Data Engineer Bootcamp (Full-Time)

HM
HIBAHMOHAMMED O SINDI
haboba1417@hotmail.com

WeCloudData

Data Engineering Diploma

Content developed by: WeCloudData Academy

Mandatory: Yes
Group: Yes
In Class: Yes
Submission: No
Grading: No
Demo: No
Duration: 120 Mins

1.1.0 Prerequisites

- An AWS account
- Knowledge of how to provision and terminate an Ubuntu EC2 instance
- Understanding of basic shell commands and Docker commands
- Experience in connecting to a database

1.1.1 Create a database on your Snowflake account

1. Create a database on your Snowflake account, we can call it `TPCDS`;
2. Create a schema under the database `TPCDS`, we can call this schema `RAW`;
3. In your schema "RAW" create a table called "inventory" with the columns and data type like this:

NAME	TYPE	NULLABLE	DEFAULT
INV_DATE_SK	# Number	No	NULL
INV_ITEM_SK	# Number	No	NULL
INV_QUANTITY_ON_HAND	# Number	Yes	NULL
INV_WAREHOUSE_SK	# Number	No	NULL

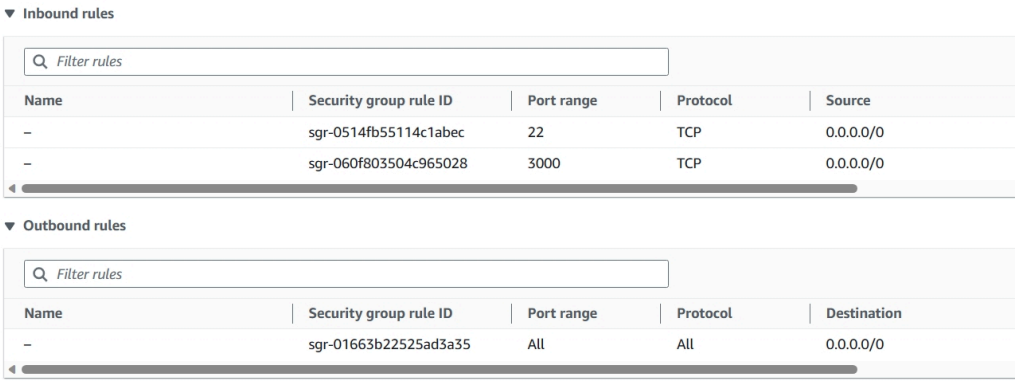
1.1.2 Launching 2 Ubuntu EC2 Instances in your AWS Console

Follow these steps to launch 2 Ubuntu EC2 instances:

Create 2 EC2 instances. One instance is in t2.small size, another is a t2.large. The small size one will be used to install **Metabase**, and the t2.large is used to install **Airbyte**. And please give the instances name in order to indentify them easily, like "Project -Metabase", "Project - Airbyte".

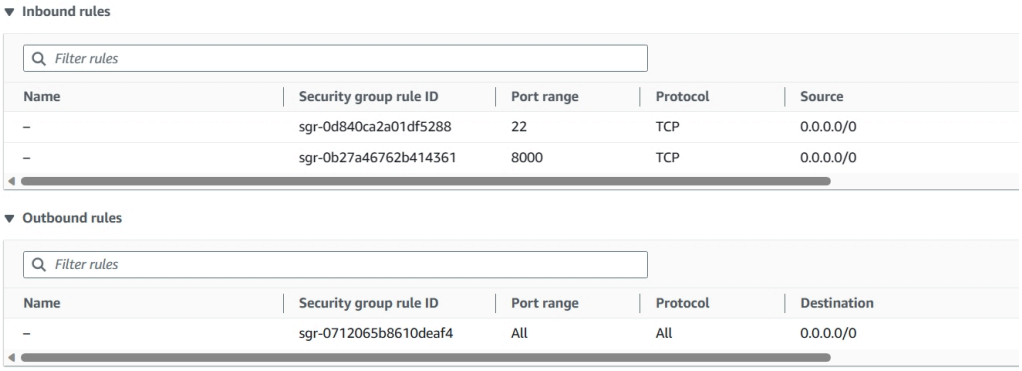
For the instances, please make sure you have the right security group settings:

For Metabase, the security group setting is:



1. Type: t2.small
2. AMI: Ubuntu Server 22.04 LTS (free tier)
3. Firewall: Allow all traffic to 0.0.0.0/0
4. Storage: **20 GB**

For Airbyte, the security group setting is:



1. Type: t2.large
2. AMI: Ubuntu Server 22.04 LTS (free tier)
3. Firewall: Allow all traffic to 0.0.0.0/0
4. Storage: **20 GB**

- You need to upload the EC2 key to the Cloud Shell before you can SSH the EC2 instance, and understand why do we need the EC2 Key.
- You should know how to exit EC2 and back to Cloud Shell.

1.1.3 Docker Installation

Installed Docker on both the created instances above by following the steps:

1. Open a new terminal and SSH into the EC2 instance you just launched using the following command:

   ```
   ssh -i <path to your pem file> ubuntu@<public IP of the EC2 instance>
   ```

2. Update the package list by running the following command:

   ```
   sudo apt update
   ```

3. Install Docker by running the following command:

   ```
   sudo apt install -y docker.io
   ```

4. Add the "ubuntu" user to the "docker" group using the following command:

   ```
   sudo usermod -aG docker ubuntu
   ```

5. Exit the current SSH session by running the following command:

   ```
   exit
   ```

6. SSH into your EC2 instance again to apply the group changes:

   ```
   ssh -i <path to your pem file> ubuntu@<public IP of the EC2 instance>
   ```

Now you have successfully installed the necessary components on your Ubuntu EC2 instance.

```
# Install Docker Compose v2 manually
# Reference: https://docs.docker.com/compose/install/linux/
DOCKER_CONFIG=${DOCKER_CONFIG:-$HOME/.docker}
mkdir -p $DOCKER_CONFIG/cli-plugins
curl -SL https://github.com/docker/compose/releases/download/v2.18.1/docker-compose-linux-x86_64 -o $DOCKER_CONFIG/cli-plugins/docker-compose
chmod +x $DOCKER_CONFIG/cli-plugins/docker-compose
docker compose version
```

When all the above tasks have been done, you can stop the EC2 instance to avoid any unnecessary server costs.

[Lecture Video] Sunday: Orientation Session
[Lecture Video] Tuesday: Introduction to Data Engineering
[Lecture Video] Wednesday: Curriculum
[Lecture Video] Thursday: Curriculum
[Lecture Slide] (Wed) Curriculum Introduction
4
Week 01 - SQL
Chapter overview
Sunday - Basic SQL
[Lecture Materials]SQL basics
[Lab] SQL Basics Exercise
[Lab] Exercise: SQL - Airbnb (Optional)
[Lecture video] SQL Basics
Monday - SQL Join and sub-select
[Lecture Materials] SQL join and sub-select
[Lab] Exercise: Join and Sub-select
[Lecture video] SQL Day 2
Tuesday - SQL Window Function
[Lecture Materials] SQL Window Function
[Lab] Exercise: Window Function
[Lecture video] SQL Day 3
[Lab Video] SQL Lab Solution
Wednesday - SQL DDL and CTE
[Lecure Materials] DDL and CTE
[Lab] SQL ddl
[Lab] SQL CTE
[Lecture Video] SQL Wednesday
Thursday - SQL Review
[Weekly Quiz] SQL - Week 1
[Lecture Video] SQL Thursday Review
[Lecture Slides] SQL Review
5
Week 02 - Python
Chapter overview
Sunday - Python data type and structure
[Lecture Materials] Python Data Structure and Data Types
[Lab] Exercise: Python Data type and structure
[Lab] Exercise: OpenAI ChatBot (Optional)
[Lecture Video] Python Sunday

Thursday - Holiday

Monday - Python Control Flow and Function

[Lecture Material] Python Control Flow and Function

[Lab] Exercise: Python Function

[Lab] Exercise: Python Control Flow

Python Quiz (Multiple-Choice)

[Lecture Video] Python Monday

Tuesday - Pandas 1

[Lecture Material] Pandas 1

[Lab] Pandas Intro

[Lecture Video] Python Tuesday

Wednesday - Pandas 2

[Lecture Material] Pandas 2

[Lab Demo] PandaSQL

[Lab] Exercise: Advanced Pandas

[Lecture Video] Python Wednesday

6

Week 03 - Client Project

Chapter overview

Sunday - Real Client Project Intro

[Lecture Material] Web Scraping

[Real Client Project] Project Requirements

[Note] Project Group Assignment

[Lecture Video] Webscraping Sunday

Monday - Real Client Project Day

Tuesday - Real Client Project

[Real Client Project] Code & Data Submission

Wednesday - RCP

Thursday - RCP

[Lecture Video] Webscraping Thursday

7

Week 04 - Linux and AWS

Chapter overview

Sunday - Linux

[Lecture Material] Linux

[Lab] Exercise: Bash Commands

[Lab] Mini Project: Riyadh Climate Data - Cron Job

[Lecture Video] - Linux Sunday

Monday - AWS Intro

[Lecture Material] AWS Intro

[Lab] AWS Account Setup

[Lab] Workshop AWS EC2

[Lab] Workshop S3

[Lecture Video] AWS Monday

Tuesday - Lambda

[Lab] Workshop: Lambda

[Lecture Material] Lambda

[Lab] Mini Project: Lambda

[Lecture Video] Lambda Tuesday

Wednesday - Practice Day

[Lecture Material] Plan For Today

[Lab video] 2024-03-06

Thursday - Practice Day

[Lecture Material] Plan For Today

[Quiz] Linux and AWS Quiz

[Lab Video] EC2, S3, Lambda workshops demo

8

Week 05 - Docker and Client Project phase 2

Chapter overview

Sunday - Docker I

[Lecture Material] Docker

[Lab] Software Installation: Docker

[Lab] Account Creation Create your Dockerhub account

[Lab] Workshop Demonstrating Hello World Example

[Lab] Workshop: Work with Docker Image

[Lab] Exercise: Basic Docker Commands

[Lecture Video] Docker Sunday

[Lab] Exercise: Basic Docker Commands Updated

Monday - Docker II

[Lab] Workshop: Install Zepplin with Docker

[Lab] Workshop: Docker Compose --Flask

[Quiz] Docker Commands Quiz

[Lecture Video] Docker II - Monday

[Lab] Workshop: Install Zepplin with Docker Updated

Tuesday - Real Client Project Phase 2

[Lecture Video] Learning Roadmap & RCP Feedback

Wednesday - Real Client Project Phase 2

Thursday - Real Client Project Phase 2

RCP project Submission (Competition)

9

Week 06 - Data Warehouse

Chapter overview

Sunday - Snowflake Data Warehouse

[Lecture Material] Snowflake

[Lab- W601]: Software Installation: DBeaver

[Lab-W602]: Account Creation: Snowflake

[Lab-W603]: Software: Connect Snowflake with DBeaver

[Lab-W604]: Exercise: Snowflake

[Lecture Video] Snowflake - Sunday

Shaohua Weekly Review [RCP] - Sunday

[Lab Video] Snowflake Demo - Monday

Monday - Data Warehouse Intro

[Lecture] Data Warehouse Intro

[Quiz-W611] Data Warehouse Concept (Grading!!)

[Lab-W612] Exercise: Use SnowSQL (Optional)

[Lecture Video] Data Warehouse - Monday

Tuesday - SQL in ETL

[Lecture Materials] SQL in ETL

[Lab] TA Exercises Review

[Lecture Video] SQL in ETL - Tuesday

Wednesday - Data Modeling and ETL

[Lecture Material] Data Modeling and ETL

[Lab-W631] Exercise: Data Modelling and ETL (Group)

[Lecture Video] Data Modelling and ETL - Wednesday

Thursday - Data Loading

[Lecture Material] Data Loading

[Lab-W641] Exercise: ETL and Data Loads (Group)

[Lecture Video] Data Loading

10

Week 07 {Project Week} - Capstone Project-1

Chapter overview

Sunday - Data Warehouse Review

[Lecture Material] Agenda For Today

[Lecture Video] Data Warehouse Lab Review - Sunday

Monday - {Capstone Project} Project Intro

[Project Material] Project Guideline

[Project Material] Project Data Overview

[Project Material] Business Requirements Overview

[Project Material] Project Infrastructure Overview

[Lab-W711] Project Task1: Setup Snowflake, EC2 and Docker

[Project Material] Project Introduction (Full-version)

[Lecture Video] Capstone Project Intro - Monday

Tuesday - {Capstone Project} Lambda Setup in Project

[Lab-W721] Project Task2: AWS Lambda Setup

Wednesday - {Capstone Project} Airbyte Setup in Project

[Lab-W731] Project Task3: Airbyte Installation and Configuration

Thursday - {Capstone Project} Self-work On Project Part 1

[Lab] Agenda for Today

[Lab-W711] Project Task1: Setup Snowflake, EC2 and Docker