# Solution – *in this project we will create dummy data with first, last name, contact, salary and gender with 100 rows and help user/stakeholder to provide answer to use cases. See below solution steps:*

* 1. Create a random data with given column names and row limit as indicated using in built rand third party libraries (i.e. random and names/faker) PySpark code. This will be done on notebook offered by AWS Glue
  2. Save/write the data into csv as “employee.csv” and save into S3 bucket under folder name “RAW\_folder”
  3. Load the employee.csv data into glue-catalogue table using crawler.
  4. Access the glue-catalogue containing employee.csv from glue Visual ETL
  5. Transform the data in visual ETL to answer questions asked.
  6. Load the transformed data into redshift.

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

* 1. Create a dummy data with 100 rows and 6 columns (first name, last name, age, contact number, salary, gender) where age is 18-50 and salary between 100k-200k.
  2. Analyse the salary and specify Y or N under new column named “status” if salary is more than 120K and N if less.
  3. Load the transformed data as parquet into same s3 bucket and save it under folder name “master folder”. Make sure the data is partitioned based on “status”
  4. Now, create a hive table and use Athena to run the simple query to see the data from hive.

OR

* 1. Load the s3 data in “master folder” into Redshift and run a simple query.

1. Database Design - List down all possible DB(Redshift) tables here

## Tables Metadata Info with Pk/FK relationship: We will have 1 table in redshift or hive and their schema is below:

## ER diagram -

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

1. Databricks
2. AWS redshift
3. S3 bucket
4. Github

\

* + - 1. Create two folder with names master and raw
      2. Go to glue, click notebook and so, to run pyspark code.
      3. Now, create dummy data here in notbook
      4. Save in into raw folder you created in step 1

## 