



Bases de Datos No Relacionales

Instituto de Computación, FING, Udelar
CC-BY Lorena Etcheverry lorenae@fing.edu.uy



Un poco de historia

Y en el comienzo, fueron los archivos ...

Cada programa resolvía la gestión de datos.

La concurrencia y consistencia se resolvían en cada caso.
Optimización, *caching*, *pre-fetching* también.

**No había separación entre
la representación (estructura de datos)
y los datos.**

Almacenamiento: en cintas de acceso secuencial

El disco magnético aparece en los 50s!

Los primeros DBMS

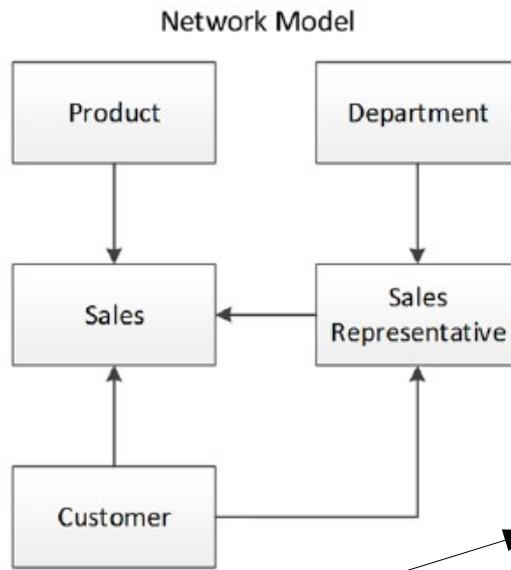
Programas dedicados a la gestión de datos:

desacoplar de la lógica de la aplicación para permitir el **reuso** de esta lógica!

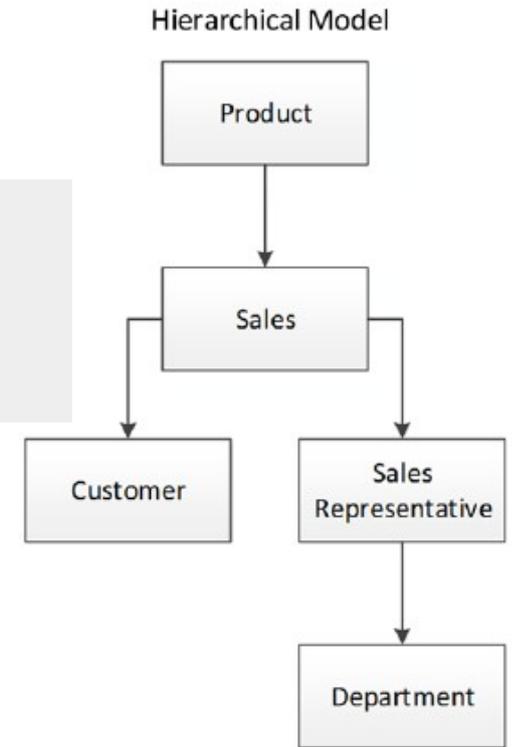
Sistemas *mainframe*

Naturaleza **navegacional**: modelos jerárquico y de red

Modelo de Red
(CODASYL, 1969)



Modelo Jerárquico
(IBM IMS, 1969)



1960

1970



IBM System/360
(IBM 1964) [1]

[1]Image By Sandstein - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=16305441>

Algunas desventajas de los primeros DBMS

Rigidez tanto en las estructuras de datos como en las consultas que podían resolver.

Enfocadas en el registro: operaciones **CRUD** (Create, Read, Update, Delete)

Hacer **consultas de análisis** demandaba hacer programas complejos.

No había **separación** entre la **representación lógica** y el almacenamiento físico

Information Retrieval

A Relational Model of Data for Large Shared Data Banks

E. F. CODD
IBM Research Laboratory, San Jose, California

(Codd, 1970-1972)

The Entity-Relationship Model—Toward a Unified View of Data

PETER PIN-SHAN CHEN
Massachusetts Institute of Technology

(Chen, 1976)

Arquitecturas cliente-servidor (90s)

OODBMS (90s)

OLAP (90s)

1960

1970

1980

1990

2000

2010

System R
SEQUEL
(1977, IBM)

SQL-86
(ANSI, 1986)

Comienzo de la Internet comercial (1995)

INGRES
(1974, U. Berkeley)



IBM PC
(1981) [1]

Algunos aciertos de los RDBMS

Basados en un **modelo formal**: el modelo relacional

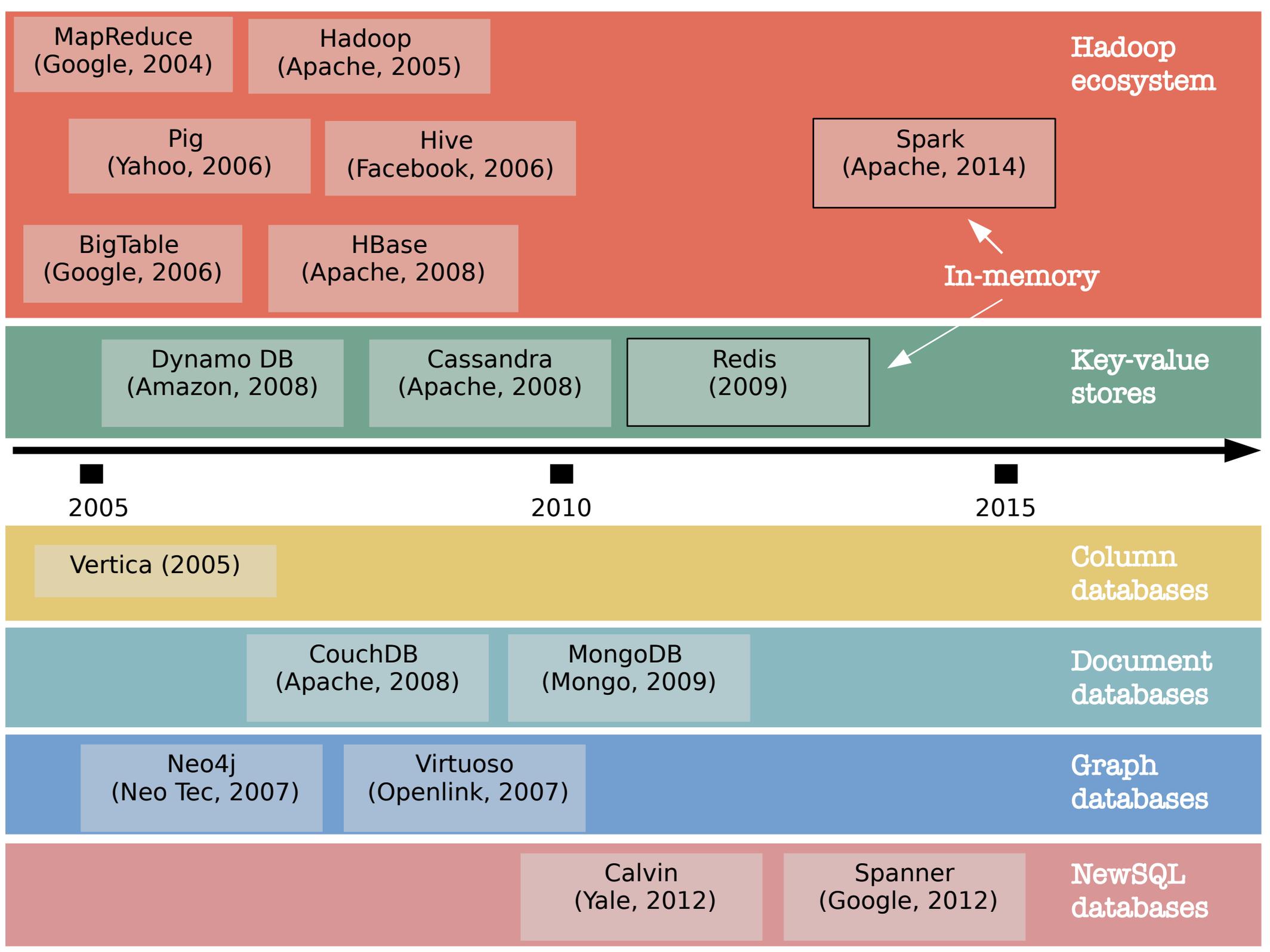
Tuplas, relaciones, restricciones

Un álgebra de operaciones

Soporte para **transacciones** (ACID)

El **lenguaje de consultas** SQL

Los RDBMS fueron LA herramienta de gestión de datos durante casi 30 años



MapReduce
(Google, 2004)

Hadoop
(Apache, 2005)

Hadoop
ecosystem

Pig
(Yahoo, 2006)

Hive
(Facebook, 2006)

Spark
(Apache, 2014)

BigTable
(Google, 2006)

HBase
(Apache, 2008)

In-memory

Dynamo DB
(Amazon, 2008)

Cassandra
(Apache, 2008)

Redis
(2009)

Key-value
stores

2005

2010

2015

Vertica (2005)

Column
databases

CouchDB
(Apache, 2008)

MongoDB
(Mongo, 2009)

Document
databases

Neo4j
(Neo Tec, 2007)

Virtuoso
(Openlink, 2007)

Graph
databases

Calvin
(Yale, 2012)

Spanner
(Google, 2012)

NewSQL
databases

Objetivos del curso

- Analizar los nuevos modelos de gestión de datos.
- Organizarlos, clasificarlos, y entender ventajas y desventajas de cada uno.
- Comprender los desafíos técnicos resueltos por los motores.
- Experimentar en su uso.

Forma de evaluación

- Dos tareas (en grupos de a dos):
 - Bases de datos documentales
 - Bases de datos de grafos
- Proyecto final (en grupos de a dos):
 - Trabajarán en el proyecto 7 semanas aprox.
 - Temas a definir (trabajaremos en la definición)
 - Informe final escrito y defensa presencial con preguntas individuales

Cronograma y forma de trabajo

- Clases hasta principios de junio
- Realizaremos actividades (cuestionarios, presentación de artículos, etc.)
- Tendrán que leer/ver el material indicado para cada tema/semana (ver EVA)
- Tendrán a disposición material para auto-estudio (videos, artículos, libro)

