

Snowflake Assignment

Pre-requisites

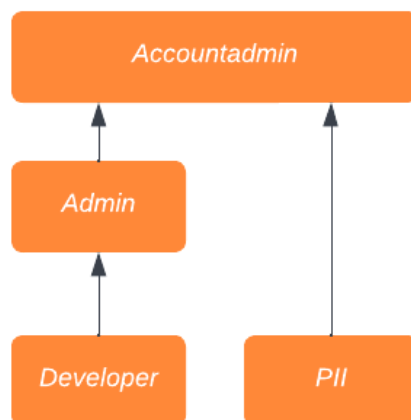
1. Snowflake trial account
2. Snowsql setup
3. Any functional cloud account
4. Public GitHub repo

General Instructions

1. Create one worksheet with all queries for this assignment
2. Keep a backup of this worksheet in a public git repo, if your Snowflake account expires there's no way to recover the worksheet.
3. While working on the assignment just consider the steps as guidelines.
4. There are a few hidden Snowflake concepts that you'll need to use to complete this assignment.
5. Few audit columns are supposed to be added to all the tables.
elt_ts timestamp - when the record was inserted into Snowflake
elt_by varchar - application name from which the record was inserted. Can be hard coded for now.
file_name varchar - The name of file used to insert data into the table.

Assignment

1. Create roles as per the below-mentioned hierarchy. Accountadmin already exists in Snowflake.



2. Create an M-sized warehouse named assignment_wh and use it for all the queries.
3. Switch to the admin role.
4. Create a database assignment_db
5. Create a schema my_schema
6. Create a table using any sample csv. You can get 1 by googling for sample csv's. Preferably search for sample employee dataset so that you have PII related columns else you can consider any column as PII.
7. Also, create a variant version of this dataset.
8. Load the file into an external and internal stage.
9. Load data into the tables using copy into statements. In one table load from the internal stage and in another from the external.
10. Upload any unrelated parquet file to the stage location and infer the schema of the file.
11. Run a select query on the staged parquet file without loading it to a snowflake table.
12. Add masking policy to the PII columns such that fields like email, phone number, etc. show as **masked** to a user with the developer role. If the role is PII the value of these columns should be visible.