# Phase 1 Documentation: Problem Definition and Design Thinking

**Project:** Big Data Analysis

**Phase 1 Submission by:** Deependra Raj Amit
**Year:** 3rd Year, Computer Science and Engineering (CSE)

**Table of Contents**

## 1. Introduction

This document represents Phase 1 of the "Big Data Analysis" project, submitted by Deependra Raj Amit, a 3rd-year student of Computer Science and Engineering (CSE). In this phase, the project focuses on defining the problem and design thinking aspects.

## 2. Problem Definition

**Objective:**

The primary objective of the "Big Data Analysis" project is to leverage big data analysis to uncover valuable insights from diverse datasets. This phase aims to clearly define the project's objectives and its relevance.

**Relevance:**
In today's data-driven world, extracting actionable insights from vast datasets is critical for informed decision-making. This project is relevant as it explores the potential of using IBM Cloud Databases for big data analysis, specifically focusing on the "rainfall in India 1901-2015" dataset. Insights from this dataset can have applications in agriculture, water resource management, and climate research.

## 3. Design Thinking

 - Data Selection

The chosen dataset, "rainfall in India 1901-2015," aligns with the project's objectives of exploring climate trends. This dataset is significant due to its potential impact on agriculture, a crucial sector in India's economy.

 - Database Setup

The IBM DB2 database has been selected as the platform for accommodating and managing the large "rainfall in India 1901-2015" dataset. The setup involves creating a dedicated database instance and configuring it for efficient data storage and retrieval.

 - Data Exploration

Data exploration involves developing queries and scripts to analyze the "rainfall in India 1901-2015" dataset. This process aims to identify patterns, trends, and anomalies within the data, laying the foundation for deeper analysis.

- Analysis Techniques

The project will employ statistical analysis techniques to derive insights from the dataset. Descriptive statistics, time series analysis, and data visualization will be used to understand historical rainfall trends in India.

- Visualization

To effectively communicate analysis results, a visualization strategy will be implemented. This will include creating graphical representations such as line charts and heatmaps to showcase rainfall patterns over time.

- Business Insights

The project's ultimate goal is to translate analysis findings into actionable business intelligence. For example, insights from this analysis can guide agricultural planning, water resource allocation, and climate adaptation strategies.

## 4. Conclusion

Phase 1 of the "Big Data Analysis" project, led by me (Deependra Raj Amit), has successfully defined the problem and outlined the design thinking process. The selection of the "rainfall in India 1901-2015" dataset, along with the setup of an IBM DB2 database, marks the beginning of a data-driven journey to uncover valuable insights. The subsequent phases will involve data exploration, analysis, visualization, and the transformation of findings into actionable business intelligence.