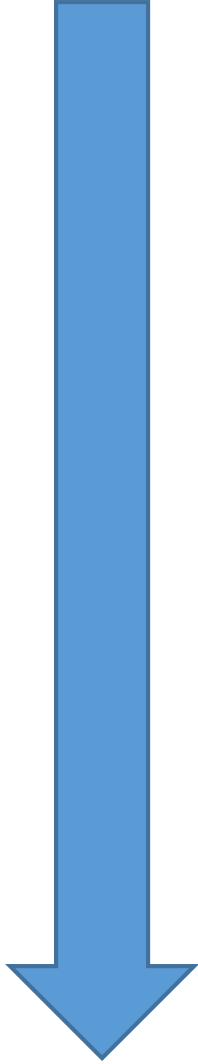


Project Flow

The background features abstract geometric shapes, primarily triangles and lines, in shades of blue and orange. These shapes are layered and overlap, creating a dynamic, modern aesthetic. The colors range from light, airy blues to deep, rich oranges and greys.

Data Pipeline



Data Ingestion



Pre-Processing



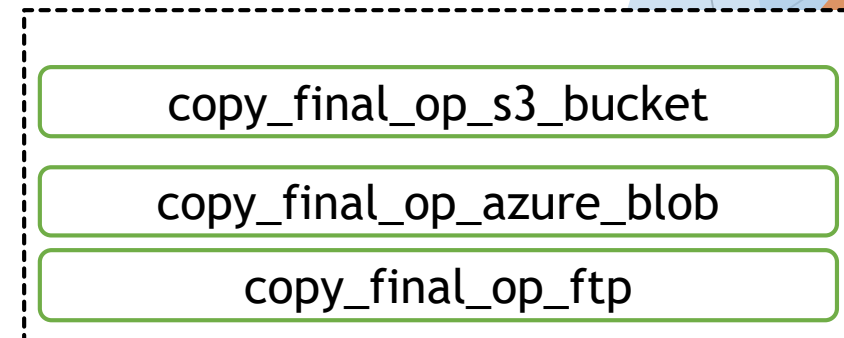
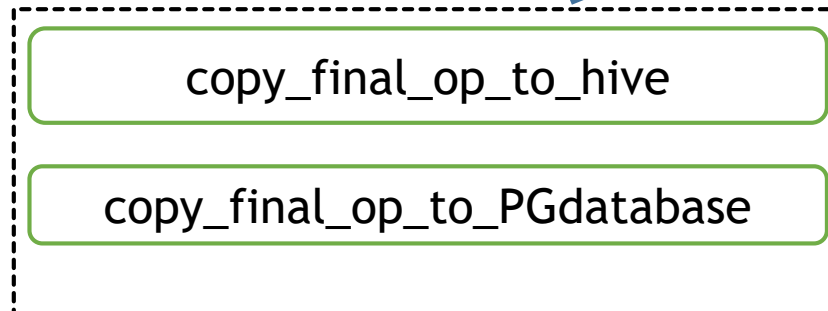
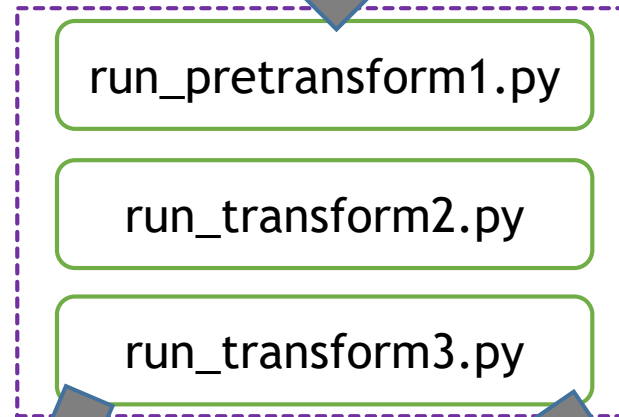
Transformation



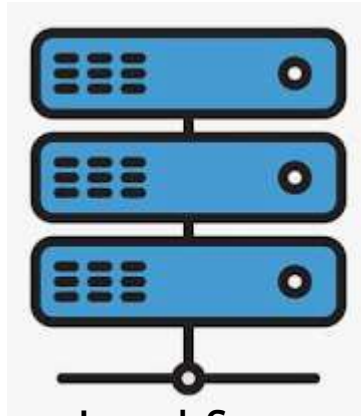
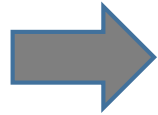
Storage

Project Flow :

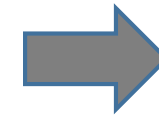
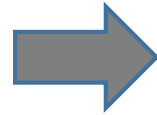
run_pipeline.py



Data Ingestion:



Local Server



Spark



Transformations



Pre-Transform 1

...



Pre-Transform n



Transform 1

...



Transform n

Storage

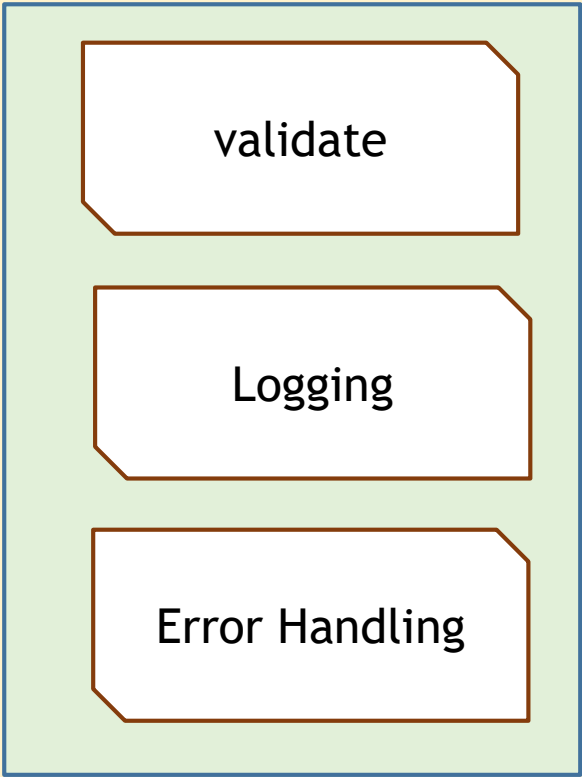
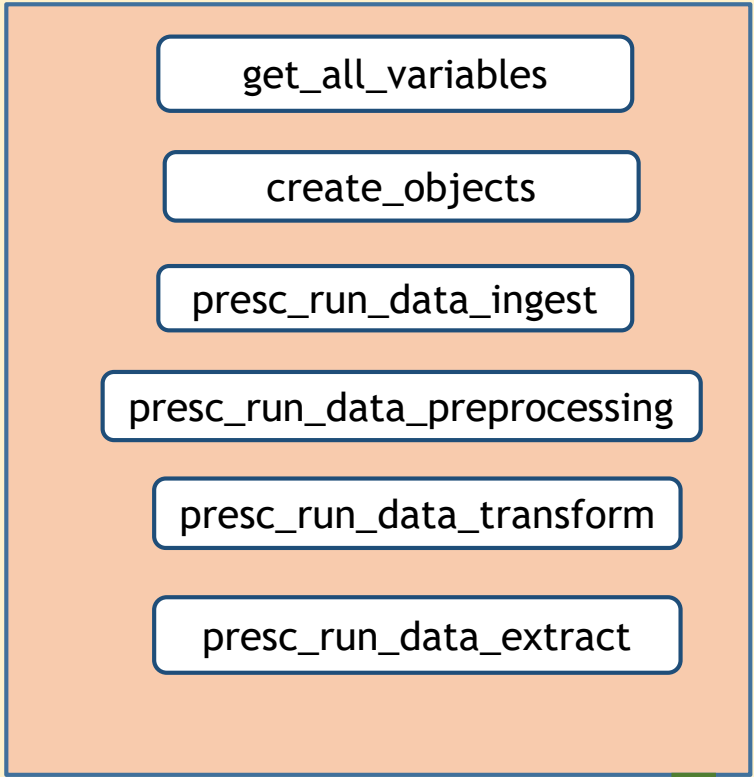


Transfer



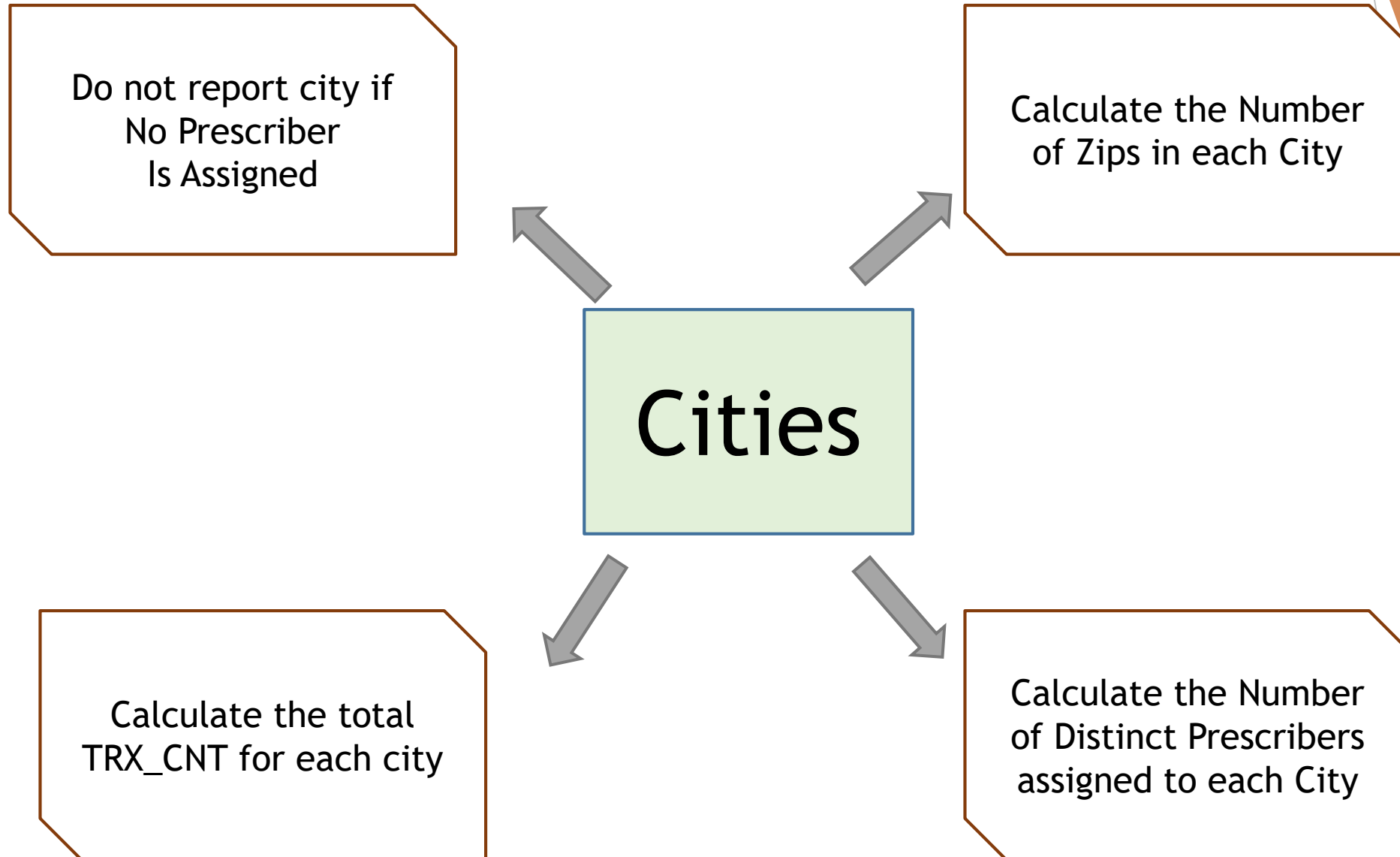
Project Flow (Code Level)

presc_run_pipeline.py



Copy to Local Server





Prescriber Report

```
graph LR; A[Prescriber Report] --> B[Apply a filter to consider the prescribers only from 20 to 50 years of experience]; A --> C[Rank the Prescribers based on their TRX_CNT for each state.]; A --> D[Select top 5 prescribers from each state.];
```

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



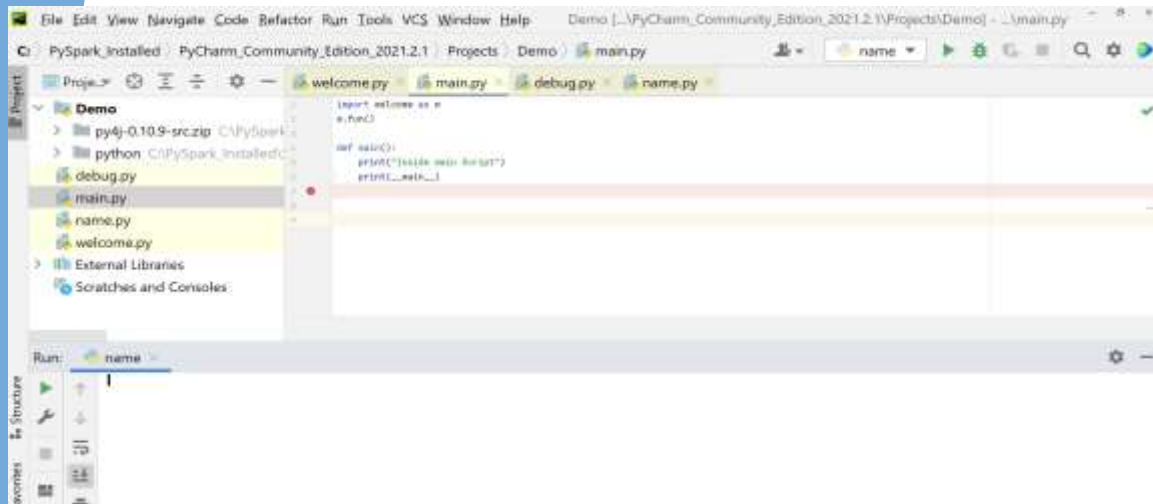
Part 2

Copy to Local Server

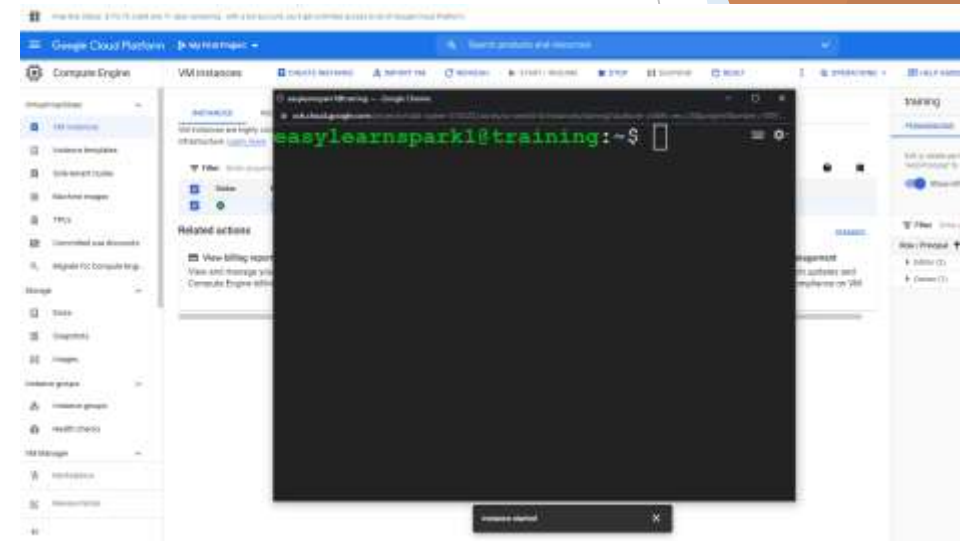


Project Approach

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=wjY0KtPvy%2BblpopBqMKAGmmSHsSvLhqL0n%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=wjY0KtPvy%2BblpopBqMKAGmmSHsSvLhqL0n%2BBGFVXOQ%3D