Elastic and High-Performance Graph Processing at Scale

Yong Guo, Alexandru Iosup, and Dick Epema
Delft University of Technology, the Netherlands
Yong.Guo@tudelft.nl

Problems & Challenges

Challenges

- Graph availability: few of graphs are publicly available
- Selection: many graphs, algorithms, and platforms
- Prediction: many programming models and infrastructures
- Platform elasticity: clouds, accelerators, multiple jobs, …

The Data Deluge

Too Many Platforms

Sources: IDC, EMC.

Research Questions

Q1: How can we build a virtual meeting space for sharing, exchanging, and analyzing graphs for the gaming community? (application domain for explicit and interaction graphs) [1]

Use GTA traces in your work!

Q2: How well do graph-processing platforms perform? [2]

Q3: How to build models to predict the performance of graph-processing platforms?

- Goal: to predict the execution time, for a given input graph and algorithm
- Method: statistical models, experimental research, synthetic and real graphs, diverse algorithms
- Targets: Hadoop and Giraph
- Benefits: help users make their selection (Q2), help us design better job scheduler (Q4)

Q4: How to efficiently process graphs on IaaS clouds with or without accelerators?

- Multi-tenancy
- Elastic environment using hybrid clouds
- Experimentally validated
- Uses results from Q1, Q2, Q3