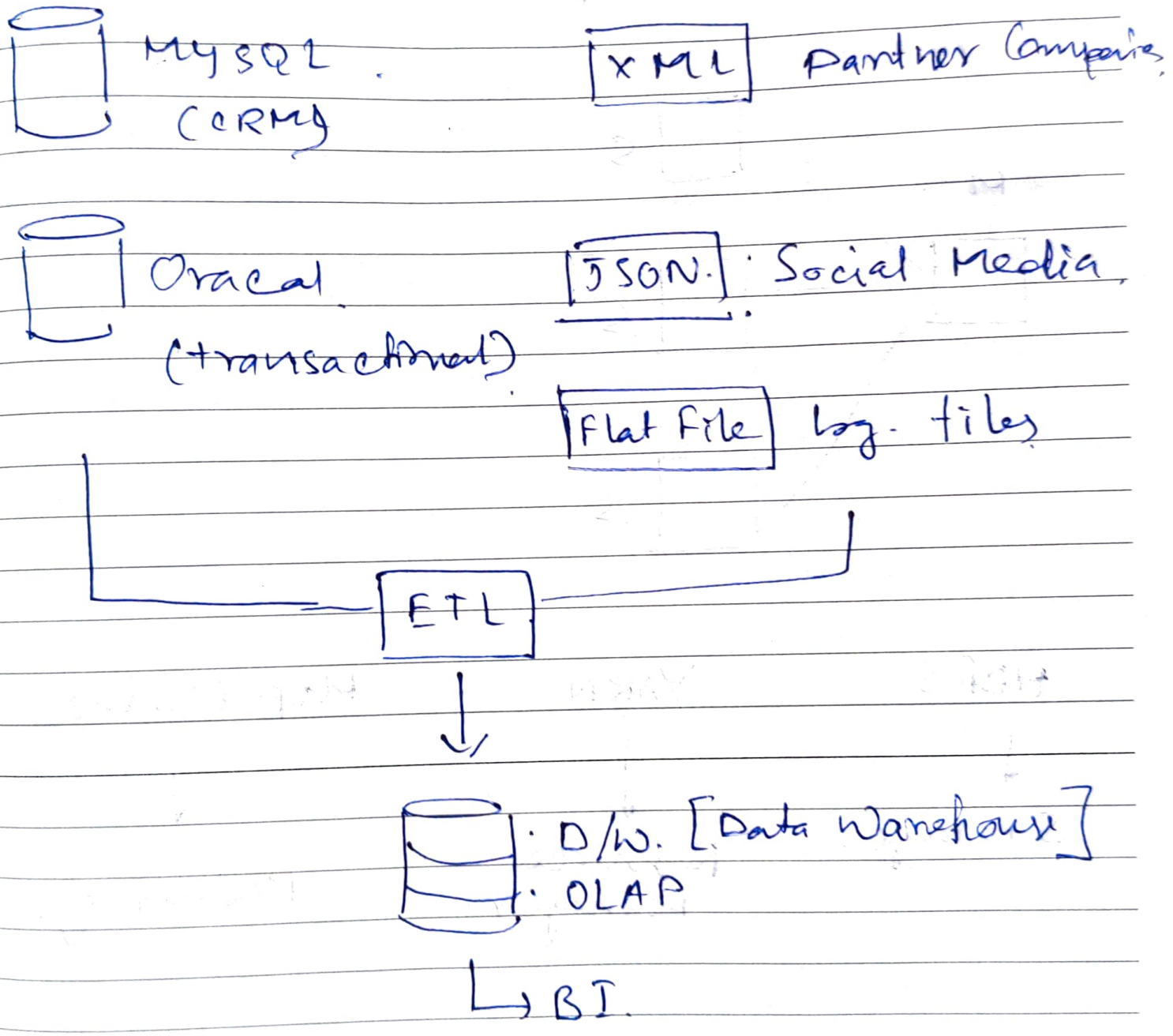
polyglot Persistence. 2

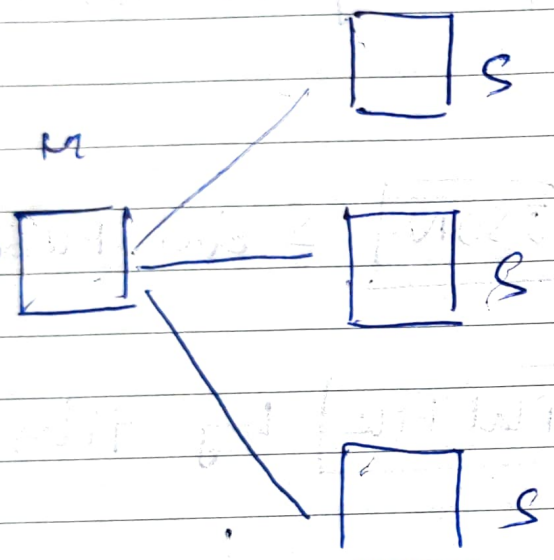
ETL -> Extract Transform Load



MPP → Massive Parallel processing.

'Cloudera' → Commercial distribution of hadoop.

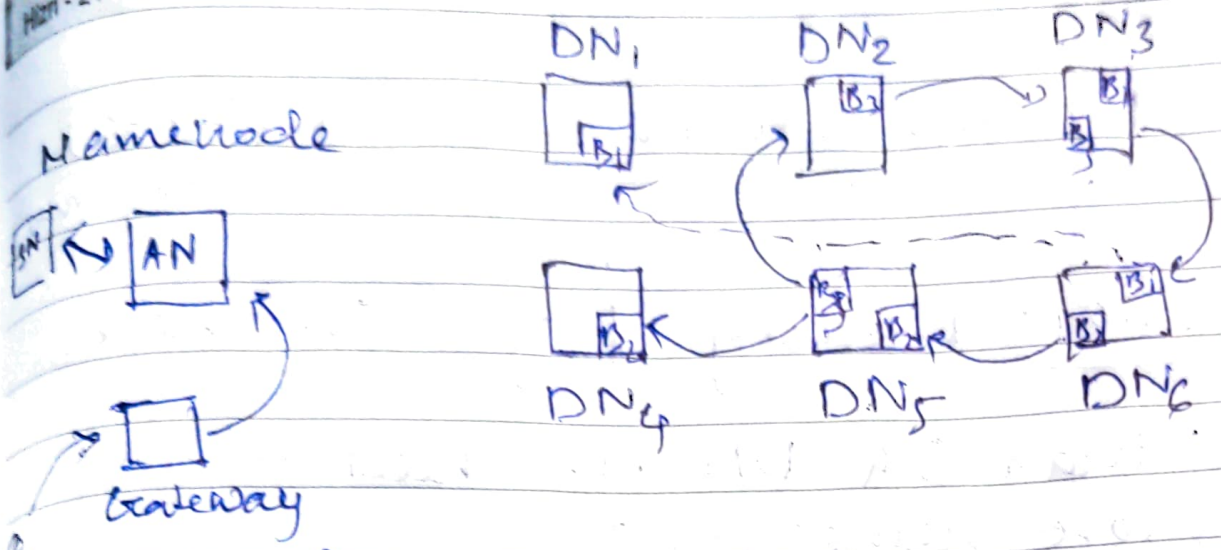
hortonworks, Mapr.



HDFS → Storage.

YARN → Resource Manager.

Map Reduce → Processing.



i) Any format of data can be stored.

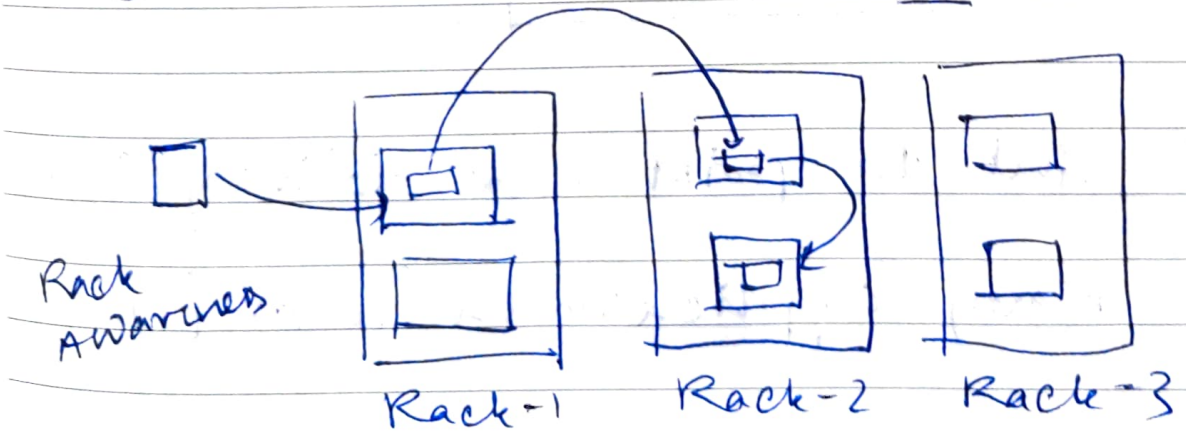
ii) Modification of file is not possible.

Block Size → Max size of data that can be stored.

192 MB → B1 B2 B3

64 64 64

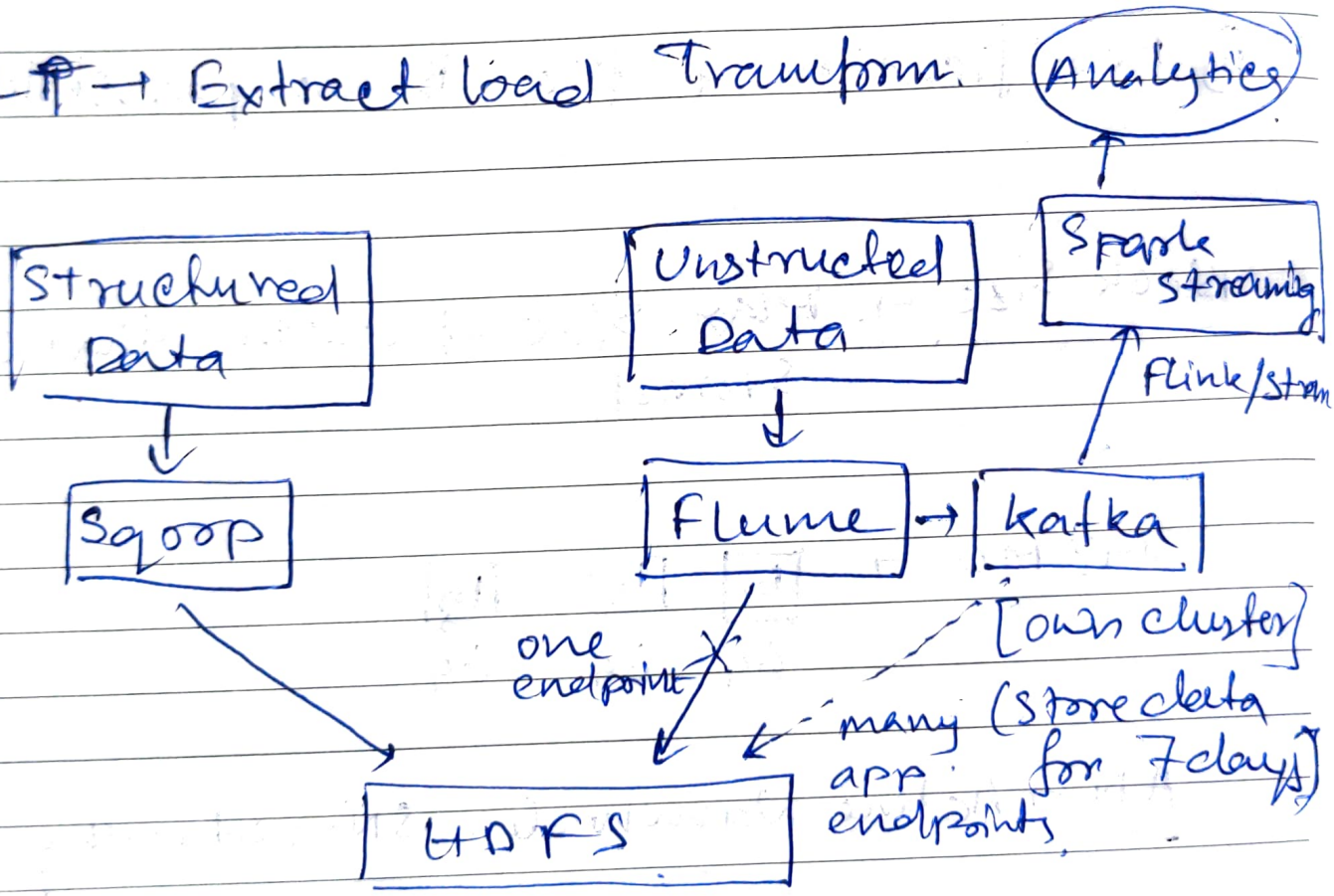
AN → Active Namenode. SN → Standby Namenode.



Lagest Hadoop Cluster → Yahoo. (42k) machines

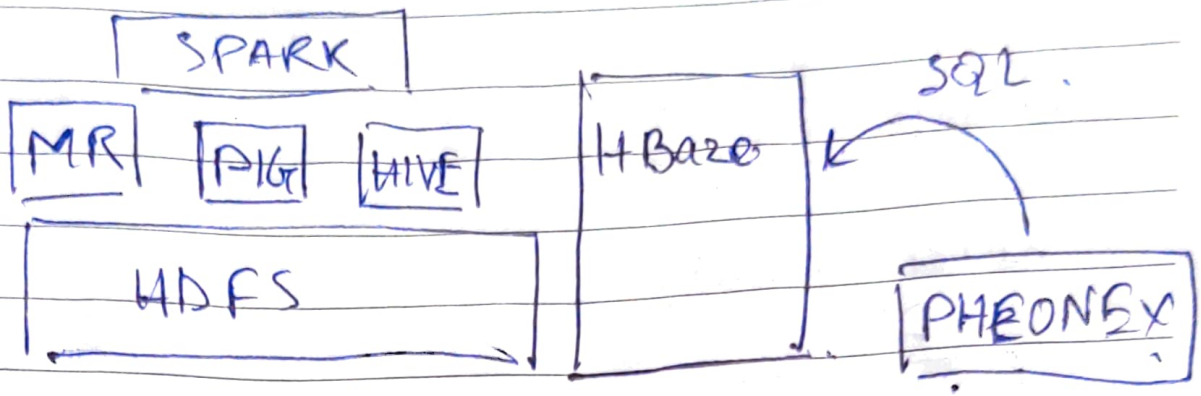
Federation: Ideally if you have more than 5k DN you need to add one more namenode.

ELT → Extract load Tranform.



Flume: Print to print data delivery System.

kafka: Many application can read from.



HIVE - SQL on top of HDFS.

SQL → HIVE (Translator) → MAP RED

SPARK - In memory execution system. (Batch processing)

Big Query → HIVE.

Small Query → Impala | HAWQ, LLAP | DRILL etc.

~~PHENEX~~ PHEONEX.