# Assignment 02: Understanding a Data Warehouse

## Due Date: see Calendar

**Points: 140**

**Introduction:**

There are 2 parts to this assignment:

**Part 1 Exploring a Data Warehouse 100 points**

In this assignment, you will review the data in an operational database and in a data

warehouse. You will **compare** the process of finding information in an operational database versus finding information in a comparable data warehouse. You will observe the data warehouse star schema and you will identify data replication in the data warehouse. Your main objective is to familiarize yourself with the structure of the data warehouse and the data you will be working with.

**Part 2 A Data Warehouse Case Study 40 points**

In this assignment, you will research the use of a real data warehouse in a real organization of your choosing. Your objective will be to understand how a data warehouse can benefit an organization.

**Please read the detailed requirements for each part of the assignment in the following pages.**

**Part 1: Exploring a Data Warehouse 100 points**

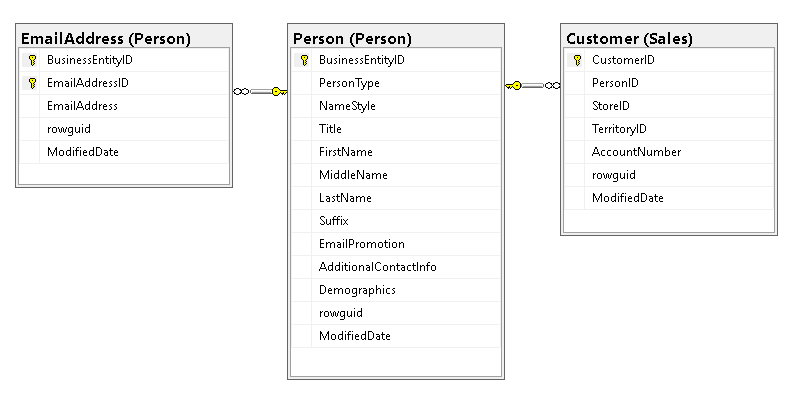
You will use the **SQL Server Management Studio** to access either the 2014 or 2016 versions of the AdventureWorks database and AdventureWorksDW data warehouse.

You must submit answers to the questions/activities listed below.

## IMPORTANT: Submit your answers in this Word document with each answer written below each question. (Do not delete the questions.)

## Note: *If you are using a server in N252, please remember to connect to the database engine in the second instance of SQL Server. Your server should be named something like: N252-0xx\MSSQLSERVER2.*

1. Create a database diagram for a portion of the **AdventureWorks** operational database. Submit a screen shot of your diagram.
   1. Include the following tables in your diagram:
      1. Sales.Customer
      2. Person.Person
      3. Person.EmailAddress



1. Create a database diagram for a portion of the **AdventureWorksDW** data warehouse. Arrange your diagram so that you see a star schema with the fact table at the center. Submit a screen shot of your diagram.

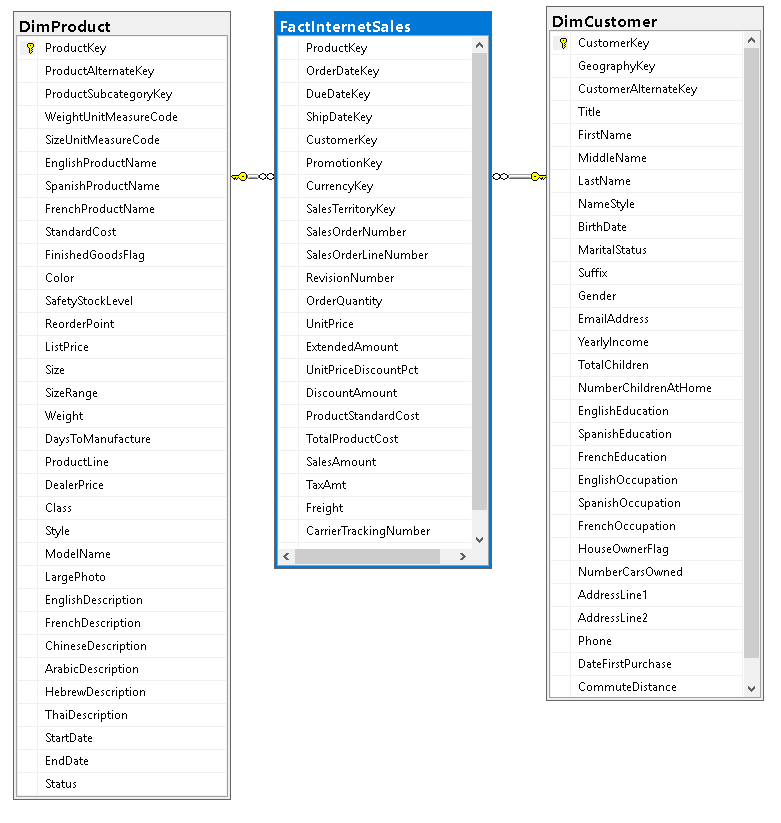
Include the following tables in your diagram:

factInternetSales

dimCustomer

dimProduct

dimDate





1. Create a database diagram for a portion of the data warehouse. Arrange your diagram so that you see a star schema. Make the fact table the center of your star. Submit a screen shot of your diagram.
   1. Include the following tables in your diagram:
      1. factResellerSales
      2. dimReseller
      3. dimProduct
      4. dimDate

****

**Using the diagrams you created, answer the following questions.**

**Examine the factInternetSales table**.

1. What does a row in the factInternetSales table represent to the business? Explain in your own words.

The Fact Internet Sales table contains records for internet sales order(s). Each record contains various elements for a customer internet sale such as sale order, sale order line number, product key, etc. An order has at least one sale order line number.

1. What is the total of the SalesAmount for all of the rows? When writing your response, format the number appropriately.

SELECT [CurrencyName], FORMAT(SUM([SalesAmount]), 'C', 'en-us') AS TotalInternetSales FROM [dbo].[FactInternetSales]

INNER JOIN [dbo].[DimCurrency] ON [dbo].[FactInternetSales].[CurrencyKey] = [dbo].[DimCurrency].[CurrencyKey]

GROUP BY [dbo].[DimCurrency].[CurrencyName]

|  |  |
| --- | --- |
| CurrencyName | TotalInternetSales |
| Australian Dollar | $9,051,988.08 |
| Canadian Dollar | $1,806,517.44 |
| Deutsche Mark | $237,784.99 |
| French Franc | $180,571.69 |
| United Kingdom Pound | $3,388,349.69 |
| US Dollar | $14,693,465.32 |

SELECT FORMAT(SUM([SalesAmount]), 'C', 'en-us') FROM [dbo].[FactInternetSales]

|  |
| --- |
| TotalInternetSales |
| $29,358,677.22 |

1. For a given sale in the fact table, how do you find that sale in the operational database? What table or tables hold the information? What columns relate the data warehouse information to the operational database information?

The Fact Internet Sales table contains records for internet sales order(s). Each record contains various elements for an internet sale such as sale order, sale order line number, product key, etc. An order have at least one sale order line number.

The details for a given internet sale in the data warehouse database can be found in the SalesOrderHeader and SalesOrderDetail tables on the operational database using the SalesOrderNumber and SalesOrderID fields.

SELECT \* FROM [AdventureWorks2014].[Sales].[SalesOrderHeader]

INNER JOIN [AdventureWorks2014].[Sales].[SalesOrderDetail]

ON [AdventureWorks2014].[Sales].[SalesOrderHeader].[SalesOrderID] = [AdventureWorks2014].[Sales].[SalesOrderDetail].[SalesOrderID]

WHERE [AdventureWorks2014].[Sales].[SalesOrderHeader].[SalesOrderNumber] = 'SO43697'

SELECT \* FROM [AdventureWorksDW2014].[dbo].[FactInternetSales]

WHERE [AdventureWorksDW2014].[dbo].[FactInternetSales].[SalesOrderNumber] = 'SO43697'



**Examine the factResellerSales table**

1. What does a row in the factResellerSales table represent to the business? Explain in your own words.

The Fact Reseller Sales table contains records for reseller sales order(s). Each record contains various elements for a reseller sale such as sale order, sale order line number, product key, etc. An order have at least one sale order line number.

1. What is the total of the SalesAmount for all of the rows? When writing your response, format the number appropriately.

SELECT [CurrencyName], FORMAT(SUM([SalesAmount]), 'C', 'en-us') AS TotalInternetSales FROM [dbo].[FactResellerSales]

INNER JOIN [dbo].[DimCurrency] ON [dbo].[FactResellerSales].[CurrencyKey] = [dbo].[DimCurrency].[CurrencyKey]

GROUP BY [dbo].[DimCurrency].[CurrencyName]

|  |  |
| --- | --- |
| CurrencyName | TotalResellerSales |
| Australian Dollar | $1,594,335.38 |
| Canadian Dollar | $14,377,925.60 |
| EURO | $2,474,535.71 |
| United Kingdom Pound | $4,279,008.83 |
| US Dollar | $57,724,791.47 |

SELECT FORMAT(SUM([SalesAmount]), 'C', 'en-us') AS TotalResellerSales FROM [dbo].[FactResellerSales]

|  |
| --- |
| TotalResellerSales |
| $80,450,596.98 |

1. For a given sale in the fact table, how do you find that sale in the operational database? What table or tables hold the information? What columns relate the data warehouse information to the operational database information?

The Fact Reseller Sales table contains records for reseller sales order(s). Each record contains various elements for an reseller sale such as sale order, sale order line number, product key, etc. An order has at least one sale order line number.

The details for a given reseller sale in the data warehouse database can be found in the SalesOrderHeader and SalesOrderDetail tables on the operational database using the SalesOrderNumber and SalesOrderID fields.



1. Are there other sales in the operational database that are not included in the 2 fact tables? Explain your answer.

Yes, there are 10 orders that are in the operational database that are not in the data warehouse. One possible reason is the fact that the transactions from the operational database have not been updated in the data warehouse.

SELECT DISTINCT [AdventureWorks2014].[Sales].[SalesOrderHeader].[SalesOrderNumber] FROM [AdventureWorks2014].[Sales].[SalesOrderHeader]

WHERE [AdventureWorks2014].[Sales].[SalesOrderHeader].[SalesOrderNumber] NOT IN (SELECT DISTINCT [AdventureWorksDW2014].[dbo].[FactInternetSales].[SalesOrderNumber] FROM [AdventureWorksDW2014].[dbo].[FactInternetSales])

AND [AdventureWorks2014].[Sales].[SalesOrderHeader].[SalesOrderNumber] NOT IN (SELECT DISTINCT [AdventureWorksDW2014].[dbo].[FactResellerSales].[SalesOrderNumber] FROM [AdventureWorksDW2014].[dbo].[FactResellerSales])

**Consider the Customer dimension (dimCustomer)**

1. What fact table(s) does the customer dimension relate to? Explain your answer with specific column names.

|  |
| --- |
| Table is referenced by foreign key |
| AdventureWorksDW2014.dbo.FactInternetSales: FK\_FactInternetSales\_DimCustomer |
| AdventureWorksDW2014.dbo.FactSurveyResponse: FK\_FactSurveyResponse\_CustomerKey |



1. How many Internet customers are there in the data warehouse?

SELECT COUNT(\*) AS TotalCustomers FROM [dbo].[DimCustomer];

|  |
| --- |
| TotalCustomers |
| 18484 |

1. How many Internet customers are there in the operational database? Explain how you got your answer.

Internet customers can be identified by querying the Sales.Customer table with condition that checks StoreID is NULL.

SELECT COUNT(\*) AS TotalCustomers FROM [AdventureWorks2014].[Sales].[Customer]

WHERE [AdventureWorks2014].[Sales].[Customer].[StoreID] IS NULL;

|  |
| --- |
| TotalCustomers |
| 18484 |

1. For a given customer in the dimension table, how do you find the customer’s name and email address in the operational database? What tables hold the information? What columns relate the data warehouse to the operational database? Give an example of a SQL query on the operational database for a specific customer in the data warehouse.

To find the customer’s name and email address in the operational database, an outer join must be performed between the DimCustomer database warehouse table and the Customer operational database table. The join is based on Sales.Customer.AccountNumber = DimCustomer.CustomerAlternateKey. Internet customer are those that do not have a store so Sales.Customer.StoreID IS NULL must be meet.

AdventureWorks contains two types of customers: resellers with stores and individuals (consumers) without stores.

SELECT [FirstName], [MiddleName], [LastName], [EmailAddress] FROM [AdventureWorksDW2014].[dbo].[DimCustomer]

FULL OUTER JOIN [AdventureWorks2014].[Sales].[Customer] ON

[AdventureWorks2014].[Sales].[Customer].[AccountNumber] = [AdventureWorksDW2014].[dbo].[DimCustomer].[CustomerAlternateKey]

WHERE [AdventureWorks2014].[Sales].[Customer].[StoreID] IS NULL;

**Consider the Reseller dimension (dimReseller)**

1. What fact table(s) does the reseller dimension relate to? Explain your answer with specific column names.

|  |
| --- |
| Table is referenced by foreign key |
| AdventureWorksDW2014.dbo.FactResellerSales: FK\_FactResellerSales\_DimReseller |



1. How many resellers are there in the data warehouse?

|  |
| --- |
| TotalReseller |
| 701 |

SELECT COUNT(\*) AS TotalReseller FROM [AdventureWorksDW2014].[dbo].[DimReseller]

1. How many resellers are there in the operational database? Explain your answer.

|  |
| --- |
| TotalReseller |
| 1336 |

1. For a given reseller in the dimension table, how do you find the customer information in the operational database? What tables hold the information? What columns relate the data warehouse to the operational database? Give an example of a SQL query on the operational database for a specific reseller in the data warehouse.

To find the reseller’s information in the operational database, an outer join must be performed between the DimReseller database warehouse table and the Customer operational database table. The join is based on Sales.Customer.AccountNumber = DimReseller. ResellerAlternateKey. Resellers are those that do have a store so Sales.Customer.StoreID IS NOT NULL must be meet.

AdventureWorks contains two types of customers: resellers with stores and individuals (consumers) without stores.

SELECT \* FROM [AdventureWorksDW2014].[dbo].[DimReseller]

INNER JOIN [AdventureWorks2014].[Sales].[Customer] ON

[AdventureWorks2014].[Sales].[Customer].[AccountNumber] = [AdventureWorksDW2014].[dbo].[DimReseller].[ResellerAlternateKey]

WHERE [AdventureWorks2014].[Sales].[Customer].[StoreID] IS NOT NULL;

**Consider the Date dimension (dimDate)**

1. What fact table(s) does the date dimension relate to? Explain your answer with specific column names.

|  |
| --- |
| Table is referenced by foreign key |
| AdventureWorksDW2014.dbo.FactCallCenter: FK\_FactCallCenter\_DimDate |
| AdventureWorksDW2014.dbo.FactCurrencyRate: FK\_FactCurrencyRate\_DimDate |
| AdventureWorksDW2014.dbo.FactFinance: FK\_FactFinance\_DimDate |
| AdventureWorksDW2014.dbo.FactInternetSales: FK\_FactInternetSales\_DimDate |
| AdventureWorksDW2014.dbo.FactInternetSales: FK\_FactInternetSales\_DimDate1 |
| AdventureWorksDW2014.dbo.FactInternetSales: FK\_FactInternetSales\_DimDate2 |
| AdventureWorksDW2014.dbo.FactProductInventory: FK\_FactProductInventory\_DimDate |
| AdventureWorksDW2014.dbo.FactResellerSales: FK\_FactResellerSales\_DimDate |
| AdventureWorksDW2014.dbo.FactResellerSales: FK\_FactResellerSales\_DimDate1 |
| AdventureWorksDW2014.dbo.FactResellerSales: FK\_FactResellerSales\_DimDate2 |
| AdventureWorksDW2014.dbo.FactSalesQuota: FK\_FactSalesQuota\_DimDate |
| AdventureWorksDW2014.dbo.FactSurveyResponse: FK\_FactSurveyResponse\_DateKey |

1. What type(s) of information is contained in the Date dimension? Be specific.

**Consider the Product dimension (dimProduct)**

1. What fact table(s) does the product dimension relate to? Explain your answer with specific column names.

|  |
| --- |
| Table is referenced by foreign key |
| AdventureWorksDW2014.dbo.FactInternetSales: FK\_FactInternetSales\_DimProduct |
| AdventureWorksDW2014.dbo.FactProductInventory: FK\_FactProductInventory\_DimProduct |
| AdventureWorksDW2014.dbo.FactResellerSales: FK\_FactResellerSales\_DimProduct |



1. For a given product in the operational database, how do you find the product information in the data warehouse? What column(s) relate the data warehouse to the operational database?

To find a given product in the operational database, an inner join must be performed between the DimProduct database warehouse table and the Production.Product operational database table. The join is based on Production.Product.ProductNumber = DimProduct.ProductAlternateKey.

SELECT \* FROM [AdventureWorks2014].[Production].[Product]

INNER JOIN [AdventureWorksDW2014].[dbo].[DimProduct]

ON [AdventureWorks2014].[Production].[Product].[ProductNumber] =

[AdventureWorksDW2014].[dbo].[DimProduct].[ProductAlternateKey]

The columns(s) related to the product in the operational and data warehouse are [AdventureWorks2014].[Production].[Product].[ProductNumber] and [AdventureWorksDW2014].[dbo].[DimProduct].[ProductAlternateKey] respectively.

1. Could there be multiple rows in the dimension table for a given product?

Yes, there are a few products that appear more than once. There are 504 products in the operational database and 606 products in the data warehouse.

SELECT \* FROM [AdventureWorksDW2014].[dbo].[DimProduct]

ORDER BY [ProductAlternateKey]

-- 606 products

SELECT \* FROM [AdventureWorks2014].[Production].[Product]

ORDER BY [ProductNumber];

-- 504 products

SELECT ProductAlternateKey, COUNT(\*) AS 'Count' FROM [AdventureWorksDW2014].[dbo].[DimProduct]

GROUP BY [ProductAlternateKey]

HAVING COUNT(\*) > 1

SELECT DISTINCT ProductAlternateKey FROM [AdventureWorksDW2014].[dbo].[DimProduct]

1. Give an example of a product in the dimension table that has both Internet and Reseller sales.

There is a total of 142 products that are both in the Internet and Reseller Sales.

SELECT P.[ProductKey] , P.[ProductAlternateKey]

FROM [AdventureWorksDW2014].[dbo].[DimProduct] AS P

WHERE P.[ProductKey] IN (SELECT P.[ProductKey] FROM [dbo].[FactInternetSales] AS T1 WHERE P.[ProductKey] = T1.[ProductKey]) AND P.[ProductKey] IN (SELECT P.[ProductKey] FROM [dbo].[FactResellerSales] AS T2 WHERE P.[ProductKey] = T2.[ProductKey])

|  |  |
| --- | --- |
| ProductKey | ProductAlternateKey |
| 593 | BK-M18S-44 |
| 355 | BK-M68S-42 |
| 570 | BK-T18Y-54 |
| 378 | BK-R89B-52 |
| 384 | BK-R64Y-40 |

1. Give an example of a product in the dimension table that has no Internet sales.

There is a total of 448 products that have no Internet Sales.

SELECT P.[ProductKey] , P.[ProductAlternateKey]

FROM [AdventureWorksDW2014].[dbo].[DimProduct] AS P

WHERE P.[ProductKey] NOT IN (SELECT P.[ProductKey] FROM [dbo].[FactInternetSales] AS T1 WHERE P.[ProductKey] = T1.[ProductKey])

|  |  |
| --- | --- |
| ProductKey | ProductAlternateKey |
| 23 | FW-1200 |
| 46 | HJ-5161 |
| 215 | HL-U509 |
| 547 | PD-R853 |
| 524 | FR-M21S-42 |

**Summarize your findings**

1. Does the factResellerSales table relate to the factInternetSales table?

The factResellerSales and factInternetSales are not related to eachother in the data warehouse. The factResellerSales contains sales for resellers while factInternetSales contains sales for internet customers (non resellers).

1. Write 1 paragraph comparing the data warehouse with the operational database.

The data warehouse tables DimCustomer and DimReseller contain details operational database table Sales.Customer. Sales.Customer contains both customer and reseller.

The data warehouse tables factInternetSales and factResellerSales contain details operational database tables Sales.SalesOrderHeader Sales.SalesOrderDetail contains both customer and reseller sales.

**Part 2 A Data Warehouse Case Study 40 points**

**General**

In this assignment, you will research the use of a real data warehouse in a real organization of your choosing.

**Details**

* Write a 3 (or more) page paper on the use of a data warehouse in a specific organization. Your paper should include:
  + A description of the organization using the data warehouse
  + A description of the data warehouse
    - The purpose of the data warehouse
    - Technical details of the data and structure of the data warehouse
  + A description of how the organization has benefited from the data warehouse
* Please do not write about one of the organizations described in the module materials.
* You must list your sources. (The list of sources does not count as part of the length of the paper.)
* Your work must be original. All writing must be your own and you may not copy text from any source unless it is appropriately noted as a quote and appropriately attributed to the original source. There should be no more than 2 quotes and neither quote should be longer than 1 sentence.
* You must use proper grammar and spelling.
* Your paper should be double-spaced with 10 or 12 point font.
* You should list your sources. (Note that the list does not count as part of the length of your paper.)

# Turn In:

1. This Word document with your answers to Part 1.
2. A separate Word document with your data warehouse paper for Part 2.
3. Submit both Word documents using the Course Assignment Tool.