



Data Management Fundamentals

Módulo 12 - Data Storage & Operations

GOLD SPONSORS



SILVER SPONSORS



Entidades académicas



Universidad de Castilla-La Mancha



Alberto Turégano



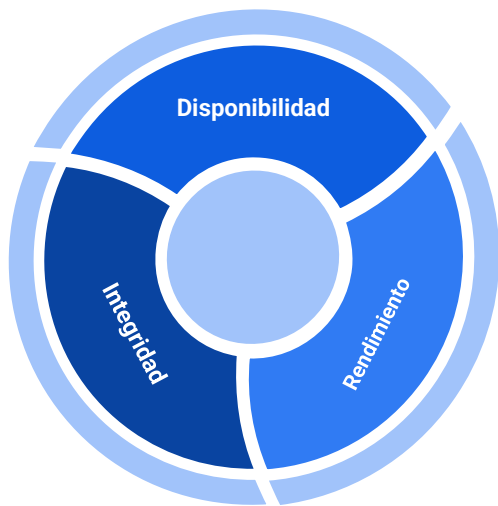
- Ingeniero informático UPM
- CDMP nivel Máster DAMA Internacional
- Máster en tecnologías Big Data e Inteligencia Artificial – Datahack School
- Global Management Program - IESE
- Programa ejecutivo Humanismo Digital - EOI
- Certificado en herramientas de IA para la productividad - The Valley

Almacenamiento de datos y operaciones



- Es el diseño, la implantación y soporte de los datos almacenados con el objetivo de maximizar su valor a lo largo de todo su ciclo de vida. Incluye:
 - El soporte de BB.DD
 - El soporte tecnológico de las BB.DD

Metas



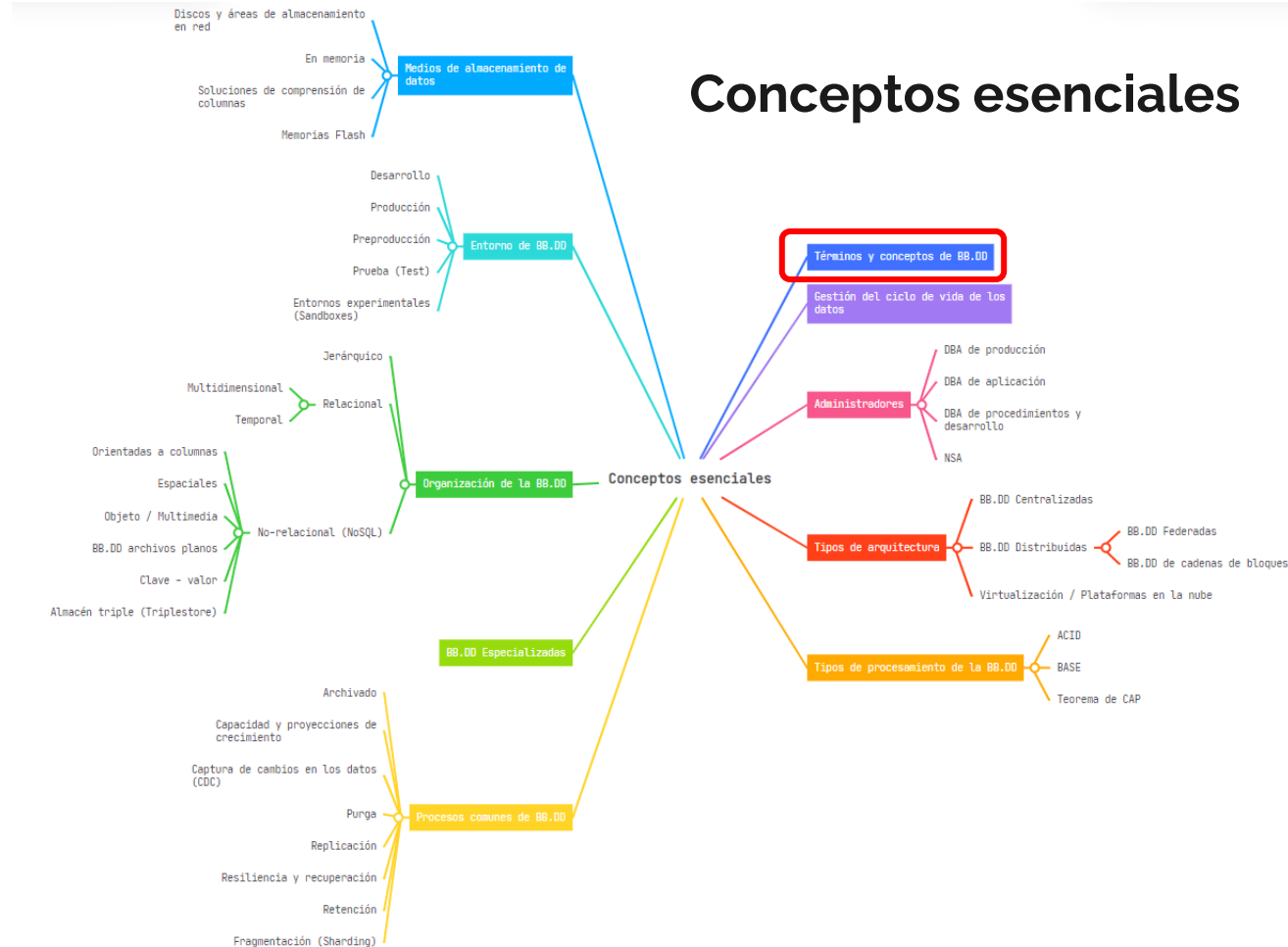
- Gestionar la **disponibilidad** de los datos a lo largo de su ciclo de vida
- Garantizar la **integridad** de los activos de datos
- Gestionar el **rendimiento** de las transacciones de datos

Conceptos esenciales

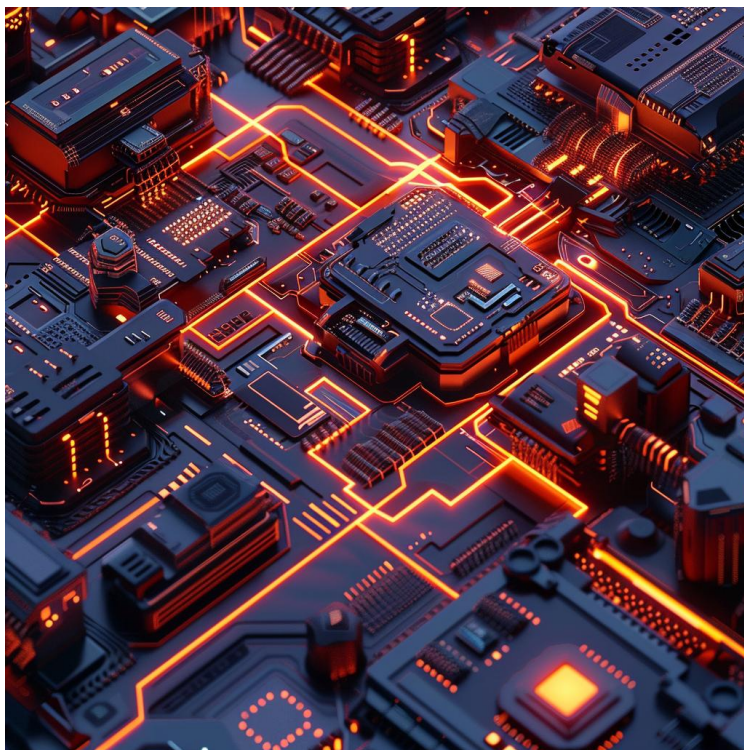
Conceptos esenciales



Conceptos esenciales

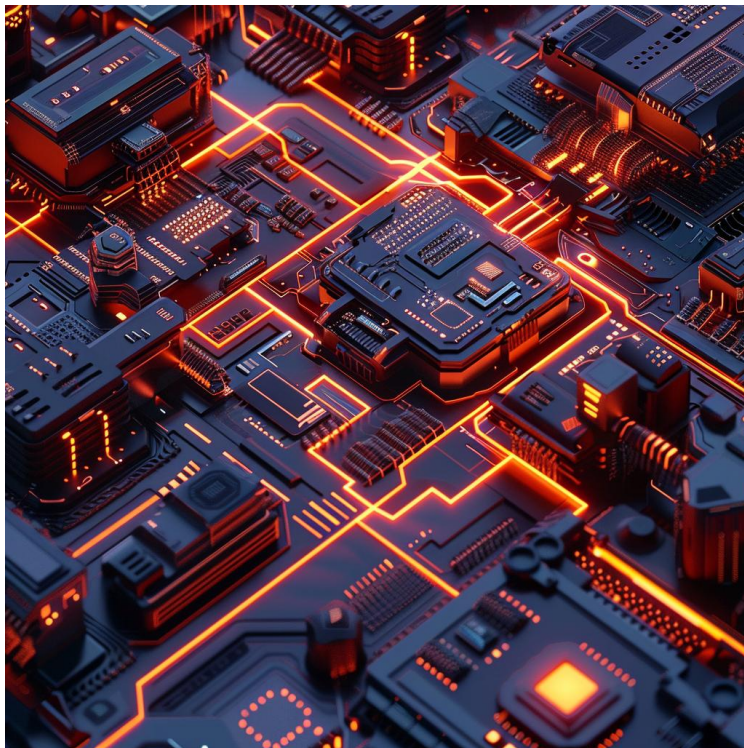


Términos y conceptos de BB.DD



Base de datos: Cualquier colección de datos almacenados, independientemente de su estructura o contenido. Algunas grandes bases de datos se refieren a **instancias** y a **esquemas**.

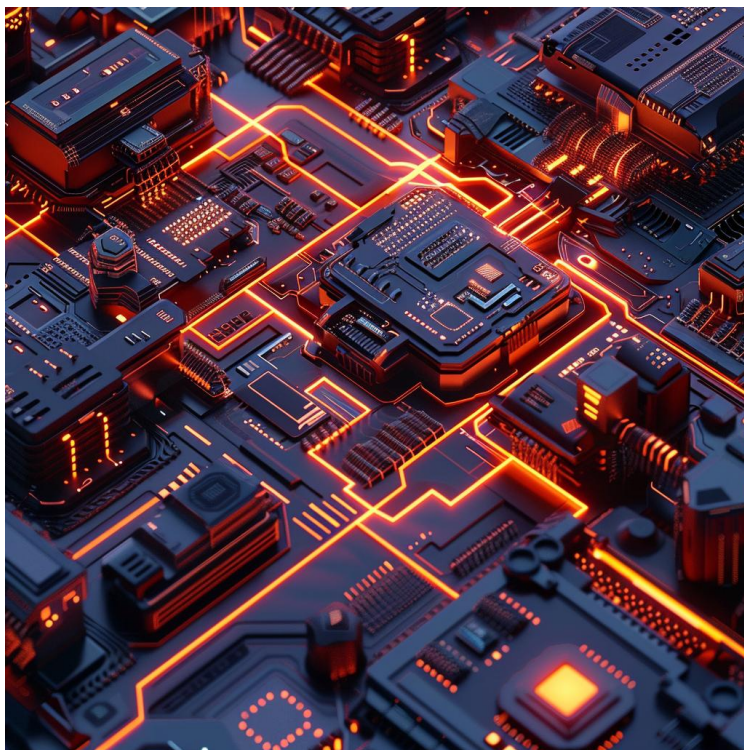
Términos y conceptos de BB.DD



Base de datos: Cualquier colección de datos almacenados, independientemente de su estructura o contenido. Algunas grandes bases de datos se refieren a **instancias** y a **esquemas**.

- **Instancia:** Una ejecución del software de la base de datos que controla el acceso a una determinada área de almacenamiento. **Una organización suele tener varias instancias que se ejecutan simultáneamente**, utilizando diferentes áreas de almacenamiento. **Cada instancia es independiente de las demás.**

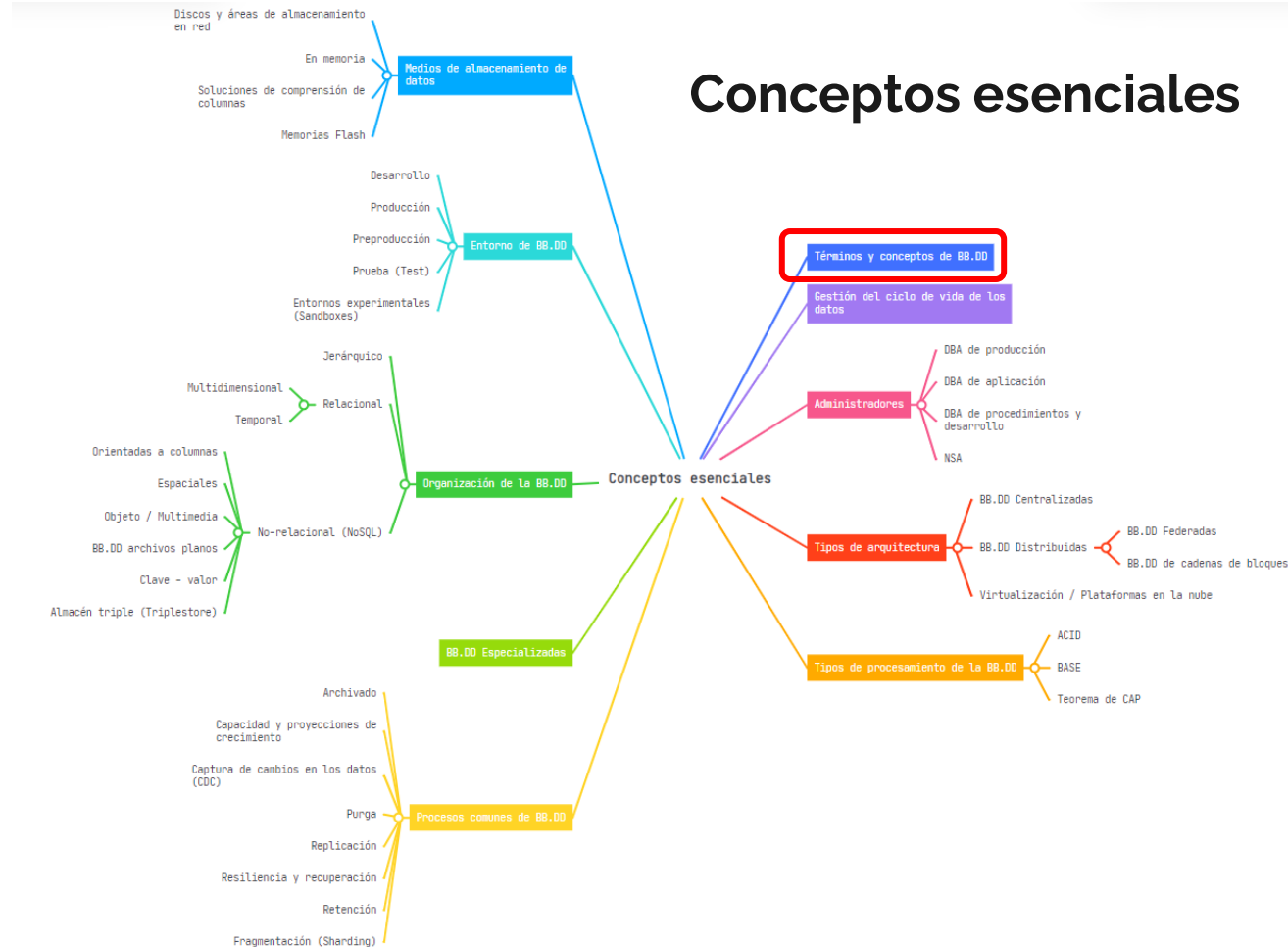
Términos y conceptos de BB.DD



Base de datos: Cualquier colección de datos almacenados, independientemente de su estructura o contenido. Algunas grandes bases de datos se refieren a **instancias y a esquemas**.

- **Instancia:** Una ejecución del software de la base de datos que controla el acceso a una determinada área de almacenamiento. **Una organización suele tener varias instancias que se ejecutan simultáneamente**, utilizando diferentes áreas de almacenamiento. **Cada instancia es independiente de las demás.**
- **Esquema:** Un subconjunto de objetos de una base de datos contenidos en la base de datos o en una instancia. **Los esquemas se utilizan para organizar los objetos en partes más manejables.** Un esquema tiene un propietario y una lista de acceso particular para el contenido del esquema. El esquema también puede utilizarse para referirse a una colección de estructuras de bases de datos con algo en común.

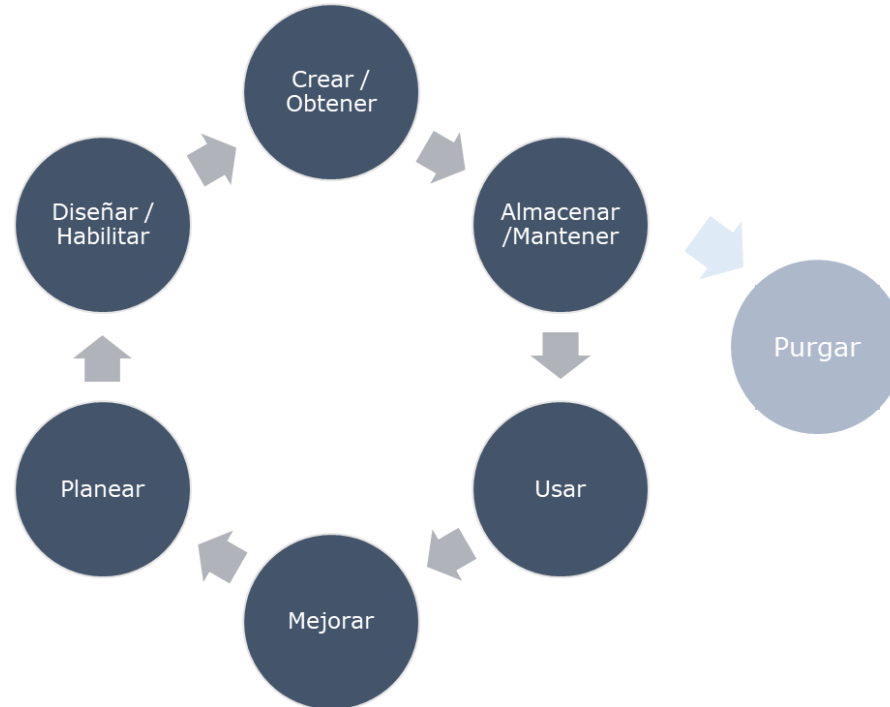
Conceptos esenciales



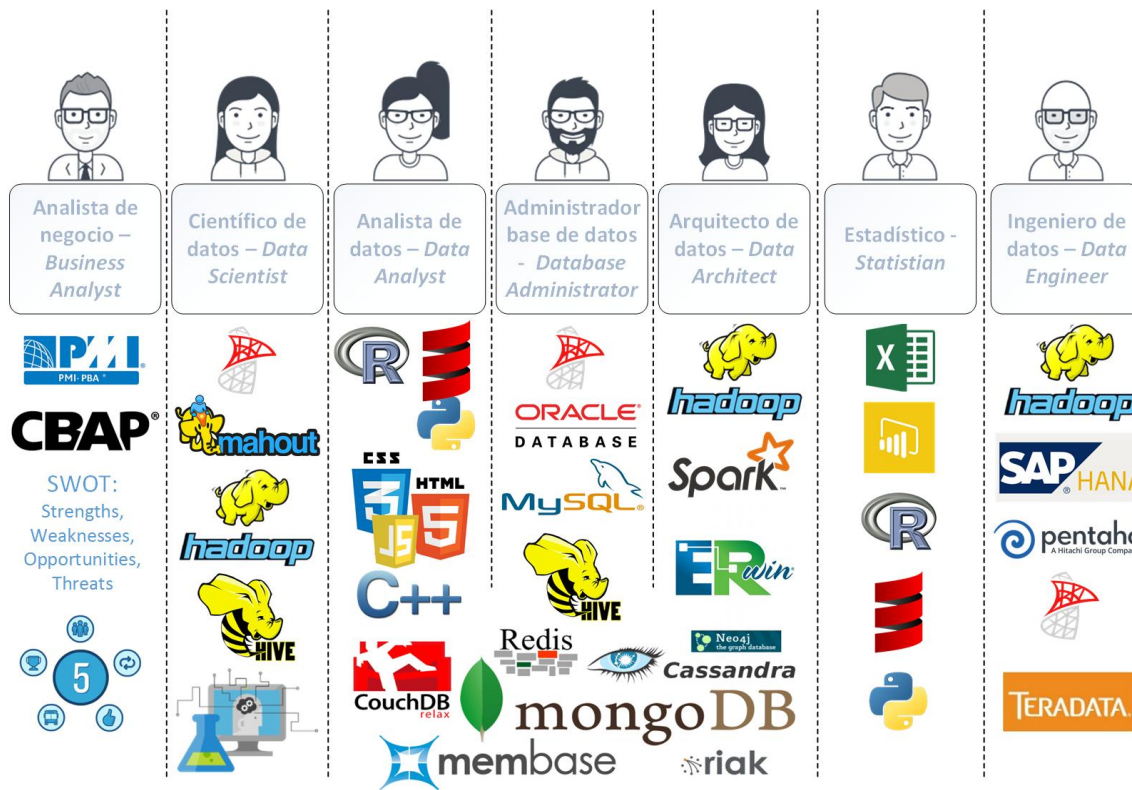
Conceptos esenciales



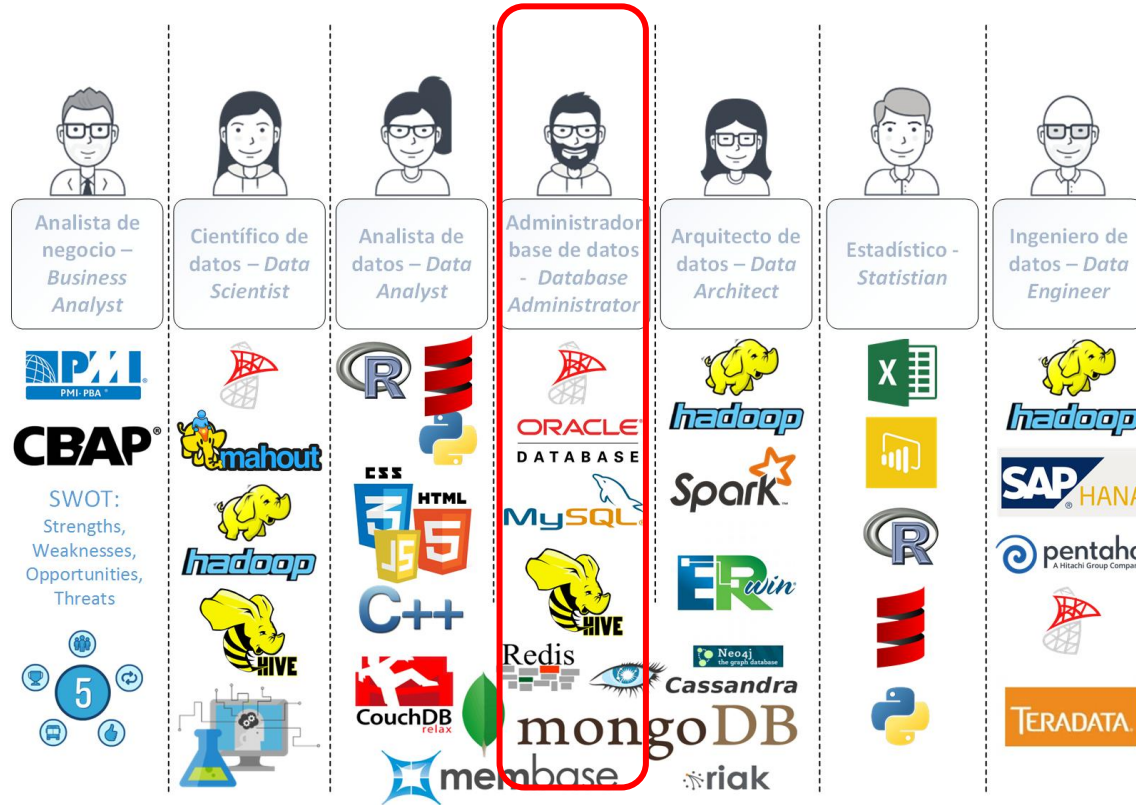
Gestión del ciclo de vida de los datos



Gestión del ciclo de vida de los datos

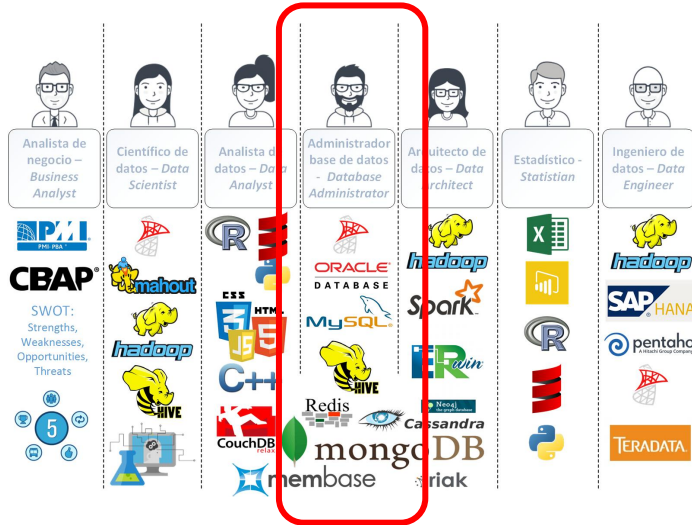


Gestión del ciclo de vida de los datos

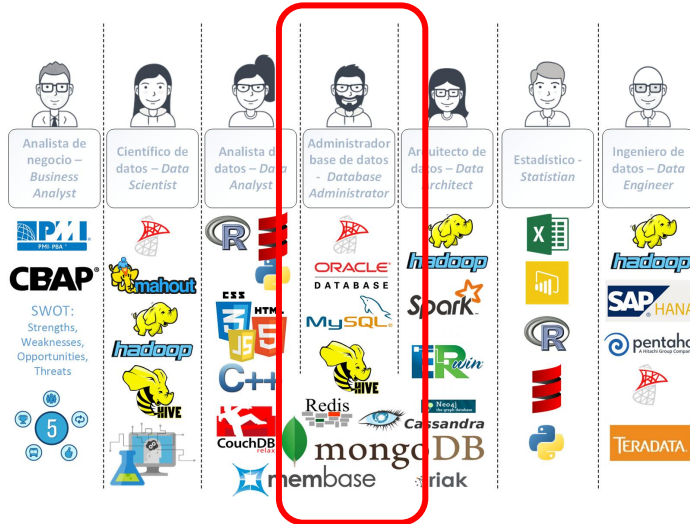


Gestión del ciclo de vida de los datos

- Los DBAs **mantienen y aseguran la precisión y consistencia de los datos durante todo su ciclo de vida** a través del diseño, implementación y uso de cualquier sistema que almacene, procese o recupere datos.

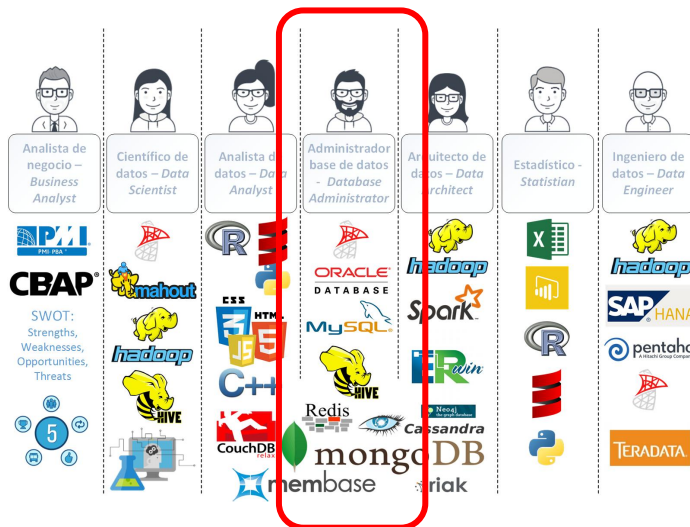


Gestión del ciclo de vida de los datos

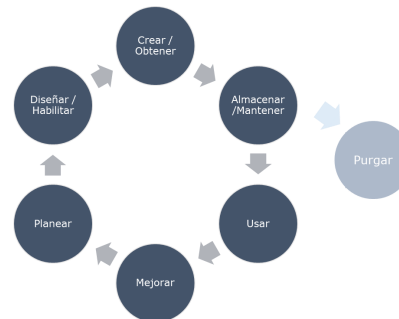


- Los DBAs **mantienen y aseguran la precisión y consistencia de los datos durante todo su ciclo de vida** a través del diseño, implementación y uso de cualquier sistema que almacene, procese o recupere datos.
- El DBA es el custodio de todos los cambios en las BB.DD.

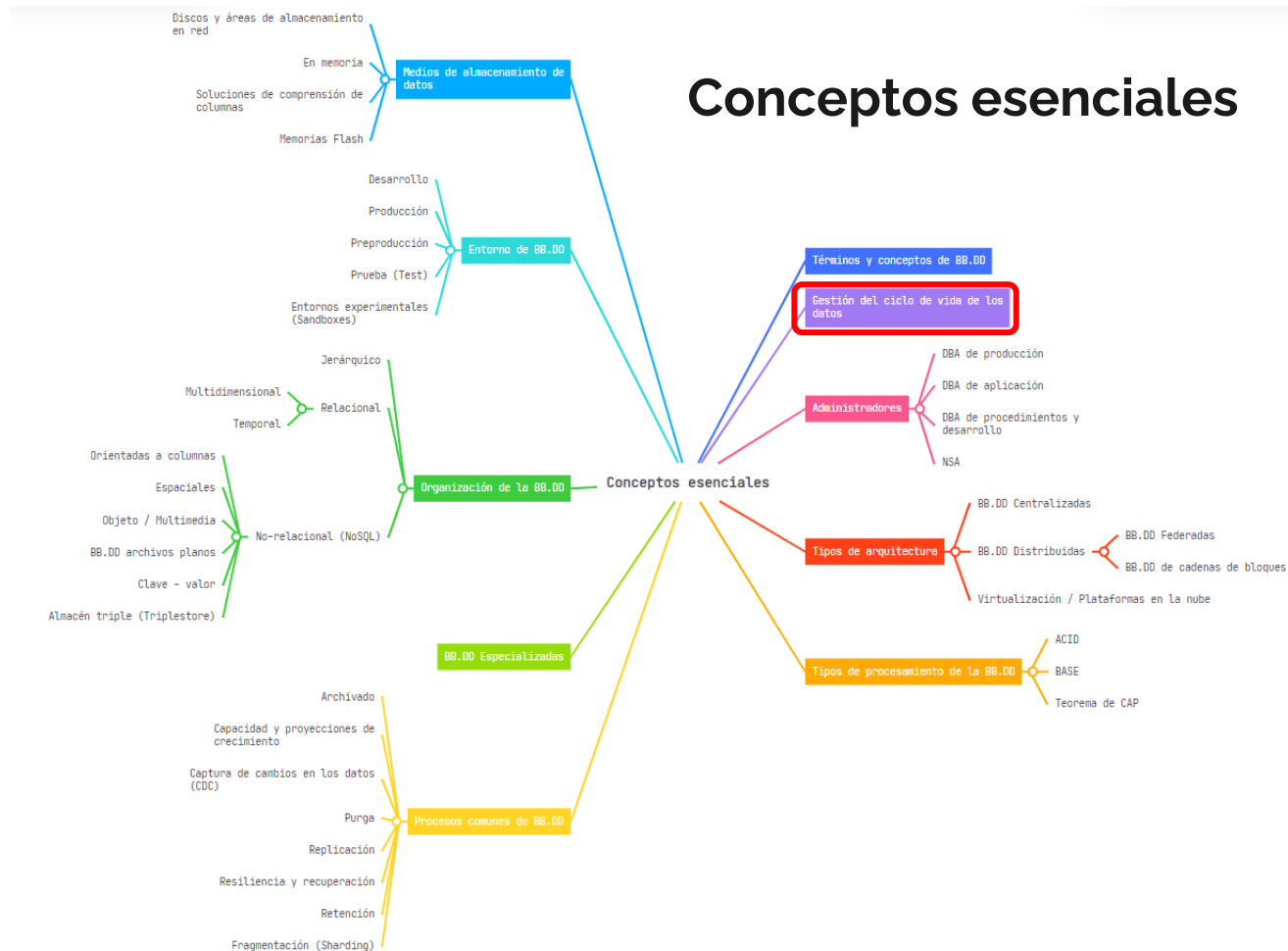
Gestión del ciclo de vida de los datos



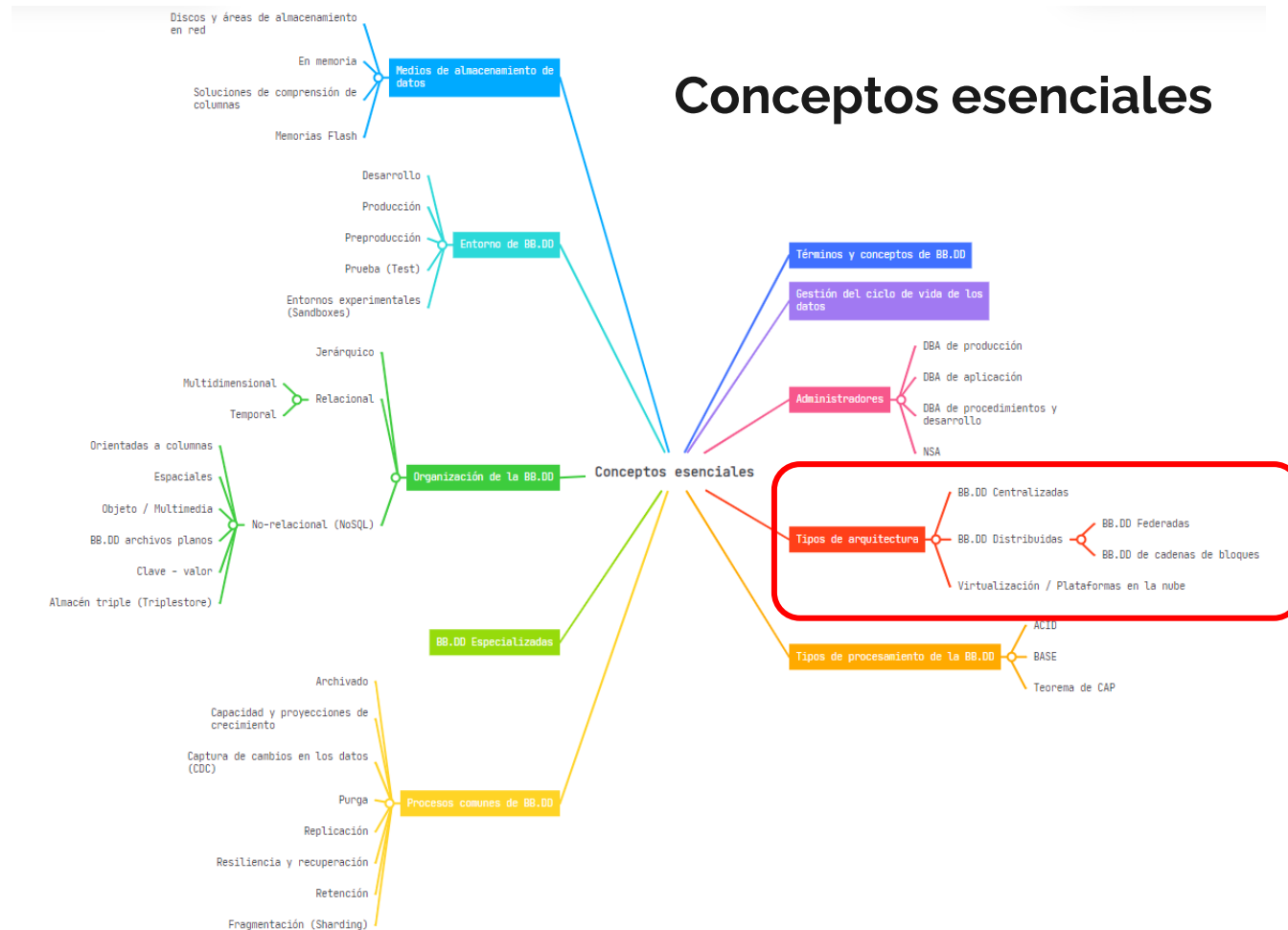
- Los DBAs **mantienen y aseguran la precisión y consistencia de los datos durante todo su ciclo de vida** a través del diseño, implementación y uso de cualquier sistema que almacene, procese o recupere datos.
- El DBA es el custodio de todos los cambios en las BB.DD.
- La gestión del ciclo de vida de los datos incluye la implementación de políticas y procedimientos para la adquisición, migración, retención, expiración y disposición de los datos:



Conceptos esenciales



Conceptos esenciales



Tipos de arquitecturas de BB.DD

Tipos de arquitecturas de BB.DD

Centralizado



Tipos de arquitecturas de BB.DD

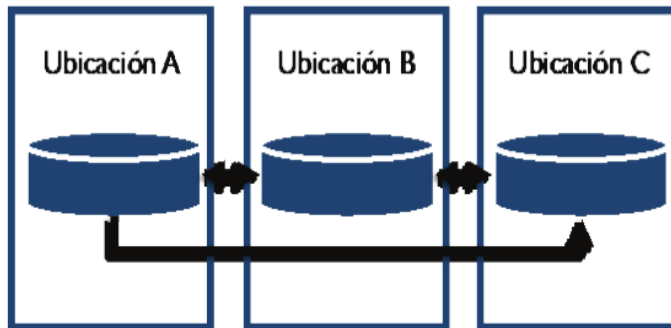
Centralizado

Vista de
Usuario

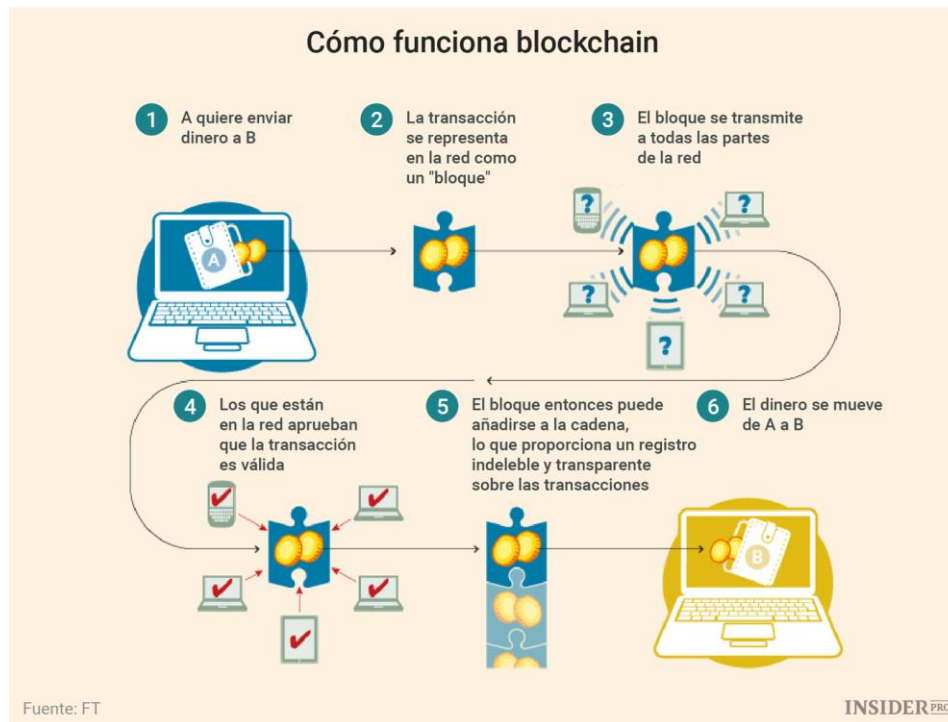


Distribuido, no Federado

Vista de Usuario



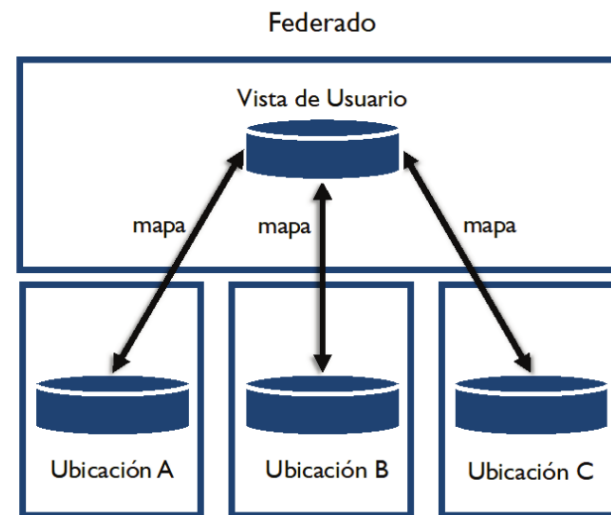
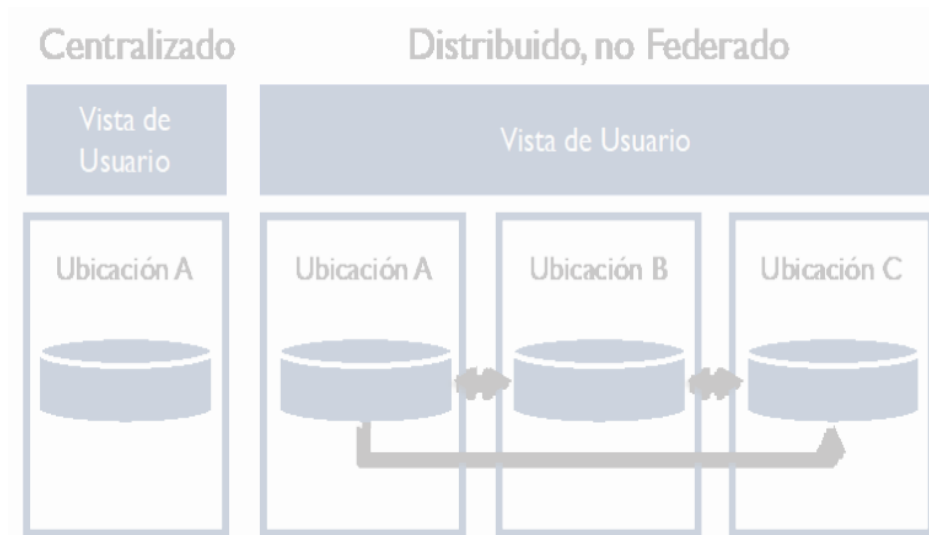
BB.DD Distribuidas: Blockchain



BB.DD Distribuidas: Blockchain



Tipos de arquitecturas de BB.DD !

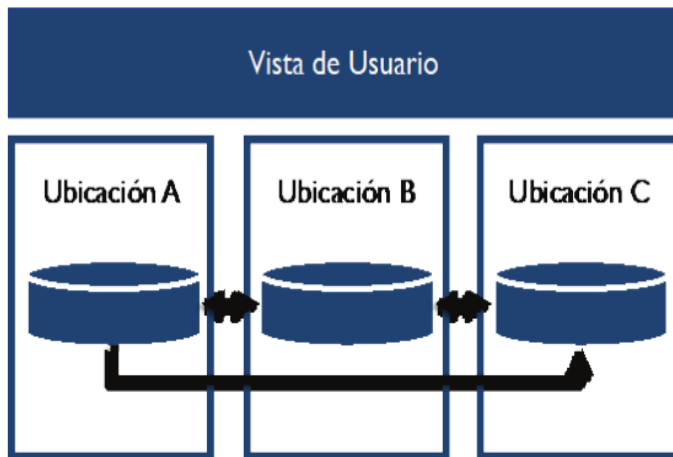


Tipos de arquitecturas de BB.DD !

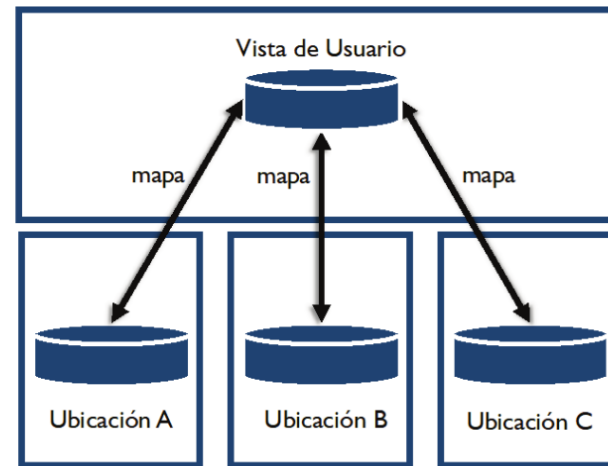
Centralizado



Distribuido, no Federado

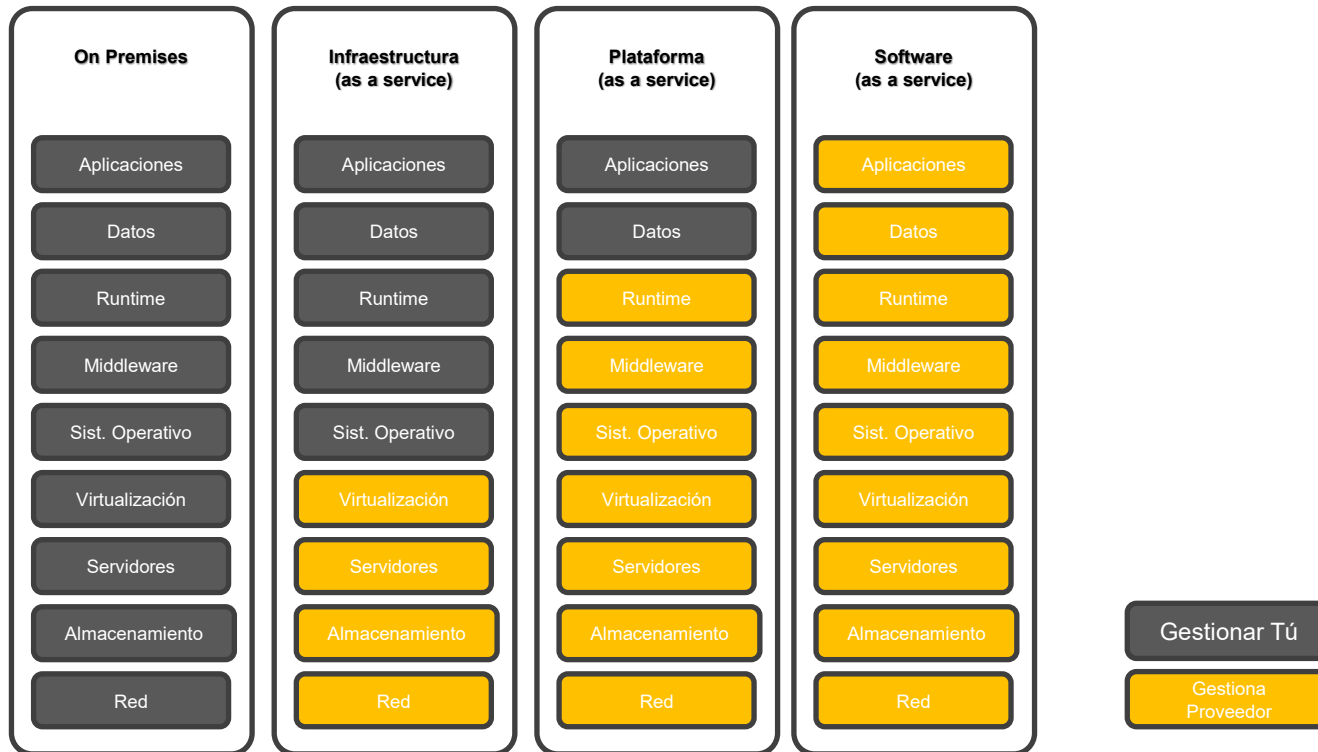


Federado



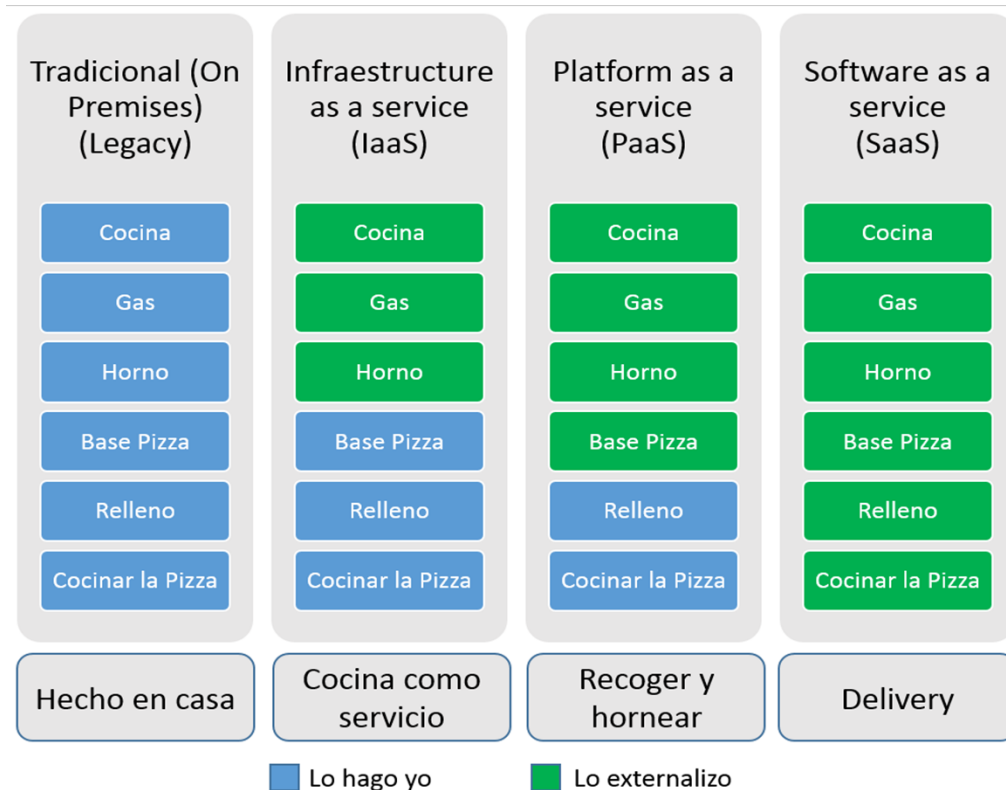
Tipos de arquitecturas de BB.DD

Virtualización / Plataformas en la nube



Tipos de arquitecturas de BB.DD

Virtualización / Plataformas en la nube

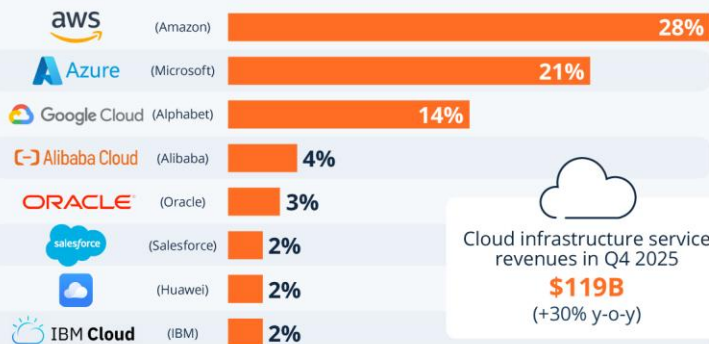


Tipos de arquitecturas de BB.DD

Virtualización / Plataformas en la nube

Big Three Hold Dominant Lead in Accelerating Cloud Market

Worldwide market share of leading cloud infrastructure service providers in Q4 2025*



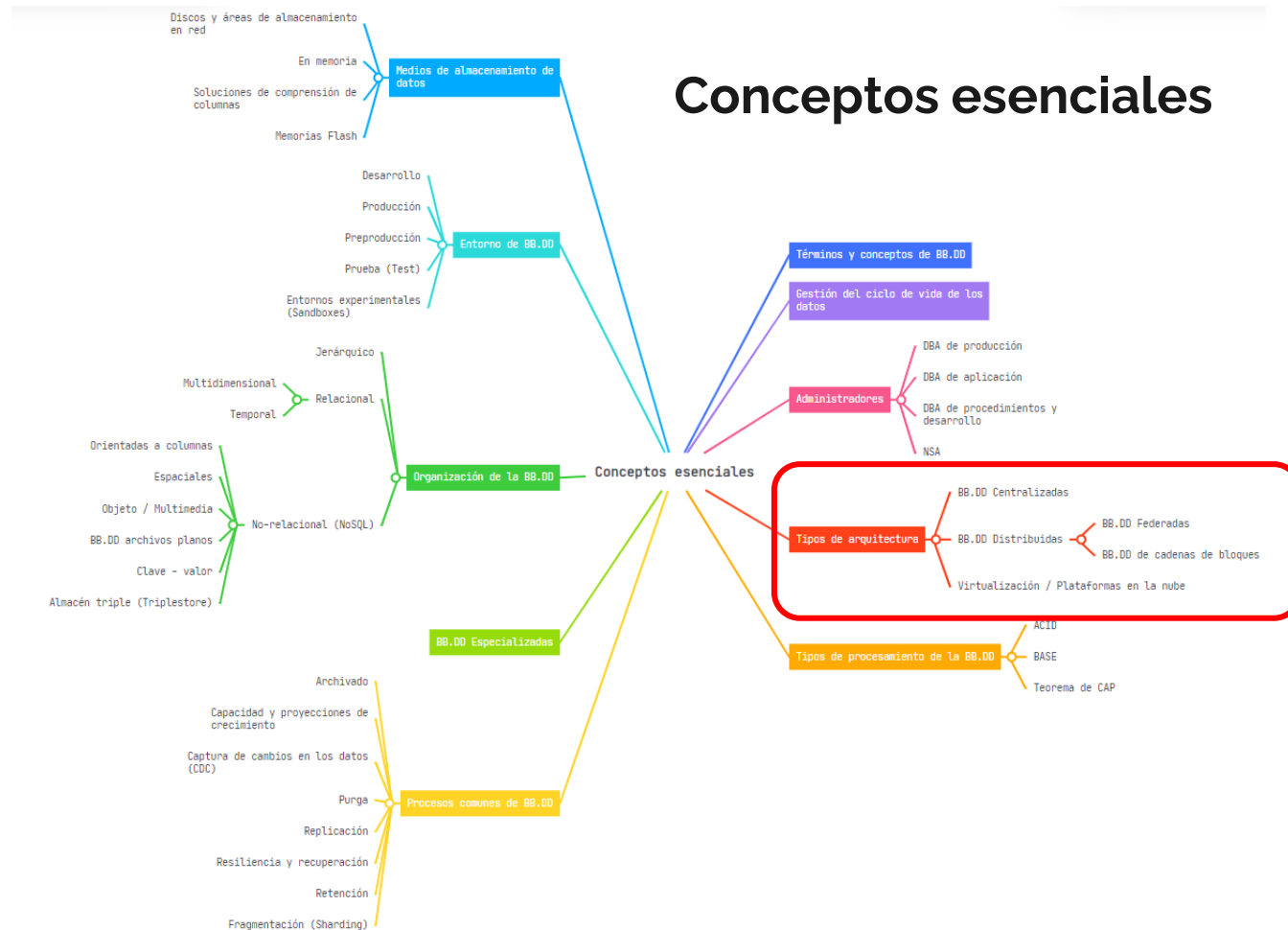
Cloud infrastructure service revenues in Q4 2025
\$119B
(+30% y-o-y)

* Includes platform as a service (PaaS) and infrastructure as a service (IaaS) as well as hosted private cloud services

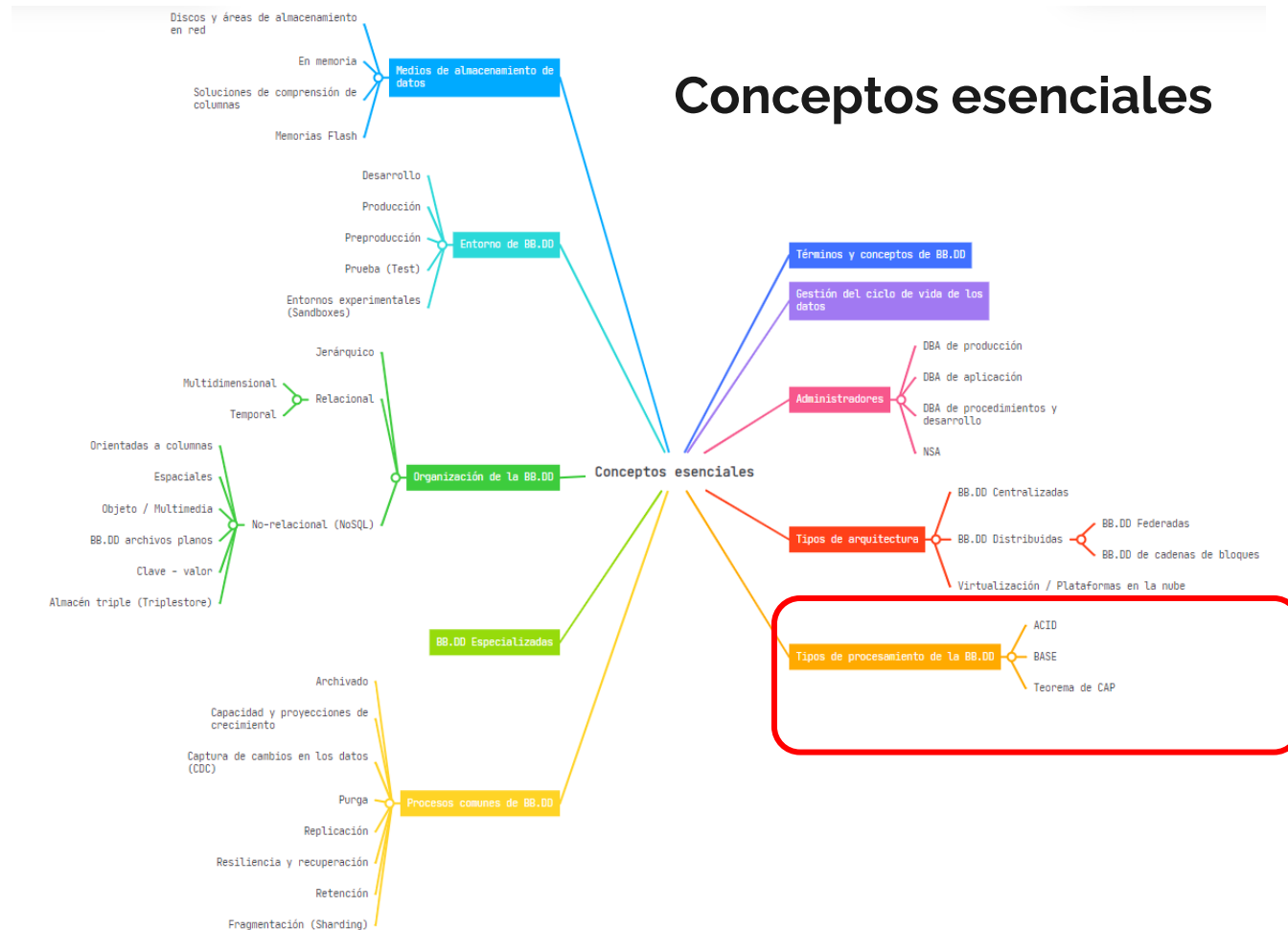
Source: Synergy Research Group

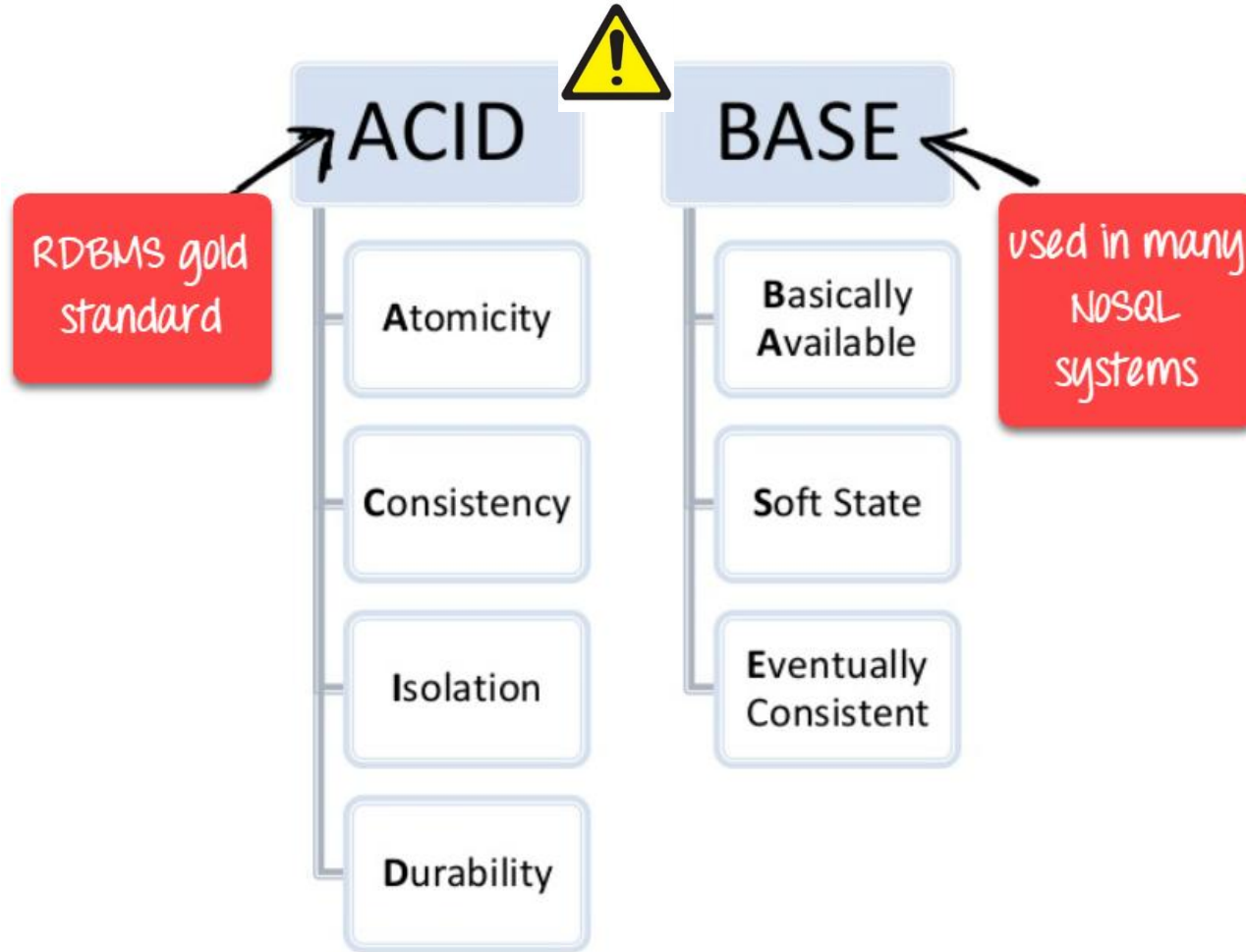


Conceptos esenciales



Conceptos esenciales





ACID

- Las propiedades **ACID** (Atomicidad, Consistencia, Aislamiento, Durabilidad) son **fundamentales para garantizar la integridad de las bases de datos, especialmente en sistemas de gestión de bases de datos relacionales.**
- Cada una de estas características ayuda a asegurar que las transacciones sean procesadas de manera confiable.

ACID

- **Atomicidad:** Se realizan todas las operaciones, o ninguna de ellas, de manera que si una parte de la transacción falla, entonces falla toda la transacción.

ACID

- **Atomicidad:** Se realizan todas las operaciones, o ninguna de ellas, de manera que si una parte de la transacción falla, entonces falla toda la transacción.
- **Consistencia:** La transacción debe cumplir todas las reglas definidas por el sistema en todo momento y debe anular las transacciones a medio completar.

ACID

- **Atomicidad:** Se realizan todas las operaciones, o ninguna de ellas, de manera que si una parte de la transacción falla, entonces falla toda la transacción.
- **Consistencia:** La transacción debe cumplir todas las reglas definidas por el sistema en todo momento y debe anular las transacciones a medio completar.
- **Aislamiento:** Cada transacción es independiente de sí misma.

ACID

- **Atomicidad:** Se realizan todas las operaciones, o ninguna de ellas, de manera que si una parte de la transacción falla, entonces falla toda la transacción.
- **Consistencia:** La transacción debe cumplir todas las reglas definidas por el sistema en todo momento y debe anular las transacciones a medio completar.
- **Aislamiento:** Cada transacción es independiente de sí misma.
- **Durabilidad:** Una vez completada, la transacción no puede deshacerse.

BASE

- **Describe las propiedades de los sistemas de bases de datos que no siguen el estricto modelo ACID tradicional**, generalmente asociado con bases de datos relacionales (NoSQL).
- Las propiedades BASE son típicamente utilizadas por sistemas de bases de datos NoSQL para **manejar grandes volúmenes de datos distribuidos de manera más flexible y eficiente** en entornos donde el rendimiento y la escalabilidad son más críticos que la consistencia inmediata.

BASE

- **Básicamente disponible:** El sistema garantiza cierto nivel de disponibilidad de los datos incluso cuando hay fallos en los nodos. Los datos pueden ser obsoletos, pero el sistema seguirá dando y aceptando respuestas.

BASE

- **Básicamente disponible:** El sistema garantiza cierto nivel de disponibilidad de los datos incluso cuando hay fallos en los nodos. Los datos pueden ser obsoletos, pero el sistema seguirá dando y aceptando respuestas.
- **Estado blando:** Los datos están en un estado de flujo constante; aunque se pueda dar una respuesta, no se garantiza que los datos sean actuales.

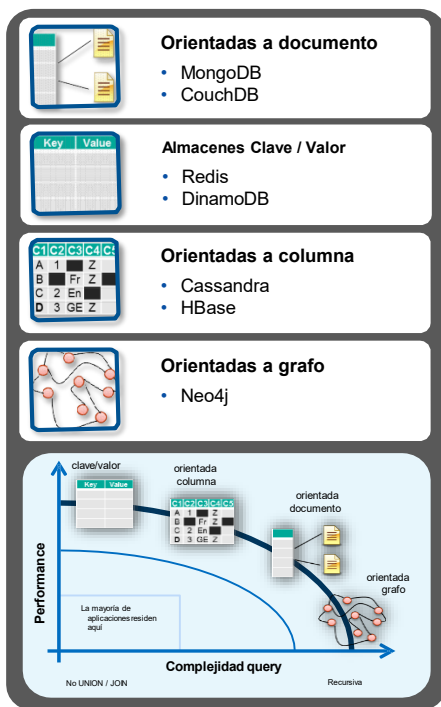
BASE

- **Básicamente disponible:** El sistema garantiza cierto nivel de disponibilidad de los datos incluso cuando hay fallos en los nodos. Los datos pueden ser obsoletos, pero el sistema seguirá dando y aceptando respuestas.
- **Estado blando:** Los datos están en un estado de flujo constante; aunque se pueda dar una respuesta, no se garantiza que los datos sean actuales.
- **Consistencia eventual:** Los datos acabarán siendo consistentes en todos los nodos y en todas las bases de datos, pero **no todas las transacciones serán consistentes en cada momento.**

ACID vs BASE

Aspecto	ACID	BASE
Manejo de tipos de datos (estructura de datos)	Esquema debe existir	Dinámica
	Estructura de la tabla existe	Ajustar sobre la marcha
	Datos de columna de un tipo determinado	Almacenar datos diferentes
Consistencia	Fuerte consistencia disponible	Fuerte, Eventual, o ninguna
Enfoque de procesamiento	Transaccional	Almacenamiento de valores clave (<i>Key-Value</i>)
Enfoque de procesamiento	Fila o columna	Almacenes de toda la columna
Historia	Almacenamiento de aplicación de la década de 1970	Almacenamiento no estructurado, década de 2000
Escala	Dependiente del producto	Separa automáticamente los datos a través de servidores
Origen	Mezcla	<i>Open-source</i>
Transacción	Sí	Posible

BB.DD NoSQL



seres.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesión: médico;
  estado_civil: soltera;
}

{
  identificador: gtr5678hh;
  nombre: C3PO;
  raza: droide;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```

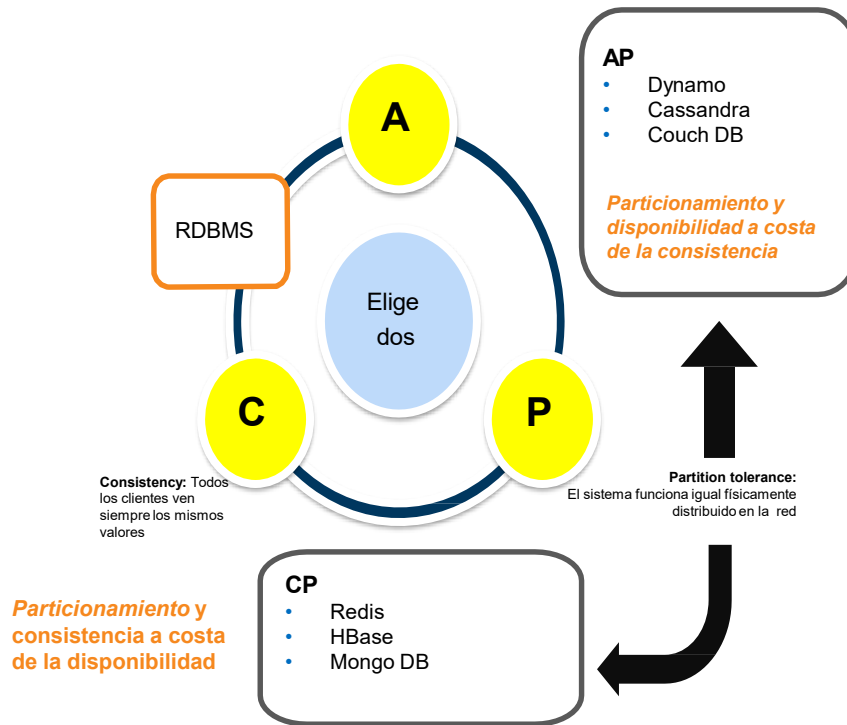
Cap Theorem (Brewer's Theorem)



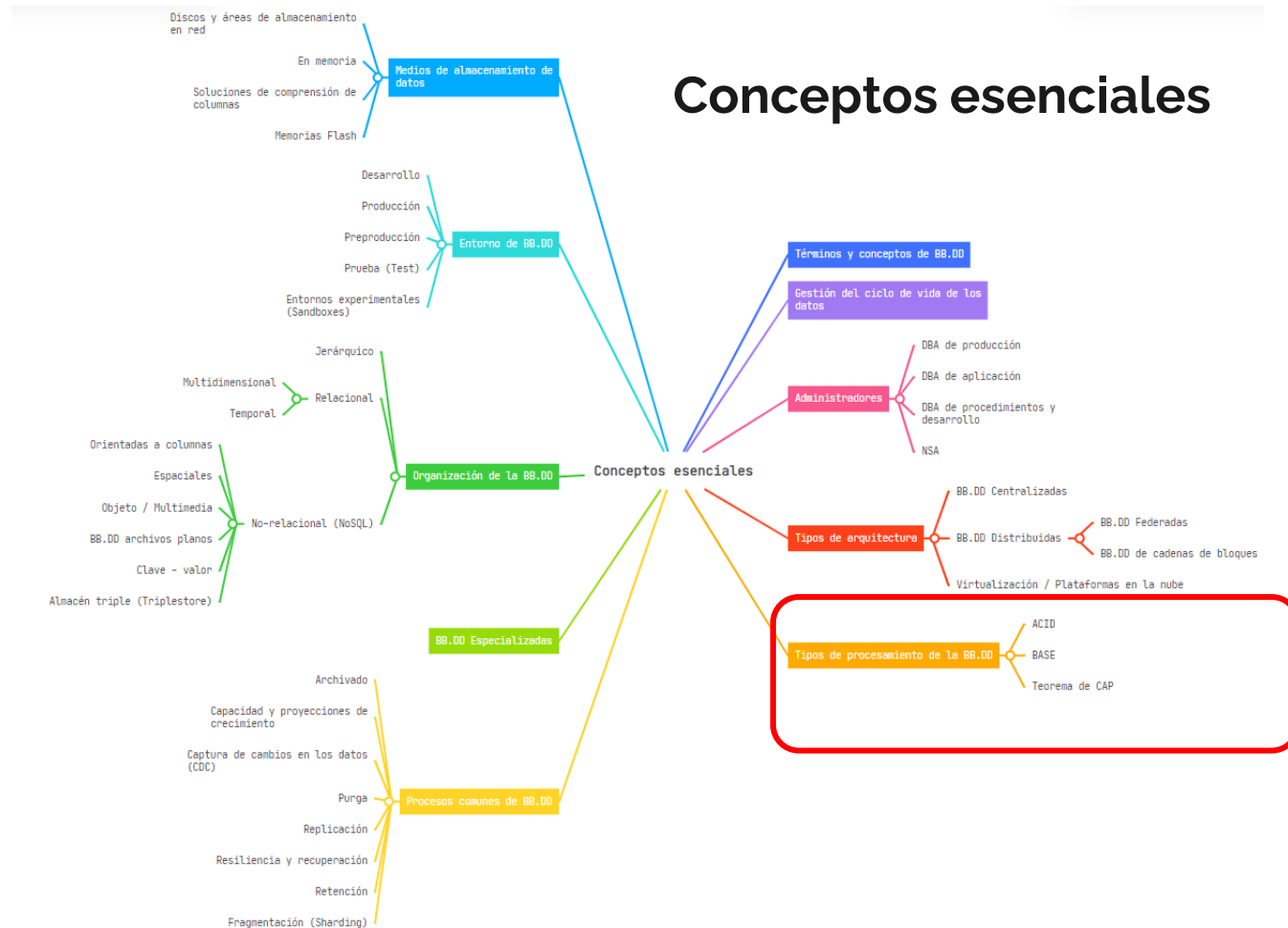
- BASE vs. ACID
- Teorema CAP

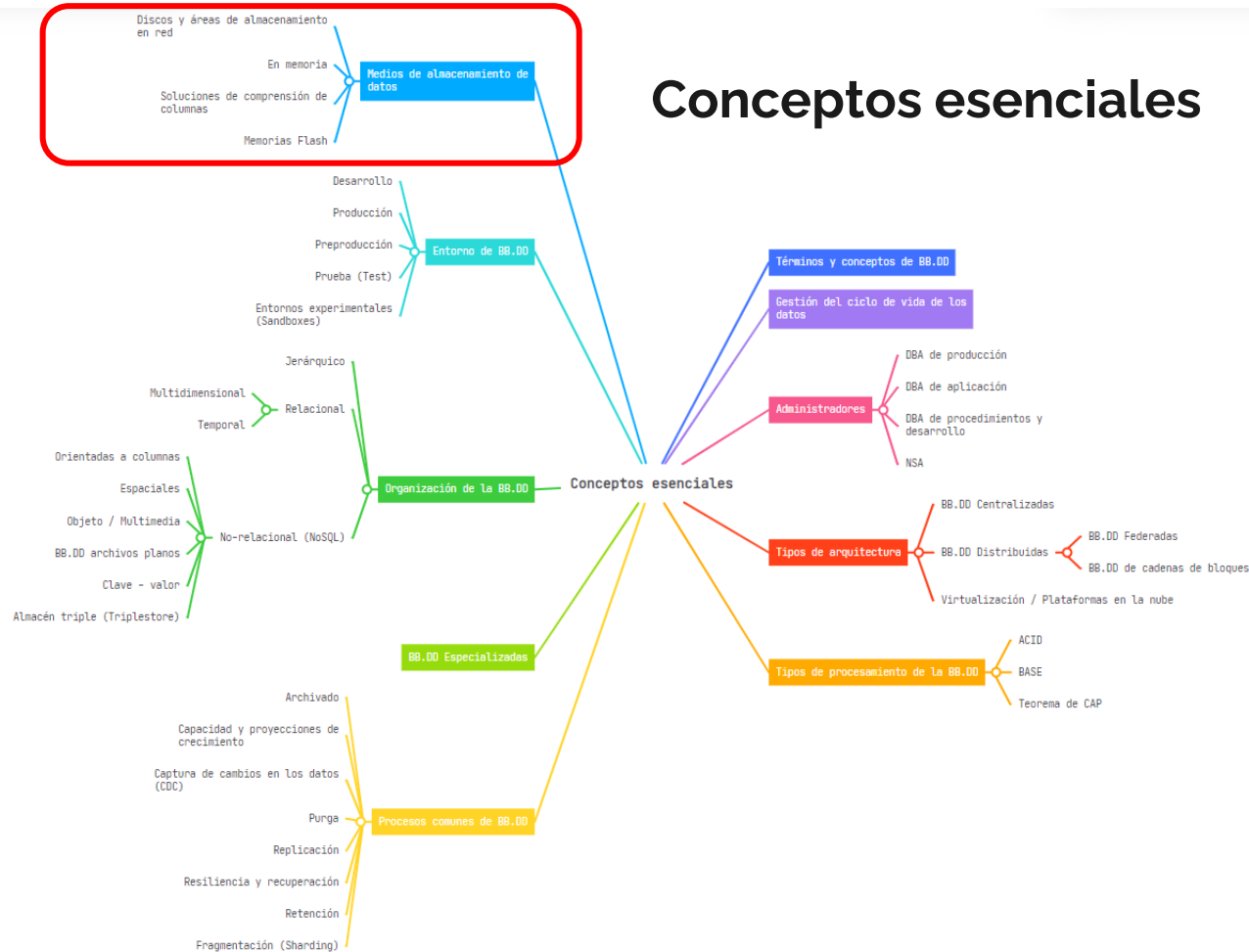
Availability:

Cualquier cliente puede leer o escribir en cualquier momento

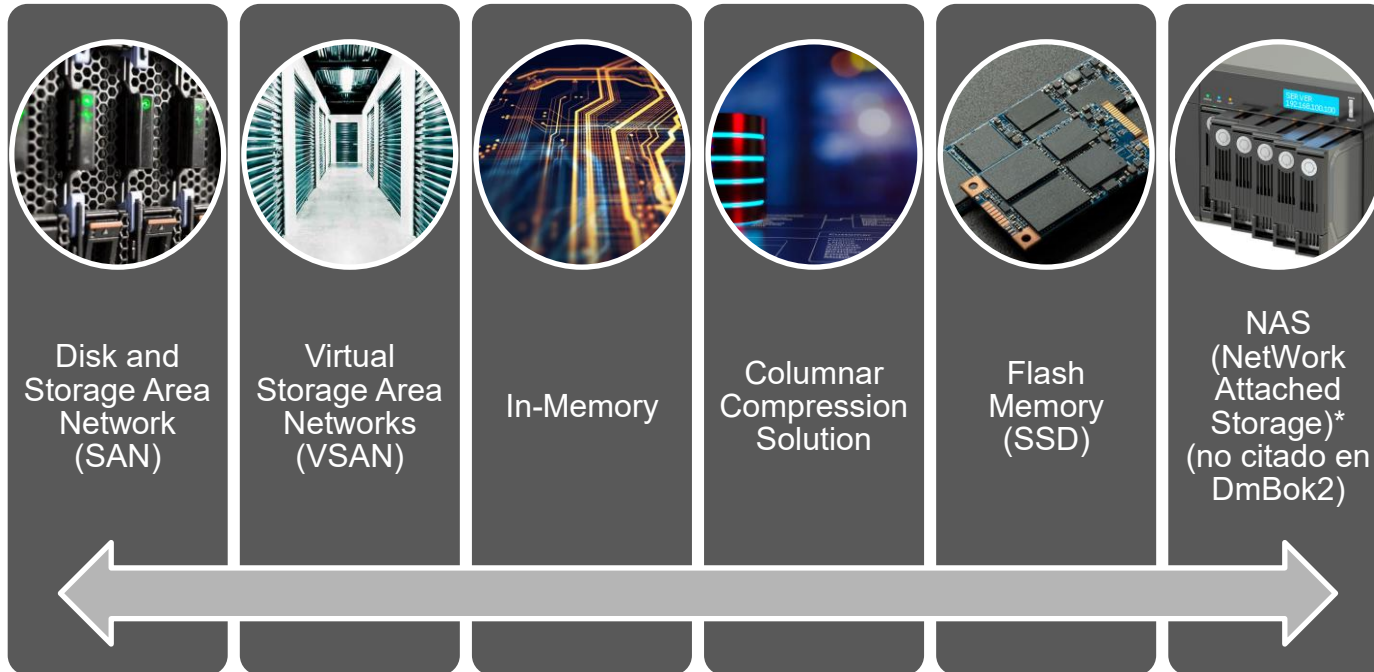


Conceptos esenciales

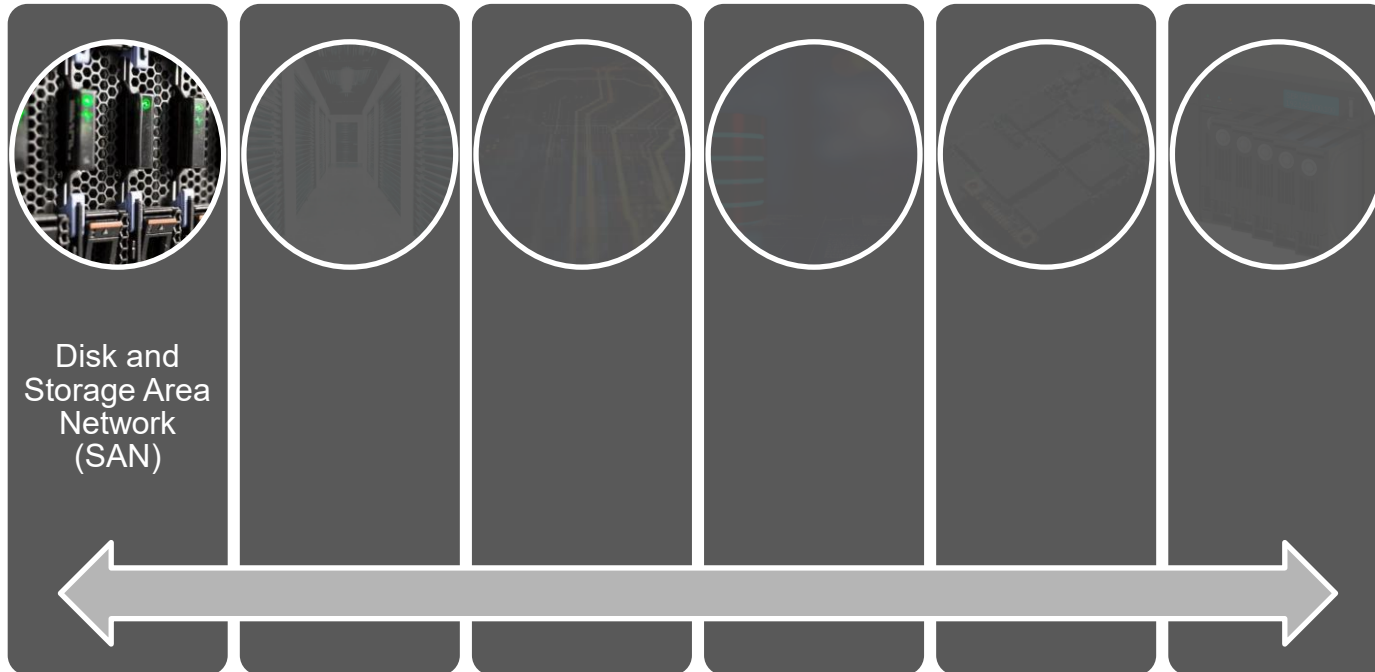




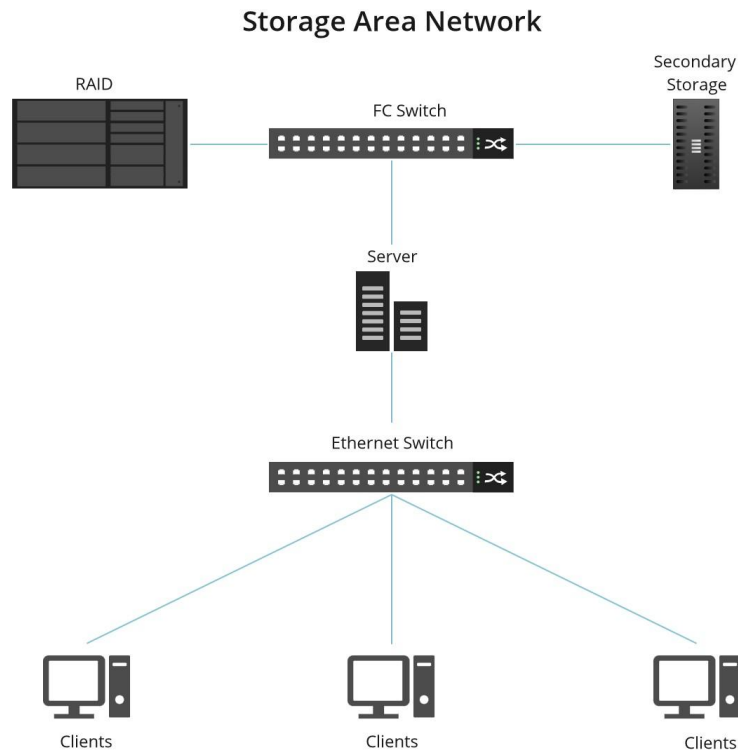
Data Storage Media



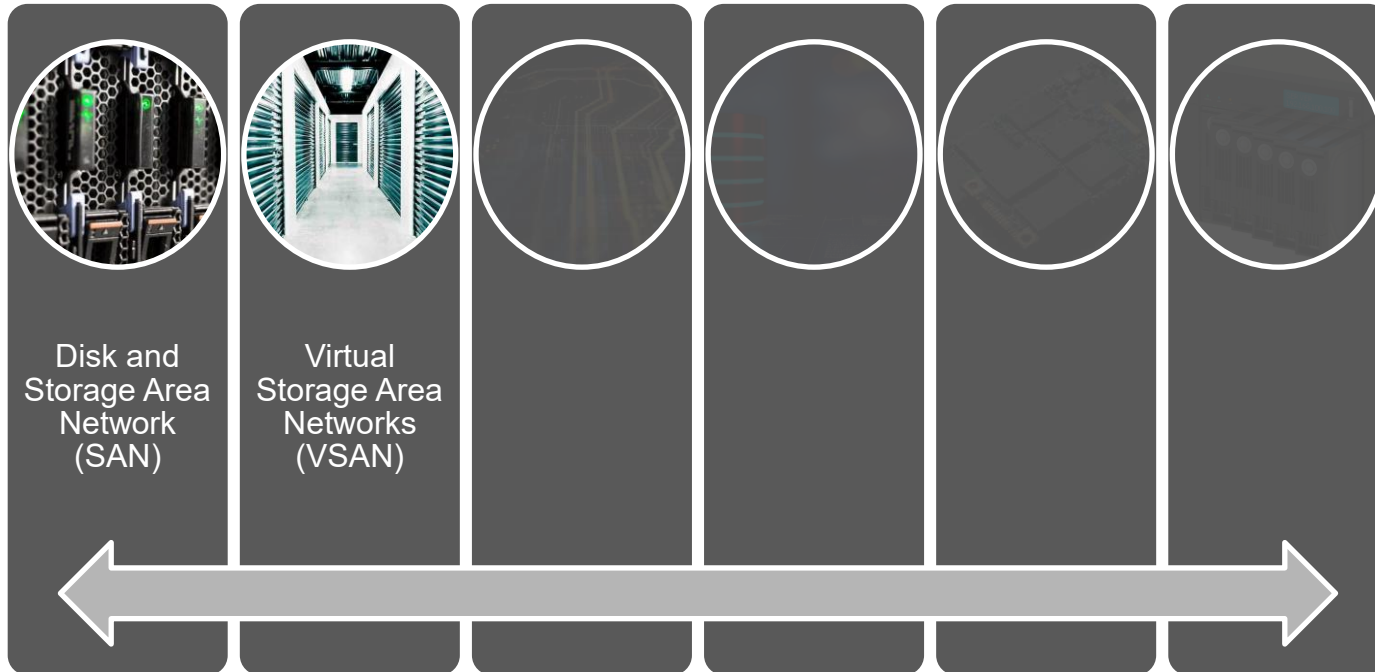
Data Storage Media



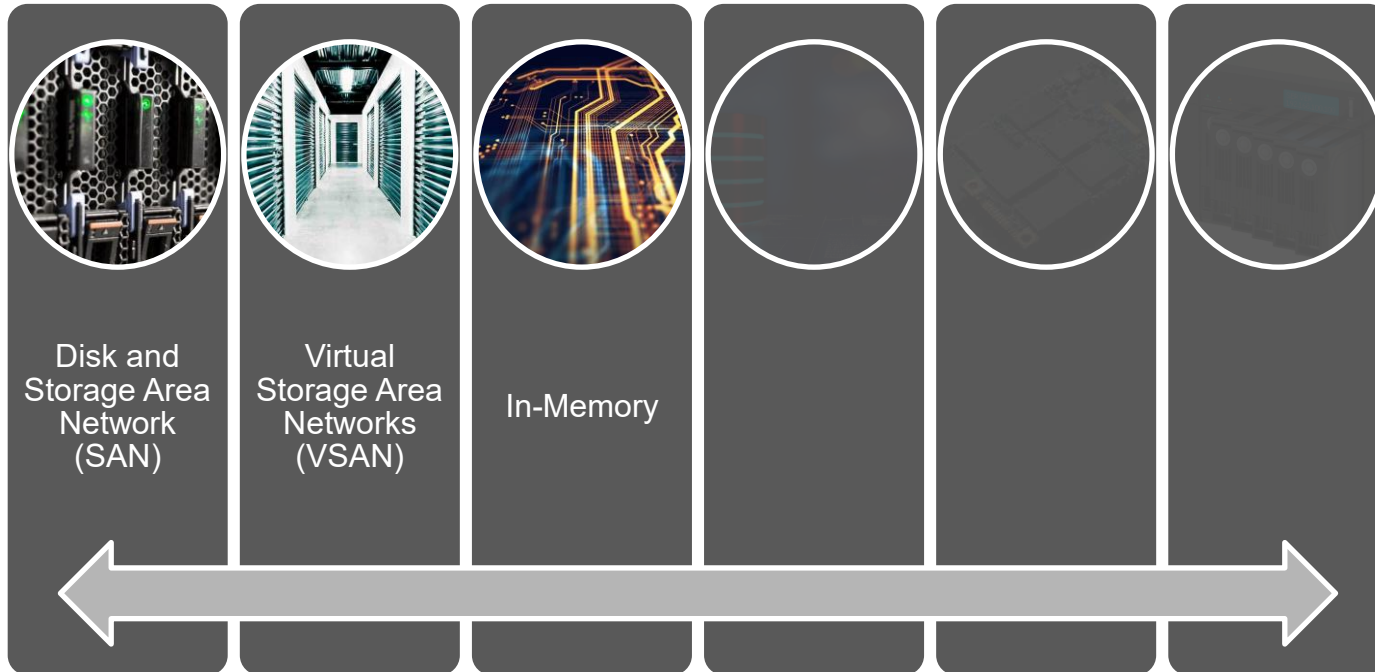
SAN



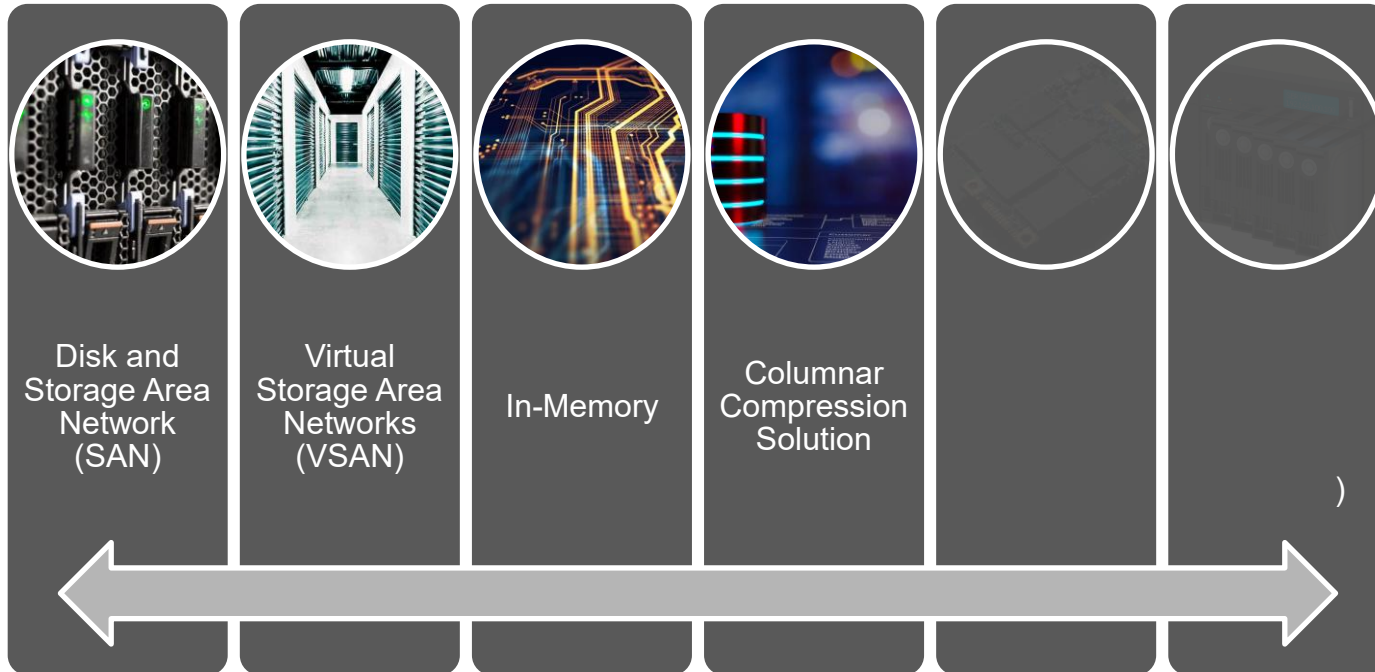
Data Storage Media



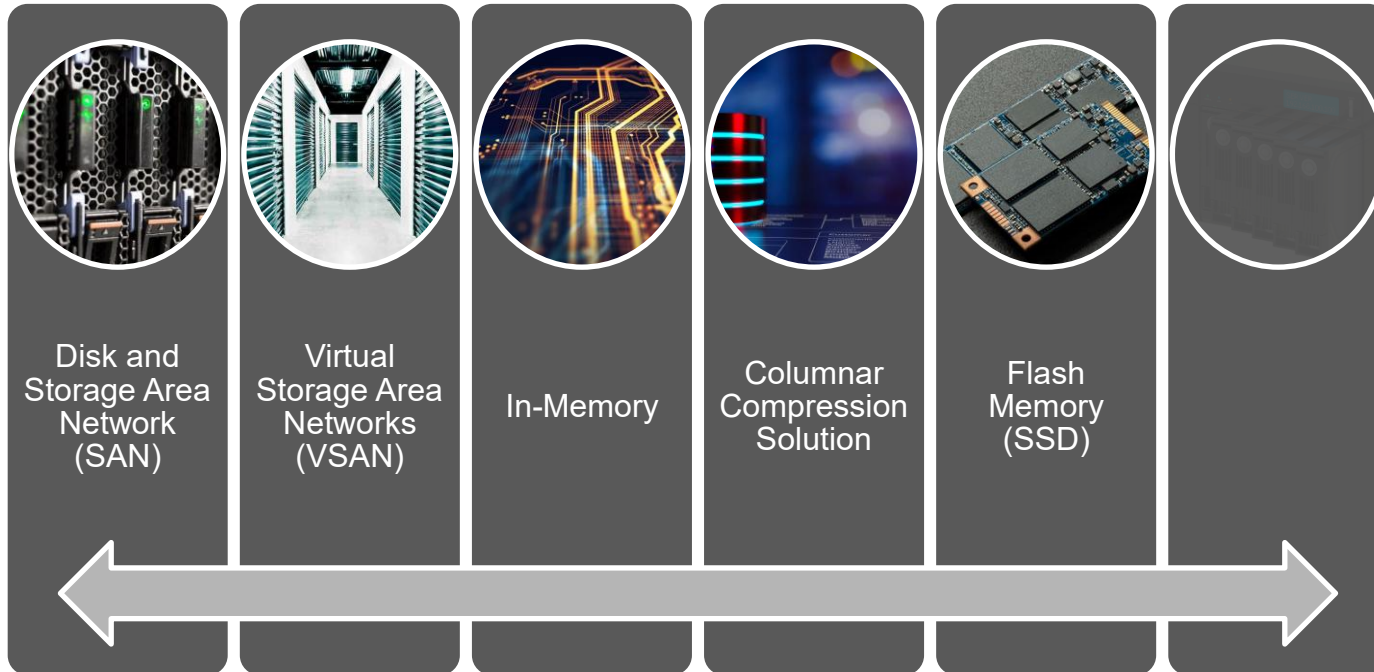
Data Storage Media



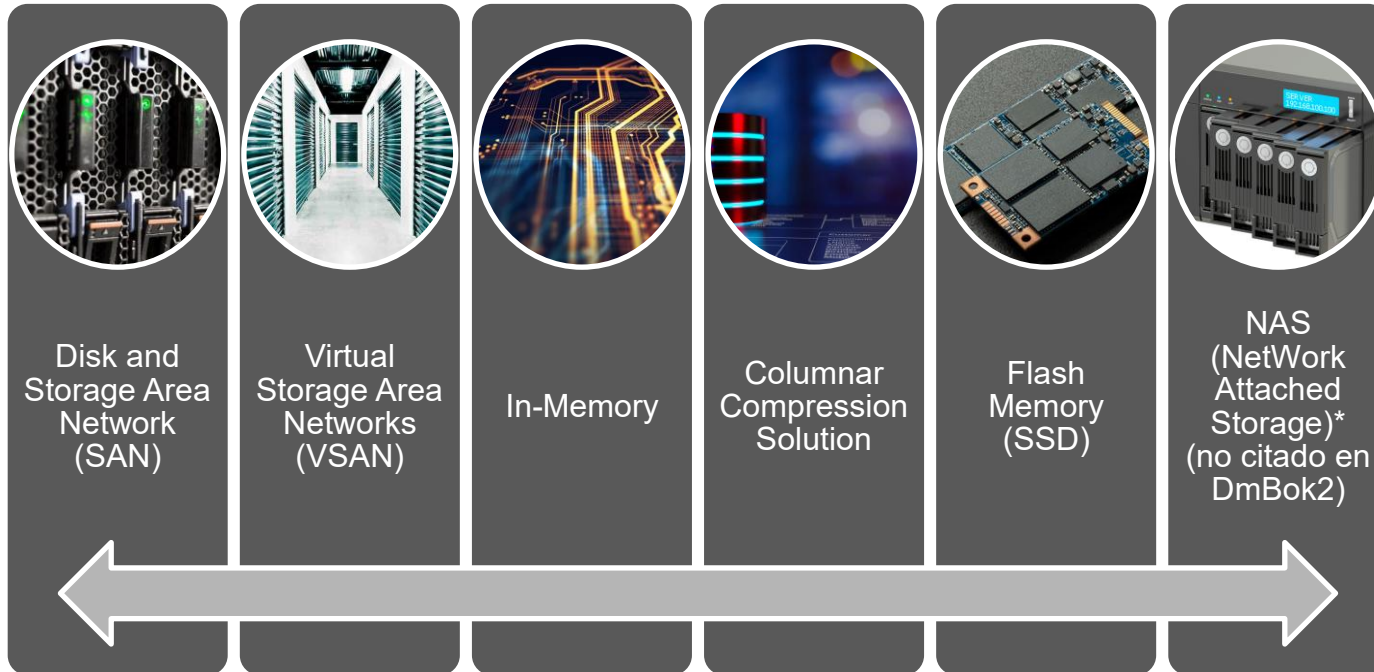
Data Storage Media

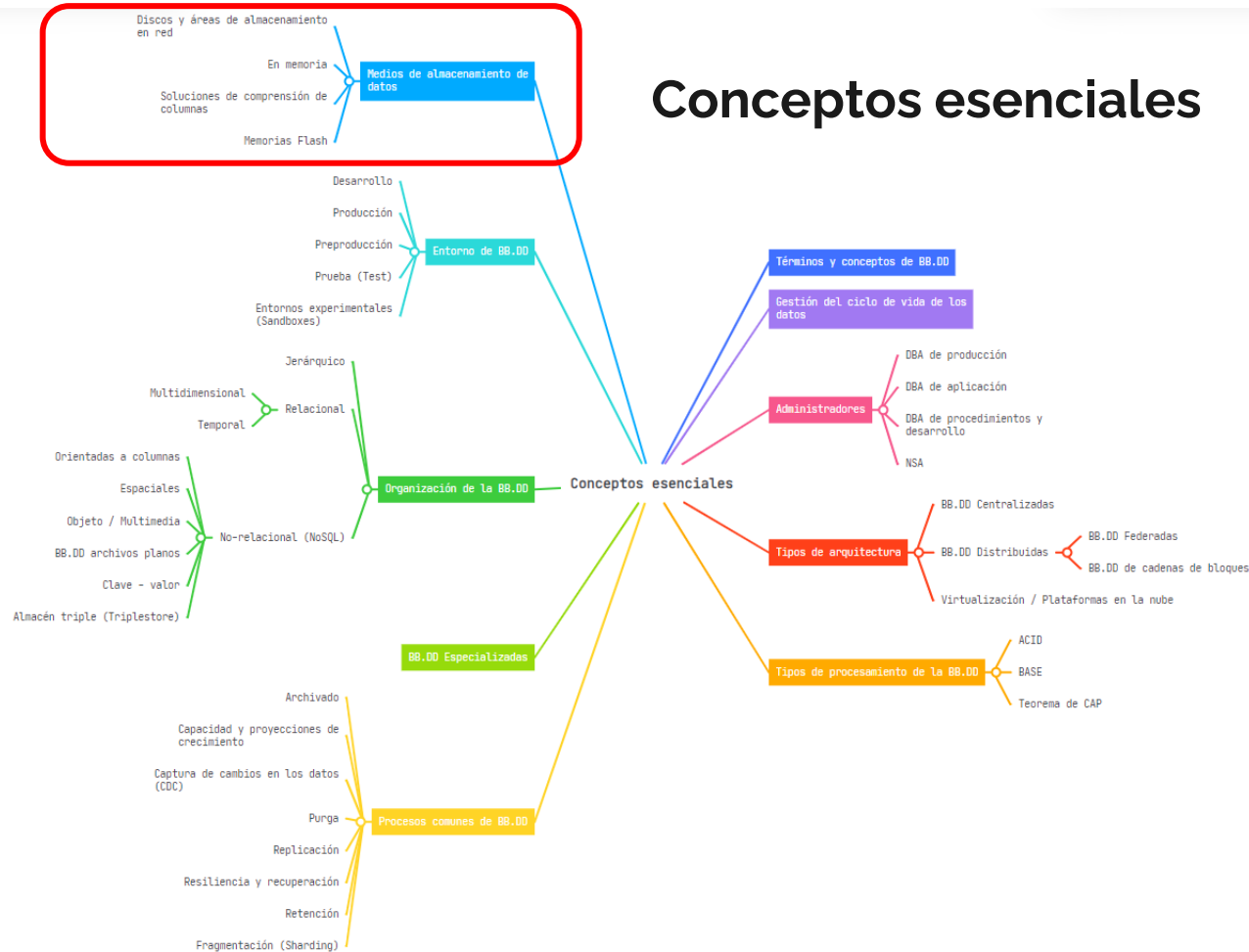


Data Storage Media



Data Storage Media





Conceptos esenciales

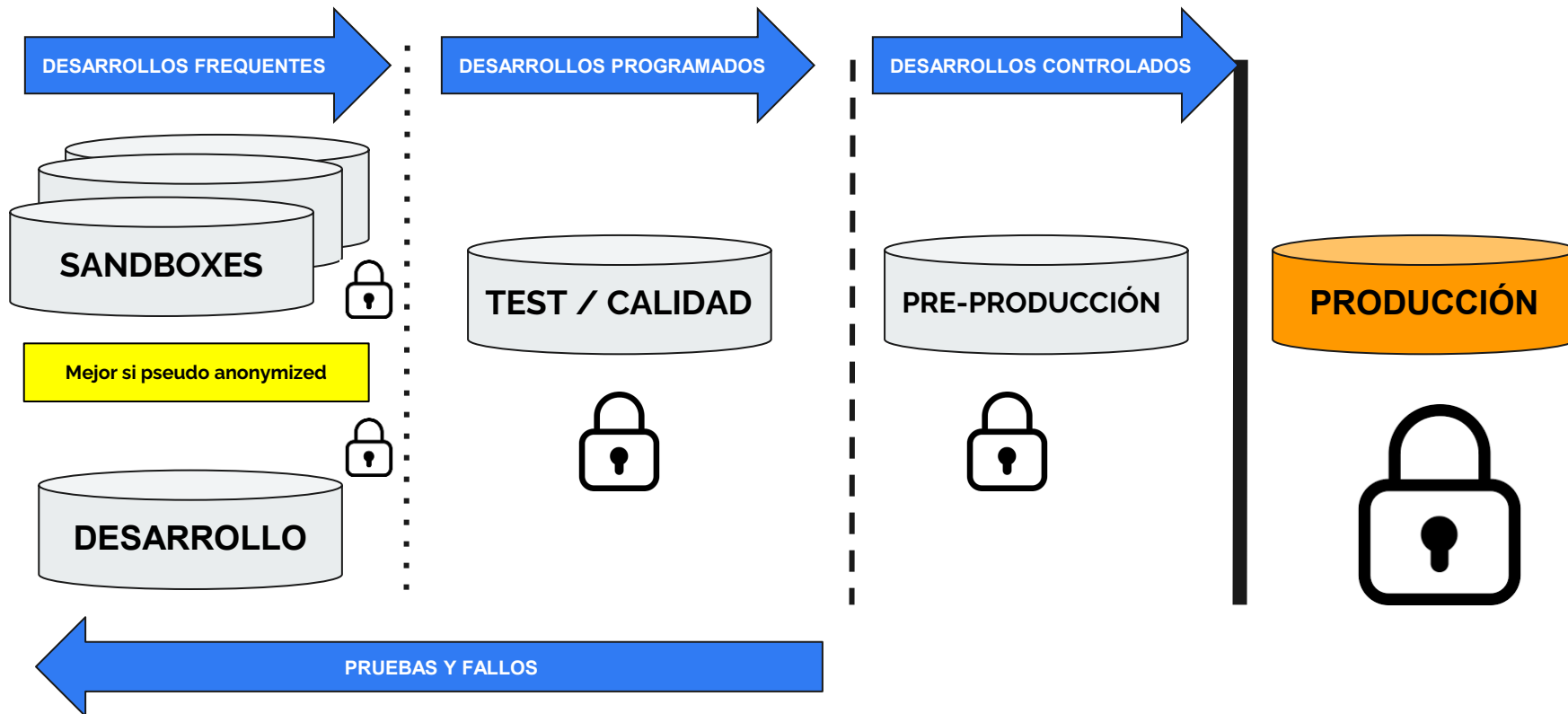


Ambientes de Base de Datos

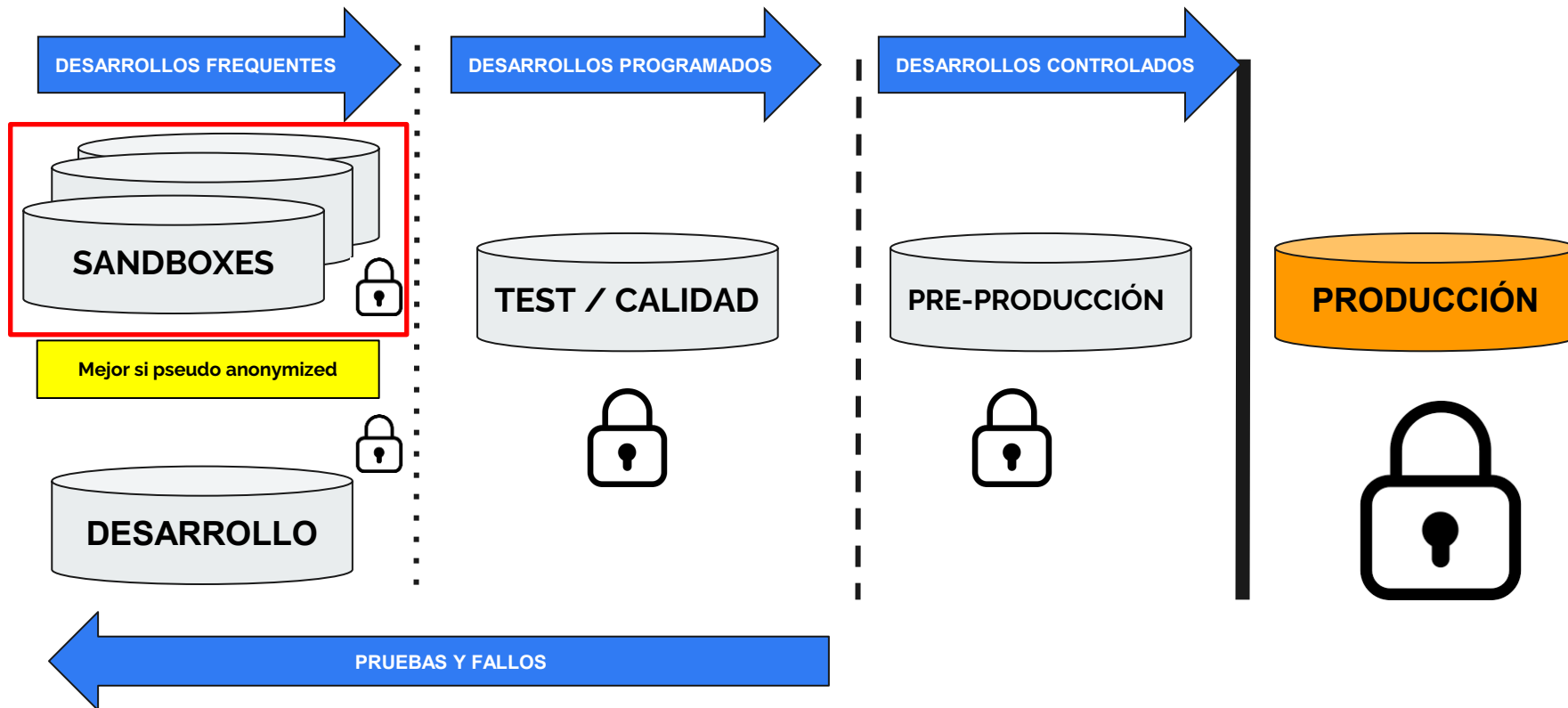
- Producción (el que no se toca)
- Pre-Producción (completo de flujos)
- Development (para desarrollo puede ser incompleto)
- Test (para lanzar tests de calidad y otros procesos)
- Sandbox (para experimentar con todo o parte de los datos)



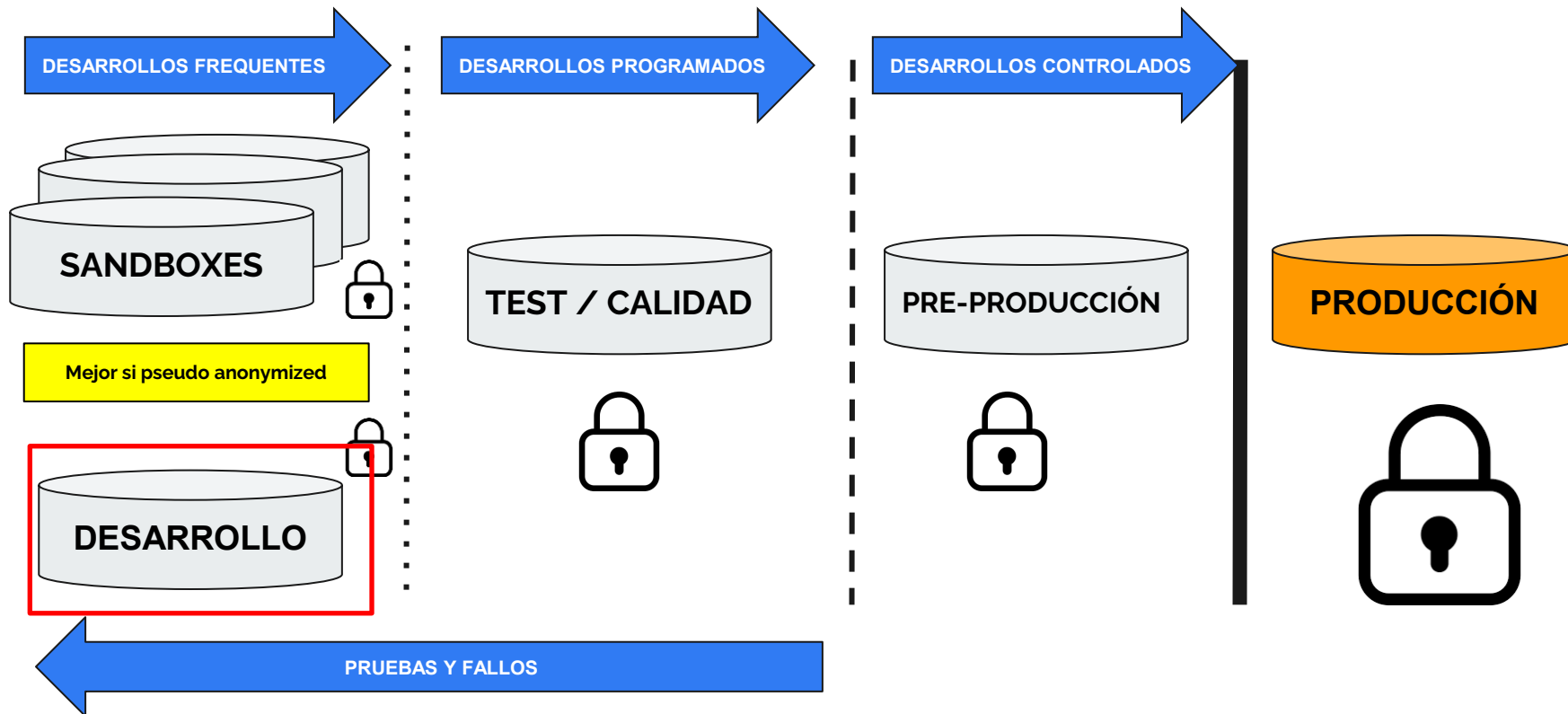
Entornos de base de datos



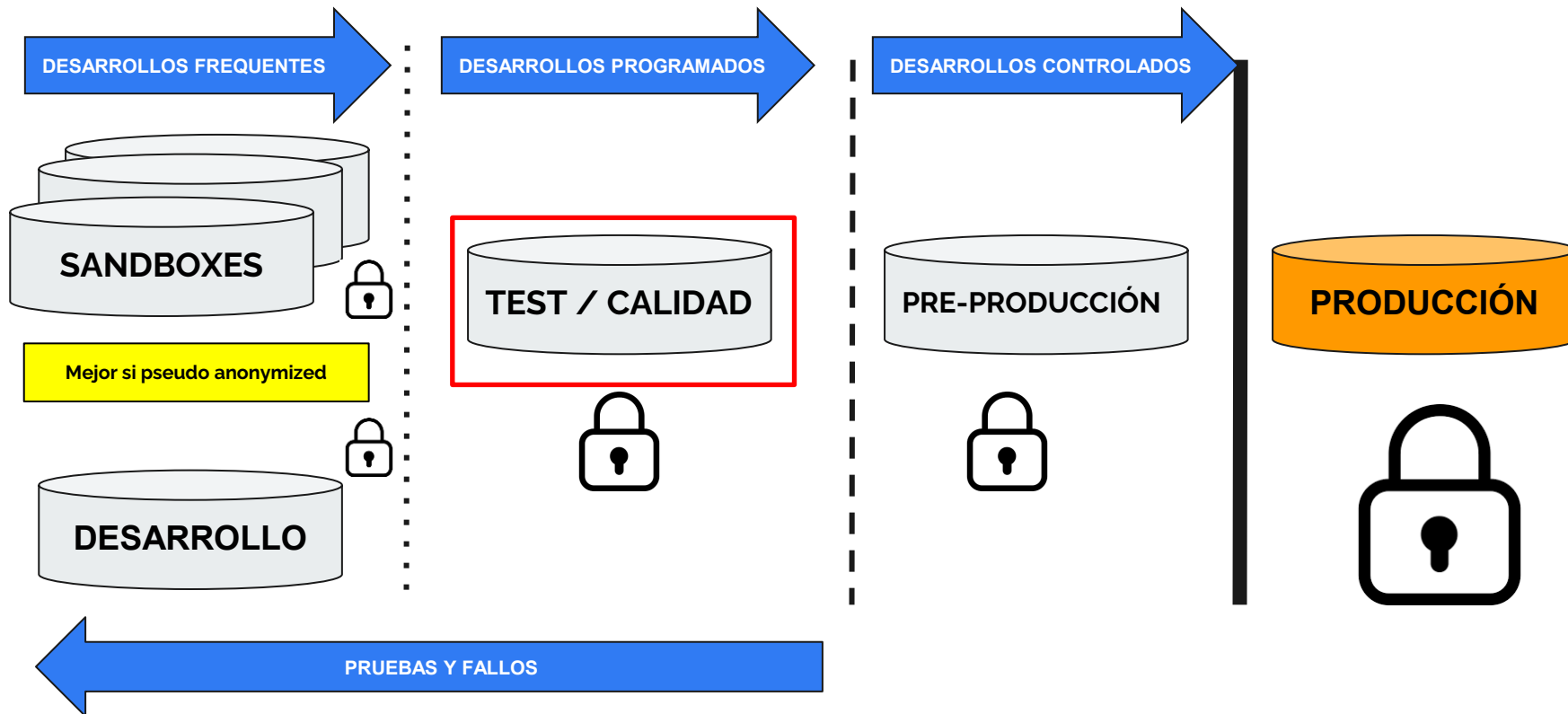
Entornos de base de datos



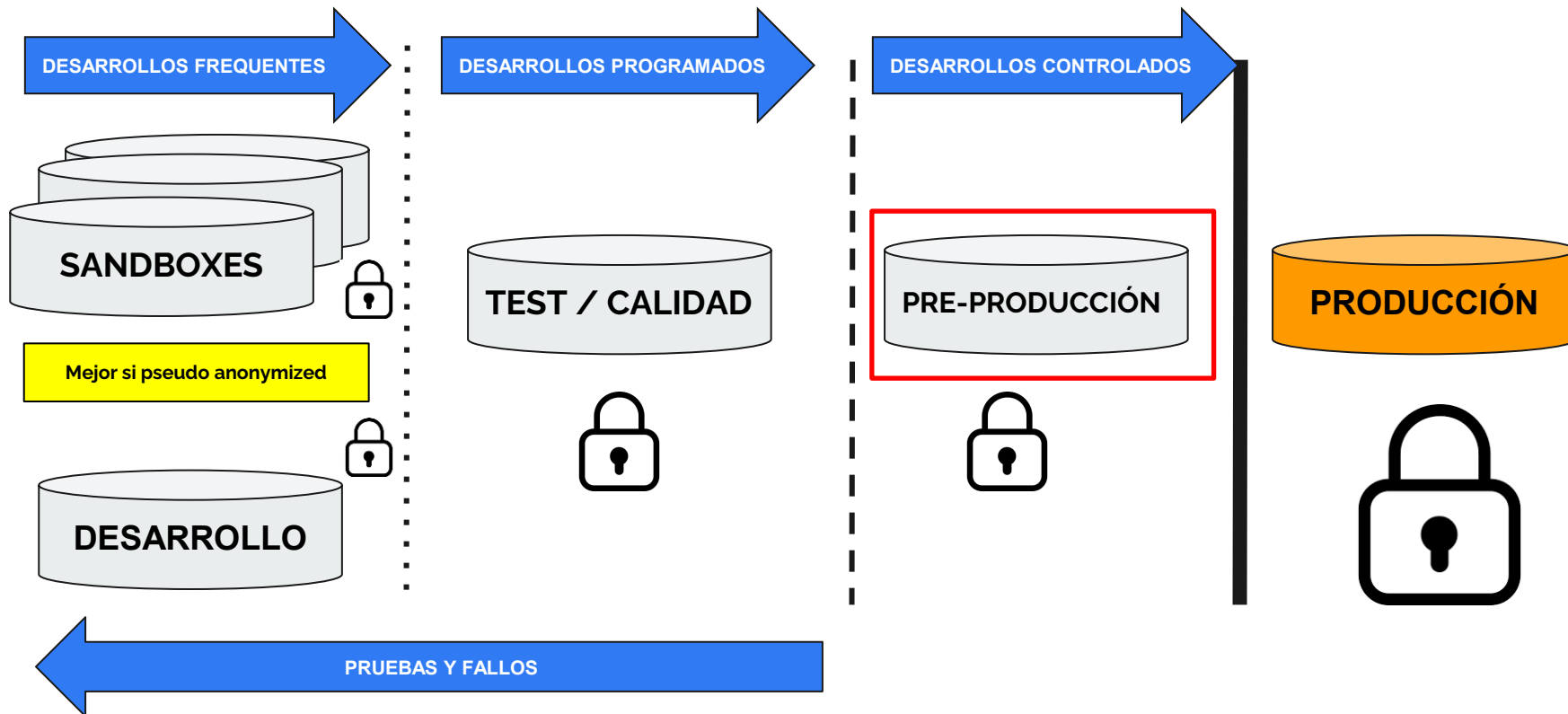
Entornos de base de datos



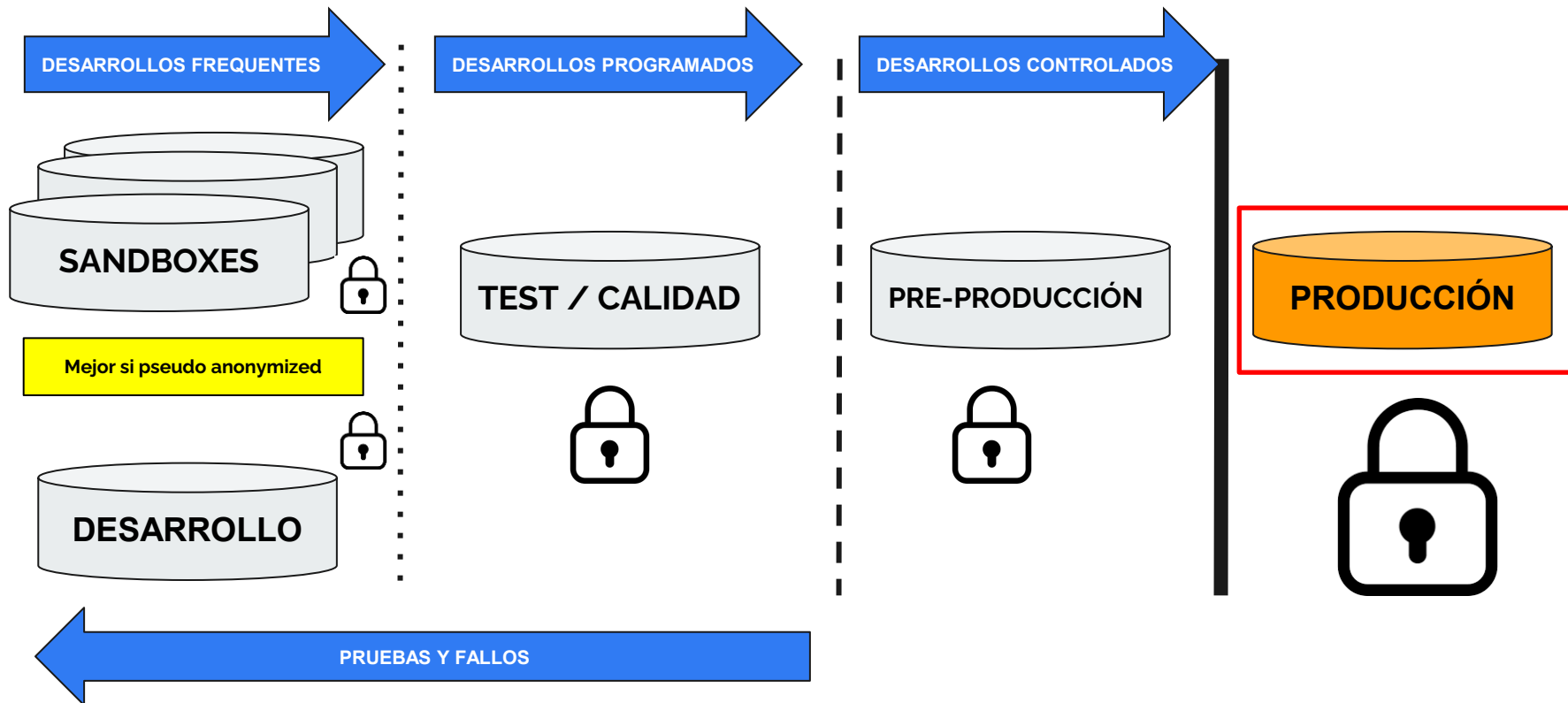
Entornos de base de datos



Entornos de base de datos



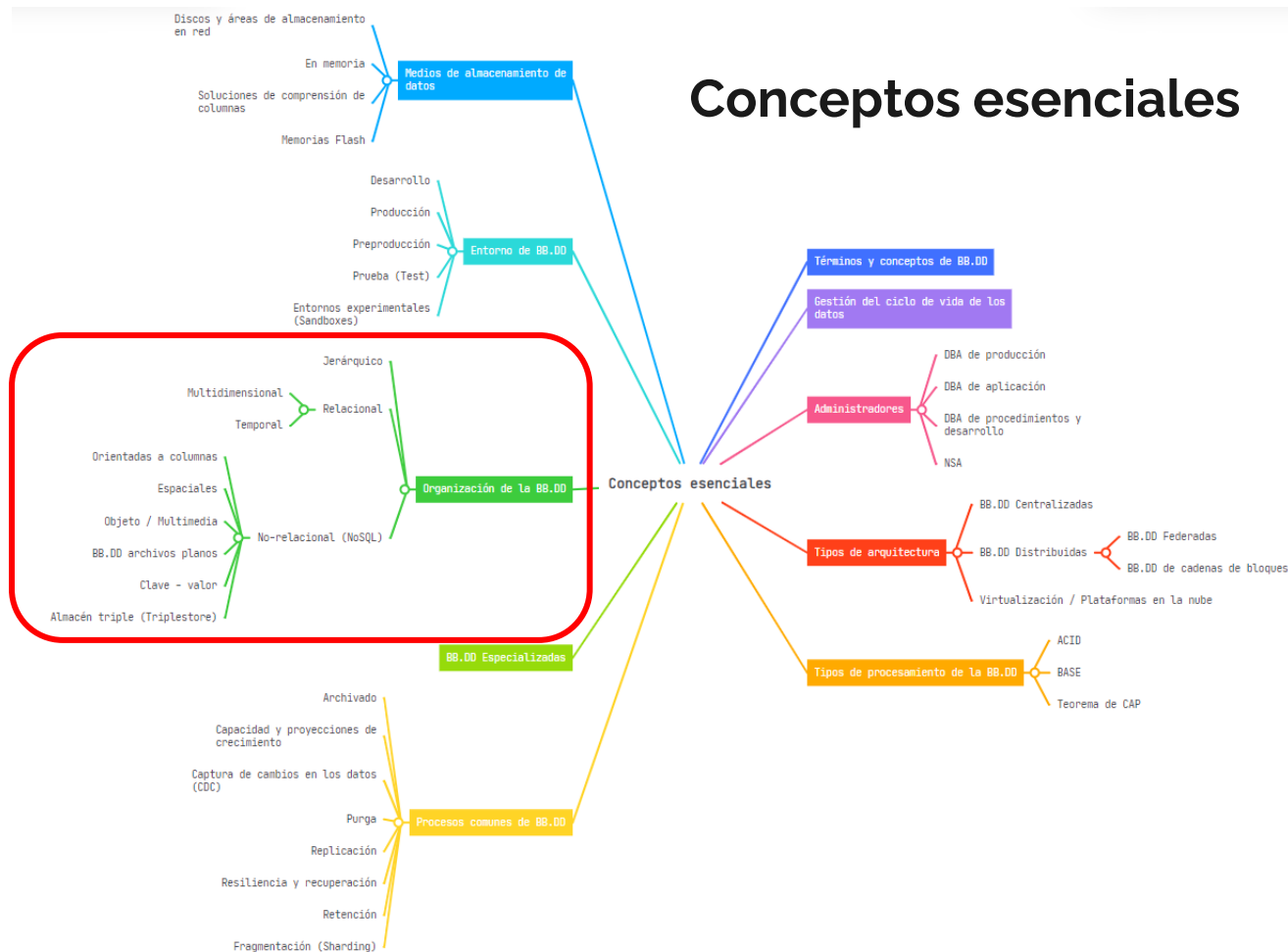
Entornos de base de datos



Conceptos esenciales



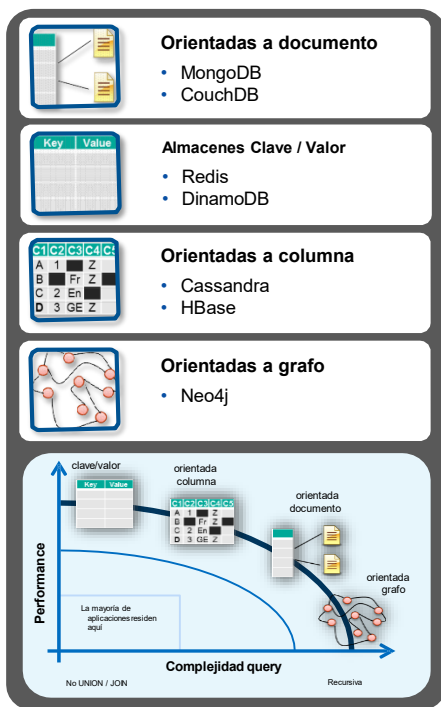
Conceptos esenciales



Organización de bases de datos

- Hierarchical
- Relational
- Multidimensional
- Temporal
- Non-Relational
- Column oriented
- Spatial
- Object / Multimedia
- Flat File Database
- Key-Value Pair
- Triplestore

BB.DD NoSQL



seres.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesión: médico;
  estado_civil: soltera;
}

{
  identificador: gtr5678hh;
  nombre: C3PO;
  raza: droide;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```

BB.DD NoSQL. Orientadas a documento



Orientadas a documento

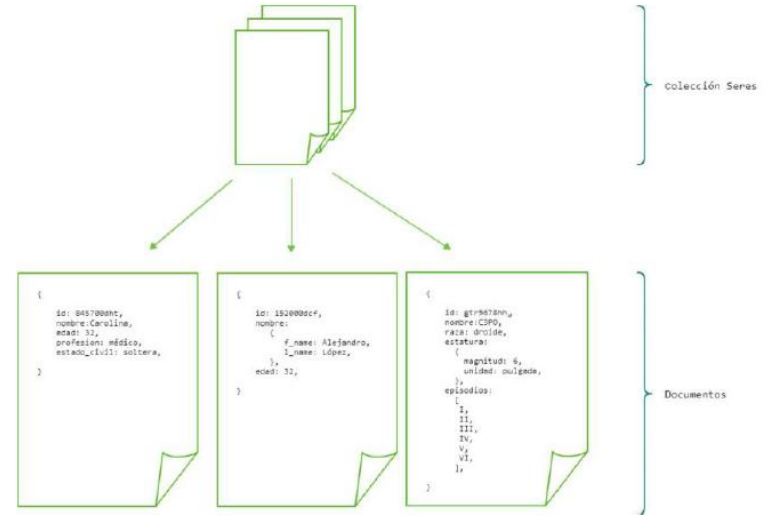
- MongoDB
- CouchDB

series.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesión: médico;
  estado_civil: soltera;
}

{
  identificador: gtn5678hh;
  nombre: C3P0;
  raza: droid;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```



BB.DD NoSQL . Clave-valor

Key | Value

Almacenes Clave / Valor

- Redis
- DinamoDB

series.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesión: médico;
  estado_civil: soltera;
}

{
  identificador: gtr5678hh;
  nombre: C3P0;
  raza: droide;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```

152000dcf	nombre: Alejandro López	edad: 32		
845700dht	nombre: Carolina	edad: 52	profesión: médico	estado_civil: soltera
gtr5678hh	nombre: C3P0	raza: droide	estatura: 6 pulgadas	episodios: I,II, III, IV, V, VI

BB.DD NoSQL. Orientadas a columna



Orientadas a columna

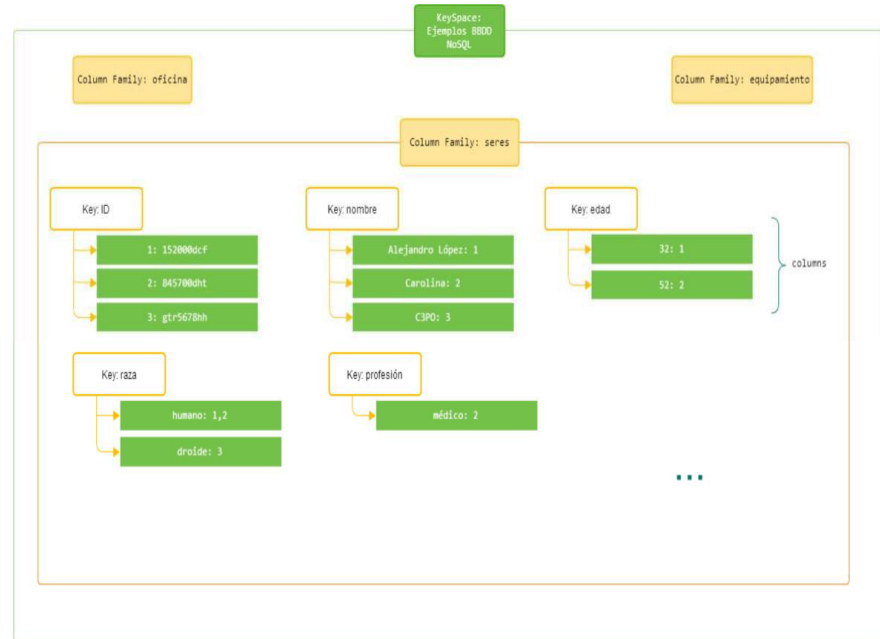
- Cassandra
- HBase

seres.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesión: médico;
  estado_civil: soltera;
}

{
  identificador: gtr5678hh;
  nombre: C3P0;
  raza: droide;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```



BB.DD NoSQL. Orientadas a grafos



Orientadas a grafo

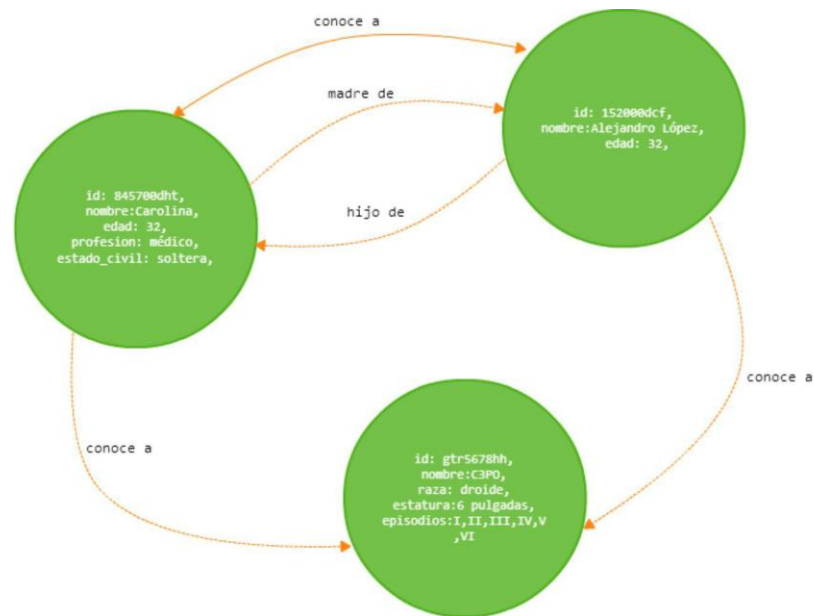
- Neo4j

seres.txt

```
{
  identificador: 152000dcf;
  nombre: Alejandro López;
  edad: 32;
}

{
  identificador: 845700dht;
  nombre: Carolina;
  edad: 52;
  profesion: médico;
  estado_civil: soltera;
}

{
  identificador: gtr5678hh;
  nombre: C3P0;
  raza: droide;
  estatura: 6 pulgadas;
  episodios: I, II, III, IV, V, VI;
}
```





#GCPsketchnotes

07.10.2021



@PVERGADIA



THECLOUDGIRL.DEV

Which Database should I use?



RELATIONAL



Cloud SQL

Managed MySQL,
PostgreSQL,
SQL Server



Cloud Spanner

Cloud-native with
large scale,
consistency,
99.999% availability



Bare Metal

Lift and shift
Oracle workloads
to Google Cloud

Good For:

General purpose
SQL DB

RDBMS+ scale,
HA, HTAP

RDBMS+ scale,
HA, HTAP

Use Case:



Web
frameworks



Gaming



Legacy
applications



ERP



Global financial
ledger



Data center
retirement



CRM



Ecommerce
and web



Supply chain/
inventory
management



SaaS
application

NON-RELATIONAL (NO SQL)

DOCUMENT

KEY VALUE



Firestore

Cloud Native, serverless,
NoSQL document database,
backend-as-a-service,
global strong consistency,
99.999% SLA



Cloud Bigtable

Cloud-native NoSQL
wide-column store
for large scale,
low-latency workloads

Good For:

Large scale, complex
hierarchical data

Heavy read + write, events

Use Case:



Mobile/web/
IoT applications



Personalization



Real-time sync



Adtech



Offline sync



Recommendation
engines



Personalized apps



Fraud detection

IN MEMORY



Memory Store

Fully managed Redis and
Memcached for sub-millisecond
data access

Good For:

In-memory and Key-value store

Use Case:



Caching



Session store



Gaming



Personalization



Leaderboard



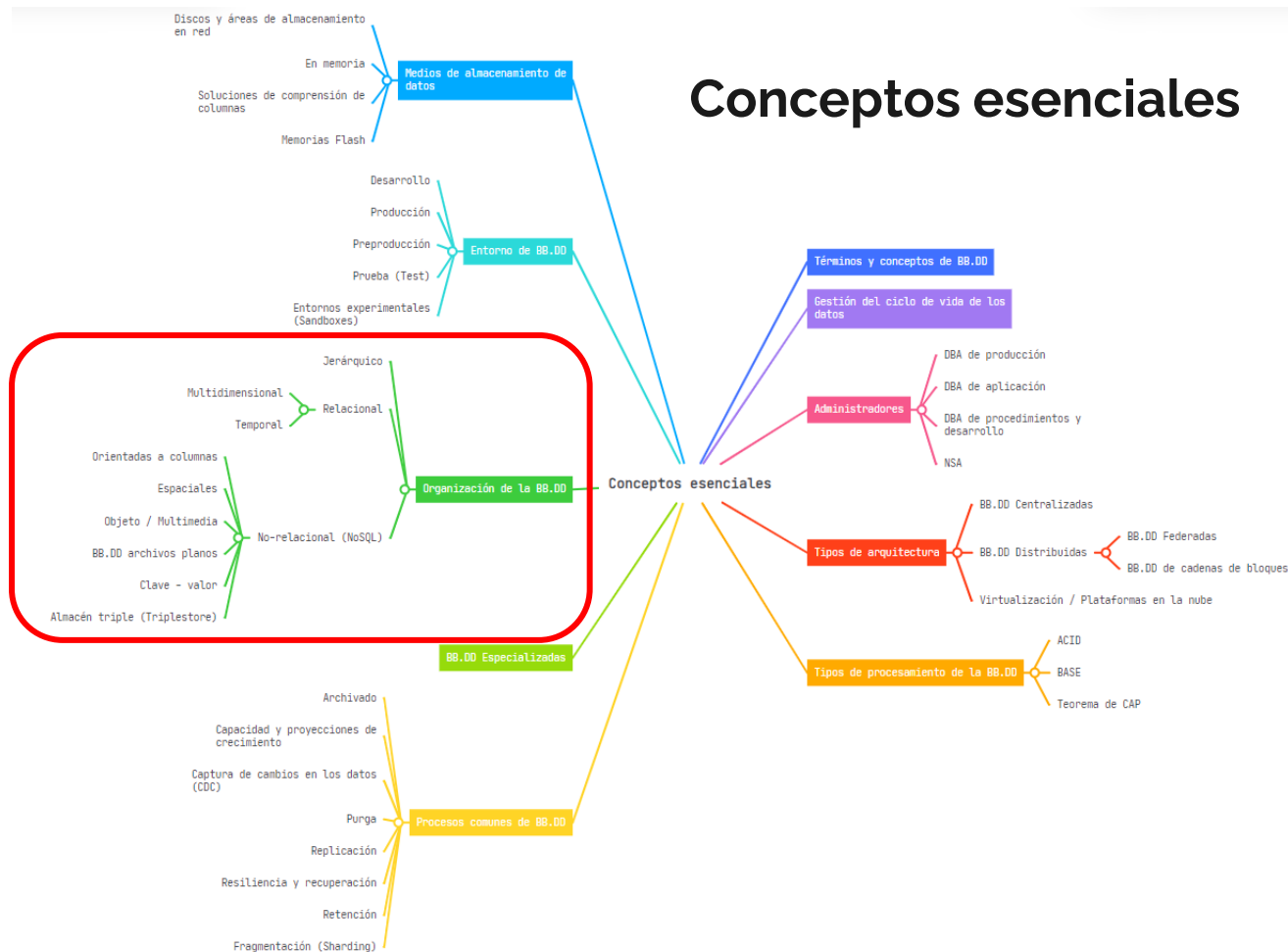
Adtech



Social chat or
news feed

Cortesía de Google

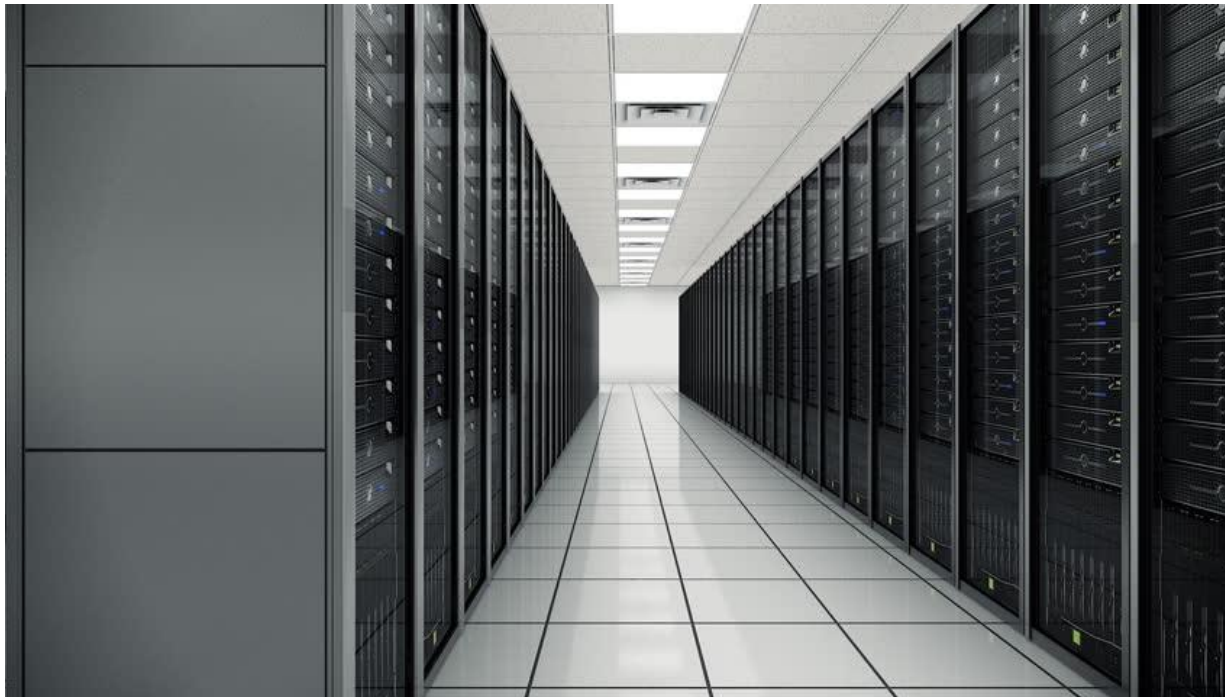
Conceptos esenciales



Conceptos esenciales



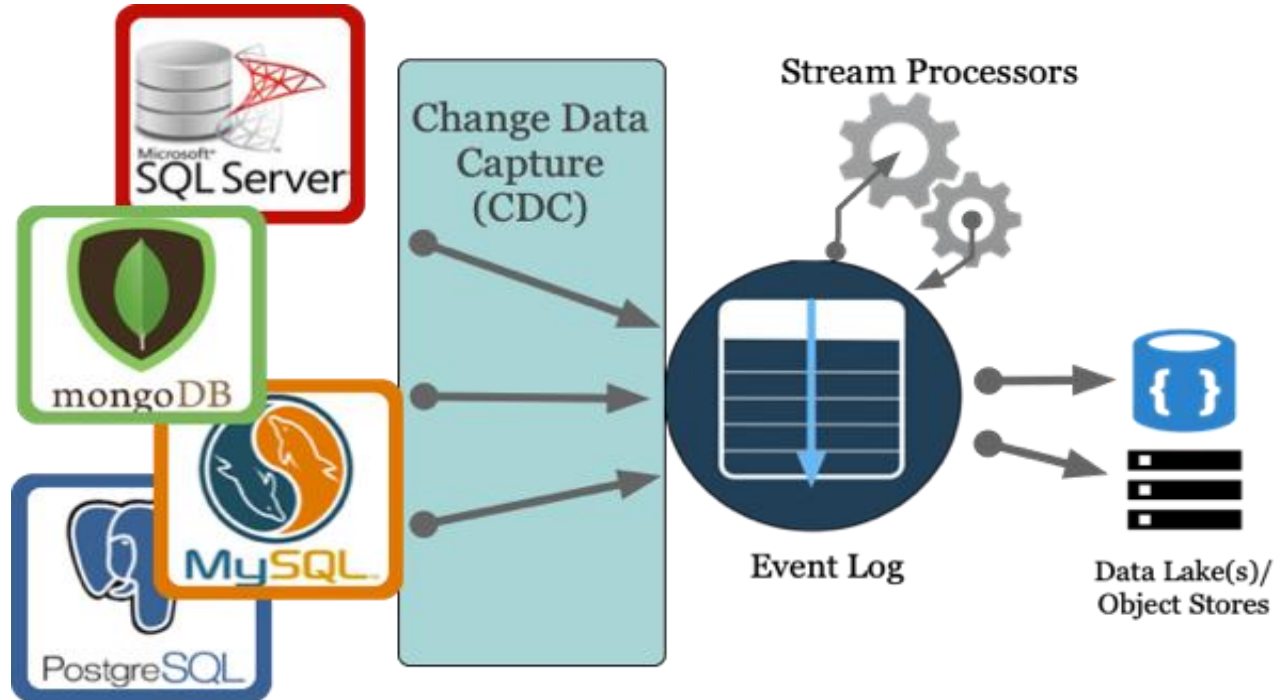
Procesos comunes: Archivado y/o purga

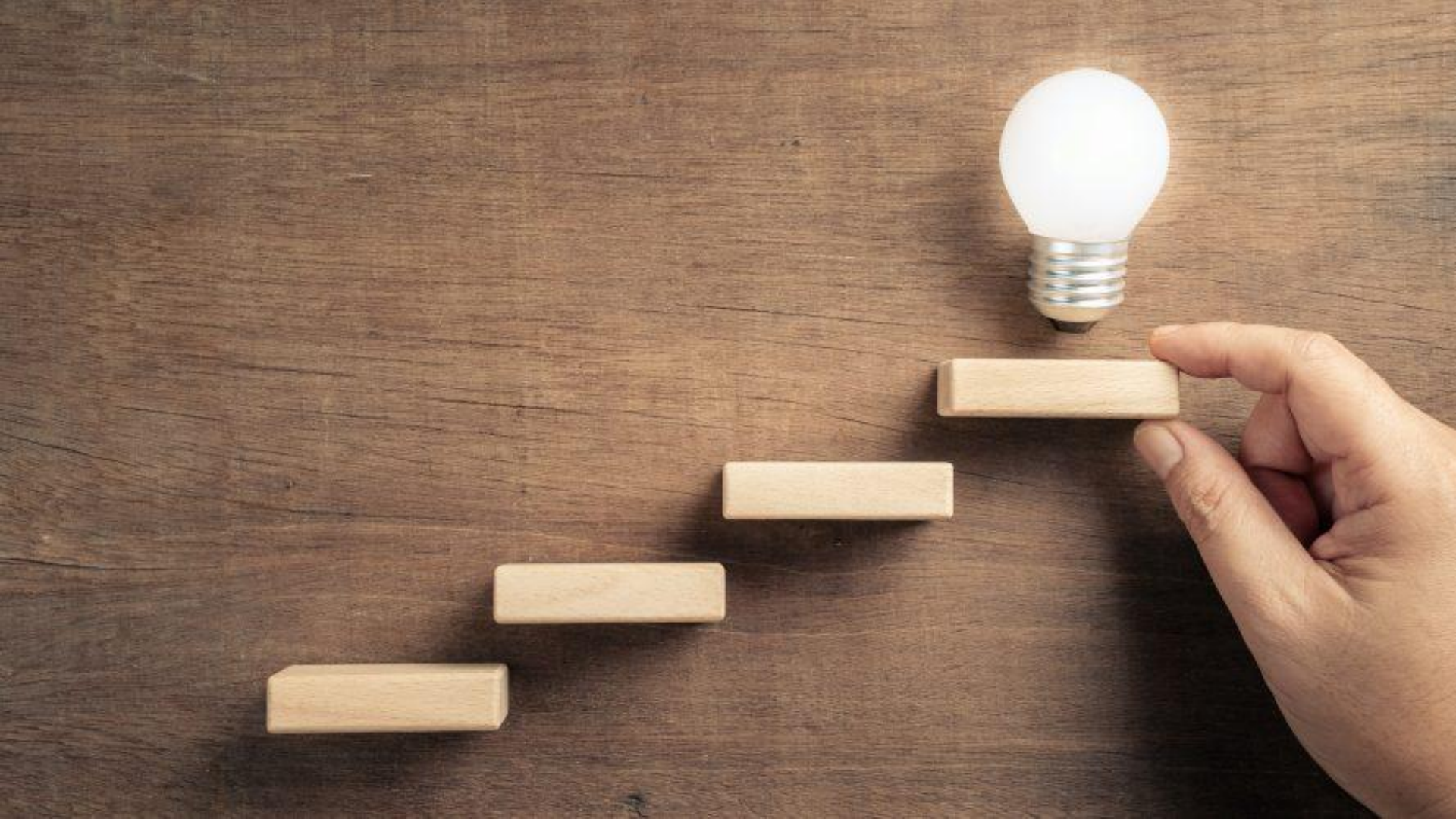


Procesos comunes: Capacity planning



Procesos comunes: CDC Change Data Capture





Performance





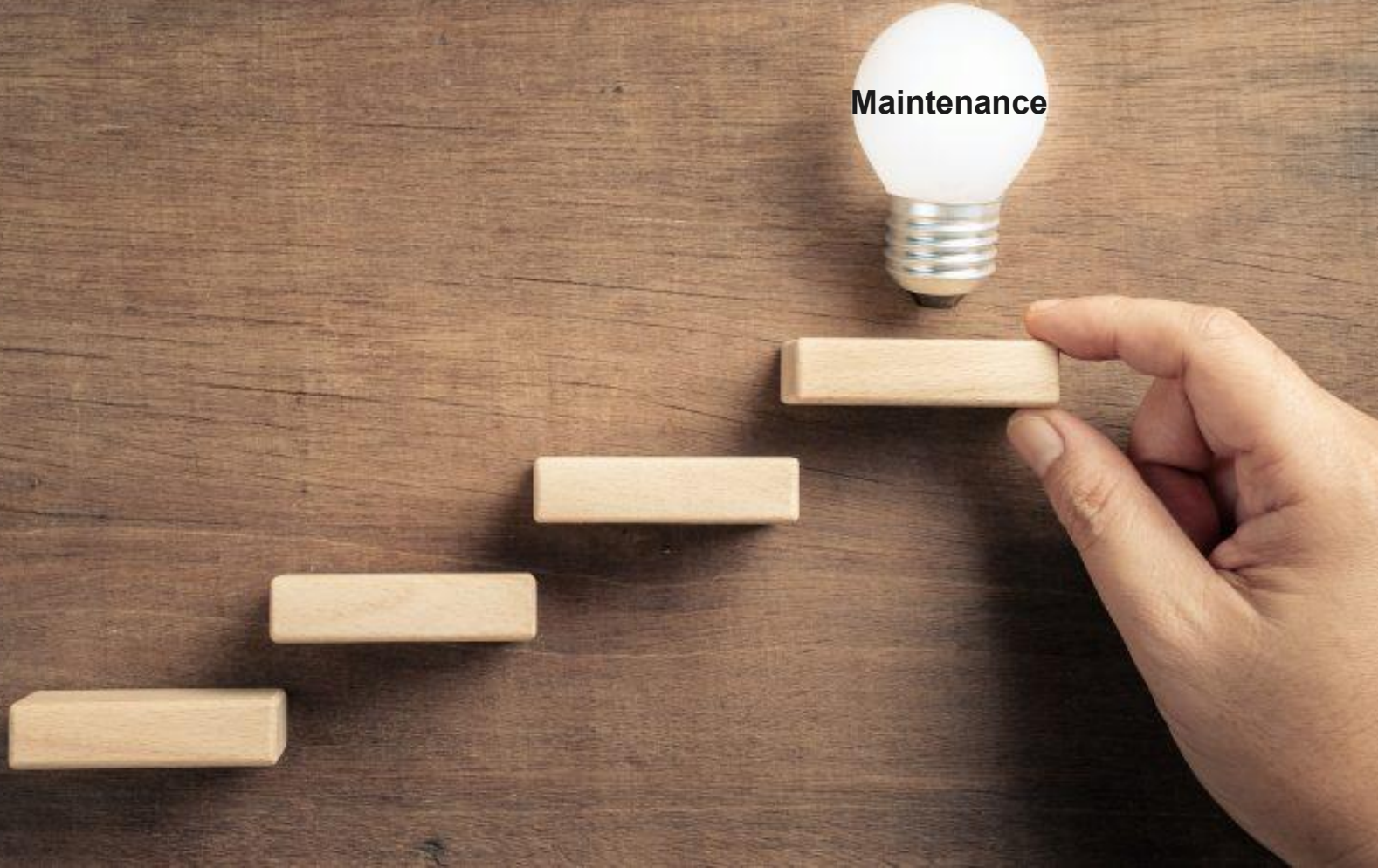
Reuse



Integrity



Maintenance





Data Storage and Operations Governance

Métricas de DS&O



- **Conteos de BB.DD.**
- **Estadísticas sobre el número de transacciones**
- **Métricas de capacidad:**
 - Capacidad de almacenamiento ocupado
 - Número de objetos en la BB.DD.
- **Costos y Complejidad de la Solución.**
 - Costo de desarrollo y de gestión de soluciones
 - Facilidad para adquirir nuevos datos
 - Complejidad de soluciones y operaciones
 - Número de sistemas que utilizan soluciones de integración de datos

Questions & Answers

6%



SLA Stands for:

- A. Service Level Agreement
- B. System Lifecycle Audit
- C. System Latitude Audit
- D. Service Lifecycle Audit

The CAP theorem asserts that the distributed system cannot comply with all the parts of the ACID. A distributed system must instead trade-off between the following properties:

- A. Consistency
- B. Utilization
- C. Availability
- D. System development
- E. Partition tolerance
- F. All of the above

The acronym BASE is made up of:

- A. Basically available
- B. Basically not available
- C. Software state
- D. Soft state
- E. Eventual consistency
- F. Everything considered

¡Gracias!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas

