A 30,000-feet Introduction to Distributed Systems for 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 24, 2013
Technion, Haifa, Israel
The Parallel and Distributed Systems Group at TU Delft

Alexandru Iosup
- Grids/Clouds
- P2P systems
- Big Data
- Online gaming

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

Ana Lucia Varbanescu
- HPC systems
- Multi-cores
- Big Data
- e-Science

Henk Sips
- HPC systems
- Multi-cores
- P2P systems

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

August 31, 2011
SPEC Research Group (RG)

The Research Group of the Standard Performance Evaluation Corporation

SPEC RG Cloud Working Group

http://research.spec.org/working-groups/rg-cloud-working-group.html
Why Is Big Data Needed? 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

Big Data = Too big, too fast, does not comply with traditional DB
Why Is Big Data Difficult?
Big Data in practice = Systems of Systems

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

2012-2013

TU Delft
What are MapReduce workloads made of?
Workload Characterization and Modeling

- Understand the global BitTorrent network, 2009—ongoing
  - Data from 1,000s of trackers, 30M+ shared files, 100M+ users
  - 15TB+ timestamped, multi-record, multi-files
  - Largest such project in the world!

- Statistical models of MR workloads
  - Data from Google, Yahoo, leading socnet operators
  - Compute WL
  - I/O WL
  - Data WL

- Future work: SPEC effort towards benchmarking suite
How Well do Graph-Processing Platforms Perform? 
An Empirical Investigation of 6 Platforms

- Large-scale graphs: socnets, Internet, online games, corporate
- Large number of platforms
  - Desktop: Neo4J, SNAP, etc.
  - Distributed: Giraph, GraphLab, etc.
  - Parallel: too many to mention

Problem: performance differences

Graphhitti

- Our work: empirical comprehensive study of 6 platforms x 5 algorithmic classes x 7 real-world & synthetic data sets
- Future work: SPEC effort towards benchmarking suite

June 24, 2013
Guo, Biczak, Varbanescu, Iosup, Martella, Willke

How Well do Graph-Processing Platforms Perform? 
An Empirical Performance Evaluation and Analysis


The State of LinkedIn

150,000,000
registered members
What is the Distributed Systems Memex?
Archiving Operational Data in Computer Science

- The Failure Trace Archive
- Peer-to-Peer Trace Archive
  - http://gwa.ewi.tudelft.nl
- Game Trace Archive
  - http://gta.st.ewi.tudelft.nl

... PWA, ITA, CRAWDAD, ...

- 1,000s of scientists now go from theory to practice
Conclusion Take-Home Message

- **Big Data**
  - Data Deluge = sink or swim
  - Big data in practice = systems of systems

- **PDS Group Work on Big Data**
  - Leads SPEC effort for benchmarking graph-processing platforms
  - Leads SPEC effort for benchmarking time-based analytics
  - Leads effort for the Distributed Systems Memex
  - Ask about our other systems-related work (Elastic MapReduce, data preservation in the cloud, ...)

- **Conclusion**: so much to do … looking for collaborators

June 24, 2013

http://www.flickr.com/photos/dimitrisotiropoulos/4934766418/
Thank you for your attention! Questions? Suggestions? Observations?

More Info:
- [http://www.st.ewi.tudelft.nl/~iosup/research.html](http://www.st.ewi.tudelft.nl/~iosup/research.html)
- [http://www.st.ewi.tudelft.nl/~iosup/research_cloud.html](http://www.st.ewi.tudelft.nl/~iosup/research_cloud.html)
- [http://www.pds.ewi.tudelft.nl/](http://www.pds.ewi.tudelft.nl/)

Alexandru Iosup

A.Iosup@tudelft.nl
[http://www.pds.ewi.tudelft.nl/~iosup/](http://www.pds.ewi.tudelft.nl/~iosup/) (or google “iosup”)
Parallel and Distributed Systems Group
Delft University of Technology
How to enable MapReduce elasticity?
Dynamic MapReduce Clusters from TUD

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

- To achieve fairness across multiple users
  - Redistribute idle resources
  - Allocate resources for new MR clusters

To isolate:
- Performance
- Failure
- Data
- Version

# Data Preservation in the Cloud: Core PaaS Services (Bit/Logical)

1. Ingest services
2. Curate services
3. Access services

<table>
<thead>
<tr>
<th>Service</th>
<th>Bit</th>
<th>Logical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add retention time+reminders; etc.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Graph exploration</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Object analysis</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Integrity Checking</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Dereferencing + Delete</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Migration (I/O) + MoveOut</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Export</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Conversion + Transformation</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Retention Admin</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

- **List Items**
- **Find Item(s)**
- **Retrieve Item(s)**
- **Emulate**
  - Stateful/-less; interactive/non-; etc.
- **Access Ctl. Admin**

- Y: Available
- -: Available
- ?: Availabe