



Big Data Technologies

Trainer: Mr. Nilesh Ghule.

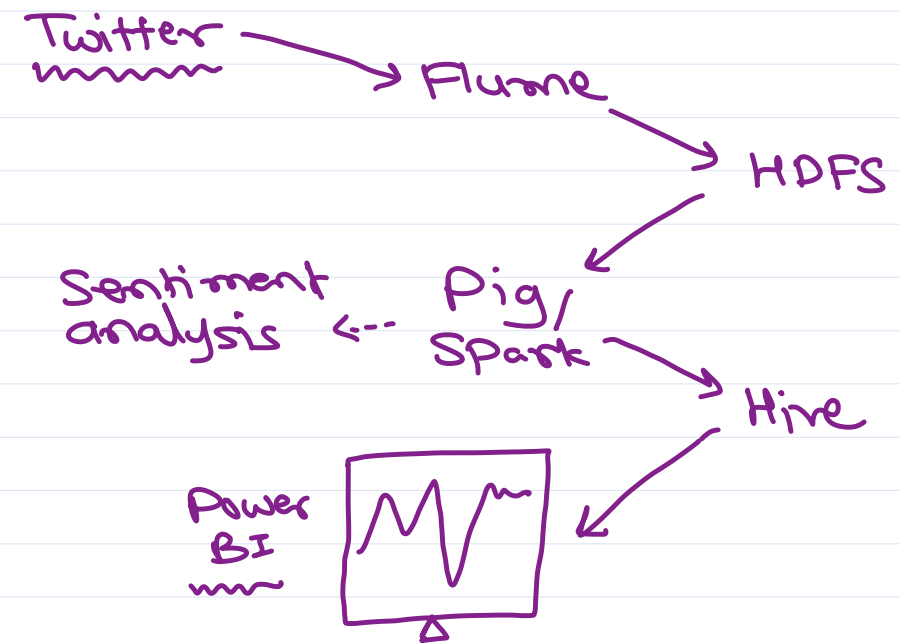


Hadoop Ecosystems

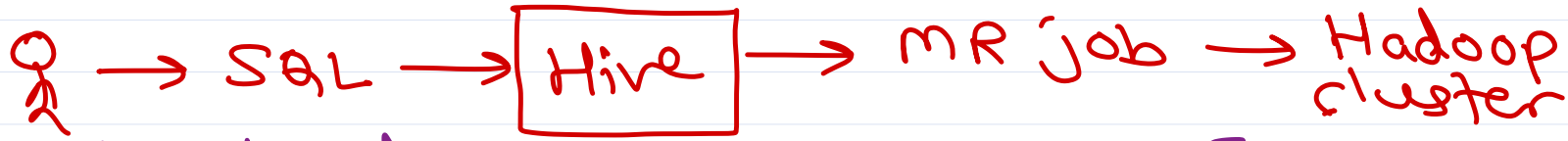
- ① Hive (Facebook) ... SQL \longrightarrow MR job
- ② Pig (Yahoo) ... Pig Latin \longrightarrow MR job
- ③ HBase ... NoSQL db fast searching
- ④ Sqoop ... RDBMS \longleftrightarrow HDFS
- ⑤ Flume ... Live Data \longrightarrow HDFS
- ⑥ Impala ... SQL \longrightarrow MR job
- ⑦ Spark ... Data processing engine
- ⑧ Tez ... Data processing engine
- ⑨ Mahout ... Machine learning Lib
- ⑩ Ambari ... Hadoop cluster mgmt
- ⑪ Oozie ... Job scheduler
- ⑫ Zookeeper ... Coordination & Sync.
- ⑬ ...

Hadoop Distributions

- ① Cloudera + Hootenworks
- ② AWS EMR
- ③ MapR
- ④ GCP



Apache Hive



Developed by Facebook.-2007

Hive QL → Extended SQL

- ① Data ingestion → LOAD DATA ...
- ② Schema - on - Read
- ③ Data formats : TEXTFILE, RCFile, ORC, PARQUE, ...
- ④ Managed & External tables.
- ⑤ Complex Data types : ARRAY, STRUCT, MAP, UNION.
- ⑥ Partitioning & bucketing
- ⑦ ...





Thank you!

Nilesh Ghule <nilesh@sunbeaminfo.com>

