Benchmarking tool for graph algorithms

Abhinaba Sarkar 201405616
Malavika Reddy 201201193
Nikita Kad 201330030
Yash Khandelwal 201302164

Abstract

DESCRIPTION:

In computer science and mathematics, graphs are abstract data structures that model structural relationships among objects. They are now widely used for data modeling in application domains for which identifying relationship patterns, rules, and anomalies is useful. These domains include the web graph, social networks,etc. The ever-increasing size of graph-structured data for these applications creates a critical need for scalable systems that can process large amounts of it efficiently. The project aims at making a benchmarking tool for testing the performance of graph algorithms like BFS, Pagerank,etc. with MapReduce, Giraph, GraphLab and Neo4j and testing which approach works better on what kind of graphs.

APPROACH & SCOPE:

First we will build various algorithms (like BFS and Pagerank) using Hadoop Mapreduce, GraphLab and Giraph then create a benchmarking tool in java which will compare the performance of the algorithms to determine which graph algorithm works better for which kind of graphs and graph related queries.

TOOLS:

- Hadoop Mapreduce Engine
- GraphLab - GraphLab is a graph-based, high performance, distributed computation framework written in C++. It was originally developed for Machine Learning tasks.

- Giraph - Apache Giraph is an iterative graph processing system built for high scalability. For example, it is currently used at Facebook to analyze the social graph formed by users and their connections.
- Neo4j - Neo4j is an open-source graph database, implemented in Java. Neo4j is an embedded, disk-based, fully transactional Java persistence engine that stores data structured in graphs rather than in tables.