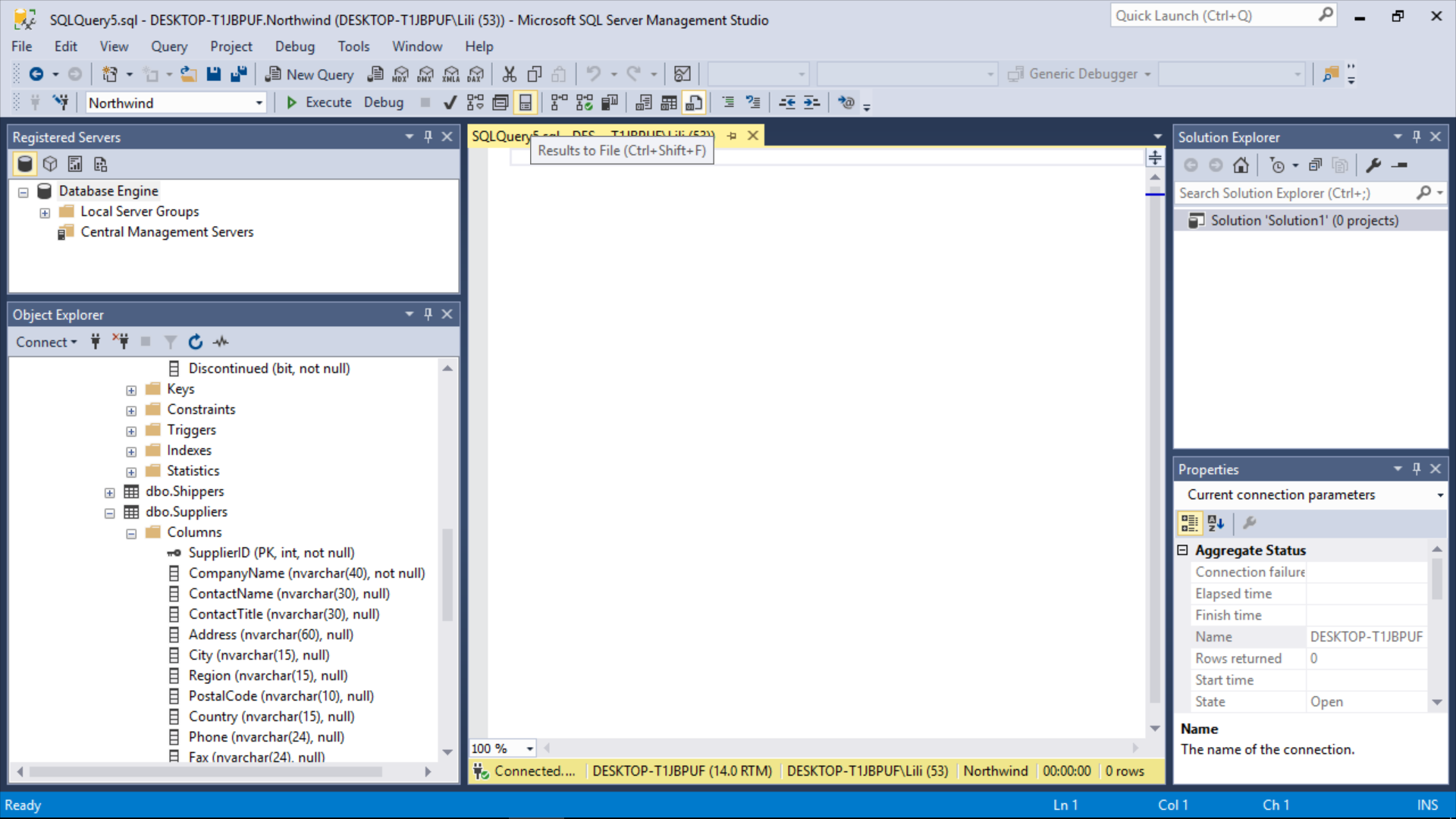
**Individual Assignment 1: Data Profiling (20 Points ≈ 10%)**

**Issue Date: 7th March 2018**

**Due Date: 18th March 2018, 5.00 pm**

**Deliverables**

1. Softcopies of your answers for the following:
   1. Part 1 in **Appendix A** template (in **ONE PDF** file).
   2. Part 2 – SQL and Result Output. *Click* **Results to File** to send your results for Queries 1 to 5 to **ONE file**. When you press **Execute** or **F5** to execute the query, SQL Server will prompt you for the filename to save your result (default is .RPT). Open the document in Word or Notepad and include the query numbering so that I can identify which result is for which question. Submit in **ONE PDF** file.



1. **Submit via iTaleem before or by 16th March 2018 (Friday), 5.00 pm**.

**Objectives**

Data profiling is a process of examining the structure and content of data sources. Getting to know your data in an important step in building the data warehouse because you must understand what you have to work with before you can leverage it as part of a solution.

You will data profile the data sources you will use for all the assignments. The objectives are:

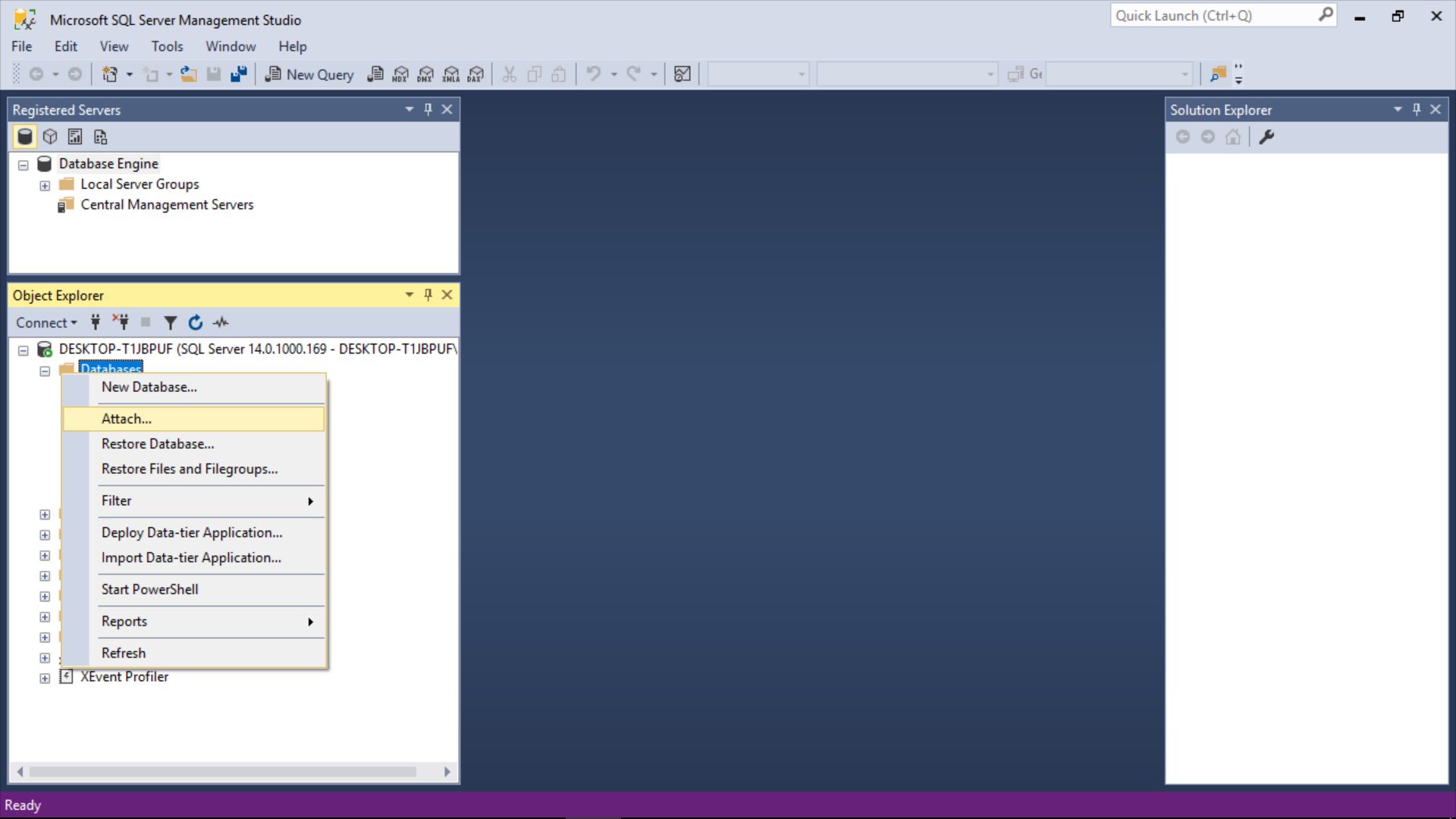
* to help you to become knowledgeable in the source data (i.e. structure, content, relationships, and quality of your data and metadata)
* to get you comfortable writing complex SQL SELECT queries against a relational database as a technique to understand the capabilities of the data and metadata.

Take note that **Microsoft SQL Server uses Transact-SQL (T-SQL)**. Though the SQL does not differ that much from ORACLE, you are required to explore the SQL functions available in SQL server.

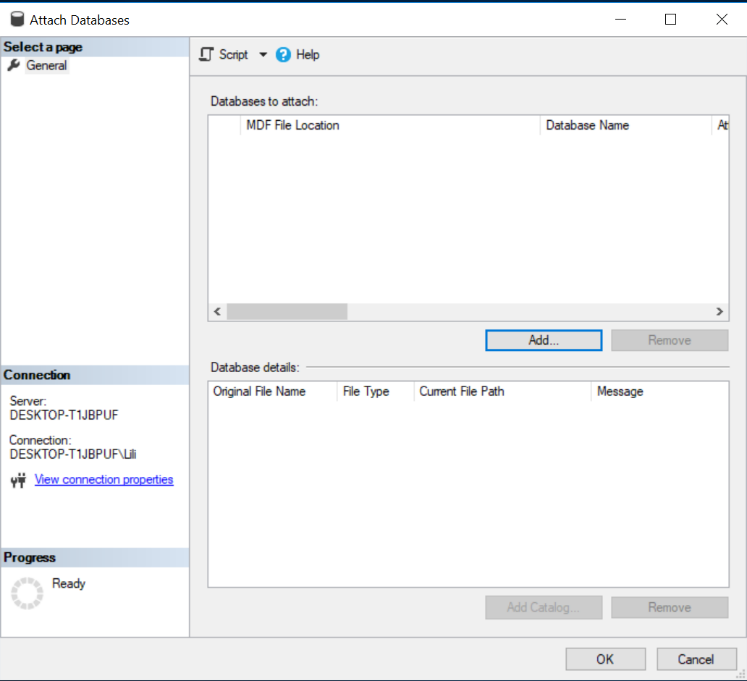
**Source of Data**

You will use **Microsoft SQL Server Northwnd** database provided in iTaleem. Download the database and attach the database – Northwnd.mdf – to your local server.

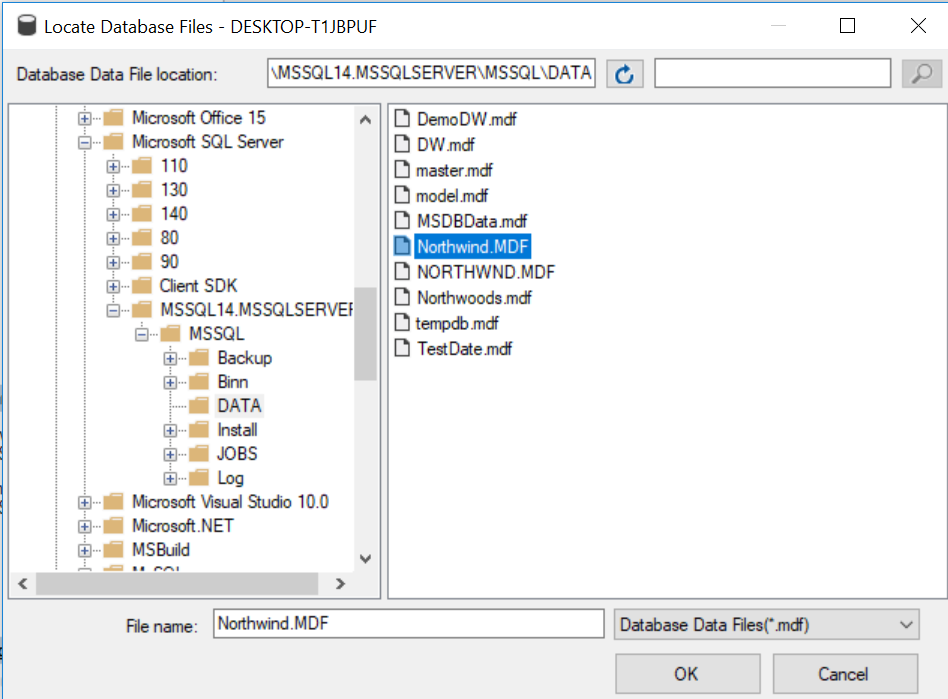
1. *Right-click* Database in the Object Explorer.
2. *Click* Attach.



1. *Click* Add.



1. *Locate the downloaded Northwnd database* in your computer and *Click* OK.



**Part 1: Identify each data source (10 points)**

For each of the data source, answer the following questions and use the template design in **Appendix A** to fill up your answers:

1. What is your each data source?
2. How many ***rows*** are in the data source?
3. What are the ***natural keys*** of the data source? In other words which column or columns from what you have been provided can act as a primary key for the data source? If there is more than one, list them both. For example in the “Orders” table it could be the Order Number, or the combination of Order Date + Customer ID.
4. What does ***one row of each data source represent****?*

For example one row in the “Orders” table represents a customer’s order of products.

**Part 2: Data Profiling Questions (10 points)**

Answer each of the following questions using SQL. Include the SQL you wrote along with the output.

1. What products were purchased in year 1997? Sort the results in ascending order. (**HINT**: DATEPART() function returns individual parts of a date)
2. Which products were the top five (5) sellers? (**HINT**: TOP() function returns the top 5 rows)
3. How many orders were received by the company each year?
4. What were the minimum and maximum days taken to deliver an order from order date to shipped date? (**HINT**: DATEDIFF() function returns the date difference in days, months, year etc.)
5. Which products were on order and who were their suppliers?

**APPENDIX A**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Source | Number of Rows | Natural Key(s) | One row represents … |
| Categories | 8 | CategoryID | Types of product categories as labelled from the CategoryName |
| Customers | 91 | CustomerID | A representative of a person for a company |
| Employees | 9 | EmployeeID | Details of the employee in the organization |
| Order Details | 2155 | OrderID | Quantity and products purchased in a single transaction |
| Orders | 830 | OrderID | Details of shipments of products to customer based on transaction made |
| Products | 77 | ProductID | Inventory of each products |
| Shippers | 3 | ShipperID | Shippers company details |
| Suppliers | 29 | SupplierID | The details of a supplier for products |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |