Lessons from Industry Standard Benchmarking

Raghunath Nambiar
Distinguished Engineer, Data Center Group, Cisco Systems, Inc

December 17-18, 2012 in Pune, India

SECOND WORKSHOP ON
BIG DATA BENCHMARKING

Center for Large-Scale Data Systems (CLDS)
San Diego Supercomputer Center, UC San Diego

Agenda

- Industry Standard Benchmarks

- A closer look the TPC

- Changing technology landscape

- Key considerations for Big Data benchmarks

Benchmarks

- Demands for them have existed since buyers were first confronted with the choice between purchasing one system over another

Historically we have seen that industry standard benchmarks enable healthy competition that results in product improvements and the evolution of brand new technologies.

Categories

- ## Industry standard benchmarks

  - Consortia driven development (Democratic process in decision making), Verifiable, Audit process, Product agnostic, Enables cross technology comparison

  TPC Transaction Processing Performance Council

  spec

  Storage Performance Council

- ## Application benchmarks

  - Vendor driven, vertical focused

  - Examples: VMMark, SAP Standard Application Benchmarks, Oracle Applications Benchmarks

- ## Synthetic Workloads

  - Open source based tools, widely adapted by engineering community

  - Examples: IOMeter, Netperf, Terasort

Viewpoints

- Vendor point of view

  Define the level playing field for competitive analysis

  Monitor release to release progress

  Product developments and enhancements

- Customer point of view

  Cross-vendor comparisons (performance, cost, power)

  Evaluate new technologies

- Researcher point of view

  Known, measurable and repeatable workloads

  Optimizations can impact products

Benchmarks and Technology Landscape

Driven by technology and industry demands

1985: Debit Credit Benchmark
1988: TPC and SPEC formed

1990s (Client server, Internet)

- Processor
- Transaction Processing
- File server, Web server
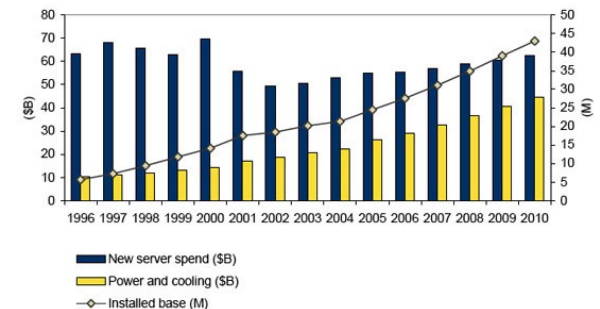- Data Warehouse

2000s (Web 2.0, Virtualization)

- Energy Efficiency
- Virtualization
- Complex Systems
- Cloud

2010s (Connected world, Analytics)

- Big Data
- Internet of things
- Software defined "everything"

Worldwide Cost to Power and Cool Server Installed Base, 1996–2010

Legend:
- New server spend ($B)
- Power and cooling ($B)
- Installed base (M)

Source: IDC, 2007

Did you know the number of virtual machine shipments now exceeds the number of physical server shipments ?

TPC Membership

Full Members

AMD	Bull	CISCO	DELL	FUJITSU
hp invent	HITACHI	HUAWEI	IBM	intel
Microsoft	NEC	ORACLE	redhat	SYBASE An SAP Company
TERADATA Raising Intelligence	UNISYS	vmware		

Associate Members

|				
IDEAS INTERNATIONAL	ITOM International Co.	SDSC SAN DIEGO SUPERCOMPUTER CENTER	TTA Telecommunications Technology Association	

Database centric

Represented by major systems and database companies

Independent audit process prior to publication

Mandatory performance, price-performance metric

Optional energy efficiency and virtualization metrics

TPC Timeline

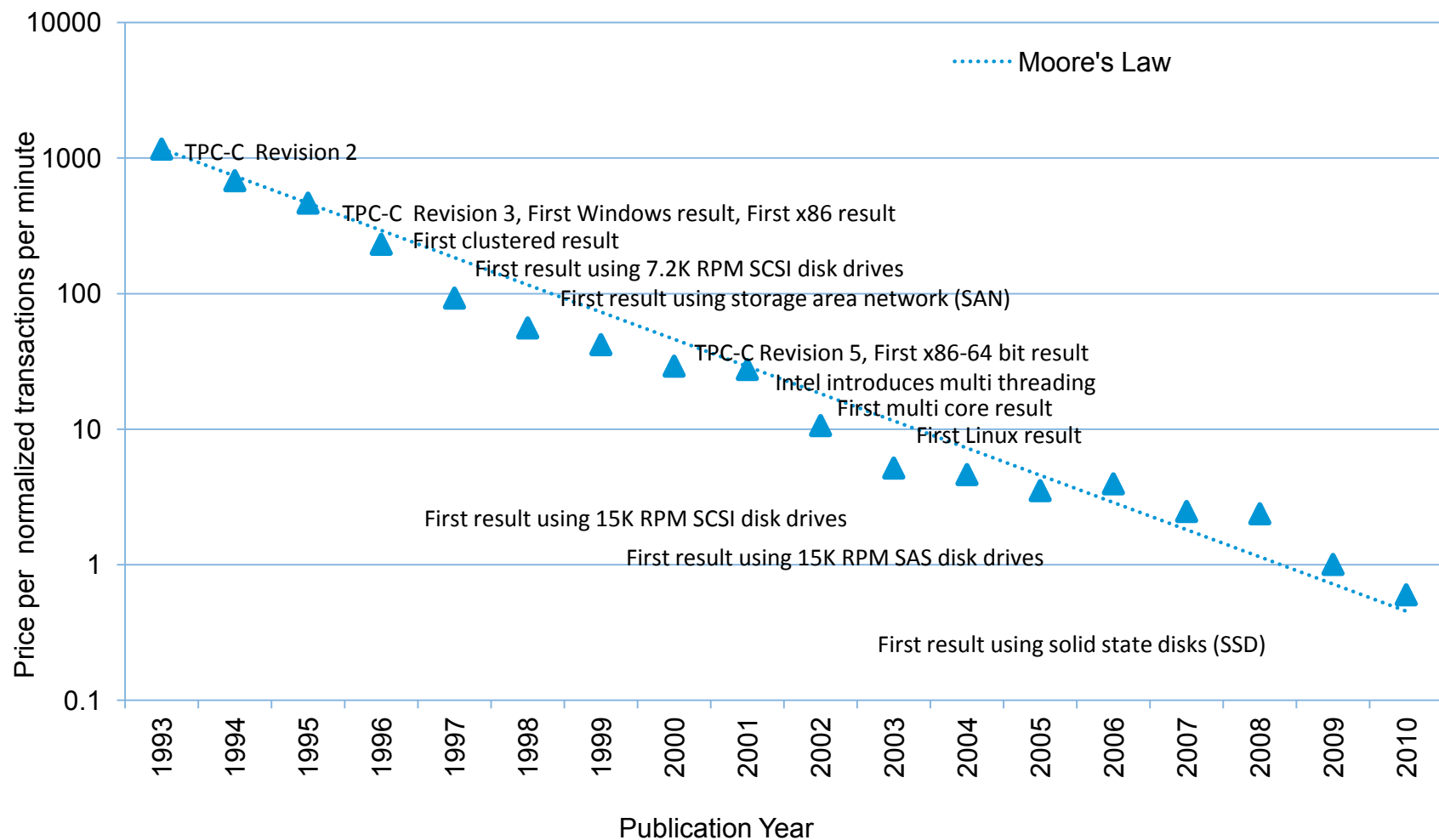| Benchmark Standards																								
TPC-A		■	■	■	■	■	■	■																
TPC-B			■	■	■	■	■	■																
TPC-C					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
TPC-D						■	■	■	■	■	■													
TPC-R											■	■	■	■	■	■	■							
TPC-H											■	■	■	■	■	■	■	■	■	■	■	■	■	■
TPC-W												■	■	■	■	■	■							
TPC-App																	■	■	■	■				
TPC-E																		■	■	■	■	■	■	■
TPC-DS																							■	■
TPC-VMS																								■
Common Specifications																								
Pricing																		■	■	■	■	■	■	■
Energy																					■	■	■	■
Developments in Progress																								
TPC-DI																					■	■	■	■
TPC-VMC																					■	■	■	■
TPC-V																					■	■	■	■
	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012

Legend:
- Obsolete
- Active
- Common Specifications
- In Progress

TPC-C Performance vs. Moore's Law
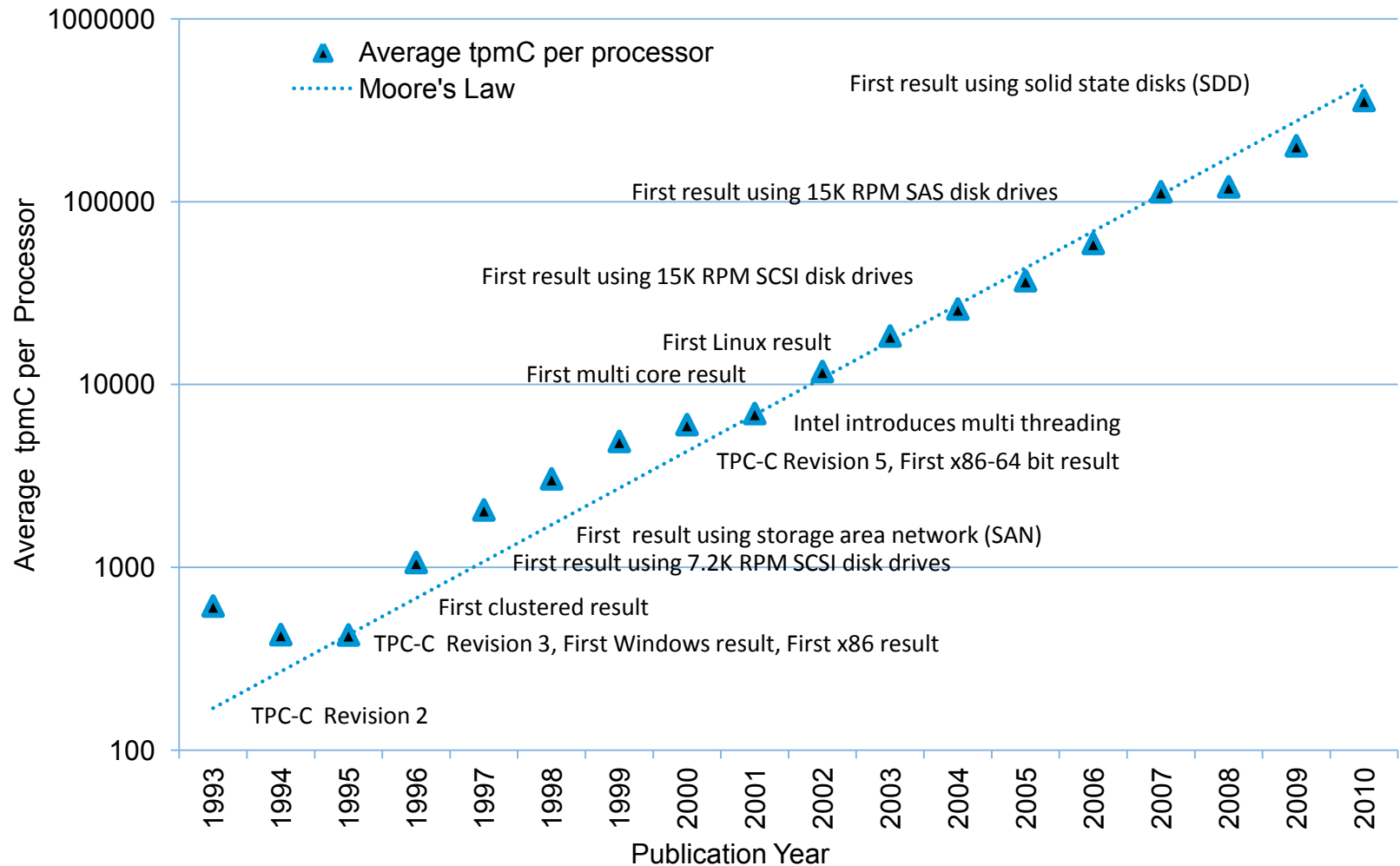
Reference: R. Nambiar, M. Poess, Transaction Performance vs. Moore's Law: A Trend Analysis: http://www.springerlink.com/content/fq6n225425151344/

TPC-C Price-Performance vs. Moore's Law

Price per normalized transactions per minute (y-axis, logarithmic: 0.1, 1, 10, 100, 1000, 10000) vs. Publication Year (x-axis: 1993–2010)

- Moore's Law
- TPC-C Revision 2
- TPC-C Revision 3, First Windows result, First x86 result
- First clustered result
- First result using 7.2K RPM SCSI disk drives
- First result using storage area network (SAN)
- TPC-C Revision 5, First x86-64 bit result
- Intel introduces multi threading
- First multi core result
- First Linux result
- First result using 15K RPM SCSI disk drives
- First result using 15K RPM SAS disk drives
- First result using solid state disks (SSD)

SPEC Membership

SPEC Members:

Acer Inc. * Action S.A. * Advanced Micro Devices * Apple Inc. * ASUSTeK Computer Inc. * Avere Systems * Bull S.A. * China Telecom Corporation Limited (Beijing Research Institute) * Cisco Systems, Inc. * Dell, Inc. * E4 Computer Engineering SPA * EMC * Fujitsu * Hitachi Data Systems * Hitachi Ltd. * HP * Huawei Technologies Co. Ltd. * IBM * IDEAS International * Intel * Itautec S/A * Lenovo * Microsoft * NEC - Japan * NetApp * NVIDIA * Oracle * Platform Computing Inc. * Principled Technologies * The Portland Group * Red Hat * Samsung * SAP AG * SGI * Super Micro Computer, Inc. * SUSE * Symantec Corporation * Unisys * Via Technologies * VMware * Voxeo Corporation *

SPEC Associates:

Acadamia Sinica, Institute of Information Science * Austrian Energy Agency * Argonne National Laboratory * Charles University * Citrix * Department of Veteran's Affairs - Corporate Data Center Operations * Dresden University of Technology ZIH * Indiana University * Institute for Information Industry Taiwan * JAIST * Karlsruhe Institute of Technology * Leibniz Rechenzentrum - Germany * National University of Singapore * Ohio State University * Purdue University * Technische Universität Darmstadt * University of Aizu - Japan * University of California - Berkeley * University of Cologne * University of Illinois at Urbana-Champaign * University of Maryland * University of Miami * University of Pavia * University of Texas at Austin * University of Tsukuba * University of Virginia * Virginia Polytechnic Institute and State University * Zhejiang University *

SPEC Research Group:

AMD * Charles University * Cisco Systems * Dell * Ghent University * Hasso Plattner Institute * Hewlett-Packard * IBM * Imperial College London * Indian Institute of Technology, Bombay * Institute for Information Industry, Taiwan * Intel * Karlsruhe Institute of Technology * McGill University * NICTA * Oracle * Purdue University * Red Hat * SalesForce.com * San Francisco State University * SAP * Siemens Corporation * Technische Universität Darmstadt * University of Alberta * University of California at Berkeley * University of Coimbra * University of Kiel * University of Mannheim * University of Minnesota * University of Pavia * University of Texas at Austin * VMware * Zhejiang University *

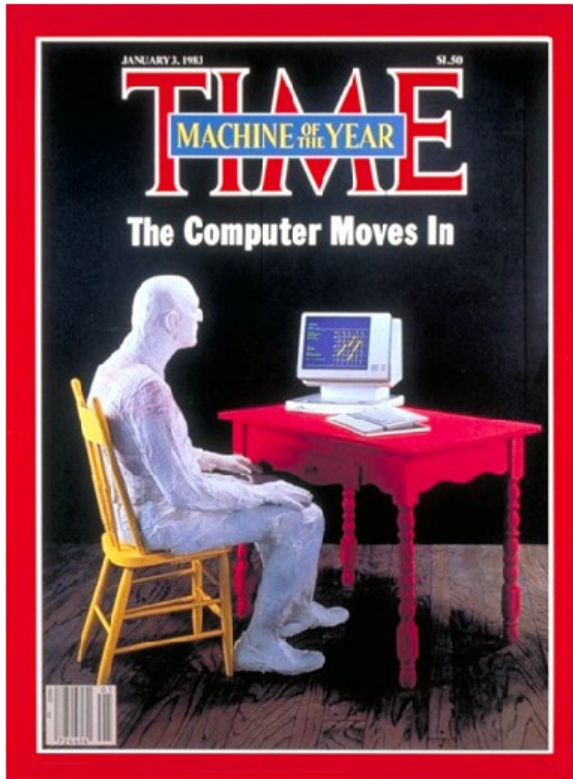System centric
Broad representation from industry and research
Peer audit process prior to publication
Primary metric is: performance
Separate benchmark for energy efficiency and virtualization

Technology Landscape is Changing

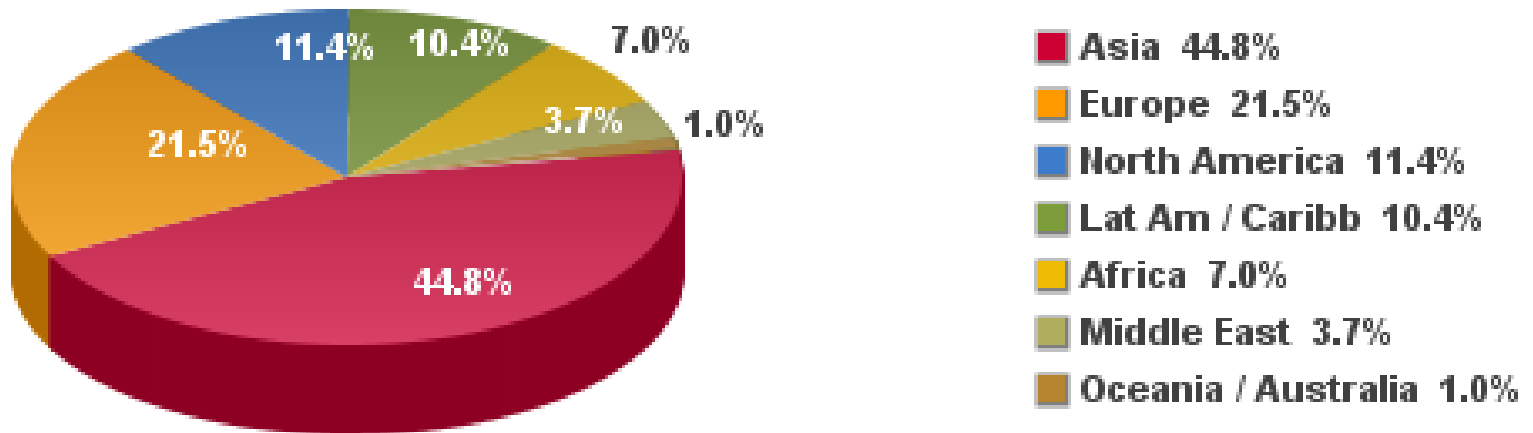the winner of Time's man of the year award for 1982 …

the winner of Time's man of the year award for 2006 …

Internet Users in the World
Distribution by World Regions - 2012 Q2

- Asia 44.8%
- Europe 21.5%
- North America 11.4%
- Lat Am / Caribb 10.4%
- Africa 7.0%
- Middle East 3.7%
- Oceania / Australia 1.0%

Source: Internet World Stats - www.internetworldstats.com/stats.htm
Basis: 2,405,518,376 Internet users on June 30, 2012
Copyright © 2012, Miniwatts Marketing Group

2,405,518,376/7,017,846,922 =34.3 %

it took …

53 years to sell 100 million radios

14 years to sell 100 million TVs

but .. in 15 years …

there are 15 billion devices connected to the Internet
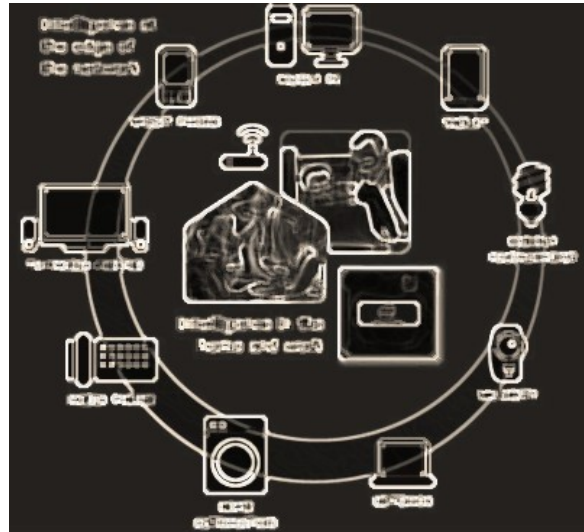
that's 2.2 devices for every man, woman, and child on the planet earth

if

facebook

were a country …

1. China (1.339 billion)

2. India (1.218 billion)

3. Facebook (900 million)

4. United States (311 million)

5. Indonesia (237 million)

6. Brazil (190 billion)

7. Pakistan (175 million)

8. Nigeria (158 million)

9. Bangladesh (150 million)

10. Russia (142 million)

Processing Movie AVATAR

4352 servers
34,816 processor-cores
102 terabytes of RAM
3 Petabytes of storage

Comprehensive statistics on more than 250 countries and entities

Unknown number of classified information
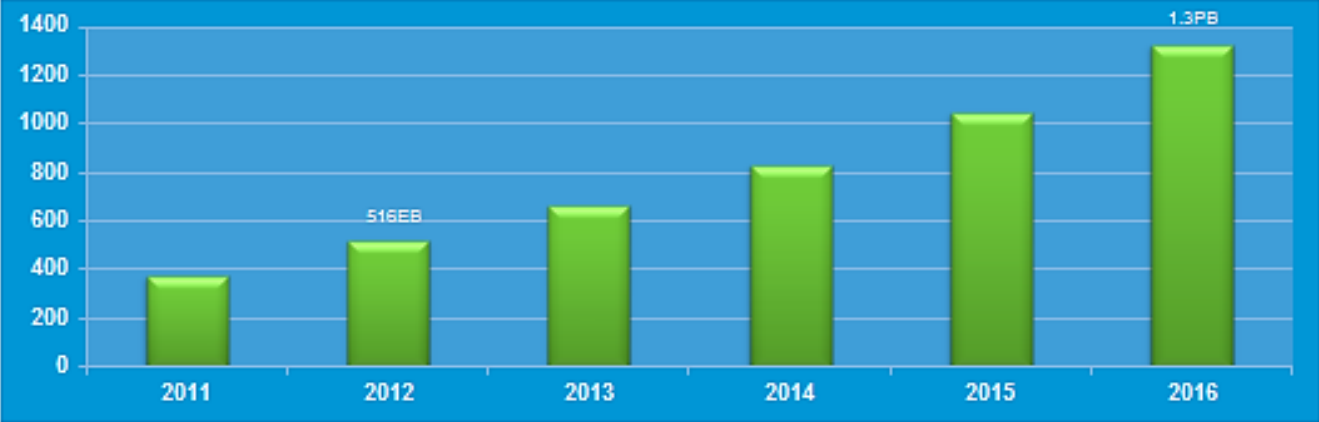
How Big is the Digital Universe ?

2008
0.5 Zettabyte

2011
2.5 Zettabytes

2020
35 Zettabytes
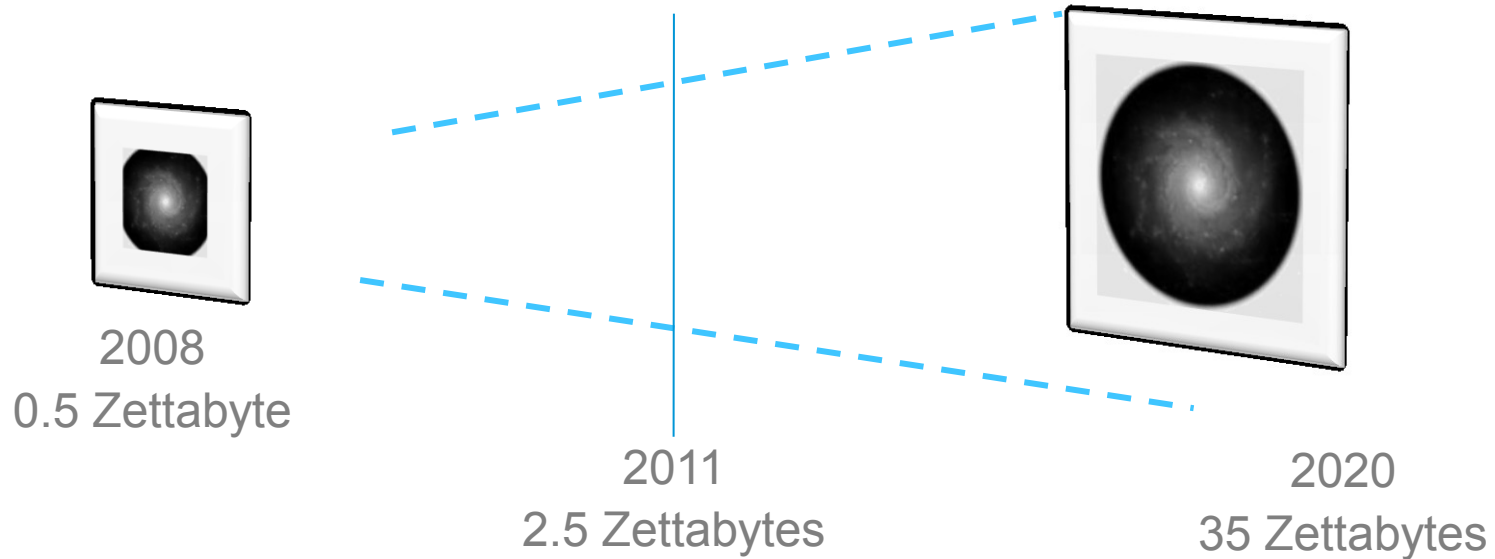
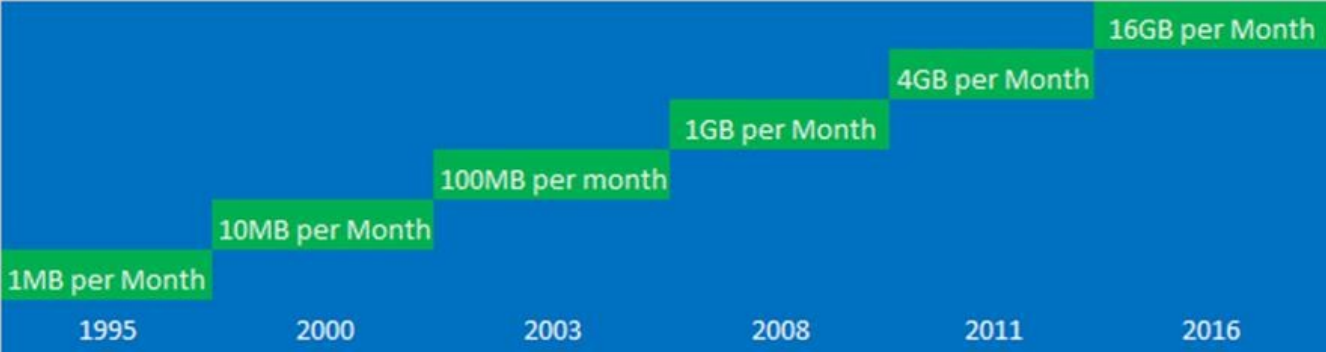| 1 Zettabyte	= 1 099 511 627 776 Gigabytes
	= 1 Billion **1TB** Disk Drives

How many disk drives were sold in 2011 ?

Global IP Traffic

Global IP Traffic

Per Capita Internet Traffic

In 2016, equivalent of all movies ever made will cross global IP networks every 3 minutes

Source: Cisco

Benchmark Acceleration Initiatives

- TPC

  Technology conference initiative on performance evaluation and benchmarking

  - 2009 (Lyon), 2010 (Singapore), 2011 (Seattle), 2012 (Istanbul)
  - 2013 (Trento) – planned

- SPEC

 - SPEC Research
  - ICPE

- WBDB

- First important step towards the development of a set of benchmarks for providing objective measures of the effectiveness of hardware and software systems dealing with big data applications.

What is Important ?

- Performance

- Cost of ownership

- Energy efficiency

- Floor space efficiency

- Manageability

- User experience

Successful Benchmark Requirements

- Relevant

- Repeatable

- Understandable

- Fair

- Verifiable

- Economical
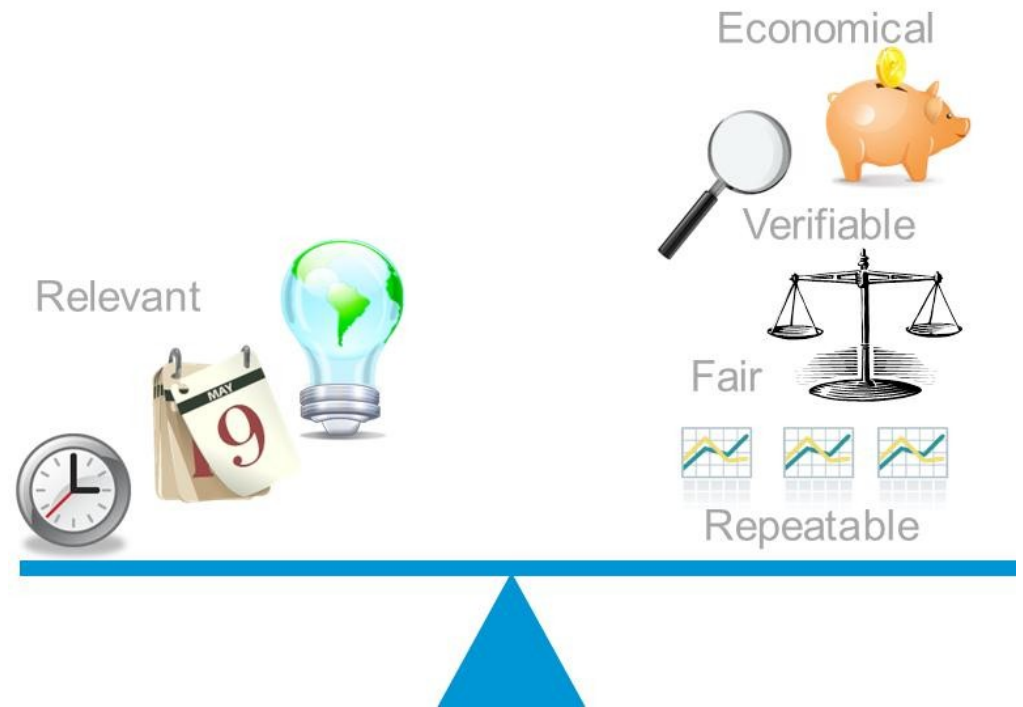
Reference: K. Huppler, The Art of Building a Good Benchmark, Performance Evaluation and Benchmarking, LNCS vol. 5895, Springer 2009

WBDB

- WBDB 2012

http://clds.ucsd.edu/wbdb2012

Chaitanya Baru, Milind Bhandarkar, Raghunath Nambiar,Meikel Poess, Tilmann Rabl: Setting the Direction for Big Data Benchmark Standards, TPCTC 2012, LNCS Vol. 7755, Springer 2012

- WBDB 2012.in

http://blogs.cisco.com/datacenter/wbdb2012-in/

- WBDB 2013

http://clds.ucsd.edu/wbdb2013.cn

- Bigdata 100 List

Thank you.