Data Engineer Bootcamp (Full-Time)

HM
HIBAHMOHAMMED O SINDI
haboba1417@hotmail.com

WeCloudData

Data Engineering Diploma

Content developed by: WeCloudData Academy

About Data

1. Data Background

The dataset utilized originates from TPCDS, a well-known dataset designed for database testing, with a specific emphasis on Retail Sales. It encompasses sales records from both websites and catalogs, along with detailed information on inventory levels for each item within every warehouse. Moreover, it incorporates 15 dimensional tables containing valuable information about customers, warehouses, items, and more.
The dataset is divided into two parts.:

Fact tables	Dimention tables
Catalog_Sales	Date_Dim
Web_Sales	Customer
Inventory	Item
FROM S3	Promotion
	Customer_Gemographics
	Call_Center
	Customer_Address
	Catalog_Page
	Warehouse
	Time_Dim
	Ship_Mode
	Household_Demographics
	Icome_Band
	Web_page
	Web_Site

- **RDS:** All tables, except for the inventory tables, are stored in the Postgres DB in AWS RDS. These tables are refreshed every day with the latest sales data, requiring daily ETL processes.
- **S3 Bucket:** The single Inventory table is stored in a S3 bucket. Each day, a new file containing the most recent data is deposited into the S3 bucket. *Note*: The inventory table typically registers data at the end of each week, leading to one entry per item per warehouse on a weekly basis.

2. Tables in the Dataset

Review the tables' schema here or find it in the Snowflake tables. Multiple tables related to customers are organized horizontally. When constructing the data model, kindly consolidate these tables into a unified customer dimension table.

Dimention Tables

Customer (c)

Column	Datatype	NULLs	Primary Key	Foreign Key
c_customer_sk	identifier	N	Y	
c_customer_id (B)	char(16)	N		
c_current_cdemo_sk	identifier			cd_demo_sk
c_current_hdemo_sk	identifier			hd_demo_sk
c_current_addr_sk	identifier			ca_addres_sk
c_first_shipto_date_sk	identifier			d_date_sk
c_first_sales_date_sk	identifier			d_date_sk
c_salutation	char(10)			
c_first_name	char(20)			
c_last_name	char(30)			
c_preferred_cust_flag	char(1)			
c_birth_day	integer			

Customer_address (ca)

Column	Datatype	NULLs
ca_address_sk	identifier	N
ca_address_id (B)	char(16)	N
ca_street_number	char(10)	
ca_street_name	varchar(60)	
ca_street_type	char(15)	
ca_suite_number	char(10)	
ca_city	varchar(60)	
ca_county	varchar(30)	
ca_state	char(2)	
ca_zip	char(10)	
ca_country	varchar(20)	
ca_gmt_offset	decimal(5,2)	

Course Content
Enter code
×
All
Lecture
Recordings
Practices
1
Program Information
Chapter overview
Program Administration
Grading and Attendance
About the projects in the bootcamp
How to use the Learning Portal
Daily Schedule
2
Surveys
Chapter overview
Surveys
Week 0 Survey - Student Background
Week 3 Survey - Client Project
Week 4 Survey
Project Group Survey
3
Week 00 (Virtual)- Program Preparation
Chapter overview
Week Plan
Week Plan
Software Installation
[Software Installation]: VsCode
[Software Installation]: Jupyter notebook
[Software Installation]: Python
[Software Installation]: MySQL
[Software Installation]: Unbuntu on Mac
[Software Installation] Ubuntu on Windows

[Online Platform Use]: Colab
Pre-bootcamp
Pre-bootcamp Material
Presentations
[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
[Project Material] Project Data Overview