

Persistencia de datos

ADR - UTN - FRBA - 2020

Persistencia de datos

► ¿Qué es?

Es la capacidad de los datos de perdurar a lo largo del tiempo.

► Tipos de persistencia

- Volátil: No necesitan ser almacenados más allá del procesamiento de los mismos.
- NO Volátil: Deben perdurar más allá del procesamiento.

Persistencia de datos

Definiciones básicas

- **Transacción:** Conjunto de instrucciones que se ejecutan como una unidad de trabajo, es decir de manera indivisible. Debe cumplir con las características ACID.
- **ACID:** Conjunto de propiedades que permiten clasificar las transacciones de los sistemas de gestión de bases de datos:
 - **Atomicity (Atomicidad):** una transacción debe completarse en su totalidad o no ejecutarse en absoluto.
 - **Consistency (Consistencia):** cualquier cambio (transacción) debe conducir de un estado válido de la base de datos a otro estado válido de acuerdo con las restricciones y el esquema de datos.
 - **Isolation (Aislamiento):** múltiples transacciones ocurren cada una de manera independiente sin interferir con ninguna otra.
 - **Durability (Durabilidad):** una vez completada la transacción, ésta debe conservarse aunque se produzcan fallos en la base de datos o el sistema completo.

Persistencia de datos

Definiciones básicas

- **Consistencia:** Todos los nodos deben garantizar la misma información al mismo tiempo, entonces si por ejemplo, insertamos datos (todos los nodos deben insertar los mismos datos), si actualizamos datos (todos los nodos deben aplicar la misma actualización a todos los datos) y si consultamos datos (todos los nodos deben devolver los mismos datos).
- **Tolerancia a partición:** Por lo general los ambientes distribuidos están divididos geográficamente, donde es normal que existan cortes de comunicación entre algunos nodos, el sistema debe permitir seguir funcionando aunque existan fallas que dividan el sistema.
- **Disponibilidad:** Independientemente si uno de los nodos se ha caído o a dejado de emitir respuestas, el sistema debe seguir en funcionamiento. El sistema debe tener la capacidad de seguir operando mientras que la falla se recupera y una vez que lo hace debe poder sincronizarse con los demás.

Persistencia de datos

- Modelos de datos

RELACIONAL

NO RELACIONAL

```
SELECT * FROM Customers_tbl WHERE  
Last_Name='Smith';
```

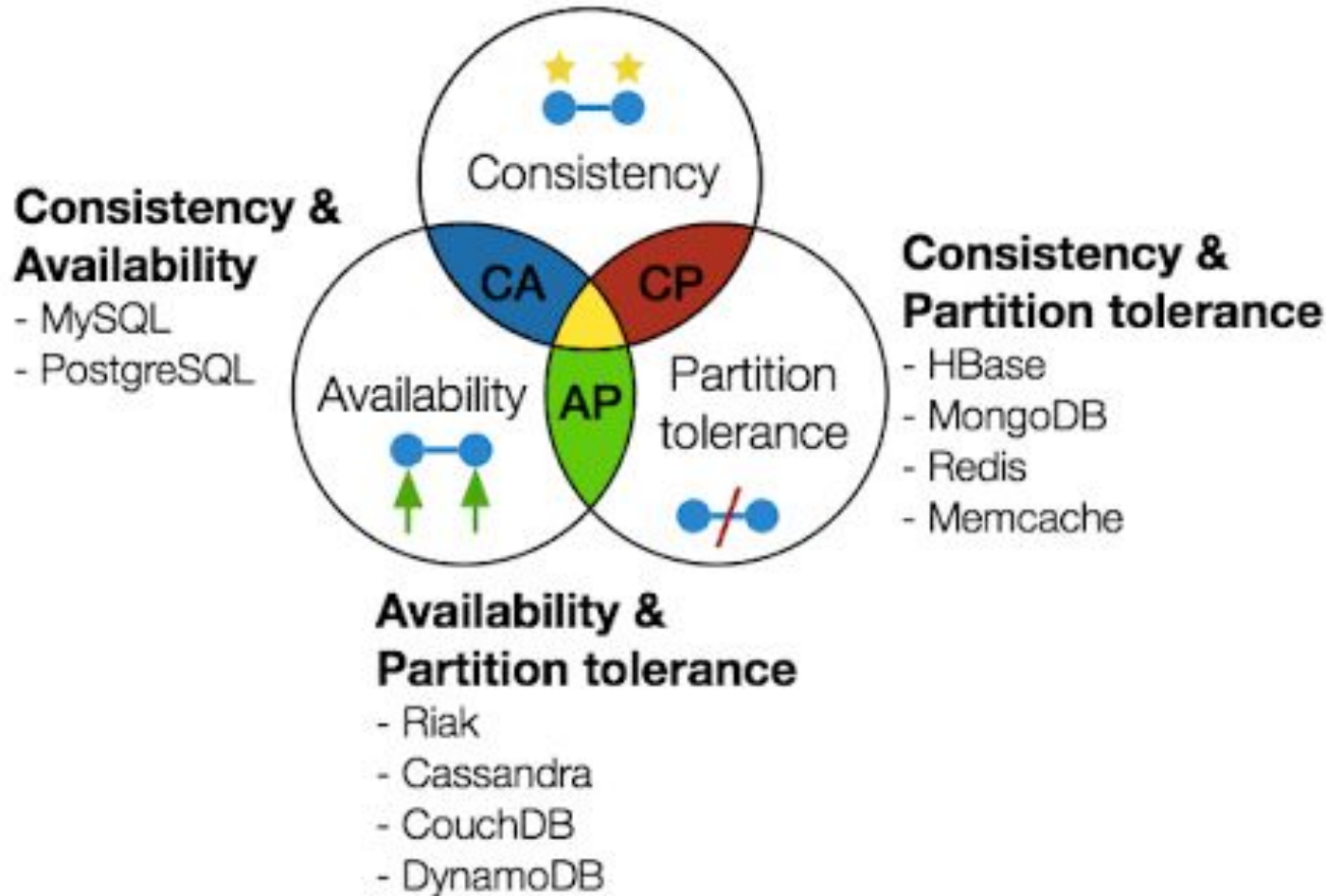
Cust_No	Last_Name	First_Name
560779	Smith	Juan
207228	Smith	George
173996	Smith	Ben
477610	Smith	Conrad

```
Get customer.firstname,customer.lastname,cust  
omer.productID.* where Last_Name='Whitelock'
```

Key	Value
746133	Firstname: George Lastname: Whitelock productID: 2012: 5
135225	Firstname: Luke Lastname: Whitelock productID: 1285: 1 1077: 5
884256	Firstname: Sam Lastname: Whitelock productID: 1442: 2

Persistencia de datos

Teorema de CAP



SISTEMAS DE PERSISTENCIA VOLÁTIL

Persistencia de datos

Sistema de CACHE

► ¿Qué es?

La caché es un buffer especial de memoria destinado a almacenar información de rápido acceso que necesite un sistema particular.

► ¿Cómo funciona?

Se accede por primera vez a caché para dejar una copia de un dato que el sistema predice que puede reutilizar en un futuro cercano. Los siguientes accesos se realizan para ejecutar la lectura de ese dato, haciendo que el tiempo de acceso medio al dato sea menor.

Persistencia de datos

Sistema MEMCACHE



► Qué es?

Es una solución de cacheo Free y Open Source utilizado en sistemas escalables diseñado para acelerar soluciones web dinámicas disminuyendo los accesos a la base de datos para almacenar o recuperar estructuras de datos.

► Hay tres comandos elementales:

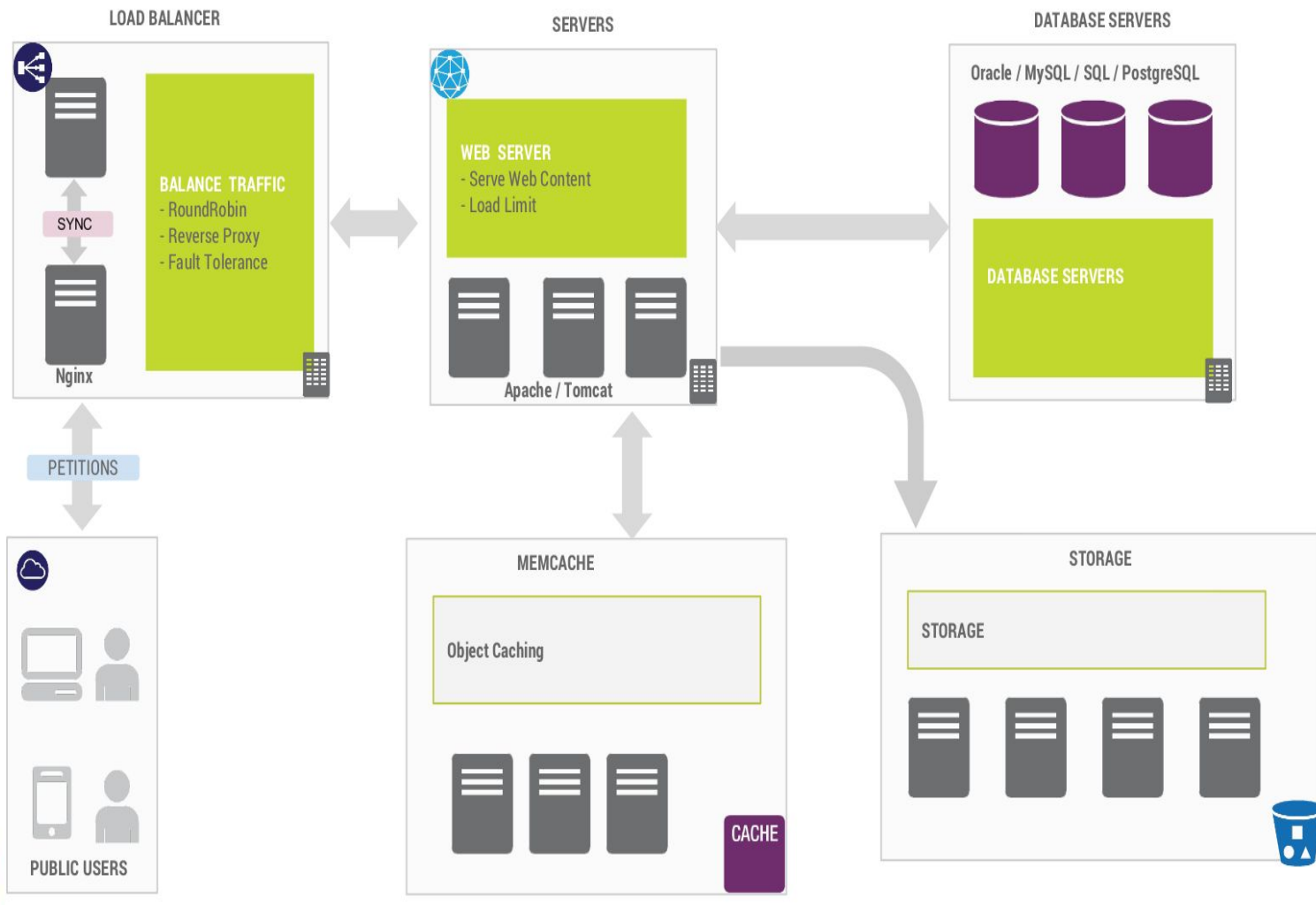
SET: Acciona la inserción de una Key

GET: Obtiene el dato identificado con la Key

DELETE: Elimina el Value y Key

Ejemplo de arquitectura

Sistema MEMCACHE



Persistencia de datos

Sistema VARNISH



► Qué es?

Es una solución para sistemas web dinámicos cuyo objetivo es cachear contenido (archivos) que se utilicen frecuentemente.

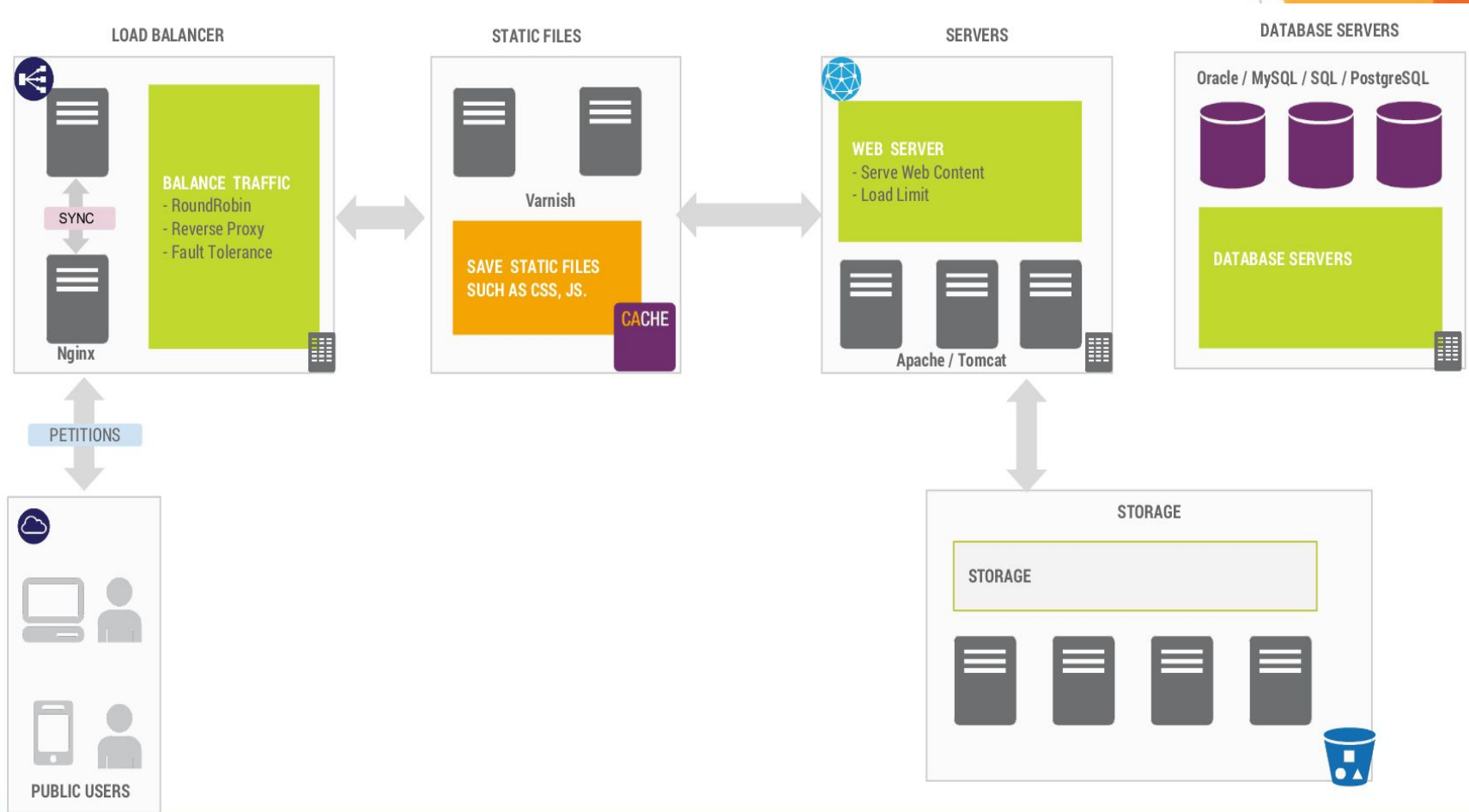
► Para qué?

El sistema está diseñado para cachear Imágenes, Scripts, CSS y cualquier archivo de contenido estático.

El Sistema decide qué información guarda en un espacio de la RAM y qué partes las guarda en el disco duro.

Procesamiento de datos

Sistema VARNISH - Ejemplo de implementación



Persistencia de datos

Sistema REDIS

Almacenamiento de datos en memoria utilizando un esquema clave-valor.



SISTEMAS DE PERSISTENCIA NO VOLÁTIL

Persistencia de datos

► Datos estructurados

Los datos estructurados son aquellos que guardan o respetan una estructura de datos que permite, más allá del posicionamiento físico, almacenarlos o recuperarlos de manera predefinida.

► Lenguaje SQL

Lenguaje estructurado de tratamiento de datos para interactuar con sistemas de bases de datos relacionales (RDBMS)

Persistencia de datos

- ▶ Bases de datos SQL

Bases de datos que implementan modelos relacionales estrictos con el objetivo de garantizar la consistencia de los datos a partir de relaciones.

- ▶ Bases de datos NO SQL

Bases de datos cuyo modelo no busca garantizar la consistencia de los datos a partir de relaciones sino que tienen por objetivo soportar modelos flexibles que NO requieran cambios estructurales a partir de la variabilidad de las estructuras de datos a lo largo del tiempo.

Persistencia de datos

NO SQL - Clasificación

Clave-Valor: los datos se almacenan en pares del tipo clave-valor. El valor es un dato de tipo blob.
Ej.: Riak, Dynamo, Azure, Redis

Column Family: permiten almacenar claves mapeadas a valores y esos valores agrupados en múltiples familias de columnas siendo cada columna un mapa de datos.
Ej.: Cassandra, HBase, Amazon SimpleDB

Basadas en documentos: La base de datos almacena y recupera documentos que pueden estar en XML, JSON o BSON.
Ej.: MongoDB, Couchbase, CouchDB, Lotus Notes, Oracle NoSQL Database

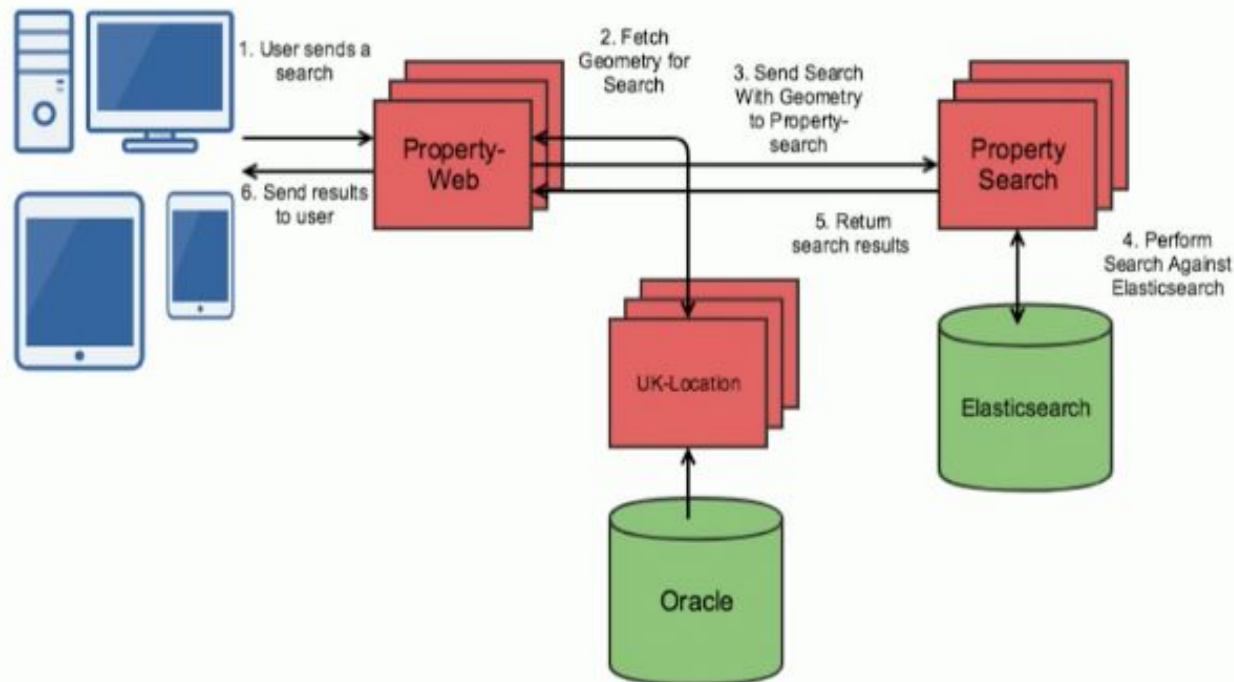
Basadas en grafos: permiten almacenar entidades y relaciones entre esas entidades. Tanto los nodos como las relaciones tienen sus propiedades asociadas.
Ej.: Neo4J, InfiniteGraph, OrientDB, FlockDB

Persistencia de datos

NO SQL - Arquitectura modelo de Sistema Elastic

Modelo utilizado para la búsqueda documental a partir de relaciones de términos.

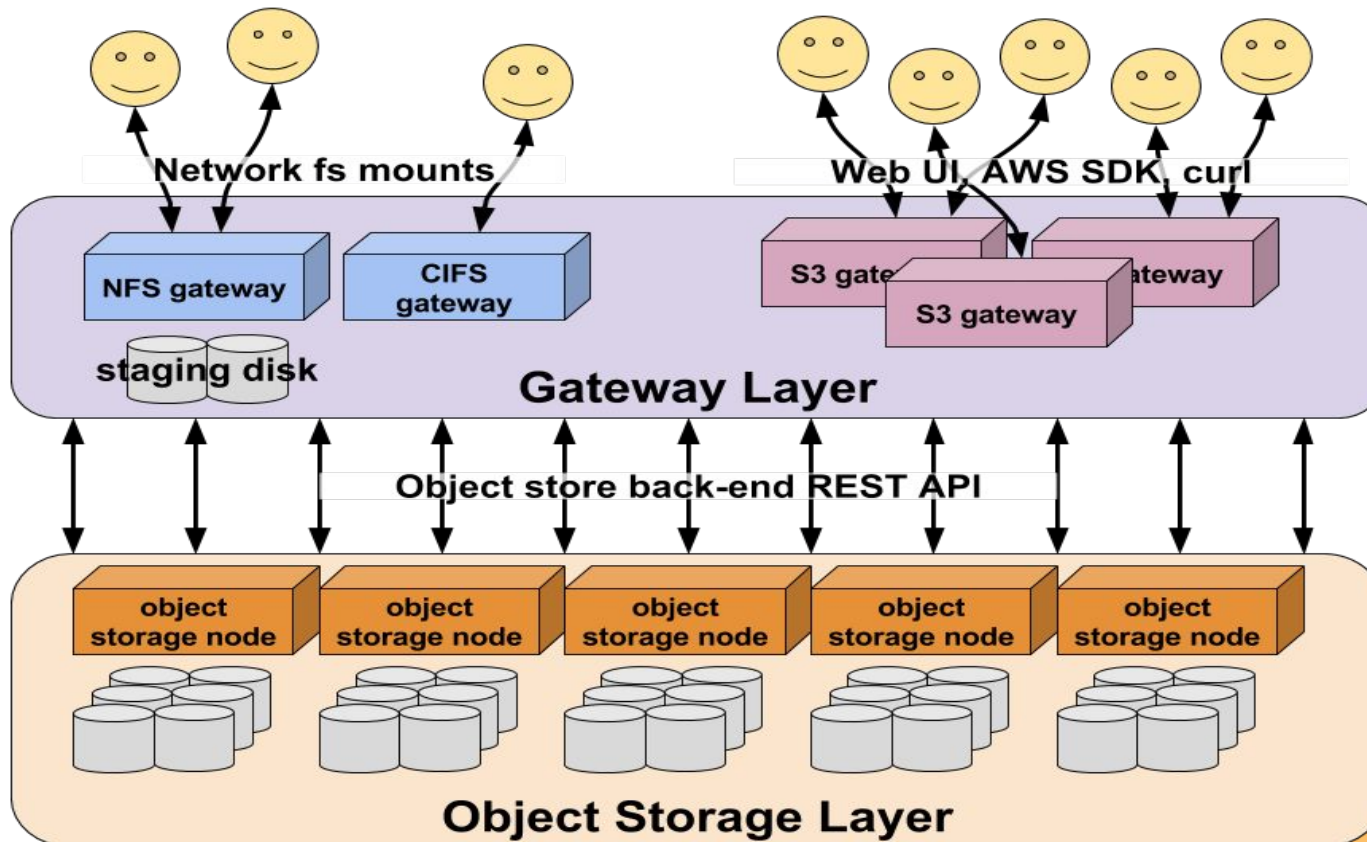
Search Flow



Persistencia de datos

Persistencia de objetos

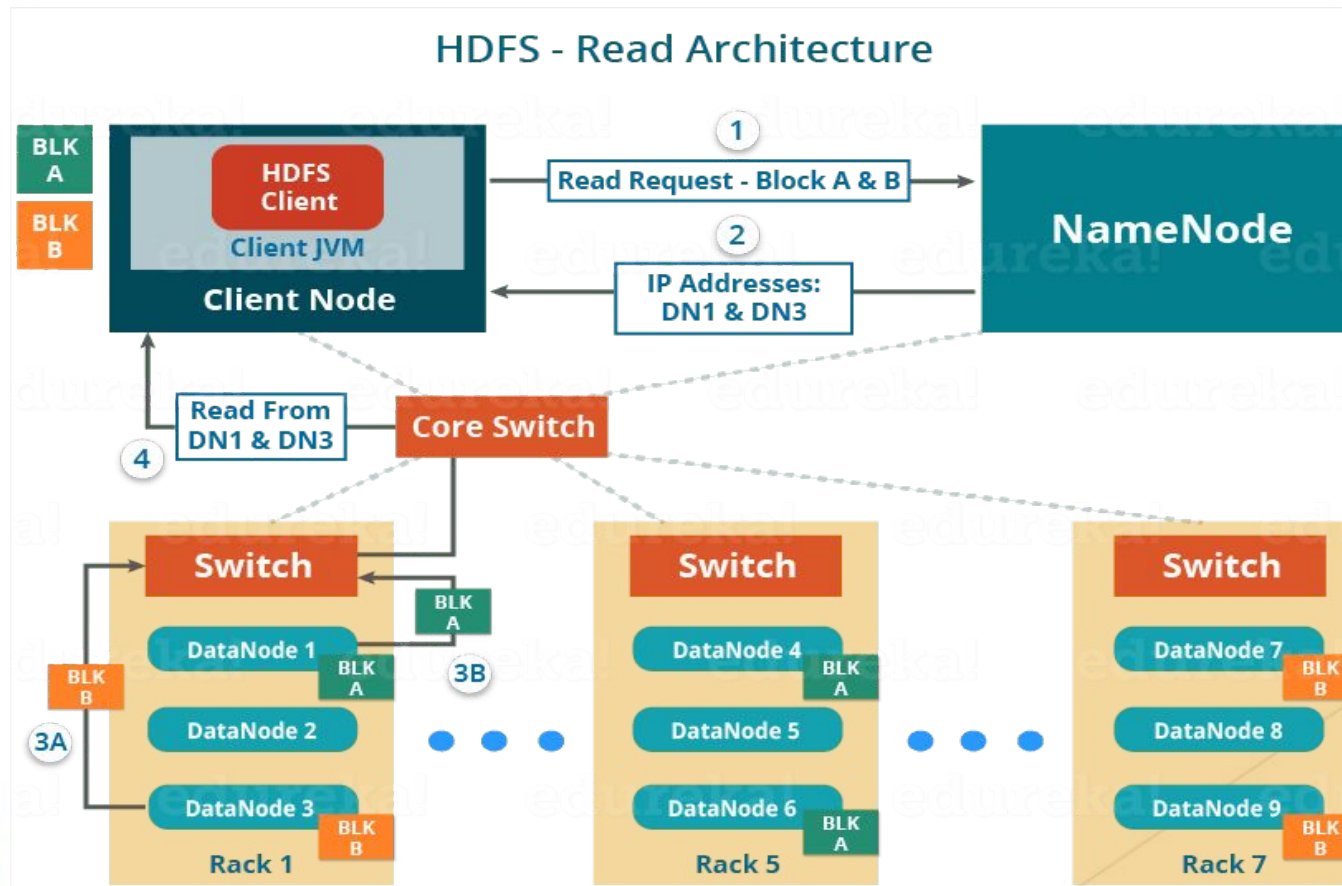
Modelo de persistencia distribuida de archivos implementado generalmente en servicios de nube pública. Ej: Amazon Simple Storage Service (S3)



Persistencia de datos

Persistencia de archivos distribuidos

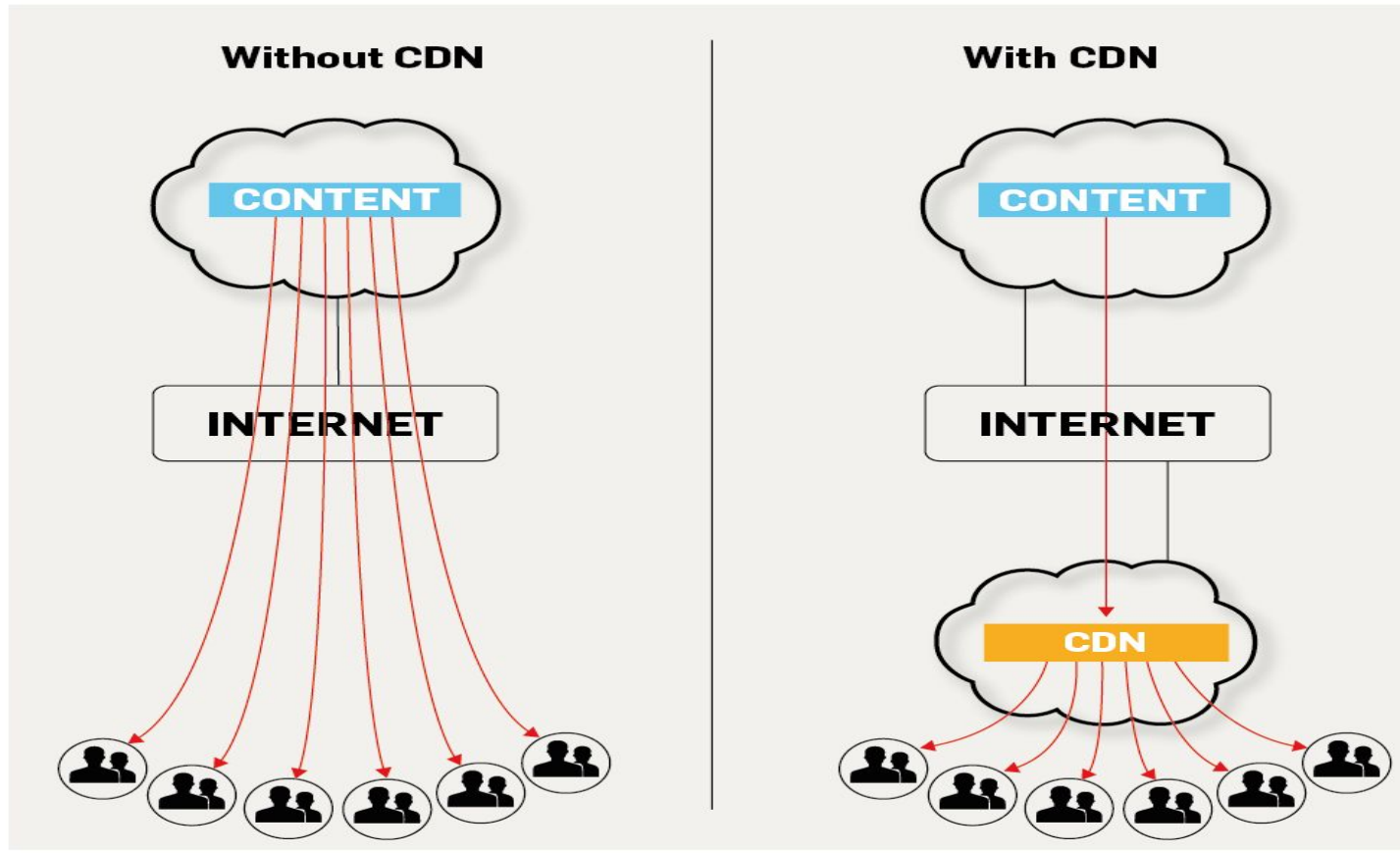
Sistema de archivos distribuido para manejo de grandes volúmenes de datos, rápido acceso y alta disponibilidad. Ej: Hadoop Distributed File System (HDFS)



Persistencia de datos

Content delivery Network (CDN)

Sistema distribuido y escalable de entrega de contenidos basado en minimizar el costo de red entre el punto de distribución y el usuario. Ej: akamai.

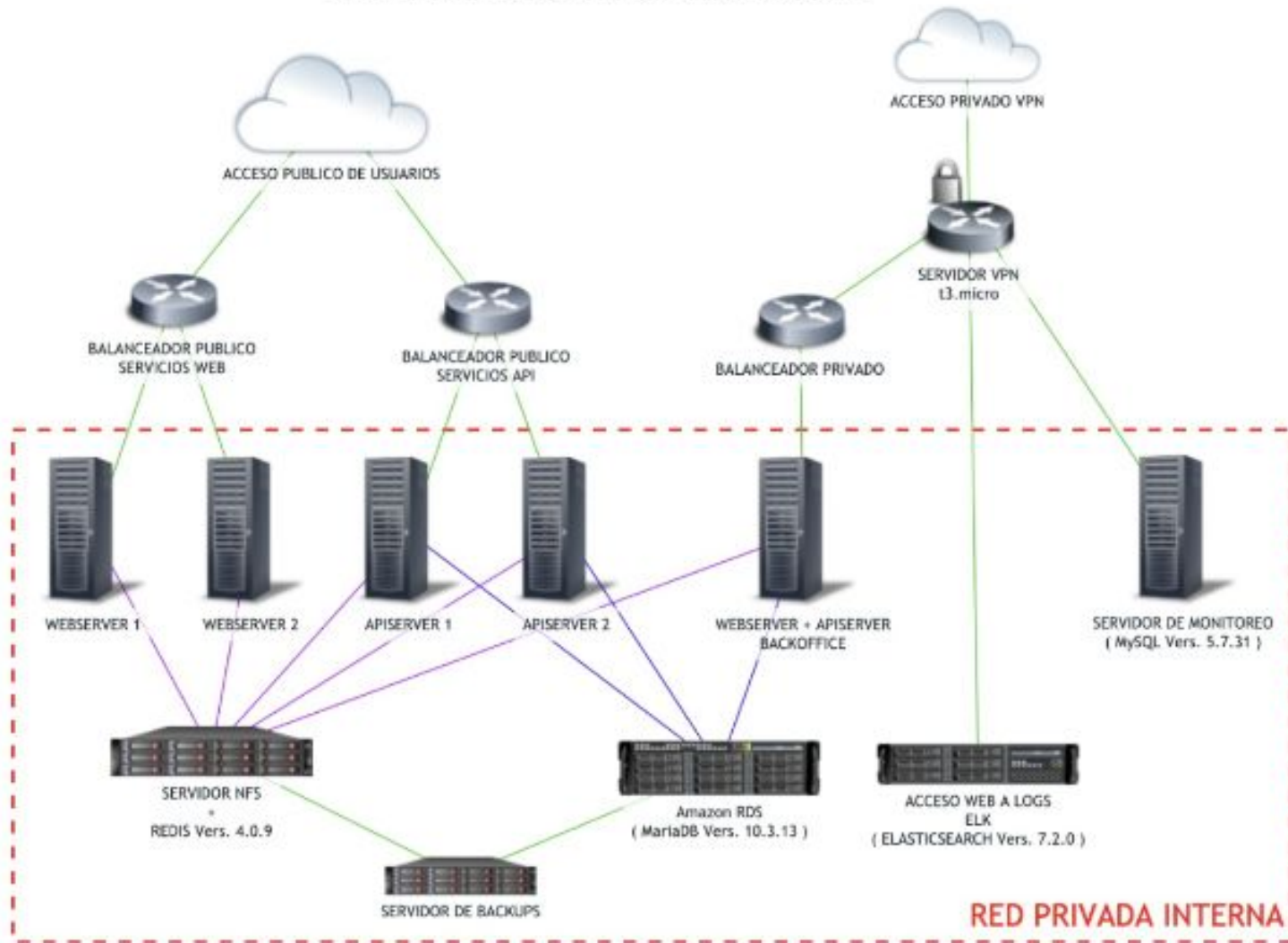


SISTEMAS DE PERSISTENCIA POLÍGLOTA

Persistencia de datos

Modelo polígota

PLATAFORMA ANGULAR + JAVA



Procesamiento de datos

Referencias

<https://aws.amazon.com/es/s3/>

<https://www.ibm.com/analytics/hadoop/hdfs>

<https://www.akamai.com/es/es/cdn/what-is-a-cdn.jsp>

<https://www.varnish-software.com/how-it-works/>

<https://memcached.org/>

<https://redis.io/topics/introduction>

Procesamiento de datos

Consultas...