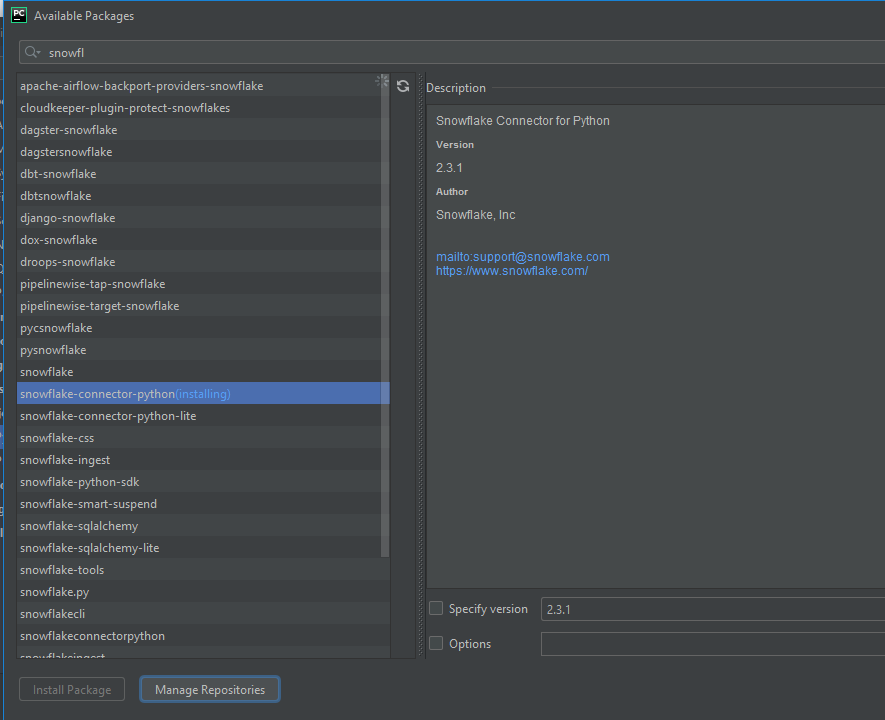
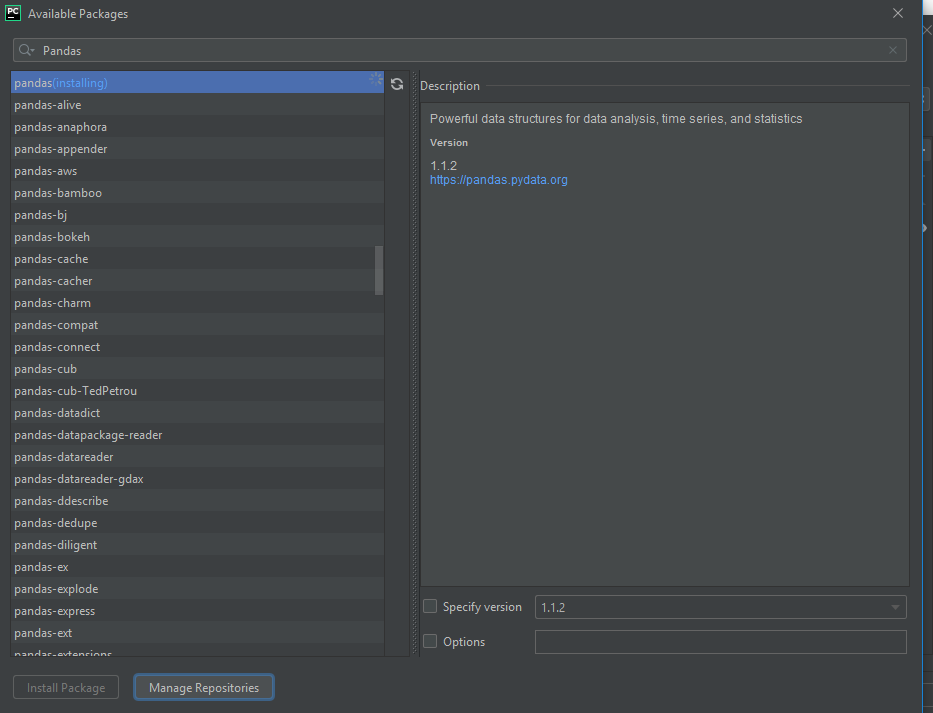
# Import snowflake connector python

File 🡪 Settings🡪 Project 🡪 Project Interpreter 🡪 click on plus mark

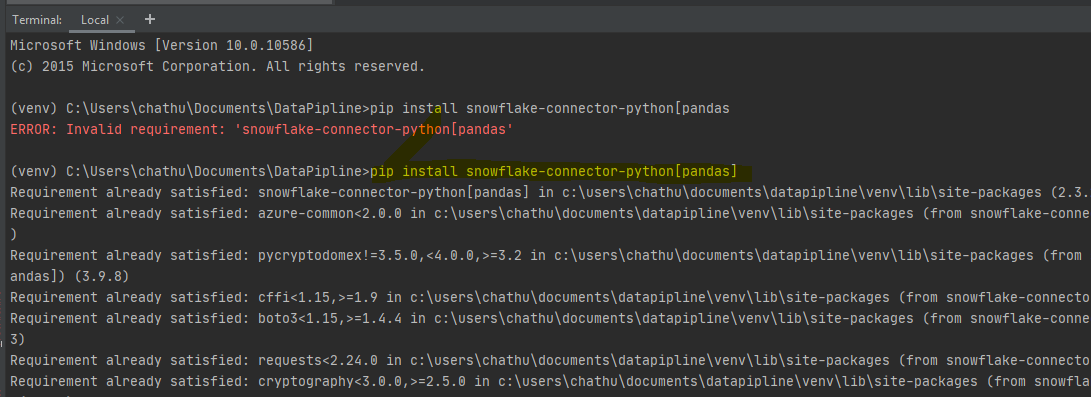


Installing Pandas

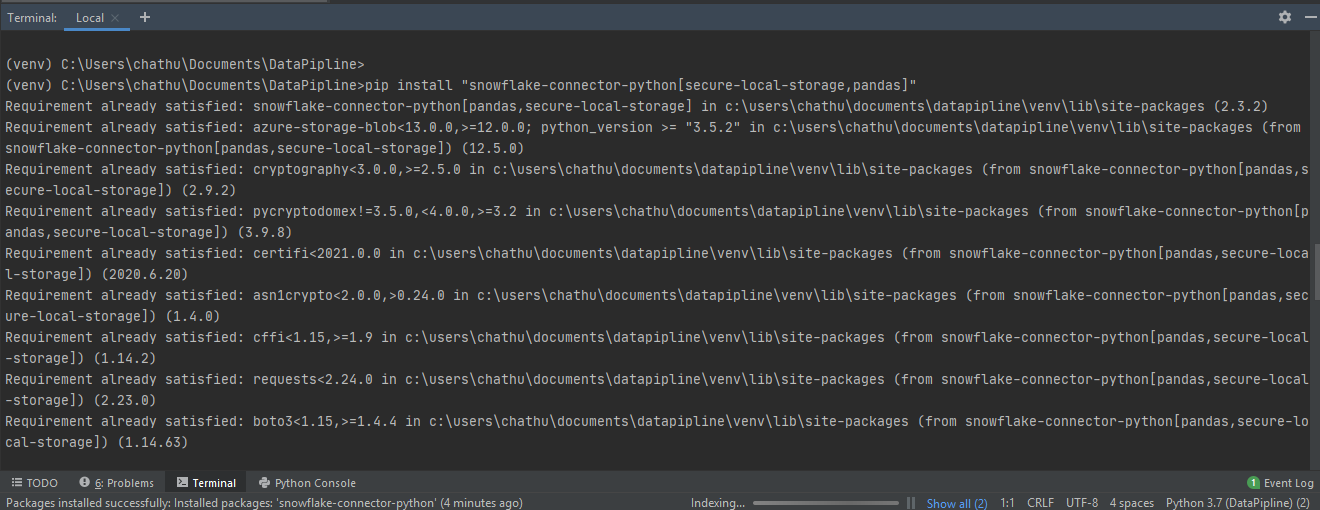


# Snowflake connector for python

pip install snowflake-connector-python[pandas]



pip install "snowflake-connector-python[secure-local-storage,pandas]"



# Q1.

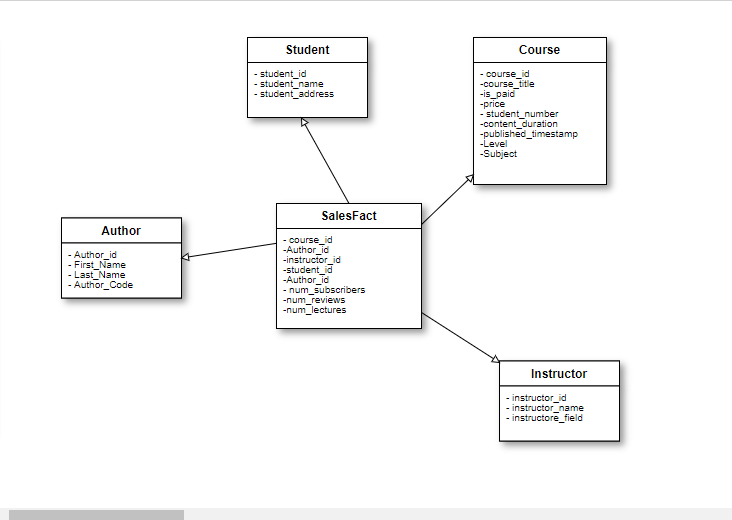
1. Create a pipeline using python and suitable python libraries. This pipeline data source should be course\_data.json file (flat files) and destination should be a database in Snowflake (<https://www.snowflake.com/>). Please create a trial account in Snowflake which provides you $400 credit for a 30-day trial. (Note: for assignment purposes you can keep the data source file in a local disk location and run the python pipeline in your local machine to load the data to Snowflake).

**Trial Account and python pipeline to load the data to Snowflake**

user='DARSHI',  
 account='ni10492.us-central1.gcp',  
 password='Chathu@1992',  
 warehouse='COMPUTE\_WH',  
 database='TEST\_DATA',  
 role='SYSADMIN',  
 schema='PUBLIC'

import pandas  
import snowflake.connector  
from snowflake.connector.pandas\_tools import write\_pandas  
  
conn = snowflake.connector.connect(  
 user='DARSHI',  
 account='ni10492.us-central1.gcp',  
 password='Chathu@1992',  
 warehouse='COMPUTE\_WH',  
 database='TEST\_DATA',  
 role='SYSADMIN',  
 schema='PUBLIC')  
  
print('connect')  
df1= pandas.read\_json(r'C:\Users\chathu\Desktop\wiley\course\_data.json')  
  
#print(abalone\_data)  
cur = conn.cursor()  
cur.execute("""TRUNCATE TABLE TEST\_DATA.PUBLIC.COURSE\_SALE """)  
  
df1 = df1[['course\_id',  
 'course\_title',  
 'is\_paid',  
 'num\_subscribers',  
 'price',  
 'author',  
 'content\_duration',  
 'level',  
 'num\_lectures',  
 'num\_reviews',  
 'published\_timestamp',  
 'subject']]  
  
  
print(df1)  
  
try:  
  
  
 write\_pandas(conn, df1, 'COURSE\_SALE')  
 conn.commit()  
finally:  
 cur.close()

1. Create a suitable star schema to analyze course sales adhering to dimensional modelling best practices and load the data into this star schema in Snowflake using your python pipeline. Also provide relevant diagrams for the star schema.



1. Write 5 analytical queries based one your star schema. Also provide explanations on what each query output denotes.

**pending**

1. Write a single-page report highlighting the challenges that you faced when completing the above-mentioned activities.

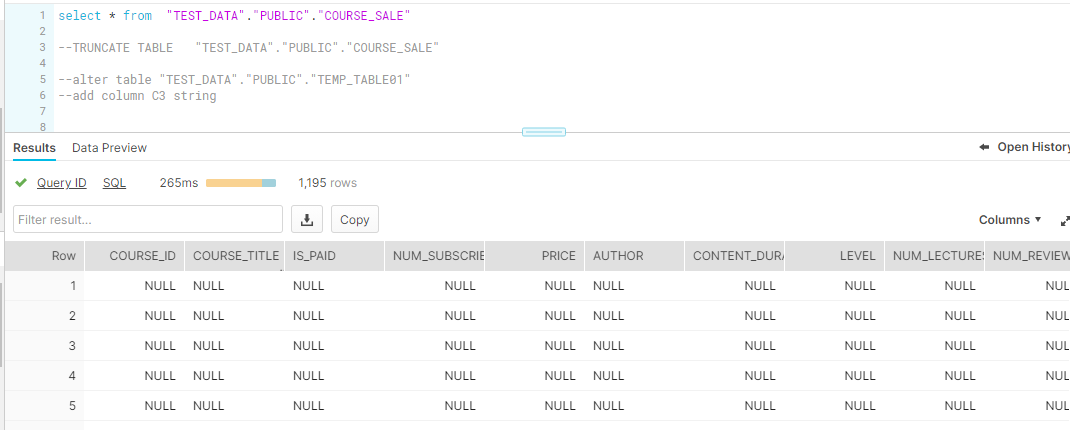
Challenges-01

Data will be displayed as null values in snowflake db because panda is a case sensitive Snowflake's default all-uppercase column names.

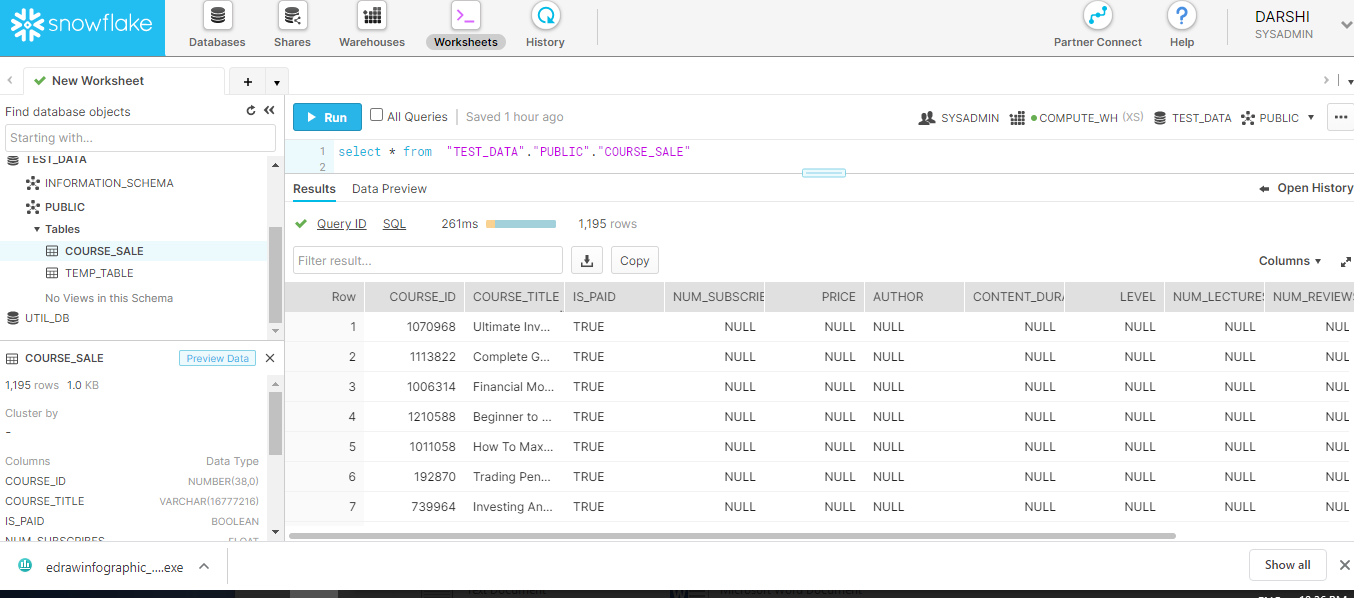
Solved upper case issue.

import pandas  
import snowflake.connector  
from snowflake.connector.pandas\_tools import write\_pandas  
  
conn = snowflake.connector.connect(  
 user='DARSHI',  
 account='ni10492.us-central1.gcp',  
 password='Chathu@1992',  
 warehouse='COMPUTE\_WH',  
 database='TEST\_DATA',  
 role='SYSADMIN',  
 schema='PUBLIC')  
  
print('connect')  
df1= pandas.read\_json(r'C:\Users\chathu\Desktop\wiley\course\_data.json')  
  
cur = conn.cursor()  
cur.execute("""TRUNCATE TABLE TEST\_DATA.PUBLIC.COURSE\_SALE """)  
  
df1 = df1[['course\_id',  
 'course\_title',  
 'is\_paid',  
 'num\_subscribers',  
 'price',  
 'author',  
 'content\_duration',  
 'level',  
 'num\_lectures',  
 'num\_reviews',  
 'published\_timestamp',  
 'subject']]  
  
df1.rename(columns={'course\_id':'COURSE\_ID',  
 'course\_title':'COURSE\_TITLE',  
 'is\_paid':'IS\_PAID',  
 'num\_subscribers':'NUM\_SUBSCRIBERS',  
 'price':'PRICE',  
 'author':'AUTHOR',  
 'content\_duration':'CONTENT\_DURATION',  
 'level':'LEVEL',  
 'num\_lectures':'NUM\_LECTURES',  
 'num\_reviews':'NUM\_REVIEWS',  
 'published\_timestamp':'PUBLISHED\_TIMESTAMP',  
 'subject':'SUBJECT'  
 },  
 inplace=True)  
print(df1)  
try:  
 write\_pandas(conn, df1, 'COURSE\_SALE')  
 conn.commit()  
finally:  
 cur.close()

Before Solved



After that Solved for Course\_id, Course\_tittle,IS\_Paid , column



Challenge 02 – Data format ,Data conversion,data type mapping with pandas and snowflake

This is pending.