de sencillez, el autor se propone fundamentar el modelo sobre una sólida base matemática.

El modelo relacional constituye una alternativa para la organización y representación de la información que se pretende almacenar en una base de datos. Se trata de un modelo teórico matemático que, además de proporcionarnos los elementos básicos de modelado (las relaciones), incluye un conjunto de operadores (definidos en forma de *álgebra relacional*) para su manipulación, sin ambigüedad posible.

Dicho modelo es habitualmente seleccionado como modelo de referencia para la elaboración del *esquema lógico* de una base de datos, como tercer paso de la habitual metodología de diseño de BD, (después del *análisis de necesidades* y la elaboración del *esquema conceptual*.)

El modelo relacional se basa en el concepto matemático de *relación*. En este modelo la información se representa en forma de “tablas”, donde cada fila de una tabla se interpreta como una colección ordenada de valores.

Formalmente, una relación se define como un conjunto de **n-tuplas**; donde una n-tupla se define a su vez como una lista ordenada de valores atómicos (esto es, no divisibles

⁷ “A relational model for large shared data banks”; publicado en Communications of the ACM.

ni descomponibles en valores más pequeños). El siguiente ejemplo, presenta una relación que representa al conjunto de los departamentos de una determinada empresa, y que recoge información sobre los mismos.

Núm.	Nombre	Localidad
D-01	Ventas	A Coruña
D-02	I+D	Ferrol

Tabla 2. Relación “Departamentos”

Elaboración propia.

La relación mostrada de la *tabla 2*, incluye dos *3-tuplas*: (D-01; Ventas; A Coruña) y (D-02; I+D; Ferrol). Cada tupla incluye información sobre los departamentos de una determinada empresa con sede en Galicia: el identificador del departamento dentro de la empresa, su nombre y la localidad donde tiene su sede. En cada tupla, los tres valores están relacionados por el hecho de describir todos ellos al mismo departamento.

Cada relación vista como una tabla, consta de un conjunto de columnas; cada una de esas columnas recibe el nombre de **atributo**. A cada atributo de una relación le corresponde un **nombre**, que debe ser único dentro de la relación, y un **dominio**: el conjunto de valores válidos para un atributo; o dicho de otra manera, el conjunto de valores que cada tupla de la relación puede tomar para ese atributo. Algunos ejemplos de dominios son:

- **NumerosTelefonosFijos**. El conjunto de los 9 dígitos que componen los números de teléfono en España.
- **DocumentoNacionalIdentidad**. El conjunto de documentos nacionales de identidad (DNI) válidos en España.
- **NombresDepartamentosAcademicos**. El conjunto de nombres de los departamentos académicos de una universidad, como Informática, Economía o Física.

Lo expuesto anteriormente se conoce como definiciones *lógicas* de dominios. Para cada uno de ellos se especifica también un *tipo de dato* o *formato*⁸. Por ejemplo el tipo de dato del dominio **Numeros-TelefonosFijos** puede declararse como una cadena de

⁸ Esto podrá verse con más claridad en el modo diseño de las tablas en el SGBD, más adelante.

caracteres de la forma *ddddddddd*, donde cada *d* es un dígito numérico (decimal) y los dos, o tres primeros especifican la provincia del número.

2.3.1 Claves

Es necesario tener una forma de especificar cómo las entidades, dentro de un conjunto de entidades dado son distinguibles. Por lo tanto, los valores de los atributos de una entidad deben ser tales que permitan *identificar unívocamente* una tupla.

La palabra *clave*, hará referencia al conjunto de atributos suficiente para distinguir las entidades entre sí.

Una *superclave* es un conjunto de uno o más atributos que, tomados colectivamente, permiten identificar de forma unívoca una tupla en el conjunto de tuplas.

En la *figura 10*, se consideran dos entidades, una débil de la otra, la clave de la débil va a ser un conjunto de dos atributos, la clave de la entidad fuerte y un discriminador. Así pues, para identificar unívocamente una tupla de la entidad débil no basta con ninguno de los dos atributos mencionados por separado, es menester la creación de una superclave, que no es más que la suma de ambos atributos.

FACTURA VENTAS (entidad fuerte).

<u>Id factura venta</u>	Fecha
-------------------------	-------

LÍNEAS DE FACTURA VENTAS (entidad débil).

<u>Id línea</u>	Importe	<u>Id factura venta</u>
-----------------	---------	-------------------------

Siendo la superclave: Id factura venta. Id línea.

Figura 10. Ejemplo de superclave.

En cambio, llamaremos *claves candidatas*, al conjunto de atributos clave, pudiendo haber en una relación más de una clave candidata, de las cuales se elegirá una que se denominará *clave primaria*.

2.3.2 Restricciones de integridad

Cada tupla de una relación debe proporcionar valores a sus atributos. Para garantizar la consistencia y la facilidad de manipulación de la información representada, existen una serie de reglas que deben ser cumplidas y que son un elemento constituyente del modelo relacional. A esas reglas de consistencia se las conoce, en la terminología del modelo, como **restricciones de integridad**. Podemos distinguir varios tipos de restricciones:

- **Restricción de DOMINIO:** “Los dominios de los atributos de una relación deben ser atómicos”. Esta restricción exige que los valores de cualquier tupla de una relación R correspondientes a los atributos A_1, \dots, A_n de R deben ser valores atómicos. Esto es, esos valores no pueden ser divididos en valores más simples. Esta condición garantiza que todas las relaciones presenten un formato regular, que pueda ser fácilmente manipulable por medio de un sencillo procedimiento o algoritmo, implementado en la forma de un programa informático.

<i>Núm.</i>	<i>Nombre</i>	<i>Localidad</i>
D-01	Ventas	A Coruña Ferrol
D-02	I+D	Ferrol

Tabla 3. Relación no válida por uso de valor múltiple.

Elaboración propia.

En este ejemplo, se muestra una relación en la que una de sus tuplas, la correspondiente al departamento de Ventas, presenta un doble valor para el atributo Localidad. De esta forma se pretende representar el hecho de que Ventas tiene dos sedes: A Coruña y Ferrol. Este formato viola la restricción de dominio.

- **Restricción de INTEGRIDAD de ENTIDAD:** “Ninguna tupla puede tomar valores nulos en los atributos que forman parte de su clave primaria”.

La necesidad de esta restricción es clara, dado que es la clave primaria la que permite distinguir a las tuplas entre sí. (Véase *tabla 4*).

Núm.	Nombre	Localidad
D-01	Ventas	
D-02	I+D	Ferrol
D-01	Ventas	

Tabla 4. Relación incorrecta.

Elaboración propia.

Siendo desconocidos los valores de Localidad en ambas tuplas, es imposible distinguir a una de otra. Se trata por tanto, de una tupla no válida, porque viola la restricción de integridad de entidad.

- Restricción de CLAVE: En una tabla “no hay dos tuplas distintas que tengan el mismo valor en la clave”. Estas columnas se denominan claves principales.

Núm.	Nombre	Localidad
D-01	Ventas	A Coruña
D-02	I+D	Ferrol
D-01	Marketing	A Coruña

Tabla 5. Violación de restricción de clave.

Elaboración propia.

En el ejemplo de la *tabla 5*, **Núm.** forma una restricción de clave principal, de este modo se garantiza que todas las tuplas de esta tabla no contengan duplicados.

- Restricción de INTEGRIDAD REFERENCIAL: Si hay una clave foránea en una tabla, es decir, que su valor no es nulo, entonces ese valor debe corresponderse con una clave primaria de la tabla con la que se relaciona. (Véase *tabla 6*).

Tabla Departamentos

Dep.	Nombre	Localidad
D-01	Ventas	La Coruña
D-02	I+D	Ferrol
D-03	Marketing	Madrid

Tabla Empleados

Núm.	Nombre	Dep.
1	David.F	D-01
2	Juan.P	D-03
3	Rebeca.G	D-01
4	Sara.A	D-07

Tabla 6. Violación de la integridad referencial

Elaboración propia.

En este ejemplo, **Dep.** en la tabla Empleados, es clave foránea que apunta al atributo **Dep.** de la tabla de origen Departamentos. No debe ser posible crear un registro de empleado con número de departamento D-07, por ejemplo, a menos que ya exista un registro para el número de departamento D-07 en la tabla Departamentos. Sin embargo si es posible, poder crear un nuevo registro de empleado con número de departamento nulo, para contemplar la situación en la que un nuevo empleado ingresa en la compañía pero todavía no se le ha asignado a un departamento concreto.

2.3.3 Reglas de transformación

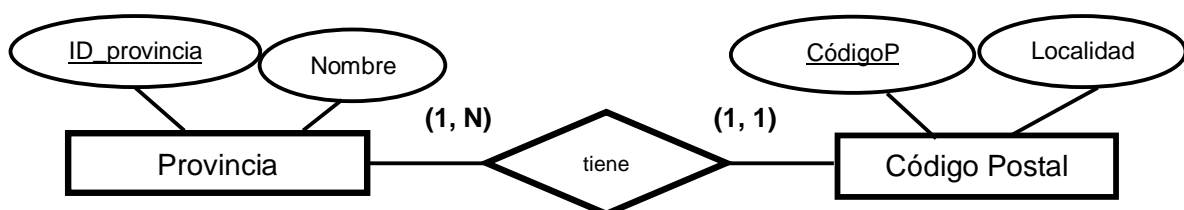
Como se ha visto en la metodología de diseño de las bases de datos⁹, un esquema conceptual basado en el modelo entidad-relación puede ser transformado, de acuerdo con unas sencillas reglas, en un esquema lógico, basado en el modelo relacional y manipulable en un SGBD. Por lo tanto:

- Por cada tipo de entidad **fuerte** *E*, del esquema ER, se crea una *tabla* *T*, que contenga todos los atributos **simples y no multivaluados** de *E*. *T* contendrá **sólo los atributos simples** que formen parte de cada atributo compuesto (no multivaluado) de *E*. Como **clave primaria** de *T* se escogerá el atributo o atributos simples que formen parte de la clave primaria de *E*. Los **atributos derivados** se ignoran, ya que pueden obtenerse a través de consultas.

⁹ Página 14.

- Por cada tipo de entidad **débil** E, del esquema ER, se sigue el mismo procedimiento que para la entidad fuerte. Con la diferencia que como **clave primaria** de T se escogerá el atributo o atributos simples que formen parte del **discriminador** de E, además de la **clave primaria** de la entidad **fuerte** E” del que dependa E. Los **atributos derivados** se ignoran, ya que es posible obtenerlos mediante las consultas.
- Las claves primarias se denotan con un subrayado continuo.
- La transformación de las relaciones dependerá del tipo de participación y cardinalidad de las mismas:
 - Las relaciones **N: M**, se transforman en una nueva tabla que tendrá como clave primaria la concatenación de los atributos clave de las entidades que relaciona. Además, cada uno de los identificadores de las entidades, por separado, se convertirá en una clave foránea que referencia la tabla creada a partir del tipo de entidad correspondiente.
 - En las relaciones **1: N**, se propaga el atributo de la entidad que tiene cardinalidad máxima N a la que tiene cardinalidad máxima 1, desapareciendo el nombre de la relación. Si existen atributos en la relación estos también se propagarán.
 - En el caso de las relaciones **1:1**, si una de las entidades posee cardinalidad (0,1) y la otra (1,1), conviene propagar la clave a la entidad con cardinalidad (1,1) de la tabla resultante de la entidad con cardinalidad (0,1). Si ambas entidades poseen cardinalidades (1,1) se puede propagar la clave de cualquiera de ellas a la tabla resultante de la otra.

Un ejemplo de la transformación de la relación **1: N** es:



El paso al modelo relacional sería:

PROVINCIA

<u>ID_provincia</u>	Nombre
---------------------	--------

CÓDIGO POSTAL

<u>CódigoP</u>	Localidad	ID_provincia
----------------	-----------	--------------

Figura 11. Transformación de una relación 1: N, al modelo relacional.

En la *figura 11*, cada provincia tiene mínimo un código postal y máximo N, en cambio, un código postal mínimo pertenece a una provincia y máximo a una. En el paso al modelo relacional, la clave primaria de la entidad Provincia del lado N, pasará a la entidad Código postal del lado 1, desapareciendo la relación tiene. Siendo ID_provincia clave foránea de la entidad PROVINCIA.

3. PLANIFICACIÓN

A la hora de planificar el proyecto, en primer lugar se establecieron las distintas etapas del trabajo con una breve descripción de las tareas a realizar. A continuación, se le asignó un tiempo de esfuerzo estimado de duración a cada tarea en horas, así como, se establecieron unos recursos tanto humanos como materiales. Se realizó una temporalización para saber la duración prevista del proyecto y finalmente se hizo un seguimiento de la planificación con las correspondientes conclusiones.

3.1 Definición de tareas

Son las tareas necesarias para conseguir el objetivo del proyecto, que consiste en crear una base de datos que almacene la información relevante para la empresa *Carlin*, cubra sus necesidades y estudie tres sectores concretos: el de las franquicias cara el público, el de los proveedores y el de los artículos. Con esta base de datos será posible explotar la información a nivel operacional y gerencial. Estas tareas están basadas en la metodología de diseño de las bases de datos descrita en el apartado **2.1** y en concreto son las siguientes:

1. Documentarse sobre el tema

Se trata de buscar, valorar y seleccionar documentos que permitan elaborar una memoria con información completa sobre el tema a tratar: Bases datos y minería de datos en entornos empresariales. Parte del tiempo dedicado a esta tarea será necesario para la formación y adquisición de conocimientos específicos y técnicos sobre el tema.

2. Definición de objetivos

Se trata de definir los objetivos que quiere la compañía obtener para su misión empresarial sobre el almacenamiento de la información. En este caso, se representará el minimundo o universo de discurso de la empresa Carlin, S.A.

3. Análisis y Diseño conceptual de la BD.

Partiendo de la realidad parcial (referente a los tres sectores en concreto) de la empresa a estudiar, se trata de analizar y extraer los requisitos de almacenamiento de información a nivel gerencial y operacional. Posteriormente se deben formalizar y modelar las necesidades en un lenguaje formal: en el modelo entidad relación.

4. Transformación en el esquema relacional.

Esta fase es el paso del diagrama entidad-relación a tablas del modelo relacional que contendrá la base de datos, es decir a un modelo matemático algebraico, según la serie de reglas descritas en el punto **2.3.3**.

5. Implementación de la base de datos.

Es la creación de las tablas obtenidas en el paso anterior, el paso del modelo relacional a un SGBD (Sistema gestor de bases de datos), es decir a un tipo de software específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan. Se ha seleccionado como SGBD Access por su fácil uso y su capacidad para almacenar el tamaño de la BD que se pretende crear. El conjunto de la información se organizará en tablas, y dichas tablas se organizarán como una matriz bidimensional de filas y columnas.

6. Alimentación de la base de datos.

Consiste en poblar los distintos campos y registros de los que consta una tabla. Para ello se utilizarán distintas técnicas, de manera semiautomática, entre ellas la de generación de datos mediante Excel y de manera manual, para tener información suficiente que ejemplifique la explotación en los pasos siguientes.

7. Explotación de datos.

Esta etapa engloba las tareas **7.1** y **7.2**. Siendo:

7.1 Consultas SQL: Para acceder a los datos del SGBD relacional nos apoyaremos en el lenguaje estructurado de consultas o SQL, que permite gracias a la utilización del álgebra y de cálculos relacionales, efectuar consultas con el fin de recuperar de forma sencilla información de interés de la base de datos.

7.2 Data Mining:

Tras la extracción de información de la base de datos operacional, con la ayuda de herramientas como las hojas de cálculo, informes, etc... se expondrán posibles decisiones acorde con los objetivos definidos de la empresa.

8. Elaboración de la Memoria.

Durante la recopilación de información, el diseño de la base de datos, la alimentación y explotación de la misma, se redactará la memoria del proyecto.

9. Elaboración de la Presentación.

Utilizando PowerPoint se elaborará la presentación del tema elegido con vista a la exposición en el tribunal correspondiente.

3.2 Esfuerzo

En este punto, se estimó el tiempo en horas que había que dedicar a cada tarea para llevar un control de las mismas. Las tareas implicaron un tiempo estimado de realización inicial de 445 horas totales. (Véase tabla 7).

ID	TAREA	DURACIÓN EN HORAS
1	<i>Documentarse sobre el tema</i>	35
2	<i>Definición de Objetivos</i>	45
3	<i>Análisis y Diseño Conceptual de la BD</i>	75
4	<i>Transformación en el esquema relacional</i>	5
5	<i>Implementación de la BD</i>	30
6	<i>Alimentación de la BD</i>	75
7	<i>Explotación de datos</i>	65
7.1	<i>Consultas SQL</i>	35
7.2	<i>Data Mining</i>	30
8	<i>Elaboración de la Memoria</i>	80
9	<i>Elaboración de la Presentación</i>	35
TOTAL		445

Tabla 7. Asignación de horas por tarea.

Elaboración propia en Excel.

3.3 Recursos

Normalmente, para el desarrollo de este proyecto, serían necesarios los siguientes recursos:

- **Recursos Humanos:** Una persona que conozca el funcionamiento, la gestión de la empresa, será la encargada de contratar a un ingeniero en informática que haga de analista, es decir que se encargue de captar las necesidades de información de la empresa en una hoja de requisitos y posteriormente implemente la base de datos en el SGBD correspondiente. En este caso, los recursos disponibles son, la alumna que hace este proyecto y los tutores asignados que servirán de guías en el desarrollo del mismo. El coste de tener a una persona cualificada que se encargue de este proyecto es de 25 euros/hora, siendo en total un coste de $(445 \cdot 25)$, **11.125 euros**.
- **Recursos materiales:** se necesitarán herramientas como un ordenador, un sistema operativo, un navegador web, un sistema gestor de base de datos, un procesador de texto, hojas de cálculo, una impresora, papel, tinta, etc...que harán posible la consecución del proyecto.
El coste de los recursos materiales fueron de un total de 1.100 euros. Siendo los materiales amortizables en cuatro años y prorrateados según la duración del proyecto (4 meses), el coste es de **91,66 euros**.

3.4 Temporalización

En el siguiente gráfico, se puede ver la fecha de comienzo (09/01/2014) y fin (15/04/2014) del proyecto, con un coste de 11.216,66 euros de manera estimada. Teniendo en cuenta que se han dedicado al proyecto 5 horas diarias, un total de 35 horas a la semana.

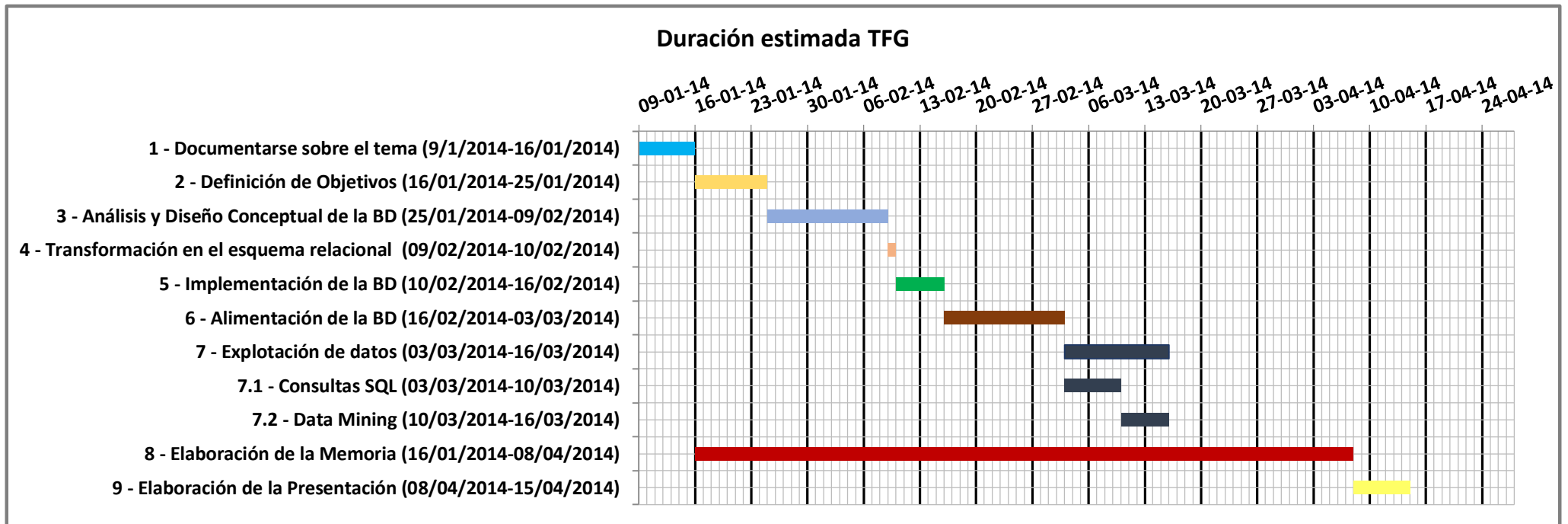


Figura 12. Diagrama de Gantt estimado.

Elaboración propia en Excel.

3.5 Seguimiento y conclusiones

Durante la elaboración del TFG se hizo un seguimiento de las tareas para garantizar que se iban completando según lo planificado y para poder actuar lo antes posible en el caso de desviaciones. De esta manera, ha sido posible analizar las desviaciones que hubo en cuanto a los tiempos estimados para cada tarea y cuáles fueron los motivos que provocaron el retraso de las mismas.

La fase de análisis y diseño conceptual de la base de datos (3), dio lugar a distintas cuestiones:

- Al tener que diseñar la base de datos operacional de la empresa Carlin y tratarse de la gestión interna de la compañía, fue difícil recabar información fehaciente, ya que este tipo de información suele ser confidencial.
- El diseño, requiere de mucho tiempo, y es una de las fases más importantes puesto que es el cimiento de la base de datos en general.
- La definición de las entidades, los atributos y relaciones, aunque parezcan sencillas de pensar, están diseñadas con el fin de representar el funcionamiento interno de la empresa y con cualquier error se obtendría una visión distinta a la esperada.
- Una vez que se diseñó el modelo, hubo que aprender a utilizar el programa (Dia.Ink), que permitía dibujar el diagrama entidad-relación con más claridad.

Por lo tanto se estimaron 15 días para la consecución de esta tarea, pero realmente acarreó 20 días.

En la fase de explotación de datos (7), las consultas SQL, se retrasaron por:

- Falta de conocimiento de la sintaxis de la consola SQL en Access 2013.
- Algunas consultas presentaban cierto grado de dificultad.
- Fue necesario en algunos casos, hacer filtros en los datos para la realización de los gráficos.

Se estimaron 13 días para la consecución de esta tarea, pero realmente acarreó 16 días.

No solo se desvió el coste sino también el tiempo estimado, ya que no fue posible hacer más que invertir más horas a la semana. En total de las 445 horas de esfuerzo estimadas para la consecución del proyecto, fueron necesarias realmente 485 horas.

Por lo tanto, las desviaciones hicieron que el proyecto finalizara, en vez del (15/04/2014), el (28/04/2014) con un coste de (485×25) , 12.216,66 euros.

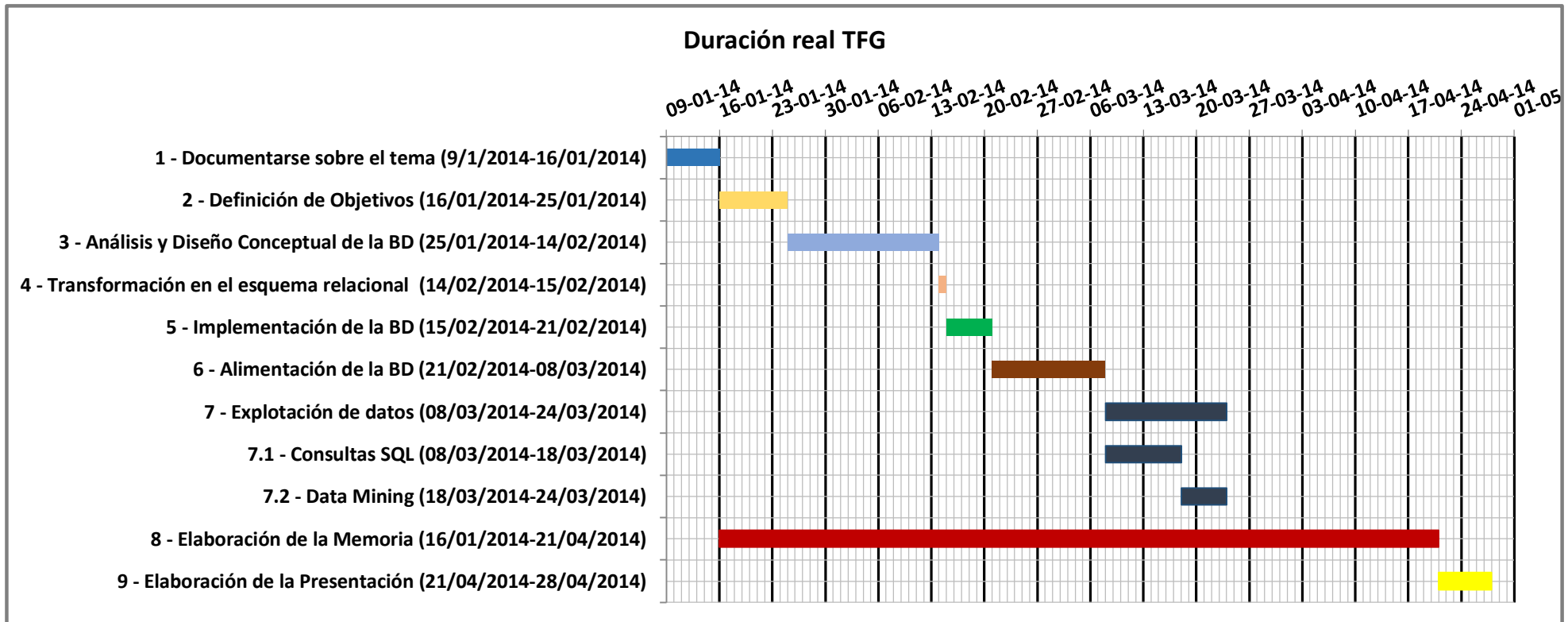


Figura 13. Diagrama de Gantt final.

Elaboración propia en Excel.

4. DISEÑO DE LA BASE DE DATOS

4.1 Selección del dominio/minimundo.

Carlin es una empresa que se posiciona como líder en el sector de Papelería, materiales de oficina y consumibles informáticos. Apuesta por las franquicias como línea de crecimiento a nivel nacional. Cerró el año 2013 con 507 franquicias repartidas por toda España.

Los clientes directos de la compañía son los franquiciados, Carlin como franquiciador se encarga de proveer a sus franquiciados de todo lo necesario para el funcionamiento de las tiendas. Esto incluye, tanto los productos, como la formación, la publicidad, el asesoramiento, el *know-how*, es decir el concepto empresarial propio, cerrado, probado y rentable.

Carlin, está asociado con distintas marcas fabricantes, con las que tiene establecido por contrato acuerdos en cuanto a precios, plazos de entrega, etc., lo que hace que goce de un poder negociador con cada proveedor homologado, con respecto a otras empresas de la competencia. Esto le permite tener un margen de beneficios importante como intermediario entre los proveedores y las tiendas franquiciadas.

Es importante aclarar que entre las marcas que ofrece, esta su propia gama de productos de marca blanca.

Los objetivos fundamentales de la empresa son:

- Analizar el rendimiento de las franquicias por zona geográfica.
- Segmentar el mercado con el fin de poder conocer mejor a los clientes y ofrecer un servicio de mayor calidad.
- Aumentar su margen de beneficios.
- Mejorar el poder negociador con los proveedores.
- Adaptar, redireccionar las campañas de marketing a un público objetivo en concreto.
- Lanzamiento de nuevos productos y posicionamiento de su marca blanca.
- Establecer una planificación estratégica a nivel de negocio con respecto al mercado.

Partiremos de la base de datos operacional de la compañía, en esta base de datos, se guardará información relativa a la gestión de la empresa, y se extraerá información que ayude a los directivos a tomar decisiones en concordancia con los objetivos definidos.

4.2 Diseño Conceptual

Como se ha definido en el punto **2.2**, el modelo que utilizaremos para representar el sistema de información de la empresa *Carlin*, es el diagrama Entidad-Relación.

El diagrama ER (*véase figura 14*), está diseñado con el fin de que los directivos de la compañía puedan seguir un proceso racional que garantice las mejores decisiones. Se ha optado por dividir el modelo en distintas partes:

- Por un lado, se visualiza la relación existente entre las franquicias y su contacto con el cliente final, es decir las ventas efectuadas en el día a día.
- Por otro lado, se visualizan las compras de las franquicias al franquiciador, de las distintas categorías de productos que ofrece la compañía.
- En un nivel inferior, el contacto del franquiciador con los proveedores y como actúa de distribuidor de esos productos.
- En la esquina superior izquierda, se refleja el ámbito territorial en el que están situadas las franquicias y la procedencia de los clientes.

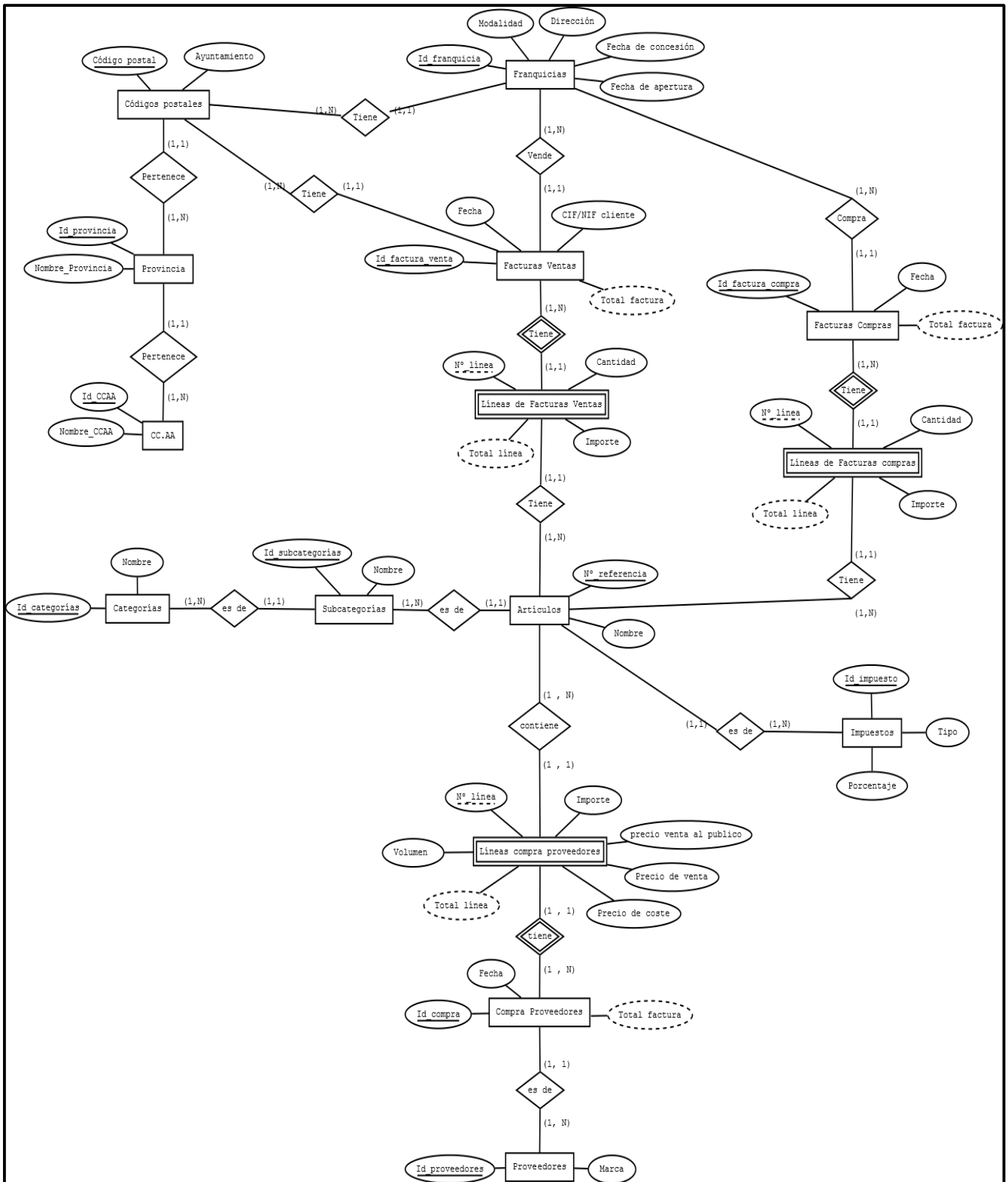


Figura 14. Modelo Entidad-Relación de Carlin.

Elaboración propia (Dia.Ink).

A continuación se procede a definir los elementos que componen el modelo ER:

La entidad **FRANQUICIAS**, representa a los clientes de Carlín, es decir los franquiciados.

Los atributos de dicha entidad serán:

- **Id franquicia**: se asignará un identificador único para cada franquicia, que permita diferenciarla de las demás.
- **Dirección**: es la dirección donde esta emplazada una determinada franquicia.
- **Modalidad**: existen distintas modalidades de franquicias. (Hiperpapelería, Ofimarket, Ofimarket mixto y Master en Franquicias.)
- **Fecha de concesión**: es la fecha en que se otorgó la franquicia.
- **Fecha de apertura**: fecha en que se abrió la tienda al público.

Estos atributos (**Fecha de concesión** y **Fecha de apertura**), nos servirán en el futuro, para generar informes que permitan saber la evolución de las compras de cada franquicia desde su apertura o el plazo que tardan en abrir las tiendas, etc...para poder tener una idea del tiempo de formación y asesoramiento en general de las tiendas.

La entidad **FACTURAS VENTAS**, representa las facturas que emite cada franquicia a sus respectivos clientes por las ventas efectuadas.

Los atributos de dicha entidad serán:

- **Id factura venta**: se asignará un identificador único para cada factura emitida.
- **Fecha**: día en el que se emitió la factura.
- **CIF/NIF cliente**: identificador de la empresa o de un cliente en particular, al que se emita la factura.
- **Total factura**: (*atributo derivado*), la suma total de las líneas de factura ventas.

La entidad débil **LÍNEAS DE FACTURA VENTAS**, representa cada una de las líneas de una factura de venta.

Los atributos de dicha entidad serán:

- **Nº línea:** *discriminador* de cada línea de una factura de venta.
- **Cantidad:** magnitud por unidades de cada artículo vendido.
- **Importe:** es el precio unitario del artículo.
- **Total línea:** (*atributo derivado*), es la cantidad por el importe.

La entidad **ARTÍCULOS**, representa a cada uno de los productos que venden las tiendas.

Los atributos de dicha entidad serán:

- **Nº referencia:** se asignará un identificador único a cada artículo.
- **Nombre:** denominación de cada uno de los artículos.

La entidad **SUBCATEGORÍA**, representa al conjunto de artículos con características similares, que solo se diferencian en la marca:

Por ejemplo:

Lapiceros: *Maped, Faber Castell, Staedtler...*

Portaminas: *Stabilo, Miilan, Bic...*

Sus atributos serán:

- **ID Subcategorías:** se le asignará un ID único para cada grupo de artículos con características similares.
- **Nombre:** calificativo de cada grupo de artículos.

La entidad **CATEGORÍAS**, representa a la familia de productos a la que pertenece cada subcategoría:

Por ejemplo:

Escritura y Corrección: *Lapiceros, Portaminas, Bolígrafos...*

Complementos de oficina: *Pegamento, Gomas de borrar, Reglas...*

Sus atributos serán:

- **ID Categorías:** se le asignará un ID único a cada categoría.

- **Nombre:** apelativo de cada familia de productos.

Esta división de los artículos, se ha hecho respetando el funcionamiento interno de la empresa. Cada artículo pertenece a una subcategoría y cada subcategoría a una categoría. Esto permitirá más adelante evaluar las compras y ventas por grupos de artículos o de manera individual.

La entidad **IMPUESTOS**, representa el IVA a pagar, la carga fiscal sobre el consumo de un determinado artículo.

Los atributos de dicha entidad serán:

- **Id impuesto:** identificador de cada impuesto.
- **Tipo:** denominación de los tipos de IVA, según sea aplicable el tipo general, el reducido o el superreducido.
- **Porcentaje:** porcentaje aplicable según el tipo de IVA.

La entidad **FACTURAS COMPRAS**, representa las facturas que emite el franquiciador a las respectivas franquicias por las compras efectuadas.

Los atributos de dicha entidad serán:

- **Id factura compra:** se asignará un identificador único para cada factura emitida.
- **Fecha:** día en el que se emitió la factura.
- **Total factura:** (*atributo derivado*), la suma total de las líneas de factura compras.

La entidad débil **LÍNEAS DE FACTURA COMPRAS**, representa cada una de las líneas de una factura de compra por parte de las franquicias.

Los atributos de dicha entidad serán:

- **Nº línea:** *discriminador* de cada línea de una factura de compra.
- **Cantidad:** magnitud por unidades de cada artículo comprado.
- **Importe:** es el precio unitario del artículo.
- **Total línea:** (*atributo derivado*), es la cantidad por el importe.

La entidad **PROVEEDORES**, representa a las distintas empresas (marcas), asociadas a Carlin, que proveen a la organización de los distintos artículos.

Los atributos de dicha entidad serán:

- **Id proveedores**: se le asignará un ID único a cada proveedor.
- **Marca**: nombre de la organización que provee a la compañía o del producto.

La entidad **COMPRA PROVEEDORES**, representa las distintas compras hechas a los proveedores.

Sus atributos serán:

- **Id compra**: se le asignará un ID único a cada compra.
- **Fecha**: se guardará en formato mes/año.
- **Total factura**: (*atributo derivado*), suma total de las líneas compra proveedores.

Con la *fecha* se pretende poder generar informes en los que se detalle el volumen de gasto por mes por ejemplo.

La entidad débil, **LÍNEAS COMPRA PROVEEDORES**, representa a cada una de las líneas de compras hechas por el franquiciador a los proveedores.

Los atributos de dicha entidad serán:

- **Nº línea**: *discriminador* de cada línea de compra a proveedores.
- **Volumen**: cantidad en unidades de las compras.
- **Importe**: es el precio unitario de cada artículo.
- **Total línea**: (*atributo derivado*), es el volumen por el importe.
- **Precio de coste**: precio al que se le compra al proveedor los artículos.
- **Precio de venta**: precio al que se vende a las franquicias los artículos.
- **Precio de venta al público**: precio que se fija en las franquicias, cara la venta al público.

La distinción de *precios*, le permite controlar a la empresa: sus compras a los proveedores, el margen de beneficio que obtiene con las ventas a las franquicias, ofrecer un mejor asesoramiento a la tienda fijando precios razonables y adaptando las campañas constantes de productos y ofertas.

La entidad **CÓDIGOS POSTALES**, permite identificar de manera unívoca distintas zonas o lugares de un país.

Sus atributos serán:

- **Código Postal**: identificador único de cada zona geográfica.
- **Ayuntamiento**: lugares a los que pertenece cada código postal.

Esta entidad, juega un papel importante ya que permitirá tomar decisiones estratégicas en cuanto a ubicación y tiempos. Se podrán generar informes y estadísticas de las ventas de cada franquicia según la zona geográfica o el mes del año, observar si la densidad de la clientela se concentra en un ayuntamiento en particular y de esta manera extraer información acerca de la situación geográfica de la tienda con respecto al grueso de su volumen de clientes, acotar zonas para focalizar campañas de marketing, etc...

La entidad **PROVINCIA**, representa a las divisiones territoriales dentro de cada CC.AA.

Sus atributos serán:

- **ID Provincia**: se establecerá un identificador único de cada provincia.
- **Nombre_Provincia**: nombre de cada Provincia.

La entidad **CC.AA**, representa a las divisiones territoriales dentro de España.

Sus atributos serán:

- **ID CCAA**: se asignará un identificador único de cada CC.AA.
- **Nombre_CCAA**: nombre de cada CC.AA.

4.3 Diseño Lógico

El objetivo es transformar el esquema conceptual obtenido en la etapa anterior, adaptándolo al modelo de datos en el que se apoya el SGBD que se va a utilizar, en este caso el modelo relacional.

FRANQUICIAS

<u>Id_franquicia</u>	Modalidad	Dirección	Fecha de concesión	Fecha de apertura	Código_postal
----------------------	-----------	-----------	--------------------	-------------------	---------------

Siendo:

- Id_franquicia, clave primaria.
- Código_postal, clave foránea de la entidad CÓDIGOS POSTALES.

FACTURAS VENTAS

<u>Id_factura_venta</u>	Fecha	CIF/NIF cliente	Id_franquicia	Código_postal
-------------------------	-------	-----------------	---------------	---------------

Siendo:

- Id_factura_venta, clave primaria.
- Id_franquicia, clave foránea de la entidad FRANQUICIAS.
- Código_postal, clave foránea de la entidad CÓDIGOS POSTALES.

CATEGORÍAS

<u>Id_categorías</u>	Nombre
----------------------	--------

Siendo:

- Id_categorías, clave primaria.

SUBCATEGORÍAS

<u>Id_subcategorías</u>	Nombre	Id_categorías
-------------------------	--------	---------------

Siendo:

- Id_subcategorías, clave primaria.
- Id_categorías, clave foránea de la entidad CATEGORÍAS.

IMPUESTOS

<u>Id_impuesto</u>	Tipo	Porcentaje
--------------------	------	------------

Siendo:

- Id_impuesto, clave primaria.

ARTÍCULOS

<u>Nº referencia</u>	Nombre	Id_subcategorías	Id_impuesto
----------------------	--------	------------------	-------------

Siendo:

- Nº referencia, clave primaria.
- Id_subcategorías, clave foránea de la entidad SUBCATEGORÍAS.
- Id_impuesto, clave foránea de la entidad IMPUESTOS.

LÍNEAS DE FACTURA VENTAS

<u>Id factura venta</u>	<u>Nº línea</u>	Cantidad	Importe	Nº referencia
-------------------------	-----------------	----------	---------	---------------

Siendo:

- Id factura venta, Nº línea, formarán la clave primaria de la entidad débil LÍNEAS DE FACTURA VENTAS.
- Nº referencia, clave foránea de la entidad ARTÍCULOS.

PROVEEDORES

<u>Id_proveedores</u>	Marca
-----------------------	-------

- Id_proveedores, clave primaria.

COMPRA PROVEEDORES

<u>Id_compra</u>	Fecha	Id_proveedores
------------------	-------	----------------

Siendo:

- Id_compra, clave primaria.
- Id_proveedores, clave foránea de la entidad PROVEEDORES.

LÍNEAS COMPRA PROVEEDORES

<u>Id compra</u>	<u>Nº línea</u>	Volumen	Importe	Precio venta al público	Precio de venta	Precio de coste	Nº_referencia
------------------	-----------------	---------	---------	-------------------------	-----------------	-----------------	---------------

Siendo:

- Id compra, Nº línea, formarán la clave primaria de la entidad débil LÍNEAS COMPRA PROVEEDORES.
- Nº_referencia, clave foránea de la entidad ARTÍCULOS.

FACTURAS COMPRAS

<u>Id factura compra</u>	Fecha	Id_franquicia
--------------------------	-------	---------------

Siendo:

- Id factura compra, clave primaria.
- Id_franquicia, clave foránea de la entidad FRANQUICIAS.

LÍNEAS DE FACTURA COMPRAS

<u>Id factura compra</u>	<u>Nº línea</u>	Cantidad	Importe	Nº_referencia
--------------------------	-----------------	----------	---------	---------------

Siendo:

- Id factura compra, Nº línea, formarán la clave primaria de la entidad débil LÍNEAS DE FACTURA COMPRAS.
- Nº_referencia, clave foránea de la entidad ARTÍCULOS.

CC.AA

<u>Id_CCAA</u>	Nombre_CCAA
----------------	-------------

Siendo:

- Id_CCAA, clave primaria.

PROVINCIA

<u>Id_provincia</u>	Nombre_provincia	Id_CCAA
---------------------	------------------	---------

Siendo:

- Id_provincia, clave primaria.
- Id_CCAA, clave foránea de la entidad CC.AA.

CÓDIGOS POSTALES

<u>Código_postal</u>	Ayuntamiento	Id_provincia
----------------------	--------------	--------------

Siendo:

- Código_postal, clave primaria.
- Id_provincia, clave foránea de la entidad PROVINCIA.

4.4 Diseño Físico

Selección del SGBD

Para llevar a cabo la implementación de la base de datos real de la empresa *Carlin*, es necesario seleccionar un Sistema Gestor de Bases de Datos.

La oferta de SGBD es muy extensa, existiendo en el mercado un amplio abanico de productos para todas las necesidades. (Oracle, postgresQL, mySQL...)

En este caso, se ha seleccionado Microsoft Access 2013 como Sistema Gestor de Base de Datos, puesto que:

- Proporciona un interfaz de usuario que facilita la consecución de las tareas más habituales de forma inmediata.
- Es posible diseñar formularios a medida, según las necesidades de cada proyecto.
- Todas las consultas que diseñemos con Access, pueden ser escritas mediante un lenguaje estándar llamado SQL (*Structured Query Language*).
- Es posible diseñar informes, que aunque con las consultas podemos ver de inmediato la información más actual almacenada en la base de datos, en muchas ocasiones debemos entregar estos datos a un tercero, para lo cual lo más

adecuado es elaborar un informe que nos permita visualizar la información de manera más clara y concisa.

- Se pueden generar gráficos y tablas que tienen un componente dinámico, de forma que podemos filtrar rápidamente por cualquier criterio y obtener de inmediato el resultado correspondiente.
- Facilita la integración, la importación y exportación de otras herramientas, como Excel, que utilizaremos más adelante en el almacenamiento y explotación de los datos.

Base de Datos de la empresa Carlin

Una vez seleccionado el gestor Access 2013, se ha implementado la base de datos de la empresa Carlin. En la *figura 15*, se pueden ver todas las tablas creadas, con sus correspondientes atributos y restricciones (restricción de clave: //ave y restricción de clave foránea: uniones), tal y como se describe en el diseño lógico¹⁰.

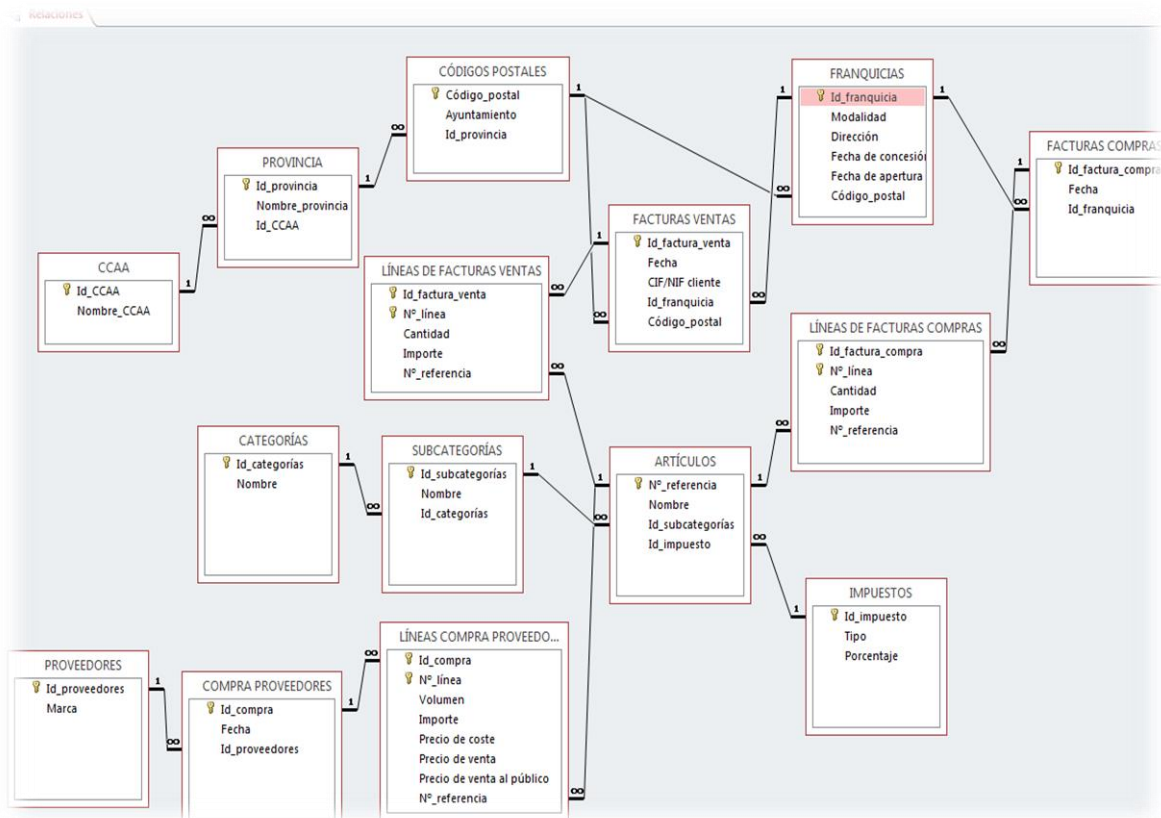


Figura 15. Relaciones BD Carlin.

¹⁰ El diseño físico se ha creado a partir del modelo lógico de la página 46-49.

A modo de ejemplo, a continuación se muestran unos pantallazos de algunas de las diferentes tablas en modo diseño para ver como se indican las claves primarias, los tipos de datos, formatos, etc.

FRANQUICIAS

Nombre del campo	Tipo de datos	Descripción (opcional)
Id_franquicia	Número	
Modalidad	Texto corto	
Dirección	Texto corto	
Fecha de concesión	Fecha/Hora	
Fecha de apertura	Fecha/Hora	

Propiedades del campo

General		Búsqueda
Tamaño del campo	Entero largo	
Formato		
Lugares decimales	Automático	
Máscara de entrada		
Título		
Valor predeterminado	0	
Regla de validación		
Texto de validación		
Requerido	Sí	
Indexado	Sí (Sin duplicados)	
Alineación del texto	General	

El tipo de datos determina la clase de valores que los usuarios pueden guardar en el campo. Presione F1 para obtener ayuda acerca de los tipos de datos.

Figura 16. Diseño de la tabla Artículos.

ARTICULOS

Nombre del campo	Tipo de datos	Descripción (opcional)
Nº_referencia	Número	
Nombre	Texto corto	
Id_subcategorías	Número	
Id_impuesto	Número	

Propiedades del campo

General		Búsqueda
Tamaño del campo	255	
Formato		
Máscara de entrada		
Título		
Valor predeterminado		
Regla de validación		
Texto de validación		
Requerido	No	
Permitir longitud cero	Sí	
Indexado	No	
Compresión Unicode	Sí	
Modo IME	Sin Controles	
Modo de oraciones IME	Nada	
Alineación del texto	General	

El tipo de datos determina la clase de valores que los usuarios pueden guardar en el campo. Presione F1 para obtener ayuda acerca de los tipos de datos.

Figura 17. Diseño de la tabla Franquicias.

PROVEEDORES

Nombre del campo	Tipo de datos	Descripción (opcional)
Id_proveedores	Número	
Marca	Texto corto	

Propiedades del campo

General	
Tamaño del campo	Entero largo
Formato	
Lugares decimales	Automático
Máscara de entrada	
Título	
Valor predeterminado	0
Regla de validación	
Texto de validación	
Requerido	Sí
Indexado	Sí (Sin duplicados)
Alineación del texto	General

El tipo de datos determina la clase de valores que los usuarios pueden guardar en el campo. Presione F1 para obtener ayuda acerca de los tipos de datos.

Figura 18. Diseño de la tabla Proveedores.

LÍNEAS DE FACTURAS COMPRAS

Nombre del campo	Tipo de datos	Descripción (opcional)
Id_factura_compra	Número	
Nº línea	Número	
Cantidad	Número	
Importe	Moneda	
Nº referencia	Número	

Propiedades del campo

General	
Formato	Euro
Lugares decimales	Automático
Máscara de entrada	
Título	
Valor predeterminado	0
Regla de validación	
Texto de validación	
Requerido	No
Indexado	No
Alineación del texto	General

El tipo de datos determina la clase de valores que los usuarios pueden guardar en el campo. Presione F1 para obtener ayuda acerca de los tipos de datos.

Figura 19. Diseño de la tabla líneas de facturas compras.

4.5 Alimentación de la Base de datos

Una vez que se han creado las distintas tablas, si se tratase de un caso real la población de la base de datos se haría en el transcurso normal del día a día de una empresa. Siendo los propios empleados los encargados de introducir los datos necesarios. En este caso, al no disponer de tal información, se han utilizado diferentes técnicas que se describirán a continuación.

4.5.1 Información real¹¹

Se ha podido obtener, de la página web de la organización, información real de interés sobre los proveedores, las categorías, las subcategorías, los artículos y las franquicias.

PROVEEDORES

El origen de los datos de la tabla proveedores es el desplegable de la página web de la empresa Carlin, en el que se muestran las distintas marcas que proveen a la compañía. Tales datos, se copiaron uno a uno manualmente, asignándoles un ID correlativo.

Id_proveedores	Marca
1	3GO
2	3M
3	APLI
4	ARCAS OLLE
5	ARCHIVO 2.000
6	ARCHIVO 2.000 BY EURODIS
7	BEAUTONE
8	BIC
9	BUNZL
10	C. DEL SUR
11	CANON
12	CARCHIVO
13	CARLIN
14	CASIO
15	CITIZEN
16	CLAIREFONTAINE
17	DOHE
18	DURACELL

Figura 20. Alimentación de la tabla proveedores.

¹¹ Disponemos de información real, puesto que se ha solicitado y se ha obtenido el permiso de la empresa Carlin para la utilización de los datos única y exclusivamente para su uso educativo.

CATEGORÍAS

La procedencia de las distintas categorías a las que pertenecen los artículos, es la página web de la organización. Se copiaron los datos manualmente y se les asignó un ID correlativo.

The image shows two side-by-side screenshots. On the left is a web interface titled 'CATEGORÍAS' with a list of categories and expandable arrows. On the right is a screenshot of a database table named 'CATEGORÍAS' with columns 'Id_categorías' and 'Nombre'.

Id_categorías	Nombre
1	Escritura y Corrección
2	Complementos de Oficina
3	Archivo y Clasificación
4	Manipulado
5	Informática, Consumibles y Accesorios
6	Embalaje
7	Etiquetaje
8	Maquinaria
9	Maletines y Carteras
10	Presentación, Señalización y Limpieza
11	Mobiliario y Sillería
*	0

Figura 21. Alimentación de la tabla Categorías.

SUBCATEGORÍAS

The image shows a screenshot of a web interface titled 'CATEGORÍAS' with the category 'Escritura y Corrección' expanded to show its subcategories.

Subcategoría
Lapiceros
Portaminas
Bolígrafos
Rotuladores
Marcadores
Correctores
Sacapuntas
Minas

Dentro de cada Categoría, existen distintas subcategorías (véase figura 22). Como en total existen 105 subcategorías, el proceso de población de esta tabla conllevaría mucho tiempo, por lo que se hizo de manera automática, utilizando herramientas incorporadas en cualquier sistema Linux actual.

Figura 22. Subcategorías.

A continuación se describirán las diferentes herramientas (programas) utilizadas:

wget: (Foundation, <https://www.gnu.org/software/wget/>, 2014), es un software libre, para descargas no interactivas de la web. Soporta protocolos HTTP, HTTPS y FTP. En este caso, se usó para descargar las páginas webs con el contenido necesario.

grep: (Foundation, <http://www.gnu.org/software/grep/>, 2014), es un software libre, que busca en un fichero o en la entrada estándar un patrón dado, y por defecto imprime aquellas líneas que contengan dicho patrón. En este caso, se usó para separar las líneas con contenido relevante en las páginas webs descargadas.

pipe: en entornos Unix, una de las herramientas más útiles es la tubería “|”. Este programa redirecciona la salida estándar de su antecesor a su sucesor.

cut: (Forums, 2014), elimina secciones de una línea dada. De este programa se usaron las siguientes opciones “-d “, que sirve para indicar el delimitador, es decir, aquel carácter que sirve de punto de referencia. “-f “, esta opción, se usa para indicar con que campo debe quedarse el programa, tomando como punto de referencia el delimitador.

cat: es un software libre, que concatena ficheros y/o los imprime por pantalla.

El uso combinado de estos programas, ha permitido obtener la información deseada, de manera casi automática, como se describe a continuación.

Una vez descargada la página web de Carlin, se detectaron las líneas de código que representaban las subcategorías. (Véase figura 23).

```

156 <li><a class="cuuchild">Escritura y Corrección<span class="down"></span></a>
157 <ul>
158 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-lapiceros">Lapiceros</a></li>
159 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-portaminas">Portaminas</a></li>
160 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-boligrafos">Boligrafos</a></li>
161 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-rotuladores">Rotuladores</a></li>
162 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-marcadores">Marcadores</a></li>
163 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-correctores">Correctores</a></li>
164 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-sacapuntas">Sacapuntas</a></li>
165 <li><a class="nochild " href="https://www.carlin.es/escritura_y_correccion-minas">Minas</a></li>
166 </ul>
167 </li>

```

Figura 23. Código web Carlin.

De esta manera, se creó un script para quedarnos sólo con el nombre de la subcategoría

```
~/Desktop/carlin$ cat carlin.html | grep "nochild" | cut -d'>' -f3 | cut -d'<' -f1
```

El resultado de esta concatenación de comandos, son 105 líneas que se corresponden con cada una de las subcategorías. Posteriormente se pasó a un fichero Excel para su correspondiente edición, y se importó a la tabla de subcategorías en Access 2013. (Véase figura 24).

Id_subcategorías	Nombre	Id_categorías
1	Lapiceros	1
2	Portaminas	1
3	Bolígrafos	1
4	Rotuladores	1
5	Marcadores	1
6	Correctores	1
7	Sacapuntas	1
8	Minas	1
9	Pegamentos	2
10	Sujección	2
11	Gomas de Borrarr	2
12	Reglas	2
13	Pesacartas	2
14	Dediles	2
15	Arandelas Adhesivas	2
16	Tizas	2
17	Sellos	2
18	Organizadores	2
19	Papeleras	2
20	Revisteros	2
21	Armarios Metálicos	2
22	Tarjeteros	2

Figura 24. Alimentación de la tabla Subcategorías.

ARTÍCULOS

```
~/Desktop/carlin$ cat carlin.html | grep "nochild" | cut -d'>' -f2 | cut -d'"' -f4 > websDeSubcategorías.txt
```

Esta concatenación de comandos funciona de la siguiente manera:

Con “cat” se ponen todas las líneas de la página web en la salida estándar, esta es redirigida mediante una tubería, a la entrada estándar de “grep”. “grep” va a poner en su salida estándar, todas aquellas líneas que contengan el patrón “nochild”, mediante otra tubería, se redirige esta salida a la entrada de “cut”, que pondrá en su salida el segundo campo después del delimitador “>”. Esta salida será a su vez entrada de otra instancia del programa “cut”, esta vez se usó como delimitador el carácter “,” (como se ve

en la *figura 23*), el carácter “ divide la línea en cinco campos, la URL que nos interesa está en el cuarto. Para finalizar se vuelcan todas las URLs en un fichero de texto.

Con “wget” se descargaron estas URLs y con el siguiente comando se obtuvieron los nombres de los artículos: (*véase figura 25*).

```
~/Desktop/carlin$ cat 1.html | grep productminitulo | cut -d'>' -f3 | cut -d'<' -f1 | cut -d'-' -f2
```

```
LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA H (112311)
LAPICERO STAEDTLER NORIS 120 HB (120)
LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA HB (112300)
LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA B (112301)
LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA 2B (112302)
LAPIZ FABER CASTELL BICOLOR FINO JANUS 2160 (116000)
```

Figura 25. Nombres de cada uno de los artículos.

Esto se fue haciendo para cada una de las subcategorías que contienen los artículos. Posteriormente se pasó a un fichero Excel, en el que se les agregó un ID, la subcategoría y el correspondiente impuesto. Una vez terminado, se importó a la tabla artículos en Access 2013. (*Véase figura 26*).

Nº_referencia	Nombre	Id_subcategorías	Id_impuesto
181	LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA H	1	1
180	LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA HB	1	1
179	LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA B	1	1
178	LAPIZ FABER CASTELL GRAFITO DESSIN 2000 DUREZA 2B	1	1
177	LAPIZ FABER CASTELL BICOLOR FINO JANUS 2160	1	1
176	LAPICERO STAEDTLER NORIS 120 HB	1	1
217	PORTAMINAS FABER CASTELL TK 1 MM.	2	2
216	PORTAMINAS FABER CASTELL TK 0.7 MM.	2	2
215	PORTAMINAS FABER CASTELL TK 0.5 MM.	2	2
214	PORTAMINAS STAEDTLER NORIS 777 0,5 M	2	2
213	PORTAMINAS MILAN CAPSULE 0,7 MM. GC	2	2
212	PORTAMINAS BIC MATIC 0,5 MM.	2	2
211	PORTAMINAS UNI,BALL SHALAKU 0,5 MM.	2	2
210	PORTAMINAS SUPER GRIP PILOT	2	2
204	MARCADOR UNI TIZA LIQUIDA PIZARRA VE	2	2
205	PORTABOLIGRAFO Q, CONNET CON CADEN	3	3
203	BOLIGRAFO BIC CRISTAL STYLUS CON PUN	3	3
202	BOLIGRAFO BIC CRISTAL STYLUS CON PUN	3	3
201	BOLIGRAFO INKJOY PAPER MATE	3	3
194	BOLIGRAFO BIC 4 COLORES FASHION	3	3
193	BOLIGRAFO BIC 4 COLORES FINE	3	3
192	BOLIGRAFO BIC 4 COLORES CLASSIC	3	3
185	BOLIGRAFO BIC CRISTAL GEL	3	3
184	BOLIGRAFO BIC CRISTAL GEL	3	3

Figura 26. Alimentación de la tabla Artículos.

Id_franquicia	Modalidad	Dirección	Fecha de concesión	Fecha de apertura	Código_postal
1 H	C.C. Espacio Medit		09/07/2010	23/11/2010	3039
2 H	C/ Entrepeñas, 4 -		06/07/1995	21/11/1995	150
3 H	C/ Pintor Seijo Rul		13/09/2003	30/01/2004	1530
4 H	C/ Oleiros Nº 4		10/10/2006	05/11/2006	1510
5 H	C/ Pintor Lorenzo		25/01/1999	11/06/1999	3500
6 H	C/ Callitx nº5- Loca		30/01/1992	16/07/1992	2860
7 H	C/ Maravall, 33		19/07/2006	15/08/2006	2800
8 H	C/ Alfonso XII nº 1		16/03/2006	14/07/2006	780
9 H	C/ José Mª Buck nº		23/12/1997	10/04/1998	700
10 H	C/ Padre Majón, n		13/10/1998	18/12/1998	750
11 H	Plaza de San Seba:		17/03/1994	22/05/1994	330
12 H	C/ Alicante, 37-39		22/12/1999	23/04/2000	770
13 H	C/ Elche nº 19		19/05/1998	25/08/1998	780
14 H	Avda. Federico Ga		28/10/1994	07/03/1995	430
15 H	C/ Instinción nº1, l		22/02/1996	21/04/1996	4380
16 H	C/ Los Campos Bol		29/07/1999	28/12/1999	3340
17 H	C/ Marques de Cas		30/08/1994	05/12/1994	3320
18 H	C/ Marques de Sar		29/06/2009	31/10/2009	3320
20 H	C/ Independencia		16/07/2008	08/11/2008	3300
21 H	C/ Francisco Pizarr		13/11/1992	13/03/1993	4690
22 H	Crta. Valverde KM		03/07/1993	21/12/1993	4690
23 H	C/ Priorat, 14 Loca		04/12/1994	27/03/1995	870

Figura 27. Alimentación de la tabla Franquicias.

En cuanto a la tabla de *franquicias* (véase figura 27), los datos meramente reales son las direcciones y el código postal correspondiente a la población. El motivo, es que la ubicación de las franquicias fuese lo más próximo a la realidad. Sin embargo las tuplas de los atributos fechas de apertura y de concesión, se poblaron mediante la función `HOY()-ALEATORIO.ENTRE()`, respetando la trayectoria de la compañía, desde 1990 hasta el 2013. El plazo que se estableció de apertura, es en torno a 15 días y 6 meses.

CÓDIGOS POSTALES

Código_postal	Ayuntamiento	Id_provincia
1001	ABREVADERO	1
1002	AMIZKARRA	1
1003	ANDALUCIA	1
1004	ANGULEMA	1
1005	AMARIÑA,Plaza	1
1006	ADURTZA	1
1007	ALTO DE ARMENTIA	1
1008	ADRIANO VI	1
1009	ARGENTINA	1
1010	AITZOL	1
1012	ALDABE	1
1013	ABETXUCO	1
1015	ARANETA	1
1110	CAMPEZO	1
1117	ALDA	1
1118	ANGOSTINA	1
1120	ARRAIA-MAEZTU	1
1128	ANTOÑANA	1
1129	ALETXA	1
1130	MURGUIA	1
1138	ACOSTA	1
1139	ALTUBE	1
1160	IBARRA	1

Figura 28. Alimentación de la tabla Códigos Postales.

Tras la descarga de los códigos postales de la página web de Correos (Sociedad Estatal Correos y Telégrafos, 2014), se importó la información obtenida a un fichero Excel y posteriormente se editó, para su importación a Access. (Véase figura 28).

4.5.2 Generación de datos

Para poder poblar las tablas *líneas compra proveedores*, *líneas de facturas ventas* y *líneas de facturas compras*, era necesario establecer distintos precios para los diferentes artículos. Por lo tanto, mediante fórmulas disponibles en Excel, generamos unos *precios de coste*, unos *precios de venta* y unos *precios de venta al público*. (Véase figura 29).

	A	B	C	D	E	F	G
		Nº_referencia	Precio de coste	Nº_referencia	Precio de venta	Nº_referencia	Precio de venta al público
1							
2		1	=ALEATORIO.ENTRE(10;35000)*0,01		3,72	1	4,09
3		2	241,4	2	260,71	2	286,78
4		3	166,29	3	179,59	3	197,55
5		4	145,58	4	157,23	4	172,95
6		5	79,72	5	86,10	5	94,71
7		6	75,23	6	81,25	6	89,37
8		7	280,27	7	302,69	7	332,96
9		8	142,67	8	154,08	8	169,49
10		9	234,66	9	253,43	9	278,78
11		10	295,67	10	319,32	10	351,26
12		11	133,13	11	143,78	11	158,16
13		12	74,57	12	80,54	12	88,59
14		13	222,08	13	239,85	13	263,83

Figura 29. Generación de precios.

El precio de coste, es el precio al que la empresa Carlin les compra a los proveedores. Estos precios se generaron mediante la fórmula `ALEATORIO.ENTRE()`, estableciéndose como mínimo un precio de 10 céntimos y como máximo un precio de 350€. Este mínimo y máximo, se ha establecido, observando el rango de los precios fijados por la organización actualmente. El precio de venta, es el precio al que Carlin le vende los artículos a los franquiciados. Se estableció un margen de beneficio del 8% sobre el precio de coste. Y por último, el precio de venta al público, es el precio al que las franquicias venden sus productos. Se estableció un margen de beneficios del 10% sobre el precio de venta.

A continuación, se explicará a modo de ejemplo, como se alimentó la tabla de *líneas compra proveedores*. (Véase figura 30). (Ídem las tablas líneas de facturas ventas y líneas de facturas compras).

	A	B	C	D	E	F	G	H
	Id_compra	Nº_línea	Volumen	Importe	Precio de coste	Precio de venta	Precio de venta al público	Nº_referencia
2	1	3	581	81537,54	140,34	151,57	166,72	588
3	1	2	327	40240,62	123,06	132,90	146,20	113
4	1	1	963	133731,81	138,87	149,98	164,98	696
5	2	4	430	84387,50	196,25	211,95	233,15	188
6	2	3	862	39014,12	45,26	48,88	53,77	77
7	2	2	94	21670,76	230,54	248,98	273,88	387
8	2	1	639	2734,92	4,28	4,62	5,08	85
9	4	1	566	12248,24	21,64	23,37	25,71	200
10	5	3	973	200934,23	206,51	223,03	245,33	341
11	5	2	697	203530,97	292,01	315,37	346,91	241
12	5	1	551	27538,98	49,98	53,98	59,38	700
13	6	2	664	166962,80	251,45	271,57	298,72	176
14	6	1	502	11355,24	22,62	24,43	26,87	100

Figura 30. Generación de datos de la tabla líneas compra proveedores.

Para establecer el *Id_compra*, se generaron de manera aleatoria, mediante la función ALEATORIO.ENTRE (), un total de 8.244 facturas. Posteriormente se ordenaron de menor a mayor, para poder establecer los números de líneas de cada factura. De esta manera, utilizando la función CONTAR.SI (), se fueron conociendo cada una de las *líneas* de cada *Id_compra*, puesto que, lo que se pretendía era obtener un acumulativo de cada factura que se repitiese.

En cuanto al *volumen* de compra, se estableció, mediante la función ALEATORIO.ENTRE (), una cantidad entre 1 y 1000 artículos. Utilizando la misma fórmula, se generaron los números de referencia de un total de 737 artículos. Como anteriormente se había creado un fichero Excel con los precios correspondientes a cada artículo, utilizando la función BUSCARV (ALEATORIO.ENTRE ()), se pudo establecer el correspondiente precio para cada número de referencia, de manera que, todos los números de referencia que se repitiesen, tuviesen el mismo precio. Finalmente el *importe* sería el volumen por el precio de coste.

5. EXPLOTACIÓN DE LA INFORMACIÓN

5.1 Almacenes de datos (data warehouses)

Las empresas grandes, tienen una estructura compleja de organización interna y por tanto puede que los diferentes datos se hallen en distintas ubicaciones, es decir provengan de diferentes sistemas operacionales de la organización y/o fuentes externas. Los encargados de adoptar las decisiones empresariales necesitan tener acceso a la información de todos esos orígenes. La formulación de consultas a cada uno de los orígenes es a la vez engorrosa e ineficiente y los almacenes de datos son los encargados de proporcionar una solución a este tipo de problemas.

Bill Inmon¹² fue el que creo la definición aceptada de lo que es un almacén de datos: *“un conjunto de datos orientado a temas, integrado, no volátil, variante en el tiempo, como soporte para la toma de decisiones”*.

Es decir, orientado al usuario, proveniente de fuentes heterogéneas, históricos, no volátil y con necesidad de actualización periódica.

De forma más general podemos definir el almacenamiento de datos como: *“una colección de tecnologías de soporte a las decisiones, que tiene como objetivo que el trabajador del conocimiento (ejecutivo, director, analista) tome decisiones mejores y más rápidas”*. (Elmasri, 2007).

Para comprender íntegramente el concepto de los almacenes de datos, es importante conocer cual es el proceso de construcción del mismo, denominado ETL

¹² Es un científico americano de la computación, reconocido por muchos como el padre del almacén de datos. Bill Inmon defiende una metodología descendente (top down) a a hora de diseñar un almacén de datos.

(Extracción, Transformación y Carga), a partir de los sistemas operacionales de una compañía:

- Extracción: obtención de la información de las distintas fuentes tanto internas como externas.
- Transformación: filtrado, limpieza, depuración, homogeneización y agrupación de la información.
- Carga: organización y actualización de los datos y los metadatos en la base de datos.

Esta herramienta se encarga de corregir y realizar un procesamiento previo de los datos, ya que los orígenes de los datos suelen entregar datos con numerosas inconsistencias, que pueden corregirse. Por ejemplo, los nombres suelen estar mal escritos o puede que los códigos postales se hayan introducido de manera incorrecta. Además también es posible transformar los datos de otras formas, como cambiar las unidades de medida o convertir los datos en un esquema diferente reuniendo datos de relaciones de varios orígenes.

Por lo tanto, los almacenes de datos, proporcionan el acceso a datos para análisis complejos, revelación de conocimientos y toma de decisiones. Dan respuesta a las demandas de alto rendimientos de datos e información de una organización.

Soportan varios tipos de aplicaciones, como OLAP (*on-line analytical processing*), DSS (*decisión support systems*) y aplicaciones de minería de datos (*Data Mining*). (Véase figura 31).

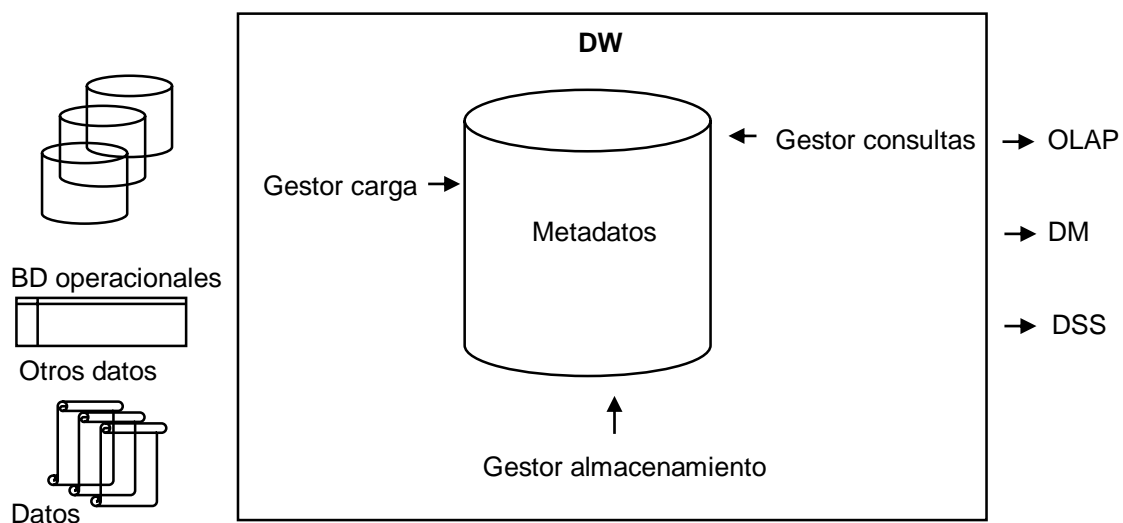


Figura 31. Arquitectura de un almacén de datos.

Las técnicas OLAP “en muchos casos se basan en jerarquías de conceptos para consolidar los datos y crear vistas lógicas a lo largo de las dimensiones de los almacenes de datos” (Vercellis, 2009). Son una solución utilizada en el campo de la llamada inteligencia empresarial (Business intelligence) cuyo objetivo es agilizar las consultas de grandes cantidades de datos. Por este motivo utiliza estructuras multidimensionales (o cubos OLAP) que contienen datos resumidos de grandes bases de datos. Permiten analizar los datos complejos mediante consultas analíticas, generación de informes, marketing y áreas similares. Algunos ejemplos de estas técnicas son:

- **Pivot** (o *rotación*): el cambio de una jerarquía (u orientación) dimensional a otra en un cubo de datos se logra fácilmente gracias a esta técnica. El cubo de datos puede ser observado como si rotara para mostrar una orientación diferente de los ejes. En el caso de la empresa en estudio, se puede pivotar el cubo para mostrar las ventas según la CCAA como filas, los totales de ventas por periodos (por ejemplo trimestres) como columnas y los artículos de la empresa en una tercera dimensión. Por tanto esta técnica equivale a tener una tabla de ventas por CCAA independiente por cada artículo, donde cada una de ellas muestra las ventas por trimestres según la CCAA.
- **Roll-up** (*compactar*): mueve hacia arriba la jerarquía, agrupando en unidades más grandes a lo largo de una dimensión. Por ejemplo, como se ha visto anteriormente, la empresa Carlin organiza sus productos por Categorías, subcategorías y artículos. Esta técnica se desplazaría desde los artículos individuales hasta una lista de categorías de productos.
- **Drill-down** (*descomponer*): ofrece la operación contraria. Por ejemplo descomponiendo las ventas de cada CCAA por provincias y estas a su vez por ayuntamiento.
- **Slice**: Consiste en una operación de “corte” dentro del cubo de datos. Por ejemplo, se quiere analizar las cifras de ventas de todas las CCAA y todas las categorías de artículos de la compañía en el año 2012. Esta técnica extrae estos datos fuera del cubo.
- **Dice**: Permite recoger valores específicos de múltiples dimensiones. El nuevo cubo que se obtiene muestra por ejemplo, las cifras de ventas de un número limitado de categorías de productos, el horizonte temporal y las dimensiones de las CCAA.

Los DSS proporcionan a las personas que han de tomar las decisiones importantes dentro de una organización, datos a nivel superior para la toma de decisiones complejas. En cambio, la minería de datos, se emplea para el descubrimiento de conocimiento: intenta descubrir reglas y estructuras a partir de los datos, es decir descubrir patrones de información a partir de grandes conjuntos de datos. Existen diversidad de técnicas descendientes de la minería de datos, como las de clasificación (clustering), análisis estadística (regresión), reglas de asociación, agrupamientos, árboles de decisión, etc. En este caso nos centraremos en los almacenes de datos y minería de datos para la explotación de la información.

5.2 Análisis de datos: Consultas SQL.

Con la explotación de datos se pretende generar una potencial ventaja competitiva que no es otra que proporcionar información privilegiada para responder a los problemas de negocio. En todas las empresas cada área acumula diferentes datos: sobre sus clientes, sus inventarios, sobre la efectividad de las campañas de marketing, información sobre proveedores y socios, además de los datos que pueden proveer del exterior, como los referentes a los competidores. Por lo tanto, se ha dividido el análisis de los datos en tres áreas concretas, siguiendo fielmente las necesidades de información de la empresa Carlin.

5.2.1 Franquicias

Conocer la marcha del negocio a través de las ventas y de las compras/pedidos, permite a las enseñanzas elaborar auténticas radiografías individualizadas de las necesidades y debilidades de sus franquiciados. Con las siguientes consultas, se pretende recoger datos valiosos sobre el comportamiento económico de las enseñanzas. De esta manera, con meses de antelación se podrán marcar los objetivos de venta de cada tienda junto con los franquiciados, analizar en el funcionamiento de una tienda que factores convergen tanto externos como internos, realizar promociones u ofertas de productos, realizar análisis de perfiles de clientes, conocer la rentabilidad de un producto en concreto, etc.

La empresa ha detectado, a pesar de que en el sector de la papelería es conocida la estacionalidad por su vínculo intrínseco con la actividad escolar y las festividades que se suceden durante todo el año (día del niño, día del estudiante, etc), que las franquicias situadas en Galicia no son capaces de optimizar el volumen de sus compras con el fin de maximizar sus beneficios sobre todo en los meses de verano. Para constatar esta sospecha, fue necesario disponer de los datos históricos de las compras y ventas de las franquicias en los meses de junio, julio y agosto.

```
SELECT franquicias.[fecha de concesión],
franquicias.[fecha de apertura],
franquicias.código_postal,
[facturas compras].fecha,
articulos.nº_referencia,
[líneas de facturas compras].cantidad,
[líneas de facturas compras].importe,
Year([facturas compras].[fecha]) AS Año_compra,
Month([facturas compras].[fecha]) AS Mes_Compra,
provincias.nombre_provincia,
ccaa.nombre_ccaa
FROM [ccaa]
INNER JOIN provincias
ON ccaa.id_ccaa = provincias.id_ccaa)
INNER JOIN ([códigos postales]
INNER JOIN franquicias
ON [códigos postales].código_postal =
franquicias.código_postal)
INNER JOIN [facturas compras]
INNER JOIN [artículos]
INNER JOIN [líneas de facturas compras]
ON artículos.nº_referencia =
[líneas de facturas compras].nº_referencia)
ON [facturas compras].id_factura_compra =
[líneas de facturas compras].id_factura_compra)
ON franquicias.id_franquicia = [facturas compras].id_franquicia)
ON provincias.id_provincia = [códigos postales].id_provincia;
```

COMPRAS	Junio	Julio	Agosto	Total general
2011	321.013,19	163.486,72	238.330,47	722.830,38
GALICIA	321.013,19	163.486,72	238.330,47	722.830,38
2012	255.874,59	306.010,71	367.625,35	929.510,64
GALICIA	255.874,59	306.010,71	367.625,35	929.510,64
2013	185.678,84	344.761,48	228.672,6	759.112,92
GALICIA	185.678,84	344.761,48	228.672,6	759.112,92
TOTAL GENERAL	762.566,63	814.258,91	834.628,42	2.411.453,96

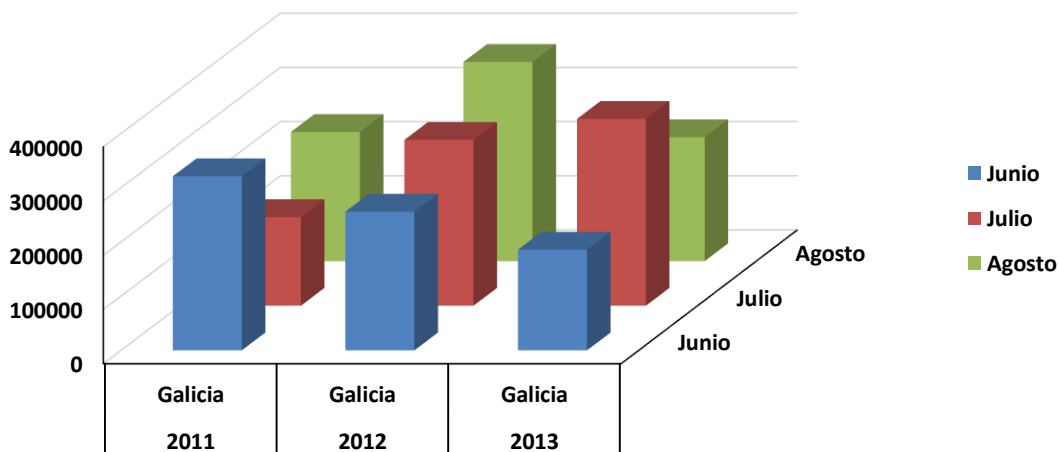


Figura 32. Consulta y gráfico de las compras en verano de las franquicias de Galicia.
Elaboración propia en Excel 2013.

```

franquicias.[fecha de apertura],
[facturas ventas].[fecha],
[lineas de facturas ventas].[cantidad],
[lineas de facturas ventas].[importe],
Month([facturas ventas].[fecha]) AS Mes_venta,
Year([facturas ventas].[fecha]) AS Año_venta,
franquicias.código_postal,
ccaa.renta_per_capita,
ccaa.nombre_ccaa,
ccaa.población
FROM (ccaa
INNER JOIN provincias
ON ccaa.id_ccaa = provincias.id_ccaa)
INNER JOIN ([códigos postales]
INNER JOIN [facturas ventas]
ON [códigos postales].código_postal =
[facturas ventas].código_postal)
INNER JOIN franquicias
ON (franquicias.id_franquicia =
facturas ventas.id_franquicia )
AND ([códigos postales].código_postal =
franquicias.código_postal))
INNER JOIN [lineas de facturas ventas]
ON [facturas ventas].id_factura_venta =
[lineas de facturas ventas].id_factura_venta)
ON provincias.id_provincia = [códigos
postales].id_provincia;
    
```

VENTAS	Junio	Julio	Agosto	Total general
2011	305.279,52	152.487,90	234.427,12	692.194,55
GALICIA	305.279,52	152.487,90	234.427,12	692.194,55
2012	238.000	297.500	331.200	866.700
GALICIA	238.000	297.500	331.200	866.700
2013	156.244,53	320.000	215.000	691.244,53
GALICIA	156.244,53	320.000	215.000	691.244,53
TOTAL GENERAL	699.524,06	769.987,90	780.627,12	2.250.139,09

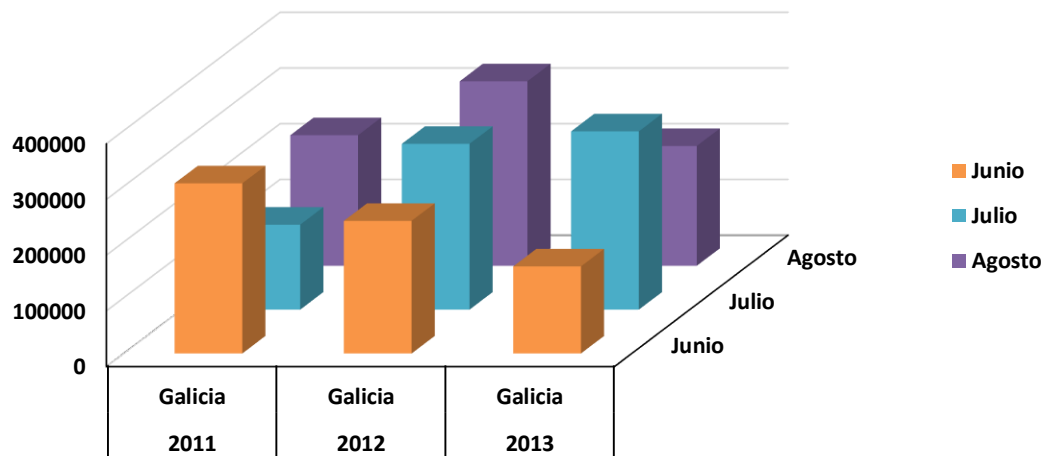


Figura 33. Consulta y gráfico de las ventas en verano de las franquicias de Galicia.

Elaboración propia en Excel 2013.

Por lo general la demanda de los artículos de las franquicias, dicta los tipos y cantidades de materiales que se compran. Por lo tanto es necesario calcular bien las cantidades para evitar el exceso de inventario o en su caso evitar las demoras ocasionadas por la inexistencia de los materiales requeridos. En este caso, comparando ambas tablas de datos (*figuras 32 y 33*), las franquicias situadas en Galicia siempre tienen exceso de inventario en los meses de junio, julio y agosto y no obtienen beneficios ¿Qué debería analizar un directivo de Carlin ante esta situación?

- La central en los meses de verano siempre prevee menos ventas por parte de sus franquiciados y por lo tanto ajusta las campañas de marketing o publicidad a nivel general. Es posible que las inversiones en publicidad, y promociones de

ventas no estén dando los frutos esperados y sea indispensable focalizar las campañas de marketing a nivel local para captar cierto público objetivo.

- Puede que sea necesario identificar de forma más precisa los segmentos de clientes y estudiar con más detalle su comportamiento, para potenciar las ventas.
- Positivamente, el soporte continuado de ayuda al franquiciado está siendo efectivo, ya que la enseña es capaz de detectar a tiempo este tipo de situaciones.

Otra información interesante derivada del almacén de datos, es poder conocer el total de facturación por CCAA, de donde provienen los clientes de una franquicia en particular o ver la evolución de dos franquicias cualesquiera desde su fecha de apertura. Para ello se han creado las siguientes consultas:

Total de ventas por CCAA.

CC.AA	TOP	FACTURACIÓN			FRANQUICIAS
	2010	2011	2012	2013	Total general
COMUNIDAD VALENCIANA	39.432.107,71	45.042.308,2	36.759.757,44	31.654.681,45	152.888.854,8
GALICIA	50.013.651,6	47.777.625,61	57.840.061,72	40.384.947,38	196.016.286,3
MADRID	82.239.895,34	74.881.225,29	77.460.719,27	80.669.602,89	315.251.442,8
TOTAL GENERAL	171.685.654,7	167.701.159,1	172.060.538,4	152.709.231,7	664.156.583,9

```

SELECT [ccaa].nombre_ccaa,
[ccaa].renta_per_capita,
[ccaa].población,
SUM([líneas de facturas ventas].importe) AS Sumalimporte,
Format([facturas ventas].fecha, 'yyyy') AS ano
FROM ((([líneas de facturas ventas]
INNER JOIN [facturas ventas]
ON ([líneas de facturas ventas].id_factura_venta =
[facturas ventas].id_factura_venta)
INNER JOIN [códigos postales]
ON [facturas ventas].código_postal =
[códigos postales].código_postal)
INNER JOIN provincias
ON [provincias].id_provincia = [códigos postales].id_provincia)
INNER JOIN ccaa
ON [ccaa].id_ccaa = [provincias].id_ccaa)
WHERE [facturas ventas].fecha BETWEEN
Format(Cdate('01/01/2009'),
'dd/mm/yyyy') AND
Format(Cdate('31/12/2013'), 'dd/mm/yyyy')
GROUP BY [ccaa].nombre_ccaa,
[ccaa].renta_per_capita,
[ccaa].población,
Format([facturas ventas].fecha, 'yyyy');

```

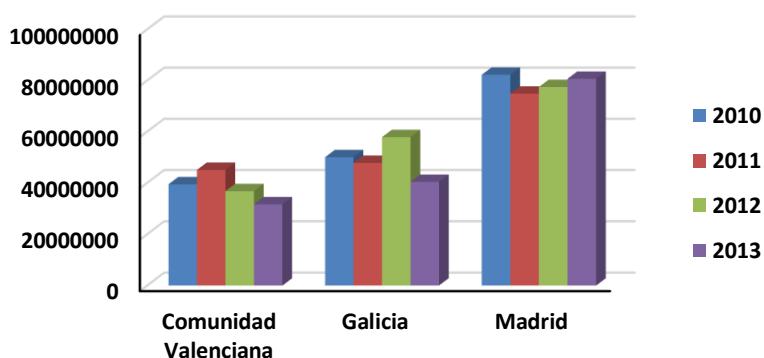


Figura 34. Consulta y gráfico del total de ventas por CC.AA.

Elaboración propia en Excel 2013.

De esta manera, veríamos en la *figura 34*, que en los últimos cuatro años Madrid es la comunidad que más ingresos obtiene en cuanto a facturación por CCAA., seguida de

Galicia y la Comunidad Valenciana. Para el directivo de la compañía sería conveniente evaluar la posibilidad de ubicar nuevas franquicias dentro de estas comunidades. En el caso de las comunidades con menos facturación, sería interesante utilizar la figura del cliente misterioso¹³, con el fin de favorecer el *feedback* y evaluar de forma objetiva el cumplimiento de las directrices de la compañía.

Procedencia de los clientes.

Si lo que se desea analizar, es de donde proceden los clientes de una determinada franquicia, es posible crear una consulta que nos devuelva el número de clientes que acuden a un ID de franquicia en particular y de que ayuntamiento proceden.

```
SELECT [códigos postales].ayuntamiento,
       Count([facturas ventas].código_postal),
       Format([facturas ventas].fecha, "yyyy") AS ano
FROM ([facturas ventas]
      INNER JOIN franquicias
            ON [facturas ventas].id_franquicia =
               [franquicias].id_franquicia)
      INNER JOIN [códigos postales]
            ON [facturas ventas].código_postal =
               [códigos postales].código_postal
WHERE [franquicias].código_postal = 15160
      AND [facturas ventas].fecha BETWEEN
         Format(Cdate('01/01/2008'), "dd/mm/yyyy") AND
         Format(Cdate('31/12/2013'), "dd/mm/yyyy"
              )
GROUP BY [códigos postales].ayuntamiento,
         Format([facturas ventas].fecha, "yyyy");
```

CLIENTES ZONA	AÑO	
	2012	Total general
BASTIAGUEIRO	5	5
BERGONDO	21	21
BETANZOS	7	7
LA CORUÑA	3	3
OLEIROS	23	23
PERILLO	1	1
SADA	55	55
SANTA CRUZ	2	2
TOTAL GENERAL	117	117

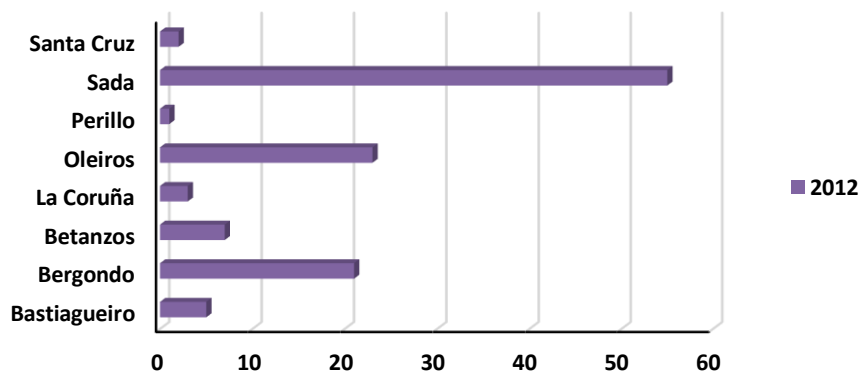


Figura 35. Consulta y gráfico de la procedencia de los clientes de la franquicia 4.

Elaboración propia en Excel 2013.

¹³ Los franquiciadores recurren a la observación por seudocompra o compra simulada mediante la figura conocida como cliente misterioso. Esta es una técnica de investigación que permite evaluar no sólo la atención al cliente, sino también el cumplimiento de las directrices de calidad en la prestación de servicios, siendo el elemento sorpresa un factor importante.

A modo de ejemplo se ha escogido la franquicia con ID: 4 situada en Sada, La Coruña. Como observamos en la *figura 35*, la mayoría de los clientes que acuden a esta franquicia son del ayuntamiento de Sada, seguido del ayuntamiento de Oleiros y Bergondo. Estos datos son totalmente lógicos, puesto que en Oleiros y Bergondo no hay ninguna franquicia de la organización. El directivo de la compañía debería barajar, como vimos anteriormente, la posibilidad de apertura de nuevas franquicias pero en este caso la ubicación sería más concreta. Por otro lado podría evaluar si las estrategias de marketing locales realmente están captando la clientela esperada.

Evolución de dos franquicias según sus ventas.

También, es posible ver la evolución de dos franquicias, desde la fecha de apertura del negocio comparando sus ventas anuales, mediante la siguiente consulta:

AÑO	VENTAS F.82	VENTAS F.133
1991	371.139,29	368.178,53
1992	619.444,92	640.562,48
1993	579.523,95	699.598,48
1994	311.825,09	280.525,10
1995	408.405,48	477.310,87
1996	372.497,45	123.803,07
1997	799.400,93	629.957,45
1998	444.733,83	568.853,74
1999	403.603,62	437.638,35
2000	580.174,68	327.336,64
2001	87.547,90	118.558,02
2002	563.347,56	440.917,14
2003	504.847,82	491.624,78
2004	208.126,24	223.910,93
2005	367.848,25	368.868,13
2006	363.888,33	343.051,20
2007	481.545,06	721.006,82
2008	347.849,56	386.661,79
2009	201.432,83	311.900,03
2010	447.485,57	208.506,16
2011	123.609,64	173.718,16
2012	98.136,77	67.411,80
2013	470.722,57	315.223,25

```
SELECT [facturas ventas].fecha,
[facturas ventas].id_franquicia,
SUM([líneas de facturas ventas].importe)
FROM ([facturas ventas]
INNER JOIN [líneas de facturas ventas]
ON [facturas ventas].id_factura_venta =
[líneas de facturas ventas].id_factura_venta)
INNER JOIN franquicias
ON [franquicias].id_franquicia = [facturas ventas].id_franquicia
WHERE [facturas ventas].fecha >= [franquicias].[fecha de apertura]
AND ( [franquicias].id_franquicia = 133
OR [franquicias].id_franquicia = 82 )
GROUP BY [facturas ventas].fecha,
[facturas ventas].id_franquicia;
```

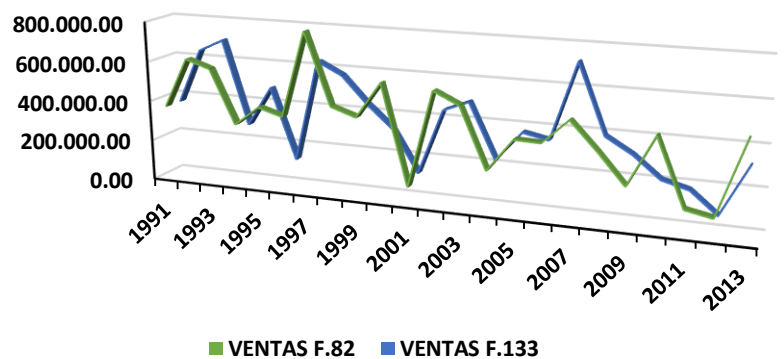


Figura 36. Consulta y gráfico de la evolución de dos franquicias según sus ventas.

Elaboración propia en Excel 2013.

En el gráfico de la *figura 36*, se han seleccionado a modo de comparación por su proximidad las franquicias número 82 situada en Betanzos y la número 133 situada en La Coruña cuya fecha de apertura coincide en 1991. De esta manera es posible ver la evolución de las tiendas que se seleccionen en concreto y llevar a cabo un análisis más detallado de las mismas.

5.2.2 Proveedores

Los proveedores, son una pieza clave, que contribuye al buen funcionamiento de cualquier compañía. Si los proveedores no cumplen con sus obligaciones, la marca ve afectada su imagen. Es importante poder mejorar el poder negociador o establecer estrategias que hagan que la relación con el suministrador sea lo más eficiente posible. Al directivo de Carlin, le interesa saber el gasto del año pasado, en el mes de septiembre en compras a proveedores para poder barajar el llevar a cabo una estrategia de integración hacia atrás, incrementando el control sobre los recursos de suministro de un determinado proveedor.

PROVEEDORES	GASTO
BIC	83.341
FABER CASTELL	66.150
FAIBO	46.723
FIXO NOTES	26.587
GRAFOPLAS	11.500
HP	55.321
MAPED	34.765
NOVUS	21.440
OFFICE BOX	12.800
PAPER MATE	67.453
PILOT	68.964
STAEDTLER	76.456
VIALUX	25.120

```
SELECT [proveedores].marca,
Sum([líneas compra proveedores].importe) AS TotalCompras,
Datepart("M", [compra proveedores].fecha) AS Mes
FROM ([líneas compra proveedores]
INNER JOIN [compra proveedores]
ON [líneas compra proveedores].id_compra =
[compra proveedores].id_compra)
INNER JOIN proveedores
ON [compra proveedores].id_proveedores =
[proveedores].id_proveedores
WHERE Datepart("yyyy", [compra proveedores].fecha) = 2012
GROUP BY [proveedores].marca,
Datepart("M", [compra proveedores].fecha); |
```



Figura 37. Consulta y gráfico del gasto en proveedores.

Elaboración propia en Excel 2013.

En la *figura 37*, se observa que los proveedores que más suministran a la empresa, son BIC y STAEDTLER, seguidos de cerca por PILOT, PAPER MATE y FABER CASTELL. Con esta información la compañía puede aprovechar su fortaleza en el mercado para ejercer cierto control y mejorar el poder negociador con el suministrador que le interese.

5.2.3. Artículos

Toda compañía ofrece una serie de productos o servicios. Carlin pone a disposición del público una serie de artículos de distintas marcas conocidas.

Con los datos históricos disponibles se puede obtener información sobre que artículos se venden más, cuales se venden menos, si se adaptan a las necesidades de los clientes, si la marca blanca de la compañía está bien posicionada en el mercado, fijar estrategias de precios, incluso son posibles análisis de patrones de compra para aprovechar coyunturas de ventas con productos asociados.

Se han analizado los datos referentes a la categoría de artículos de escritura y corrección, más en concreto las subcategorías de bolígrafos, marcadores, minas y portaminas.

ESCRITURA Y	CORRECCIÓN				
ARTÍCULOS	2010	2011	2012	2013	Total general
BOLÍGRAFOS	8.613.947,90	6.006.404,23	8.073.732,82	4.643.183,80	27.337.268,76
MARCADORES	1.766.650,20	2.337.875,26	2.015.511,80	1.915.495,49	8.035.532,77
MINAS	1.423.174,23	2.174.196,62	2.879.282,72	1.388.918,02	7.865.571,61
PORTAMINAS	3.066.375,05	3.377.253,27	3.433.413,95	5.089.658,84	14.966.701,13
TOTAL	14.870.147,39	13.895.729,4	16.401.941,31	13.037.256,18	58.205.074,29

```
SELECT artículos.nº_referencia,
artículos.nombre,
subcategorías.nombre,
categorías.nombre,
[lineas de facturas ventas].importe,
[facturas ventas].fecha,
Year([facturas ventas] ! [fecha]) AS Año_venta
FROM [facturas ventas]
INNER JOIN ((categorías
INNER JOIN (subcategorías
INNER JOIN artículos
ON subcategorías.id_subcategorías =
artículos.id_subcategorías)
ON categorías.id_categorías =
subcategorías.id_categorías)
INNER JOIN [lineas de facturas ventas]
ON artículos.nº_referencia =
[lineas de facturas ventas].nº_referencia)
ON [facturas ventas].id_factura_venta =
[lineas de facturas ventas].id_factura_venta;
```

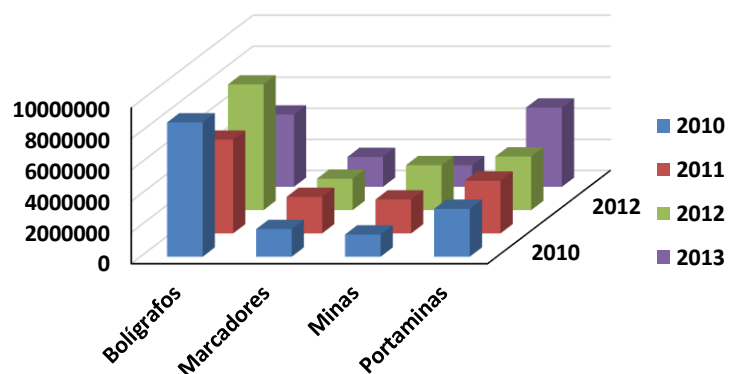


Figura 38. Consulta y gráfico de ventas por subcategoría.

Elaboración propia en Excel 2013.

Y como vemos en el gráfico de la *figura 38*, la subcategoría que más se vende en los últimos cuatro años es la de los bolígrafos. Ahora, si lo que interesa es ver con más detalle que marca de bolígrafos se vende más, podemos utilizar la técnica *drill-down* descomponiendo la subcategoría de bolígrafos y viendo cuales son las marcas en concreto que más se venden.

ESCRITURA Y CORRECCIÓN SUBCATEGORÍA	AÑOS				Total general
	2010	2011	2012	2013	
BOLÍGRAFOS	8.613.947,90	6.006.404,23	8.073.732,82	4.643.183,80	27.337.268,76
BOLIGRAFO BIC 4 COLORES CLASSIC	938.121,97	1.021.667,16	261.959,48	528.951,82	2.750.700,45
BOLIGRAFO BIC 4 COLORES FASHION	375.466,25	551.181,15	976.101,76	948.749,91	2.851.499,09
BOLIGRAFO BIC 4 COLORES FINE	183.461,15	392.715,33	238.618,54	114.415,21	929.210,23
BOLIGRAFO BIC CRISTAL	2.068.243,58	830.095,12	989.426,98	607.985,82	4.495.751,52
BOLIGRAFO BIC CRISTAL GEL	314.332,12	256.050,07	812.927,90	274.352,06	1.657.662,17
BOLIGRAFO BIC CRISTAL STYLUS CON PUNTERO CAPUCHON PARA PANTALLAS TACTILES PUNTA 1 MM...	432.206,93	515.769,60	306.263,19	542.157,81	1.796.397,55
BOLIGRAFO BIC CRISTAL STYLUS CON PUNTERO RETRACTIL PARA PANTALLAS TACTILES PUNTA 1 MM...	1.461.870,97	856.072,96	1.579.135,56	178.175,88	4.075.255,38
BOLIGRAFO CRISTAL CLIC	626.987,24	461.363,88	392.083,52	406.589,10	1.887.023,76
BOLIGRAFO G,2 RETRACTIL PILOT	1.398.485,1	821.168,17	911.496,67	478.776,74	3.609.926,69
BOLIGRAFO INKJOY PAPER MATE	129.511,37	152.245,09	99.049,47	228.625,25	609.431,20
PORTABOLIGRAFO Q,CONNET CON CADENA DE SOBREMESA NEGRO	685.261,18	148.075,64	1.506.669,70	334.404,16	2.674.410,70
TOTAL GENERAL	8.613.947,90	6.006.404,23	8.073.732,82	4.643.183,80	27.337.268,76

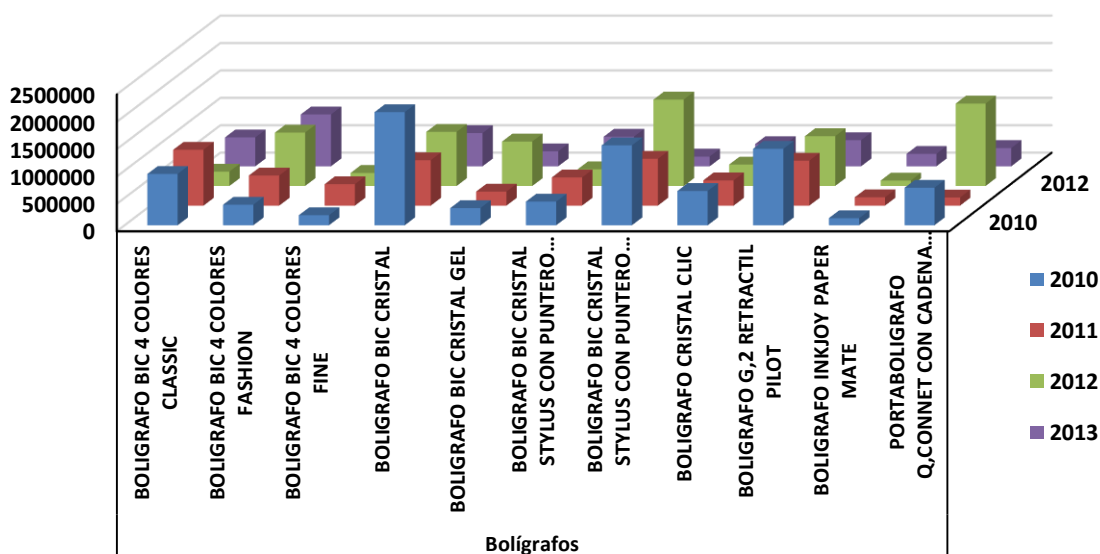


Figura 39. Marcas que más se venden en la subcategoría de bolígrafos.

Elaboración propia en Excel 2013.

Así en la *figura 39*, observamos que los bolígrafos BIC son los más vendidos: BIC CRISTAL en el año 2010, BIC 4 COLORES CLASSIC en el 2011, BIC CRISTAL STYLUS en 2012 y en este último año los más vendidos son los BIC 4 COLORES FASHION. Lo cual guarda relación con la consulta de proveedores vista anteriormente.

VENTAS ESCRITURA Y CORRECCIÓN SUBCATEGORÍA	AÑOS		
	2012	2013	Total general
MINAS	2.024.361,65	1.747.721,66	3.772.083,32
MINAS LIDERPAPEL TUBO DE 12 0,5 MM. HB	284.593,12	46.150,23	330.743,35
MINAS LIDERPAPEL TUBO DE 12 0,7 MM. HB	298.013,43	357.093,55	655.106,98
LIDERPAPEL			
TUBO DE 12 MINAS FABER CASTELL 0.7 MM. HB	280.036,31	284.467,03	564.503,35
TUBO DE 12 MINAS FABER CASTELL 0.5 MM. HB	1.161.718,78	1.060.010,83	2.221.729,62
PORTAMINAS	3.498.837,17	3.777.774,32	7.276.611,49
PORTAMINAS BIC MATIC 0,5 MM.	516.980,21	338.796,33	855.776,54
PORTAMINAS FABER CASTELL TK 0.5 MM.	1.233.941,67	1.122.805,60	2.356.747,27
PORTAMINAS FABER CASTELL TK 0.7 MM.	290.023,06	138.405,80	428.428,86
PORTAMINAS FABER CASTELL TK 1 MM.	636.398,31	783.022,10	1.419.420,42
PORTAMINAS MILAN CAPSULE 0,7 MM. GOMA +	448.456,39	743.424,37	1.191.880,76
PORTAMINAS			
PORTAMINAS STAEDTLER NORIS 777 0,5 MM.	21.332,35	39.473,25	60.805,17
PORTAMINAS SUPER GRIP PILOT	134.660,69	76.577,29	211.237,98
PORTAMINAS UNIBALL SHALAKU 0,5 MM.	217.044,46	535.269,55	752.314,02
CUERPO NEGRO			
TOTAL GENERAL	5.523.198,83	5.525.495,98	11.048.694,82

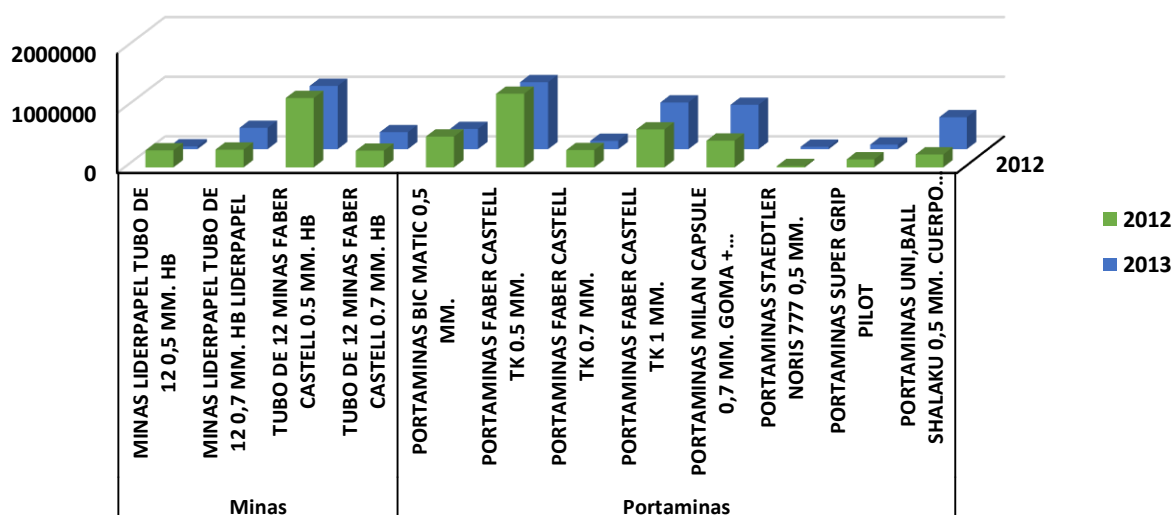


Figura 40. Regla de asociación entre los portaminas y las minas FABER CASTELL.

Elaboración propia en Excel 2013.

También es posible descubrir reglas de asociación entre dos o más artículos. En este caso, veremos que ocurre con la venta de portaminas y minas en los últimos dos años. En el *gráfico 40*, se puede intuir una regla de asociación entre el portaminas FABER CASTELL 0,5 MM y las minas FABER CASTELL 0,5 MM. Esto nos dice que hay una probabilidad bastante elevada de que cuando se venda un portaminas también se vendan las minas de dicha marca, tratándose de un producto no fácilmente sustitutivo.

6. CONCLUSIONES

Una empresa capaz de cumplir sus objetivos en la sociedad actual debe ser “abierta” en el más amplio sentido de la palabra. Actualmente las empresas compiten por satisfacer las necesidades de los clientes con la mayor calidad y agilidad posible, ofreciendo servicios y productos diferenciados y competitivos en calidad/precio. Para todo ello resulta fundamental el alineamiento de las arquitecturas de los sistemas de información con las estrategias corporativas, por lo que los sistemas de información deben constituir una herramienta eficaz que contribuya al rediseño de los procesos de negocio.

Los activos más valiosos de una empresa tienden a no ser los activos tangibles o los depósitos en los bancos, sino los conocimientos, habilidades, valores y actitudes de las personas que forman parte de la empresa. De hecho, para generar riqueza sería suficiente con tener conocimiento sobre un tema determinado y explotarlo de la mejor manera posible. Los factores de la producción como capital, tierra y trabajo, han sido sustituidos por el capital intelectual, que comprende todos aquellos conocimientos tácitos o explícitos que generan valor económico para la empresa.

Como ejemplo práctico, se expuso el caso de Carlin S.A. Esta empresa gallega presentaba ciertas necesidades de información en cuanto a las franquicias, los proveedores y los artículos. Siguiendo la metodología descrita, ha sido posible la generación, desarrollo y análisis de alternativas para adoptar una trayectoria en la toma de decisiones. Se ha visto como a través de la información interna generada por la organización en el funcionamiento rutinario de la empresa, se produce conocimiento y se descubre información no evidente. Según Vendrell (2001), “el conocimiento tiene un gran valor, porque los seres humanos crean a partir de él, nuevas ideas, visiones e interpretaciones que aplican directamente al uso de la información y la toma de decisiones”.

Sin el análisis oportuno, hubiese resultado difícil reconocer ciertas pautas de comportamiento, tendencias, observar la evolución del negocio, detectar cambios en el

consumo o simplemente localizar a tiempo desviaciones, adoptando las acciones oportunas en cada momento.

De esta manera, se ha contemplado como con el análisis de datos históricos, es posible que la empresa aprenda de su historia, de sus mejores prácticas y que pueda evitar tropezarse con los mismos errores del pasado. La acuciante tendencia a explotar la información, está marcando cada vez más la diferencia en el ámbito empresarial.

Bibliografía

Ángeles Saavedra Places, M. R. (2012/2013). Bases de datos como soporte para la toma de decisiones. La Coruña: Universidad de La Coruña.

Bramer, M. (2013). *Principles of Data Mining* (2nd ed.). New York: Springer. ISBN 978-1-4471-4883-8.

Charte Ojeda, F. (2010). *Microsoft Office Access 2010*. España: ANAYA MULTIMEDIA. ISBN 978-84-415-2800-0.

Chen, P. (marzo de 1976). The entity relationship model-toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 9-36.

Connolly, t. M., & Begg, C. E. (2005). *SISTEMAS DE BASES DE DATOS* (Cuarta ed.). Addison Wesley. ISBN 84-7829-075-3.

De Miguel, A. y. (1993). *Concepción y Diseño de Bases de datos del modelo entidad/relación al modelo relacional*. Madrid: Ra-Ma. ISBN 84-7897-083-5.

Elmasri, R., (2007). *Fundamentos de Sistemas de Bases de Datos* (Quinta ed.). Addison Wesley. ISBN 978-84-7829-085-7.

Forums, T. U. (01 de 04 de 2014). <http://www.unix.com/man-page/FreeBSD/1/CUT/>.
Obtenido de <http://www.unix.com/man-page/FreeBSD/1/CUT/>

Foundation, F. S. (01 de 04 de 2014). <http://www.gnu.org/software/grep/>. Obtenido de <http://www.gnu.org/software/grep/>

Foundation, F. S. (01 de 04 de 2014). <https://www.gnu.org/software/wget/>. Obtenido de <https://www.gnu.org/software/wget/>

<https://www.carlin.es/>. (02 de marzo de 2014). Obtenido de <https://www.carlin.es/>

Molina López, J. M., & García Herrero, J. (2006). TÉCNICAS DE ANÁLISIS DE DATOS. *Universidad Carlos III*, 9.

Shackle, G. L. (1966). *Decisión, orden y tiempo en las actividades humanas*. Ed. Tecnos, Madrid. ISBN 978-8-4309-0219-4.

Silberschatz Abraham, Korth Henry, Sudarshan S. (2007). *Fundamentos de diseño de bases de datos* (Quinta ed.). McGraw-Hill. ISBN 978-84-481-5671-8.

Sociedad Estatal Correos y Telégrafos, S. (1 de 04 de 2014). <http://www.correos.es>. Obtenido de http://www.correos.es/ss/Satellite/site/servicio-bd_codigos_postales_inicio/detalle_servicio-sidioma=es_ES-submenu=no.

Vendrell, P. (2001). *Conocimiento: el oro gris de las organizaciones*. Madrid: Fundación DINTEL. ISBN 84-931933-6-4.

Vercellis, C. (2009). *Business Intelligence, Data Mining and Optimization for Decision Making*. WILEY. ISBN 978-0-470-51138-1.