Department of Computer Science and Informatics
CSID6853 – Data Warehousing – 2024
Practical Project - Part C

---

## Instructions

- Ensure your data warehouse has been pre-created on CSI server with all the relevant fact and dimension tables.
- All the tables must be empty.
- Create a stored procedure (in your database as well), called spCreateDM, that will drop and create all the tables in the database. Upload the SQL file also onto Blackboard.
  (format: spCreateDM_studentnumber_InitialsSurname.zip).
- Upload in **one zip file,** all the files required to run your **SSIS** Visual Studio Solution onto Blackboard. Include the store procedure as well: (format: SSIS_studentnumber_InitialsSurname.zip)
- Book a 45-minute timeslot on **05-Jun-2024, 06-June-2024** or **07-Jun-2024** to demonstrate the SSIS ETL to me. First-come-first-serve basis for the bookings.

---

## SQL Server Integration Services Packages [75 marks]

- You need to develop the necessary SQL Server Integration Services packages to load the dimensional model (see figure 1) for the Maker Inc business.
- The focus for the ETL project is to ensure the data warehouse is loaded with data that would enable the senior manager of sales to analyse **sales information**.
- Make sure you can run your package(s) multiple times. In other words, cleanup / delete the data from the tables before you load the necessary data.
- Use the following data to link stores with countries:

| | STOREID | COUNTRY_CODE |
|---|---|---|
| 1 | 1 | US |
| 2 | 2 | US |
| 3 | 3 | US |
| 4 | 4 | US |
| 5 | 5 | IN |
| 6 | 6 | JP |
| 7 | 7 | DE |

- Your SSIS solution should consist of the following packages:
  - Master
  - LoadDimension
  - LoadFact

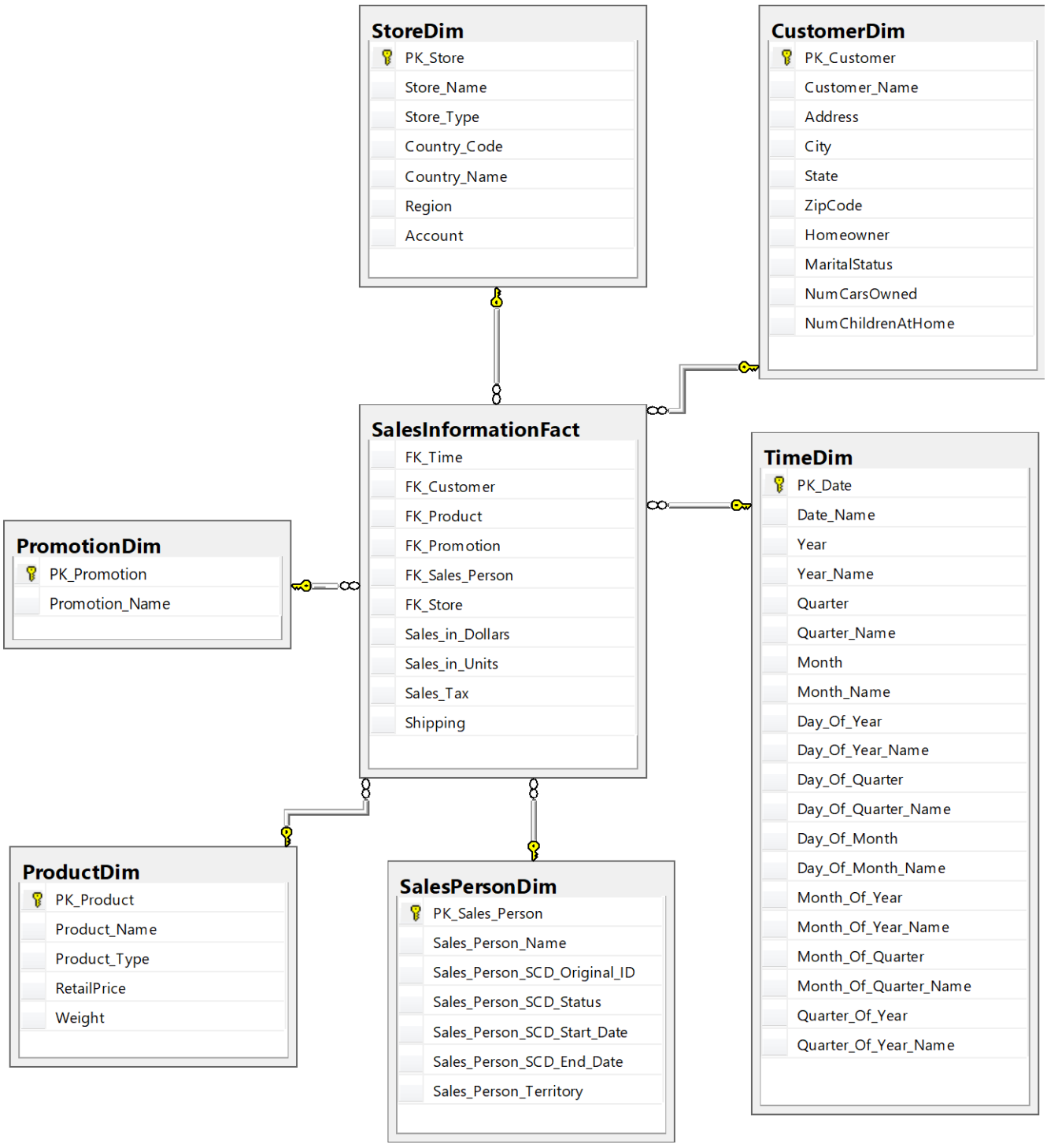


**Figure 1: Dimensional Model**

| Package | Purpose |
|---|---|
| MasterPackage.dstx | To run all the individual packages in sequence. |

| | |
|---|---|
| LoadDimensions.dtsx | 1. Load all the dimension tables in this package.<br>2. Use all possible data sources to load the dimension tables.<br>3. You must use separate Data flows each dimension table.<br>4. For loading TimeDime, you will need to use data from the SalesTimeTable20132016.csv file.<br>a. Date_Name must be the full English month name (for example: Thursday, January 01 2013)<br>b. Year_Name must be in the format Calendar YYYY (for example: Calendar 2013)<br>c. Quarter_Name must be in the format Quarter qq, YYYY (for example: Quarter 1, 2013)<br>d. Month_Name must be in the format Month YYYY (for example: January 2013)<br>5. For loading CustomerDim, you will need make use of derived column for Customer_Name (Firstname + MiddleInitials + Lastname).<br>6. For loading ProductDim, you will need to convert Weight to datatype Float.<br>7. For loading StoresDim, you will need to incorporate a hierarchy.<br>a. ExternalAccountData.csv contains the hierarchy.<br>b. At the lowest level there is a 'Country', which belongs to a 'Region' which belongs to an 'Account'.<br>c. The **StoresDim** table must contain for each country, the sales region it belongs to (Region column) and the name of the Account it reports to (Account column).<br>d. Use ExternalLocationsData.csv to link the **Locations** with the **Stores**.<br>8. Think about SalesPersonDim – implement a Slowly Changing Dimension (SCD) to accommodate staff changes. **HINT**: Compare the two *Staff* tables in the operational source to give you an idea of how to solve this.<br>9. Ensure that the data warehouse connection is converted to a Project Connection in the context menu. |

| | |
|---|---|
| LoadFacts.dtsx | 1. You must use separate Data flows for each step.<br>2. Load the SalesInformationFact table by using the Order Processing System as the source for the fact table.<br>a. Load the fact table in such a way that the correct sales_person is linked to a fact for the relevant time_period.<br>b. Make use of Lookup items to ensure only data loads that are present in the ProductDim and TimeDim tables. Redirect rows to no match output (for handle rows with no matching entries).<br>c. Make use of a Union All for all unmatched entries from step 2.a. and write it to a delimited text file.<br>3. You will notice there is data lacking for the following measures:<br>a. Set ShippingTax = 0 for all entries.<br>b. Set SalesTax = TotalStoreSales * TaxRate (use 15%) for all entries. |