IaaS Cloud Benchmarking: Approaches, Challenges, and Experience

Alexandru Iosup

Parallel and Distributed Systems Group
Delft University of Technology
The Netherlands

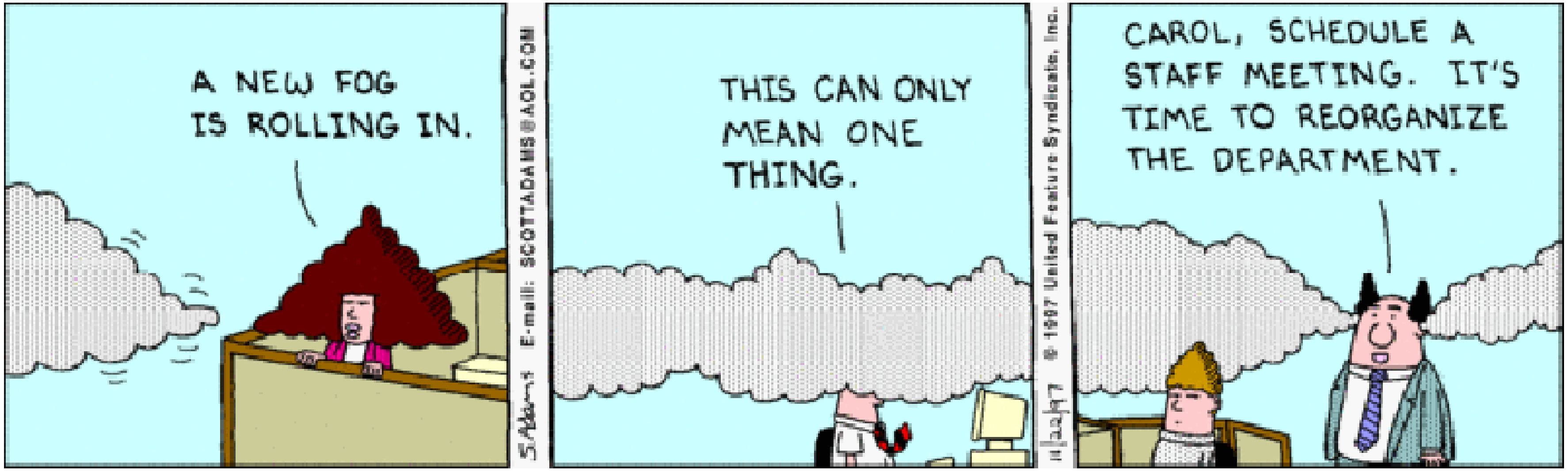
Our team: **Undergrad** Nassos Antoniou, Thomas de Ruiter, Ruben Verboon, … **Grad** Siqi Shen, Nezih Yigitbasi, Ozan Sonmez **Staff** Henk Sips, Dick Epema, Alexandru Iosup **Collaborators** Ion Stoica and the Mesos team (UC Berkeley), Thomas Fahringer, Radu Prodan (U. Innsbruck), Nicolae Tapus, Mihaela Balint, Vlad Posea (UPB), Derrick Kondo, Emmanuel Jeannot (INRIA), Assaf Schuster, Mark Silberstein, Orna Ben-Yehuda (Technion), ...

November 12, 2012

MTAGS, SC'12, Salt Lake City, UT, USA

What is Cloud Computing?
1. A Cloudy Buzzword

- 18 definitions in computer science (ECIS'10). NIST has one. Cal has one. We have one.
- "We have redefined cloud computing to include everything that we already do." Larry Ellison, Oracle, 2009


A NEW FOG IS ROLLING IN.
THIS CAN ONLY MEAN ONE THING.
CAROL, SCHEDULE A STAFF MEETING. IT'S TIME TO REORGANIZE THE DEPARTMENT.
S.Adams
E-mail: SCOTTADAMS@AOL.COM
© 1997 United Feature Syndicate, Inc.
11/22/97

Source: http://dilbert.com/strips/comic/1997-11-22/

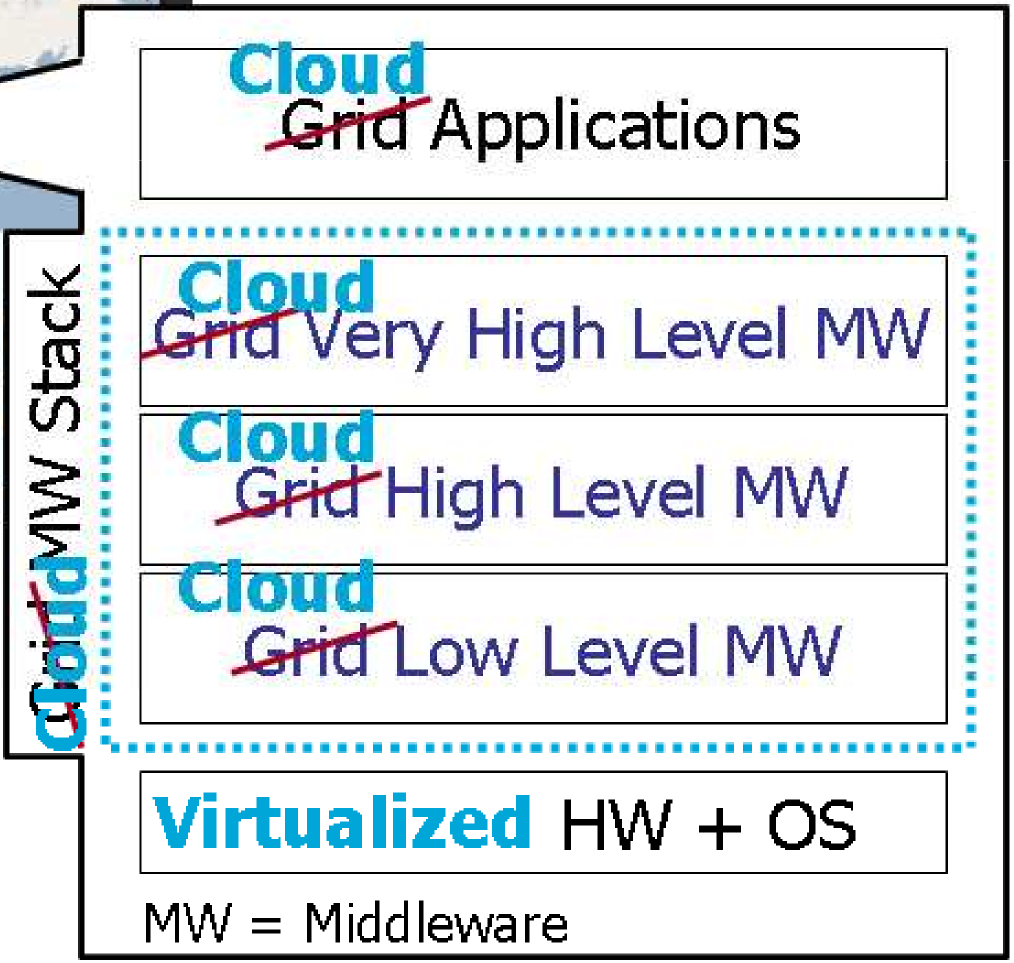
What is Cloud Computing?
2. A Descendant* of the Grid Idea

* Subset.

Asia
Europe
Atlantic Ocean
Africa
South America
Pacific Ocean
Indian Ocean
Austral
7
4
14
15
17
31
32
33
35
36

Source: http://royal.pingdom.com/2008/04/11/map-of-all-google-data-center-locations/

"A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities [+ for] nontrivial QoS." I. Foster, 1998 + 1999



~~Grid~~ Cloud MW Stack

Cloud ~~Grid~~ Applications

Cloud ~~Grid~~ Very High Level MW

Cloud ~~Grid~~ High Level MW

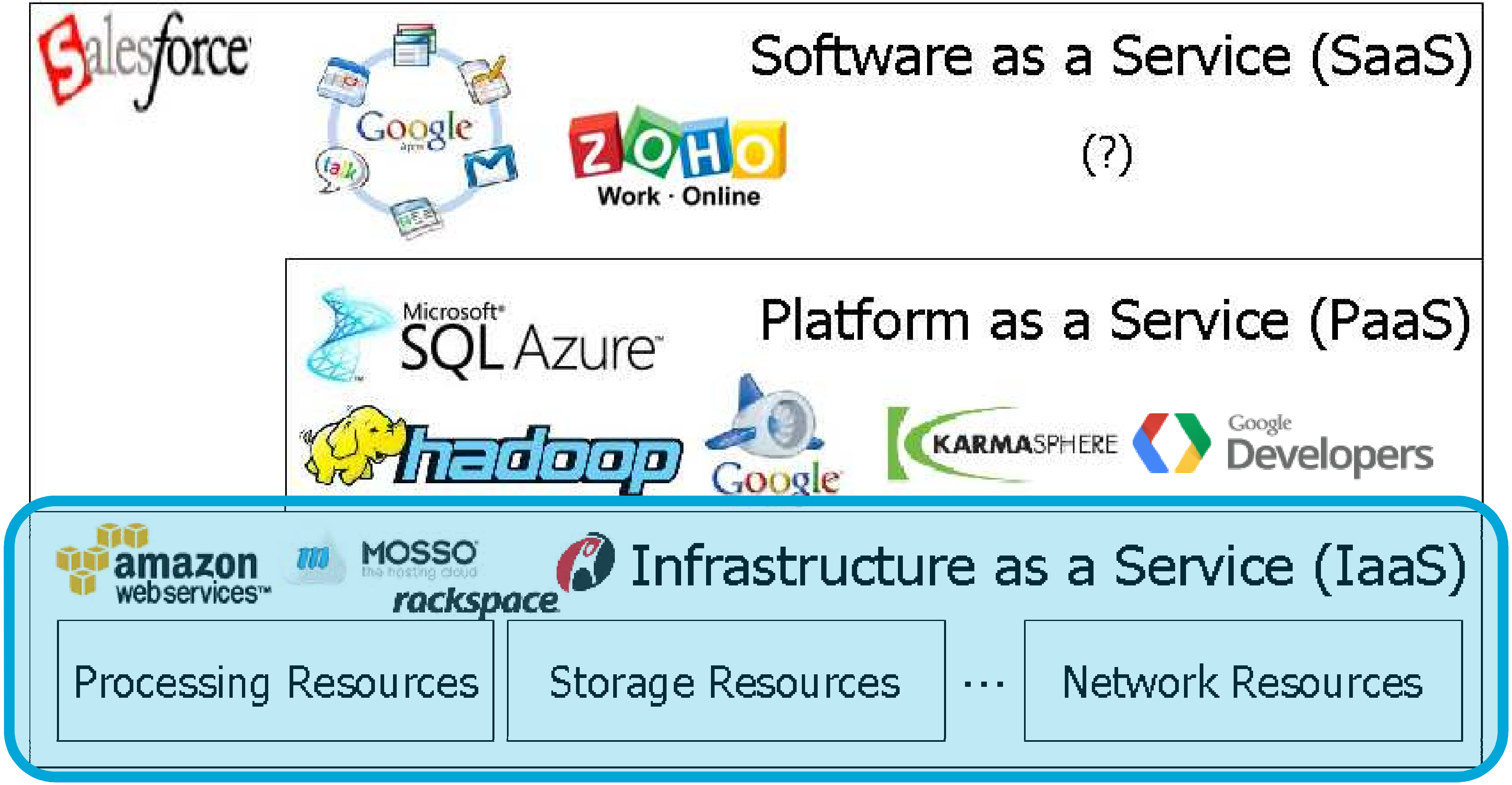
Cloud ~~Grid~~ Low Level MW

Virtualized HW + OS

MW = Middleware

What is Cloud Computing?
3. A Useful IT Service

"Use only when you want! Pay only for what you use!"



Software as a Service (SaaS)

(?)

Platform as a Service (PaaS)

Infrastructure as a Service (IaaS)

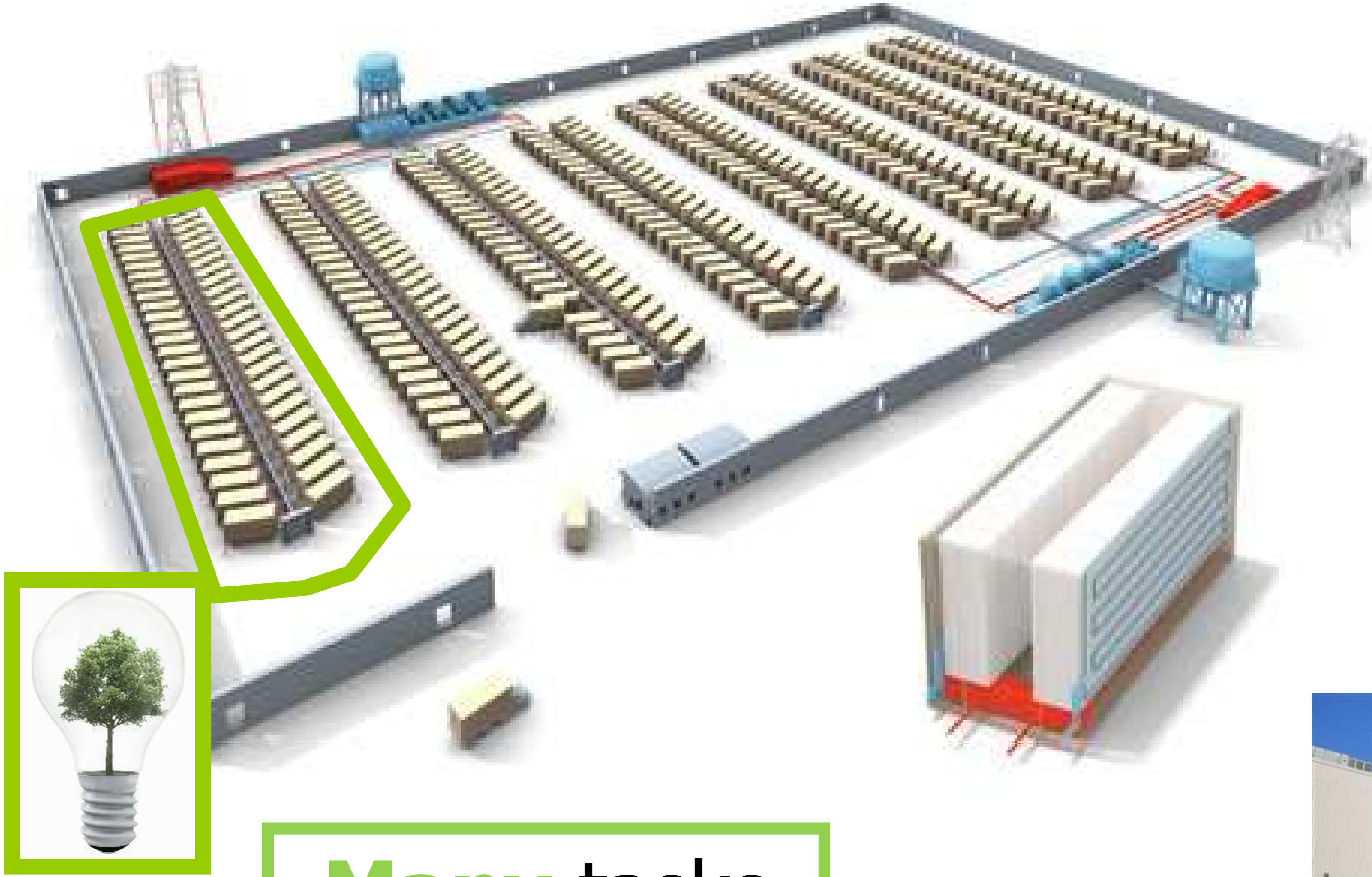
Processing Resources

Storage Resources

…

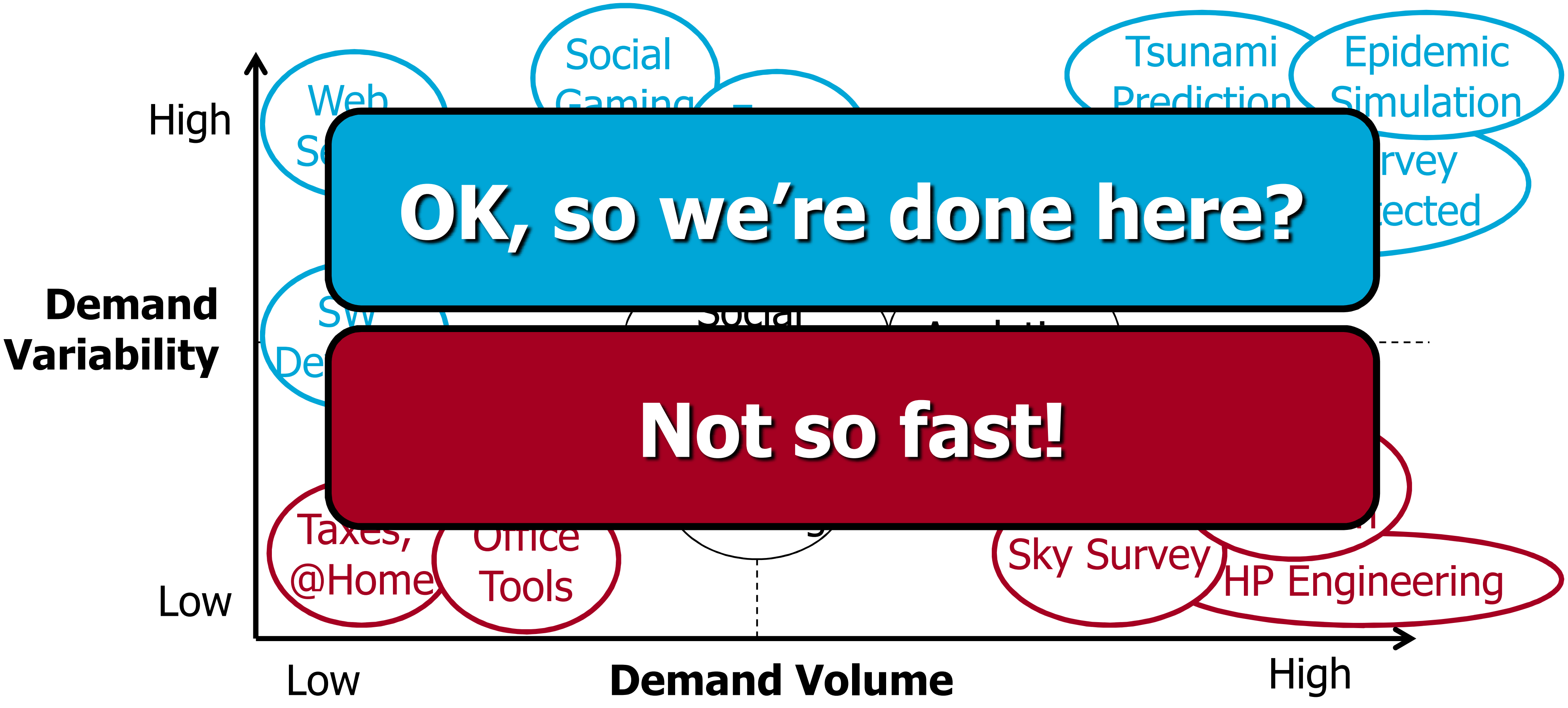
Network Resources

IaaS Cloud Computing



Many tasks

Which Applications Need Cloud Computing? A Simplistic View…



OK, so we're done here?

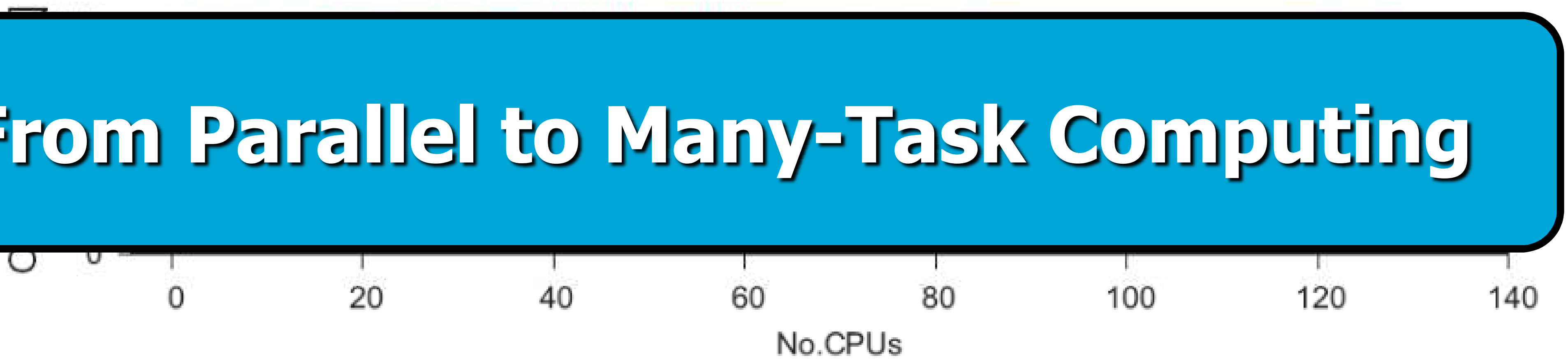
Not so fast!

November 12, 2012

After an idea by Helmut Krcmar

What I Learned from Grids

- Average job size is 1 (that is, there are **no [!] tightly-coupled, only conveniently parallel jobs**)



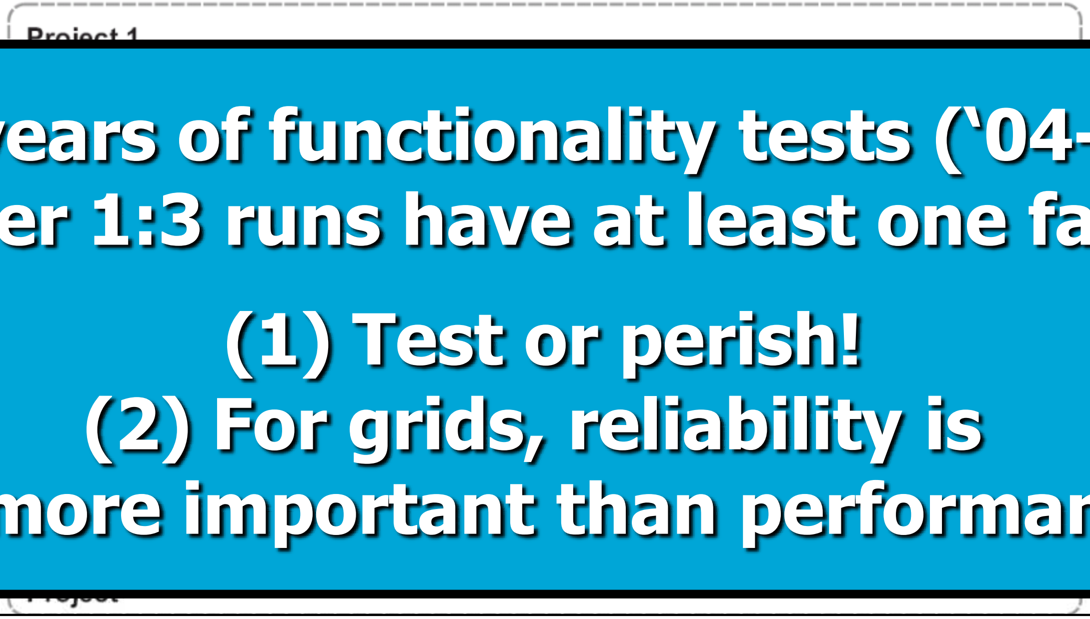
From Parallel to Many-Task Computing

A. Iosup, C. Dumitrescu, D.H.J. Epema, H. Li, L. Wolters, How are Real Grids Used? The Analysis of Four Grid Traces and Its Implications, Grid 2006.

A. Iosup and D.H.J. Epema, Grid Computing Workloads, IEEE Internet Computing 15(2): 19-26 (2011)

What I Learned from Grids?

- NMI Build-and-Test Environment at U.Wisc.-Madison: 112 hosts, >40 platforms (e.g., X86-32/Solaris/5, X86-64/RH/9)
- Serves >50 **grid middleware packages**: Condor, Globus, VDT, gLite, GridFTP, RLS, NWS, INCA(-2), APST, NINF-G, BOINC …



Two years of functionality tests ('04-'06): over 1:3 runs have at least one failure!

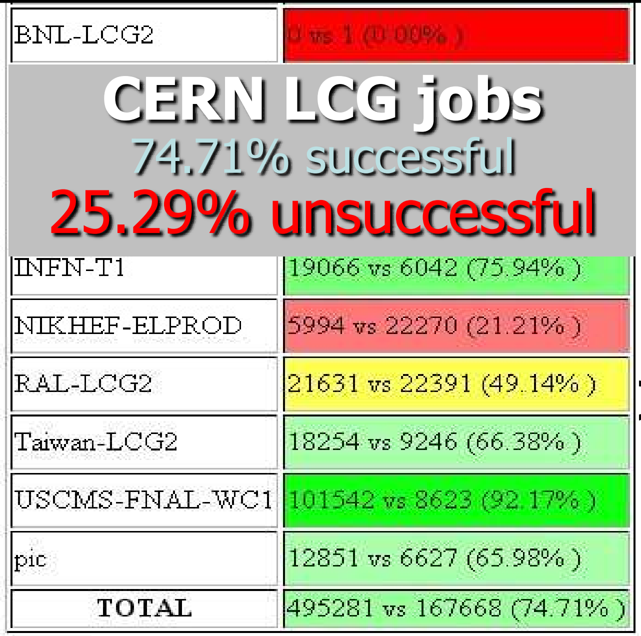
(1) Test or perish!
(2) For grids, reliability is more important than performance!

A. Iosup, D.H.J.Epema, P. Couvares, A. Karp, M. Livny, Build-and-Test Workloads for Grid Middleware: Problem, Analysis, and Applications, CCGrid, 2007.

What I Learned from Grids

- Server
- 99.99999% reliable

Grids are unreliable infrastructure



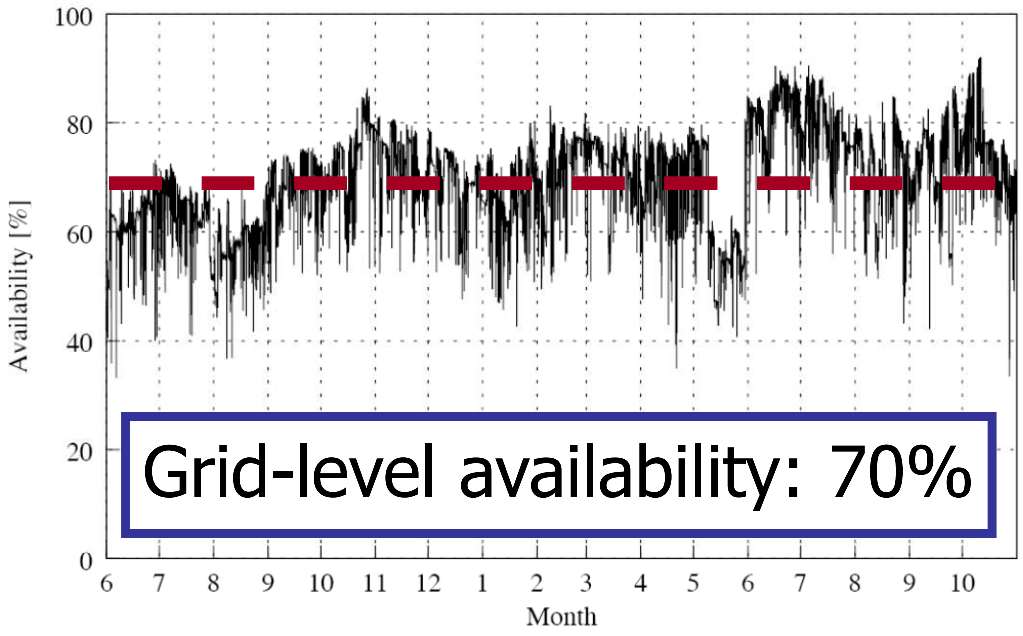
CERN LCG jobs

74.71% successful

25.29% unsuccessful

BNL-LCG2	0 vs 1 (0.00%)
INFN-T1	19066 vs 6042 (75.94%)
NIKHEF-ELPROD	5994 vs 22270 (21.21%)
RAL-LCG2	21631 vs 22391 (49.14%)
Taiwan-LCG2	18254 vs 9246 (66.38%)
USCMS-FNAL-WC1	101542 vs 8623 (92.17%)
pic	12851 vs 6627 (65.98%)
TOTAL	495281 vs 167668 (74.71%)

Source: dboard-gr.cern.ch, May'07.



Grid-level availability: 70%

n,

06]

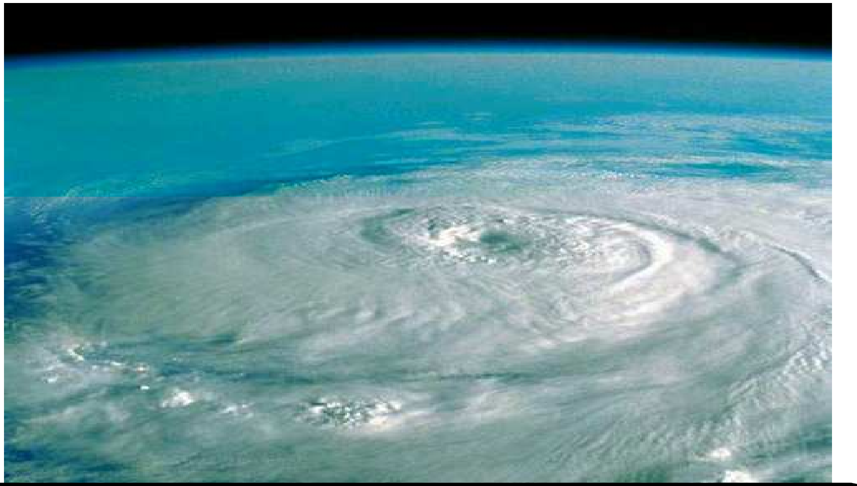
)6]

ı et

A. Iosup, M. Jan, O. Sonmez, and D.H.J. Epema, On the Dynamic Resource Availability in Grids, Grid 2007, Sep 2007.

What I Learned From Grids, Applied to IaaS Clouds

or



We just don't know!

- "The path to abundance"
- On-demand capacity
- Cheap for short-term tasks
- Great for web apps (EIP, web crawl, DB ops, I/O)

- "The killer cyclone"
- Performance for scientific applications (compute- or data-intensive)
- Failures, Many-tasks, etc.

This Presentation: Research Questions

Q0: What are the workloads of IaaS clouds?

Q1: What is the performance of production IaaS cloud services?

Q2: How variable is the performance of widely used production cloud services?

Q3: How do provisioning and allocation policies affect the performance of IaaS cloud services?

Other questions studied at TU Delft: How does virtualization affect the performance of I

But … This is benchmarking = process of quantifying the performance and other non-functional properties of the system

Delft University of Technology

Why IaaS Cloud Benchmarking?

- Establish and share best-practices in answering important questions about IaaS clouds

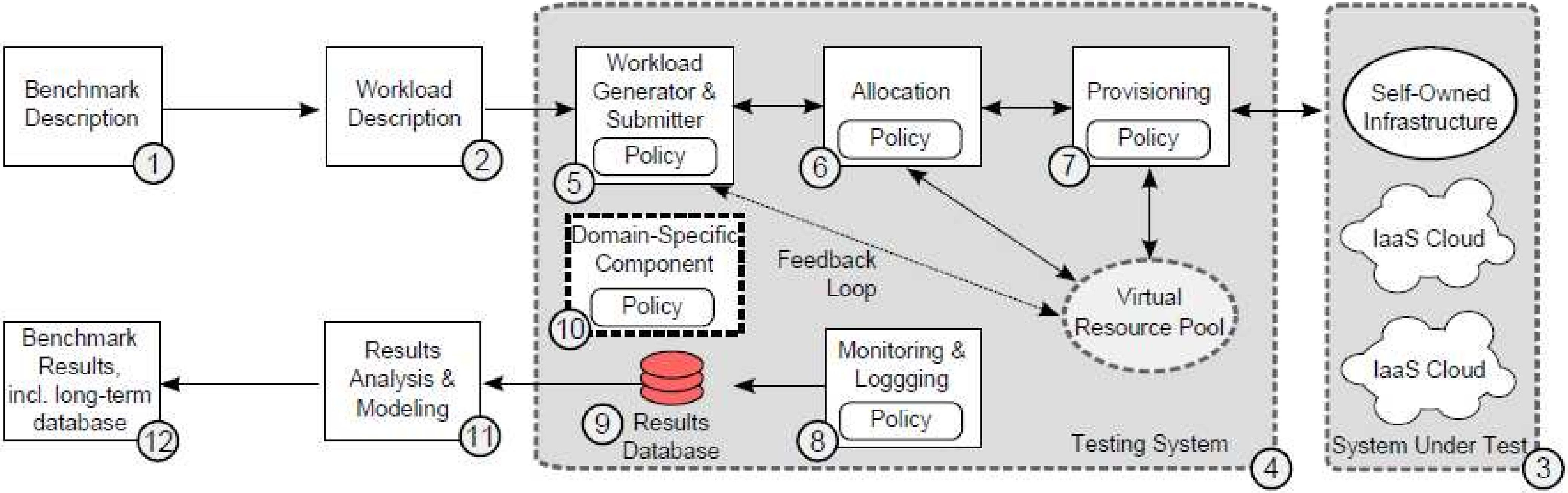
- Use in procurement
- Use in system design
- Use in system tuning and operation
- Use in performance management

- **Use in training**

Agenda

1. An Introduction to IaaS Cloud Computing
2. Research Questions or Why We Need Benchmarking?
3. **A General Approach and Its Main Challenges**
4. IaaS Cloud Workloads (Q0)
5. IaaS Cloud Performance (Q1) and Perf. Variability (Q2)
6. Provisioning and Allocation Policies for IaaS Clouds (Q3)
7. Conclusion

A General Approach for IaaS Cloud Benchmarking



Approach: Real Traces, Models, Real Tools, Real-World Experimentation (+ Simulation)

- Formalize real-world scenarios
- Exchange real traces
- Model relevant operational elements
- Scalable tools for meaningful and repeatable experiments
- Comparative studies
- Simulation only when needed (long-term scenarios, etc.)

Rule of thumb:

Put 10-15% project effort into benchmarking

10 Main Challenges in 4 Categories*

* List not exhaustive

- **Methodological**
  1. Experiment compression
  2. Beyond black-box testing through testing short-term dynamics and long-term evolution
  3. Impact of middleware

- **System-Related**
  1. Reliability, availability, and system-related properties
  2. Massive-scale, multi-site benchmarking
  3. Performance isolation

- **Workload-related**
  1. Statistical workload models
  2. Benchmarking performance isolation under various multi-tenancy models

- **Metric-Related**
  1. Beyond traditional performance: variability, elasticity, etc.
  2. Closer integration with cost models

Iosup, Prodan, and Epema, IaaS Cloud Benchmarking: Approaches, Challenges, and Experience, MTAGS 2012. (invited paper)

Read our article

Agenda

1. An Introduction to IaaS Cloud Comput
2. Research Questions or Why We Need B
3. A General Approach and Its Main Challe
4. **IaaS Cloud Workloads (Q0)**
5. **IaaS Cloud Performance (Q1) & Perf. Variability (Q2)**
6. **Provisioning & Allocation Policies for IaaS Clouds (Q3)**
7. Conclusion

IaaS Cloud Workloads: Our Team

Alexandru Iosup
TU Delft

BoTs
Workflows
Big Data
Statistical modeling

Dick Epema
TU Delft

BoTs
Grids

Mathieu Jan
TU Delft/INRIA

BoTs
Statistical modeling

Ozan Sonmez
TU Delft

BoTs

Thomas de Ruiter
TU Delft

MapReduce
Big Data
Statistical modeling

Radu Prodan
U.Isbk.

Workflows

Thomas Fahringer
U.Isbk.

Workflows

Simon Ostermann
U.Isbk.

Workflows

What I’ll Talk About

IaaS Cloud Workloads (Q0)

1. BoTs
2. Workflows
3. Big Data Programming Models
4. MapReduce workloads

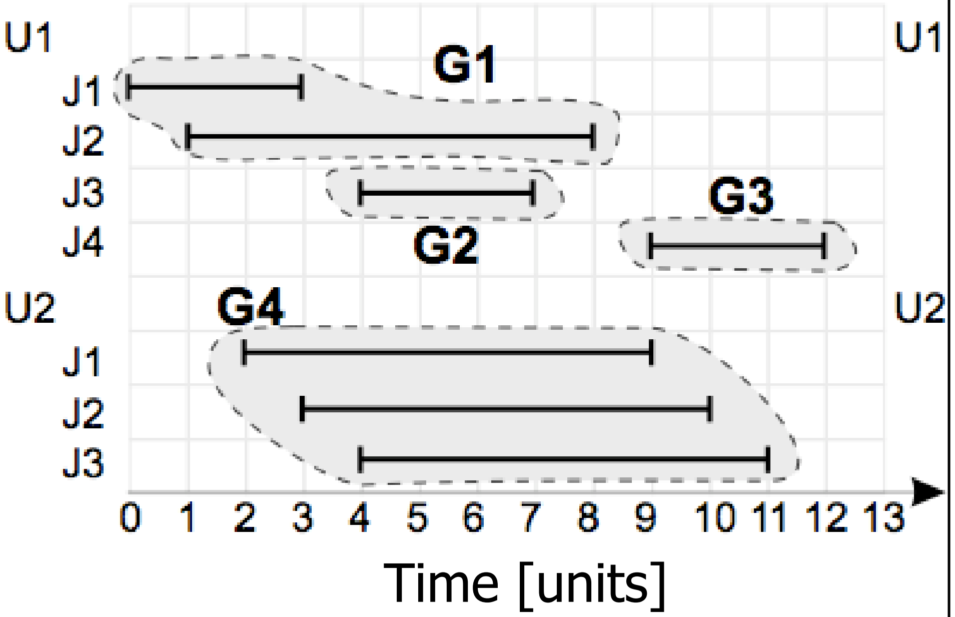
What is a Bag of Tasks (BoT)? A System View

BoT = set of jobs sent by a user…

$$W_u = \{J_i | user(J_i) = u\}$$

…that is submitted at most Δs after the first job

$$ST(J') \leq ST(J) + \Delta$$



- Why Bag of *Tasks*? From the perspective of the user, jobs in set are just tasks of a larger job
- A single useful result from the complete BoT
- Result can be combination of all tasks, or a selection of the results of most or even a single task

Iosup et al., The Characteristics and Performance of Groups of Jobs in Grids, Euro-Par, LNCS, vol.4641, pp. 382-393, 2007.

Q0

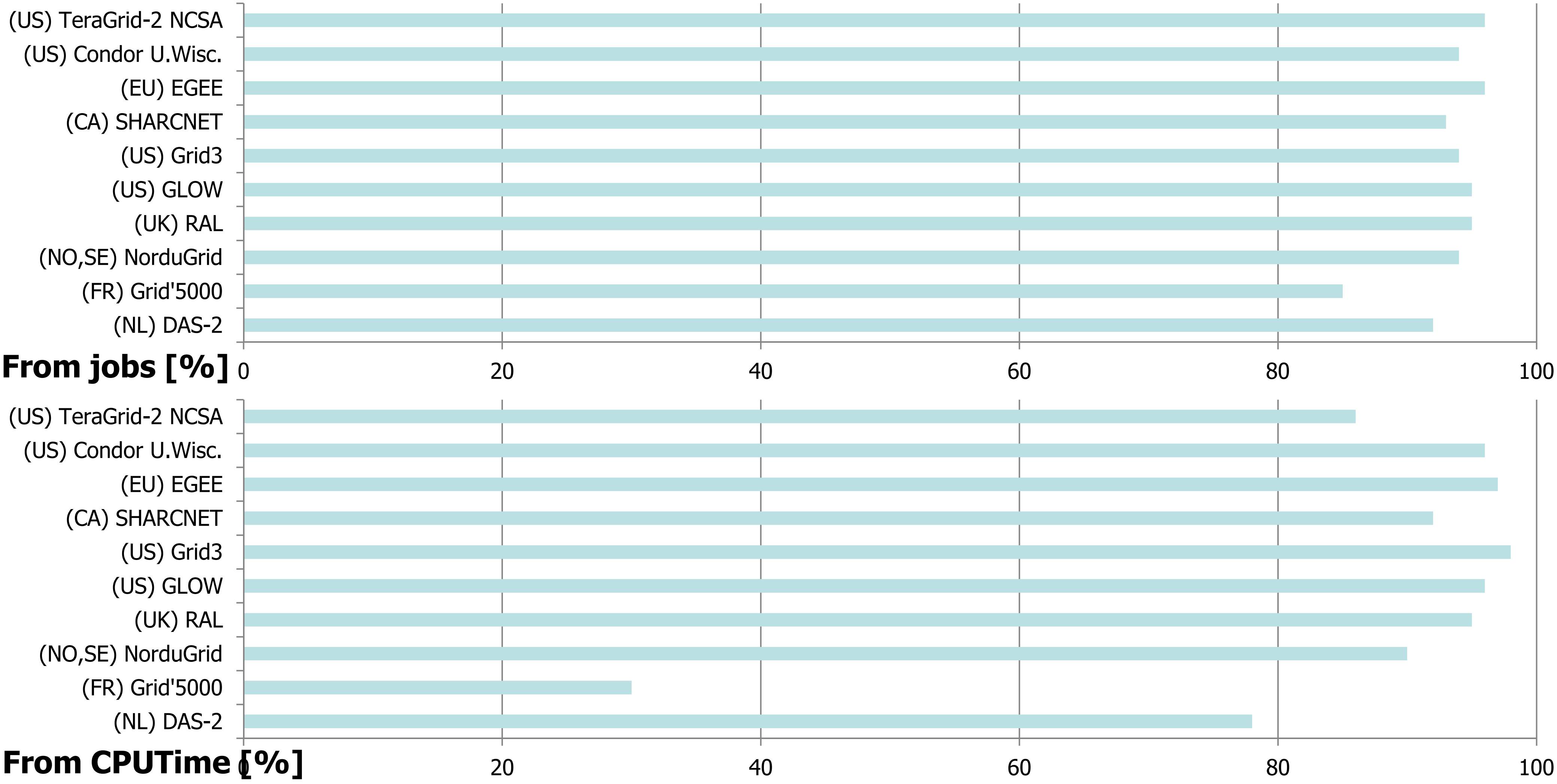
Applications of the BoT Programming Model

- Parameter sweeps
  - Comprehensive, possibly exhaustive investigation of a model
  - Very useful in engineering and simulation-based science

- Monte Carlo simulations
  - Simulation with random elements: fixed time yet limited inaccuracy
  - Very useful in engineering and simulation-based science

- Many other types of batch processing
  - Periodic computation, Cycle scavenging
  - Very useful to automate operations and reduce waste

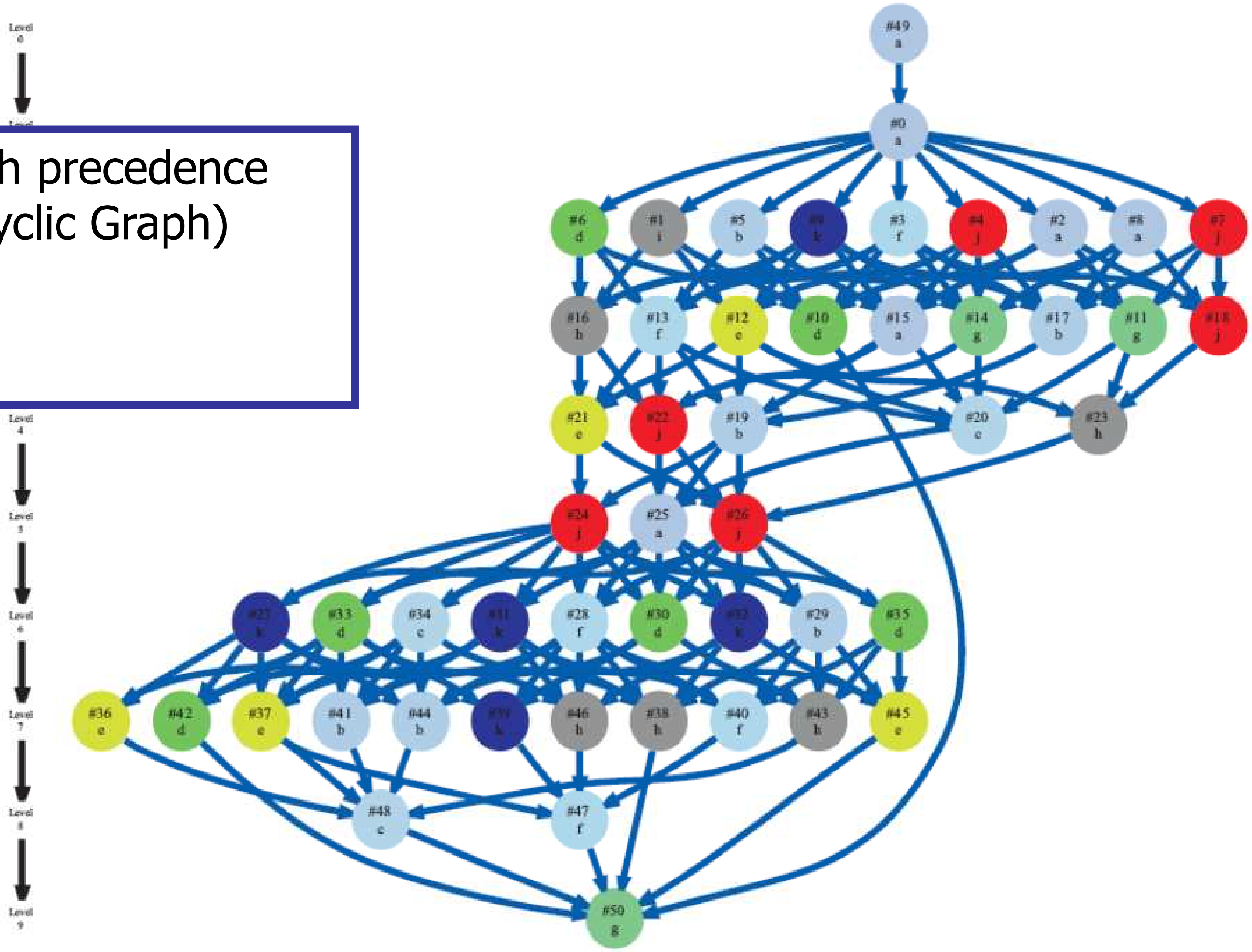
BoTs Are the Dominant Programming Model for Grid Computing (Many Tasks)



Iosup and Epema: Grid Computing Workloads.
IEEE Internet Computing 15(2): 19-26 (2011)

What is a Wokflow?

WF = set of jobs with precedence
(think Direct Acyclic Graph)



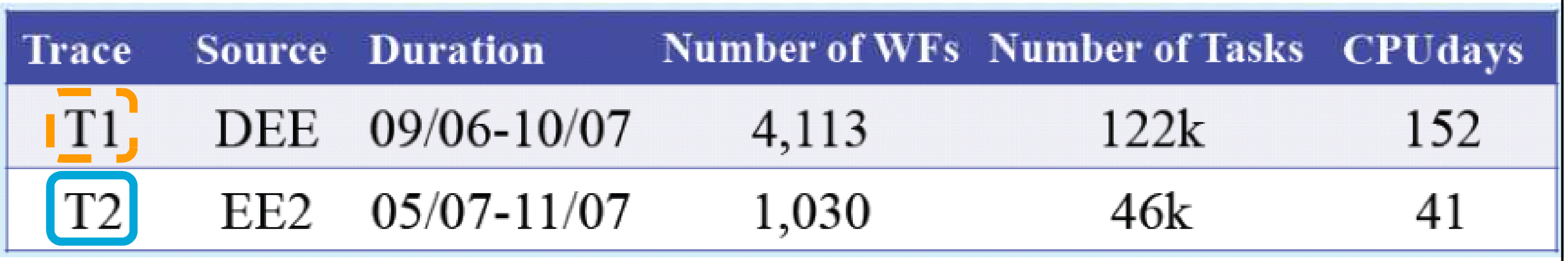
Applications of the Workflow Programming Model

- Complex applications
  - Complex filtering of data
  - Complex analysis of instrument measurements
- Applications created by non-CS scientists*
  - Workflows have a natural correspondence in the real-world, as descriptions of a scientific procedure
  - Visual model of a graph sometimes easier to program
- Precursor of the MapReduce Programming Model (next slides)

*Adapted from: Carole Goble and David de Roure, Chapter in "The Fourth Paradigm", http://research.microsoft.com/en-us/collaboration/fourthparadigm/

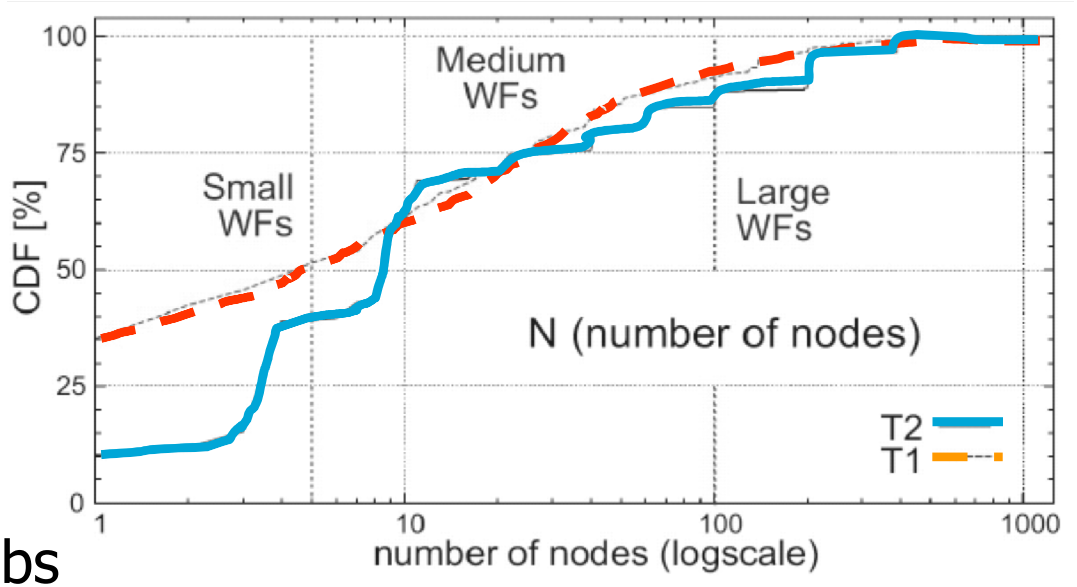
Workflows Exist in Grids, but Did No Evidence of a Dominant Programming Model

- Traces



Trace	Source	Duration	Number of WFs	Number of Tasks	CPUdays
T1	DEE	09/06-10/07	4,113	122k	152
T2	EE2	05/07-11/07	1,030	46k	41

- Selected Findings



  - Loose coupling
  - Graph with 3-4 levels
  - Average WF size is 30/44 jobs
  - 75%+ WFs are sized 40 jobs or less, 95% are sized 200 jobs or less

Ostermann et al., On the Characteristics of Grid Workflows, CoreGRID Integrated Research in Grid Computing (CGIW), 2008.

Q0

What is "Big Data"?

- Very large, distributed aggregations of loosely structured data, often incomplete and inaccessible
- Easily exceeds the processing capacity of conventional database systems
- Principle of Big Data: *"When you can, keep everything!"*
- Too big, too fast, and doesn't comply with the traditional database architectures

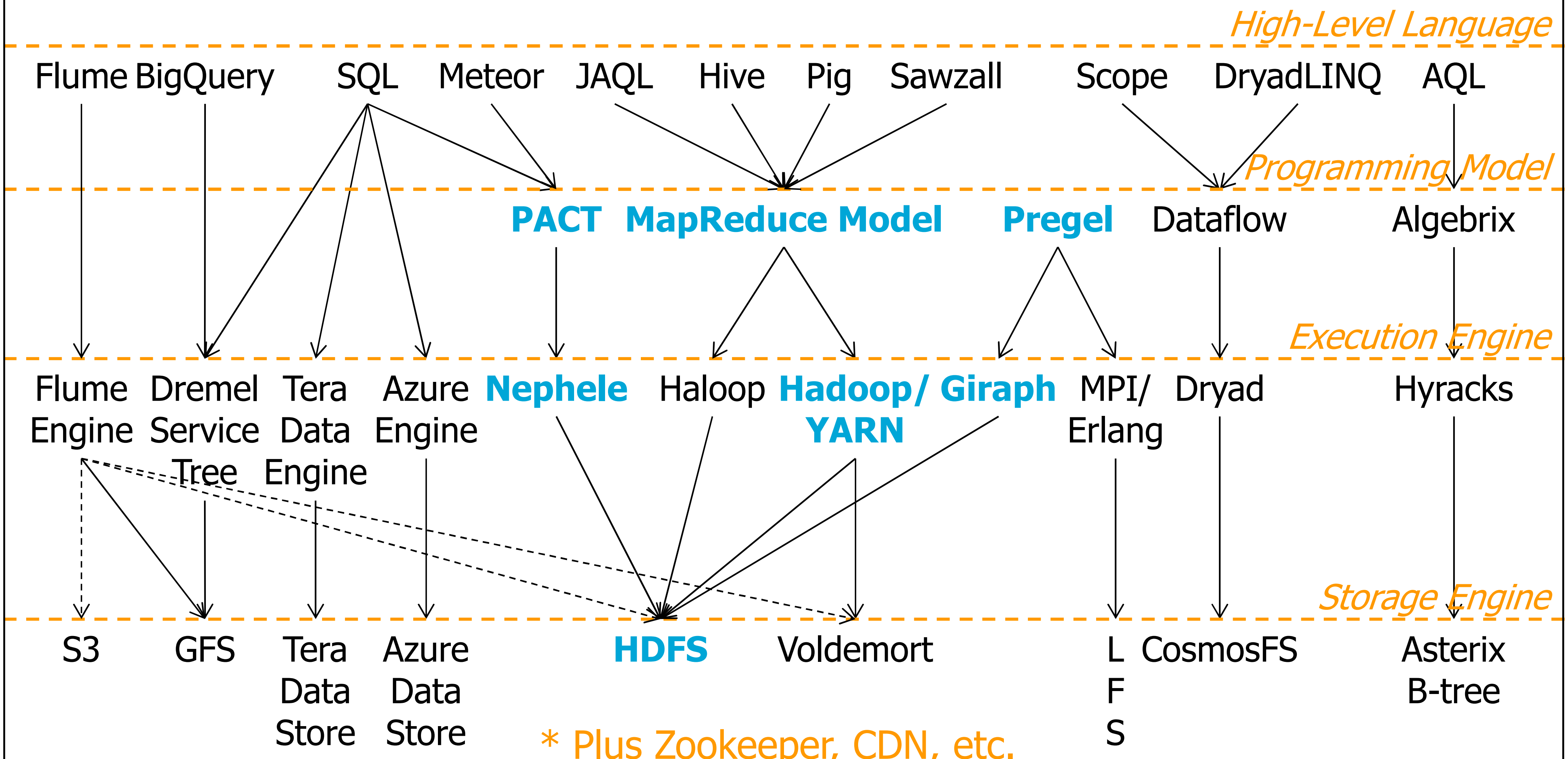
The Three “V”s of Big Data

- Volume
  - More data vs. better models
  - Data grows exponentially
  - Analysis in near-real time to extract value
  - Scalable storage and distributed queries
- Velocity
  - Speed of the feedback loop
  - Gain competitive advantage: fast recommendations
  - Identify fraud, predict customer churn faster
- Variety
  - The data can become messy: text, video, audio, etc.
  - Difficult to integrate into applications

Adapted from: Doug Laney, “3D data management”, META Group/Gartner report, Feb 2001. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Q0

Ecosystems of Big-Data Programming Models



High-Level Language

Flume BigQuery SQL Meteor JAQL Hive Pig Sawzall Scope DryadLINQ AQL

Programming Model

PACT **MapReduce Model** **Pregel** Dataflow Algebrix

Execution Engine

Flume Engine, Dremel Service Tree, Tera Data Engine, Azure Engine, **Nephele**, Haloop, **Hadoop/ YARN**, **Giraph**, MPI/ Erlang, Dryad, Hyracks

Storage Engine

S3, GFS, Tera Data Store, Azure Data Store, **HDFS**, Voldemort, LFS, CosmosFS, Asterix B-tree

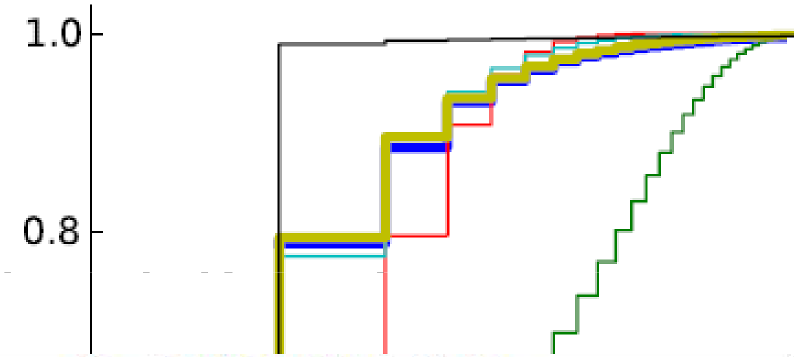
* Plus Zookeeper, CDN, etc.

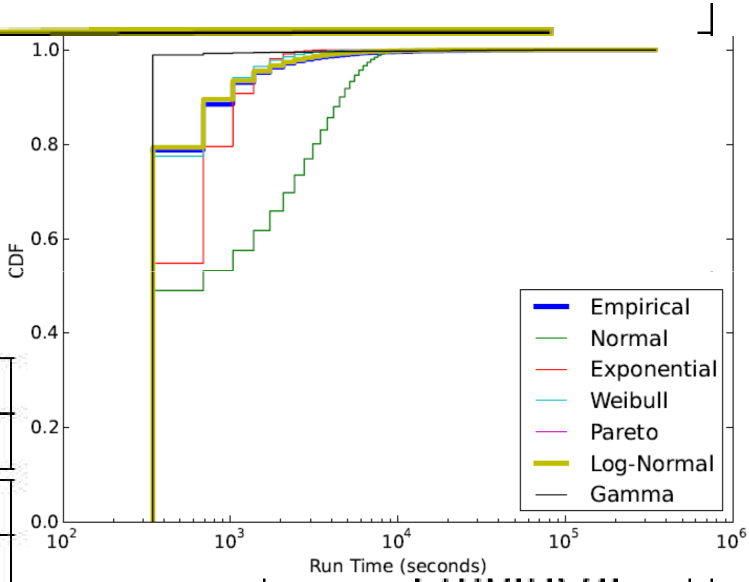
2012-2013

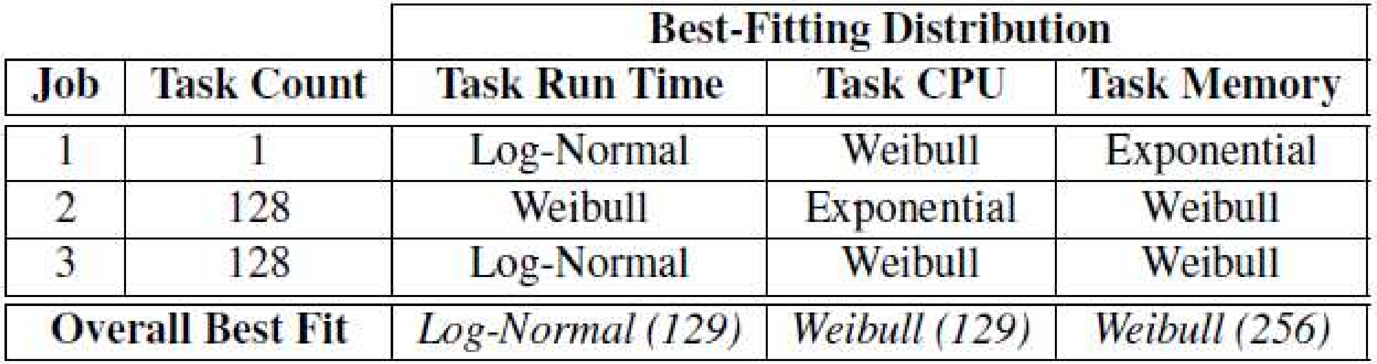
Adapted from: Dagstuhl Seminar on Information Management in the Cloud,
http://www.dagstuhl.de/program/calendar/partlist/?semnr=11321&SUOG

Our Statistical MapReduce Models

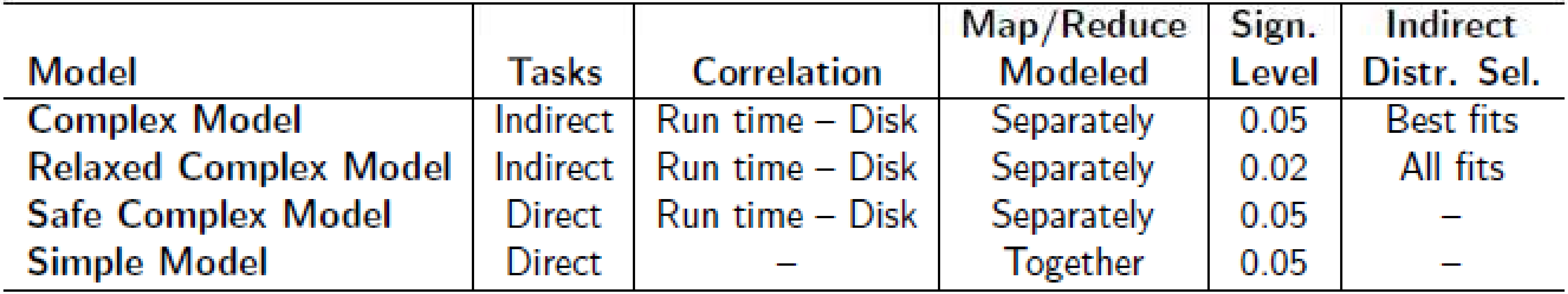
- Real traces
  - Yahoo
  - Google
  - 2 x Social N







		Best-Fitting Distribution		
Job	Task Count	Task Run Time	Task CPU	Task Memory
1	1	Log-Normal	Weibull	Exponential
2	128	Weibull	Exponential	Weibull
3	128	Log-Normal	Weibull	Weibull
Overall Best Fit		*Log-Normal (129)*	*Weibull (129)*	*Weibull (256)*



Model	Tasks	Correlation	Map/Reduce Modeled	Sign. Level	Indirect Distr. Sel.
Complex Model	Indirect	Run time – Disk	Separately	0.05	Best fits
Relaxed Complex Model	Indirect	Run time – Disk	Separately	0.02	All fits
Safe Complex Model	Direct	Run time – Disk	Separately	0.05	–
Simple Model	Direct	–	Together	0.05	–

de Ruiter and Iosup. A workload model for MapReduce. MSc thesis at TU Delft. Jun 2012. Available online via TU Delft Library, http://library.tudelft.nl .

Agenda

1. An Introduction to IaaS Cloud Comput
2. Research Questions or Why We Need B
3. A General Approach and Its Main Challe
4. **IaaS Cloud Workloads (Q0)**
5. **IaaS Cloud Performance (Q1) & Perf. Variability (Q2)**
6. **Provisioning & Allocation Policies for IaaS Clouds (Q3)**
7. Conclusion

IaaS Cloud Performance: Our Team

Alexandru Iosup
TU Delft

Performance
Variability
Isolation
Multi-tenancy
Benchmarking

Dick Epema
TU Delft

Performance
IaaS clouds

Nezih Yigitbasi
TU Delft

Performance
Variability

Athanasios Antoniou
TU Delft

Performance
Isolation

Radu Prodan
U.Isbk.

Benchmarking

Thomas Fahringer
U.Isbk.

Benchmarking

Simon Ostermann
U.Isbk.

Benchmarking

What I’ll Talk About

IaaS Cloud Performance (Q1)

1. Previous work
2. Experimental setup
3. Experimental results
4. Implications on real-world workloads

Some Previous Work (>50 important references across our studies)

Virtualization Overhead

- Loss below 5% for computation [Barham03] [Clark04]
- Loss below 15% for networking [Barham03] [Menon05]
- Loss below 30% for parallel I/O [Vetter08]
- Negligible for compute-intensive HPC kernels [You06] [Panda06]

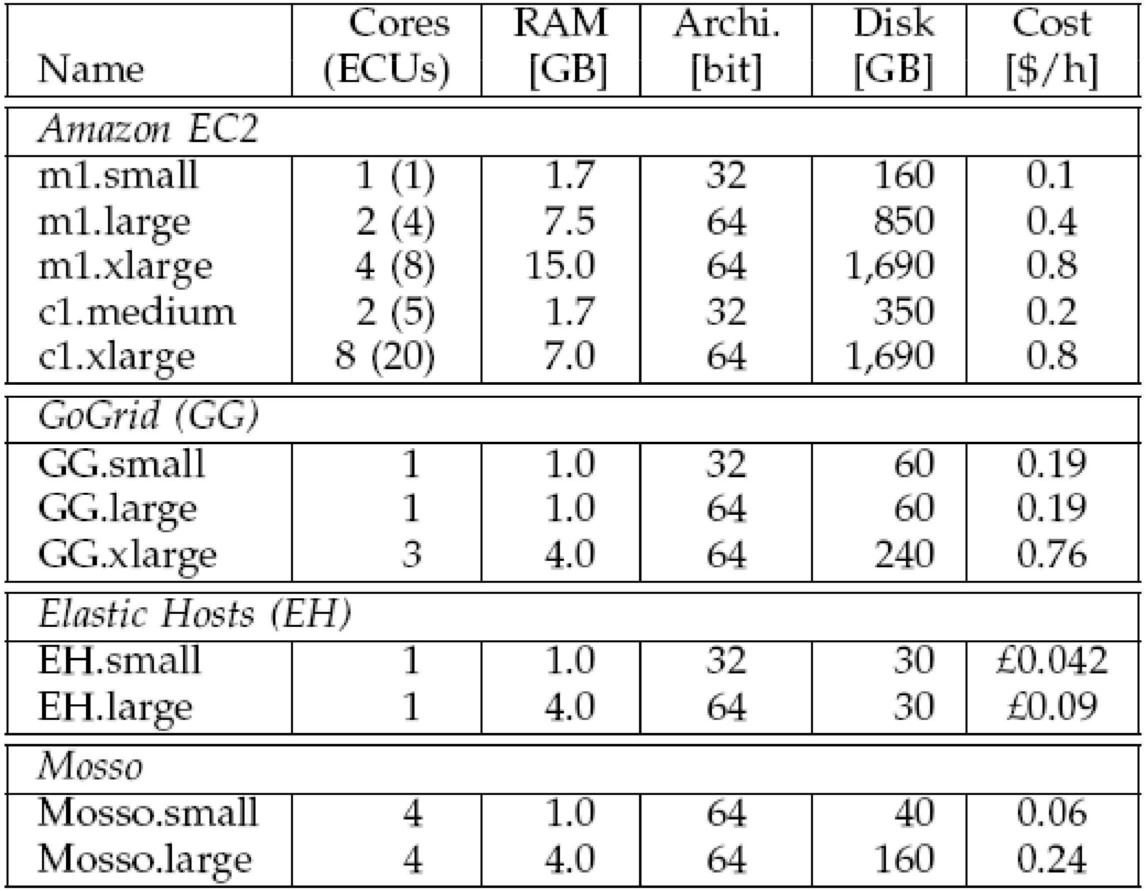
Cloud Performance Evaluation

- Performance and cost of executing a sci. workflows [Dee08]
- Study of Amazon S3 [Palankar08]
- Amazon EC2 for the NPB benchmark suite [Walker08] or selected HPC benchmarks [Hill08]
- CloudCmp [Li10]
- Kosmann et al.

Production IaaS Cloud Services

Q1

- **Production IaaS cloud:** lease resources (infrastructure) to users, operate on the market and have active customers



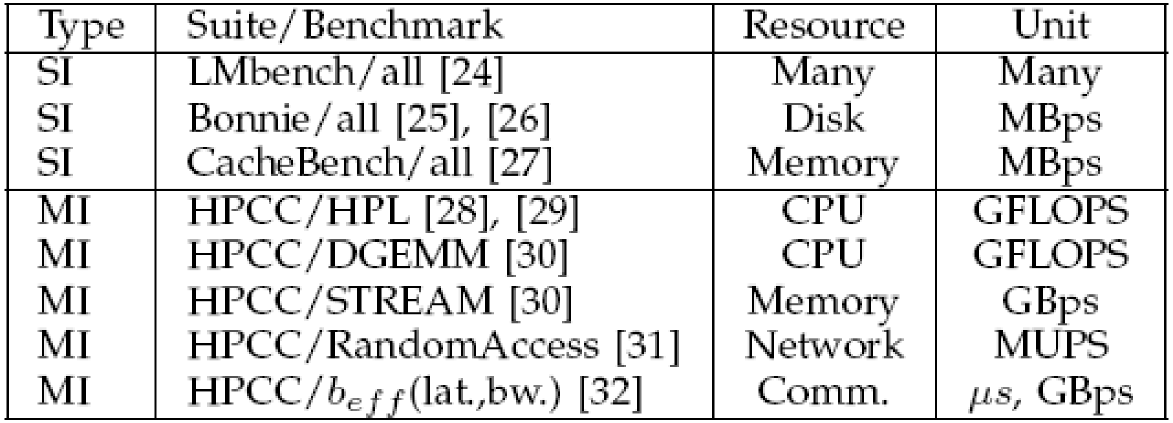
Name	Cores (ECUs)	RAM [GB]	Archi. [bit]	Disk [GB]	Cost [$/h]
Amazon EC2					
m1.small	1 (1)	1.7	32	160	0.1
m1.large	2 (4)	7.5	64	850	0.4
m1.xlarge	4 (8)	15.0	64	1,690	0.8
c1.medium	2 (5)	1.7	32	350	0.2
c1.xlarge	8 (20)	7.0	64	1,690	0.8
GoGrid (GG)					
GG.small	1	1.0	32	60	0.19
GG.large	1	1.0	64	60	0.19
GG.xlarge	3	4.0	64	240	0.76
Elastic Hosts (EH)					
EH.small	1	1.0	32	30	£0.042
EH.large	1	4.0	64	30	£0.09
Mosso					
Mosso.small	4	1.0	64	40	0.06
Mosso.large	4	4.0	64	160	0.24

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Our Method

Q1

- Based on general performance technique: model performance of individual components; system performance is performance of workload + model [Saavedra and Smith, ACM TOCS'96]
- Adapt to clouds:
  1. Cloud-specific elements: resource provisioning and allocation
  2. Benchmarks for single- and multi-machine jobs
  3. Benchmark CPU, memory, I/O, etc.:

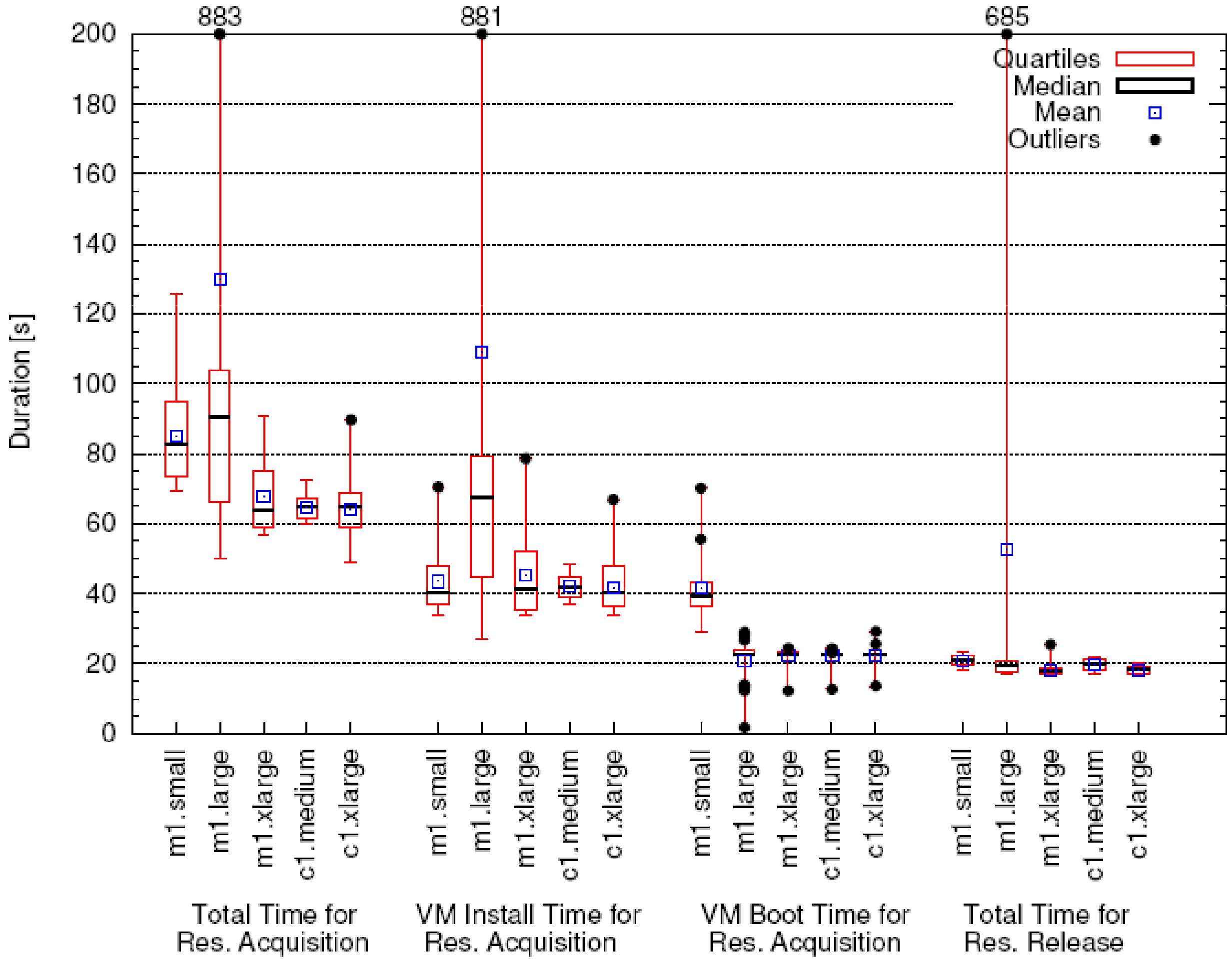


Type	Suite/Benchmark	Resource	Unit
SI	LMbench/all [24]	Many	Many
SI	Bonnie/all [25], [26]	Disk	MBps
SI	CacheBench/all [27]	Memory	MBps
MI	HPCC/HPL [28], [29]	CPU	GFLOPS
MI	HPCC/DGEMM [30]	CPU	GFLOPS
MI	HPCC/STREAM [30]	Memory	GBps
MI	HPCC/RandomAccess [31]	Network	MUPS
MI	HPCC/b_{eff}(lat.,bw.) [32]	Comm.	μs, GBps

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Single Resource Provisioning/Release

Q1

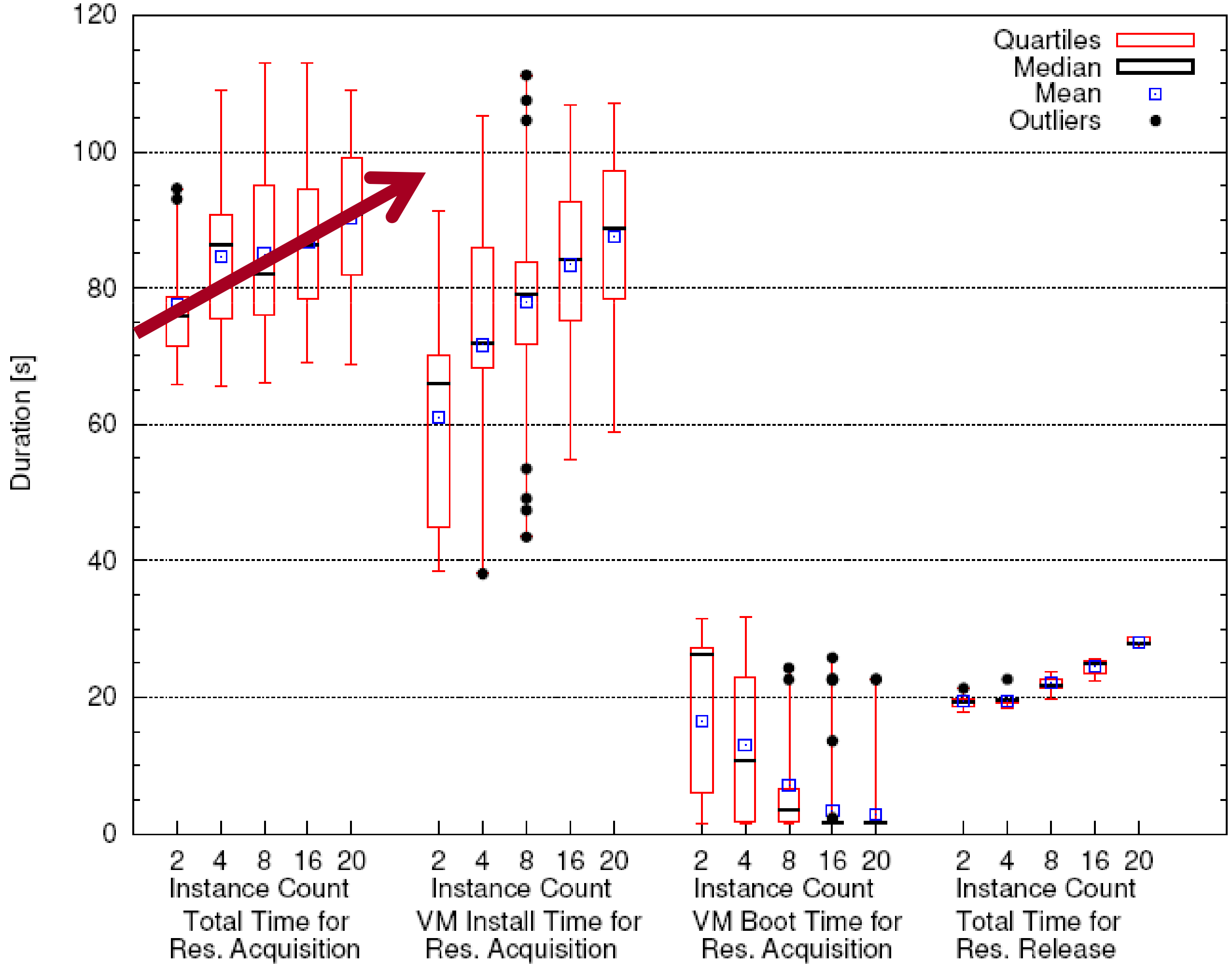


- Time depends on instance type
- Boot time non-negligible

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Multi-Resource Provisioning/Release

Q1



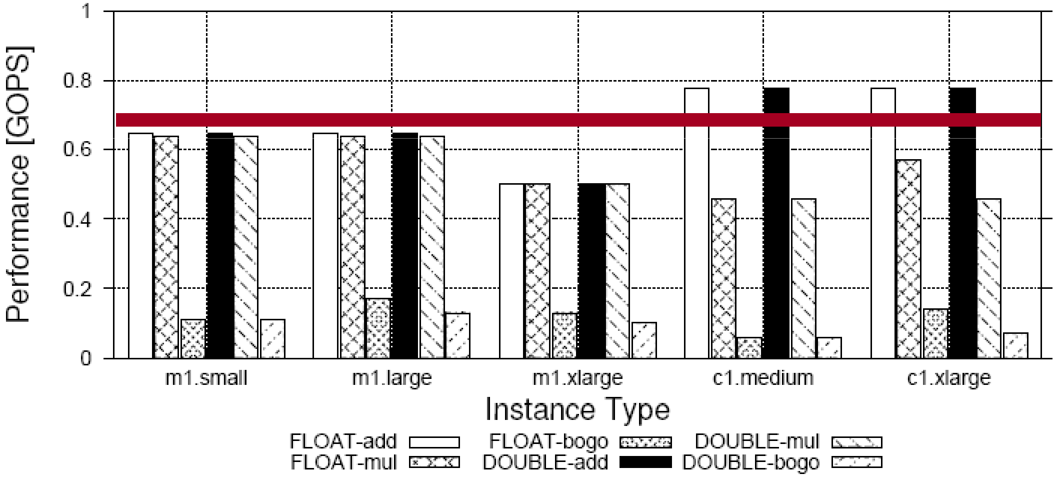
- Time for ***multi***-resource increases with number of resources

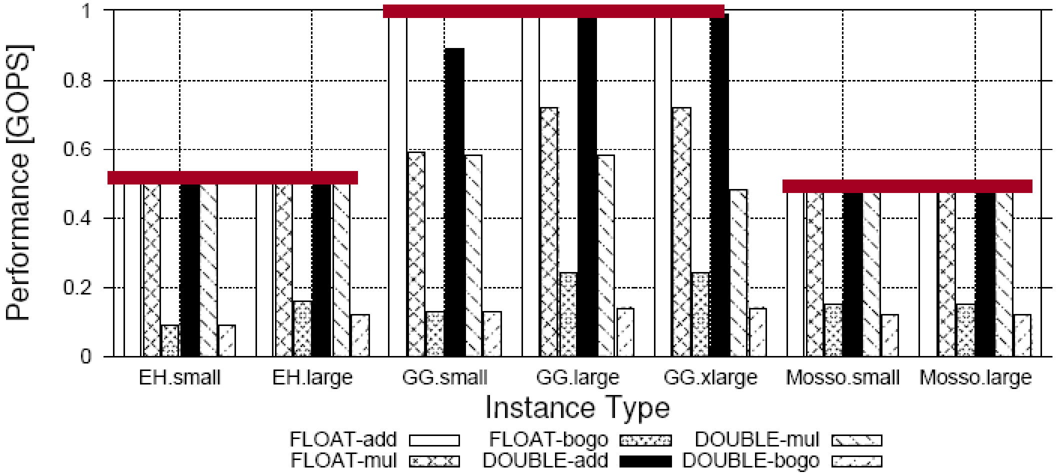
Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

CPU Performance of Single Resource

Q1

- ECU definition: “a 1.1 GHz 2007 Opteron” ~ 4 flops per cycle at full pipeline, which means at peak performance one ECU equals 4.4 gigaflops per second (GFLOPS)
- Real performance 0.6..0.1 GFLOPS = ~1/4..1/7 theoretical peak

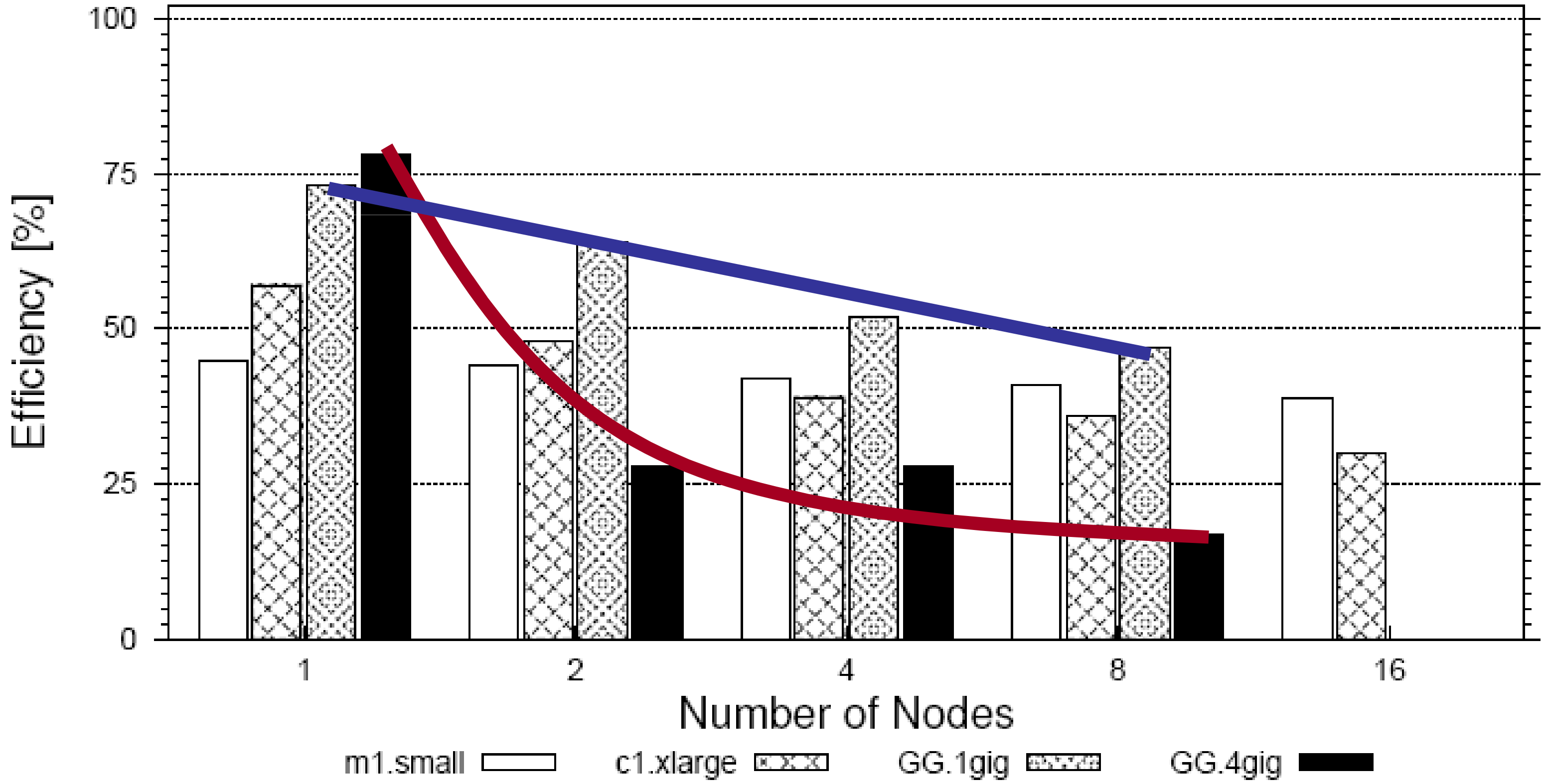




Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

HPLinpack Performance (Parallel)

Q1



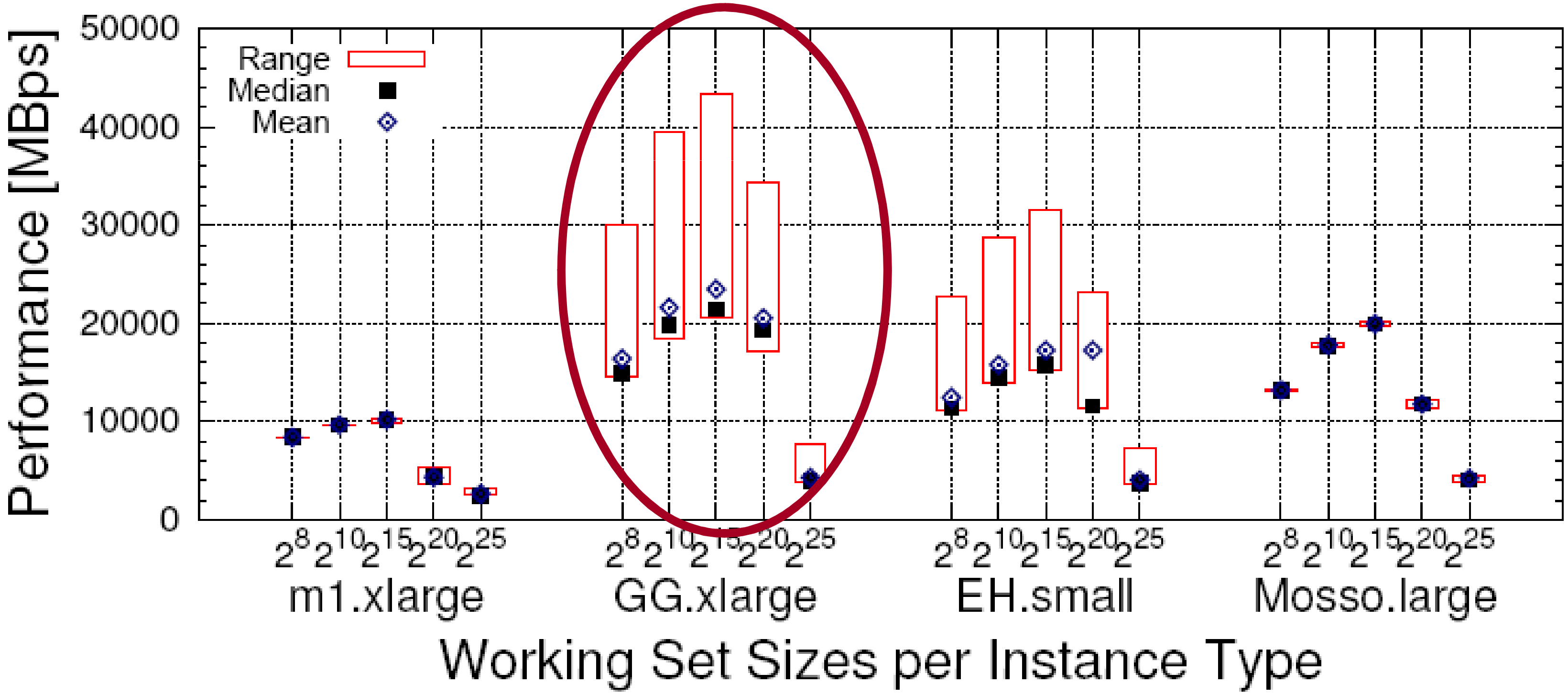
- Low efficiency for parallel compute-intensive applications
- Low performance vs cluster computing and supercomputing

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Performance Stability (Variability)

Q1

Q2



- High performance variability for the best-performing instances

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Summary

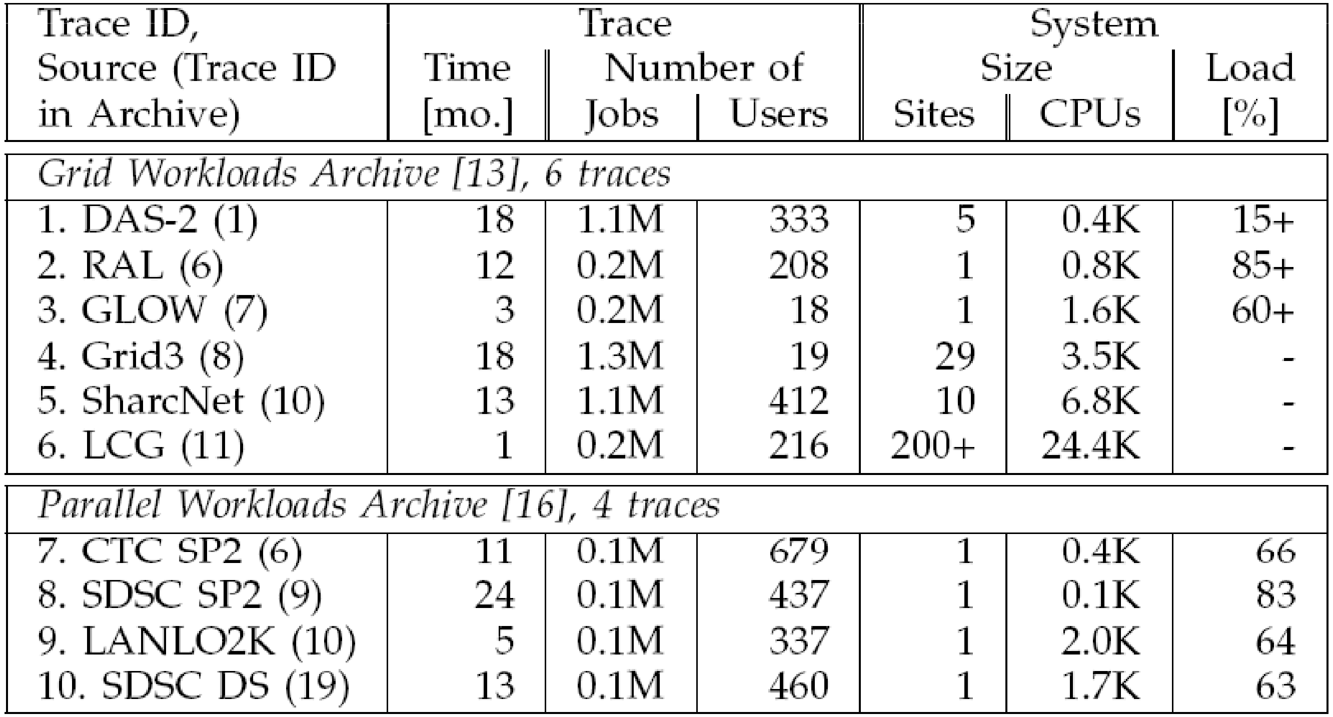
Q1

- Much lower performance than theoretical peak
    - Especially CPU (GFLOPS)
- Performance variability
- Compared results with some of the commercial alternatives (see report)

Q1

Implications: Simulations

- Input: real-world workload traces, grids and PPEs
- Running in
  - Original env.
  - Cloud with source-like perf.
  - Cloud with measured perf.
- Metrics
  - WT, ReT, BSD(10s)
  - Cost [CPU-h]

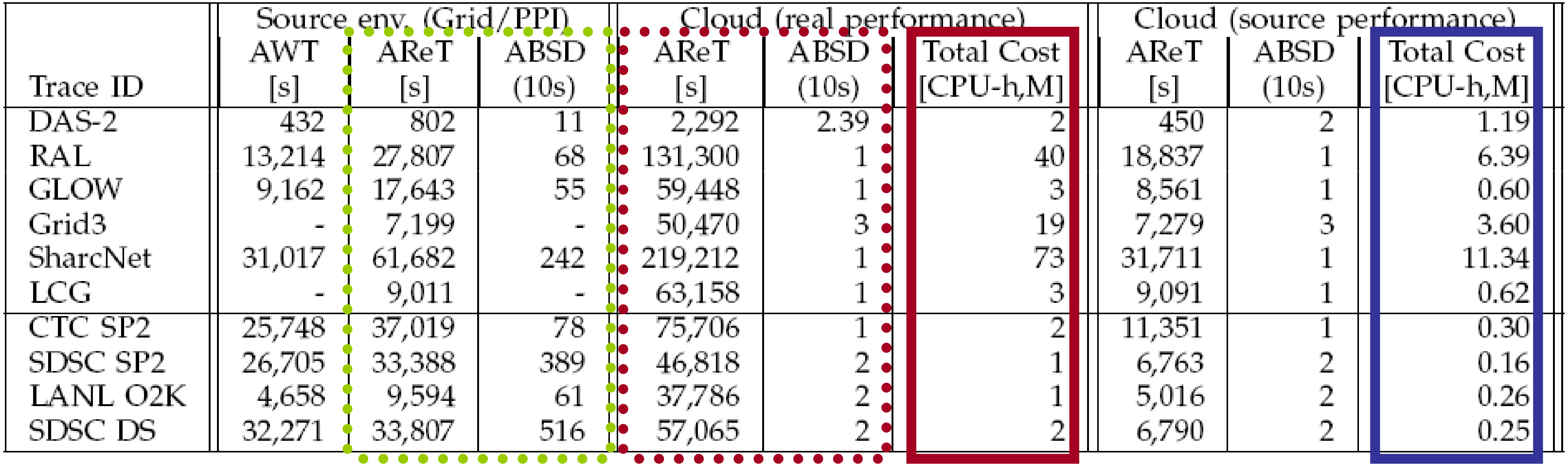


Trace ID,	Trace			System		
Source (Trace ID	Time	Number of		Size		Load
in Archive)	[mo.]	Jobs	Users	Sites	CPUs	[%]
Grid Workloads Archive [13], 6 traces						
1. DAS-2 (1)	18	1.1M	333	5	0.4K	15+
2. RAL (6)	12	0.2M	208	1	0.8K	85+
3. GLOW (7)	3	0.2M	18	1	1.6K	60+
4. Grid3 (8)	18	1.3M	19	29	3.5K	-
5. SharcNet (10)	13	1.1M	412	10	6.8K	-
6. LCG (11)	1	0.2M	216	200+	24.4K	-
Parallel Workloads Archive [16], 4 traces						
7. CTC SP2 (6)	11	0.1M	679	1	0.4K	66
8. SDSC SP2 (9)	24	0.1M	437	1	0.1K	83
9. LANLO2K (10)	5	0.1M	337	1	2.0K	64
10. SDSC DS (19)	13	0.1M	460	1	1.7K	63

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Q1

Implications: Results



	Source env. (Grid/PPI)			Cloud (real performance)			Cloud (source performance)		
Trace ID	AWT [s]	AReT [s]	ABSD (10s)	AReT [s]	ABSD (10s)	Total Cost [CPU-h,M]	AReT [s]	ABSD (10s)	Total Cost [CPU-h,M]
DAS-2	432	802	11	2,292	2.39	2	450	2	1.19
RAL	13,214	27,807	68	131,300	1	40	18,837	1	6.39
GLOW	9,162	17,643	55	59,448	1	3	8,561	1	0.60
Grid3	-	7,199	-	50,470	3	19	7,279	3	3.60
SharcNet	31,017	61,682	242	219,212	1	73	31,711	1	11.34
LCG	-	9,011	-	63,158	1	3	9,091	1	0.62
CTC SP2	25,748	37,019	78	75,706	1	2	11,351	1	0.30
SDSC SP2	26,705	33,388	389	46,818	2	1	6,763	2	0.16
LANL O2K	4,658	9,594	61	37,786	2	1	5,016	2	0.26
SDSC DS	32,271	33,807	516	57,065	2	2	6,790	2	0.25

- Cost: Clouds, real >> Clouds, source
- Performance:
  - AReT: Clouds, real >> Source env. (**bad**)
  - AWT,ABSD: Clouds, real << Source env. (**good**)

Iosup et al., Performance Analysis of Cloud Computing Services for Many Tasks Scientific Computing, (IEEE TPDS 2011).

Agenda

1. An Introduction to IaaS Cloud Comput
2. Research Questions or Why We Need B
3. A General Approach and Its Main Challe
4. **IaaS Cloud Workloads (Q0)**
5. **IaaS Cloud Performance (Q1) & Perf. Variability (Q2)**
6. **Provisioning & Allocation Policies for IaaS Clouds (Q3)**
7. Conclusion

IaaS Cloud Performance: Our Team

Alexandru Iosup
TU Delft

Performance
Variability
Isolation
Multi-tenancy
Benchmarking

Dick Epema
TU Delft

Performance
IaaS clouds

Nezih Yigitbasi
TU Delft

Performance
Variability

Athanasios Antoniou
TU Delft

Performance
Isolation

Radu Prodan
U.Isbk.

Benchmarking

Thomas Fahringer
U.Isbk.

Benchmarking

Simon Ostermann
U.Isbk.

Benchmarking

What I'll Talk About

IaaS Cloud Performance Variability (Q2)

1. Experimental setup
2. Experimental results
3. Implications on real-world workloads

Production Cloud Services

Q2

- **Production cloud:** operate on the market and have active customers

- **IaaS/PaaS:**
  Amazon Web Services (AWS)
  - EC2 (Elastic Compute Cloud)
  - S3 (Simple Storage Service)
  - SQS (Simple Queueing Service)
  - SDB (Simple Database)
  - FPS (Flexible Payment Service)

- **PaaS:**
  Google App Engine (GAE)
  - Run (Python/Java runtime)
  - Datastore (Database) ~ SDB
  - Memcache (Caching)
  - URL Fetch (Web crawling)

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Our Method [1/3]
Performance Traces

Q2

- CloudStatus*
  - Real-time values and weekly averages for most of the AWS and GAE services
- Periodic performance probes
  - Sampling rate is under 2 minutes

* www.cloudstatus.com

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Our Method [2/3] Analysis

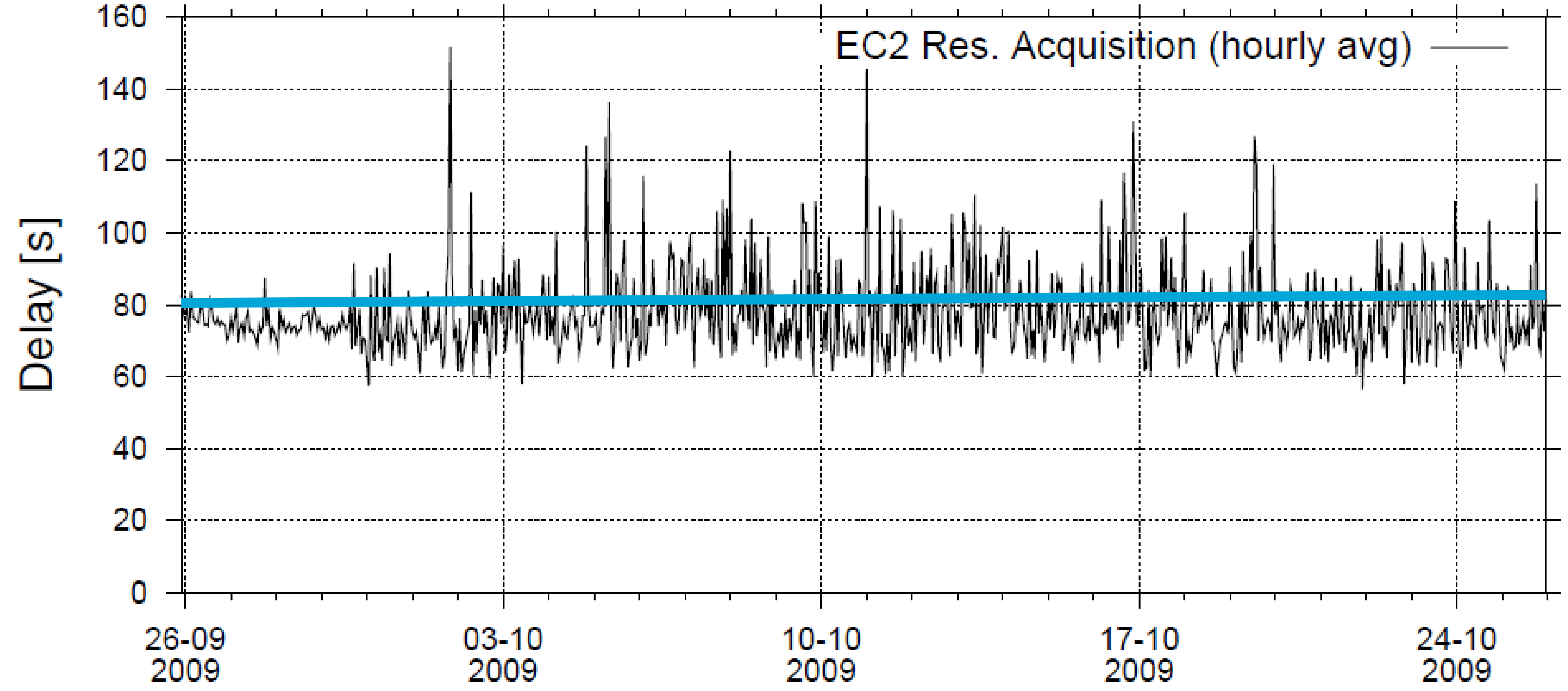
Q2

1. Find out whether variability is present
   - Investigate several months whether the performance metric is highly variable
2. Find out the characteristics of variability
   - Basic statistics: the five quartiles (Q_0-Q_4) including the median (Q_2), the mean, the standard deviation
   - Derivative statistic: the IQR (Q_3-Q_1)
   - CoV > 1.1 indicate high variability
3. Analyze the performance variability time patterns
   - Investigate for each performance metric the presence of daily/monthly/weekly/yearly time patterns
   - E.g., for monthly patterns divide the dataset into twelve subsets and for each subset compute the statistics and plot for visual inspection

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Our Method [3/3]
Is Variability Present?

- **Validated Assumption:** The performance delivered by production services is variable.

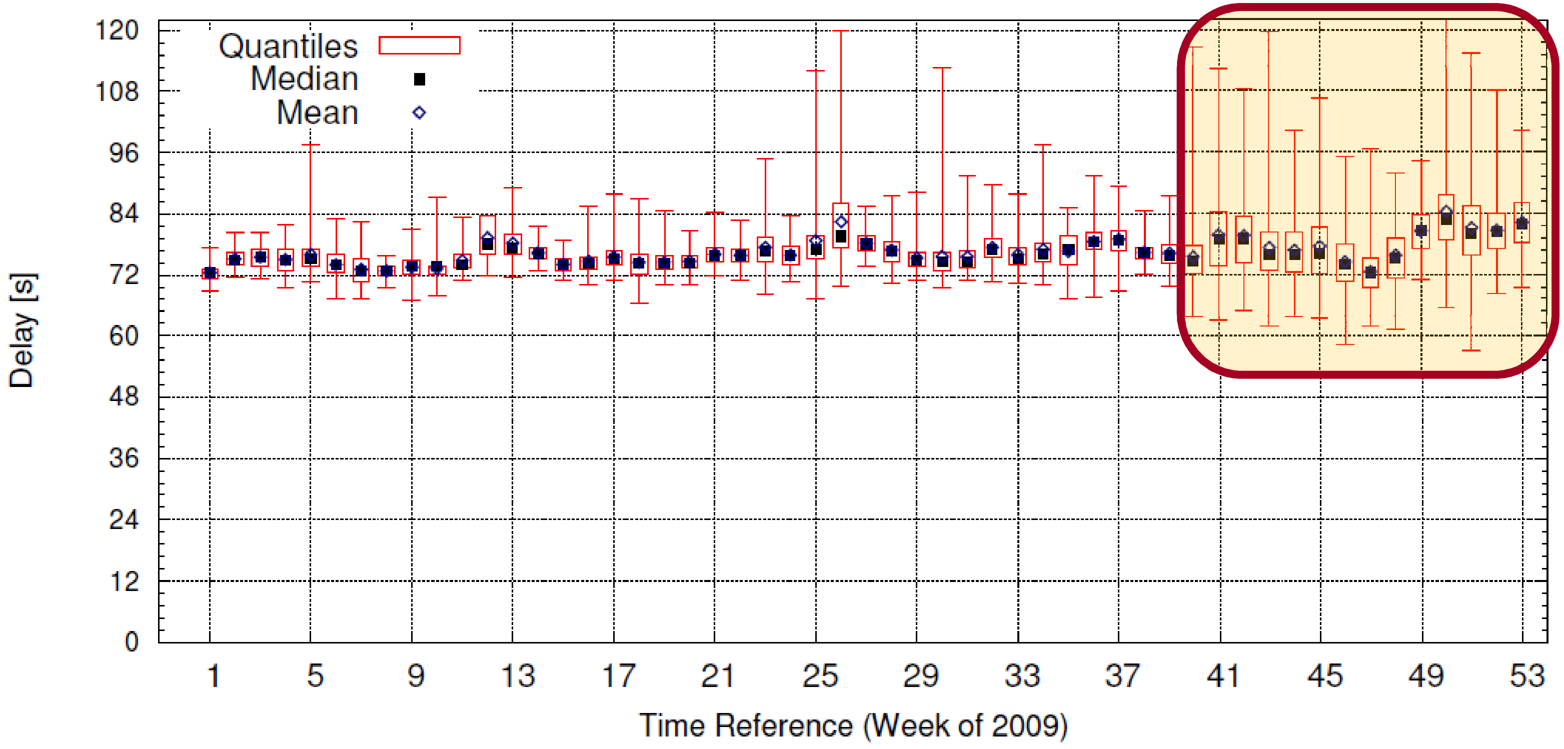


Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

AWS Dataset (1/4): EC2

Variable Performance

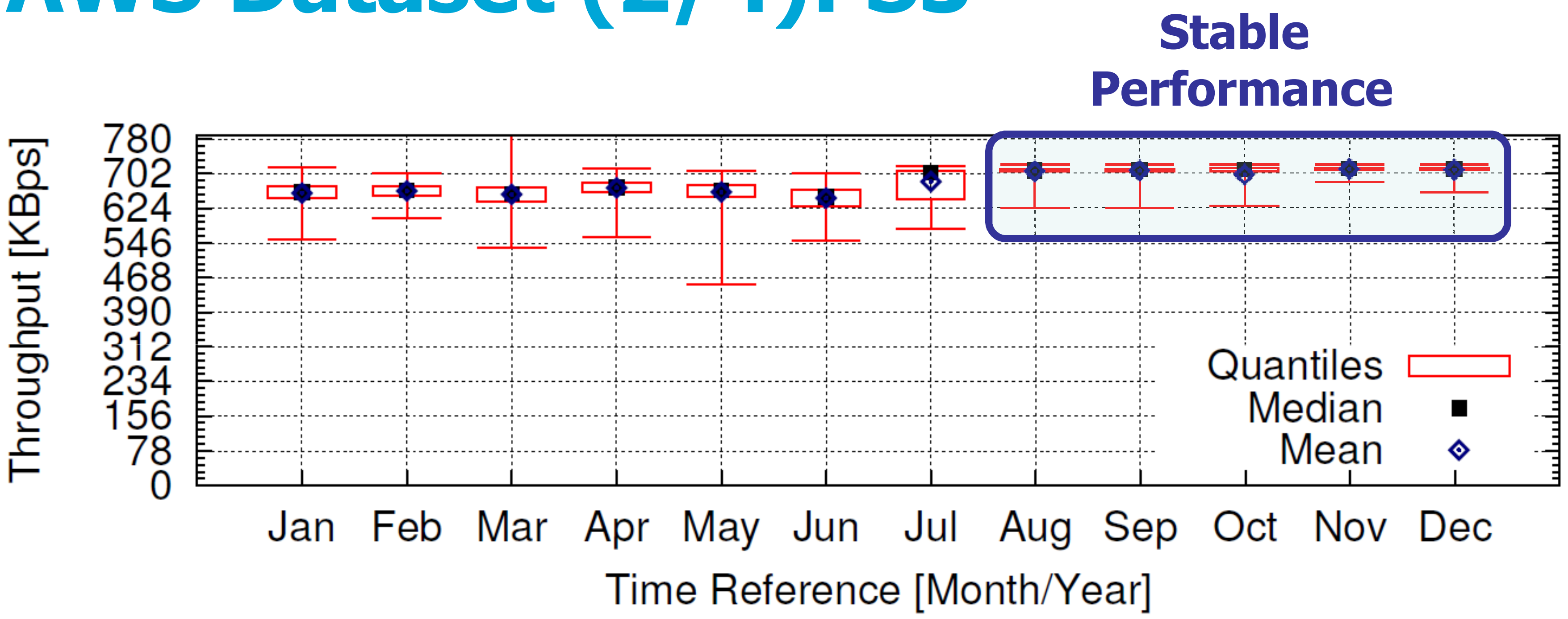
Q2



- **Deployment Latency [s]:** Time it takes to start a small instance, from the startup to the time the instance is available
- Higher IQR and range from week 41 to the end of the year; possible reasons:
  - Increasing EC2 user base
  - Impact on applications using EC2 for auto-scaling

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

AWS Dataset (2/4): S3

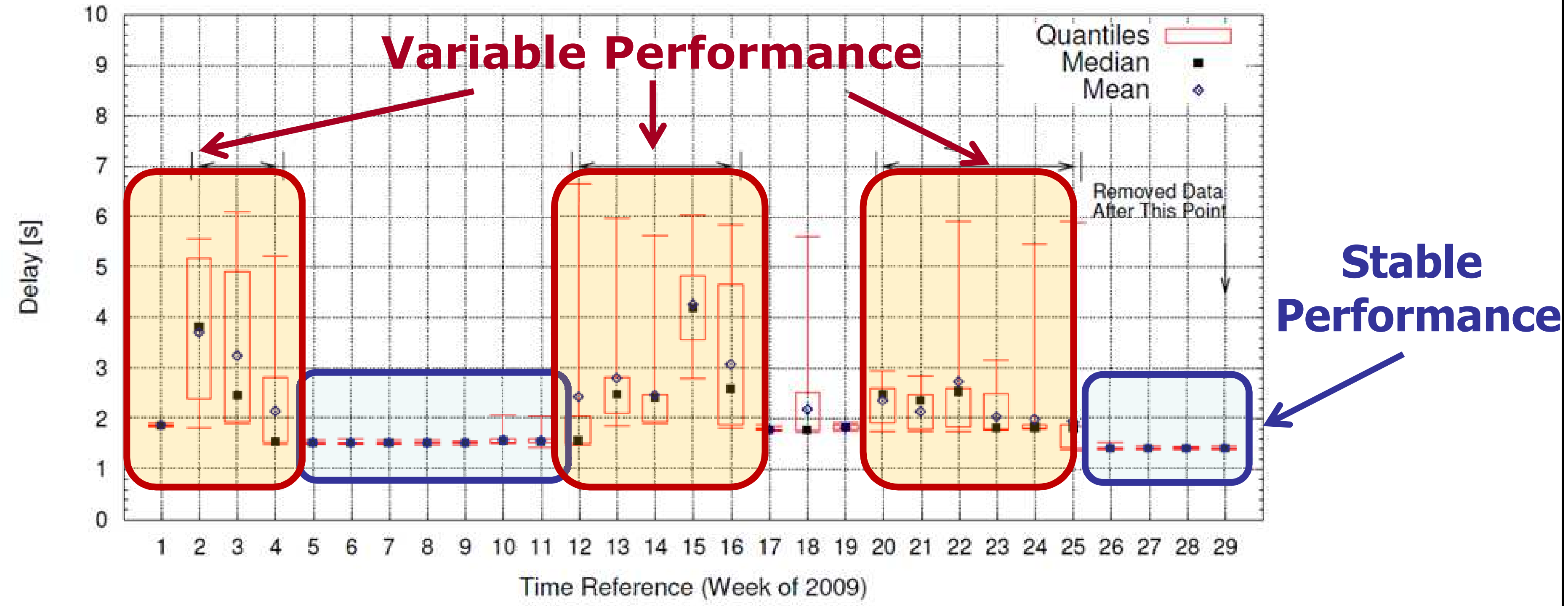


- **Get Throughput [bytes/s]:** Estimated rate at which an object in a bucket is read
- The last five months of the year exhibit much lower IQR and range
  - More stable performance for the last five months
  - Probably due to software/infrastructure upgrades

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

AWS Dataset (3/4): SQS

Q2



- **Average Lag Time [s]:** Time it takes for a posted message to become available to read. Average over multiple queues.
- Long periods of stability (low IQR and range)
- Periods of high performance variability also exist

AWS Dataset (4/4): Summary

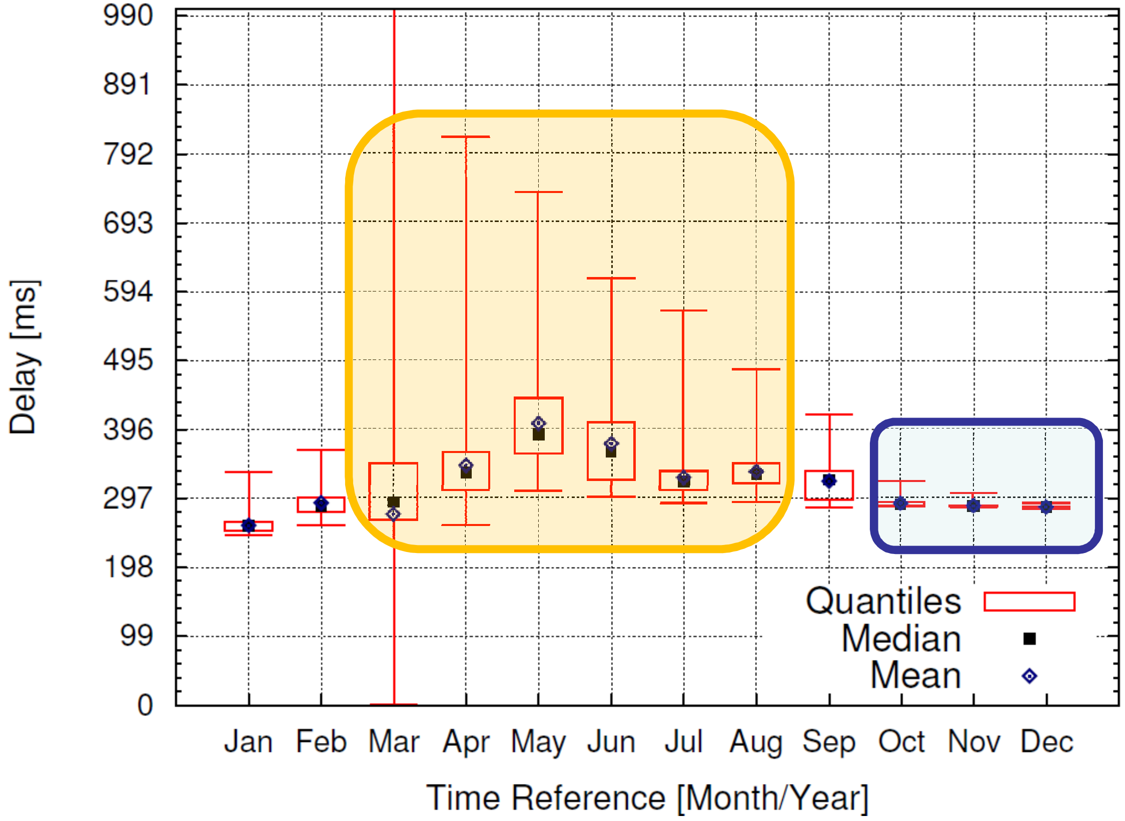
Q2

- **All services exhibit time patterns in performance**
- EC2: periods of special behavior
- SDB and S3: daily, monthly and yearly patterns
- SQS and FPS: periods of special behavior

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

GAE Dataset (1/4): Run Service

Q2

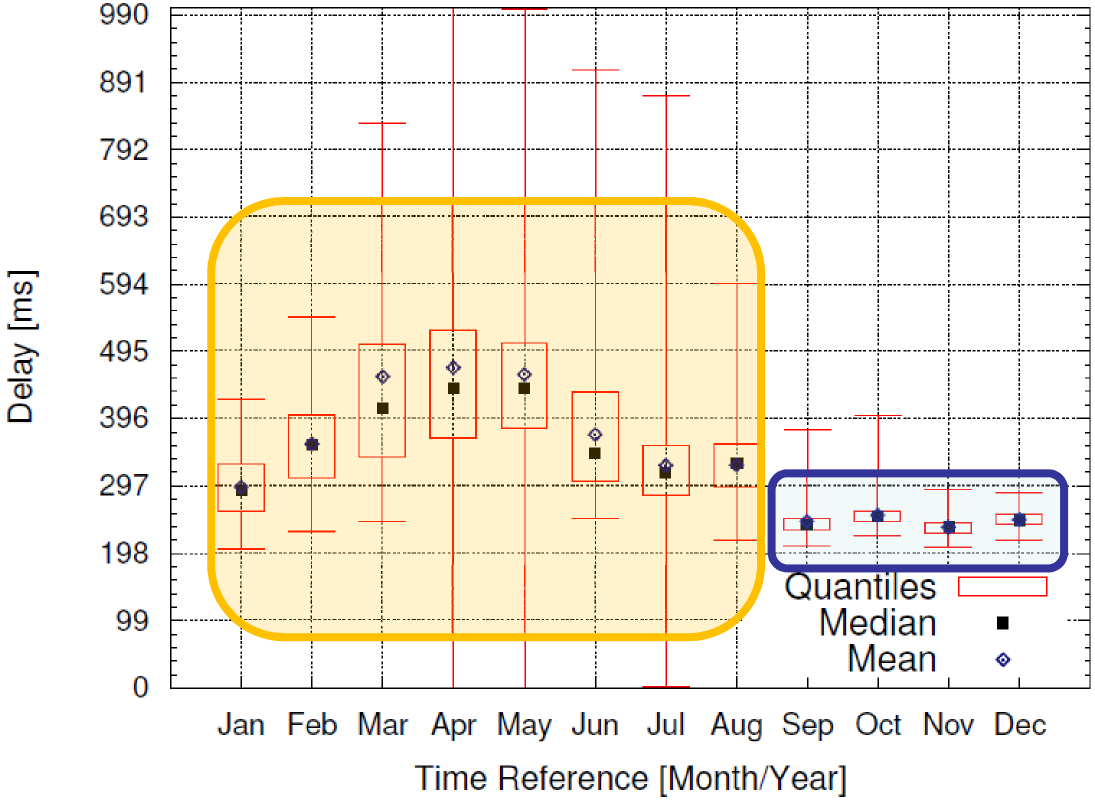


- **Fibonacci [ms]:** Time it takes to calculate the 27th Fibonacci number
- Highly variable performance until September
- Last three months have stable performance (low IQR and range)

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

GAE Dataset (2/4): Datastore

Q2

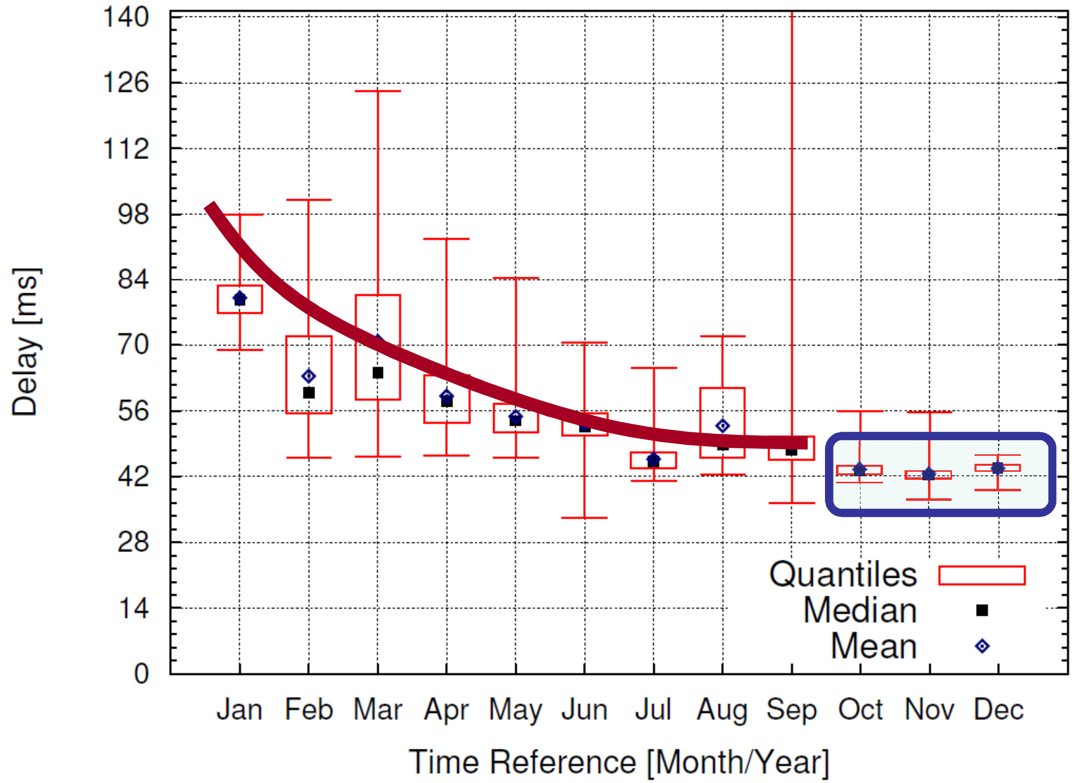


- **Read Latency [s]:** Time it takes to read a "User Group"
- Yearly pattern from January to August
- The last four months of the year exhibit much lower IQR and range
  - More stable performance for the last five months
  - Probably due to software/infrastructure upgrades

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

GAE Dataset (3/4): Memcache

Q2



- **PUT [ms]:** Time it takes to put 1 MB of data in memcache.
- Median performance per month has an increasing trend over the first 10 months
- The last three months of the year exhibit stable performance

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

GAE Dataset (4/4): Summary

Q2

- All services exhibit time patterns

- Run Service: daily patterns and periods of special behavior
- Datastore: yearly patterns and periods of special behavior
- Memcache: monthly patterns and periods of special behavior
- URL Fetch: daily and weekly patterns, and periods of special behavior

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Experimental Setup (1/2): Simulations

Q2

- Trace based simulations for three applications
- **Input**
  - GWA traces
  - Number of daily unique users
  - Monthly performance variability

Application	**Service**
Job Execution	GAE Run
Selling Virtual Goods	AWS FPS
Game Status Maintenance	AWS SDB/GAE Datastore

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Experimental Setup (2/2): Metrics

Q2

- Average Response Time and Average Bounded Slowdown
- Cost in millions of consumed CPU hours
- **Aggregate Performance Penalty** -- APP(t)

$$\frac{P(t)}{P_{ref}} \times \frac{U(t)}{\max U(t)}$$

  - Pref (Reference Performance): Average of the twelve monthly medians
  - P(t): **random** value sampled from the distribution corresponding to the current month at time t (*Performance is like a box of chocolates, you never know what you're gonna get ~ Forrest Gump*)
  - max U(t): max number of users over the whole trace
  - U(t): number of users at time t
  - **APP—the lower the better**

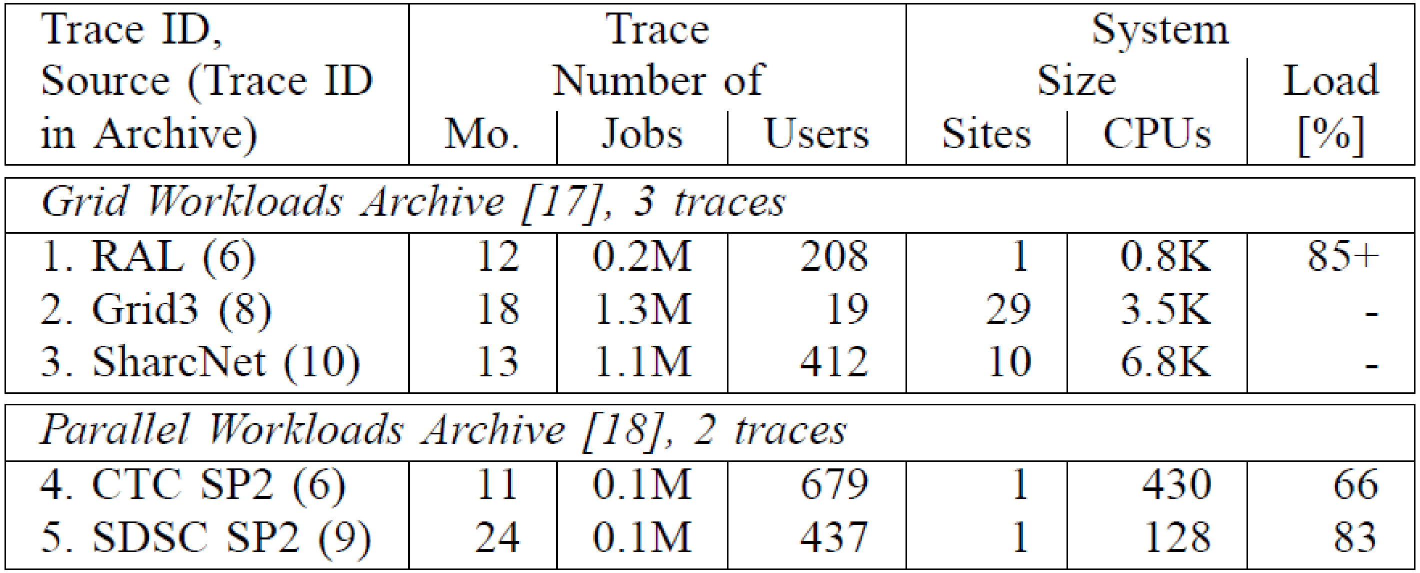
TUDelft
Delft University of Technology

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Q2

Grid & PPE Job Execution (1/2): Scenario

- Execution of compute-intensive jobs typical for grids and PPEs on cloud resources
- **Traces**



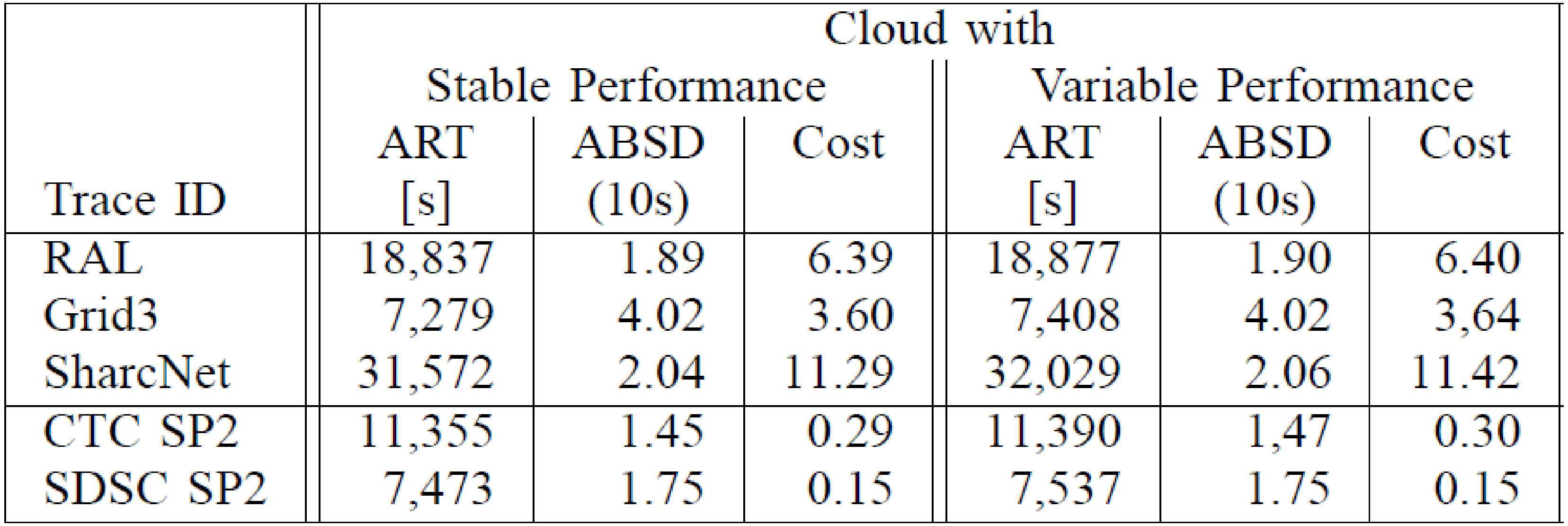
Trace ID,	Trace			System		
Source (Trace ID		Number of		Size		Load
in Archive)	Mo.	Jobs	Users	Sites	CPUs	[%]
Grid Workloads Archive [17], 3 traces						
1. RAL (6)	12	0.2M	208	1	0.8K	85+
2. Grid3 (8)	18	1.3M	19	29	3.5K	-
3. SharcNet (10)	13	1.1M	412	10	6.8K	-
Parallel Workloads Archive [18], 2 traces						
4. CTC SP2 (6)	11	0.1M	679	1	430	66
5. SDSC SP2 (9)	24	0.1M	437	1	128	83

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Q2

Grid & PPE Job Execution (2/2): Results

- All metrics differ by less than 2% between cloud with stable and the cloud with variable performance
- Impact of service performance variability is low for this scenario



	Cloud with					
	Stable Performance			Variable Performance		
Trace ID	ART [s]	ABSD (10s)	Cost	ART [s]	ABSD (10s)	Cost
RAL	18,837	1.89	6.39	18,877	1.90	6.40
Grid3	7,279	4.02	3.60	7,408	4.02	3,64
SharcNet	31,572	2.04	11.29	32,029	2.06	11.42
CTC SP2	11,355	1.45	0.29	11,390	1,47	0.30
SDSC SP2	7,473	1.75	0.15	7,537	1.75	0.15

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Selling Virtual Goods (1/2): Scenario

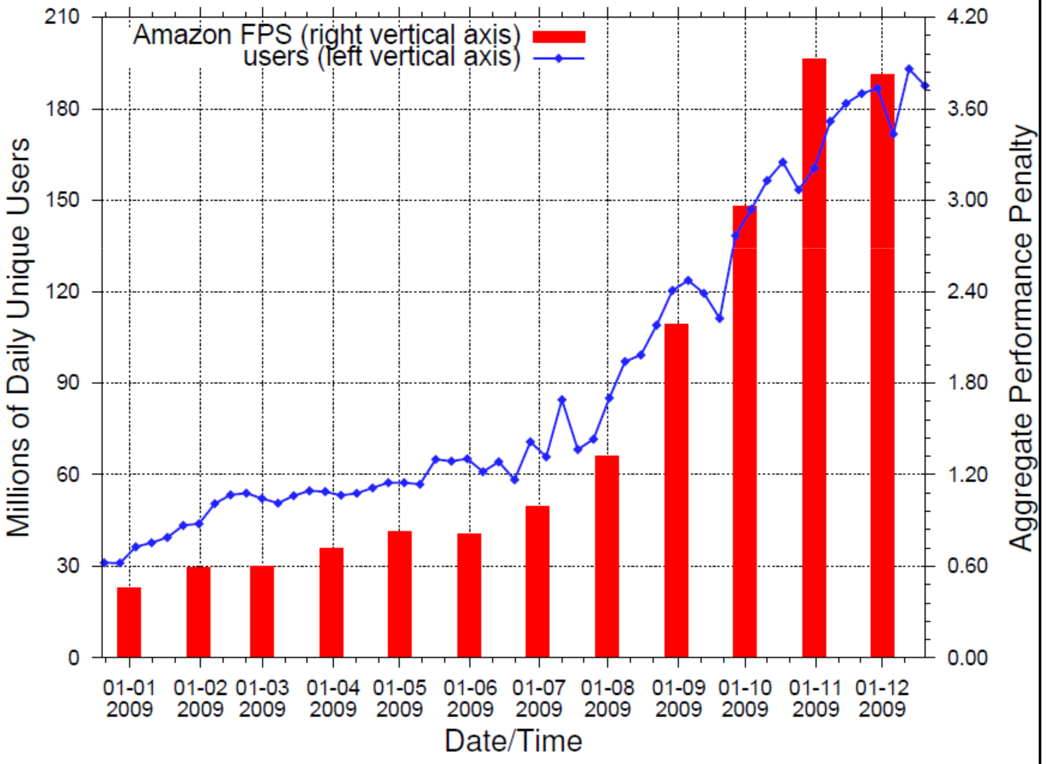
- Virtual good selling application operating on a large-scale social network like Facebook
- Amazon FPS is used for payment transactions
- Amazon FPS performance variability is modeled from the AWS dataset
- **Traces:** Number of daily unique users of Facebook*

* Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Q2

Selling Virtual Goods (2/2): Results

- Significant cloud performance decrease of FPS during the last four months + increasing number of daily users is well-captured by **APP**



- APP metric can trigger and motivate the decision of switching cloud providers

TUDelft
Delft University of Technology

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Game Status Maintenance (1/2): Scenario

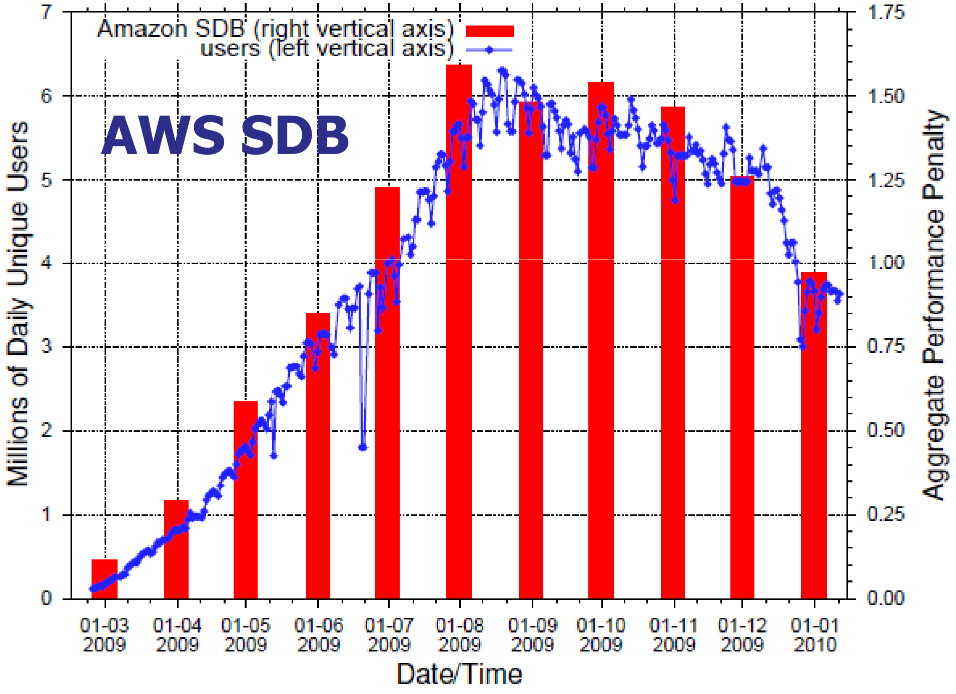
Q2

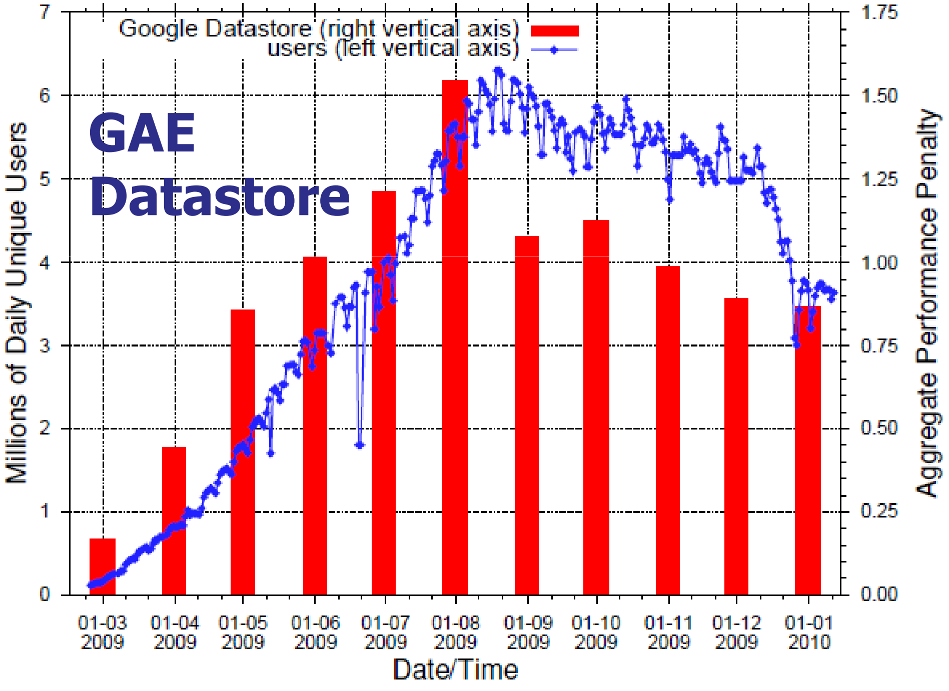
- Maintenance of game status for a large-scale social game such as Farm Town or Mafia Wars which have millions of unique users daily
- AWS SDB and GAE Datastore
- We assume that the number of database operations depends linearly on the number of daily unique users

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Game Status Maintenance (2): Results

Q2





- Big discrepancy between SDB and Datastore services
- **Sep’09-Jan’10:** APP of Datastore is well below than that of SDB due to increasing performance of Datastore
- APP of Datastore ~**1** => no performance penalty
- APP of SDB ~**1.4** => %40 higher performance penalty than SDB

Iosup, Yigitbasi, Epema. On the Performance Variability of Production Cloud Services, (IEEE CCgrid 2011).

Agenda

1. An Introduction to IaaS Cloud Comput
2. Research Questions or Why We Need B
3. A General Approach and Its Main Challe
4. **IaaS Cloud Workloads (Q0)**
5. **IaaS Cloud Performance (Q1) & Perf. Variability (Q2)**
6. **Provisioning & Allocation Policies for IaaS Clouds (Q3)**
7. Conclusion

IaaS Cloud Policies: Our Team

Alexandru Iosup
TU Delft

Provisioning
Allocation
Elasticity
Utility
Isolation
Multi-Tenancy

Dick Epema
TU Delft

Provisioning
Allocation
Koala

Bogdan Ghit
TU Delft

Provisioning
Allocation
Koala

Athanasios Antoniou
TU Delft

Provisioning
Allocation
Isolation
Utility

Orna Agmon-Ben Yehuda
Technion
Elasticity, Utility

David Villegas
FIU/IBM
Elasticity, Utility

What I’ll Talk About

Provisioning and Allocation Policies for IaaS Clouds (Q3)

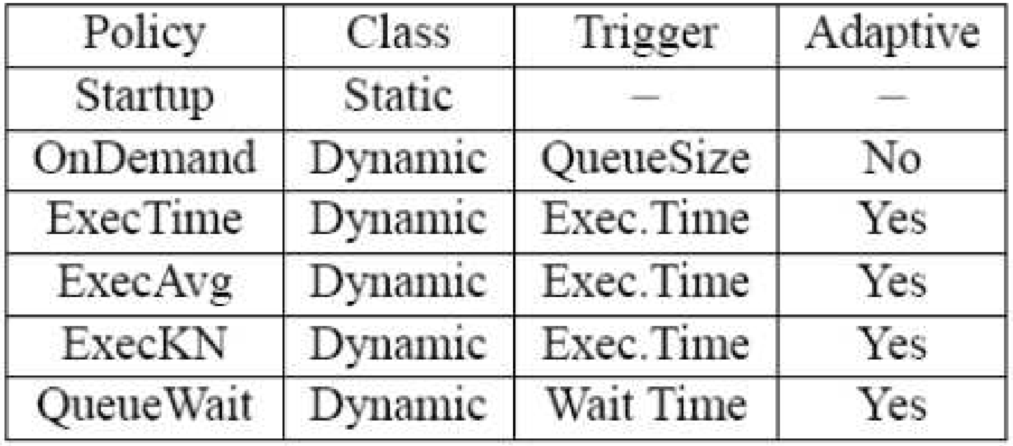
1. Experimental setup
2. Experimental results
3. Ad for Bogdan’s lecture (next)

Q3

Provisioning and Allocation Policies*

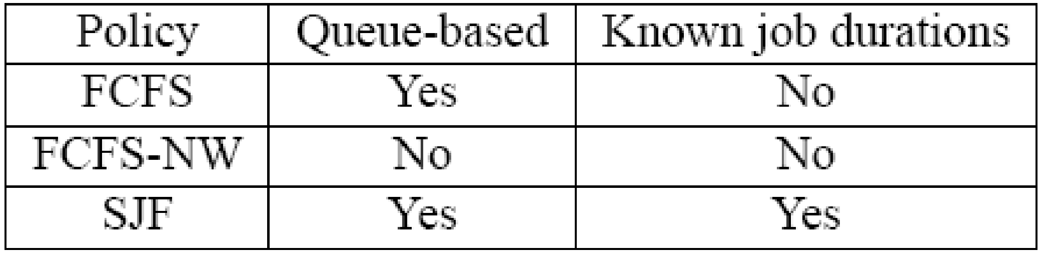
* For User-Level Scheduling

- Provisioning



Policy	Class	Trigger	Adaptive
Startup	Static	–	–
OnDemand	Dynamic	QueueSize	No
ExecTime	Dynamic	Exec.Time	Yes
ExecAvg	Dynamic	Exec.Time	Yes
ExecKN	Dynamic	Exec.Time	Yes
QueueWait	Dynamic	Wait Time	Yes

- Allocation



Policy	Queue-based	Known job durations
FCFS	Yes	No
FCFS-NW	No	No
SJF	Yes	Yes

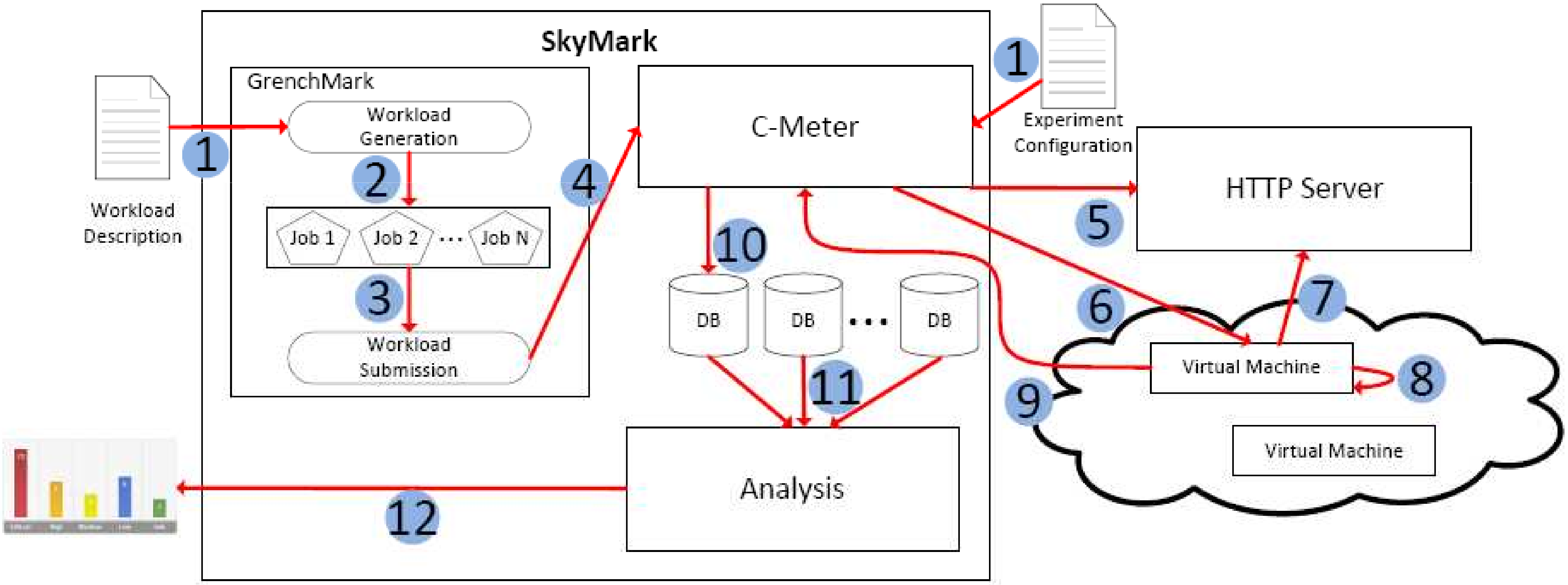
- Also looked at combined Provisioning + Allocation policies

The SkyMark Tool for IaaS Cloud Benchmarking

Villegas, Antoniou, Sadjadi, Iosup. An Analysis of Provisioning and Allocation Policies for Infrastructure-as-a-Service Clouds, CCGrid 2012

Q3

Experimental Tool: SkyMark



Provisioning and Allocation policies steps 6+9, and 8, respectively

Delft University of Technology

Villegas, Antoniou, Sadjadi, Iosup. An Analysis of Provisioning and Allocation Policies for Infrastructure-as-a-Service Clouds, PDS Tech.Rep.2011-009

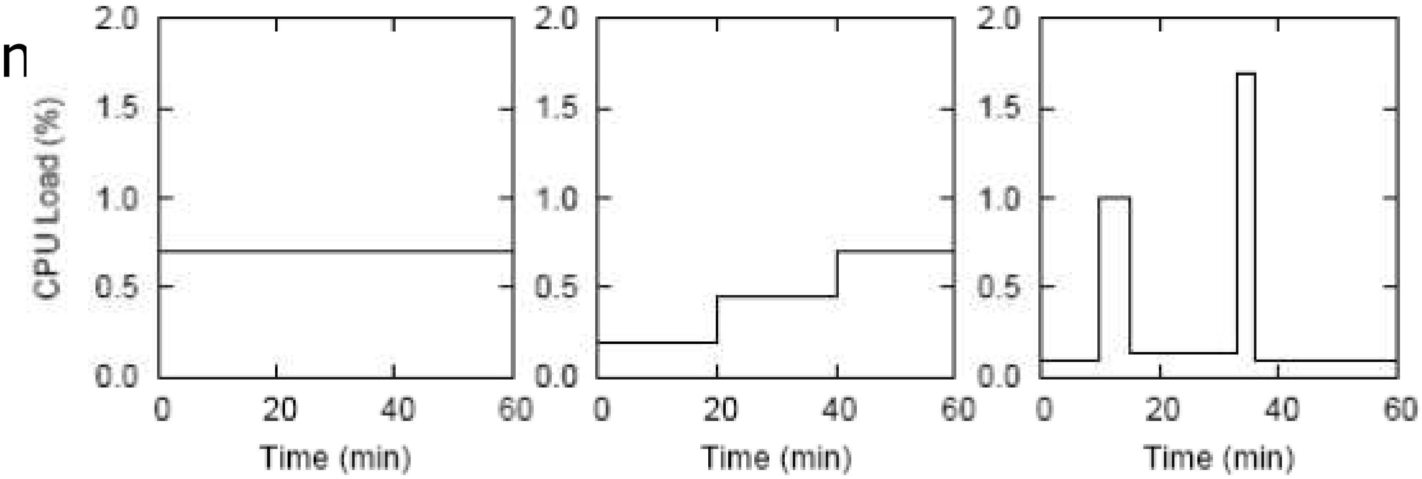
Q3

Experimental Setup (1)

- Environments
  - DAS4, Florida International University (FIU)
  - Amazon EC2

Workload Unit	CPU	Memory	I/O	Appears in
WU1	X			WL1
WU2		X		WL2,WL4
WU3			X	WL3,WL4

- Workloads
  - Bottleneck
  - Arrival pattern



Villegas, Antoniou, Sadjadi, Iosup. An Analysis of Provisioning and Allocation Policies for Infrastructure-as-a-Service Clouds, CCGrid2012 + PDS Tech.Rep.2011-009

Q3

Experimental Setup (2)

- **Performance Metrics**
  - Traditional: Makespan, Job Slowdown
  - Workload Speedup One (SU1)
  - Workload Slowdown Infinite (SUinf)

$$SU_1(W) = \frac{MS(W)}{\sum_{i \,\in\, W} t_R(i)}$$

$$SU_\infty(W) = \frac{MS(W)}{\max_{i \,\in\, W} t_R(i)}$$

- **Cost Metrics**
  - Actual Cost (Ca)
  - Charged Cost (Cc)

$$C_a(W) = \sum_{i \,\in\, leased\ VMs} t_{stop}(i) - t_{start}(i)$$

$$C_c(W) = \sum_{i \,\in\, leased\ VMs} \lceil t_{stop}(i) - t_{start}(i) \rceil$$

- **Compound Metrics**
  - Cost Efficiency (Ceff)
  - Utility

$$C_{eff}(W) = \frac{C_c(W)}{C_a(W)}$$

$$U(W) = \frac{SU_1(W)}{C_c(W)}$$

Villegas, Antoniou, Sadjadi, Iosup. An Analysis of Provisioning and Allocation Policies for Infrastructure-as-a-Service Clouds, CCGrid 2012