

TEMA 1. FUNDAMENTOS DE BASE DE DATOS

ORGANIZACIÓN LÓGICA DE LOS DATOS. Estructurada

- Sistemas orientados al proceso.
- Sistemas orientados a los datos.

Sistemas orientados al proceso FICHEROS.-

La información contenida en los dispositivos de memoria secundaria se estructura en ficheros. **Un FICHERO es un conjunto de bytes almacenados de forma organizada en un dispositivo de memoria secundaria.** Desde el punto de vista del fichero, su información se almacena en unidades llamadas REGISTROS.

FICHERO O ARCHIVO.- Un fichero almacena información referente a un mismo tema, de forma estructurada con el fin de manipular los datos de forma individual.

De las definiciones anteriores se deduce que un fichero está formado por registros, agrupados o no en bloques, que forman una colección almacenada en un soporte físico de almacenamiento.

TIPOS DE FICHEROS:

Los ficheros pueden clasificarse según diferentes criterios:

a) Por su contenido:

- Texto
- Binario

b) Por la utilidad:

- Permanentes
 - . Maestros o de Situación
 - . Movimientos.
 - . Históricos
- Temporales

c) Por la organización de los datos en el fichero acceso que se realiza al fichero:

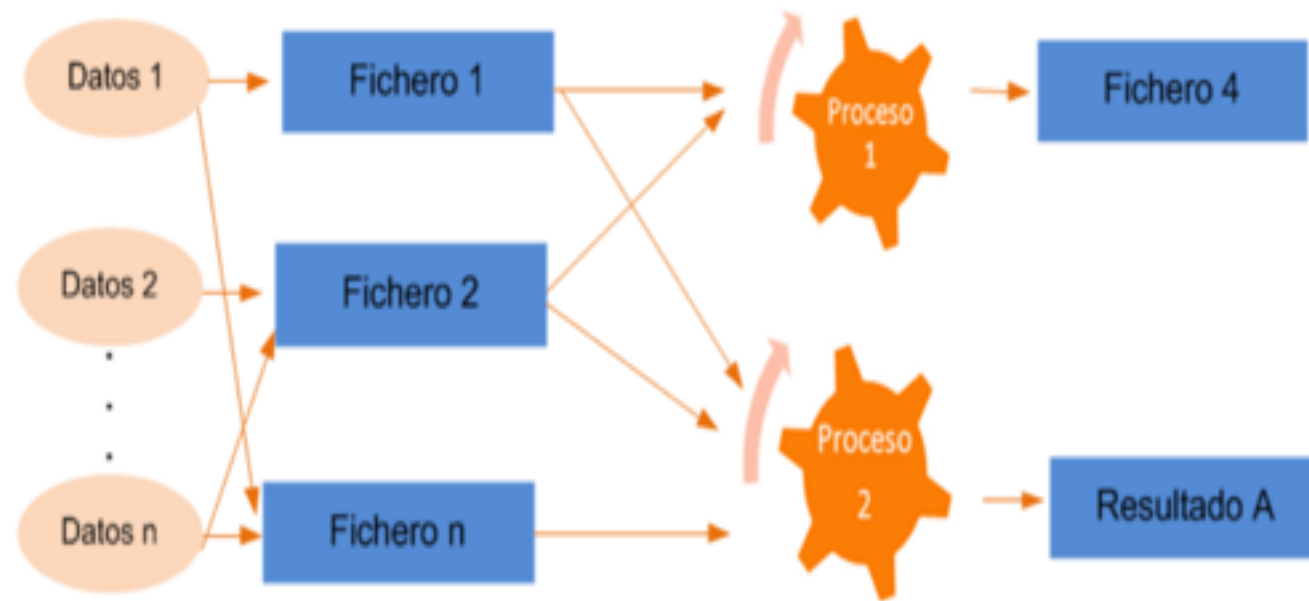
- Archivos secuenciales.
 - Archivos de acceso directo o aleatorio
 - Archivos indexados
- Hoy en día estas clasificaciones están en desuso. Actualmente un sistema operativo trata un fichero desde dos puntos de vista:
- Según su contenido (texto o datos binarios)
 - Según su tipo (imágenes, ejecutables, clips de video)

Los ficheros de texto suelen tener distintas extensiones para conocer que tipo de texto se halla dentro del fichero.

- Ficheros de configuración: .ini, .inf, .cfg
- F. de código fuente: .c, .java, .sql
- F. de páginas web: .html, php, .css, .xml

Los ficheros binarios son los que requieren un formato para ser interpretado.

- De imagen: .jpg
- De video: .avi
- Comprimidos: .rar
- Ejecutables: .exe
- Procesadores de texto: .doc



- Estos eran los sistemas tradicionales, que ponían énfasis en el tratamiento de los datos.
- Las aplicaciones se diseñan e implantan con total independencia unas de otras y los datos no se suelen transferir entre ellas, sino que se duplican siempre que los distintos trabajos los necesitan.
- El sistema clásico de gestión de ficheros genera además de una ocupación excesiva de memoria secundaria, un aumento de los tiempos de proceso, al repetirse los mismos controles y operaciones en los distintos ficheros y, sobre todo, inconsistencias

Sistemas orientados a los datos.

- Para solventar esta dependencia entre los ficheros que almacenan los datos y las aplicaciones que los utilizan surgen, a finales de los 60, las Bases de Datos.

CONCEPTO DE BASES DE DATOS

- “Colección o depósito de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos, y su definición (estructura de la base de datos) única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos.”
- Una base de datos es un juego centralizado y estructurado de datos almacenados en un sistema de computadoras.
- Proporciona utilidades para recuperar, agregar, modificar y suprimir los datos cuando sea necesario.
- También proporciona utilidades para transformar los datos recuperados en información útil.
- Normalmente, una base de datos está gestionada por un administrador de base de datos (DBA).

Cronología de la Historia de la Base de Datos

Años 1960: las computadoras se convierten en rentables para las compañías privadas y aumenta la capacidad de almacenamiento.

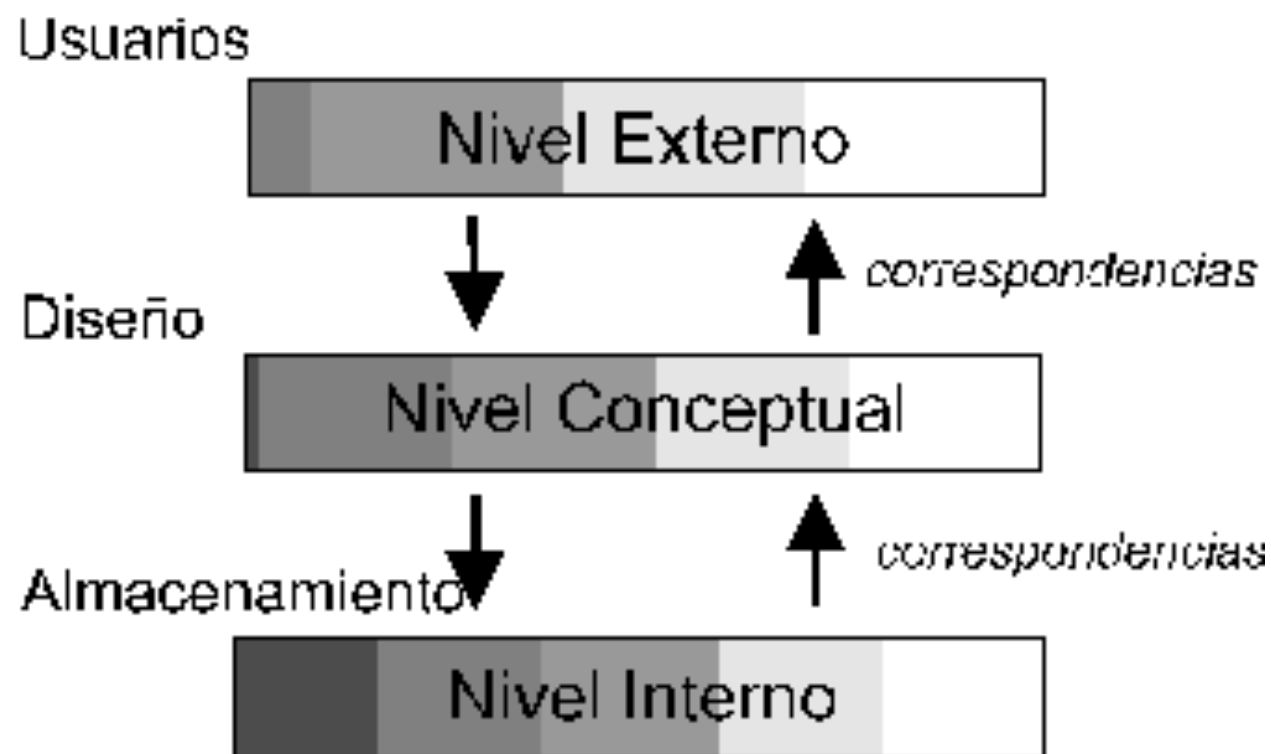
1970-72: E. F. Codd propone el modelo relacional para las bases de datos, desconectando la organización lógica del almacenamiento físico.

1976: P. Chen propone el modelo de relación de entidades (ERM) para el diseño de base de datos.

- Principios de los años 1980: empiezan a aparecer los primeros sistemas de bases de datos relacionales disponibles en el mercado a principios de los años 1980 con Oracle Versión 2.*
- Medios de los años 1980: SQL (lenguaje de consulta estructurado) se convierte en el "estándar intergaláctico".*
- Principios de los años 1990: una revolución del sector comienza con menos compañías supervivientes. Oracle sobrevive.*
- Medios de los años 1990: ¡Boom! Aparece Internet/WorldWide Web (WWW), de gran utilidad. Se produce una terrible lucha para permitir el acceso remoto a sistemas informáticos con datos heredados.*
- Finales de los años 1990: la gran inversión en compañías de Internet ayuda a crear un auge del mercado de las herramientas para conectores web/Internet/bases de datos.*
- Principios del siglo XXI: continúa el constante crecimiento de las aplicaciones de bases de datos. Ejemplos: sitios web comerciales (yahoo.com, amazon.com, google.com), sistemas del gobierno (Oficina de Servicios de Ciudadanía y e Inmigración, Oficina del censo), museos de arte, hospitales, escuelas, etc.*

ARQUITECTURA DE BASES DE DATOS

- En 1975, el comité ANSI-SPARC propuso una arquitectura de tres niveles para los sistemas de bases de datos, que resulta muy útil a la hora de conseguir estas tres características.
- La definición de un sistema de información es la descripción detallada de la arquitectura del sistema. Las arquitecturas de bases de datos han evolucionado mucho desde sus comienzos, aunque la considerada estándar hoy en día es la descrita por el comité ANSI/X3/SPARC (Standard Planning and Requirements Committee of the American National Standards Institute on Computers and Information Processing), que data de finales de los años setenta.



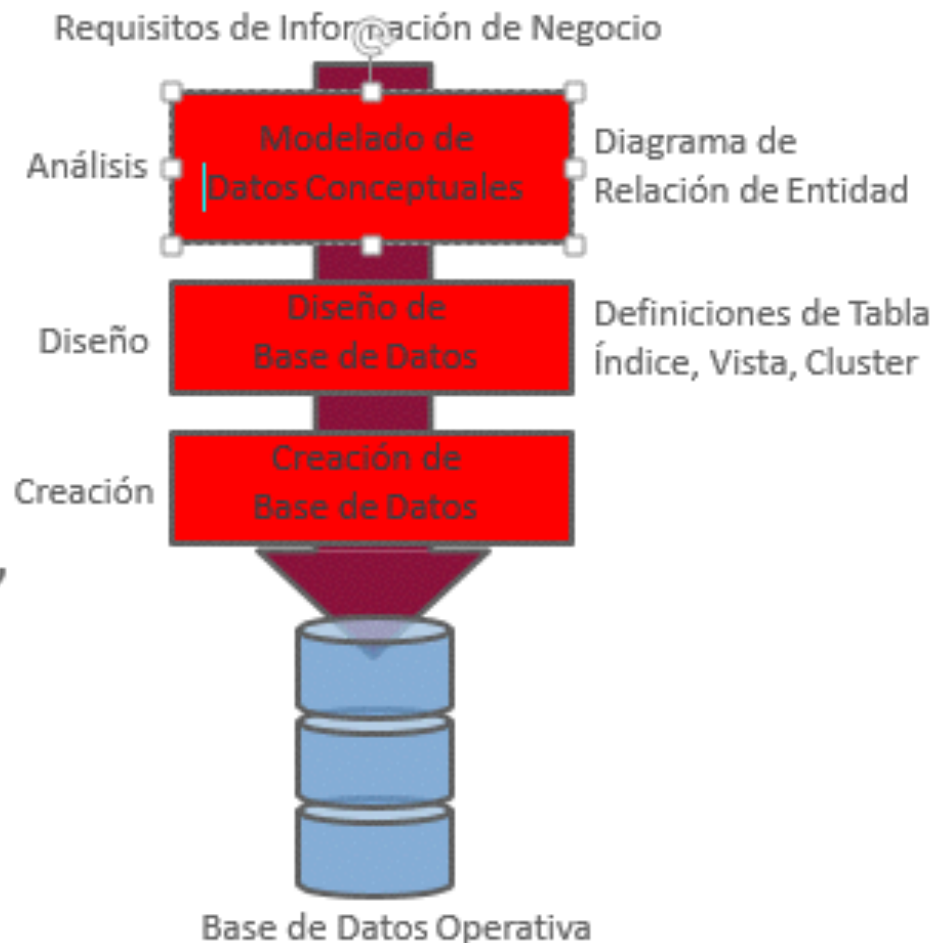
Modelado de los datos

- La modelización de datos es una actividad que se realiza a lo largo del proceso de desarrollo del sistema de información. Cuenta con varias fases :
- Modelo conceptual: Análisis del proceso
- Modelo Lógico de los datos: Se describe mediante el modelo E/R
- Modelo físico: El diagrama E/R se refina convirtiéndose en un modelo físico de datos. Para ello utilizamos el modelo relacional en las bases de datos relacionales
- Finalmente se implanta en el SGBD elegido utilizando SQL para ello.

Modelado de una base de datos

- El modelado de datos es el primer paso en el proceso de desarrollo de bases de datos.
- Implica recopilar y analizar los datos que necesita un negocio para realizar un seguimiento y, a continuación, traza la organización de los datos en un diagrama de relación de entidad.

Proceso de Desarrollo de Base de Datos



CONCEPTO DE SISTEMA GESTOR DE BASES DE DATOS

- El sistema de gestión de la base de datos (SGBD) es una aplicación que permite a los usuarios definir, crear y mantener la base de datos, y proporciona acceso controlado a la misma.
- El SGBD es la aplicación que interacciona con los usuarios de los programas de aplicación y la base de datos.
- En general, un SGBD proporciona los siguientes servicios:
 - Permite la definición de la base de datos mediante el lenguaje de definición de datos. Este lenguaje permite especificar la estructura y el tipo de los datos, así como las restricciones sobre los datos. Todo esto se almacenará en la base de datos.
 - Permite la inserción, actualización, eliminación y consulta de datos mediante el lenguaje de manejo de datos.

CONCEPTO DE SISTEMA GESTOR DE BASES DE DATOS

Conjunto de programas que nos permiten gestionar la B.D. que es lo que nos facilita la manipulación de gran cantidad de información.

Un SGBD nos va a permitir la implantación, acceso y mantenimiento de la B.D.

$S.G.B.D + B.D = SISTEMA DE BASES DE DATOS$

Cualquier operación que queramos realizar contra la B.D. ha de estar procesada por el gestor.

COMPONENTES DE UN SGBD

Un SGBD tiene varios módulos, cada uno de los cuales realiza una función específica. El sistema operativo proporciona servicios básicos al SGBD, que es construido sobre él.

- **El procesador de consultas** es el componente principal de un SGBD. Transforma las consultas en un conjunto de instrucciones de bajo nivel que se dirigen al gestor de la base de datos.
- **El gestor** de la base de datos es el conjunto de programas que se encargan de garantizar la privacidad, seguridad e integridad de los datos, el acceso concurrente a ellos así como interaccionar con el sistema operativo.

El gestor de la base de datos es el interface con los programas de aplicación y las consultas de los usuarios. El gestor de la base de datos acepta consultas y examina los esquemas externo y conceptual para determinar qué registros se requieren para satisfacer la petición. Entonces el gestor de la base de datos realiza una llamada al gestor de ficheros para ejecutar la petición.

- **El gestor de ficheros** maneja los ficheros en disco en donde se almacena la base de datos. Este gestor establece y mantiene la lista de estructuras e índices definidos en el esquema interno. Pero el gestor de ficheros no realiza directamente la entrada y salida de datos. Lo que hace es pasar la petición a los métodos de acceso del sistema operativo que se encargan de leer o escribir los datos en el sistema.
 - **El preprocesador del LMD** convierte las sentencias del LMD embebidas en los programas de aplicación, en llamadas a funciones estándar escritas en el lenguaje anfitrión. El preprocesador del LMD debe trabajar con el procesador de consultas para generar el código apropiado.
 - **El compilador del LDD** convierte las sentencias del LDD en un conjunto de tablas que contienen metadatos. Estas tablas se almacenan en el diccionario de datos.
 - **El diccionario de datos** es una “*metabase de datos*” es decir es una base de datos que contiene información sobre otra base de datos. En el d.d. se almacena toda la descripción de los datos, información referente a la estructura de datos, relaciones entre ellos, gestión e implantación de la B.D.

EL D.D. debe contener todo lo que cualquier usuario quiera saber sobre la B.D.:

- Descripción externa, conceptual e interna de la B.D
- Transformación entre los tres niveles
- Restricciones sobre datos
- Acceso a datos
- Descripciones de las cuentas de usuario
- Las autorizaciones de cada usuario
- Los esquemas externos de cada programa

LENGUAJES

Los SGBD deben ofrecer lenguajes e interfaces apropiadas para cada tipo de usuario: administradores de la base de datos, diseñadores, programadores de aplicaciones y usuarios finales. Se tratan de lenguajes que aportan una gramática sencilla, fácil de entender por usuarios no expertos.

Lenguaje de definición de datos (LDD)

Para garantizar la independencia de los datos es necesaria la definición a los diferentes niveles de abstracción, por lo tanto el SGBD necesita un lenguaje que nos permita la realización de esta tarea.

Una vez finalizado el diseño de una base de datos y escogido un SGBD para su implementación, el primer paso consiste en especificar el esquema conceptual y el esquema interno de la base de datos, y la correspondencia entre ambos. En muchos SGBD no se mantiene una separación estricta de niveles, por lo que el administrador de la base de datos y los diseñadores utilizan el mismo lenguaje para definir ambos esquemas, es el lenguaje de definición de datos (LDD). ***en SQL createtable/createindex...***

Después de escribirlos habrá un compilador LDD que será el que se va a encargar de generar el conjunto de ficheros para almacenar en el disco. (Diccionario de datos).

Lenguaje de modificación de datos (LMD)

Una vez creados los esquemas de la base de datos, los usuarios necesitan un lenguaje que les permita manipular los datos de la base de datos: realizar consultas, inserciones, borrados y modificaciones. Este lenguaje es el que se denomina lenguaje de manejo de datos (LMD).

Dependiendo del modelo de datos y el SGBD existen dos tipos de DML:

- No Procedimentales: solo requieren que en las instrucciones o sentencias del lenguaje se especifiquen que datos se van a manipular siendo el propio DML el encargado de determinar los procedimientos más efectivos para ello. (por ejemplo: SQL)
- Procedimentales: los cuales requieren que en las sentencias del lenguaje se especifiquen que acciones deben realizarse. (por ejemplo: PL/SQL)

Lenguaje de control de datos (LCD)

Se utiliza para controlar el acceso a la información de la B.D. definiendo privilegios y tipos de acceso así como para el control de la seguridad de los datos. De esta tarea se encarga el administrador.

CLASIFICACIÓN DE LAS BASES DE DATOS

Podemos realizar una clasificación atendiendo a varios criterios:

1. Número de usuarios a los que da servicio el sistema.

Los sistemas **monousuario** sólo atienden a un usuario a la vez, y su principal uso se da en los ordenadores personales.
Los sistemas **multiusuario**, entre los que se encuentran la mayor parte de los SGBD, atienden a varios usuarios al mismo tiempo.

2. Donde se encuentra la base de datos:

Centralizados: Casi todos los SGBD son centralizados, sus datos se almacenan en un solo computador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la base de datos en sí residen por completo en una sola máquina.

Distribuidos En los SGBD distribuidos la base de datos real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red. Los SGBD distribuidos homogéneos utilizan el mismo SGBD en múltiples sitios. Tipos de Datos:

- Sistemas relacionales estándar(int, char...)
- XML(para el caso de las bases de datos que trabajan con documentos XML)
- Objetos –relaciones. Tipos complejos de datos.
- Objetos .Tipos de datos objetos y métodos asociados.

4. Lenguajes soportados:

- SQL estandar.
- NoSQL(bd- documentales)Gestores de base de datos de pago

5. Gestores de base de datos de pago

Los mejores gestores de base de datos y más utilizados, con licencia de pago, son los siguientes

- **Oracle:** Es uno de los sistemas de gestión de base de datos relacional más fiable y usado. Pertenece a Oracle Corporation y se desarrolló en 1977
- **SQL Server:** En contraposición y como competencia directa a Oracle, está SQL Server de Microsoft. Ambos gestores ocupan gran parte de la cuota de mercado del sector de base de datos. SQL Server comparte muchas características con Oracle aunque, por supuesto, también existen diferencias. SQL Server se ejecuta en **Transact-SQL**, un conjunto de programas que añaden características al programa, como pueden ser el tratamiento de errores y excepciones, procesamiento de datos, extracción de datos directos de Web, uso de distintos lenguajes de programación, etc... y más características que hacen de SQL Server un gestor muy completo. Además, destaca por su carácter administrativo tanto en sus funciones y seguridad como en la flexibilidad de las bases de datos.

6. Gestores de base de datos de Open Source

En contrapunto a los anteriores gestores de pago, mencionaremos, a continuación, dos de los principales gestores de Open Source:

- **MySQL:** Gestor de simple instalación que actúa del lado del cliente (servidor) y de código abierto con licencia comercial disponible. Actualmente, pertenece a Oracle Corporation. Gestiona bases de datos relacionales, es multiusuario y el más usado dentro del software libre. Destaca por requerir de poca memoria y procesador para funcionar, dando lugar además a una mayor velocidad en sus operaciones. Es usado principalmente para el desarrollo web.
- **FireBird:** Potente y, a la vez, sencillo sistema de gestión de base de datos relacional SQL. Es uno de los mejores gestores de código abierto (Open Source) compatible con Windows y Linux. Entre otras funciones, da soporte completo para procedimientos almacenados, transacciones compatibles con las características ACID y métodos de acceso múltiple (nativo, Python, .NET, etc...).
- **MariaDB** es un sistema de gestión de bases de datos derivado de [MySQL](#) con licencia GPL (General Public License). Es desarrollado por Michael (Monty) Widenius(fundador de [MySQL](#)), la fundación MariaDB y la comunidad de desarrolladores de software libre.¹ Introduce dos motores de almacenamiento nuevos, uno llamado [Aria](#) -que reemplaza con ventajas a [MyISAM](#)- y otro llamado [XtraDB](#) -en sustitución de [InnoDB](#). Tiene una alta compatibilidad con MySQL ya que posee las mismas órdenes, interfaces, APIs y bibliotecas, siendo su objetivo poder cambiar un servidor por otro directamente

7. Bases de datos Nosql

- En informática, NoSQL (a veces llamado "no sólo SQL") es una amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico del sistema de gestión de bases de datos relacionales (RDBMS) en aspectos importantes, el más destacado es que no usan SQL como el principal lenguaje de consultas.
- Los datos almacenados no requieren estructuras fijas como tablas, normalmente no soportan operaciones JOIN, ni garantizan completamente ACID (atomicidad, consistencia, aislamiento y durabilidad), y habitualmente escalan bien horizontalmente. Los sistemas NoSQL se denominan a veces "no sólo SQL" para subrayar el hecho de que también pueden soportar lenguajes de consulta de tipo SQL.
- Por lo general, los investigadores académicos se refieren a este tipo de bases de datos como almacenamiento estructurado, término que abarca también las bases de datos relacionales clásicas.
- A menudo, las bases de datos NoSQL se clasifican según su forma de almacenar los datos, y comprenden categorías como clave-valor, las implementaciones de BigTable, bases de datos documentales, y Bases de datos orientadas a grafos
- Cuando pensamos en bases de datos relacionales a nuestra mente suelen acudir los mismos nombres. En la parte comercial tenemos Oracle y Microsoft SQL Server. Del lado del software libre, tenemos opciones como Postgre SQL o MySQL. Al final todo son tablas, columnas, claves primarias, y sobre todo, consultas SQL. La decisión de cuál elegir, se basará en sus características y precio.
- Si hablamos de bases de datos NoSQL, la cosa se complica. A día de hoy existen unos 150 sistemas de bases de datos NoSQL. Elegir uno de ellos puede ser muy difícil, ya que ninguno ha obtenido todavía la fama que sí han conseguido las bases de datos relacionales.
- Pero el problema principal que encontramos, es que aunque todas se denominan NoSQL, en realidad hay diferentes tipos. Dependiendo de lo que necesitemos, deberemos decantarnos por una u otra.

En esta categoría encontramos:

- [MongoDB](#): probablemente la base de datos NoSQL más famosa del momento. En octubre del año pasado, MongoDB conseguía 150 millones de dólares en financiación, convirtiéndose en una de las startups más prometedoras. Algunas compañías que actualmente utilizan MongoDB son Foursquare o eBay.
- [CouchDB](#): es la base de datos orientada a documentos de Apache. Una de sus interesantes características es que los datos son accesibles a través de una API Rest. Este sistema es utilizado por compañías como Credit Suisse y la BBC.