Spring 2025 CS307 Project Part1

> Main Contributors:
>
> Leader and Overall Design: ZHU Yueming
>
> Data Preparation and Documentation: WANG Lishuang, WANG Zhuo
>
> Review: MA Yuxin

General Requirement:

- It is a group project with **only 2 teammates** who are **in the same lab session**. Each group should finish the project independently and submit only one report written by the teammates.
    - The teammate you select for Project 1 will also be your teammate for Project 2. It is not allowed to change teammates once paired.
- You should submit the report before the deadline. All late submissions after the deadline will receive a score of zero.
- DO NOT copy ANY sentences and figures from the Internet and your classmates. Plagiarism is strictly prohibited in this course.
- The text description should be rigorous, the overall design should be logical organised, the report structure and the layout of diagram should be clear and easy to read, otherwise, you will receive a penalty in the scoring stage.
- The number of pages for your report should be between **6** and **11**. Reports **only or less than 6 pages** and **more than 11 pages** will receive a penalty in the scoring stage.

DBMS can help us manage data in a convenient manner and improve the efficiency of data retrieval. Your work of Project 1 is mainly divided into three parts below:

1. Design an E-R diagram based on the provided data file and data relationships.
2. Design a relational database using PostgreSQL according to the provided data file.
3. Import all data into the database.

Background

This project is based on the requirements for a fictional international trading company called Southern Union manufacture of Science and Technology of China (SUSTech). In this project, you need to design a database for the Marketing Department of SUSTech. The database is used for storing the organizational structure of the Marketing Department and details of the sales contracts.

SUSTech is a top-tier international trading company based in China. The customers of SUSTech come from 8 main areas inside China and around the world including Europe, America, Asia, Eastern China, Northern China, Southern China, and Southwestern China, Hong Kong, Macao and Taiwan regions of China, each of which is handled by a Supply Center. It should be noted that the Asia Supply Center manages businesses in all countries in Asia except for China. Each Supply

Center has a manager who is responsible for the daily goings-on of the entire center. The names of the managers are listed below:

Europe	America	Asia	Eastern China	Northern China	Southern China	Southwestern China	Hong Kong, Macao and Taiwan regions of China
Qian Qianqiu	Kong Yibo	David Robinson	You Xiangxing	Jiang Feiqing	Liz Jones	Zheng Jiaxuan	Gaston Harris

Data Description

The data includes **50,000** contracts, where a contract contains one or more orders. Each order represents a **single kind of** product that has been ordered, i.e., different products should be separated into multiple orders. Here is the explanation of the columns:

contract num	client enterp	supply cente	country	city	industry	product code	product name	product model	unit price	quantity	contract dat	estimated de	lodgement d	director	salesman	salesman nu	gender	age	mobile phone
CSE0000000	Allianz	Europe	Germany		Insurance	MC32085	Money Counter	MoneyCounterA	780	940	2017/3/11	2017/4/29	2017/3/16	Qian Qianqiu	Shen Haifeng	12419695	Female	38	15620510105
CSE0000000	Allianz	Europe	Germany		Insurance	S6I1543	Software Encrypti	SoftwareEncryp	690	510	2017/3/11	2017/5/2	2017/4/10	Qian Qianqiu	Xie Ercui	12135700	Female	46	18981092712
CSE0000000	Allianz	Europe	Germany		Insurance	L8N0649	Laptop	LaptopA9	7357	780	2017/3/11	2017/3/23	2017/3/19	Qian Qianqiu	Chen Chenjia	12524934	Male	41	15098824410
CSE0000000	Allianz	Europe	Germany		Insurance	P02C741	Paper Cutter	PaperCutter64	100	700	2017/3/11	2017/3/26	2017/4/23	Qian Qianqiu	Leonard Clar	12422761	Male	38	18904797951
CSE0000000	Allianz	Europe	Germany		Insurance	W1473B8	Water Heater	WaterHeater57	290	540	2017/3/11	2017/5/9	2017/4/5	Qian Qianqiu	Feng Zhenm	12137182	Male	45	18910019042
CSE0000000	Allianz	Europe	Germany		Insurance	D1G4290	Digitizer	Digitizer39	271	130	2017/3/11	2017/4/29	2017/3/24	Qian Qianqiu	Shui Nuobo	12335892	Male	39	13913540869
CSE0000000	Allianz	Europe	Germany		Insurance	EZ12538	Electric Fan	ElectricFanU8	980	870	2017/3/11	2017/4/28	2017/4/7	Qian Qianqiu	Qi Lvwei	12024317	Female	38	13983501780
CSE0000000	Allianz	Europe	Germany		Insurance	W1473B8	Water Heater	WaterHeater71	520	970	2017/3/11	2017/4/10	2017/4/19	Qian Qianqiu	Danny Smith	12135186	Male	39	15656054124
CSE0000000	Allianz	Europe	Germany		Insurance	V56301I	Video Capture Ca	VideoCaptureCa	470	360	2017/3/11	2017/4/27	2017/5/7	Qian Qianqiu	Xie Ercui	12135700	Female	46	18981092712
CSE0000000	Allianz	Europe	Germany		Insurance	S03485W	Sound Card	SoundCard93	10	860	2017/3/11	2017/3/26	2017/3/27	Qian Qianqiu	Joanne Taylo	12322095	Unknown	46	18929478204
CSE0000000	Allianz	Europe	Germany		Insurance	T7E2913	Toll Machine	TollMachineH9	880	610	2017/3/11	2017/4/20	2017/4/1	Qian Qianqiu	Shi Yuzhou	12336560	Male	43	15659504876
CSE0000000	Allianz	Europe	Germany		Insurance	CG63821	Computer Termin	ComputerTermi	360	680	2017/3/11	2017/4/11	2017/4/10	Qian Qianqiu	Zhang biqin	12212675	Female	43	13913466199
CSE0000000	Allianz	Europe	Germany		Insurance	C48967S	Charger	ChargerF5	31	760	2017/3/11	2017/3/15	2017/5/6	Qian Qianqiu	Lou White	12517344	Male	48	15622710919
CSE0000000	Allianz	Europe	Germany		Insurance	GC81302	Gas Water Heate	GasWaterHeate	540	800	2017/3/11	2017/4/9	2017/4/19	Qian Qianqiu	Qian Qianqiu	12225484	Female	43	13826275879
CSE0000000	Allianz	Europe	Germany		Insurance	B239N58	Building Intercom	BuildingIntercoi	710	700	2017/3/11	2017/3/16	2017/4/3	Qian Qianqiu	Adolf Davies	12422793	Male	38	15075997121
CSE0000000	Allianz	Europe	Germany		Insurance	C91D367	Camera	Camera95	386	880	2017/3/11	2017/3/26	2017/4/29	Qian Qianqiu	Zheng Biyan	12514710	Female	41	15031117856
CSE0000000	Allianz	Europe	Germany		Insurance	M6328D7	Mouse	Mouse94	62	310	2017/3/11	2017/4/23	2017/4/21	Qian Qianqiu	Hector White	12415692	Male	44	13932852743
CSE0000000	Allianz	Europe	Germany		Insurance	AE80564	Air Purifier	AirPurifierO6	990	120	2017/3/11	2017/4/25	2017/4/2	Qian Qianqiu	Dou Yishuanq	12331382	Unknown	42	18697569508
CSE0000000	Allianz	Europe	Germany		Insurance	R2I4560	Real Estate	RealEstateW0	70	750	2017/3/11	2017/3/25	2017/5/7	Qian Qianqiu	Hua Wenjing	12215245	Male	41	13841249541
CSE0000000	Allianz	Europe	Germany		Insurance	P9L0284	Program-Controll	Program-Contro	610	130	2017/3/11	2017/4/10	2017/4/12	Qian Qianqiu	Lindsay Davi	12433009	Female	44	18943286924
CSE0000000	Allianz	Europe	Germany		Insurance	M9J6453	Mp4	Mp437	223	590	2017/3/11	2017/3/19	2017/5/4	Qian Qianqiu	Rosalind Rob	12321373	Female	41	15066206015
CSE0000000	Allianz	Europe	Germany		Insurance	C02614Q	Cpu	Cpu38	940	720	2017/3/11	2017/3/20	2017/4/12	Qian Qianqiu	Danny Smith	12135186	Male	39	15656054124
CSE0000000	Allianz	Europe	Germany		Insurance	T0E9518	Tester	TesterO7	124	160	2017/3/11	2017/4/11	2017/5/1	Qian Qianqiu	Li Linglian	12520613	Female	49	13834907428
CSE0000001	HypoVereins	Europe	Germany		Bank	R54810N	Repeater	RepeaterB1	105	910	2020/9/3	2020/10/17	2020/10/5	Qian Qianqiu	Barbara Hall	12115041	Female	41	15001561594
CSE0000001	HypoVereins	Europe	Germany		Bank	FC40169	Function Server	FunctionServerI	720	350	2020/9/3	2020/10/21	2020/10/23	Qian Qianqiu	Shi Tingping	12515554	Female	39	15061819814

column name	discreption	Chinese discreption
Contract_number	Contract number, starting from CSE0000000	合同编号，从CSE0000000开始。
client_enterprise	Client company name	客户公司
country	Country where the company is located	公司所在的国家
city	City where the company is located (can be empty)	公司所在的城市，可以为空
industry	Industry sector of the company	公司所属的行业
product_code	Product code. A contract may contain multiple products	产品编码。一个合同中可以包含多个产品
product_name	Product name corresponding to the product code	产品编码对应的产品名称
Product_model	Product model. One product code may contain multiple product models	产品型号。一个产品编码对应多个产品型号
unit_price	Unit price corresponding to the product model	产品型号对应的单价
quantity	Quantity of ordered products	订单产品数量
order_date	Contract order date	合同下单日期
estimated_delivery_date	Estimated product delivery date	产品预计到达日期
lodgement_date	Actual arrival date (empty if later than 2025-3-24)	实际到达日期，如果晚于2025-3-24，则为空
director	Regional administrator where the company is located	公司所在区域的管理员
supply_center	Supply center in the company's region	公司所在区域的供应中心
salesman	Salesperson. Each product in a contract corresponds to one salesperson	销售员。合同中，一个产品对应一个销售员
Salesman number	Salesperson ID number	销售员编号
gender	Salesperson's gender	销售员性别
age	Salesperson's age	销售员年龄
mobile_number	Salesperson's phone number	销售员电话

Notes:

- the value of **city** will be **NULL** if the client enterprise is **not** in China;
- the value of **lodgement date** will be **NULL** if the date is later than 2025-3-24;

Requirement of the Project Report

Note: Some tasks consist of basic requirements and advanced requirements. You may not get full points if you only meet the basic ones.

Basic Information of Your Group

1. Names, student IDs, and the lab session of the group members
2. You are required to write down the contributions and the percentages of contributions for each group member. **Please clearly state which task(s)/part of the task(s) is/are done by which member in the group.**
    - **If you failed to link a task/part of a task to one of the group members, we will not count the score for the task** (since we don't know who accomplished this task; maybe it was done by an elf while you were sleeping at night?).

About one page

Task 1: E-R Diagram (30%)

Make an E-R Diagram of your database design with any diagram software. Hand-drawn results will not be accepted. Please follow the standard of E-R diagrams.

In the report, you are required to provide a snapshot of the E-R diagram. Also, please specify the name of the software/online service you use for drawing the diagram.

About one page

Task 2: Database Design (30%)

Design the tables and columns based on the background provided above. Generate the E-R diagram via the "Show Visualization" feature if you are using DataGrip (if you plan to use other GUI database clients, please find the corresponding way by yourself to generate the E-R diagram according to the client you use.). Briefly describe the design of the tables and columns including (but not limited to) the meanings of tables and columns

In the report, you are required to provide the following content:

1. Attach the snapshot of the E-R diagram generated by DataGrip.
2. Briefly describe the table designs and the meanings of each table and column.

In addition, please submit an **SQL file** as an attachment that contains the DDLs (`create table` statements) for all the tables you created. **Please make it into a separate file but not copy and paste the statements into the report.**

About one or two pages

Notes for the database design:

Following requirements you must be accomplished.

1. All data items should base on csv file.
2. Your design needs to follow the requirements of the three normal forms.
3. Every row in each table should be uniquely identified by its primary key. (You may use a simple or a composite primary key).
4. Every table should be involved in a foreign key. No isolated table is allowed. (每个表要有外键，或者有其他表的外键指向。)
5. Your design should contain no circular foreign-key links. (对于表之间的外键方向，不能有环。例如：A表有外键关联B表，B表有外健关联C表，C表有外健关联A表)
6. You should use appropriate data types for different fields.

Try to complete the following requirements

1. Each table should contain at least one mandatory (“Not Null”) column (including the primary key but not the id column).
2. Other than the system-generated self-increment ID column, there should be at least one column with the "unique" constraint. (除了主键自增的id之外，需要有其他unique约束的列)
3. Your design should be easy to expand when requirements change.

Task 3: Data Import (40%)

In this task, you should write scripts to import the content in csv into the database you have designed before. After importing the data, you should also make sure all data is successfully imported.

Task 3.1 Basic Requirements: 10%

Introduce the scripts that being used to import data. You can list a similar table below to describe the functions of all the scripts you submit in the attachment.

Script name	Author	Description
Script1.py	xxx	Run this script to complete the import of passenger data.

A description of how you use the script to import csv file. In this part, You should clearly state the steps.

- You can create an intermediate file based on the original csv file, and then import the intermediate file through a script.
- You can also directly import csv file in a programming language and import it line by line.
- If you manually modified the file, please also explain the specific modifications.

About one or two pages

Task 3.2 Data Accuracy checking: 15%

According to the data in csv file, we will **give several questions** in presentation week to check whether all data have been correctly imported into your database. Please prepare SQL queries for the following requirements **in the report**, and you will run the queries by **datagrip** in this part **during your presentation week**.

10% for accuracy

5% for Whether to prepare SQL statements before and complete the process of accuracy checking on time

1. How many salesmans are there for each gender?
2. How many companies are there in each supply center?
3. How many salesmans are there in each supply center?
4. How many salesmans are there within a given age range (lower and upper bounds)?
5. How many countries are there in each supply center?
6. Given a country name, list all company names and their respective industries.
7. Given a product code, list all product models and their corresponding unit prices.
8. Given a contract number, list all order details in the contract, including (product model, order quantity, salesmans name, and lodgement_date).

1. 不同性别的销售员分别有多少个。
2. 每一个supply center中有多少公司。
3. 每一个supply center中有多少个销售员。
4. 年龄在某个上界与下界区间的销售员有多少个。
5. 每一个suppy center中有多少个国家。
6. 根据某一个国家名称，列出其所有的公司名称，及其公司所属行业。
7. 根据某一个产品编码，列出其所有的产品型号、及对应的单价。
8. 根据某一个合同编码，列出合同中所有订单内容，每个订单包含（产品型号、购买数量、销售员名称、实际到达日期）

About one page

Task 3.3 Advanced requirements: 15%

You may also need to finish the following advanced requirement to get the remaining points.

1. Try to **optimize your script**, and find **more than one ways** to import data, and provide a comparative analysis of the computational **efficiencies** between these ways.
2. Try to import data across multiple systems (e.g., Windows, MacOS, Linux).
3. Try to import data using various programming languages (e.g., Java, Python, C++).
4. Experiment with other databases; we recommend use **OpenGauss**.
5. Try to import data with different data volumes.

For the advanced points, please make sure to describe your test environment, procedures, and actual time costs. It is required to write a paragraph or two to analyze the experiment results.

About 0 to 4 pages

How to Submit Your Report

Submit the report in PDF format with necessary attachments (such as SQL scripts and source code files) on the bb website before **23:30 on April 28th, 2025, Beijing Time (UTC+8)**. For attachments, please put them into separate directories based on the task, and compress them into a `.zip` archive.

Disclaimer

The names, characters, businesses, and events in the background of this project are purely fictional. The items in the files are randomly-generated fake data. Any resemblance to actual events, entities or persons is entirely coincidental and should not be interpreted as views or implications of the teaching group of CS307.