

INICIACIÓN A LAS BASES DE DATOS RELACIONALES CON ACCESS Y POSTGRESQL

24 horas.

6 sesiones formativas.

Días 15, 18, 21, 25, 27 y 29 de noviembre de 9.00 a 13.00 h.



Formador: Juan Pedro Pérez Alcántara.



Es el
momento de ser
mucho más
globales
humanos
productivos
actuales
creativos

¿Por qué son importantes las bases de datos?

- Son uno de los sectores de las ciencias de la computación más transversales y con más impacto global en nuestro día a día.
- Permite la creación de almacenes de datos distribuidos y multiconcurrentes. Propicia la computación distribuida.
- Enorme flexibilidad ante el planteamiento de problemas complejos. Mímesis de los procesos lógicos que ocurren en la realidad gracias a la modelización de sistemas reales.
- Últimas depositarias de los datos producto de procesos de recogida que pueden ser extraordinariamente costosos.

¿Por qué son importantes las bases de datos?

- Se encuentran en la base de la gran mayoría de las aplicaciones informáticas presentes en la administración y en todo tipo de organizaciones, independientemente de su tamaño.
- Guardan lo máspreciado: los datos. Las aplicaciones van y vienen, pero los datos son para siempre.

En conclusión: nadie en tareas gestivas, científicas, técnicas o administrativas puede ser ajeno a ellas.

Bases de datos y sistemas gestores de bases de datos

- Sistema gestor de bases de datos (SGBD)
 - Software especializado en la creación, gestión, administración, utilización y mantenimiento de bases de datos. Ejemplos: PostgreSQL, Oracle, MySQL, OpenOffice Base, SQL Server, Access, etc.
- Bases de datos
 - Estructuras de datos contenidas y manipuladas por un sistema gestor de bases de datos.
- Estructura de datos
 - Es un modo particular de almacenar y organizar los datos para que puedan ser utilizados con eficacia.

Propósito de las bases de datos

- Albergar la información del problema a tratar en una forma conveniente para su tratamiento. Esta forma es un modelo entidad-relación que reproduce los flujos y la estructura de la información del problema real.
- Controlar la introducción de datos en la base, de forma que la información contenida en ella sea consistente, atómica y no redundante:
 - consistente: la información almacenada no puede ser contradictoria entre sí;
 - atómica: la información debe estar en el nivel de desagregación más conveniente;
 - no redundante: un dato en concreto debe estar recogido en la base de datos una sola vez, y no varias, so pena de incurrir en inconsistencias.

Propósito de las bases de datos

- Permitir el análisis y recuperación de datos en condiciones adecuadas y con flexibilidad.

En conclusión: utilizamos bases de datos para poner orden allí donde no lo hay y propiciar de esta manera la resolución de preguntas y análisis sobre los datos que se recogen en nuestra actividad científico-técnica o de gestión diaria.

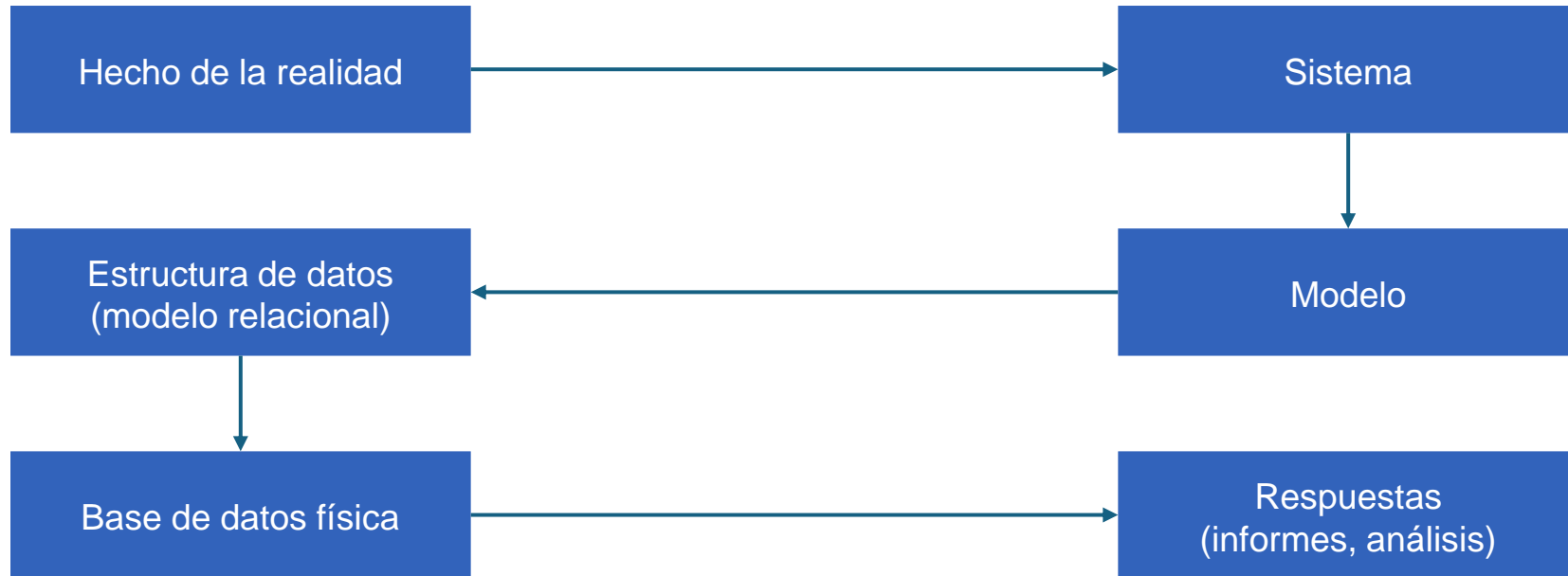
Propósito de las bases de datos

- Plantear la infraestructura software necesaria para apoyar todas las necesidades de uso y gestión de las bases de datos que contiene.
- Tareas de definición de datos. Creación de las estructuras de datos (modelos) que albergarán la información. Modelado visual. Lenguaje SQL / DDL (Data Definition Language).
- Tareas administrativas. Gestión de accesos, permisos, usuarios, cuotas, copias de seguridad, exportaciones, importaciones, encriptación, etc. Lenguaje SQL / DBA (Database Administration).
- Tareas de edición de datos. Inserción, modificación y borrado de datos. Control de concurrencia por transacciones. Lenguaje SQL / DML (Data Manipulation Language).

Propósito de las bases de datos

- Tareas de recuperación y consulta de datos. Creación de consultas sobre los datos. Lenguaje SQL / DML. Ordenación (índices) de los datos y recuperación selectiva de los mismos.
- Tareas programáticas avanzadas. Creación de programas complejos dentro de la base de datos para la manipulación avanzada de datos. Lenguajes como PL/SQL, PL/pgSQL, SQL/PSM, T-SQL, Java, etc.

Génesis de una base de datos



Sistemas

Un sistema es una realidad:

- Compuesta por partes organizadas, denominadas componentes;
- Los componentes se relacionan e interactúan entre sí;
- Ni las propiedades ni el funcionamiento del conjunto del sistema pueden deducirse por completo de las propiedades aisladas de alguna de sus partes;
- Los sistemas intercambian energía, información y/o materia con su entorno.

Los sistemas racionalizan y organizan una realidad compleja, ayudándonos a comprenderla y con vocación de control y predicción de su deriva futura.

Sistemas

La realidad de los procesos naturales y humanos que nos rodea puede ser comprendida de una forma sistémica:

- el clima global y los climas locales funcionan como un sistema;
- la actividad económica funciona según sistemas económicos;
- la gestión de la logística de una cadena de supermercados funciona como un sistema;
- el sistema tributario de un país sirve para recaudar impuestos para su funcionamiento;
- un reglamento administrativo también es un sistema basado en la tramitación y validación de un proceso de base legal.

Es propio de la actividad científica el identificar y explicar los componentes y el funcionamiento de estos sistemas.

Es propio de la actividad técnico / científica el controlar el estado de los sistemas y predecir su deriva.

Modelos

Un modelo:

- es una representación simbólica y esquemática de un sistema;
- se construye para reproducir, controlar y predecir su estado **de acuerdo a una tarea en concreto;**
- sólo se tiene en cuenta aquellos elementos y relaciones del sistema a modelar que son realmente relevantes para la tarea en cuestión (se desprecian los "efectos mariposa").

De esta manera, se puede decir que un modelo es una representación simplificada, parcial y conveniente de un sistema que puede ser mucho más complejo.

Control de la información y de los datos

Para comprender el estado actual, pasado y futuro (predicción) de un sistema:

- se necesita contar con datos sobre el estado de cada uno de sus componentes y las relaciones que se establecen entre ellos;
- debe contarse con datos sobre los flujos de energía, materia y/o información que intercambian los componentes del sistema;
- gracias a los modelos matemáticos y/o a la experiencia adquirida, se puede procesar estos datos para intentar predecir la evolución futura del sistema en su conjunto y/o tomar decisiones estratégicas.

En este contexto, y teniendo en cuenta a los procesos computacionales como el método idóneo para el tratamiento masivo de datos, es necesario aprender a concebir, preparar, implementar y explotar la información que describe un sistema en un entorno computacional.

Base de DATOS (Concepto de dato)

Unidad mínima de información que:

- está organizada especialmente para su análisis o para razonar y tomar decisiones;
- está usualmente formateada de un modo especial.

Los datos pueden existir en una variedad de formas, por ejemplo:

- como números o texto en papel;
- como bits y bytes almacenadas en memorias electrónicas;
- como hechos almacenados en la mente de una persona.

En ciencias de la computación, un dato es cualquier cosa en una forma que un ordenador puede utilizar.

Datos: organizados para su análisis y propiciar la toma de decisiones

El analista debe seleccionar el grupo de variables necesarias y suficientes para caracterizar el sistema que desea analizar.

Por ejemplo:

- para crear un listado de personas destinado al cobro de impuestos, Hacienda debería incluir, entre otros datos, el nombre, los apellidos, la fecha de nacimiento, la profesión, el domicilio, etc. de dichas personas;
- para estudiar y caracterizar el clima de una región, los climatólogos deberían recoger en sus estaciones datos sobre presión atmosférica, precipitación, humedad, insolación, temperatura, etc.

Datos: unidades mínimas de información

Desagregación: los datos son desagregables a varios niveles:

Población total de un país

└─> Suma de la población de sus provincias

└─> Suma de la población de los núcleos de población y
población rural diseminada de cada provincia

Composición: pueden ser datos compuestos según un FORMATO ESPECIAL que tenga una semántica implícita:

41001002301 ──────────> Provincia (2 dígitos) - Municipio (3 dígitos) - Orden (4
dígitos) - Entidad singular (2 dígitos)

Derivación: pueden ser datos derivados de un proceso aritmético o lógico aplicado a datos básicos:

$$Densidad = \frac{Masa}{Volumen}$$

Dato: un dato es cualquier cosa en una forma que un ordenador pueda utilizar

Aplicable a los datos que se pretenden analizar y tratar en un sistema computacional, ya que debemos tener en cuenta que:

Las máquinas carecen de capacidad de abstracción, sólo tienen datos

Hay que atomizar, ordenar, estructurar y desagregar los datos lo necesario como para que el computador pueda analizarlos

Los computadores son máquinas discretas, carecen del sentido de continuidad de los seres humanos

Los números que maneja la máquina siempre están sujetos a una precisión finita

Las máquinas manejan estupendamente números, pero son muy malas con los textos complejos

Tenemos que convertir, de alguna manera, las palabras en números

Hechos —————→ *Consecuencias*

**Dato: un dato es cualquier cosa en una forma
que un ordenador pueda utilizar**

Las máquinas carecen de capacidad de abstracción

Para nosotros, es trivial que “Miguel de Cervantes”, “Cervantes” o “Cervantes, Miguel de” son la misma cosa, para la máquina no. Si queremos que la máquina reconozca siempre a Miguel de Cervantes, hay que llegar a un consenso de cual es el nombre a utilizar de esta persona y usarlo en todos los casos.

Las máquinas son máquinas discretas

Para nosotros, es “comprensible” entender que entre el 0 y el 1 hay infinitos números. Para la máquina, con una memoria finita, llega un momento que no se puede diferenciar un número infinitesimalmente pequeño de otro.

Las máquinas manejan bien números, pero no palabras

Para nosotros, es trivial que las palabras “Alcalá de Henares”, “Alcalá de Enares” y “Allcalá de Henares” hacen referencia a la misma cosa, pero que las dos últimas contienen un error ortográfico. Para la máquina, estas tres cadenas de letras son diferentes.

Dato: un dato es cualquier cosa en una forma que un ordenador pueda utilizar

Ejemplo de datos que un ordenador NO puede utilizar con eficacia

Imaginemos una encuesta para turistas, que es recogida por encuestadores sobre una plantilla predefinida en una hoja de papel:

Encuesta 1 (encuestador A)

Nacionalidad

Española

¿Qué destacarías de Córdoba?

El salmorejo, el tiempo, la mezquita, las cordobesas.

¿Qué crees que hay que mejorar?

Profesionalidad del servicio, pocos conciertos.

Encuesta 2 (encuestador B)

Nacionalidad

España

¿Qué destacarías de Córdoba?

La comida, el sol, monumentos, la gente.

¿Qué crees que hay que mejorar?

El personal del hotel, oferta cultural.

**Dato: un dato es cualquier cosa en una forma
que un ordenador pueda utilizar**

El ejemplo de antes, mejor planteado

País de procedencia: _____

A destacar (marque tantas X como sea necesario):

- | | | |
|--|--|---|
| <input type="checkbox"/> Gastronomía | <input type="checkbox"/> Trato con residentes | <input type="checkbox"/> Monumento (Mezquita) |
| <input type="checkbox"/> Tiempo | <input type="checkbox"/> Monumento (Torre Calahorra) | <input type="checkbox"/> Monumento (Alcázar) |
| <input type="checkbox"/> Monumento (otros) | | |

A mejorar (marque tantas X como sea necesario):

- | | | |
|--|---|--|
| <input type="checkbox"/> Profesionalidad en restauración | <input type="checkbox"/> Cultura (conciertos) | <input type="checkbox"/> Información turística |
| <input type="checkbox"/> Profesionalidad en hostelería | <input type="checkbox"/> Cultura (museos) | <input type="checkbox"/> Transportes |

Datos: conclusión

Es tarea del analista:

- determinar cuales son las variables relevantes para describir la realidad que pretende analizar;
- determinar cual es el nivel de desagregación relevante para el problema que se tiene entre manos;
- determinar el formato en el que dichos datos deben ser codificados, para que la codificación de los mismos sea segura, sencilla y eficaz.

BASE DE DATOS

- "Una base de datos es la organización de una colección de datos que se interrelacionan, se comparten y se controlan" (**G. W. y J. V. Hansen**, "Diseño y administración de bases de datos", 1997)
- "Una base de datos es una colección de datos organizada de forma sistemática, de manera que un programa de ordenador puede consultarlos para ofrecer respuestas" (**Wikipedia**, 2006)
- De las definiciones anteriores extraemos que una base de datos debe...
 - albergar datos según una estructura conveniente diseñada de antemano que propicie la interrelación de hechos para producir información;
 - tener control sobre la naturaleza y formato de los datos que se incorporan a ella, de manera que se garantice la calidad de los mismos;
 - controlar el acceso de diferentes perfiles de usuario a sus datos y estructura.

Base de datos: mecanismos de control sobre la naturaleza de los mismos

- **Dominios:** La edad de una persona no debería ser codificada con un número negativo o un valor mayor a, por ejemplo, 110 años.
- **Integridad referencial:** Los valores posibles que puede tomar el dato de provincia de nacimiento de una persona debería ser controlado por una lista de provincias, de forma que no puedan introducirse provincias inventadas, ajenas al país o que se incluyan con faltas de ortografía o de criterio en la base de datos.
- **Tipo de datos y precisión:** Si queremos controlar la posición por GPS de una flota de camiones que se mueven por Europa, bastará con que las coordenadas UTM de cada uno de ellos se codifique con una precisión de centenas de metros, sin decimales. Sin embargo, para el levantamiento topográfico de un yacimiento arqueológico necesitaremos una precisión centimétrica, con lo que usaremos dos decimales.

Sistemas actuales de bases de datos computacionales

- **Bases de datos relacionales**

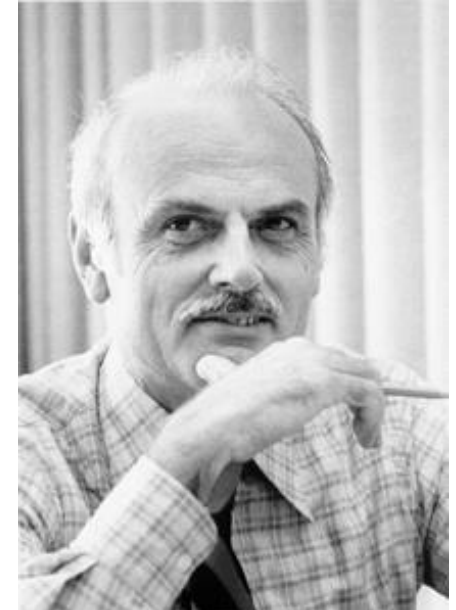
- se diseñan según la metodología entidad-relación
- muy robustas y versátiles
- énfasis en la versatilidad del modelado, muy adaptables, así como en la integridad del dato

- **Bases de datos NoSQL**

- metodologías de diseño muy diversas
- Se centran en el rendimiento Big Data
- inflexibles, estudiar bien el caso de uso

Base de datos relacionales

- Teoría desarrollada por Edgar Frank "Ted" Codd (1923 - 2003) en su publicación seminal *A Relational Model of Data for Large Shared Data Banks* (1970).
- Se basa en:
 - los datos se recogen en formato tabular, de forma que cada tabla del sistema recoge una colección de objetos homogéneos en sus propiedades y estructura. Es muy importante destacar que cada uno de estos objetos de la tabla debe ser unívocamente identificado por una o un conjunto de sus propiedades;
 - cada tabla describe los objetos que contiene en base a una colección arbitraria de propiedades, llamadas campos;
 - cada colección de objetos (es decir, cada tabla) puede establecer relaciones con otros objetos (es decir, con otras tablas) siempre que se relacionen gracias a sus propiedades (sus campos).



Base de datos relacionales: ventajas

- Elimina las redundancias de datos, reduciendo en la práctica la peligrosidad de inconsistencias en los mismos;
- gran capacidad de atomización de los datos, redundando en una gran eficacia en el almacenamiento de los mismos;
- enorme capacidad de interrelación natural (y no artificial, como en otras arquitecturas más antiguas) de los datos, propiciando de esta manera la aparición de gran cantidad de nueva información, incluso con bases de datos vecinas. Esta capacidad de interrelación de datos es imbatible;
- optimización del procesamiento de los datos;
- aproximación en lenguaje natural al tratamiento de los datos gracias al lenguaje SQL (Structured Query Language).

¡Fatal, peligroso!

ID	Ciudad	País
1	Madrid	España
2	Sevilla	Espanya
3	Valladolid	España
4	Munich	Alemania
5	Berlin	Alemania
6	Heidelberg	Alemania

Segun Codd...

ID	Ciudad	País
1	Madrid	34
2	Sevilla	34
3	Valladolid	34
4	Munich	49
5	Berlin	49
6	Heidelberg	49

Existe una relación
natural entre los datos

ID	Ciudad
34	España
49	Alemania

Sistemas de información y de gestión del conocimiento

Los **Sistemas de información** son sistemas diseñados para:

- permite almacenar datos de forma controlada;
- permite editar, eliminar y añadir datos al sistema;
- recuperación selectiva de los mismos;
- controlar de acceso a los datos.

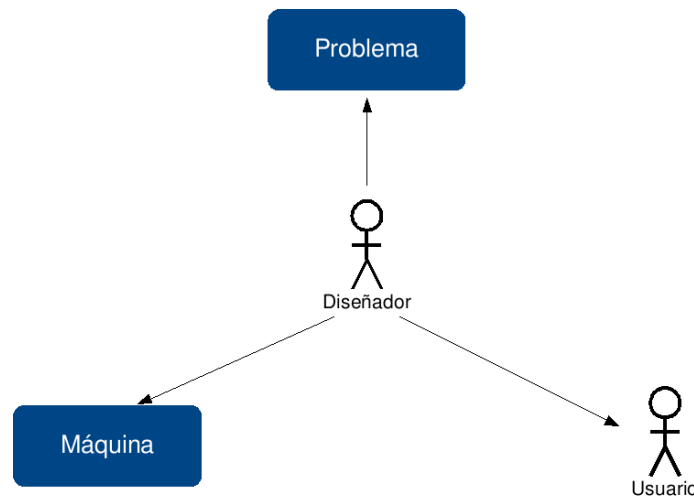
Las bases de datos son elementos principales dentro de muchos sistemas de información.

Los **sistemas de gestión del conocimiento** interrelacionan múltiples sistemas de información bajo unos estándares de estructuración de la información o presentación comunes para permitir a las personas (o incluso, en casos muy sofisticados, a las computadoras) acceder al máximo nivel de abstracción de la información: el conocimiento.

Diseño de base de datos relacionales: el equipo de diseño

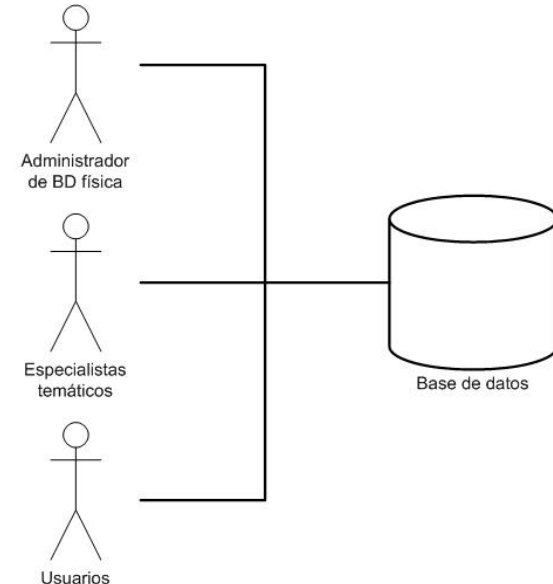
- Suele ser un equipo multidisciplinar de expertos temáticos e ingenieros de software.
- Debe entender el problema y tratar con los que lo conocen para extraer toda la información necesaria para crear el modelo.
- Debe conocer la máquina, para adaptar el problema a un lenguaje que ésta entienda.
- Debe entender las necesidades de los usuarios finales: qué esperan del sistema y cómo se supone que deben interaccionar con la información.

En resumen: debe poner los medios para expresar el problema en el lenguaje de la máquina para que sea entendido en lenguaje humano.



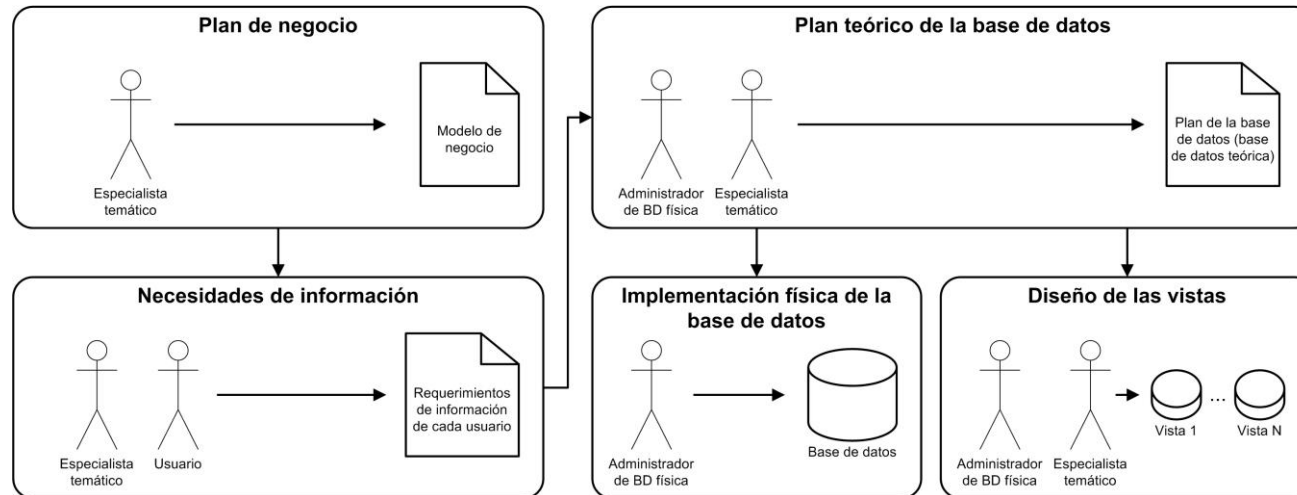
Actores en el diseño, implementación y utilización de una base de datos

- **Administradores y especialistas en la base de datos física:** administran la base de datos gracias a un sistema gestor de bases de datos.
- **Especialistas en los procesos y temática de base de datos:** son los especialistas en la realidad que se pretende modelar en la base de datos.
- **Usuarios:** usuarios finales de la base de datos encargados de la digitalización, edición, procesamiento y aprovechamiento de los datos.



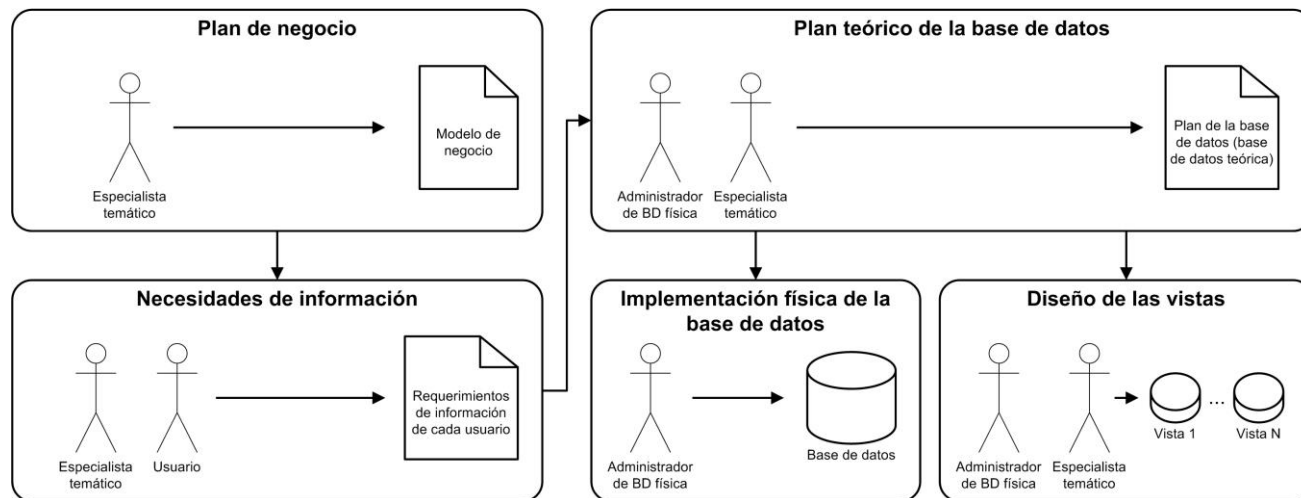
Planificación estratégica de la base de datos

- **Plan de negocio:** descripción de las funciones y objetivos de la organización, así como de los procesos operativos de la misma. Descripción de los flujos de trabajo y de información dentro y fuera de la organización. (¡Ojo a los POYAQUE!)
- **Necesidades de información:** identificación de las necesidades de información de cada departamento y área de la organización para cumplir con sus responsabilidades dentro del esquema general del plan de negocio.

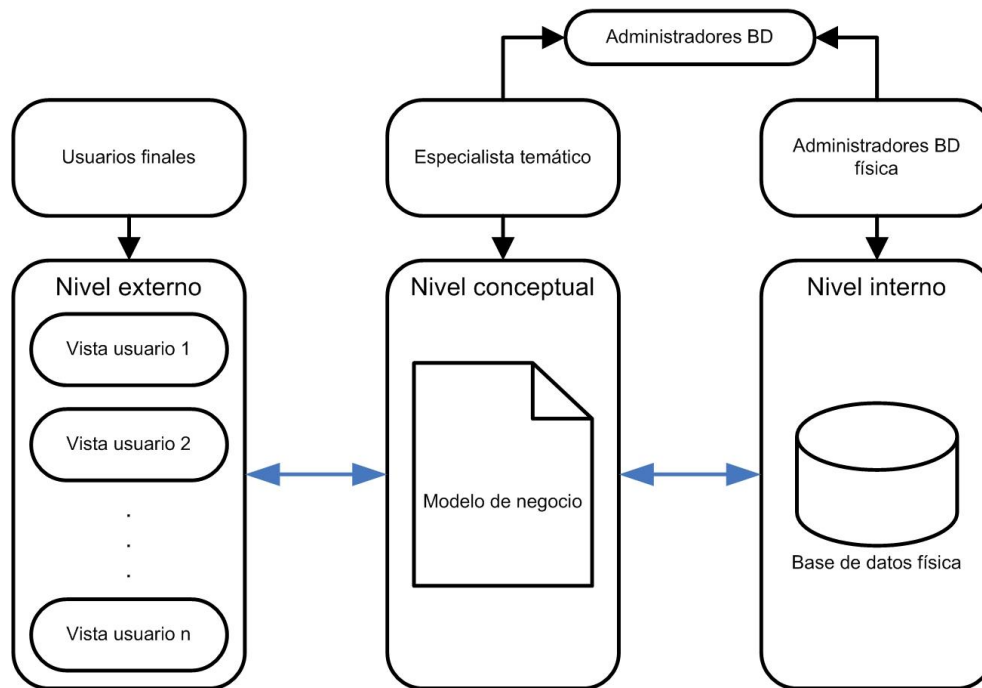


Planificación estratégica de la base de datos

- **Plan teórico de la base de datos:** planificación teórica de la base de datos. Reuniones de expertos en procesos, actividades y entidades del plan de negocio junto a especialistas en modelado de datos.
- **Implementación y diseño de consultas:** concreción de los diferentes proyectos de bases de datos dentro de la organización. Implantación física del sistema, formación del personal y diseño de las vistas del sistema para cada grupo de usuarios.



Interacciones entre los usuarios y la información



Formas de acceso a la base de datos

