

# FinanceDW Assignment

## DAT 701, 2018 Assignment #2 (60%)

Hand Out Date: Monday, October 15th 2018  
Final Due Date: 5PM, Monday November 12th, 2018  
Total Marks: 60, making 60% of the course.  
Submit via email to:  
Note: there will be no extensions to this assignment.

---

For this assignment you are to write your answers in a word document and all t-SQL should be in suitably named script files. **You should submit the word document as a PDF**, and the scripts in a zip folder, emailed to me directly please.

This assignment is in six parts: Part 1 (data warehouse design), Part 2 (data warehouse DDL), Part 3 (data warehouse ETL), Part 4 (views and indexing), Part 5 (PowerBI), Part 6 (Peer Review). Each part has specific milestone dates which you are required to meet as set out below:

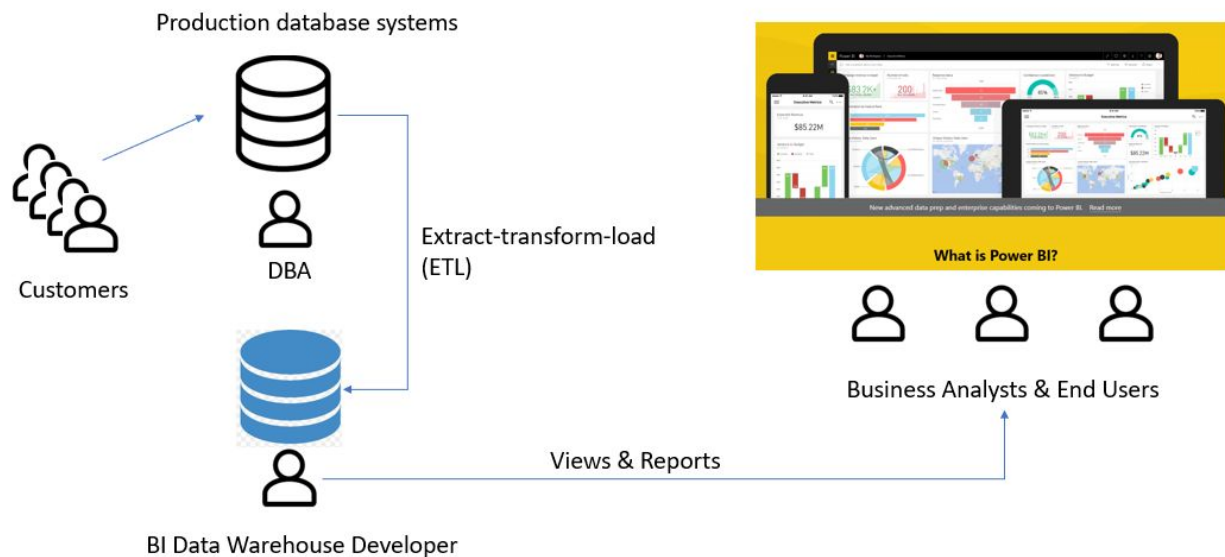
Milestone	Description	Due Date
Part 1 & Part 2	Design and create (DDL) FinanceDW. The design should include a full logical entity relationship diagram. The DDL script should be included.	5PM, Saturday October 20th, 2018
Part 3	ETL: t-sql to populate FinanceDW with data from FinanceDB	5PM, Wednesday October 24th, 2018
Part 4	Views and indexes created. T-sql to be submitted.	5PM, Friday November 2nd, 2018
Part 5 & 6	PowerBI dashboard created & in-class demo. Peer review done.	<b>In-class</b> Wednesday November 10th, 2018

Final milestone	Full design, all scripts and PowerBI pbix file submitted.	5PM Friday November 16th, 2018
-----------------	---	--------------------------------

## Background

For this assignment, you will be required to create a reporting data warehouse for FinanceDB. We have created a virtual machine for each of you, which you will use to create your data warehouse and PowerBI reports. This assignment is divided into size parts, described below. The marking schedule is also below for you.

A typical enterprise reporting architecture is made up of: production online transactional (OLTP) databases, reporting data warehouses and reporting tools, such as PowerBI. A simple system is shown in Figure 1 below:



**Figure 1:** a “typical” enterprise reporting architecture

In a large organisation, the OLTP systems are usually maintained by teams of DBAs and application developers. These are “mission critical” systems, responsible for delivering business services (e.g. Amazon sales) to customers. Transactional data is lifted from the production systems into a reporting data warehouse via a process known as Extract-Transform-Load (ETL).

It is usually the job of a data warehouse developer to design the data warehouse and develop the ETL solution. Finally, the data warehouse developer will also provide views and reports within the data warehouse, which feed reports for business analysts and end users.

In this assignment you will play the role of the DBA, the data warehouse developer and the business analyst at various stages. You will be setup in a team as shown in Figure 2:

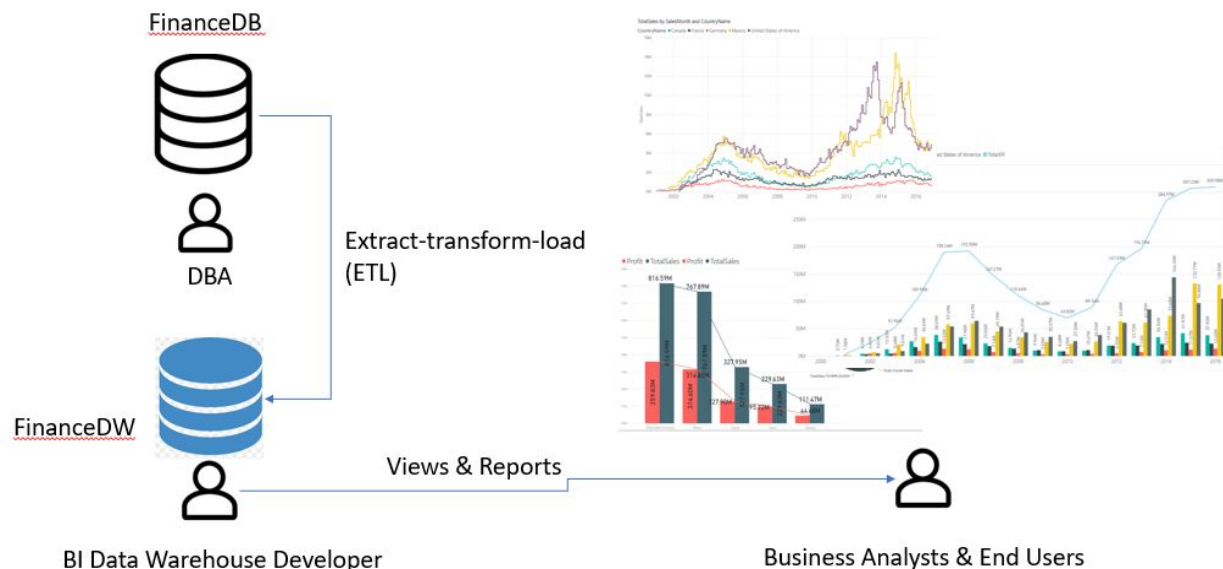


Figure 2: Setup and Roles for Assignment #2

You will all get to play each of these roles at different stages. You will all host a “FinanceDB” production database, for which you will act as the DBA. You will all get to be the data warehouse developer and develop and ETL solution from someone else’s FinanceDB into your own FinanceDW. You will all create views in your data warehouse to feed reports. Finally, you will all get to be the business analyst and develop a PowerBI report using the views in someone else’s FinanceDW.

Mark Caukhill and I have setup virtual machines on TALOS to allow you all to do this. The first class will be a little experimental while we set these up and get them working. You can access your machines from the following URL: <https://nt-vcsa.talos.net.nz/>

This assignment is made up of six parts, each worth between 8 and 12 marks (or 8% - 12% of the DAT701 course). Each part, including the marking schedule is briefly described throughout the following sections.

## Part One: Data Warehouse Design (10 marks)

You are required to design a dimensional data model to support a reporting data warehouse based on FinanceDB. The data warehouse should support (at a minimum) all of the queries and reports that we presented in Assignment #1. The outputs of this part will include:

- a conceptual entity-relationship diagram which shows all relevant dimension and fact tables, their attributes, data types and relationships.
- FinanceDB restored to your own virtual machine

The marking schedule is as follows:

0	4	7	10
No design	FinanceDB is restored and a partial, incomplete or inaccurate design has been submitted	FinanceDB is restored and a complete dimensional model, with all relevant dimensions, fact tables, relationships, attributes and data types.	FinanceDB is restored and a complete dimensional model, as per the previous. Also includes augmented dimensions or facts to support the business reporting requirements above and beyond the basics.

## Part Two: Data Warehouse DDL (10 marks)

For this part, you should create the data warehouse you designed in Part One. This should include the DDL (data definition language) to:

- Create the FinanceDW database in SQL Server
- Create all dimensions and fact tables

- Any additional aspects (for example, defining the recovery models, table partitioning, definition of primary keys and foreign keys, constraints etc.)

The marking schedule is as follows:

0	4	7	10
No DDL submitted	A partial, incomplete or inaccurate t-sql script has been submitted	A complete and accurate DDL script which successfully creates the data warehouse, dimensions, fact tables and relationships.	A complete and accurate DDL script as per previous. To excel, students should also consider the scale of this data warehouse, implement table partitioning where appropriate, set the recovery model and database owner.

## Part Three: ETL (12 marks)

For this part, you will be assigned partners. Each partnerships will consist of:

- One person, the “DBA”, whose version of FinanceDB will be the “production source database”
- The other person, the “data warehouse developer”, will host FinanceDW, the “reporting data warehouse”.
- Everyone will act as the “production source database” for one other person. Everyone will also have to build their own “reporting data warehouse”.

The tasks for this part include:

- On the production source database system, create a database login for your partner and give them read-only access to your version of FinanceDB
- Help your partner to create a linked server from their “reporting data warehouse” to your “production source database”
- Everyone must create an extract-transform-load (ETL) routine using t-sql to extract data from the “production source database” into their own FinanceDW

The marking schedule is as follows:

0	4	8	12
No ETL solution.	Database logins and linked servers are successfully created.  A partial, incomplete or failing ETL workflow.	Database logins and linked servers are successfully created.  An ETL workflow which successfully and accurately populates data from the “production source” FinanceDB into your own FinanceDW.	Extends on the previous by encapsulating the ETL workflow in a set of well-designed, modular stored procedures, which are then callable from a SQL Agent Job or simple t-sql script.  An excellent ETL solution should also include logic to handle new data and potential UPSERT operations.

## Part Four: Indexing and Views (8 marks)

For this part, you are to implement two or more of the reporting queries from Assignment #1 as views in FinanceDW. The choice of which reporting queries is completely up to you, but they must be distinct (i.e. the cannot be largely the same, except for minor details). You may also create additional reporting views if you would like to.

The tasks for this part include:

- Create 2 or more reporting views in FinanceDW to support the reporting requirements
- You should briefly document the purpose of these views i.e. what types of reports will they support, what types of business questions can be answered from these views?
- You should also consider suitable indexes to support these views.

The marking schedule is as follows:

0	4	8
No views created.	At least one view created, but with poor / limited documentation and no evidence of suitable index design.	Two or more distinct views created, suitably indexed and well documented.

## Part Five: PowerBI (12 marks)

For this part, you will be assigned partners. Each partnerships will consist of:

- One person, the “data warehouse developer”, will host FinanceDW, the “reporting data warehouse”.
- The other person will be the “business analyst” and create a simple PowerBI report using the views provided in their partner’s version of FinanceDW.

The tasks for this part include:

- Ensure that your “business analyst” has a database login with read-only privileges to your version of FinanceDW.
- You should work with your “business analyst” to describe your views and make any adjustments where appropriate
- As the business analyst, you should create a simple PowerBI dashboard using your partner’s FinanceDW views. You should consider the business question that your dashboard is intended to answer, appropriate use of visualisations, colours, labels, filters and drill-down.
- You should briefly document the “theme” or business questions that your dashboard allows end-users to explore.

The marking schedule is as follows:

0	4	8	12
No dashboard solution..	A simplistic or incomplete PowerBI dashboard.	A good PowerBI dashboard, with a clear theme / business focus and concise documentation.	An advanced PowerBI dashboard, with appropriate selection of colours, visualisations, labels, filters and drill-down to facilitate self-service exploration of the data.

## Part Six: Peer Review (8 marks)

For this part, you will be reviewed by your “data warehouse developer” partner and your “business analyst” partner. You will be assessed on your communication, helpfulness and time management.

I (Nick) will interview each of your partners to get these scores as shown below:

0	4	8
Significant problems communicating, collaborating or with time management.	Overall, a reasonable partnership but some difficulties with communication, problem-resolution or time management.	A good collaboration with all partners. Delivered on-time, resolved difficulties or challenges quickly without hassle.