MAZ Học Data

PRACTICAL SQL FOR DATA ANALYSIS

"In Data We Trust"

Course Objectives

Fundamental knowledge

Understand **key concepts** about data science industry, relational database, data pipeline, data analysis process, benefits of SQL

T-SQL statements

Use **basic & advanced T-SQL statements** to structure, query, transform data and customize databases

Cases studies

Apply **problem solving skill** and **logical thinking** to complete challenging cases studies

...

Course Outline

01 **Getting started & exploring general concepts**

SQL– Relational Database – Schema - Normalization

02 **Writing simple SELECT queries**

Retrieve the results by Select – From - Where

03 **Using built-in functions & expressions**

Use built-in functions to cleaning data

04 **Joining tables & unions**

Expand your data by combining and merging multi tables

05 **Subqueries, CTE and intermediate practices**

Enhance your queries with common table expression and subqueries

06 **Grouping, summarizing data and discovering advanced analytics queries**

Use advanced analytics queries with GROUP BY and WINDOW FUNCTION

07 **Apply problem solving skills into data analytics querying case studies (1)**

User retention with Cohort Analysis

08 **Apply problem solving skills into data analytics querying case studies (2)**

Improve user retention rate with User Segmentation

01

Database concepts

SQL– Relational Database –
Schema

Introduction to SQL

SQL: Structured Query Language

- SQL is a standard language that was designed to query and manage data in relational database management systems (RDBMSs)
- SQL is both an ANSI and ISO standard language based on the relational model

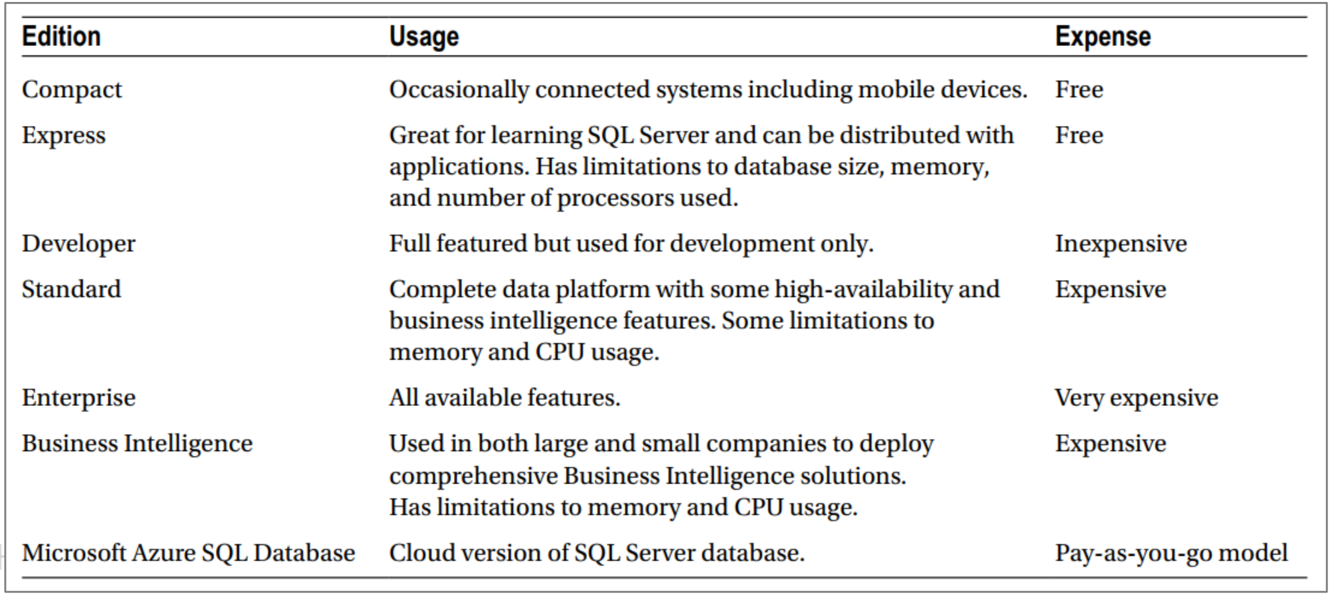
T-SQL is Microsoft’s implementation

- T-SQL is based on standard SQL, but it also provides some nonstandard/proprietary extensions
- Other company ‘ implementation: MySQL, SQLite, PostgreSQL, PL-SQL, ...

Introduction to SQL

SQL Sever

- SQL Server is Microsoft's relational database management system (RDBMS). An RDBMS stores data in tables according to the relational model.
- Microsoft makes SQL Server available in many editions.

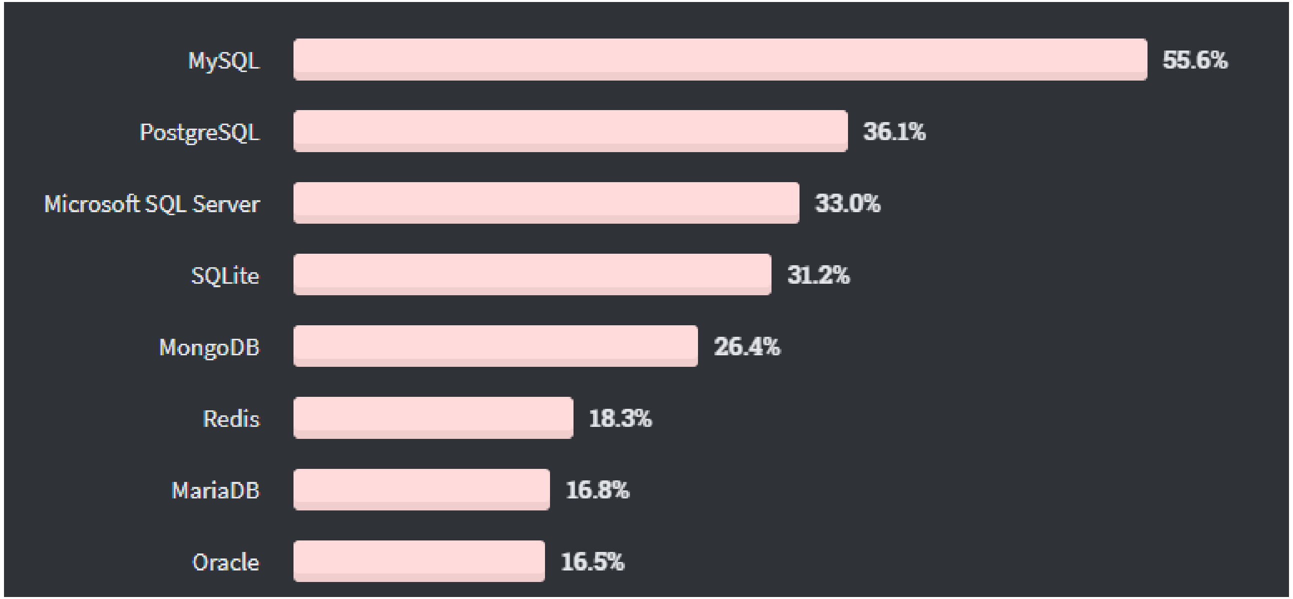


Edition	Usage	Expense
Compact	Occasionally connected systems including mobile devices.	Free
Express	Great for learning SQL Server and can be distributed with applications. Has limitations to database size, memory, and number of processors used.	Free
Developer	Full featured but used for development only.	Inexpensive
Standard	Complete data platform with some high-availability and business intelligence features. Some limitations to memory and CPU usage.	Expensive
Enterprise	All available features.	Very expensive
Business Intelligence	Used in both large and small companies to deploy comprehensive Business Intelligence solutions. Has limitations to memory and CPU usage.	Expensive
Microsoft Azure SQL Database	Cloud version of SQL Server database.	Pay-as-you-go model

Introduction to SQL

Most popular databases in 2021

Source: StackOverflow Developer Survey Results 2021



Datasbases

Database

- You can think of a database as a container of objects such as tables, views, stored procedures, and other objects.

Relational Database

- Entities are represented as relations (tables), in which their attributes are represented as domains (columns).
- Most relational databases are normalized, with relationships defined between tables through primary and foreign keys.
- An RDBMS is supposed to implement the relational model and provide the means to store, manage, enforce the integrity of, and query data.

Primary Key & Foreign Key

Primary Key

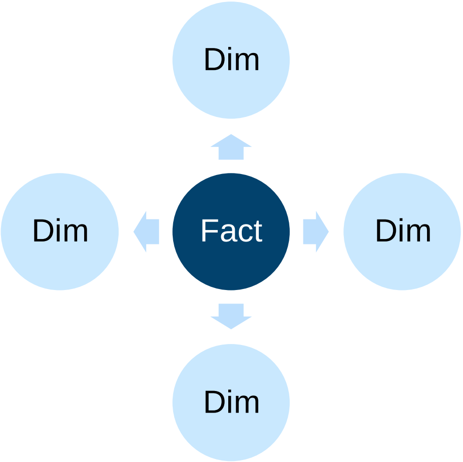
- **One** or **more columns** that can be used as a **unique identifier** for each row in a table

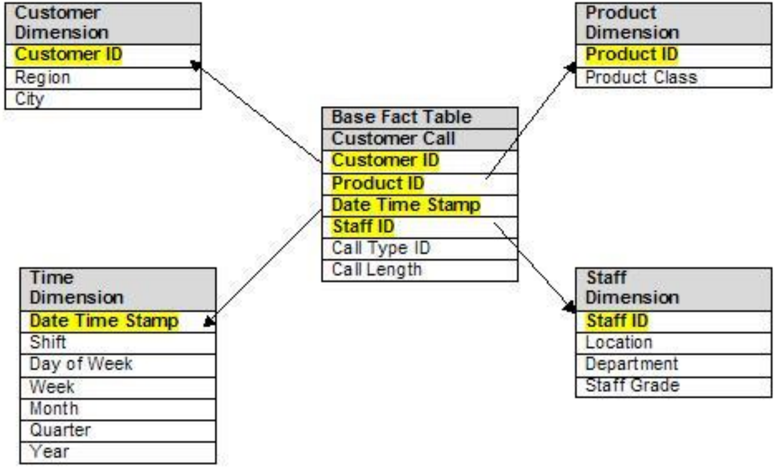
Foreign Key

- One or more columns that can be used together to identify a single row in another table

Star schema

- …





Customer Dimension
Customer ID
Region
City

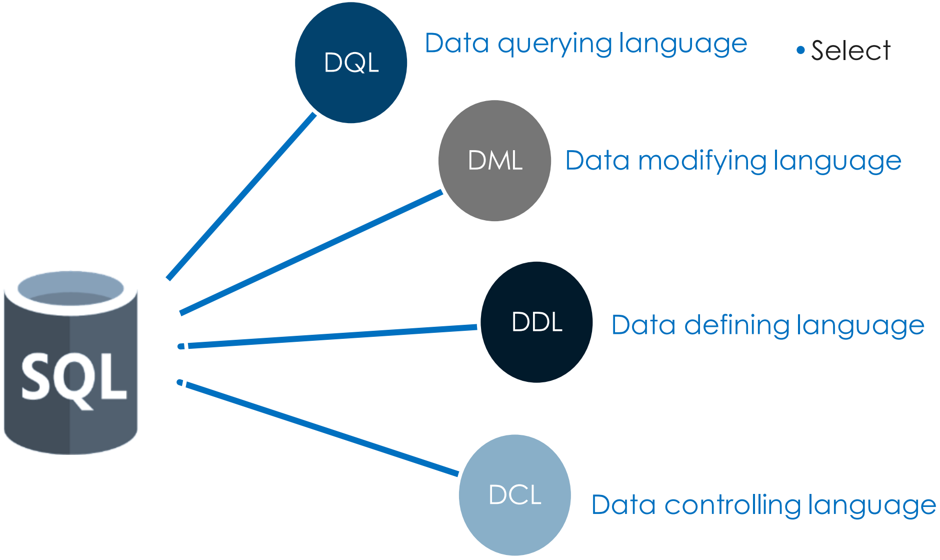
Product Dimension
Product ID
Product Class

Base Fact Table
Customer Call
Customer ID
Product ID
Date Time Stamp
Staff ID
Call Type ID
Call Length

Time Dimension
Date Time Stamp
Shift
Day of Week
Week
Month
Quarter
Year

Staff Dimension
Staff ID
Location
Department
Staff Grade

SQL Commands



Data querying language

- Select

Data modifying language

- Insert
- Update
- Delete

Data defining language

- Create
- Alter
- Drop
- Truncate

Data controlling language

- Grant
- Revoke

02

Writing simple SELECT queries

Retrieve the results by
Select – From - Where

Using the SELECT Statement

Selecting a Literal Value

```
SELECT 1
SELECT ‘Toi ten la: Your name'
SELECT N‘Tôi tên là Hiếu Nguyễn'
```

Retrieving from a Table

```
-- Show specific columns
SELECT  <column1>
        , <column2>
        , ...
FROM  <schema>.<table>;
--Show all columns in a table
SELECT  *
FROM  <schema>.<table>;
```

Mixing Literals and Column Names

- Let try it

How to write in a query editor?

Comment and take note

- **Single line comments**

  -- This is my notes

- **Multiple-line comments**

  /* This is my notes.

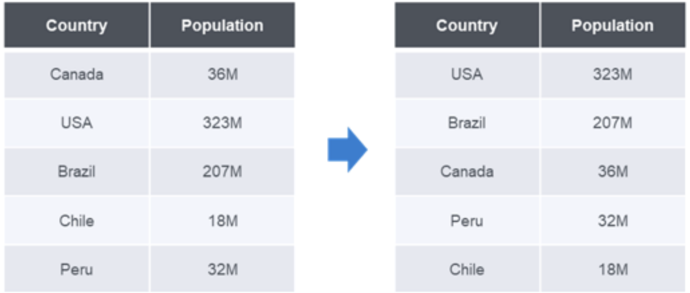
  With many rows … */

Sorting Data

So far, you have learned how to retrieve a list of columns from a table and filter the results. This section covers how to sort the data that are retrieved using the **ORDER BY** clause.

Here is the syntax:

```
SELECT      <column1>, <column2>
FROM        <table_name>
ORDER BY    <column1> ASC,  <column2> DESC
```



Country	Population
Canada	36M
USA	323M
Brazil	207M
Chile	18M
Peru	32M

Country	Population
USA	323M
Brazil	207M
Canada	36M
Peru	32M
Chile	18M

Try it:

```
SELECT *
FROM SalesLT.Customer
ORDER BY FristName ASC
```

Filtering Data

Adding a WHERE Clause

```
-- Syntax:
SELECT      <column1>,<column2>
FROM        <schema>.<table>
WHERE       <condition> ;
```

Try it:

```
SELECT *
FROM SalesLT.Product
WHERE ProductCategoryID = 6
AND StandardCost < 1000
```

Exercise 1: Write a query using a WHERE clause that displays all the products listed in the **SalesLT.Product** table which have the **color "black"** and **size 50**. Display the ProductID, Name, Color, Size, for each one.

Then sorts in ascending order of ProductCategoryID and descending order of ListPrice.

Using WHERE Clauses with Alternate Operators

Group	Operator	Description
Comparison operators	=	Equal
	<> Or !=	Not equal
	>	Greater than
	<	Less than
	>=	Greater than or equal
	<=	Less than or equal
Logical operators	AND	Return records that meet all the conditions separated by AND in WHERE clause.
	OR	Return records that meet any of the conditions separated by OR in WHERE clause.
	NOT	Return records that do not satisfy any of the conditions in WHERE clause.
SQL operators	[NOT] BETWEEN	Returns values [NOT] within a given range
	[NOT] IN	Specify multiple values in a WHERE clause (a shorthand for multiple OR conditions)
	IS [NOT] NULL	Returns records having [NOT] NULL values in the given fileds.
	[NOT] LIKE	Returns records that [DO NOT] match a specified pattern in a column.

Using the IN

```
-- Syntax:
SELECT      <column1>,<column2>
FROM        <schema>.<table>
WHERE       <column> IN (<value1>,<value2>, ...);
```

```
-- Syntax:
SELECT      <column1>,<column2>
FROM        <schema>.<table>
WHERE       <column> IN ( SELECT STATEMENT );
```

Exercise 2: Write a query using a WHERE clause that displays all the products listed in the SalesLT.Product table which have the color “black” or ‘red’ and size is greater than 50. Display the ProductID, Name, Color, Size, for each one.

Using the LIKE Operator (Pattern matching)

Syntax:

```
SELECT      <column1>,<column2>
FROM        <schema>.<table>
WHERE       <column> LIKE pattern ;
```

Wildcard

%	Represent zero or more characters
_	Represent exactly one character
[ ]	Restrict a list or range of characters
^	Include the caret (^) symbol to list characters or the range of characters that you don’t want to use as replacements

Exercise 3: Retrieve the ProductID, ProductNumber and Name of the products, that must

- Have Product number begins with ‘FR-‘

- Have Product name contains ‘HL’ or ‘Mountain’

Time for practicing

Exercise 4: Retrieve the ProductID, ProductNumber and Name of the products,

- That must have Product number begins with ‘FR-' followed by any character other than ‘M' and ends with a '-' followed by any two numerals.

- And satisfy one of the following conditions: color of black, red, or white size is S or M and */