Requirements Specifications Document

# Introduction – Big data eco system plays a vital role in transforming the structured, semi structured and unstructured data to make it accessible to data engineers, data scientist, business analyst as crucial information to generate valuable insights for business, hospitals and governments etc. These transformation can help the business to make decisions on the marketing campaigns or new product lunch. In today s world it is very important to analyze the data using various big data applications.

## Purpose –The project description clearly states to take the help of big data eco system to analyze the data of its competitors which has been collected from various sources to enhance the revenue of the given health care insurance company. The company wants to analyze the behavior and the conditions of the customers, from the data of various health care insurance companies in order to customize their offers to make the customers by their insurance policies and finally which will definitely help them to increase their revenue.

## Intended Audience and Use – The given data from the various sources are first given to the data engineers who will perform the transformations and actions, and make the refined data available to data scientists, business analysts, and company management to take right decisions.

## Product Scope – The product will be very beneficial for the insurance company, the given data when will be transformed for an analytical view, it will help the company to make the decisions on their campaigns and benefits that should be given to the policy buyers. The transformed data will help the data scientists to gather information about the behaviors of the customer, it will help to understand the key strategies that made the customers to buy the insurance policies from their competitors. It can be used to gather basic information’s of customers so that they can be targeted for social media ad campaigns directly.

## Definitions and Acronyms -Clearly define all key terms, acronyms, and abbreviations used in the SRS. This will help eliminate any ambiguity and ensure that all parties can easily understand the document.

# Overall Description – As per the given requirements, there are various things to consider while developing this product, there are sets of thirteen queries that is needed to be answered from the given data, the data itself needs a certain level of cleaning like removal of null values, elimination of duplicate values, creation of dimensional model, and overall ETL processes has to be done. This product is needed to get the results for given quires which will be crucial for data scientists, business analysts. This product is a complete new product and will might be integrated with other products of the company to make the comparison between the data of internal and external sources.

## User Needs – Data experts like data engineers, data scientists, and business analysts requires such product in order to understand the consumer behaviors and organizational policies that is required to enhances the products, services, implementation strategies and finally the revenue of the organization. The extracted output will help the experts to get hands on knowledge to make certain kinds of strategies and which will be used by all the company product or service development teams, marketing team to make a creative product and marketing strategies.

## Assumptions and Dependencies – Our assumptions are regarding the outcomes of the product which will help us to get the answers of certain types of questions in order to increase the revenue of the health insurance company, writing the data pipelines will help us to answer questions like number of claims on a particular disease, from which city the most of the claims are coming from etc., these questions will be answered by using the whole big data eco systems which has various dependencies on the software’s, clusters, cloud platform etc. These are the latest technologies that we are dependent on and assume that it will help us to find the answers to all the questions, so choosing the right kind of platforms is the most important decision as well.

# System Features and Requirements -In order to complete this product it requires a team of data engineers and manager who will manage the overall project execution, as this project requires various systems and platforms like cluster, warehouses, data lakes which will help the data to be properly distributed and processed and stored to be made accessible later for the data scientists, business analysts in future.

## Functional Requirements – The technologies that we required must have multiple functionalities like storage capacity, computation capacity, and processing capacities, there are other basics functions like project management, version control that must be required to properly manage the workflow of the development which must interact with each other to make the suitable environment for development.

## External Interface Requirements – In order to fulfill the functional requirements there are certain tools that are required such as:

### User

### Hardware

### Software: Amazon AWS cloud for Data lake S3, EMR servers for writing complex pipelines with spark, python, SQL etc, Data Warehouse like red shift.

### Communications: Zoom or microsoft teams, Gmail, git for sharing the code, zira for managing the sprint tasks to communicate about the project progress, updates, and further requirement.

## System Features - System features are a type of functional requirements. These are features that are required in order for a system to function.

## Nonfunctional Requirements - Nonfunctional requirements, which help ensure that a product will work the way users and other stakeholders expect it to, can be just as important as functional ones. These may include:

### Performance requirements

### Safety requirements

### Security requirements: I am role for user roles and permissions for certain functionalities.

### Usability requirements

### Scalability requirements: Hardware in AWS, Data Bricks, Azure etc.

## 