Crawling few pages from Amazone 5 pages

Create a crawler to get the Data from list of books

Title

Amount

Author

Rating

Link\_book

Load it to a Table

Create another crawler

Review

User name

Title

Description if present

Rating

Reviewed Date

Get the review of the book max 5 to 6 per book

Error handling must be done

Follow the pep8 formats

Made as modularized as you can

Now The Source Data has been prepared

We need to make pipeline to load the data from Mysql Onprem to Snowflake Using Airflow and Snowsql

CREATE DATABASE amazonebooks;

USE DATABASE amazonebooks;

USE SCHEMA public;

CREATE TABLE amazone\_books (

book\_id INT NOT NULL AUTO\_INCREMENT

,book\_title TEXT

,book\_amount FLOAT

,book\_author TEXT

,book\_rating FLOAT

,book\_link TEXT

,business\_date DATE DEFAULT(CURRENT\_DATE)

,PRIMARY KEY (book\_id)

);

CREATE OR REPLACE STAGE amazone\_books\_stage

FILE\_FORMAT = (TYPE= csv FIELD\_DELIMITER = ',' FIELD\_OPTIONALLY\_ENCLOSED\_BY = '"' SKIP\_HEADER = 1);

In only using SnowSQL we can use PUT command

USE DATABASE amazonebooks;

PUT 'file:///<PATH>/L\_AIRPORT\_ID.csv' @amazone\_books\_stage;

USE DATABASE dev\_lnd;

COPY INTO amazone\_books FROM @amazone\_books\_stage;

Once Completed remove

REMOVE @amazone\_books\_stage;