

## ***1.##What is Spark SQL?***

***Spark SQL is a Spark module for structured data processing.***

***It provides a programming abstraction called DataFrames and can also act as a distributed SQL query engine.***

***It enables unmodified Hadoop Hive queries to run up to 100x faster on existing deployments and data.***

## ***2.##Is there a module to implement SQL in Spark? How does it work?***

***PySpark SQL is a module in Spark which integrates relational processing with Spark's functional programming API.***

***We can extract the data by using an SQL query language. We can use the queries same as the SQL language.***

### ***3.##What is a Parquet file?***

***Apache Parquet is an open source,column-oriented data file format designed for efficient data storage and retrieval.***

***It provides efficient data compression and encoding schemes with enhanced performance to handle complex data in bulk.***

### ***4.##List the functions of Spark SQL.***

***1.String Functions.***

***2.Date & Time Functions.***

***3.Collection Functions.***

***4.Math Functions.***

***5.Aggregate Functions.***

***6.Window Functions.***

### ***5.##How is Spark SQL different from HQL and SQL?***

***Hive,on one hand,is known for its efficient query processing by making use of SQL-like HQL(Hive Query Language)***

***and is used for data stored in Hadoop Distributed File System whereas Spark SQL makes use of structured query language***

***and makes sure all the read and write online operations are taken care of.***

### ***6.##Why is Spark SQL used?***

***Spark provides a faster and more general data processing platform.***

***Spark lets you run programs up to 100x faster in memory, or 10x faster on disk, than Hadoop.***

***7.## Is Spark SQL faster than Hive?***

***Speed:—The operations in Hive are slower than Apache Spark***

***in terms of memory and disk processing as Hive runs on top of Hadoob.***