YARN:

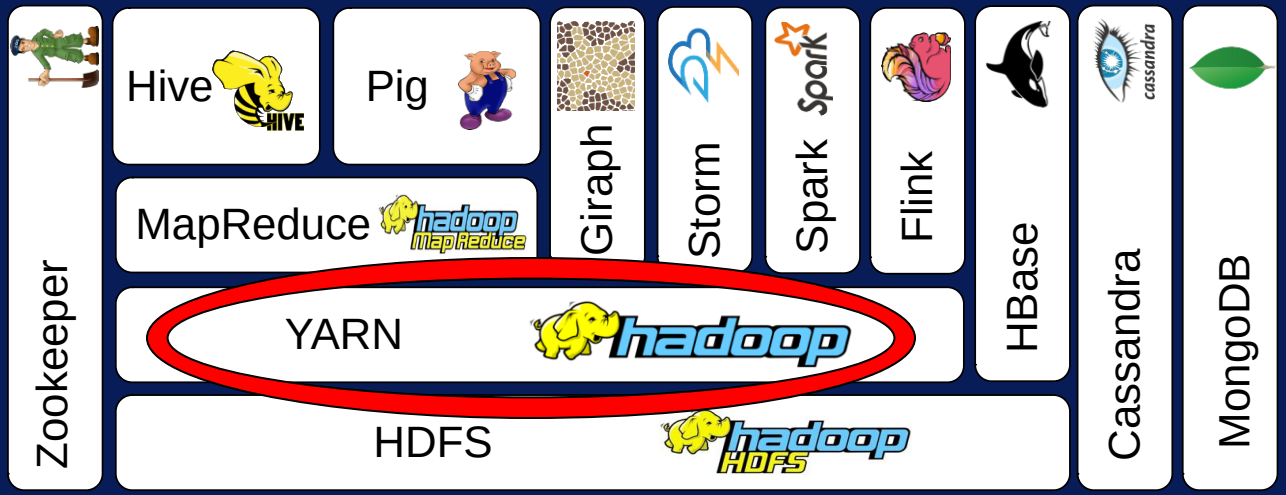
The Resource Manager for Hadoop

After this video you will be able to..

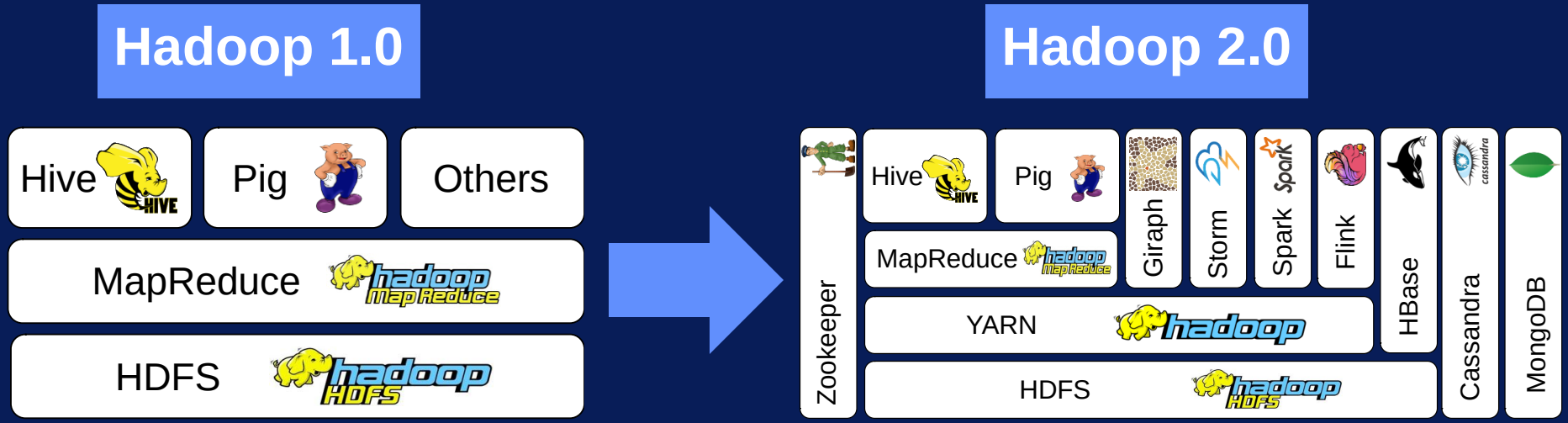
- Outline how YARN provides flexible resource management for Hadoop cluster
- Explain how YARN extends Hadoop to enable multiple frameworks such as MapReduce, Giraph, Spark and Flink

HDFS Cluster Utilization

Share Hadoop across applications



Hadoop evolved over time!



Hadoop 1.0

Only MapReduce jobs

Hive | Pig | Others

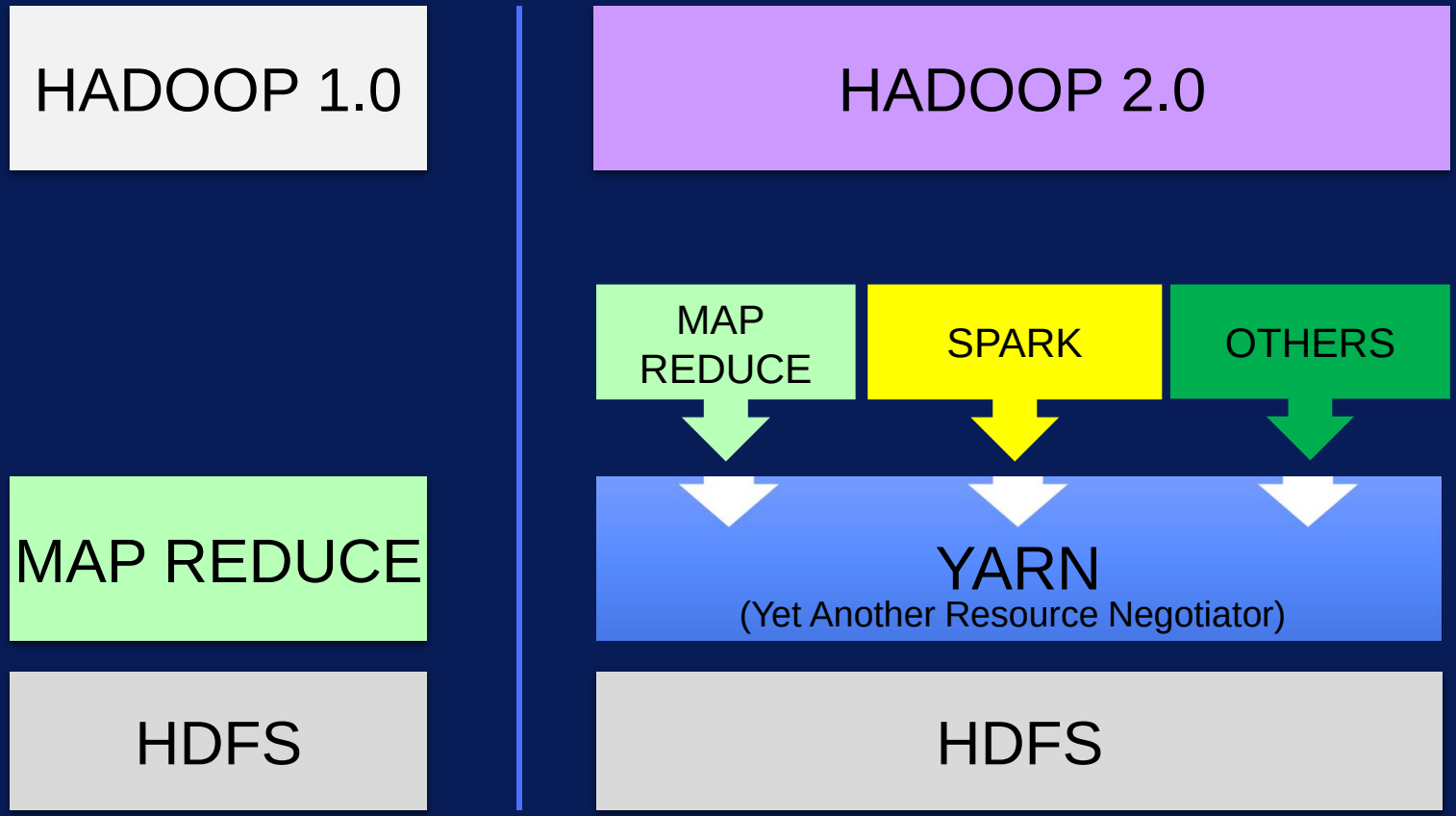
MapReduce

HDFS

Other applications not supported

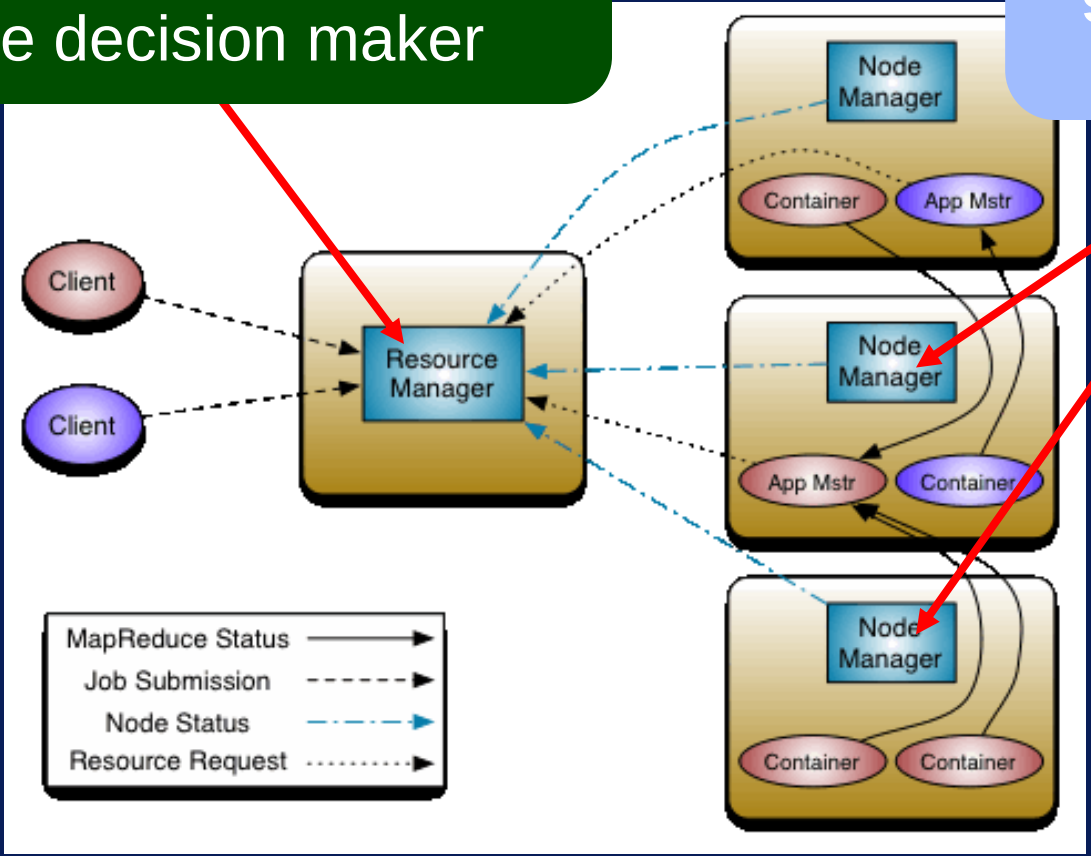
Poor Resource utilization

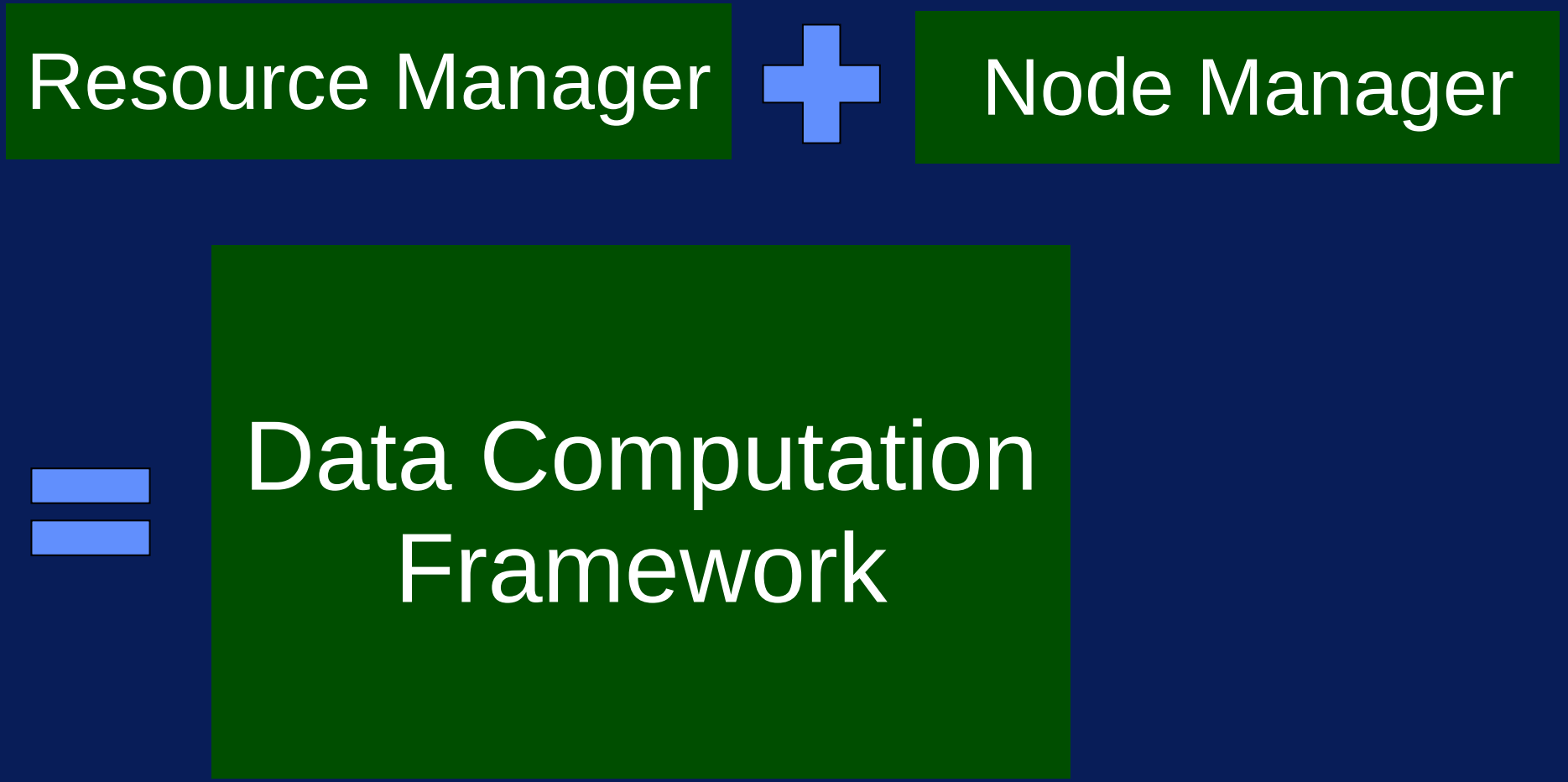
One dataset → many applications



Central Resource Manager
==
ultimate decision maker

Each machine gets a Node Manager

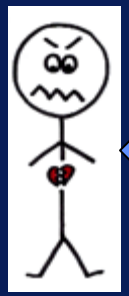



Resource Manager
Node Manager
Data Computation Framework

Application Master = personal negotiator

Negotiates

Resource Manager

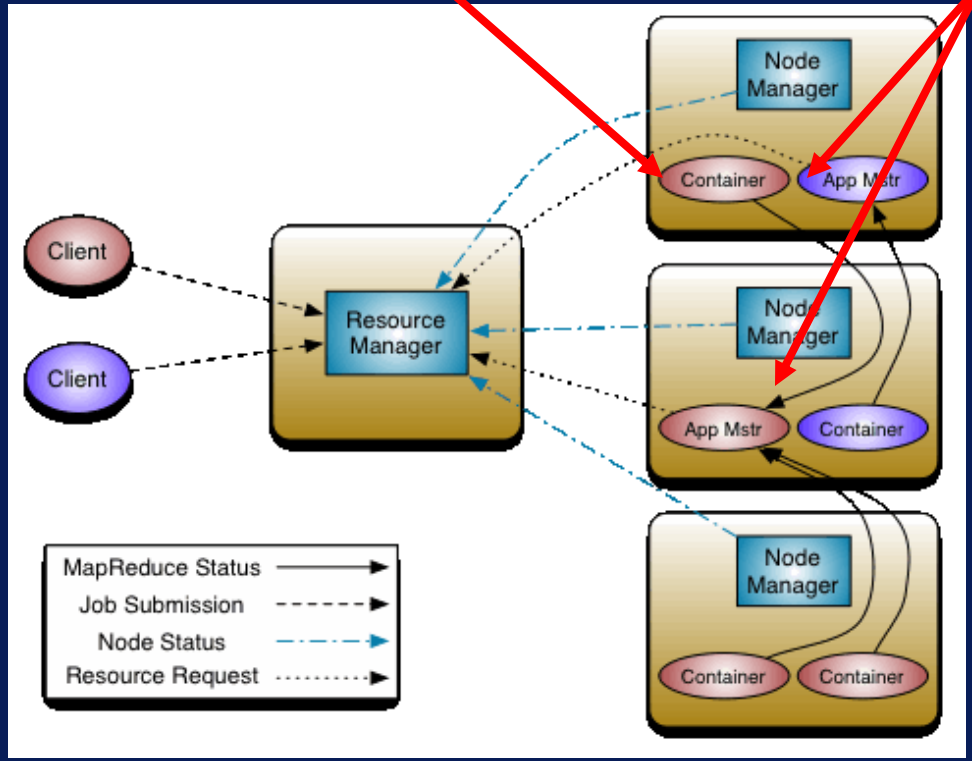


Gets the job done

Node Manager

Container = a machine

Application Master = Personal Negotiator



Essential gears in YARN engine

Resource Manager

Applications Master

Node Manager

Container

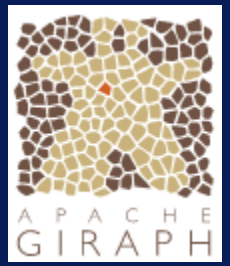
2X ↑ Jobs per day

2X ↑ CPU utilization

2.5X ↑ Number of tasks from all jobs

* Source: Apache Hadoop YARN: Yet Another Resource Negotiator.” In Proceedings of the 4th Annual Symposium on Cloud Computing, 5:1–5:16. SOCC ’13.

YARN → More Applications



and growing …

Data → Value

Many choices in Hadoop 2.0

One dataset → Many applications

Higher Resource Utilization → Lower Cost