

Hive

Hive

Hive

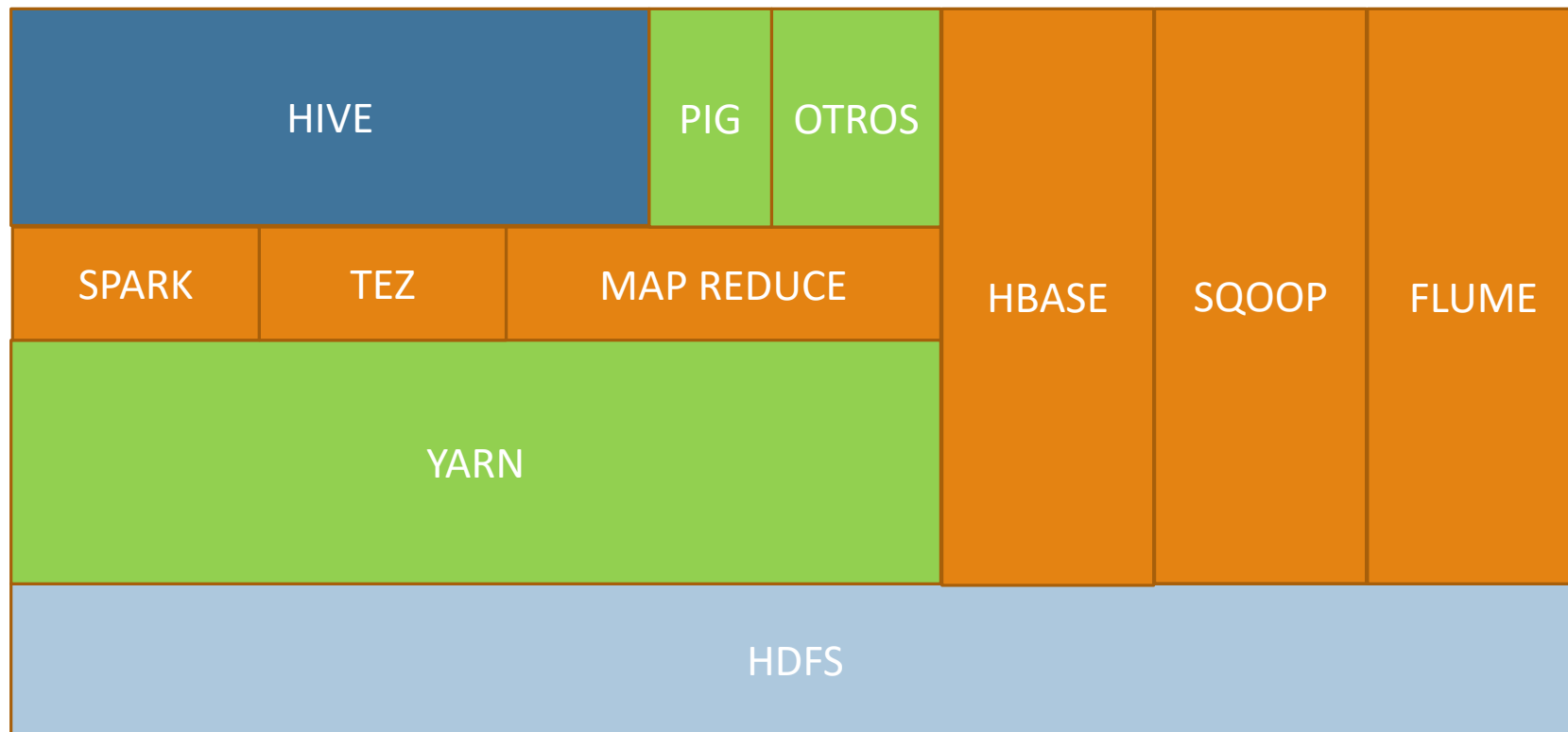
- ❑ Permite acceder a HDFS como si fuera una Base de datos
- ❑ Podemos ejecutar comandos muy parecidos a SQL para recuperar valores (HiveSQL)
- ❑ Esto simplifica enormemente el desarrollo y la gestión con Hadoop

<http://hive.apache.org/>



Hive

❑ Ecosistema hadoop



Hive

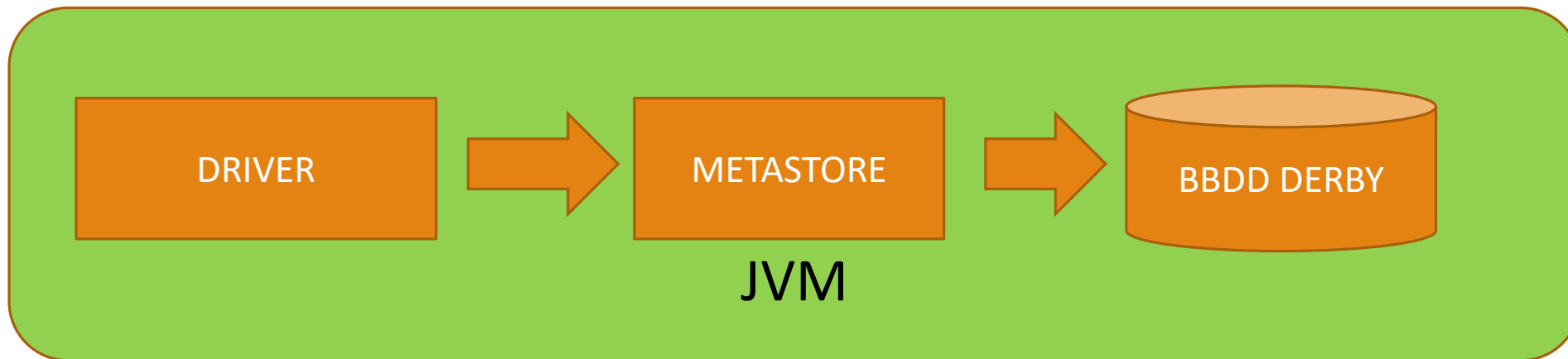
- ❑ Aunque en principio estaba diseñado para el procesamiento por lotes ahora se integra con frameworks de tiempo real Tez y Spark
- ❑ Utiliza un lenguaje de consulta llamado HiveQL
- ❑ Crear una especie de frontal que permite acceder a los datos almacenados en HDFS de forma muy sencilla sin necesidad de conocer Java, Map Reduce u otras tecnologías

Hive

- Metastore

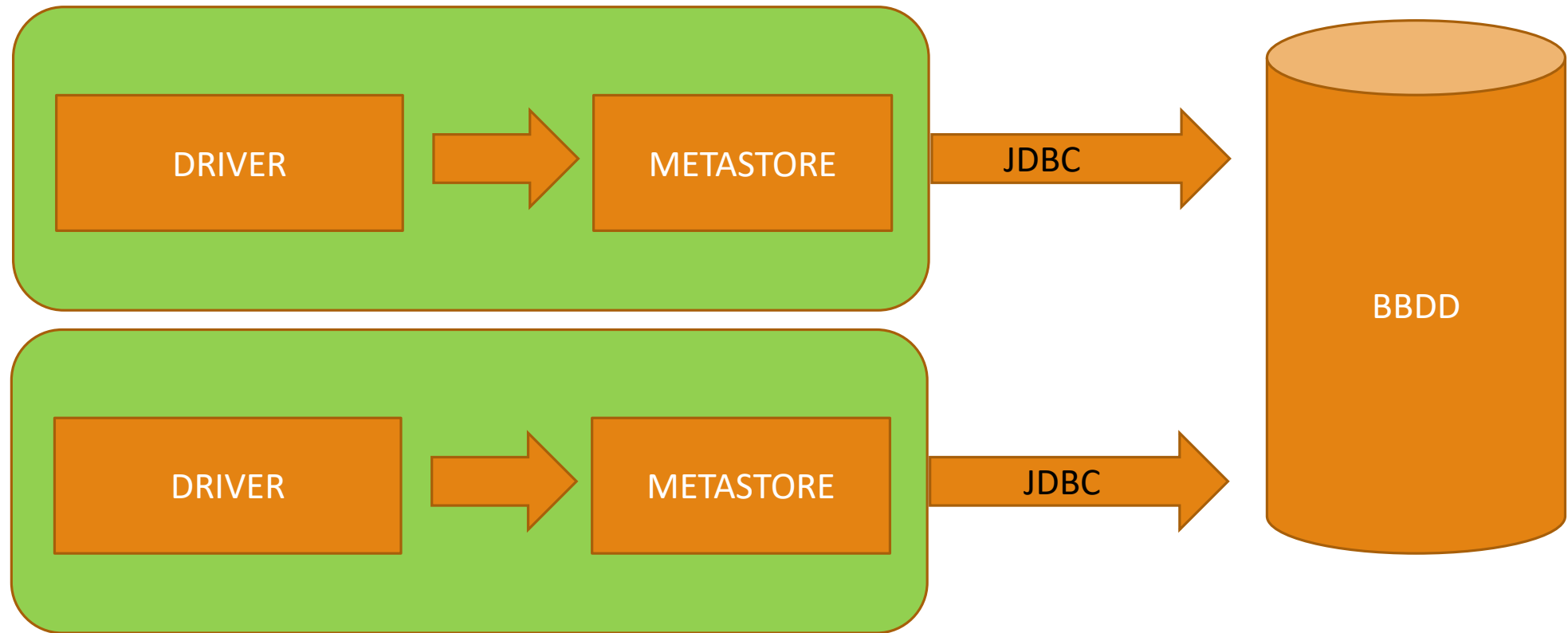
- Tiene un Metastore que es un repositorio central para los metadatos de Hive

- Embebido



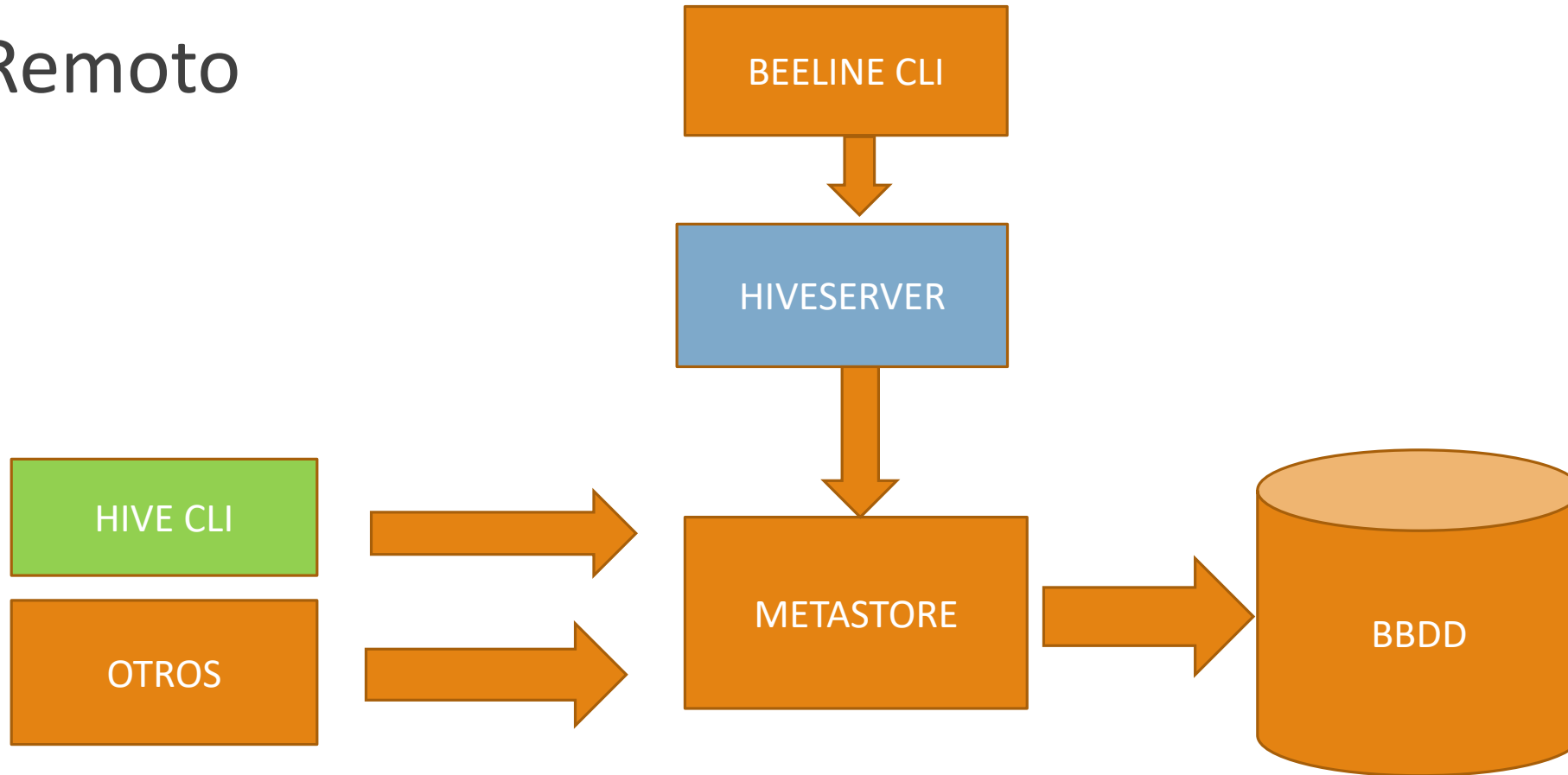
Hive

- Metastore
 - Local



Hive

- Metastore
 - Remoto



Hive

□ Tipos básicos

TYPE	DESCRIPTION	EXAMPLE
TINYINT	8-bit signed integer	1
SMALLINT	16-bit signed integer	1
INT	32-bit signed integer	1
BIGINT	64-bit signed integer	1
FLOAT	32-bit single precision floating point number	1.0
DOUBLE	64-bit double precision floating point number	1.0
BOOLEAN	true/false value	TRUE
STRING	Character string	'a','a'
TIMESTAMP	Timestamp with nanosecond precision	'2012-01-02 03:04:05.123456789'

Hive

❑ Tipos complejos

TYPE	DESCRIPTION	EXAMPLE
ARRAY	Colección ordenada de campos. Los campos deben ser del mismo tipo	<code>array(1, 2)</code>
MAP	Colección no ordenada de pares clave-valor. Las claves deben ser primitivas y los valores de cualquier tipo.	<code>map('a', 1, 'b', 2)</code>
STRUCT	Colección de campos con nombre. Los campos pueden ser de distinto tipo.	<code>struct('a', 1, 1.0)</code>