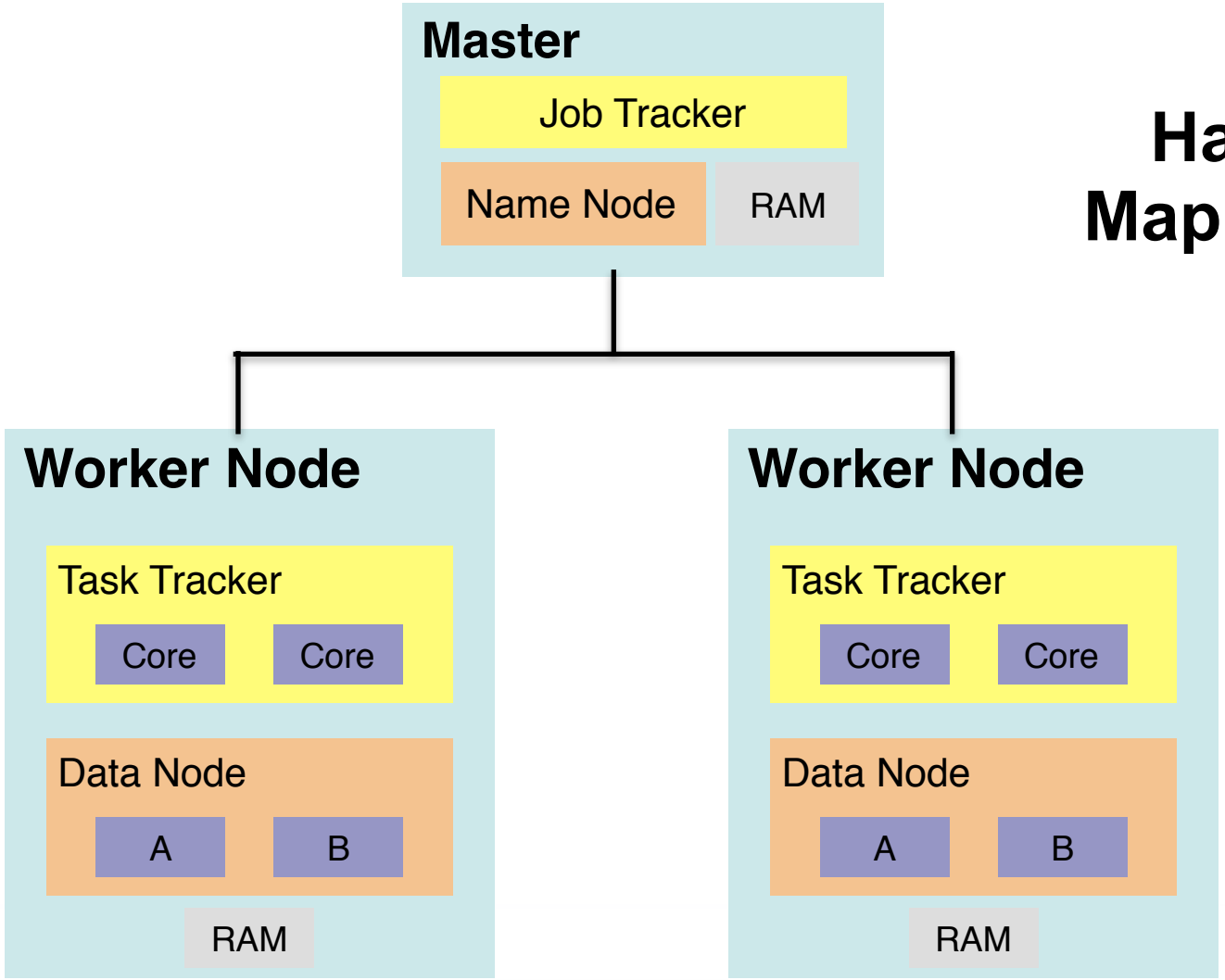
livelessons
video instruction from technology experts

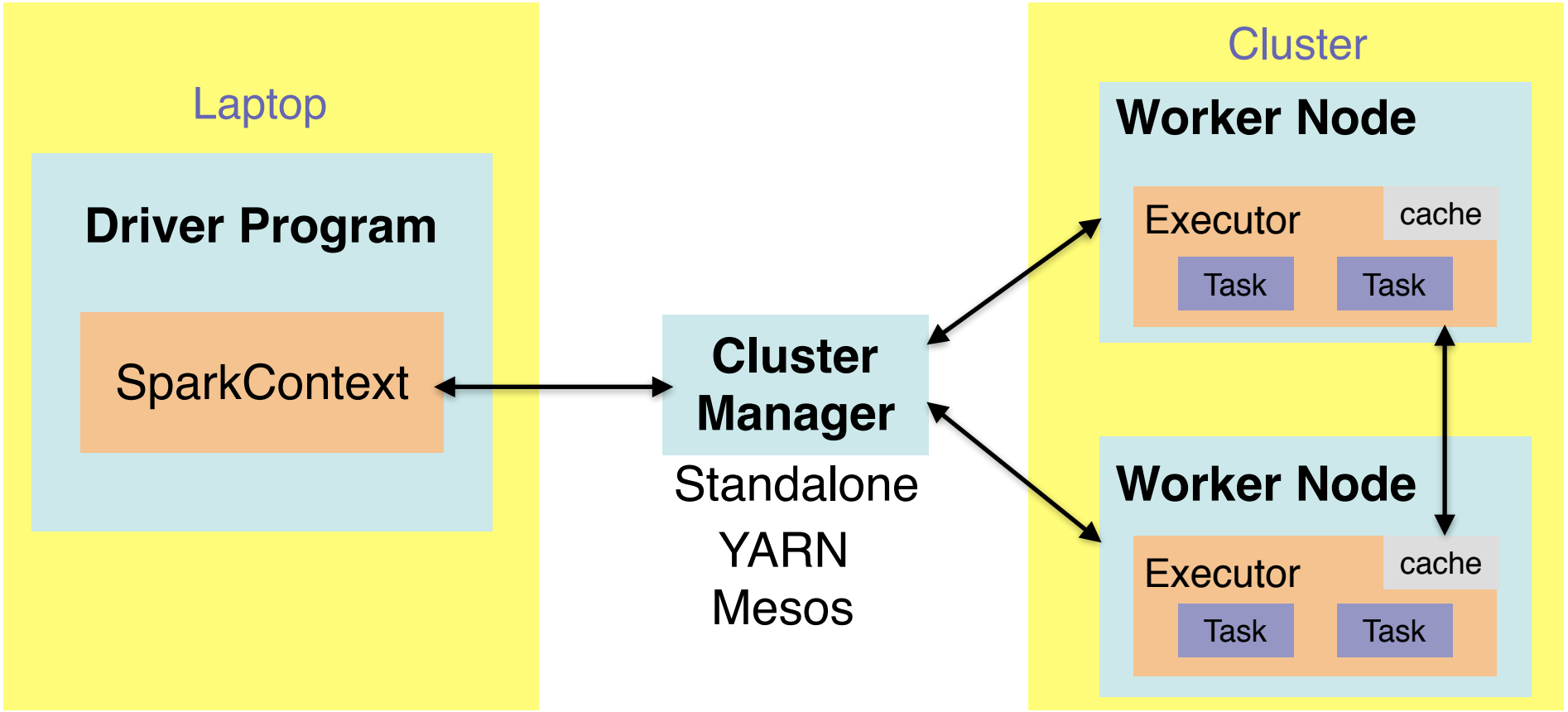
Lesson 4: Spark Internals

4.3 The Spark Execution Context



Hadoop MapReduce

Spark Execution Context



Terminology

Term	*Meaning*
Driver	*Process that contains the SparkContext*
Executor	*Process that executes one or more Spark tasks*
Master	*Process that manages applications across the cluster*
Worker	*Process that manages executors on a particular node*

Spark vs. Hadoop

- Spark only replaces MapReduce (computation)
- Still need a data store: HDFS, HBase, Hive, etc.
- Spark has a more flexible/general programming model
- Spark often faster for iterative computation