# Big Data and Cloud Computing Workshop

## Morning Session: Overview of Big Data Technologies

## 1. Introduction to Big Data Technologies

Comprehensive overview of major big data technologies according to the resources provided, focusing on Hadoop, Spark, AWS, and Google Cloud.

## Afternoon Session: Hands-On with Cloud Services

By looking up to various videos on youtube about big data on Googlle Cloud I could summarize the following :-

### 1. Setting Up Data Pipelines on Google Cloud

The afternoon session was a practical, hands-on workshop where I focused on setting up data pipelines and processing large datasets using Google Cloud services. The main steps and learnings are summarized below.

2. Google Cloud Setup

-Service Account: Created a service account with appropriate permissions for managing data and running data pipelines.

### 3. Google Cloud Storage

- Bucket Creation: Created a Google Cloud Storage bucket for storing raw data files.

- Data Upload: Uploaded sample large datasets into the bucket.

### 4. Google BigQuery

- Dataset Creation: Created a BigQuery dataset within the project.

- Data Import: Imported the data from Google Cloud Storage into BigQuery tables using the console and CLI.

### 5. Google Dataflow

- Pipeline Setup: Set up a Dataflow pipeline for processing and transforming data.

- Apache Beam: Used Apache Beam SDK to define the pipeline steps.

- Job Execution: Executing the Dataflow job, monitoring its progress and performance via the Google Cloud Console.

### Conclusion

The hands-on session provided practical experience in setting up and managing data pipelines on Google Cloud. It highlighted the ease of use, flexibility, and power of Google Cloud services for big data analytics. The combination of theoretical knowledge and practical skills gained from this workshop has significantly enhanced my understanding of big data technologies and cloud-based data analytics platforms.