# Solution – *The project will be help us analyse the historic data from our competitors and help us create better offering to attract new customers and retain existing one. The detailed steps on how we will achieve this are as follows:*

1. Store the raw data in AWS S3.
2. Create access key and secret key in AWS so, I can access S3 data from Databricks.
3. Establish connection between S3 and Databricks so, the raw data are available for transformation in Databricks.
4. Use PySpark in Databricks to transform the data to help answer all the use cases mentioned below.
5. Create connection between.
6. Store the final tables into AWS redshift so, users can access it in format that is relevant to further analyse.

# Use Cases – *Below are all the use cases:*

1. Which disease has a maximum number of claims. >> no. of claims and disease
2. Find those Subscribers having age less than 30 and they subscribe any subgroup >> subscriber age, subgroup for subscriber less than 30yrs age
3. Find out which group has maximum subgroups. >> group and max number of subgroups in each group.
4. Find out hospital which serve most number of patients >> hospital and total number of patients.
5. Find out which subgroups subscribe most number of times >> most subscribed subgroups.
6. Find out total number of claims which were rejected >> total num claims rejected.
7. From where most claims are coming (city) >> city and number of claims.
8. Which groups of policies subscriber subscribe mostly Government or private. > groups with subscriber that are from govt offered health plan
9. Average monthly premium subscriber pay to insurance company. >> avg monthy premium and name of insurance company
10. Find out Which group is most profitable >> name of group and profit amt arranged in desc order
11. List all the patients below age of 18 who admit for cancer >> patients with age < 18 and disease = cancer.
12. List patients who have cashless insurance and have total charges greater than or equal for Rs. 50,000. >> name of patient where cashpayment = 0 or “cashless” and total charges > Rs 50,000
13. List female patients over the age of 40 that have undergone knee surgery in the past year >> patient name when sex= female and age > 40 and surgery type = knee
14. Database Design - List down all possible db(Redshift) tables here

## Tables Metadata Info with Pk/FK relationship: We will have 12 tables as we will be providing final data that will be readily usable and answer each of the thirteen use cases.

## ER diagram - *Optional*

# Technologies and Platforms to be used in this solution -*here are the list of tech we wil use :*

1. Amazon EMR.
2. AWS redshift
3. Jupyter Notebook
4. Jira
5. Github

## 