

A Revised Implimentation of the GraphZ Graph Processing System

Ballmer, Alexander

alexandersballmer@gmail.com

Moudgalya, Shreyas

smoudgal@hawk.iit.edu

June 3, 2016

Abstract

GRAPH/Z is a distributed parallel graph processing system running on top of ZHT, a zero hop distributed hash table. It uses an iterative vertex-centric model to store a large graph and run a variety of algorithms over it. GRAPH/Z has some performance issues, and cannot scale the same as graphlab, a similar commercial framework. We hope to rewrite GRAPH/Z to be cometicitive with Graphlab on a single node with multiple threads.

1 Background

GRAPH/Z was based off of a graph processing paradigm called Pregel. Pregel uses a model centered around the vertexes of the graph. Each vertex has an update function that is run in the vertex's context in the graph. The update function allows the vertex to modify its edges, perform calculations, and send messages to other vertexes. The computation occurs in iterations called supersteps, with each iteration calling every vertex's update function theoretically in parallel. Even if there are fewer threads than vertexes, all of the update functions behave as if they are called in parallel.

Messages sent by vertexes are used to communicate between vertexes, sending data to the next iteration. At each iteration, vertexes can vote to halt and disable themselves. A halted vertex can be re-enabled by receiving a message. If all the vertexes are disabled at the start of an iteration, the entire system halts and returns.

GRAPH/Z adapts the Pregel model, but adds the concept of a distributed hash table. The hash table stores both the graph's edges and vertexes, and also provides a platform for running the distributed message queue. The hash ta-

ble serves as the only means of communication between nodes. The hash table used is ZHT, a DHT implementation that is fault-tolerant and can scale to 32000 cores. ZHT abstracts away the physical hardware, exposing only a key-value store.

2 Problem

3 Related Work

4 Proposed Solution

5 Evalution

6 Timeline

7 Deliverables

8 Conclusion