



CLOUD COMPUTING APPLICATIONS

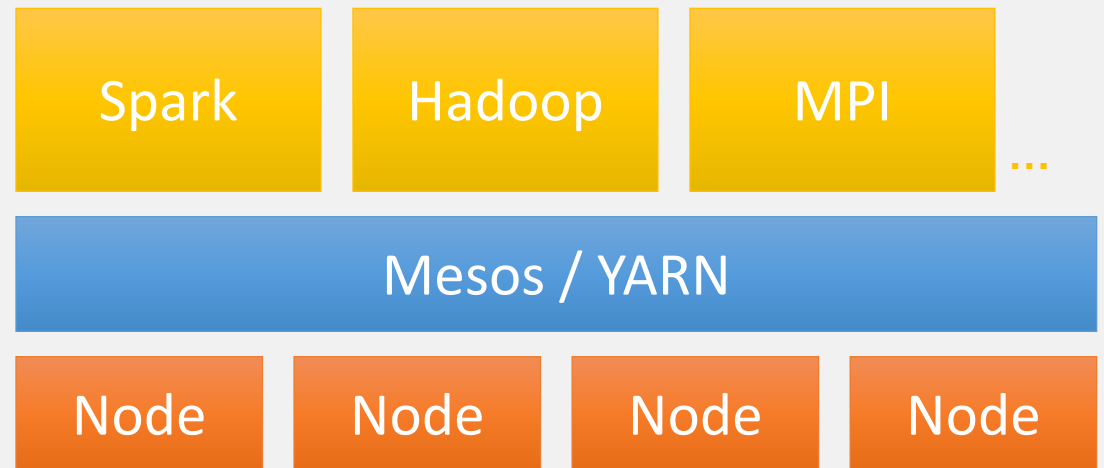
Spark Implementation

Roy Campbell & Reza Farivar

Implementation

Runs on Apache Mesos or YARN
to share resources with Hadoop &
other apps

Can read from any Hadoop input
source (e.g. HDFS)



Spark Scheduler

Dryad-like DAGs

Pipelines functions within a stage

Cache-aware work reuse & locality

Partitioning-aware to avoid shuffles

