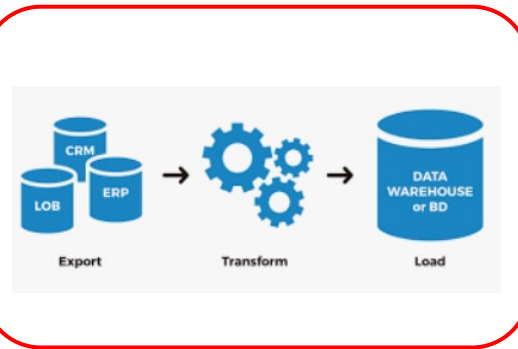# Business Model

Vendors

Asher Healthcare
Solutions (AHS)

Karis Pharma

Business Model

Asher Healthcare Solutions:

- ✓ Provider of technology solutions, analytics and high-value data to different biopharmaceutical manufacturers, healthcare providers and government agencies.

- ✓ Buy Raw Medical Data from vendors, Develop Analytical Reports/Products and sell it to Top Pharma Companies.

Karis Pharma:

- ✓ One of the World's premier Bio-Pharmaceutical company.

- ✓ Successfully developed a Vaccine for Covid and got all their approval to launch across the globe.

- ✓ Approached AHS to provide COVID Vaccine Related Analytical Reports.

Functional Specification:

To start with, Karis wants below 2 Reports :

1. USA City report.

- ✓ **Number of distinct prescribers assigned for each city.**
- ✓ **Total TRX_CNT prescribed in each city.**
- ✓ **Number of zips in each city**
- ✓ **Do not report a city if no prescriber is assigned to.**

File Type :json
Compression Type: bzip2
Location Type: S3, Azure Blob
Number of split Files: 1

2. USA Prescriber Report.

Top 5 Prescribers with the highest TRX_CNT in each state. Consider the Prescribers only with working experience from 20 to 50 Years.
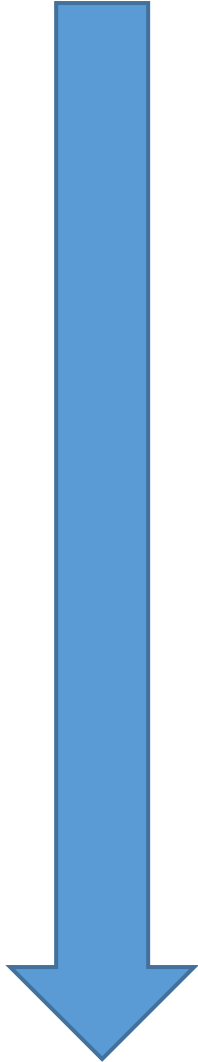
File Type :orc
Compression Type: snappy
Location Type: S3, Azure Blob
Number of split Files: 2

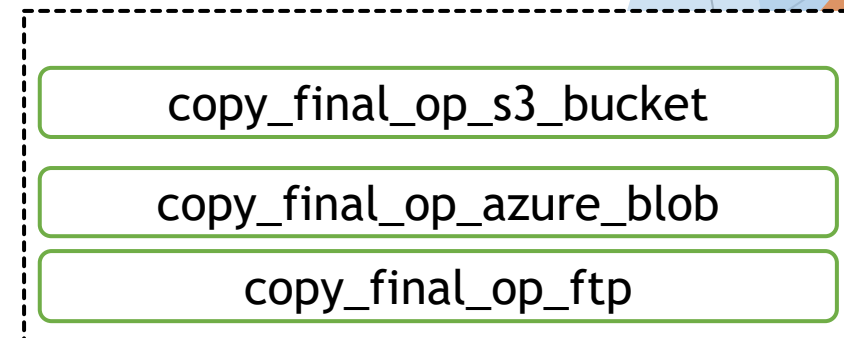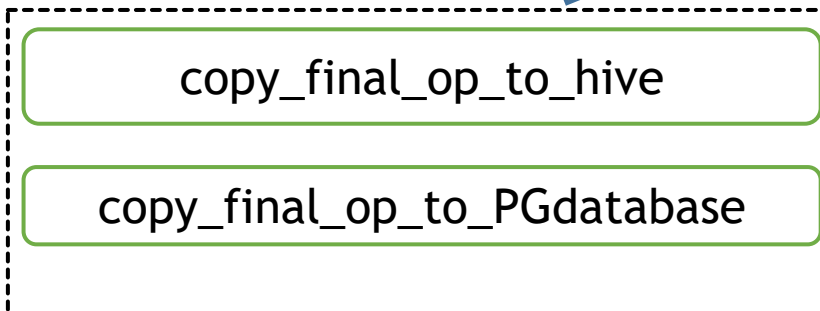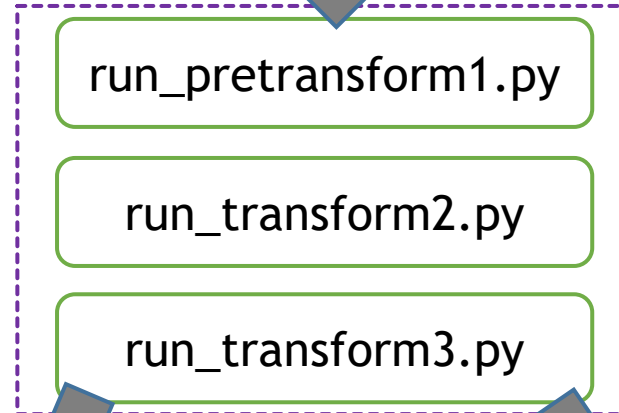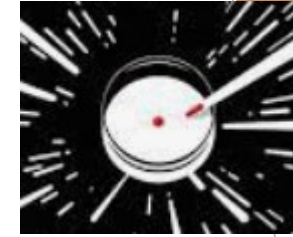# Project Flow

Data Pipeline

Data Ingestion

Pre-Processing

Transformation

Storage

Project Flow :

run_ingest.py

run_pipeline.py

run_pretransform1.py

run_transform2.py

run_transform3.py

copy_final_op_to_hive

copy_final_op_to_PGdatabase

copy_final_op_s3_bucket

copy_final_op_azure_blob

copy_final_op_ftp
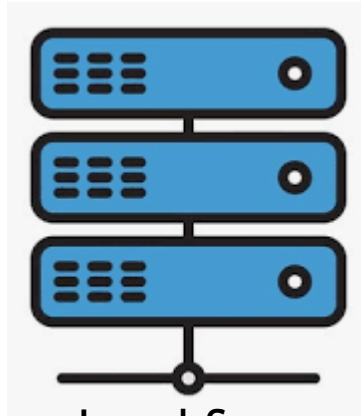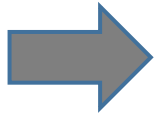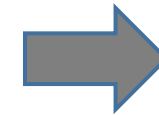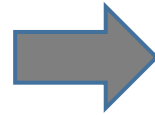
Data Ingestion:

Transformations

Pre-Transform 1   ...   Pre-Transform n   Transform 1   ...   Transform n
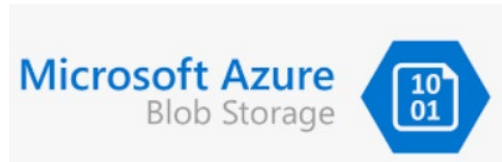
Storage

Transfer

Project Flow (Code Level)

presc_run_pipeline.py

get_all_variables

create_objects

presc_run_data_ingest

presc_run_data_preprocessing

presc_run_data_transform

presc_run_data_extract

validate

Logging

Error Handling

Copy to Local Server

Do not report city if
No Prescriber
Is Assigned

Calculate the Number
of Zips in each City

Cities

Calculate the total
TRX_CNT for each city

Calculate the Number
of Distinct Prescribers
assigned to each City

Prescriber Report

Apply a filter to consider the prescribers only from 20 to 50 years of experience

Rank the Prescribers based on their TRX_CNT for each state.

Select top 5 prescribers from each state.

Part 1

presc_run_pipeline.py

Copy Input files to HDFS

get_all_variables

create_objects

presc_run_data_ingest

presc_run_data_preprocessing

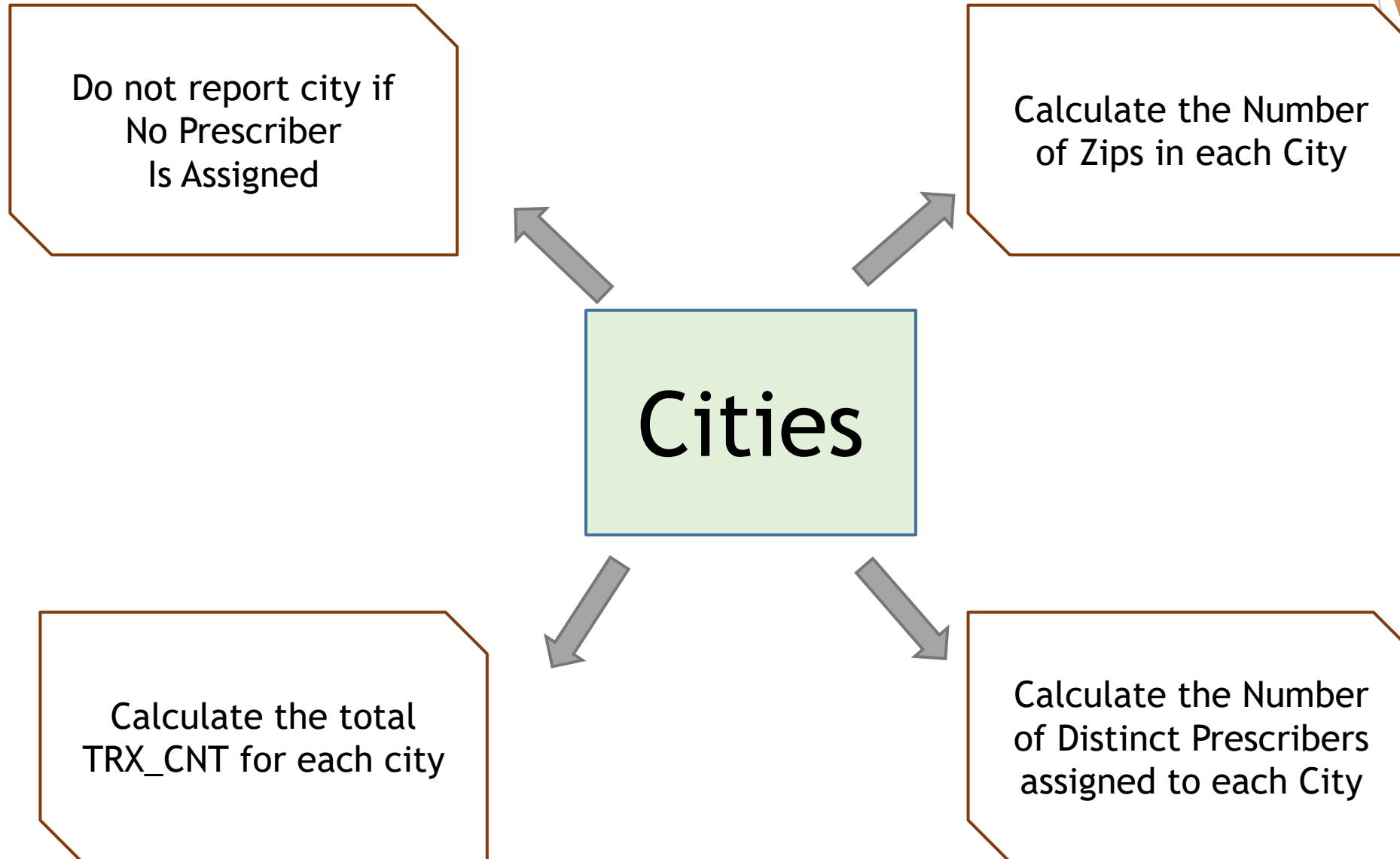presc_run_data_transform
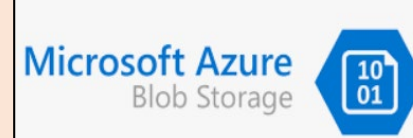
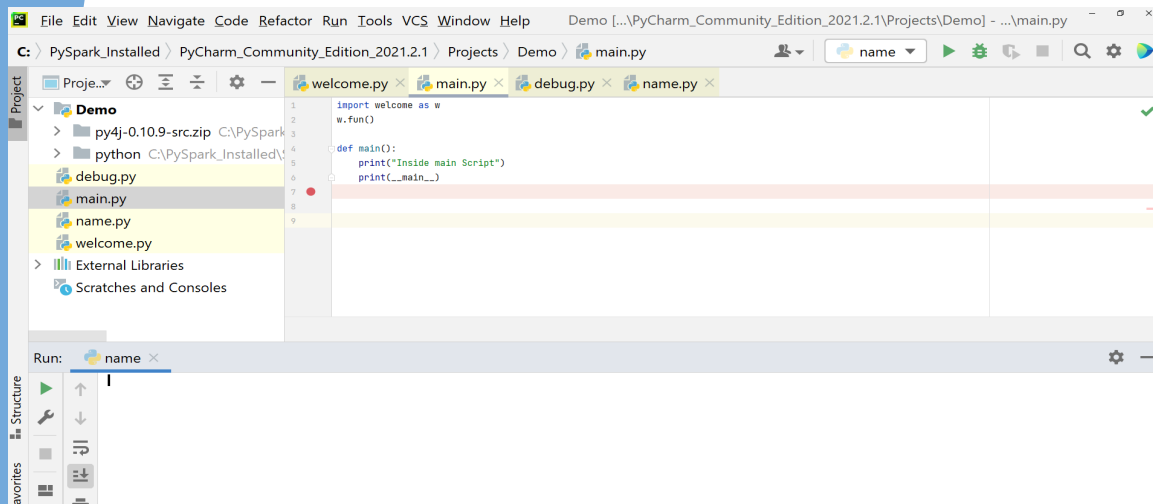presc_run_data_extract

validate

Logging

Error Handling

Part 3

Part 2

Copy to Local Server

amazon S3
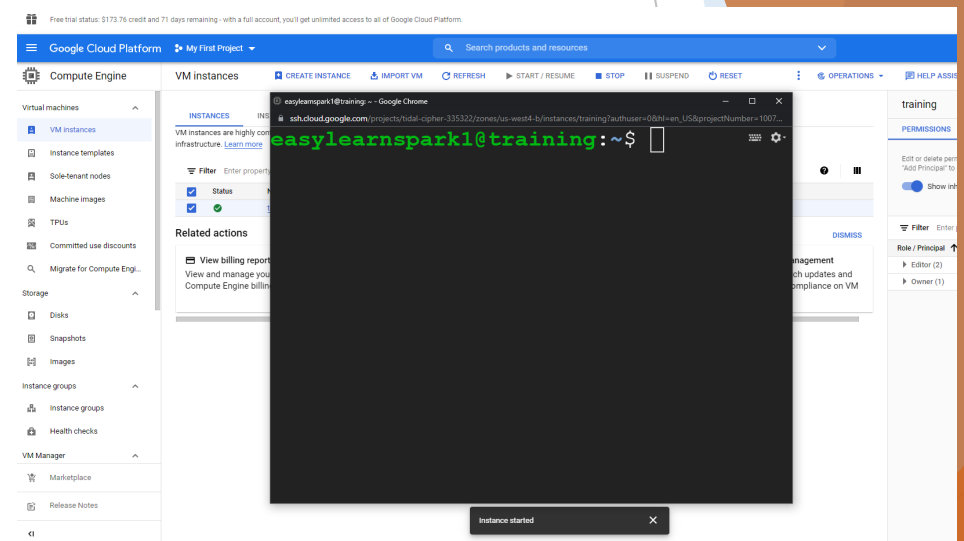
Microsoft Azure Blob Storage

Project Approach

Test and Debug at Windows PyCharm

Deployment at Single Node Cluster

Download City Dimension File at below Link:

https://prescpipeline.blob.core.windows.net/input-vendor-data/city/us_cities_dimension.parquet?st=2022-04-21T14:19:25Z&se=2022-12-31T22:19:25Z&si=read&spr=https&sv=2020-08-04&sr=c&sig=wjY0KtPvyy%2BbIpopBqMKAGmmSHsSvLhqL0n%2BBGFVXOQ%3D

Download Prescriber Fact File at below Link:

https://prescpipeline.blob.core.windows.net/input-vendor-data/presc/USA_Presc_Medicare_Data_2021.csv?st=2022-04-21T14:19:25Z&se=2022-12-31T22:19:25Z&si=read&spr=https&sv=2020-08-04&sr=c&sig=wjY0KtPvyy%2BbIpopBqMKAGmmSHsSvLhqL0n%2BBGFVXOQ%3D