



CALIFORNIA STATE UNIVERSITY  
**FULLERTON**

**ADVANCED DATABASE MANAGEMENT  
Fall 22 CPSC 531-03**

**Project Proposal**

Prepared by

Vaibhav Monpara  
(885198523)  
(vaibhavmonpara@csu.fullerton.edu)

Vasukumar Lakhani  
(885226365)  
(vasulakhani@csu.fullerton.edu)

## Project Details

**Project Name:** Big Data Analysis using MapReduce in Hadoop

**Problem Statement:**

Data Analysis is very critical in understanding the data, and what we can do with the data. For small datasets it is easier to process and obtain the results. But as for big companies, it becomes crucial for them to obtain the trends of the company for any changes need to be made. During the last decade, the most challenging problem the world envisaged was big data problem. The big data problem means that data is growing at a much faster rate than computational speeds. And it is the result of the fact that storage cost is getting cheaper day by day, so people as well as almost all business or scientific organizations are storing more and more data. There are many systems and tools in the market which does help to analyze the data and help the commercial or any other larger organizations, we aim to reduce the processing time by implementing our big data technology which would process the data and help improve efficiency of the system.

**Approach:**

By implementing the Big Data technology using Hadoop we can realize the importance of framework and can learn in-depth of the framework at the same time. We will implement parallel processing and thus we can reduce the overhead quantifiably. We will use the results to analyze the data and get the maximum usage from the data. Also, we can analyze data through word clouds that will make the data simpler to visualize and to present. For example, the word count program can give us the most frequently occurring words, which can be implemented to get the current trends w.r.t a particular topic. Similarly, we can use word co-occurrence program that would give us results that we can show some relevant terms and can be used together.