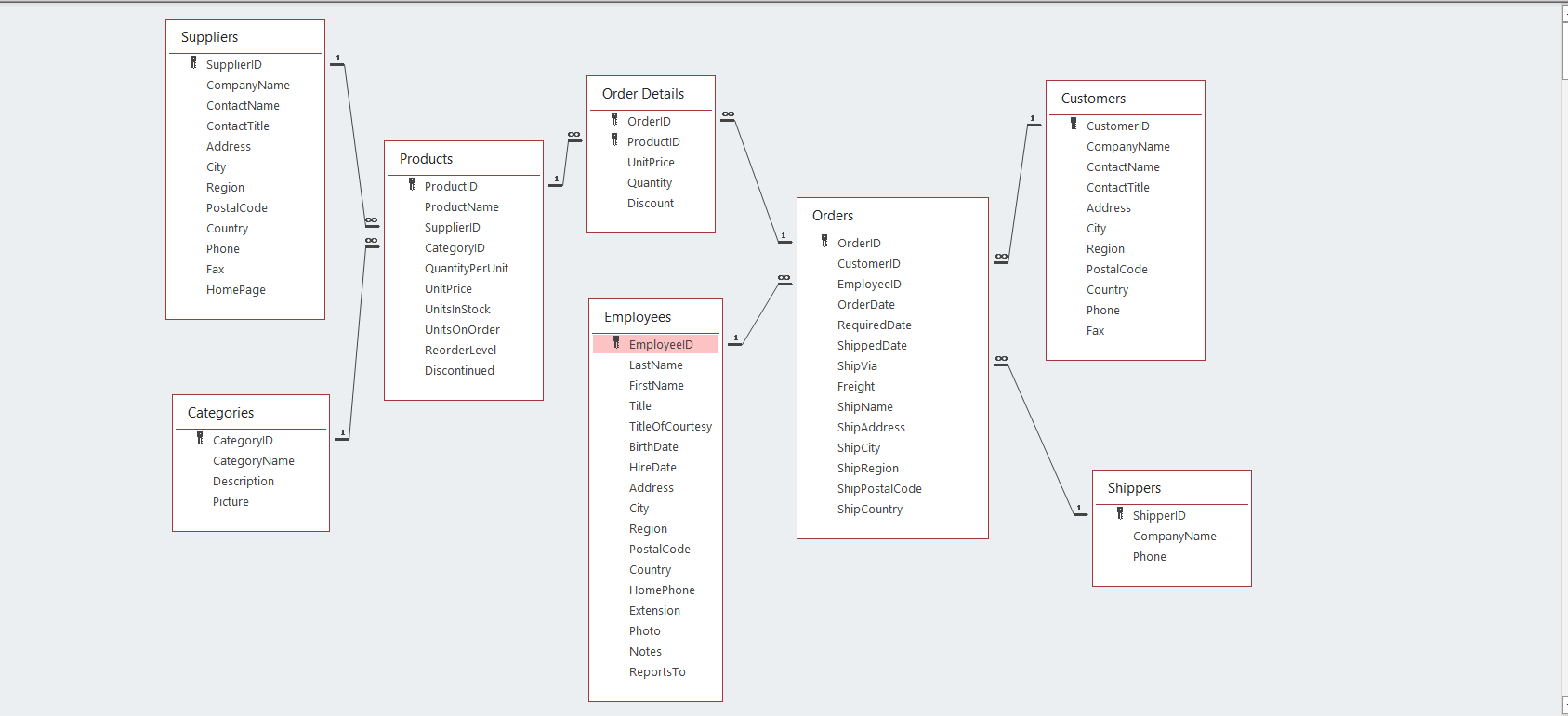
# Option 1: Data Warehouse Dimensional Modeling Project

This project aims to give students the experience of developing a working data warehouse using a commercial data warehousing system based on a current transactional database. See the ERD of Jensen Co. below:



The Jensen company is a distributor of foods around the world. They have hired you and your team as data engineering consultants to design, create, and populate a data warehouse given their current database management system. They have given you access to some of their historical data. It is currently located in the Githib repository . Which will need to be opened in Access. The most important business processes their stakeholders care about as of right now are

1. The process of a customer placing an order.
2. The process of tracking inventory.
3. The process of ordering supplies from their suppliers.

COMMENTS:  
- We would like to have data about the warehouse that the products are stored

Begin by utilizing the 4-step dimensional modeling process we have learned and applied in class.

The deliverables of this project are as follows:

**Project Outline**

* 1. Overview
* By the end of this project, we will have populated a new data warehouse for Jensen Co. As their data engineers, our job is to improve their business process and make their database more efficient and easier to use. For that, we will update their transactional database system and implement three star-schemas for the (1) customer placing order, (2) the inventory tracking process, and (3) the ordering supplies from the supplier process.
* These three start schemas will then be put together into one data warehouse.
* Our purpose will also be to move their database tool from Microsoft Access to Snowflake.
  1. Technology requirements:
* Initially, data is stored in Microsoft Access. The data needs to be migrated into Snowflake using Excel CSV files.
* Lucid will be used to create ERD Diagrams.
* Snowflake to create and design the data warehouse.
  1. Deadlines:
* Milestone 1: March 31st
* Milestone 2: April 7th
* Presentation and Final Deliverable: April 18th

**Milestone 1**

**Process 1: The process of a customer placing an order.**

* + 1. Select the business process.

We are modeling the process of individuals placing an online order distributed by Jensen Co. An order is placed by the customer, then an employee packages the order and finally, a shipper delivers the order to the customer.

* + 1. Declare the grain.

Each row in this table represents a single item within the order placed by the customer. It does not represent an entire order, it is one item at a time.

* + 1. Identify the Dimensions.

Customer Dimension

Shipper Dimension

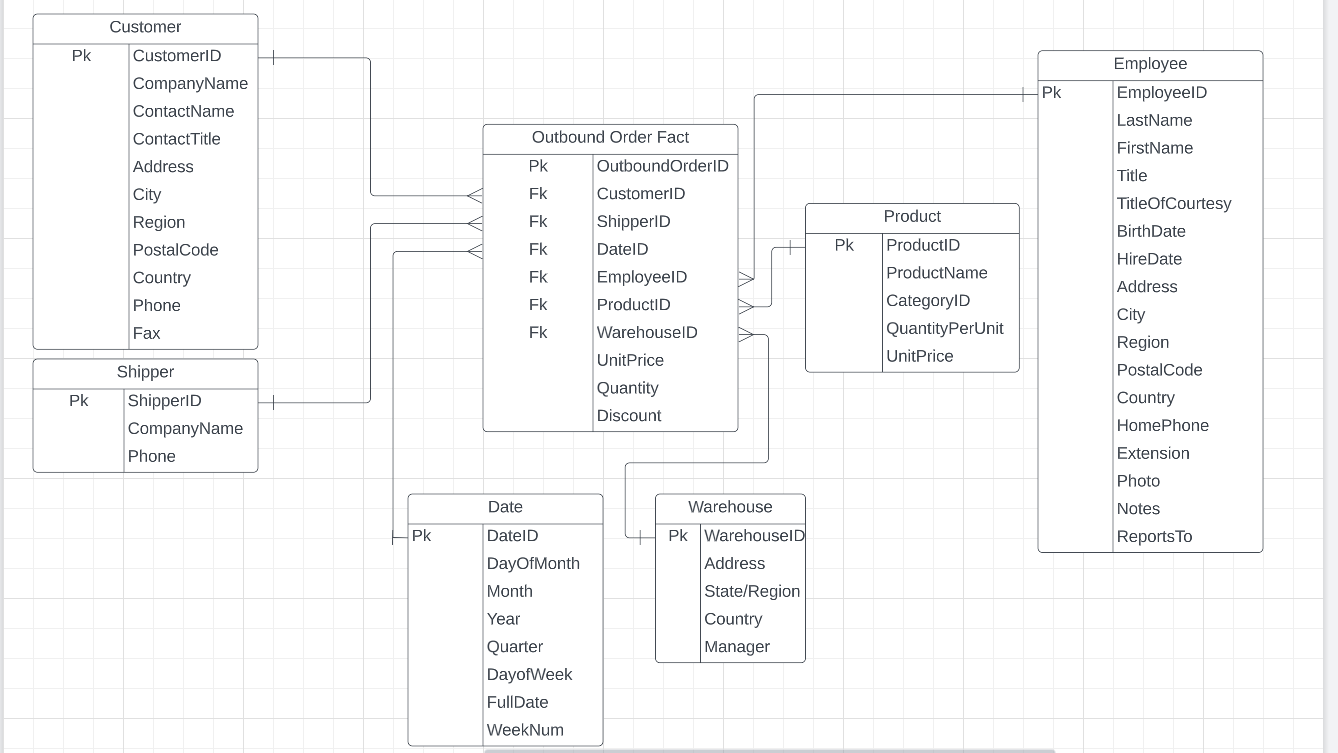
Employee Dimension

Date Dimension

Product Dimension - Categories

* + 1. Identify the Facts.

Outbound Order Detail: will have all the dimensions + UnitPrice, Quantity, Discount, Freight, ShipName, ShipAddress, ShipCity, ShipRegion, ShipPostalCode, ShipCountry



**Process 2: The process of tracking inventory.**

1. Select the business process.

To keep track of the amount of quantity of each product

1. Declare the grain.

Although this is totally a business decision that needs to be made by Jensen Co, each row in this table represents the quantity of a particular item on-hand at a given warehouse during a specific week

1. Identify the Dimensions.

Supplier Dimension

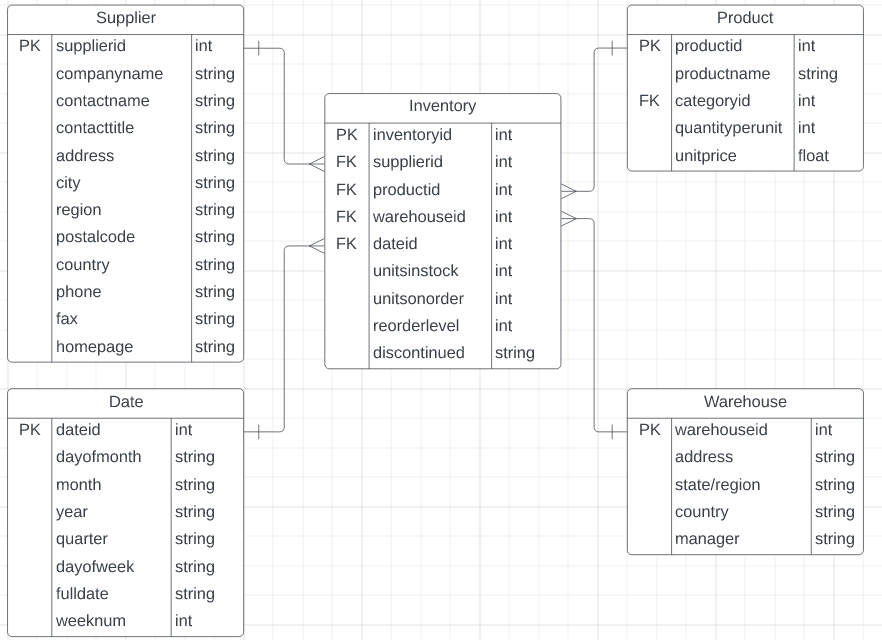
Product Dimension – Linked to category

Date Dimension

Warehouse

1. Identify the Facts.

Inventory: All the dimensions + (Units in stock, Units on order, ReorderLevel, Discontinued) from the Product dimension



**Process 3: The process of ordering supplies from their suppliers.**

1. Select the business process.

The process of ordering supplies from their suppliers.

1. Declare the grain.

Each row in this table represents the quantity of a particular item that needs to be ordered from the supplier at a given warehouse.

1. Identify the Dimensions.

Supplier Dimension

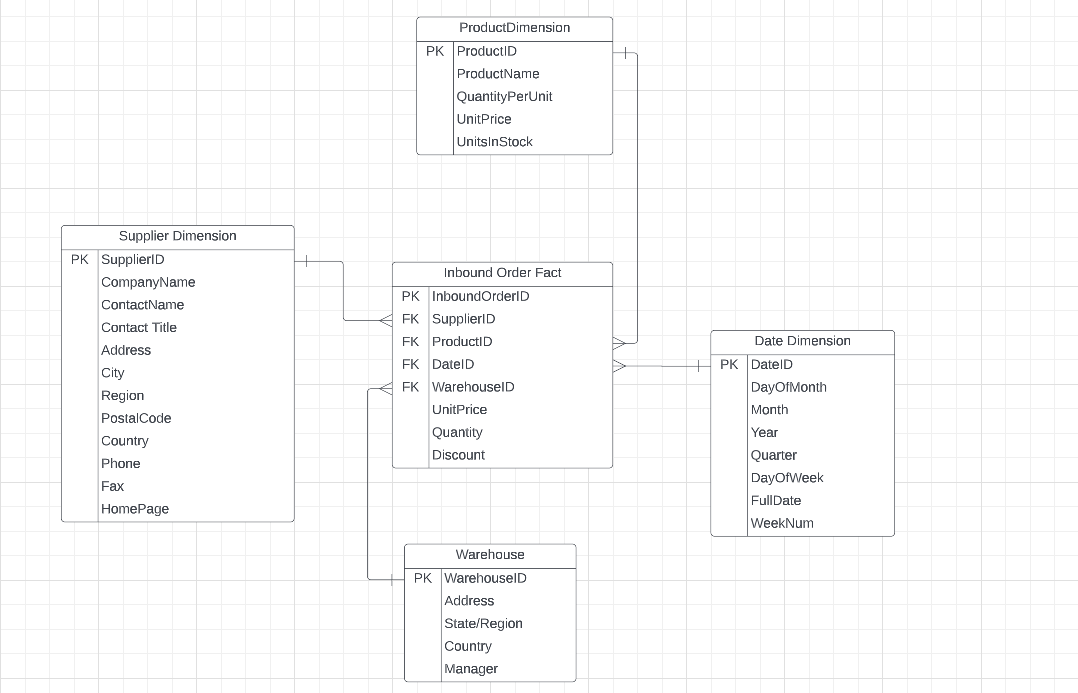
Date Dimension

Product Dimension

Warehouse Dimension

1. Identify the Facts.

Inbound Order Detail: All the dimensions keys + Unit Price, Quantity, and Discount



* Create a draft bus matrix given the above prompt.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Business process / dimension** | **Customer** | **Shipper** | **Employee** | **Date** | **Product** | **Supplier** | **Warehouse** |
| **Customer placing an order** | x | x | x | x | x |  | x |
| **Tracking Inventory** |  |  |  | x | x | x | x |
| **Placing an order to a supplier** |  |  |  | x | x | x | x |

* Create a data warehouse on Snowflake and import the Microsoft access data.

S3 bucket info:

https://finalproject5360.s3.us-west-1.amazonaws.com/JensenCo.MDB

**Credentials for GROUP ACCOUNT**

Username: coolgroup123

Password: Datafinal1

CQB86705

Email: [a02334309@usu.edu](mailto:a02334309@usu.edu)

<https://qyyfogg-cqb86705.snowflakecomputing.com>

**Milestone 2:**

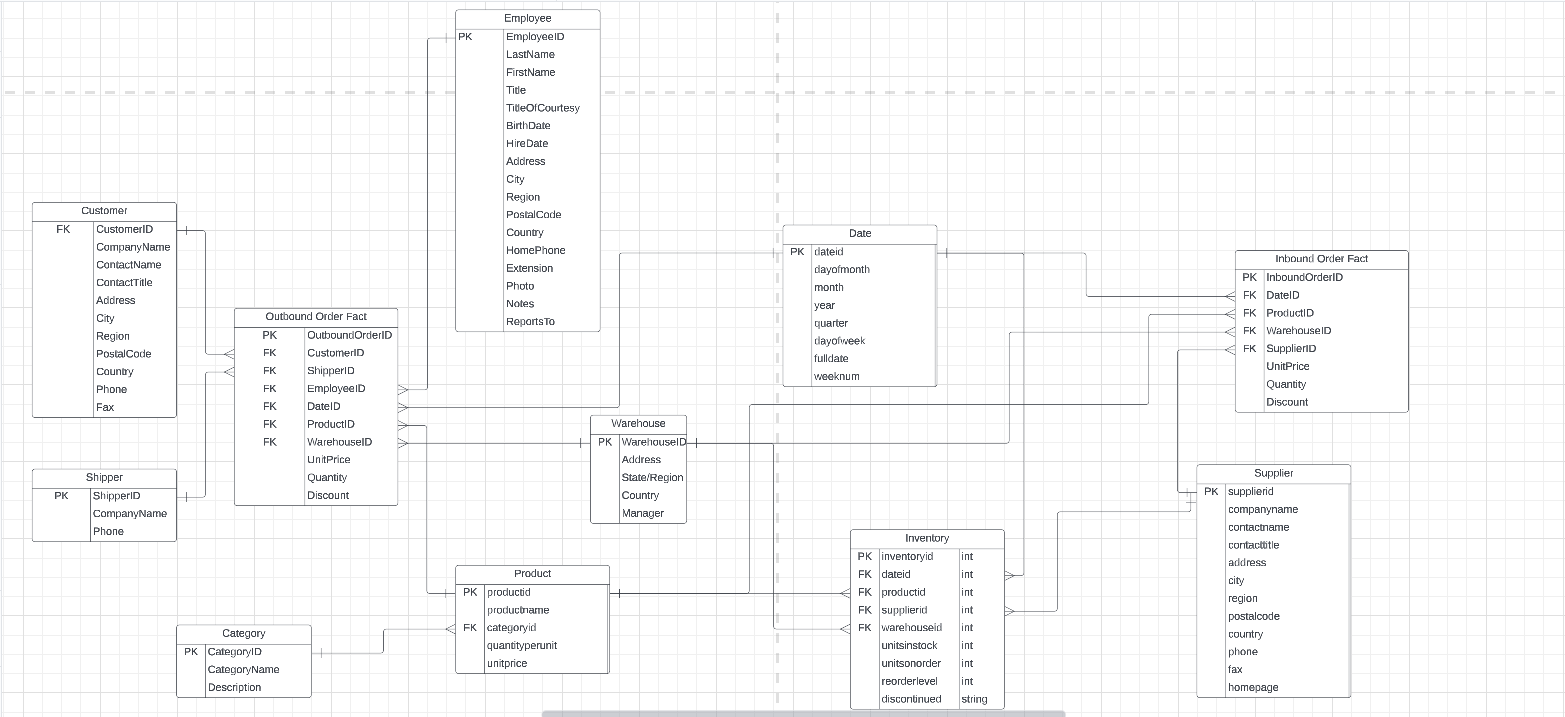
1. Based on the feedback given for Milestone 1. Combine the three-star schemas to create one large enterprise DW and identify conformed dimensions. Turn in a screenshot of your combined star schemas. Use LucidChart.

Feedback from Milestone 1

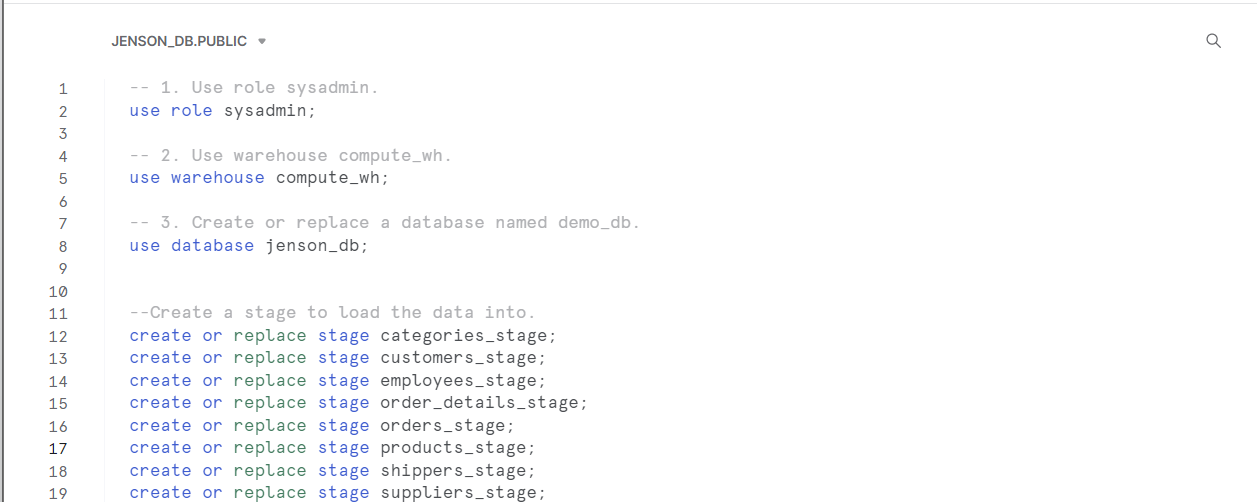
A few notes: Continue focusing primarily on the sales fact table (outbound order fact) and its dimensions. For the other fact tables, you may need to present something like, "This is how we recommend building them and populating them in the future...." Consider, do you want to create any derived facts? Do you want to create an extended\_sales\_price, which could be the sales price \* quantity? You do not have to, but this could be an additional fact to consider creating from the given data. Should the freight measure be included in the Outbound Order Fact table? If so, where? How are you going to populate the report\_to field in the employee\_dim table? This is a bit out of the scope of the course, but I want you to consider how (in the future) you would populate a field like this. They report to an employee, but it is in the employee table. Page 217 of the textbook discusses this. Again, this is just something to note; you do not need to create other tables to accommodate this. However, you can if you would like. As you move forward, think about how you will populate these tables? Should you load the data into Snowflake into a temporary database, and then write queries, as needed, to load the data into the warehouse schema? Further, I would consider tracking down 10 transactions in the relational data, finding all associated data to those transactions (customers, employees, etc.), and loading that into Snowflake, to begin with. Then, write queries with that small amount of data until you get the warehouse to populate correctly, then load in all of the data. It can be easier to figure out processes with fewer data (and less costly!). OR you could consider writing queries against the current tables, exporting the output, and then importing it into Snowflake. Great work on this! This is an iterative process!

Steps done in Milestone 2:

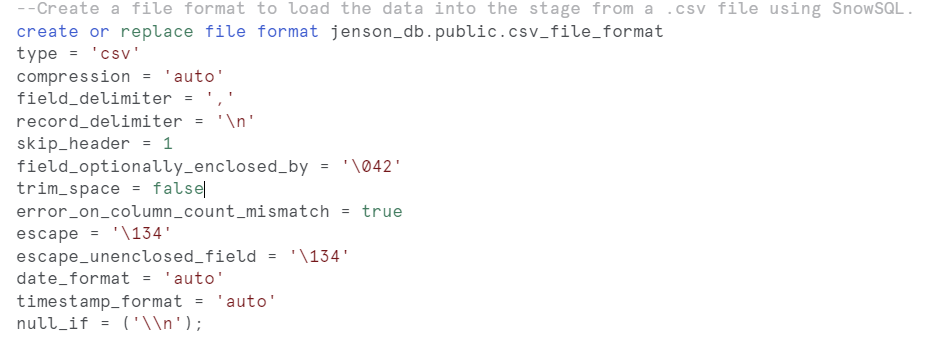
1. Combined the three ERDs to create one big star schema ERD



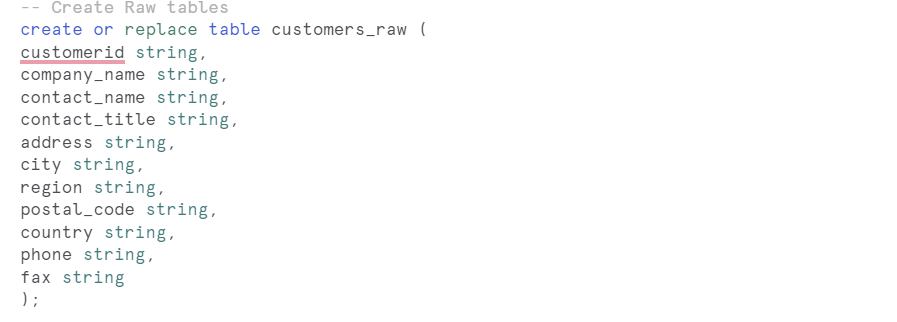
1. In SnowSQL, imported each CSV file into the stages
   1. Sample Code: --put file://C:\Users\Owner\Downloads\final\_project\%name%.csv @%name%\_stage;
2. In Snowflake, created a stage for each table:
   1. Sample Code: create or replace stage customers\_stage;
   2. Picture:



1. Created a file format to load the data into the stage from a .csv file using SnowSQL.



1. Next, created a raw table for each stage



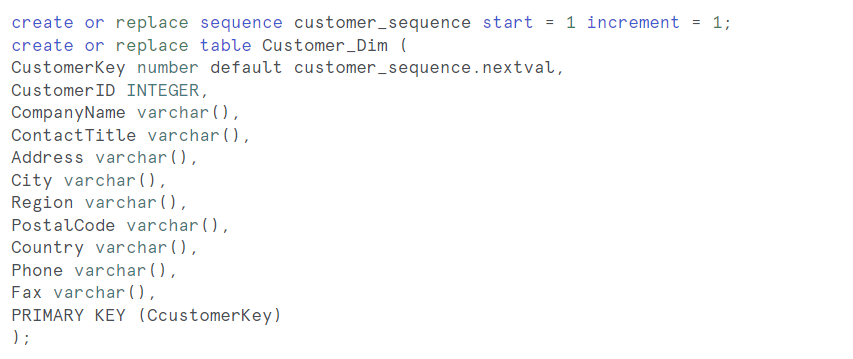
1. Copied the data from the stage into the raw and loaded it

Sample code: copy into categories\_raw from @categories\_stage/categories.csv.gz

file\_format = (format\_name = 'csv\_file\_format');



1. Defined each table and the data types that each column will have. Also linked are the primary and foreign keys.
   1. Picture:



1. Other:

* Got rid of the columns Employee[Photo] and Category[Picture]
* Created and defined a Warehouse table, which will be empty but ready to use.
* Created a code storage file so that our code does not get deleted if the Snowflake account’s credits run out

1. Still to do:

* the tables are not populated. Will need to use the INSERT INTO and upload the data from the adequate tables.
* For the added dimensions and fields, we need to specify how we will track and record the data - Explain how they would work.

**Presentation & Final Deliverable:**

1. Document code is used to transform data from the current format to be input into the data warehouse schema given. Turn in the documentation of your code. All applicable data should be included in the data warehouse.
   1. CSV file upload into Snowflake
2. Prepare a 8-minute presentation on your project. For the first five minutes, present to Jensen Co's data team. For the last ten minutes, describe your process to the class. Describe the ah-ahs, oh-nos, and key takeaways.
3. Turn in your portfolio (which should include Milestones 1-3) which should be a documentation of your project (i.e., someone should be able to recreate your project by working through your documentation).

<https://us04web.zoom.us/j/71398861002?pwd=C8S8QuENWYbkPebJ0h6PEXrCxbbzbP.1>