LAB: String Patterns, Sorting & Grouping

The practice problems for this Lab will provide hands on experience with string patterns, sorting result sets and grouping result sets. You will also learn how to run SQL scripts to create several tables at once, as well as how to load data into tables from .csv files.

HR Database

We will be working on a sample HR database for this Lab. This HR database schema consists of 5 tables called EMPLOYEES, JOB_HISTORY, JOBS, DEPARTMENTS and LOCATIONS. Each table has a few rows of sample data The following diagram shows the tables for the HR database.

SAMPLE HR DATABASE TABLES

EMPLOYEES

Emp_id	F_name	L_name	SSN	B_date	Sex	Address	Job_id	Salary	Manager_id	Dep_id
E1001	John	Thomas	123456	1976-01-09	M	5631 Rice, OakPark,IL	100	100000	30001	2
E1002	Alice	James	123457	1972-07-31	F	980 Berry ln, Elgin,IL	200	80000	30002	5
E1003	Steve	Wells	123458	1980-08-10	M	291 Springs, Gary,IL	300	50000	30002	5

JOB_HISTORY

Emp_id	Start_date	Job_id	Dep_id
E1001	2000-01-30	100	2
E1002	2010-08-16	200	5
E1003	2016-08-10	300	5

JOBS

Job_id	Job_title	Min_salary	Max_salary
100	Sr. Architect	60000	100000
200	Sr.SoftwareDeveloper	60000	80000
300	Jr.SoftwareDeveloper	40000	60000

DEPARTMENTS

Dept_id	Dep_name	Manager_id	Loc_id
2	Architect Group	30001	L0001
5	Software Development	30002	L0002
7	Design Team	30003	L0003
5	Software	30004	L0004

LOCATIONS

Loc_id	Dep_id
L0001	2
L0002	5
L0003	7

To complete this lab you will utilize Db2 database service on IBM Cloud as you did for the previous lab. There are three parts to this lab:

I. Creating tables
II. Loading data into tables
III. Composing and running queries

Copyright ©2018 IBM Corporation

Part I: CREATING TABLES

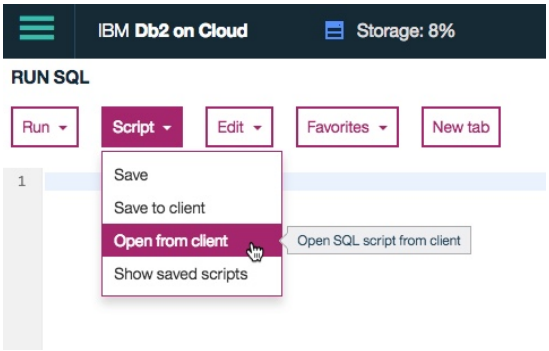
If you do not yet have access to Db2 on IBM Cloud, please refer to Lab Instructions in the Module/Week 1.

Rather than create each table manually by typing the DDL commands in the SQL editor, you will execute a script containing the create table commands for all the tables. Following step by step instructions are provided to perform this:

1) Download the script file " Script_Create_Tables.sql" provided on the Lab page

2) Login to IBM Cloud and go to the Dashboard: https://console.bluemix.net/dashboard/apps where you can find the Db2 service that you created in a previous Lab. Click on the service. Next, open the Db2 Console by clicking on "Open Console" button. Go to the Run SQL page. The Run SQL tool enables you to run DDL and SQL statements.

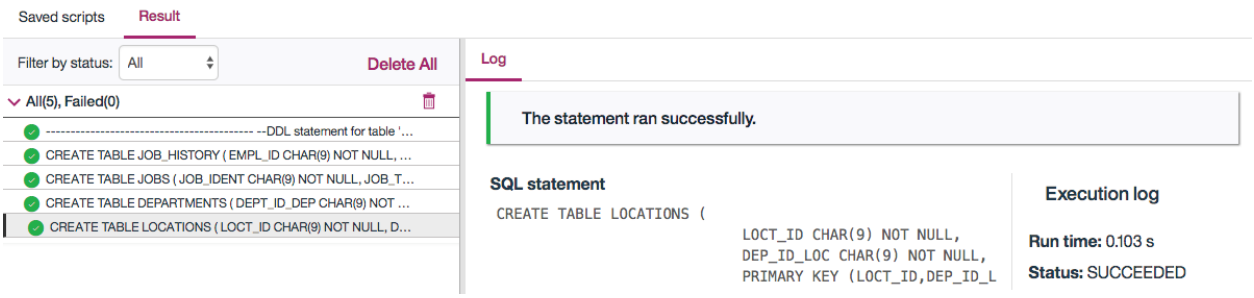
3) Click on the Folder icon to "Open script". Locate the file Script_Create_Tables.sql that you downloaded to your computer earlier and open it.



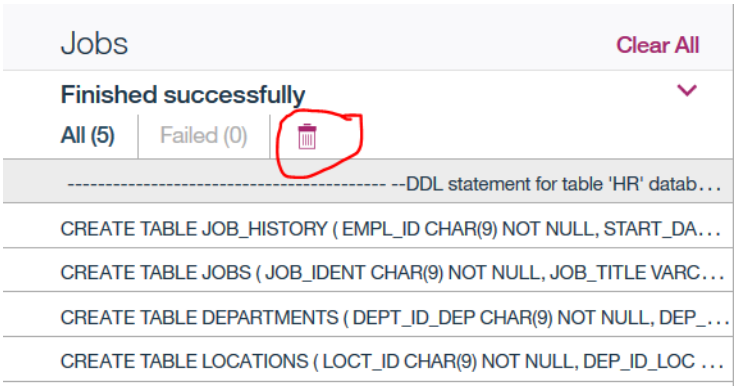
4) Once the statements are in the SQL Editor tool , you can run the queries against the database by selecting the option "Run" then "Run All".

Copyright ©2018 IBM Corporation

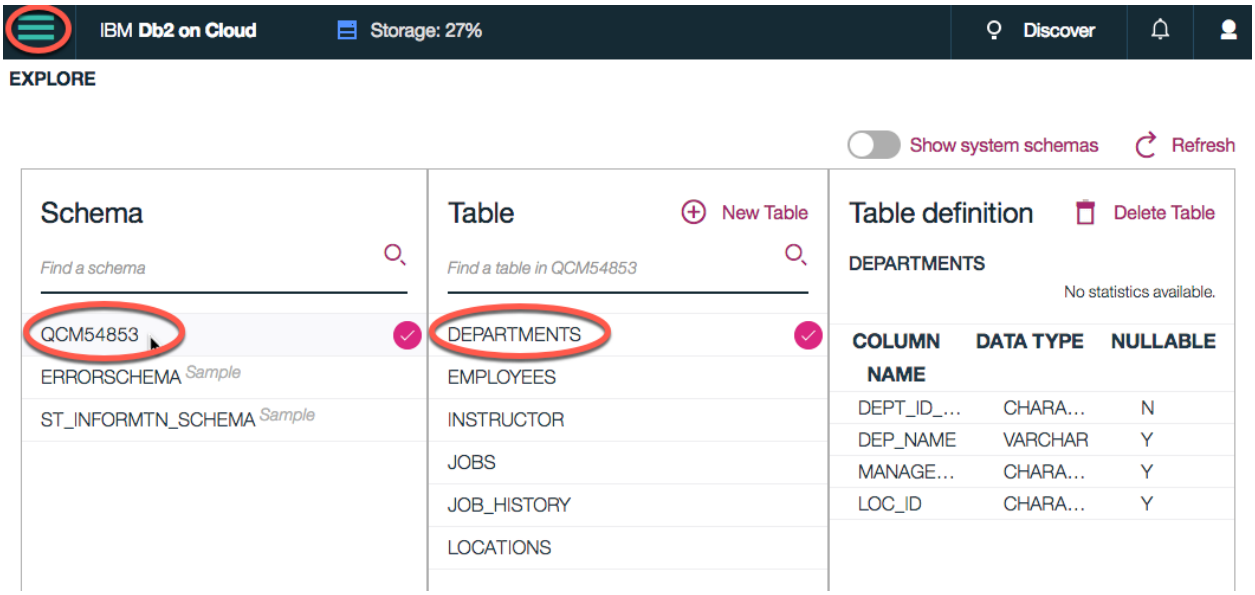
5) At the bottom of the screen you will see a Result section. Clicking on a query will show the execution details of the job - whether it ran successfully, or had any errors or warnings. Ensure your queries ran successfully and created all the tables.



6) You can delete the Jobs history by clicking on the Delete icon as shown below.



7) Now you can look at the tables you created. Navigate to the three bar menu icon and select "Explore". Select the correct Schema to see the newly created tables.



Copyright ©2018 IBM Corporation

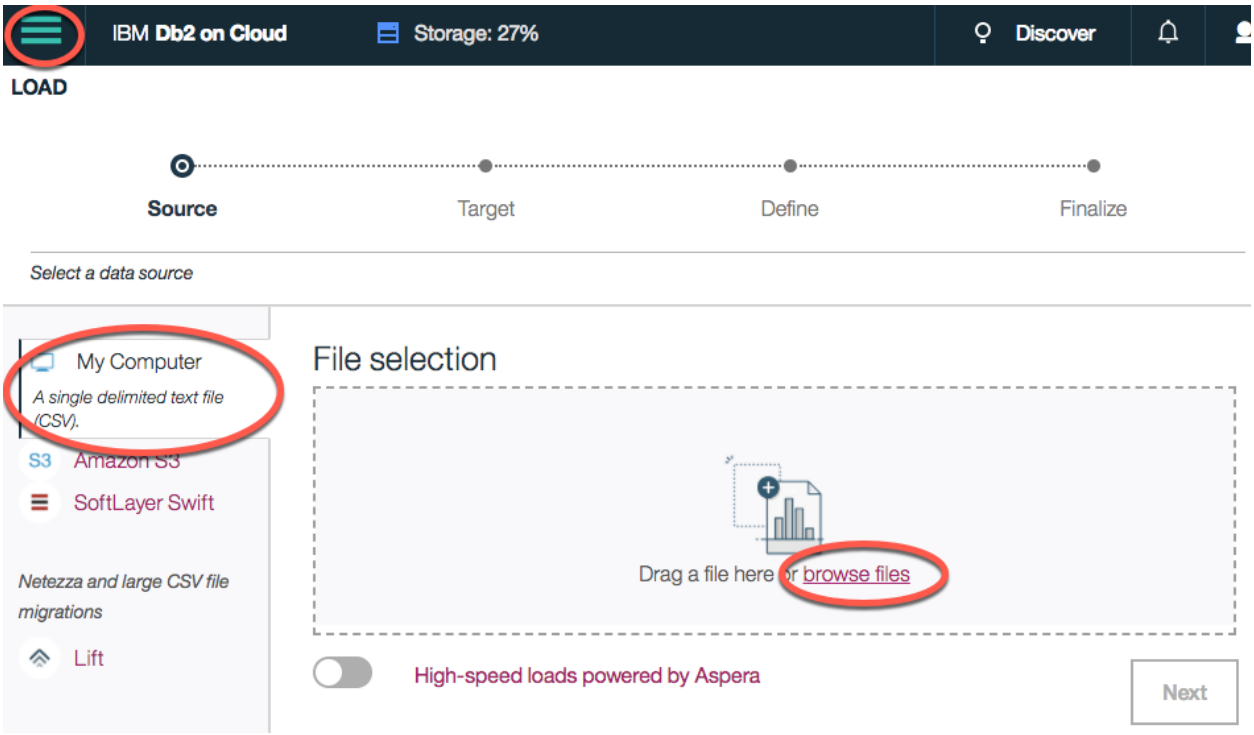
Part II: LOADING DATA

Now let us see how data can be loaded into Db2. We could manually insert each row into the table one by one but that would take a long time. Instead, Db2 (and almost every other database) allows you to Load data from csv files.

Please follow the steps below which explains the process of loading data into the tables we created earlier.

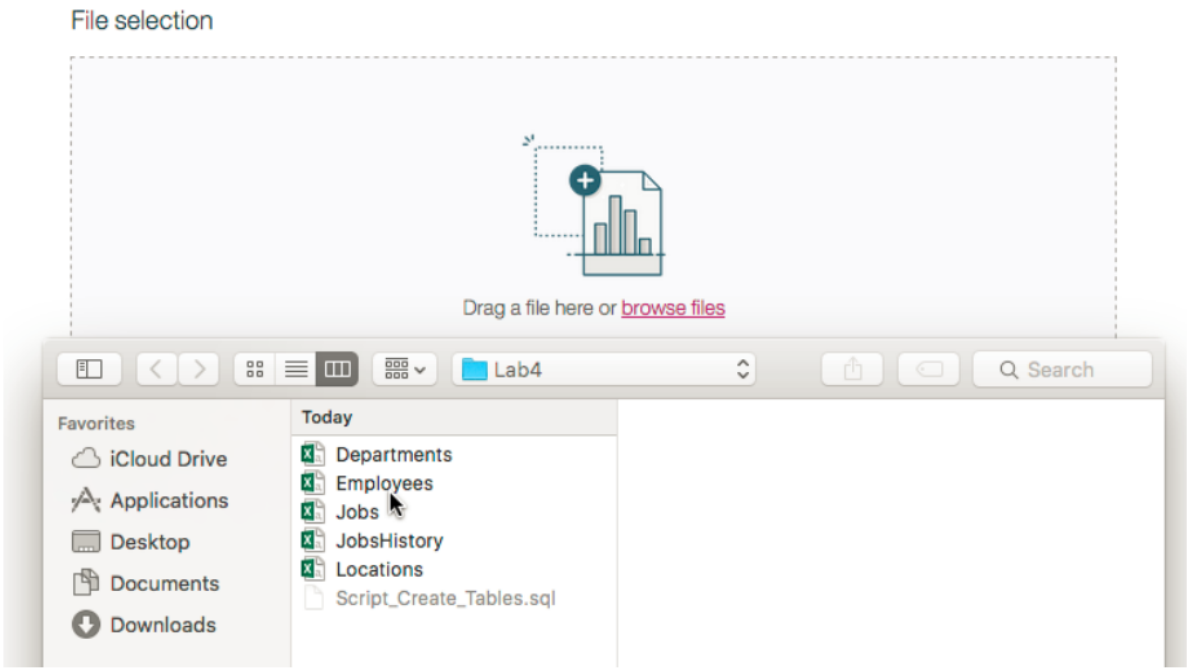
1) Download the 5 required data source files: (“Employees.csv”,”Departments.csv”,”Jobs.csv”,”JobsHistory.csv”,”Locations.csv”) to your computer:

2) First let us learn how to load data into the Employees table that we created earlier. From the 3 bar menu icon, navigate to the "Load" page and ensure "My Computer" is selected as the source. Click on the "browse files" link.

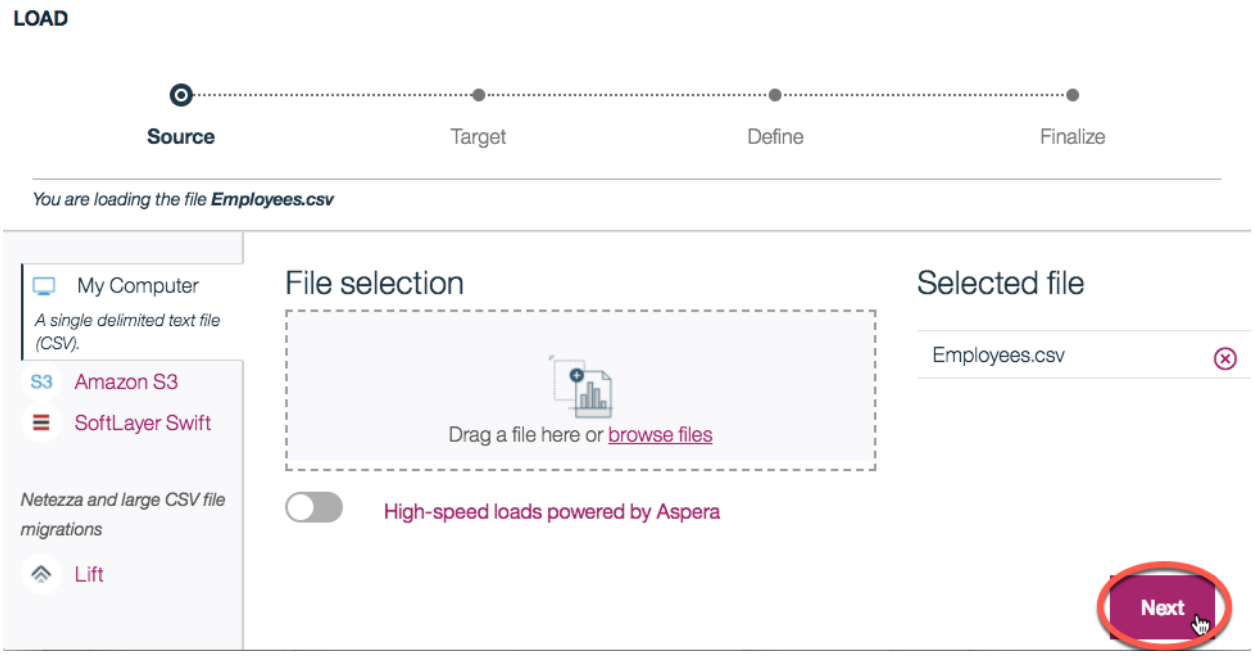


3) Choose the file "Emloyees.csv" that you downloaded to your computer and click "Open".

Copyright ©2018 IBM Corporation

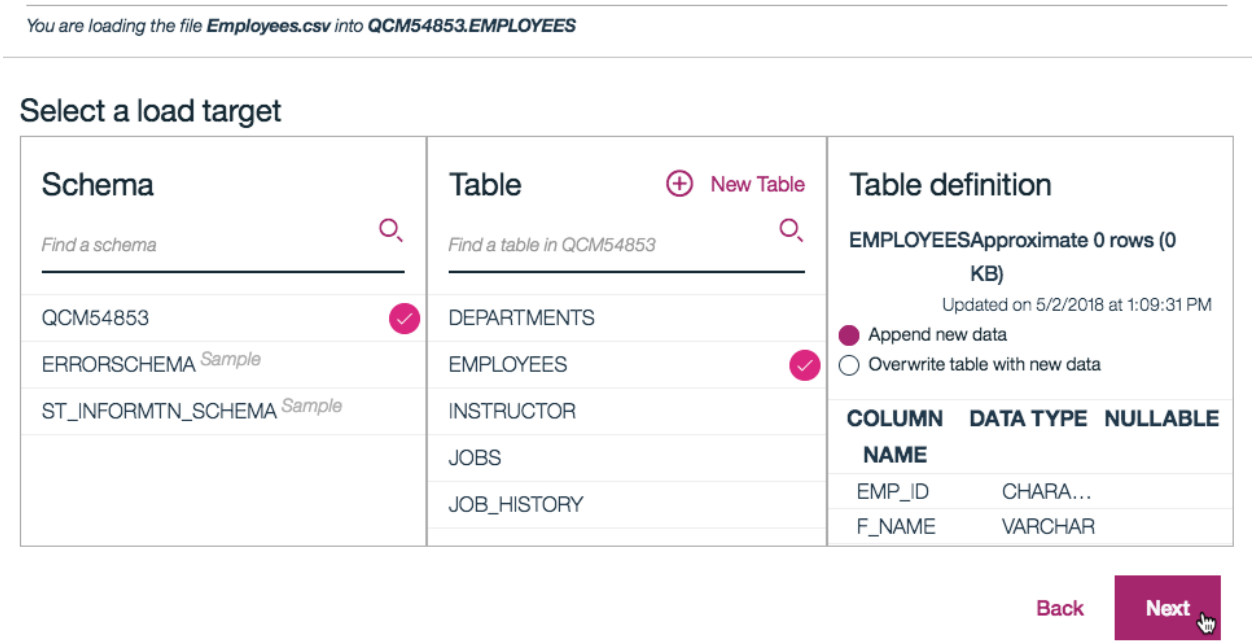


4) Once the File is selected click "Next" in the bottom right corner.

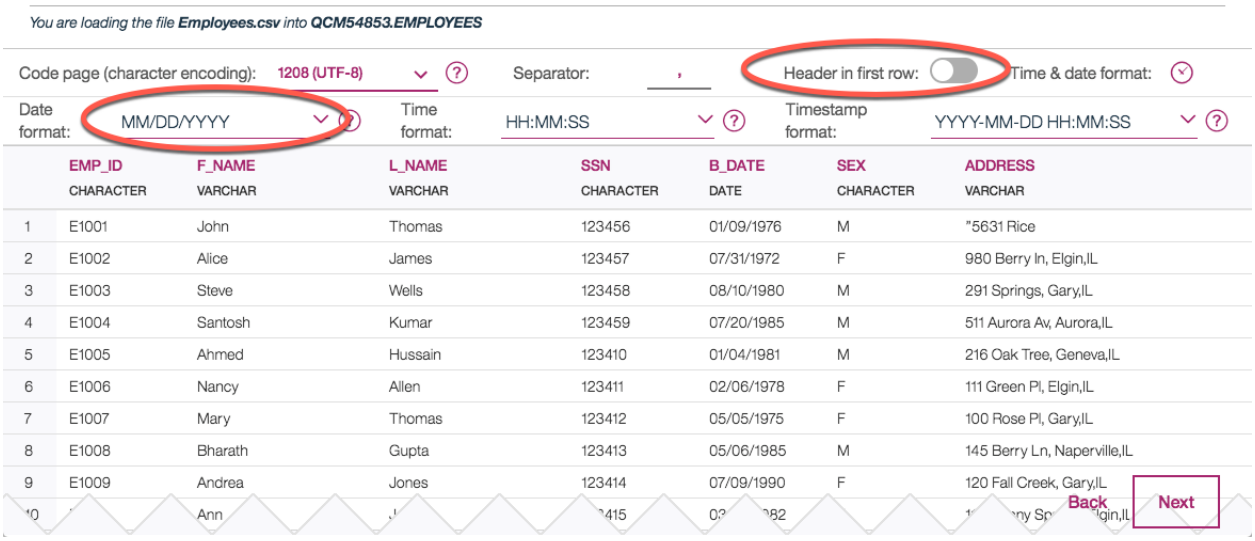


5) Select the schema for your Userid. It will show all the tables that have been created in this Schema previously, including the Employees table. Select the EMPLOYEES table.

Copyright ©2018 IBM Corporation



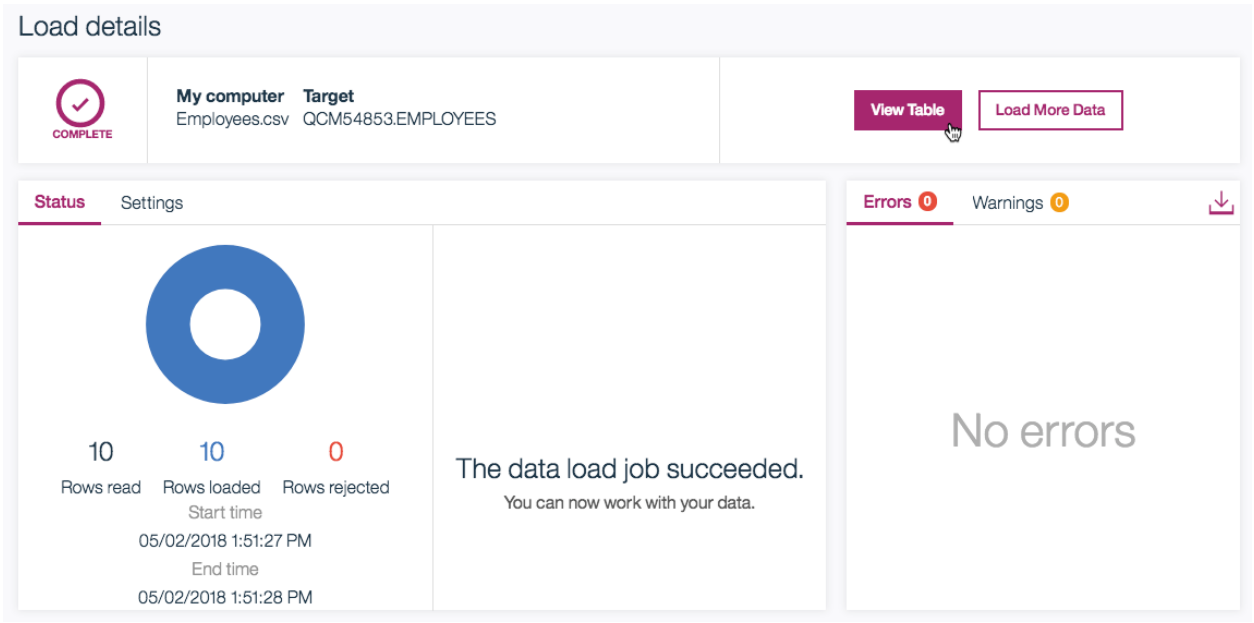
6) Since our source data files do not contain any rows with column labels, turn off the setting for "Header in first row". Also, click on the down arrow next to "Date format" and choose "MM/DD/YYYY" since that is how the date is formatted in our source file.



7) Click "Next". Review the Load setting and click "Begin Load" in the top-right corner.

Copyright ©2018 IBM Corporation

8) After Loading is complete you will notice that we were successful in loading all 10 rows of the Employees table. If there are any Errors or Warnings you can view them on this screen.



9) You can see the data that was loaded by clicking on the View Table. Alternatively you can go into the Explore page and page select the correct schema, then the EMPLOYEES table, and click "View Data".

Copyright ©2018 IBM Corporation

QCM54853.EMPLOYEES

Delete Table Export to CSV

	EMP_ID CHARACTER(9)	F_NAME VARCHAR(15)	L_NAME VARCHAR(15)	SSN CHARACTER(9)	B_DATE DATE	SEX CHARACTER(1)	ADDRESS VARCHAR(30)	JOB_ID CHARACTER(9)
1	E1001	John	Thomas	123456	1976-01-09	M	5631 Rice, OakPark,	100
2	E1002	Alice	James	123457	1972-07-31	F	980 Berry ln, Elgin,Il	200
3	E1003	Steve	Wells	123458	1980-08-10	M	291 Springs, Gary,IL	300
4	E1004	Santosh	Kumar	123459	1985-07-20	M	511 Aurora Av, Aurora	400
5	E1005	Ahmed	Hussain	123410	1981-01-04	M	216 Oak Tree, Geneva	500
6	E1006	Nancy	Allen	123411	1978-02-06	F	111 Green Pl, Elgin,IL	600
7	E1007	Mary	Thomas	123412	1975-05-05	F	100 Rose Pl, Gary,IL	650
8	E1008	Bharath	Gupta	123413	1985-05-06	M	145 Berry Ln, Naper	660
9	E1009	Andrea	Jones	123414	1990-07-09	F	120 Fall Creek, Gary,	234
10	E1010	Ann	Jacob	123415	1982-03-30	F	111 Britany Springs,E	220

10. Now its your turn to load the remaining 4 tables of the HR database – Locations, JobHistory, Jobs, and Departments. Please follow the steps above to load the data from the remaining source files.

Question 1: Were there any warnings loading data into the JOBS table? What can be done to resolve this?

Hint: View the data loaded into this table and pay close attention to the JOB_TITLE column.

Question 2: Did all rows from the source file load successfully in the DEPARTMENT table? If not, are you able to figure out why not?

Hint: Look at the warning. Also, note the Primary Key for this table.

Copyright ©2018 IBM Corporation

Part III: COMPOSING AND RUNNING QUERIES

You created the tables for the HR database schema and also learned how to load data into these tables. Now try and work on a few advanced DML queries that were introduced in this module.

Follow these steps to create and run the queries indicated below

1) Navigate to the Run SQL tool in Db2 on Cloud

2) Compose query and run it.

3) Check the Logs created under the Results section. Looking at the contents of the Log explains whether the SQL statement ran successfully. Also look at the query results to ensure the output is what you expected.

Query 1: Retrieve all employees whose address is in Elgin,IL

[Hint: Use the LIKE operator to find similar strings]

Query 2: Retrieve all employees who were born during the 1970's.

[Hint: Use the LIKE operator to find similar strings]

Query 3: Retrieve all employees in department 5 whose salary is between 60000 and 70000 .

[Hint: Use the keyword BETWEEN for this query]

Query 4: Retrieve a list of employees ordered by department name, and within each department ordered alphabetically in descending order by last name.

[Hint: Use the 'order by" clause for this query]

Query 5: Retrieve the department number, the number of employees in the department, and their average salary.

[Hint: Use COUNT(*) to retrieve the total count of a column, and AVG() function]

In this lab you learned how to work with string patterns, sorting result sets and grouping result sets.

Thank you for completing this lab!

Copyright ©2018 IBM Corporation

Lab Solutions

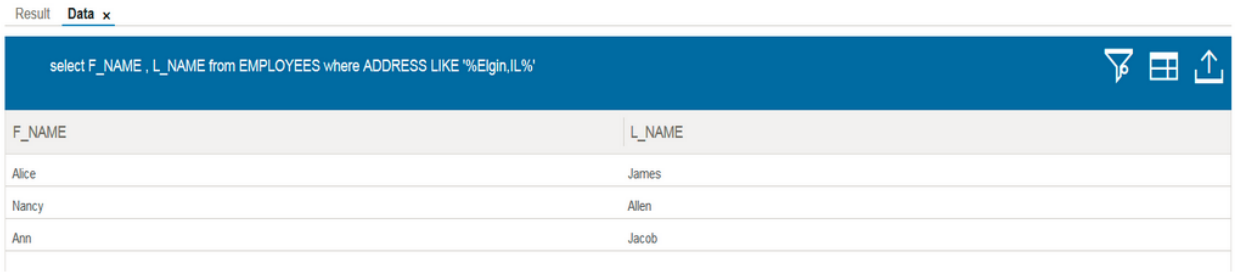
Please follow these steps to get the answers to the queries:

1) Navigate to the Run SQL page on Db2 on Cloud.

2) Download the script file(“Module4_Queries.txt”) or text files(“Modules4_Queries.sql”). Open the file with extension .sql in the editor

3) Run the queries. Looking at the contents of the Log explains that the SQL statement that we ran was successful. Here are the results for the queries:

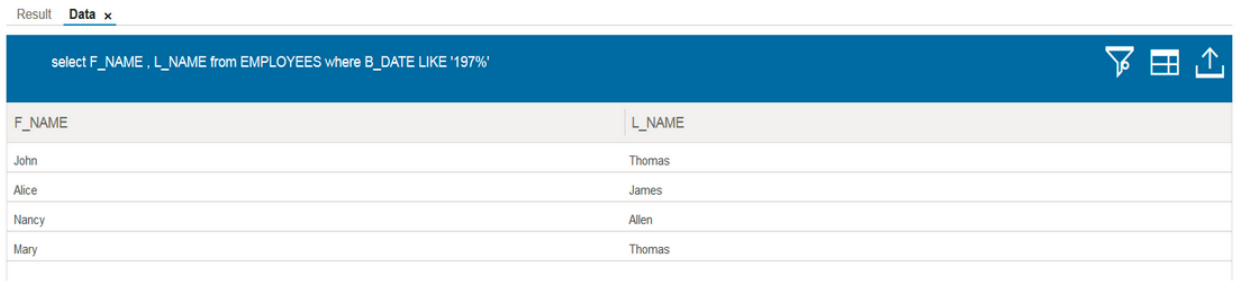
Query 1: Output


Result Data ×

select F_NAME , L_NAME from EMPLOYEES where ADDRESS LIKE '%Elgin,IL%'

F_NAME	L_NAME
Alice	James
Nancy	Allen
Ann	Jacob

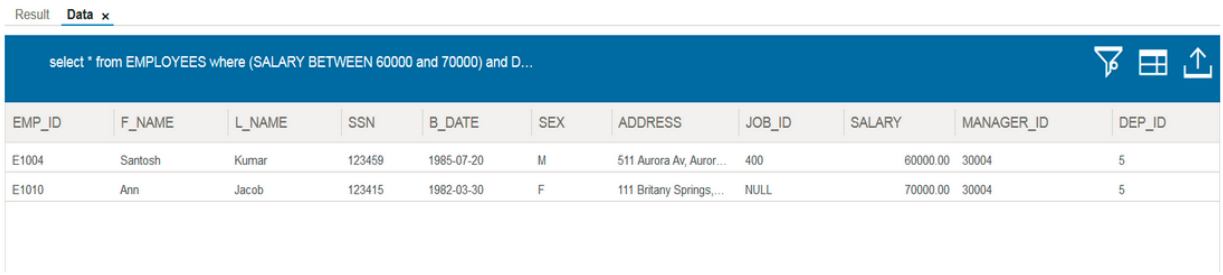
Query 2: Output


Result Data ×

select F_NAME , L_NAME from EMPLOYEES where B_DATE LIKE '197%'

F_NAME	L_NAME
John	Thomas
Alice	James
Nancy	Allen
Mary	Thomas

Query 3: Output

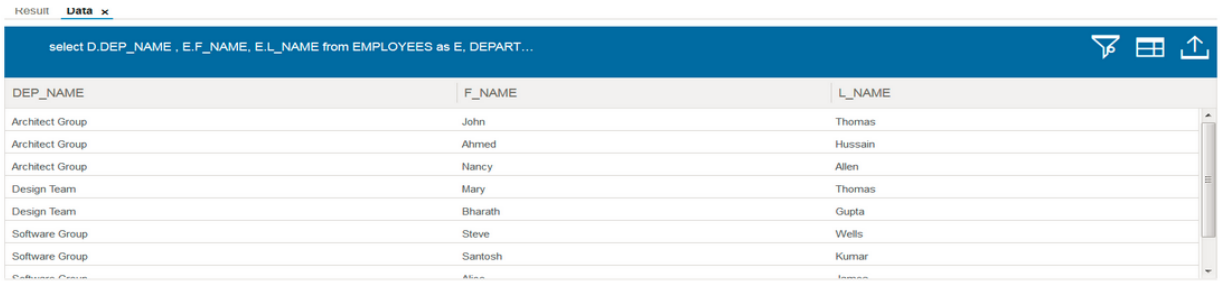

Result Data ×

select * from EMPLOYEES where (SALARY BETWEEN 60000 and 70000) and D...

EMP_ID	F_NAME	L_NAME	SSN	B_DATE	SEX	ADDRESS	JOB_ID	SALARY	MANAGER_ID	DEP_ID
E1004	Santosh	Kumar	123459	1985-07-20	M	511 Aurora Av, Auror...	400	60000.00	30004	5
E1010	Ann	Jacob	123415	1982-03-30	F	111 Britany Springs,...	NULL	70000.00	30004	5

Copyright ©2018 IBM Corporation

Query 4: Output

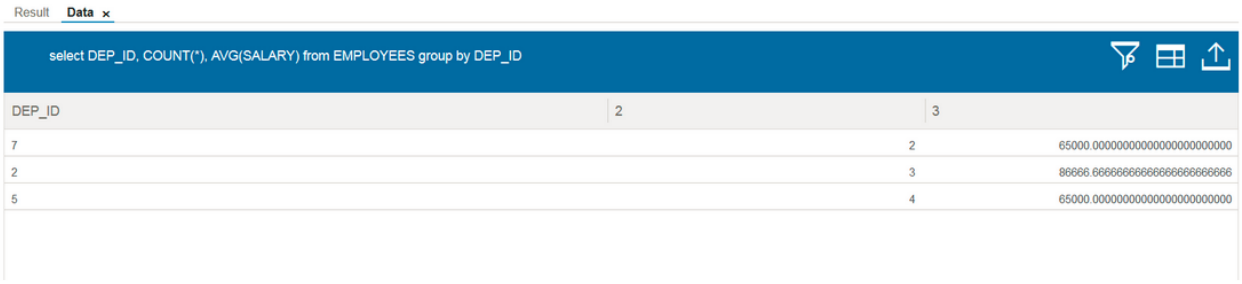


Result Data ×

select D.DEP_NAME , E.F_NAME, E.L_NAME from EMPLOYEES as E, DEPART...

DEP_NAME	F_NAME	L_NAME
Architect Group	John	Thomas
Architect Group	Ahmed	Hussain
Architect Group	Nancy	Allen
Design Team	Mary	Thomas
Design Team	Bharath	Gupta
Software Group	Steve	Wells
Software Group	Santosh	Kumar

Query 5: Output



Result Data ×

select DEP_ID, COUNT(*), AVG(SALARY) from EMPLOYEES group by DEP_ID

DEP_ID	2	3
7	2	65000.000000000000000000000
2	3	86666.666666666666666666666
5	4	65000.000000000000000000000

Copyright ©2018 IBM Corporation