Big Data Platforms for Artifical Intelligence

Keijo Heljanko

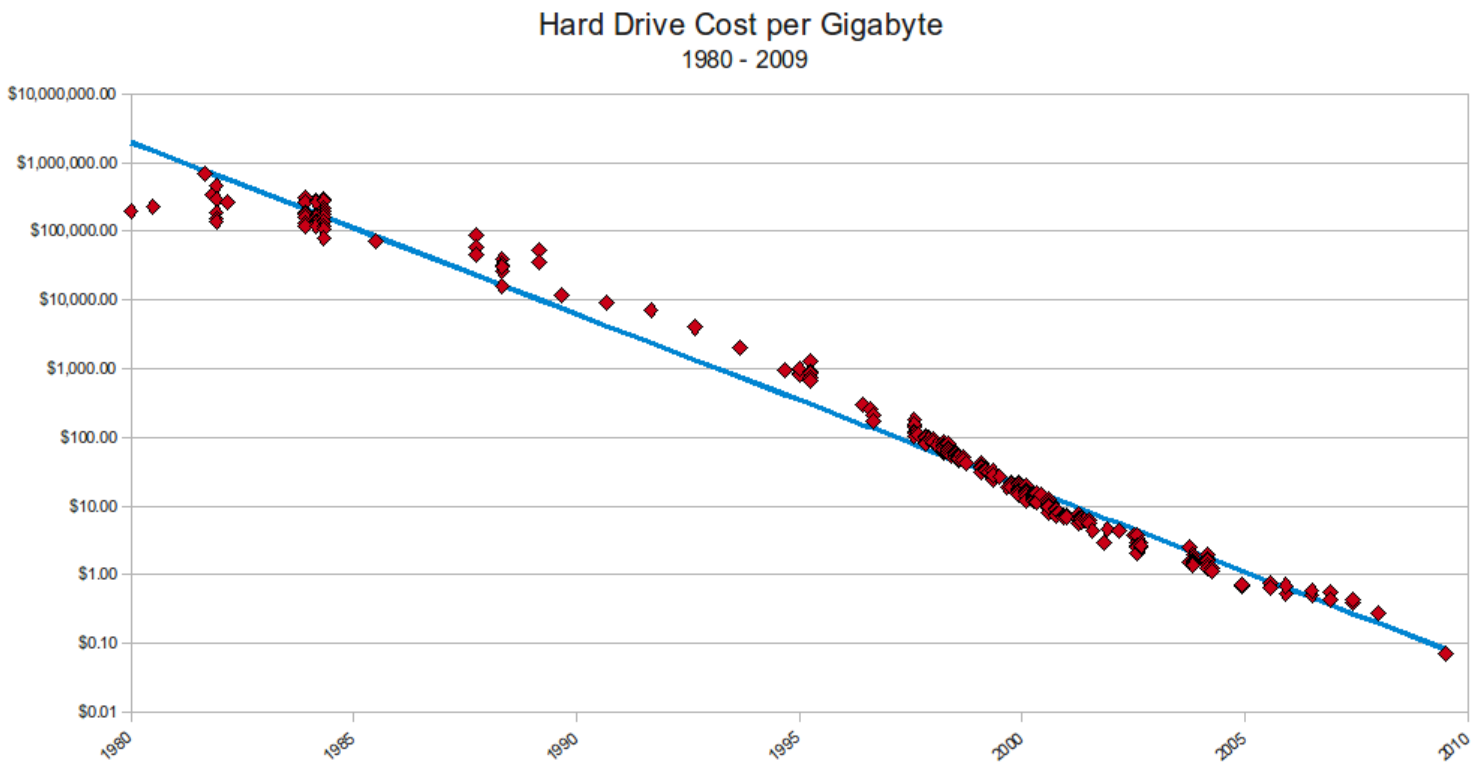
Department of Computer Science and
Helsinki Institute for Information Technology HIIT
School of Science, Aalto University
firstname.lastname@aalto.fi

4.5-2017

Big Data

- EMC sponsored IDC study estimates the Digital Universe to be 4.4 Zettabytes (4 400 000 000 TB) in 2013
- The amount of data is estimated to grow to 44 Zettabytes by 2020
- Data comes from: Video, digital images, sensor data, biological data, Internet sites, social media, Internet of Things, . . .
- Some examples:
  - Netflix is collecting 1 PB of data per month from its video service user behaviour, total data warehouse 20+ PB
  - Rovio is collecting in the order of 1 TB of data per day of games logs
- The problem of such large data massed, termed Big Data calls for new approaches to analyzing these data masses

Kryder's Law

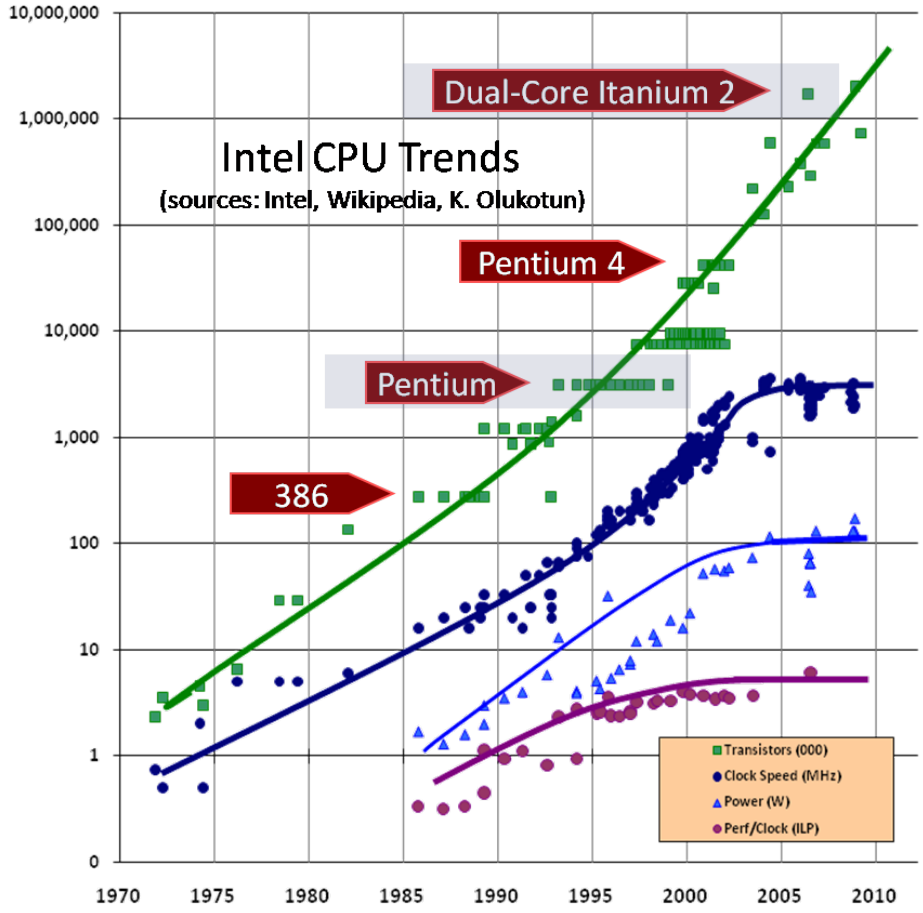


- Cost of storing a bit on a hard disk halves every 14 months
  `http://www.mkomo.com/cost-per-gigabyte`

Big Data approach to Business Intelligence

- Storing data is becoming cheaper every year
  - The cost of storage of raw data is quite small for most applications
  - Just store all the raw data for future use
  - Do “schema on read” - Make the data usable by structuring it when needed
  - Because the raw data has already been collected, there is a history of data to analyze when an opportunity arrises
  - Big Data platforms are needed to handle the vast masses of (potentially unstructured) data
  - The Gartner 3 Vs of Big Data: Volume, Variety, Velocity

No Single Threaded Performance Increases



- Herb Sutter: The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software. Dr. Dobb's Journal, 30(3), March 2005 (updated graph in August 2009).

Implications of the End of Free Lunch

- The clock speeds of microprocessors are not going to improve much in the foreseeable future
  - The efficiency gains in single threaded performance are going to be only moderate
- The number of transistors in a microprocessor is still growing at a high rate
  - One of the main uses of transistors has been to increase the number of computing cores the processor has
  - The number of cores in a low end workstation (as those employed in large scale datacenters) is going to keep on steadily growing
- Programming models need to change to efficiently exploit all the available concurrency - scalability to high number of cores/processors will need to be a major focus

Dark Silicon - End of Moore's law in Sight?

- Even worse news ahead, computing will be hitting a wall: Silicon is becoming energy constrained
- We will have much more transistors than what we can switch on and off at each clock cycle
- This is called: Dark Silicon
- For implications of dark silicon, see:
  - Hadi Esmaeilzadeh, Emily R. Blem, Rene St. Amant, Karthikeyan Sankaralingam, Doug Burger: Dark Silicon and the End of Multicore Scaling. IEEE Micro 32(3): 122-134 (2012)
  - Michael B. Taylor: A Landscape of the New Dark Silicon Design Regime. IEEE Micro 33(5): 8-19 (2013)

Dark Silicon - Improving Energy Efficiency

- To combat the power limits more energy efficient computing is needed
- GPGPU computing is improving the energy efficiency of training deep neural networks
- Special purpose hardware (ASICs) for inference of deep neural networks: Google Tensor Prosessing Unit
- Software/hardware co-design: Moving from 32 bit floating point to 8 bit integers for neural networks improves energy efficiency
- New hardware for low latency non-volatile memories and GPGPUs are hard to keep fully utilized, wasting energy: Luiz Barroso, Mike Marty, David Patterson, Parthasarathy Ranganathan: Attack of the Killer Microseconds. Comm. of the ACM, Vol. 60 No. 4, Pages 48-54.

Bump in Kryder's Law since 2010

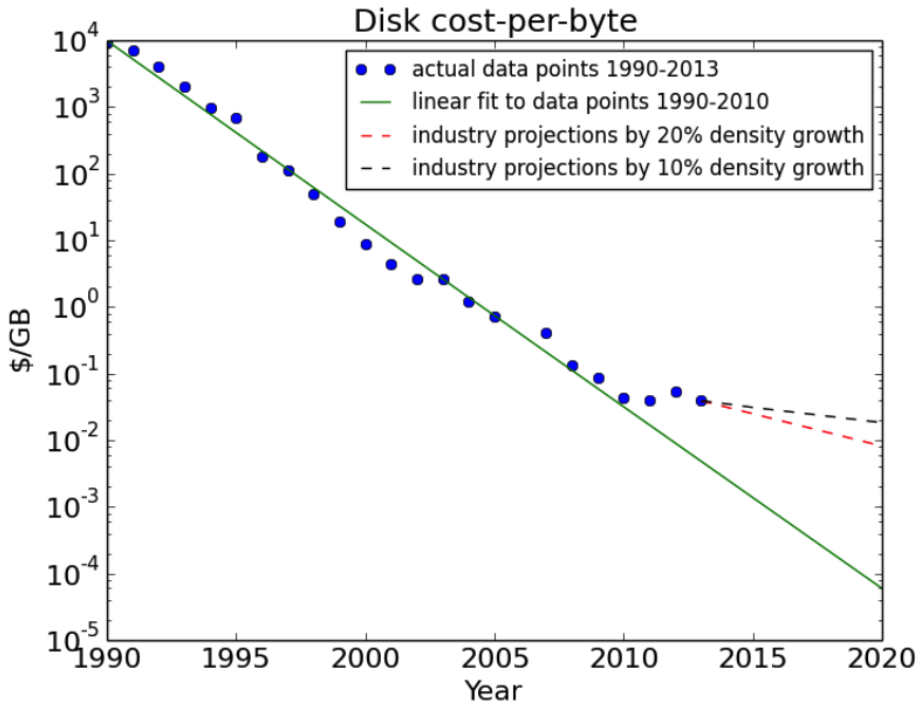


Fig. 1. Cost-per-byte decrease slowed dramatically in 2010 [2].

- Preeti Gupta, Avani Wildani, Ethan L. Miller, Daniel C. Rosenthal, Ian F. Adams, Christina E. Strong, Andy Hospodor: An Economic Perspective of Disk vs. Flash Media in Archival Storage. MASCOTS 2014: 249-254

Distributed Warehouse-scale Computing (WSC)

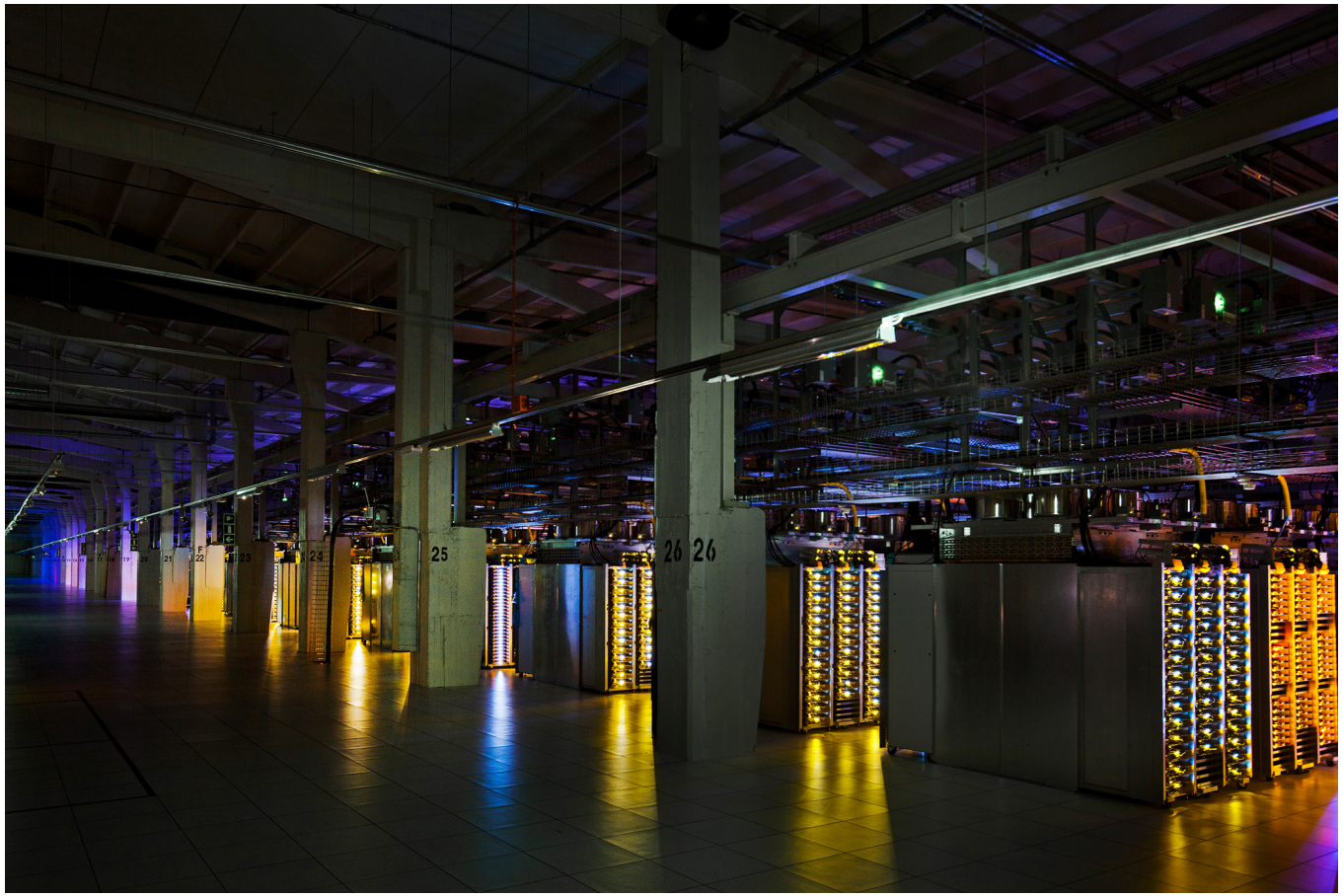
- Google is one of the companies who has had to deal with vast datasets too big to be processed by a single computer - the scale of data processed is truly Big Data
- The smallest unit of computation in Google scale is: Warehouse full of computers
- [WSC]: Luiz André Barroso, Jimmy Clidaras, and Urs Hölzle: *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second edition* Morgan & Claypool Publishers 2013

  http://dx.doi.org/10.2200/S00516ED2V01Y201306CAC024
- The WSC book says:
  "...we must treat the datacenter itself as one massive warehouse-scale computer (WSC)."

Google Summa Warehouse Scale Computer

Google Summa Sea Water Cooling

Google Summa Computer Hall



Google Servers from Year 2012

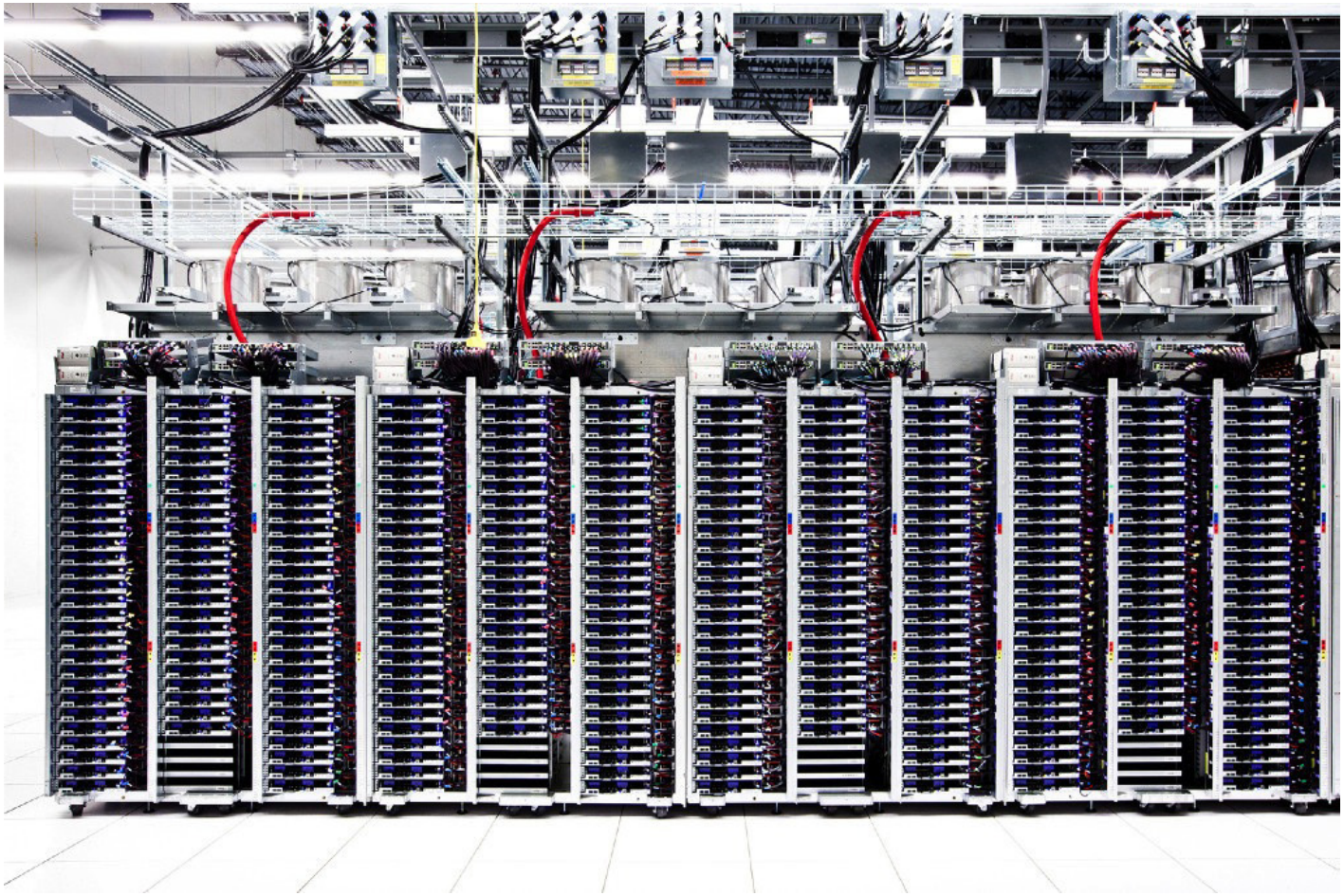


Figure: Google Server Racks form 2012, Figure 1.2 of [WSC]

Jeffrey Dean (Google): Joys of Real Hardware

Typical first year for a new cluster:

- $\approx$ 0.5 overheating (power down most machines in $<$ 5 mins $\approx$ 1-2 days to recover)
- $\approx$ 1 PDU failure ($\approx$ 500-1000 machines suddenly disappear, $\approx$ 6 hours to come back)
- $\approx$ 1 rack-move (plenty of warning, $\approx$ 500-1000 machines powered down, $\approx$ 6 hours)
- $\approx$ 1 network rewiring (rolling $\approx$5 % of machines down over 2-day span)
- $\approx$ 20 rack failures (40-80 machines instantly disappear, 1-6 hours to get back)
- $\approx$ 5 racks go wonky (40-80 machines see 50% packetloss)
- $\approx$ 8 network maintenances (4 might cause $\approx$ 30-minute random connectivity losses)

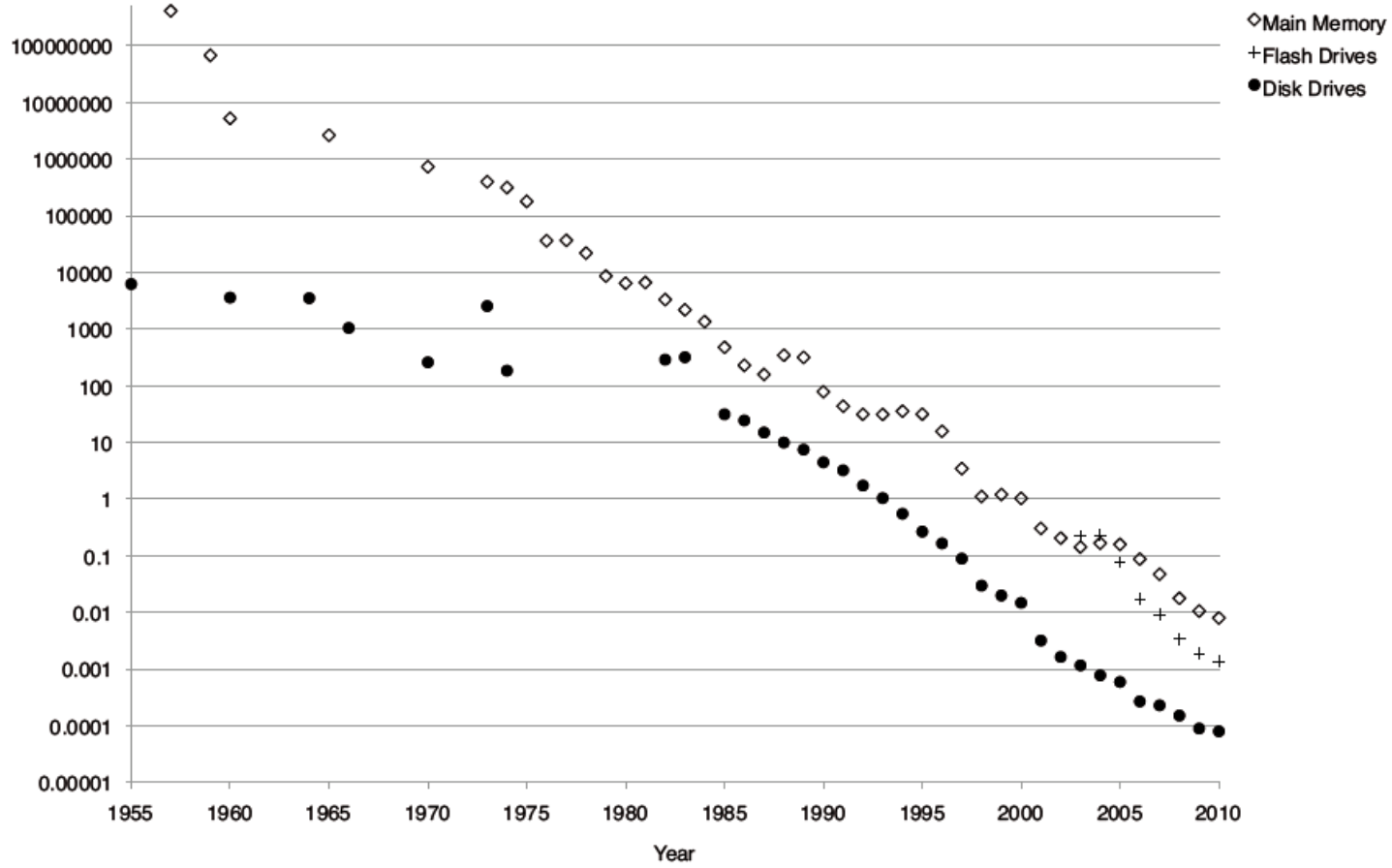
Jeffrey Dean (Google): Joys of Real Hardware (cnt.)

- ≈ 12 router reloads (takes out DNS and external vips for a couple minutes)
- ≈ 3 router failures (have to immediately pull traffic for an hour)
- ≈ dozens of minor 30-second blips for dns
- ≈ 1000 individual machine failures
- ≈ thousands of hard drive failures slow disks, bad memory, misconfigured machines, flaky machines, etc.
- Long distance links: wild dogs, sharks, dead horses, drunken hunters, etc.

Warehouse Scale Computers: Automation

- In order to maintain the warehouse scale computer with minimal staff, fault tolerance has to be fully automated:
  - All data is replicated to several hard disks in several computers
  - If one of the hard disks or computers breaks down, the software stack automatically reconfigures itself around the failure
  - For many applications this can be done fully transparently to the user of the system
- Thus warehouse scale computers can be maintained with minimal staff

Tape is Dead, Disk is Tape, RAM locality is King



- Trends of RAM, SSD, and HDD prices. From: H. Plattner and A. Zeier: In-Memory Data Management: An Inflection Point for Enterprise Applications

Tape is Dead, Disk is Tape, RAM locality is King

- RAM (and SSDs) are radically faster than HDDs: One should use RAM/SSDs whenever possible
- RAM is roughly the same price as HDDs were a decade earlier
  - Workloads that were viable with hard disks a decade ago are now viable in RAM
  - One should only use hard disk based storage for datasets that are not yet economically viable in RAM (or SSD)
  - The Big Data applications (HDD based massive storage) should consist of applications that were not economically feasible a decade ago using HDDs

Google MapReduce

- A scalable batch processing framework developed at Google for computing the Web index
- When dealing with Big Data (a substantial portion of the Internet in the case of Google!), the only viable option is to use hard disks in parallel to store and process it
- Some of the challenges for storage is coming from Web services to store and distribute pictures and videos
- We need a system that can effectively utilize hard disk parallelism and hide hard disk and other component failures from the programmer

Apache Hadoop Background

- An Open Source implementation of the MapReduce framework, originally developed by Doug Cutting and heavily used by e.g., Yahoo! and Facebook
- "Moving Computation is Cheaper than Moving Data" - Ship code to data, not data to code.
- Project Web page: `http://hadoop.apache.org/`

Hadoop - Linux of Big Data

- Hadoop = Open Source Distributed Operating System Distribution for Big Data
  - Based on “cloning” the Google architecture design
  - Fault tolerant distributed filesystem: HDFS
  - Batch processing for unstructured data: Hadoop MapReduce and Apache Pig (HDD), Apache Spark (RAM)
  - Distributed SQL database queries for analytics: Apache Hive, Spark SQL, Cloudera Impala, Facebook Presto
  - Fault tolerant real-time distributed databases: HBase, Kudu
  - Distributed machine learning libraries, text indexing & search, etc.
  - Data import and export with traditional databases: Sqoop

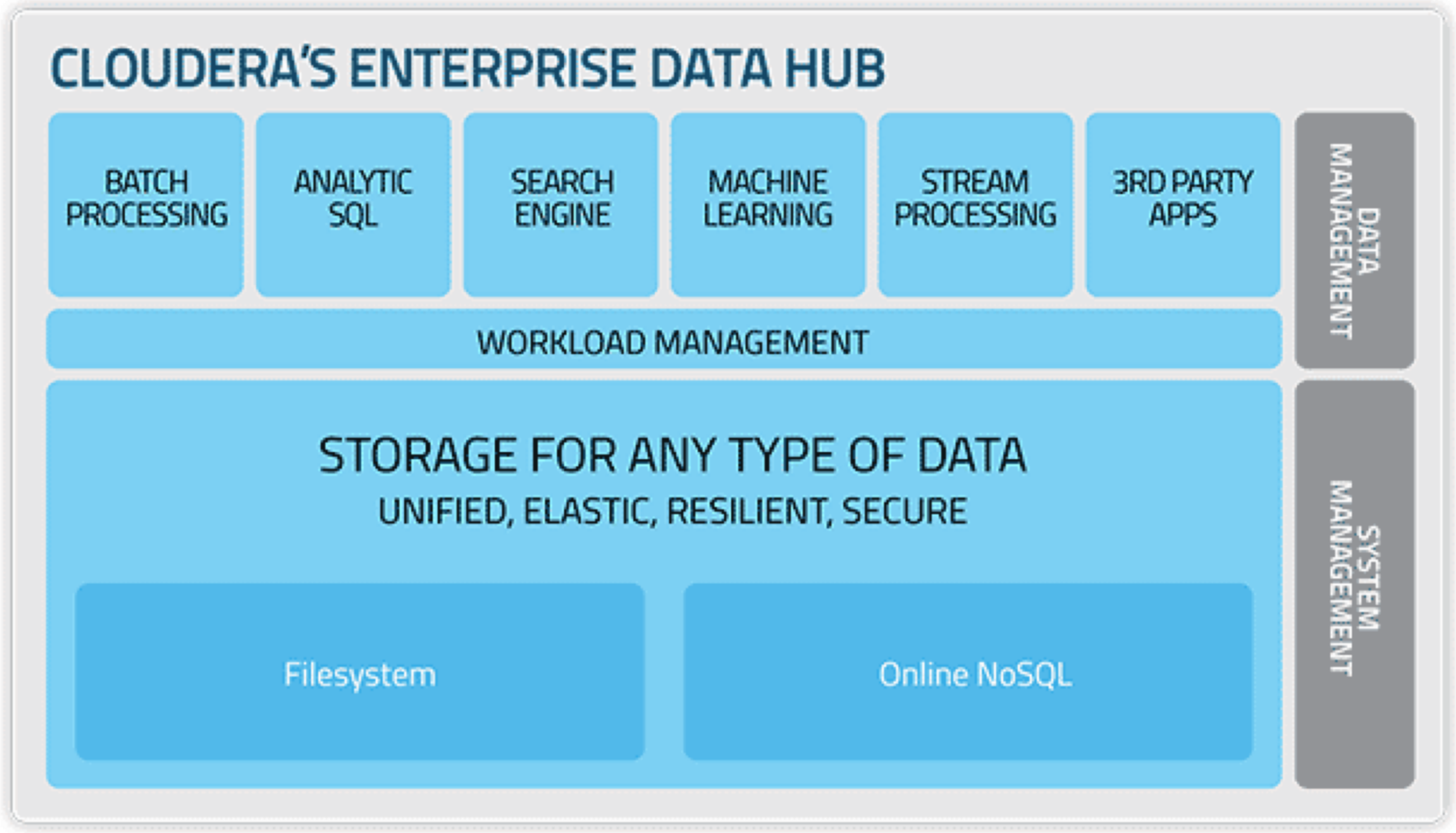
“SQL on Hadoop is the New Black”

- Many new massively parallel SQL implementations on top of Hadoop - A typical approach to implementing an Enterprise Data Warehouse
  - Apache Hive
  - Cloudera Impala and Kudu
  - Berkeley Spark SQL
  - Presto from Facebook

Commercial Hadoop Support

- Cloudera: Probably the largest Hadoop distributor, partially owned by Intel (740 million USD investment for 18% share). Available from:
  `http://www.cloudera.com/`
- Hortonworks: Yahoo! spin-off from their large Hadoop development team:
  `http://www.hortonworks.com/`
- MapR: A rewrite of much of Apache Hadoop in C++, including a new filesystem. API-compatible with Apache Hadoop.
  `http://www.mapr.com/`

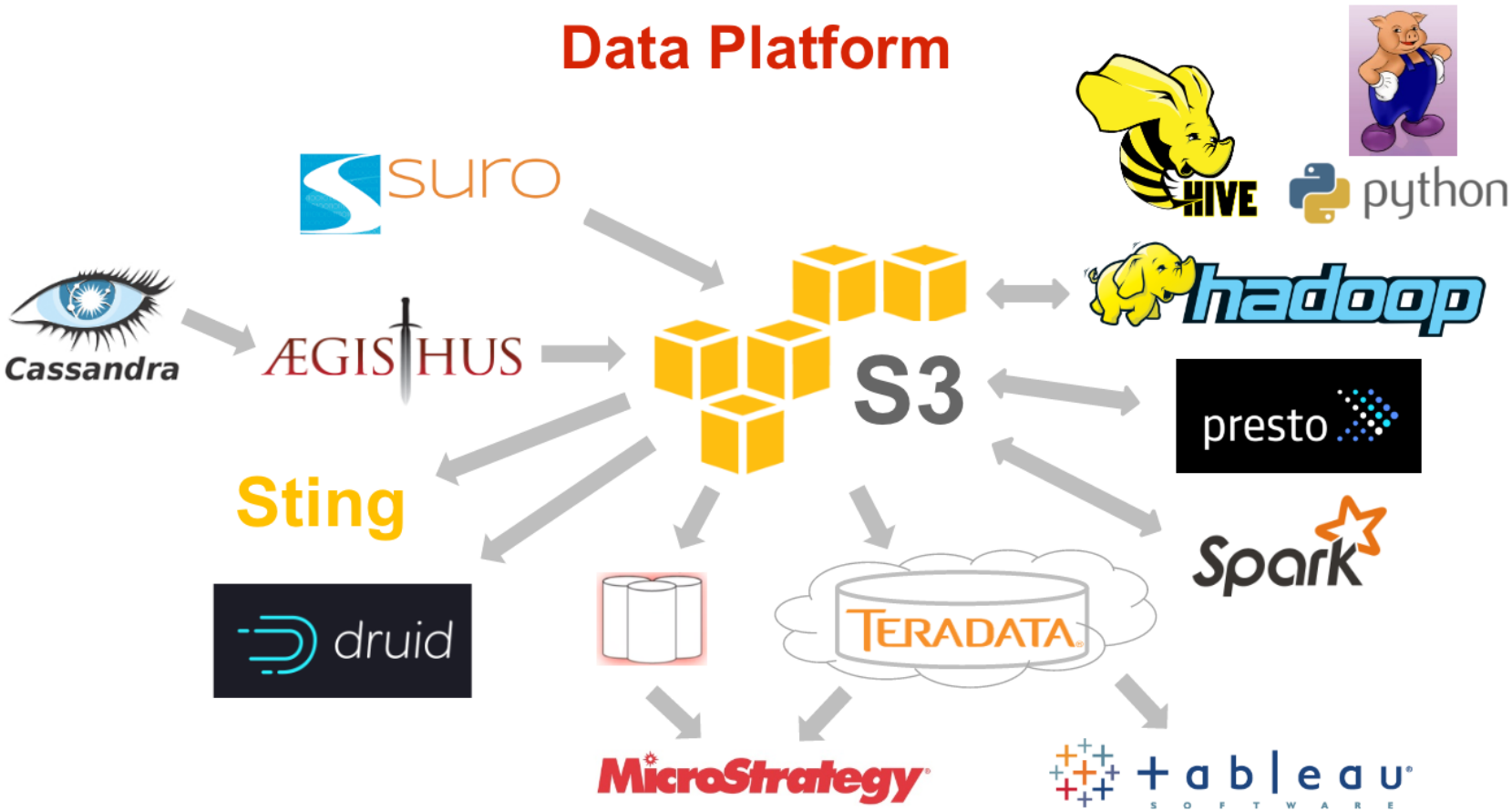
Example: Cloudera's Hadoop Stack



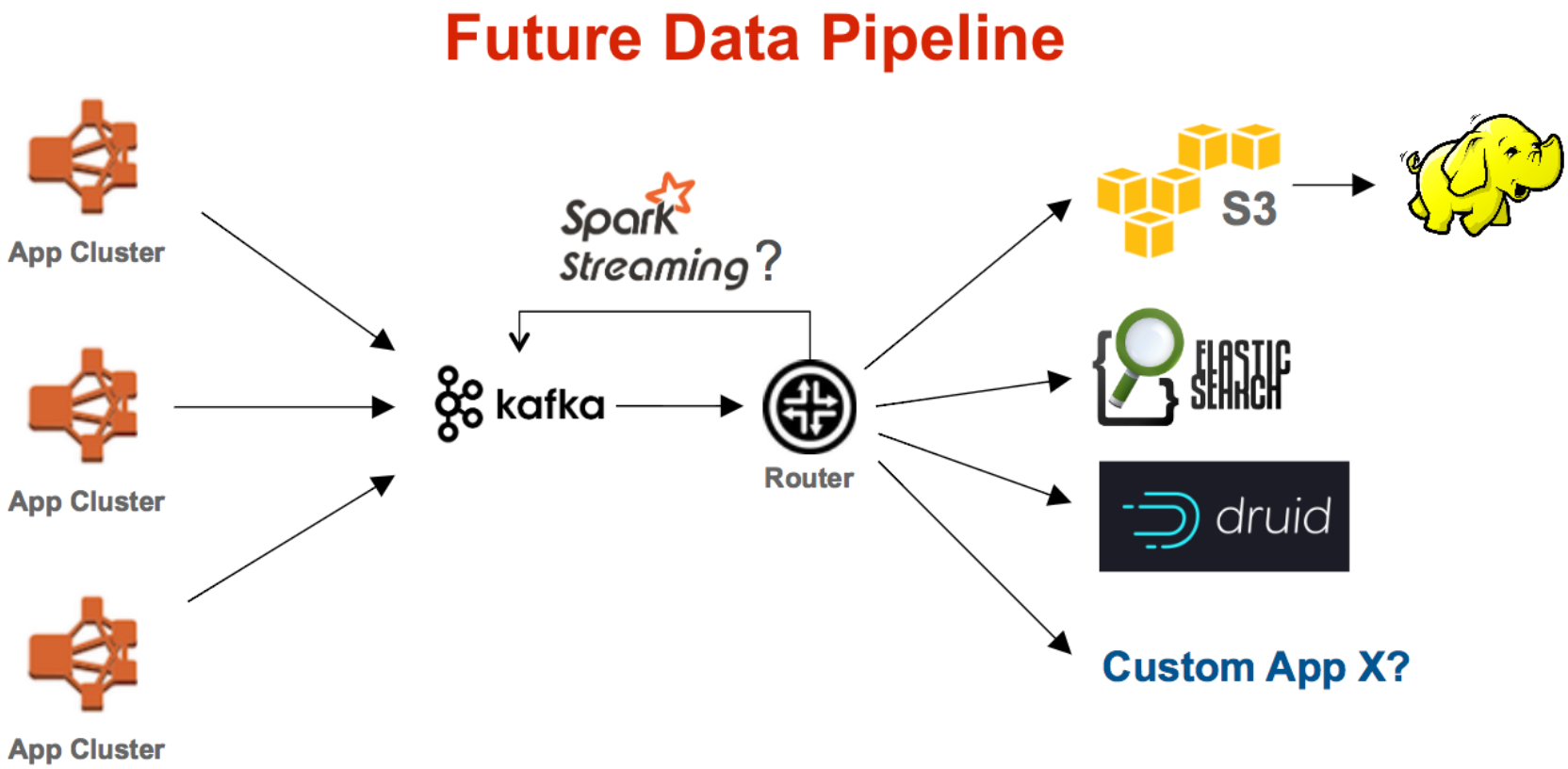
Example: Netflix Big Data Architecture

- Netflix is the largest Web based video service with more that 1/3 peak Internet traffic in USA
- Netflix has a 20 PB+ of data (in 2015) collected from all its operations
- Company policy is to only do data driven business decisions
- Data is used to:
  - Recommend films
  - Choose which content to purchase
  - Improve user interface through A/B testing
  - Do ad-hoc customer analytics, etc.
- See Kurt Brown: Big data at Netflix: Faster and easier
  `https://dl.dropboxusercontent.com/u/58977236/Netflix%20-%20Big%20Data%20-%20Faster%20and%20Easier%20-%20Strata%20NY%202015.pdf`

Netflix Big Data Architecture



Netflix Data Collection



Conclusions

- Artificial Intelligence applications require a Big Data backend for data collection and analytics
- Hadoop is becoming the "Linux distribution for Big Data", including also other components such as Apache Spark for main memory computing
- Hadoop consists of a number of interoperable open source components
- Commercial support is available from commercial vendors: Cloudera, HortonWorks, and MapR
- There is a move to hosted Big Data Applications: Billing is done on data volume being processed instead of number of computers used
- This allows special purpose hardware (e.g., GPGPUs and ASIC accelerators) to be used for improved energy efficiency