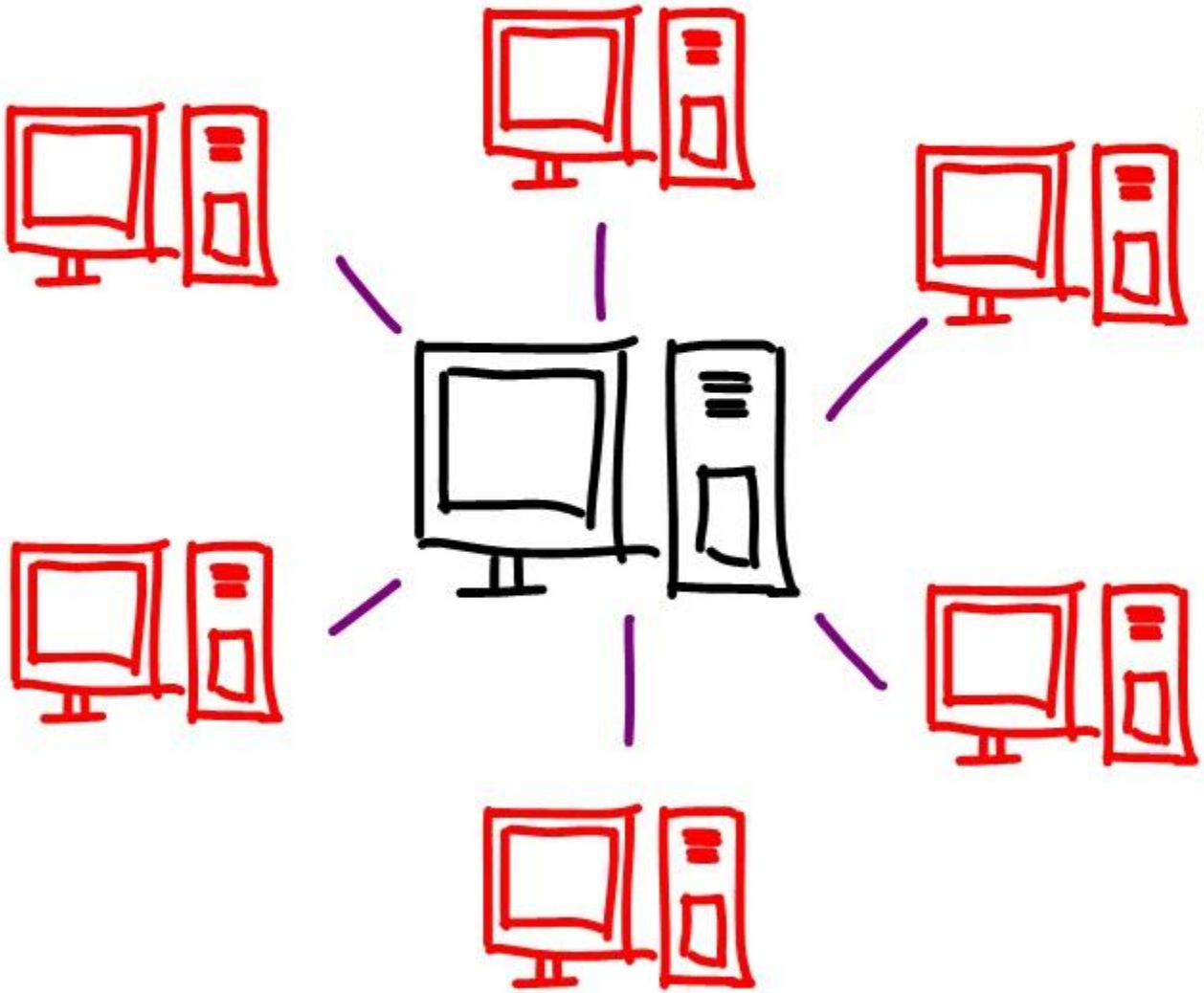
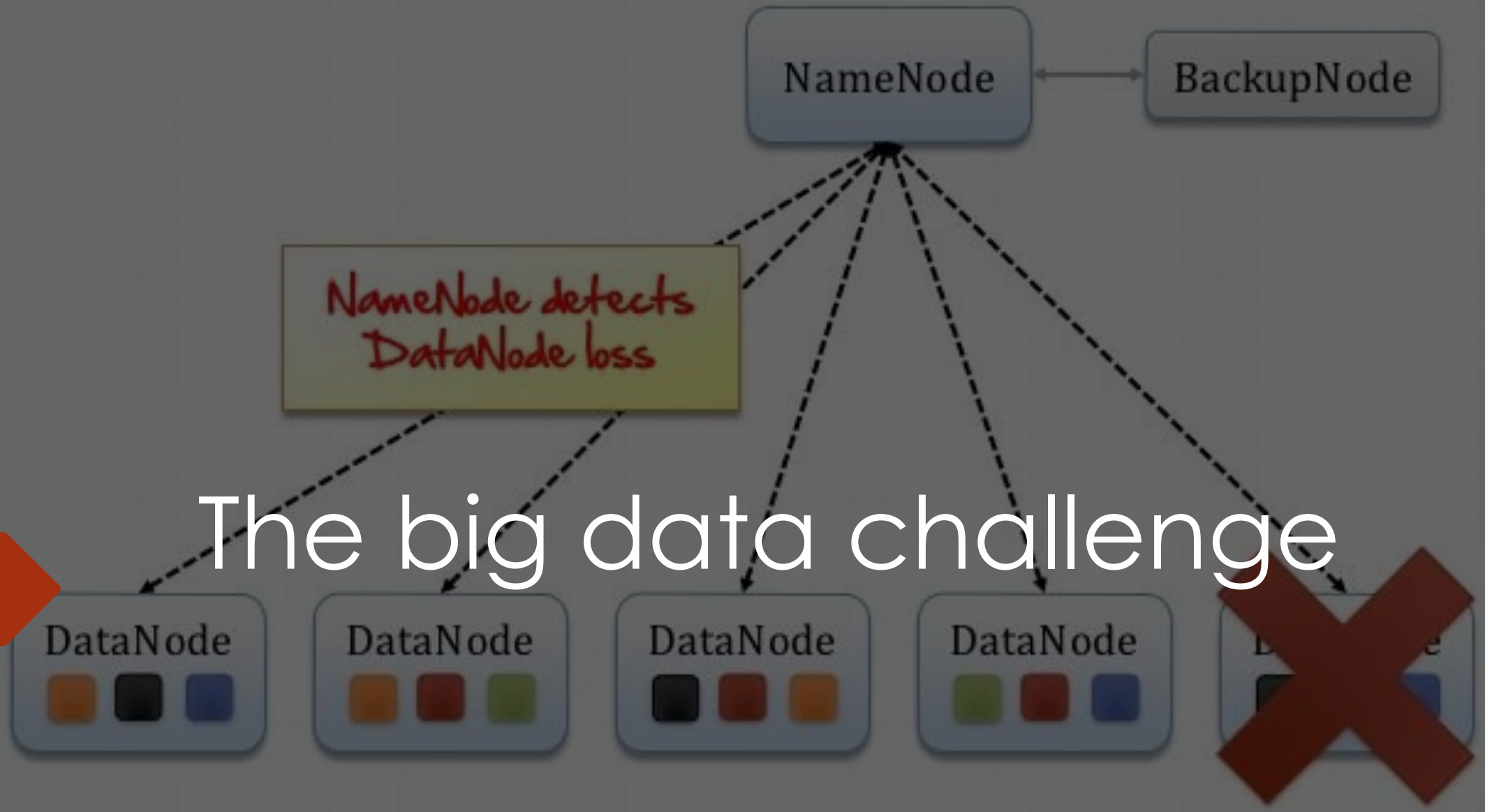
Distributed computing and data processing frameworks

Distributed Computing



The big data challenge



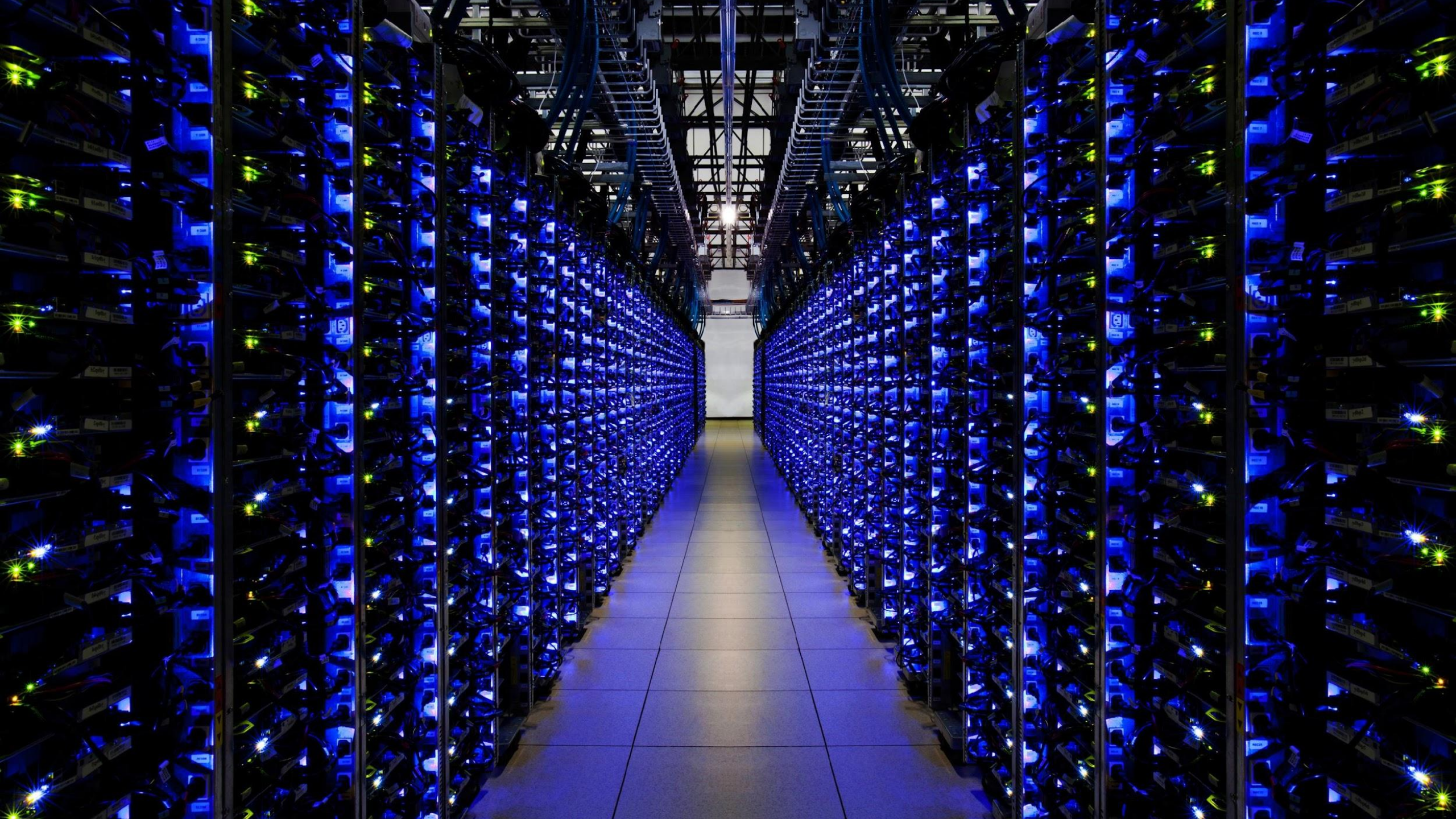
The major hardware restrictions

- Processor time
- Memory
- Hard drive space
- Network bandwidth

Memory (RAM)

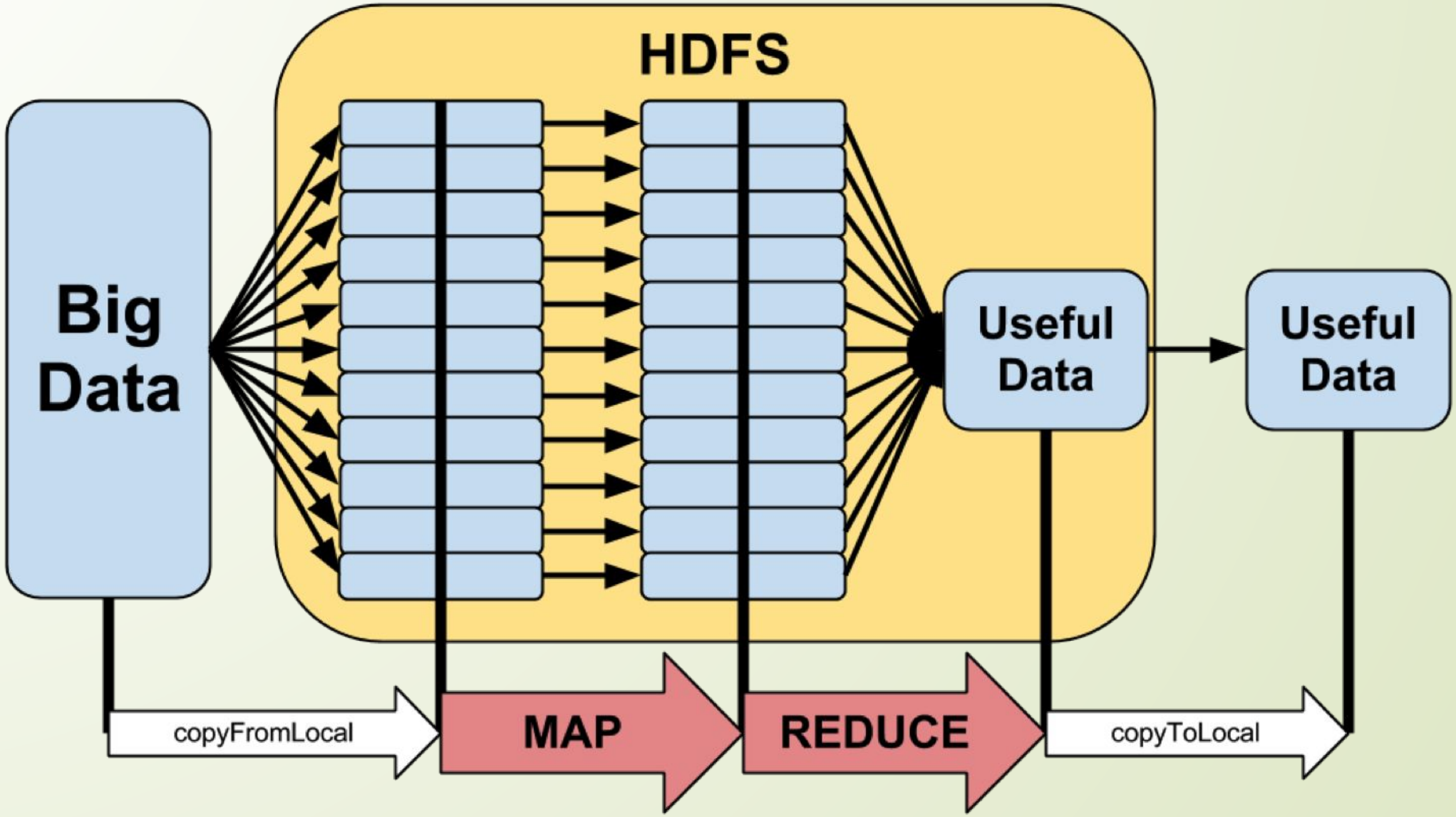
Hard drive space

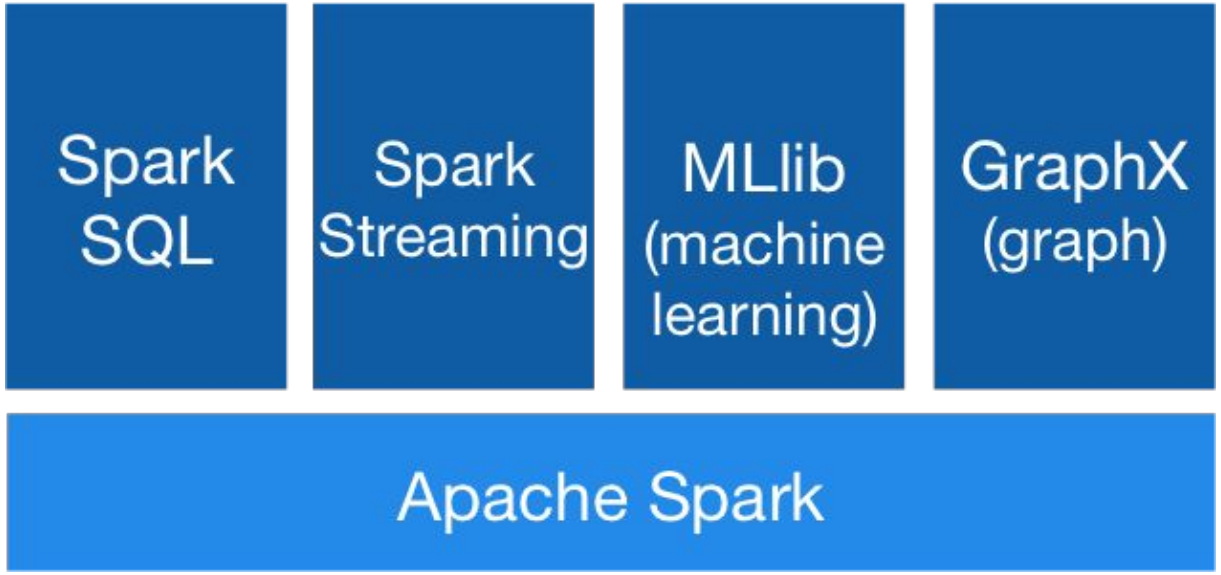
BANDWIDTH



Hadoop

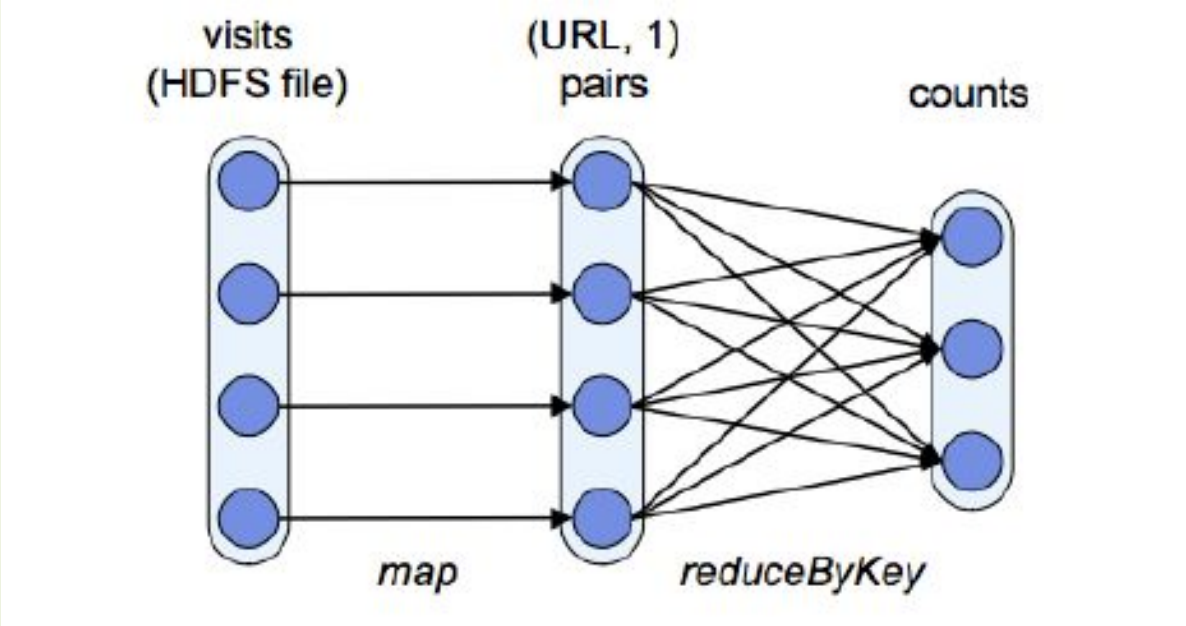
MapReduce


HDFS
Big Data
Useful Data
Useful Data
copyFromLocal
MAP
REDUCE
copyToLocal


Spark SQL
Spark Streaming
MLlib (machine learning)
GraphX (graph)
Apache Spark

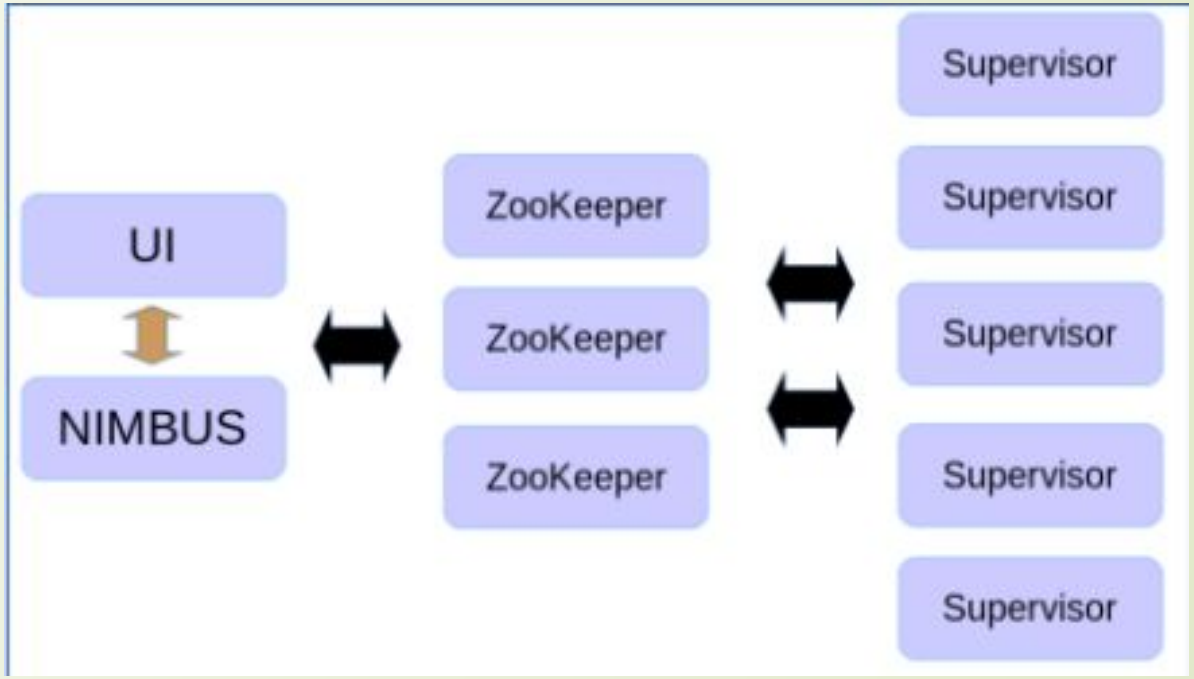
Spark

Resilient Distributed Datasets (RDDs)

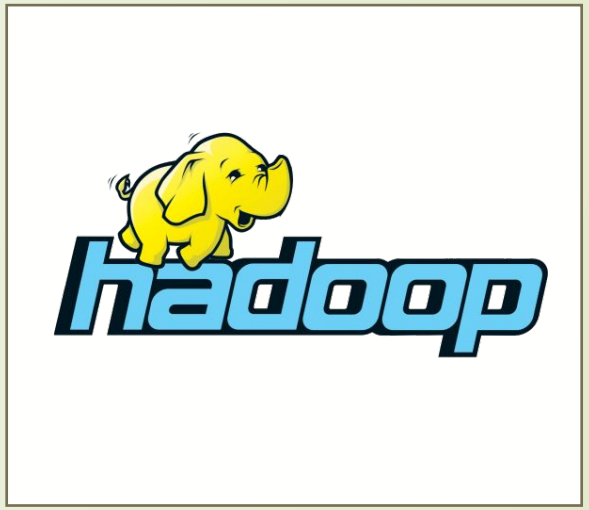


Spark Streaming

Storm framework



Discussion



THANKS FOR
WATCHING