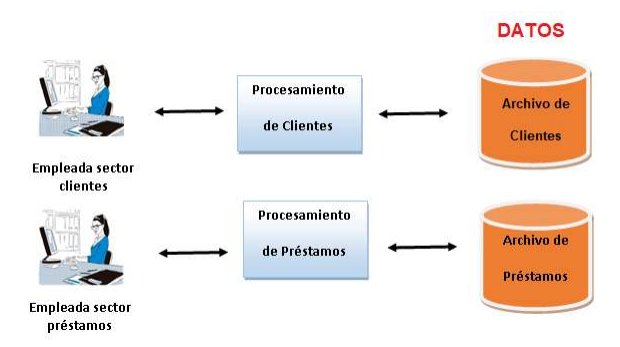
INTRODUCCIÓN AL SOFTWARE DE LOS SISTEMAS DE COMPUTACIÓN

**Bases De Datos**

Si se considera que una de las aplicaciones más importantes de una computadora es el almacenamiento, recuperación y mantenimiento de grandes cantidades de información, los archivos como aplicación convencional no es del todo eficiente. Los archivos se diseñan de acuerdo con los programas, donde el programador decide si debe haber archivos, cuántos deben ser. Qué organización tendrán, que información contendrán, qué programas actuarán sobre ellos y cómo lo harán.

En la siguiente figura puede verse un ejemplo abreviado referido a un Banco.



En este Banco, hay un sector de **Clientes** que guarda los datos necesarios de todas las personas que son clientes del Banco y allí deberán dirigirse aquellos que quieran serlo. Por otra parte, se encuentra el sector de **Préstamos** donde cada persona que necesite solicitar un préstamo deberá dirigirse allí. En este sector, la persona que atiende le solicita también los mismos datos personales que en el sector de clientes, además de asignarle el monto del préstamo, el número de cuotas, la fecha en que solicita el préstamo y la fecha de vencimiento de las cuotas entre otros.

Esta modalidad tiene la ventaja de ser lo suficientemente eficiente, ya que el archivo está pensado para el programa que lo va a usar. Sin embargo, con un sistema tradicional de archivos, cada sector de una misma empresa será el responsable de crear y mantener los datos necesarios, aún cuando éstos estén duplicados.

De esta manera, se pueden encontrar los siguientes problemas:

|  |  |
| --- | --- |
| Problemas | Características |
| **Actualización de la Información** | La actualización puede resultar costosa cuando se tiene información total o parcialmente duplicada en archivos diferentes. Esto conduce a inconsistencia de datos |
| **Redundancia** | Consiste en tener datos que no aportan información, porque pueden ser deducidos de otros. |
| **Rigidez en la búsqueda** | No siempre el modo de acceso que tiene el archivo es el más eficiente, no pudiendo cambiarse. |
| **Dependencia con los programas** | Cualquier cambio en la estructura del archivo implica una modificación de los programas que lo tratan. |
| **Confidencialidad y seguridad** | La confidencialidad consiste en evitar el acceso a determinados usuarios. La seguridad consiste en que los datos no puedan ser modificados por usuarios no autorizados. Ambas cosas deben hacerse por programa. |

Las bases de datos surgen como una alternativa válida a los sistemas de archivos, que facilita la manipulación de grandes cantidad de información, para hacerla segura y accesible a una variedad de usuarios para una variedad de propósitos.

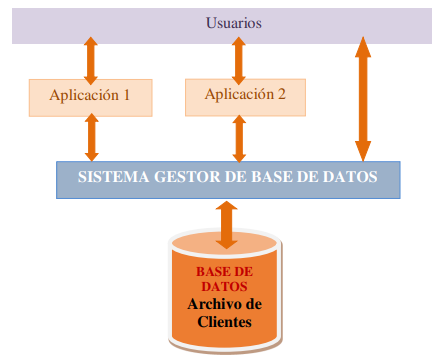
*Una base de datos es un sistema formado por un conjunto de datos organizados de tal manera  
que se controla el almacenamiento de datos redundantes, los datos resultan independientes de los programas que los usan y se pueden acceder a ellos de diversas formas.*

Los requisitos de una buena base de datos son:

* Varios usuarios accediendo a la base de datos y cada uno accederá a determinada información.
* Se controlará el acceso de los usuarios asegurando confiabilidad y seguridad.
* Los datos se almacenan sin redundancia, excepto en casos especiales (redundancia aceptable).
* Se accede de distintas maneras, flexibilizando las búsquedas.
* Deben existir mecanismos concretos de recuperación de información en caso de fallos.
* Se puede cambiar el soporte físico sin repercusión en los programas que usan la base.
* Se puede modificar los contenidos, las relaciones o agregar nuevos datos sin afectar los programas que usan la base de datos.
* Existe una interfaz de la base de datos que permite usarla de forma cómoda y sencilla.

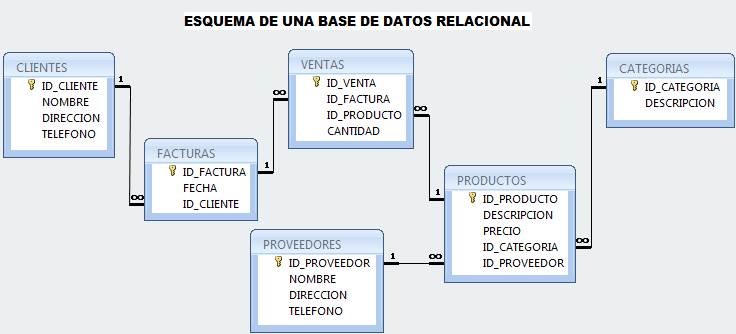
Sistema gestor de base de datos

Se denomina Sistema de Gestión de Base de Datos (SGBD o DBMS, Data Base Management System), al conjunto de software destinado a la creación, gestión, control y manipulación de la información sobre una base de datos. Los SGBD tienen como propósito registrar y mantener información.



Bases de datos relacionales

Una base de datos relacional es una colección de elementos de datos organizados en un conjunto de tablas formalmente descritas desde la que se puede acceder a los datos.



**Componentes**

Tablas o Entidades:

Las tablas son sistemas de clasificación constituidos por filas horizontales y columnas verticales que permiten agrupar datos y presentarlos de forma ordenada. La tabla lleva el nombre de la entidad que representa.

Es un objeto real o abstracto con características particulares, capaces de hacerse distinguir de otros objetos, y del cual se almacena información en la base de datos. Una entidad toma como significado a conceptos u objetos que tienen una importancia en el sistema u organización.

Si tomamos el ejemplo de una base de datos que guarde la información de la Facultad de Ingeniería de la UNJu, acerca de los alumnos, los docentes, las materias y las aulas donde se dictan, se tomaría como Tablas o entidades ALUMNO, PROFESOR, MATERIA y AULA.

Atributos

Un atributo es una unidad básica e indivisible de información acerca de una entidad que sirve para identificarla o describirla.

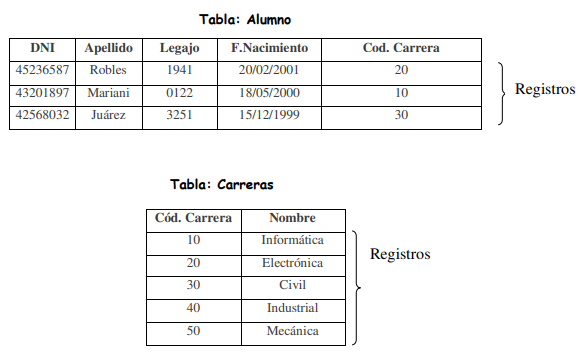
Si continuamos con el ejemplo anterior, atributos de la entidad ALUMNO son: DNI, Apellido y nombre, fecha de nacimiento, domicilio, carreta que cursa (Informática, Industrial, química, etc), teléfono, código de materia que cursa, etc.

Registros o Tuplas

En una base de datos, la información de cada tabla o entidad se almacena en registros, y cada atributo, en campos de dicho registro. Existen distintos tipos de registros dentro de la misma base, ya que cada entidad necesitará una estructura distinta.

En una base de datos habrá tantos tipos de registros como entidades haya, mientras que en un archivo sólo hay un tipo único de registro.

Siguiendo con el ejemplo anterior, la estructura del registro de la entidad ALUMNO va a ser muy distinta del registro de la entidad CARRERAS. ya que ambas entidades tienen distintos atributos.

****

Claves: Superclave, clave candidato, clave principal o primaria y clave foránea o ajena.

Se llama **superclave** de una entidad a un atributo o conjunto de atributos que permite identificar de forma única a un registro de una entidad.

Si de una superclave no se puede obtener *ningún* subconjunto que a su vez sea superclave. se dice que dicha superclave es **clave candidata**.

De todas las claves candidatas existentes, el diseñador de la base de datos, escogerá una que individualizará de forma inequívoca a cada registro de la entidad. Esta clave se denomina **clave principal o primaria.**

Cuando existe una referencia entre dos entidades, esto es cuando mi campo o conjunto de campos de una de las entidades es la clave de otra. se las llama **clave foránea o ajena**. Esta clave foránea permite localizar una entidad a partir de otra.

En el ejemplo anterior, en la entidad ALUMNO algunas superclaves pueden ser:

* DNI
* Legajo (LU)
* DNI + Apellido + nombre
* DNI + Apellido + nombre + fecha de nacimiento

Las claves candidatas serían todas aquellas que no permitan encontrar un subconiunto que pueda ser a su vez superclave, por ejemplo

* DNI
* Legajo

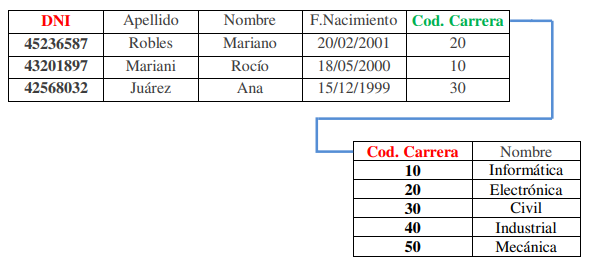
Entre estas claves candidatos, se debe escoger DNI, ya que podría haber dos alumnos con el mismo apellido y nombre. Entonces DNI es la clave primaria o principal.

La entidad ALUMNO tiene un atributo código de carrera que cursa. En la entidad CARRERAS el código de la materia es la clave primaria o principal. Por lo tanto, en la entidad ALUMNO el atributo código de materia que cursa es una clave foránea, ya que relaciona a esta entidad con la entidad CARRERA.

Relaciones

Las entidades por sí solas no describen la realidad de un sistema de información. Además de  
identificar objetos, hay que establecer las asociaciones existentes entre ellos. Esto es una relación: la existencia de algo común entre entidades.

Siguiendo con el ejemplo, existe la relación entre las entidades ALUMNO - CARRERAS (el alumno cursa una carrera).



En el ejemplo se observa en color rojo a las claves principales de cada una de las tablas (“DNI” en la tabla ALUMNO y “Cod. Carrera” en la tabla CARRERAS) y en verde a la clave foránea, en este caso se encuentra en la tabla ALUMNO y es “Cod. Carrera”. Este campo es el que permite la relación entre las tablas.

Cardinalidad

La cardinalidad de una relación representa la participación en la relación de cada una de las entidades afectadas. La cardinalidad siempre se evalúa de a dos entidades.

Existen tres tipos posibles

1:1 (una a una): a cada registro de una entidad le corresponde no mas de un registro de la otra y viceversa. Es biunívoca.



La relación EMPLEADO – CONYUGE (Un empleado puede estar casado con una única persona).

1:N (una a muchas): a cada registro de la primera entidad le pueden corresponder varios registros de la segunda, y a cada registro de la segunda le corresponde no mas de uno de la primera.



La relación PADRE – HIJO (un padre puede tener muchos hijos, pero estos tendrán un solo padre).

N:N (muchas a muchas): a cada registro de la primera entidad le pueden corresponder varios registros de la segunda y viceversa.



La relación PROFESOR – ALUMNO (un profesor da clases a muchos alumnos y un alumno tiene varios profesores)

Usuarios de una base de datos

|  |  |  |
| --- | --- | --- |
| Tipo de Usuario | Características | Responsabilidades |
| Administrador de Base de Datos | Encargado control general del sistema base de datos. | * Decidir el contenido de la información. * Decidir las estructuras de almacenamiento y la estrategia de acceso. * Vincularse con el resto de los usuarios de la base de datos. * Definir controles de autorización y validación. * Definir estrategia de respaldo y recuperación de datos por fallas del sistema. * Controlar el rendimiento y utilización de la base. - Responder a los cambios en los requerimientos. |
| Programador de aplicaciones | Escriben los programas que utilizan las Bases de Datos. | * Realizar tareas de creación, borrado, consulta, recuperación o modificación de datos. - Diseñan aplicaciones para apoyar al usuario final. * Diseñan aplicaciones para apoyar al usuario final. |
| Usuario Final | Accede a la Base de Datos desde una terminal. | * Utiliza los programas creados por el programador. |

Aplicaciones De Las Bases De Datos

Big Data:

Si bien el concepto es nuevo, el manejo de grandes volúmenes de datos se remonta hacia los años 1960/1970 cuando se comenzaron a estructurar los datos en las bases de datos relacionales.

Hacia 2005, se comenzó a observar la gran cantidad de datos que se generaban a través de las redes sociales y servicios online. Eso hizo que se comenzara a trabajar para ver cuál podría ser la utilidad que se le podía dar a esos datos, resultando ser útiles para gran cantidad de sectores tanto públicos como privados.

A

Con la llegada de Internet de las cosas, hay muchos dispositivos conectados a Internet que generan datos sobre patrones de uso de los clientes y el rendimiento de los productos, generando que estos datos bien procesados permitan tomar decisiones más certeras.

¿Qué es Big Data?

*“Big data son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior”.* Gartner (aprox. 2001)

En términos generales se puede decir que es la tendencia existente en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis.

De tal manera que, el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.

El Big Data está basado en tres características denominadas “Las tres V”:

* Volumen:

La cantidad de datos importa.

Con Big Data, hay que procesar grandes volúmenes de datos no estructurados que pueden tratarse de datos de valor desconocido, como feeds de datos de Twitter, flujos de clics de una página web o de aplicación para móviles, o de equipos con sensores.

Para algunas organizaciones, esto puede suponer decenas de terabytes de datos, para otras, incluso cientos de Petabytes.

* Velocidad:

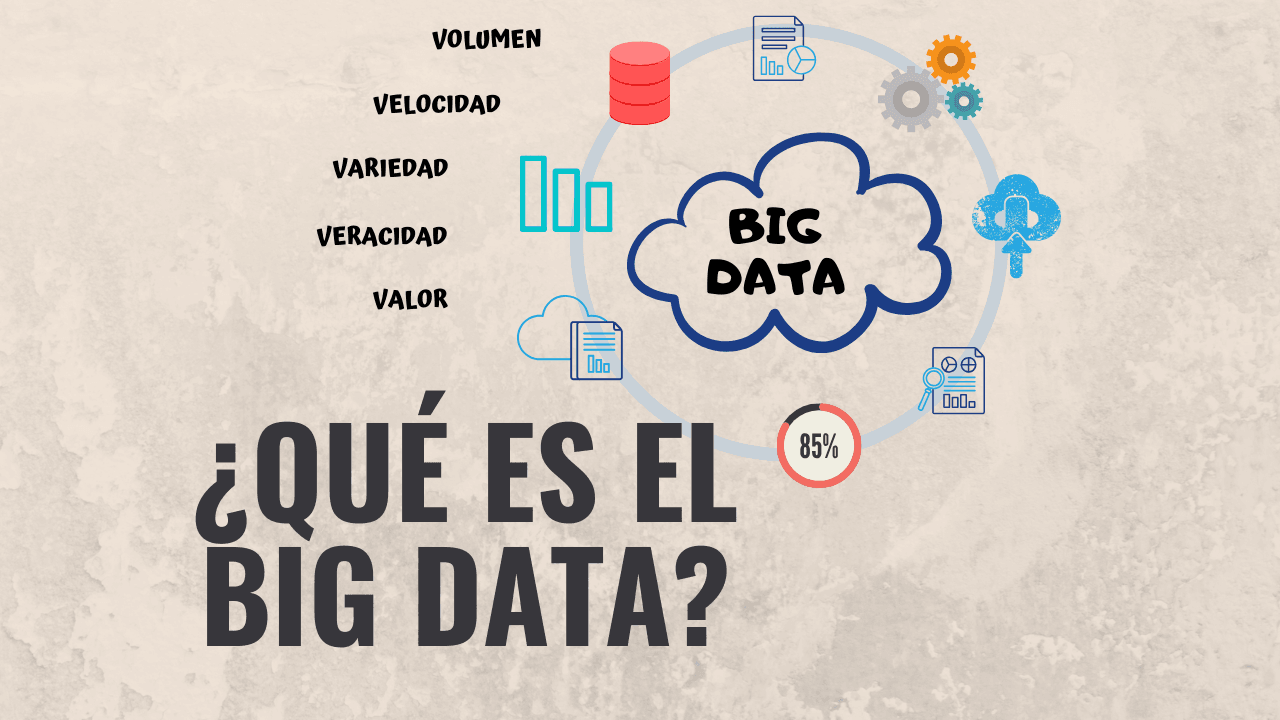
La velocidad es el ritmo al que se reciben los datos y (posiblemente) al que se utilizan.

Por lo general, la mayor velocidad de los datos se transmite directamente a la memoria, en vez de escribirse en un disco. Algunos productos inteligentes habilitados para Internet funcionan en tiempo real o prácticamente en tiempo real y requieren una evaluación y actuación en tiempo real.

* Variedad

La variedad hace referencia a los diversos tipos de datos disponibles.

Los tipos de datos convencionales eran estructurados y podían organizarse claramente en una base de datos relacional. Con el auge del big data, los datos se presentan en nuevos tipos de datos no estructurados. Los tipos de datos no estructurados y semiestructurados, como el texto, audio o video, requieren un preprocesamiento adicional para poder obtener significado y habilitar los metadatos.



Hace un tiempo que, a las tres V, se le han agregado dos más:

* Valor:

Convertir la información en conocimiento debe ser el propósito de cualquier estrategia de Big Data y esto se consigue a través de la analítica que ayuda en la toma de decisiones.

Se trata en definitiva de utilizar los datos de manera rentable y a la vez eficiente. Tener datos es fácil, lo complicado y el factor verdaderamente diferencial es el uso que se da de estos para comprender, por ejemplo, el comportamiento de los clientes o alcanzar niveles óptimos en competitividad.

* Veracidad:

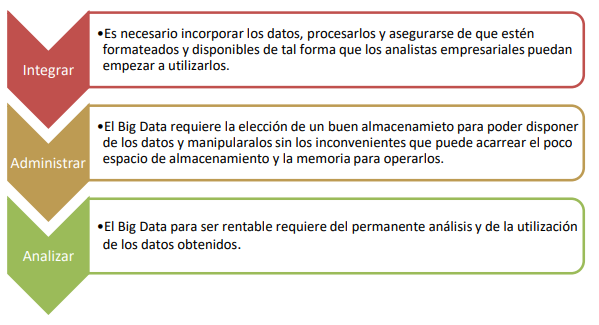
El Big Data hay que controlar necesariamente la integridad de los datos, ya que de esto dependerá del acierto de las decisiones.

Una de las dimensiones más importantes también del Big Data señala a la fiabilidad de la información recogida eliminando cualquiera inexactitud o incertidumbre en la recopilación de estos, vinculándose este punto con el tratamiento de los datos o limpieza de estos.

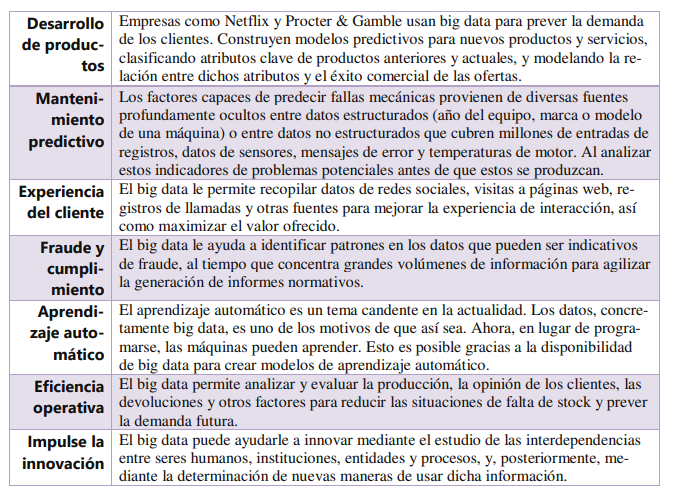
Esto supone un gran reto a medida que los datos crecen.

Cómo funciona Big Data

Para que el Big Data aporte nuevas perspectivas a cualquier negocio que lo quiera implementar, requiere de tres acciones básicas:



Algunos Ejemplos de su uso

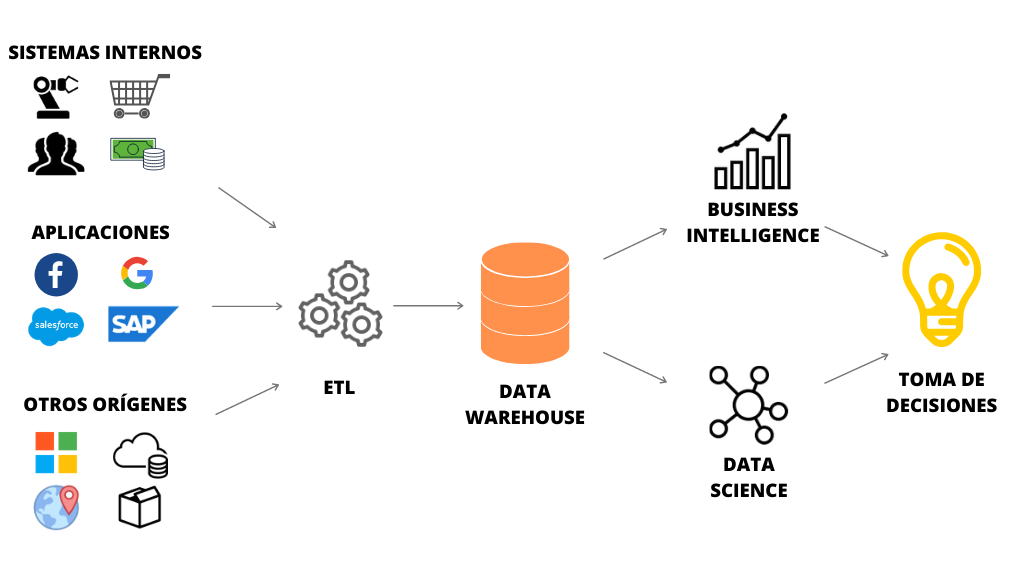


¿Qué es Data Warehouse?

Un data Warehouse (almacén de datos), es un sistema de gestión de datos, diseñado para permitir y respaldar actividades de inteligencia empresarial (BI), especialmente análisis.

Los almacenes de datos se han diseñado exclusivamente para realizar consultas y análisis y, a menudo, contienen grandes cantidades de datos históricos. Los datos de un almacén de datos se obtienen normalmente de una amplia variedad de fuentes, como archivos de registro de las aplicaciones y aplicaciones de transacciones.

Sus capacidades analíticas permiten que las organizaciones obtengan información empresarial valiosa de sus datos para mejorar la toma de decisiones.



Un almacén de datos típico está compuesto por:

* Una base de datos relacional para almacenar y gestionar los datos.
* Una solución de extracción carga y transformación (ELT) para preparar los datos para el análisis.
* Funciones de análisis estadísticos, elaboración de informes y extracción de datos.
* Herramientas de análisis de clientes para visualizar y presentar los datos a los usuarios de empresa.
* Otras aplicaciones analíticas más avanzadas que generan información procesable aplicando algoritmos de aprendizaje automático e inteligencia artificial (IA).