



Apache Hive

HDFS query engine



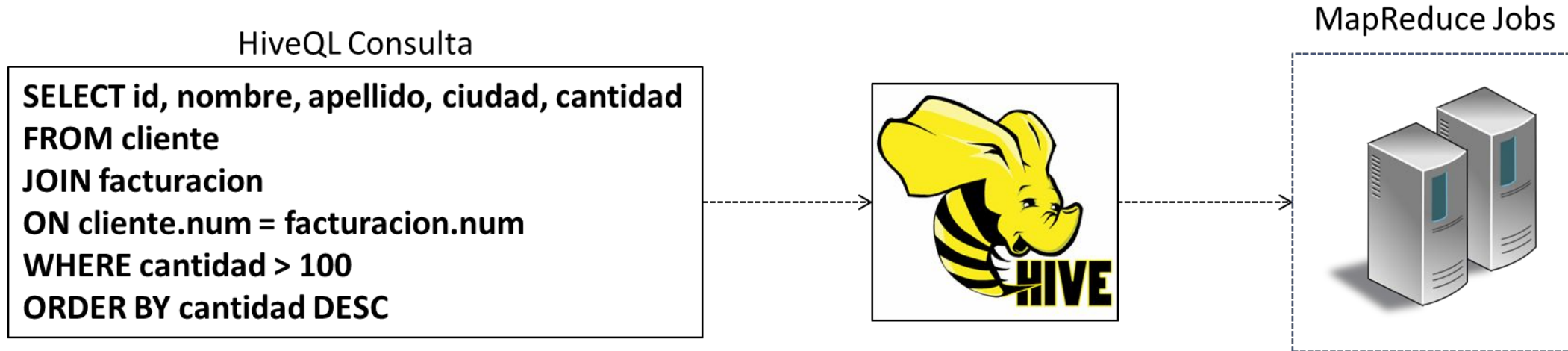
❖ ¿Qué es Apache Hive?

❖ Accediendo a Hive

❖ Creación y población de tablas



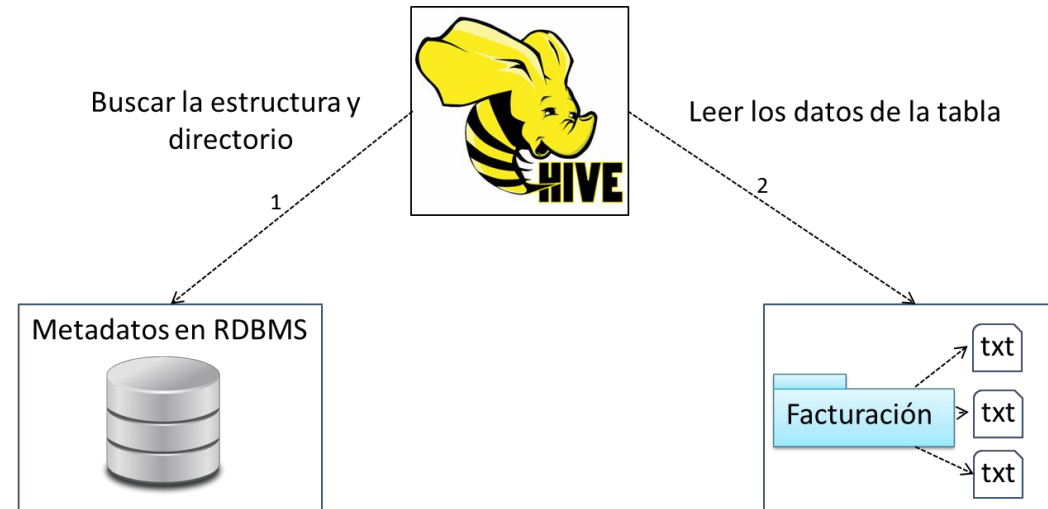
- Hive es una abstracción de alto nivel de MapReduce:
 - Utiliza un lenguaje llamado HiveQL.
 - Genera código MapReduce que es ejecutado en un clúster Hadoop.





Apache Hive – Almacenamiento de datos

- Las consultas de Hive operan con tablas, como en una base de datos relacional.
 - Las tablas de Hive se corresponden con un directorio de datos en HDFS.
- En cuanto a la localización y estructura:
 - Son especificados en la creación de la tabla.



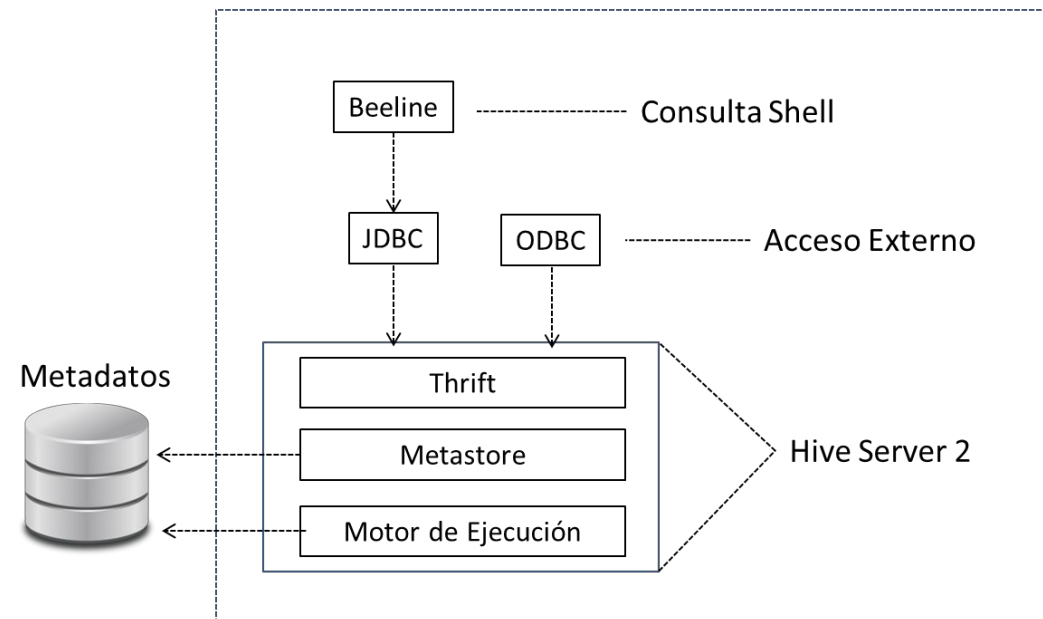


- ❖ ¿Qué es Apache Hive?
- ❖ Accediendo a Hive
- ❖ Creación y población de tablas



Apache Hive – Arquitectura Básica

- Se dispone de un sistema llamado *HiveServer 2* que es un contenedor del motor de ejecución de Hive:
- Accesible vía Beeline (comandos de consola), JDBC, ODBC, o Hue.





- **SHOW DATABASES:** lista todas las bases de datos
- **USE DATABASE:** ‘entrar’ en una bbdd para explorar sus tablas
- **SHOW TABLES:** lista todas las tablas que se encuentra en la base de datos
- **DESCRIBE TABLE:** lista los campos de una tabla especifica



- ❖ ¿Qué es Apache Hive?
- ❖ Accediendo a Hive
- ❖ Creación y población de tablas



- Crear una tabla con ficheros secuenciales también es sencillo.
 - Por defecto es *STORED AS TEXTFILE*, por lo que no es necesario especificarlo.

```
CREATE TABLE productos (  
    id INT,  
    nombre STRING,  
    precio INT,  
    fecha_añadido TIMESTAMP  
)  
    STORED AS SEQUENCEFILE; --STORED AS TEXTFILE  --STORED AS PARQUET
```