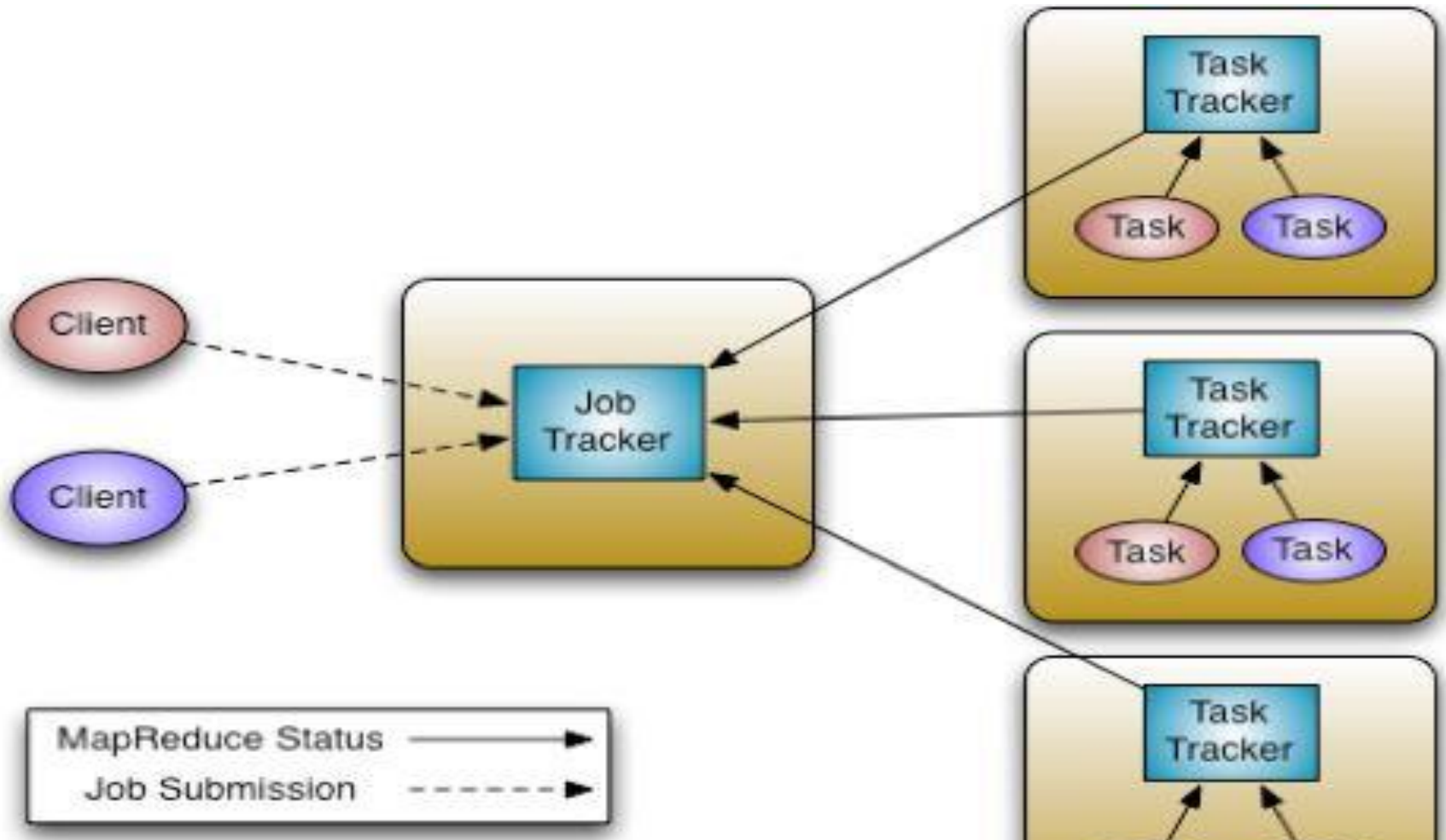# Hadoop 2.0 (YARN)

Agenda

-

- Recap of Hadoop 1.0 architecture

- Job scheduling in Hadoop 1.0 and its limitations

- Hadoop 2.0 (YARN) and its salient features

- Architectural overview of Hadoop 2.0

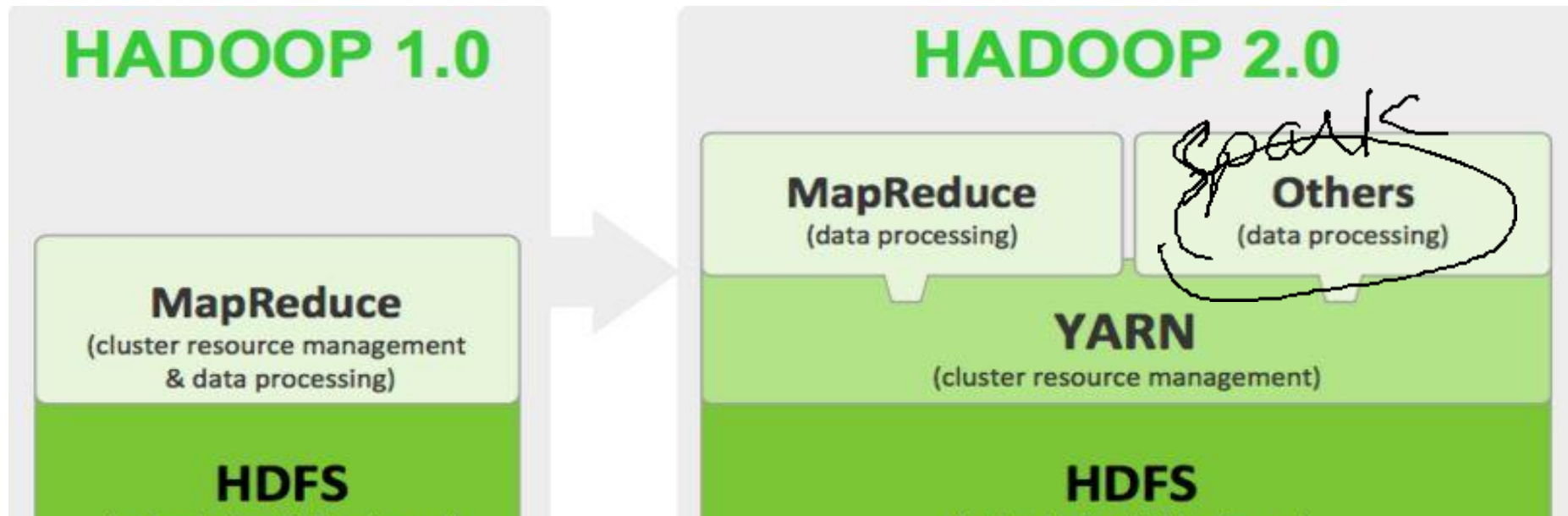# Job tracker (centralized scheduler in Hadoop 1.0)

YARN: Yet another resource negotiator

YARN (Hadoop 2.0) is the new scheduler and centralized resource manager in the cluster.

It replaces the Job tracker in Hadoop 2.0

The started back in 2012 as an apache sub project and a beta version was relased in mid 2013

stable versions became available from 2014 onwards

Components of YARN

Its a 2 tiered model with some components (deamons) in master mode and some operating in slave mode.

Resource manager works in master mode (runs on a dedicated hardware in production setup)

Node manager works in slave mode & its services are run in the data nodes.

Resource manager comprises of 2 components

    1)Scheduler

    2)Applications Manager

Node manager consists of a container (an encapsulation of resources for running a job) and an app Master (application master)