**ETL Documentation**

The purpose of the ETL documentation is to outline the process of extract, transform and load from the Fudgemart and Fudgeflix databases into our Datawarehouse. In this project, we have focused on the business process of “Order Fulfillment” in which we are calculating the lag days for each order placed by the customer of the Fudge Corporation Ltd. (The merger of Fudgemart and Fudgeflix) for the products belonging to the original companies.

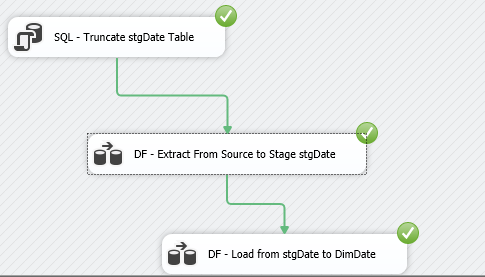
We created the following packages:

1. **DateDimensionImport:** Extracted data of dates from ExternalSources2, staged the data in the staging database and loaded data in the data warehouse.
2. **StageImport:** Extracted data of products, customers, orders from databases of Fudgemart and Fudgeflix to conncet 4 subpackages (DF – Stage Customer, DF – StageProduct, DF – Stage Orders, DF – Stage FactOrderFulfillment) to SQL – Truncate Stage Tables and loaded the data in the staging area, separately for Fudgmart and Fudgeflix
3. **WarehouseImport:** Created 4 sub – packages (DF – Stage to FMFFWarehouseDimProduct, DF – Stage to FMFFWarehouseDimCustomer, DF – Stage to FMFFWarehouseDimOrder, DF – Stage to FactOrderFulfillment) to import data from both, Fudgemart and Fudgeflix in each of these packages for Dimensions and used referential integrity for creating tables in FactOrderFulfillment

**DateDimensionImport:**

The date dimension import included a data flow task that truncated the stgDate table, and two data flow tasks: extracting the data from the source and loading the data from the stage to the data warehouse. The data was extracted into the staging environment in the same format from the ExternalSources2 database. The data was then transferred from the stage environment into the data warehouse as a Type 1 slowly changing dimension, with the DateKey as the business key.

**DateDimension Import Control Flow:**



**Source to Stage Control Flow:**

**A screenshot of a cell phone

Description automatically generated**

The SQL – Truncate Stage Tables package includes truncate statements for Customers, Product, Orders and OrderFulfillment for respective tables in both Fudgemart and Fudgeflix

**DF – Stage Customer:**

**Fudgemart:** Customer table (fm\_customers) from Fudgemart were imported to get the data of customers such as id, first name, last name, address etc.

**Fudgeflix:** Accounts table (ff\_accounts), and ff\_zipcodes from Fudgeflix were imported to get the data of its customers

**DF – Stage Product:**

**Fudgemart:** Here, fm\_products and fm\_departments\_lookup is used to get the data of Products and the respective departments that they belong to in Fudgemart

**Fudgeflix:** For this, ff\_titles is used to get the data of the services that users in Fudgeflix use.

**DF – Stage Orders:**

**Fudgemart:** fm\_orders, fm\_shipvia\_lookup and fm\_order\_details from Fudgemart are used to to gets the details of the orders placed in Fudgemart

**Fudgeflix:** ff\_account\_titles is used to obtain the information about the orders placed in Fudgeflix

**DF – Stage Fact Order Fulfillment:**

1. **For both (Fudgemart and Fudgeflix), the order date and shipped day were used to count the lag days**
2. **The data obtained from both the tables is also used to calculate the quantity, which is the amount of quantity in each order placed by the customer**

**Stage to Data Warehouse Control Flow:**

**A screenshot of a cell phone

Description automatically generated**

The above control flow shows the connection between packages that were created during staging. The data loaded into the 3 staging packages, namely for Customers, products and Order Details is used to create 3 dimensions FMFFWarehouse.DimCustomer, FMFFWarehouse.DimProduct and FMFFWarehouse.DimOrder. FMFFWarehouse.DimDate is directly loaded in the data warehouse as mentioned earlier.

The data that gets loaded from the stage to these 4 tables in the data warehouse is then used to create the fact, which in our case is the FMFFWarehouse.FactOrderFulfillment

Below is the star schema of the entire process:

**Star Schema:**

**A screenshot of a computer

Description automatically generated**

**Stage Tables and Data Warehouse Tables :**

**A screenshot of a cell phone

Description automatically generated**