



YARN

School of Information Studies
Syracuse University

Hadoop 1 (MapReduce) vs. Hadoop 2 (YARN)

Hadoop 1 (MapReduce)

- Batch processing only.
- No high availability or failover capability.
- Only MapReduce applications.
- Data nodes and name nodes are dependent.
- Applications execute and quit.

Hadoop 2 (YARN)

- Suited to a variety of distributed tasks.
- High availability and failover.
- Runs any distributed application, including MapReduce.
- Data node and name nodes (node managers) are independent.
- Applications can remain resident.

YARN: The Data Operating System

- Hadoop 2.0 introduces YARN (yet another resource negotiator).
- Orchestrates processing over the nodes.
- Uses HDFS for storage.
- Runs a variety of applications.
- TEZ: optimization of MapReduce jobs over YARN.
- Slider: a shim to enable applications over YARN.

