



Azure HD Insight

Concepts

- Big Data processing services
 - Spark processing engine
 - Clustered distributed model
 - Used instead of Azure Synapse Analytics
- Streaming => Apache Kafka/Hadoop (open source framework with MPP) [Hadoop Map/Reduce/Spark/Hive/Kafka/Storm/R]
- Stores data in Azure Data Lake

Technologies

- Hadoop Map/Reduce => MPP, Java code -> Hadoop cluster (replaced Spark)
- Apache Hive => interactive SQL-like query facility (polybase), separate nodes
- Apache Kafka => clustered stream & scaling
- Apache Storm => real-time processing, scale, fault tolerance, reliable
 - Increased data volume @ moderate CPU
 - Replay unsuccessful data procedure
 - Interoperable with Azure Event Hubs/ IoT Hub/Kafka/RabbitMQ (Message queue service)
 - Write data to HDFS/HBase/Redis/SQL DB/APIs



Compiled by Dr T Oberholster
DP-900 version: March 2021

[Website & Blog](#)
[Amazon Books](#)
[Teepublic Shirts](#)
[RedBubble Merch](#)
[SpreadShirt Apparel](#)
[Contrado Fashion](#)