

atmantree.com

El autor del presente documento lo ha publicado
bajo las condiciones que especifica la licencia



Creative Commons
Attribution-NonCommercial-ShareAlike 3.0

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

En caso de dudas escriba a:
info@atmantree.com

Bases de Datos Libres desde 40.000 pies de altura



Ing. Carlos Gustavo Ruiz
@atmantree | atmantree.com
tubasededatoslibre.org

Agenda



Taxeo y Despegue
Breve Presentación
Persistencia de Datos
Hardware
Almacenes de Datos
Gestores de Bases de Datos
Mapa de los Gestores
Bases de Datos
Cache – Llave/Valor -
Documentos – Relacionales
Otras Bases de Datos
Aproximación Final
Aterrizaje

Taxeo y Despegue



Taxeo y Despegue

Por favor apague su celular (o colóquelo en silencio)

Por regulaciones de la aeronáutica civil esta prohibido fumar en cualquier etapa del vuelo

A 40.000 pies de altura tendremos una vista panorámica del mundo de las bases de datos, si desea tomar fotos puede hacerlo.

Sin embargo daremos una dirección para que puedan descargar esta presentación y sus ejemplos

Breve Presentación

Nombre:

Carlos Gustavo Ruiz
(arahat)

Roles:

Consultor,
Programador,
Activista,
Voluntario

Internet:

<http://atmantree.com>
@atmantree

Comunidades:

TuBaseDeDatosLibre.org
Python Venezuela

Presentador sobre:

Bases de Datos
Geomática
Programación

Otros intereses:

Lectura
Escritura
Fotografía Amateur
Guitarra

Persistencia de Datos

- Desde la antigüedad se sabe que la memoria es frágil y es difícil recordar cosas cada vez más complejas
- Aunque hay sus excepciones:

“Entrenando catorce horas al día, Jaime García (un colombiano que vive en Brunete) ha recitado 150.000 dígitos de π en la Facultad de Matemáticas de la Universidad Complutense de Madrid, batiendo el anterior récord del mundo (100.000 dígitos) de Akira Haraguchi estuvo rodeado de cierta polémica y ni siquiera fue considerado válido”

Microsiervos (<http://is.gd/VH5YLn>)

Para ver otros records visite:

<http://www.recordholders.org/en/list/memory.html>



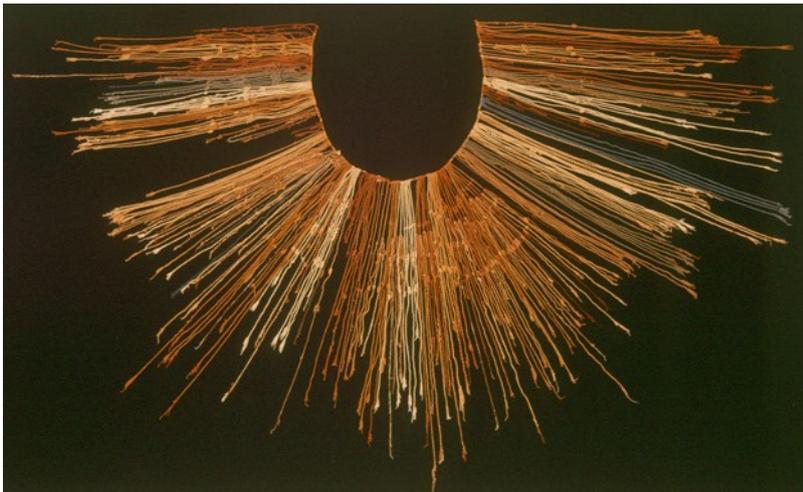
Persistencia de Datos

- 0 excepciones de excepciones:

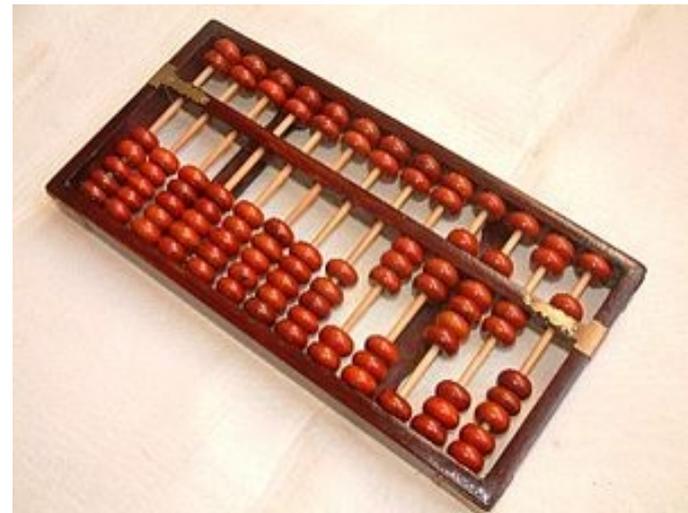


Persistencia de Datos

- Es por ello que desde tiempos antiguos se inventaron mecanismos para mantener registros de la información



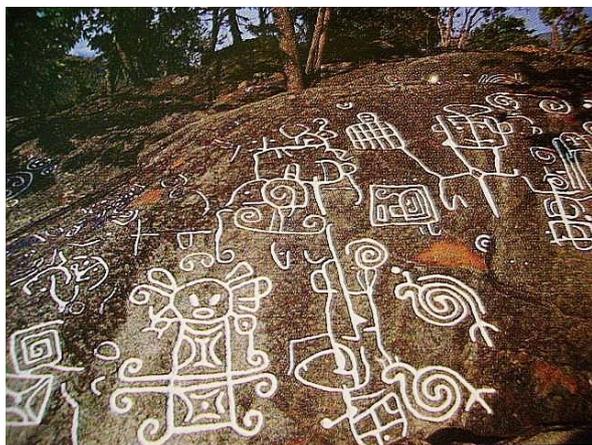
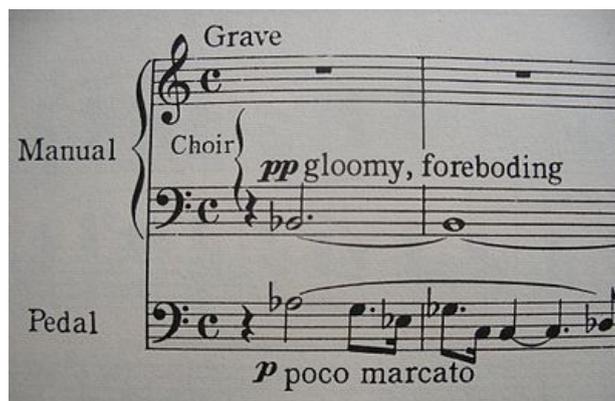
<http://es.wikipedia.org/wiki/Quipu>



<http://es.wikipedia.org/wiki/%C3%81baco>

Persistencia de Datos

Información de todo tipo



Persistencia de Datos

Sin embargo, desde el punto de vista de la informática y la computación es importante poder contar con los valores almacenados en un grupo de variables luego de finalizada la ejecución de un programa.



Persistencia de Datos

A la capacidad de mantener un dato luego de finalizada la ejecución de un programa (o de un determinado módulo del mismo) se le conoce como:

Persistencia de Datos

A la capacidad de mantener un dato luego de finalizada la ejecución de un programa (o de un determinado módulo del mismo) se le conoce como:

**Persistencia de
Datos**

Persistencia de Datos

La persistencia es importante en algún modo en todas las áreas de nuestro quehacer, es especial en tareas críticas o de alto riesgo, por ejemplo:

En el mes de Julio de 1969 un equipo con un procesador de 0,043 Mhz de velocidad y 64 Kbytes de memoria guió el hombre a la luna.

Todo lo que necesitaba para esta tarea estaba en esos 64 Kb

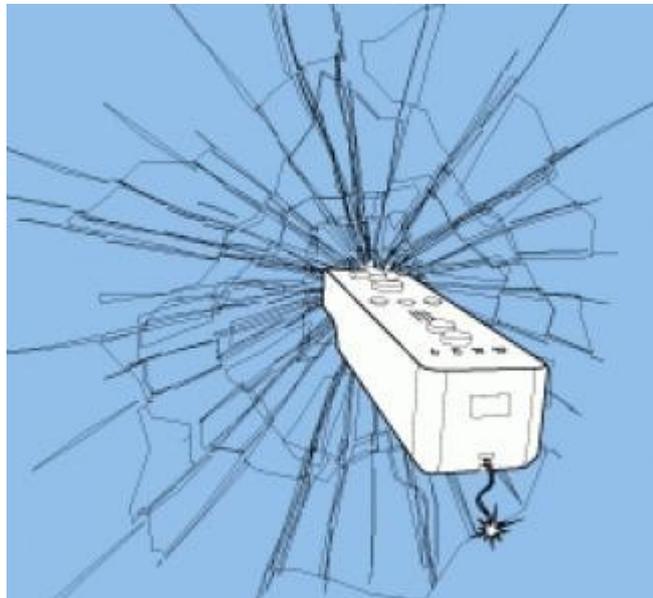
<http://is.gd/uMbToN>

Si le gusta las teorías sobre la misión Apolo 11 visite: <http://is.gd/bmxGp1>



Hardware

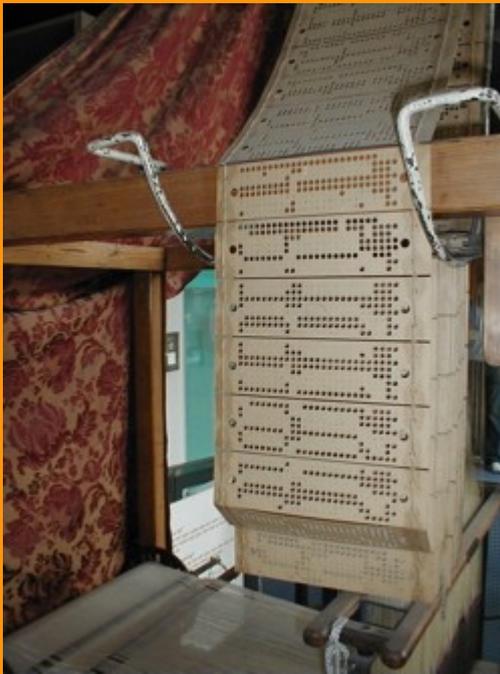
Por un lado la persistencia de datos puede ser vista como la evolución del hardware disponible. Así que veamos por un momento como han cambiado los dispositivos de almacenamiento de datos.



Hardware es todo lo que puedes golpear

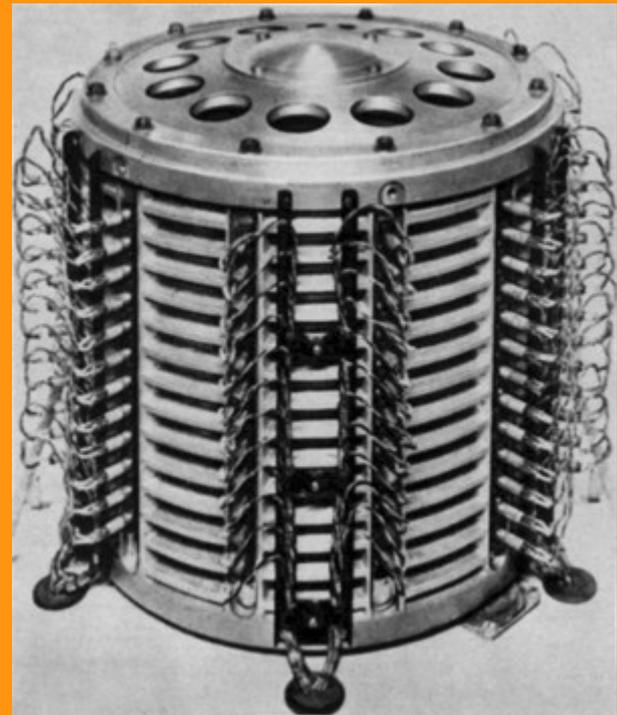
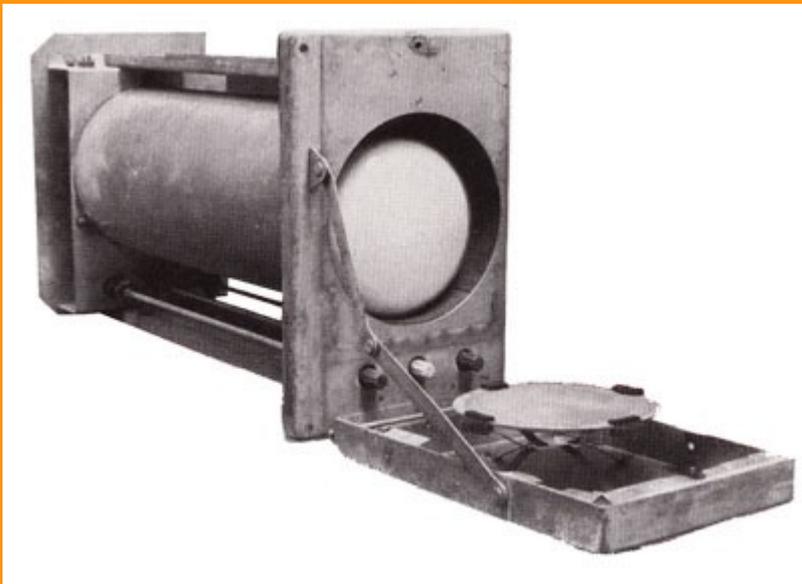
Hardware

Tarjetas Perforadas (1725-1880)



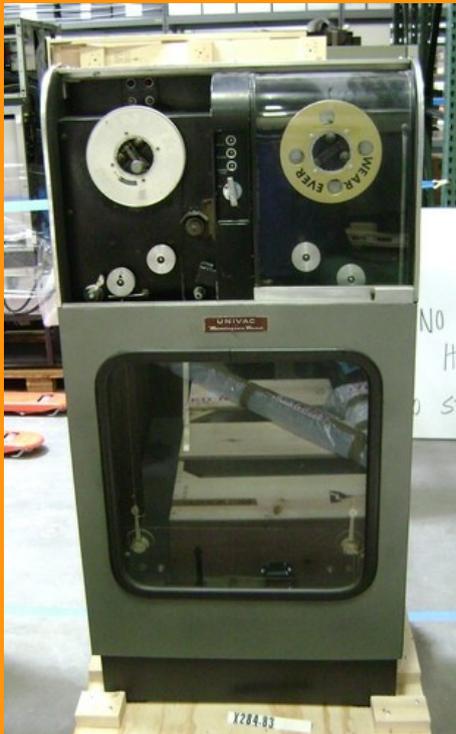
Hardware

Tubo de Williams (1945) y Memoria de Tambor (1950)



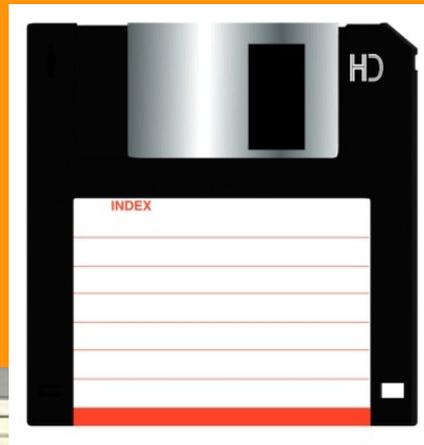
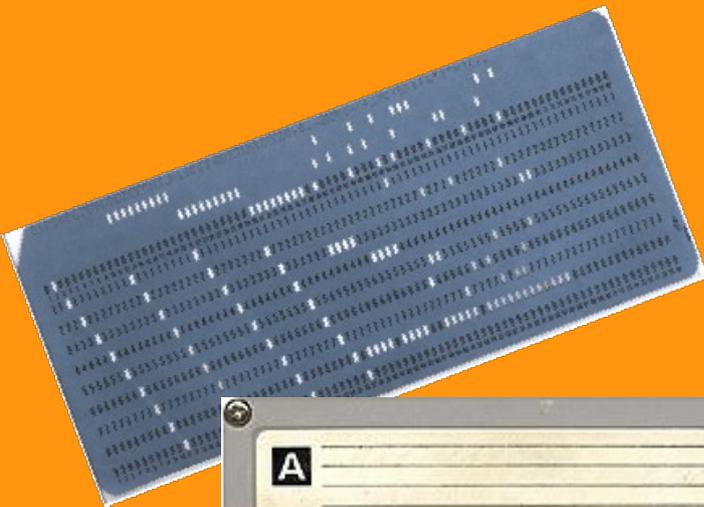
Hardware

Cintas Magnéticas (1951) y Discos Duros (1956)



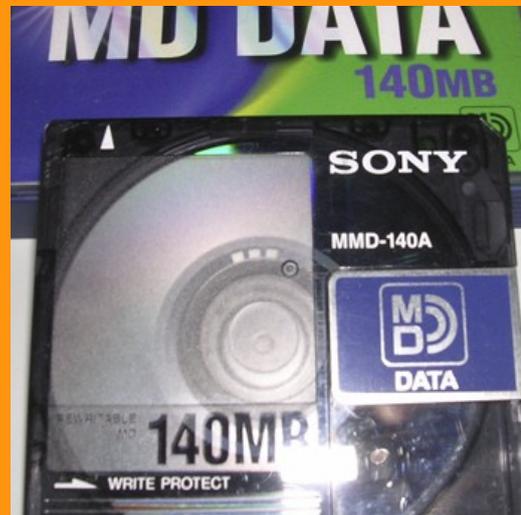
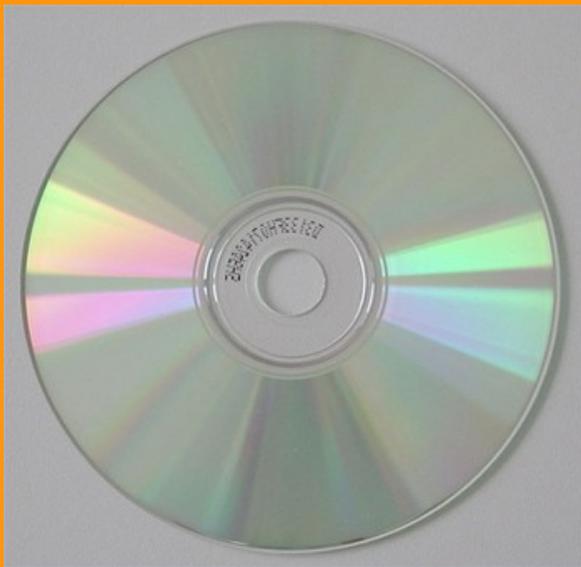
Hardware

Etapa de la Miniaturización (1972 - 1987)



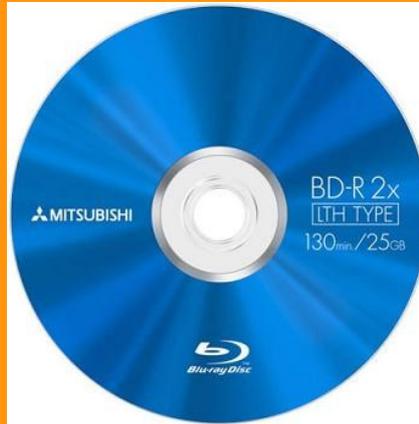
Hardware

La Etapa del Laser (1990 - 2008)



Hardware

La Nueva Etapa del Almacenamiento



Anuncio

Damas y Caballeros anunciamos que hemos culminado la etapa de despegue exitosamente.

Ya puede soltarse el cinturón, sin embargo trate de mantenerse en su asiento.

Nos encontramos sobrevolando sobre los
“Almacenes de datos”

Almacén de Datos



Almacenes de Datos

Los almacenes, son colecciones de datos orientados a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones.

Los más conocidos son los corporativos conocidos como *data warehouse*.

Almacenes de Datos

Los almacenes, son colecciones de datos orientados a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones.

Los más conocidos son los corporativos conocidos como *data warehouse*.

Almacenes de datos

Sin embargo podemos iniciar un almacén de datos a partir de soluciones mucho más simples como archivos de texto plano, archivos con valores separados por (comas, tabulador o caracteres especiales), o con herramientas como las suites ofimáticas.

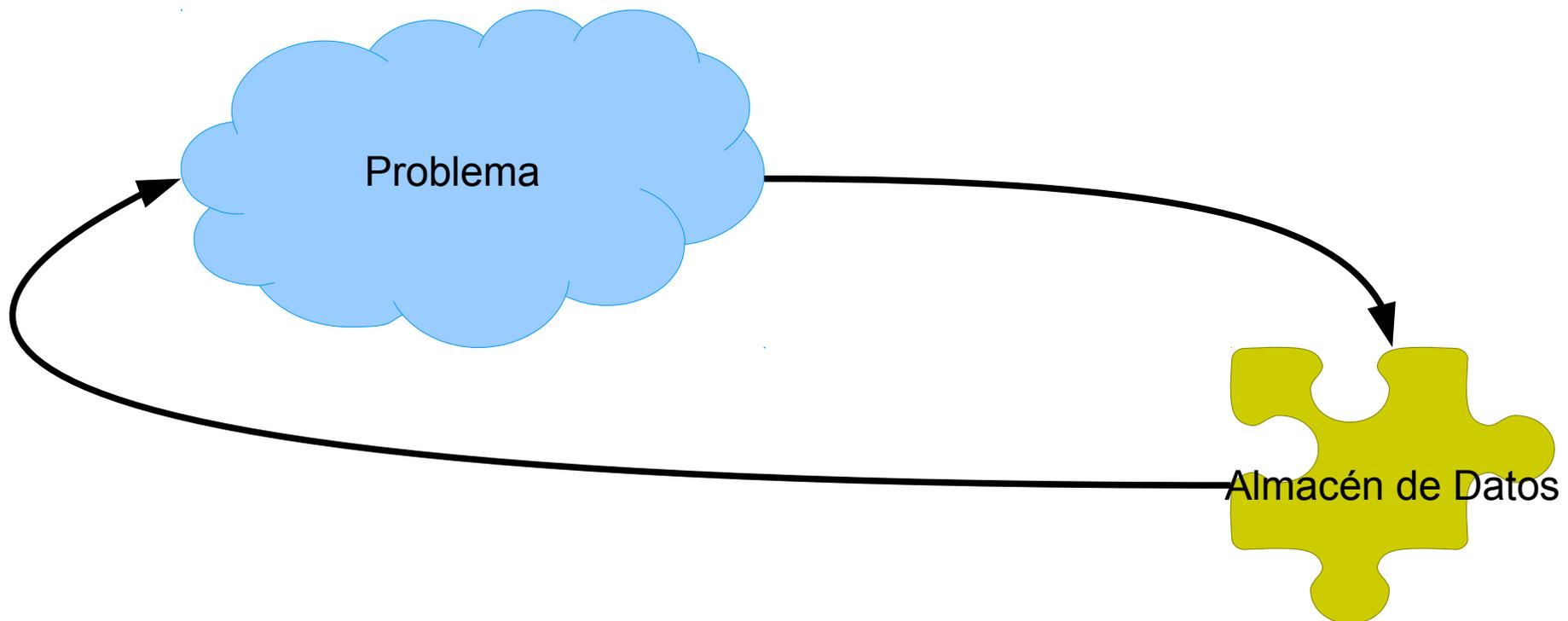


Almacenes de Datos

Veamos un ejemplo

Gestores de Bases de Datos

El problema con los almacenes de datos es que siempre estamos muy cerca de la implementación de la solución. No hay capacidad de abstraerse al dominio del problema.



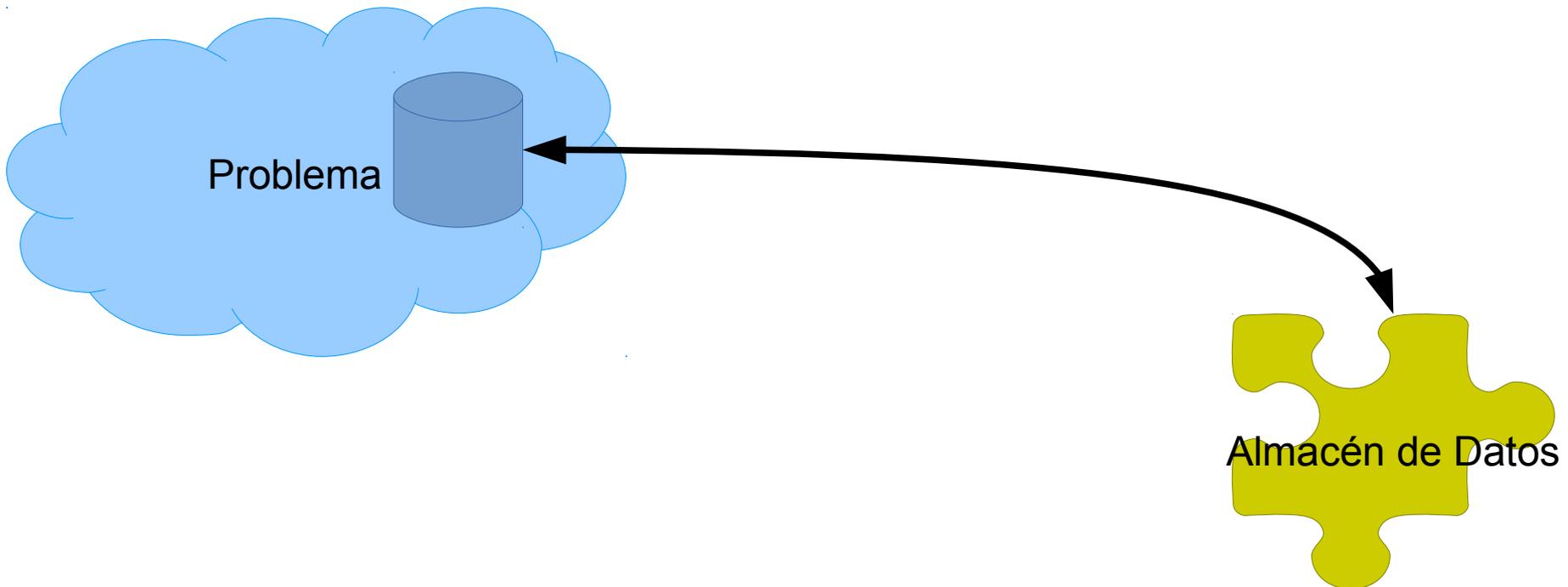
Gestores de Bases de Datos

Los gestores de bases de datos son una capa de abstracción que permite que nos concentremos en el problema y bajo ciertas condiciones confiemos a ellos el guardado y recuperación de la información.

En otras palabras los gestores de bases de datos nos permiten delegar la implementación de los almacenes a un programa.

Gestores de Bases de Datos

Esquemáticamente podemos decir en que nuestros datos hablan el mismo idioma que nuestro problema. Por lo tanto siempre conversamos al mismo nivel y en el mismo lenguaje.



Gestores de Bases de Datos

Hoy en día a estas condiciones en las cuales confiamos al usar un gestor de base de datos se les conoce por los nombres ACID y BASE. (en inglés ácidos y bases)



Gestores de Bases de Datos

- ACID – Acrónimo de los términos:

- Atomicity
- Consistency
- Isolation
- Durability



- Descrito en la norma ISO/IEC 10026-1: 1992 sección 4.

Gestores de Bases de Datos

Atomicity (atomicidad)

asegura que la operación se ha realizado o no, y por lo tanto ante un fallo del sistema no puede quedar a medias.

Consistency (Integridad)

asegura que sólo se empieza aquello que se puede terminar. No permite romper reglas de consistencia de la BD

Isolation (Aislamiento)

asegura que una operación no puede afectar a otras. Las transacciones son independientes.

Durability (Durabilidad)

asegura que una vez realizada la operación, ésta persistirá y no se podrá deshacer aunque falle el sistema

Gestores de Bases de Datos

Entonces surge el problema del Teorema de CAP de Sistemas Distribuidos:

Consistency

Availability

Partition Tolerance

Gestores de Bases de Datos

Entonces surge el problema del Teorema de CAP de Sistemas Distribuidos:

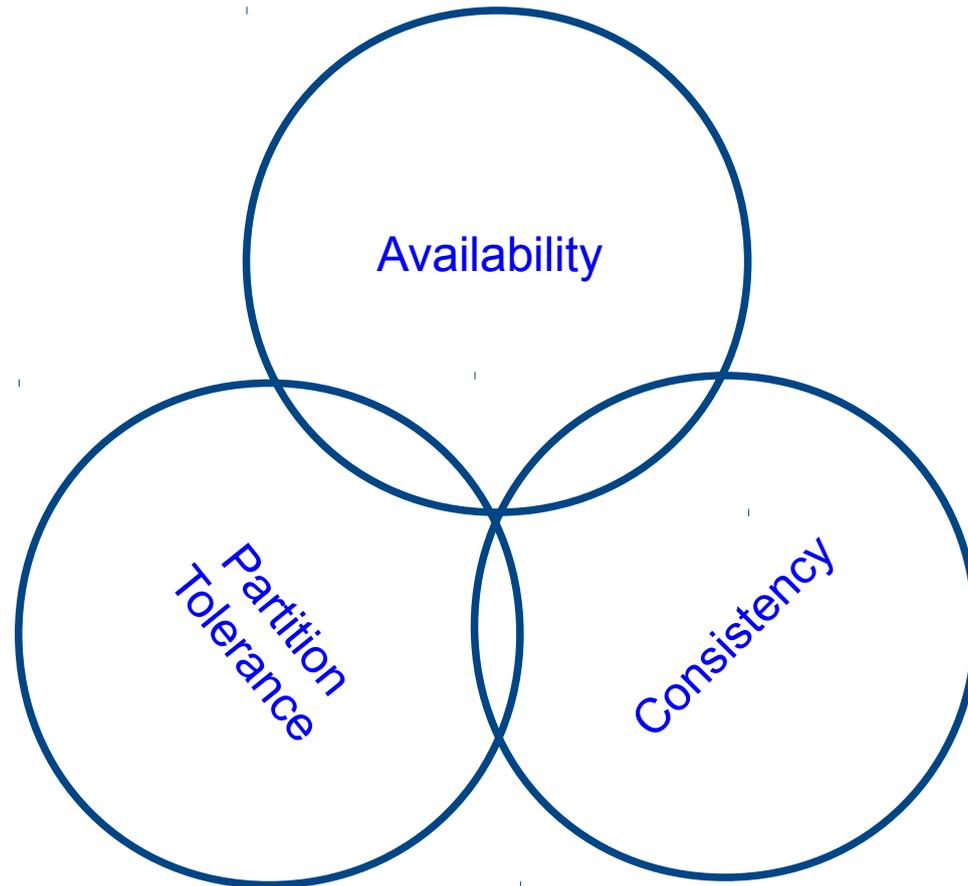
Consistencia

Disponibilidad

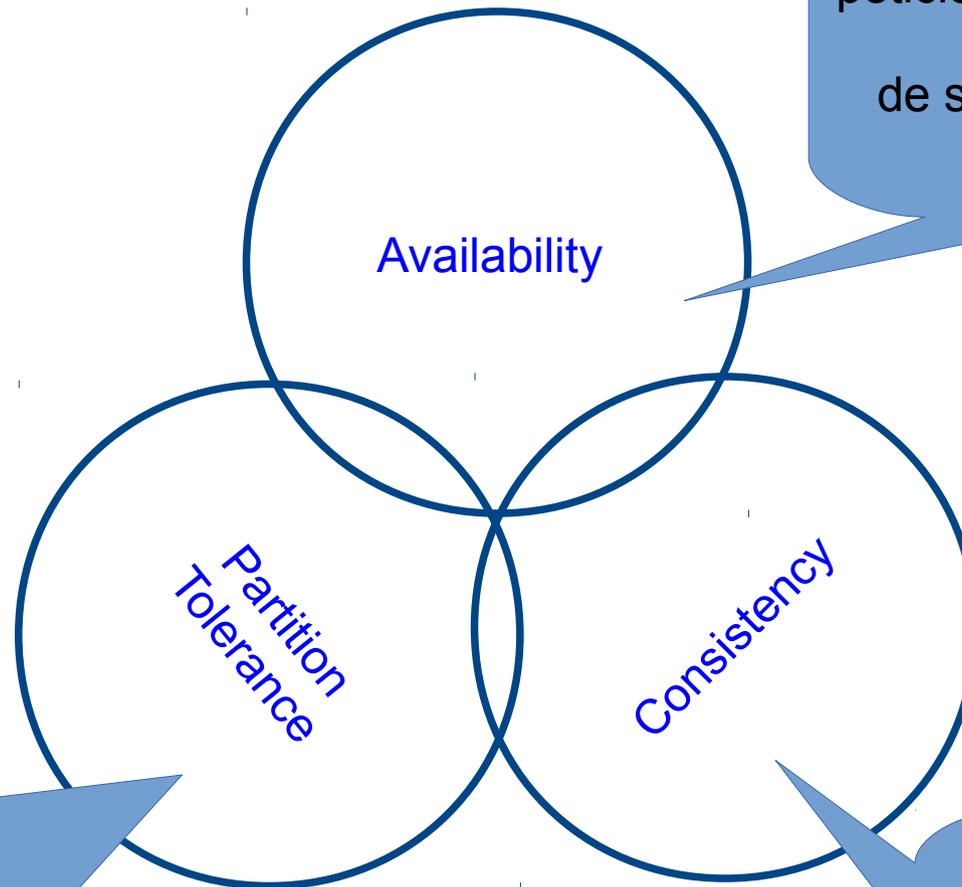
Tolerancia a Particionamiento

Solo pueden coexistir dos de ellas.

Gestores de Bases de Datos



Gestores de Bases de Datos

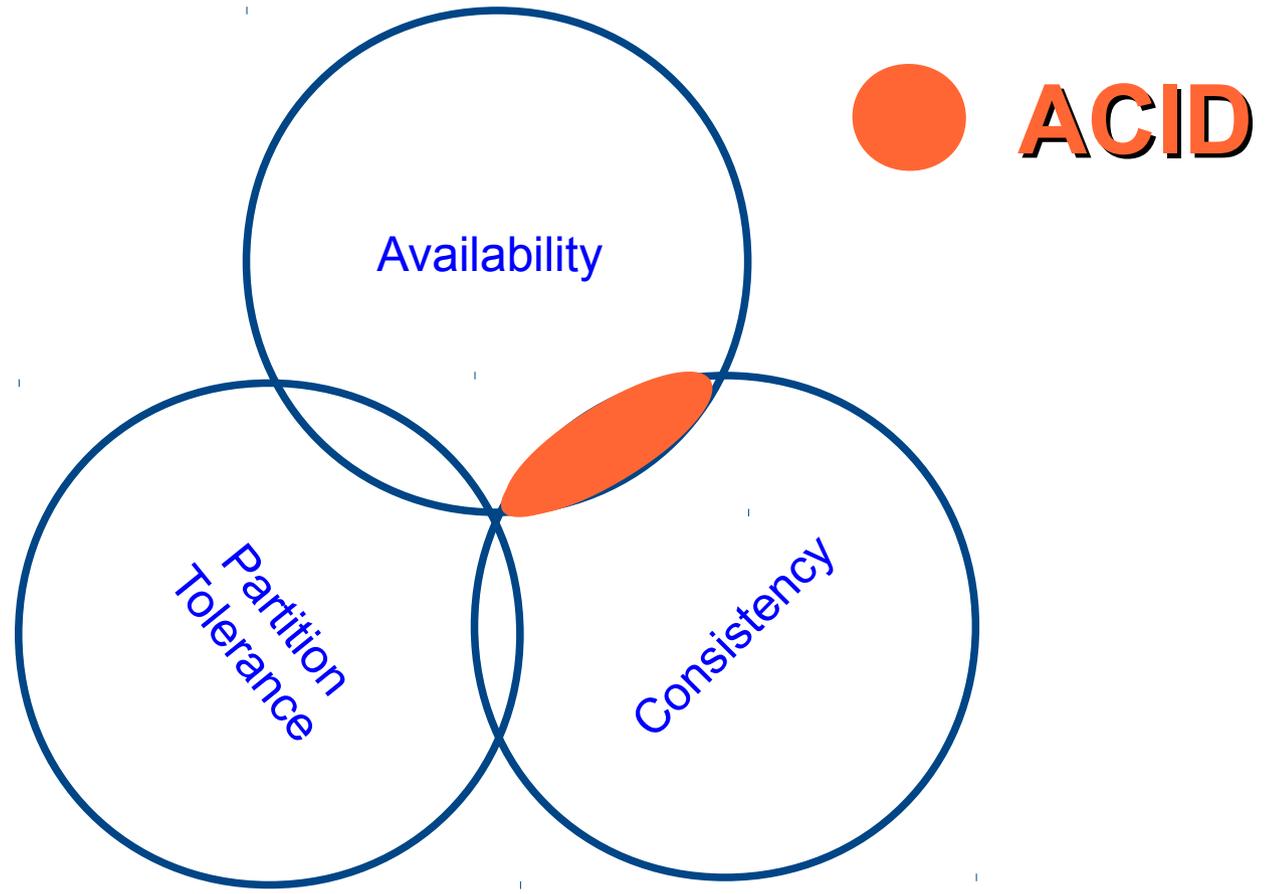


la garantía de que cada petición a un nodo reciba una confirmación de si ha sido o no resuelta satisfactoriamente

que el sistema siga funcionando a pesar de algunas pérdidas arbitrarias de información o fallos parciales del sistema a menos que todos los nodos fallen el sistema puede seguir funcionando

todos los nodos vean la misma información al mismo tiempo

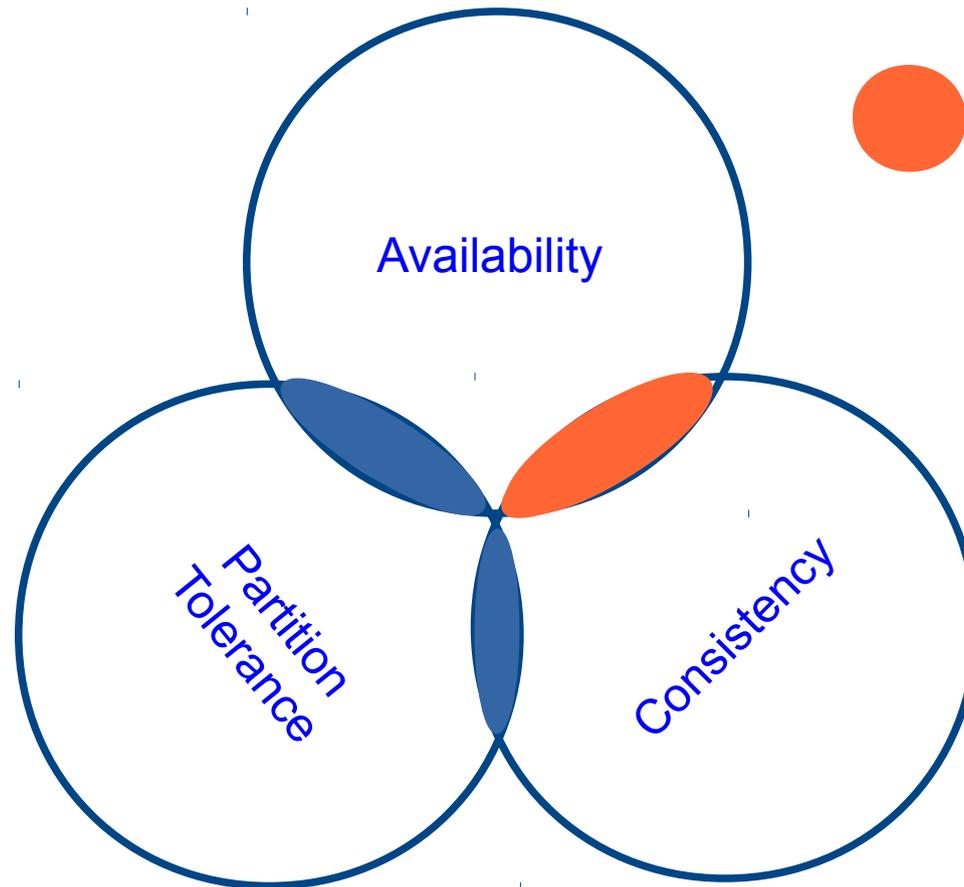
Gestores de Bases de Datos



Gestores de Bases de Datos

● **BASE**

● **ACID**



Gestores de Bases de Datos

- BASE – Acrónimo de términos
 - Basically Available
 - Soft state
 - Eventual consistency
- BASE se justifica con un a sola palabra
 - Escalabilidad



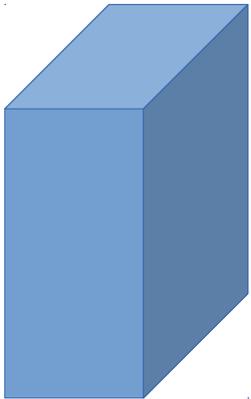
Gestores de Bases de Datos

BASE utiliza un enfoque optimista permitiendo que la consistencia se entienda como un estado de cambio continuo. Este enfoque es ideal para manejo de gran volumen de datos.



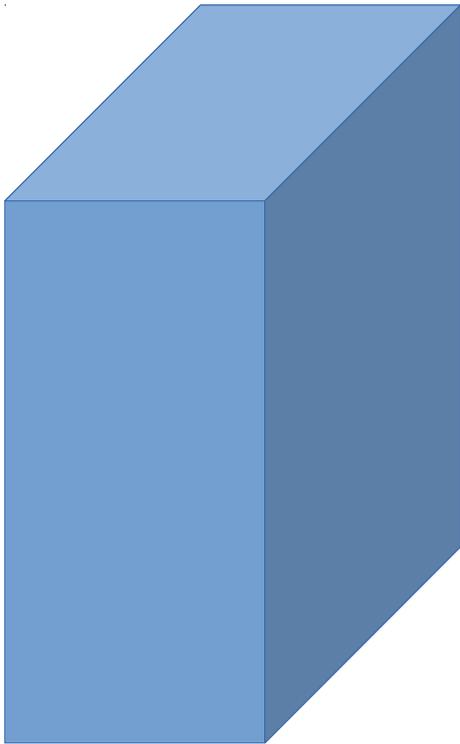
Gestores de Base de Datos

Escalar una Base de Datos



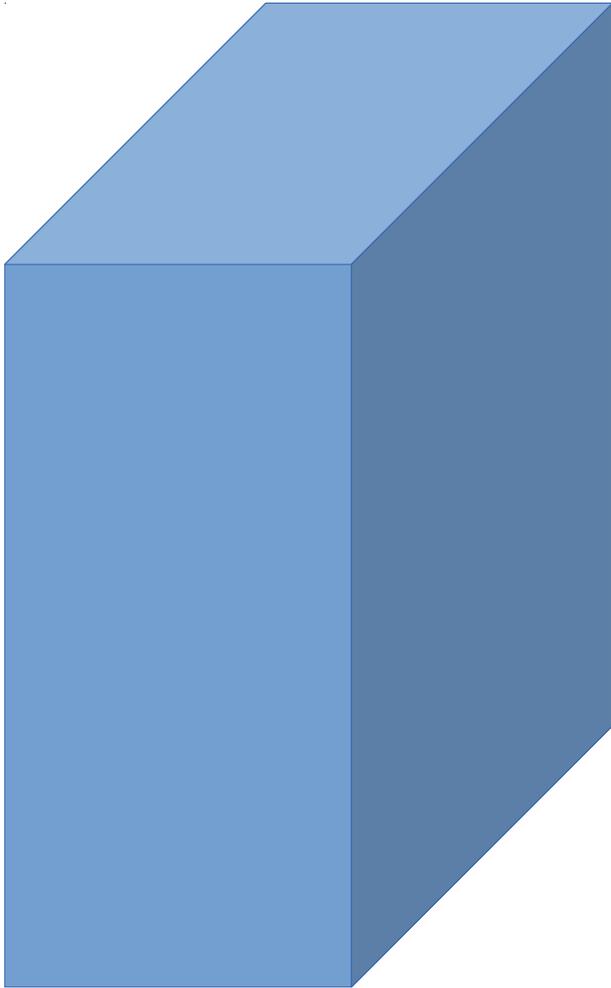
Gestores de Base de Datos

Escalar una Base de Datos



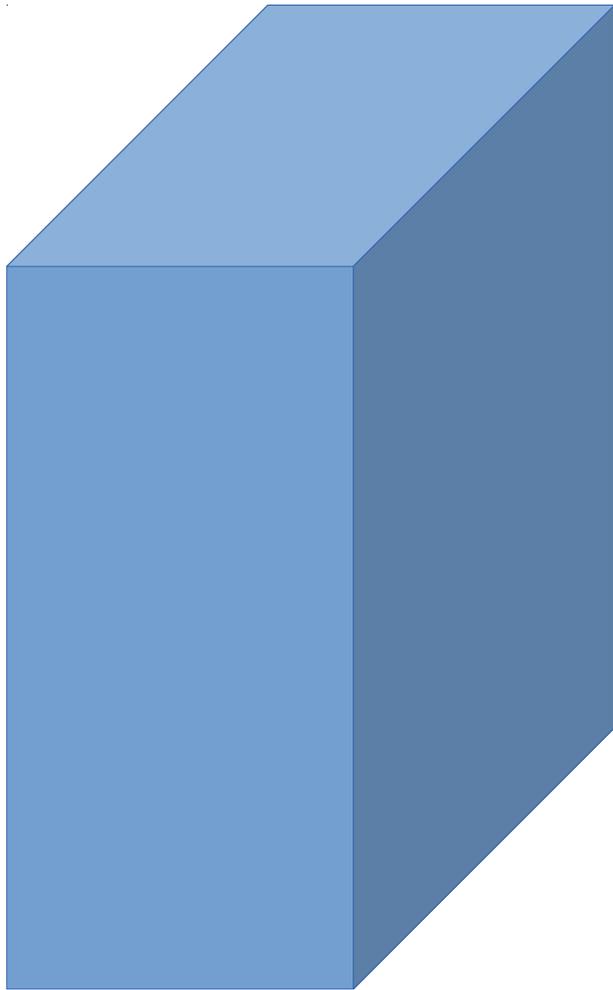
Gestores de Base de Datos

Escalar una Base de Datos



Gestores de Base de Datos

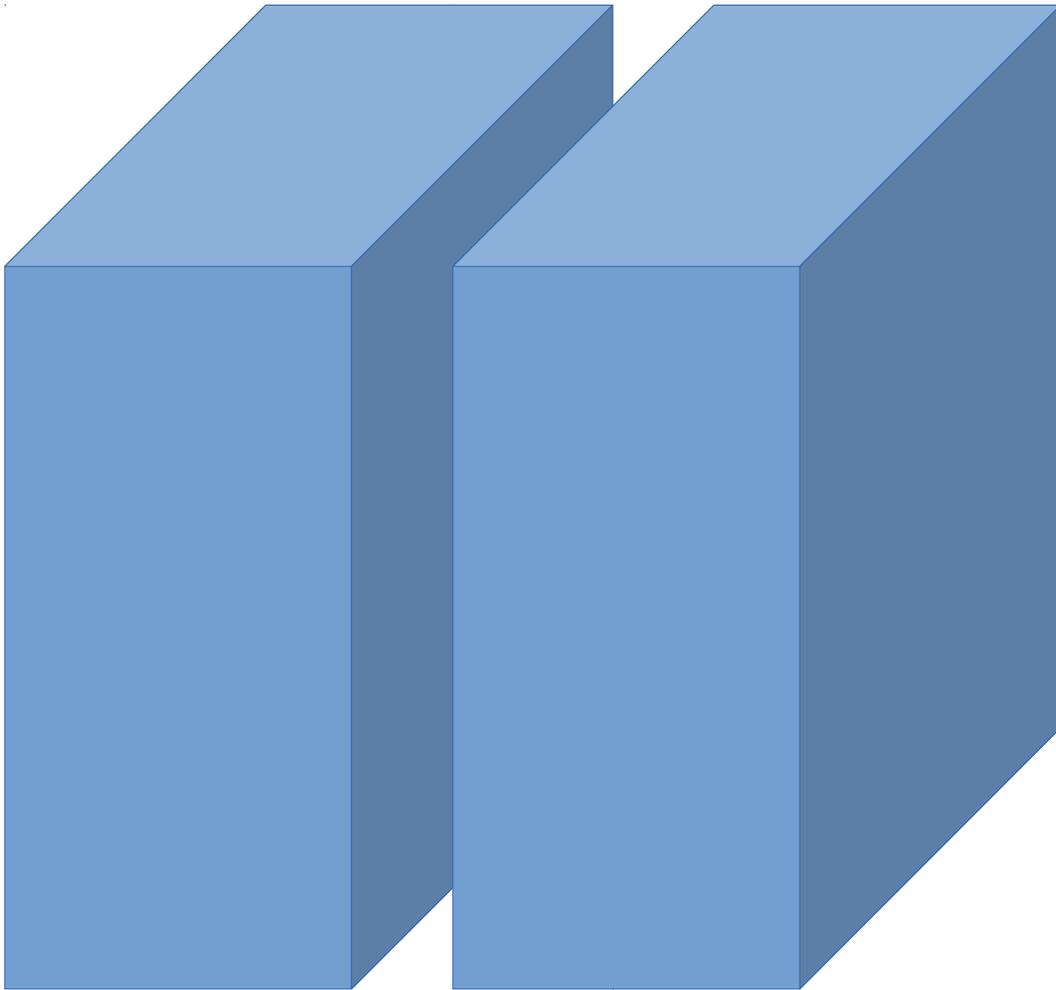
Escalar una Base de Datos



Scale Up

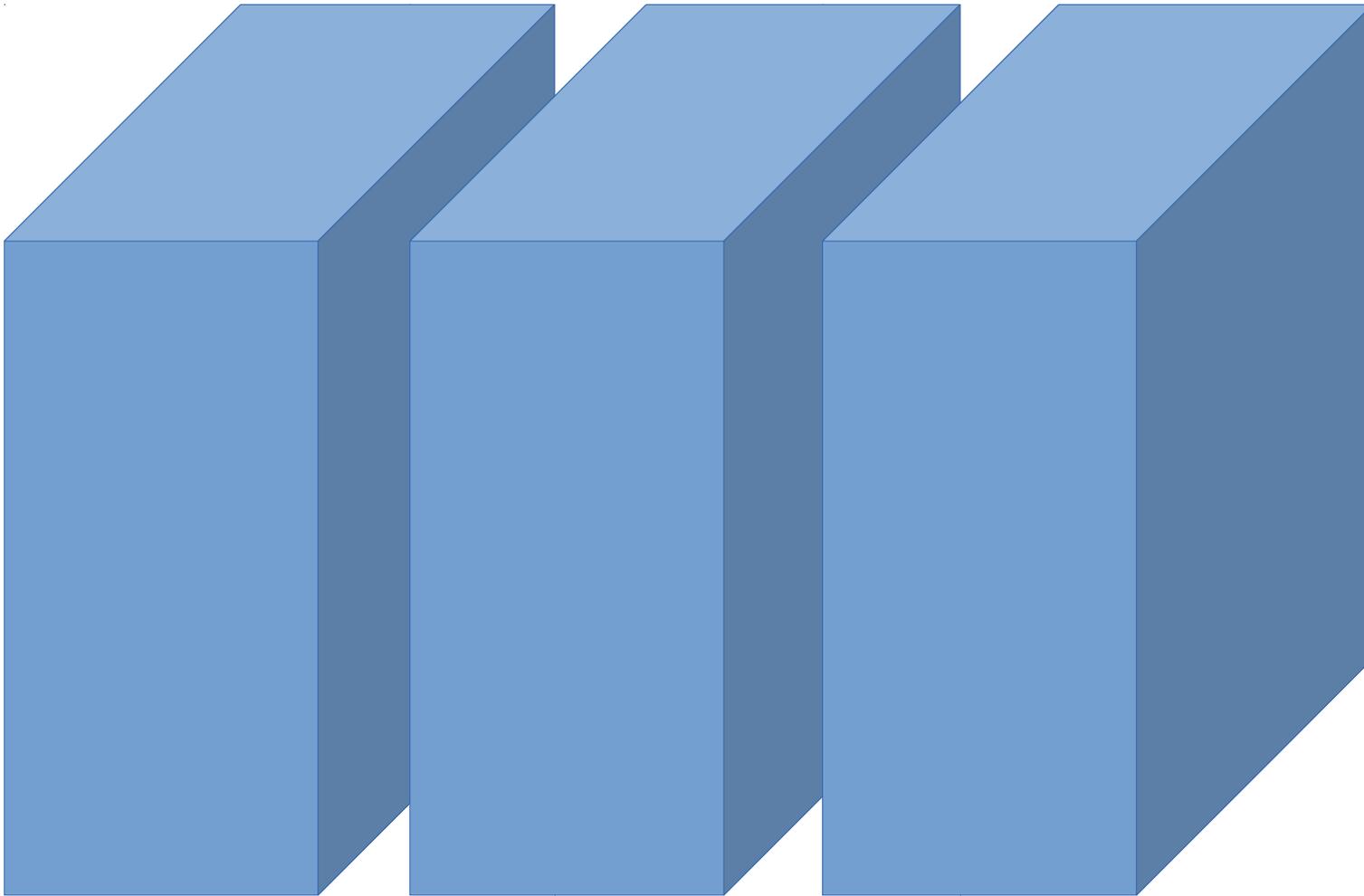
Gestores de Base de Datos

Escalar una Base de Datos



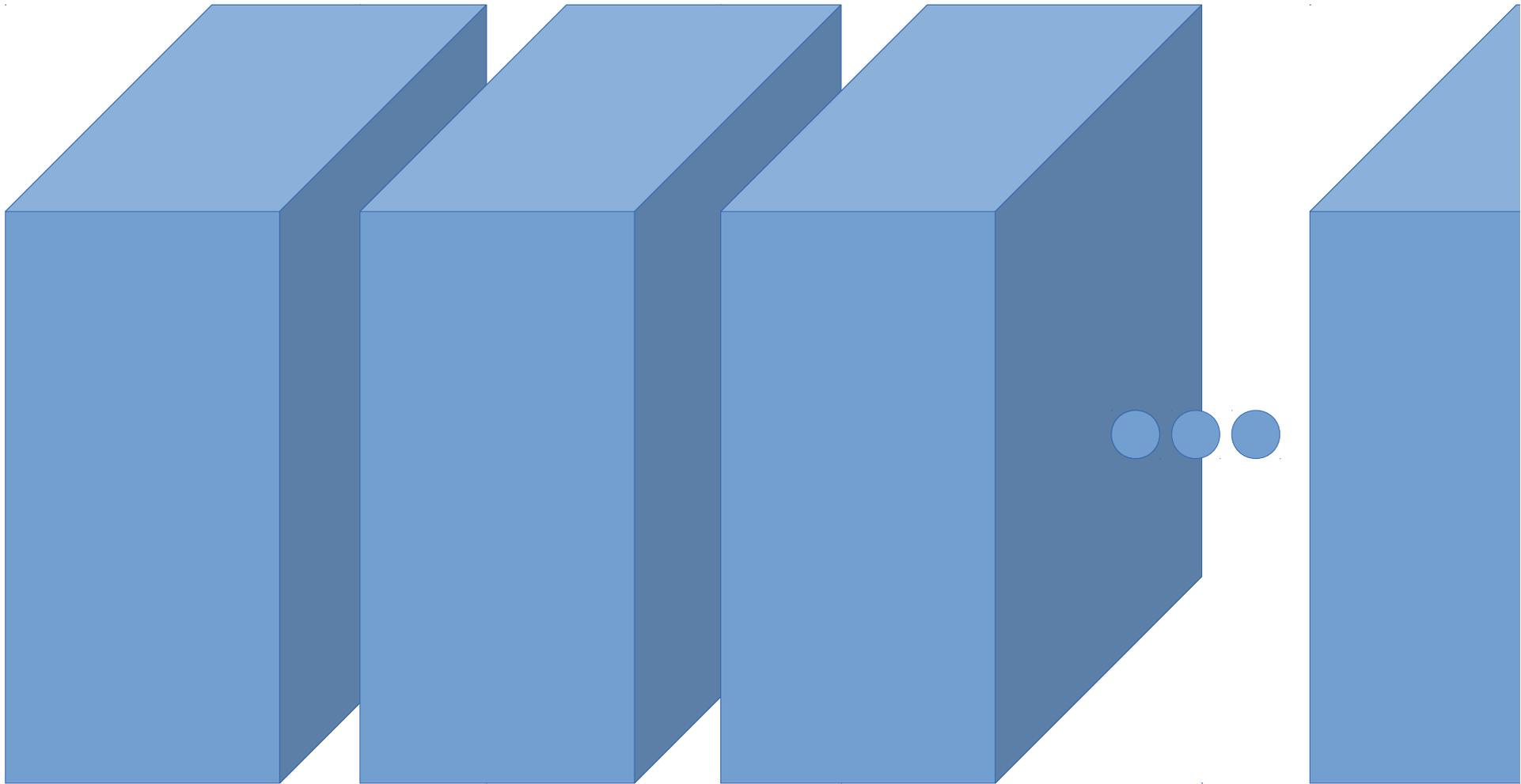
Gestores de Base de Datos

Escalar una Base de Datos



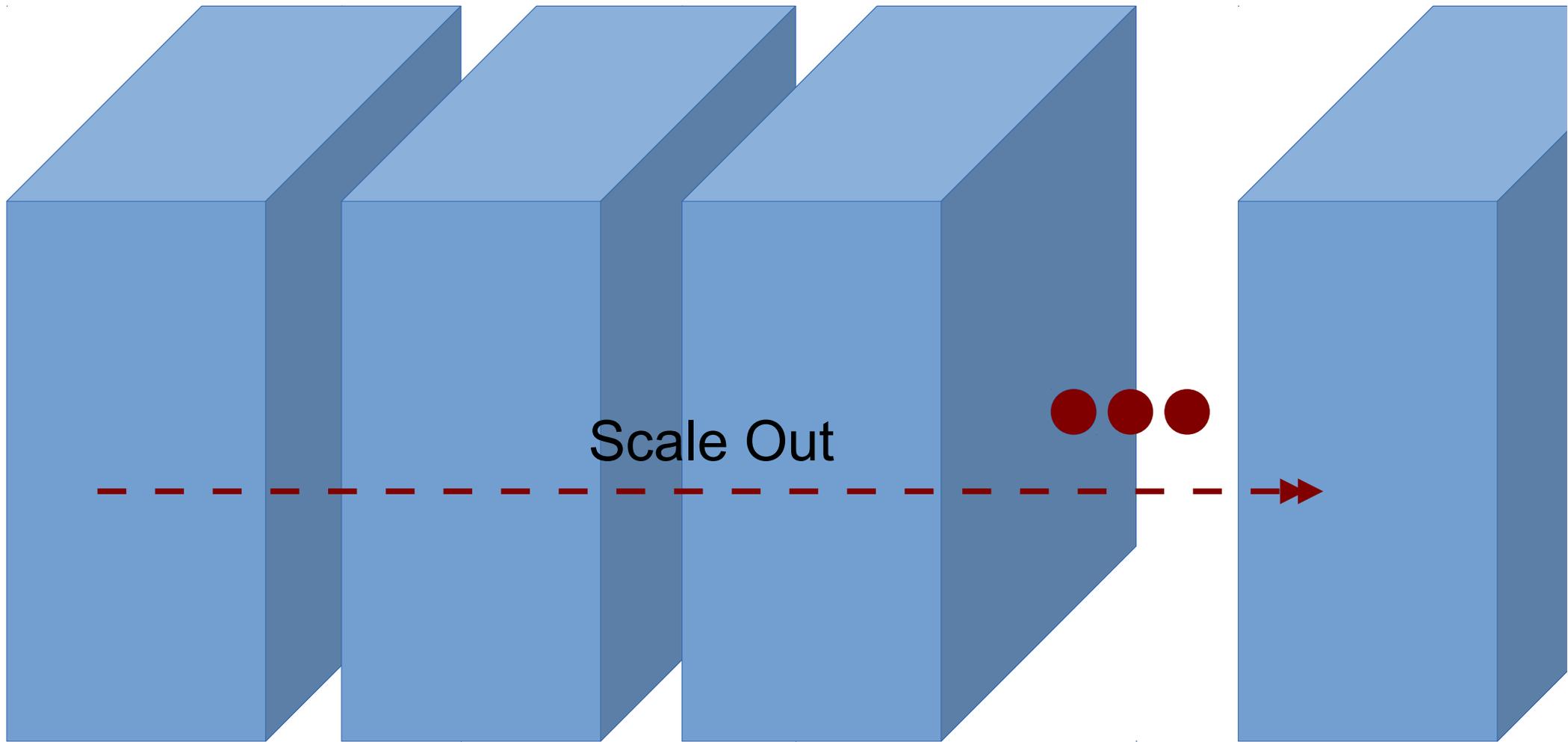
Gestores de Base de Datos

Escalar una Base de Datos

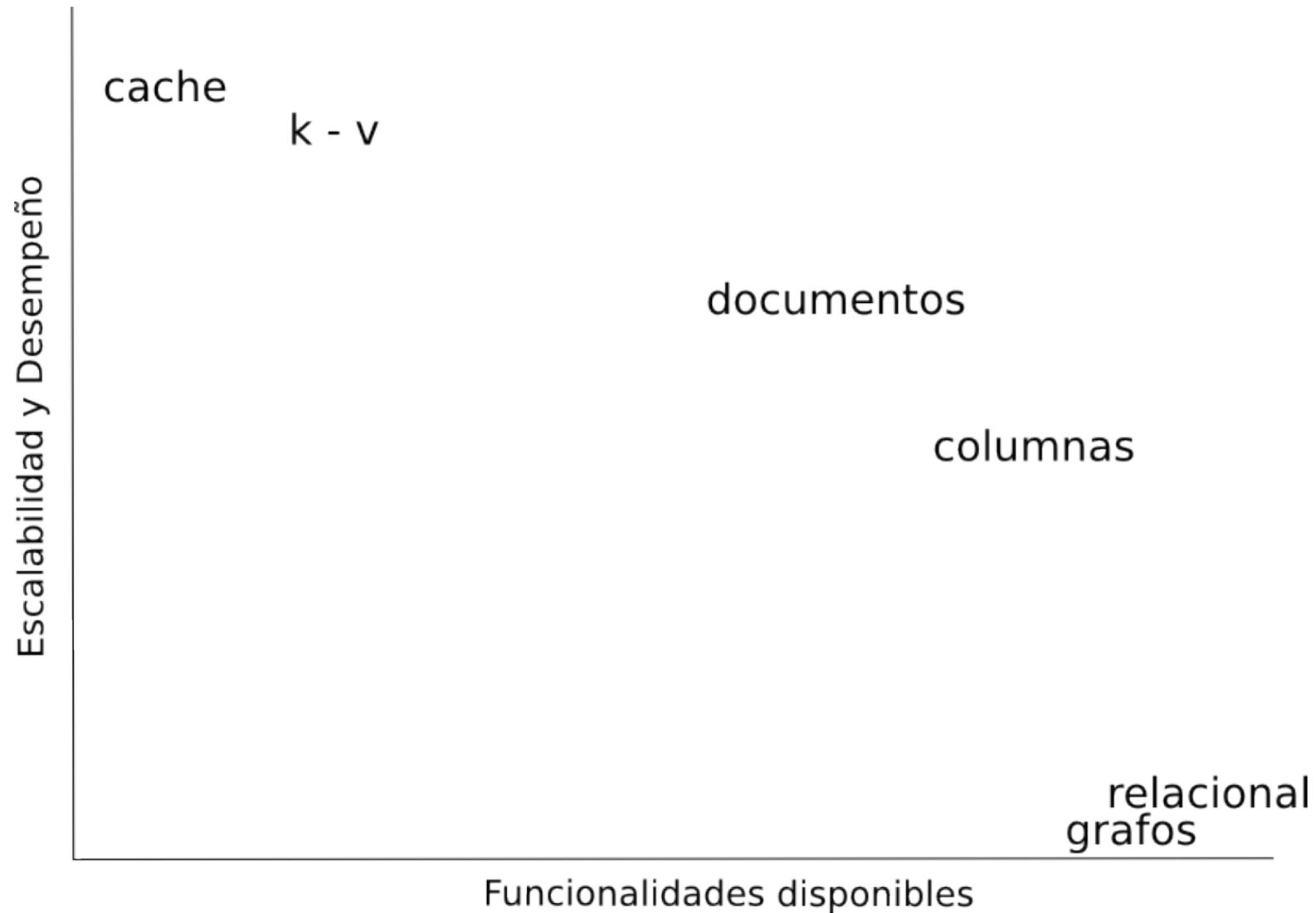


Gestores de Base de Datos

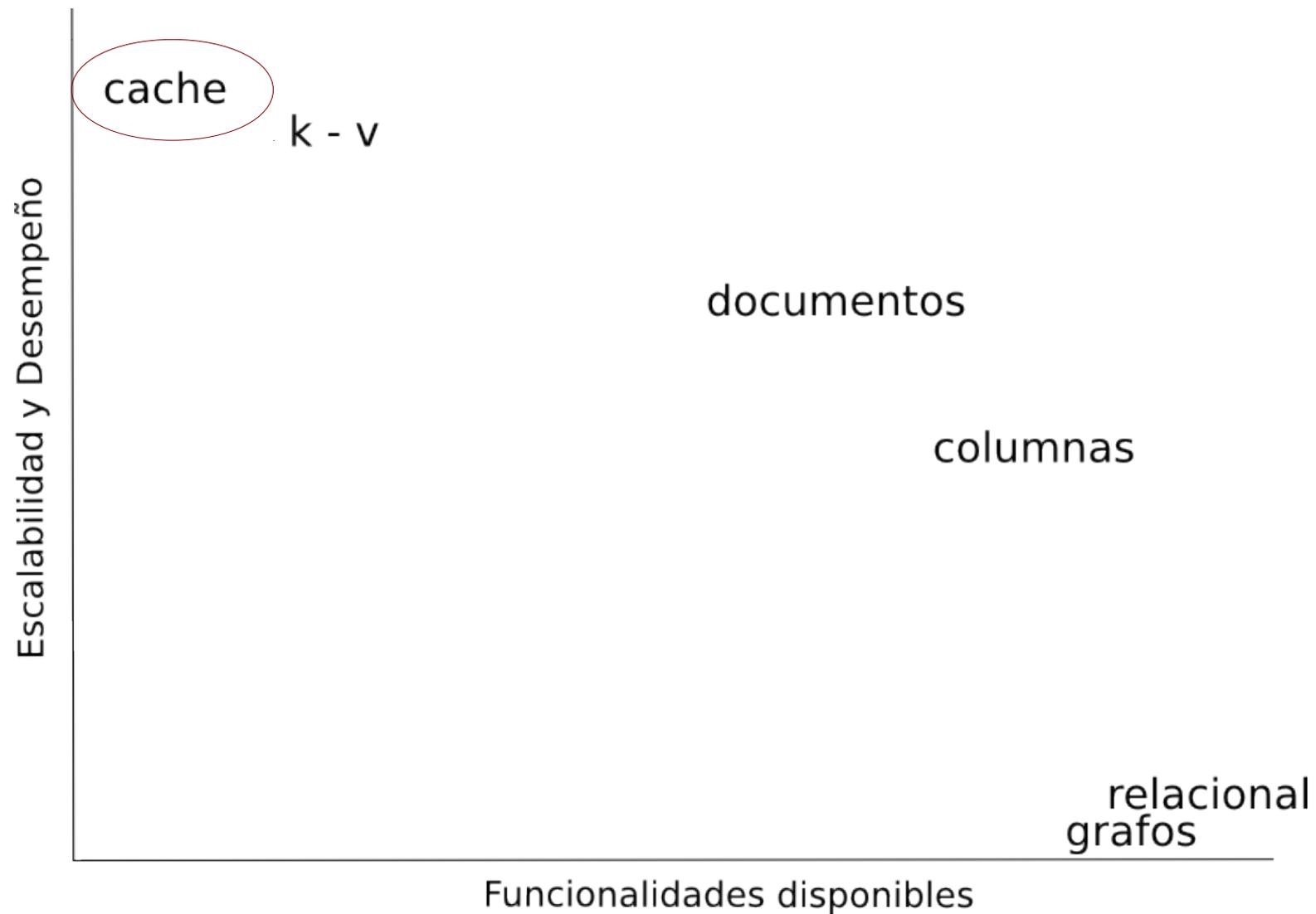
Escalar una Base de Datos



Gestores de Bases de Datos



Gestores de Bases de Datos



Mapa de los Gestores de Bases de Datos

La primera medida que se suele tomar para mejorar el desempeño de una aplicación que ha escalado horizontalmente es establecer mecanismos de memoria intermedia o cache.

Para hacer cache de una serie de valores dentro de una aplicación usamos:

Memcached

Mapa de los Gestores de Bases de Datos

- Memcached
 - Almacena llaves y valores en memoria
 - Es sumamente rápido
 - Su memoria es volátil
 - Ayuda a descargar las consultas directas a otros ámbitos de datos



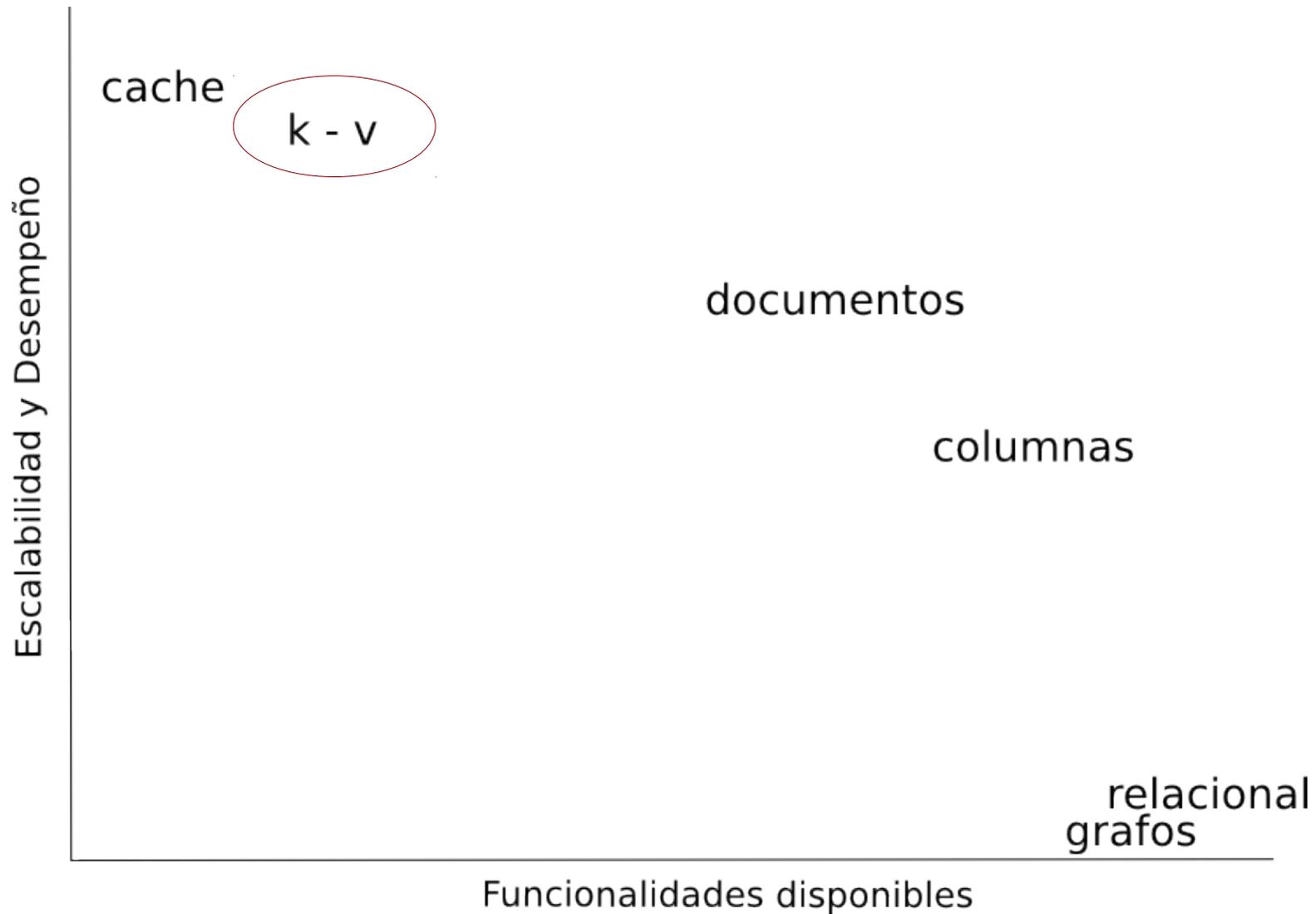
Mapa de los Gestores de Bases de Datos

- Memcached

Veamos un ejemplo



Gestores de Bases de Datos



Mapa de los Gestores de Bases de Datos

La forma tal vez más simple de organizar un grupo de datos es mediante un conjunto de llaves y valores.

En tal sentido, que una base de datos ofrezca más funcionalidades es un valor para quien usa este tipo de base de datos.

Mapa de los Gestores de Bases de Datos

Existen multitud de bases de datos libres que aplican con distintos énfasis este tipo de repositorio de datos, entre ellos:

Riak

Voldemort

Redis

Mapa de los Gestores de Bases de Datos

Existen multitud de bases de datos libres que aplican con distintos énfasis este tipo de repositorio de datos, entre ellos:

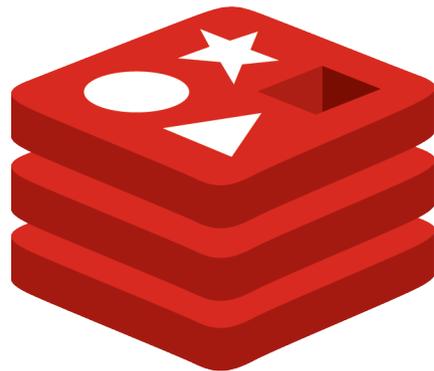
Riak

Voldemort

Redis

Mapa de los Gestores de Bases de Datos

- Redis
 - Es un gestor de base de datos de Llave-Valor
 - De alto rendimiento
 - Sin embargo no guarda a disco inmediatamente
 - Tiene soporte para tipos de datos complejos como, hashes, listas, conjuntos y conjuntos ordenados



redis

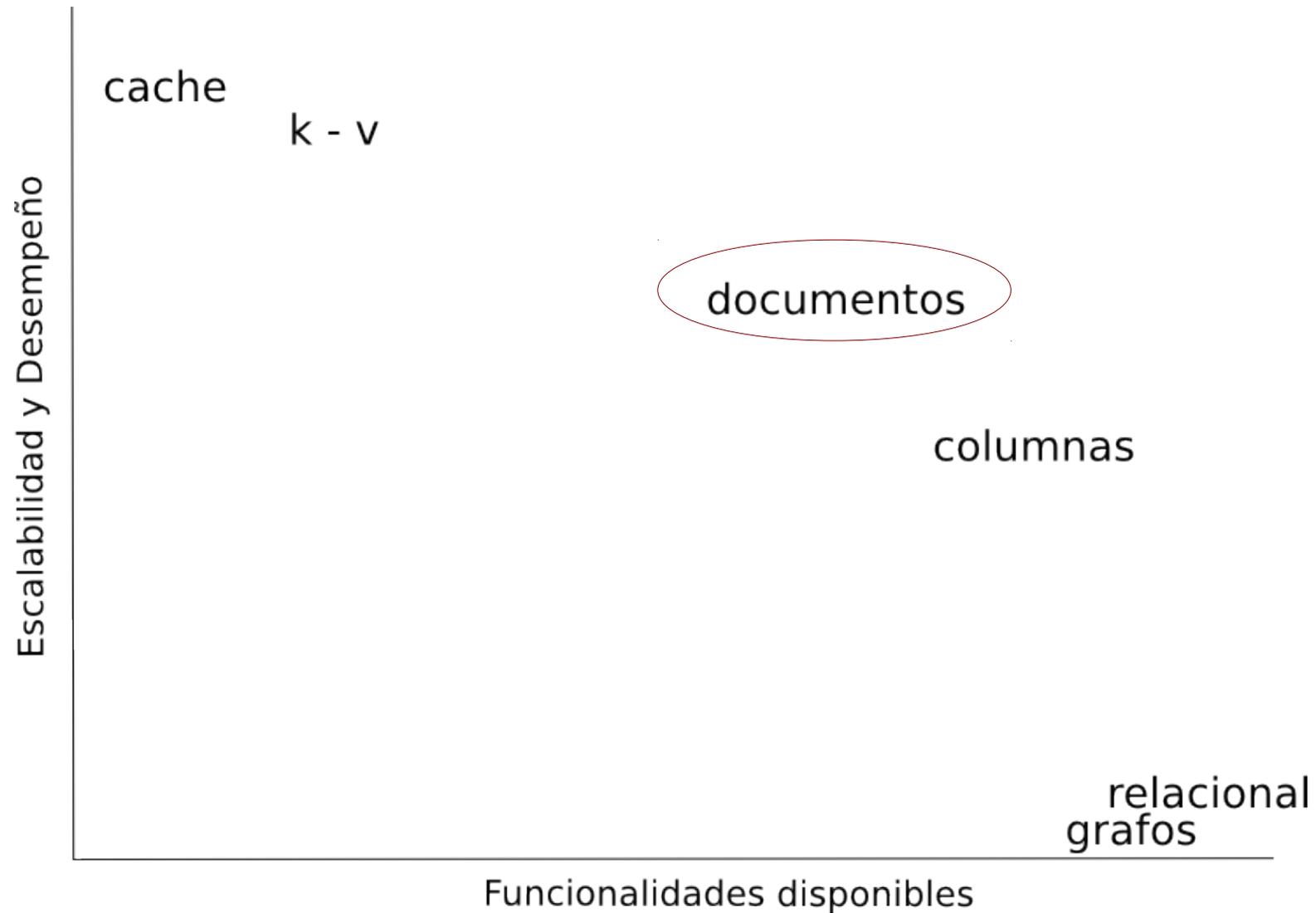
Mapa de los Gestores de Bases de Datos

- Redis

Veamos un ejemplo



Gestores de Bases de Datos



Mapa de los Gestores de Bases de Datos

En algunos casos es deseable mantener junta la mayor cantidad de data relacionada. Como vimos con los hashes de redis es posible asociar un grupo de llaves y valores a una llave principal.

Cuando estos hashes son más complejos entonces recurrimos a estructuras más adecuadas para la gestión de documentos, como pueden ser; XML o JSON

Mapa de los Gestores de Bases de Datos

Cuando se habla de documentos, es en el sentido de los documentos JSON o XML, no documentos ofimáticos como LibreOffice u otro equivalente.

Mapa de los Gestores de Bases de Datos

- MongoDB
 - Diseñado para ser grande
 - Esquema flexible
 - Basado en Documentos JSON
 - Listo de paquete para escalar horizontalmente
 - Fácil de aprender



Mapa de los Gestores de Bases de Datos

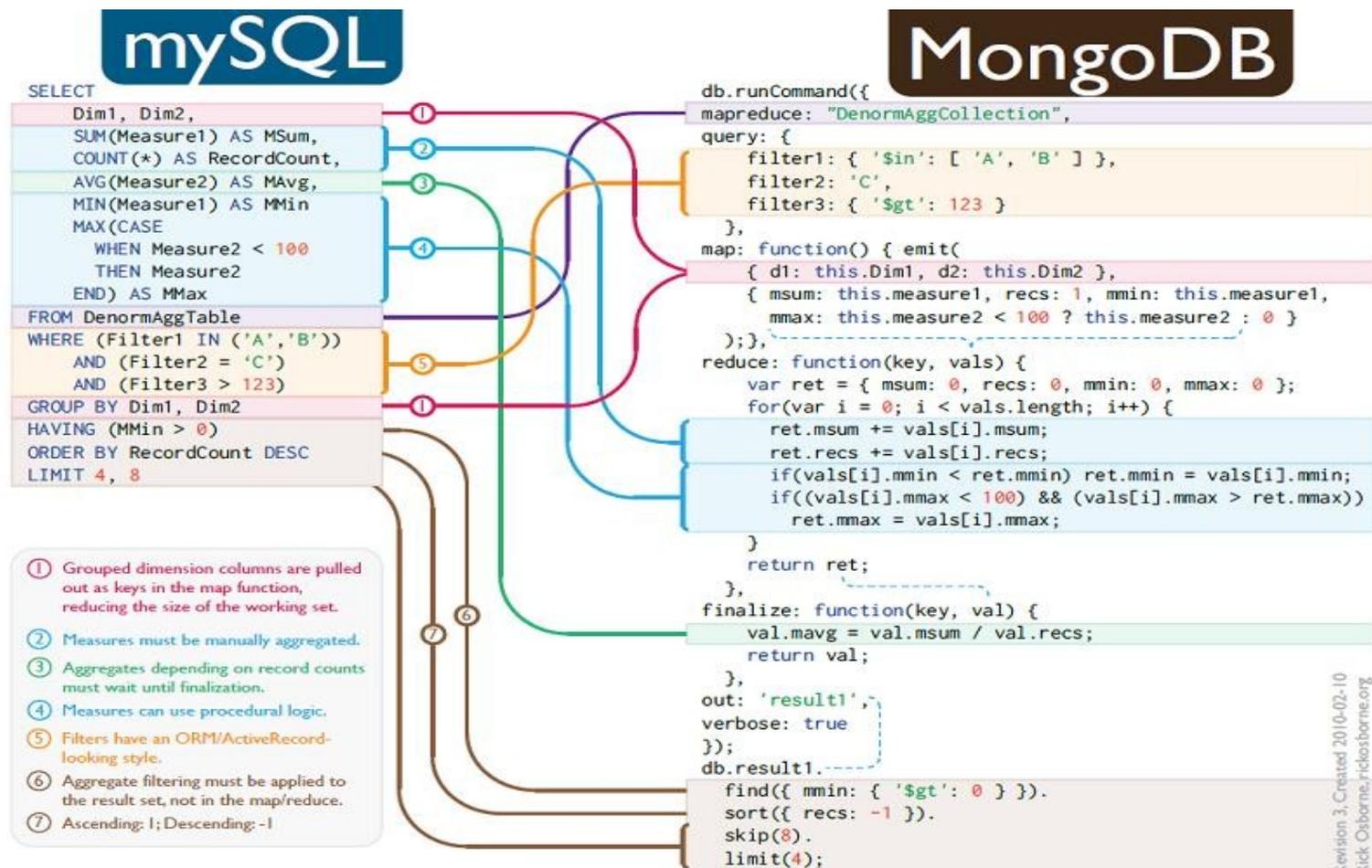
- MongoDB

Veamos un ejemplo



Mapa de los Gestores de Bases de Datos

- MongoDB

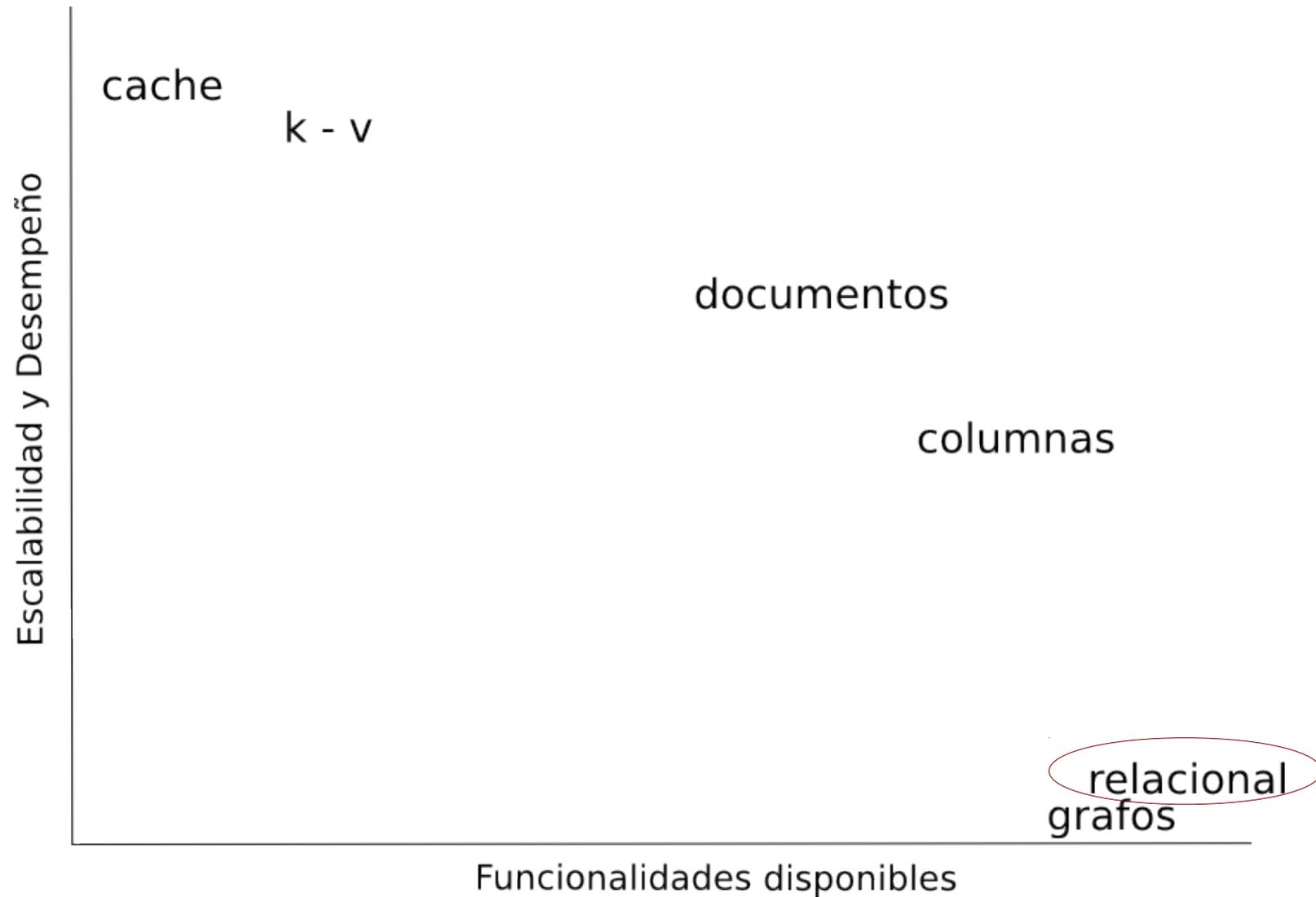


Anuncio

En breves momentos estaremos aterrizando, nuestro recorrido termina donde normalmente comienzan todos los cursos de base de datos.

Nos encontramos sobrevolando sobre las
“Bases de Datos Relacionales”

Gestores de Bases de Datos



Mapa de los Gestores de Bases de Datos

En el mundo de las bases de datos existe un modelo predominante que se conoce como relacional.

Su lenguaje de consultas es “casi” universal

Se basa en la lógica de predicados y la teoría de los conjuntos, postulado inicialmente por Edgar Frank Codd en 1970.

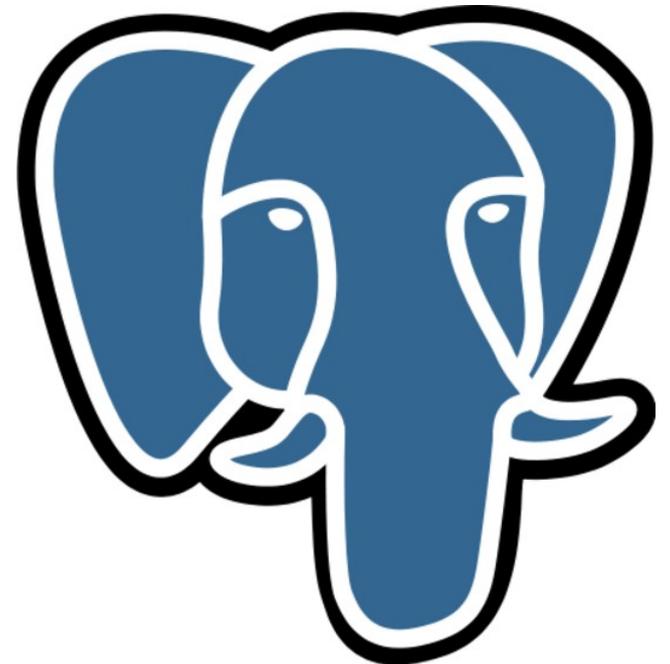
Mapa de los Gestores de Bases de Datos

Las relaciones (o conjuntos de datos) se representan como tablas bidimensionales de filas y columnas.

Codd también definió las reglas de normalización de estas bases de datos.

Mapa de los Gestores de Bases de Datos

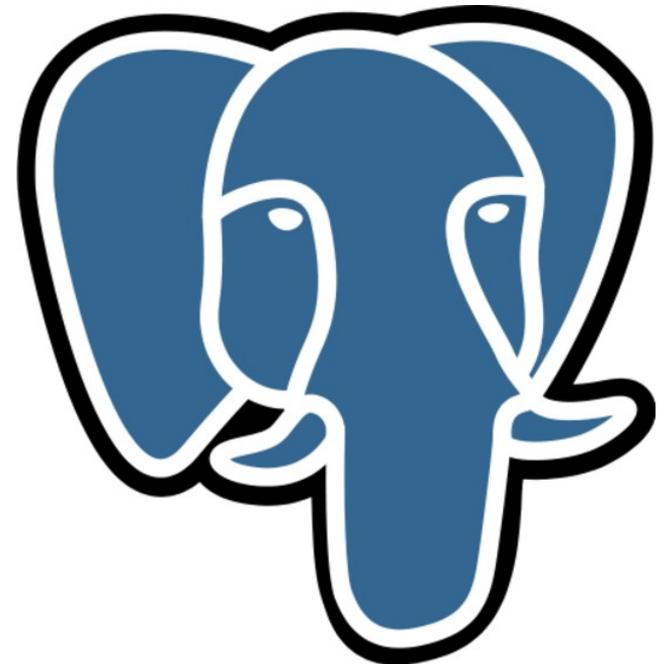
- PostgreSQL
 - Implementa una forma de SQL apegada al estándar
 - Extensible
 - Gran conjunto de tipos de datos
 - Integridad Referencial



Mapa de los Gestores de Bases de Datos

- PostgreSQL

Veamos un ejemplo



Otras Bases de Datos

- Bases de Datos de Grafos

- Para datos altamente conectados
- El programa se llama Neo4J



- Bases de Datos en Columnas

- Similares a las relacionales sin embargo se concentran en las columnas y familias de columnas para gestionar los datos.
- Su lenguaje de consulta es CQL (equivalente al SQL)



Aproximación Final



Aproximación Final

- Para la reflexión:
 - Existe una base de datos libre que se ajusta a cada tipo de problema u organización de la información.
 - Aunque puedas hacer la mayor parte del trabajo con una base de datos relacional, es posible que este enfoque te genere retrabajo por la migración de esquemas o inconvenientes al escalar.
 - “no digas que no, si no lo has probado”, en otras palabras evita decir “no” por defecto. Atrévete a usar otras tecnologías

Aproximación Final

- Para la reflexión:
 - No existe una solución perfecta, todos los gestores de base de datos requieren compensaciones.
 - Si quieres muchas funcionalidades, probablemente no escale bien
 - Si quieres mucha velocidad, probablemente no puedas guardar a disco con frecuencia
 - Si quieres mayor flexibilidad, probablemente no puedas normalizar del todo los datos

Aterrizaje

- Agradecimientos
 - Prof. Julimar García
 - Ing. Degli Pino
 - Amigos del CNTI
 - Jóvenes y docentes de la UPTOSCR

Gracias a todos por la oportunidad, por su tiempo y su atención

