Distributed Storage System for Optimum Data Access

Aditya Joshi
Department of Computer Science
North Carolina State University
ajoshi6@gmail.com

Karthikeyan Vaideswaran
Department of Computer Science
North Carolina State University
kvaides@ncsu.edu

Sachin Saligram
Department of Computer Science
North Carolina State University
ssaligr@ncsu.edu

1 Introduction

The aim of any distributed system is to provide consistency, availability and partition tolerance. We want to build such a distributed system that provides fast and reliable access to data.

2 Project Description

This project aims at providing data at faster rates without compromising on server performance and network bandwidth.

The idea is to set up multiple servers to form a distributed system with data replication. Data requested by the user is provided from the most optimal server.

The definition of an optimal server is dependant of multiple factors such as latency, bandwidth, server load, and storage capacity to name a few. The algorithm we implement takes these factors into account while determining the optimal access path to the data requested.

3 Infrastructure

For our infrastructure, we plan on using the following.

1. EC2 on Amazon Web Services (AWS) with S3 object storage.
2. A distributed system manager like Zookeeper.
3. Peer-to-peer distribution and communication protocol and algorithm for load balancing and optimum data transfer.

4 Goals

The goal of this project is to provide a user with access to data with least latency and time while not affecting load and performance of the distributed system.