# **FINAL PROJECT PLAN**

**Title:** Live twitter data analysis using Cloud services

**Team members:**

1. Dhanunjaya Lakshman, Koneru (Student ID: 16336828)
2. Lakshmi Durga Likhitha, Tellakula (Student ID: 16335481)
3. Sushmitha Reddy, Pollapalli (Student ID: 16336844)
4. Siddartha Rao, Damarla (Student ID: 16337125)

**Roles and Responsibilities:**

* Lakshman and Likhitha are responsible for research, pre-processing, transforming, and visualizing the data. Checking the compatibility of tools used and preprocessing the data in Lambda using Pandas, handling CloudWatch and hosting interactive dashboards using EBS are the typical tasks handled by them.
* Sushmitha and Siddhartha are responsible for installing libraries using EC2, extracting tweets from twitter, producing visualizations using Plotly and storing the results in S3 are the typical tasks handled by them.

**Motivation and Purpose:**

The motivation behind the "Live Twitter Data Analysis using Cloud Services" project is to harness the power of cloud computing to extract, process, and analyze real-time data from Twitter. Twitter is one of the most popular social media platforms with millions of users tweeting every day. By leveraging cloud services, the project aims to provide a scalable, efficient, and cost-effective solution to analyze tweets and gain insights into trending topics, sentiment analysis, and social media behavior.

The purpose of this project is to create a pipeline that can extract tweets from Twitter, store them in an S3 bucket, perform data transformations and pre-processing using pandas in Lambda, generate visualizations using plotly, and store the results back in S3. This pipeline will be triggered using CloudWatch, which will monitor for incoming tweets and trigger the processing pipeline when new tweets are detected. An interactive dashboard will be hosted using Elastic Beanstalk (EBS), which will allow users to explore the data and gain insights. EC2 will be used to install additional libraries and dependencies that may be required for processing and analysis.

In summary, the primary objective of this project is to create a scalable, cost-effective, and efficient solution for real-time data analysis using cloud services. By leveraging the power of cloud computing, the project aims to extract valuable insights from Twitter data and enable users to make data-driven decisions.

**Cloud Technologies used:**

1. **Amazon S3 (Simple Storage Service):** This is a cloud storage service used for storing and retrieving data. In this project, S3 is used to store the extracted tweets, processed data, and visualizations.
2. **AWS Lambda:** This is a serverless computing service used to run code in response to events. In this project, Lambda is used to perform data transformations and pre-processing on the tweets.
3. **Plotly:** This is a cloud-based data visualization platform used to create interactive charts and graphs. In this project, Plotly is used to generate visualizations of the processed data.
4. **Amazon CloudWatch:** This is a monitoring service used to monitor resources and applications in real-time. In this project, CloudWatch is used to trigger the processing pipeline when new tweets are detected.
5. **Amazon Elastic Beanstalk (EBS):** This is a cloud-based platform used for deploying and scaling web applications. In this project, EBS is used to host an interactive dashboard for exploring the processed data.
6. **Amazon EC2 (Elastic Compute Cloud):** This is a cloud-based virtual machine used for running applications in the cloud. In this project, EC2 is used to install additional libraries and dependencies that may be required for processing and analysis.

These cloud technologies work together to create a scalable, efficient, and cost-effective solution for analyzing live Twitter data.