

# Modelado y Procesamiento de Grandes Volúmenes de Datos

CPAP, FING, Udelar – 2020



Un poco de historia

# 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**

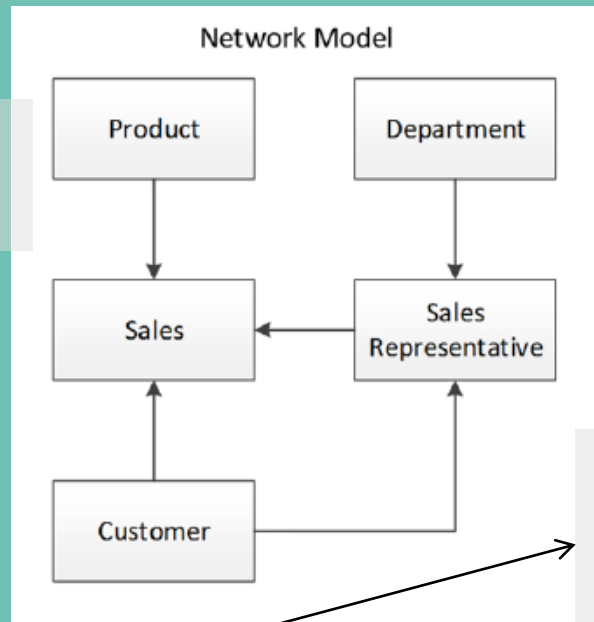
- el disco magnético aparece en los 50s!

# Los primeros DBMS (DataBase Management Systems)

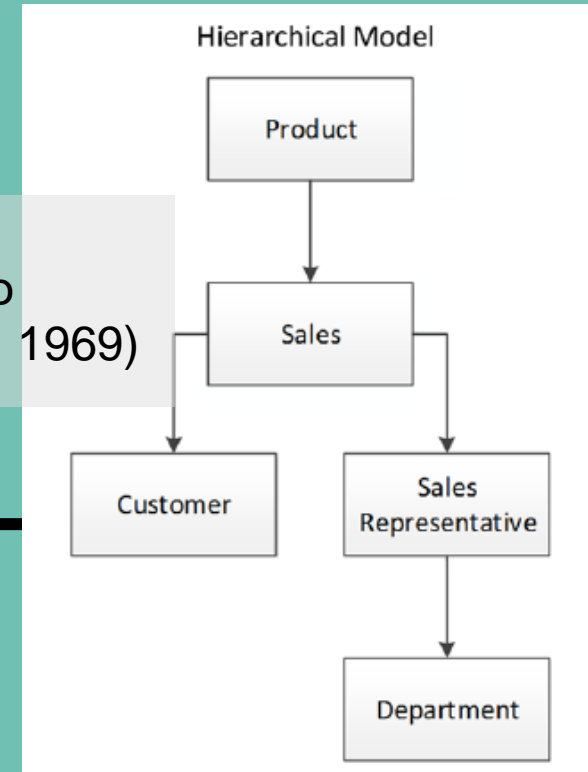
Programas dedicados a la gestión de datos:  
**desacoplar** de la lógica de la aplicación para permitir el **reuso**  
de esta lógica

Sistemas *mainframes*

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.
- Enfocados 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)

System R  
SEQUEL  
(1977, IBM)

IBM PC  
(1981) [1]

SQL-86  
(ANSI, 1986)

Comienzo de la Internet comercial (1995)

INGRES  
(1974, U. Berkeley)



[1] CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=51833>

1960

1970

1980

1990

2000

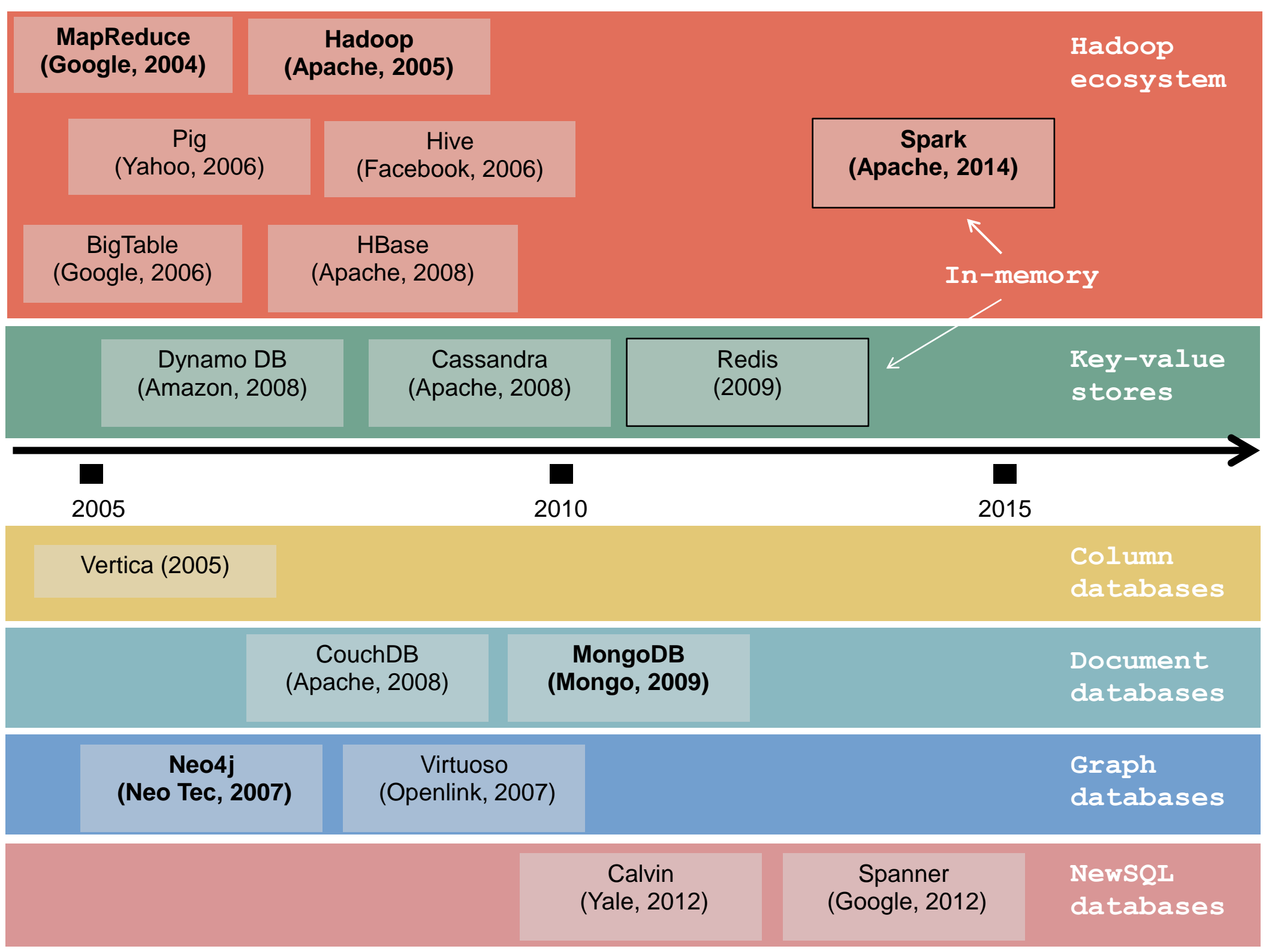
2010

# Algunos aciertos de los RDBMS

- Basados en un **modelo formal**: el modelo relacional
  - tuplas, relaciones, restricciones
  - un álgebra de operaciones (**álgebra relacional**)
- Soporte para **transacciones** (ACID)
- El **lenguaje de consultas** SQL

Los RDBMS han sido LA herramienta de gestión de datos durante más de 30 años





# Objetivos del curso

- Fundamentos de modelado relacional.
- Analizar 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.

# Cronograma y forma de trabajo

## •Clases teóricas y prácticas (remotas)

- Se recomienda leer el material indicado de forma previa a cada clase (ver EVA)

## •Clases de laboratorio (remotas)

## •Actividades

- Laboratorios (en grupos de 2 o 3 personas)
- Lectura de artículos

## •Defensa/Prueba final

- defensa con todo lo visto durante el curso (conforme parte la evaluación del curso junto con las evaluaciones de las actividades)