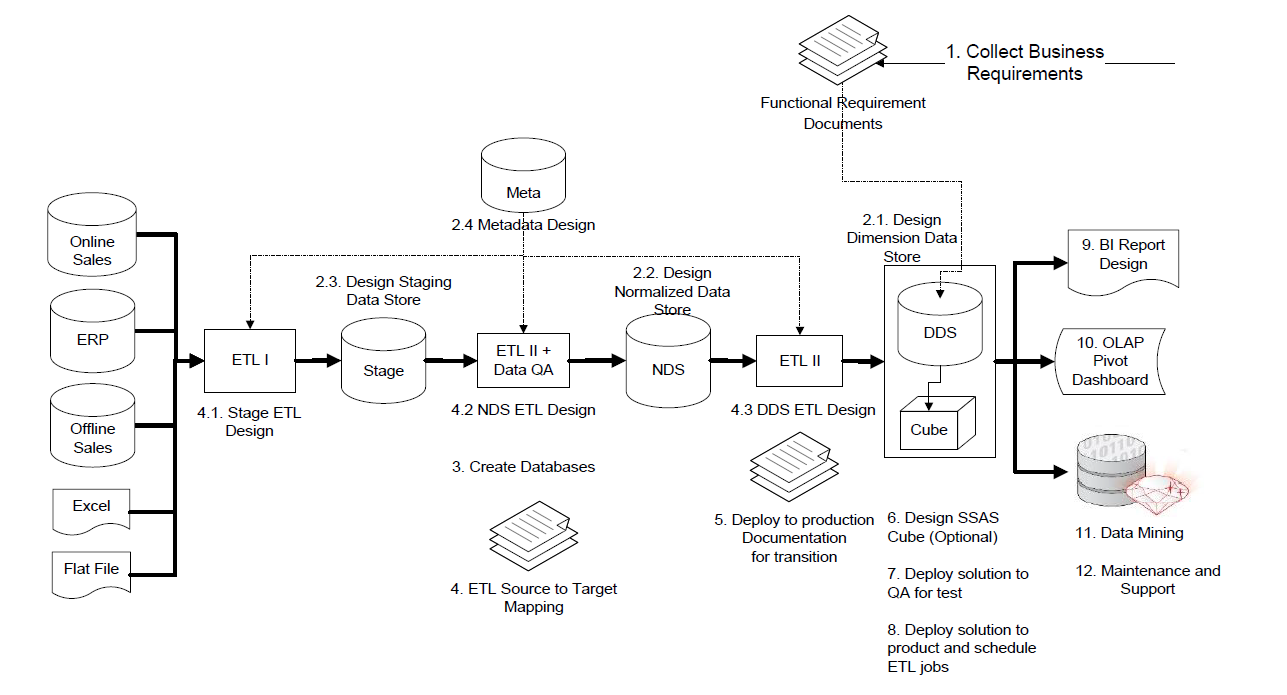
Active BI Project Notes

Contents

[1. Overview 1](#_Toc133337027)

[ 1](#_Toc133337028)

[2. Collect Business Requirements 1](#_Toc133337029)

[2.1. Functional requirements 1](#_Toc133337030)

[2.2. Non-Functional (system) requirements 1](#_Toc133337031)

[3. Design Data Warehouse 1](#_Toc133337032)

[3.1. Dimensional Data Store (DDS) or subject oriented databases with the following subjects: 1](#_Toc133337033)

[ Product Sales 1](#_Toc133337034)

[ Subscription Sales 1](#_Toc133337035)

[ Supplier Performance 1](#_Toc133337036)

[ CRM Data Marts 1](#_Toc133337037)

[ Campaign Result 1](#_Toc133337038)

[3.2. Data Integration layer, or Normalized Data Store (NDS) 1](#_Toc133337039)

[ Using ER modeling technology to design 1](#_Toc133337040)

[3.3. Staging database to unify the source data and apply data quality rules prior data loading into NDS 1](#_Toc133337041)

[4. Build a Data Warehouse 1](#_Toc133337042)

[4.1. Source to target mapping 1](#_Toc133337043)

[4.2. Load Files into Staging Database 1](#_Toc133337044)

[4.2.1. ETL Preparation 1](#_Toc133337045)

[4.2.2. Creating an End-To-End ETL Package 1](#_Toc133337046)

[4.3. Load Daily Transactions into Staging Database 1](#_Toc133337047)

[ Stage weekly External Data 1](#_Toc133337048)

[ Stage Daily Incremental 1](#_Toc133337049)

[ Stage Daily Full Reload 1](#_Toc133337050)

[ Stage Monthly Incremental 1](#_Toc133337051)

[ Stage Adhoc Full reload 1](#_Toc133337052)

[4.4. Slowly Changing Dimension 1](#_Toc133337053)

[4.5. Load NDS 1](#_Toc133337054)

[4.5.1. NDS Customer (24 tables) 1](#_Toc133337055)

[4.5.2. NDS Products & Supplier 1](#_Toc133337056)

[4.5.2.1 NDS Artist (fuzzy lookup) 1](#_Toc133337057)

[4.5.2.2 NDS Products 1](#_Toc133337058)

[4.5.3. NDS Store 1](#_Toc133337059)

[4.5.4. NDS Order 1](#_Toc133337060)

[4.6. Design package with Data Quality consideration 1](#_Toc133337061)

[4.6.1. SSIS Profiling 1](#_Toc133337062)

[ The record count of the source file must be greater than 0 1](#_Toc133337063)

[ NULL customer names are not allowed 1](#_Toc133337064)

[ ALL order dates must be less than or equal to the current date 1](#_Toc133337065)

[ No duplicate unique identifiers are allowed 1](#_Toc133337066)

[4.6.2. NDS and Stage tables row counts comparison 1](#_Toc133337067)

[4.6.3. Data quality/validation issue from the source: 1](#_Toc133337068)

[4.6.4. Source data are not used but part of NDS entities 1](#_Toc133337069)

[4.7. Load DDS – Product Sales star schema 1](#_Toc133337070)

[4.7.1. Load dimension tables 1](#_Toc133337071)

[4.7.2. Load time dimension 1](#_Toc133337072)

[4.7.3. Load fact tables 1](#_Toc133337073)

[4.8. Deploy SSIS packages 1](#_Toc133337074)

[5. Test Data Warehouse 1](#_Toc133337075)

[6. Consume Data Warehouse 1](#_Toc133337076)

[5.1. SSRS 1](#_Toc133337077)

[5.1.1. Reporting Lifecycle 1](#_Toc133337078)

[7. Power BI 1](#_Toc133337079)

# Overview

We recommend HYBRID approach for this project

# 

# Collect Business Requirements

## Functional requirements

Total 16 requirements, see details at Functional Requirements document

|  |  |  |
| --- | --- | --- |
| No. | Requirements | Priority |
| 1 | The business users need to be able to analyze “product sales” (that’s when a customer is buying a product rather than subscribing to a package) over time by geographic area, by customer demographic, by stores and sales territory, and by product hierarchy. The users also need to know the revenues, the costs, and the margin in U.S. dollars. Local currencies will be converted to U.S. dollars. | High |
| 2 | The business users need to be able to analyze “subscription sales” (that’s when a customer is subscribing to a package rather than purchasing a product) over time by geographic area, by customer demographic, by stores and sales territory, and by product hierarchy. In addition, the users also need to analyze subscription sales by subscription duration and lead source (in other words, where the customers are coming from). The users need to know the revenues, costs, and margins, evaluated every day for a period of one month. | High |

## Non-Functional (system) requirements

Total 22 requirements, see details at System Requirements document

|  |  |
| --- | --- |
| No. | Requirements |
| 1 | All data warehouse users must be able to access the data warehouse front-end applications (reports and OLAP) without logging in again. They just need to log in once to Windows on their PC or laptop. |
| 2 | The data warehouse front-end applications must not be accessible from outside the company network. |
| 3 | All front-end applications are ideally web-based, accessible from anywhere within the company network. |
| 4 | All users access the front-end applications from a central portal. The user interface layout and design need to follow company standards and guidelines. |

# Design Data Warehouse

## Dimensional Data Store (DDS) or subject oriented databases with the following subjects:

### Product Sales

### Subscription Sales

### Supplier Performance

### CRM Data Marts

### Campaign Result

## Data Integration layer, or Normalized Data Store (NDS)

### Using ER modeling technology to design

## Staging database to unify the source data and apply data quality rules prior data loading into NDS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Target\_Table** | **Source DB** | **Source\_Table** | **ETL Package Name** | **Note** |
| order\_detail | jade | order\_detail | Stage Daily Incremental | Use Meta.data\_flow |
| product | jade | product | Stage Daily Incremental | Use Meta.data\_flow |
| store | jade | store | Stage Daily Incremental | Use Meta.data\_flow |
| customer | jade | customer | Stage Daily Incremental | Use Meta.data\_flow |
| order\_header | jade | order\_header | Stage Daily Incremental | Use Meta.data\_flow |
| country | Flat file | country.dat | Stage weekly External Data |  |
| language | Excel file | langurage.xls | Stage weekly External Data |  |
| package | Jade & webtower | package |  | Jade & webtower |
| package\_type | Jade & webtower | package\_type |  | Jade & webtower |
| product\_type | jade | product\_type |  |  |
| product\_category | jade | product\_category |  |  |
| household\_income | jade | household\_income |  |  |
| product\_status | jade | product\_status |  |  |
| artist | jade | artist |  |  |
| interest | jade | interest |  |  |
| division | Jade & jupiter | division |  | Jade & jupiter |
| customer\_type | jade | customer\_type |  |  |
| email\_address\_type | jade | email\_address\_type |  |  |
| customer\_status | jade | customer\_status |  |  |
| address\_type | jade | address\_type |  |  |
| phone\_number\_type | jade | phone\_number\_type |  |  |
| format | jupiter | format |  |  |
| permission | webtower | permission |  |  |
| region | jupiter | region |  |  |
| occupation | webtower | occupation |  |  |
| state | jupiter | state |  |  |
| currency\_rate | jupiter | currency\_rate |  |  |
| currency | jupiter | currency |  |  |

# Build a Data Warehouse

## Source to target mapping

## Load Files into Staging Database

### ETL Preparation

#### Create databases

##### Create databases by using script

* Stage – Temporary database for verify source data.
* Meta – Metadata management database
* NDS – Normalized Data Store
* DDS – Dimensional Data Store

|  |  |
| --- | --- |
| Databases | # of Tables |
| DDS | 19 |
| NDS | 57 |
| Stage | 28 |
| Meta | 24 |
| jade | 24 |
| jupiter | 9 |

##### Retore Stage database from backup

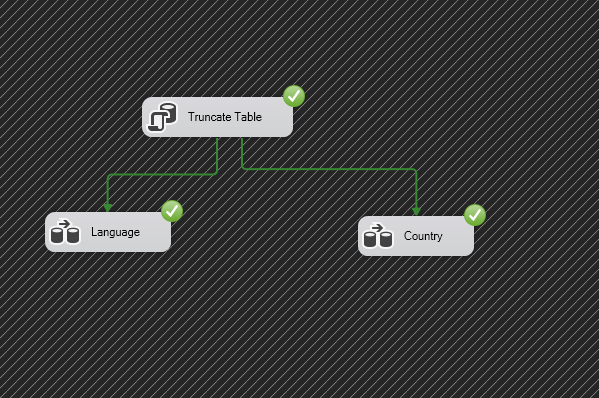
~

### Creating an End-To-End ETL Package

#### Verify source data

* Country.dat – flat file
* Language.xls – excel file

#### Implement ETL from Excel file and Flat file to the Stage database at SQR-Server



## Load Daily Transactions into Staging Database

### Stage weekly External Data

### Stage Daily Incremental

### Stage Daily Full Reload

### Stage Monthly Incremental

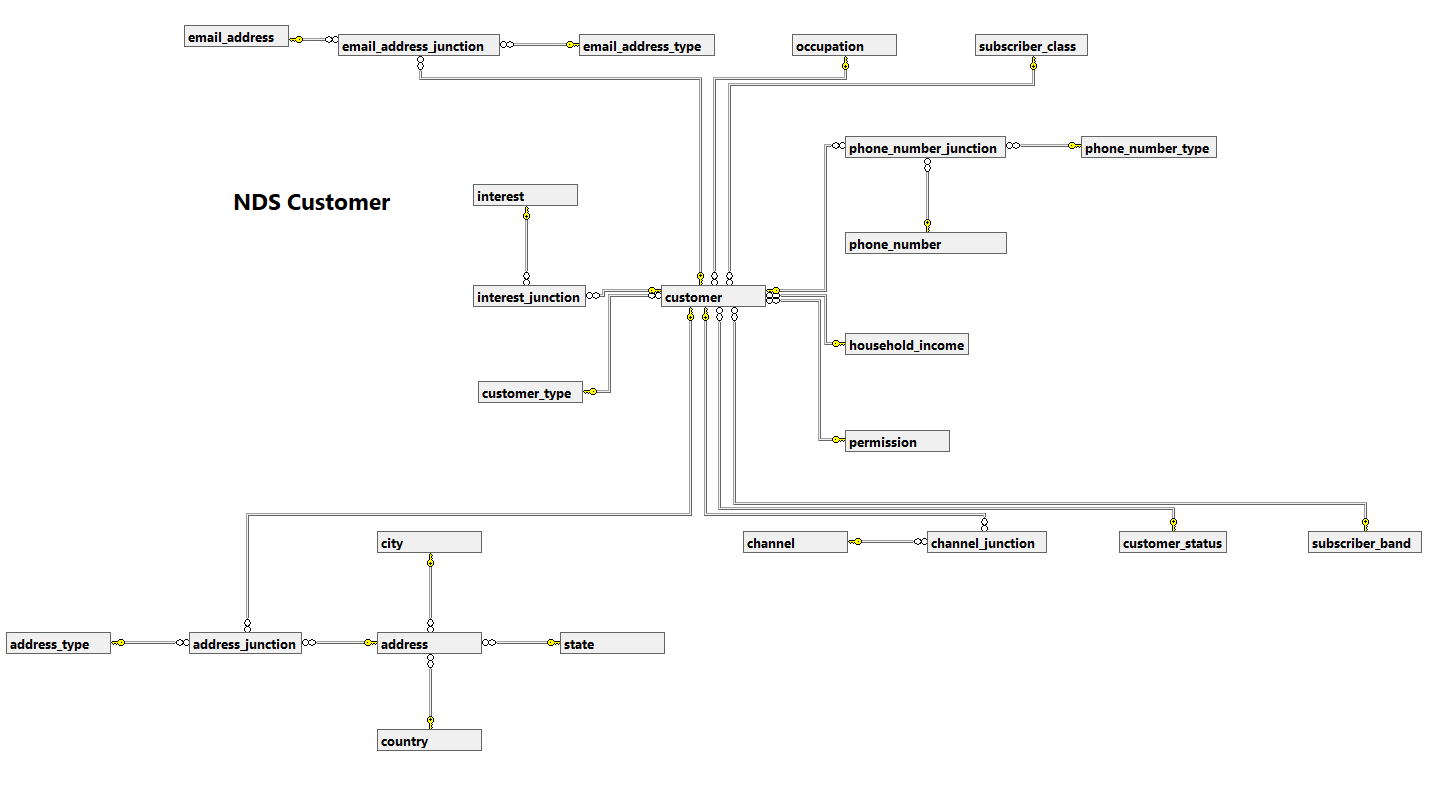
### Stage Adhoc Full reload

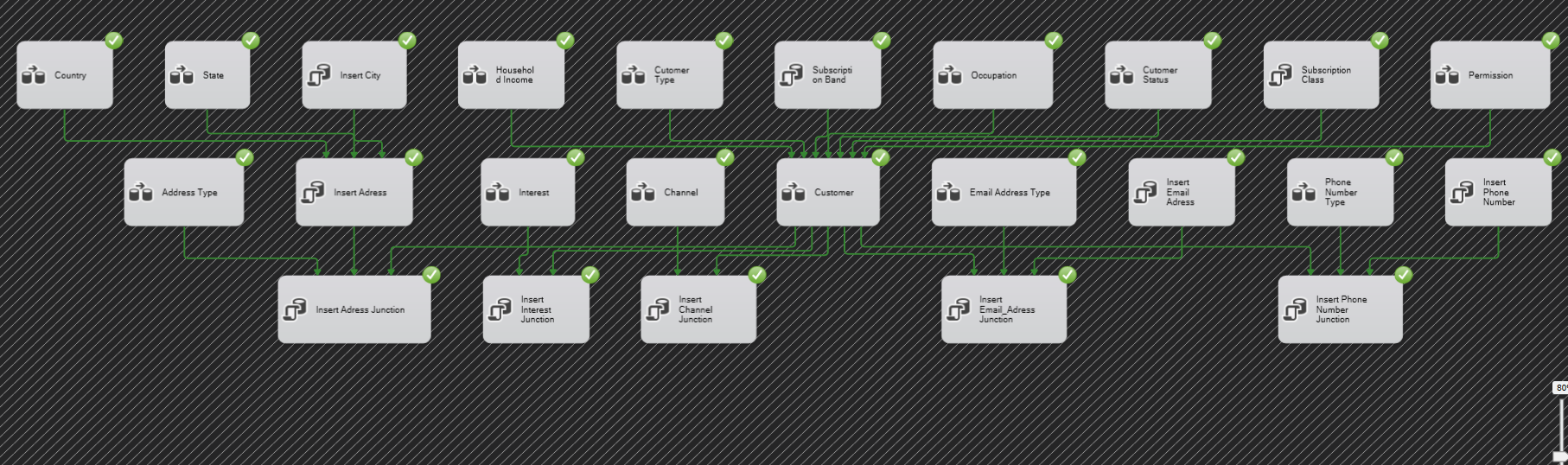
## Slowly Changing Dimension

## Load NDS

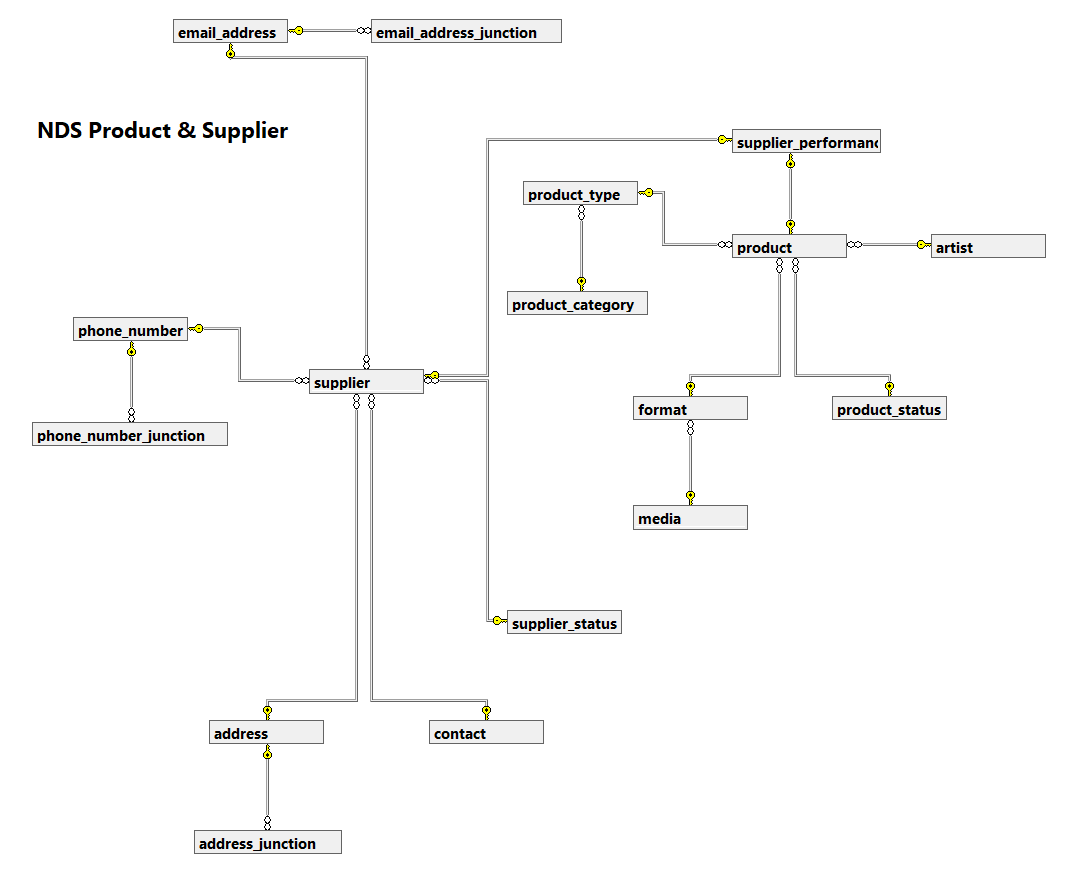
* + - Customer
    - Product
    - Store
    - Order

### NDS Customer (24 tables)

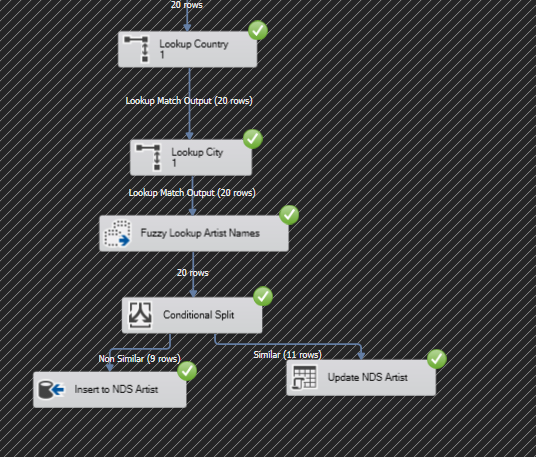




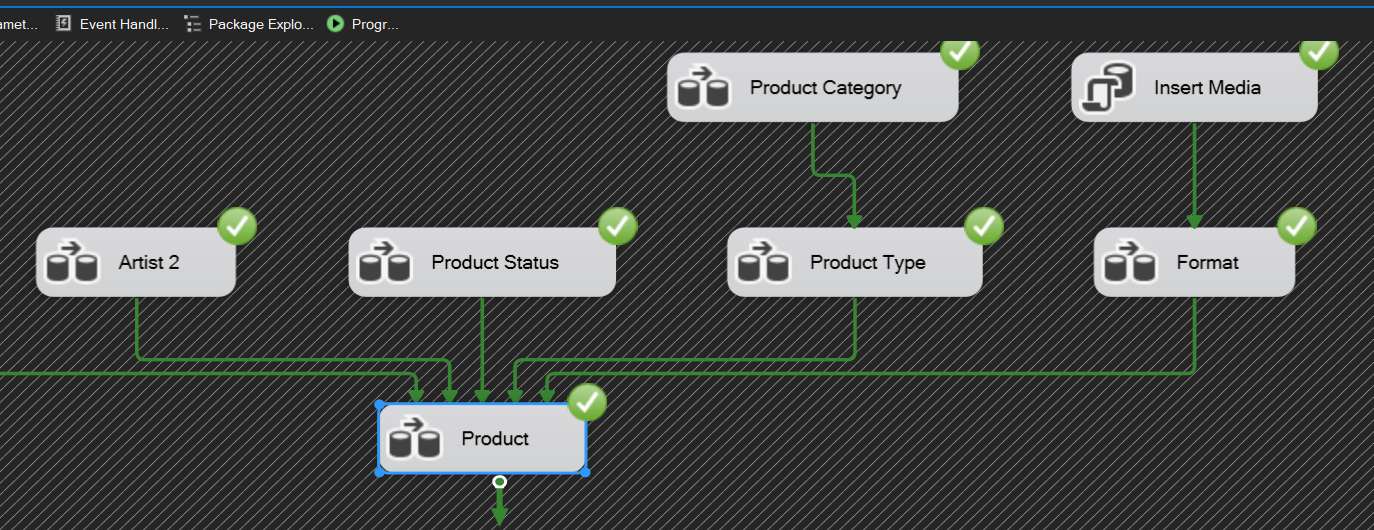
### NDS Products & Supplier



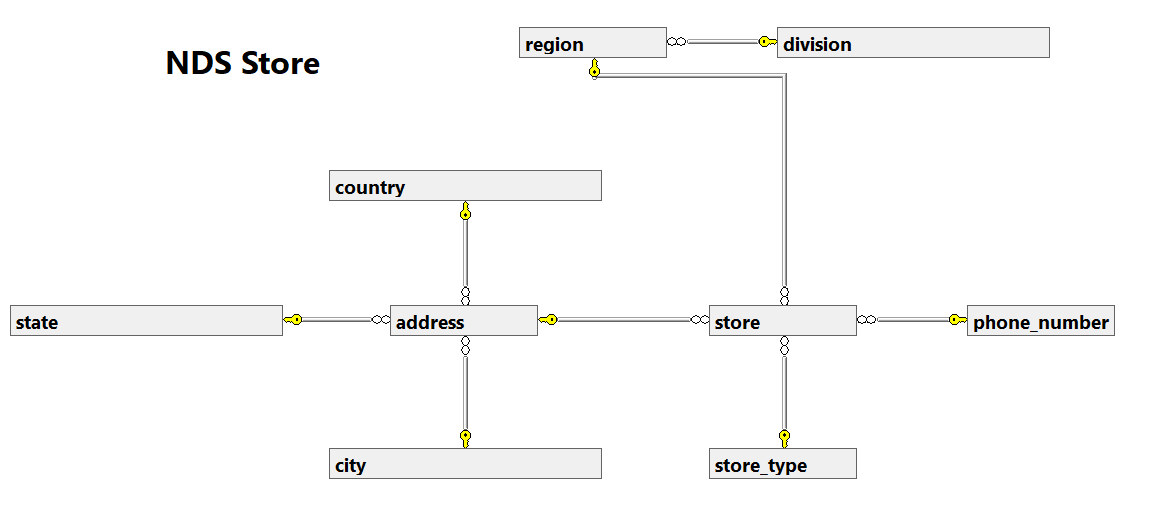
### NDS Artist (fuzzy lookup)

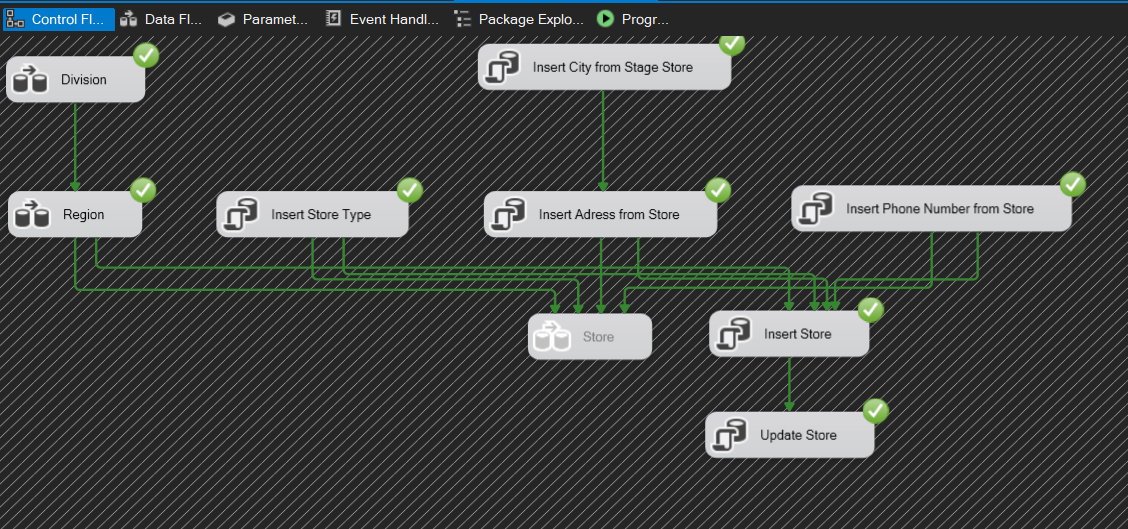


### NDS Products

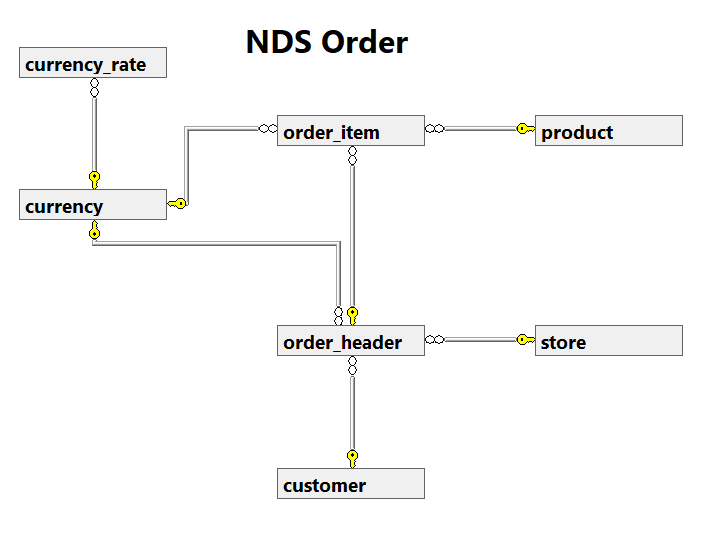


### NDS Store





### NDS Order



## Design package with Data Quality consideration

### SSIS Profiling

#### Baseline quality indicator

### The record count of the source file must be greater than 0

### NULL customer names are not allowed

### ALL order dates must be less than or equal to the current date

### No duplicate unique identifiers are allowed

### NDS and Stage tables row counts comparison

| table\_name | NDS | Stage | nds\_Stage\_diff |
| --- | --- | --- | --- |
| address | 108412 | NULL | 108412 |
| address\_junction | 49592 | NULL | 49592 |
| address\_type | 2 | 2 | 0 |
| artist | 597 | 576 | 21 |
| business\_unit | 0 | NULL | 0 |
| campaign | 0 | NULL | 0 |
| campaign\_result | 0 | NULL | 0 |
| channel | 6 | NULL | 6 |
| channel\_junction | 108316 | NULL | 108316 |
| city | 341 | NULL | 341 |
| communication | 0 | NULL | 0 |
| communication\_category | 0 | NULL | 0 |
| communication\_subscription | 0 | NULL | 0 |
| contact | 0 | NULL | 0 |
| country | 243 | 243 | 0 |
| currency | 175 | 175 | 0 |
| currency\_rate | 121 | 121 | 0 |
| customer | 54158 | 54158 | 0 |
| customer\_status | 4 | 4 | 0 |
| customer\_type | 3 | 3 | 0 |
| delivery\_status | 0 | NULL | 0 |
| delivery\_status\_category | 0 | NULL | 0 |
| division | 3 | 3 | 0 |
| email\_address | 54155 | NULL | 54155 |
| email\_address\_junction | 54158 | NULL | 54158 |
| email\_address\_type | 2 | 2 | 0 |
| format | 28 | 27 | 1 |
| household\_income | 11 | 11 | 0 |
| interest | 38 | 38 | 0 |
| interest\_junction | 160351 | NULL | 160351 |
| language | 0 | 505 | -505 |
| lead | 0 | NULL | 0 |
| media | 6 | NULL | 6 |
| occupation | 101 | 101 | 0 |
| order\_header | 44425 | 44425 | 0 |
| order\_item/ order\_detail | 63770 | 63778 | -8 |
| package | 0 | 24 | -24 |
| package\_type | 0 | 6 | -6 |
| permission | 2 | 2 | 0 |
| phone\_number | 52822 | NULL | 52822 |
| phone\_number\_junction | 54158 | NULL | 54158 |
| phone\_number\_type | 2 | 2 | 0 |
| product | 2459 | 2458 | 1 |
| product\_category | 4 | 3 | 1 |
| product\_status | 6 | 4 | 2 |
| product\_type | 66 | 65 | 1 |
| region | 11 | 11 | 0 |
| state | 50 | 50 | 0 |
| store | 102 | 102 | 0 |
| store\_type | 4 | NULL | 4 |
| subscriber\_band | 4 | NULL | 4 |
| subscriber\_class | 4 | NULL | 4 |
| subscription\_sales | 0 | NULL | 0 |
| subscription\_status | 0 | NULL | 0 |
| supplier | 0 | NULL | 0 |
| supplier\_performance | 0 | NULL | 0 |
| supplier\_status | 0 | NULL | 0 |

### Data quality/validation issue from the source:

#### Jade Order Detail

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| order\_id | line\_no | product\_code | qty | price | unit\_cost |
| 26521 | 1 | NULL | 1.00 | 1.09 | 0.87 |
| 42306 | 1 | NULL | 1.00 | 4.99 | 3.93 |
| 49526 | 1 | NULL | 2.00 | 2.59 | 1.52 |
| 63315 | 2 | NULL | 1.00 | 4.99 | 3.57 |
| 67395 | 2 | NULL | 1.00 | 1.09 | 0.65 |
| 83434 | 1 | NULL | 1.00 | 2.59 | 1.71 |
| 83919 | 1 | NULL | 1.00 | 1.99 | 1.48 |
| 95054 | 1 | NULL | 2.00 | 2.59 | 1.34 |

#### 

#### Jade Address Type

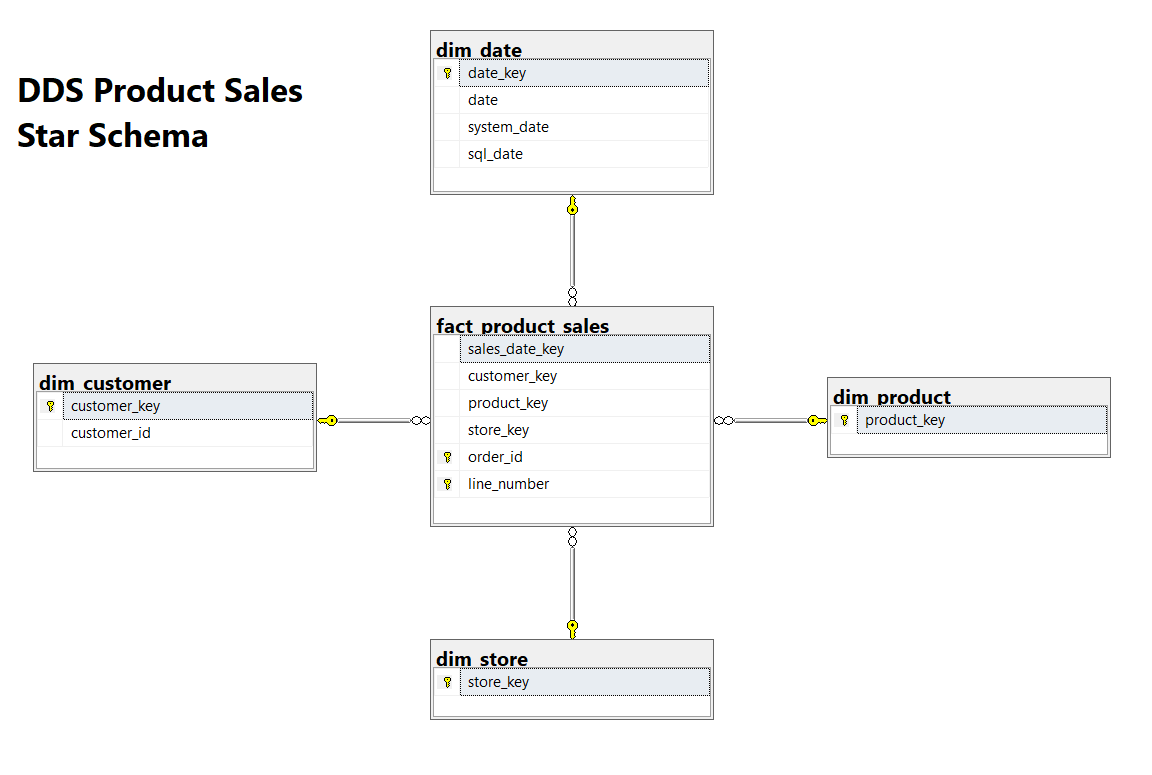
|  |  |  |
| --- | --- | --- |
| address\_type\_code | address\_type | description |
| H | Home | Home Email Address |
| W | Work | Work Email Address |

### Source data are not used but part of NDS entities

#### Package

#### Package Type

## Load DDS – Product Sales star schema



## Load dimension tables

## Load time dimension

## Load fact tables

## Deploy SSIS packages

# Test Data Warehouse

# Consume Data Warehouse

## SSRS

### Reporting Lifecycle

# Power BI