**Instructions**

1. **Team information**

|  |  |
| --- | --- |
| **Student Name** | **Student ID** |
| Wu, Hao | 1446122 |
| Lu, Wang |  |
| Yang, Langhu | 1452377 |
|  |  |

1. **Case study identification and requirement analysis**

**Sales – Sales focus**

* Sales person with Customers details for each territory

This can be used to identify by the human resource departments to know the details of their employed sales persons, the territory dimension is added to separate each territory.

* Total sales for each sales person for each territory quarterly

This can be an important indicators to evaluate the performance of each sales person within one territory.

* List of credit sales for each sales person

This can be used to identify the credit sales of each sales person so that we know the percentage of the credit sales and percentage of cash sales.

* List of all discounted sales for each territory

This could be used to identify that which territory is more sensitive for the price change and then help the business to make price strategies towards different territories.

* Total sales for all products with discount sales quarterly

This can be used to help the business to evaluate whether the product positioning of each products is proper or not.

1. **Proposed solution**

In order to answer above questions in a quicker and understandable way, we need to create our own data ware house OLAP database. We will follow ETL (**Extract, Transformation, and Load**) process to do that.

**Extract -> Transformation**

In order to extract data from OLTP database, we should create OLAP database to move data OLTP from new database. So we should decide the DW database schemas to do so.

Depends on different requirements (dimensions) from the questions 1 to question 5, we decided to use both star schema and snowflake. In detail, we are going to use star schema for question 1 to question 3, to use snowflake schema for question 4 to question 5:

1. Sales person with Customers details for each territory (star schema)

We have three dimensions in this question: sales person, customer details, and territory. These three tables could be linked to each other by Sales & order table, so we could apply start schema by Sales & order table as fact table. (Refer to Fig 1)

1. Total sales for each sales person for each territory quarterly (star & snowflake)

We could see four dimensions for this question: Sales person, territory and date table and quarter table. These three tables: sales person, territory and date table could be linked by sales & order table in the centre. But we will have quarterly result from quarter table which linked with date table only. The relationship in between date table and quarter table could be one to more, so we could use hybrid schema that created from star and snowflake schemas to achieve it. (Refer to Fig 1)

1. List of credit sales for each sales person (star schema)

Similar with above, we could find sales person as one dimensions and sales & order as fact table, so we could have one dimension with one fact table. (Refer to Fig 1)

1. List of all discounted sales for each territory (snowflake)

In this case, we will need to have progressive relationship to fetch discount sales. In order to do that, we could have promotion table, and product table and territory table. So the relationship could be sales & order table link to product table and territory table, and product table link to promotion table. In this way, we will have snowflake schema because of the product table could have one to more relationship with promotion table. (Refer to Fig 1)

1. Total sales for all products with discount sales quarterly (snowflake)

Similar with the questions above, we will link sales & order table with product and promotion tables, and link sales & order with date and quarter tables as well. As we discussed above, product and promotion tables have relationship of one to many; date and quarter tables also have relationship of one to many. So it could be another case that we should apply snowflake schema. (Refer to Fig 1)



Fig 1. DW database design with Hybrid schema (star & snowflake schema)

**Transformation**

After the OLAP database has been defined, we will need to do data cleaning. This data cleaning is to make sure the data in our OLAP database are useful, meaningful, readable and correct. In practise, we will need to clean up data such as NULL, Caps, duplicated, with in any invalid formats and index and so on.

Other than these, in order to improve performance, we will add index, views and so on in our OLAP database, and we will try to use less joins and optimize our query as we can for quicker result fetching.

**Transformation -> Load**

In this part, we will use SQL 2014 BI version to process these DW tasks include load data. The main programming language here to perform all these tasks could be SQL queries.

1. **Expected outcome**

The goal of this project is creating a data warehouse based on SQL server to answer business questions fast and easily, with clear and self-explain data structure and data base table design. The aim also includes easy to maintain, update, and even change the question queries in existing query dimensions as well.

The business query questions will be tried to fully implemented with best efficient and maintainability, those report will help business manager to analysis their business better, thus grow their business in their strength area.

As a class project, it doesn't involve in any commercial deployment or application. However, because of this project, it is a great chance for us to learn the whole process of data warehouse as well as utility of important terms in it. Besides, we also could discover a potential pattern that could help to know a data warehouse in sales industry. More details that have impact on the sales analysis will be discussed in the project.