**HW3 due by Wednesday Oct. 12 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 for week 5 before doing this Homework 3.**

|  |  |
| --- | --- |
| **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.**

**Table**: employees, trips, dept\_codes, expenses

USE expense;

SELECT first\_name, last\_name, dept\_name, COUNT(t.employee) AS total\_number\_trips,

SUM(gross\_amount + tax) AS total\_expenses

FROM employees e JOIN dept\_codes d

ON e.dept = d.dept\_id JOIN trips t

ON e.ssn = t.employee JOIN expenses ex

ON t.trip\_id = ex.trip\_id AND t.employee = ex.employee

GROUP BY e.ssn

ORDER BY total\_number\_trips DESC

LIMIT 25;

Table

Description automatically generated

2. (40 points) Create **four** SQL queries on your data warehouse that answer interesting 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.**

**1)**

-- find which department has the most employees

SELECT dept\_name, COUNT(ssn) AS total\_employees

FROM employees e JOIN dept\_codes d

ON e.dept = d.dept\_id

GROUP BY dept\_name

ORDER BY total\_employees DESC;

-- Sales department has the most employees with a number of 19 people.

Graphical user interface, application

Description automatically generated

**2)**

-- show the total expenses of each trip

SELECT e.trip\_id, SUM(gross\_amount + tax) AS total\_expense

FROM expenses e JOIN trips t

ON e.trip\_id = t.trip\_id AND e.employee = t.employee

GROUP BY e.trip\_id

ORDER BY e.trip\_id

LIMIT 25;

-- I calculated the total expenses of each trip and the result is shown below. And I found that the trip id 4604 has the least total expense.

Table

Description automatically generated with low confidence

**3)**

-- display employees name and who traveled in 2017

SELECT first\_name, last\_name

FROM employees e JOIN trips t

ON e.ssn = t.employee

WHERE YEAR(STR\_TO\_DATE(start\_date,"%m/%d/%Y")) = 2017

GROUP BY ssn

LIMIT 25;

-- There are totally 99 employees traveled in 2017.

Graphical user interface, table

Description automatically generated with medium confidence

**4)**

-- display employees who do not have reimbursement

SELECT DISTINCT(first\_name), last\_name

FROM employees e JOIN reimbursements r

ON e.ssn = r.employee

WHERE ssn IN (SELECT employee

FROM reimbursements r

GROUP BY employee

HAVING SUM(reimbursement\_amount) = 0);

-- There are totally 7 employees who do not have reimbursement.

Graphical user interface, application, table

Description automatically generated