**BIG DATA ANALYSIS USING IBM CLOUD DATABASES**

**TEAM LEADER:** DHIVYA T (822721106012)

**PHASE 2:** Innovation

**Problem Title:** Big Data Analysis

**Problem Statement:** Dive into the world of big data analysis with IBM Cloud Databases. Uncover hidden insights from vast datasets, from climate trends to social patterns. Visualize your findings and derive valuable business intelligence. Embark on data-driven adventures, exploring the endless possibilities of big data!

**FLOW CHART:**

**START**

**DATA COLLECTION**

**DATA PREPROCESSING**

**ANALYSIS**

**VISUALIZATION**

**INTERPRETATION**

**REPORTING**

**END**

**Objective:** The project involves delving into big data analysis using IBM Cloud Databases. The objective is to extract valuable insights from extensive datasets, ranging from climate trends to social patterns. The project includes designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and visualizing the results for business intelligence.

**DATA COLLECTION:**

Sea level datasets, in the context of climate and environmental science, refer to collections of data that record and measure various aspects of sea level in different locations and over time. These datasets are essential for monitoring changes in sea level, understanding the causes of those changes, and assessing the potential impacts on coastal areas.

**Reference:** The Impact of Sea-Level rise on Developing Countries: A Comparative Analysis by Dasgupta, Laplante, Meisner, Wheeler and Yan

**DATA PREPROCESSING:**

**Data Quality Assessment:**

Conduct a preliminary assessment of data quality to identify potential issues. This may include checking for data gaps, erroneous values, or inconsistencies.

**Handling Missing Data:**

* Identify and handle missing data points:
* If missing data is limited, you can interpolate or fill gaps using appropriate methods, such as linear interpolation for time series data.
* For data with extensive missing values, consider using data imputation techniques, although caution is needed to avoid introducing bias.

**Outlier Detection and Handling:**

* Detect and address outliers or anomalous data points that may skew your analysis. Outliers can result from equipment malfunctions, extreme weather events, or other factors.
* Depending on the context and cause of outliers, you may choose to:
* Remove them if they are clearly erroneous and not representative of the underlying pattern.
* Transform them (e.g., winsorizing) to reduce their impact on the analysis.
* Investigate and potentially correct the underlying issue if it's a systematic problem.

**Unit Conversion and Standardization:**

Ensure that all data variables are in consistent units. For sea level data, this may involve converting measurements to a common reference point (e.g., mean sea level) or standardizing units (e.g., meters, millimeters).

To work with sea level data effectively, it's essential to gather information from various sources like tide gauges and satellites while ensuring consistent formatting. Conduct preliminary data quality checks to identify issues and address missing data and outliers accordingly. Standardize units if needed and, optionally, apply data smoothing or aggregation. Format the data for analysis tools, maintain comprehensive documentation, and then explore the cleaned data. Depending on your analysis goals, consider splitting the dataset for model development and evaluation.

**ANALYSIS:**

To analyze sea level data effectively, start with clean and organized data. Look for trends and patterns visually, calculate basic statistics, and examine long-term trends and extreme events. Use statistical models or tests if you have specific questions. Consider data integration and create clear visuals to present your findings. Summarize your insights, discuss implications, and, if relevant, project future sea level changes. Adapt your methods to your research goals and seek expert guidance as needed.

**VISUALIZATION:**

"Visualization of sea level data involves creating graphical representations, such as charts, graphs, maps, and other visual formats, to convey sea level measurements, trends, and patterns, aiding in the interpretation and communication of data related to sea level changes in coastal regions and oceans."

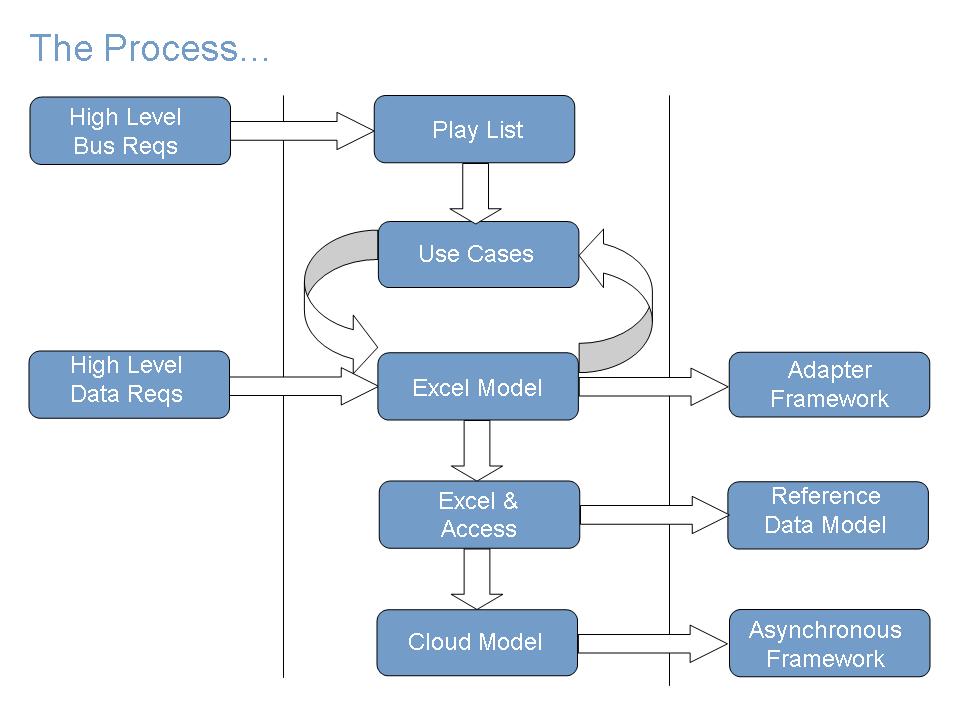
Key elements of sea level data visualization may include time series plots to show changes over time, spatial maps to display regional variations, and other graphical representations that help convey the important aspects of sea level data. Effective visualization enhances data-driven decision-making and public awareness regarding sea level rise and its implications.

**INTERPRETTION:**

"Interpretation of sea level data involves the examination, analysis, and comprehension of sea level measurements and associated information to extract valuable insights, draw conclusions, and gain a deeper understanding of factors influencing sea level changes and their potential impacts on coastal regions and environments."

In essence, interpretation of sea level data goes beyond raw data analysis, aiming to provide context and meaning to the observations and measurements, ultimately contributing to our understanding of sea level variations and their consequences.

**REPORTING:**



"Reporting sea level data entails the organized and clear presentation of sea level measurements, trends, and analysis outcomes in written, visual, or oral formats to convey important information about sea level changes, their causes, and their potential impacts, facilitating informed decision-making and public awareness."

Effective reporting of sea level data may involve the creation of reports, scientific papers, presentations, visualizations, and other communication methods to share research findings and their implications with various audiences. It plays a vital role in raising awareness of sea level rise and supporting actions to mitigate and adapt to its effects.