MD5 HASH CRACKING USING A MAPREDUCE

Hengjie Wang & Jordan Thoms

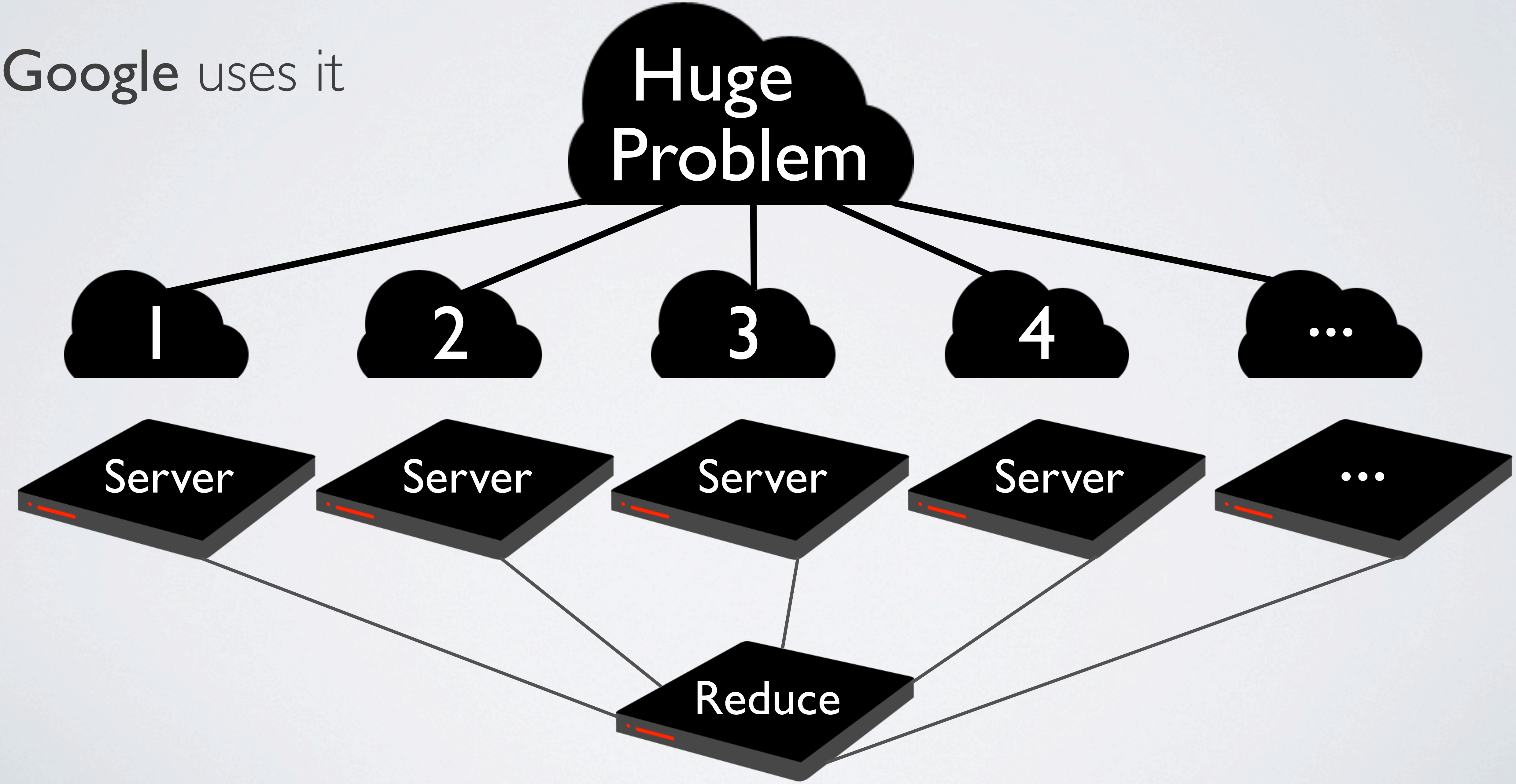
INTRODUCTION

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

MAPREDUCE

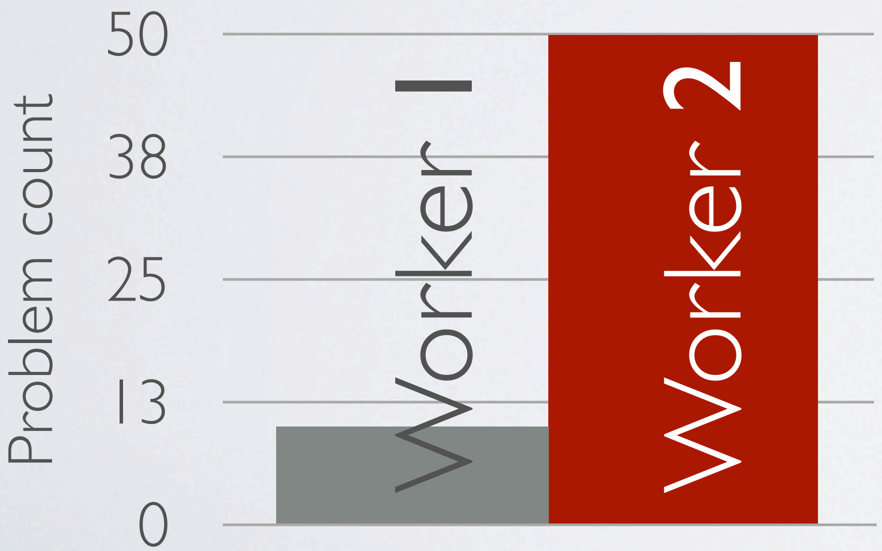
- Paradigm for **distributing work** on highly parallelizable problems with huge datasets **over a cluster of servers**.

- **Google** uses it

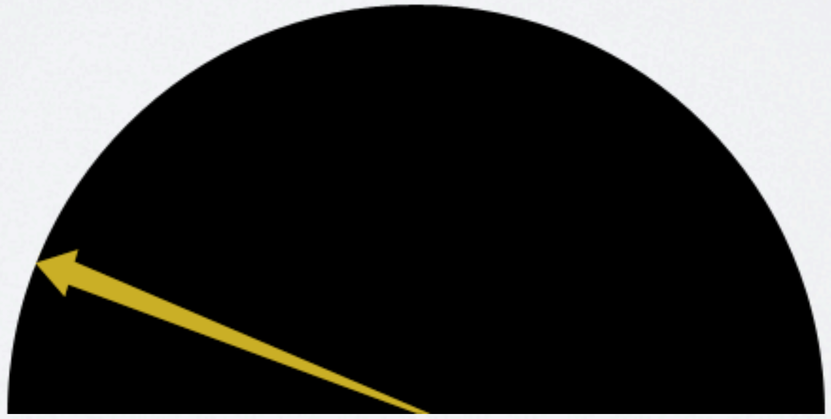


TECHNICAL ISSUES

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



Uneven work distribution



Slow single node performance

Actual	Expected
?	100

Unpredictable no. of workers

SOLUTION:
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

- We need to **deploy our own servers**
- JUJU: **Deploy thousands** of servers easily
- Use Amazon Web Service's **EC2** server instances

1	5	9	13	17
2	6	10	14	18
3	7	11	15	19
4	8	12	16	20

juju

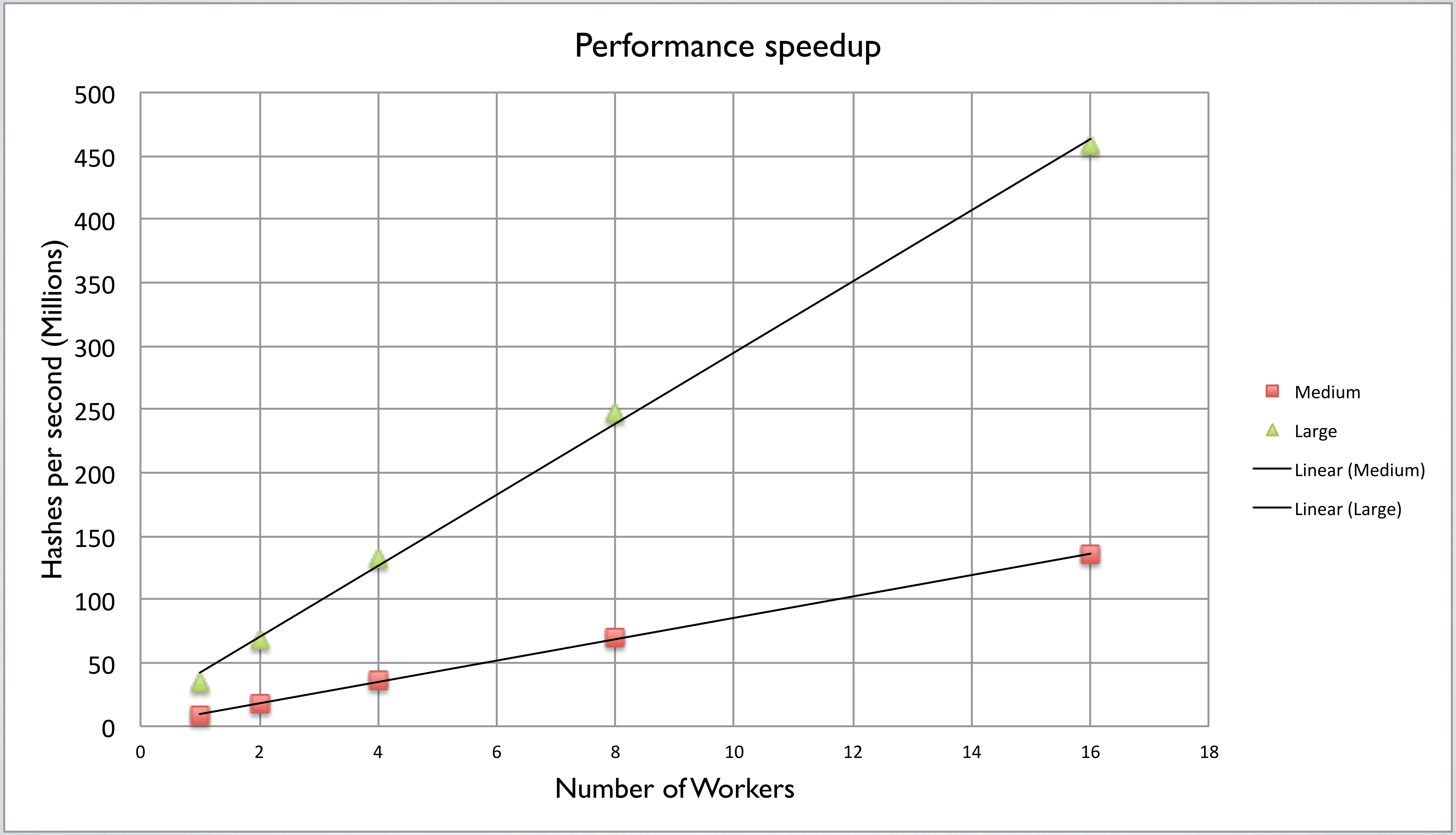
BENCHMARK METHODOLOGY

- **Separate tests** on different hardware
- Hardware (Amazon EC2):
  - High CPU **Medium** instance
  - High CPU **Extra Large** instance

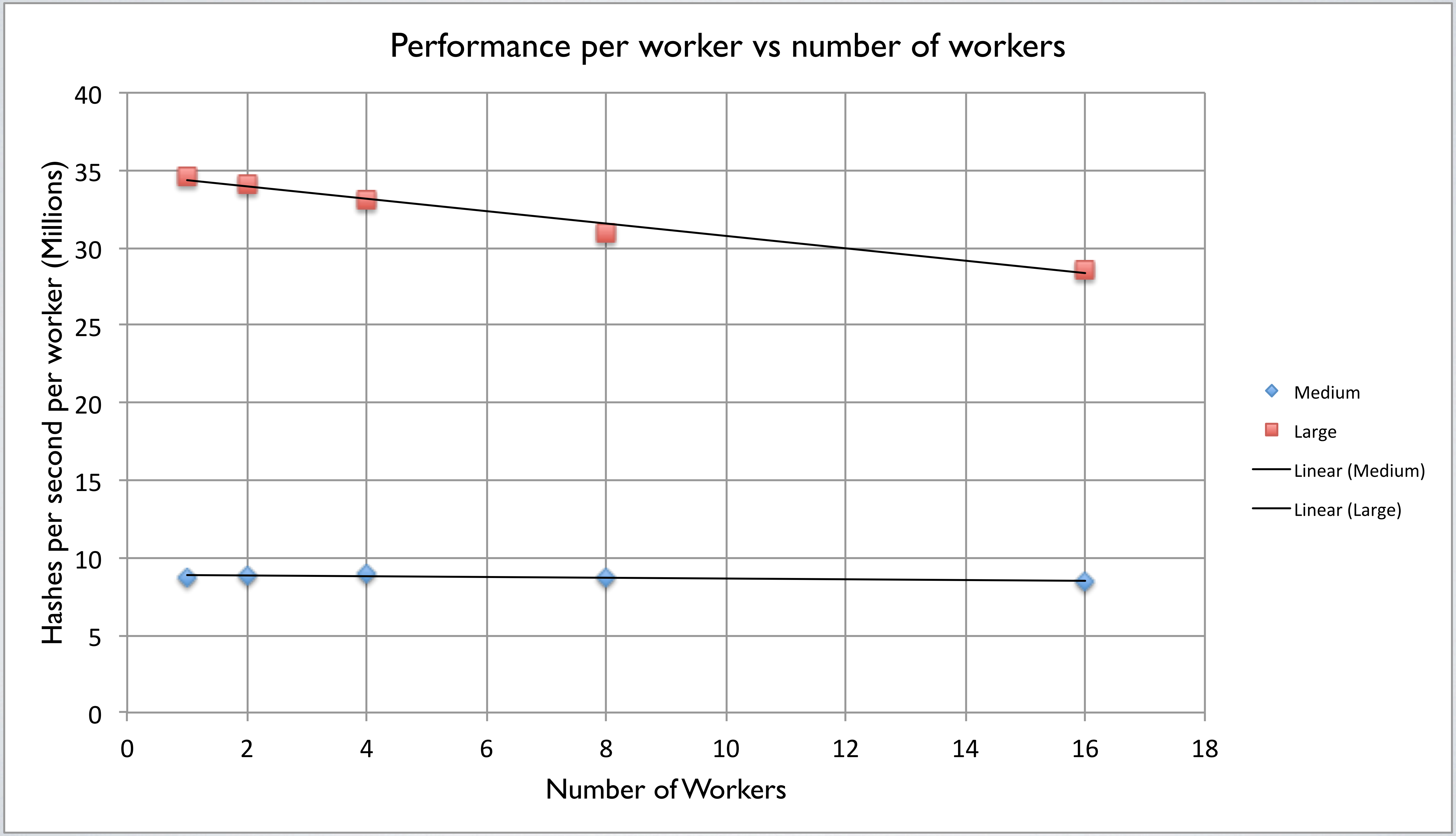
	Virtual Cores*	Threads	RAM
Medium	2	2	1.7GB
Large	8	12	7GB

*One Virtual Core is 2.5 EC2 Compute Unit. One EC2 Compute Unit provides the equivalent CPU capacity of a **1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor**.

RESULTS


Performance speedup
Hashes per second (Millions)
0
50
100
150
200
250
300
350
400
450
500
Number of Workers
0
2
4
6
8
10
12
14
16
18
Medium
Large
Linear (Medium)
Linear (Large)

RESULTS


Performance per worker vs number of workers
Hashes per second per worker (Millions)
40
35
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18
Number of Workers
◆ Medium
■ Large
Linear (Medium)
Linear (Large)

AND YOUR PASSWORD IS...

SUMMARY

- Hadoop is better
- Linear increase in performance as server increases
- Hackers can effectively crack MD5 hashes
- Typical user: make your passwords long!
- Website devs: Use SHA512 instead