## Data Warehousing with IBM Cloud Db2 Warehouse

## Phase 2: Innovation

Consider incorporating advanced analytics tools or machine learning models for predictive analysis within the data warehouse.

## Our Solution:

## IBM SPSS Statistics: Best for Dashboard Capabilities

**IBM SPSS Statistics** is a popular predictive analytics tool. It offers a user-friendly interface and a strong set of features including the SPSS modeler, which provides advanced statistical procedures, helps ensure precision, and provides positive decision-making. All of the analytics lifecycle features are included, such as data preparation and management to analysis and reporting.

## Pricing:

IBM provides multiple **payment options**, including a subscription plan, term licenses, and academic plans. IBM also offers a **free trial** to allow users to try it before they buy it.

### **Features:**

* **Linear Elastic Net Regression:** Estimates linear regression models for dependent variables with one or more independent variables.
* **Custom Tables:** Gives customization to customers’ data sets.
* **Bootstrapping:** Approximation of sampling distributions by an estimator by resampling the original data set.

### **Pros:**

* Easy statistical analysis of large data for beginners.
* Improves efficiency with coding.
* Interactive dashboard.

### **Cons:**

* Expensive software.

## STEPS TO IMPLEMENT DATA WAREHOUSE IN IBM CLOUD:

1. **CREATE AN IBM CLOUD ACCOUNT**

**Create a Cloud Foundry App**: We can use the IBM Cloud Foundry service to host our web application. Create a Cloud Foundry app and specify the runtime environment we want to use (Node.js, Python, etc.)

1. **Install and Set Up the IBM Cloud CLI (Command Line Interface)**: To interact with Cloud Foundry from our command line, we'll need to install the IBM Cloud CLI. Follow the installation instructions provided by IBM Cloud.
2. **Log in to IBM Cloud CLI**: Open our command line interface and run the following command to log in to IBM Cloud.

**Select the Right Data Warehousing Service:** IBM Db2 Warehouse on Cloud.

**Define Data Warehouse Schema:**

**The Schema used in our database is Star Schema.**

* In multiple dimension there is no sub-dimension tables.
* Data Redundancy is high.
* Increased Performance.
* It is easy to understand.

**Data Modeling:** Star Schema is used in data modeling.

Code:

**Create the Time Dimension table**

CREATE TABLE Time (

TimeKey INT PRIMARY KEY,

Date DATE,

Day INT,

Month INT,

Quarter INT,

Year INT

);

**Data Ingestion:** Code for Inserting or loading the data into the data warehouse.

**Steps involves in Data Ingestion:**

* Data Loading Tools: Many IBM Cloud data warehousing services provide tools for bulk data loading.
* ETL (Extract, Transform, Load): Utilize ETL processes to extract data from source systems, transform it as needed, and load it into the data warehouse.

Data Integration Services: IBM Cloud offers data integration services that can help you move and transform data into the data warehouse.

**Code for Inserting or loading the data into the data warehouse:**

**Insert data into the Time Dimension**

INSERT INTO Time (TimeKey, Date, Day, Month, Quarter, Year)

VALUES

(1, '2023-01-01', 1, 1, 1, 2023),

(2, '2023-01-02', 2, 1, 1, 2023),

-- Add more rows as needed

(N, 'YYYY-MM-DD', D, M, Q, Y);

**Data Transformation:** If required, perform data transformations within the data warehouse to ensure data consistency and compatibility with our schema.

**Data Security and Access Control:** Implement security measures to protect our data warehouse and configure access controls. Ensure that only authorized users can access and manipulate the data.

**Optimize Query Performance:** Depending on our data volume and query requirements, optimize the performance of our data warehouse by creating indexes, partitions, and analyzing query execution plans.

**Set Up Backup and Recovery:** Implement backup and recovery mechanisms to safeguard your data in case of unexpected data loss or system failures.

**Monitoring and Maintenance:** Regularly monitor the performance of our data warehouse, track data usage, and ensure that the system is running efficiently. Perform routine maintenance tasks like updates and data purging.

**User Training and Documentation:** Train users and data analysts on how to work with the data warehouse. Create documentation to guide users on how to query and analyze data.

**Integration with BI Tools and Applications:** If you plan to use business intelligence (BI) tools or analytics applications, ensure that they are integrated with your data warehouse for easy access to data.

**Scaling and Growth Strategy:** Plan for the scalability and growth of our data warehouse. Consider how to handle increasing data volumes and performance demands.

**Compliance and Regulations:** Ensure that our data warehouse implementation complies with relevant data privacy and compliance regulations.

**TEAM MEMBERS:**

**JEYALAKSHMI SUBIKSHA S**

**HUDHA N**

**GAYATHRI S**