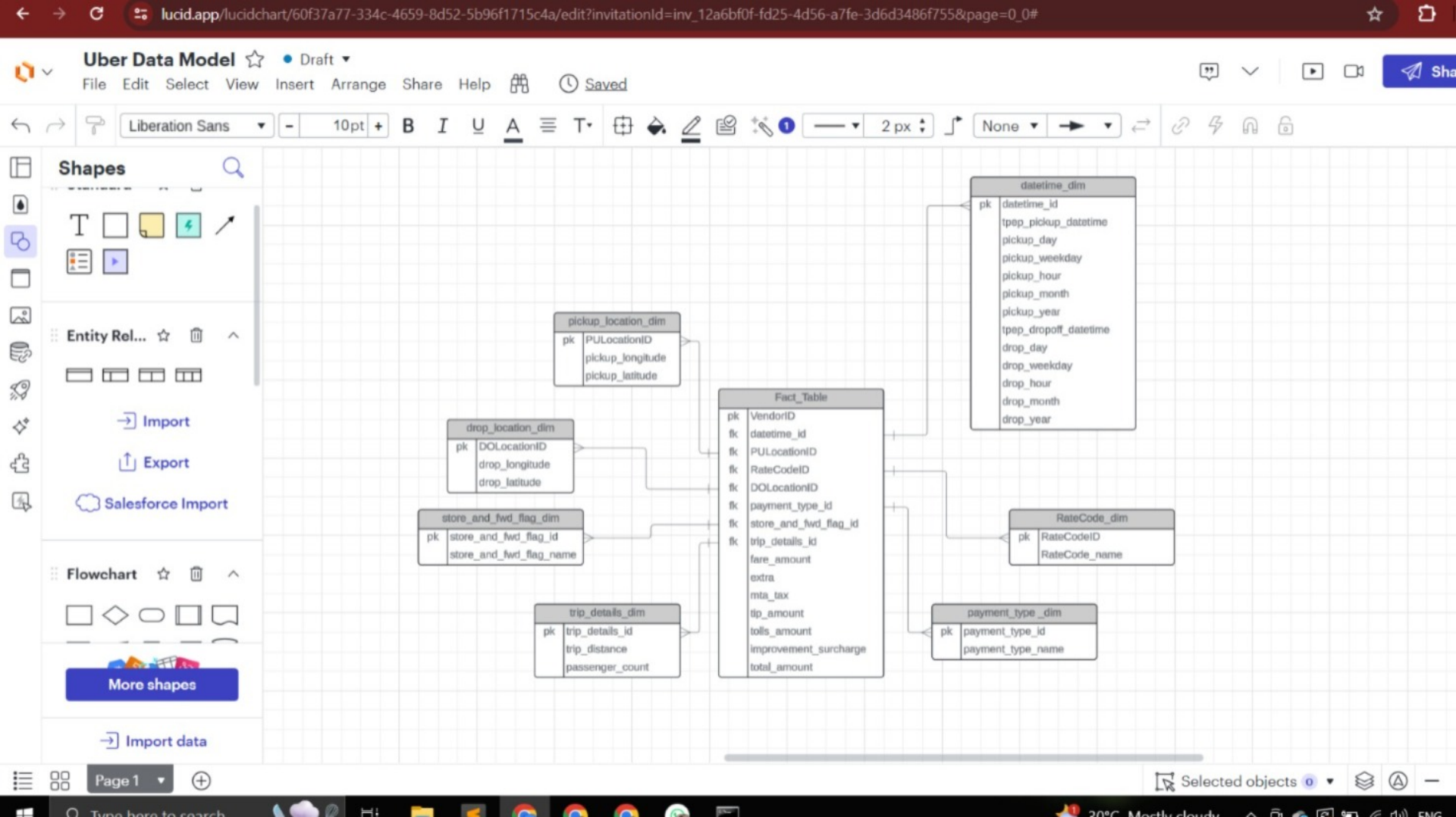

lucid.app/lucidchart/60f37a77-334c-4659-8d52-5b96f1715c4a/edit?invitationId=inv_12a6bf0f-fd25-4d56-a7fe-3d6d3486f755&page=0_0#
Uber Data Model
Draft
File Edit Select View Insert Arrange Share Help
Saved
Liberation Sans
10pt
2 px
None
Shapes
Entity Rel...
Import
Export
Salesforce Import
Flowchart
More shapes
Import data
datetime_dim
pk datetime_id
tpep_pickup_datetime
pickup_day
pickup_weekday
pickup_hour
pickup_month
pickup_year
tpep_dropoff_datetime
drop_day
drop_weekday
drop_hour
drop_month
drop_year
pickup_location_dim
pk PULocationID
pickup_longitude
pickup_latitude
drop_location_dim
pk DOLocationID
drop_longitude
drop_latitude
store_and_fwd_flag_dim
pk store_and_fwd_flag_id
store_and_fwd_flag_name
trip_details_dim
pk trip_details_id
trip_distance
passenger_count
Fact_Table
pk VendorID
fk datetime_id
fk PULocationID
fk RateCodeID
fk DOLocationID
fk payment_type_id
fk store_and_fwd_flag_id
fk trip_details_id
fare_amount
extra
mta_tax
tip_amount
tolls_amount
improvement_surcharge
total_amount
RateCode_dim
pk RateCodeID
RateCode_name
payment_type_dim
pk payment_type_id
payment_type_name
Page 1
Selected objects 0

Free trial status: ₹24,736.28 credit and 87 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.
DISMISS ACTIVATE

Attention Customers from India: As of October 1st, due to regulatory changes by the Reserve Bank of India, we will not be able to charge your card for recurring payments if you have not set up an e-mandate or if the transaction is above 5000 INR. If your payments are being declined, Billing administrators will have to log into their Payments account and manually reauthorize the payments. Read more
DISMISS

Bucket details

GO TO PATH REFRESH LEARN

uber-aataanalytics-kaleem

Location	Storage class	Public access	Protection
asia-south1 (Mumbai)	Standard	Subject to object ACLs	Soft Delete

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser

uber-aataanalytics-kaleem

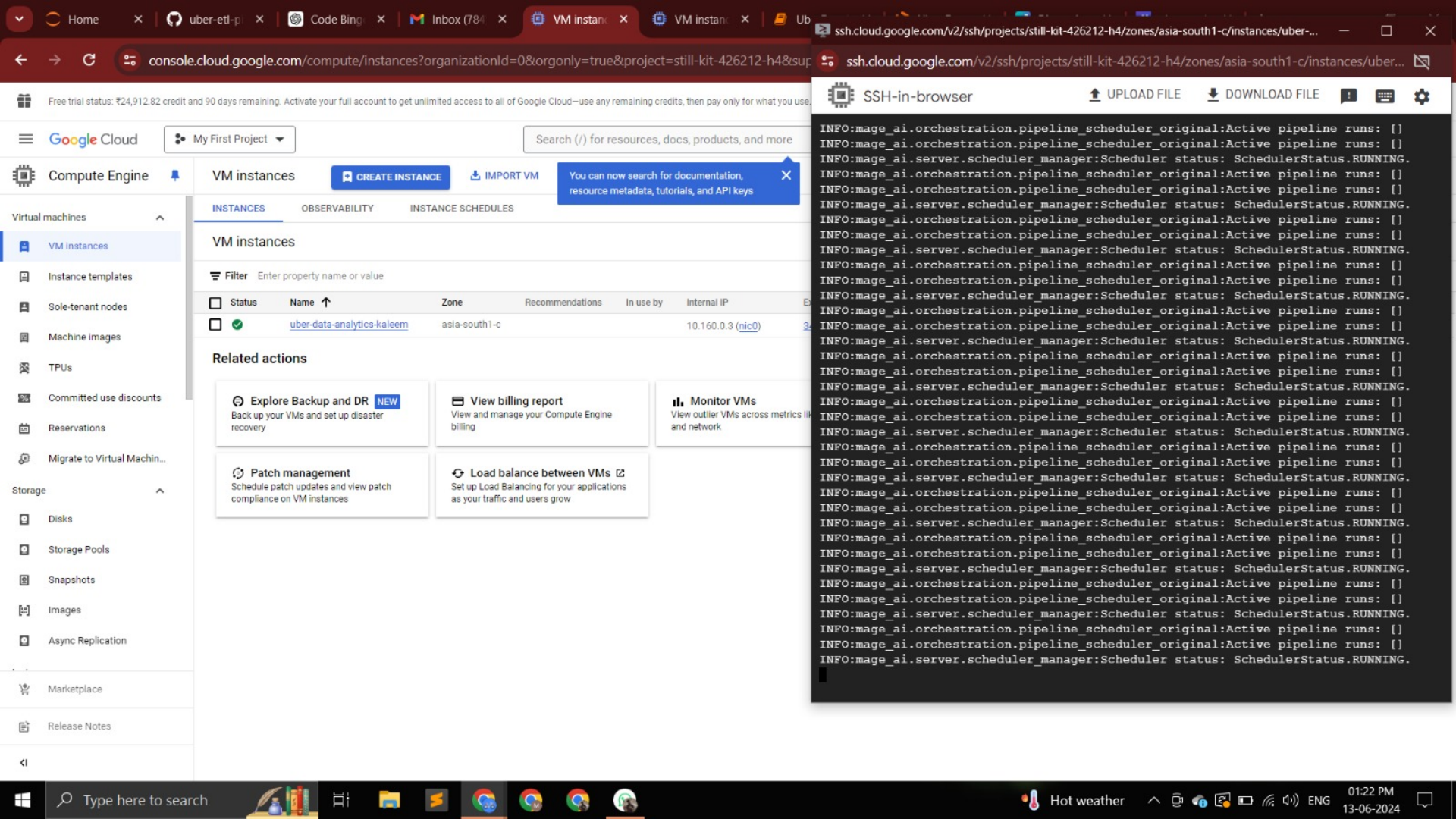
Buckets > uber-aataanalytics-kaleem

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS EDIT RETENTION
DOWNLOAD DELETE

Filter by name prefix only | Filter | Filter objects and folders | Show Live objects only

Name	Last modified	Public access	Version history
uber_data.csv	Jun 12, 2024, 11:10:12 PM	Public to internet Copy URL	–

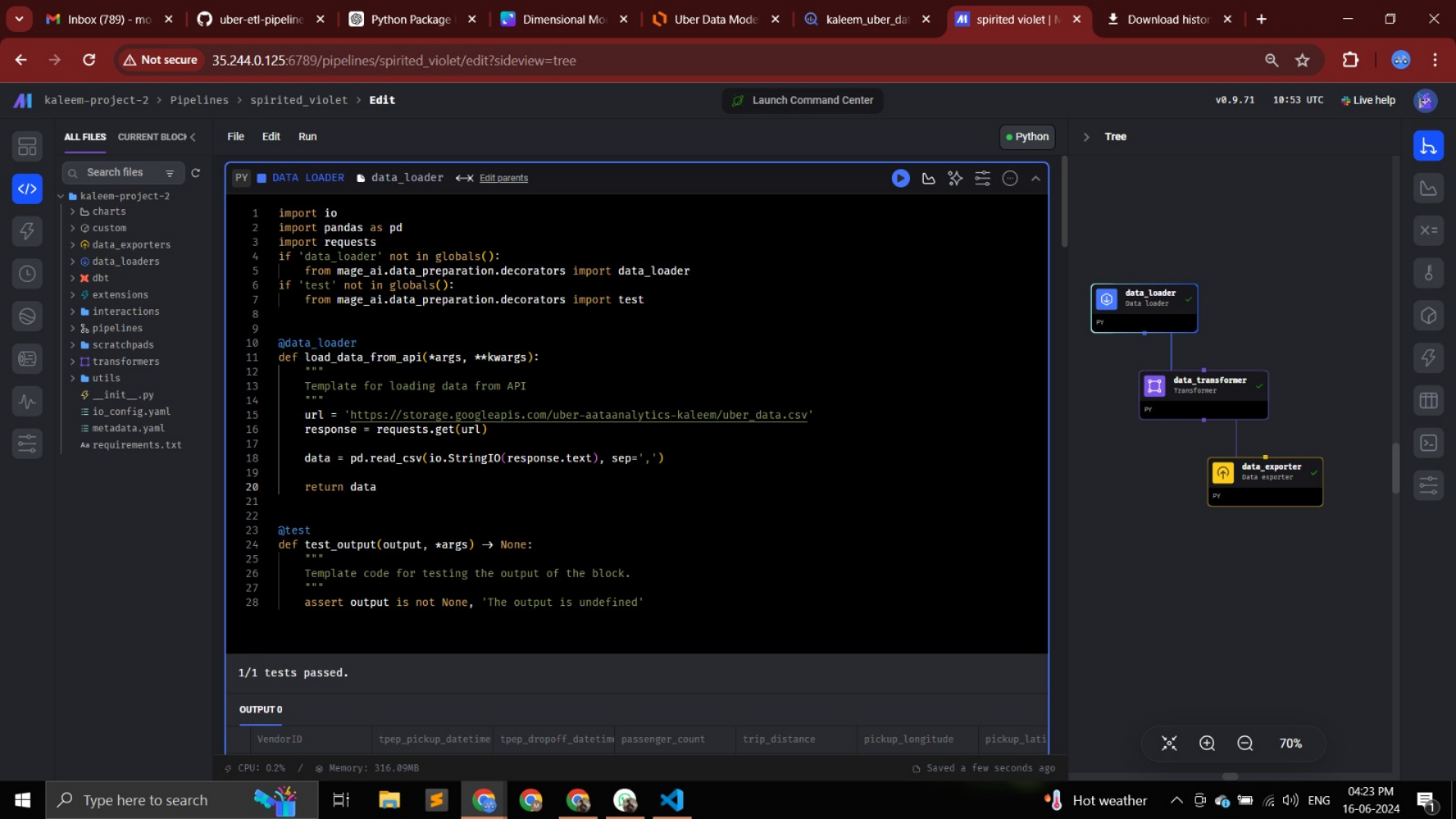
Type here to search 30°C Partly sunny ENG 03:03 PM 19-06-2024


Home
uber-etl-p
Code Bing
Inbox (784
VM instanc
VM instanc
console.cloud.google.com/compute/instances?organizationId=0&orgonly=true&project=still-kit-426212-h4&sup
Free trial status: ₹24,912.82 credit and 90 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.
Google Cloud
My First Project
Search (/) for resources, docs, products, and more
Compute Engine
VM instances
CREATE INSTANCE
IMPORT VM
You can now search for documentation, resource metadata, tutorials, and API keys
INSTANCES
OBSERVABILITY
INSTANCE SCHEDULES
Virtual machines
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Committed use discounts
Reservations
Migrate to Virtual Machin...
Storage
Disks
Storage Pools
Snapshots
Images
Async Replication
Marketplace
Release Notes
Filter
Enter property name or value
Status
Name
Zone
Recommendations
In use by
Internal IP
uber-data-analytics-kaleem
asia-south1-c
10.160.0.3 (nic0)
Related actions
Explore Backup and DR
NEW
Back up your VMs and set up disaster recovery
View billing report
View and manage your Compute Engine billing
Monitor VMs
View outlier VMs across metrics
and network
Patch management
Schedule patch updates and view patch compliance on VM instances
Load balance between VMs
Set up Load Balancing for your applications as your traffic and users grow
ssh.cloud.google.com/v2/ssh/projects/still-kit-426212-h4/zones/asia-south1-c/instances/uber-...
SSH-in-browser
UPLOAD FILE
DOWNLOAD FILE
INFO:mage_ai.orchestration.pipeline_scheduler_original:Active pipeline runs: []
INFO:mage_ai.server.scheduler_manager:Scheduler status: SchedulerStatus.RUNNING.
Type here to search
Hot weather
ENG
01:22 PM
13-06-2024

ssh.cloud.google.com/v2/ssh/projects/still-kit-426212-h4/zones/asia-south1-c/instances/uber-data-analytics-kaleem?authuser=0&hl=en_US&projectNumber=705867784738&useAdminProxy=true - Google Chrome
ssh.cloud.google.com/v2/ssh/projects/still-kit-426212-h4/zones/asia-south1-c/instances/uber-data-analytics-kaleem?authuser=0&hl=en_US&projectNumber=705867784738&useAdminProxy=true
SSH-in-browser
UPLOAD FILE
DOWNLOAD FILE

```
Linux uber-data-analytics-kaleem 6.1.0-21-cloud-amd64 #1 SMP PREEMPT_DYNAMIC Debian 6.1.90-1 (2024-05-03) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
mohammedkaleemullah06@uber-data-analytics-kaleem:~$ sudo apt update
sudo apt install pipx
sudo pipx ensurepath
Get:1 file:/etc/apt/mirrors/debian.list Mirrorlist [30 B]
Get:2 file:/etc/apt/mirrors/debian-security.list Mirrorlist [39 B]
Hit:7 https://packages.cloud.google.com/apt google-compute-engine-bookworm-stable InRelease
Hit:3 https://deb.debian.org/debian bookworm InRelease
Get:4 https://deb.debian.org/debian bookworm-updates InRelease [55.4 kB]
Get:5 https://deb.debian.org/debian bookworm-backports InRelease [56.5 kB]
Hit:6 https://deb.debian.org/debian-security bookworm-security InRelease
Hit:8 https://packages.cloud.google.com/apt cloud-sdk-bookworm InRelease
Fetched 112 kB in 1s (105 kB/s)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
All packages are up to date.
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
pipx is already the newest version (1.1.0-1).
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
/root/.local/bin has been been added to PATH, but you need to open a new terminal or re-login for this PATH change to take effect.

You will need to open a new terminal or re-login for the PATH changes to take effect.

Otherwise pipx is ready to go! ✨★✨
mohammedkaleemullah06@uber-data-analytics-kaleem:~$ pipx install mage-ai
'mage-ai' already seems to be installed. Not modifying existing installation in '/home/mohammedkaleemullah06/.local/pipx/venvs/mage-ai'. Pass '--force' to force installation.
mohammedkaleemullah06@uber-data-analytics-kaleem:~$ source myenv/bin/activate
(myenv) mohammedkaleemullah06@uber-data-analytics-kaleem:~$ pip install pandas
pip install google-cloud
pip install google-cloud-bigquery
pip install botocore
Requirement already satisfied: pandas in ./myenv/lib/python3.11/site-packages (2.2.2)
Requirement already satisfied: numpy>=1.23.2 in ./myenv/lib/python3.11/site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in ./myenv/lib/python3.11/site-packages (from pandas) (2.8.2)
```

Type here to search
36°C Mostly cloudy
ENG
02:19 PM
13-06-2024


Inbox (789) - mo
uber-etl-pipeline
Python Package
Dimensional Mo
Uber Data Mode
kaleem_uber_da
spirited violet | M
Download histor
Not secure
35.244.0.125:6789/pipelines/spirited_violet/edit?sideview=tree
kaleem-project-2 › Pipelines › spirited_violet › Edit
Launch Command Center
v0.9.71
10:53 UTC
Live help
ALL FILES
CURRENT BLOCK
Search files
kaleem-project-2
charts
custom
data_exporters
data_loaders
dbt
extensions
interactions
pipelines
scratchpads
transformers
utils
__init__.py
io_config.yaml
metadata.yaml
requirements.txt
File
Edit
Run
Python
PY
DATA LOADER
data_loader
Edit parents
import io
import pandas as pd
import requests
if 'data_loader' not in globals():
from mage_ai.data_preparation.decorators import data_loader
if 'test' not in globals():
from mage_ai.data_preparation.decorators import test
@data_loader
def load_data_from_api(*args, **kwargs):
"""
Template for loading data from API
"""
url = 'https://storage.googleapis.com/uber-aataanalytics-kaleem/uber_data.csv'
response = requests.get(url)
data = pd.read_csv(io.StringIO(response.text), sep=',')
return data
@test
def test_output(output, *args) → None:
"""
Template code for testing the output of the block.
"""
assert output is not None, 'The output is undefined'
1/1 tests passed.
OUTPUT 0
VendorID
tpep_pickup_datetime
tpep_dropoff_datetim
passenger_count
trip_distance
pickup_longitude
pickup_lati
CPU: 0.2% / Memory: 316.09MB
Saved a few seconds ago
Tree
data_loader
Data loader
data_transformer
Transformer
data_exporter
Data exporter
70%
Type here to search
Hot weather
ENG
04:23 PM
16-06-2024

Inbox (789) - mo | uber-etl-pipeline | Python Package | Dimensional Mo | Uber Data Mode | kaleem_uber_da | spirited violet | N | Download histor

Not secure 35.244.0.125:6789/pipelines/spirited_violet/edit?sideview=tree

kaleem-project-2 > Pipelines > spirited_violet > Edit

Launch Command Center v0.9.71 10:53 UTC Live help

ALL FILES CURRENT BLOCK

Search files

- kaleem-project-2
  - charts
  - custom
  - data_exporters
  - data_loaders
  - dbt
  - extensions
  - interactions
  - pipelines
  - scratchpads
  - transformers
  - utils
  - __init__.py
  - io_config.yaml
  - metadata.yaml
  - requirements.txt

File Edit Run

Python

PY TRANSFORMER data_transformer 1 parent

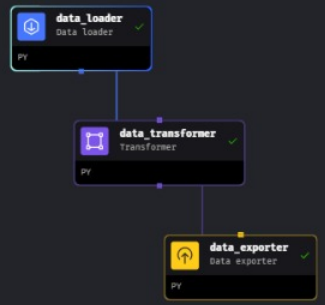
Positional arguments for decorated function:

```
@transformer
def transform(data):
    data → data_loader
```

```python
import pandas as pd
if 'transformer' not in globals():
    from mage_ai.data_preparation.decorators import transformer
if 'test' not in globals():
    from mage_ai.data_preparation.decorators import test


@transformer
def transform(df, *args, **kwargs):

    df['tpep_pickup_datetime'] = pd.to_datetime(df['tpep_pickup_datetime'])
    df['tpep_dropoff_datetime'] = pd.to_datetime(df['tpep_dropoff_datetime'])
    df = df.drop_duplicates().reset_index(drop=True)
    df['trip_id'] = df.index



    # datetime_dim = datetime_dim.rename(columns={'tpep_pickup_datetime': 'datetime_id'}).reset_index(drop=True)

    trip_distance_dim = df[['trip_distance']].reset_index(drop=True)
    trip_distance_dim['trip_distance_id'] = trip_distance_dim.index
    trip_distance_dim = trip_distance_dim[['trip_distance_id','trip_distance']]


    pickup_location_dim = df[['pickup_longitude', 'pickup_latitude']].reset_index(drop=True)
    pickup_location_dim['pickup_location_id'] = pickup_location_dim.index
    pickup_location_dim = pickup_location_dim[['pickup_location_id','pickup_latitude','pickup_longitude']]


    dropoff_location_dim = df[['dropoff_longitude', 'dropoff_latitude']].reset_index(drop=True)
    dropoff_location_dim['dropoff_location_id'] = dropoff_location_dim.index
    dropoff_location_dim = dropoff_location_dim[['dropoff_location_id','dropoff_latitude','dropoff_longitude']]
```

CPU: 0.2% / Memory: 316.09MB Saved a few seconds ago

Tree



70%

Type here to search 36°C Mostly sunny ENG 04:23 PM 16-06-2024

Inbox (789) - mo | uber-etl-pipeline | Python Package | Dimensional Mo | Uber Data Mode | kaleem_uber_da | spirited violet | N | Download histor

Not secure 35.244.0.125:6789/pipelines/spirited_violet/edit?sideview=tree

kaleem-project-2 > Pipelines > spirited_violet > Edit

Launch Command Center v0.9.71 10:53 UTC Live help

ALL FILES CURRENT BLOCK

Search files

- kaleem-project-2
  - charts
  - custom
  - data_exporters
  - data_loaders
  - dbt
  - extensions
  - interactions
  - pipelines
  - scratchpads
  - transformers
  - utils
  - __init__.py
  - io_config.yaml
  - metadata.yaml
  - requirements.txt

File Edit Run

Python

PY DATA EXPORTER data_exporter 1 parent

Positional arguments for decorated function:

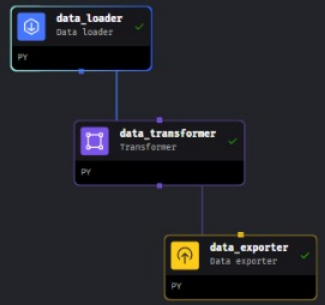
```
@data_exporter
def export_data(data):
    data → data_transformer
```

```python
from mage_ai.data_preparation.repo_manager import get_repo_path
from mage_ai.io.bigquery import BigQuery
from mage_ai.io.config import ConfigFileLoader
from pandas import DataFrame
from os import path

if 'data_exporter' not in globals():
    from mage_ai.data_preparation.decorators import data_exporter


@data_exporter
def export_data_to_big_query(data, **kwargs) -> None:
    """
    Template for exporting data to a BigQuery warehouse.
    Specify your configuration settings in 'io_config.yaml'.

    Docs: https://docs.mage.ai/design/data-loading#bigquery


    """
    config_path = path.join(get_repo_path(), 'io_config.yaml')
    config_profile = 'default'

    for key, value in data.items():
        table_id = 'still-kit-426212-h4.kaleem_uber_data_analytics.{}'.format(key)
        BigQuery.with_config(ConfigFileLoader(config_path, config_profile)).export(
            DataFrame(value),
            table_id,
            if_exists='replace',  # Specify resolution policy if table name already exists
        )
```

CPU: 0.2% / Memory: 316.09MB

Saved a few seconds ago

Tree



70%

Type here to search

36°C Mostly sunny ENG 04:23 PM 16-06-2024

Inbox (8 | uber-etl | Data Eng | Dimensi | Uber Da | BigQuer | Untitled | Downlo | Uber-Da | Home | (2) Linke

console.cloud.google.com/bigquery?project=still-kit-426212-h4&ws=!1m11!1m10!12m5!1m3!1sstill-kit-426212-h4!2sus-central1!3sa20a0958-8ee1-4047-8e88-d8969d14bbd6!2e1!14m3!1...

Free trial status: ₹24,736.28 credit and 84 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. DISMISS ACTIVATE

Attention Customers from India: As of October 1st, due to regulatory changes by the Reserve Bank of India, we will not be able to charge your card for recurring payments if you have not set up an e-mandate or if the transaction is above 5000 INR. If your payments are being declined, Billing administrators will have to log into their Payments account and manually reauthorize the payments. Read more DISMISS

Google Cloud | My First Project | Search (/) for resources, docs, products, and more | Search

Explorer + ADD

Type to search

Viewing resources.

SHOW STARRED ONLY

- Notebooks
- Data canvases
- External connections
- kaleem_uber_data_analy...
  - dropoff_location_dim
  - fact_table
  - payment_type_dim
  - pickup_location_dim
  - trip_distance_dim

SUMMARY ACTIVITY

Jun 16, 2024

Created - Mohammed Kalee... 4:37 PM

Untitled query | *Payment ...mes

Payment type names

RUN | SAVE QUERY | DOWNLOAD | SHARE | SCHEDULE | MORE | Query completed.

```
1 SELECT b.payment_type_name
2 from still-kit-426212-h4.kaleem_uber_data_analytics.fact_table a
3 join still-kit-426212-h4.kaleem_uber_data_analytics.payment_type_dim b
4 on a.payment_type_id = b.payment_type_id
5 group by b.payment_type_name
```

Press Alt+F1 for Accessibility Options.

Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	payment_type_name
1	Credit card
2	Cash
3	No charge

Job history

REFRESH

Type here to search

NIFTY +0.28%

ENG 03:04 PM 19-06-2024

UBER

DATA VISUALIZATION

Presented by Kaleemullah

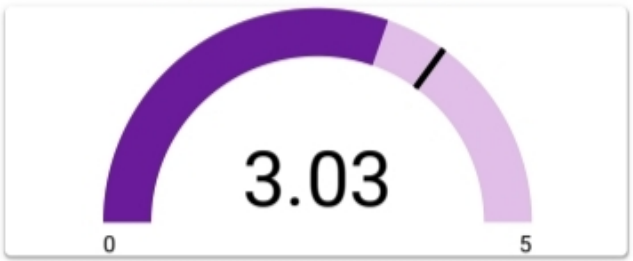
passenger_count	Record Count
0	3
1	65.5K
2	13.7K
3	4.1K
4	1.9K
5	8.7K
6	6.1K

Jan 1, 2016 - Dec 31, 2016

payment_type	total_amount
1	1,202,467.81
2	434,002.97
3	1,838.95
4	762.36

1 - 4 / 4

Average Trip Distance



Total Amount

1.6M

Average Tip Amount

1.87

VendorID	fare_amount	extra	tip_amount	tolls_amount
1	162,428.91	5,055.5	21,122.6	3,986.11
2	1,162,831.93	5,114	166,130.5	32,755.34

1 - 2 / 2