Module 1 - Getting Started with SQL

In this module, you will learn some basic SQL statements and practice them hands-on on a live database.

Learning Objectives

- Describe SQL and Databases
- Explain the syntax of basic SQL statements - Select, Insert, Update, Delete
- Compose and execute basic SQL statements hands-on on a live database
- Demonstrate how to write basic SQL statements

1. Introduction to Databases

Introduction to Databases

IBM Developer   SKILLS NETWORK

Course Overview

- Basics of SQL
- Relational Database Model

- At the end of this course, you will be able to discuss SQL basics and explain aspects of the relational database model

- At the end of this lesson, you will be able to:
  - Describe SQL, data, database, relational database
  - List five basic SQL commands

IBM Developer   SKILLS NETWORK

Hello and welcome to SQL for data science. First, we will talk a little bit about what you'll learn in this course.
- This course teaches you the basics of the SQL language and the relational database model.
- There will be some lab exercises, and at the end of each section there are a few review questions.
- And at the end, there is a final exam.
By the end of this course, you will be able to discuss SQL basics and explain various aspects of the relational database model.

In this video, we will learn about SQL and relational databases. By the end of this video, you will be able to describe SQL, data, database, a relational database, and list five basic SQL commands.

What is SQL?

- A language used for relational databases
- Query data

IBM Developer   SKILLS NETWORK

What is data?

- Facts (words, numbers)
- Pictures
- One of the most critical assets of any business
- Needs to be secure

IBM Developer   SKILLS NETWORK

But wait, what is SQL and what is a relational database? What is SQL? SQL is a language used for relational databases to query or get data out of a database. SQL is also referred to as SQL and is short for its original name Structured English Query Language. So SQL is a language used for a database to query data.

But what is data and what is a database?
Data is a collection of facts in the form of words, numbers, or even pictures. Data is one of the most critical assets of any business. It is used and collected practically everywhere. Your bank stores data about you, your name, address, phone number, account number et cetera. Your credit card company and your paypal accounts also store data about you. Data is important; so, it needs to be secure, and it needs to be stored and accessed quickly. The answer is a database.

总结：

What is a database?

- A repository of data
- Provides the functionality for adding, modifying and querying that data
- Different kinds of databases store data in different forms

Relational Database

- Data stored in tabular form - columns and rows
- Columns contain item properties e.g. Last Name, First Name, etc.
- Table is collection of related things e.g. Employees, Authors, etc.
- Relationships can exist between tables
  (hence: "relational")

Student ID	First Name	Last Name
34933	Victoria	Slater
93759	Justin	McNeil
20847	Jessica	Bennett
65947	Michelle	Dolin
24956	David	Price
65692	Franklin	Mullins
24271	Alissa	Lee

So, what is a database? Databases are everywhere and used every day, but they are largely taken for granted.
- A database is a repository of data. It is a program that stores data.
- A database also provides the functionality for adding, modifying, and querying that data.
- There are different kinds of databases of different requirements.

The data can be stored in various forms.
- When data is stored in tabular form, the data is organized in tables like in a spreadsheet, which is columns and rows. That's a relational database.
- The columns contain properties about the item such as last name, first name, email address, city.
- A table is a collection of related things like a list of employees or a list of book authors.
- In a relational database, you can form relationships between tables.

DBMS

- Database: repository of data
- DBMS: Database Management System - software to manage databases
- Database, Database Server, Database System, Data Server, DBMS - often used interchangeably

What is RDBMS?

- RDBMS = Relational database management system
- A set of software tools that controls the data
  - access, organization, and storage
- Examples are: MySQL, Oracle Database, IBM Db2, etc.

So a database is a repository of data. A set of software tools for the data in the database is called a database management system or DBMS for short. The terms database, database server, database system, data server, and database management systems are often used interchangeably.

For relational databases, it's called a relational database management system or RDBMS. RDBMS is a set of software tools that controls the data such as access, organization, and storage. And RDBMS serves as the backbone of applications in many industries including banking, transportation, health, and so on. Examples of relational database management systems are my SQL, Oracle Database, DB2 Warehouse, and DB2 on Cloud.

Basic SQL Commands

- Create a table
- Insert
- Select
- Update
- Delete

Summary

- You can now describe:
  - SQL
  - Data
  - Database
  - Relational Databases
  - RDBMS
  - 5 basic SQL commands:
    - Create, Insert, Select, Update, Delete

For the majority of people using a database, there are five simple commands to create a table, insert data to populate the table, select data from the table, update data in the table, delete data from the table.

So those are the building blocks for SQL for data science. You can now describe what is SQL, what is data, what is a database, and what is a relational database. You know that RDBMS stands for Relational Database Management System, and you can list five basic SQL commands to create a table, insert data to populate the table, select data from the table, update data in the table, and delete data from the table.

总结:

2. SELECT Statement

SELECT statement
Retrieving data from a table

IBM Developer   SKILLS NETWORK   IBM Developer   SKILLS NETWORK

Retrieving rows from a table

- After creating a table and inserting data into the table, we want to see the data
- SELECT statement
  - A Data Manipulation Language (DML) statement used to read and modify data

```
Select statement:    Query
Result from the query:  Result set/table

Select * from <tablename>
```

IBM Developer   SKILLS NETWORK

The main purpose of a database management system, is not just to store the data but also facilitate retrieval of the data.
So, after creating a relational database table and inserting data into the table, we want to see the data. To see the data, we use the SELECT statement.

- The SELECT statement is a data manipulation language statement. Data Manipulation Language statements or DML statements are used to read and modify data.
- The SELECT statement is called a query, and the output we get from executing this query is called a result set or a result table. In its simplest form, a SELECT statement is select star from table name.

Using the SELECT Statement
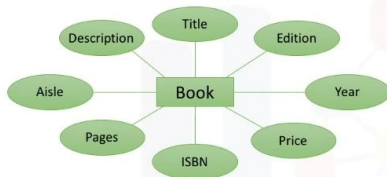
Title	Edition	Year	Price	ISBN	Pages	Aisle	Description
Database Fundamentals	1	2010	24.99	978-0-9800628-3-1	300	DB-A02	Teaches you the fundamentals of databases
Getting started with DB2 Express-C	1	2010	24.99	978-0-9866628-3-5-1	280	DB-A01	Teaches you the essentials of DB2 using DB2 Express-C

Example: select * from Book

```
db2 => select * from Book

Book_ID  Title                      Edition  Year  Price  ISBN           Pages  Aisle   Description
B1       Getting started with DB2   1        2010  24.99  978-0-         280    DB-A01  Teaches you the essentials of DB2
         Express-C                                        98666283-5-1                  using DB2 Express-C
B2       Database Fundamentals      1        2010  24.99  978-0-         300    DB-A02  Teaches you the fundamentals of
                                                          98006283-1-1                  databases
B3       Getting started with DB2 App 1      2011  35.99  978-0-         345    DB-A03  Teaches you the essentials of
         Dev                                              98086283-4-1                  developing applications for DB2.
B4       Getting started with WAS CE 1       2010  49.99  978-0-         458    DB-A04  Teaches you the essentials of
                                                          98946283-3-1                  WebSphere Application Server

    4 record(s) selected.
```

Example: select <column 1, column 2, ..., column n from Book

```
db2 => select book_id, title, edition, year, price, ISBN, pages, aisle, description from Book
```

IBM Developer   SKILLS NETWORK

Based on the book entity example, we would create the table using the entity name book and the entity attributes as the columns of the table. The data would be added to the book table by adding rows to the table using the insert statement.

In the book entity example, select star from book gives the result set of four rows. All the data rows for all columns in the table book are displayed.

In addition, you can also retrieve all the rows for all columns by specifying the column names individually in the SELECT statement.

总结：

Retrieving a subset of the columns

- You can retrieve just the columns you want
- The order of the columns displayed always matches the order in the SELECT statement
  - `SELECT <column 1>, <column 2> from Book`

```
db2 => select book_id, title from Book

Book_ID     Title
----------------------------------------
B1          Getting started with DB2 Express-C
B2          Database Fundamentals
B3          Getting started with DB2 App Dev
B4          Getting started with WAS CE

    4 record(s) selected.
```

IBM Developer

SKILLS NETWORK

You don't always have to retrieve all the columns in a table. You can retrieve just a subset of columns. If you want, you can retrieve just two columns from the table book.

For example book_id and title. In this case, the select statement is select book_id, title from book. In this case, only the two columns display for each of the four rows.

Also notice that the order of the columns displayed always matches the order in the SELECT statement.

Restricting the Result Set: WHERE Clause

- Restricts the result set
- Always requires a Predicate:
  - Evaluates to: True, False or Unknown
  - Used in the search condition of the Where clause

```
select book_id, title from Book
        WHERE predicate


db2 => select book_id, title from Book
        WHERE book_id='B1'


Book_ID     Title
----------------------------------------
B1          Getting started with DB2 Express-C

    1 record(s) selected
```

IBM Developer

SKILLS NETWORK

What if we want to know the title of the book whose book_id is B1. Relational operation helps us in restricting the result set by allowing us to use the clause WHERE. The WHERE clause always requires a predicate.
A predicate is conditioned evaluates to true, false or unknown. Predicates are used in the search condition of the WHERE clause.
So, if we need to know the title of the book whose book_id is B1, we use the WHERE clause with the predicate book_id equals B1. Notice the result set is now restricted to just one row whose condition evaluates to true.

WHERE Clause Comparison Operators

```
select book_id, title from Book
    WHERE book_id = 'B1'
```

Summary

Now you can:
- Retrieve data from a relational database table
- Define the use of a predicate
- Identify the syntax of the SELECT statement using the WHERE clause
- List the comparison operators supported by a RDBMS

Equal to	=
Greater than	>
Lesser than	<
Greater than or equal to	>=
Less than or equal to	<=
Not equal to	<>

IBM Developer

SKILLS NETWORK

总结:

Reading: SELECT statement examples (2 mins)

![IBM Developer Skills Network logo]

Objectives

At the end of this lab you will be able to:

- use SELECT queries to retrieve data from the database

Effort: 2 min

The general syntax of SELECT statments is:

select COLUMN1, COLUMN2, ... from TABLE1 ;

To retrieve all columns from the COUNTRY table we could use "*" instead of specifying individual column names:

select * from COUNTRY ;

The WHERE clause can be added to your query to filter results or get specific rows of data. To retrieve data for all rows in the COUNTRY table where the ID is less than 5:

select * from COUNTRY where ID < 5 ;

In case of character based columns the values of the predicates in the where clause need to be enclosed in single quotes. To retrieve the data for the country with country code "CA" we would issue:

select * from COUNTRY where CCODE = 'CA';

In the lab that follows later in the module, you will apply these concepts and practice SELECT queries hands-on.

Good luck!

Author(s)

Rav Ahuja

Changelog

Date	Version	Changed by	Change Description
2020-08-31	2.0	Malika Singla	Markdown file creation

总结:

How does the syntax of a SELECT statement look?

```
SELECT column1, column2, ...
FROM table_name
WHERE condition
;
```

What do the keywords / clauses of a SQL statement shown above do?

- **FROM**: Specifies from which table to get the data. The clause can include optional JOIN subclauses to specify the rules for joining tables.
- [Optional Clause] **WHERE** : Specifies which rows to retrieve.

Why is there a semicolon after the SQL statements?

- Some database systems require a semicolon at the end of each SQL statement for execution. It is a standard way to separate one SQL statement from another which allows more than one SQL statement to be executed in the same call to the server. So, it is good practice to use a semicolon at the end of each SQL statement.

. Problem:

Retrieve the names, production company names, filming locations, and release years of the films which are not written by James Cameron.

▲ Click here for Hint
▶ Click here for Solution

```
SELECT Title, ProductionCompany, Locations, ReleaseYear FROM FilmLocations WHERE Writer<>"James Cameron";
```

3. COUNT, DISTINCT, LIMIT

COUNT

COUNT() - a built-in function that retrieves the number of rows matching the query criteria.

Number of rows in a table:

```
select COUNT(*) from tablename
```

COUNT

Rows in the MEDALS table where Country is Canada:

```
select COUNT(COUNTRY) from MEDALS
       where COUNTRY='CANADA'
```

Result:

```
        1
---------------
        29
```

In this video, we'll briefly present a few useful expressions that are used with select statements.
The first one is COUNT, COUNT is a built-in database function that retrieves the number of rows that match the query criteria.
For example, get the total number of rows in a given table, **select COUNT(*) from tablename**. Let's say you create a table called MEDALS which has a column called COUNTRY, and you want to retrieve the number of rows where the medal recipient is from Canada. You can issue a query like this:
Select COUNT(COUNTRY) from MEDALS where COUNTRY='CANADA.'

DISTINCT

DISTINCT is used to remove duplicate values from a result set.

Retrieve unique values in a column:

```
select DISTINCT columnname from tablename
```

DISTINCT

List of unique countries that received GOLD medals:

```
select DISTINCT COUNTRY from MEDALS
       where MEDALTYPE = 'GOLD'
```

Result:

```
        1
---------------
        21
```

The second expression is DISTINCT. DISTINCT is used to remove duplicate values from a result set.
Example, to retrieve unique values in a column, **select DISTINCT columnname from tablename**. In the MEDALS table mentioned earlier, a country may have received a gold medal multiple times. Example, retrieve the list of unique countries that received gold medals. That is, removing all duplicate values of the same country. **Select DISTINCT COUNTRY from MEDALS where MEDALTYPE = 'GOLD'**.

LIMIT

LIMIT is used for restricting the number of rows retrieved from the database.

Retrieve just the first 10 rows in a table:

```
select * from tablename LIMIT 10
```

LIMIT

Retrieve 5 rows in the MEDALS table for a particular year:

```
select * from MEDALS
       where YEAR = 2018 LIMIT 5
```

Result:

COUNTRY	GOLD	SILVER	BRONZE	TOTAL	YEAR
Norway	14	14	11	39	2018
Germany	14	10	7	31	2018
Canada	11	8	10	29	2018
United States	9	8	6	23	2018
Netherlands	8	6	6	20	2018

The third expression is LIMIT, LIMIT is used for restricting the number of rows retrieved from the database. Example, retrieve just the first 10 rows in a table. **Select * from tablename LIMIT 10**. This can be very useful to examine the results set by looking at just a few rows instead of retrieving the entire result set which may be very large. Example, retrieve just a few rows in the MEDALS table for a particular year. **Select * from MEDALS where YEAR = 2018 LIMIT 5**.

In this video we looked at some useful expressions that are used with select statements, namely the COUNT, DISTINCT, and LIMIT built-in functions.

总结:

4. INSERT Statement

Hello, and welcome to the INSERT statement. In this video, we will learn about populating a relational database table. At the end of this video, you'll be able to

- Identify the syntax of the INSERT statement.
- Explain two methods to add rows to a table. One row at a time or multiple rows.

After table is created, the table needs to be populated with data. To insert data into a table, we use the INSERT statement. The INSERT statement is used to add new rows to a table. The INSERT statement is one of the data manipulation language statements. Data manipulation language statements or DML statements are used to read and modify data.

Based on the author entity example, we created the table using the entity name, author, and the entity attributes as the columns of the table. Now we will add the data to the author table by adding rows to the table.
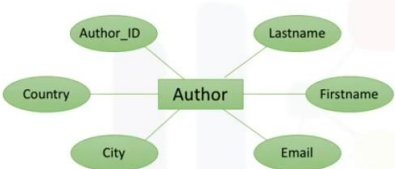
Using the INSERT Statement

Author_ID	Lastname	Firstname	Email	City	Country
A1	Chong	Raul	rfc@ibm.com	Toronto	CA
A2	Ahuja	Rav	ra@ibm.com	Toronto	CA
A3	Hakes	Ian	ih@ibm.com	Toronto	Ca
A4	Sharma	Neeraj	ns@ibm.com	Chennai	IN
A5	Perniu	Liviu	lp@ibm.com	Transylvania	RO

INSERT INTO [TableName]
  <([ColumnName],...)>
VALUES ([Value],...)

INSERT INTO AUTHOR
  (AUTHOR_ID, LASTNAME, FIRSTNAME, EMAIL, CITY, COUNTRY)
VALUES ('A1', 'Chong', 'Raul', 'rfc@ibm.com', 'Toronto', 'CA')

To add the data to the author table, we use the INSERT statement. The syntax of the INSERT statement looks like this:
INSERT INTO [TableName] <([ColumnName],...)> VALUES ([Value],...)
In this statement, table name identifies the table, the column name list identifies each column in the table, and the values clause specifies the data values to be added to the columns in the table. To add a row with the data for Raul Chong, we insert a row with an author underscore ID of A one, the last name is Chong, the first name as Raul, the email as RFC@IBM.com, the city as Toronto, and the country as CA for Canada.
The author table has six columns, so the INSERT statement lists the six column names separated by commas, followed by a value for each of the columns also separated by commas. It is important that the number of values provided in the values clause is equal to the number of column names specified in the column name list. This ensures that each column has a value.

Inserting multiple rows

Author_ID	Lastname	Firstname	Email	City	Country
A1	Chong	Raul	rfc@ibm.com	Toronto	CA
A2	Ahuja	Rav	ra@ibm.com	Toronto	CA
A3	Hakes	Ian	ih@ibm.com	Toronto	Ca
A4	Sharma	Neeraj	ns@ibm.com	Chennai	IN
A5	Perniu	Liviu	lp@ibm.com	Transylvania	RO

INSERT INTO AUTHOR
  (AUTHOR_ID, LASTNAME, FIRSTNAME, EMAIL, CITY, COUNTRY)
VALUES
  ('A1', 'Chong', 'Raul', 'rfc@ibm.com', 'Toronto', 'CA')
  ('A2', 'Ahuja', 'Rav', 'ra@ibm.com', 'Toronto', 'CA')

Tables do not need to be populated one row at a time. Multiple rows can be inserted by specifying each row in the values clause. In the values clause, each row is separated by a comma. For example, in this INSERT statement we are inserting two rows, one for Raul Chong and one for Rav Ahuja.

总结：

5. UPDATE and DELETE Statements

Hello and welcome to the UPDATE Statement and the DELETE Statement. In this video, we will learn about altering and deleting data in a relational database table. At the end of this lesson, you will be able to

- identify the syntax of the UPDATE statement and DELETE statement.
- explain the importance of the WHERE clause in these statements.

To alter or modify the data in the author table, we use the UPDATE statement. The syntax of the UPDATE statement looks like this: **UPDATE [TableName] SET [[ColumnName]=[Value]]** In this statement, table name identifies the table, the column name identifies the column value to be changed as specified in the WHERE condition.

Let's look at an example. you want to update the first name and last name of the author with Author_Id A2 from Rav Ahuja to Lakshmi Katta. In this example, to see the update statement in action, we start by selecting all rows from the author table to see the values. To change the first name and last name to Lakshmi Katta, where the author ID is equal to A2, enter the UPDATE statement as follows: **UPDATE AUTHOR SET LASTNAME='KATTA' FIRSTNAME='LAKSHMI' WHERE AUTHOR_ID='A2'**
Now, to see the result of the update, select all rows again from the author table and you will see that in row two the name changed from Rav Ahuja to Lakshmi Katta. Note that if you do not specify the WHERE clause, all the rows in the table will be updated. In this example, without specifying the WHERE clause, all rows in the table would have changed the first and last names to Lakshmi Kata.

Sometime later, there might be a need to remove one or more rows from a table. The **rows** are removed with the DELETE statement. The DELETE statement is one of the data manipulation language statements used to read and modify data. The syntax of the DELETE statement looks like this: **DELETE FROM [TableName] WHERE [Condition]**
The rows to be removed are specified in the WHERE condition. Based on the author entity example, we want to delete the rows for author ID A2 and A3. Let's look at an example. **DELETE FROM AUTHOR WHERE AUTHOR_ID ID IN ('A2', 'A3')**
Note that if you do not specify the WHERE clause, all the rows in the table will be removed.

总结:

Summary & Highlights

Congratulations! You have completed this lesson. At this point in the course, you know:

- You can use Data Manipulation Language (DML) statements to read and modify data.

- The search condition of the WHERE clause uses a predicate to refine the search.

- COUNT, DISTINCT, and LIMIT are expressions that are used with SELECT statements.

- INSERT, UPDATE, and DELETE are DML statements for populating and changing tables.

总结：

Which of the following statements are correct about databases?

○ A database is a repository of data

○ There are different types of databases - Relational, Hierarchical, No SQL, etc.

○ A database can be populated with data and be queried

◉ **All of the above**

✔

True or False: A SELECT statement is used to retrieve data from a table.

◉ **True**

○ False

✔

What is the function of a WHERE clause in an UPDATE statement?

◉ **A WHERE clause enables you to specify which rows will be updated.**

○ A WHERE clause enables you to specify a new table to receive the updates.

○ A ~~WHERE~~ **SET** clause enables you to list the column and data to be updated.

○ A WHERE clause is never used with an UPDATE statement.

✔

总结：