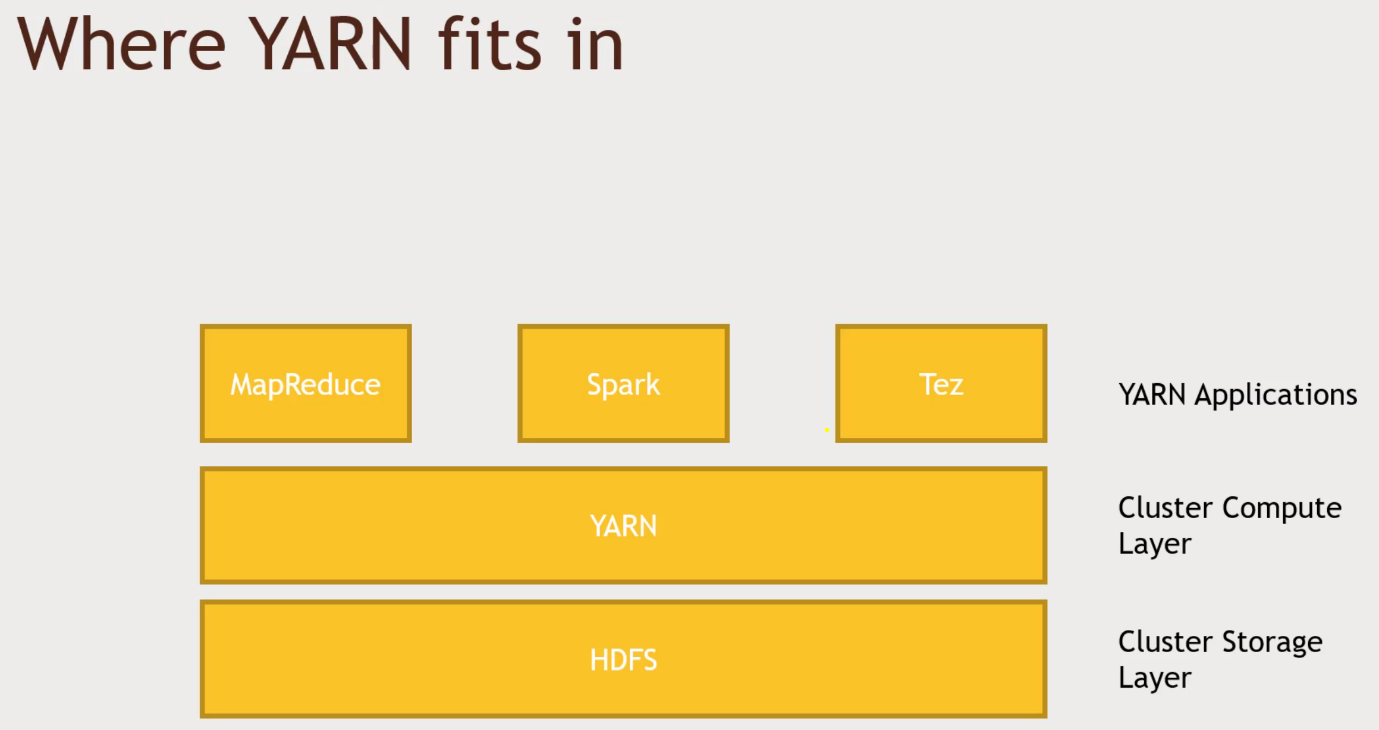
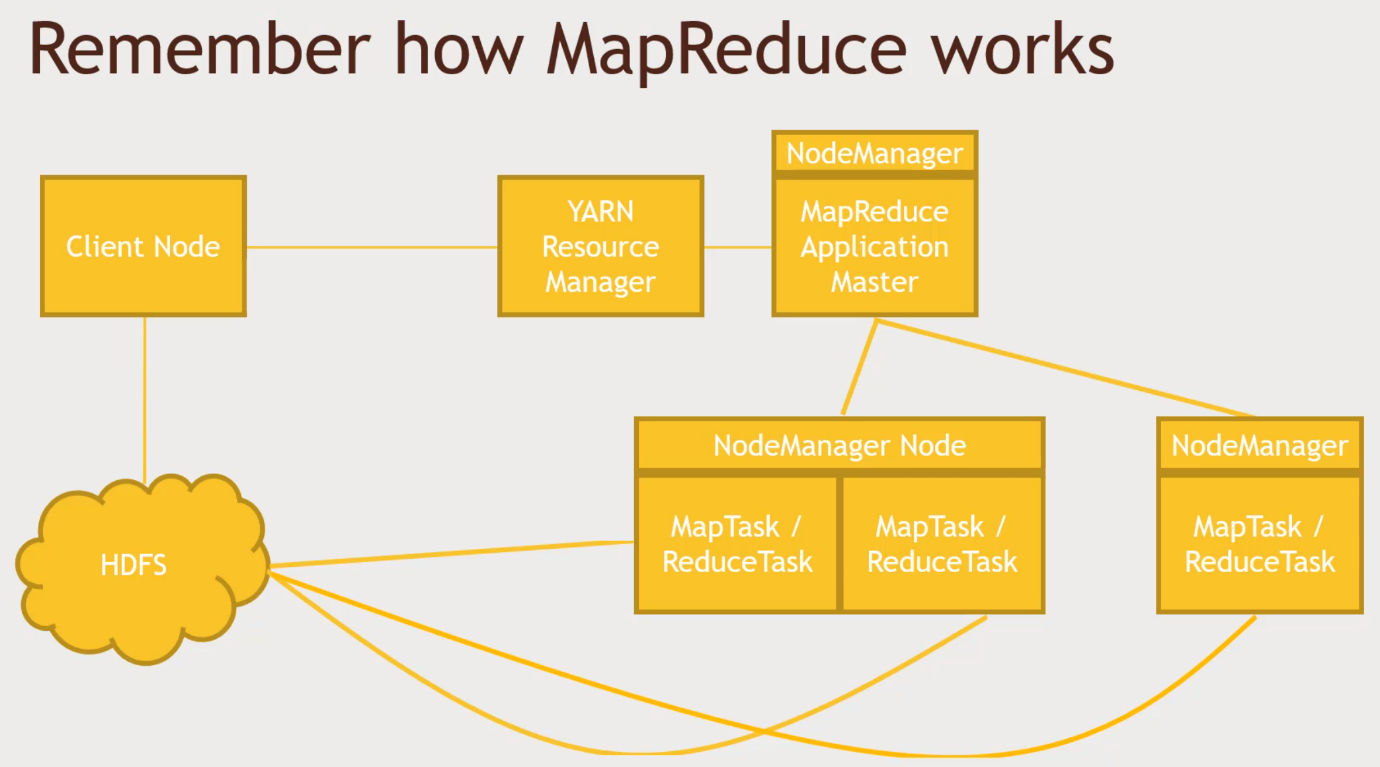
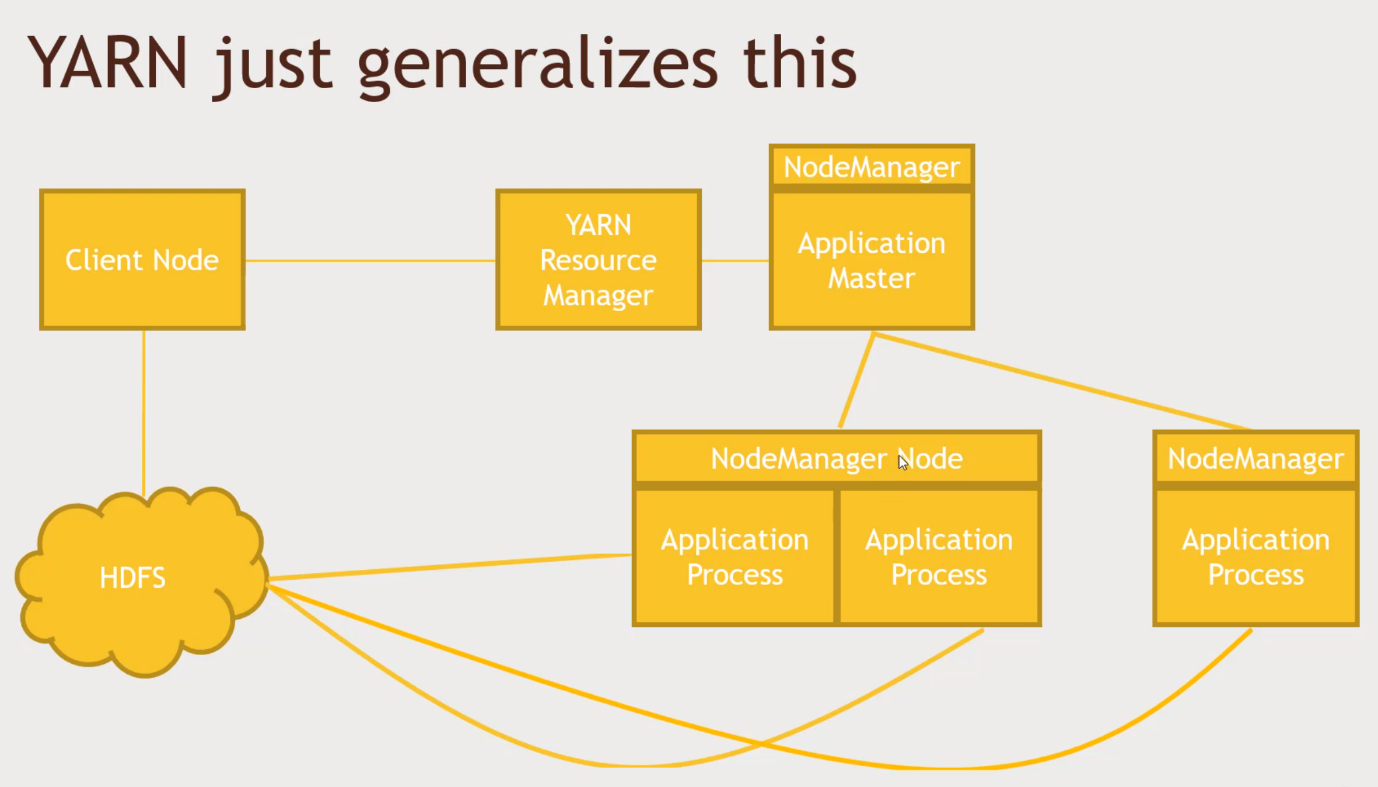
HADOOP YARN

You won’t need to code in YARN. It is like a Wheel.

YARN Architecture

Compute layer for your application. That connect between the Data. Split up the actual computation across your cluster. HDFS- data storage across your cluster.

YARN tries to maintain localicality. Manages computation resources.

How YARN works

* Your application talks to the Resource Manager to distribute work to your cluster
* You can specify data locality- which HDFS block(s) do you want to process?
  + YARN will try to get your process on the same node that has your HDFS blocks
* You can specify different scheduling options for applications
  + So you can run more than one application at once on your cluster
  + FIFO, Capacity and Fair schedulers
    - FIFO runs jobs in sequence, first in first out
    - Capacity may run jobs in parallel if there’s enough spare capacity
    - Fair may cut into a larger running job if you just want to squeeze in a small 1