Requirements Specifications Document

- Introduction -
  - Purpose - The major purpose of this document is to track the behavior, condition of customers so that they can customize offers for them to buy insurance policies and also calculate royalties to those customers who bought policies in the past, this in turn will enhance their revenues.
  - Intended Audience and Use - Project Managers,Cloud Architect, Data Engineers, Data Analyst and Business Analyst and IT Administrators.
  - Product Scope - To leverage the Big Data ecosystem to enhance revenue, understand customers better, and improve overall business performance.
- Overall Description -

    In this project, data pipelines will be developed for the healthcare insurance company. These pipelines will enable the company to analyze customer behaviors and send targeted offers and royalties, thus informing strategic business decisions aimed at revenue enhancement.
  - User Needs -The product will be used by following users.
    - Cloud Architect: To design secure,scalable and cost-effective cloud platform for data services.
    - Data Engineers : To design, implement, and maintain the data pipelines, ensuring efficient extraction, transformation, and loading of data from various sources into the target data warehouse or storage system.
    - Data analysts: To extract insights from the data to support decision-making processes
    - Business managers : To gain a deeper understanding of customer behaviors, market trends, and competitive landscapes for strategic decision-making.
    - IT administrators: To manage the infrastructure and to ensure the reliability and security of the data pipelines.
    - Executives and stakeholders: To monitor key performance indicators (KPIs) and track the overall success of the business performance evaluation
    - Compliance officers : To ensure regulatory adherence and data privacy laws.

  - Assumptions:
    - Availability of data sources including third party sources
    - Infrastructure Availability(AWS services (S3, Redshift, EMR), andDatabricks)
    - Budget
    - Platform secure infrastructure
    - Data quality: nulls, uncleaned data
    - Regulatory Compliance

  - Dependencies

        - Availability of data sources : availability and accessibility of data from various sources, including competitor data and third-party sources. Latency, or reliability of those data could cause impact the functionality and performance of the data pipelines
        - Technical expertise : The effective deployment of the project relies on having proficient data engineers skilled in technologies such as AWS services (S3, Redshift, EMR), Databricks, and PySpark which could impact project timelines and solution quality, especially if there are gaps in knowledge or experience among team members.
        - Infrastructure Availability and Performance : Availability of resources plays vital for supporting the execution and scalability of data pipelines.

- System Features and Functional Requirements
    - Functional Requirements
        - Data integration from various sources.
        - Data cleaning and preprocessing.
        - Data analysis and visualization.
        - Automation of tasks.
        - Scalable architecture
    - External Interface Requirements
        - User: Project Managers,Cloud Architect, Data Engineers, Data Analyst and Business Analyst and IT Administrators.
        - Hardware: Laptop
        - Software: Cloud services like AWS(EMR, Redshift,Glue,S3), databricks,snowflakes,Python,PySpark,hadoop,Hive
        - Communications : Daily standup, following agile methodology within the above mentioned user
    - Non-functional Requirements
        - Performance Requirements
            - Efficient data processing speed.
            - High throughput.
            - Prompt response time.
            - Minimized latency.
            - Support for concurrent tasks.

        - Safety Requirement
            - Protection of sensitive data.
            - Compliance with regulations.
            - Implementation of encryption and access controls.

- Security Requirement
  - Authentication and authorization mechanisms.
  - Intrusion detection and prevention measures.
  - Protection against cyber threats and unauthorized access.

- Usability requirements
  - Intuitive user interfaces.
  - Clear documentation and training materials.
  - Accessibility for both technical and non-technical users.

- Scalability requirements :
  - Accommodation of data volume growth.
  - Handling increasing user demands.
  - Horizontal scaling and load balancing strategies.to achieve fault tolerance and availability.