Análise de Desempenho de Plataformas de Processamento de Grafos

Performance Evaluation of Graph Processing Platforms

Daniel N. R. da Silva, Klaus Wehmuth, Carla Osthoff, Ana Paula Appel, Artur Ziviani
{dramos, klaus, osthoff, ziviani}@lncc.br and appel@br.ibm.com
Performance Evaluation of Graph Processing Platforms
Graphs

Set of entities (nodes) and Relationships (edges)

Technology, Biology, Sociology, Chemistry, etc.
Graphs

Set of entities (nodes) and Relationships (edges)

Technology, Biology, Sociology, Chemistry, etc.
Graphs
Set of entities (nodes) and Relationships (edges)
Technology, Biology, Sociology, Chemistry, etc.
Graphs

Set of entities (nodes) and Relationships (edges)

Technology, Biology, Sociology, Chemistry, etc.
To be pretentious, we can say that the **whole universe** is a single (really huge, possibly infinite) network, where the nodes are interactions between elementary particles, and the edges are the particles themselves. This is a network with perhaps $10^{80}$ nodes.

Large networks and graph limits. László Lovász. 2012
How to process large graphs?
Mizan  GPS  Apache Giraph  GraphX  OpenG  Flink

graphchi  PGX  ComBLAS  Pregel+  TITAN

SocialLite  GraphMat  Galois  GraphLab  Spark

neo4j  Haloop  Pegasus  dex  OrientDB  Hadoop

MORE THAN 80 PLATFORMS

Algorithm Analysis Metric Scope
Memory Network Processing
Clustering Degree Diameter
Language Model Optimizations
Related Works by Year

- 2008: 1
- 2009: 0
- 2010: 3
- 2011: 1
- 2012: 4
- 2013: 7
- 2014: 12
- 2015: 19
Summing up

• Performance evaluation in this paper:
  • 1 computer environment
  • 3 algorithms
  • 3 platforms
  • 23 graphs
Algorithms

Connected Components
Page Rank
Single Source Shortest Paths
Algorithms

Connected Components

Page Rank

Single Source Shortest Paths
Algorithms

Connected Components

Page Rank

Single Source Shortest Paths
Algorithms

Connected Components

Page Rank

Single Source Shortest Paths

Performance Evaluation of Graph Processing Platforms
Networks

Node Color: Cluster
Node Size: Degree

Barabási Albert

Performance Evaluation of Graph Processing Platforms
Networks

number of nodes

degree

Node Color: Cluster
Node Size: Degree

Erdos Renyi

Performance Evaluation of Graph Processing Platforms
Networks

23 Nets:
- 9 BA
- 9 ER
- 5 RE

$\log_{10}(|V| + |E|)$ FROM
- 0
- 7.0
- 7.5
- 8.0
- 8.5
- 9.0
- 9.5

Performance Evaluation of Graph Processing Platforms
Iosup et al. PVLDB 2016.
Platforms

Shared Memory

Distributed Memory
Platforms

- Shared Memory
  - Database
  - Library
    - Based on Disk Access
  - Graph Specific
    - Matrix Abstraction
    - General Purpose
- Distributed Memory
  - Graph Centric
  - Vertex Centric
Platforms

Shared Memory
- Database
  - Neo4J
- Library
- Based on Disk Access
- GraphChi
- Graph Centric
- Vertex Centric
- Giraph++

Distributed Memory
- Graph Specific
- Matrix Abstraction
- General Purpose
- PowerGraph
- Apache Giraph
- KDT
- Apache Spark
- GraphX

Performance Evaluation of Graph Processing Platforms
Platforms

GraphX
D. Memory
Framework
Graph & Table
## Platforms

<table>
<thead>
<tr>
<th>GraphX</th>
<th>NetworKit</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. Memory</td>
<td>S. Memory</td>
</tr>
<tr>
<td>Framework</td>
<td>Library</td>
</tr>
<tr>
<td>Graph &amp; Table</td>
<td>C. N. Analysis</td>
</tr>
</tbody>
</table>
## Platforms

<table>
<thead>
<tr>
<th>GraphX</th>
<th>NetworKit</th>
<th>PowerGraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. Memory</td>
<td>S. Memory</td>
<td>D. Memory</td>
</tr>
<tr>
<td>Framework</td>
<td>Library</td>
<td>Framework</td>
</tr>
<tr>
<td>Graph &amp; Table</td>
<td>C. N. Analysis</td>
<td>Graph &amp; ML</td>
</tr>
</tbody>
</table>

Performance Evaluation of Graph Processing Platforms
Experiments

• Ubuntu 14.10 Server
  • 2 Intel Xeon 2.27GHz (8 cores)
  • 47 GB RAM
  • 1 TB HD

• Make span
  • Average 5 executions
  • 95 % confidence interval
  • 14 hours threshold
Platform Comparison

Page Rank

Performance Evaluation of Graph Processing Platforms
Algorithm Comparison

GraphX

<table>
<thead>
<tr>
<th>BA</th>
<th>ER</th>
<th>RE</th>
<th>BA</th>
<th>ER</th>
<th>RE</th>
<th>BA</th>
<th>ER</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Page Rank</td>
<td>SSSP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Algorithm Comparison

Networkit

BA  ER  RE  BA  ER  RE  BA  ER  RE

CC  Page Rank  SSSP
Algorithm Comparison

PowerGraph

Performance Evaluation of Graph Processing Platforms
Network Comparison

GraphX

<table>
<thead>
<tr>
<th>CC</th>
<th>PR</th>
<th>SP</th>
<th>CC</th>
<th>PR</th>
<th>SP</th>
<th>CC</th>
<th>PR</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barabási Albert</td>
<td>Erdos Renyi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Network Comparison

Networkit

Barabási Albert

Erdos Renyi

Performance Evaluation of Graph Processing Platforms
Network Comparison

PowerGraph

Performance Evaluation of Graph Processing Platforms
Conclusions

• Small server applicability;
• NetworKit better in 54% of cases
• PowerGraph
  • Better for BA networks
  • The less iterations, the more efficient
Future works

• Increase experimental coverage
  • Algorithms, Metrics, Nets, and Platforms
• Taxonomy
• Use a larger computation environment
  • Private Cloud
Análise de Desempenho de Plataformas de Processamento de Grafos
Performance Evaluation of Graph Processing Platforms

Daniel N. R. da Silva, Klaus Wehmuth, Carla Osthoff, Ana Paula Appel, Artur Ziviani
{dramos, klaus, ostoff, ziviani}@lncc.br and appel@br.ibm.com