



UAB

Universitat Autònoma
de Barcelona

Andrés Cencerrado

Tecnologías de BBDD NoSQL (I)

Origen y necesidad

- Recordando las “4 V”



- Volumen: Enorme cantidad de datos.
- Velocidad: Generación y movimiento a gran velocidad.
- Variedad: Distintos tipos.
- Veracidad: El grado de confianza de los datos.

Origen y necesidad

- Necesidad de implementación sistemas NoSQL a gran (enorme) escala



Google
BigTable

"A Bigtable is a sparse, distributed, persistent multidimensional sorted map. The map is indexed by a row key, column key, and a timestamp; each value in the map is an uninterpreted array of bytes."

Fay Chang, Jeffrey Dean, et al. "Bigtable: A Distributed Storage System for Structured Data" – OSDI 2006

Origen y necesidad

- Necesidad de implementación sistemas NoSQL a gran (enorme) escala



"Dynamo is used to manage the state of services that have very high reliability requirements and need tight control over the tradeoffs between availability, consistency, cost-effectiveness and performance."

Giuseppe DeCandia, Deniz Hastorun, et al. "Dynamo: Amazon's Highly Available Key-value Store" – SOSP 2007

Origen y necesidad



Key	Value
2843	{ name: "st_23", location: "York", month: "11", day: "23", deg_C: "5.4" }
4567	{ name: "st_344", alt: "450", month: "4", day: "15", hum: "78.2" }

Entradas arbitrarias

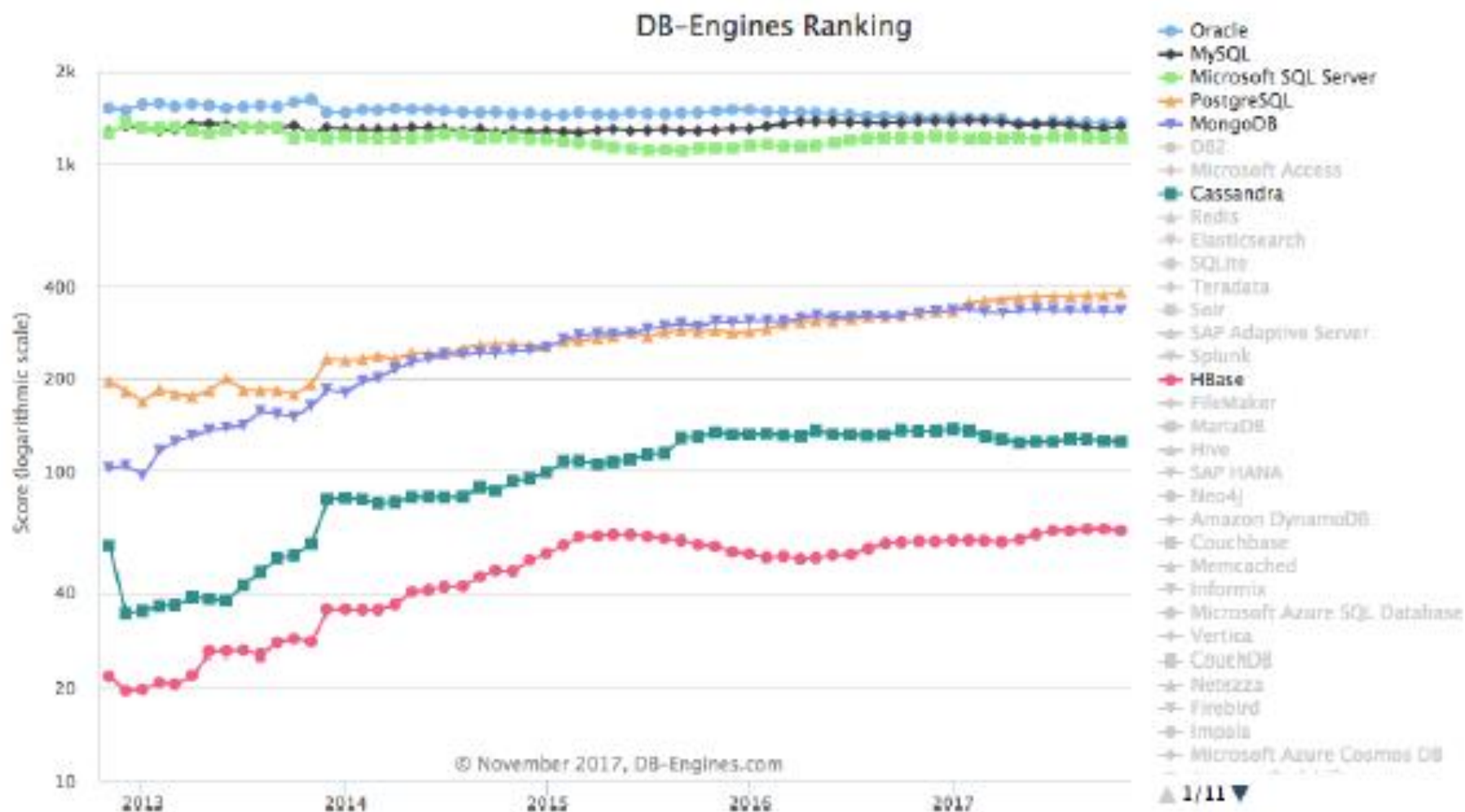
Familias de columnas

ID	Weather Station Data		Measurement data		
2843	name="st_23"	location="York"	month="11"	day="23"	deg_C="5.4"
4567	name="st_344"	alt="450"	month="4"	day="15"	hum="78.2"

Claves de columna

Claves de fila/registro

Actualmente hay 300+ tecnologías DBMS

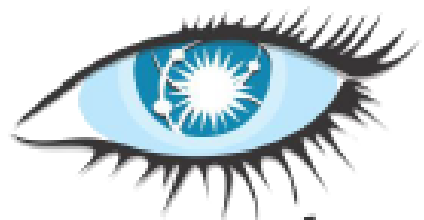




- 2007, Apache
- Integración Hadoop, HDFS
- Wide column store
- Escrita en Java, desarrollo a partir de BigTable
- Sistema CP (consistencia y tolerancia a particiones)



Tecnologías representativas



cassandra

ebay



- 2007 Facebook – 2009 Apache
- Alta escalabilidad y disponibilidad
- Wide column store
- Escrita en Java, muy buena integración con Hadoop y otras plataformas Big Data
- Sistema AP (disponibilidad y tolerancia a particiones)

NETFLIX

Tecnologías representativas



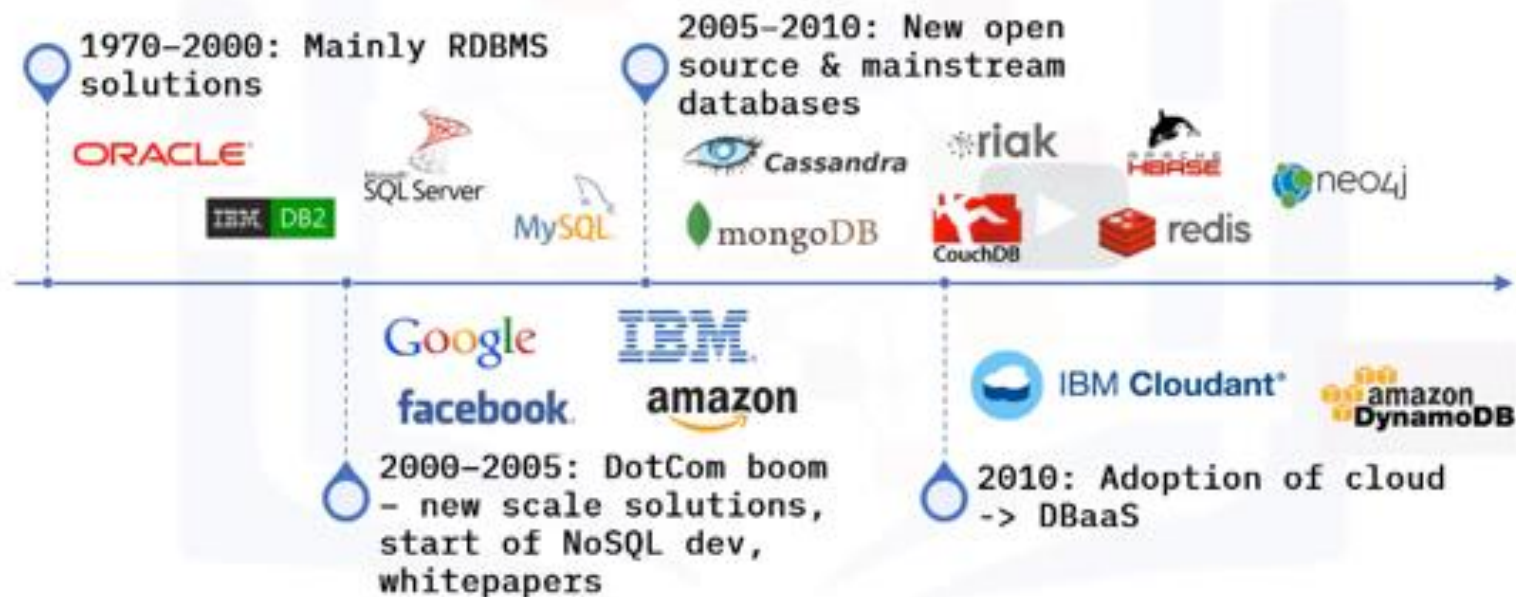
- 2007 10gen Inc. (MongoDB Inc.)
- Simplicidad en el desarrollo
- Document store
- Escrita en C++, muy buena integración en distintos casos de uso
- Sistema CP (consistencia y tolerancia a particiones)



BOSCH

Create, Read, Update, and Delete (CRUD)

History of NoSQL



NoSQL Database Categories

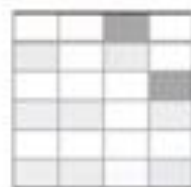
General consensus is...
...NoSQL databases fit
into four categories



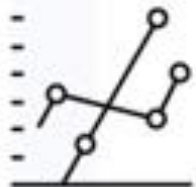
Key-Value



Document



Column



Graph

Key-Value NoSQL Database Examples



Document-Based NoSQL Database Examples



IBM **Cloudant**®



mongoDB



CouchDB



terrastore



OrientDB®



Couchbase

RAVENDB

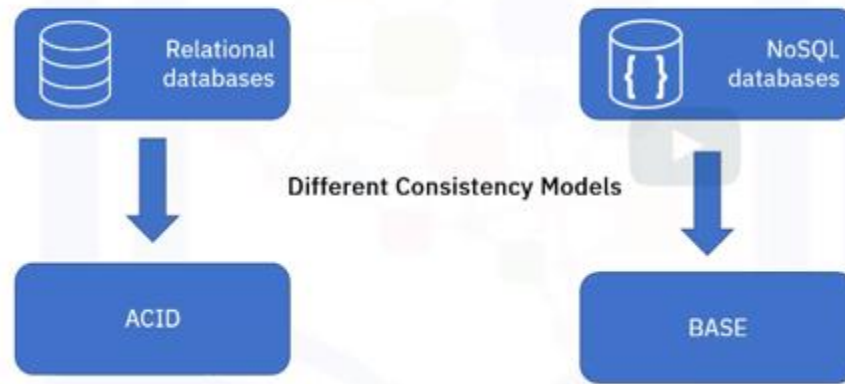
Column-Based NoSQL Database Examples



Graph NoSQL Database Examples



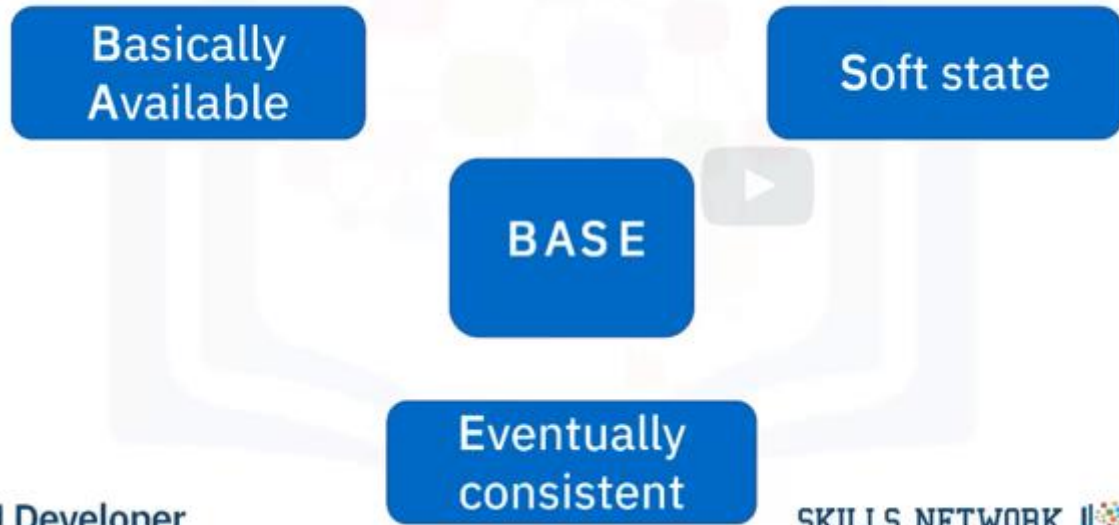
ACID vs. BASE consistency models



ACID definition



BASE definition



M Developer

SKILLS NETWORK 

