Cloud Computing: Project Proposal (Team C)

Section 1: Team Members

Anahita Bhiwandiwalla: ab3744
Digvijay Singh: ds3161
Kuber Kaul: kk2872
Payal Rani: pr2417

Section 2: Abstract:

In today's time, data management is a crucial need in every organization. The main questions that arise are - How do we store vast data? Where do we store this data? How do we ensure data availability and minimize data loss due to any intermediate failures?

In this project we wish to create a cloud based application that allows people across different geographic locations to access shared data from the Cloud.

In order to improve performance and usability, we plan to intelligently pre-fetch the files that might be required the users. We propose an innovative way to intelligently pre-fetch data using Google Prediction API and cache replacement Algorithm. We plan to explore both Randomized Marking Algorithm and Least Recently Used (LRU) Algorithm used by Dropbox and other cloud based applications for cache management. Randomized Marking Algorithm is an online algorithm that guarantees better performance than LRU in worst case. Google Prediction API that provides pattern-matching and machine learning capabilities. The application would learn the user's behavior from the past file access patterns and would intelligently pre-fetch the files required by the user to reduce the overhead in the data transfer as much as possible.

In order to reduce network bandwidth, we plan to tailor the application depending on the client. If the client is a mobile device, we do not transfer the original large size image but a smaller thumbnail that uses less bandwidth. For desktop clients, the user can download the original large sized files.

To summarize, we propose an innovative, user friendly and robust cloud based application that would provide rich user experience by reducing network bandwidth and intelligent data pre-fetch using Google Prediction API and an efficient online algorithm for cache replacement.

Section 3: Technology to be used:

Google App Engine to host the system
BlobStore for storage of large files on the Cloud
Programming Language – Python
Front end Languages: HTML5,CSS,Javascript
Frameworks: Django (Python Web framework)
Google Prediction API

Section 4: Team name:

GoData