MBIS623: Tutorial W10

Snowflake: Cloud Data Warehousing Platform Introduction

In This Tutorial

- Snowflake video introduction
- Logging into Snowflake
- Locating data and tools
- Basic queries to create dimension tables

Task 1

Snowflake at a Glance

Developed in 2012, Snowflake is a fully managed SaaS (software as a service) that provides a single platform for data warehousing, data lakes, data engineering, data science, data application development, and secure sharing and consumption of real-time / shared data.

Snowflake features out-of-the-box features like separation of storage and compute, on-the-fly scalable compute, data sharing, data cloning, and third-party tools support to handle the demanding needs of growing enterprises.

- Go to the Videos section in Snowflake Getting Started page: https://docs.snowflake.com/en/other-resources.html#videos and watch these tree brief introductions:
  - Architecture & Key Concepts: https://www.youtube.com/watch?v=dZlBCLLL7UA
  - Introduction to Snowflake: https://www.youtube.com/watch?v=fEtoYweBNQ4
  - Introduction to Worksheets & Queries: https://www.youtube.com/watch?v=mH4pvxcfNYw

Task 2

Logging into Snowflake

Use the following credentials to login to Snowflake:

- Login URL: https://uc-teaching.snowflakecomputing.com/console/login#/
- Username: **`<usercode>_MBIS623`**; e.g.: **`xyz123_MBIS623`**
- Password: same as the username
  - IMPORTANT: you will be required to reset the password upon the first login
  - IMPORTANT: have your web browser store the credentials for you
- Schema: **`<usecrode>_MBIS623`**

One you've logged in, locate your schema — it will be empty at this stage; this is where you can create your tables.

Locate the NYC311 schema, with the data that will be familiar to you now — it is a copy of the data that from the NYC311 schema from the MySQL Server we've been using earlier in this course.

Note the overall organization of data in Snowflake:

- The top-level of the data hierarchy is the database level, which in our case is **NYC311**
- Within the database there are schemas, where the actual data objects live

Task 3

Snowflake Worksheet Introduction — Basic Queries

IMPORTANT! Set your current schema by entering this at the top of your new worksheet:

```
use schema NYC311.<uc_usercode>_MBIS623;
```

- For this part of the tutorial explore the **`NYC311/NYC311.SERVICE_REQUEST_1M`** and **`SERVICE_REQUEST_32M`** tables. Pay attention to the datatypes and other (mainly syntax) differences you notice, as compared to SQLite.

Task 4

NYC311 Data Warehouse Dimension Tables

- For this part of the tutorial try and reproduce the dimension tables, based on the instructions provided in last week's tutorial.

  This is an exploratory task — you'll need to consider appropriate datatypes, syntax, and the location of the tables.

Task 5

NYC311 Data Warehouse Fact Table

- For this part of the tutorial try and reproduce the fact table, based on the instructions provided in the las week's tutorial.
- Then, try and populate the table. During this stage use the 1M table as the data source, not the 32M one.

ANSWERS/QUERIES WILL BE PROVIDED LATER THIS WEEK