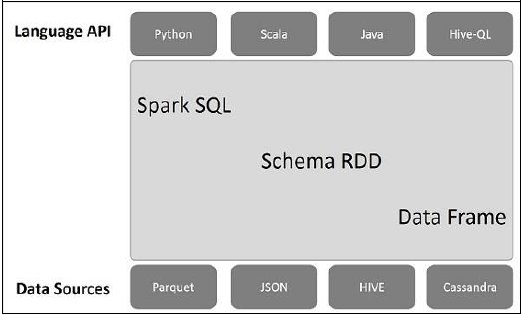
# Spark SQL

**Intro**

* Provides a programming abstraction called DataFrame and can act as distributed SQL query engine



* A DataFrame is a distributed collection of data, which is organized into named columns.
* SQLContext is a class and is used for initializing the functionalities of Spark SQL.
* Hive is a data warehouse infrastructure tool to process structured data in Hadoop.