

Develop ELT and Staging in Snowflake

Questions

- Name of your database: **IMT577_DW_JWALA_SURESH**
- What went well for you?
 - Created the data warehouse
 - Created stages and staging tables
 - Successfully loaded data from Azure Blob Storage to the tables

- What didn't go well?

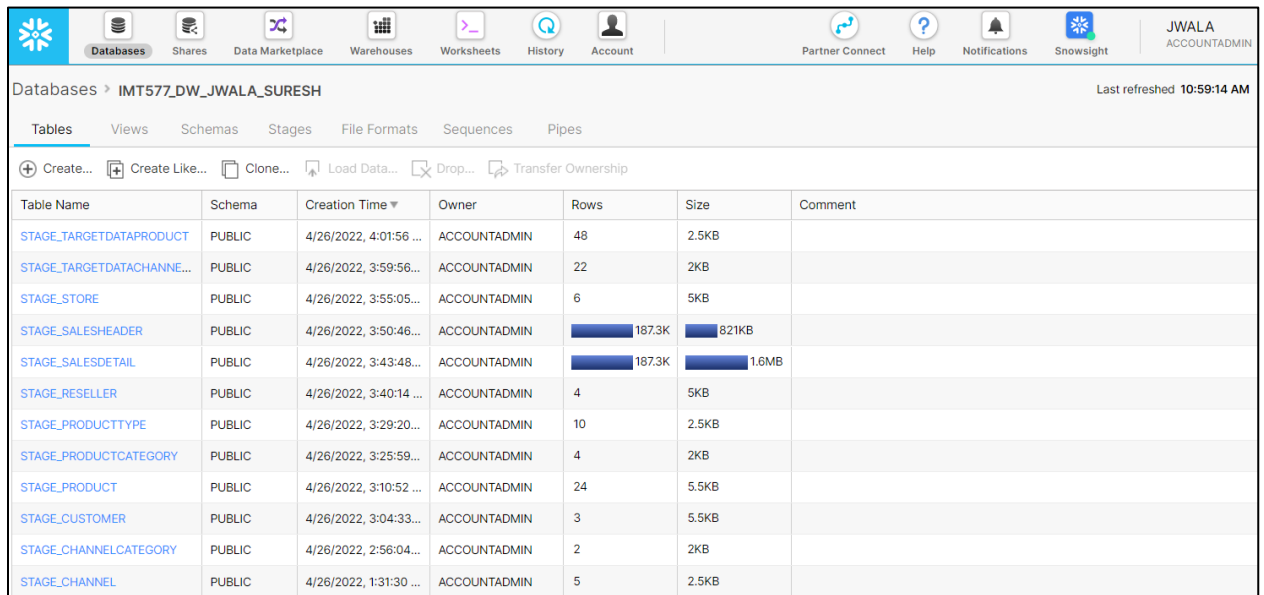
While loading data from Azure Blob Storage faced some issues due to data type mismatch for Created Date and Modified Date columns in the provided data. Verified the values in the CSV and tried using Timestamp and Date data types. Reached out to Professor and learned that VARCHAR was the correct data type to be used and thus fixed the issue.

- Did you learn something significant to share with your peers?
 - I learned that while creating the staging tables we should enter the columns in the exact order as given in the CSV files. I got an error while creating one of the stage tables and that's when I realized this. Also, I tried out giving a different column name than in the CSV to see if that will cause any issue. Trying out like this helped me understand what I need to focus on while creating the stage tables.
 - Stages and staging tables can have different names, but it is easier to use the same name to avoid confusion.
- Do you have any tips or tricks to share with the class?
 - To ensure that I am not missing any steps while creating the staging tables, I created a document with the step-by-step procedure explained in the video. This helped me to avoid rewatching the videos and save time. Also, I believe this will be useful for future reference.
 - Including all CSV files as separate sheets, I created a single Google sheet document. This helped me to easily toggle between different data sheets and ensure that I am not missing out on any columns while creating the stage tables.

- Referring to the normalized ERD model while creating the staging tables was also helpful to ensure that the right data type is used for columns.

Screenshots

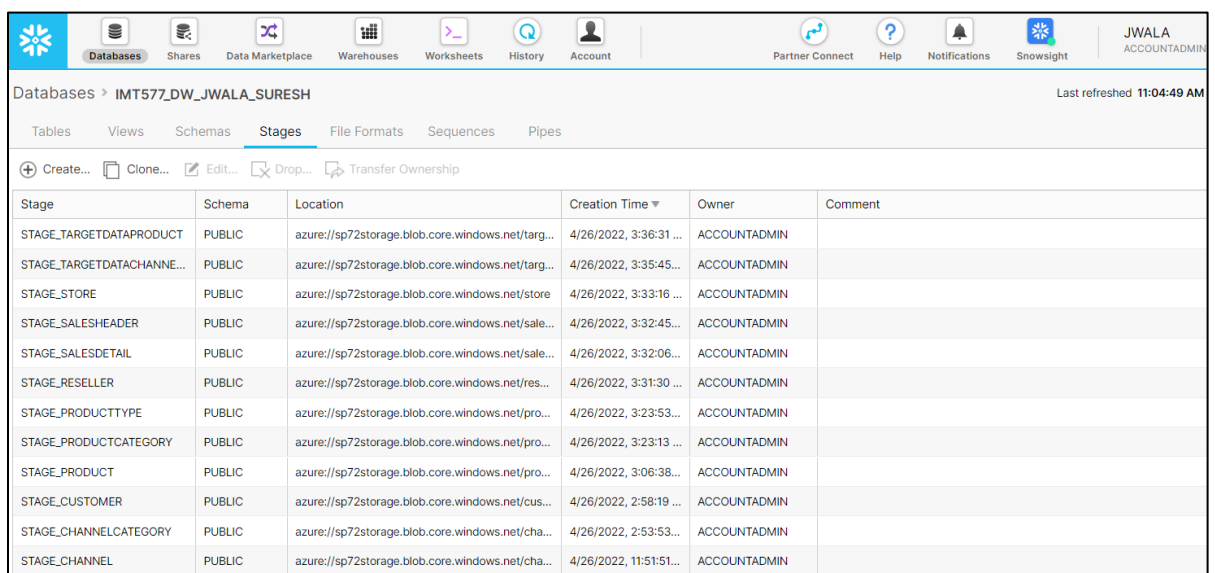
- Screenshot of your database in Snowflake. Make sure it includes the Staging tables



The screenshot shows the Snowflake Databases interface for the database 'IMT577_DW_JWALA_SURESH'. The 'Tables' tab is selected, displaying a list of staging tables. The table includes columns for Table Name, Schema, Creation Time, Owner, Rows, Size, and Comment. The tables listed are STAGE_TARGETDATAPRODUCT, STAGE_TARGETDATACHANNE..., STAGE_STORE, STAGE_SALESHEADER, STAGE_SALESDETAIL, STAGE_RESELLER, STAGE_PRODUCTTYPE, STAGE_PRODUCTCATEGORY, STAGE_PRODUCT, STAGE_CUSTOMER, STAGE_CHANNELCATEGORY, and STAGE_CHANNEL.

Table Name	Schema	Creation Time	Owner	Rows	Size	Comment
STAGE_TARGETDATAPRODUCT	PUBLIC	4/26/2022, 4:01:56 ...	ACCOUNTADMIN	48	2.5KB	
STAGE_TARGETDATACHANNE...	PUBLIC	4/26/2022, 3:59:56...	ACCOUNTADMIN	22	2KB	
STAGE_STORE	PUBLIC	4/26/2022, 3:55:05...	ACCOUNTADMIN	6	5KB	
STAGE_SALESHEADER	PUBLIC	4/26/2022, 3:50:46...	ACCOUNTADMIN	187.3K	821KB	
STAGE_SALESDETAIL	PUBLIC	4/26/2022, 3:43:48...	ACCOUNTADMIN	187.3K	1.6MB	
STAGE_RESELLER	PUBLIC	4/26/2022, 3:40:14 ...	ACCOUNTADMIN	4	5KB	
STAGE_PRODUCTTYPE	PUBLIC	4/26/2022, 3:29:20...	ACCOUNTADMIN	10	2.5KB	
STAGE_PRODUCTCATEGORY	PUBLIC	4/26/2022, 3:25:59...	ACCOUNTADMIN	4	2KB	
STAGE_PRODUCT	PUBLIC	4/26/2022, 3:10:52 ...	ACCOUNTADMIN	24	5.5KB	
STAGE_CUSTOMER	PUBLIC	4/26/2022, 3:04:33...	ACCOUNTADMIN	3	5.5KB	
STAGE_CHANNELCATEGORY	PUBLIC	4/26/2022, 2:56:04...	ACCOUNTADMIN	2	2KB	
STAGE_CHANNEL	PUBLIC	4/26/2022, 1:31:30 ...	ACCOUNTADMIN	5	2.5KB	

- Screenshot of your database Stage. This is different from your staging tables



The screenshot shows the Snowflake Stages interface for the database 'IMT577_DW_JWALA_SURESH'. The 'Stages' tab is selected, displaying a list of staging stages. The table includes columns for Stage, Schema, Location, Creation Time, Owner, and Comment. The stages listed are STAGE_TARGETDATAPRODUCT, STAGE_TARGETDATACHANNE..., STAGE_STORE, STAGE_SALESHEADER, STAGE_SALESDETAIL, STAGE_RESELLER, STAGE_PRODUCTTYPE, STAGE_PRODUCTCATEGORY, STAGE_PRODUCT, STAGE_CUSTOMER, STAGE_CHANNELCATEGORY, and STAGE_CHANNEL.

Stage	Schema	Location	Creation Time	Owner	Comment
STAGE_TARGETDATAPRODUCT	PUBLIC	azure://sp72storage.blob.core.windows.net/targ...	4/26/2022, 3:36:31 ...	ACCOUNTADMIN	
STAGE_TARGETDATACHANNE...	PUBLIC	azure://sp72storage.blob.core.windows.net/targ...	4/26/2022, 3:35:45...	ACCOUNTADMIN	
STAGE_STORE	PUBLIC	azure://sp72storage.blob.core.windows.net/store	4/26/2022, 3:33:16 ...	ACCOUNTADMIN	
STAGE_SALESHEADER	PUBLIC	azure://sp72storage.blob.core.windows.net/sale...	4/26/2022, 3:32:45...	ACCOUNTADMIN	
STAGE_SALESDETAIL	PUBLIC	azure://sp72storage.blob.core.windows.net/sale...	4/26/2022, 3:32:06...	ACCOUNTADMIN	
STAGE_RESELLER	PUBLIC	azure://sp72storage.blob.core.windows.net/res...	4/26/2022, 3:31:30 ...	ACCOUNTADMIN	
STAGE_PRODUCTTYPE	PUBLIC	azure://sp72storage.blob.core.windows.net/pro...	4/26/2022, 3:23:53...	ACCOUNTADMIN	
STAGE_PRODUCTCATEGORY	PUBLIC	azure://sp72storage.blob.core.windows.net/pro...	4/26/2022, 3:23:13 ...	ACCOUNTADMIN	
STAGE_PRODUCT	PUBLIC	azure://sp72storage.blob.core.windows.net/pro...	4/26/2022, 3:06:38...	ACCOUNTADMIN	
STAGE_CUSTOMER	PUBLIC	azure://sp72storage.blob.core.windows.net/cus...	4/26/2022, 2:58:19 ...	ACCOUNTADMIN	
STAGE_CHANNELCATEGORY	PUBLIC	azure://sp72storage.blob.core.windows.net/cha...	4/26/2022, 2:53:53...	ACCOUNTADMIN	
STAGE_CHANNEL	PUBLIC	azure://sp72storage.blob.core.windows.net/cha...	4/26/2022, 11:51:51...	ACCOUNTADMIN	

- Screenshot of two stage tables datasets demonstrating that they are populated

1. STAGE_STORE Table

The screenshot shows the Snowflake SQL Editor interface. The top navigation bar includes icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, History, and Account. The user is logged in as JWALA ACCOUNTADMIN. The current session is 'IT Session 5 - Steps' with a query titled 'Assignment 5.12'. The query is as follows:

```
-- Create Stage for loading data
21 CREATE OR REPLACE FILE FORMAT CSV_SKIP_HEADER
22 TYPE = 'CSV'
23 field_delimiter = ','
24 FIELD_OPTIONALLY_ENCLOSED_BY = ''
25 skip_header = 1;
26
27 --View Data
28 SELECT * FROM Stage_Store;
```

The query has been executed successfully, returning 6 rows in 84ms. The results are displayed in a table with the following columns: Row, STOREID, SUBSEGMENTID, STORENUMBER, STOREMANAGER, ADDRESS, CITY, STATEPROVINCE, COUNTRY, POSTALCODE, PHONENUMBER, CREATEDDATE, CREATEDBY, MODIFIEDDATE, and MODIFIEDBY. The data is as follows:

Row	STOREID	SUBSEGMENTID	STORENUMBER	STOREMANAGER	ADDRESS	CITY	STATEPROVINCE	COUNTRY	POSTALCODE	PHONENUMBER	CREATEDDATE	CREATEDBY	MODIFIEDDATE	MODIFIEDBY
1	1	1	5	Freida Williams	90021 State ...	Atlanta	Georgia	United States	30301	404-902-1067	1/1/13 21:23	company(SQ...	NULL	NULL
2	2	2	34	Michelle Miller	16371 14th S...	St Louis	Missouri	United States	63103	557-704-9940	1/1/13 21:23	company(SQ...	NULL	NULL
3	3	2	39	Keith Chum	16205 138th ...	Jackson	Mississippi	United States	39201	601-250-4658	1/1/13 21:23	company(SQ...	NULL	NULL
4	4	3	21	Mark Weber	8365 10th Av...	Bentonville	Arkansas	United States	72716	479-950-66...	1/1/13 21:23	company(SQ...	NULL	NULL
5	5	4	8	James Trunk	437 Main Av...	Little Rock	Arkansas	United States	72202	501-168-3940	1/1/13 21:23	company(SQ...	NULL	NULL
6	6	6	10	Joann Chan	1037 2nd Str...	Jefferson City	Missouri	United States	65103	573-398-30...	1/1/13 21:23	company(SQ...	NULL	NULL

The screenshot shows the Snowflake Table Properties for the STAGE_STORE (PUBLIC) table. The table is located in the Databases > IMT577_DW_JWALA_SURESH > STAGE_STORE (PUBLIC) schema. The table was last refreshed at 11:10:33 AM. The table has 14 columns, all of which are nullable and have a default value of NULL. The columns are:

Column Name	Ordinal	Type	Nullable	Default	Comment
STOREID	1	NUMBER(38,0)	false	NULL	
SUBSEGMENTID	2	NUMBER(38,0)	true	NULL	
STORENUMBER	3	NUMBER(38,0)	true	NULL	
STOREMANAGER	4	VARCHAR(255)	true	NULL	
ADDRESS	5	VARCHAR(255)	true	NULL	
CITY	6	VARCHAR(255)	true	NULL	
STATEPROVINCE	7	VARCHAR(255)	true	NULL	
COUNTRY	8	VARCHAR(255)	true	NULL	
POSTALCODE	9	NUMBER(38,0)	true	NULL	
PHONENUMBER	10	VARCHAR(255)	true	NULL	
CREATEDDATE	11	VARCHAR(255)	true	NULL	
CREATEDBY	12	VARCHAR(255)	true	NULL	
MODIFIEDDATE	13	VARCHAR(255)	true	NULL	
MODIFIEDBY	14	VARCHAR(255)	true	NULL	

2. Stage_ProductType Table

IT Session 5 - Steps | Assignment 5.12 | + | -

Run | All Queries | Saved 7 seconds ago | ACCOUNTADMIN | IMT577_DW (XS) | IMT577_DW_JWAL... | PUBLIC

```
17 ***** /
18
19 -- Create Stage for loading data
20 CREATE OR REPLACE FILE FORMAT CSV_SKIP_HEADER
21 TYPE = 'CSV'
22 field_delimiter = ','
23 FIELD_OPTIONALLY_ENCLOSED_BY = '"'
24 skip_header = 1;
25
26 --View Data
27
28 SELECT * FROM Stage_ProductType;
```

Results | Data Preview | Open

Query ID | SQL | 83ms | 10 rows

Filter result... | Download | Copy

Row	PRODUCTTYPEID	PRODUCTCATEGORYID	PRODUCTTYPE	CREATEDDATE	CREATEDBY	MODIFIEDDATE	MODIFIEDBY
1	1	1	Cosmetics	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
2	2	1	Jewelry	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
3	3	2	Baby 0-2	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
4	4	2	Kids 2-6	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
5	5	2	Kids 7-14	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
6	6	3	Women's Casual	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
7	7	3	Women's Eveningwear	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
8	8	3	Women's Formal	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
9	9	4	Men's Casual	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL
10	10	4	Men's Formal	2013-01-01 20:55:47	company\SQLServerSalesAcco...	NULL	NULL

Databases > IMT577_DW_JWALA_SURESH > STAGE_PRODUCTTYPE (PUBLIC) | Last refreshed 11:19:19 AM

Tables | Views | Schemas | Stages | File Formats | Sequences | Pipes

Load Table

Column Name	Ordinal	Type	Nullable	Default	Comment
PRODUCTTYPEID	1	NUMBER(38,0)	false	NULL	
PRODUCTCATEGORYID	2	NUMBER(38,0)	true	NULL	
PRODUCTTYPE	3	VARCHAR(255)	true	NULL	
CREATEDDATE	4	VARCHAR(255)	true	NULL	
CREATEDBY	5	VARCHAR(255)	true	NULL	
MODIFIEDDATE	6	VARCHAR(255)	true	NULL	
MODIFIEDBY	7	VARCHAR(255)	true	NULL	