

# Health and Insurance – Project Explanation

## #Project Flow

My previous project name is Health and insurance and it is in pharma domain.

In this project the source was structured data which was stored in RDBMS(Oracle- version 12g)

we imported this data using sqoop ETL and stored this into HDFS.

after that we created the table in hive on this imported data and did the analysis as per the business requirement.

here we applied various transformation like join and other aggregation.

once data is duplicated or we need to filter data based on latest record

we do the deduplication on hive data.

And extracted data we stored into S3 and further team was accessing data from s3 for their purpose like data visualization.

## #UT(Unit Testing)

Also implementing the business logic and code we do the unit testing as ,

- 1.count check : check source and target data validation.
- 2.Null check : we check the mandatory fields should not be null in target
- 3.Duplicate Check : we check the data should not be duplicated in Target
- 4.Detailed Check : we check the imported data should be accurate

## #Project Role

Developer/Data Engineer/Bigdata Developer

## #Project Responsibilities :

- 1.To write the script to import and process the data
- 2.To do the automation

- 3.To write the script for to create required tables
- 4.To monitor the job
- 5.to optimize the Job
- 6.To validate the Data is it populated as expected
- 7.To provide the KT to new joiner
- 8.To report the progress to mgr.

This is all about project flow,UT and responsibilities

\*\*\*\*\*END\*\*\*\*\*