

Big Data in Cloud: Challenges & Opportunities

Wei Wang

Department of Computer Science & Engineering

HKUST

Fall 2015

Data, data, data!



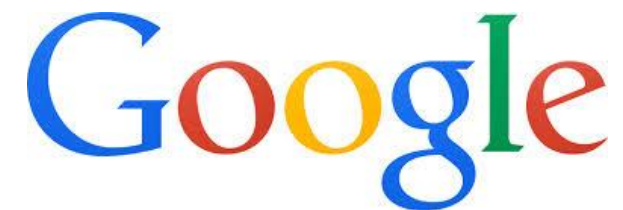
40 TB per second



1.1M requests per second (2013)



10 TB every 30 minutes



Hadoop cluster: 330K nodes, 365 PB (2014)

20B web pages a single day (2012)



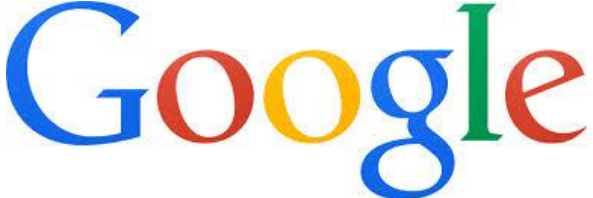
“640K ought to be enough for anybody.”
— Bill Gates (1981)

How can we manage and process big data?



Datacenter

Estimated # servers

 > 1M

 Microsoft ~ 1M





Several 100,000s each





“I think there is a world market for maybe **five** computers.”

— Thomas Watson (1943)

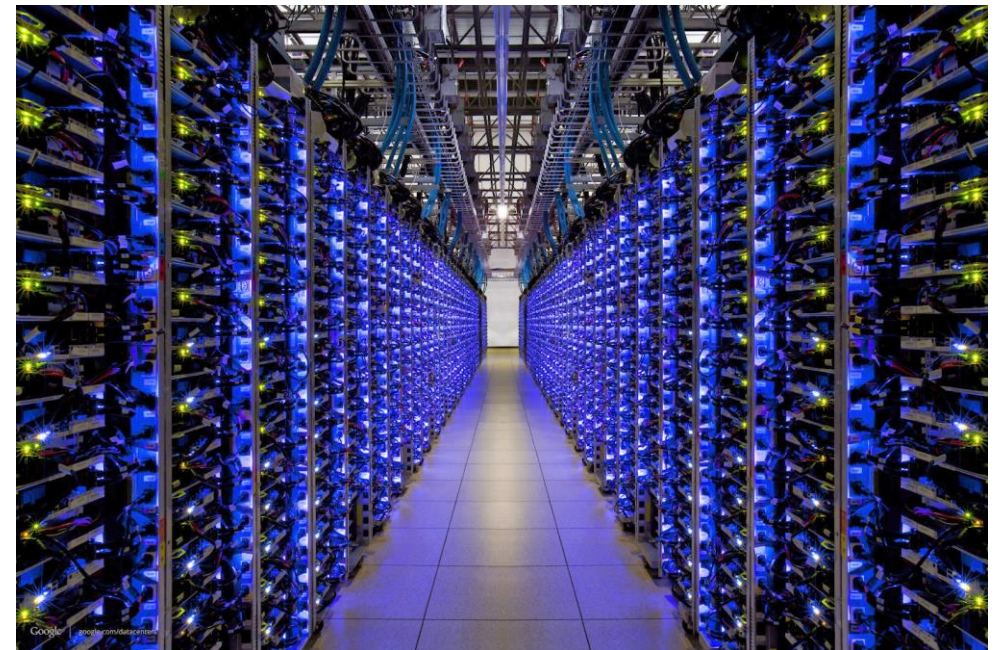
How do we program this thing?



The datacenter *is* a computer



OS X Yosemite



Challenges & Opportunities

Limited bisection b/w

New netw. topology

Diverse workload

General-purpose engine

Frequent failure

Fault tolerance

Poor utilization

Efficient scheduling

Unpredictable QoS

Fair sharing w/ predictable service isolation



“Simple things
should be
simple, complex
things should be
possible.”

— Alan Kay