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 1 Introduction

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

4 2 Project Description

- 5 This project aims at providing data at faster rates without compromising on server performance and
- 6 network bandwidth.
- 7 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.
- 9 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.

12 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.

8 4 Goals

14

15

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.