****

**INDIVIDUAL ASSIGNMENT 1**

**Shuya Chen**

**CIS467: DATA MANAGEMENT, WAREHOUSING, AND VISUALIZATION**

**Professor: Mikhail Lysyakov**

**Spring A 2025**

**HW3 due by Monday Feb. 24 by 11:59 PM**

The create\_expense.sql script (you can find it on Blackboard in the Homework 3 folder) creates a database which contains the 7 tables described below. The data contains a tracking system for expense reports filed by employees at a manufacturing company. **Please watch the Panopto “video 5 for HW3” before doing this Homework 3.**

**If you use Chat GPT, please use the “Share” button (looks like ‘upward arrow’) in the right corner of ChatGPT chat, and ‘copy link’ and share the link to that chat in this Word document and briefly explain how you used it for your Homework (for each HW question if you used it). No points off will be taken for using ChatGPT (it is allowed to use it for Homework) but you are required to share the link to a chat if you used it.**

**More information on how to share a chat here:** <https://help.openai.com/en/articles/7925741-chatgpt-shared-links-faq>

|  |  |
| --- | --- |
| **employees** | **Field Description** |
| Ssn (pk) | Unique SSN ID# for employee |
| First\_name | Employee first name |
| Last\_name | Employee last\_name |
| Dept (fk) | Dept ID# |
| Start\_year | Year of employment |

|  |  |  |
| --- | --- | --- |
| **trips** | | **Field Description** |
| Employee (pk, fk) | | SSN of employee travelling |
| Trip\_ID (pk) | | Unique Trip ID# |
| Start\_date | | Start date of trip |
| End\_date | | End date of trip |
| Reason\_code (fk) | | Code for reason for trip |
|  | |  |
| **expenses** | **Field Description** | | |
| Employee (pk, fk)  Trip\_id (pk, fk) | SSN of employee travelling  Unique Trip ID# | | |
| Expense\_seq (pk) | Sequence# for expense report line item | | |
| Account\_no (fk) | Account number for line item | | |
| Gross\_amount | Gross dollar amount of line item | | |
| tax | Sales tax (if applicable) of line item | | |
|  |  | | |

|  |  |
| --- | --- |
| **dept\_codes** | **Field Description** |
| Dept\_ID (pk) | Dept ID# |
| Dept\_name | Name of department |

|  |  |  |
| --- | --- | --- |
| **Reason\_codes** | **Field Description** | |
| Reason\_code (pk) | Reason ID# | |
| Reason\_description | Description of reason for trip | |
|  |  | |
| **account\_codes** | **Field Description** |
| Account\_no (pk) | Account ID# |
| Account\_description | Description of account |
| Account\_type | Category of account |
| **reimbursements** | **Field Description** |
| Employee (pk, fk) | SSN of employee travelling |
| Trip\_id (pk, fk) | Unique Trip ID# |
| Auditor | Auditor last name |
| Reimbursement\_amount | Amount of reimbursement |
| Reimbursement\_date | Date of reimbursement |

Diagram

Description automatically generated

Please put all of your work into **this** **single Word doc**.

1. (60 points) Design and create a data warehouse for the Expense database. The decisions about which fields to include and how to aggregate the data are left to you. You do not need to include every single data point from the 7 tables given. Use your judgement as to what will be interesting/useful for the organization. But please make sure that you pull (combine) data from **at least four tables** and compute relevant aggregate statistics. Please see many examples from class lectures and you may adapt those codes for your purpose (for this dataset).

**Submit a screenshot of the first 25 rows of your data warehouse (paste into this Word document) and the SQL code that you used to create it. Please copy and paste your SQL code into this Word document.**

CREATE TABLE expense\_summary\_dw AS

SELECT

e.ssn AS employee\_id,

e.first\_name,

e.last\_name,

d.dept\_name AS department,

YEAR(STR\_TO\_DATE(t.start\_date, '%m/%d/%Y')) AS year, -- Convert date format

COUNT(DISTINCT t.trip\_id) AS total\_trips,

SUM(ex.gross\_amount) AS total\_expense\_amount,

SUM(ex.tax) AS total\_tax,

COALESCE(SUM(r.reimbursement\_amount), 0) AS total\_reimbursement,

(SUM(ex.gross\_amount) - COALESCE(SUM(r.reimbursement\_amount), 0)) AS net\_expense,

(SUM(ex.gross\_amount) / NULLIF(COUNT(DISTINCT t.trip\_id), 0)) AS avg\_expense\_per\_trip

FROM employees e

JOIN dept\_codes d ON e.dept = d.dept\_id

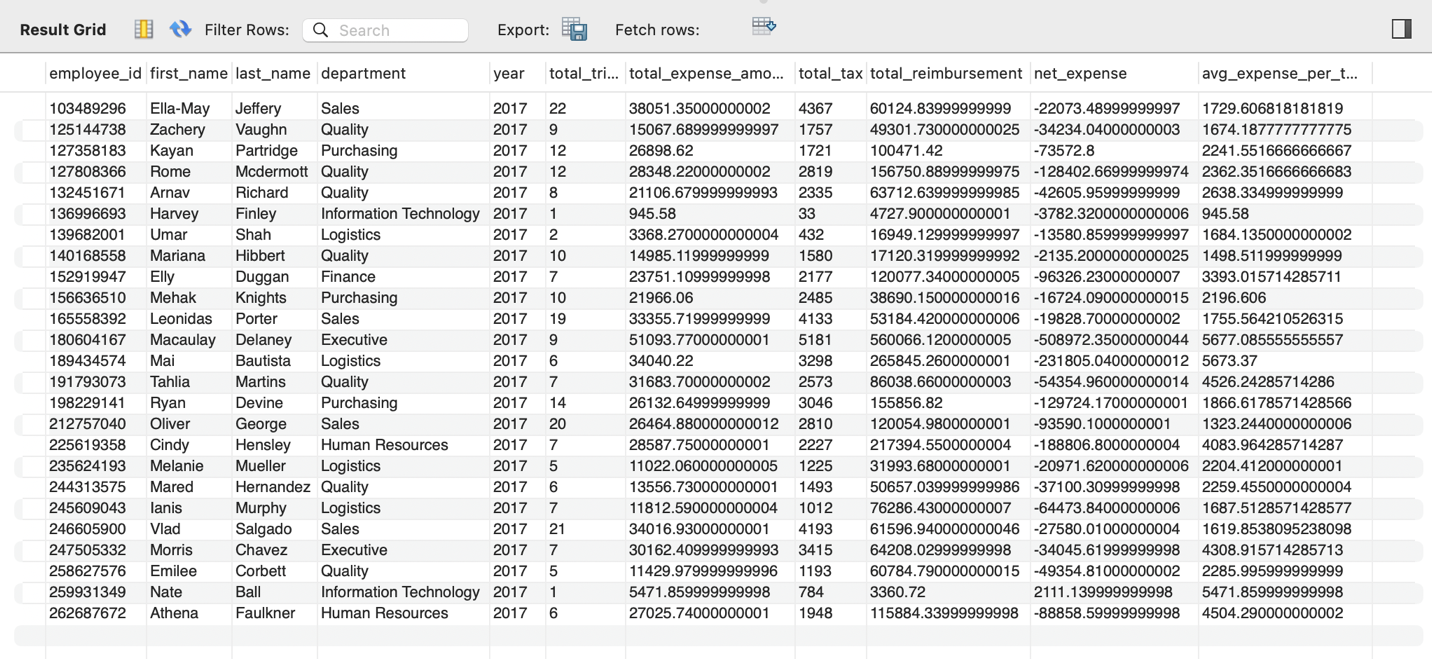
JOIN trips t ON e.ssn = t.employee

LEFT JOIN expenses ex ON t.trip\_id = ex.trip\_id

LEFT JOIN reimbursements r ON t.trip\_id = r.trip\_id

GROUP BY e.ssn, e.first\_name, e.last\_name, d.dept\_name, YEAR(STR\_TO\_DATE(t.start\_date, '%m/%d/%Y'));

SELECT \* FROM expense\_summary\_dw LIMIT 25;



2. (40 points) Create **four** SQL queries on your data warehouse that answer interesting important questions. At least two queries should be more than simple queries. For example, more complex queries could include Joins, a Group By element or a subquery or use some aggregate functions and summary calculations (see examples in the class lectures’ slides).

**Submit a copy of each query SQL code (paste into this Word document), and the screenshot of each query results (or the first 25 rows if it is longer) and a one or two sentence description of the question your SQL code was addressing and what you found in the results.**

**Query 1: Top 5 Employees with the Highest Total Expenses**

SELECT employee\_id, first\_name, last\_name, department, total\_expense\_amount

FROM expense\_summary\_dw

ORDER BY total\_expense\_amount DESC

LIMIT 5;

表格

AI 生成的内容可能不正确。

**Query 2: Yearly Expense Trends per Department**

SELECT year, department, SUM(total\_expense\_amount) AS department\_expense

FROM expense\_summary\_dw

GROUP BY year, department

ORDER BY year DESC, department\_expense DESC;

表格

AI 生成的内容可能不正确。

**Query 3: Employees with Higher Expenses Than Reimbursements**

SELECT employee\_id, first\_name, last\_name, department,

total\_expense\_amount, total\_reimbursement,

(total\_expense\_amount - total\_reimbursement) AS unreimbursed\_expense

FROM expense\_summary\_dw

WHERE total\_expense\_amount > total\_reimbursement

ORDER BY unreimbursed\_expense DESC;

表格

AI 生成的内容可能不正确。

**Query 4: Average Expense Per Trip by Department**

SELECT department,

COUNT(DISTINCT employee\_id) AS total\_employees,

COUNT(total\_trips) AS total\_trips,

SUM(total\_expense\_amount) AS total\_department\_expense,

AVG(avg\_expense\_per\_trip) AS avg\_expense\_per\_trip

FROM expense\_summary\_dw

GROUP BY department

ORDER BY avg\_expense\_per\_trip DESC;

表格

AI 生成的内容可能不正确。

**Summary of Queries**

1. **Top spending employees:** Identify employees with the highest expenses to optimize cost management.
2. **Yearly department expense trends:** Track spending changes across departments to improve budgeting.
3. **Employees with unreimbursed expenses:** Detect cases of insufficient reimbursements and refine policies.
4. **Average travel expenses by department:** Compare trip costs across departments to enhance expense control.

These analyses help businesses manage expenses more effectively and improve financial efficiency.