



Towards an Industry Standard for Performance Evaluation and Benchmarking Big Data Workloads

Raghunath Nambiar

Strategist, Performance and Solution Engineering
Data Center Group, Cisco Systems, Inc

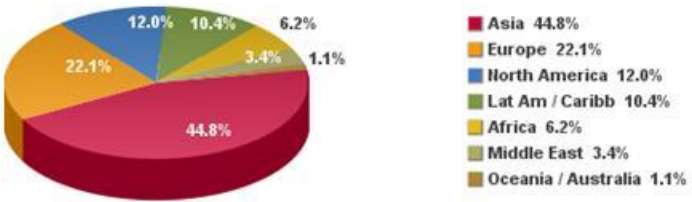
May 8-9, 2012 | San Jose, CA

WBDB2012

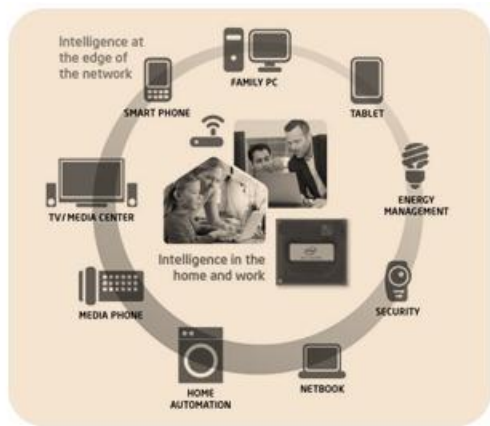
Workshop on
Big Data Benchmarking

Big Data Landscape

**Internet Users in the World
Distribution by World Regions - 2011**



Source: Internet World Stats - www.internetworldstats.com/stats.htm
 Basis: 2,267,233,742 Internet users on December 31, 2011
 Copyright © 2012, Miniwatts Marketing Group

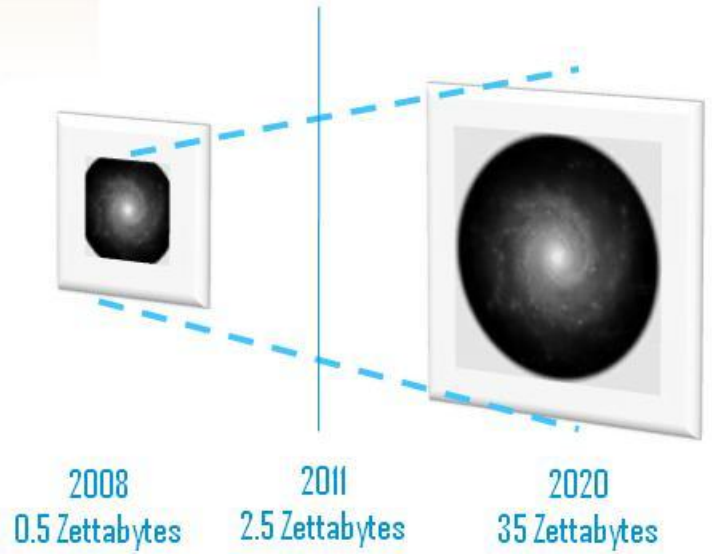


There are 15 billion devices connected to the Internet
 That's 2.2 devices for every man, woman, and child on the planet earth



was a country, it would be the 3rd largest in the world

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)



Big Data in Enterprises

Almost every business is conducted over internet

Business generate more data,
Store more data,
Store them for longer period,
often required due to compliance

More data will improves predictive analytics

Sales Products Process Inventory
Finance Payroll Shipping Tracking
Authorization Customers Profile

Machine logs Sensor data
Call data records Web click stream data
Satellite feeds GPS data Sales data Blogs
Emails Pictures Video

Structured
Semi-structured
Un-structured

Industry Standard Benchmark Landscape

- Industry standard benchmarks

Transaction Processing Performance Council (TPC)

Standard Performance Evaluation Corporation (SPEC)

Storage Performance Council (SPC)

- Application benchmarks

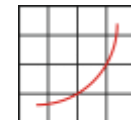
VMWare VMMark

SAP Standard Application Benchmarks

Oracle Applications Benchmark

TPC Transaction Processing
Performance Council

Storage Performance Council



spec

Why Benchmarks are Important ?

- Vendor point of view
 - Define the playing field (measurable, repeatable)
 - Enable competitive analysis
 - Monitor release to release progress
 - Result understood by engineering, sales and customers
 - Accelerate focused technology development
- Customer point of view
 - Cross-vendor comparisons (performance, cost, energy)
 - Evaluate new technologies
 - Eliminate costly in-house characterization

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

Industry Standard Benchmarks

- Broad Industry representation**
(all decision taken by the board)
- Verifiable (audit process)**
- Domain specific standard tests**
- Resolution of disputes and challenges**

- Performance
- Cost of Ownership
- Energy Efficiency
- Floor Space Efficiency
- Manageability

Towards an Industry Standard for Performance Evaluation and Benchmarking Big Data Workloads



⋮

- **Open Source Model**
- 3 Major Components – Dataset, Workload, Metric
- Experts are welcome to join development
- Industry standard committees are allowed to pick “a version” to standardize (audit, pricing etc) the benchmark
- The open source project will continue

*4th International Conference on Performance Evaluation Benchmarking

Survey

1. Is there an eminent need for benchmark(s) in the Big Data Space ? [Y/N]
2. What are the top 5 Big Data workloads used in your organization ?
3. Focus for next 1 year - considering the limited resources [choose one]
 - Technology Benchmark ?
 - Competitive Benchmark ?
4. What would you like to measure ? [choose one]
 - Hardware (server, storage, network)
 - Software (operating system, processing engine)
 - Application (how fast a common set of problems can be solved)
5. How many of you would join an industry standard committee to develop a benchmark standard (2-3 year commitment of 1-2 fulltime people): [Y,N, Unknown]
6. How many of you would join an informal group to develop a benchmark workload (1-2 year commitment): [Y,N, Unknown]

TPCTC 2012

4th International Conference on Performance Evaluation and Benchmarking

This year's focus will be “Benchmarking Big Data Workloads”

More information: www.tpc.org/tpctc2012

Thank you.

