**SSIS Technical Test - Solution document**

Contents

[Requirement Understanding 3](#_Toc65607096)

[Solution: 3](#_Toc65607097)

[Loading data in to Dimension tables 4](#_Toc65607098)

[Loading data into Fact\_sales\_USD\_NZD 5](#_Toc65607099)

[Assumptions: 5](#_Toc65607100)

[What can be done better? 6](#_Toc65607101)

[Testing evidence after loading fact table 7](#_Toc65607102)

## Requirement Understanding

* Create new fact table which allows business to run query which returns the US Dollar orders with following columns
  + Total Sales amount in New Zealand Dollars
  + Total Sales amount in US Dollars
  + Total Quantity ordered
  + Grouped by: Calendar year, Product color
* Conversation rate of US Dollars to New Zealand Dollars using API

## Solution:

* Based on the requirement, the Fact\_Internet\_Sales table is used as the reference to build my target table
* Target table structure is show as below

|  |  |
| --- | --- |
| Table Columns | Data Type |
| [ProductKey] | [int] |
| [OrderDateKey] | [int] |
| [DueDateKey] | [int] |
| [ShipDateKey] | [int] |
| [CustomerKey] | [int] |
| [PromotionKey] | [int] |
| [CurrencyKey] | [int] |
| [SalesTerritoryKey] | [int] |
| [SalesOrderNumber] | [nvarchar](30) |
| [SalesOrderlineNumber] | [nvarchar](20) |
| [Color] | [nvarchar](20) |
| [RevisionNumber] | [tinyint] |
| [OrderDate] | [datetime] |
| [DueDate] | [datetime] |
| [ShipDate] | [datetime]NOT |
| [ProductStandardCost] | [money] |
| [OrderQuantity] | [smallint] |
| [UnitPrice] | [money] |
| [UnitPriceDiscountPct] | [float] |
| [FinalUnitAmount] | [money] |
| [FinalProductCost] | [money] |
| [TaxAmt] | [money] |
| [Freight] | [money] |
| [SalesAmount\_USD] | [money] |
| [NZD\_Current\_rate] | [money] |
| [SalesAmount\_NZD] | [money] |

From understanding the source tables, Sales orders are the transactional tables which has all the necessary attributes to build the relationship with all master tables

Dimension tables which are needed are,

* Dim\_Product for Product\_key
* Dim\_Customer for Customer\_key
* Dim\_Date for Orderdate\_key,shipdatekey,Duedatekey
* Dim\_promotion for Promotion\_key
* Dim\_currency for Currency\_key
* Dim\_salesterritory for salesterritory\_key

## Loading data in to Dimension tables

1. **Dim\_Product**

* Based on the analysis, found the history of products data is maintained
* As there were some data quality issues, created some process to update the status correct for all historical record. This is implemented using SQL Execute task in Control flow which updates the status of history records
* Implemented SCD logic to load all new data from source table and maintain history in database
* Status flag is also set to identify the active record
* Considerations in Dim\_Product, few source columns could not be identified, hence populated them as NULL or DUMMY as these columns are not going to impact the Fact table

1. **Dim\_Customer**

* Based on the analysis, found the only latest data is maintained
* Implemented SCD Type 1 logic to insert all new data and update all existing date from source table
* Considerations in Dim\_Customer: Few source columns are having more column length and current dimension table are having less than source, as these columns are not going to impact the Fact table no changes implemented

1. **Dim\_Currency**

* Based on the analysis, found the only latest data is maintained
* Implemented SCD Type 1 logic to insert all new data and update all existing date from source table

1. **Dim\_Promotion**

* Based on the analysis, found the only latest data is maintained
* Implemented SCD Type 1 logic to insert all new data and update all existing date from source table

1. **Dim\_Salesterritory**

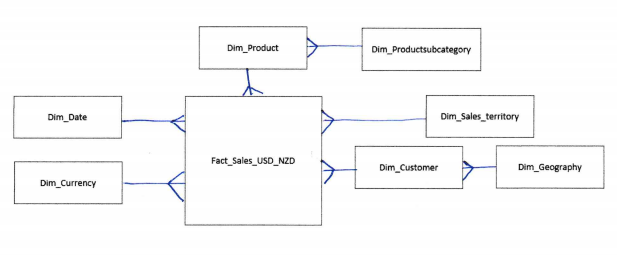
* Based on the analysis, found the only latest data is maintained
* Implemented SCD Type 1 logic to insert all new data and update all existing date from source table

1. **Dim\_Date**

* Considered this as one-time load and not implemented any package to load

## Loading data into Fact\_sales\_USD\_NZD

* I have implemented SSIS Package to build Fact and Dimension relation as below



* Obtaining Dimension key – Implemented look up transformation to take ID from Sales transactions and bringing the necessary keys from dimension tables
  + Dim\_Product for Product\_key
  + Dim\_Customer for Customer\_key
  + Dim\_Date for Orderdate\_key,shipdatekey,Duedatekey
  + Dim\_promotion for Promotion\_key
  + Dim\_currency for Currency\_key
  + Dim\_salesterritory for salesterritory\_key
* Implemented a script task to read the API data, Live conversion rate of USD to NZD will be extracted from API and stored in variable. This variable is used as conversion rate when loading fact table
  + Note: Implemented conversion rate only from USD to NZD, no other data conversion implemented
  + As all the transactions amount shown in source are USD and the requirement in to convert only one currency, I implemented directly in fact table without creating separate table.

STTM: Please find the attached STTM document for all the packages which has been created for more details.

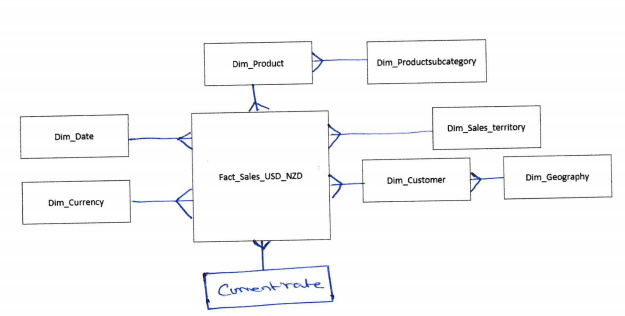


## Assumptions:

1. There are lot of transaction records with CurrencyrateID- NULL value in in Sales order header table
   1. I assume all these are original USD transaction hence, no currency rate entry in currenyrate table.
2. All Amounts in Sales order are having base currency as USD. They are not the local currency
3. While implementing packages for Dimension tables, I could not get the source columns details hence, I have passed NULL or Dummy values to those columns as it was not used in fact table
4. Dimension Date is one time load and no package has been built

## What can be done better?

I have just focused on the fact table creation and loading table to return the desired output. There are lot of improvements which has to be implemented

1. Creating separate table to hold all the currency rate conversion from USD to Local currency
   * This will help us have conversation rate for all currency
   * Can have FK relation with Dim\_Currency
   * Use Currencyrate\_key directly in fact table to link the record and get current rate for conversion

* + Can store all the historical rate information

1. The tables in DWH are not having problem date columns to capture the changes details. Even if it is SCD 1 there should be date column in dimension table to identify the updated date
2. Build Audit & Control process
   * Audit process has to be implemented to do all the audit check after completion of data load
     1. Count records in source
     2. Count records loads in target
     3. Key columns validations
   * Control process to maintain ETL Process can be implemented
     1. ETL Control Tables to capture the run details of packages
     2. ETL control table to load only incremental data
3. Error Handling
   * Error handling has to be implemented
   * All the records which do not meet the business requirement has to be captured in the exception flow and send to customers for their correction
4. Notification
   * Email notifications has to be setup
5. Run Package in groups
   * As of now, I have given individual packages to run
   * Ideally, I want to create a master package to with sequence container and have multiple execute package tasks to run all these packages and then run fact table package.

## Testing evidence after loading fact table

