APACHE SPARK

Apache Spark is a unified analytics engine for large-scale data processing.

Apache Spark achieves high performance for both batch and streaming data, using a state-of-the-art DAG scheduler, a query optimizer, and a physical execution engine.

Spark offers over 80 high-level operators that make it easy to build parallel apps. And you can use it *interactively* from the Scala, Python, R, and SQL shells.

Apache spark	Hadoop
Faster Good Nice	Slow
Real time processing	Batch processing
It has RDD	It doesn't have RDD
Low latency	High latency Does not process data interactively External job scheduler
Fault tolerant	Fault tolerant

Cost

- **Apache Spark –** As spark requires a lot of RAM to run in-memory. Thus, increases the cluster, and also its cost.
- **Hadoop MapReduce –** MapReduce is a cheaper option available while comparing it in terms of cost.

xv. Language Developed

- **Apache Spark –** Spark is developed in **Scala**.
- **Hadoop MapReduce –** Hadoop MapReduce is developed in **Java.**