Big Data Technologies
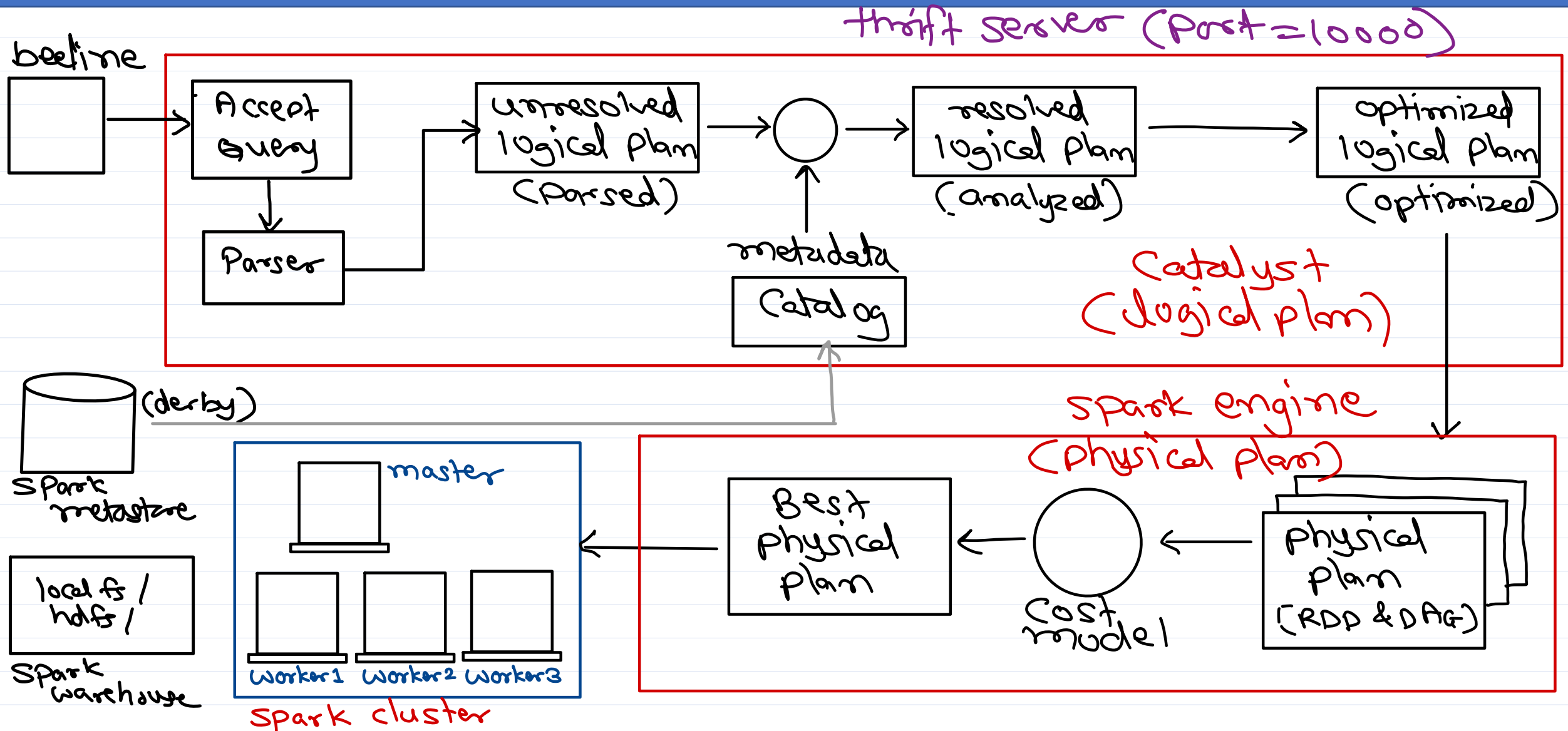
Trainer: Mr. Nilesh Ghule.

Spark SQL architecture

thrift server (port=10000)

beeline

Accept query → Parser

unresolved logical plan (Parsed) → ● → resolved logical plan (analyzed) → optimized logical plan (optimized)

metadata
Catalog

Catalyst (logical plan)

(derby)
Spark metastore

local fs / hdfs /
Spark warehouse

master
Worker1 Worker2 Worker3
Spark cluster

Spark engine (Physical plan)

Best Physical Plan ← Cost model ← Physical Plan [RDD & DAG]

Thank you!

Nilesh Ghule <nilesh@sunbeaminfo.com>