IS490DB and IS490DB2

Assignment 4

Working with the Structured Query Language

Advice: The best way to approach this assignment is to re-read chapters 6 and 8 and review the slides for DML and hierarchical queries. You can complete part 1 after the DML lectures, but you will probably want to wait for the hierarchical queries lecture before starting Part 2.

Part 1) Practice standard SQL

Consider the COMPANY relational schema that we have been using throughout the course. We have written SQL statements to create and populate the COMPANY database and the TA has posted details of the tables in moodle. Write SQL statements that will provide an answer to the following questions. You don't need to show the results of the queries, just submit your SQL statements.
Feel free to copy the tables to your tablespace and add more records so test that your queries work as expected. Lastly, you must demonstrate that you understand how to write both explicit and implicit joins by including at least one example of each in your solution, i.e. use explicit joins to answer at least 1 query and implicit joins to answer at least 1 query.

a) Retrieve the names of all employees in Department 5 who work more than 10 hours per week on the 'Product X' project.

b) List the names of all employees who have a dependent with the same first name as themselves.

c) Find the names of all employees who are directly supervised by 'Franklin Wong'.

d) For each project, list the project name and the total hours per week (by all employees) spent on that project.

e) For each department, retrieve the department name and the average salary of all employees working in that department.

f) Retrieve the name of employees who work on EVERY project.

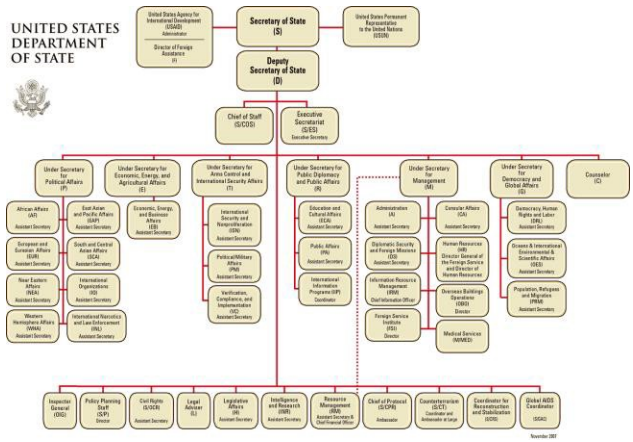
g) Retrieve the average salary of all female employees

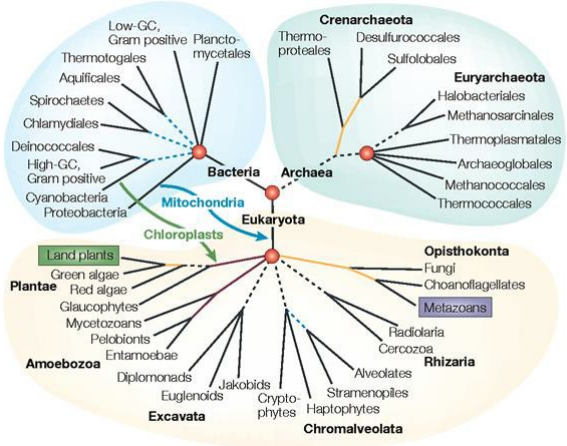
h) Find the names and addresses of all employees who work on at least one project located in Houston, but whose department has no location in Houston.

i) List the last names of all department managers who have no dependents.

Part 2) Working with Hierarchical Data

Motivation: Hierarchies are a common way of organizing data, from human resource management to scientific taxonomies:





Support for hierarchical queries was added to the SQL 3 standard in 1999 but implementation in off-the-shelf relational database management systems (RDBMS) has lagged. Oracle's support for hierarchical queries predates the SQL 3 by almost 20 years and has since developed into a rich, though non-standard, syntax for handling hierarchical data. The Oracle reference manual for hierarchical query support begins at Hierarchical Queries. It describes the CONNECT BY clause and all of its supporting pseudo-columns (e.g., LEVEL, CONNECT_BY_ISLEAF), operators (e.g., PRIOR, START WITH, NOCYCLE, ORDER SIBLINGS BY), and functions (e.g., SYS_CONNECT_BY_PATH). Several of these syntactic features will be needed to complete this assignment. The CONNECT BY clause alone is insufficient.

Part 2 Instructions:

The goal of this assignment is to find the longest path in a branch of the Medical Subject Headings (MeSH) hierarchy. MeSH was developed by the National Library of Medicine to facilitate indexing of PubMed articles. A branch of the MeSH hierarchy containing anatomical structures and abnormalities has been extracted from the Unified Medical Language System (UMLS) and placed in the LIS490DB_UMLS_MESH_RELS table. This table contains the parent-child relations of each MeSH concept, e.g.:

PAR	CHD
C0206187	C0009437
C0009437	C0042425
C0005400	C0206187

The description of each MeSH is in the LIS490DB_UMLS_MESH table, e.g.:

CUI	MSH	CONCEPT
C0042425	D014670	AMPULLA VATER
C0005400	D001652	Bile Ducts
C0206187	D017734	Bile Duct, Extrahepatic
C0009437	D003135	Bile Duct, Common

You can browse the MeSH hierarchy online using the unique identifiers in the MSH field of this table.

The deliverable is a brief report describing how you completed this assignment. It must contain the queries that you used to find paths in the hierarchy, the longest path(s) that you found, and the MeSH strings for this path. For example, the tables above show an actual path in the hierarchy. If this were the longest path, your report would contain the following table:

Level	Parent CUI	Parent Concept	Child CUI	Child Concept
1	C0005400	Bile Ducts	C0206187	Bile Duct, Extrahepatic
2	C0206187	Bile Duct, Extrahepatic	C0009437	Bile Duct, Common
3	C0009437	Bile Duct, Common	C0042425	AMPULLA VATER

Hint: The Oracle LEVEL pseudo-column and SYS_CONNECT_BY_PATH function are useful in this assignment.

Your report must also answer the following questions:

- Does LIS490DB_UMLS_MESH_RELS contain a strict hierarchy (a child can only have one parent) or do some concepts have more than one parent? For example, the U.S. State Department shown above is not a strict hierarchy (see the dotted line).
- How did you determine the answer to this question?

Hint: The Oracle NOCYCLE operator can help you answer these questions.

Useful URLs:

- Oracle documentation for hierarchical queries: http://docs.oracle.com/cd/B19306_01/server.102/b14200/queries003.htm
- U.S. National Library of Medicine
  - Medical Subject Headings: https://www.nlm.nih.gov/mesh/
  - MeSH browser: https://www.nlm.nih.gov/mesh/2015/mesh_browser/MBrowser.html
  - Celko, J. (2005). *SQL for Smarties: Advanced SQL Programming* (3rd Ed.), Chapter 28: Trees and Hierarchies in SQL, Morgan Kaufman, San Francisco, CA. (E-book available online through the UIUC library.)

Instructions: Upload your SQL statements to answer each of the queries in part 1 and your report from part 2 into moodle. Don't forget to grant the TA and instructor access to your tables.

Last updated 23 October, 2017