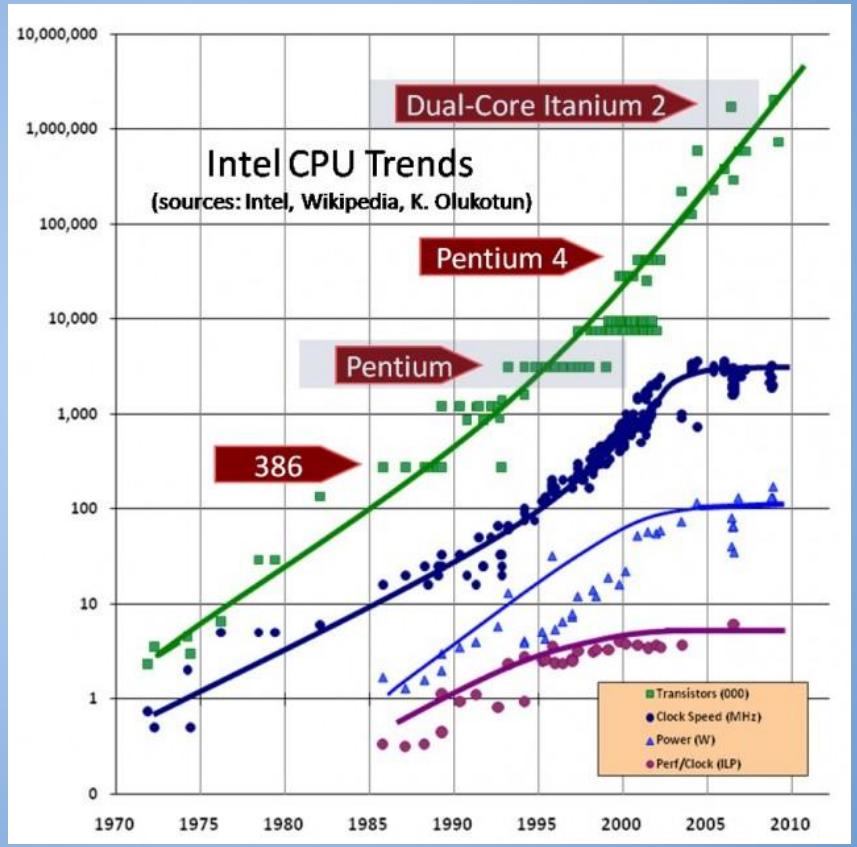
There's Plenty of Room in the Cloud

[Shameless reference to Feynman's talk from 1959]

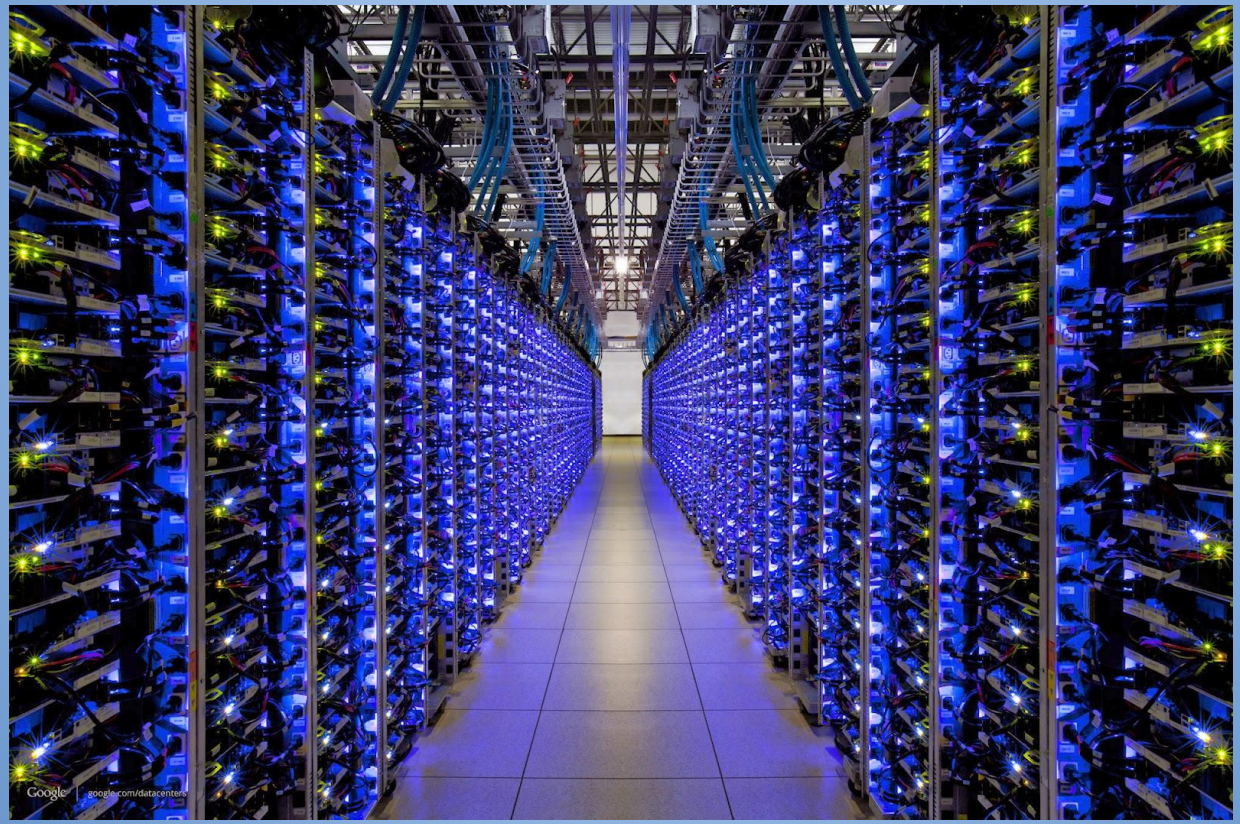
Lecturer: *Zoran Dimitrijevic <zorand@gmail.com>*
Altiscale, Inc.

Spring 2015 CS290B -- Cloud Computing

50 Years of Moore's Law



Data center



Ubiquitous Access to Internet

- Free or Public
- But highly available, and high speed
- UCSB guest web?

Computation in Cloud

- MPI (Message-passing Interface)
- Map-reduce model (Google MapReduce)
- Spark (Resilient Distributed Datasets)
- No-SQL (bigtable-like) & scanlets
- PubSub model (Apache Kafka)

Computational Models

- Sequential, select_server/event based
- Multithreaded, shared-memory
  - locks, semaphores,queues, dynamic threads, ...
- Remote Procedure Calls (RPC)
- Message-Passing Interface
  - Point-to-point (send, receive)
  - Collective (broadcast, reduce, barrier)
  - Dynamic process management, fault-tolerance

- Web services, mySQL
- Clusters of Linux Machines
- Google Map-Reduce (c++)
- Hadoop Map-Reduce (java)
- Google BigTable, Scanlets, NoSQL
- Flume
- PubSub model, Apache Kafka
- Apache Spark (UC Berkeley)

[Map-reduce, Jeff Dean,Sanjay Ghemawat, Google, OSDI, 2004]

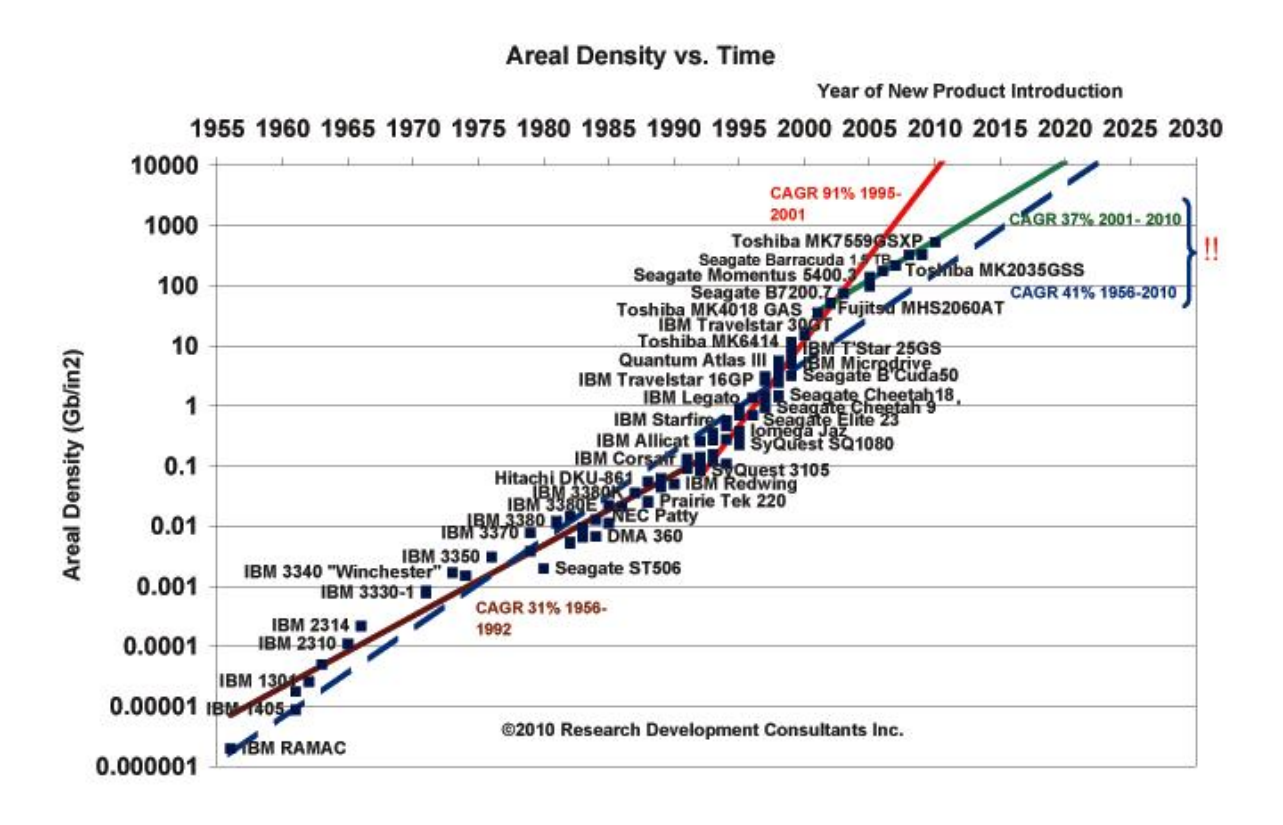
```
map(String key, String value):
  // key: document name
  // value: document contents
  for each word w in value:
    EmitIntermediate(w, "1");

reduce(String key, Iterator values):
  // key: a word
  // values: a list of counts
  int result = 0;
  for each v in values:
    result += ParseInt(v);
  Emit(AsString(result));
```

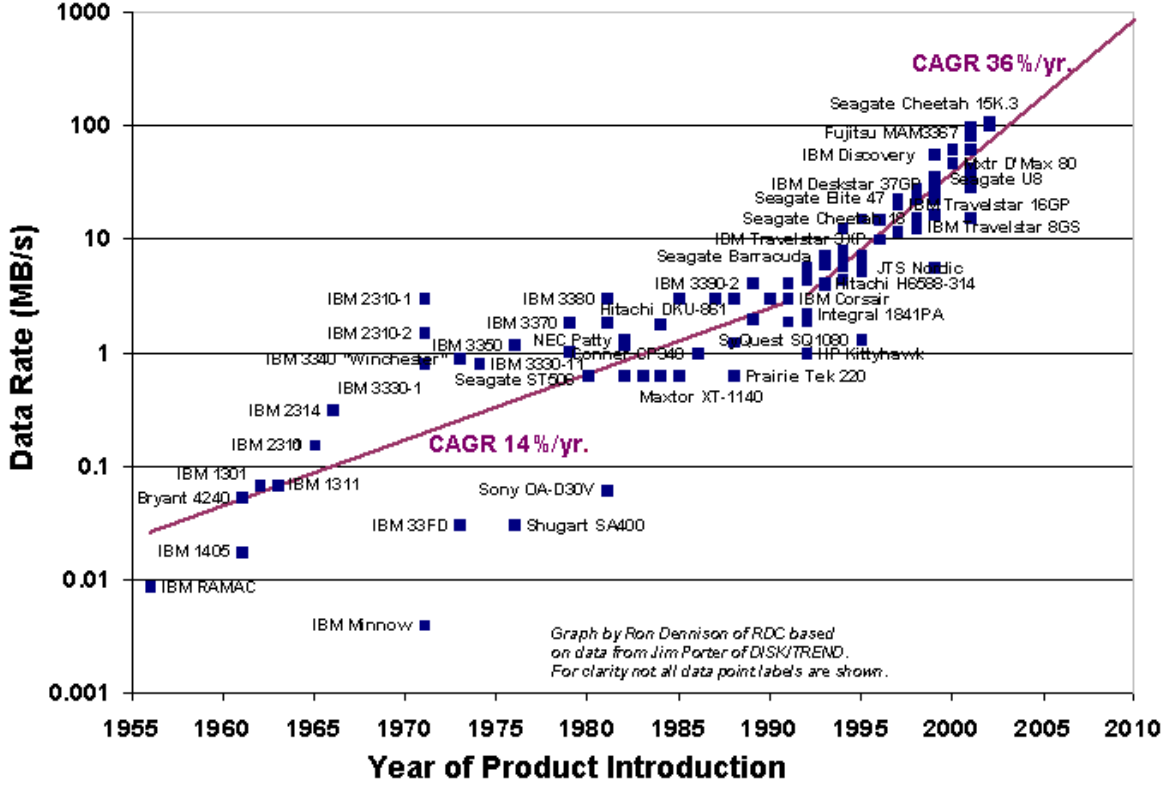
Storage

- File systems
- RAIDs
- NFS vs. Active Storage
  - DAS (direct-attached storage
  - NAS (network-attached storage)
  - SAN (storage area network)
- Distributed File Systems

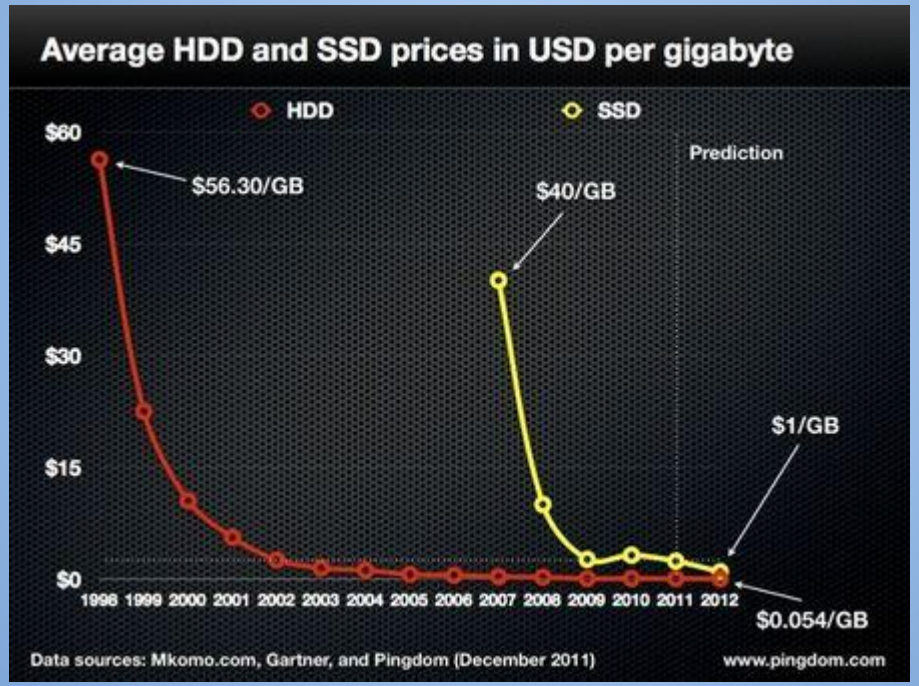
Hard disk Density



IDEMA Data Rate vs. Time



Harddisks vs. SSD



Storage in Cloud

- *Collocate computation and storage*
- Google File System (GFS)
- Hadoop DFS (HDFS)
- Bigtable/ NoSQL storage / Spanner
- Service-oriented storage (dropbox, etc.)

Distributed File Systems

- Name-node (manages meta-data)
  - Replica management
  - Sharding of namenode, master-election
  - Lease management
  - Single or multiple writers? Append-only?
- Data-node (manages data chunks)
  - Local checksums
  - Client libs access data nodes
  - Copy-on-write, snapshots...

Archival Storage

- HDFS RAID
- Reed-solomon coding / Erasure coding
- Object-based storage
- Read/Write/Modify tradeoffs
- Overhead for recovery
  - Replication: fast, many-to-many
  - RS coding (n+m): n x reads, CPU, one write

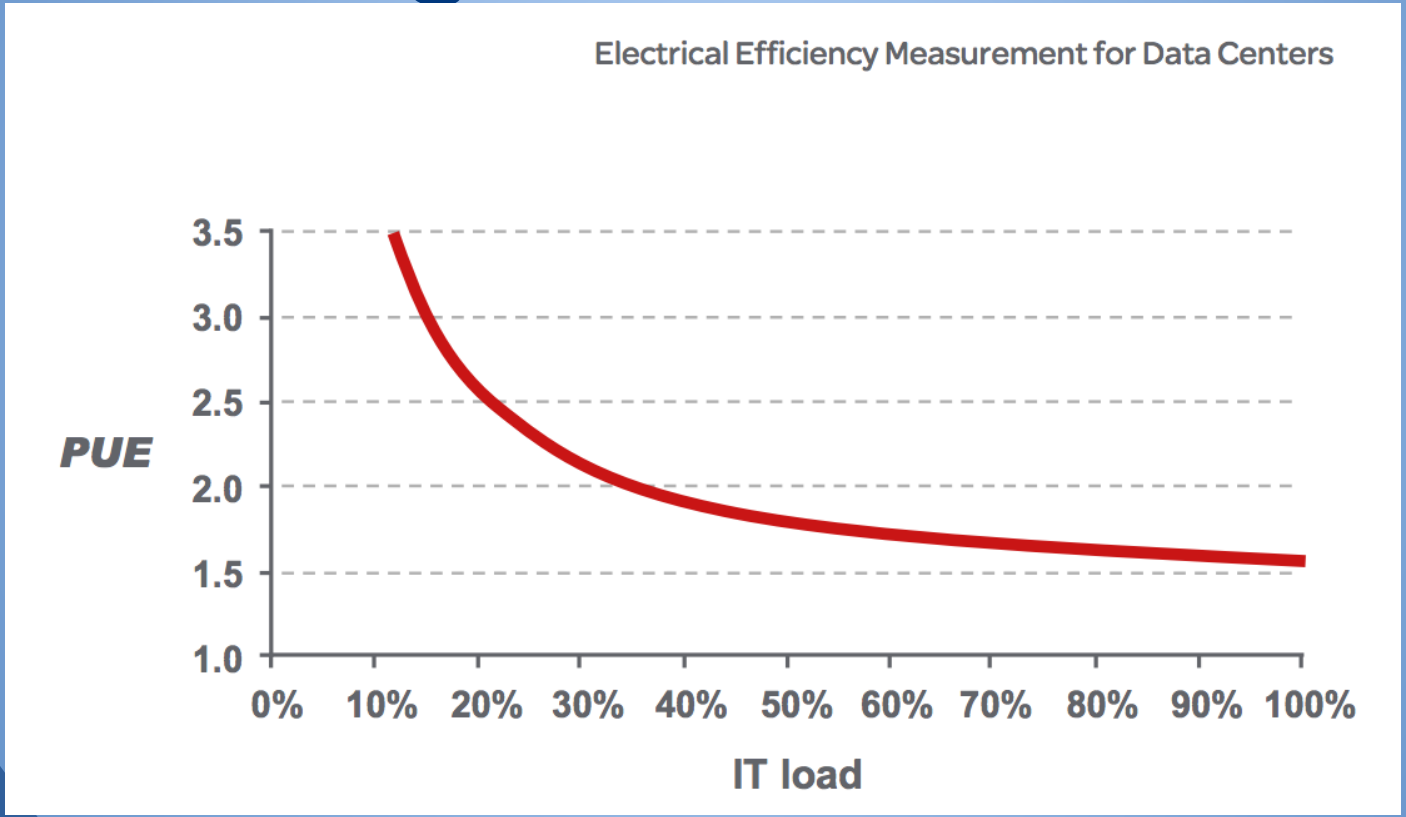
Sharing nodes

- Multitenancy
- Google BORG
- Hadoop YARN
- Apache Mesos
- Docker / Orchestration

Cost-efficiency

- Utilization (CPU, storage, memory, network)
- Power consumption (datacenter level)
- Sharing clusters
  - quota oversubscription
  - scaling
- Operation cost
  - sharing SRE/OPS resources, on-call
  - global 24/7 support (time zones, holidays)

Power Usage Effectiveness



Open Source Community

- Linux, Apache
- Github, Jira
- Jenkins
- Chef, Docker
- Startups
- Academia
- Large companies and Licenses

Apache Hadoop

- Yahoo, Hortonworks, Cloudera
- Apache
  - Hadoop MR, HDFS, YARN, tools, zookeeper
  - Avro, Cassandra, HBase, Hive
  - Pig, Tez, Spark, Mahout
  - Kafka, HDFS with erasure coding, ...

Questions?

Security, Privacy, Encryption...

- Private-public key cryptography
- Certificates
- Kerberos and authentication
- Delegation tokens
- Encryption at rest
- SSL and data transfer
- Computation (who gets the keys?)

Can we run secure cloud-computing platform?

- Docker/linux containers vs. VMs
- Local file system storage (/tmp)
- Unix root: who can read memory?
- Key distribution, who has root?
- Fault-tolerance
  - Moving data is expensive
  - Who has the keys to restart nodes

References

- http://www.extremetech.com/extreme/203490-moores-law-is-dead-long-live-moores-law
- http://en.wikipedia.org/wiki/PDP-11
- http://en.wikipedia.org/wiki/Nexus_6
- http://en.wikipedia.org/wiki/IPhone_6
- https://courses.cs.washington.edu/courses/cse590s/03au/grochowski-trends.pdf
- http://www.google.com/about/datacenters/efficiency/internal/#servers
- https://www.cs.berkeley.edu/~matei/papers/2012/nsdi_spark.pdf
- http://rondennison.com/presentations.htm
- http://en.wikipedia.org/wiki/Message_Passing_Interface
- http://research.google.com/archive/mapreduce.html
- http://static.googleusercontent.com/media/research.google.com/en/us/archive/bigtable-osdi06.pdf
- http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/spanner-osdi2012.pdf
- http://static.googleusercontent.com/media/research.google.com/en/us/archive/gfs-sosp2003.pdf
- http://en.wikipedia.org/wiki/Reed%E2%80%93Solomon_error_correction
- http://smahesh.com/HadoopUSC/
- https://code.facebook.com/posts/536638663113101/saving-capacity-with-hdfs-raid/
- https://blog.docker.com/2015/02/orchestrating-docker-with-machine-swarm-and-compose/
- http://blog.docker.com/tag/orchestration/
- http://mesos.apache.org/
- http://www.apcmedia.com/salestools/NRAN-72754V/NRAN-72754V_R2_EN.pdf?sdirect=true
- https://hadoop.apache.org/