MD5 HASH CRACKING USING A MAPREDUCE

Hengjie Wang & Jordan Thoms

INTRODUCTION

INTRODUCTION

- Volunteer your Password

INTRODUCTION

- Volunteer your Password
- What is a MapReduce

INTRODUCTION

- Volunteer your Password
- What is a MapReduce
- Technical Issues

INTRODUCTION

- Volunteer your Password
- What is a MapReduce
- Technical Issues
- Implementing our own MapReduce Architecture

INTRODUCTION

- Volunteer your Password
- What is a MapReduce
- Technical Issues
- Implementing our own MapReduce Architecture
- Benchmarks

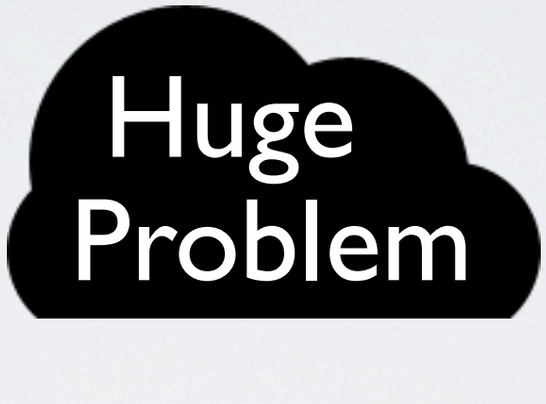
MAPREDUCE

- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.

- **Google** uses it

MAPREDUCE

- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.
- **Google** uses it

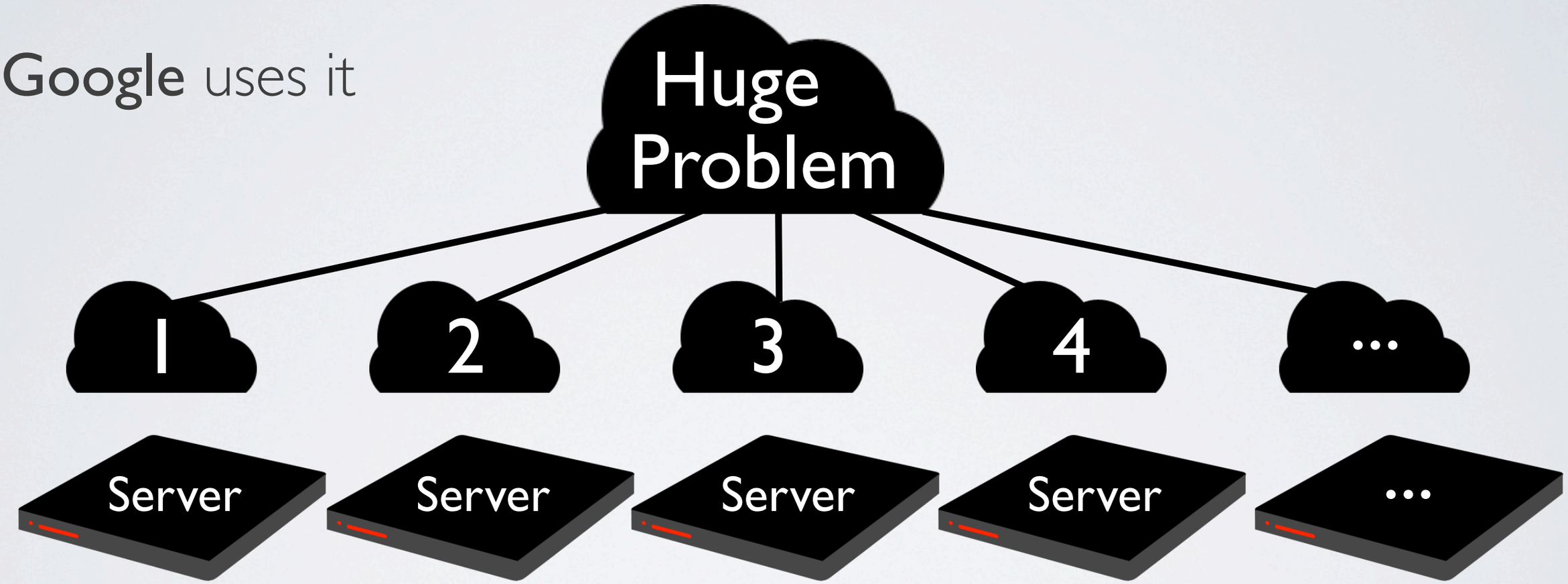


MAPREDUCE

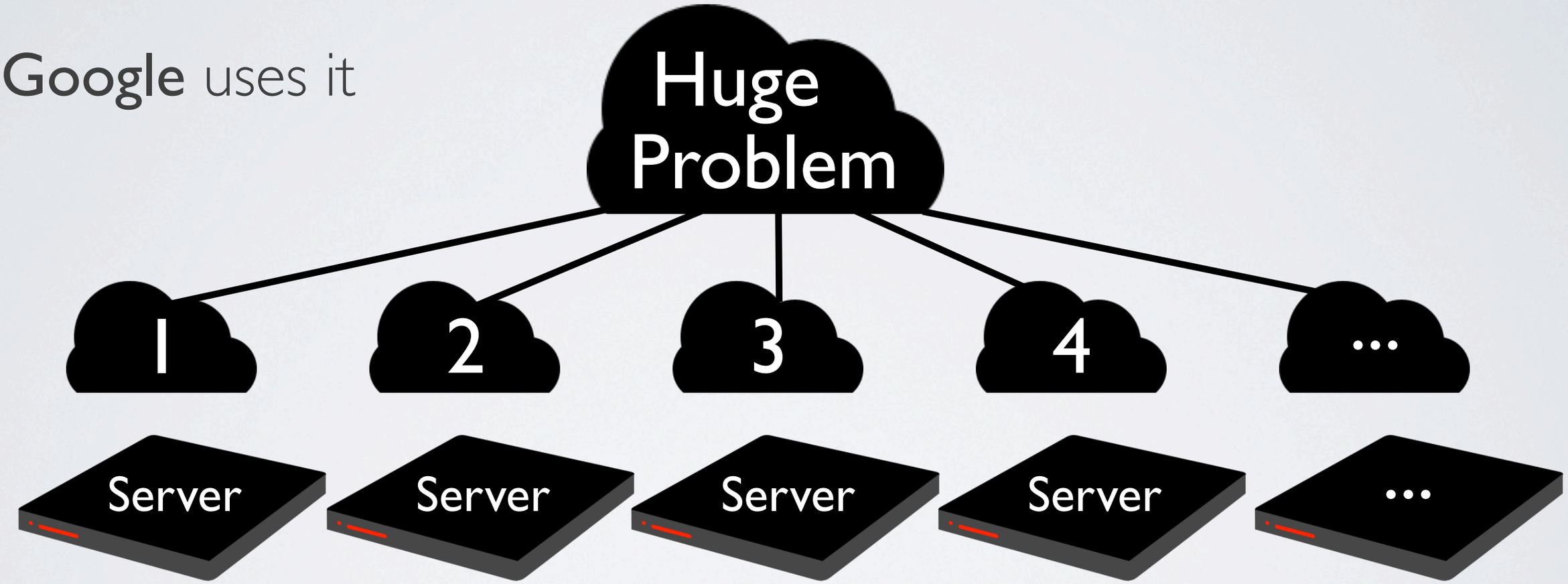
- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.
- **Google** uses it

MAPREDUCE

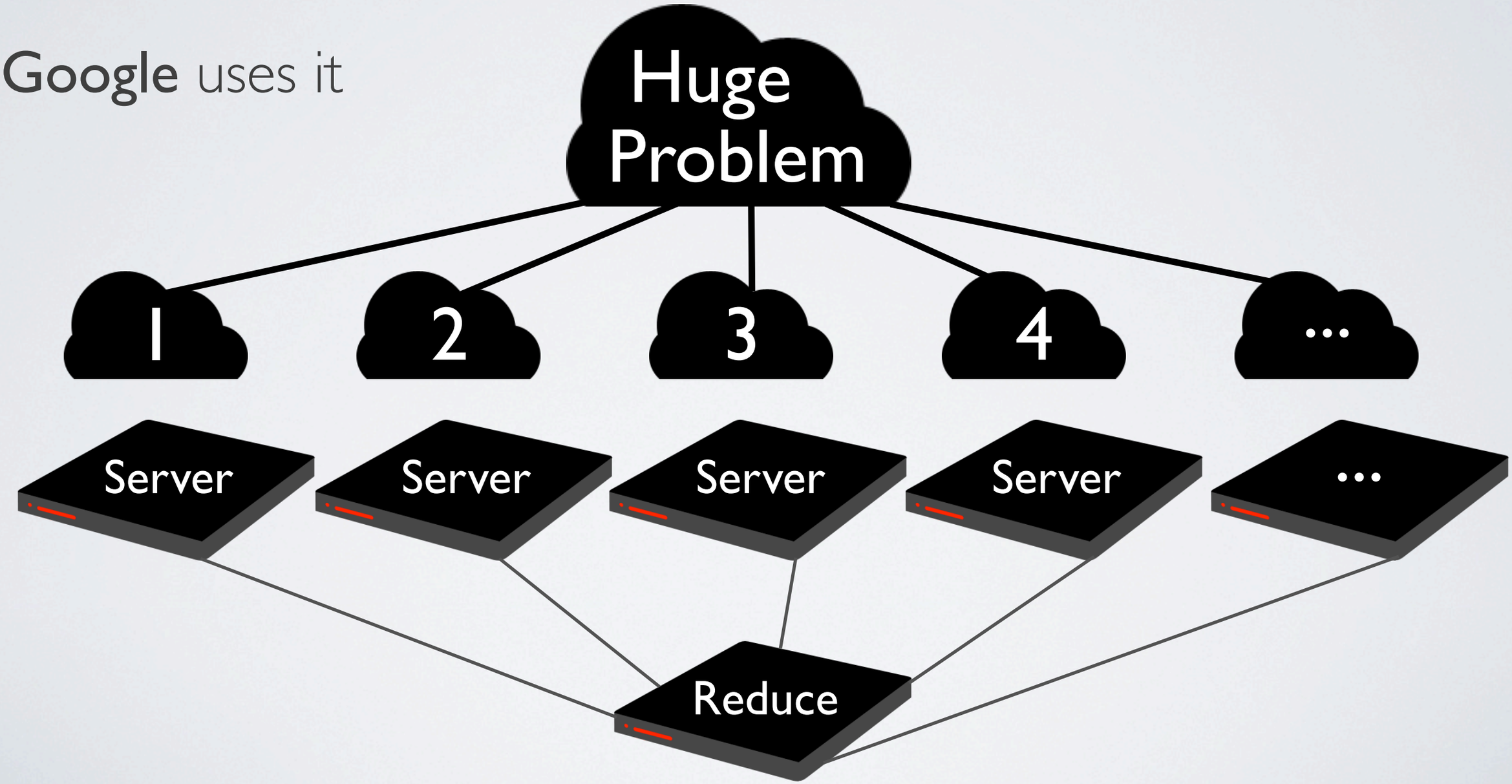
- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.
- **Google** uses it



- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.
- **Google** uses it

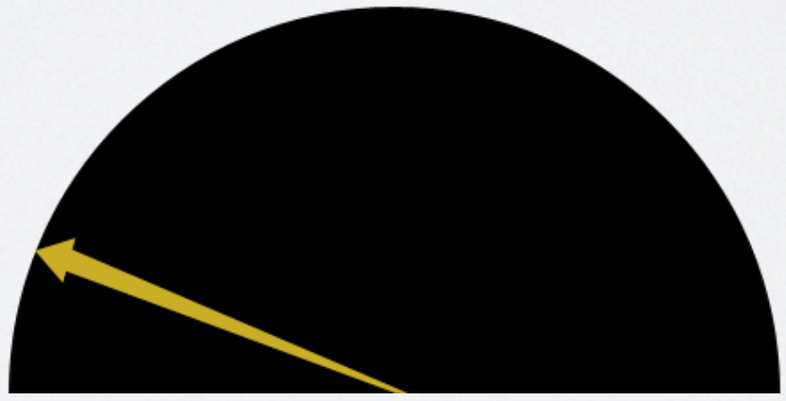


- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.
- **Google** uses it



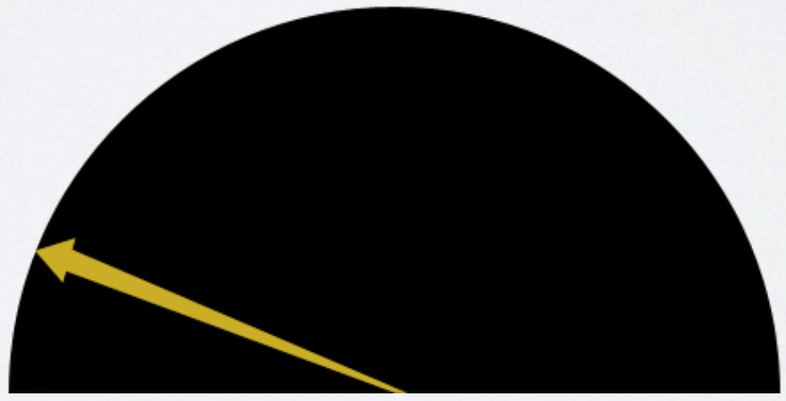
TECHNICAL ISSUES

TECHNICAL ISSUES



Slow single node performance

TECHNICAL ISSUES



Slow single node performance

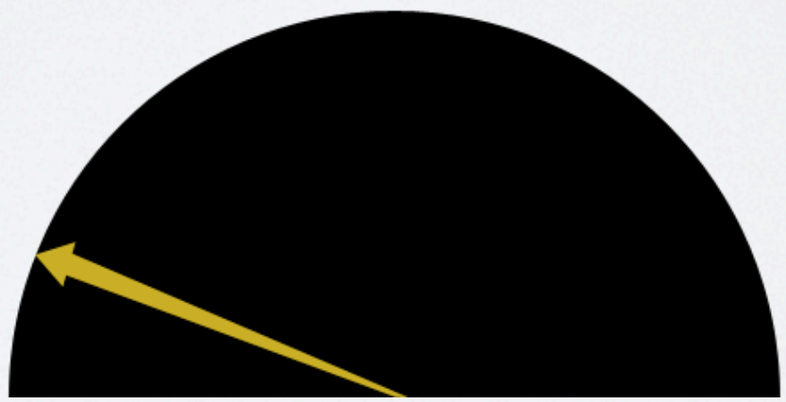
Actual	Expected
?	100

Unpredictable no. of workers

TECHNICAL ISSUES

- Previous used **Google App Engine** (and it sucked)
  - **Uneven** problem mapping
  - **Slow** single worker node performance
  - **Unpredictable** number of worker nodes

Uneven problem mapping



Slow single node performance

Actual	Expected
?	100

Unpredictable no. of workers

SOLUTION

- Apache **Hadoop**
- Similar to **Google App Engine**
- Complete control of our own **servers**

APACHE HADOOP

- Similar to Google App Engine
- Distributed File System
- Map Reduce Engine built on top of file system
- Highly available and scalable
  (e.g. Facebook uses it to store 21 petabytes of data)

SERVERS

- Deploy servers
  - JUJU
    - Deploy thousands of servers easily
- Using Amazon Web Service's EC2 server instances