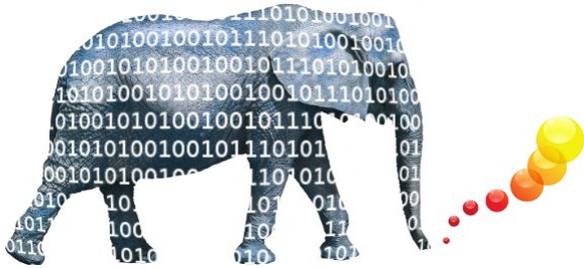


José Antonio Salazar Carmona

Tecnología de base de datos de tipo NoSQL

Concepto



Volumen



Variedad



Velocidad

Conjunto de estrategias que permiten el uso eficiente de datos con un enfoque en el rendimiento, confiabilidad y agilidad.

De la complejidad a la simplicidad



Inicios

2003



Toma fuerza el concepto



“Bigtable: A Distributed Storage System for Structured Data”

Google



Toma fuerza el concepto



“Dynamo: Amazon’s Highly Available Key-value Store”

amazon.com[®]



Características NoSQL

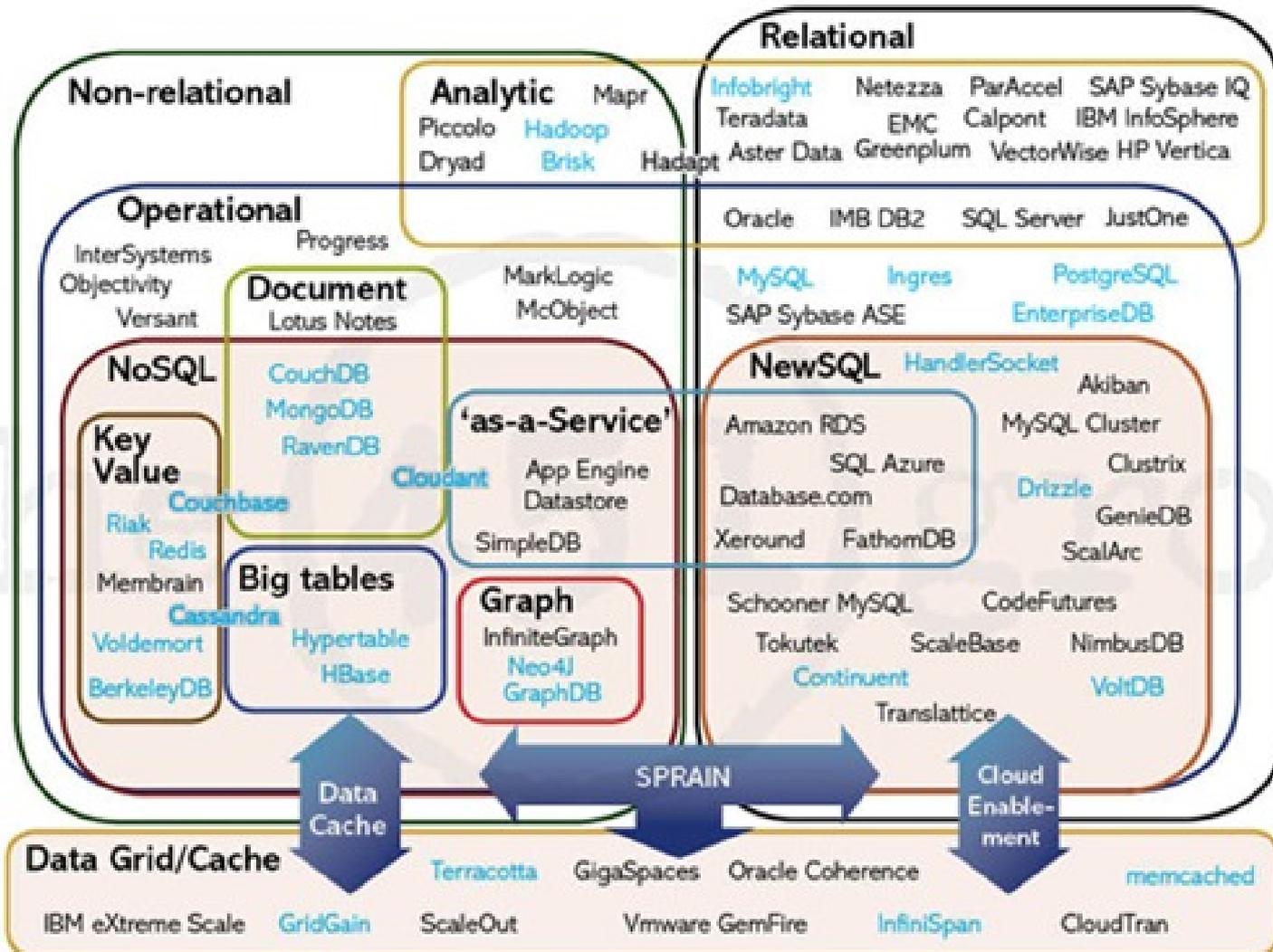
Mitos

- Sólo open source
- Sólo big data
- Sólo computación nube
- Sólo para las minorías

Realidades

- Más que renglones
- No existen Joins
- Libres de esquemas
- Alto desempeño
- Escalable
- Innovador

Opciones



Teorema Brewer

(2000)

Visual Guide to NoSQL Systems

Availability:
Each client can
always read
and write.

A

Data Models

Relational (comparison)
Key-Value
Column-Oriented/Tabular
Document-Oriented

CA

RDBMSs
(MySQL,
Postgres,
etc)

Aster Data
Greenplum
Vertica

AP

Dynamo
Voldemort
Tokyo Cabinet
KAI

Cassandra
SimpleDB
CouchDB
Riak

Pick Two

C

Consistency:
All clients always
have the same view
of the data.

CP

BigTable
Hypertable
Hbase

MongoDB
Terrastore
Scalaris

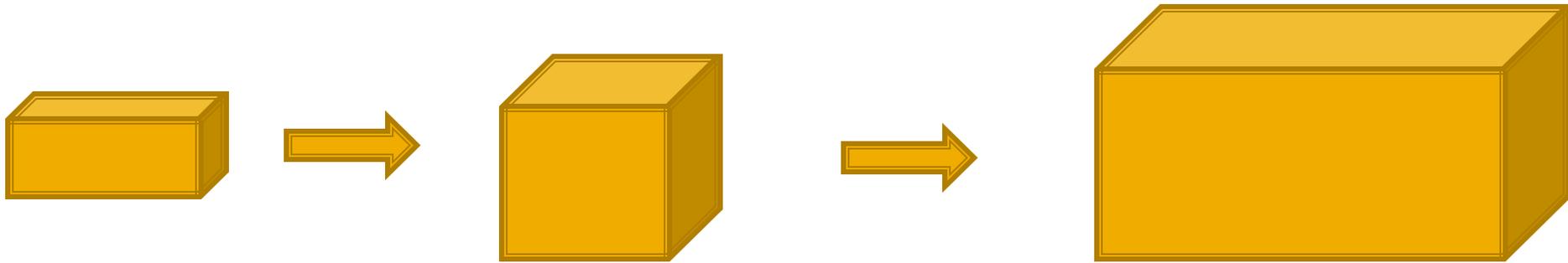
Berkeley DB
MemcacheDB
Redis

P

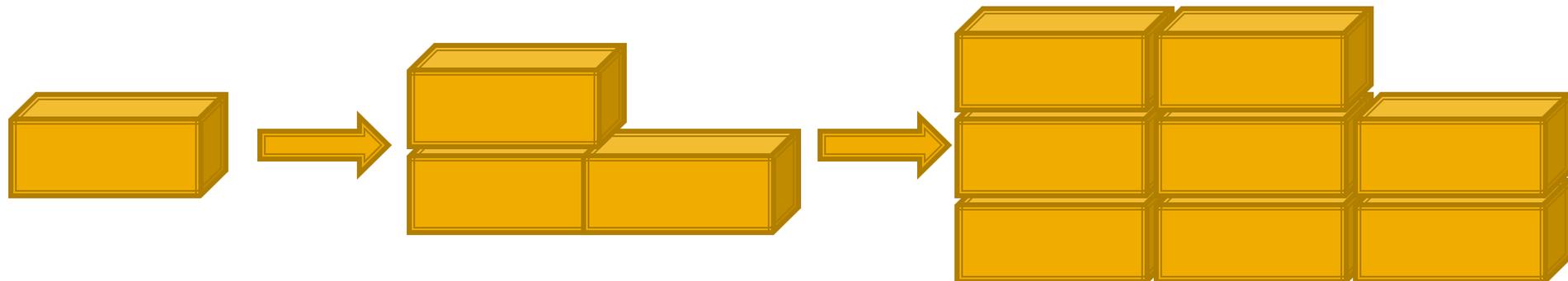
Partition Tolerance:
The system works
well despite physical
network partitions.

Estrategias

Escalabilidad vertical

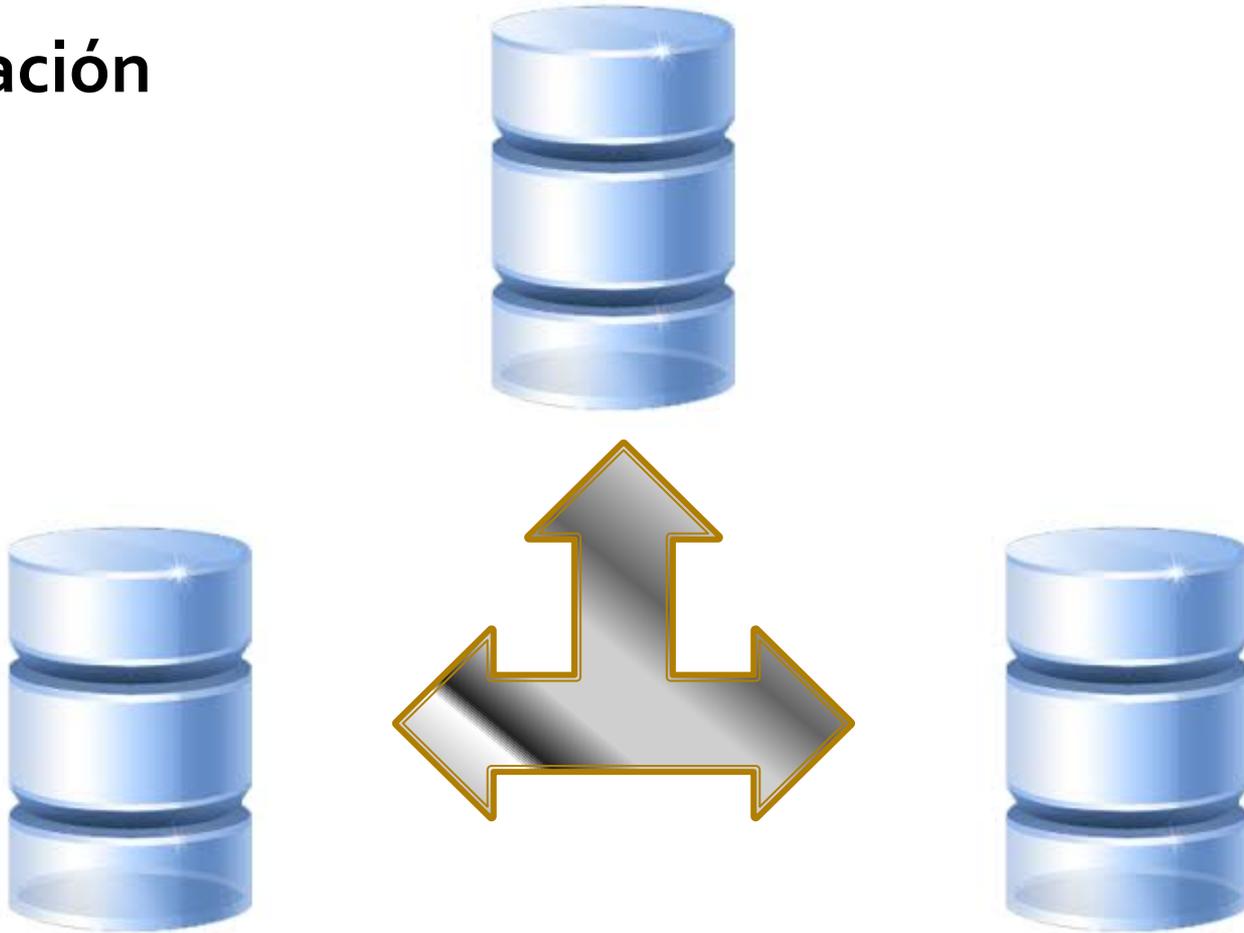


Escalabilidad horizontal



Estrategias

Replicación



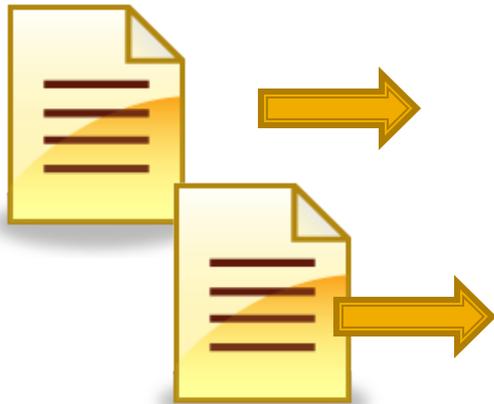
Estrategias

Particionamiento



Estrategias

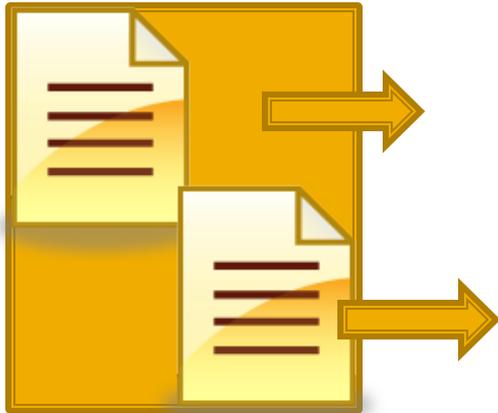
Índices invertidos



Palabra	Documento: ubicación
análisis	doc1: [18, 75]
base	doc3: [12]
complejo	doc1: [23, 57, 99]
datos	doc2: [48], doc3:[59, 83], doc1: [78, 94]
internet	doc2:[17, 67], doc1:[18, 39, 107]
tecnología	doc2:[63]

Estrategias

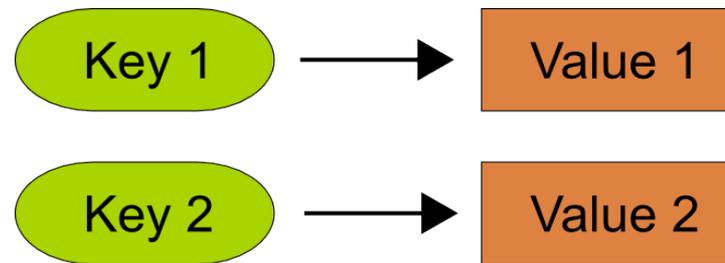
Índices invertidos



Palabra	Documento: ubicación
análisis	n1.doc1: [18, 75]
base	n2.doc3: [12]
complejo	n1.doc1: [23, 57, 99]
datos	n2.doc2: [48], n2.doc3:[59, 83], n1.doc1: [78, 94]
internet	n2.doc2:[17, 67], n1.doc1[18, 39, 107]
tecnología	n2.doc2:[63]

Categorías de bases de datos

clave-valor

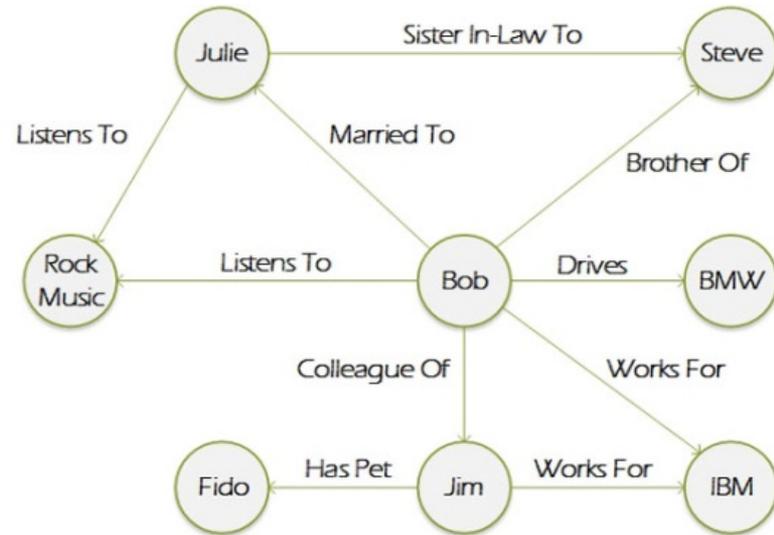


Columnar

Key	Driver Information		Car Information		
123546	Name:John	Insurance: Geico	Car: Speed3	Year:2013	Warranty:Yes
123547	Name:Jen	Insurance:State Farm	Car:626	Year:2008	
123548	Name:Tonv				

Categorías de bases de datos

grafos



documentos

```
{  
  "ID": 1,  
  "FIRST": "Frank",  
  "LAST": "Weigel",  
  "ZIP": "94040",  
  "CITY": "MV",  
  "STATE": "CA"  
}
```

JSON

=

User Info			
KEY	First	Last	ZIP_id
1	Frank	Weigel	2
2	Ali	Dodson	2
3	Mark	Azad	2
4	Steve	Yen	3

+

Address Info			
ZIP_id	CITY	STATE	ZIP
1	DEN	CO	30303
2	MV	CA	94040
3	CHI	IL	60609
4	NY	NY	10010

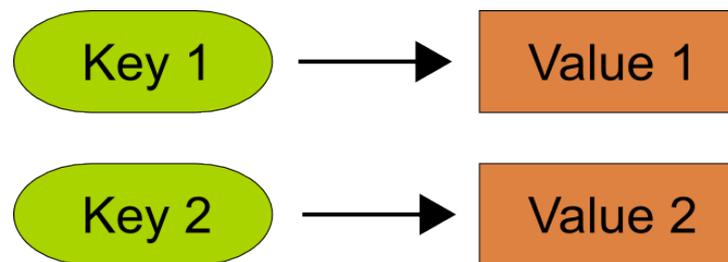
Categorías de bases de datos

[lista completa](#)

Tipo	Usos	Ejemplos
clave-valor: Usa una clave para acceder a los valores.	<ul style="list-style-type: none">- Almacenar archivos- Sistemas de archivos basados en llaves- Cache de objetos- Sistemas para escalar	<ul style="list-style-type: none">- Berkeley DB- Memcache- Redis- Riak- DynamoDB
Columnares: Uso de matrices distribuidas con renglones y columnas como llaves.	<ul style="list-style-type: none">- Web crawler- Problemas con grandes volúmenes de datos	<ul style="list-style-type: none">- Apache Hbase- Apache Cassandra- Hypertable- Apache Accumulo
Grafos: Uso de relaciones.	<ul style="list-style-type: none">- Redes sociales- Detección de fraudes- Relaciones fuertes entre datos	<ul style="list-style-type: none">- Neo4J- AllegroGraph- BigData (RDF Store)- InfiniGraph (Objectivity)
Documentos: Uso de estructuras de datos jerárquicas.	<ul style="list-style-type: none">- Alta disponibilidad de datos- Búsqueda de documentos- Ejes de integración- Administración de contenido web- Publicación	<ul style="list-style-type: none">MongoDBCouchDBCouchbaseeXist-dbBerkeley DB XML



- Muy veloz
- Alta disponibilidad
- Tolerante a fallas
- Simplicidad operacional
- Escalable
- Distribuye los datos en cada nodo del cluster
- Interfaz REST
- Integrable con Apache Solr



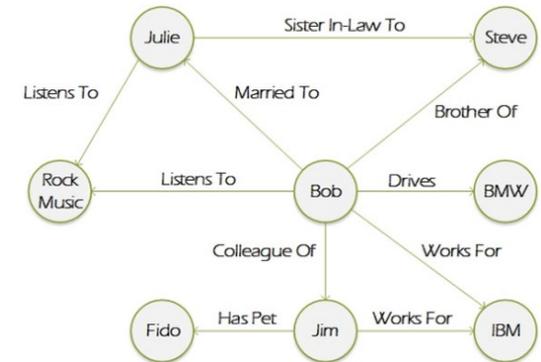


- Alta consistencia
- Escalable
- Control de versiones
- Compresión de datos
- Tablas en memoria
- Eficiente para big data
- Puede convivir en el ecosistema Hadoop
- Un clúster en producción por lo menos requiere cinco nodos

Key	Driver Information	Car Information
123546	Name:John Insurance: Geico	Car: Speed3 Year:2013 Warranty:Yes
123547	Name:Jen Insurance:State Farm	Car:626 Year:2008
123548	Name:Tonv	



- Puede ser pequeño o grande
- Sin esquemas
- Interfaz REST
- Interfaz Web
- Integrado a Apache Solr
- Los nodos esclavo aceptan escrituras
- Versión empresarial con transacciones
- Los datos viven en todos los nodos



- Escalable
- Alto desempeño
- Facilidad de acceso
- Sin esquemas
- Campos con tipo
- Referencias entre documentos
- Mapeos en memoria
- Índices secundarios
- Consultas completas
- Particionamiento configurable



[Ir al ejemplo](#)

Popularidad de bases de datos

- <http://db-engines.com/en/ranking>

