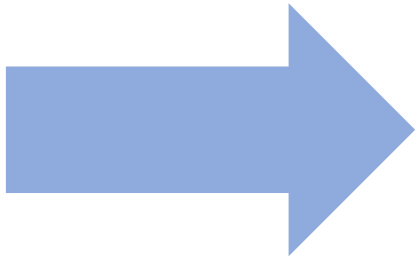
SQL Fundamentals

For Analyst and Data Analyst

KOTAK SAKTI

Leverage SQL for data analytics and business intelligence decision

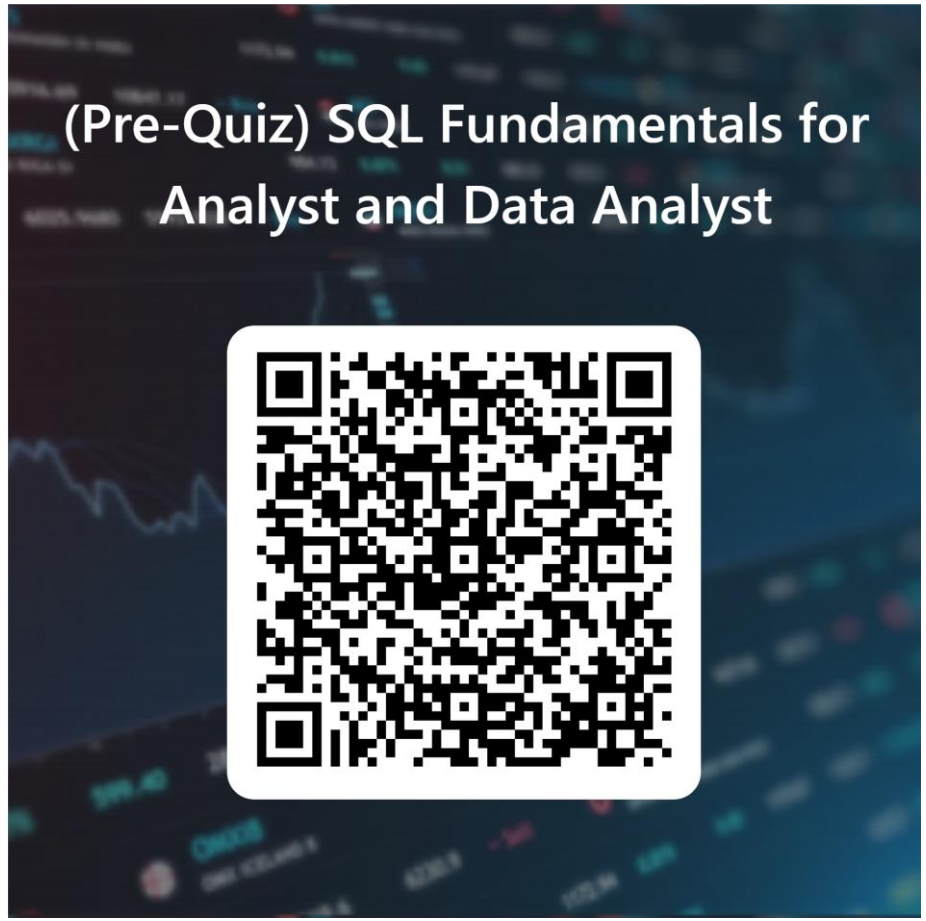
SQL Server Management Studio



Excel

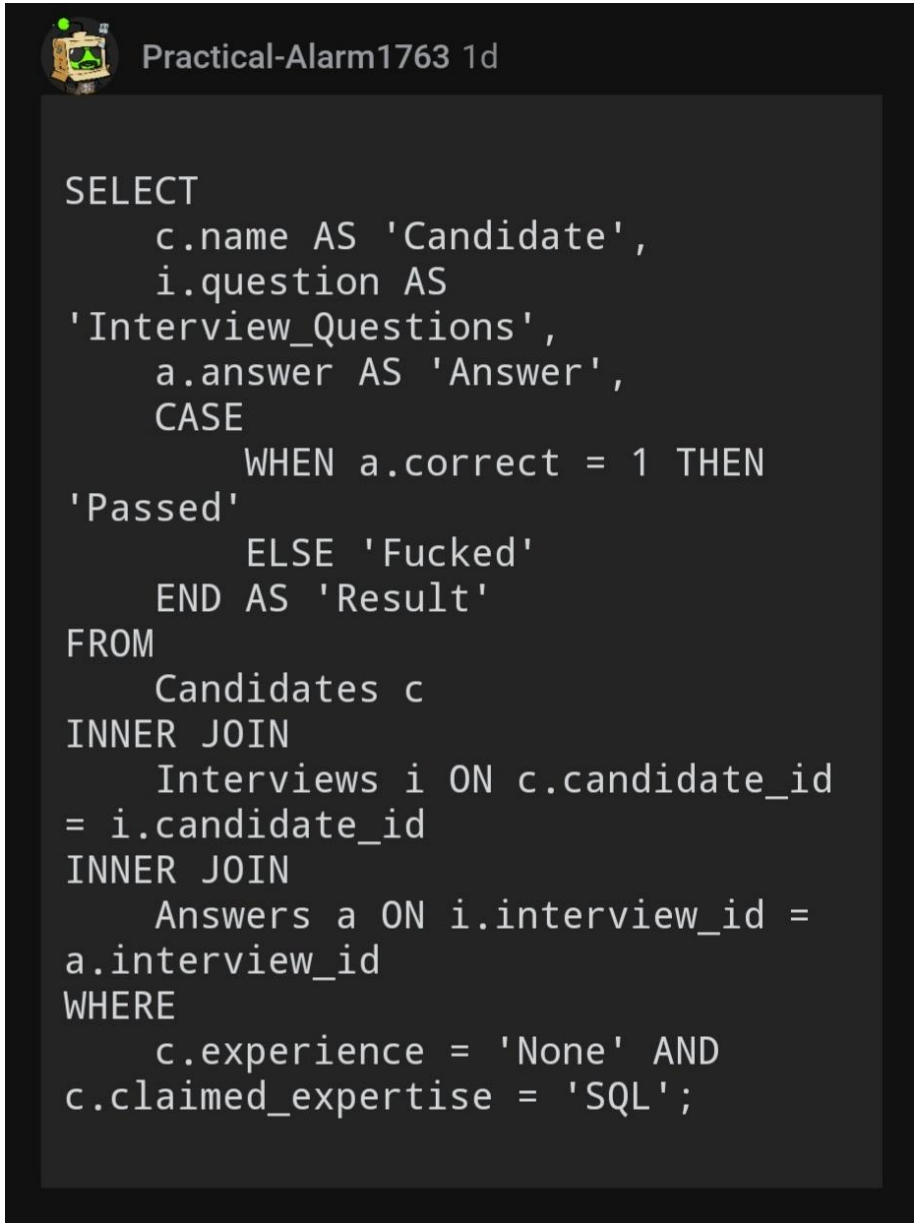
Tableau

PowerBI



https://forms.office.com/r/AYbYykY4J8

Can anyone guess what is the output of this query?



Learning Objectives

By the end of this session, you'll be able to:

- Understand what SQL is and how it is used
- Set up SQL Server and SQL Server Management Studio
- Fundamentals of database and how to query them
- How to get popular database in your local machine for training
- Connect your SQL Server to Excel and PowerBI (optional)
- Expand learning curve using documentations, Github and MOOCs

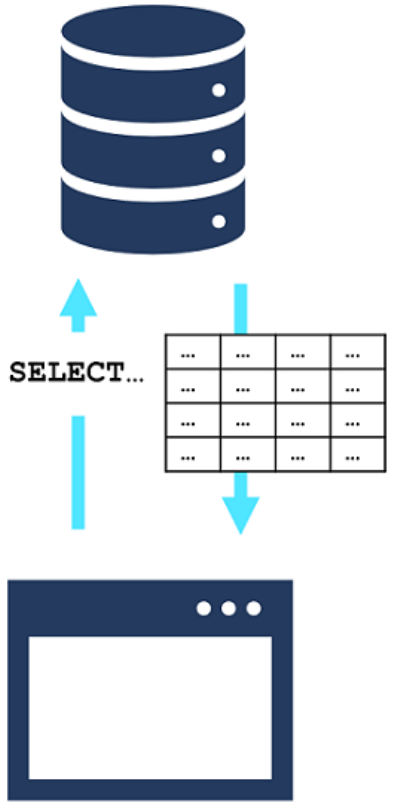
Prerequisites and setup steps

- A virtual machine or computer running Windows 10, Windows 11, with at least 4 CPUs and 8GB RAM.
- **Display:** At least 1440x900 or 1600x900 (16:9) recommended.
- **Internet connectivity:** You must have the ability to connect to the Internet
- Microsoft Excel (recommended) and Microsoft Power BI Desktop installed (optional)
- You will need Administrator rights on the virtual machine or computer. It is recommended to use your own personal laptop, do not use company laptop or owned by others.

Module 1

What is SQL?

Introduction to SQL: SQL in brief



- SQL stands for Structured Query Language, and was designed in the early 1970s at IBM
- SQL is used to communicate and manipulate with relational databases
- There is universal standard for SQL set by the American National Standards Institute (ANSI), with updates released every ~3-5 years
- Vendors are constantly adding new features on top of the standards, which creates different “flavors” of SQL (MySQL, PostgreSQL, SQLite, etc.)
- SQL is the standard language for relational database management

Introduction to SQL: Different flavors of SQL

These flavors of SQL are much more similar than they are different – all are based on the same universal standard, with slight variations in syntax. Each vendors add their own variations and extensions.

Introduction to SQL: Microsoft SQL Server

Microsoft database systems such as SQL Server, Azure SQL Database, Azure Synapse Analytics, and others use a dialect of SQL called **Transact-SQL, or *T-SQL*.**

T-SQL includes language extensions for writing stored procedures and functions, which are application code that is stored in the database, and managing user accounts.

Introduction to SQL: Important Technology / Concepts

Technology / Concept	Description
SQL Server 2022	The latest major version of SQL Server. Microsoft SQL Server is a relational database management system (RDBMS). Applications and tools connect to a SQL Server instance or database, and communicate using Transact-SQL (T-SQL).
SQL Server Management Studio (SSMS)	Graphical User Interface Management and Query Tool
Microsoft Azure	Microsoft's cloud platform for computing, data, and applications.
Database	A database in SQL Server is made up of a collection of tables that stores a specific set of *structured* data.
Tables	Tables are database *objects* that contain all the data in a database. In table, data is logically organized in a row-and-column format similar to a spreadsheet
Schema	In a SQL database, a schema is a list of logical structures of data
Instance	The name of the server or instance. For example, MyServer or MyServer\MyInstance.
Database normalization	Normalization is the process of structuring the tables and columns in a relational database to minimize redundancy and preserve data integrity.
Cardinality	Cardinality refers to the uniqueness of values in a column of a table, commonly described as how two tables relate (one-to-one, one-to-many, or many-to-many). • Primary keys are unique, cannot contain duplicates and null • Foreign keys are non-unique, may contain duplicates and null

Introduction to SQL :Database, Schemas and Table

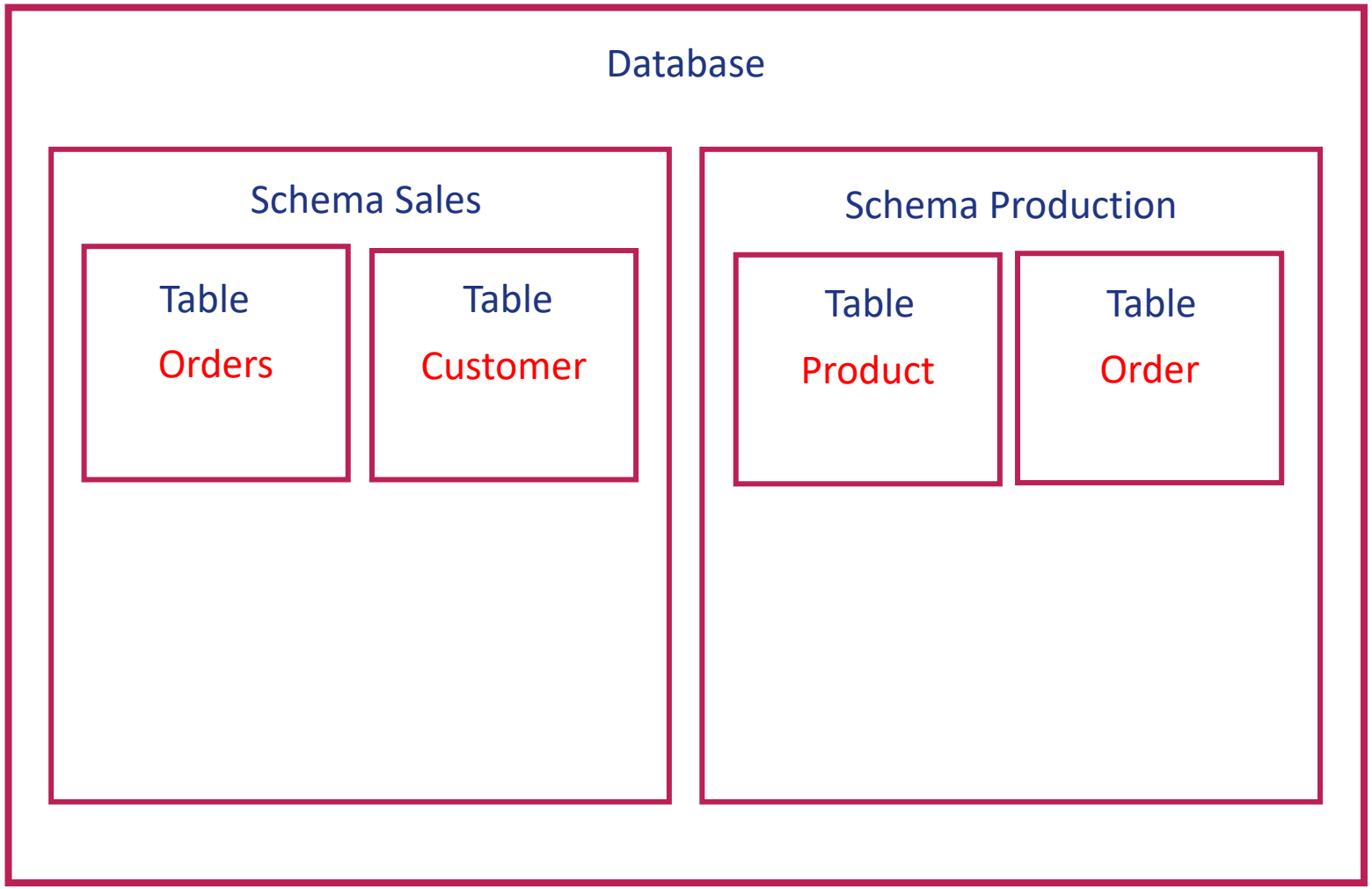
A set as "a collection of definite, distinct objects considered as a whole."

In terms applied to SQL Server databases, you can think of a set as a collection of distinct objects containing zero or more members of the same type.

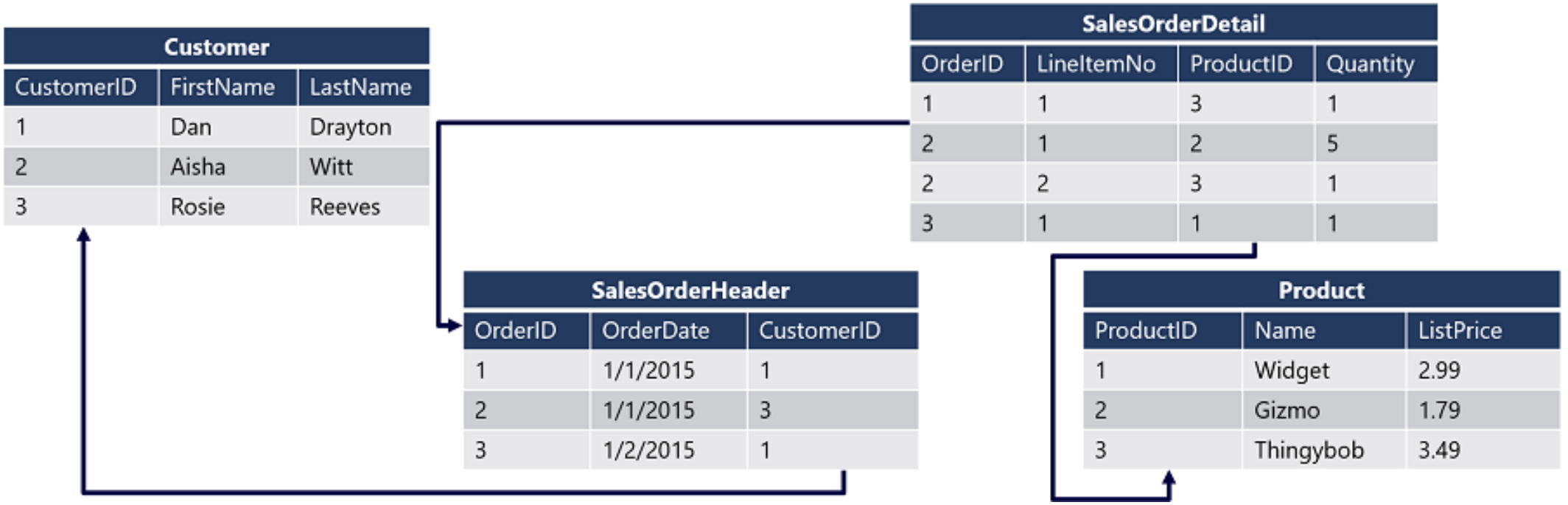
For example, the **Customer** table represents a set: specifically, the set of all customers.

There is no concept of a *first* row, a *second* row, or a *last* row.

You will see that the results of a SELECT statement also form a set.



Introduction to SQL: Relational Database



Customer

CustomerID	FirstName	LastName
1	Dan	Drayton
2	Aisha	Witt
3	Rosie	Reeves

SalesOrderDetail

OrderID	LineItemNo	ProductID	Quantity
1	1	3	1
2	1	2	5
2	2	3	1
3	1	1	1

SalesOrderHeader

OrderID	OrderDate	CustomerID
1	1/1/2015	1
2	1/1/2015	3
3	1/2/2015	1

Product

ProductID	Name	ListPrice
1	Widget	2.99
2	Gizmo	1.79
3	Thingybob	3.49

The diagram shows a relational database that contains four tables:
Customer | SalesOrderHeader | SalesOrderDetail | Product

Introduction to SQL: Important Concept

Technology / Concept	Description
The “Big 6” statements and clauses	SELECT, FROM, WHERE, GROUP BY, HAVING, and ORDER BY
“GROUP BY” aggregate functions	COUNT, COUNT(DISTINCT), MIN, MAX, AVG, and SUM
Query Multiple Tables using JOINS	INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL OUTER JOIN, and UNION

Introduction to SQL: Common T-SQL Data Types

Data Types	Specification	Storage
INT	INT(-2,147,483,648 to 2,147,483,647)	4 bytes
BIGINT	INT(-9,223,372,036,854,775,808) to (9,223,372,036,854,775,807)	8 bytes
FLOAT	Decimal (precise to 23 digits)	Depends on value of n
DECIMAL	DECIMAL Decimal (to 65 digits – most precise)	

Data Types	Specification	Storage
CHAR	String (0 – 255)	
VARCHAR	String (0 – 255)	
TEXT	String (0 – 65535)	
Date	yyyy-MM-dd	
Datetime2	yyyy-MM-dd HH:mm:ss[.nnnnnnn]	
Time	hh:mm:ss[.nnnnnnn]	

Module 2

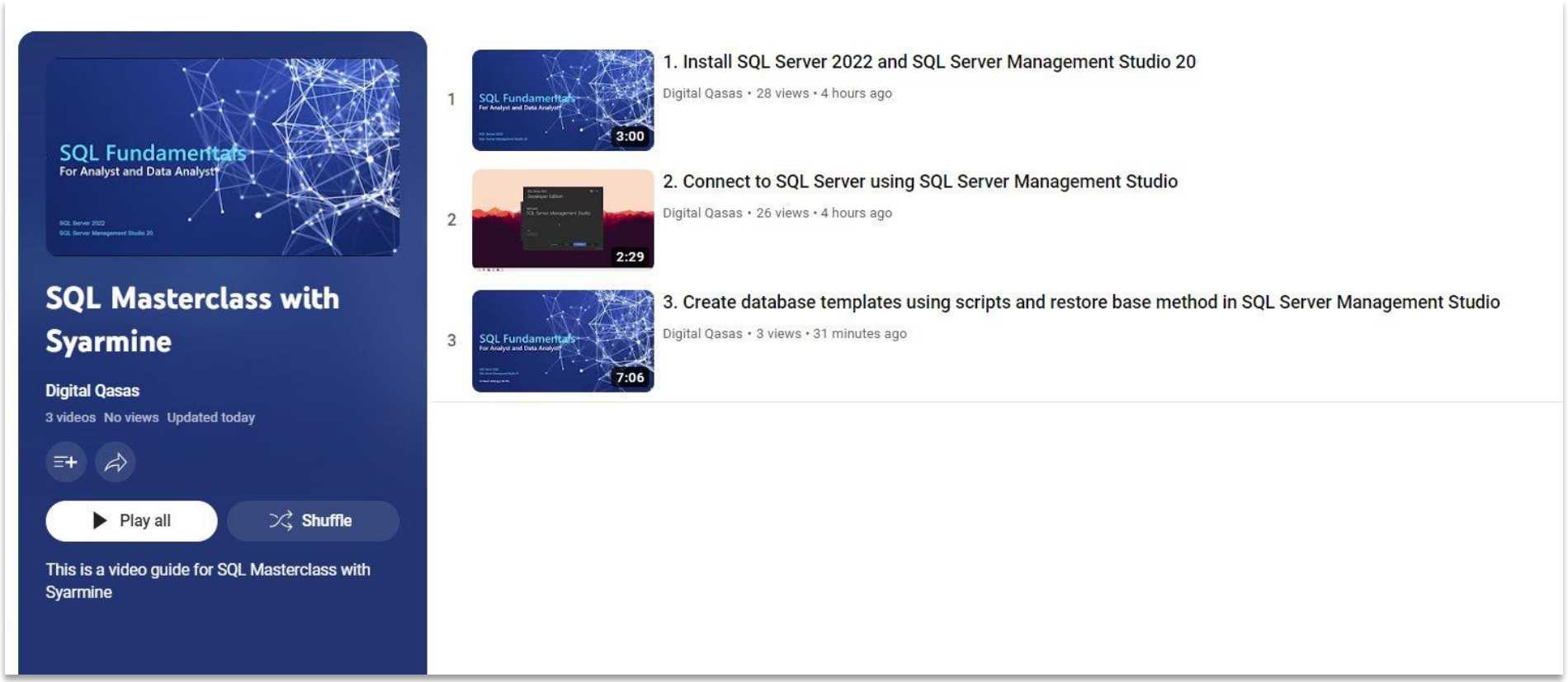
Setting SQL Server and SQL Server Management Studio

Setting up SQL Server (Developer) and SQL Server Management Studio

Learn how to install SQL Server Management Studio

- Install SQL Server 2022 (Developer):
  - Choose basic installation
- Install SQL Server Management Studio:
  - Installation instruction:
- SQL Documentation Navigation Tips
- Tools: SQL Server, SQL Server Management Studio, Excel, PowerBI (optional)
- PC Requirement: https://learn.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16

Youtube Playlist: Setting up SQL Server (Developer) and SQL Server Management Studio



[SQL Masterclass with Syarmine - YouTube]

Module 3

Creating a Database and Table

Creating a Database and Table - From Excel to SQL

Excel

PowerBI

SQL SSMS

Creating a Database and Table

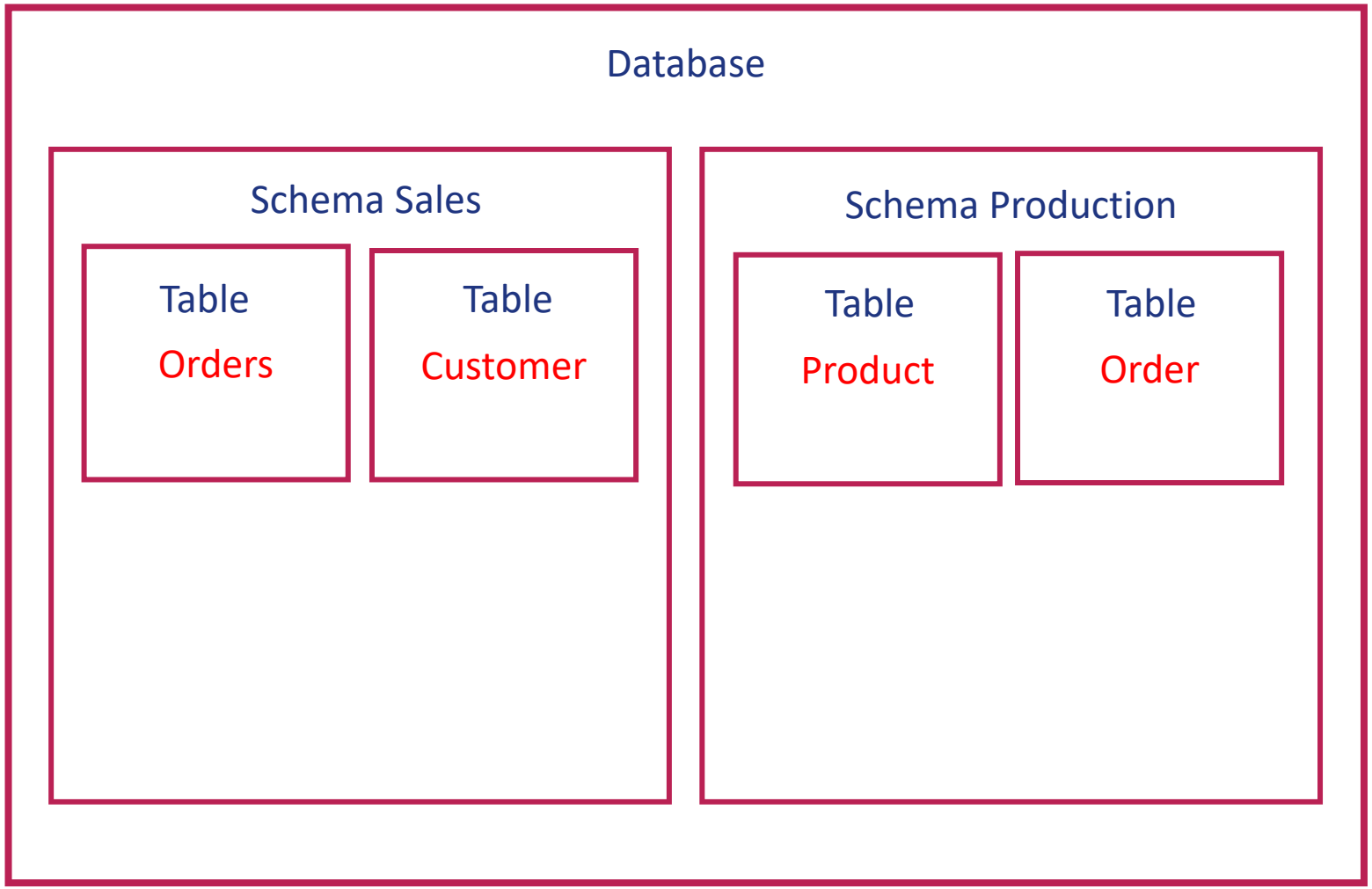
In SQL Server database systems, tables are defined within *schemas* to create logical namespaces in the database.

For example, a **Customer** table might be defined in a **Sales** schema, while a **Product** table is defined in a **Production** schema.

The database might track details of orders that customers have placed in an **Order** table in the **Sales** schema. You then might also need to track orders from suppliers for product components in an **Order** table in the **Production** schema.

Database systems such as SQL Server use a hierarchical naming system:

For example:
Server1.StoreDB.Sales.Order.



Different statements in SQL

In any SQL dialect, the SQL statements are grouped together into several different types of statements. These different types are:

- **Data Manipulation Language** (DML) is the set of SQL statements that focuses on querying and modifying data. DML statements include SELECT, the primary focus of this training, and modification statements such as INSERT, UPDATE, and DELETE. ← we focus on this
- **Data Definition Language** (DDL) is the set of SQL statements that handles the definition and life cycle of database objects, such as tables, views, and procedures. DDL includes statements such as CREATE, ALTER, and DROP. ← and this
- **Data Control Language** (DCL) is the set of SQL statements used to manage security permissions for users and objects. DCL includes statements such as GRANT, REVOKE, and DENY.

DML statements are also used by application developers to perform "CRUD" operations to create, read, update, or delete application data.

Deconstructing SQL statement

```
SELECT OrderDate, COUNT(OrderID) AS Orders
FROM Sales.Order
WHERE Status = 'Shipped'
GROUP BY OrderDate
HAVING COUNT(OrderID) > 1
ORDER BY OrderDate DESC;
```

Deconstructing SQL statement

Column(s)

```
SELECT OrderDate, COUNT(OrderID) AS Orders
FROM Sales.Order
WHERE Status = 'Shipped'
GROUP BY OrderDate
HAVING COUNT(OrderID) > 1
ORDER BY OrderDate DESC;
```

Table

Deconstructing SQL statement

SQL STATEMENT	REMARK
SELECT OrderDate, COUNT(OrderID) AS Orders	The SELECT clause returns the **OrderDate** column, and the count of **OrderID** values, to which is assigns the name (or *alias*) **Orders**:
FROM Sales.Order	The FROM clause identifies which table is the source of the rows for the query; in this case it's the **Sales.SalesOrder** table:
WHERE Status = 'Shipped'	The WHERE clause filters rows out of the results, keeping only those rows that satisfy the specified condition; in this case, orders that have a status of "shipped":
GROUP BY OrderDate	The GROUP BY clause takes the rows that met the filter condition and groups them by **OrderDate**, so that all the rows with the same **OrderDate** are considered as a single group and one row will be returned for each group:
HAVING COUNT(OrderID) > 1	After the groups are formed, the HAVING clause filters the groups based on its own predicate. Only dates with more than one order will be included in the results:
ORDER BY OrderDate DESC;	For the purposes of previewing this query, the final clause is the ORDER BY, which sorts the output into descending order of **OrderDate**:

The “Big 6” Elements of SQL SELECT Statement

START OF STATEMENT

SELECT	Identifies the column(s) you want your	SELECT columnName
FROM	Identifies the table(s) your query will pull data from	FROM tableName
WHERE	(Optional) Specifies record-filtering criteria for filtering your results	WHERE logicalCondition
GROUP BY	(Optional) Specifies how to group the data in your results	GROUP BY columnName
HAVING	(Optional) Specifies group-filtering criteria for filtering your results	HAVING logicalCondition
ORDER BY	(Optional) Specifies the order in which your query results are displayed	ORDER BY columnName

END OF STATEMENT

Execution of SQL SELECT Statement

START OF STATEMENT

FROM	The FROM clause is evaluated first, to provide the source rows for the rest of the statement. A virtual table is created and passed to the next step.
WHERE	The WHERE clause is next to be evaluated, filtering those rows from the source table that match a predicate. The filtered virtual table is passed to the next step.
GROUP BY	GROUP BY is next, organizing the rows in the virtual table according to unique values found in the GROUP BY list. A new virtual table is created, containing the list of groups, and is passed to the next step. From this point in the flow of operations, only columns in the GROUP BY list or aggregate functions may be referenced by other elements.
HAVING	The HAVING clause is evaluated next, filtering out entire groups based on its predicate. The virtual table created in step 3 is filtered and passed to the next step.
SELECT	The SELECT clause finally executes, determining which columns will appear in the query results. Because the SELECT clause is evaluated after the other steps, any column aliases (in our example, **Orders**) created there cannot be used in the GROUP BY or HAVING clause.
ORDER BY	The ORDER BY clause is the last to execute, sorting the rows as determined by its column list.

END OF STATEMENT

Writing a readable T-SQL query

Consider the following guidelines to make your T-SQL code easily readable (and therefore easier to understand and debug!):

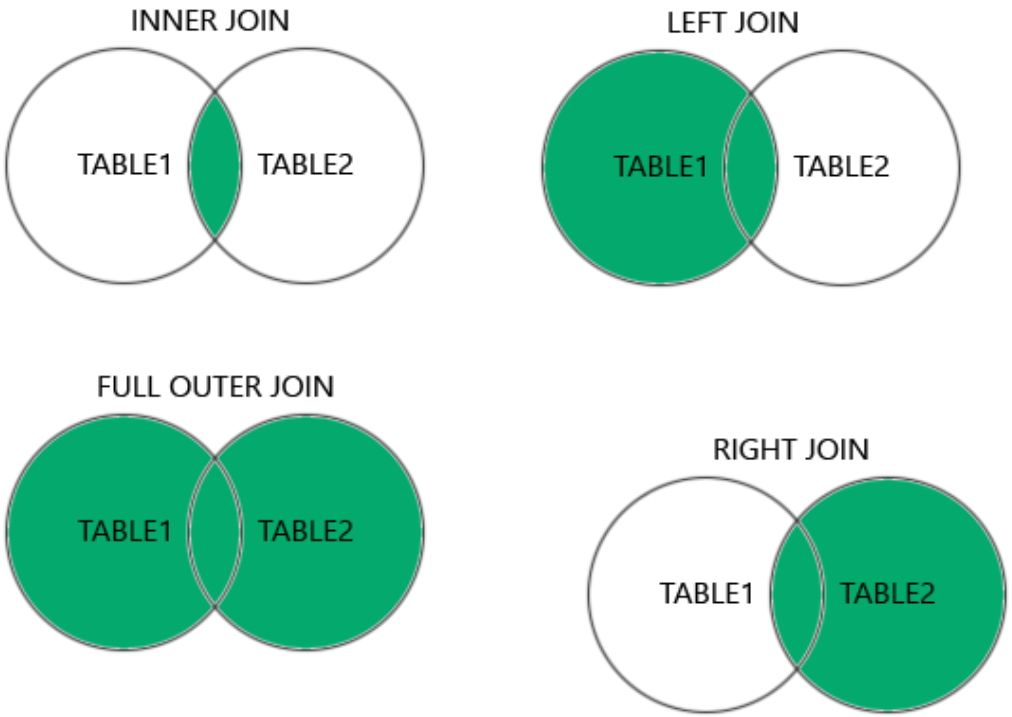
- Capitalize T-SQL keywords, like SELECT, FROM, AS, and so on. Capitalizing keywords is a commonly used convention that makes it easier to find each clause of a complex statement.
- Start a new line for each major clause of a statement.
- If the SELECT list contains more than a few columns, expressions, or aliases, consider listing each column on its own line.
- Indent lines containing subclauses or columns to make it clear which code belongs to each major clause

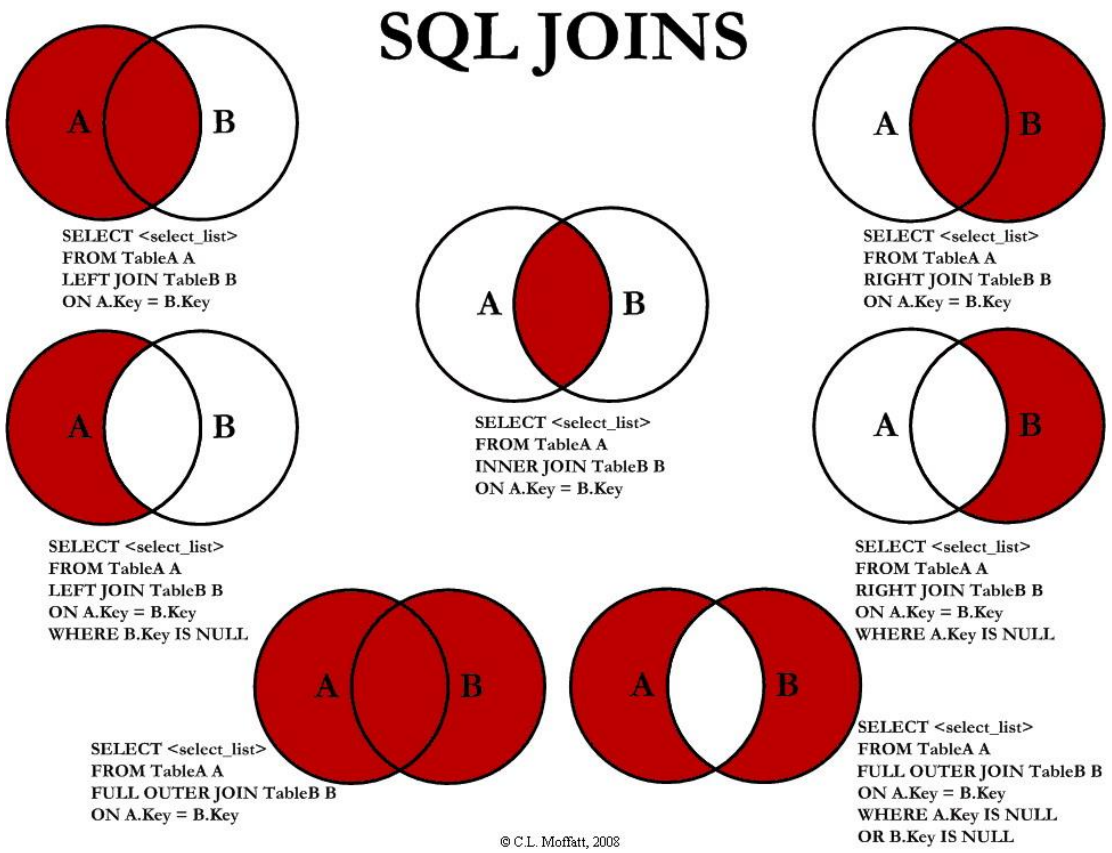
```
SELECT FirstName
FROM Sales
```

```
SELECT
    FirstName,
    ID,
    Age
FROM Sales
ORDER BY Age
```

```
SELECT
    FirstName,
    ID,
    Age
FROM Sales
ORDER BY Age
```

Database Join





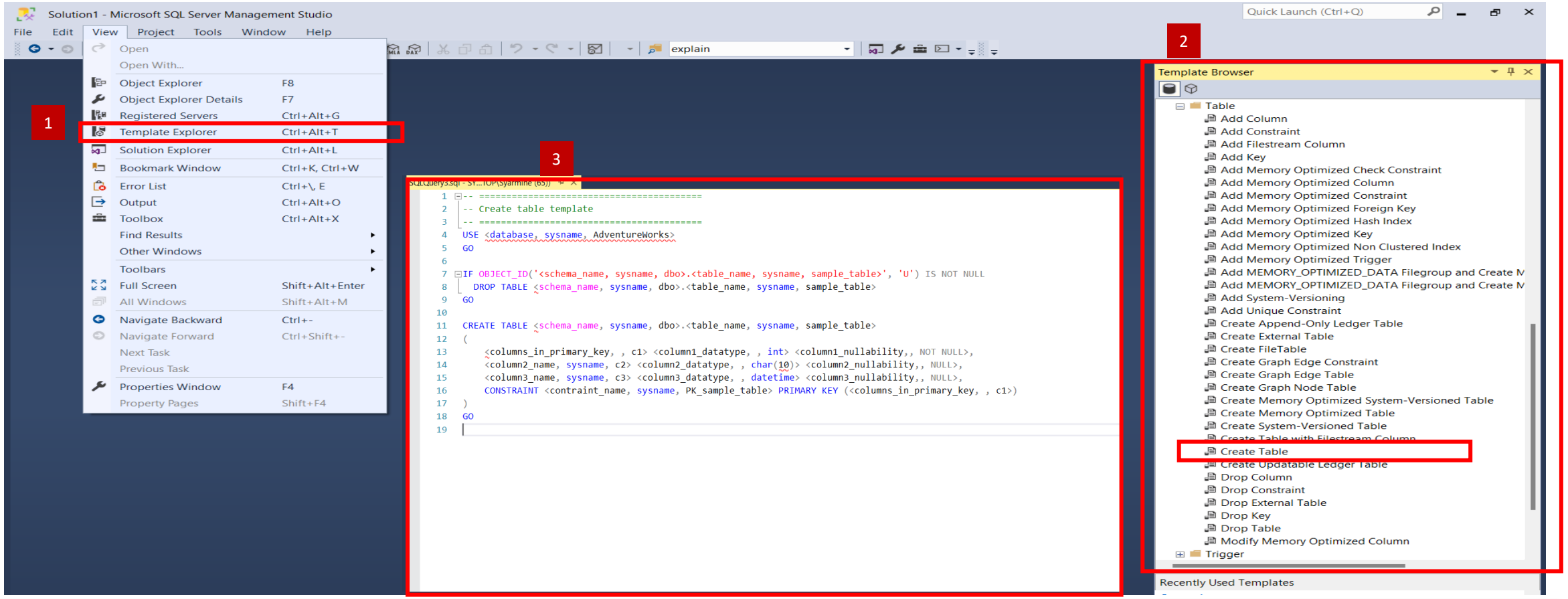
Creating a Database and Table - From Excel to SQL

Create Database | **Create Table** | **Insert Data** | **Query Data** | **More**

- Select
- Where
- Update
- Delete
- Alter – new column, update data

- Primary Key
- More Tables
- Foreign Key
- Index
- JOINS
- Functions
- GROUP BY

Extra: Use TEMPLATES in SQL Server Management Studio



1. Select “Template Explorer”
2. Select available templates. For example, select “Create table”
3. Sample of “Create Table” template

Extra: Use TEMPLATES in SQL Server Management Studio

- There are pre-built Transact-SQL Templates that are available in SSMS.
- Use Template Browser to locate the template. You can:
  - Open a template
    - Create a database
    - Create a table
  - Edit a template
  - Locate the template on disk
  - Save the template
  - Save custom template
- Read more here: https://learn.microsoft.com/en-us/sql/ssms/template/templates-ssms?view=sql-server-ver16

Extra: Use VIEWS in SQL Server Management Studio

- A view is a virtual table whose contents are defined by a query. Like a table, a view consists of a set of named columns and rows of data.
- Views are generally used to focus, simplify, and customize the perception each user has of the database. Unless indexed, a view does not exist as a stored set of data values in a database.
- A view acts as a filter on the underlying tables referenced in the view. The query that defines the view can be from one or more tables or from other views in the current or other databases (i.e using joins)
- Security: Views can be used as security mechanisms by letting users access data through the view, without granting the users permissions to directly access the underlying base tables of the view.
- Types of view: Indexed view, Partitioned view, System view
- The CREATE VIEW must be the first statement in a query batch. A view can have a maximum of 1,024 columns.

Extra: Use VIEWS in SQL Server Management Studio

```
CREATE VIEW view_name

AS

SELECT statement
```

Extra: Use SUBQUERIES in SQL Server Management Studio

- A subquery is a SELECT statement nested within another query.
- The nested query, which is the subquery, is referred to as the inner query. The query containing the nested query is the outer query.
- The purpose of a subquery is to return results to the outer query. The form of the results will determine whether the subquery is a scalar or multi-valued subquery:
  - Scalar subqueries return a single value. Outer queries must process a single result.
  - Multi-valued subqueries return a result much like a single-column table. Outer queries must be able to process multiple values.

Extra: Use SUBQUERIES in SQL Server Management Studio

SCALAR SUBQUERIES

```
SELECT
        SalesOrderID,
        ProductID,
        OrderQty
FROM
        Sales.SalesOrderDetail
WHERE  SalesOrderID =
        (SELECT MAX(SalesOrderID)
        FROM Sales.SalesOrderHeader);
```

MULTI-VALUED SUBQUERIES

```
SELECT
        c.CustomerID,
        o.SalesOrderID
FROM
        Sales.Customer AS c
JOIN
        Sales.SalesOrderHeader AS o ON
        c.CustomerID = o.CustomerID
WHERE
        c.CountryRegion = 'Canada';
```

Module 4

Understanding scripts, restore a database and templates for database creation

Understanding scripts and database restore from popular databases

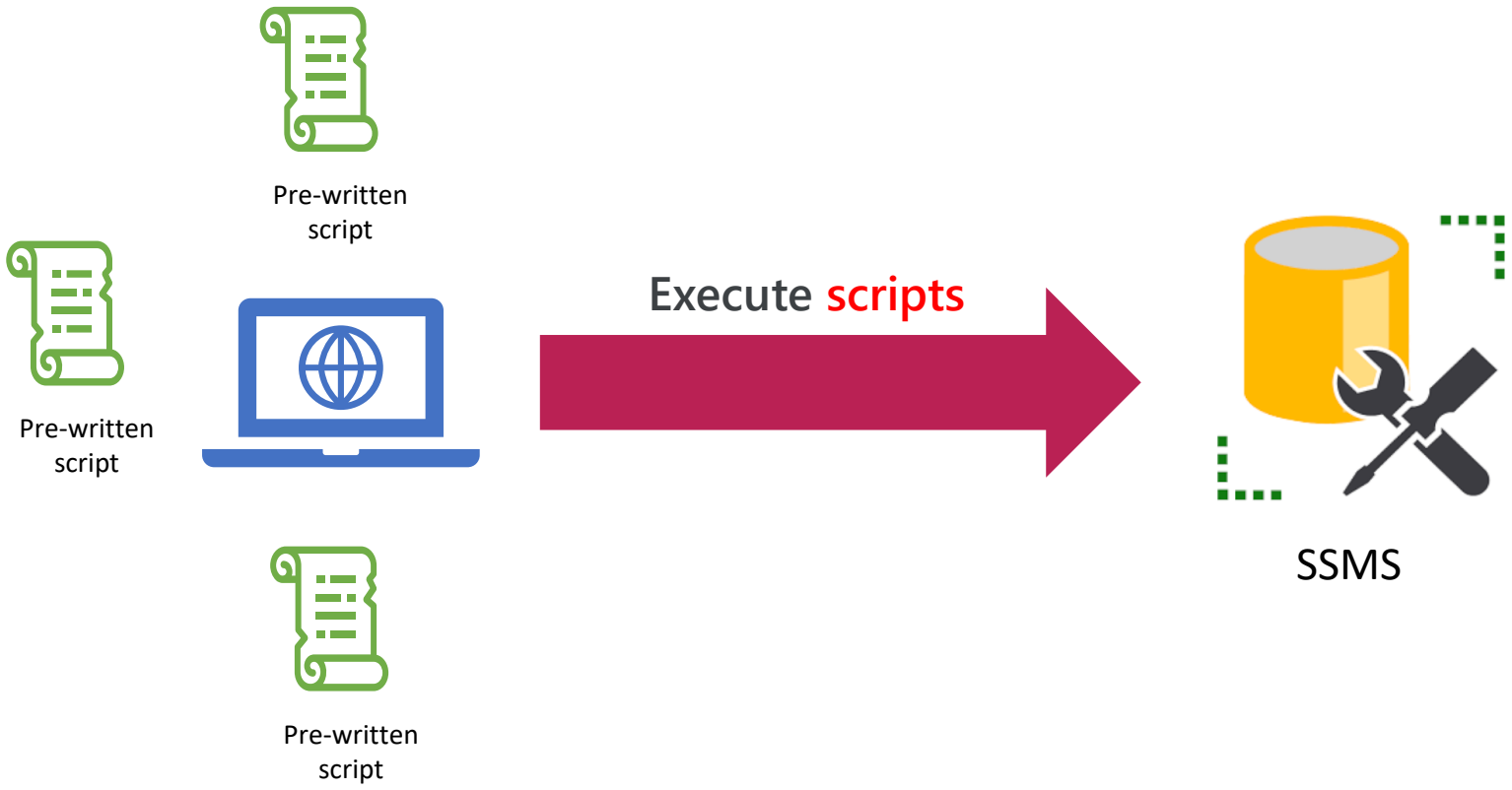
- Go to this Github repositories and download the sample databases:
  - Adventureworks
  - Contoso
  - Northwind
  - Wide World Importers
- You can replicate these databases by installing the database or installing from a backup.

- Read more here: https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver16&tabs=ssms

Understanding scripts from popular databases

- Installing AdventureWorks
  - Download AdventureWorks-oltp-install-script.zip and extract the zip file to the C:\Samples\AdventureWorks folder.
  - Open C:\Samples\AdventureWorks\instawdb.sql in SQL Server Management Studio and follow the instructions at the top of the file.
  - *Enable SQLCMD mode via Query tab*
- Run a pre-written script (i.e by content creator or database administrator) within SSMS

Execute scripts from safe and popular databases

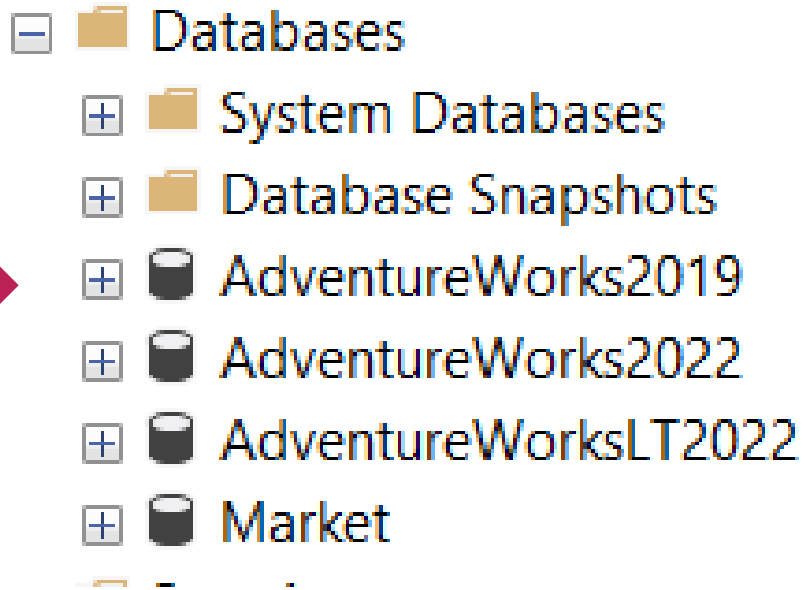


Example: Execute scripts from AdventureWorks database

```
1  ================================================================================
2  File:     instawdb.sql
3
4  Summary:  Creates the AdventureWorks sample database. Run this on
5  any version of SQL Server (2008R2 or later) to get AdventureWorks for your
6  current version.
7
8  Date:     October 26, 2017
9  Updated:  October 26, 2017
10
11 --------------------------------------------------------------------------------
12 This file is part of the Microsoft SQL Server Code Samples.
13
14 Copyright (C) Microsoft Corporation.  All rights reserved.
15
16 This source code is intended only as a supplement to Microsoft
17 Development Tools and/or on-line documentation.  See these other
18 materials for detailed information regarding Microsoft code samples.
19
20 All data in this database is ficticious.
21
22 THIS CODE AND INFORMATION ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY
23 KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE
24 IMPLIED WARRANTIES OF MERCHANTABILITY AND/OR FITNESS FOR A
25 PARTICULAR PURPOSE.
26 ================================================================================*/
27
28
29  HOW TO RUN THIS SCRIPT:
30
31  1. Enable full-text search on your SQL Server instance.
32
33  2. Open the script inside SQL Server Management Studio and enable SQLCMD mode.
34     This option is in the Query menu.
35
36  3. Copy this script and the install files to C:\Samples\AdventureWorks, or
37     set the following environment variable to your own data path
```

AdventureWorks script leading to creation of AdventureWorks database

Execute scripts



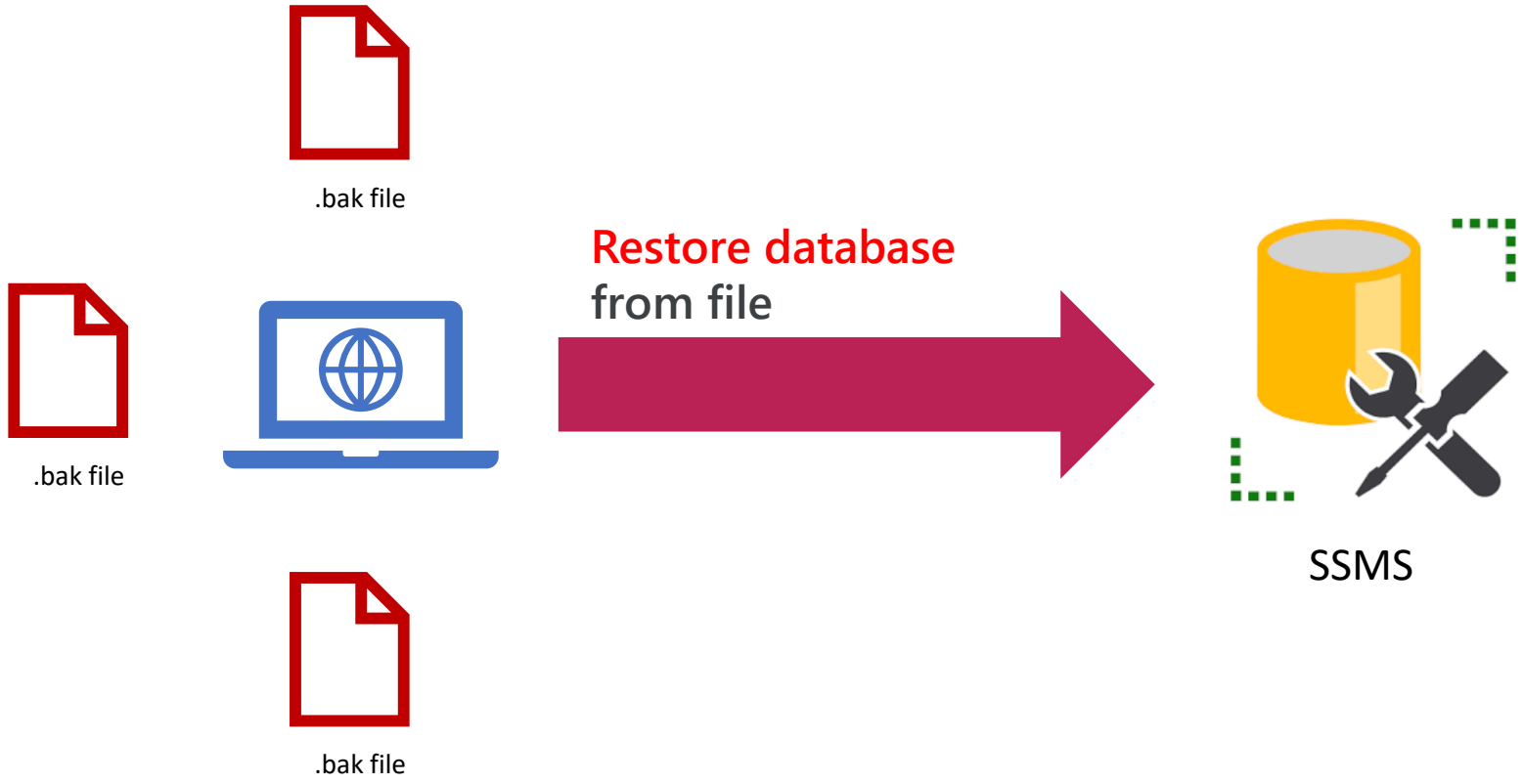
AdventureWorks database created

Understanding database restore from popular databases

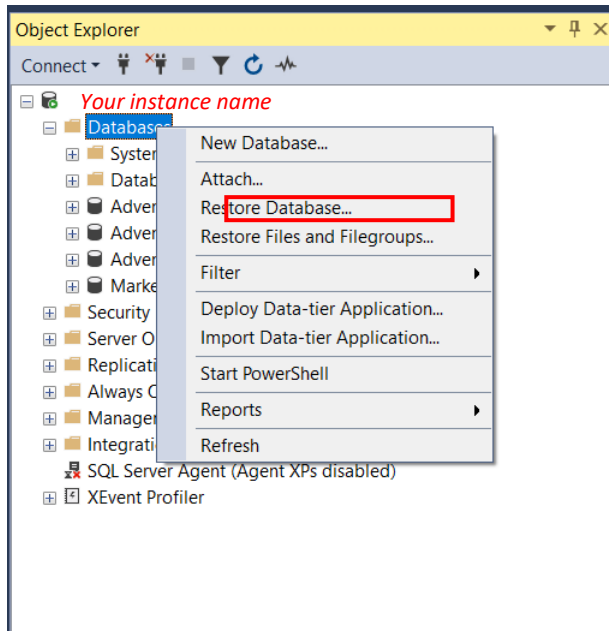
Install a database from database backup

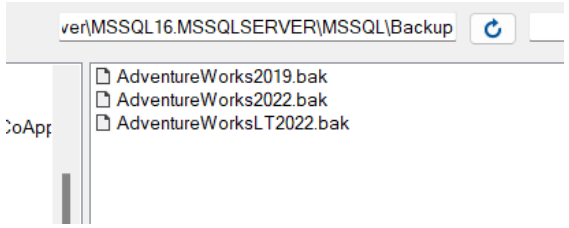
- Locate the Backup folder for your SQL Server instance. The default path for 64-bit SQL Server 2016 is C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Backup. The MSSQL value is MSSQL14 for SQL Server 2017, MSSQL13 for SQL Server 2016, MSSQL12 for SQL Server 2014, MSSQL11 for SQL Server 2012, and MSSQL10 for SQL Server 2008R2.
- Download the .bak file from AdventureWorks release and save it to the Backup folder for your SQL Server instance.
- Open SQL Server Management Studio and connect to your SQL Server instance.
- Restore the database using the SQL Server Management Studio user interface. For more information, see Restore a database backup using SSMS.
- Or, run the RESTORE DATABASE command in a new query Window. On the Standard toolbar, click the New Query button.

Restore database using file from safe and popular databases

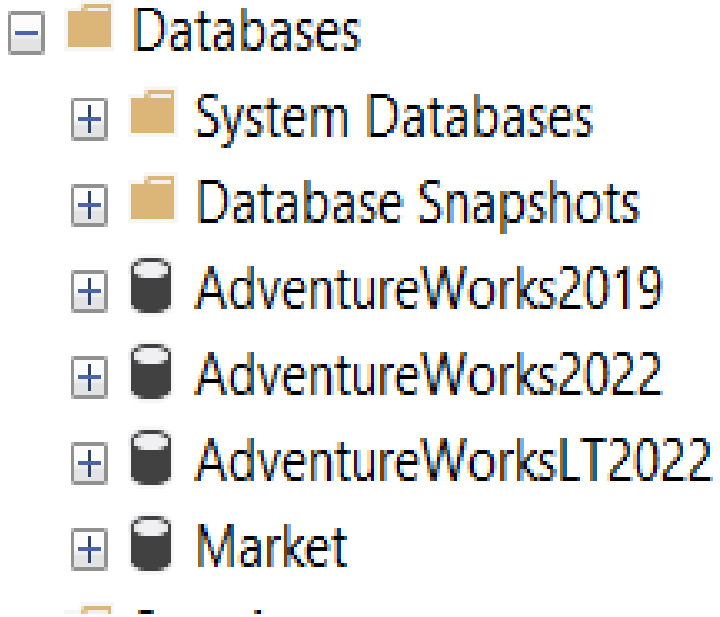


Example: Restore database using AdventureWorks .bak file





Restore database using **file**



Module 5

Connecting SQL Server to Excel and PowerBI

Connect to the Database Engine

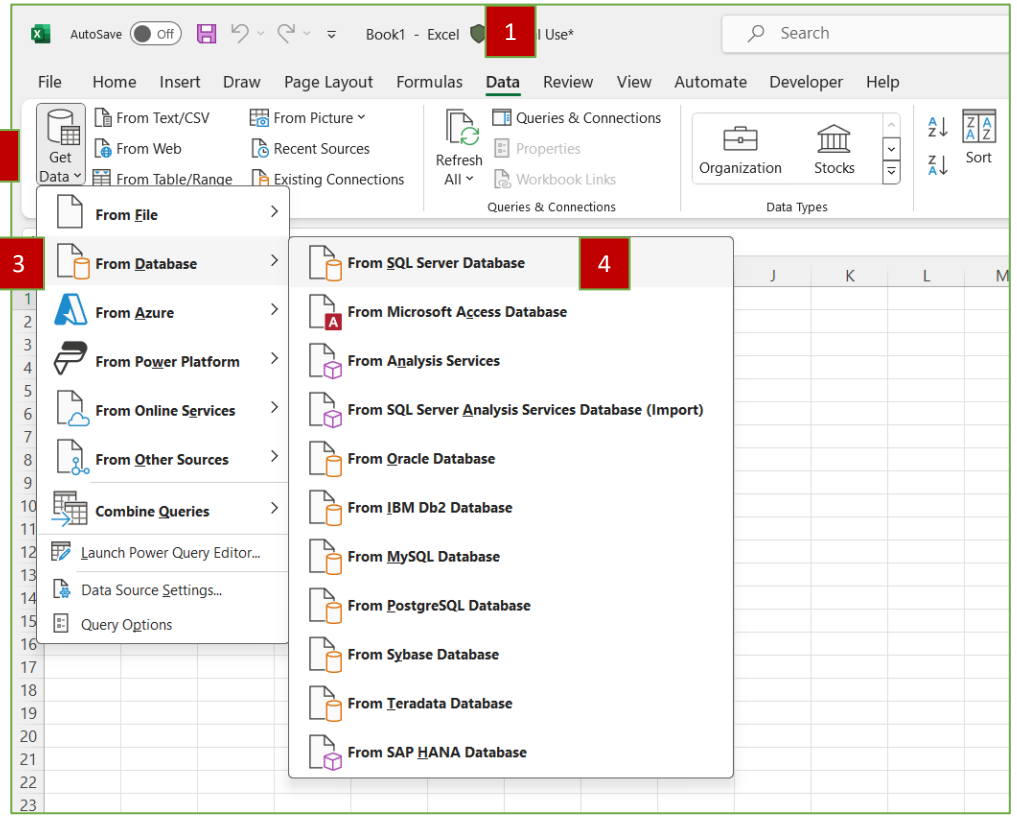
- **Connection options:** When you connect to the Database Engine, you must provide an instance name (that is, the server or instance where the Database Engine is installed), a network protocol, and a connection port, in the following format:

- **Connect to SQL Server on the same machine as the client (this session):** If you're connecting to a server configured with default settings, use one of the following options:
  - localhost
  - 127.0.0.1
  - . (a single period)
- Use get data from Excel or PowerBI and connect in Tableau

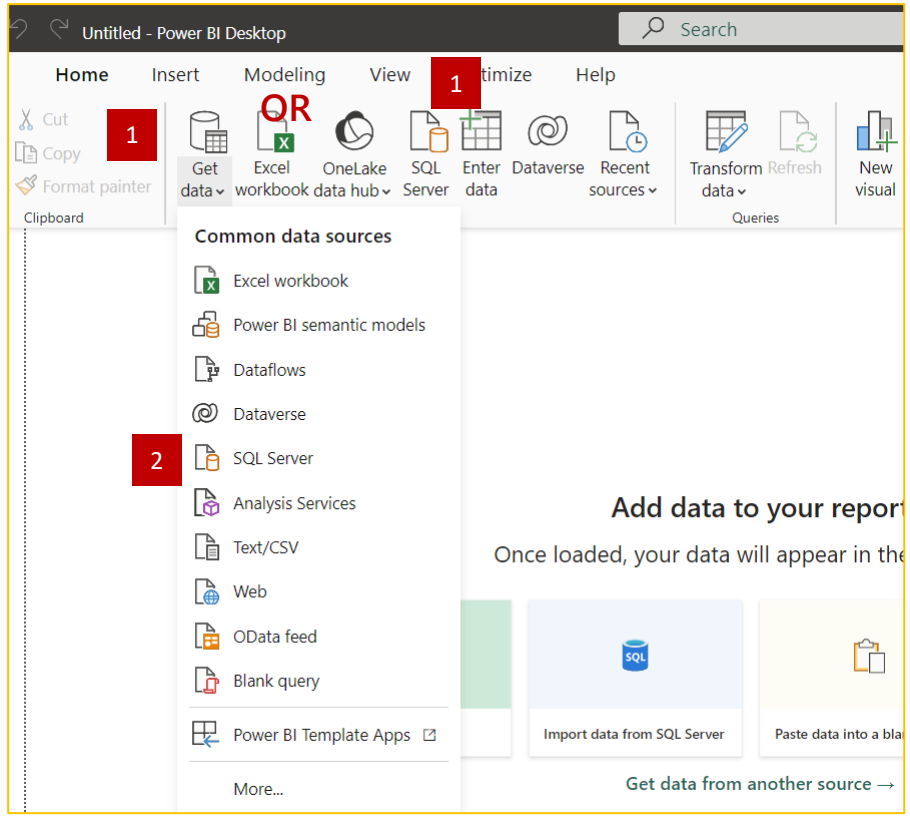
Excel PowerBI

Connect to the Database Engine

Excel



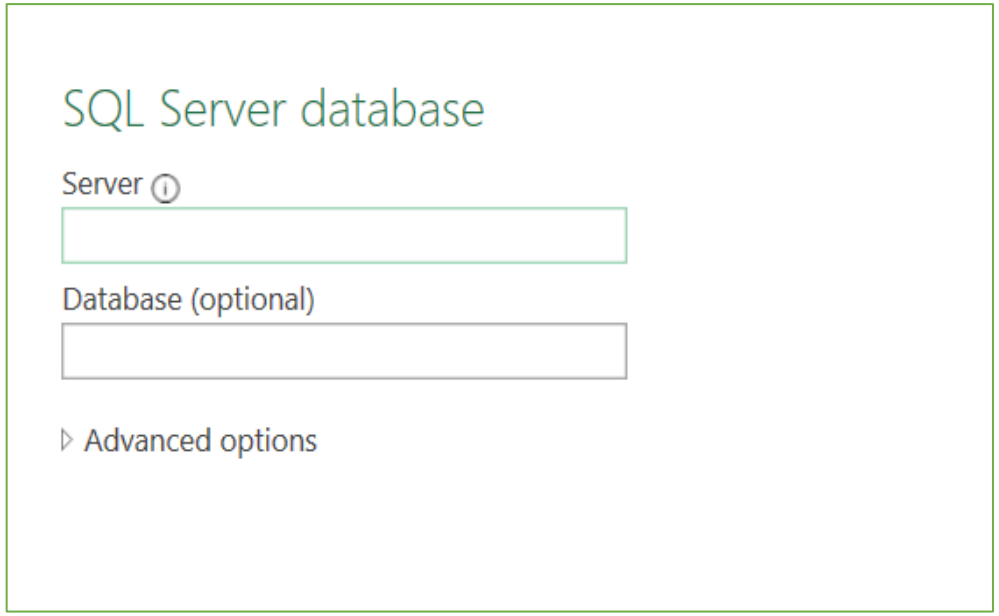
PowerBI



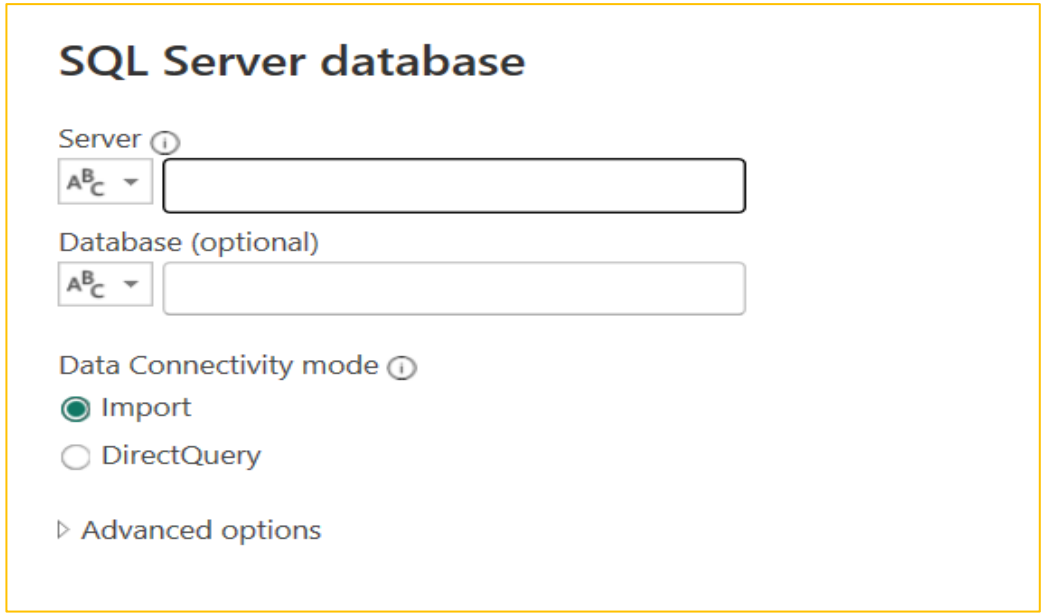
Excel PowerBI

Connect to the Database Engine

Excel



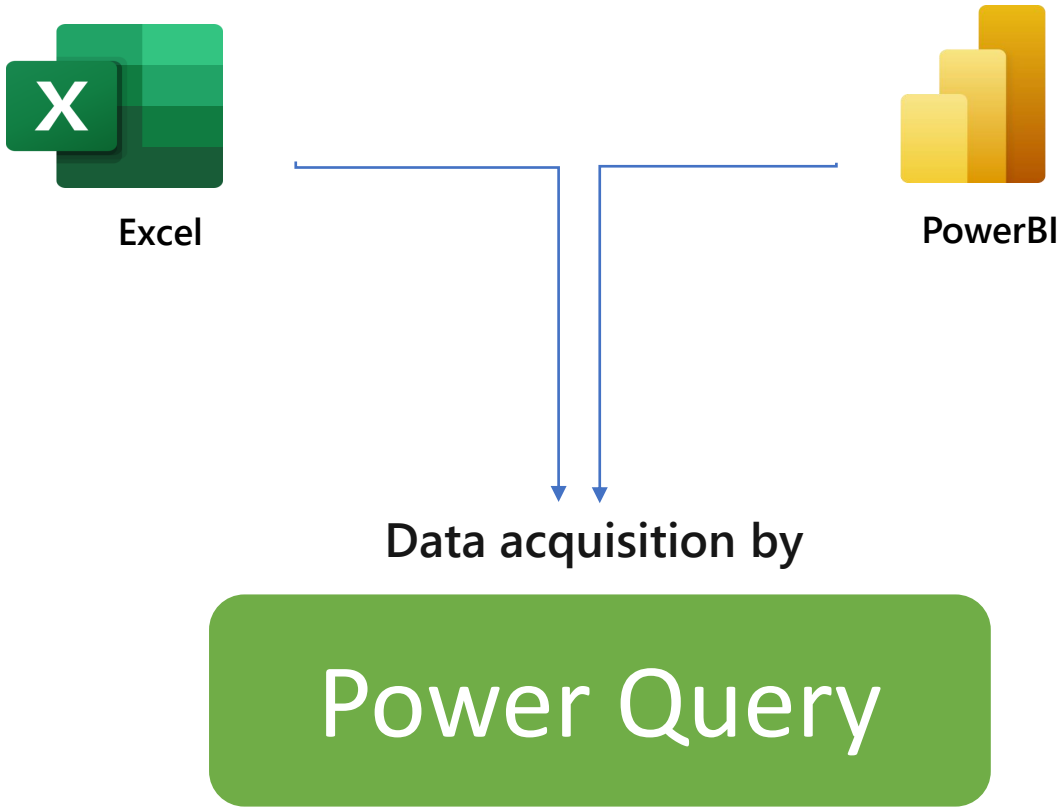
PowerBI



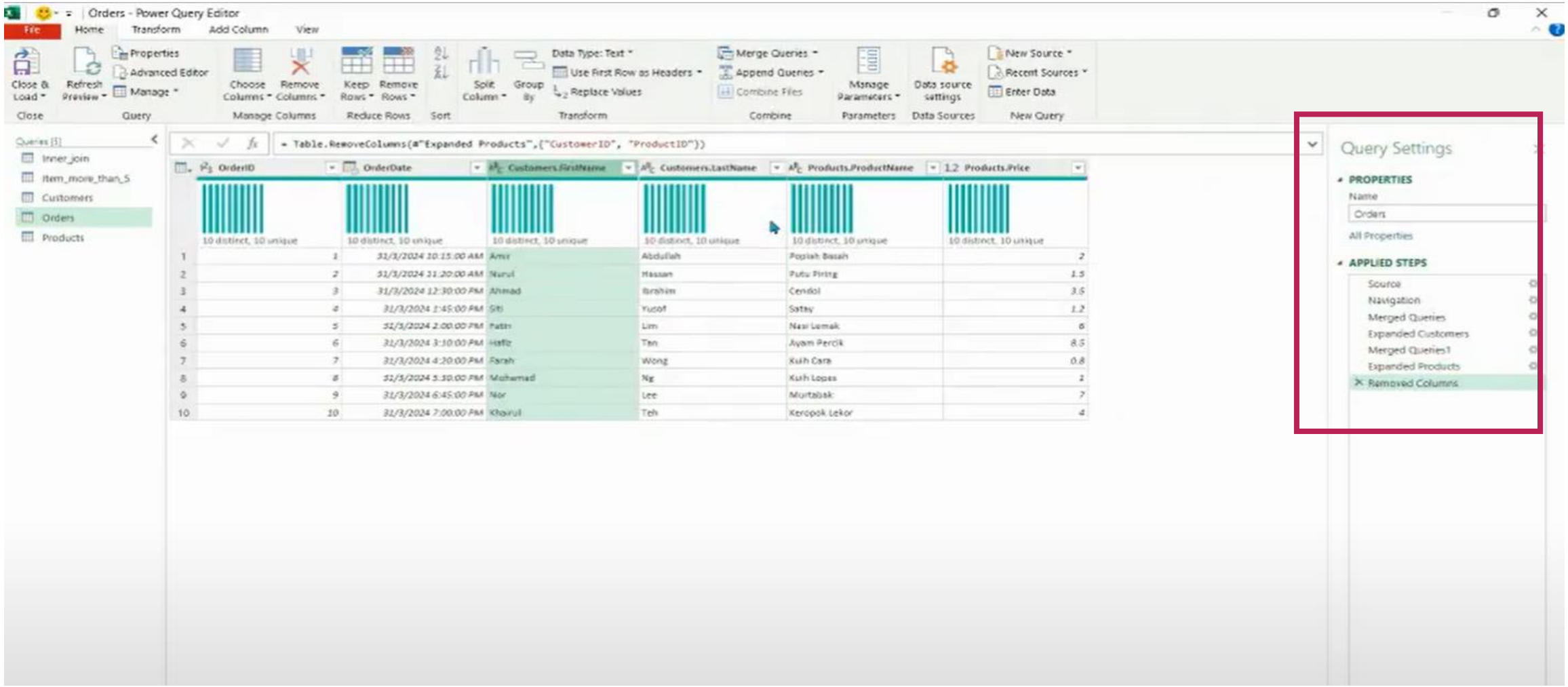
Server name: If you use local server, can use *localhost*, *. Or ServerName*
Database (optional): You now may use AdventureWorks2019, AdventurWorks2022 or AdventureWorksLT2022 or Market

Excel PowerBI

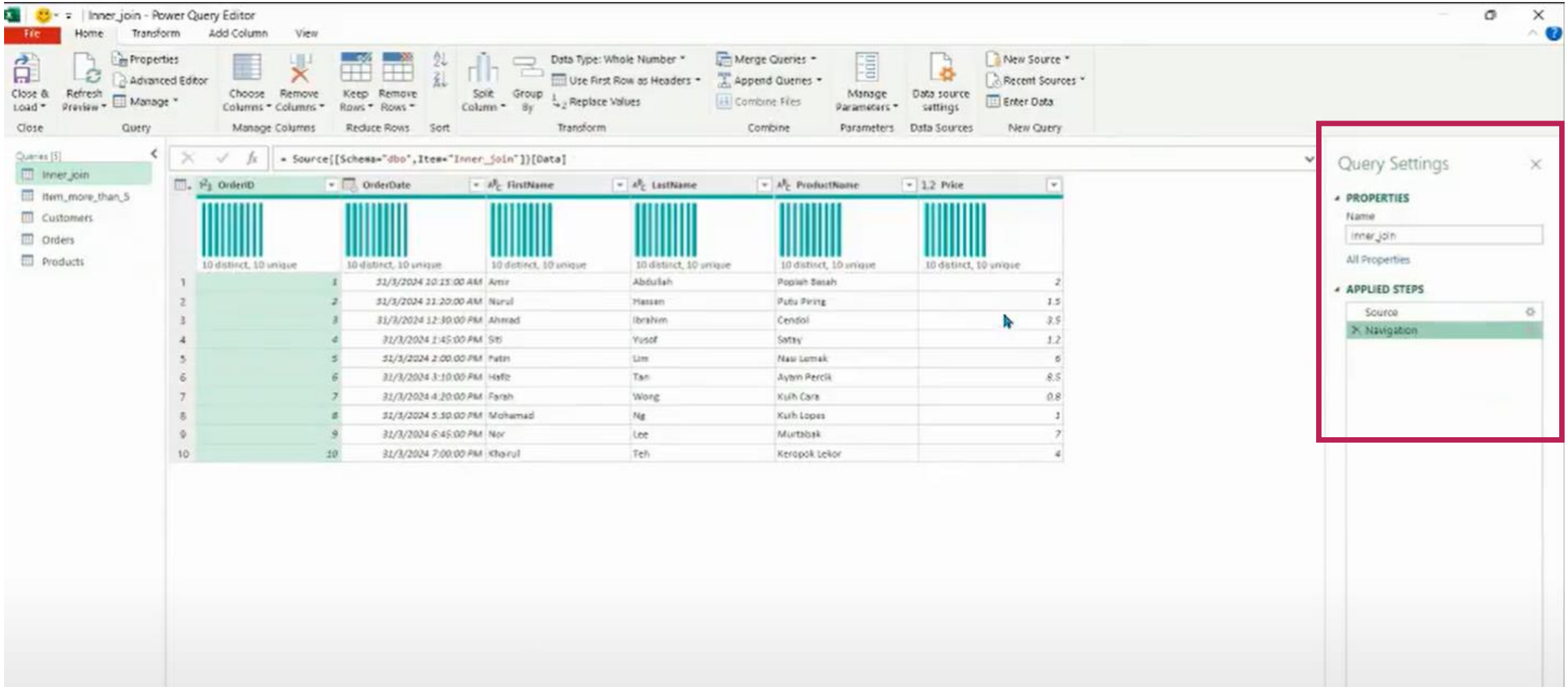
Connect to the Database Engine



Data transformation: Using Power Query

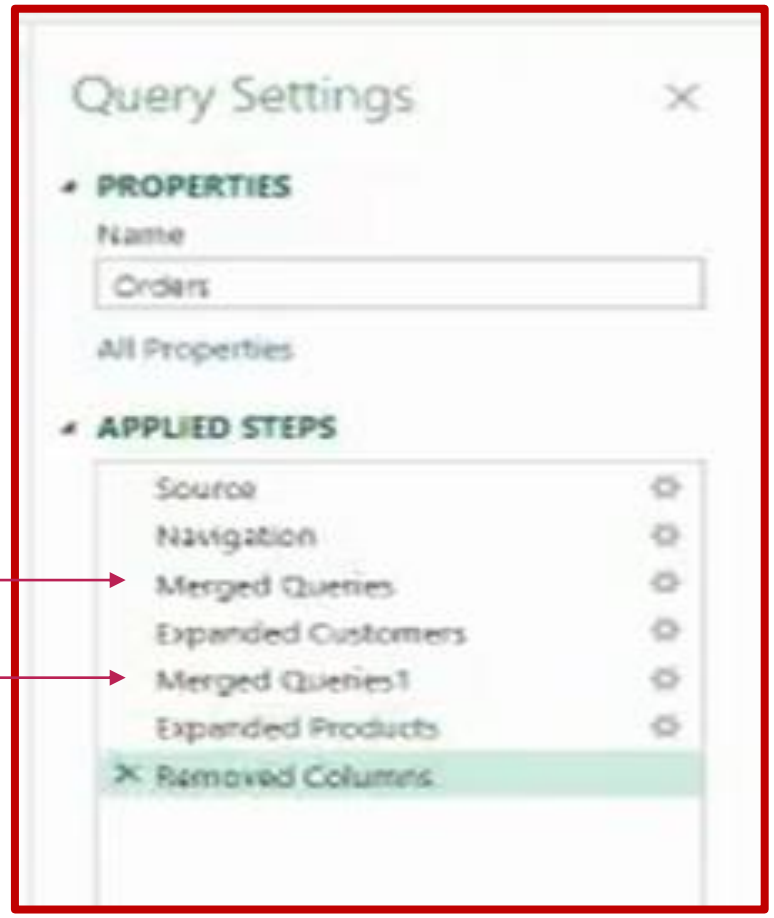


Data transformation: Fetch transformed data from SQL



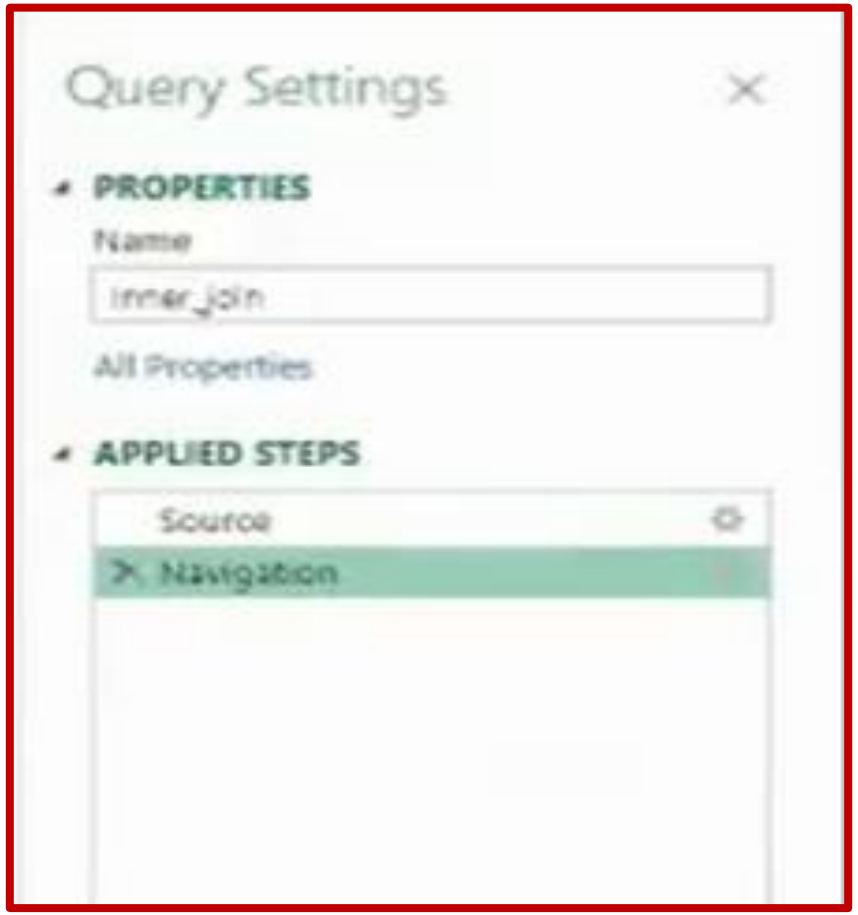
What is the difference?

All transformation inside Power Query



Merge happened 2x

All transformation happened prior SQL, only need to load the table / view



Less step → improved performance

The process in the background…

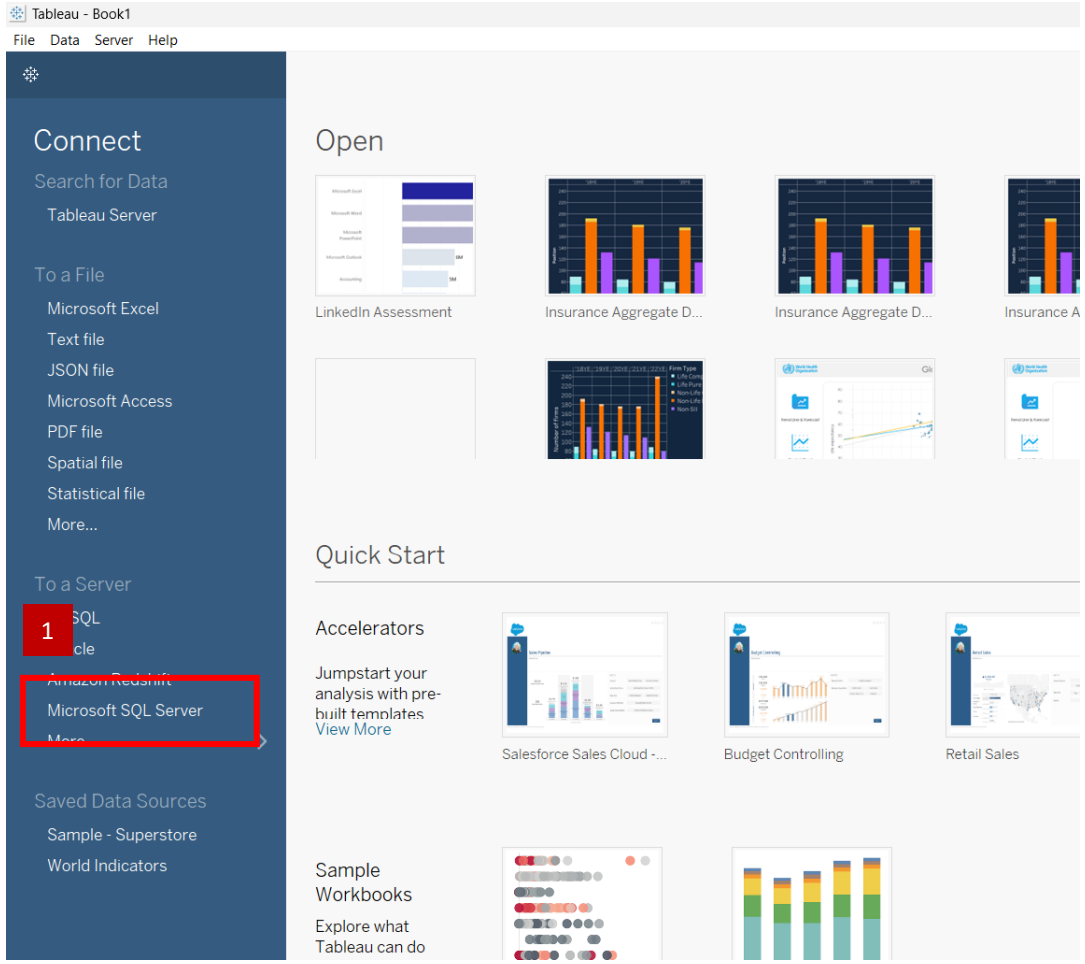
Excel / PowerBI: Recorded automatically as M-language as you click in Power Query

SSMS (T-SQL): More control, get the output set based on your input, happened BEFORE pre-loaded in Power Query.

```
--JOINS
--INNER JOINS
--An inner join will return rows where there is at least one match in both tables.
SELECT Orders.OrderID, Orders.OrderDate, Customers.FirstName, Customers.LastName, Products.ProductName, Products.Price
FROM Orders
INNER JOIN Customers ON Orders.CustomerID = Customers.CustomerID
INNER JOIN Products ON Orders.ProductID = Products.ProductID;
```

Tableau

Connect to the Database Engine (Tableau)



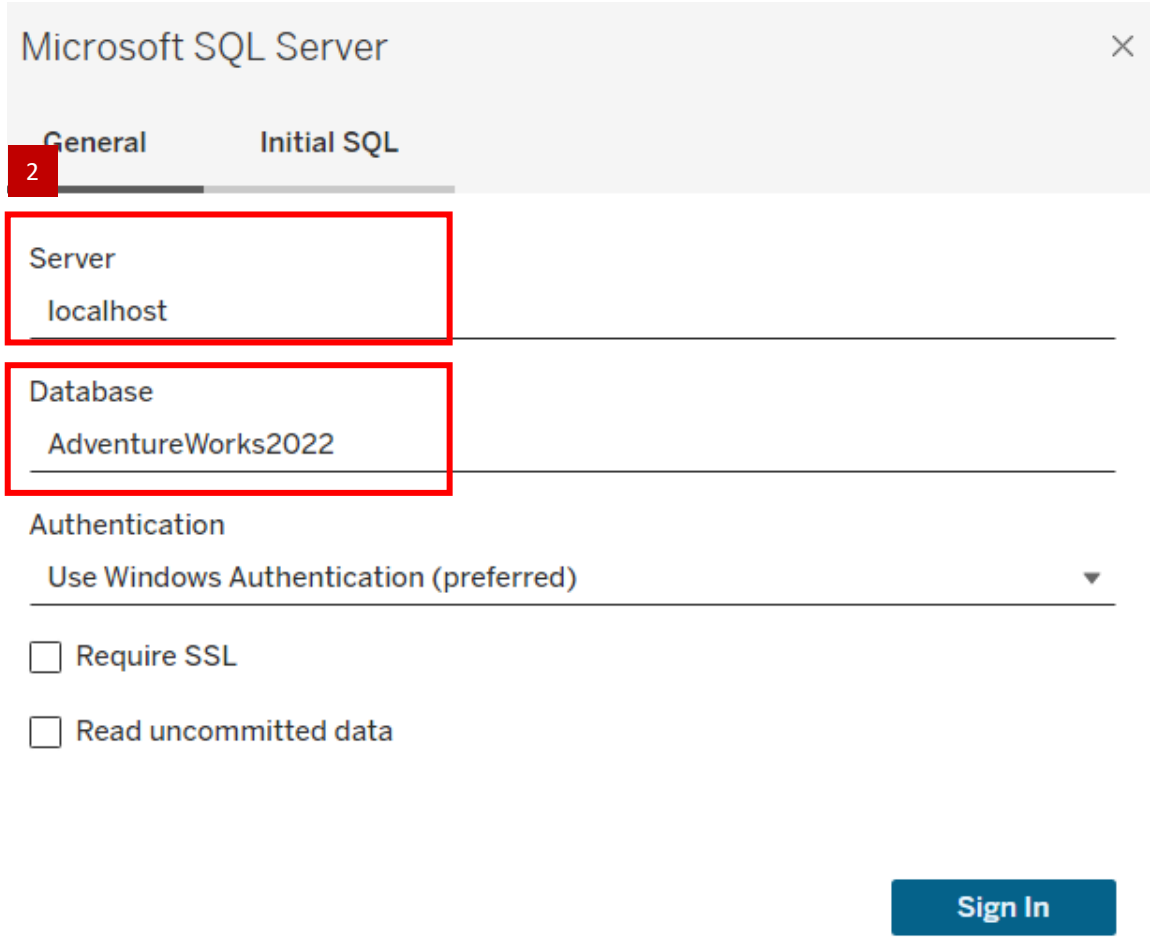
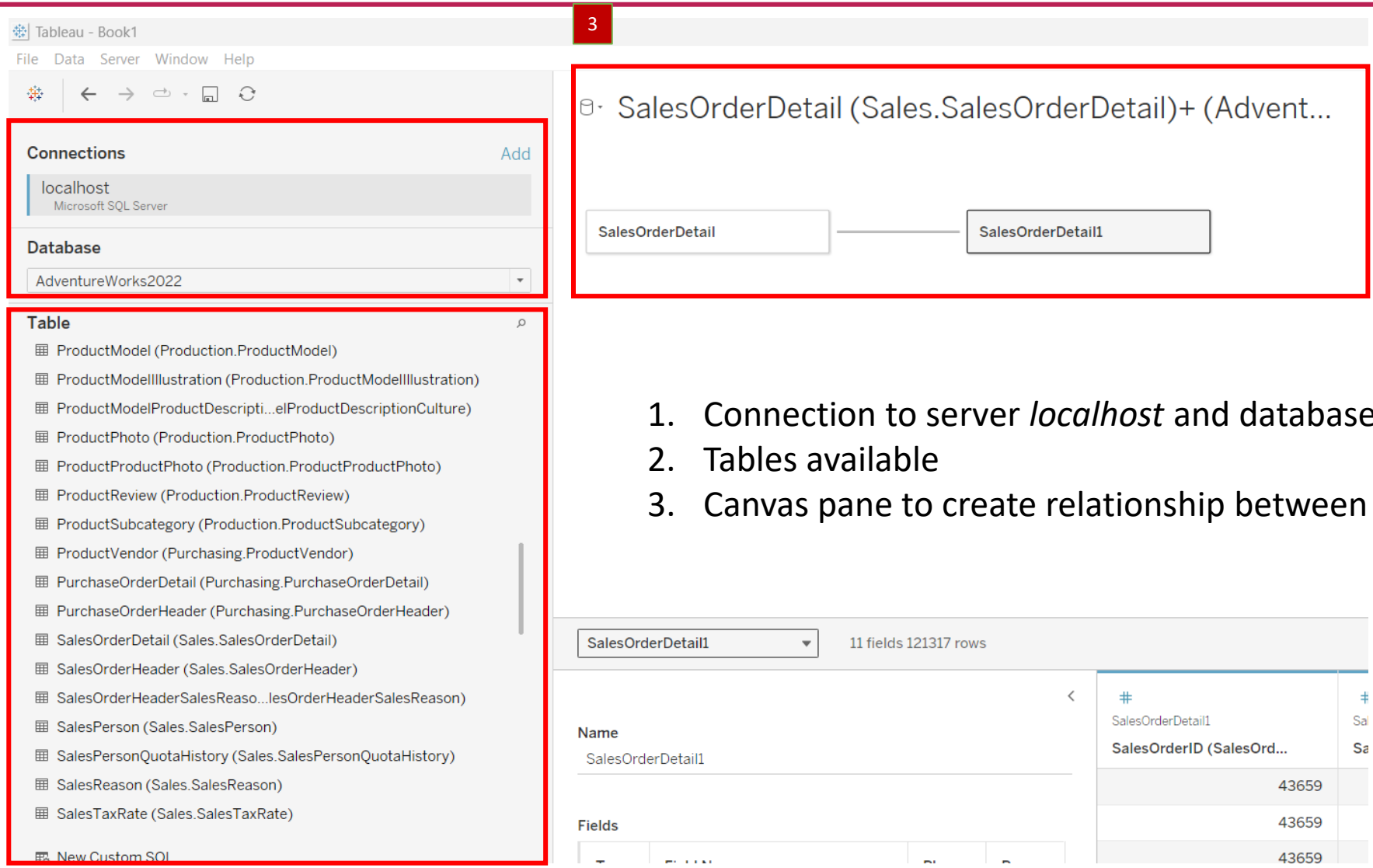


Tableau Data Source Pane

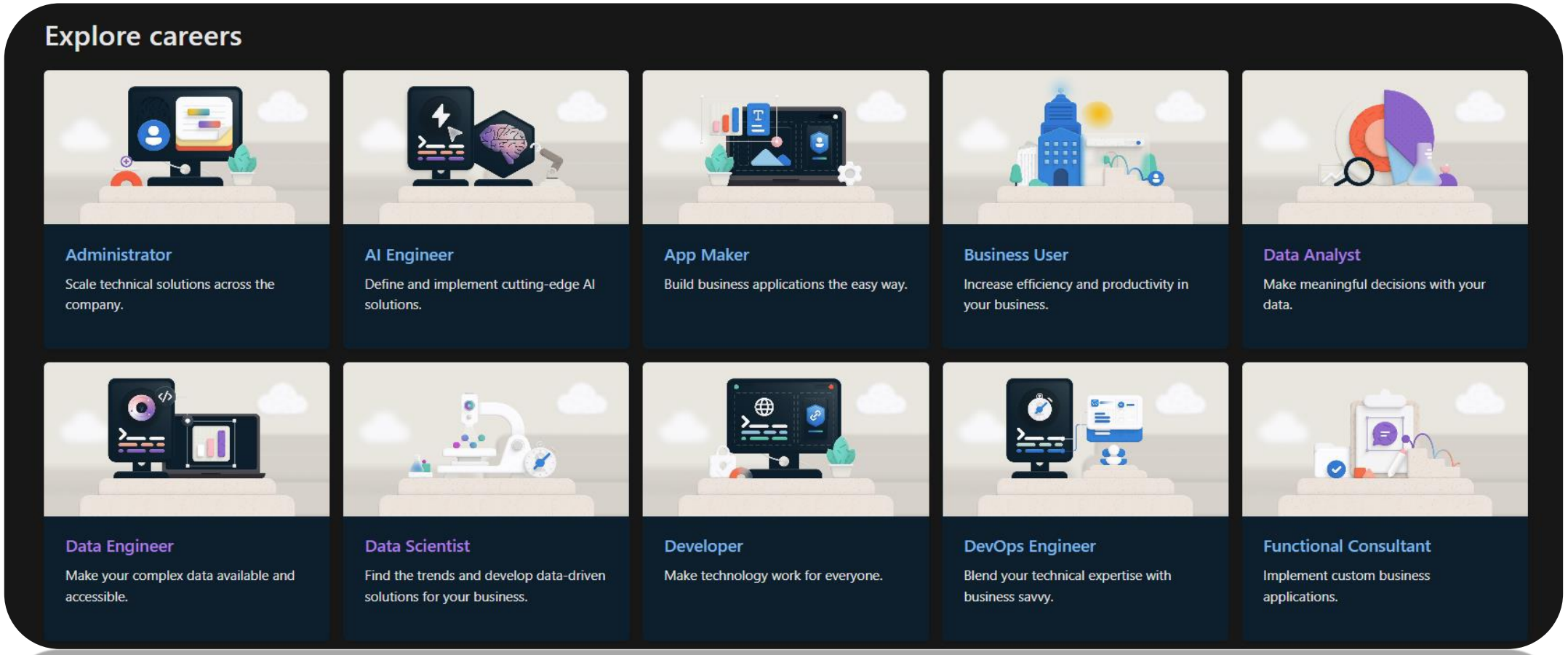


1. Connection to server *localhost* and database *AdventureWorks2022*
2. Tables available
3. Canvas pane to create relationship between logical tables

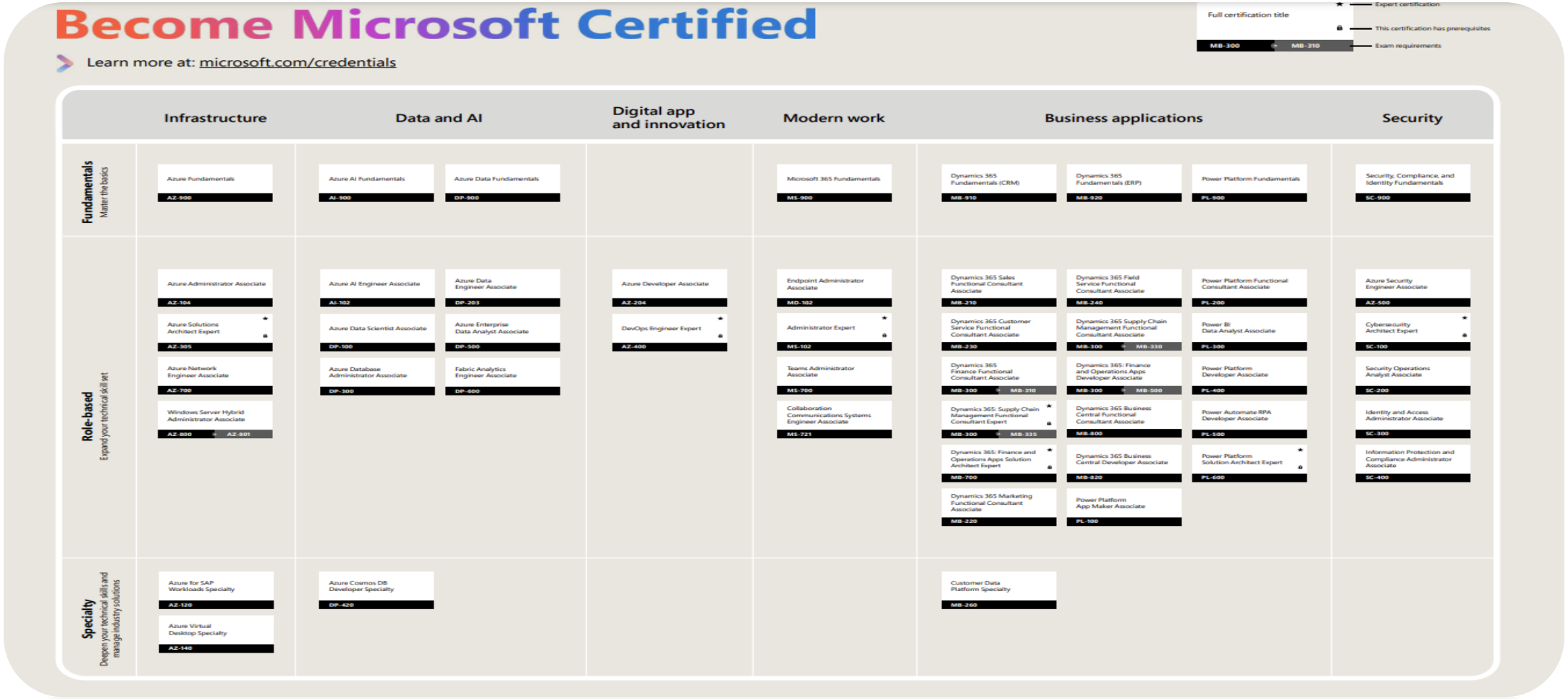
Module 6

Expand learning curve using MOOCs and Github for learning

Microsoft Career Path



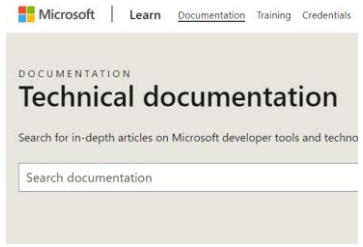
Microsoft Certifications



Learn more

Get this presentation and code

https://github.com/Syarmine/SQL-Masterclass



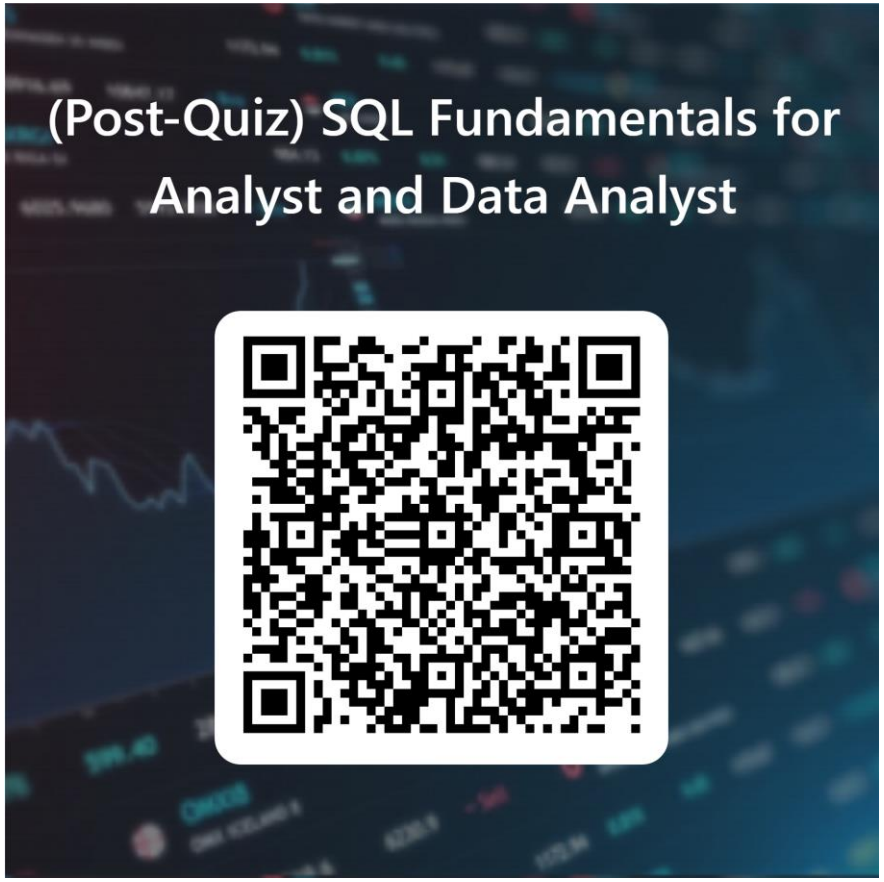
Learn more about SQL Server 2022

https://www.microsoft.com/en-my/sql-server/sql-server-2022

Take the Microsoft Learning Path

https://learn.microsoft.com/en-us/sql/sql-server/educational-sql-resources

Post-Quiz

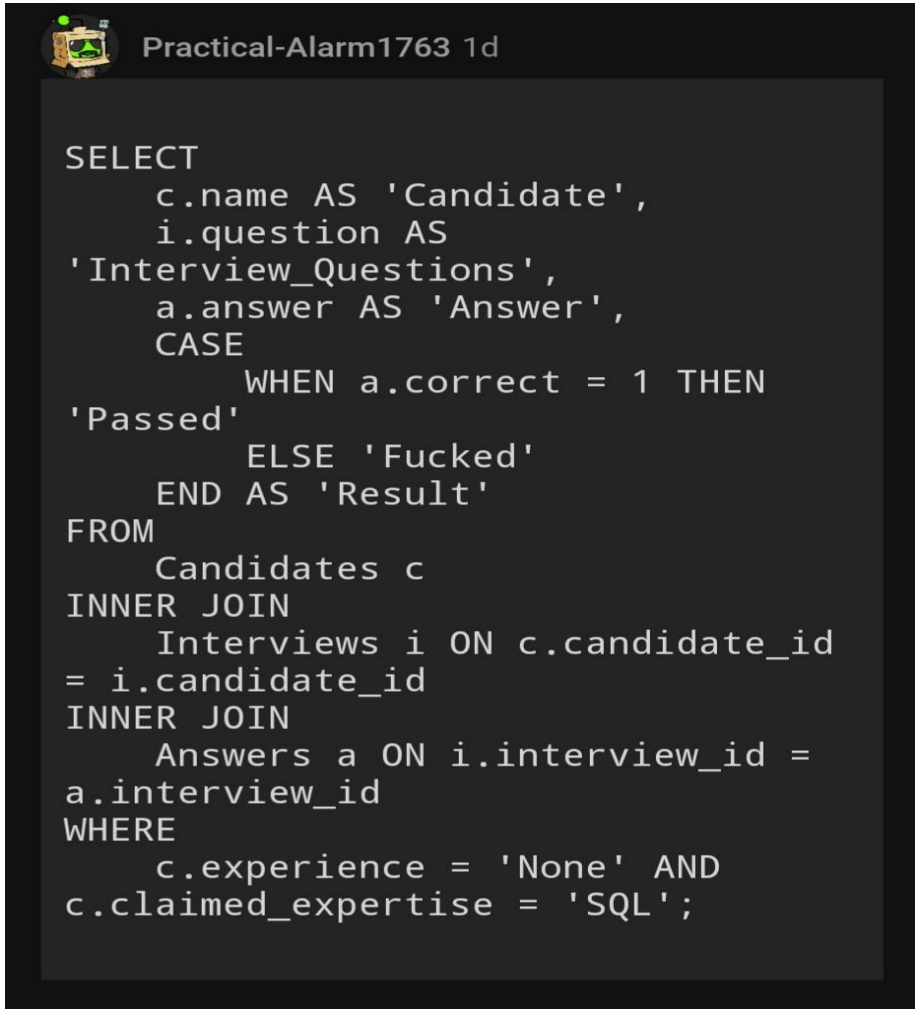


https://forms.office.com/r/QVGuSzE9ir

Summary

Do you understand this now?

No need to lie about your SQL proficiency anymore



Thank you for your attention!

Follow me on:

@SyarmineS

@ Syarmine Shah – connect and say hi!