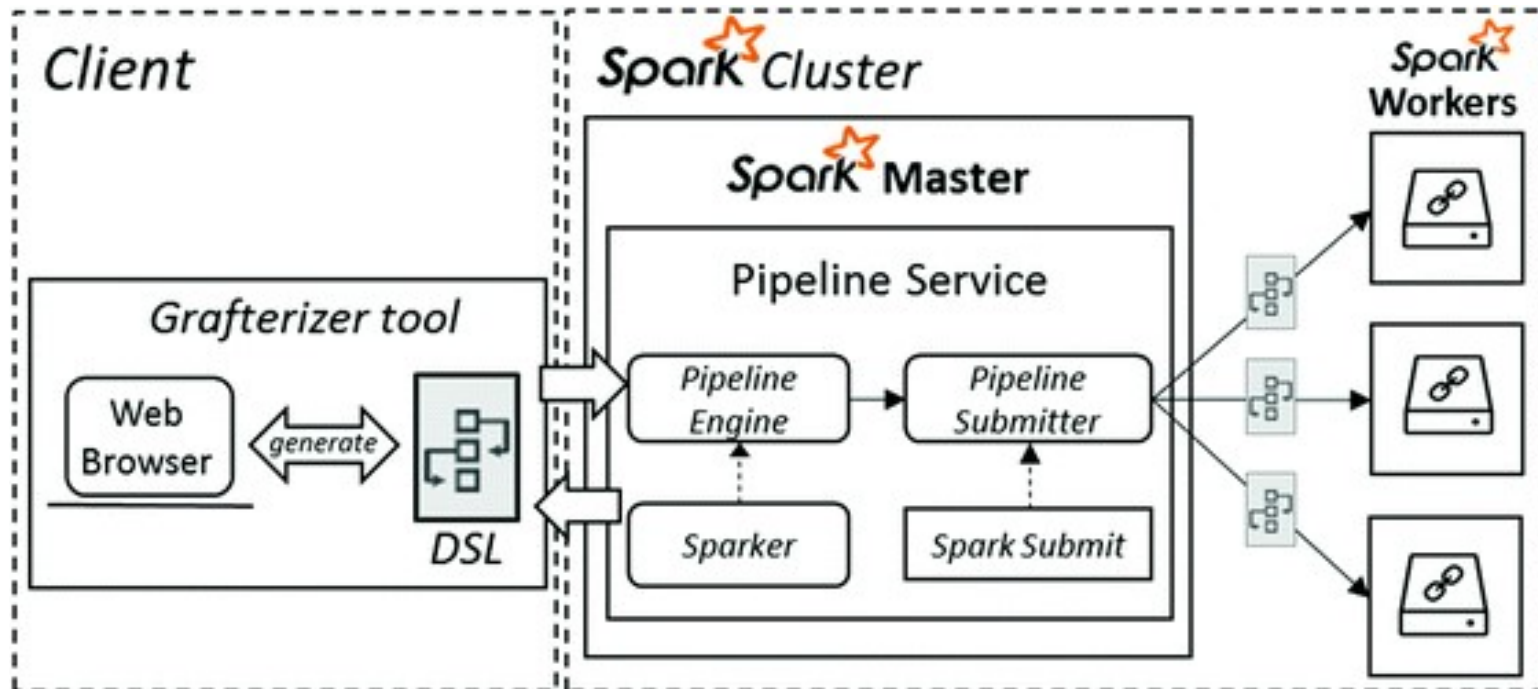


Apache Spark

Apache Spark

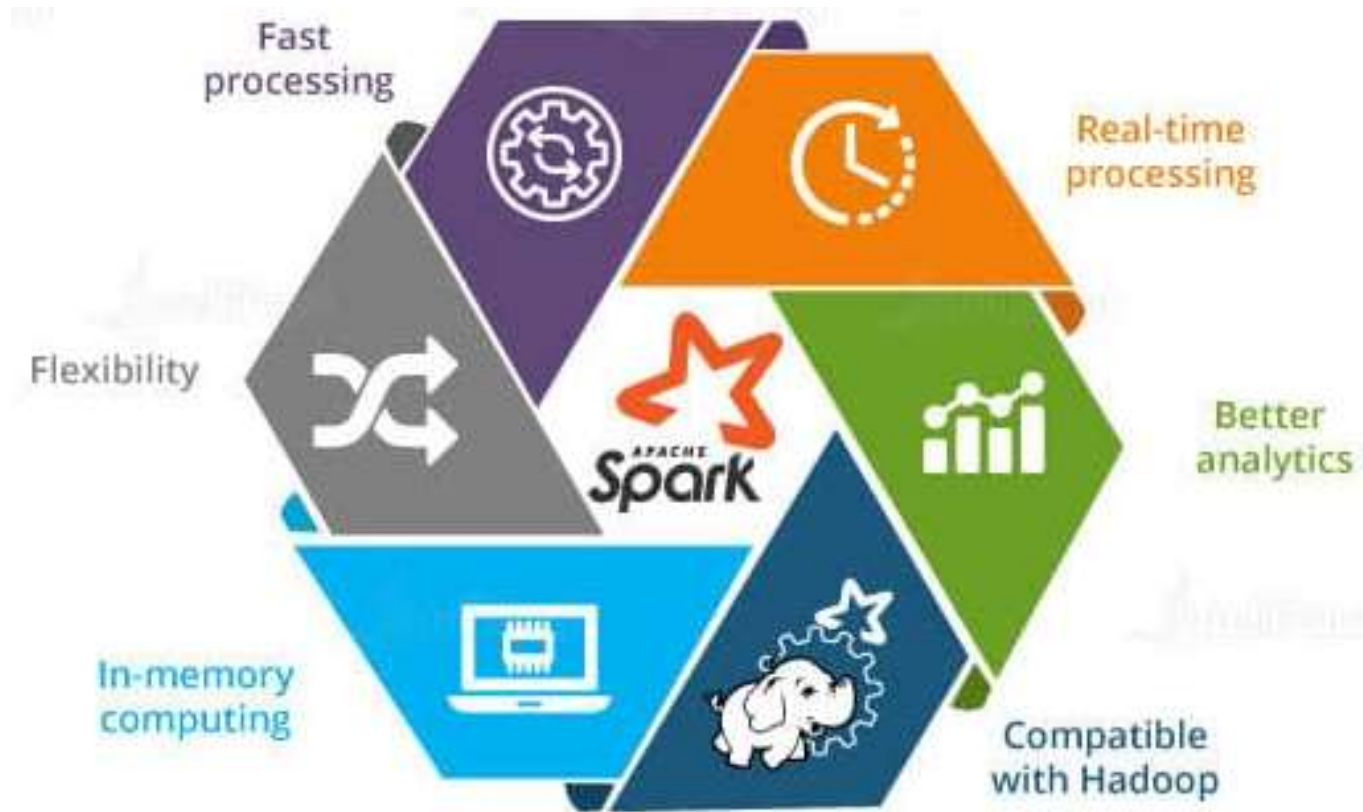
- Apache Spark is a lightning-fast cluster computing technology, designed for fast computation.
- The main feature of Spark is its in-memory cluster computing that increases the processing speed of an application.



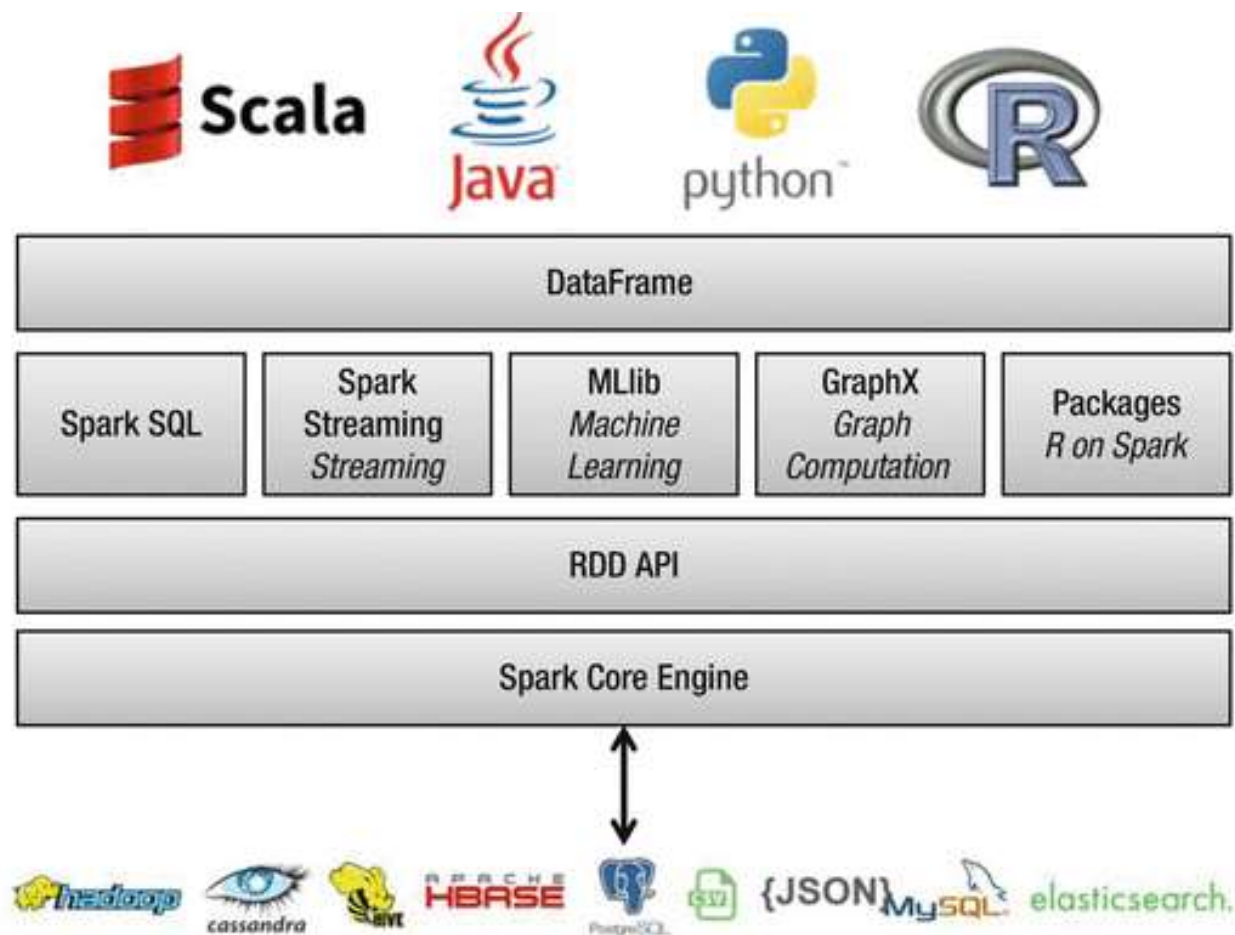
Features of Apache Spark

- Speed
 - Helps to run an application up to 100 times faster
- Supports multiple languages
 - Provides built-in APIs in Java, Scala, or Python.
- Advanced Analytics
 - Supports 'Map' and 'reduce', SQL queries, Streaming data, Machine learning (ML), and Graph algorithms.

Features of Apache Spark



Components of Spark



Components of Spark

- Apache Spark Core
 - Is the underlying general execution engine for spark platform that all other functionality is built upon.
- Spark SQL
 - A component on top of Spark Core that introduces a new data abstraction which provides support for structured and semi-structured data.
- Spark Streaming
 - Performs streaming analytics.
 - It ingests data in mini-batches and performs RDD (Resilient Distributed Datasets) transformations on those mini-batches of data.
- MLlib (Machine Learning Library)
 - MLlib is a distributed machine learning framework.
- GraphX
 - GraphX is a distributed graph-processing framework.

Thanks