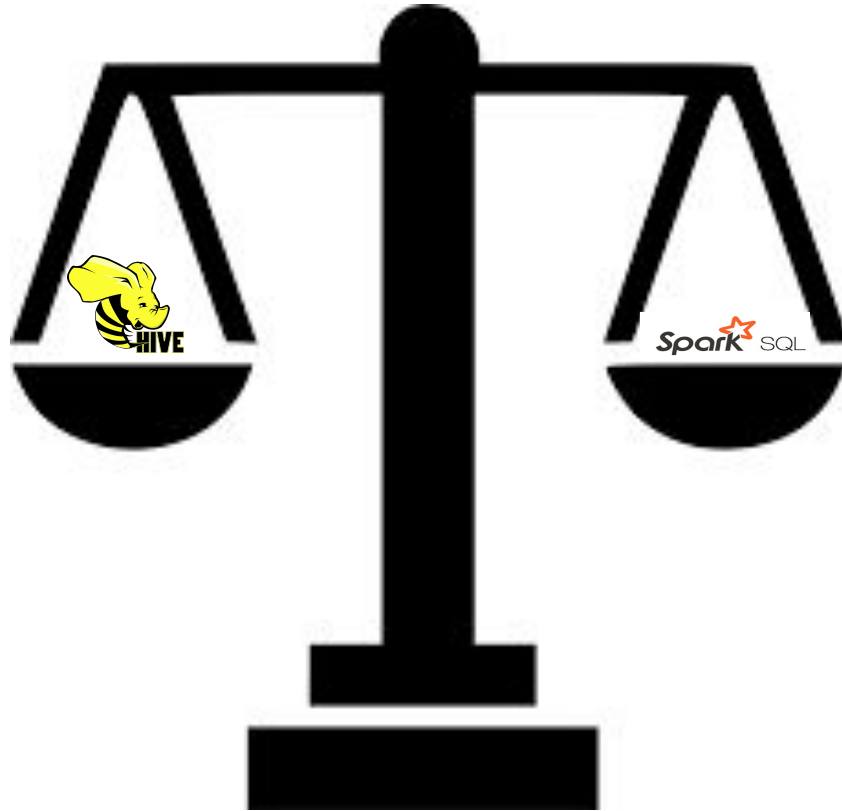


Why do we need Spark SQL?

Why do we need Spark SQL?



Spark SQL was built to overcome the limitations of Apache Hive running on top of Apache Spark

To provide a easy programming interface for non programmers

Limitations of Hive



Hive uses Mapreduce which lags in performance with medium and small sized datasets

Hive cannot drop encrypted databases

Spark SQL advantages over Hive

- Faster execution

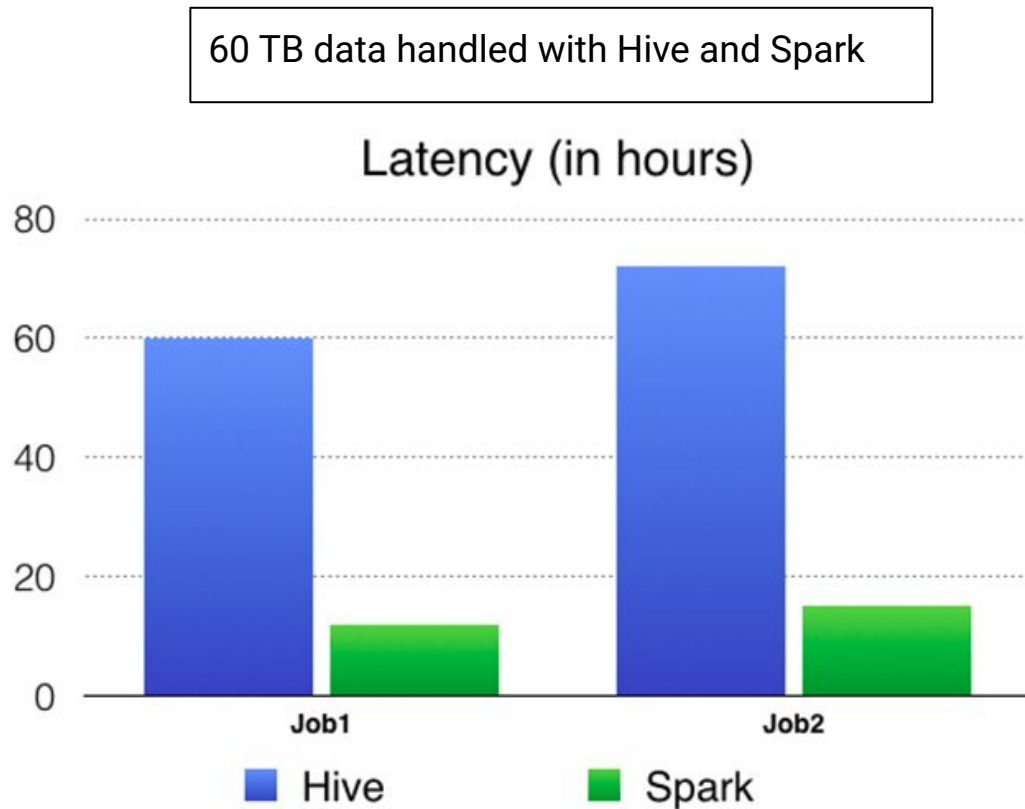


60 hours



10 hours

Apache Spark: A 60 TB+ production use case



Spark SQL advantages over Hive

- Faster execution
- No migration hurdles
- Real time querying



60 hours



10 hours



Spark SQL uses metastore service of Hive to query the data stored and managed by Hive.

Thank You!!

