A TU Delft View on Big Data Processing and Preservation

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, **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, Orna Ben-Yehuda (Technion), Ted Willke (Intel), Claudio Martella (Giraph), Ana Lucia Varbanescu (UvA, NL)…

June 4, 2013

Technion, Haifa, Israel

TU Delft

Delft University of Technology

Lectures at the Technion Computer Engineering Center (TCE), Haifa, IL

IaaS Cloud Benchmarking	May 7
Massivizing Online Social Games	May 9
Gamification in Higher Education	May 27
Lectures at IBM Haifa, Intel Haifa	June 2,3
Scheduling in IaaS Clouds	Actually, HUJI June 5

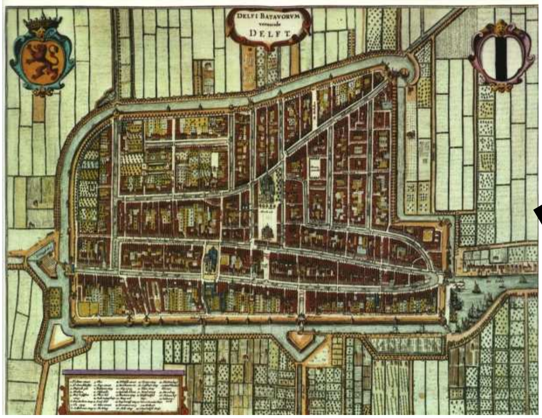
A TU Delft perspective on Big Data Processing and Preservation **June 6** **10am Taub 337**

Grateful to Orna Agmon Ben-Yehuda, Assaf Schuster, Isaac Keslassy.

Also thankful to Bella Rotman and Ruth Boneh.

TU Delft

(TU) Delft – the Netherlands – Europe



founded 13th century
pop: 100,000

founded 1842
pop: 13,000

pop: 16.5 M

pop.: 100,000 (We are here) אנחנו כאן

TU Delft

The Parallel and Distributed Systems Group at TU Delft

VENI

Alexandru Iosup

Grids/Clouds
P2P systems
Big Data
Online gaming

Dick Epema

Grids/Clouds
P2P systems
Video-on-demand
e-Science

VENI

Ana Lucia Varbanescu

HPC systems
Multi-cores
Big Data
e-Science

Henk Sips

HPC systems
Multi-cores
P2P systems

VENI

Johan Pouwelse

P2P systems
File-sharing
Video-on-demand

Home page

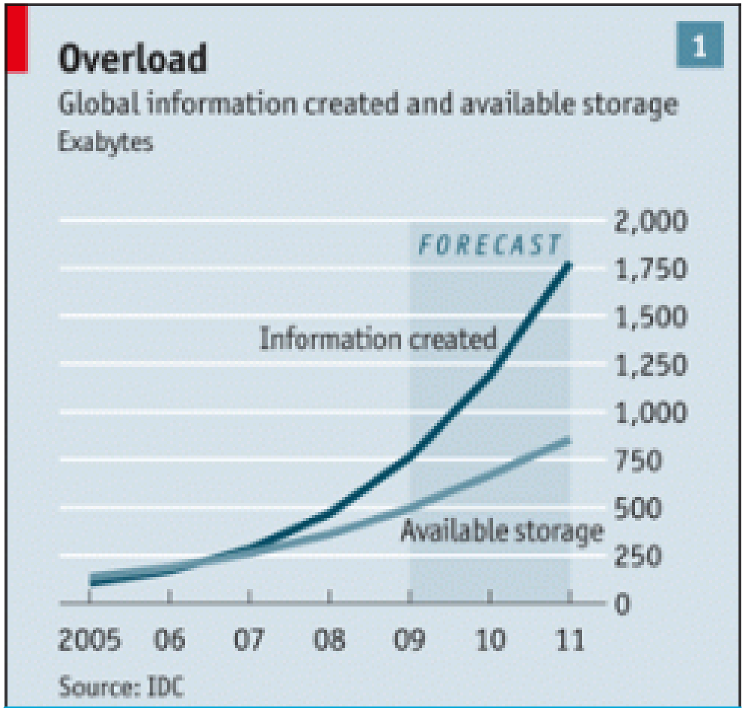
- www.pds.ewi.tudelft.nl

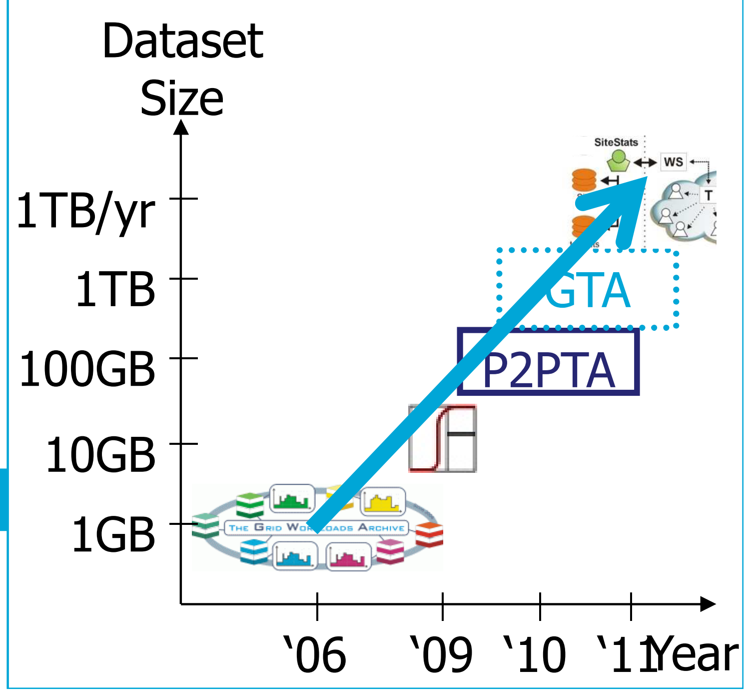
Publications

- see PDS publication database at publications.st.ewi.tudelft.nl

The Data Deluge

- **All human knowledge**
  - Until 2005: 150 Exa-Bytes
  - 2010: 1,200 Exa-Bytes
- **Online gaming (Consumer)**
  - 2002: 20TB/year/game
  - 2008: 1.4PB/year/game (only stats)
- **Public archives (Science)**
  - 2006: GBs/archive
  - 2011: TBs/year/archive





The Data Deluge
The Professional World Gets Connected

Feb 2012

100M Mar 2011, 69M May 2010

Source: Vincenzo Cosenza, The State of LinkedIn,
http://vincos.it/the-state-of-linkedin/

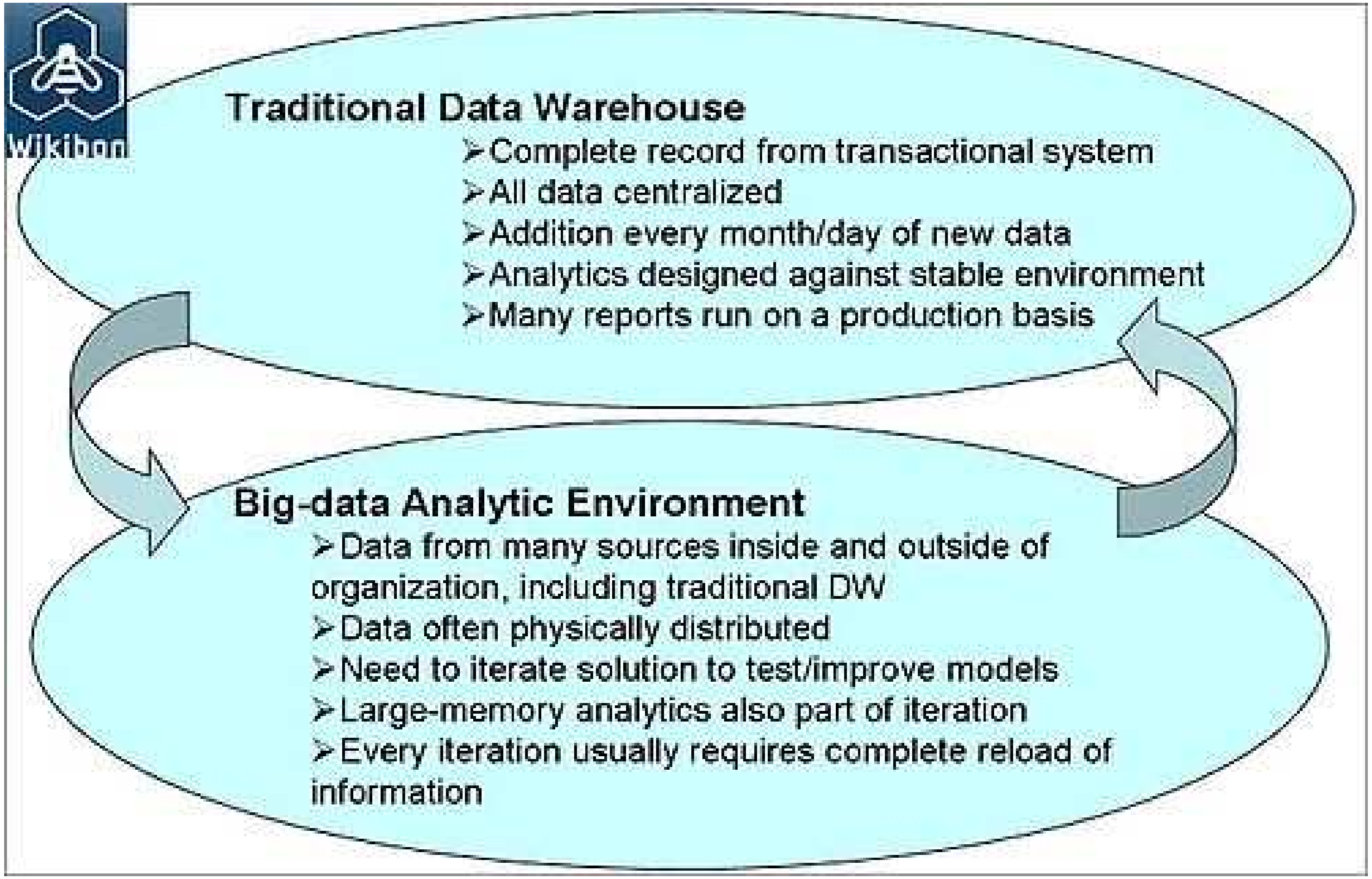
The Three "V"s of Big Data
When you can, keep *and* process everything

- Volume
  - More data vs. better models
  - Data grows exponentially + iterative mod
  - 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

Too big, too fast, does not comply with traditional DB

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

Data Warehouse vs. Big Data

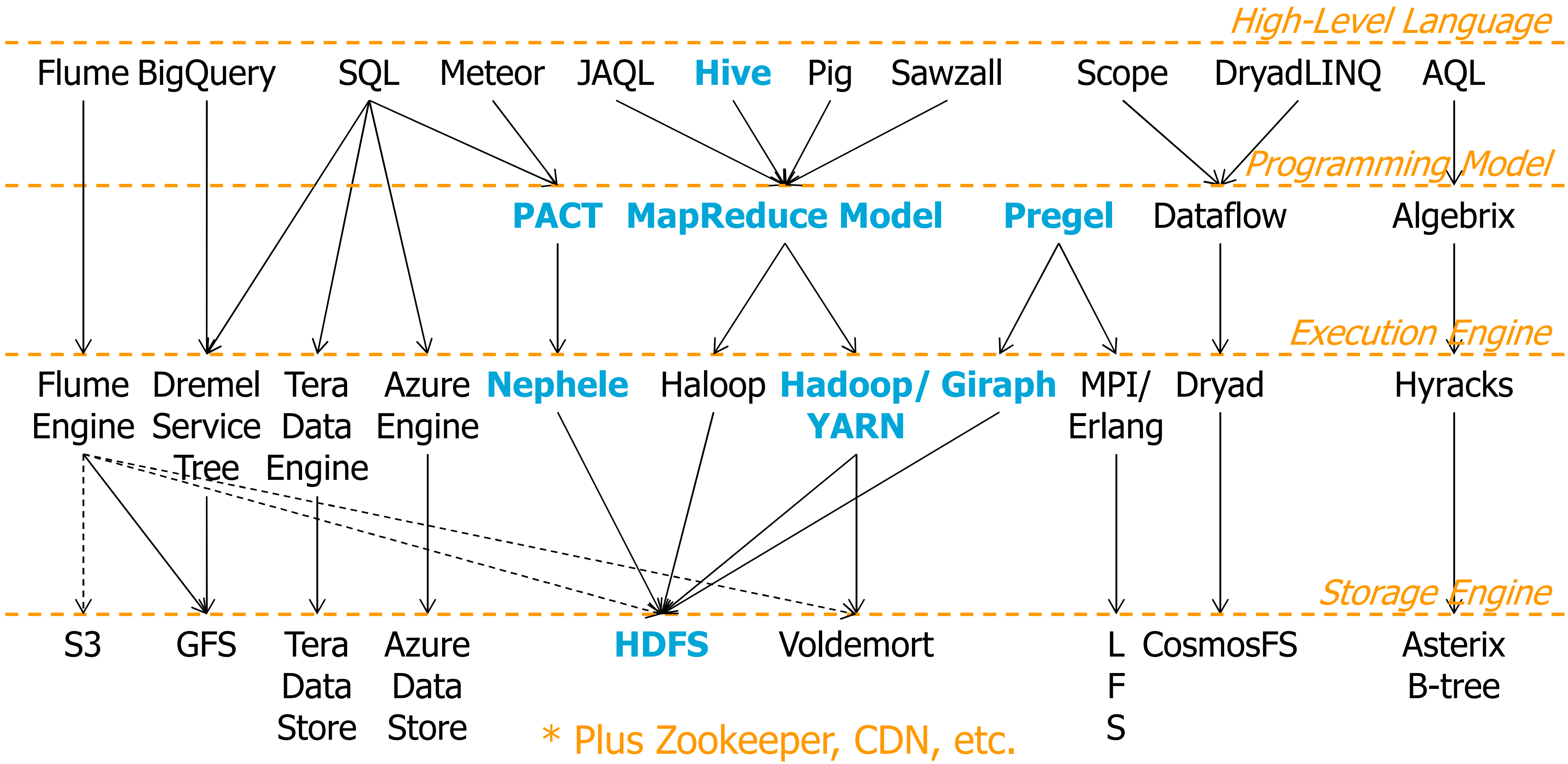


Source: http://wikibon.org/

Agenda

1. Introduction to Big Data
2. **Programming Models for Big Data**
3. PDS Group Work on Big Data
4. Summary

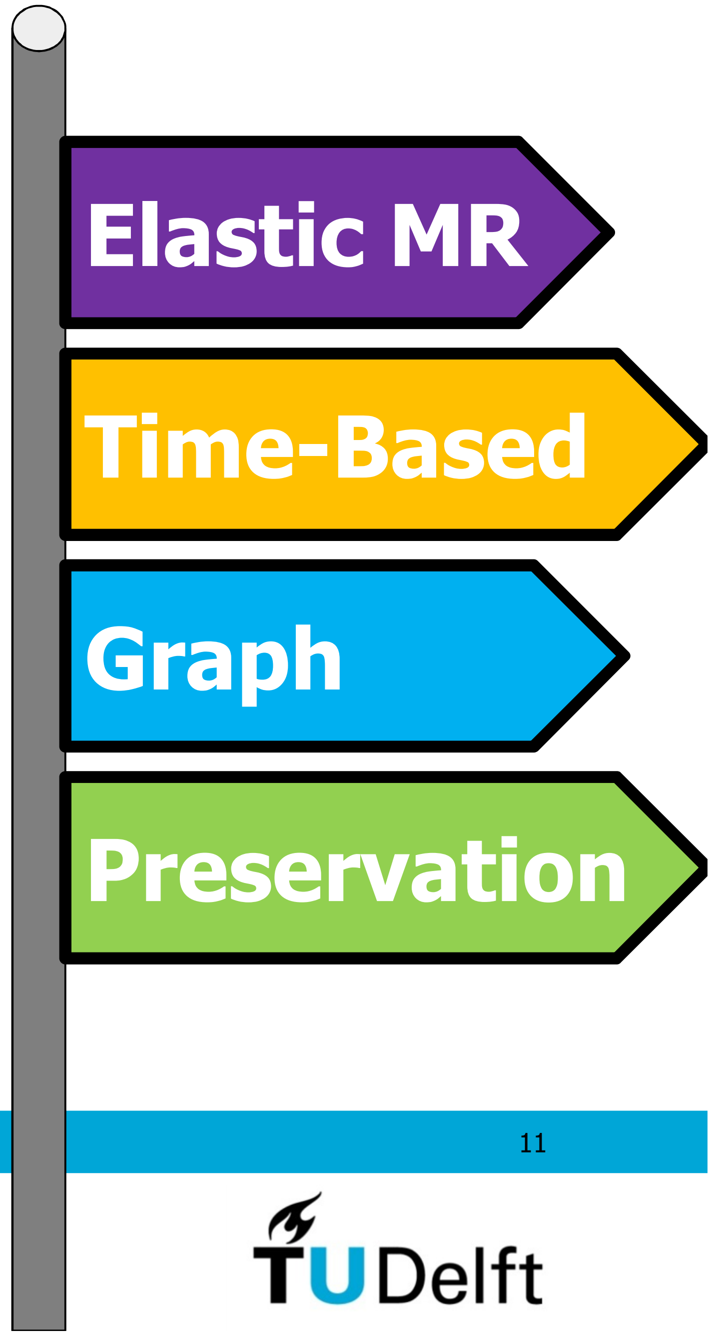
Programming Models for Big Data: Systems of Systems (Why Big Data is Difficult)



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

Agenda

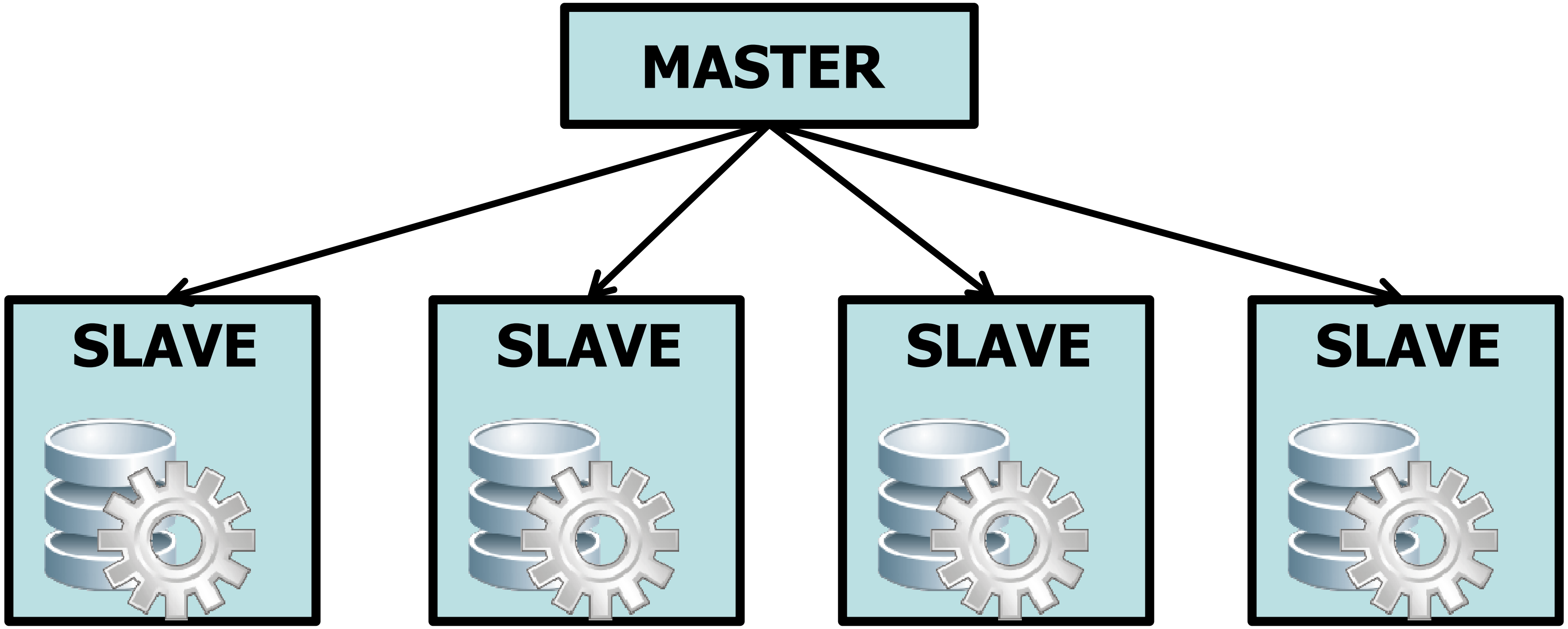
1. Introduction
2. Programming Models for Big Data
3. **PDS Group Work on Big Data**
    1. **MapReduce:**
       Elastic MR and
       Time-Based Analytics
    2. Graph Processing
    3. Preservation
4. Summary



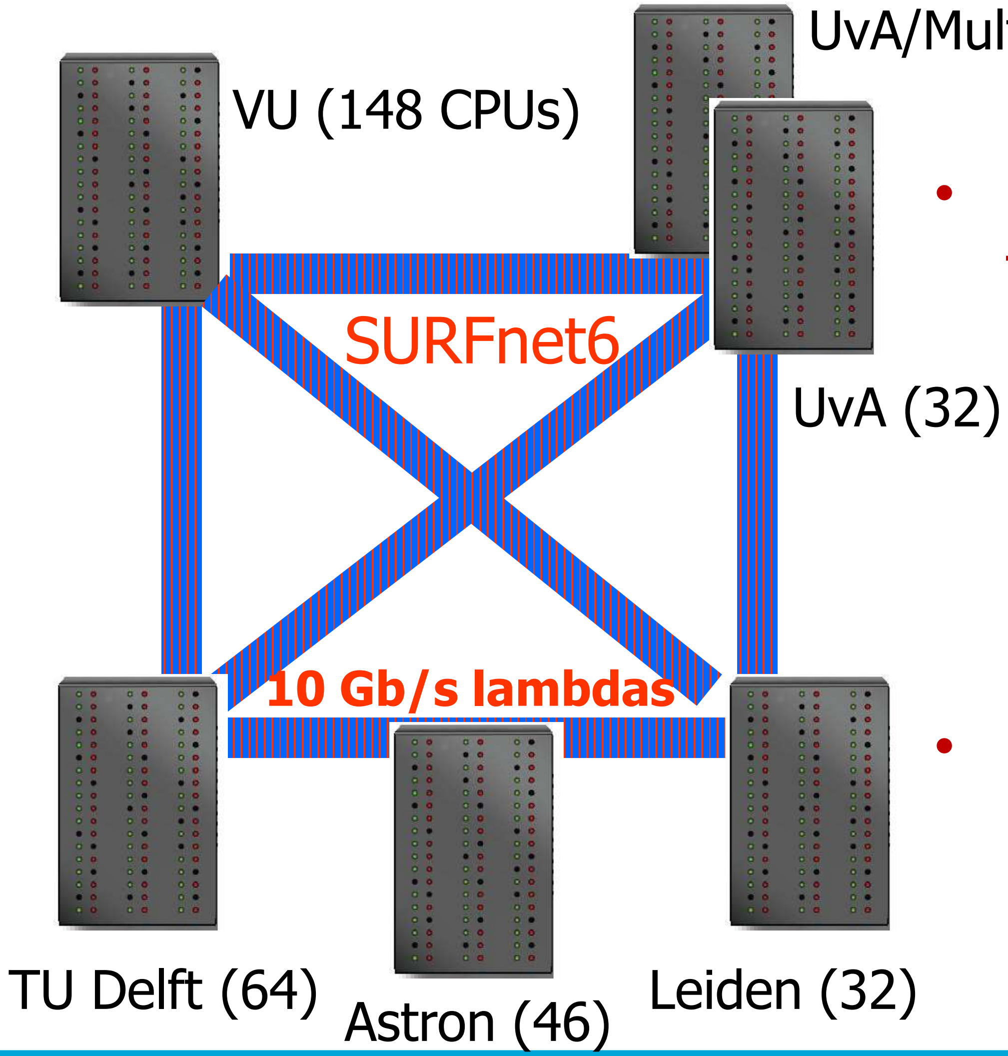
TU Delft

MapReduce Overview

- MR cluster
  - Large-scale data processing
  - Master-slave paradigm
- Components
  - Distributed file system (storage)
  - MapReduce framework (processing)



The DAS-4 Infrastructure



- Used for research in systems for over a decade
  - 1,600 cores (quad cores)
  - 2.4 GHz CPUs, GPUs
  - 180 TB storage
  - 10 Gbps Infiniband
  - 1 Gbps Ethernet
- Koala grid scheduler

The Koala Grid Scheduler

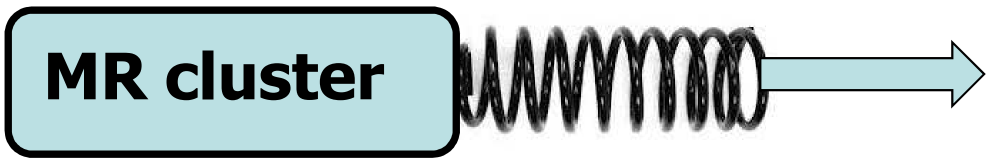
TU Delft

Why Dynamic MapReduce Clusters?

- Improve resource utilization
  - Grow when the workload is too heavy
  - Shrink when resources are idle

- Fairness across multiple MR clusters
  - Redistribute idle resources
  - Allocate resources for new MR clusters

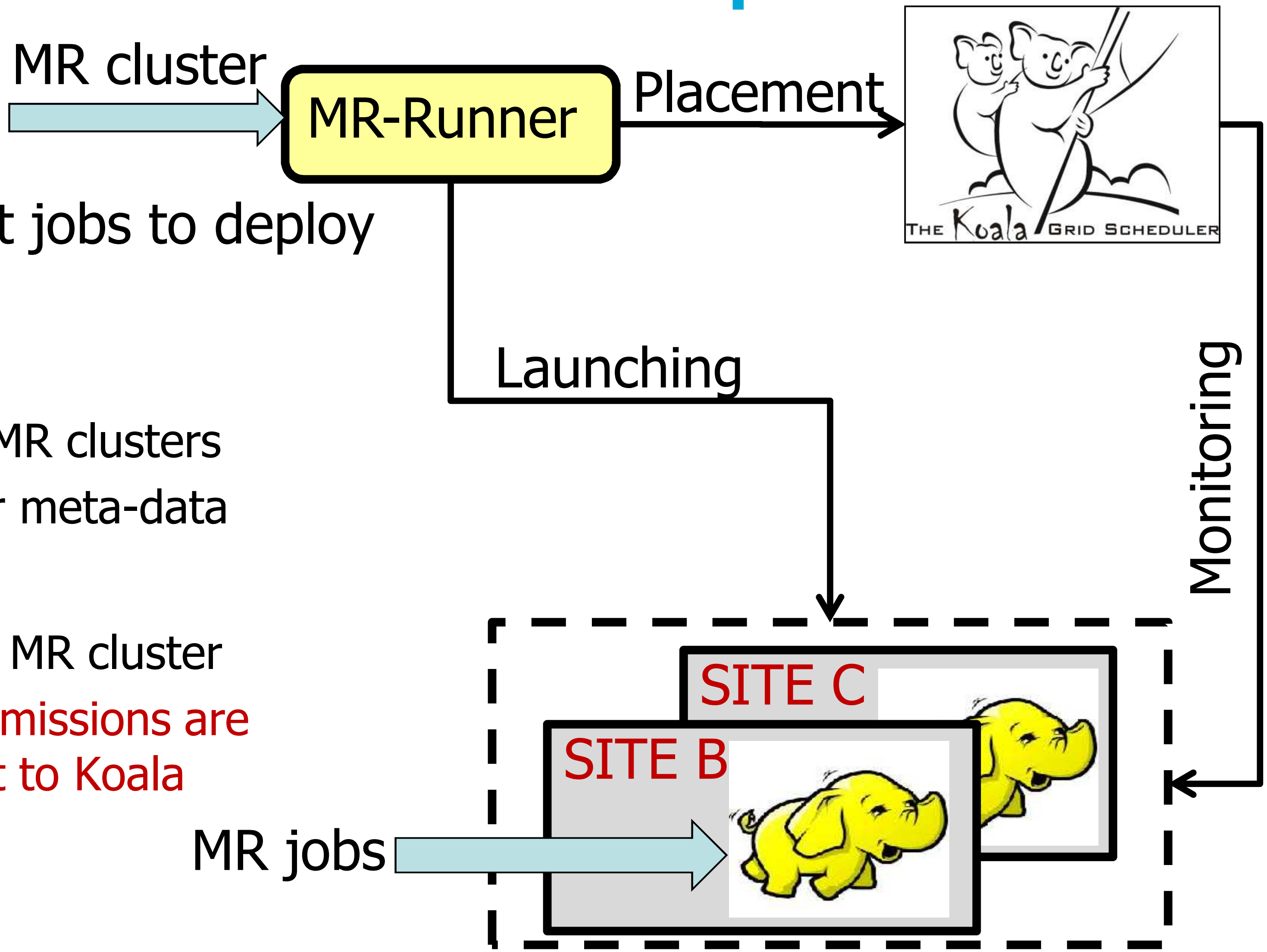
Isolation
- Performance
- Failure
- Data
- Version



Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

KOALA Grid Scheduler and MapReduce

- Users submit jobs to deploy MR clusters
- Koala
  - ➢ Schedules MR clusters
  - ➢ Stores their meta-data
- MR-Runner
  - ➢ Installs the MR cluster
  - ➢ MR job submissions are transparent to Koala

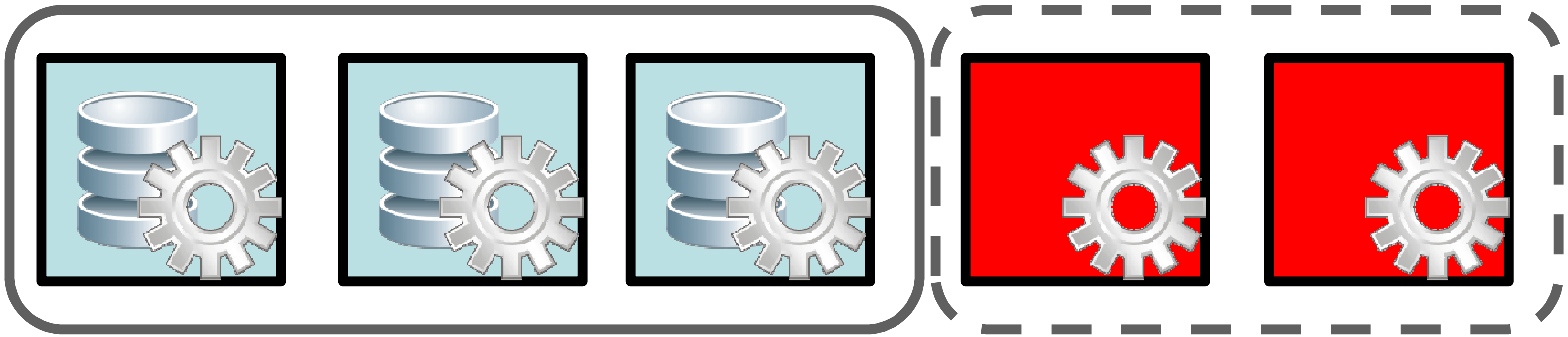


Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

TU Delft

System Model

- Two types of nodes
  - Core nodes: TaskTracker and DataNode
  - Transient nodes: only TaskTracker

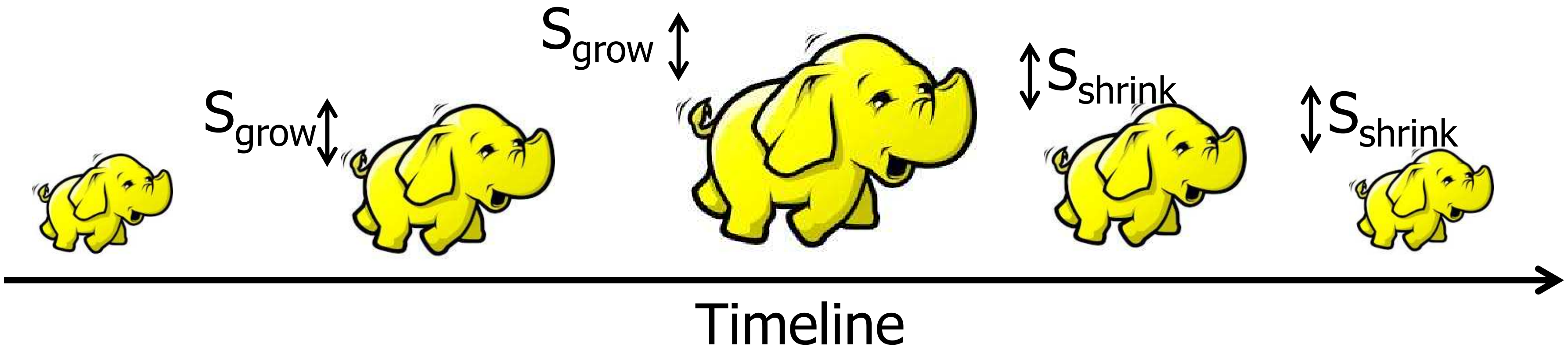


Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

Resizing Mechanism

- Two-level provisioning
  - Koala makes resource offers / reclaims
  - MR-Runners accept / reject request

- Grow-Shrink Policy (GSP)
  - MR cluster utilization: $$F_{\min} \leq \frac{totalTasks}{availSlots} \leq F_{\max}$$
  - Size of grow and shrink steps: $\mathbf{S_{grow}}$ and $\mathbf{S_{shrink}}$

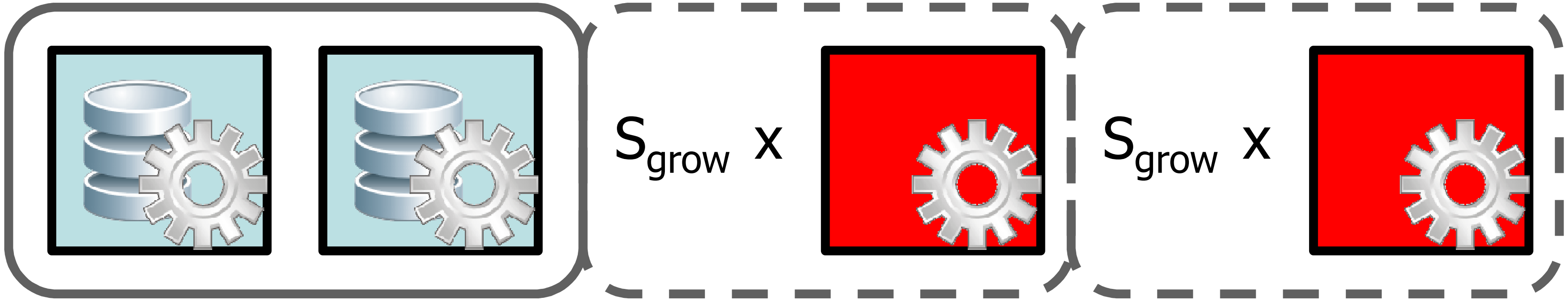


Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

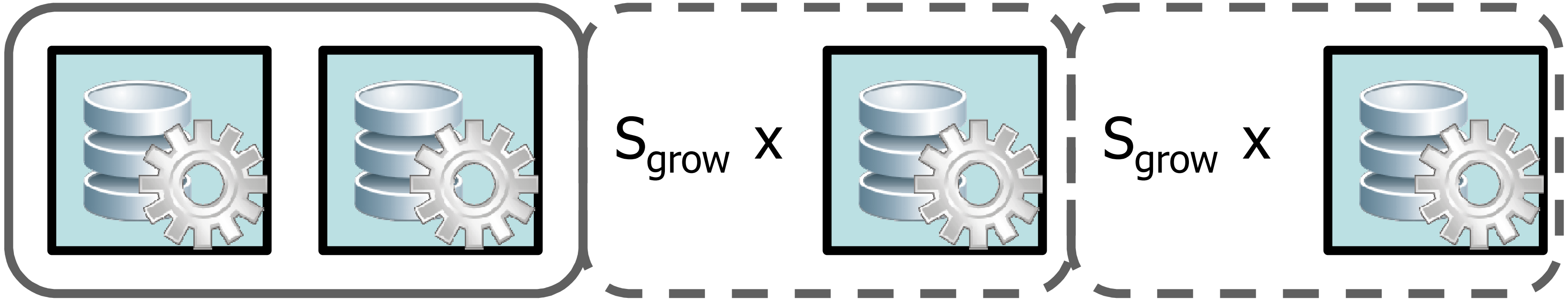
TU Delft

Baseline Policies

- Greedy-Grow Policy (GGP)—only grow with transient nodes:



- Greedy-Grow-with-Data Policy (GGDP)—grow, core nodes:



Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

Setup

- *98% of jobs @ Facebook take less than a minute*
- *Google reported computations with TB of data*

- DAS-4
- Two applications: Wordcount and Sort

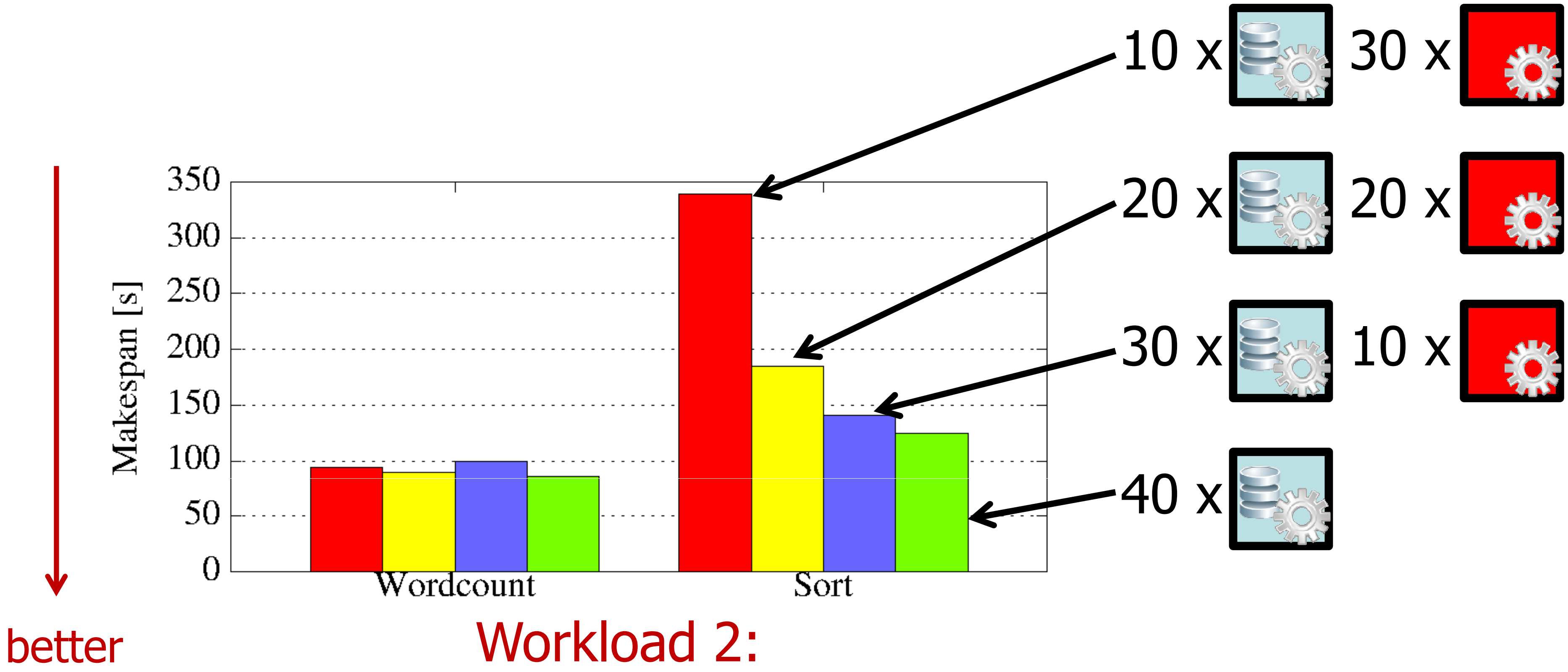
Workload 1
- Single job
- 100 GB
- Makespan

Workload 2
- Single job
- 40 GB, 50 GB
- Makespan

Workload 3
- Stream of 50 jobs
- 1 GB → 50 GB
- Average job execution time

Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

Transient Nodes



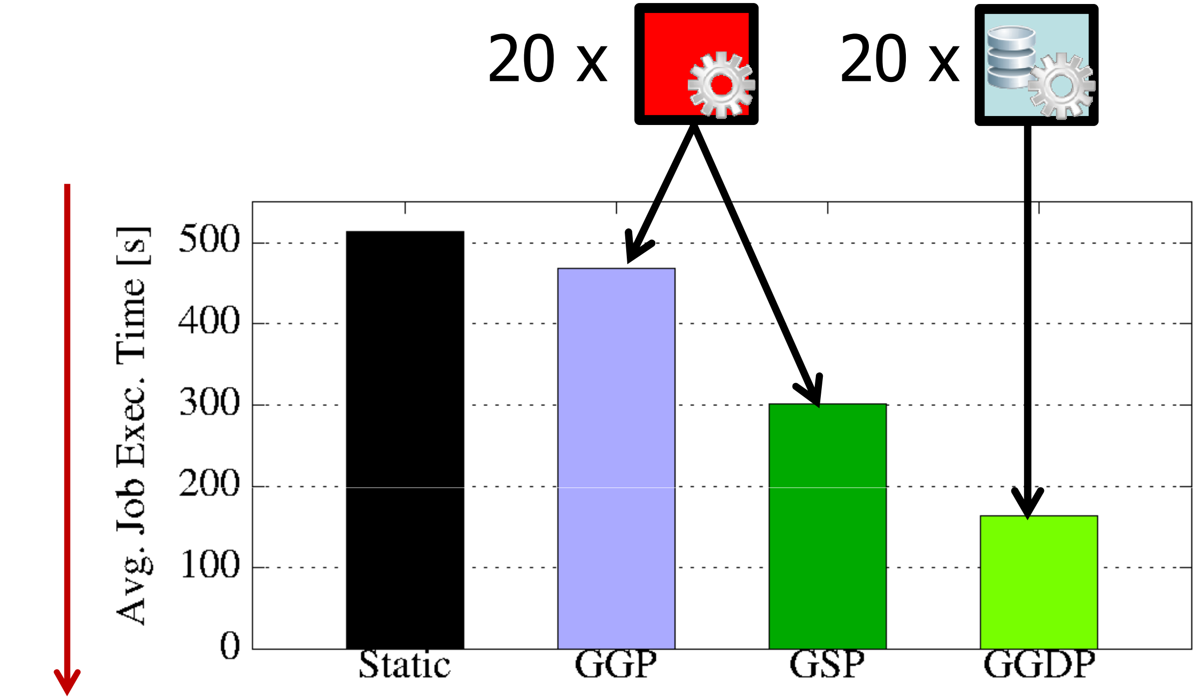
Workload 2:
40GB, 50GB

- Wordcount scales better than Sort on transient nodes

Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

TUDelft

Performance of Resizing using Static, Transient, and Core Nodes



Workload 3 => 20 x

(50 jobs, 1GB to 50GB)

- Resizing bounds

  $F_{min} = 0.25$

  $F_{max} = 1.25$

- Resizing steps
  - GSP

    $S_{grow} = 5$

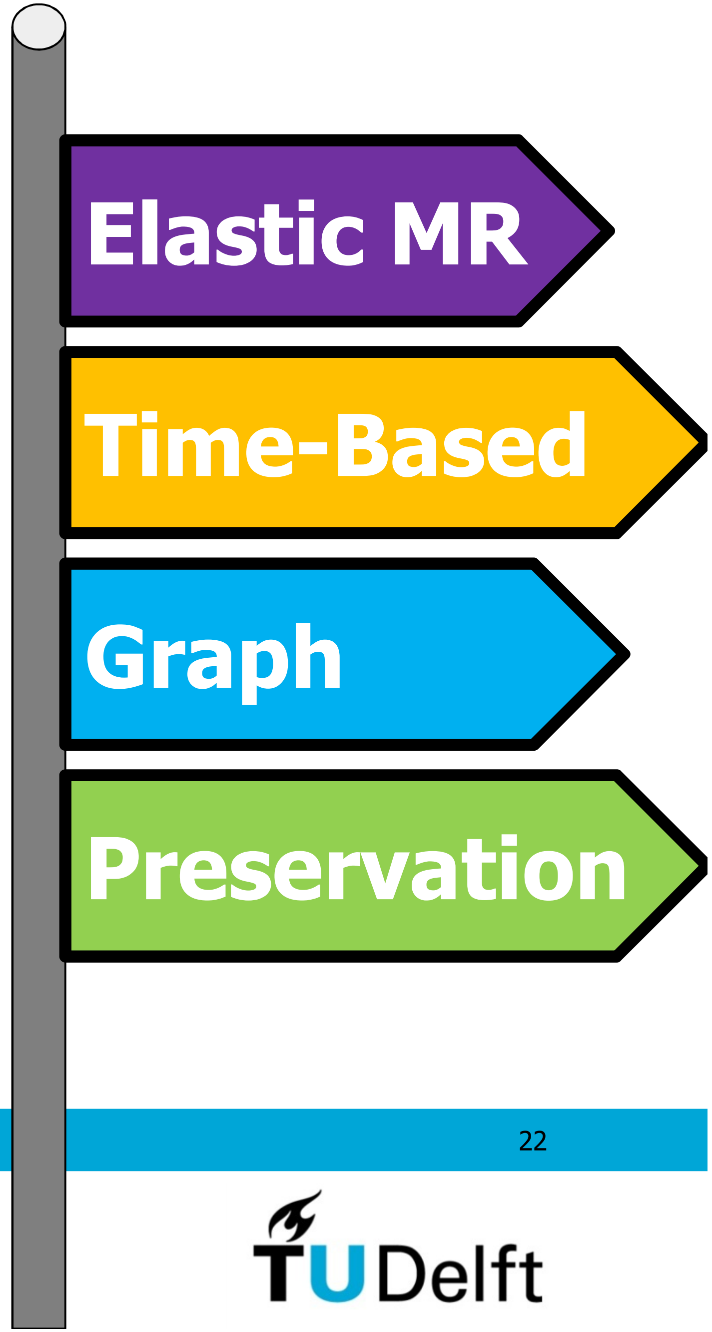
    $S_{shrink} = 2$
  - GG(D)P

    $S_{grow} = 2$

Ghit and Epema. Resource Management for Dynamic MapReduce Clusters in Multicluster Systems. MTAGS 2012. Best Paper Award.

Agenda

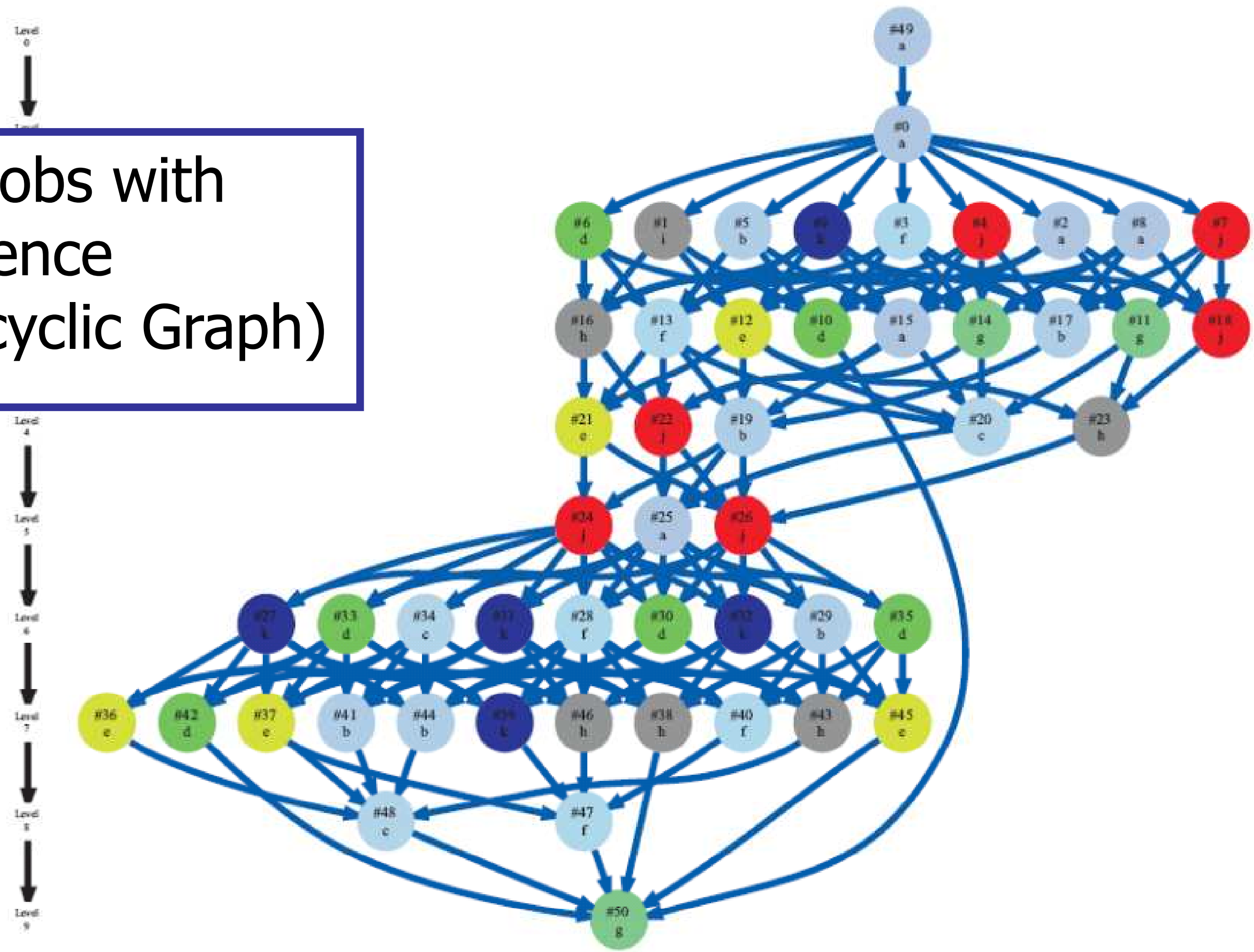
1. Introduction
2. Programming Models for Big Data
3. **PDS Group Work on Big Data**
   1. **MapReduce:**
      Elastic MR and
      Time-Based Analytics
   2. Graph Processing
   3. Preservation
4. Summary



TU Delft

What is a Wokflow?

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



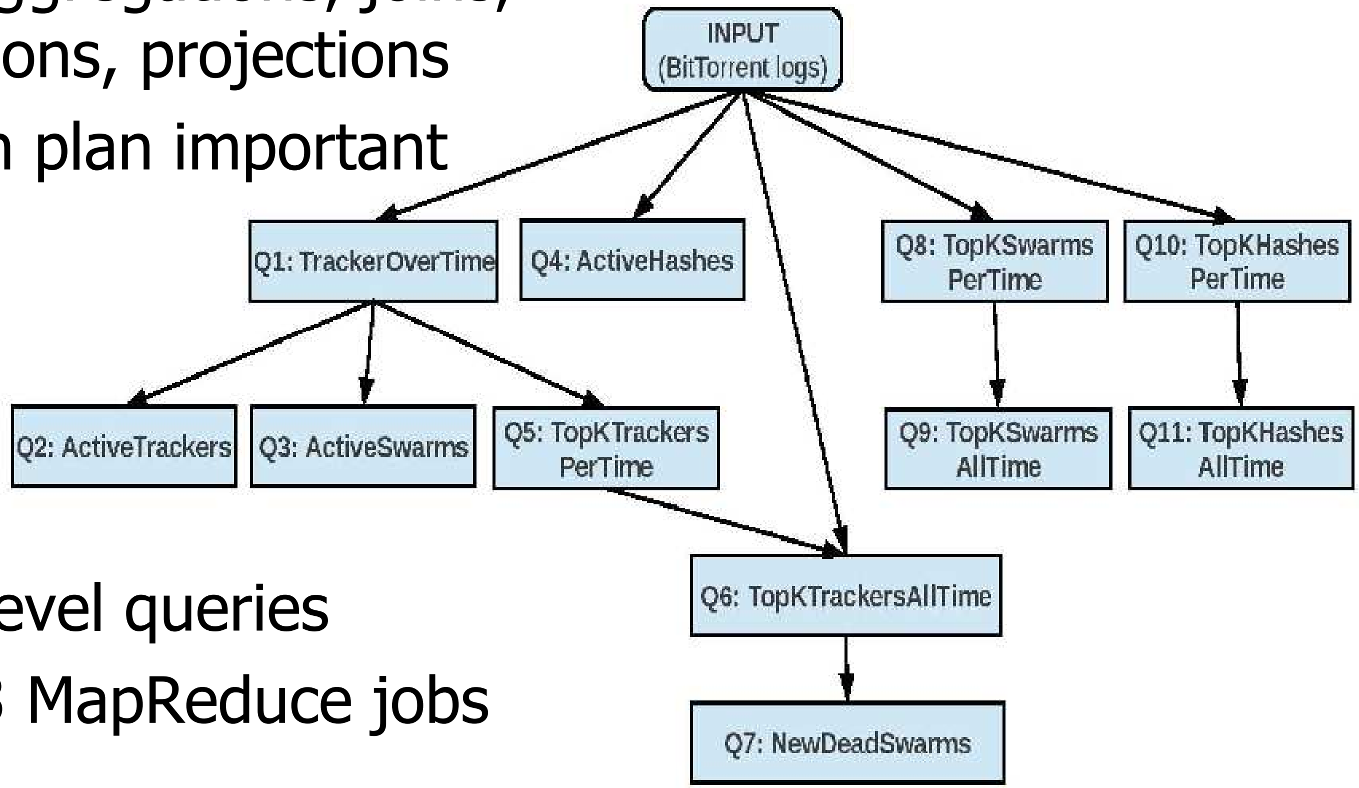
TU Delft

The BTWorld Project
“Observe the Global BitTorrent Network”

- Started 2009
  - Collect data from 1,000s of trackers
  - Over 30M shared files (swarms)
  - Over 100M BT clients
- Data set
  - 15TB of stored multi-files, 1 file/tracker/sample
  - Timestamped, multi-record files
    - Hash: unique id for file
    - Tracker: unique id for tracker
    - Information per file: seeders, leechers

Queries for the BTWorld Project

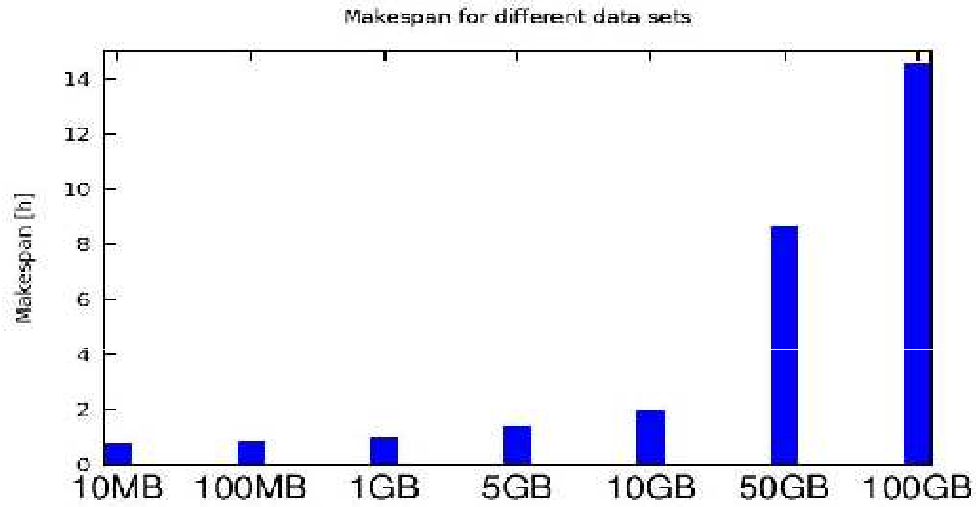
- Non-trivial algorithms
  - SQL aggregations, joins, selections, projections
- Execution plan important

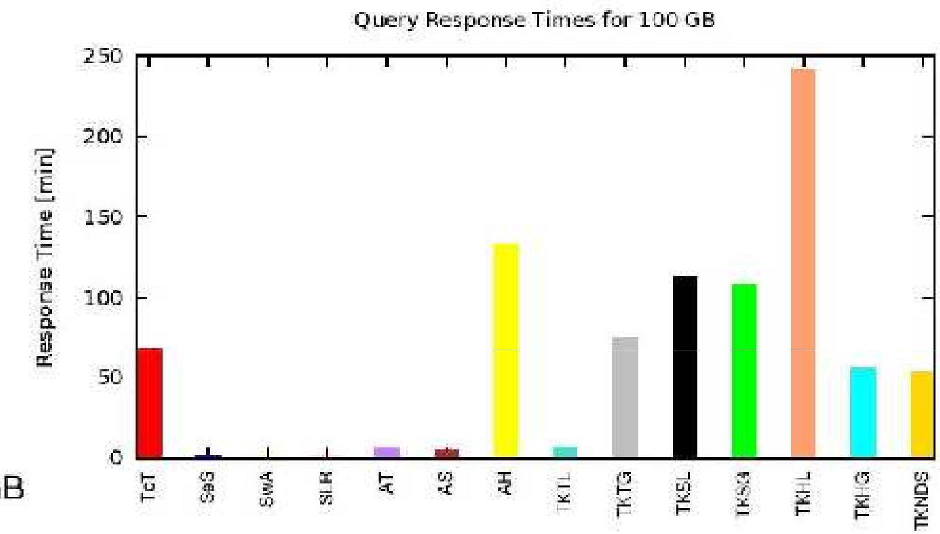


- 14 high-level queries
- Pig -> 33 MapReduce jobs

TUDelft

Preliminary results



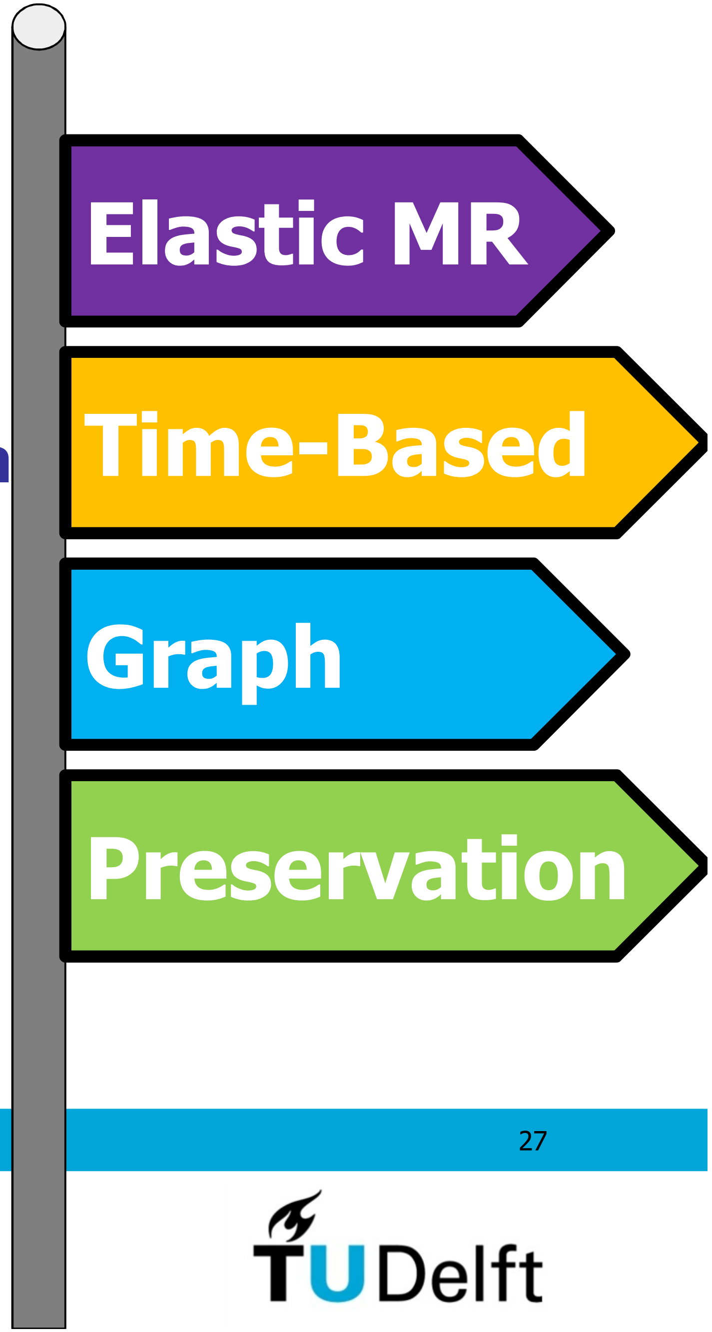


- (left) Up to 100GB: 14 hours for the workflow
- (right) Large variation in query response time
- Also profiled resource utilization (CPU, memory, disk, …)

TU Delft

Agenda

1. Introduction
2. Programming Models for Big Data
3. **PDS Gruoup Work on Big Data**
    1. MapReduce
    2. **Graph Processing**
    3. Preservation
4. Summary



TU Delft

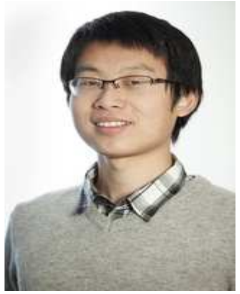
Big Data/Graph Processing: Our Team

Alexandru Iosup
TU Delft

Cloud Computing
Gaming Analytics
Performance Eval.
Benchmarking
Variability

Ana Lucia Varbanescu
UvA

Parallel Computing
Multi-cores/GPUs
Performance Eval.
Benchmarking
Prediction



Yong Guo
TU Delft

Cloud Computing
Gaming Analytics
Performance Eval.
Benchmarking

Marcin Biczak
TU Delft

Cloud Computing
Performance Eval.
Development

Graphitti

http://www.pds.ewi.tudelft.nl/graphitti/

Claudio Martella
VU Amsterdam
All things Giraph

Ted Willke
Intel Corp.
All things graph-processing

Consultant for the project.
Not responsible for issues related to this work. Not representing official products and/or company views.

TU Delft

Why "How Well do Graph-Processing Platforms Perform?"

- Large-scale graphs exists in a wide range of areas: social networks, website links, online games, etc.

- Large number of **platforms** available to developers
  - Desktop: Neo4J, SNAP, etc.
  - Distributed: Giraph, GraphLab, etc.
  - Parallel: too many to mention

Problem: Large differences in performance profiles across different graph-processing **algorithms** and **data sets**

Some Previous Work

Graph500.org: BFS on synthetic graphs

Performance evaluation in graph-processing (limited algorithms and graphs)

- Hadoop does not perform well [Warneke09]
- Graph partitioning improves the performance of Hadoop [Kambatla12]
- Trinity outperforms Giraph in BFS [Shao12]
- Comparison of graph databases [Dominguez-Sal10]

Performance comparison in other applications

- Hadoop vs parallel DBMSs: grep, selection, aggregation, and join [Pavlo09]
- Hadoop vs High Performance Computing Cluster (HPCC): queries [Ouaknine12]
- Neo4j vs MySQL: queries [Vicknair10]

Problem: Large differences in performance profiles across different graph-processing **algorithms** and **data sets**

Our Method

A benchmarking suite for the performance evaluation of graph-processing platforms

1. Multiple Metrics, e.g.,
    - Execution time
    - Normalized: EPS, VPS
    - Utilization

2. Representative graphs with various characteristics, e.g.,
    - Size, Density
    - Directedness

3. Typical graph algorithms, e.g.,
    - BFS
    - Connected components
    - …

http://bit.ly/10hYdIU

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

Benchmarking suite
Data sets

Graphs	# V	# E	d ($\times 10^{-5}$)	$\bar{D}$	Size	Directivity
Amazon	262.1 K	1.2 M	1.8	4.7	18 MB	directed
WikiTalk	2.4 M	5.0 M	0.1	2.1	87 MB	directed
KGS	293.3 K	16.6 M	38.5	112.9	210 MB	undirected
Citation	3.8 M	16.5 M	0.1	4.4	297 MB	directed
DotaLeague	61.2 K	50.9 M	2,719.0	1,663.2	655 MB	undirected
Synth	2.4 M	64.2 M	2.2	53.6	964 MB	undirected
Friendster	65.6 M	1.8 B	0.1	55.1	31 GB	undirected

Graph500
http://www.graph500.org/

The Game Trace Archive
http://gta.st.ewi.tudelft.nl/

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis

Benchmarking Suite
Algorithm classes

1. General Statistics (STATS: # vertices and edges, LCC)
2. Breadth First Search (BFS)
3. Connected Component (CONN)
4. Community Detection (COMM)
5. Graph Evolution (EVO)

Benchmarking suite
Platforms and Process

- Platforms

YARN

Giraph

- Process
  - Evaluate baseline (out of the box) and tuned performance
  - Evaluate performance on fixed-size system
  - Future: evaluate performance on elastic-size system
  - Evaluate scalability

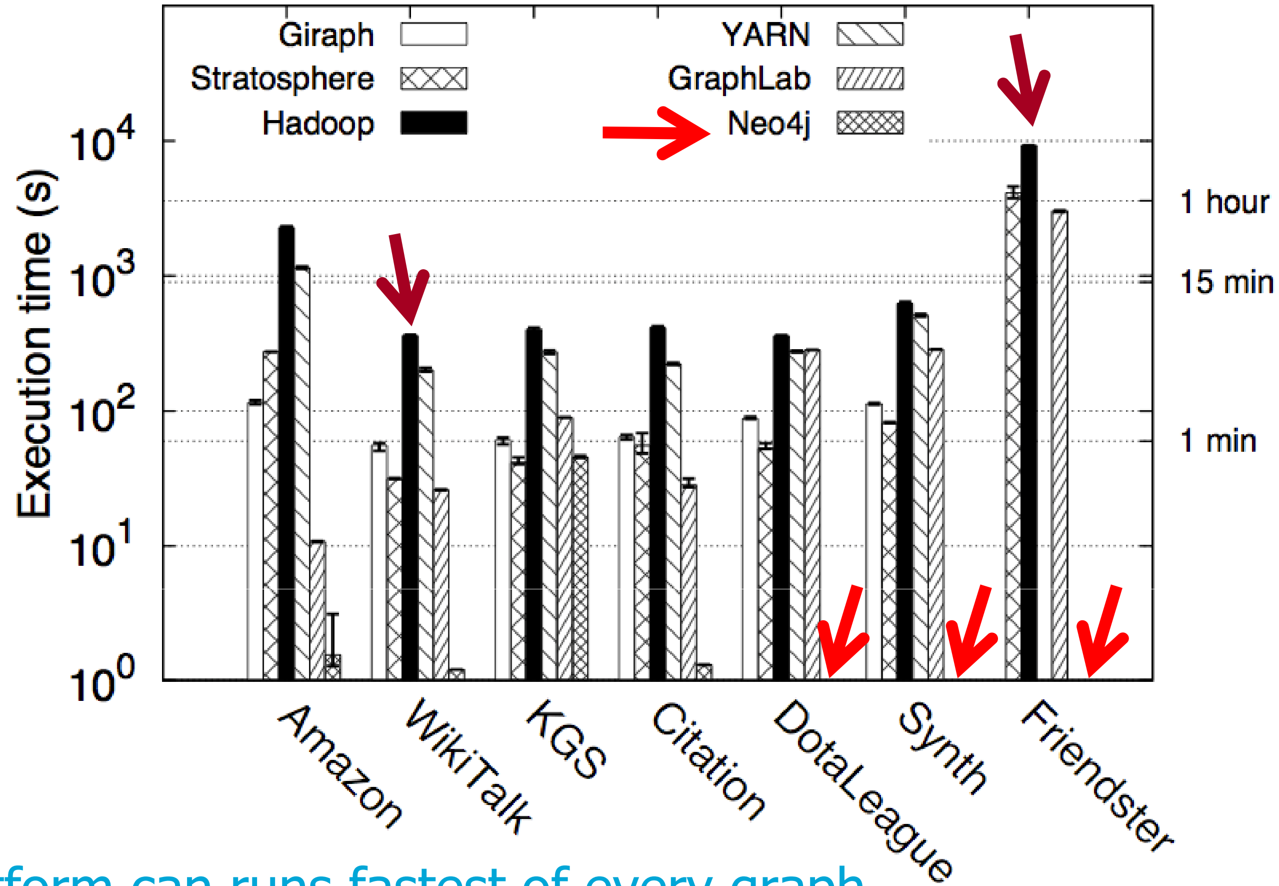
Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

Experimental setup

- Size
  - Most experiments take 20 working nodes
  - Up to 50 working nodes
- DAS4: a multi-cluster Dutch grid/cloud
  - Intel Xeon 2.4 GHz CPU (dual quad-core, 12 MB cache)
  - Memory 24 GB
  - 10 Gbit/s Infiniband network and 1 Gbit/s Ethernet network
  - Utilization monitoring: Ganglia
- HDFS used here as distributed file systems

BFS: results for all platforms, all data sets



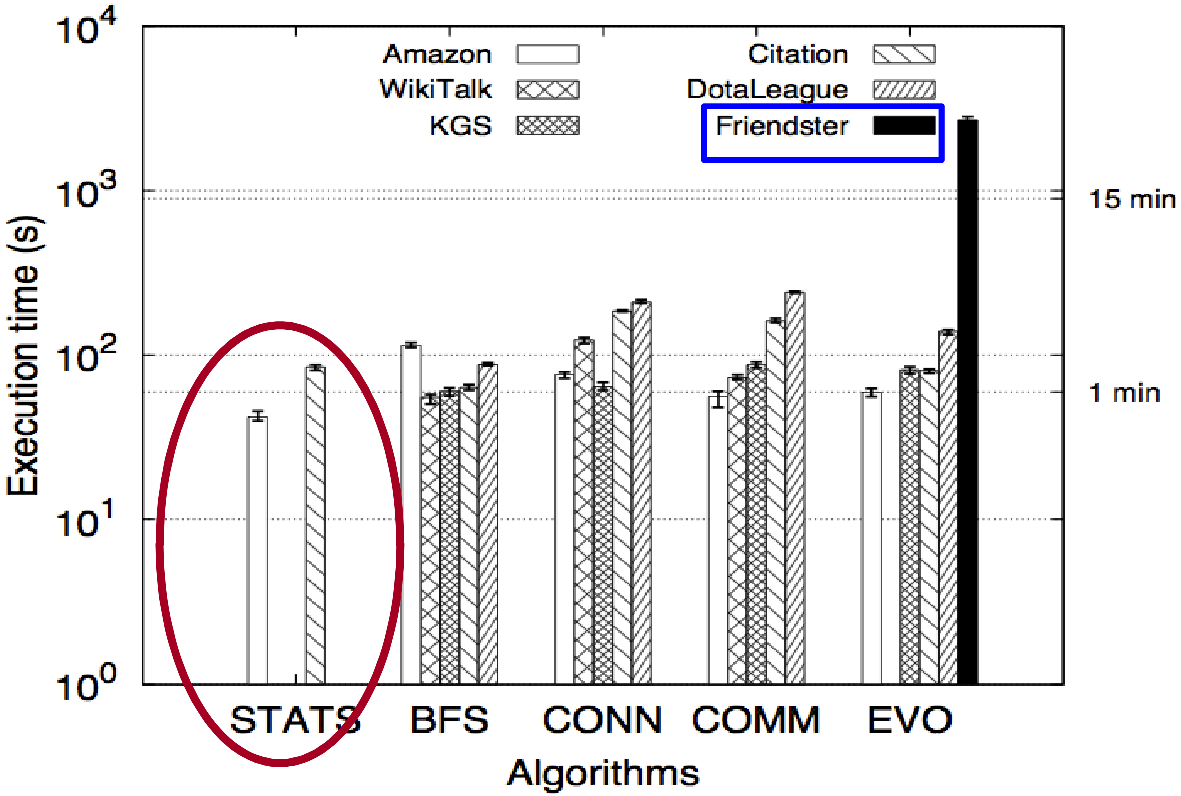
- No platform can runs fastest of every graph
- Not all platforms can process all graphs
- Hadoop is the worst performer

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

Giraph: results for all algorithms, all data sets

Q4

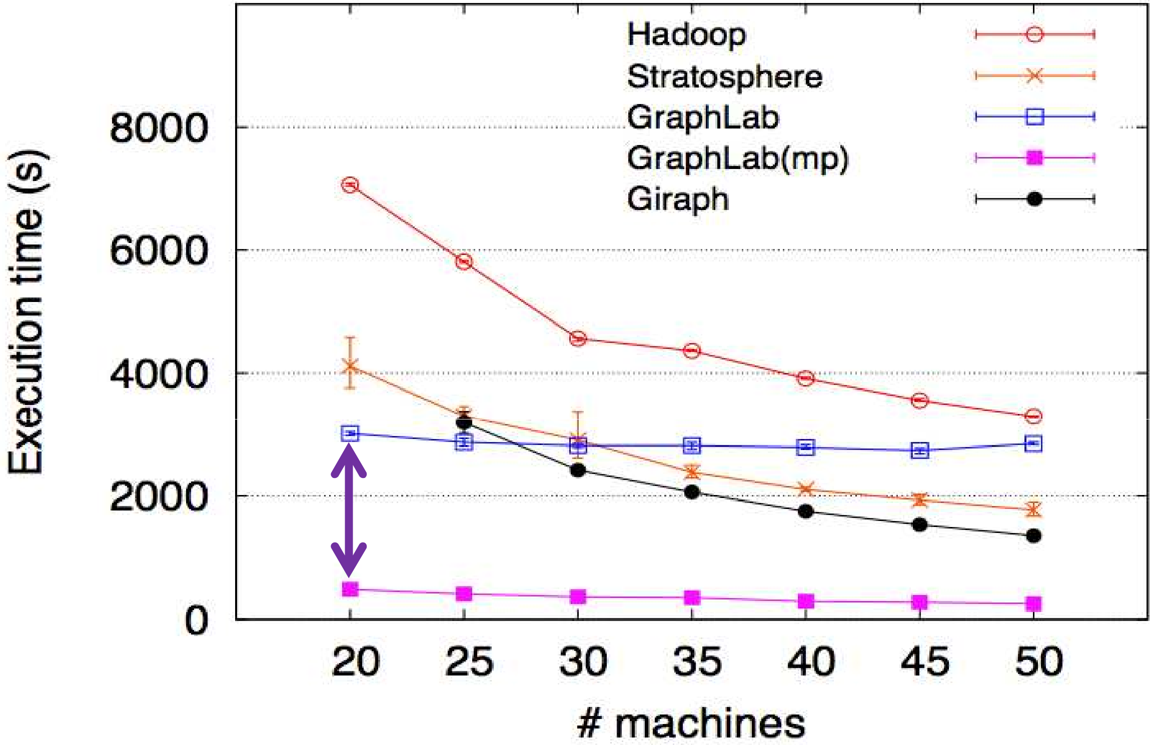


- Storing the whole graph in memory helps Giraph perform well
- Giraph may crash when graphs or messages become larger

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

Horizontal scalability: BFS on Friendster (31 GB)



- Using more computing machines can reduce execution time
- Tuning needed for horizontal scalability, e.g., for GraphLab, split input

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

Q4

Additional Overheads
Data ingestion time

- Data ingestion
  - Batch system: one ingestion, multiple processing
  - Transactional system: one ingestion, one processing

- Data ingestion matters even for batch systems

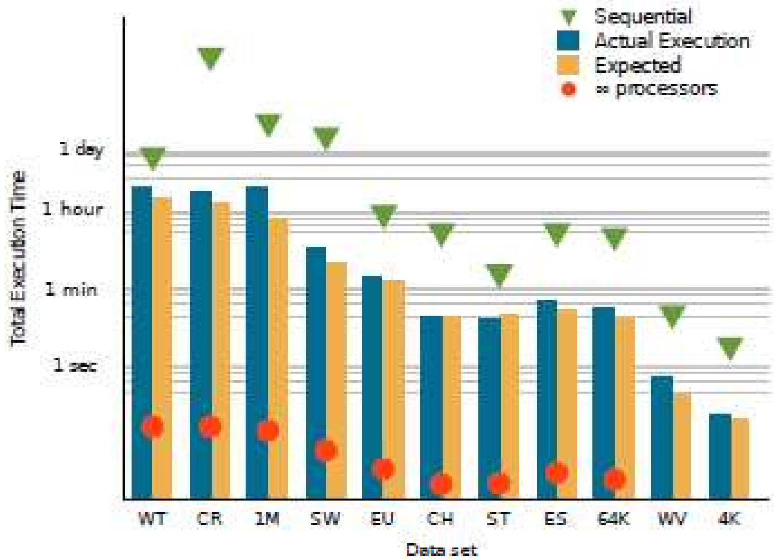
	Amazon	DotaLeague	Friendster
HDFS	1 second	7 seconds	5 minutes
Neo4J	4 hours	6 **days**	**n/a**

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis

Graphitti

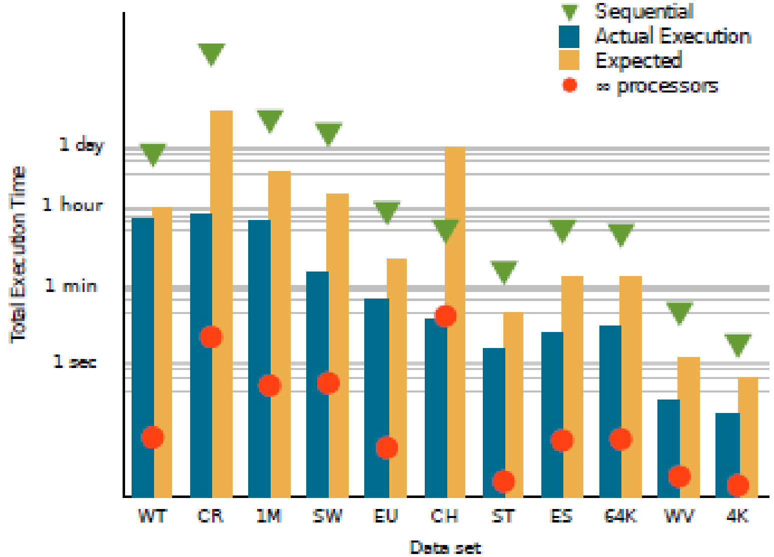
GPUs vs CPUs: All-Pairs Shortest Path

Pender and Varbanescu. MSc thesis at TU Delft. Jun 2012. TU Delft Library, http://library.tudelft.nl .

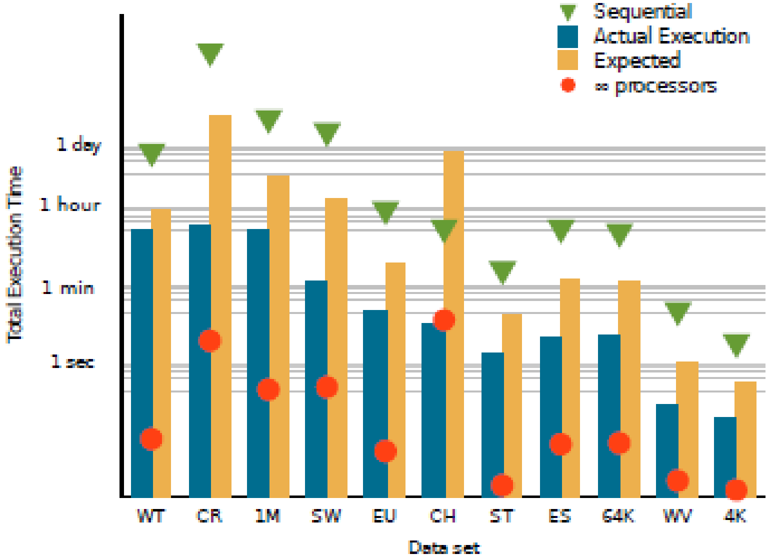


(a) Intel Xeon E5620

	Dataset
WT	Wikipedia Talk Network
CR	California Road Network
1M	Graph 1M
SW	Stanford Web Graph
EU	EU Email Communication Network
CH	Chain 100K
ST	Star 100K
ES	Epinions Social Network
64K	Graph 64K
WV	Wikipedia Vote
4K	Graph 4K



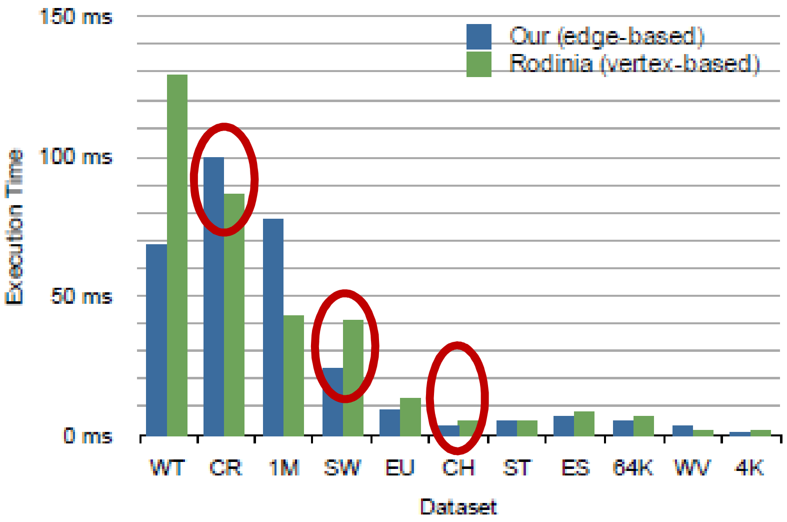
(c) Nvidia Tesla C2050/ C2070



(d) Nvidia GeForce GTX480

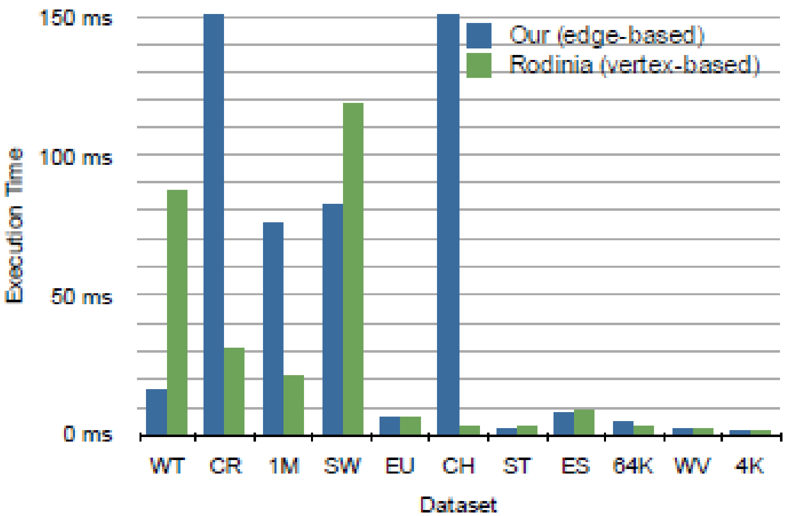
GPUs vs CPUs: BFS vs Data Format, E/V-based

Pender and Varbanescu. MSc thesis at TU Delft. Jun 2012. TU Delft Library, http://library.tudelft.nl .

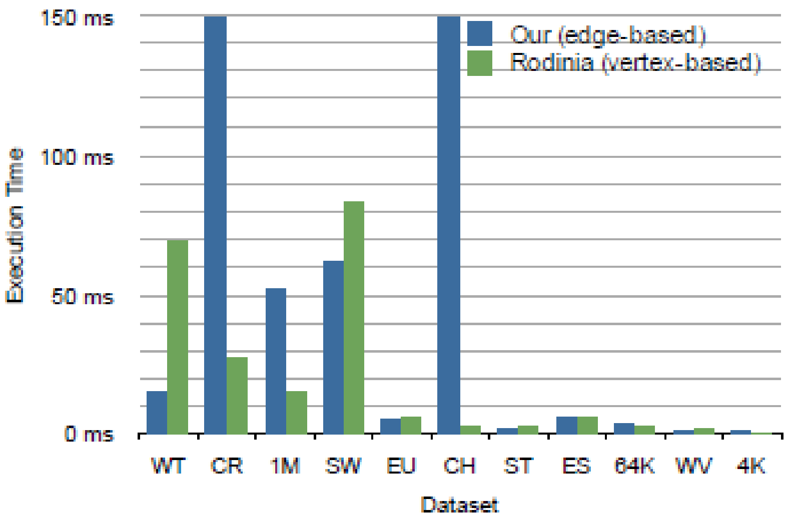


(a) Intel Xeon E5620

	Dataset
WT	Wikipedia Talk Network
CR	California Road Network
1M	Graph 1M
SW	Stanford Web Graph
EU	EU Email Communication Network
CH	Chain 100K
ST	Star 100K
ES	Epinions Social Network
64K	Graph 64K
WV	Wikipedia Vote
4K	Graph 4K



(c) Nvidia Tesla C2050/ C2070



(d) Nvidia GeForce GTX480

TU Delft

Q4

Conclusion and ongoing work

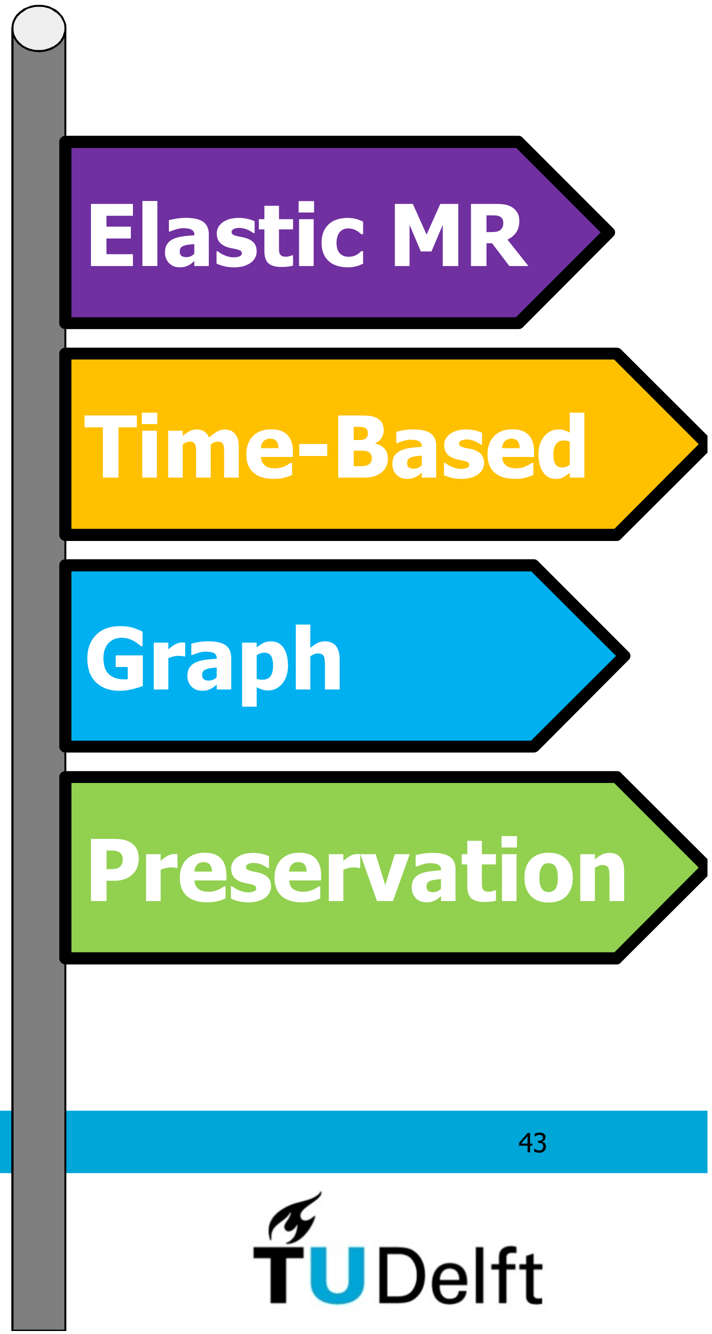
- Performance is f(Data set, Algorithm, Platform, Deployment)
- Cannot tell yet which of (Data set, Algorithm, Platform) the most important (also depends on Platform)
- Platforms have their own drawbacks
- Some platforms can scale up reasonably with cluster size (horizontally) or number of cores (vertically)

- Ongoing work
  - *Benchmarking* suite
  - Build a performance boundary model
  - Explore performance variability

http://bit.ly/10hYdIU

Agenda

1. Introduction
2. Programming Models for Big Data
3. **PDS Group Work on Big Data**
    1. MapReduce
    2. Graph Processing
    3. **Preservation**
4. Summary



The Personal Memex

- Vannevar Bush in the 1940s: record your life
- MIT Media Laboratory: The Human Speechome Project/TotalRecall, data mining/analysis/visio
  - Deb Roy and Rupal Patel "record practically every waking moment of their son's first three years" (20% privacy time…Is this even legal?! Should it be?!)
  - 11x1MP/14fps cameras, 14x16b-48KHz mics, 4.4TB RAID + tapes, 10 computers; 200k hours audio-video
  - Data size: 200GB/day, 1.5PB total

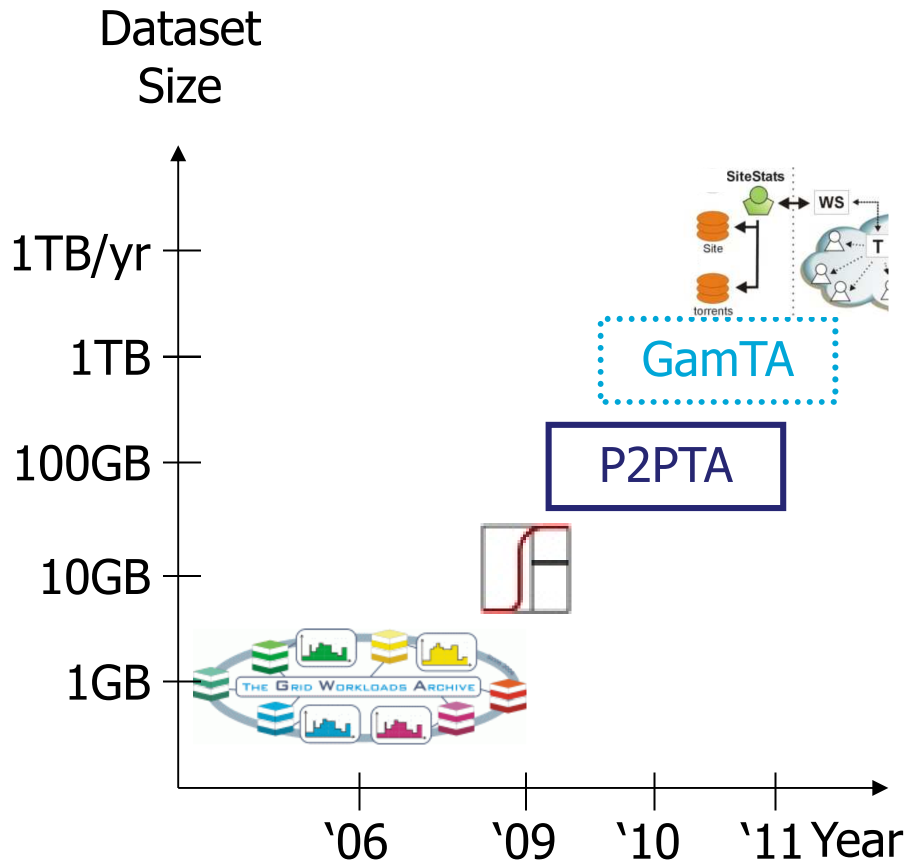
What is the Distributed Systems Memex ?

Data sets in Comp.Sci.

http://gwa.ewi.tudelft.nl

The Failure Trace Archive

http://fta.inria.fr



Peer-to-Peer Trace Archive
… PWA, ITA, CRAWDAD, …

- 1,000s of scientists: From theory to practice

The Grid Workloads Archive
Content

ID	System	Period	Number of observed Sites	CPUs	Jobs	Groups	Users
GWA-T-1	DAS-2	02/05-03/06	5	400	602K	12	332
GWA-T-2	Grid'5000	05/04-11/06	15	~2500	951K	10	473
GWA-T-3	NorduGrid	05/04-02/06	~75	~2000	781K	106	387
GWA-T-4	AuverGrid	01/06-01/07	5	475	404K	9	405
GWA-T-5◇	NGS	02/03-02/07	4	~400	632K	1	379
GWA-T-6◇							206
GWA-T-7‡							18
GWA-T-8‡							19
GWA-T-9‡	TeraGrid	08/05-03/06	1*	96	1.1M	26	121
	Total	13.51 yrs	136	>10000	>7M	191	2340
	Average	1.5 yrs	15	1151	>750K	21	>250

http://gwa.ewi.tudelft.nl

6 traces online

A. Iosup, H. Li, M. Jan, S. Anoep, C. Dumitrescu, L. Wolters, D. Epema, The Grid Workloads Archive, FGCS 24, 672–686, 2008.

The Failure Trace Archive
Content

System	Type	# of Nodes	Target Component	Period	Year
SETI@home	Desktop Grid	226,208	CPU	1.5 years	2007-2009
Overnet	P2P	3,000	host	2 weeks	2003
Microsoft	Desktop	51,663	host	35 days	1999
LANL	SMP, HPC Clusters	4750	host	9 years	1996-2005
HPC2	HPC Clusters	256	IO	2.5 years	1996-2005
HPC4					006
PNNL					007
NERSC					006
Skype					
Web sites	Web servers	129	host	8 months	2001-2002
DNS	DNS servers	62,201	host	2 weeks	2004
PlanetLab	P2P	200-400	host	1.5 year	2004-2005
Grenouille03	DSL	4800	host	1 year	2003
Grenouille05	DSL	4800	host	1 year	2005
EGEE	Grid	2500 queues	CE queue	1 month	2007
Grid'5000	Grid	1288	host	1.5 years	2005-2006
deug05	Desktop Grid	40	CPU	1 month	2005

http://fta.inria.fr

25 traces online

D. Kondo, B. Javadi, A. Iosup, D. Epema, The Failure Trace Archive: Enabling Comparative Analysis of Failures in Diverse Distributed Systems, CCGrid 2010 (accepted)

TU Delft

The Game Trace Archive
Content

Name	Period	Size (GB)	Node (M)	Edge (M)	Category
KGS	2002/02-2009/03	2	0.8	27.4	Chess Game
FICS	1997/11-2011/09	168	0.4	144.2	Chess Game
BBO	2009/11-2009/12	10	3.9	12.9	Card Game
XFire	2008/05-2011/12	58	7.7	34.7	OMGN
Dot[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	RTS
DotaLicious	2010/04-2012/02	1	0.1	0.0	RTS
Dota Garena	2009/09-2010/05	1	0.3	0.1	RTS

http://gta.st.ewi.tudelft.nl

2 traces online +1/month

Guo and Iosup, The Game Trace Archive, ACM NETGAMES 2012.

- Share gaming traces and best-practices on using them

TU Delft

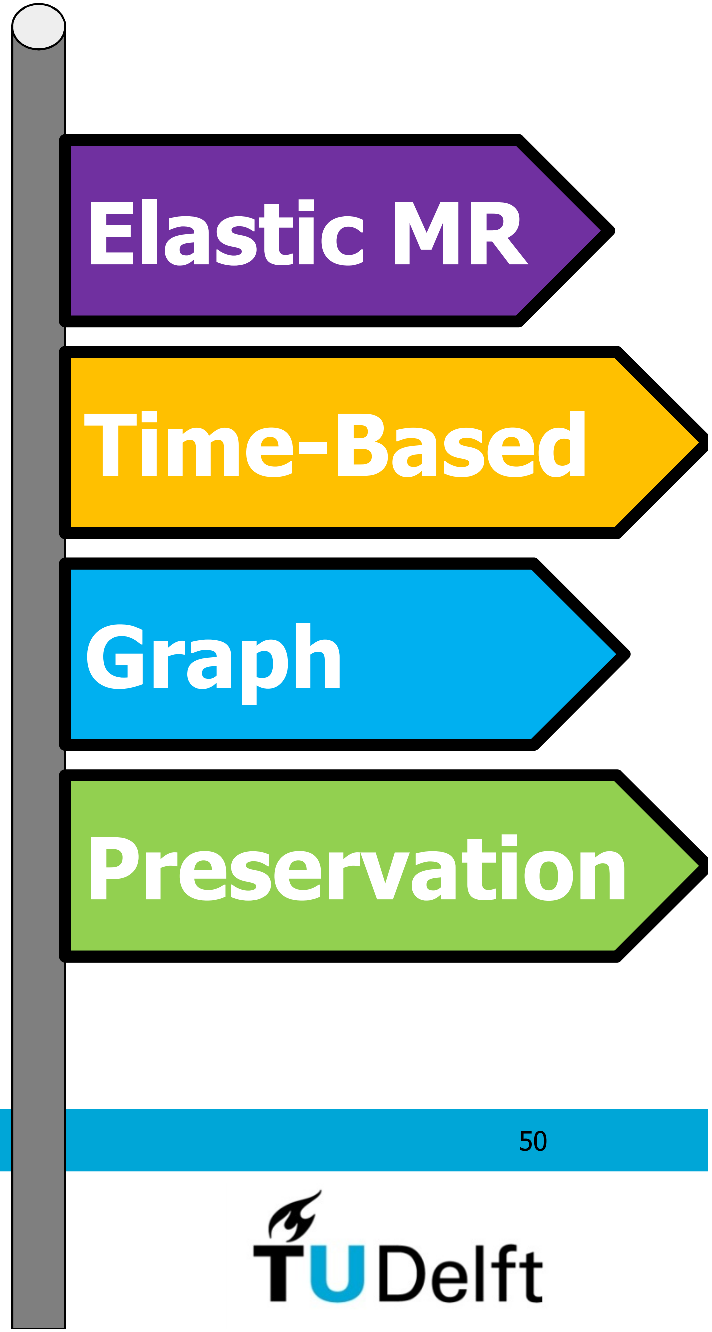
The Cloud Workloads Archive (ongoing)

Trace	System	Size	Period	Notes
CWA-01	Facebook	1.1M/-/-	5m/2009	Time & IO
CWA-02	Yahoo M	28K/28M/	20d/2009	~Full
CWA-03	Facebook	61K/10M/	10d/2009	Full detail
CWA-04	Facebook	?/?/-	10d/01-	Full detail
CWA-05	Facebook	?/?/-	3m/02+2	Full detail
CWA-06	Google 2			Future
CWA-07	eBay			Need
CWA-08	Twitter			Need
CWA-09	Google	9K/177K/4M	7h/2009	Coarse

- Own traces: 3+ years of observation of Amazon WS and Google AE
- Tools
  - Convert to CWA format
  - Analyze and model automatically → **Report**

Agenda

1. Introduction
2. Programming Models for Big Data
3. PDS Group Work on Big Data
    1. MapReduce
    2. Graph Processing
    3. Preservation
4. Summary



~~Conclusion~~ Take-Home Message

- **Programming Models for Big Data**
  - **Big data programming models have ecosystems**
  - Many trade-offs, many programming models
  - Models: MapReduce, Pregel, PACT, Dryad, …
  - Execution engines: Hadoop, Koala+MR, Giraph, PACT/Nephele, Dryad, …
- **PDS Group Work on Big Data**
  - Elastic Map Reduce
  - Map Reduce for time-based analytics: a use case
  - Towards a benchmarking suite for graph-processing platforms
  - Archives: Grid, P2P, Failures, Online Games
- **Conclusion: a thousand flowers already bloomed, so much to do … looking for collaborators**

Thank you for your attention! Questions? Suggestions? Observations?

HPDC 2013

More Info:

- http://www.st.ewi.tudelft.nl/~iosup/research.html
- http://www.st.ewi.tudelft.nl/~iosup/research_cloud.html
- http://www.pds.ewi.tudelft.nl/

Do not hesitate to contact me…

Alexandru Iosup

A.Iosup@tudelft.nl
http://www.pds.ewi.tudelft.nl/~iosup/ (or google "iosup")
Parallel and Distributed Systems Group
Delft University of Technology

TUDelft

Reading Material

- **Workloads**
  - Alexandru Iosup, Dick H. J. Epema: Grid Computing Workloads. IEEE Internet Computing 15(2): 19-26 (2011)
- **The Fourth Paradigm**
  - "The Fourth Paradigm", http://research.microsoft.com/en-us/collaboration/fourthparadigm/
- **Programming Models for Big Data**
  - Jeffrey Dean, Sanjay Ghemawat: MapReduce: Simplified Data Processing on Large Clusters. OSDI 2004: 137-150
  - Jeffrey Dean, Sanjay Ghemawat: MapReduce: a flexible data processing tool. Commun. ACM 53(1): 72-77 (2010)
  - Matei Zaharia, Andy Konwinski, Anthony D. Joseph, Randy Katz, and Ion Stoica. 2008. Improving MapReduce performance in heterogeneous environments. In Proceedings of the 8th USENIX conference on Operating systems design and implementation (OSDI'08). USENIX Association, Berkeley, CA, USA, 29-42.
  - Matei Zaharia, Dhruba Borthakur, Joydeep Sen Sarma, Khaled Elmeleegy, Scott Shenker, Ion Stoica: Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. EuroSys 2010: 265-278
  - Tyson Condie, Neil Conway, Peter Alvaro, Joseph M. Hellerstein, Khaled Elmeleegy, Russell Sears: MapReduce Online. NSDI 2010: 313-328
  - Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, Grzegorz Czajkowski: Pregel: a system for large-scale graph processing. SIGMOD Conference 2010: 135-146
  - Dominic Battré, Stephan Ewen, Fabian Hueske, Odej Kao, Volker Markl, Daniel Warneke: Nephele/PACTs: a programming model and execution framework for web-scale analytical processing. SoCC 2010: 119-130

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis

http://bit.ly/10hYdIU