

Big Data Aplicado

Josep Garcia

garcia_jos19@ieseduardoprimo.es



IES EDUARDO
PRIMO MARQUÉS





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

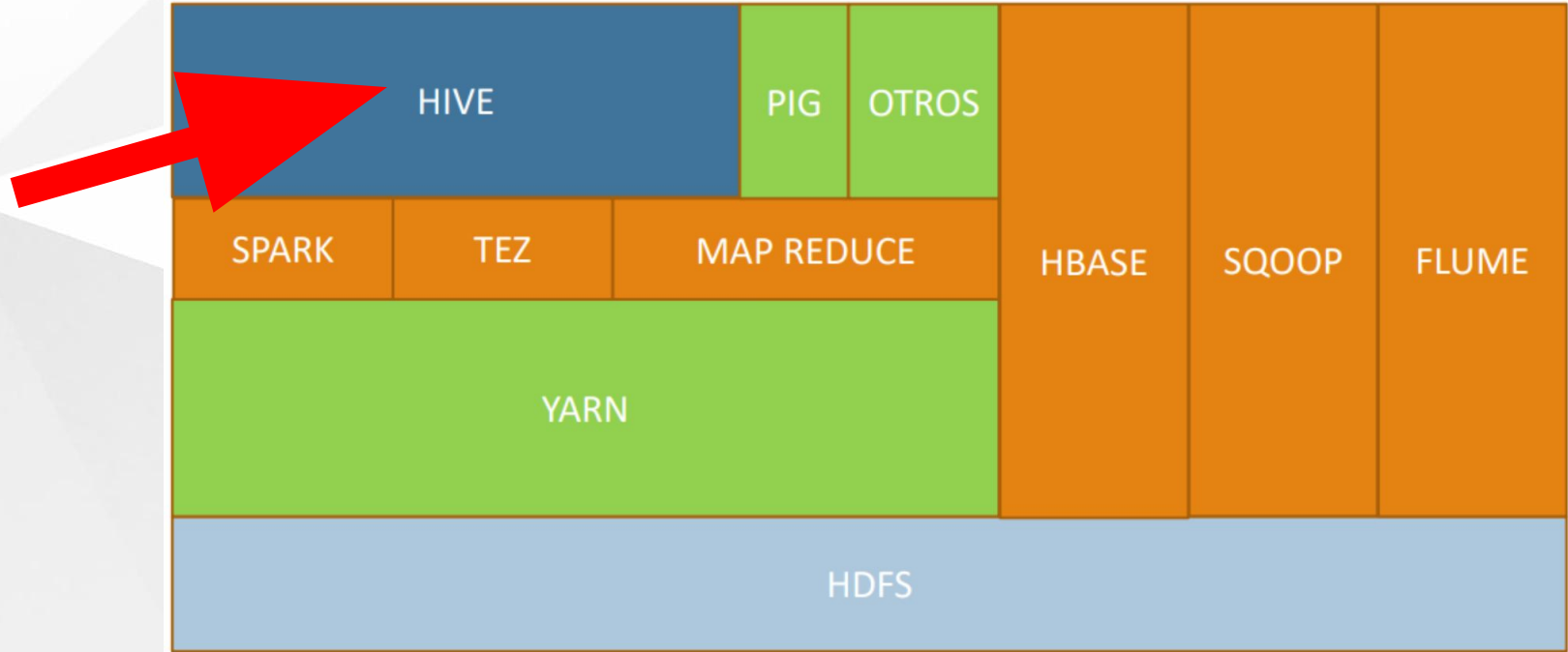
<http://hive.apache.org>



IES EDUARDO
PRIMO MARQUÉS



HIVE





HIVE

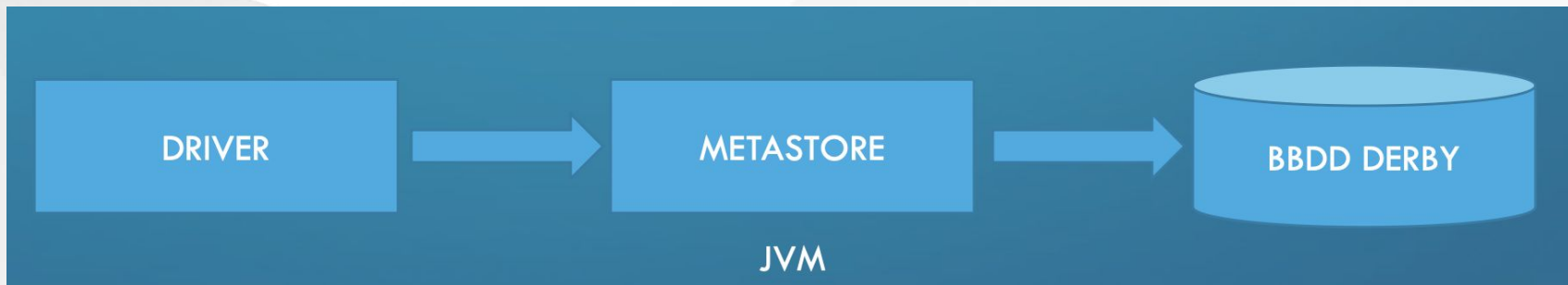
- Aunque en principio estaba diseñado para el procesamiento por lotes ahora se integra con frameworks de tiempo real Tez y Spark.
- Utiliza un lenguaje de consulta llamado HiveSQL
- Crear una especie de frontal que permite acceder a los datos almacenados en HDFS de forma muy sencilla sin necesidad de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



HIVE

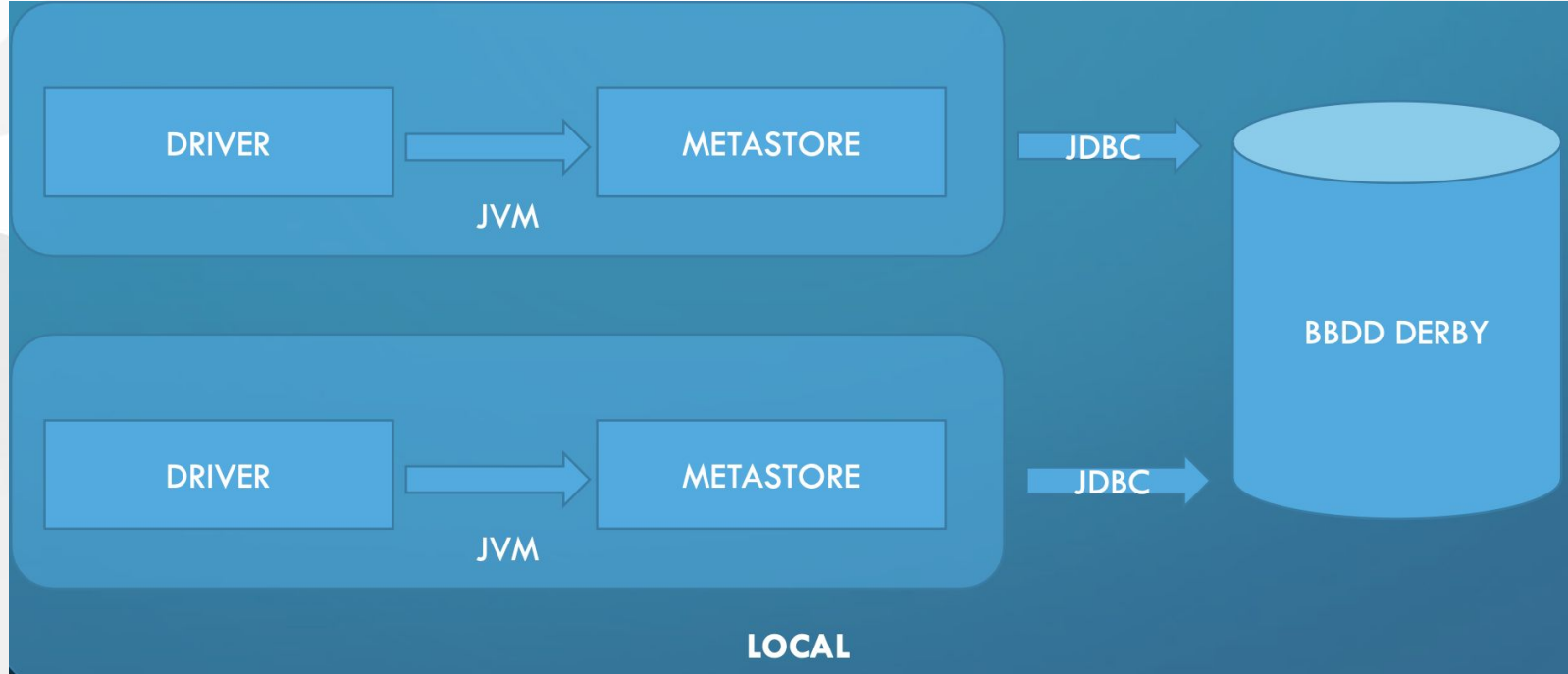
- **Metastore:** es un repositorio central para los metadatos de Hive.

Modo embebido (1 usuario)



HIVE

Modo local, varios usuarios



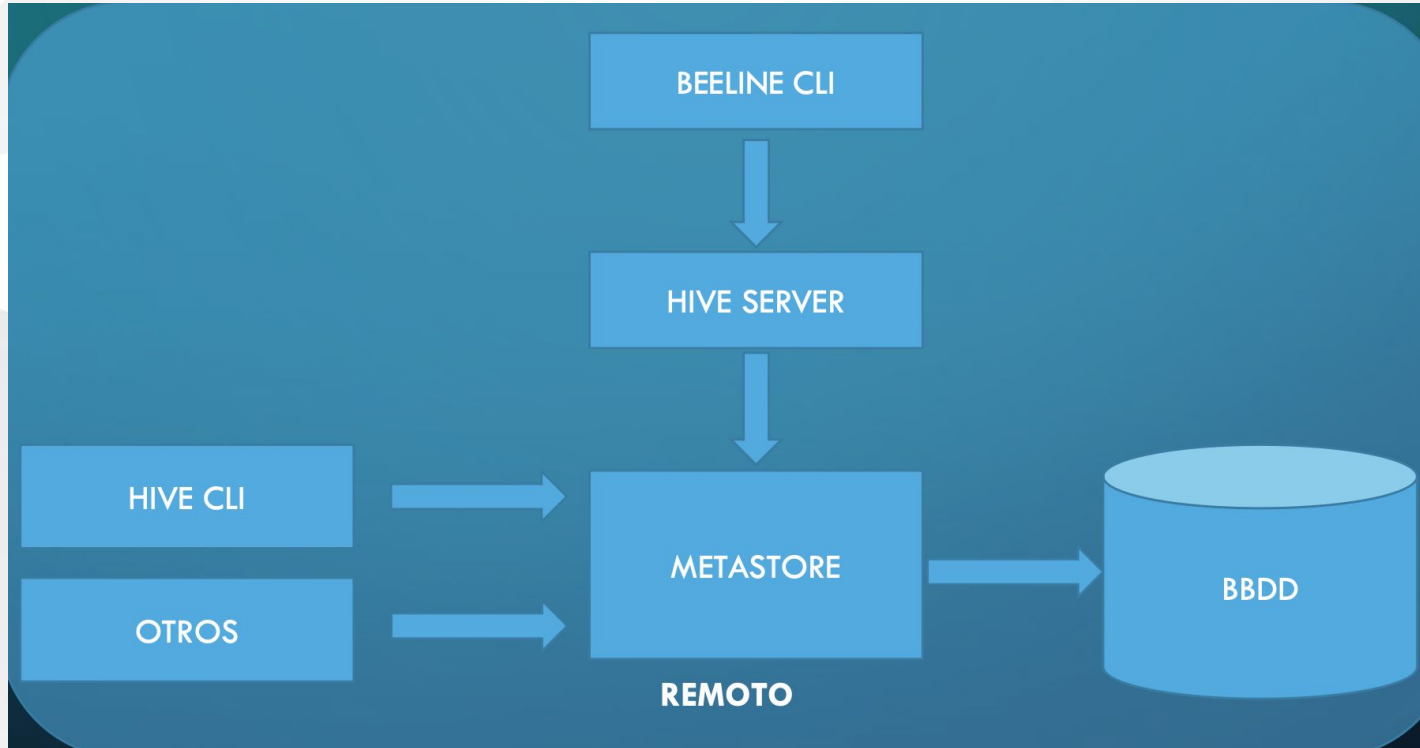
IES EDUARDO
PRIMO MARQUÉS



HIVE



Modo remoto



IES EDUARDO
PRIMO MARQUÉS





HIVE - Tipos de datos

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 de datos

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>





HIVE - Comandos HiveSQL

DDL (Data Definition Language):

- CREATE TABLE
- CREATE/DROP/ALTER/USE DATABASE
- CREATE/DROP/TRUNCATE TABLE
- ALTER TABLE/PARTITION/COLUMN
- CREATE/DROP/ALTER VIEW
- CREATE/DROP/ALTER INDEX
- CREATE/DROP MACRO

...



IES EDUARDO
PRIMO MARQUÉS





HIVE - Comandos HiveSQL

DML (Data Manipulation Language):

- LOAD
- INSERT
- UPDATE
- DELETE
- MERGE

...





HIVE - Comandos HiveSQL

CONSULTA

SELECT

[ALL | DISTINCT]

select_expr,

select_expr, ...

FROM table_reference

FROM table_reference

[WHERE where_condition]

[GROUP BY col_list]

[HAVING having_condition]

[CLUSTER BY col_list] | [DISTRIBUTE BY col_list]

[SORT BY col_list]] [LIMIT number] ;

