**SQL Manual**

You can have a scenario where you have multiple databases each having their own tables. You may have a database to manage customer tables, and another for product, etc.

You can create a table using two ways. One is a query and one is a table.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Click save, and then give the table a name. Click refresh, and then you will see that dbo.Employee has been created as a table. Right click on the table and you can add values to the table by going to edit top 200 rows.

A screenshot of a computer program

Description automatically generated

After adding the values:

A screenshot of a computer

Description automatically generated

Right click the table and select top 1000 rows. You will now see the rows:

A screenshot of a computer

Description automatically generated

In the same way, you can right click and create a database as well.

**Query to create database**



**Query to create table**



Make sure that all queries or scripts are run in the database that you want to execute them in. You can select the database in the dropdown.

**Making comments**

You can add two dashes to make anything into a comment.

**Inserting data into a table**

A screenshot of a computer

Description automatically generated

You can insert rows or multiple rows of data as above.

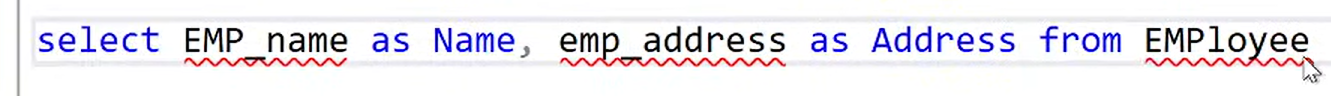
**Selecting columns**

A close-up of words

Description automatically generated

You can select all columns, or you can choose to select only some columns.

**Using an alias when selecting columns**



This only changes the name for the selection purpose, but not the name of the column in the DB.

**Deleting rows from a table**

A blue text on a white background

Description automatically generated

Above we are saying to delete all rows from the Employe table.

Deleting tables

A close up of a sign

Description automatically generated

You can delete tables with the query above.

**CRUD Operations (Create, Read, Update, and Delete).**

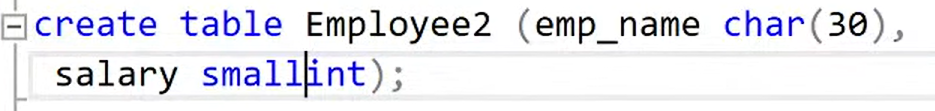
**Data Types**

You must design your database with the correct data types

**A screenshot of a computer

Description automatically generated**

So based on the type of data, you can can state the data type.

****

**Precision**

When defining the data type, you can state the precision. So if you want to store 123, you want to store it as five digits and 2 decimal places.

**A screenshot of a computer

Description automatically generated**

So a practical example of this is:

**A screenshot of a computer

Description automatically generated**

**Storing the data considering UNICODE**

A close up of a computer screen

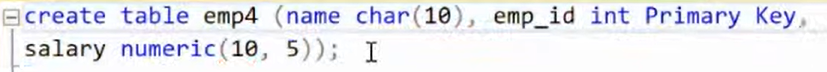
Description automatically generated

Sometimes people have letters which don’t fall in ASCII, and may be foreign names. You can use nchar data type and you must use the N keyword when entering the value into the table.

**Identity**

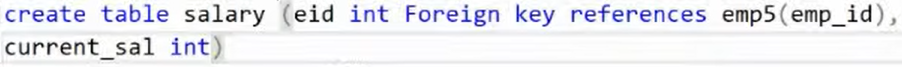
We all have some sort of a unique identity, for example employee id’s or passport numbers. There could be many davids, or johns in this world. When you store information in a table, you have to be able to uniquely identify each and every row. A unique ID is essentially a primary key constraint. You cant have two rows with the same ID in a table. Also, a null ID is not allowed.

Using the primary key constraint.



Referential Identity concept

When creating a table which holds a foreign key, it is used to reference a different table.



So we create a new table named salary and give it a column named eid which is a foreign key to the emp5 table which column emp\_id. You can only insert values into the eid column in the foreign key table if the emp\_id in the emp5 table exists, otherwise it will not be allowed.

**Unique, not null, and Default Constraints**

Lets create a new table in a database.

CREATE TABLE Citizens (

Citizen\_ID INT PRIMARY KEY,

Citizen\_Name VARCHAR(100) NOT NULL,

NI\_Number VARCHAR(20) UNIQUE,

Country VARCHAR(50) DEFAULT 'United Kingdom');

Notice how the national insurance number has a UNIQUE constraint and how the country has a default constraint. That is because we want the NI number to be unique (without being the primary key) and we want the country to have a default value of united kingdom. We have stated that the name cannot be a NULL value, so the user is forced to insert a value for it. We then insert values into the table:

INSERT INTO Citizens (Citizen\_ID, Citizen\_Name, NI\_Number, Country)

VALUES

(1, 'John Doe', 'AB123456C', 'United Kingdom'),

(2, 'Jane Smith', 'CD789012D', 'United Kingdom'),

(3, 'Alice Johnson', 'EF345678E', 'United Kingdom'),

(4, 'Bob Wilson', 'GH901234F', 'United Kingdom'),

(5, 'Eva Davis', 'IJ567890G', 'United Kingdom');

**Identity constraint**

Notice how you have been putting in the values 1,2,3,4,5 for the primary key. However, you do not have to do that. You can apply the identity constraint to the primary key to force it to auto increment.

CREATE TABLE Citizens (

Citizen\_ID INT PRIMARY KEY IDENTITY(1,1),

Citizen\_Name VARCHAR(100),

NI\_Number VARCHAR(20) UNIQUE,

Country VARCHAR(50) DEFAULT 'United Kingdom'

);

You can now insert into without putting the primary key in there:

INSERT INTO Citizens (Citizen\_Name, NI\_Number, Country)

VALUES

('Mary Johnson', 'KL123456H', 'United Kingdom'),

('Tom Williams', 'MN789012I', 'United Kingdom'),

('Grace Brown', 'OP345678J', 'United Kingdom');

Lets say that you wanted to alter the table that was created before and you wanted to add the identity constraint to it. You can do this with the use of the ALTER key word.

ALTER TABLE Citizens

ALTER COLUMN Citizen\_ID INT PRIMARY KEY IDENTITY(1,1);

**Difference between primary key and unique**

The unique constraint allows a NULL value, whereas the primary key does not.

**Foreign Key Constraint**

Lets assume that we are creating a voters table that needs to be linked to the citizen table through the citizen\_id. We state the foreign key references keywords and connect the citizen\_id.

A screenshot of a computer code

Description automatically generated

**Check Constraint**

We can check if certain columns have a condition. For the example below we have added a check if the age is greater than or equal to 18. We have also added a check to state that the height must be greater or equal to 120. Also, the mobile number must have a length which is equal to 11. If these conditions are not met, then a value cannot be added into the respective columns. Finally, we have declared the Citizen\_ID column so that it exists in both tables. We then add a foreign key constraint linking the two columns which have the same name.

CREATE TABLE Citizen\_info (

Info\_ID smallint PRIMARY KEY,

Age smallint CHECK (Age >= 18),

Height smallint CHECK (Height >= 120),

Mobile\_no INT UNIQUE CHECK (LEN(Mobile\_no) = 11),

Citizen\_ID INT,

FOREIGN KEY (Citizen\_ID) REFERENCES Citizens(Citizen\_ID)

);

**Foreign Key Constraint with the correct constraints**

CREATE TABLE Citizen\_info (

Info\_ID smallint PRIMARY KEY IDENTITY (1,1),

Age smallint CHECK (Age >= 18),

Height smallint CHECK (Height >= 120),

Mobile\_no bigint UNIQUE CHECK (LEN(Mobile\_no) = 11),

Citizen\_ID int,

FOREIGN KEY (Citizen\_ID) REFERENCES Citizens(Citizen\_ID)

);

INSERT INTO Citizen\_info (Age, Height, Mobile\_no, Citizen\_ID)

VALUES

(25, 160, 12345678901, 1),

(30, 175, 23456789012, 2),

(22, 150, 34567890123, 3);

select \* from Citizens

select \* from Citizen\_info

In the code above, I have created a table named Citizen info which will be connected to the citizens table. The info id column is the primary key and the age, height, and mobile number are all checked for specific inputs. The citizen ID has a foreign key constraint. The values are inserted into the table, and notice how I do not need to add the info id as it is auto incremented. I add the citizen id which exists in the citizens table. If I try to add a citizen Id in the citizens info table that does not exist in the citizens info table, an error will be received. The results look like this:

A screenshot of a computer

Description automatically generated

**Creating a Schema**

You can create tables and put them inside schemas. Schemas are just like folders or containers. Inside a database, you can have different schemas, and inside that schema, you will have tables. Lets now create a schema named bus drivers and then put two tables inside the bus\_drivers schema named employees and salary.

CREATE SCHEMA BusDrivers

CREATE TABLE BusDrivers.Employees (

Employee\_id SMALLINT PRIMARY KEY IDENTITY (1,1),

Age SMALLINT CHECK (Age >= 18),

Height SMALLINT CHECK (Height >= 100)

);

CREATE TABLE BusDrivers.Salary (

Salary INT,

CURRENCY NVARCHAR(10) DEFAULT N'£',

Employee\_id SMALLINT,

FOREIGN KEY (Employee\_id) REFERENCES BusDrivers.Employees(Employee\_id)

);

-- Inserting three rows into the BusDrivers.Employees table

INSERT INTO BusDrivers.Employees (Age, Height)

VALUES

(25, 175),

(30, 160),

(22, 180);

-- Inserting corresponding rows into the BusDrivers.Salary table

INSERT INTO BusDrivers.Salary (Salary, Employee\_id)

VALUES

(50000, 1),

(60000, 2),

(45000, 3);

SELECT \* FROM BusDrivers.Employees

SELECT \* FROM BusDrivers.Salary

A screenshot of a computer

Description automatically generated

A schema is a great way of putting your tables into categories.

**GetDate Function and smalldatetime data type.**

If we want to get the current date or todays date, then we can do that by using the GetDate function. Lets say we want to insert values into a table, but we want want to insert the current date, we can do that. Lets create a table named prisoners in the MinistryOfJustice schema.

CREATE SCHEMA MinistryOfJustice

Create TABLE MinistryOfJustice.Prisoners (

Pristoner\_id SMALLINT PRIMARY KEY IDENTITY (1,1),

Prisoner\_name varchar (20),

Date\_of\_entry smalldatetime

);

-- Inserting three rows into the MinistryOfJustice.Prisoners table

INSERT INTO MinistryOfJustice.Prisoners (Prisoner\_name, Date\_of\_entry)

VALUES

('John Doe', '2023-12-01 10:30:00'),

('Jane Smith', '2023-12-01 11:45:00'),

('Bob Wilson', '2023-12-02 09:15:00');

select \* from MinistryOfJustice.Prisoners

INSERT INTO MinistryOfJustice.Prisoners (Prisoner\_name, Date\_of\_entry)

VALUES

('Jeremy Salsa', GETDATE()),

('Laura Bailey', GETDATE()),

('Hannah Winter', GETDATE());

SELECT \* FROM MinistryOfJustice.Prisoners

I first create the schema and then create a table in that schema. I add values to the table by specifying the date. However, the second time I add rows to the table, I use the getdate() function to add todays date for the entry. We could even add a check to ensure that todays date is used. We could also put a default value of todays date as well if we wish.

This is the result:

A screenshot of a computer

Description automatically generated

The TOP function

If we want to see the top number of rows in a table we can use:

select TOP (3) \* FROM MinistryOfJustice.Prisoners

The result is as below:

A screenshot of a computer

Description automatically generated

**Seeing the distinct of unique entries in a specific column**

You can use the DISTINCT keyword to see the distinct values in a column.

select DISTINCT(CompanyName) FROM [SalesLT].[Customer]

A screenshot of a computer

Description automatically generated

This aggregation will give you all of the distinct values in a column. If you choose more than one column, It will give you the distinct combination of both rows.

**Selecting all columns**

You use the asterisk to select all columns

SELECT \* FROM dbo.Employee

**Selecting specific columns in a table**

SELECT Citizen\_ID, Citizen\_Name FROM [dbo].[Citizens]

The second query is much better performance wise. When you have large tables, this will be better when running this query.

**Altering a table to add a column**

ALTER TABLE dbo.Citizens ADD SALARY INT

When altering any table, you have to use the alter table command and then you can add a column by using the add command.

A screenshot of a computer

Description automatically generated

**Altering a table to drop a column**

As you want to drop a column, you are essentially altering the table, so you have to state the alter table command first, followed by drop column.

ALTER TABLE dbo.Citizens DROP COLUMN SALARY

A screenshot of a computer

Description automatically generated

The table has been altered, and the column named salary has been dropped.

**Altering a table to change a datatype for a column**

Lets say that we first alter a table and add a column named salary with a data type of smallint.

ALTER TABLE dbo.Citizens ADD SALARY SMALLINT

We then realise that the data type is wrong, and we actually needed the data type INT. We can alter the table and the column by doing the following.

ALTER TABLE dbo.Citizens ALTER COLUMN SALARY INT

**Ordering values**

Lets say we first select three columns from a table:

select ProductID, Name, StandardCost from [SalesLT].[Product]

A screenshot of a table

Description automatically generated

We then want to order by the standard cost. We can do this with:

select ProductID, Name, StandardCost from [SalesLT].[Product] ORDER BY StandardCost

A screenshot of a product list

Description automatically generated

If you want to change the order by high to low you just use the DESC command:

select ProductID, Name, StandardCost from [SalesLT].[Product] ORDER BY StandardCost DESC

**Where clause for conditional statements.**

If I want to see all the rows where the standard cost is above 2000, then you can use the where clause:

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE StandardCost >= 2000

A screenshot of a computer

Description automatically generated

You can also use the between clause, which allows you to see the number of rows that fall between certain amounts:

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE StandardCost between 1500 and 2500

A screenshot of a table

Description automatically generated

**Where clause combining conditional statements.**

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE StandardCost >= 2000 and ProductID > 752

A screenshot of a computer

Description automatically generated

We can also use the OR operator.

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE StandardCost >= 2000 or ProductID > 752

A screenshot of a table

Description automatically generated

**Like statement**

If you want the query to return values where the Name column starts with letters that match with mountain, then you would write this query:

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE Name LIKE 'Mountain%'

A screenshot of a table

Description automatically generated

However, if you want to look for the letter’s mountain inside any part of a record, then you would use two percentages.

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE Name LIKE '%Mountain%'

A screenshot of a data

Description automatically generated

**Not like statement**

Lets now say that we want to find records where the letters for mountain as one word do not occur anywhere in the Name column, we can use the NOT LIKE statement:

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE Name NOT LIKE '%Mountain%'

**A screenshot of a table

Description automatically generated**

**IN statement to check for multiple values in a column**

Lets say that I want to search for two specific product id’s in a column, I can use the IN clause to do this. I can also use the same concept in any other column when searching for things, as long as I am looking for a specific value.

select ProductID, Name, StandardCost from [SalesLT].[Product] WHERE ProductID IN (680, 706)

**A screenshot of a computer

Description automatically generated**

**COUNT statement to count the number of items**

Lets say I wanted to count the number of products in the table. I can do this by using the count function.

select COUNT (ProductID) AS Product\_Count from [SalesLT].[Product]

The count function counts the number of rows in a column. If I was to put in

select COUNT (\*) AS Product\_Count from [SalesLT].[Product]

we would still have the same count, as we are now counting the number of rows for every column, which is still the same count as the number of rows in the ProductID column.

**COUNT statement to count the number of distinct items**

Lets say that I want to count the distinct number of colours in the table. If I was to count all of the rows, the answer would be 225. However, if was to count the number of distinct items in the colour column, the answer is 9.

select COUNT (DISTINCT Color) AS Color\_count from [SalesLT].[Product]

**COUNT statement when having NULL values**

Remember that the count function will not count NULL values. This means that if you have a column with loads of null values, the count function will only count the rows that do not have these null values.

The statements below can be used to count the number of null values and not null values respectively.

select COUNT (DISTINCT Color) AS Color\_count from [SalesLT].[Product] WHERE Color IS NULL

select COUNT (DISTINCT Color) AS Color\_count from [SalesLT].[Product] WHERE Color IS NOT NULL

**Min and Max values.**

**You can select the minimum and maximum values in a column with:**

select min(StandardCost) from [SalesLT].[Product]

select max(StandardCost) from [SalesLT].[Product]

You can use the above in more creative ways as shown below.

SELECT \* FROM [SalesLT].[Product] WHERE StandardCost IN (SELECT MIN(StandardCost) FROM [SalesLT].[Product])

The query above first executes the inner query and then executes the outer. We are first saying to select the minimum value in the standard cost column. We are then saying to select all columns where standard cost contains that minimum value. In essence, the above is like saying the following.

SELECT \* FROM [SalesLT].[Product] WHERE StandardCost IN (0.85)

The approach above is called subqueries, as shown below.

A white board with red text and black text

Description automatically generated

**Average and Standard Deviation**

SELECT AVG(StandardCost) as Average\_Cost FROM [SalesLT].[Product]

SELECT STDEV(StandardCost) as Average\_Cost FROM [SalesLT].[Product]

**Mathematical Operations**

You can use SQL to perform maths as well as shown below. You can also use mathematical functions.

SELECT 10\*2 +5

SELECT SQRT(100)

So if you want the sqrt of each row in the standard cost column, you can do this with:

SELECT sqrt(StandardCost) FROM [SalesLT].[Product]

**Groupby function**

Let’s say that we want to find out the number of units sold for each of the different products in the example below. We have to group each of the products in order to find this out.

A screen shot of a computer

Description automatically generated

Lets now do the same example for a real example in my database.

In the statement below, we are only grouping the colours by their name.

SELECT Color FROM [SalesLT].[Product] group by Color

In the query below, we are grouping the colours by their name and we are getting the count of each of the colours that have been grouped.

SELECT Color, COUNT(Color) AS color\_count FROM [SalesLT].[Product] group by Color

A screenshot of a computer

Description automatically generated

You must made sure that the columns that you want to group by must be the only columns that are selected.

You can now get additional information with additional aggregate functions:

SELECT Color, COUNT(Color) AS color\_count, SUM (StandardCost) AS sum\_of\_standard\_cost FROM [SalesLT].[Product] group by Color

A table with numbers and text

Description automatically generated

We can group the colours, and then perform additional aggregate functions to get the colour count and then sum of the standard cost for each colour.

**Having Clause**

When you perform a group by statement, you MUST use the having clause if you want to filter or perform any conditional statements on the table. In normal cases where there is no group by function applied, you would use the WHERE clause, however, in this case you have to use the having clause.

Let’s say that we want to only get the records where the sum of the standard cost is less than 200 from the example above. We can do this as:

SELECT Color, COUNT(Color) AS color\_count, SUM (StandardCost) AS sum\_of\_standard\_cost

FROM [SalesLT].[Product] group by Color HAVING SUM(StandardCost) <200

Notice how you cannot use the ALIAS when using the having clause. For example, I cannot write HAVING sum\_of\_standard\_cost < 200.

**Inner Join**

When understanding joins, you must understand the concept of left and right tables. Without understanding this concept, it has hard to understand joins.

Right table.

Left table.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

You have been assigned a task to get the product categories for each of the product id’s and product names in the right table, as the product category name does not exist in the right table. In order to do this, you have to join the tables. You can join the tables in different ways, but we will first look at the inner join.

In the inner join, you start by selecting the columns that you want to see in both tables, and you use an alias so that you know which column you are seeing from which table (left and right table). You then write ‘INNER JOIN’ and then write the name of the right table. You then join them on their primary/foreign keys.

SELECT S.ProductCategoryID, S.Name, P.Name, P.ProductID FROM [SalesLT].[ProductCategory] AS S

INNER JOIN [SalesLT].[Product] AS P

ON S.ProductCategoryID = P.ProductCategoryID

You get the following result:

A screenshot of a table

Description automatically generated

Essentially, the inner join matches the rows where the product category id’s exist in both tables. As the product category ID of 5 (any many others that are not shown above) exist in the left table, they will be joined to the right table. However, notice how product categories 1,2,3, and 4 do not exist in the joint table as these product category ids’ do not exist in the right table. Lets say we want to see product categories in the joint table, even though their corresponding right table values are NULL. We can do this with a left join. A left join includes all primary key values in the joint table, even if they don’t exist in the right table.

**Left Join**

SELECT S.ProductCategoryID, S.Name, P.Name, P.ProductID FROM [SalesLT].[ProductCategory] AS S

LEFT JOIN [SalesLT].[Product] AS P

ON S.ProductCategoryID = P.ProductCategoryID

You can now see from the results that Product Categories 1,2,3, and 4 have been included, even though the columns Name and Product ID in the right table have values of NULL (as they don’t exist in the right table).

A screenshot of a computer

Description automatically generated

**Right Join**

Lets now look at different tables to depict the concept of a right join. The

Right table.

Left table.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

You can already see that the CustomerID column in the right table has values that are not present in the left table. Lets join the tables so that we get the CustomerID, AddressID from the left table, and the CustomerID, and first Name on the right table. However, lets perform a right join so that all CustomerID’s that exist in the right table end up in the joint table.

SELECT L.CustomerID, L.AddressID, R.CustomerID, R.FirstName FROM [SalesLT].[CustomerAddress] AS L

RIGHT JOIN [SalesLT].[Customer] AS R

ON L.CustomerID = R.CustomerID

By performing the query above, you can see the join below.

A screenshot of a table

Description automatically generated

If you go down, you can see that the values in the left table are also included, as they exist in the right table as well.

A screenshot of a table

Description automatically generated

At this point it is important to clarify that the left and right joins also match the primary/foreign key items between both tables, the only difference is that the left table also includes the primary/foreign key items in the left table that don’t exist in the right, and the right join does vice versa.

**Full outer Join**

The full outer join matches the primary/foreign key between two tables, but also includes the primary/foreign key records that don’t match and exist in the left table and those that don’t match and exist in the right table. Therefore, this shows a full join of both tables whether or not the values match or not.

SELECT L.CustomerID, L.AddressID, R.CustomerID, R.FirstName FROM [SalesLT].[CustomerAddress] AS L

FULL OUTER JOIN [SalesLT].[Customer] AS R

ON L.CustomerID = R.CustomerID