

# Formación senior en HBASE

PARTE DE LA FORMACIÓN BIG DATA ENGINEER DE BIG DATA  
ACADEMY PERÚ

---

# Concepto

---

BIG DATA  
ACADEMY

# HBase

---

Es una base de datos de **alta concurrencia**, **baja latencia**, **distribuida** y **NoSQL** que **se ejecuta sobre HDFS**, orientada a almacenar billones de registros. Se encuentra basada en la base de datos Big Table de Google y está **desarrollada en Java**.



---

# Objetivo fundamental

---

# Objetivo fundamental de HBase

---

1. Escritura rápida de múltiples peticiones sobre el disco duro, en el orden de micro-segundos
2. Lectura rápida de múltiples peticiones sobre el disco duro, en el orden de micro-segundos

---

# Naturaleza de funcionamiento

---



# Namespaces

Las tablas sobre HBase son almacenadas sobre un “namespace”. Físicamente estos namespace se encuentran como carpetas en HDFS bajo la ruta “/hbase/data”

## NAMESPACE

Row Key	contactinfo		profilephoto
	fname	lname	image
jdupont	Jean	Dupont	
jsmith	John	Smith	<smith.jpg>
mrossi	Mario	Rossi	<mario.jpg>

Row Key	contactinfo		profilephoto
	fname	lname	image
jdupont	Jean	Dupont	
jsmith	John	Smith	<smith.jpg>
mrossi	Mario	Rossi	<mario.jpg>

Row Key	contactinfo		profilephoto
	fname	lname	image
jdupont	Jean	Dupont	
jsmith	John	Smith	<smith.jpg>
mrossi	Mario	Rossi	<mario.jpg>

Row Key	contactinfo		profilephoto
	fname	lname	image
jdupont	Jean	Dupont	
jsmith	John	Smith	<smith.jpg>
mrossi	Mario	Rossi	<mario.jpg>

# Tablas

Una tabla es una agrupación de registros, sin embargo a diferencia de las tablas SQL, las tablas sobre HBase tienen una estructura distinta

Column Families		contactinfo		profilephoto
Column Names	Row Key	fname	lname	image
Rows	jdupont	Jean	Dupont	
	jsmith	John	Smith	<smith.jpg>
	mrossi	Mario	Rossi	<mario.jpg>

En cada familia de columnas, cada registro puede tener diferentes columnas

Para una misma tabla, cada familia de columna es un archivo diferente sobre HDFS



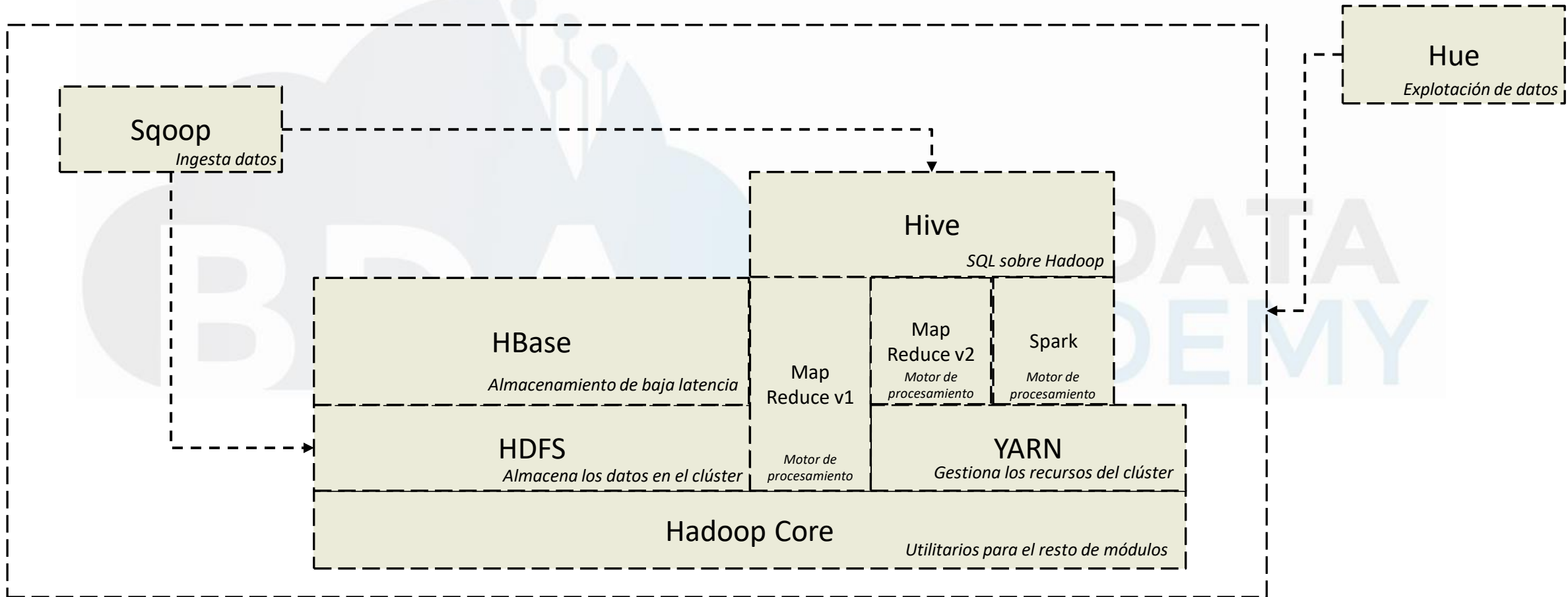
---

# Arquitectura de componentes

---



# Arquitectura de componentes

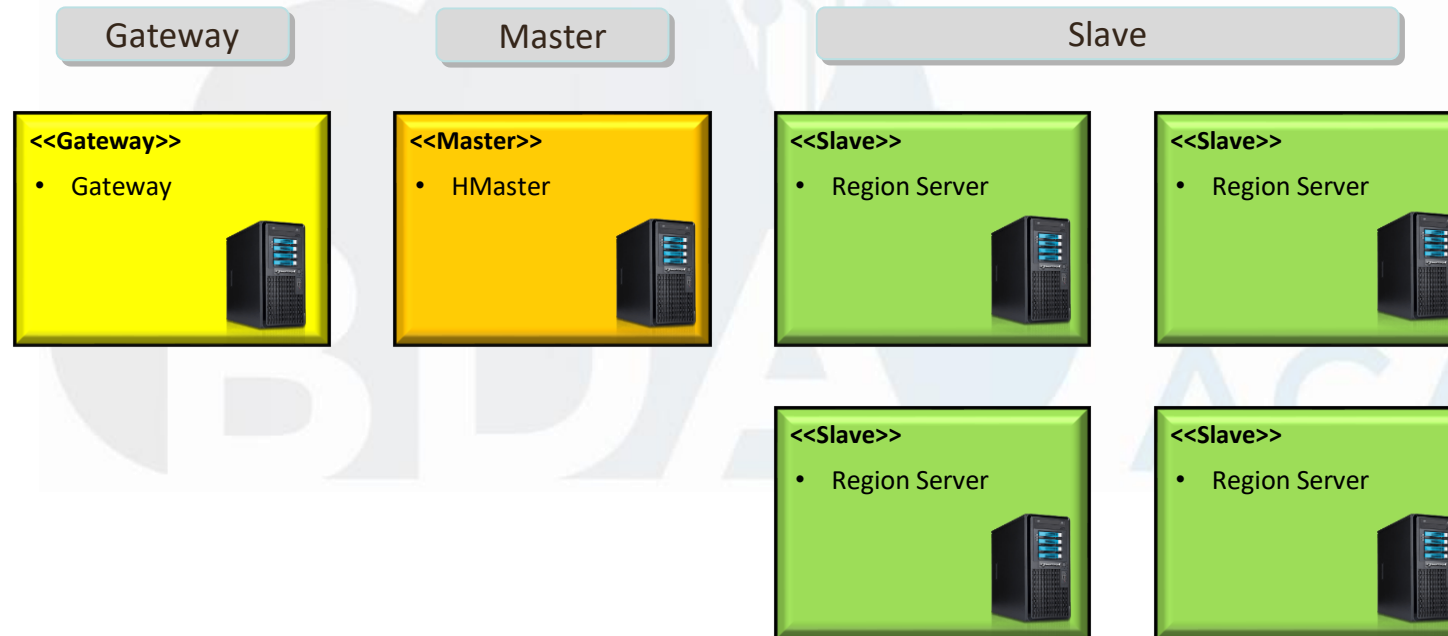


---

# Arquitectura de servicios

---

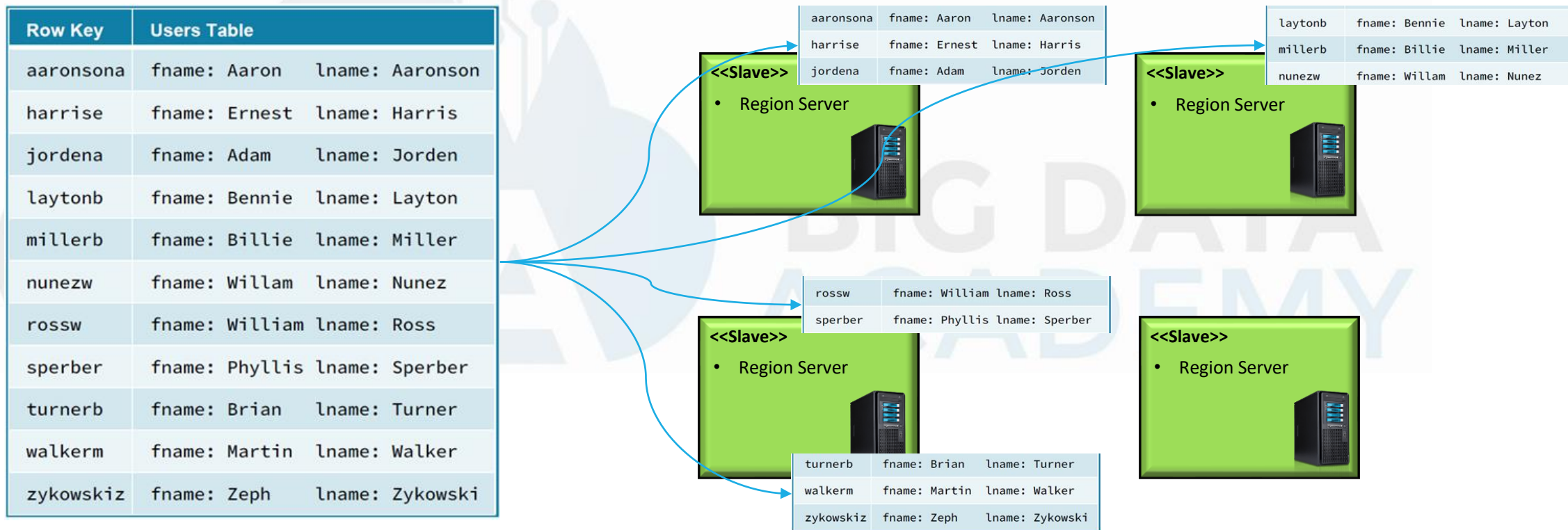
# Servicios de HBase



## DESCRIPCIÓN

- **Gateway:** Nodo con la configuración de HBase y acceso a las shells
- **HMaster:** Recibe las peticiones de almacenamiento y contiene la metadata
- **Regions Server:** Almacena los datos

# ¿Cómo almacena HBase?



---

# Ejercicios teóricos

---



# Ejercicios teóricos

---

1. ¿Qué es HBase?
- 2.Cuál es el objetivo de HBase
3. Describa cómo es una tabla en HBase
4. ¿Cuál es la importancia de la RowKey?
5. Dibuje la arquitectura de componentes del ecosistema Hadoop + HBase
6. ¿Cuáles son los servicios disponibles en HBase?
7. ¿Qué pasa si un Region Server cae?

---

# Resumen

---

BIG DATA  
ACADEMY



# Resumen

---

Hablemos...