

Amreeta Choudhury

Lecture Notes 06/13/2017 Module 4

Class: 8:30-10:00 pm EST

Article:

SQL Article: <http://www.skilledup.com/articles/learn-sql-it-most-in-demand-skill-in-single-day>

Announcements:

Academic Integrity

All work is expected to be original either from team effort or individual design; cheating and plagiarism will not be tolerated. Columbia University expects that its students will act with honesty and propriety at all times and will respect the rights of others. It is fundamental University policy that academic dishonesty in any guise or personal conduct of any sort that disrupts the life of the University or denigrates or endangers members of the University community is unacceptable and will be dealt with severely. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources and assistance upon which that work is based.

Next Requirements Discussion forum due **June 17th 11:59 pm EST**

Final Draft due **June 18th 11:59 pm EST**

Session Objectives:

- Relational DB overview and evolution
- Basic Structure
- On Line Analytical Processing (OLAP)
- Extraction, Transformation, Load
- Data Model
- Achieving Scale
- Achieving Speed
- SQL - Structured Query Language
- ACID – Atomicity, Consistency, Isolation, Durability
- SQL and NoSQL
- Presentation Logistics

Discussion:

Discussed history and influential players for the Database

Codd's Amendments

Foundation rule

Information rule

Guaranteed Access Rule

System treatment of null values rule

Dynamic online catalog based on the relational model

Comprehensive Data Sub language rule

View Updating Rule

High level insert, update and delete rule

Physical data independence rule

Logical data independence rule

Integrity independence rule

Distribution independence rule

Non-subversion rule

Resources with SQL Commands, Examples and Rules:

https://www.tutorialspoint.com/dbms/dbms_codds_rules.htm

<https://www.w3schools.com/sql/>

<https://community.modeanalytics.com/sql/tutorial/introduction-to-sql/> (This is a nice one)

https://www.tutorialspoint.com/dbms/relational_algebra.htm

<https://www.databasemanagement.com>

Concepts:

Separate the data from the application

Predictable and standard structure

Tables, rows and columns

Structured Query Language

Crash Recovery

Concurrency

Class Exercise

With your group, please work on this exercise. You may use outside resources to help such as <http://www.w3schools.com/sql>

Table A

Sid	Students	Major	City
1	John	Statistics	Beijing
2	Joe	Engineering	New York City
3	Jim	Art	New York City
4	Amy	Mathematics	New Delhi
5	Teresa	Statistics	Rome

Table B

lid	Instructors	City	Background	Sid
1	Amreeta	New York City	Statistics	1
2	Mark	Waltham	Computer Science	2
3	Jim	Waltham	Management	10
4	Jean	New York City	Mathematics	12

Write out the code that does the following:

1. Select all students from New York City from the first table.
2. Select only the distinct values from the city column from the first table.
3. Rename Sid and select all students from the first table.
4. Select all of the different cities (only distinct values) from the first table and the second table ordered by City

(Answers)

1.

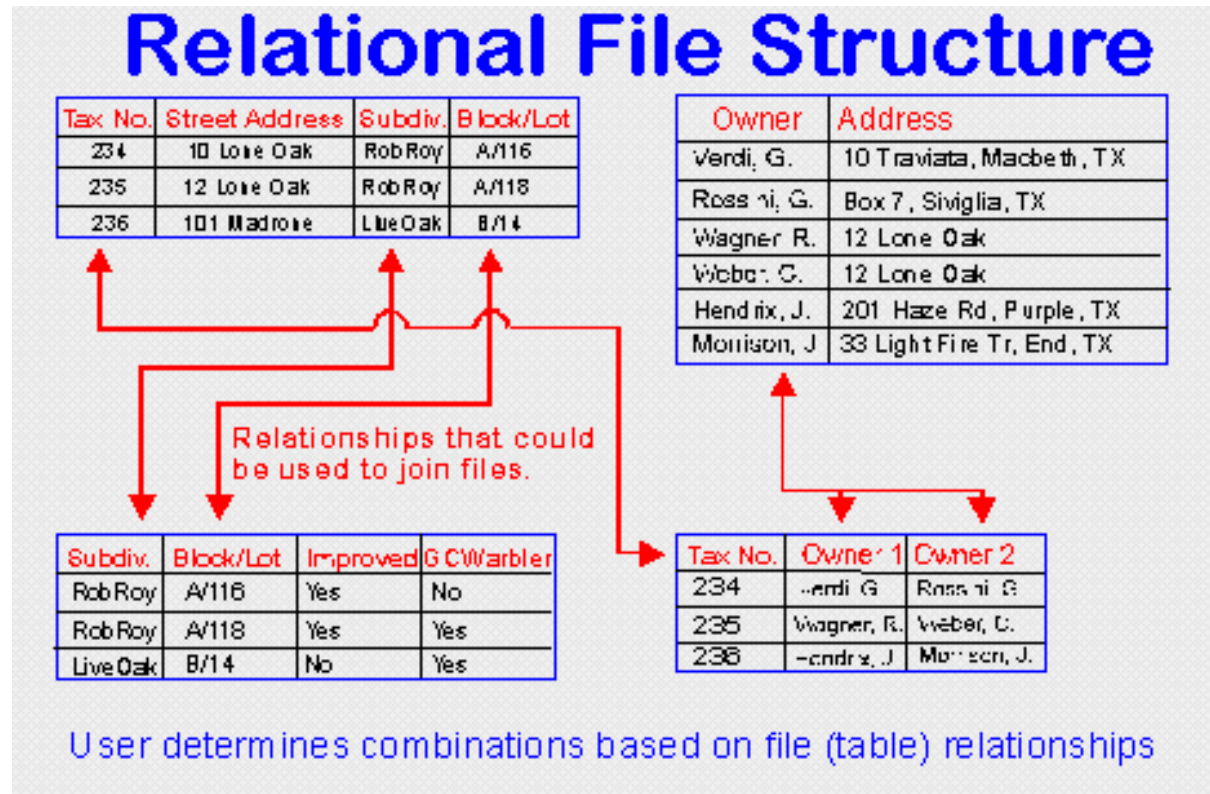
```
SELECT Students
FROM A
WHERE City = 'New York City';
```
2.

```
SELECT DISTINCT City
FROM A;
```
3.

```
SELECT Sid AS 'Student ID'
FROM A;
```
4.

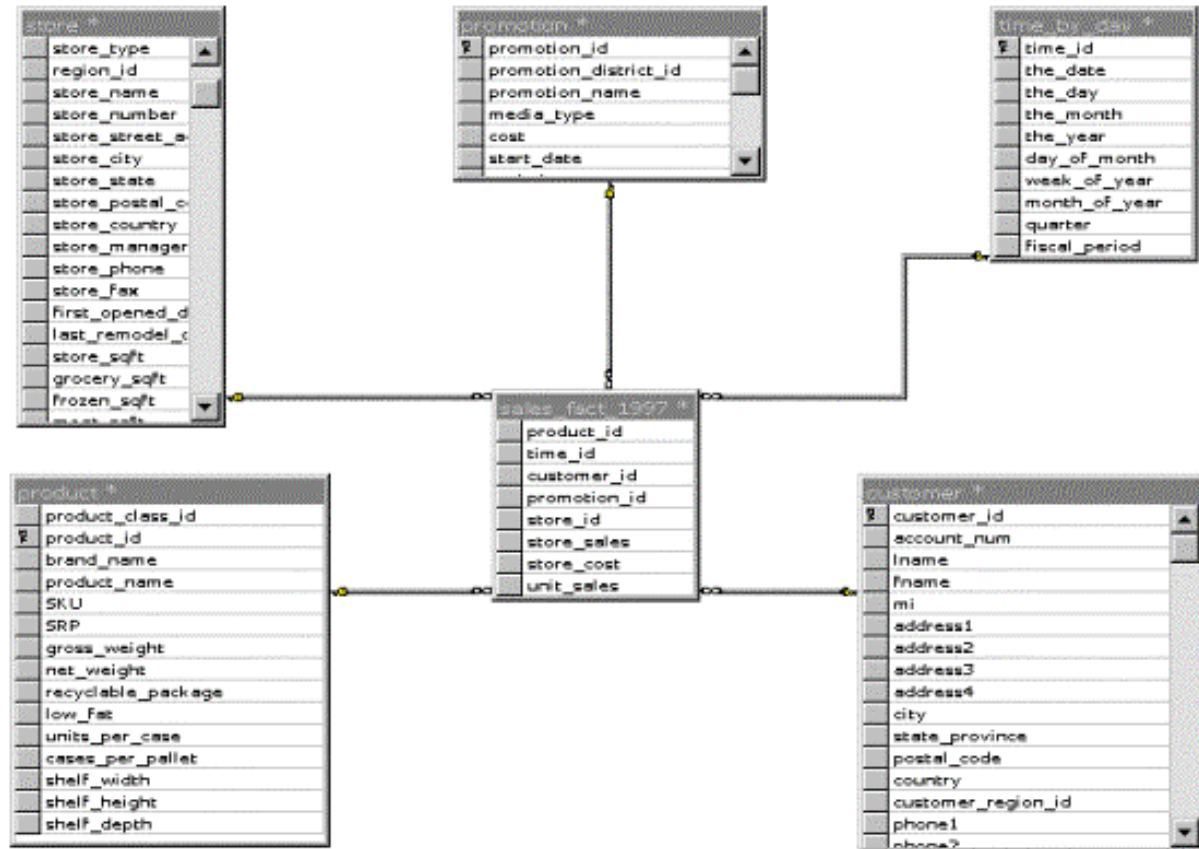
```
Select City from A
INTERSECT
Select City from B
ORDER BY City;
```

Structures of Database:



Resource: Univ of Colo Boulder

<http://www.colorado.edu/geography/gcraft/notes/datacon/gif/relafile.gif>



Resource: Microsoft

Codd's Rules for OLAP: Online Analytical Processing

<http://olap.com/learn-bi-olap/codds-paper/>

ETL Process:

Mapping ; Source -> target

Staging Area

Cleansing

Transformation

Transport

Scale achieved through parallel processing and partitioning

Speed is achieved by:

- Partitioning
- Query optimization
- Indexing
- In memory
- Appliance approach

Interesting article: <http://bradhedlund.com/2012/08/27/the-appliance-approach-to-big-data-and-private-cloud/>

ØSELECT DISTINCT column_name(s) FROM table_name(s) -- The DISTINCT label in this statement means that a single value of the attribute will be returned upon a query. (A great statement to execute if you would like to see a school name display once in your query rather than the twenty-five times it is recorded in your database.

DELETE FROM – The DELETED statement is most commonly used to delete entire tables or data held within tables.

ØDELETE FROM table_name (NOTE: Deletes the entire table!!!) or DELETE FROM table_name WHERE condition -- This statement deletes either an entire table or, if specified by the WHERE conditional clause, deletes the records (also referred to as rows or tuples) where a condition is met.

TRUNCATE TABLE (Only deletes the data inside of the table.) - The TRUNCATE TABLE statement is most commonly used to delete the information contained within a table.

ØTRUNCATE TABLE table_name -- This statement will delete/clear the information contained within a table.

UPDATE - The UPDATE statement is most commonly used to update a particular cell within a table with new/updated data.

ØUPDATE table_name SET column_name =new_value WHERE column_name=some_value -- This statement will update the data value in a given cell for a particular table given the existence of a current data value in the cell.

(Source: <http://www.databasemanagement.com>)

Other Concepts:

Stored procedures

ACID properties

Data lake vs Data warehouse

SQL vs NoSQL: http://www.thegeekstuff.com/2014/01/sql-vs-nosql-db/?utm_source=tuicool