

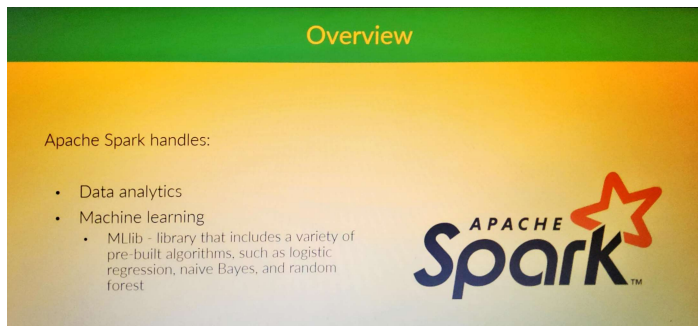
Spark in databricks

Thursday, 5 January 2023 3:47 PM

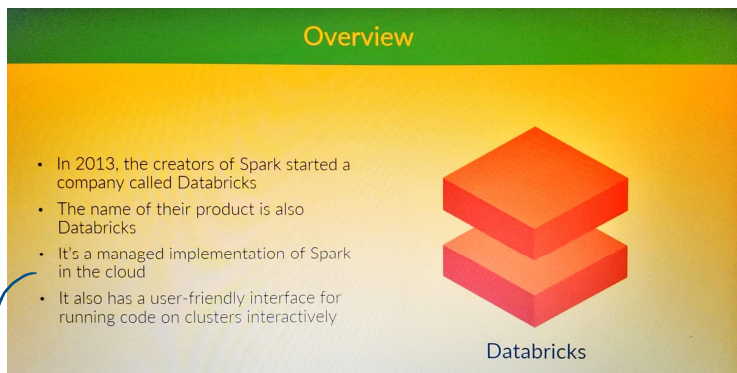
Overview

Spark.

- It is an open source framework for doing big data processing.
- It was developed as a replacement for Apache Hadoop's Map Reduce Framework.
 - Both Spark and MapReduce process data on compute clusters.
 - But one of Spark's big advantages is that it does in-memory processing, which can be orders of magnitude faster than the disk based processing that MapReduce uses.

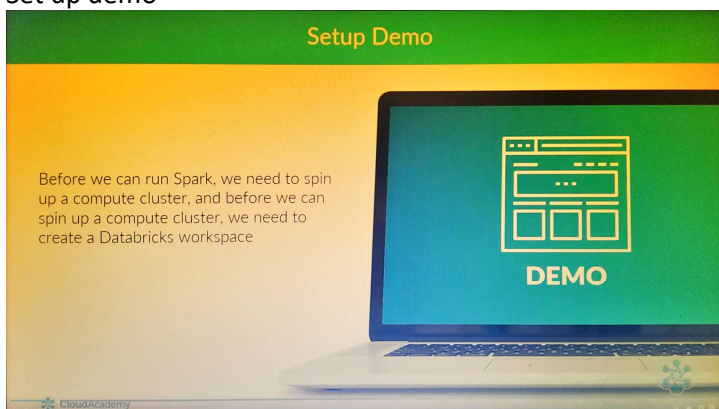


It doesn't include M.N. but you can use another framework on it, such as TensorFlow.



→ You don't need to worry about clusters. Microsoft is partner of Databricks, to integrate it with Azure Services.

Set up demo



<https://github.com/cloudacademy/azure-databricks>.

From <https://cloudacademy.com/course/running-spark-on-azure-databricks/notebooks/?context_resource=lp&context_id=3191>