Pregel: A System for Large-Scale Graph Processing

Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn,
Naty Leiser, and Grzegorz Czajkowski
Google, Inc.
{malewicz,austern,ajcbik,dehnert,ilan,naty,gczaj}@google.com

Bryan Rockwood

11/24/13

What is Pregel?

- A scalable, fault-tolerant platform with an accessible and flexible API to express arbitrary graph algorithms

- So in english -- a large framework that processes excessive amounts of graphing equations, updating the graphs. That still doesn't help does it?

Pregel Implementation

- Clusters of computers partition graphs to optimize work flow.

- Masters are not assigned parts of the graph, but direct orders to the worker and to coordinate assignments.

-While the workers are on the front-lines, the Master ensures the big picture is fully operational.

Analyzing Pregel

- Although Pregel itself has work to do to improve it’s ability to keep up with the growing graph sizes, it does make a huge step toward computing massive algorithms.

- Breaking down these large graphs and dividing them amongst workers and aggregators allows for important tasks to be completed efficiently.

Advantages and Disadvantages

Advantages

- Intuitive and easy to use API
- Can handle graphs consisting of billions of vertices.
- Flexible for further evolution of the framework

Disadvantages

- Cannot change the API
- Assigning vertices to machines is a challenge.

Real World Uses

- Social Graphs, such as Google+

-Web-Link Graphs

-Location Graphs