Peer-graded Assignment: Submit your Work and Review your Peers

It looks like this is your first peer-graded assignment. Learn more

Instructions | My submission | Discussions

DataCapstone Project

Submitted on December 2, 2024

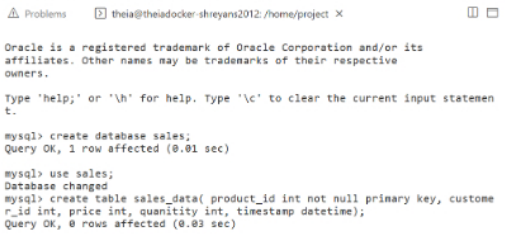
Shareable Link

PROMPT

OLTP Task 2: Design a table named sales_data. **(3pts)**

Upload the **createtable.jpg** (or .png) file for your peers to review.

Design a table named sales_data



RUBRIC

Did the learner design and create the table sales_data correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

TIP:
If the screenshot appears small and is hard to read try zooming in by pressing "Ctrl" and "+" keys together (Mac: "Command" and "+"), or Right-click on the image and "View Image" (Firefox) or "Open Image in new Tab" (Chrome).

- 0 points
  Incorrect. The learner did not submit the screenshot, or the "CREATE TABLE" command was not used correctly.
- 1 point
  Partially correct. The learner has correctly used the "CREATE TABLE" command and specified data types for at least 2 fields correctly.

  product_id int
  customer_id int
  price decimal or int
  quantity int

timestamp timestamp

2 points
Partially correct. The learner has correctly used the "CREATE TABLE" command and specified data types for at least 3 fields correctly.

product_id int
customer_id int
price decimal or int
quantity int
timestamp timestamp

3 points
Correct. The learner has correctly used the "CREATE TABLE" command and specified data types for all fields correctly.

product_id int
customer_id int
price decimal or int
quantity int
timestamp timestamp

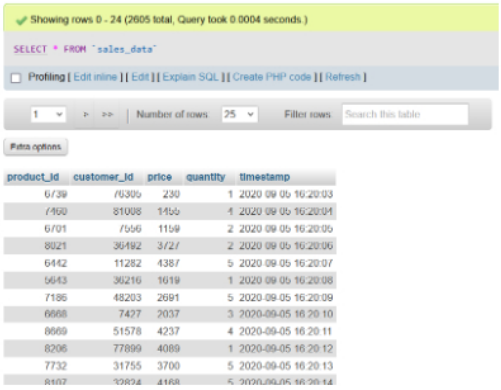
MD

PROMPT

OLTP Task 3: Import the data in the file oltpdata.csv. **(1pt)**

Upload the **importdata.jpg** (or .png) file for your peers to review.

Import data in oltpdata.csv



RUBRIC

Did the learner import the oltpdata.csv into sales_data table correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points
Incorrect. The learner did not submit the screenshot, or the output does not contain a successful import message.

1 point
Correct. The learner has correctly used the phpMyAdmin tool to import the data and the screenshot shows a success message.

MD

PROMPT

OLTP Task 8: Write a bash script to export data. **(3pts)**

Upload the **exportdata.jpg** (or .png) file for your

RUBRIC

Did the learner write the datadump.sh correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria

peers to review,

bashscript to export data

datadump.sh

```
#!/bin/sh

mysql -u root -p sales_data > salesdata.sql
```

below.

- 0 points
  Incorrect. The screenshot is not submitted or shows none of the items below.
  - datadump.sh starts with a shebang
  - uses mysqldump command with correct table and database names
  - destination file name is sales_data.sql

- 1 point
  Partially correct. The screenshot shows any 1 of the items below.
  - datadump.sh starts with a shebang
  - uses mysqldump command with correct table and database names
  - destination file name is sales_data.sql

- 2 points
  Partially correct. The screenshot shows any 2 of the items below.
  - datadump.sh starts with a shebang
  - uses mysqldump command with correct table and database names
  - destination file name is sales_data.sql

- **3 points**
  Correct. The screenshot shows all the items below.
  - datadump.sh starts with a shebang
  - uses mysqldump command with correct table and database names
  - destination file name is sales_data.sql

PROMPT

NoSQL Task 1: Import 'Catalog.json' into mongodb server server into a database named 'catalog' and a collection named 'electronics'. **(2pts)**

Upload the **mongoimport.jpg (or .png)** file for your peers to review.

import catalog.json into mongodb

```
ieia@theiadocker-shreyans2012:/home/project$ mongoimport --db cata
g --collection electronics --file catalog.json --jsonArray

connected to: mongodb://localhost/
438 document(s) imported successfully. 0 document(s) failed to import.
```

RUBRIC

Did the learner correctly import the catalog.json file into electronics collection in catalog database?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot, or the command or parameters were not correct.
- 1 point
  Partially correct. The screenshot shows the correct usage of the mongoimport command with at least 2 of the following parameters:

  Database = catalog
  Collection = electronics
  File = catalog.json
- **2 points** MD
  Correct. The screenshot shows the correct usage of the mongoimport command with the following parameters:

  Database = catalog
  Collection = electronics
  File = catalog.json

PROMPT

NoSQL Task 4: Create an index on the field "type" **(2pts)**

Upload the **create-index.jpg** (or .png) file for your peers to review.

create index on the field "type"

```
> db.electronics.createIndex({"type":1})
{
        "createdCollectionAutomatically" : false,
        "numIndexesBefore" : 1,
        "numIndexesAfter" : 2,
        "ok" : 1
```

RUBRIC

Did the learner correctly create the index?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the command used is not correct.
- 1 point
  Partially correct. The learner has correctly used the createIndex() command, but the field mentioned is not correct.
- **2 points** MD
  Correct. The learner has correctly used the createIndex() command with the field mentioned as "type".

PROMPT

NoSQL Task 7: Write a query to find out the average screen size of smart phones. **(2pts)**

Upload the **mongo-query-mobiles2.jpg** (or .png) file for your peers to review.

find out avg screen size of smartphone

```
> db.electronics.aggregate([
... { "$match": { "type": "smart phone" } },
... { "$group": { "_id":"$type", "average":{"$avg":"$screen size"} } } ])
{ "_id" : "smart phone", "average" : 6 }
```

RUBRIC

Did the learner correctly find the average screen size of smart phones?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the command used and the results are not correct.
- 1 point
  Partially correct. The answer has got any two of the items below correct:
  - Used the aggregate() method
  - Used the $group operator
  - Used the $match operator
  - The result shows the average size as 6
- **2 points**
  Correct. The answer has got all of the items below correct:
  - Used the aggregate() method
  - Used the $group operator
  - Used the $match operator
  - The result shows the average size as 6

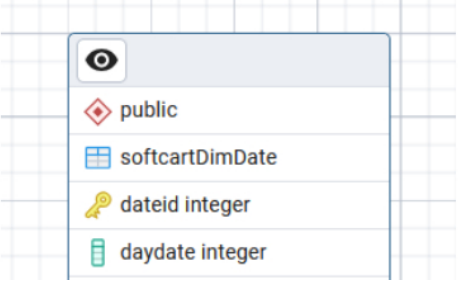
MD

PROMPT

Data Warehousing Task 1: Design the dimension table softcartDimDate. **(2pts)**

Upload the **softcartDimDate.jpg** (or .png) for your peers to review.

desiging dimension table softcartDimDate

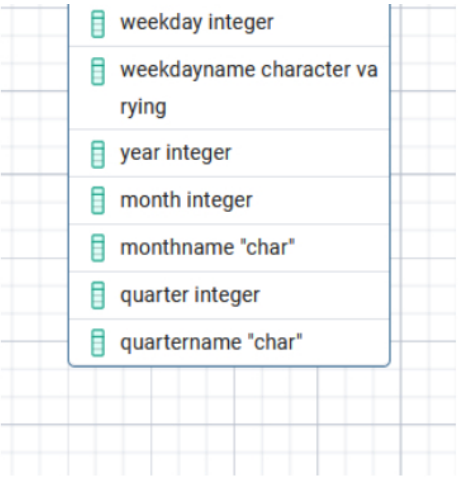


RUBRIC

Did the learner identify the fields for softcartDimDate table correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or less than 6 fields below have been correctly identified. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)



- dateid
- date
- year
- quarter
- quartername
- month
- monthname
- day
- weekday
- weekdayname

1 point

Partially correct. The learner has identified between 6 and 9 of the fields below. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

- dateid
- date
- year
- quarter
- quartername
- month
- monthname
- day
- weekday
- weekdayname

2 points

MD

Correct. The learner has identified all of the 10 fields below. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

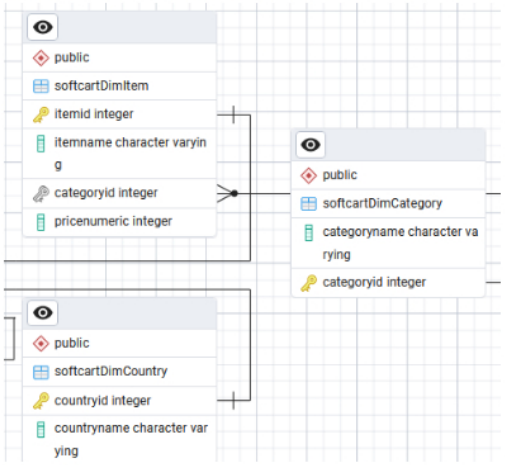
- dateid
- date
- year
- quarter
- quartername
- month
- monthname
- day
- weekday
- weekdayname

PROMPT

Data Warehousing Task 2: Design the dimension table softcartDimCategory, softcartDimItem, SoftcartDimCountry. **(3pts)**

Upload the **dimtables.jpg** (or .png) for your peers to review.

design dimension table softcartdimcategory, softcartDimitem, softcartDimCountry



RUBRIC

Did the learner design the tables softcartDimCategory, softcartDimItem, softcartDimCountry correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points
Incorrect. The learner did not submit the screenshot or only 1 field below has been correctly identified in the corresponding tables. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

- categoryid
- category
- itemid
- item
- countryid
- country

1 point
Partially correct. The learner has correctly identified at least 2 of the fields below in the corresponding tables. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

- categoryid
- category
- itemid
- item
- countryid
- country

2 points
Partially correct. The learner has correctly identified at least 4 of the fields below in the corresponding tables. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

- categoryid
- category

- itemid
- item
- countryid
- country

3 points MD

Correct. The learner has correctly identified all of the fields below in the corresponding tables. (Ignore any spelling mistakes, usage of underscores and hyphens in the field names)

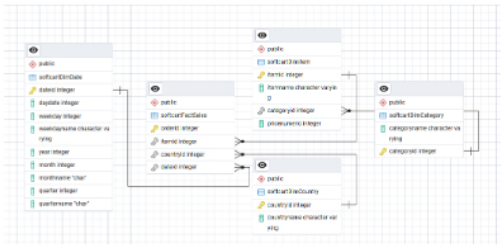
- categoryid
- category
- itemid
- item
- countryid
- country

PROMPT

Data Warehousing Task 3: Design the relationships **(2pts)**

Upload the **softcartrelationships.jpg** (or .png) for your peers to review.

design relationships



RUBRIC

Did the learner design the relationships correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points

Incorrect. The learner did not submit a screenshot or no relationships were identified.

1 point

Partially correct. The learner has correctly identified at least 2 of the relationships below:

- softcartDimDate.dateid(many) to softcartFactSales.dateid(one)
- softcartDimCountry.countryid (many) to softcartFactSales. countryid (one)
- softcartDimItem.itemid (many) to softcartFactSales. itemid (one)
- softcartDimCategory.categoryid(many) to softcartFactSales.categoryid (one)

2 points MD

Correct. The learner has identified all the relationships below

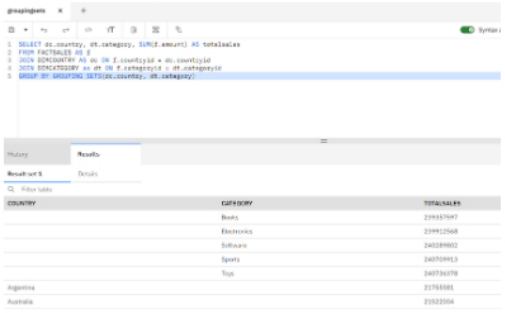
- softcartDimDate.dateid(many) to softcartFactSales.dateid(one)
- softcartDimCountry.countryid (many) to softcartFactSales. countryid (one)
- softcartDimItem.itemid (many) to softcartFactSales. itemid (one)
- softcartDimCategory.categoryid(many) to softcartFactSales.categoryid (one)

PROMPT

Data Warehouse Reporting Task 1: Create a grouping sets query using the columns country, category, totalsales. **(2pts)**

Upload the **groupingsets.jpg** (or .png) for your peers to review.

create grouping sets query



RUBRIC

Did the learner write the sql aggregation grouping sets query correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points
Incorrect. The learner did not submit the screenshot, or the sql query has less than 2 of the items below, or the sql query is not correct.

- Contains country, category, sum(amount) in the select
- Joins tables factsales and dimcountry on countryid
- Joins tables factsales and dimcategory on categoryid
- Has a grouping sets clause with fields country,category

1 point
Partially correct. The sql query is correct and has at least 2 of the items below:

- Contains country, category, sum(amount) in the select
- Joins tables factsales and dimcountry on countryid
- Joins tables factsales and dimcategory on categoryid
- Has a grouping sets clause with fields country,category

2 points
Correct. The sql query is correct and has all of the items below:

MD

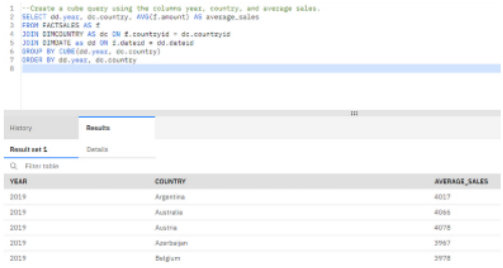
- Contains country, category, sum(amount) in the select
- Joins tables factsales and dimcountry on countryid
- Joins tables factsales and dimcategory on categoryid
- Has a grouping sets clause with fields country,category

PROMPT

Data Warehouse Reporting Task 2: Create a cube query using the columns year, country, and average sales. **(2pts)**

Upload the **cube.jpg** (or .png) for your peers to review.

creating cube query



RUBRIC

Did the learner write the sql aggregation cube query correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points
Incorrect. The learner did not submit the screenshot, or the sql query is not correct.

1 point
Partially correct. The sql query is correct and has at least 2 of the items below:

- Contains year, country, avg(amount) in the select
- Joins tables factsales and dimcountry on countryid
- Joins tables factsales and dimdate on dateid
- Has a cube clause with fields year, country

2 points
Correct. The sql query is correct and has all of the items below:

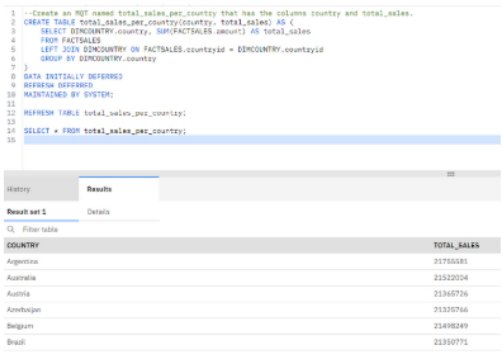
- Contains year, country, avg(amount) in the select
- Joins tables factsales and dimcountry on countryid
- Joins tables factsales and dimdate on dateid
- Has a cube clause with fields year, country

PROMPT

Data Warehouse Reporting Task 3: Create Materialized Query Table/View:
(You may use either Db2 or PostgreSQL to complete this task.)
Create a Materialized Query Table (in case of Db2) or Materialized View (in case of PostgreSQL) named total_sales_per_country using the columns country and total sales. **(4pts)**

Upload the **mqt.jpg** (or .png) for your peers to review.

Materialized Query Table



```
--Create an MQT named total_sales_per_country that has the columns country and total_sales.
CREATE TABLE total_sales_per_country(country, total_sales) AS (
    SELECT DIMCOUNTRY.country, SUM(FACTSALES.amount) AS total_sales
    FROM FACTSALES
    LEFT JOIN DIMCOUNTRY ON FACTSALES.countryid = DIMCOUNTRY.countryid
    GROUP BY DIMCOUNTRY.country
)
DATA INITIALLY DEFERRED
REFRESH DEFERRED
MAINTAINED BY SYSTEM;

REFRESH TABLE total_sales_per_country;

SELECT * FROM total_sales_per_country;
```

History | Results

Result set 1 | Details

Filter table

COUNTRY	TOTAL_SALES
Argentina	21755581
Australia	21522004
Austria	21366726
Azerbaijan	21335766
Belgium	21498249
Brazil	21350771

RUBRIC

Note: The learner may use either Db2 or PostgreSQL to complete this task.

Did the learner write the create Materialized Query Table (in case of Db2) or Materialized View (in case of PostgreSQL) statement correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the sql is not correct.

- 2 points
  Partially correct. The sql query is correct and has at least 2 of the items below:
  - Contains country, total_sales in *create table* statement (in case of Db2) or *create materialized view* statement (in case of PostgreSQL)
  - Contains country, sum(amount) in the select clause
  - Joins tables factsales, dimcountry on countryid
  - Has a group by clause with country

- **4 points**
  Correct. The sql query is correct and has all 4 of the items below:
  - Contains country, total_sales in *create table* statement (in case of Db2) or *create materialized view* statement (in case of PostgreSQL)
  - Contains country, sum(amount) in the select clause
  - Joins tables factsales, dimcountry on countryid
  - Has a group by clause with country

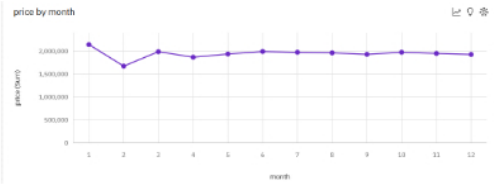
MD

PROMPT

Dashboard Creation Task 4: Create a line chart. **(2pts)**

Upload the **linechart.jpg** (or .png) for your peers to review.

Draw line chart



RUBRIC

Did the learner create the line chart correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot.
- 1 point
  Partially correct. The screenshot shows the line chart that shows month wise sales and the chart is NOT properly titled and labelled.
- **2 points**
  Correct. The screenshot shows the line chart that shows month wise sales and the chart is properly titled and labelled.

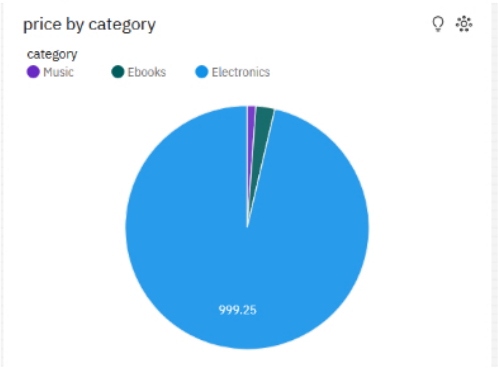
MD

PROMPT

Dashboard Creation Task 6: Create a pie chart. **(2pts)**

Upload the **piechart.jpg** (or .png) for your peers to review.

create pie chart



RUBRIC

Did the learner create the pie chart of category wise total sales correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit a screenshot.
- 1 point
  Partially correct. category wise sales and the chart is NOT properly titled and labelled.
- **2 points**
  Correct. The screenshot shows the pie chart that shows category wise sales and the chart is properly titled and labelled.

MD

PROMPT

Dashboard Creation Task 5: Create a bar chart. **(1pt)**

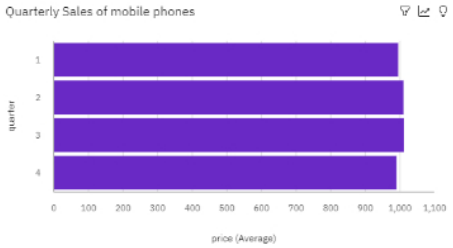
RUBRIC

Did the learner create the bar chart correctly?

Review the screenshot uploaded by the learner

Upload the **barchart.jpg** (or .png) for your peers to review.

create bar chart



and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit a screenshot.
- 1 point
  Partially correct. The screenshot shows the bar chart of quarterly sales of mobile phones and the chart is NOT properly titled and labelled.
- **2 points**
  Correct. The screenshot shows the bar chart of quarterly sales of mobile phones and the chart is properly titled and labelled.

MD

PROMPT

ETL Task 1: Implement the function get_last_rowid() **(2pts)**

(You may use either Db2 or PostgreSQL to complete this task.)

Upload the **get_last_rowid.jpg** (or .png) for your peers to review.

implement function get_last_rowid()

```
def get_last_rowid():
    SQL = "SELECT MAX(ROWID) FROM sales_data"
    stmt = ibm_db.exec_immediate(conn, SQL)
    res = ibm_db.fetch_both(stmt)
    print(res)
    return int(res[0])

last_row_id = get_last_rowid()
print("Last row id on production datawarehouse = ", last_row_id)
```

RUBRIC

Did the learner implement the function get_last_rowid correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the function implementation is not correct.
- 1 point
  Partially correct. The screenshot shows 1 of the items below:
  - A SQL statement to select the last_rowid
  - A return statement that returns the last_rowid
- **2 points**
  Correct. The screenshot shows all the items below:
  - A SQL statement to select the last_rowid
  - A return statement that returns the last_rowid

MD

PROMPT

RUBRIC

ETL Task 2: Implement the function get_lastest_records() **(2pts**

Upload the **get_latest_records.jpg** (or .png) for your peers to review.

Implement get_latest_records()

```
def get_latest_records(rowid):
    SQL = "SELECT * FROM sales_data WHERE rowid > %s"
    cursor.execute(SQL, [rowid])
    new_recs = cursor.fetchall()
    for row in new_recs:
        print(row)
    return new_recs

new_records = get_latest_records(last_row_id)

# print("New rows on staging datawarehouse = ", len(new_records))
```

Did the learner implement the function get_latest_records correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the function implementation is not correct.
- 1 point
  Partially correct. The screenshot shows 1 of the items below:
  - A SQL statement to select all records greater than the given row _id
  - A return statement that returns the latest records
- **2 points**
  Correct. The screenshot shows all the items below:
  - A SQL statement to select all records greater than the given row _id
  - A return statement that returns the latest records

MD

PROMPT

ETL Task 3: Implement the function insert_records() **(2pts)**

(You may use either Db2 or PostgreSQL to complete this task.)

Upload the **insert_records.jpg** (or .png) for your peers to review.

function insert_records()

```
def insert_records(records):
    SQL = "INSERT INTO sales_data(rowid,product_id,customer_id,quantity) VALUES(?,?,?,?);"
    stmt = ibm_db.prepare(conn, SQL)

    for record in records:
        ibm_db.execute(stmt, record)

insert_records(new_records)
print("New rows inserted into production datawarehouse = ", len(new_records))
```

RUBRIC

Did the learner implement the function insert_records correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the function implementation is not correct.
- 1 point
  Partially correct. The screenshot shows 1 of the items below:
  - A SQL statement to insert all the records
  - A ibm_db.exec_immediate() function (in case of Db2) or cursor.execute() function (in the case of PostgreSQL)

the case of PostgreSQL)

2 points MD
Correct. The screenshot shows all the items below:

- A SQL statement to insert all the records
- A ibm_db.exec_immediate() function(in case of Db2) or cursor.execute() function (in the case of PostgreSQL)

PROMPT

Pipelines Task 2: Define the DAG **(2pts)**

Upload the **dag_definition.jpg** (or .png) for your peers to review.

define DAG

```
# DAG Definition
dag = DAG(
    'process_web_log',
    description='SoftCart access log ETL pipeline',
    default_args=default_args,
    schedule_interval=dt.timedelta(days=1),
)
```

RUBRIC

Did the learner define the DAG correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points
Incorrect. The learner did not submit the screenshot or there is no entry for the DAG parameters visible in the screenshot.

1 point
Partially correct. The screenshot shows at least 2 parameters defined with proper values from the list below.

- DAG id
- Schedule
- default_args
- description

2 points MD
Correct. The screenshot shows all 4 parameters defined with proper values from the list below.

- DAG id
- Schedule
- default_args
- description

PROMPT

Pipelines Task 3: Create a task to extract data

RUBRIC

Did the learner create the extract_data task

Pipelines Task 3: Create a task to extract data **(2pts)**

Upload the **extract_data.jpg** (or .png) for your peers to review.

Create task to extract data

Did the learner create the extract_data task correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or there is no entry for the unzip task visible in the screenshot.
- 1 point
  Partially correct. The screenshot shows the use of 1 of the below:
  - BashOperator
  - the cut command with proper options
- **2 points**
  Correct. The screenshot shows the use of all of the below:
  - BashOperator and
  - the cut command with proper options

MD

PROMPT

Pipelines Task 4: Create a task to transform the data in the txt file **(2pts)**

Upload the **transform_data.jpg** (or .png) for your peers to review.

Create task to transform the data in txt file

RUBRIC

Did the learner create the transform_data task correctly?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or there is no entry for the extract task visible in the screenshot.
- 1 point
  Partially correct. The screenshot shows the use of 1 of the below:
  - BashOperator
  - the grep command with proper options
- **2 points**
  Correct. The screenshot shows the use of all of the below:

MD

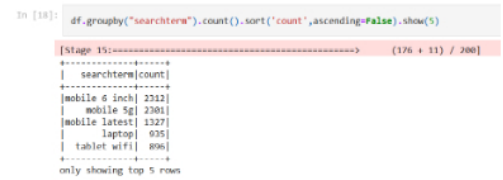
- BashOperator and
- the grep command with proper options

PROMPT

Big Data Task 5 - Print the top 5 most frequently used search terms. **(2pts)**

Upload the **top5terms.jpg** (or .png) file for your peers to review.

print 5 most frequently used search terms



RUBRIC

Did the learner correctly identify the top 5 search terms?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the output is not correct.

- 1 point
  Partially Correct. The screenshot shows at least 3 of the terms below.
  - mobile 6 inch
  - mobile 5g
  - mobile latest
  - laptop
  - tablet wifi

- **2 points**
  Correct. The screenshot shows the terms below in the order given.
  - mobile 6 inch
  - mobile 5g
  - mobile latest
  - laptop
  - tablet wifi

MD

PROMPT

Big Data Task 6 - Load the sales forecast model. **(1pt)**

Upload the **loadmodel.jpg** (or .png) file for your peers to review.

load sales forecast model

```
from pyspark.ml.regression import LinearRegressionModel

model=LinearRegressionModel.load('sales_prediction.model')
```

RUBRIC

Did the learner correctly load the sales forecast model?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

- 0 points
  Incorrect. The learner did not submit the screenshot or the command used

is not correct.

1 point MD

Correct. The screenshot shows the correct usage of LinearRegressionModel.load() command to load the forecast model.

PROMPT

Big Data Task 7 - Using the sales forecast model, predict the sales for the year of 2023. **(1pt)**

Upload the **forecast.jpg** (or .png) file for your peers to review.

upload forecast

```
from pyspark.ml.feature import VectorAssembler

def predict(year):
    assembler = VectorAssembler(inputCols=["year"],outputCol="features")
    data=[[year,0]]
    columns=["year","sales"]
    _ = spark.createDataFrame(data,columns)
    __ = assembler.transform(_).select('features','sales')
    predictions = model.transform(__)
    predictions.select('prediction').show()

predict(2023)
```

```
+------------------+
|        prediction|
+------------------+
|175.16564294006457|
+------------------+
```

RUBRIC

Did the learner correctly load the sales forecast model?

Review the screenshot uploaded by the learner and grade this question based on the criteria below.

0 points

Incorrect. The learner did not submit the screenshot or the command used is not correct.

1 point MD

Correct. The screenshot shows the correct usage of predict() command to predict the sales for the year 2023.

Start new attempt

Comments

Comments left for the learner are visible only to that learner and the person who left the comment.

SS

Share your thoughts...

Like Dislike Report an issue