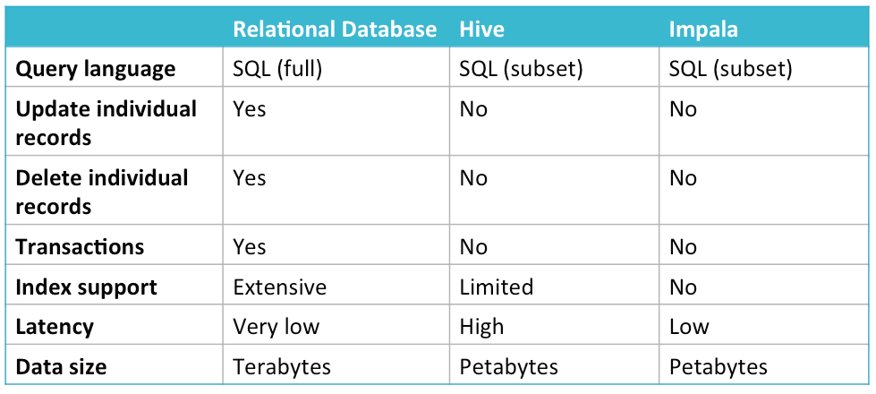
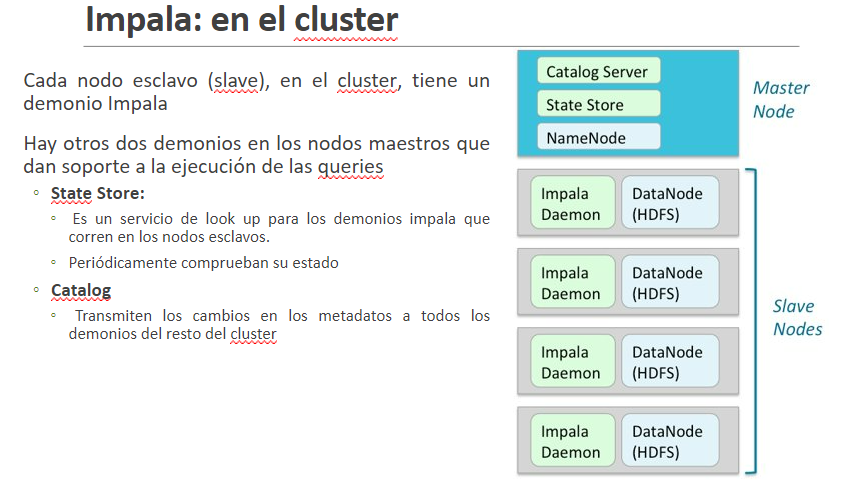
**APuntes**

* La diferencia entre Impala y Hive es que Impala ejecuta las queries directamente en un clúster y hive en un MapReduce. Como podemos recordar, en los ejericicos de Hijve, ejecutaban un reducer por cada agrupación (sum, count, etc).
* Por ello Impala es mucho más rápido de Hive o Pig
* Para utilizar Impala no es necesario saber programar. Únicamente es necesario saber realizar queries.
* Impala, como Hive, operan sobre los mismos datos: tablas en HDFS, y metadatos en el Metastore.





Procedimiento de ejecución Impala:

* Es el demonio Impala quien planifica la query.
* El cliente (impala-Shell o Hue) se conecta con un demonio local impala, llamado coordinador
* El coordinador pide una lista de demonios impala activos al State Store
* El coordinador distribuye la query entre los demonios impala disponibles, siempre intentando mantener el Data Locality
* Cada demonio va devolviendo el resultado al coordinador y este lo combina y lo muestra al cliente