Intro:

For the Assignment, we will build off the work we have done in module 5 Lab during class. On Canvas, you will find that a physical data model is loaded along with the facts and dimensions of a star schema database. Additionally, the conceptual model for the beginning of the architecture was developed.

Now, you are going to build this architecture and deploy both the transactional database as an Operational Data Store (ODS) and the OLAP Data warehouse.

Here are the links to the necessary documentation to complete the assignment

1. Facts & Dimensions Bus Matrix
2. Physical Data Model Template
3. Database documentation – OLTP – dav6100_db_2:
   a. Layout File
   b. DDL
4. An update file

1. Deploy the architecture in AWS (10 pts)

Deploy:

Using the Conceptual AWS Architecture diagram below as a guide, implement a development environment with the following components (Note: you should have most or all of these elements in place already from previous course work and can reuse them here)

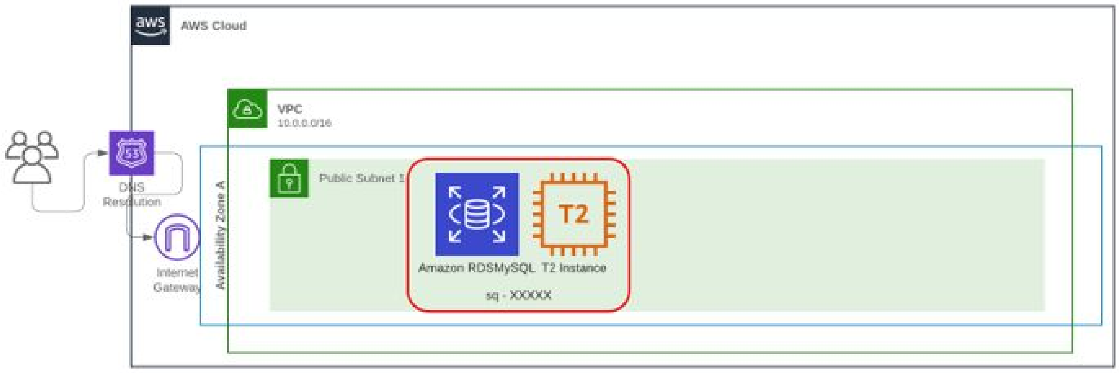
Infrastructure requirements:

- One Virtual Private Cloud (VPC) – AWS East -1 Regions
- One Availability Zones – One for Master, One for Slave Architecture
- Deploy EC2 Instances as needed
    - Note: RDS will deploy an EC2 instance when you deploy a database

- Create a single MySQL instance as your DB for both your OLTP and OLAP databases
    - Use dav6100_db_2 as the name for your OLTP database
    - Use dav6100_db_2_rpt as the name for your OLAP database

Document:

Once you have completed your infrastructure, please create a Conceptual AWS Architecture Diagram with physical attributes and submit your final architecture diagram as part of a submission to the lab as a PDF document. You should have a database in one resulting MySQL instance with an EC2 t2.micro instance deployed. You should specify specific details (physical attributes) about the security group, VPC, and EC2 instance. Also, add your endpoint and port for your MySQL instance to the diagram! A starting reference point is provided below.



We Recommend you use Lucid Charts to create the diagram. Please use the following conventions for submissions:

- <<Task_No>>_<<Last_Name>>_<<First_Name>>_<<YU_Student_ID>>_Assignment_2
- Example: 1_Chiazza_Brandon_2009390_Assignment_2

2. Develop the schema. (30 pts)

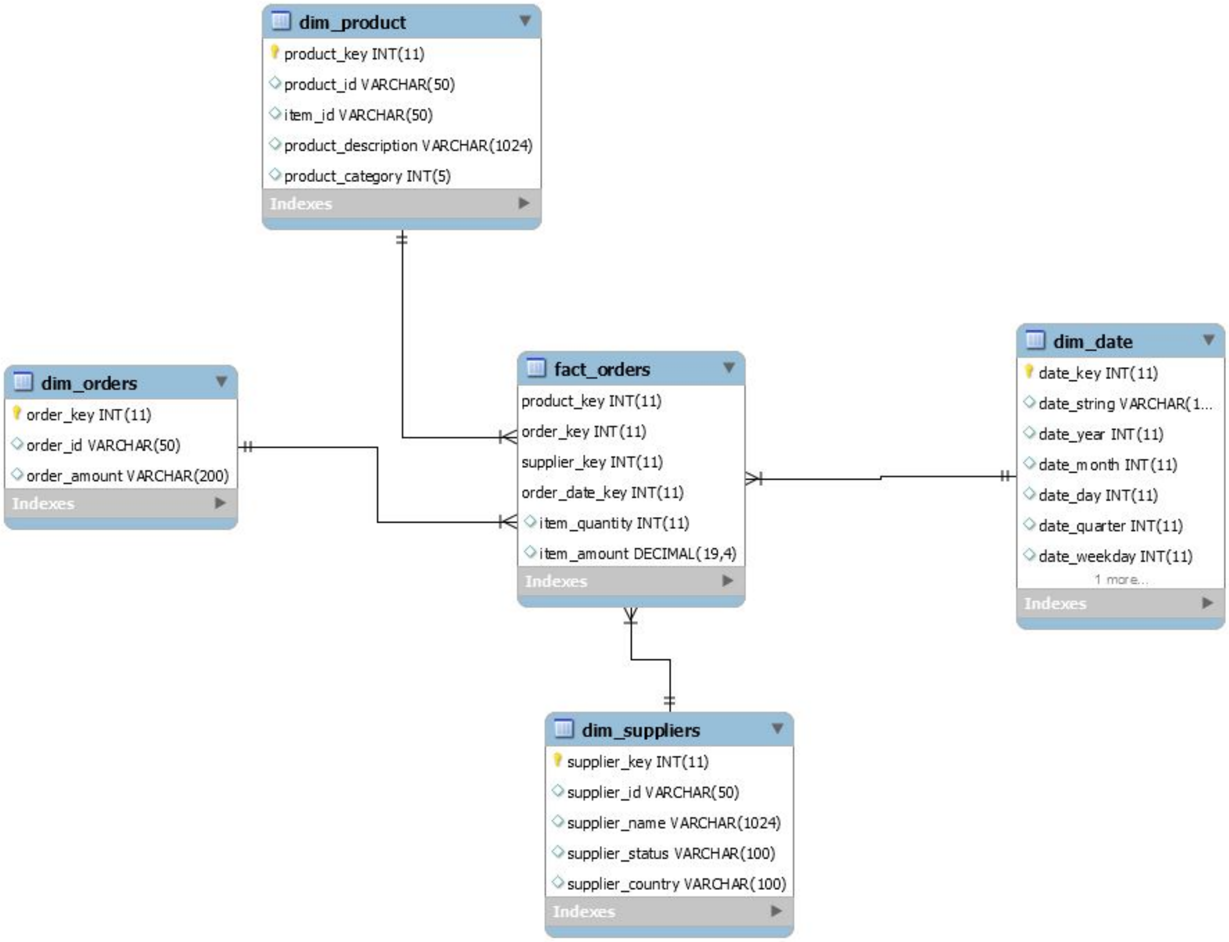
Now that you have the infrastructure and the physical data model, you are going to develop the OLAP database. Follow the steps below to complete the deployment (.5 pts)

1. First, you should ensure that you could establish a connection to your two MySQL primary databases. You can review guidance here on how to establish a connection.

Note: you may have to troubleshoot your ports and connection strings. You must make sure that you appropriately assigned your local machine to the correct security group.

2. Next, connect to your OLTP database, dav6100_db_2. Construct the database by running the DDL file in the folder online.
3. Run the following test queries to ensure that it is properly compiled.
    - select count(*) from `dav6100_db_2`.`r_base_stat`;
        i. Expected return = 33
    - select count(*) from `dav6100_db_2`.`r_ctry`;
        i. Expected return = 247
    - select format(sum(`ord_item_amt`*`ord_item_qty`),0) from `dav6100_db_2`.`t_ord_item`;
    - Expected return = 206,615,960,315
4. After everything checks out, recall that in Lab #1, we decided on three Facts with full atomic granularity that were important to the business user. The following are the facts we care about for this assignment:
    i. Order Count by Supplier
    ii. count of orders over time
    iii. The total amount of orders by supplier.

    Using the documentation listed above (bus matrix, physical data model template, the conceptual data model), construct a star schema DDL called dav6100_db_2_rpt. If updates are needed to the physical data model, please make them and insert them into your DDL. The following is an example of the layout of the file



5. Once completed, run the DDL file you created in (4) in your MySQL OLAP database instance, dav6100_db_2_rpt. Export the DDL file from your MySQL instance and upload a copy as a .txt file and as a submission to this assignment. Use the following naming convention:
    - <<Task_No>>_<<Last_Name>>_<<First_Name>>_<<YU_Student_ID>>_Assignment_2
    - Example:2_Chiazza_Brandon_2000993_Assignment_2

3. Transform and Load the Data. (30 pts)

3.1 Develop transformations (20 pts):

Now that you have created the schema, you will have to execute the transformations to the OLTP data to create the dimensional tables used in the schema.

1. Referring to your physical data model template logic, write a series of scripts that will insert data into your dimensional tables from your OLTP tables.

2. Remember, you are moving from a third normal form to a denormalized form or star schema. De-normalization is important here, and your transformations will use GROUP BY and JOINS to flatten the OLTP database tables. After you have transformed the transactional data, test the data with several queries.

3. Compile your transformations in one sequential .txt file. Once compiled, submit your work to Canvas using the following naming convention:

- <<Task_No>>_<<Last_Name>>_<<First_Name>>__<<YU_Student_ID>>_Assignment_2
- Example: 3.1_Chiazza_Brandon_200090_Assignment_2

3.2 Query the result. With the database built, answer the following questions by querying the database. (10 pts)

· How many Orders by Supplier by Quarter are there? Use the Order Dispatched Date to identify the appropriate date in the date dimension.

· What is the count of orders by the supplier? Who are the top 5 suppliers that receive orders based on total orders?

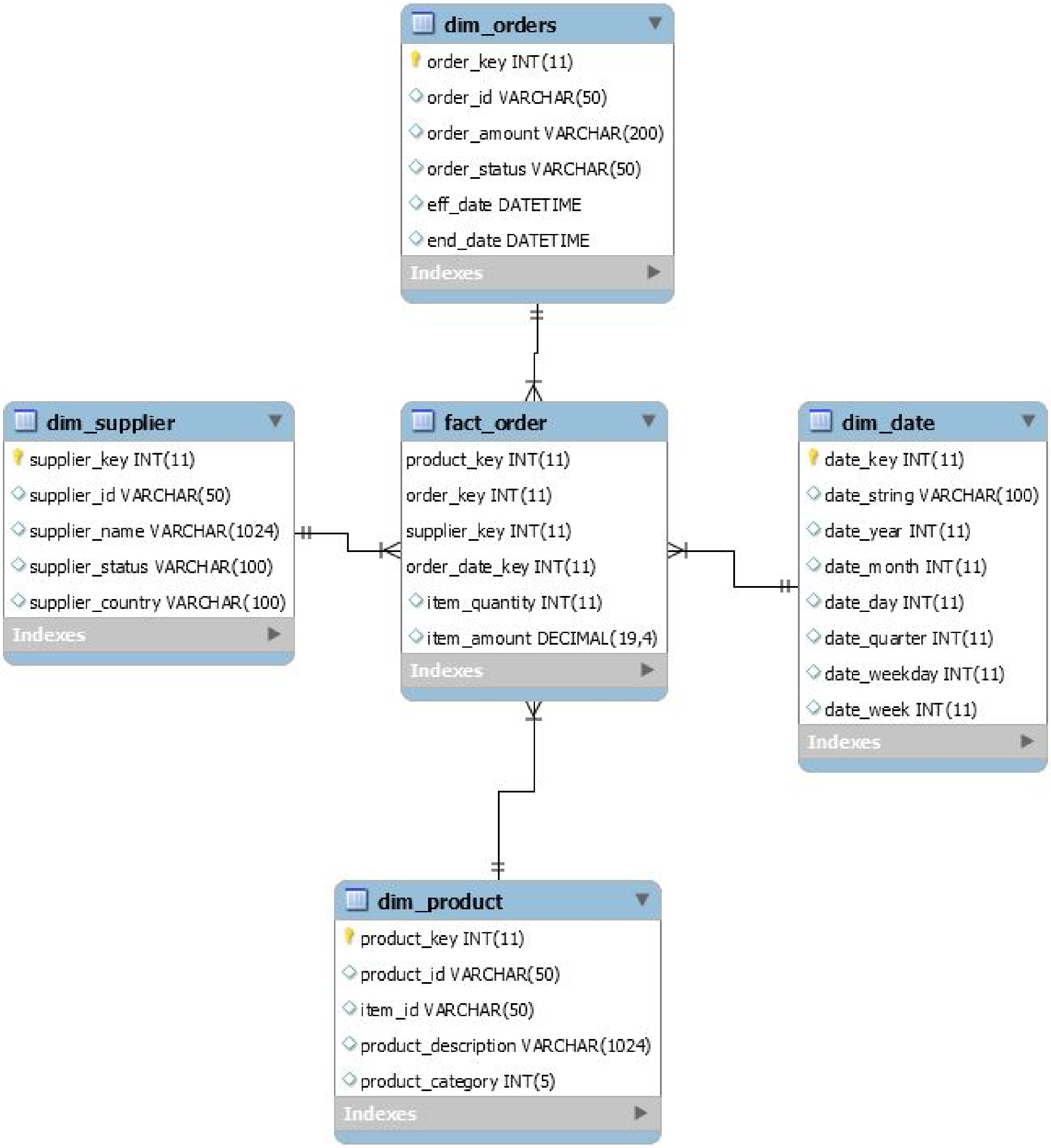
· How many orders per quarter?

Compile all your queries into one sequential .txt file. Once compiled, submit your work to Canvas using the following naming convention:

- <<Task_No>>_<<Last_Name>>_<<First_Name>>_<<YU_Student_ID>>_Assignment_2
- Example: 3.2_Chiazza_Brandon_2000930_Assignment_2

4. Type Dimensions. (30 pts)

Now that we have constructed the basic database, you will update the schema to account for status changes to orders. You will note that in the database, dav6100_db_2, the t_ord_item table has status ‘cur’ and ‘ini’. A status of ‘cur’ means current status of the line item on an order whereas a status of ‘ini’ just means draft. Additionally, you will note that on the t_ord_order table there are unique IDs for orders. In this step, you will update your schema to create a Type II Dimension with an effective date, end date, and a current column to retain the historical information of the line items on the order. This means you will have to update your schema from above to account for these new fields. Below is an example of what the updated schema could look like:



Step 1. Update your schema

Step 2. Next, write an update to your transformation logic to account for updates from the OLTP database. The logic should *insert* a new record for each change to the status of the order. The logic should also retain the historical status for that order and update the table.

Step 3. Next, run the update_query.txt in the folder on Canvas

Step 4. Now, execute the transformation script.

Your result should look something like the following:

Order_key	Order_id	Order_amount	Order_status	Effective_Date	End_Date
1000	1543	223.9	Dis	2/20/2020 12:00	NULL
1001	1543	223.9	venr	09/19/2019 12:00	2/20/2020 12:00

Note: run the following script: *update_query.txt* in Canvas. Before you run above.

Compile your Update DDL and transformation scripts into a single .txt file. Once compiled, submit your work to Canvas using the following naming convention:

- ○ <<Task_No>>_<<Last_Name>>_<<First_Name>>_<<YU_Student_ID>>_Assignment_2
- ○ Example: 4_Chiazza_Brandon_2000930_Assignment_2