D-MART ETL PROCESS PROJECT

Introduction:

About this Project:

This document contains all the information about the DW developed for the D-mart ETL project.

SOURCE Details

Database Details

Here are the details to connect to retail db running under MySQL Database Server.

* Hostname
* Port Number
* Database Name: retail\_db
* User Name
* Password

Here are the details related to tables from which we need to pull data.

* Transactional
* Master Data Tables

Data Pipeline

Read Data

* Frequency: DAILY
* Schedule: 10 AM Every Day
* Dimensions
  + Extract data from products, categories and departments, etc
* Facts
  + Extract data from orders and order\_items based on order\_date

Process Data

Mapping Document - Master Data Tables to Dimensions

* Type 1 dimension
* Apply transformation rules based on the mapping.
* Approach 2 - Get data from all the master data tables, apply processing (join) as part of the pipeline.

We have different database

Dmart\_Ecommerce\_db

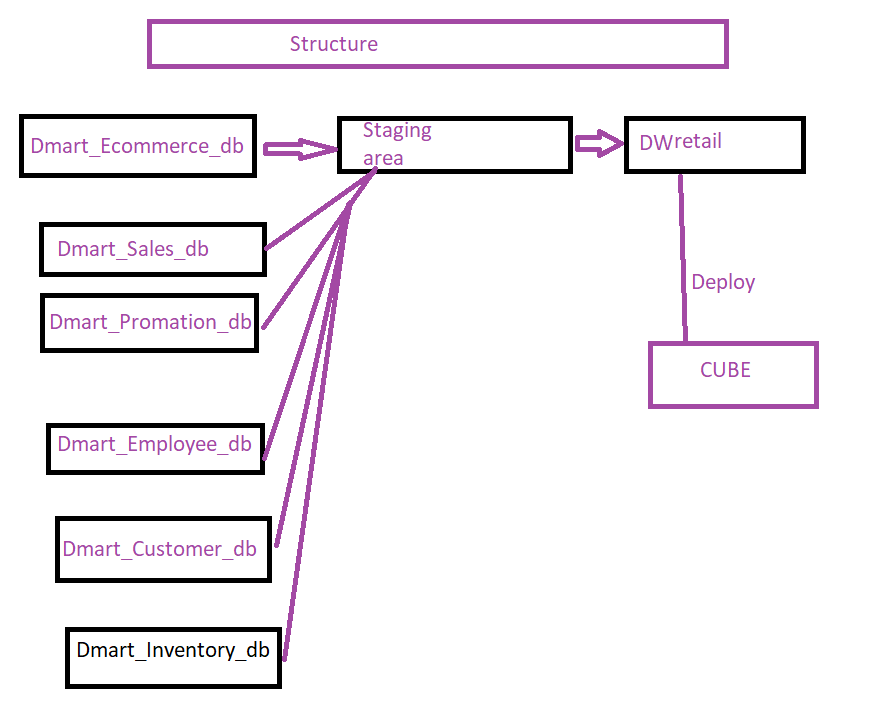
Dmart\_Sales\_db

Dmart\_Promation\_db

Dmart\_Employee\_db

Dmart\_Customer\_db

# Architecture:



## Source Data

The source data section shows a high level view of the data sources for the Data Warehousenamely the INITECH\_SQL database and Symposium storage database.

## Extraction, Transformation and Load

The extraction, transformation and load section shows a high level of the data processing and storage process. The main components involved are the FTP process which involves the Support3 server, the SSIS packagesexecutionfor the ETL, the relational database INITECH\_SQL\_STAGING and INITECH\_SQL\_DW and OLAP cube MindtreeGMSSupport.DW Components

## Database Engine

The Database Engine is the core service for storing, processing, and securing data. The Database Engine provides controlled access and rapid transaction processing to meet the requirements of the most demanding data consuming applications within your enterprise.

The Database Engine manages relational databases for online transaction processing or online analytical processing data. This includes creating tables for storing data, and database objects such as indexes, views, and stored procedures for viewing, managing, and securing data.

A relational database is a collection of data items organized as a set of formally described tables from which data can be accessed easily.

Load Data

* Frequency: DAILY
* SLA: By 4 PM every day
* Dimensions
  + Delete the existing dimension data
  + Load the snapshot from the source
* Facts
  + Facts should be loaded after dimensions are loaded

Target Details

Database Details

Here are the details related to the data warehouse into which data should be loaded.

* Hostname
* Port Number
* Database Name: retail\_dw
* User Name
* Password

Here are the details about Facts and Dimensions.

* Facts
  + fact\_product\_revenue\_dly
* Dimensions
  + dim\_product
  + dim\_customer