CENTENNIAL COLLEGE

School of Engineering Technology and Applied Science

Big Data and Hadoop

Reza Dibaj

Learning Objective

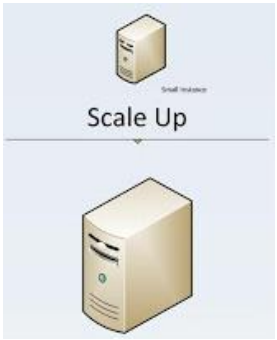
- History SQL and NoSQL Databases
- Hadoop

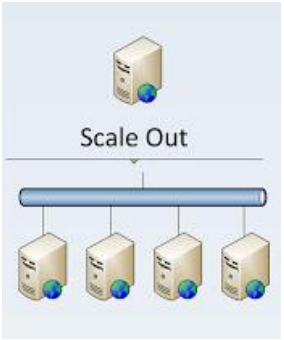
History

1980s
SQL, a dominant database environment

Amazon
Google } **High Traffic**

2000s
1st Solution: Scale up the hardware
Buy Bigger Boxes (BBB)
2nd Solution: Scale out the hardware
Buy Lots of Little Boxes

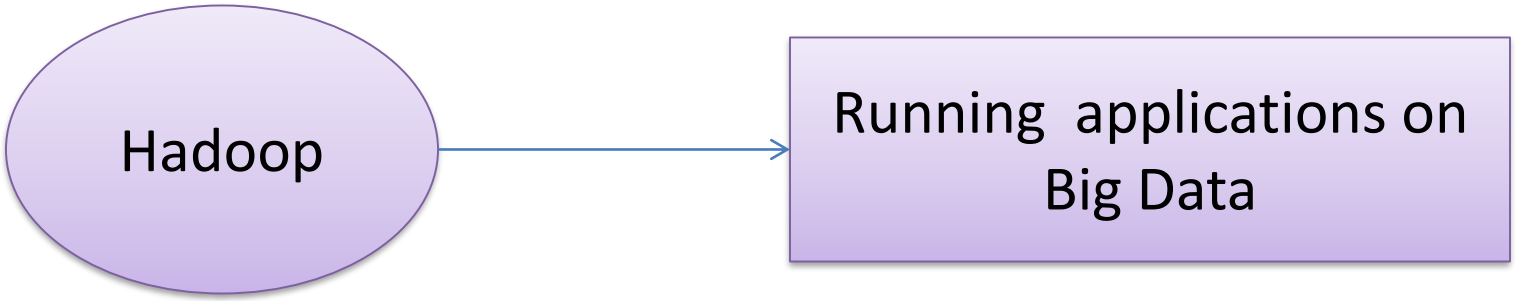




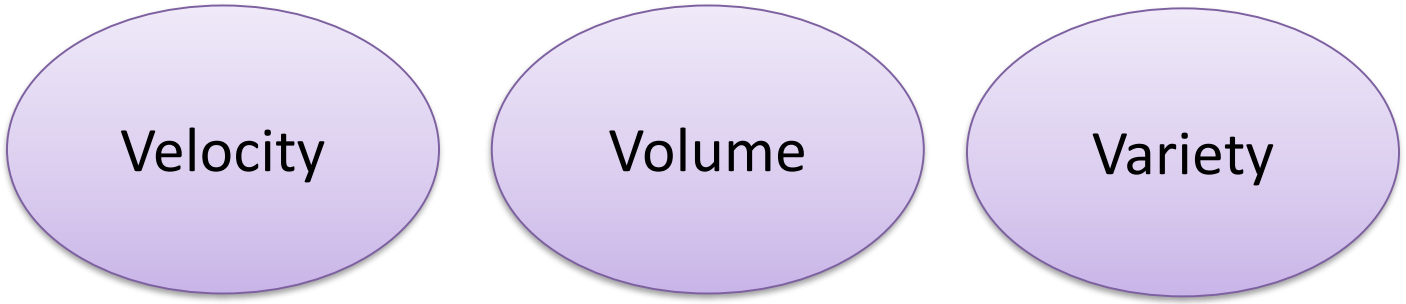
Problem: SQL was designed to work on a large box, and not on a large amount of small boxes.

What is Hadoop?

Hadoop: a framework of open-source set of tools distributed under Apache License

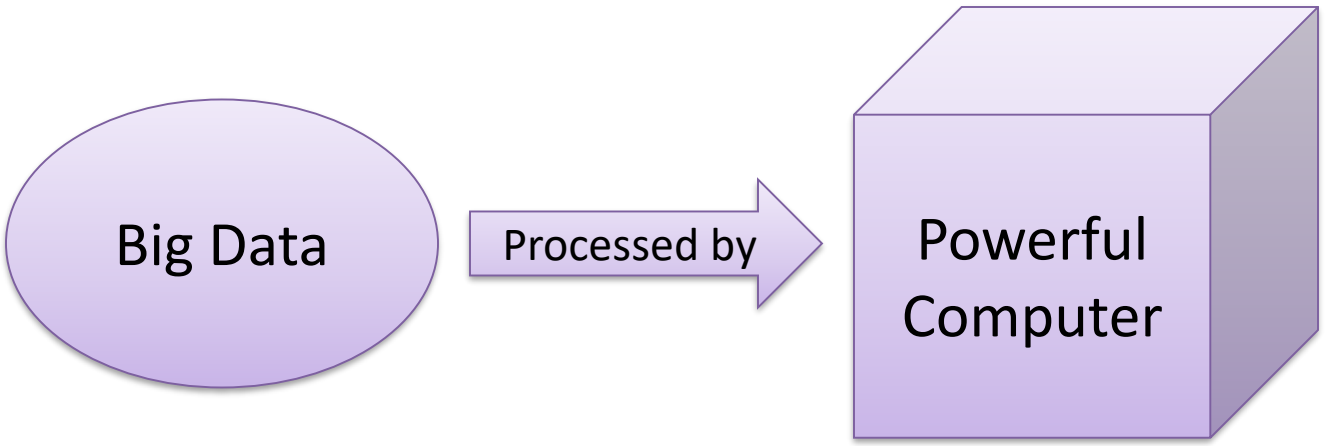


Big Data Challenge Points



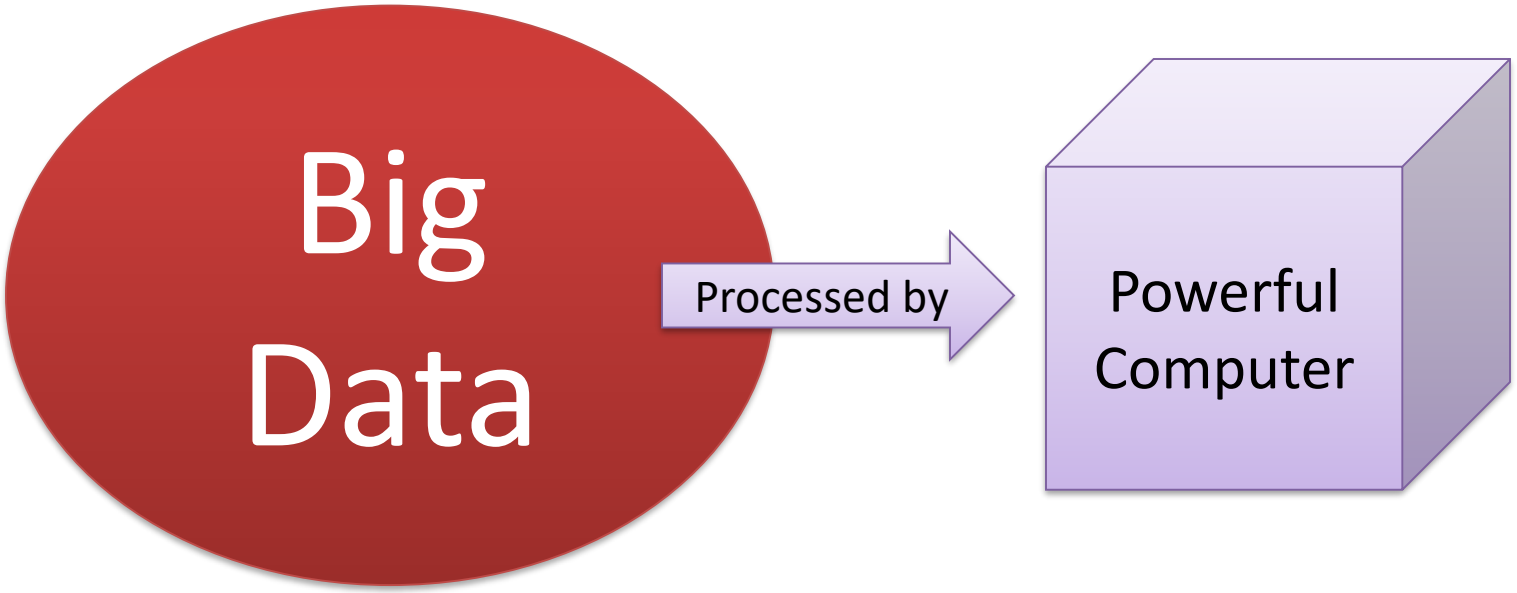
Traditional Approach

Enterprise Approach



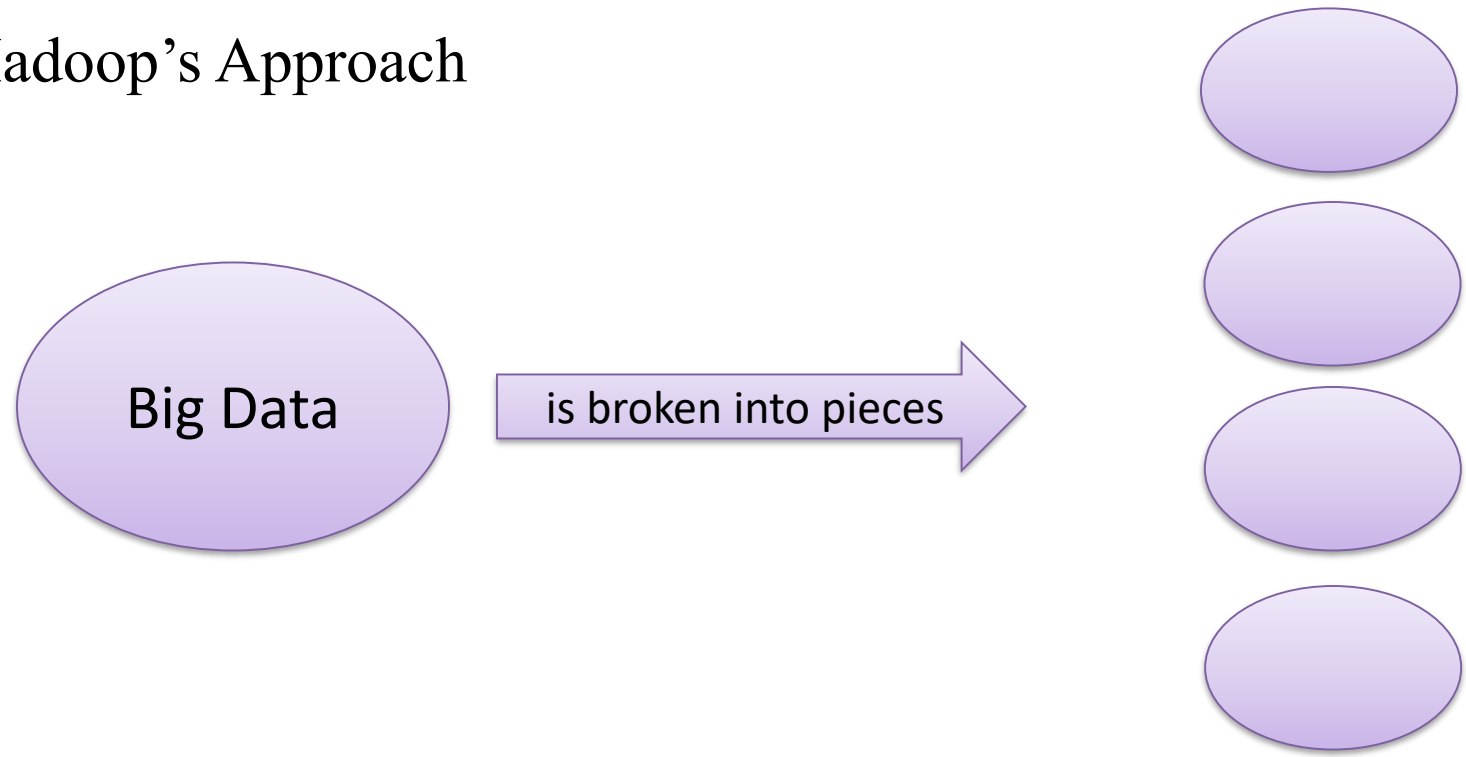
Traditional Approach

Enterprise Approach



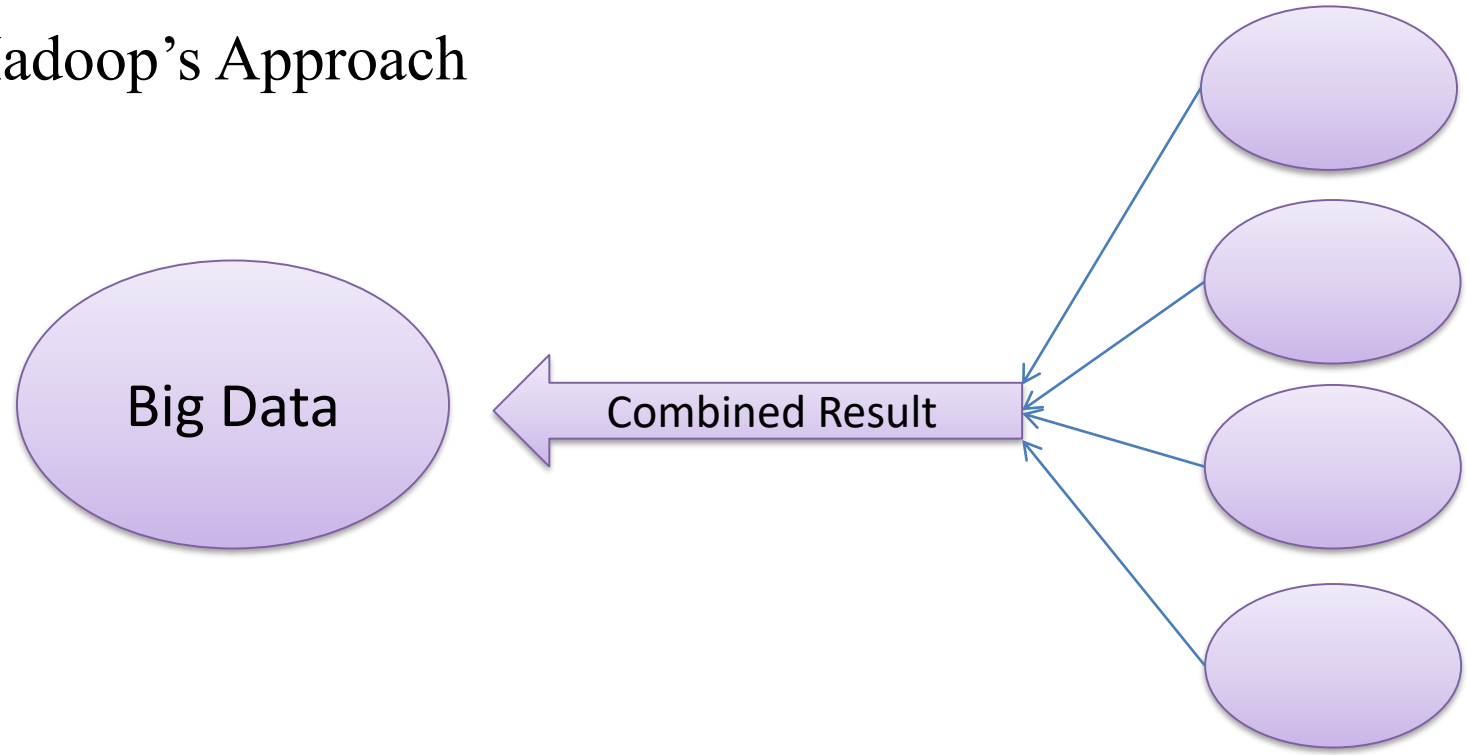
NoSQL Approach

Hadoop's Approach

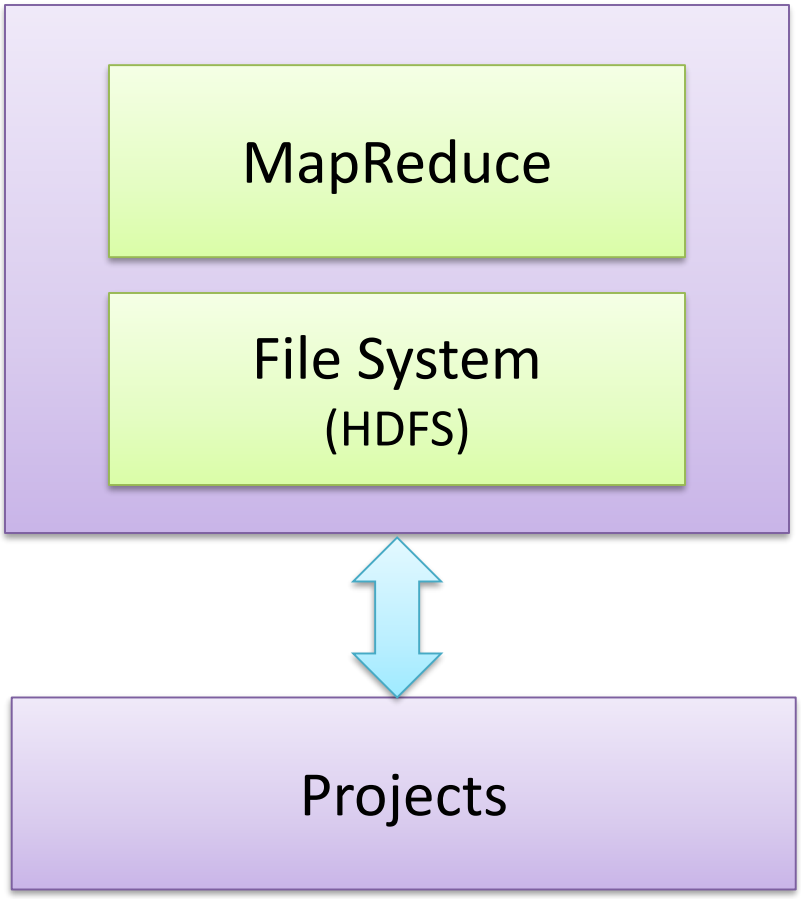


NoSQL Approach

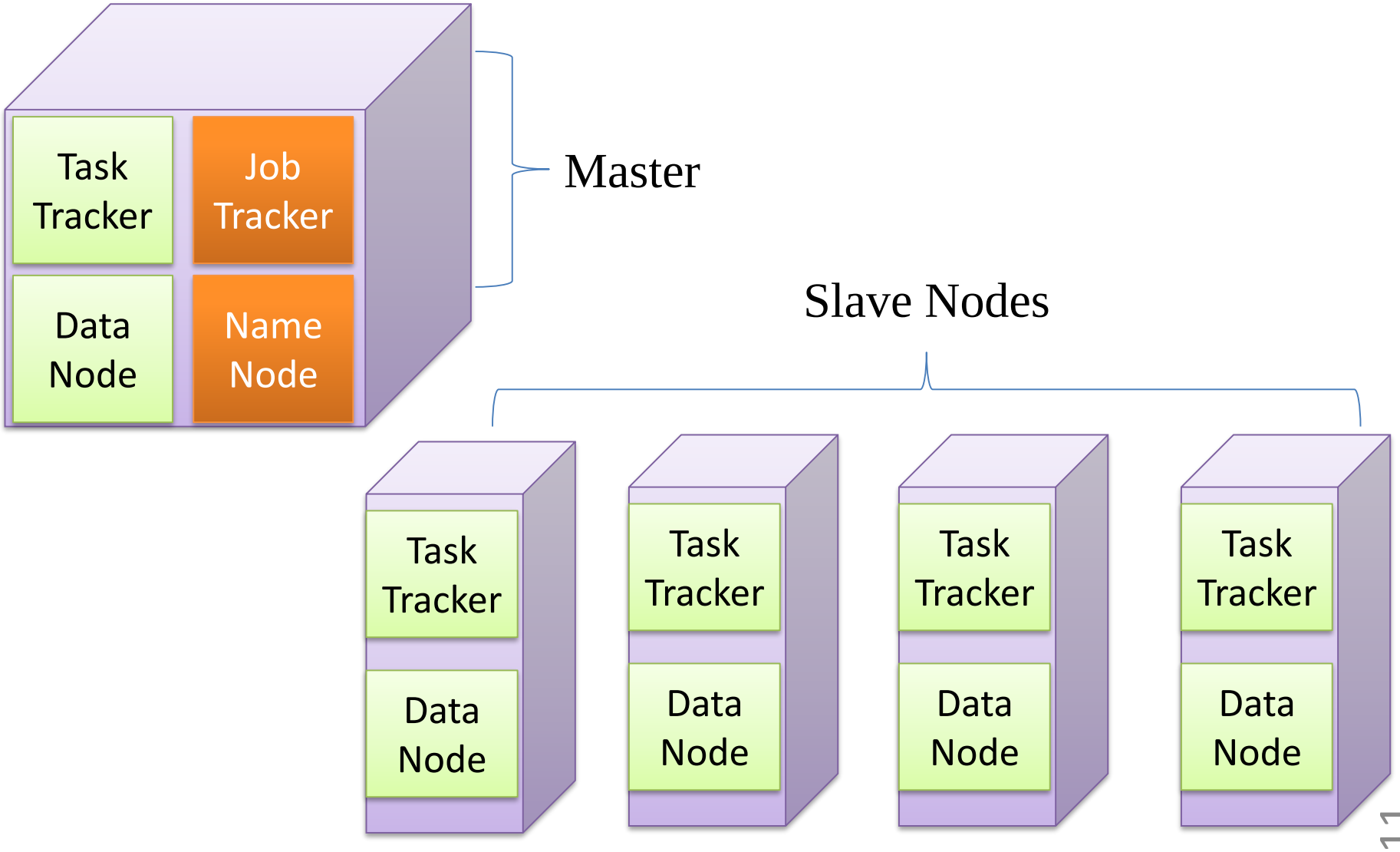
Hadoop's Approach



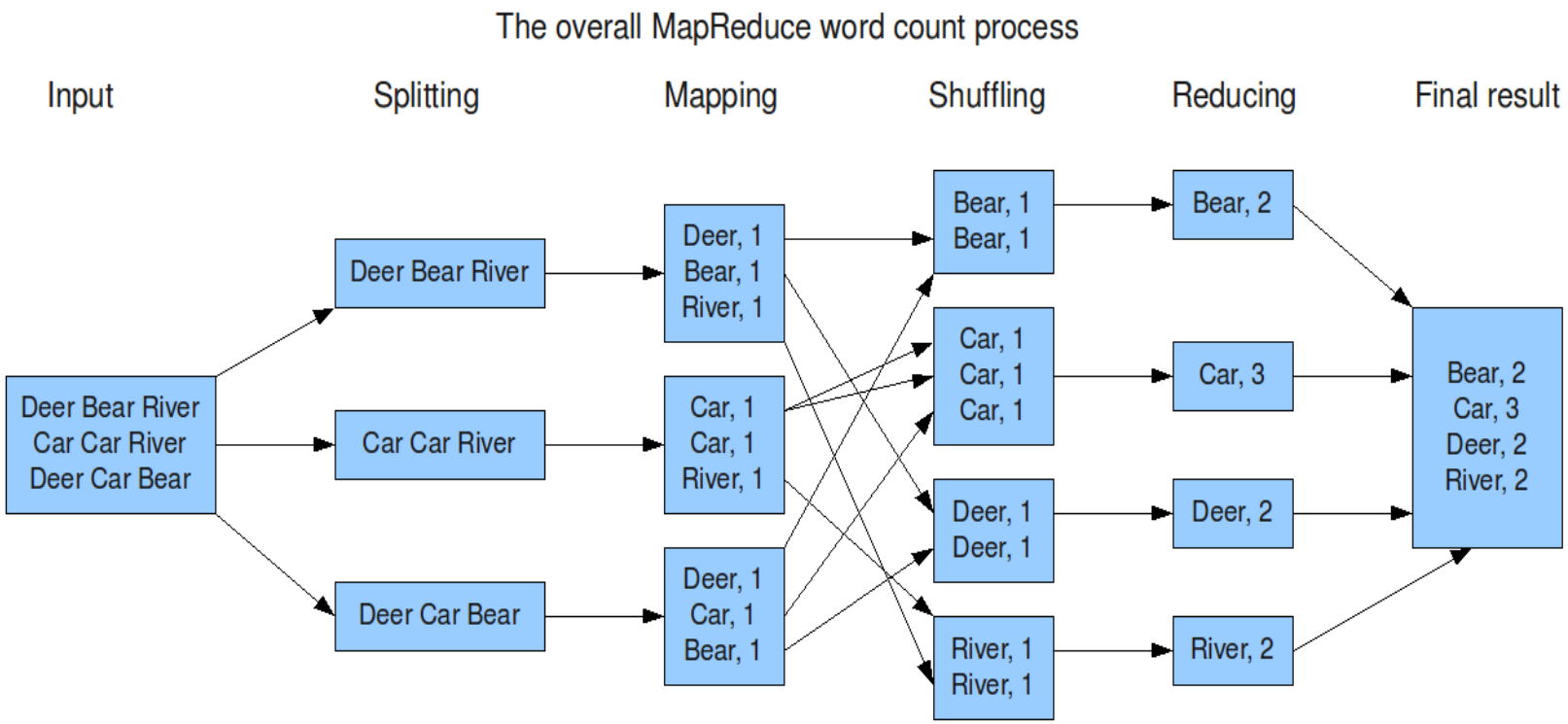
Hadoop's Architecture



Master and Slave Nodes



MapReduce Architecture



Question

- Hadoop is designed to work with commodity hardware in a distributed environment in the analysis of big data.

❑ True
❑ False

Summary and future work

- Traditional SQL environments are here to stay.
- Hadoop is just a beginning, and not the end.
- Spark, Storm and other frameworks are going to enhance Hadoop capabilities.

Thank you