



UAB

Universitat Autònoma
de Barcelona

Tomás Margalef

El Ecosistema Hadoop

Introducción al Ecosistema Hadoop

Las cinco “V”s del **Big Data**:

- **Volumen**: Enorme cantidad de datos.
- **Velocidad**: Los datos se generan y se mueven a gran velocidad.
- **Variedad**: Los datos son de distintos tipos.
- **Veracidad**: El grado de confianza de los datos.
- **Valor**: La dificultad de explotación.




Introducción al Ecosistema Hadoop

En 2004 **Google** publicó un artículo describiendo la problemática para indexar la web y propuso una solución.

Esta solución se basaba en un sistema de ficheros (GFS) y en un modelo de programación denominado MapReduce



Introducción al Ecosistema Hadoop

Doug Cutting vio el potencial de la solución y en 2005 **YAHOO!** lanzó una implementación basada en esta propuesta a la que llamó  **hadoop**.

Hadoop es desarrollado y mantenido por The Apache software Foundation.



Introducción al Ecosistema Hadoop

Objetivos del ecosistema

- **Escalabilidad:** Capacidad de almacenar grandes volúmenes de datos.
- **Tolerancia a fallos:** Capacidad de recuperarse de errores hardware.
- **Tipos de datos:** Capacidad de gestionar datos de diferentes tipos.
- **Entorno compartido:** Capacidad de gestionar múltiples tareas de forma simultánea.
- **Aportar valor:** Capacidad de extraer valor de los datos.



Introducción al Ecosistema Hadoop

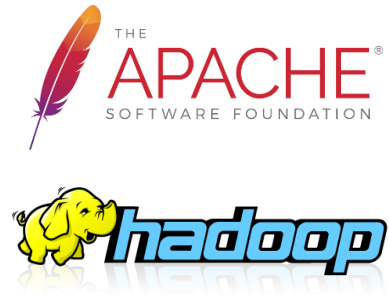
El ecosistema **Hadoop** incluye un conjunto de herramientas y aplicaciones para alcanzar estos objetivos.

Estas aplicaciones se pueden organizar en un diagrama de capas.

Las capas representan distintas interfaces, desde el almacenamiento a los lenguajes de alto nivel.



El Ecosistema Hadoop



Modelo ("*Framework*") de programación

Gestor de recursos en el cluster

MapReduce



YARN Yet Another Resource Negotiator



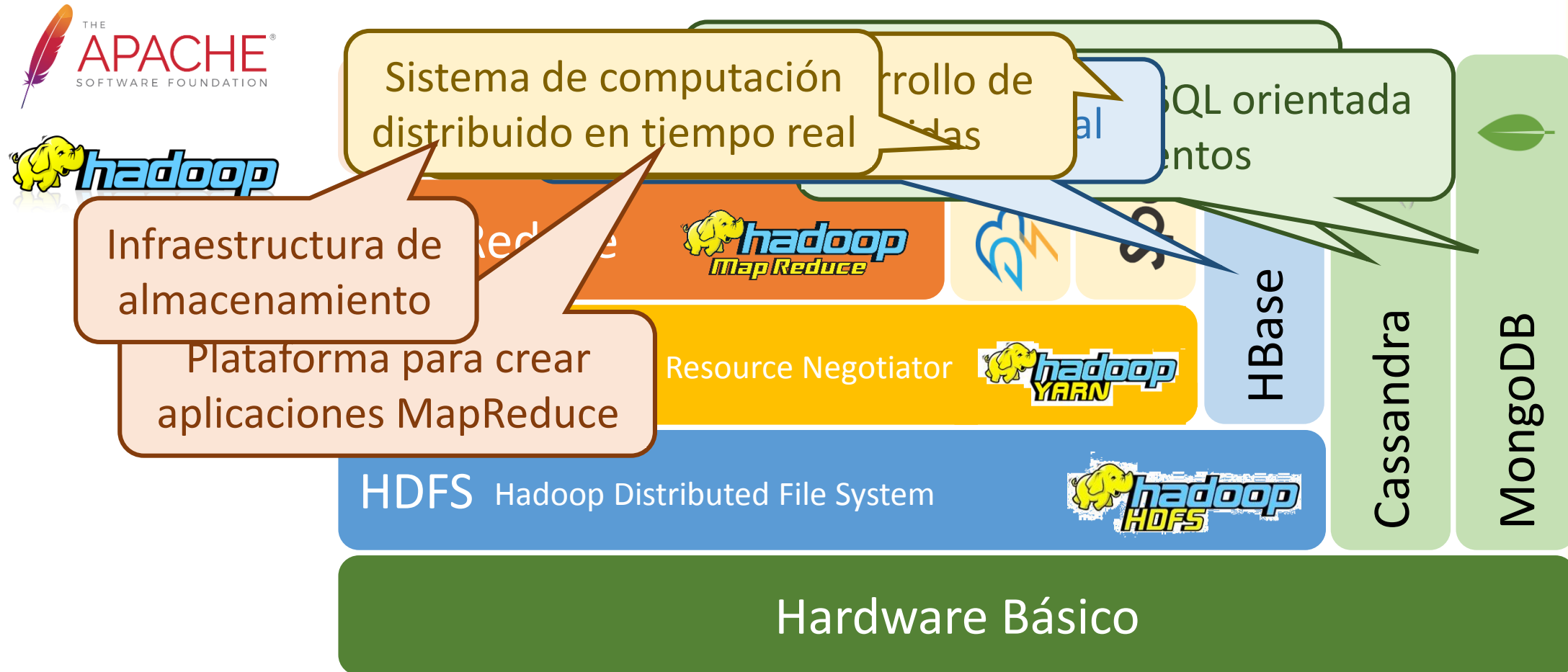
Sistema de archivos distribuido

HDFS Hadoop Distributed File System

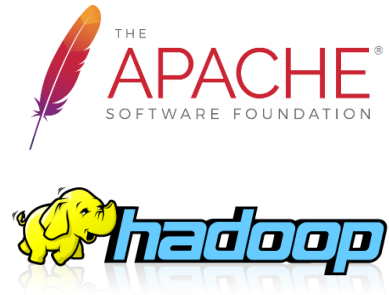


Hardware Básico

El Ecosistema Hadoop



El Ecosistema Hadoop



BIG DATA

El Ecosistema Hadoop



A green-tinted image of a tablet displaying the word "MOOC" in large, bold, white letters. The background is a solid light orange color.

MOOC

El Ecosistema Hadoop

A green-tinted image of a laptop keyboard with hands typing. Overlaid on the image is the UAB MOOC logo, which includes the text "UAB", "Universitat Autònoma de Barcelona", "MOOC", and "Escola de Postgrau".

UAB
Universitat Autònoma de Barcelona

MOOC
Escola de
Postgrau

UAB

Universitat Autònoma
de Barcelona