The Product Company

~ Final Data Mart Development Report ~

Team # 1

Team Members

Samruddhi Deshpande

Vikram Parmar

Mithila Nigudkar

Date : 05/08/2020

ISTE-DW Data Warehousing
Prof. Jai Kang

~ Table of Contents ~

I. Data Mart Design Definition

1. Universe of Discourse

This data mart defines the sales of a company within its three departments (TPC-W, TPC-E and PEC) which is calculated by the sales of the products on daily, monthly, quarterly & Yearly basis and also the other factors such as quantity, and method of shipping, delivery or payment.

Scope : This document is all about the 'SalesOrder' data mart, its grain, facts, dimensions and SCD's. It also covers aggregated data marts, few queries and ETL process performed on the data mart.

2. Information Package

Process Name: _Product sales of 3 departments in a company

Grain: Sales of each department (TPCW, TPCE and PEC) - Yearly, monthly,quarterly, weekly or daily.

Customer_Dim	Product_Dim	Supplier_Dim	OrderDate_dim	SalesDate_dim	Transaction_Details_Junk_dim
Customer_SK	Product_SK	Supplier_SK	OrderDate_SK	SalesDate_SK	TransactionDetails_SK
Cust_id(NK)	Product_ID(NK)	Supplier_id(NK)	Year	Year	MethodOfPayment
CustType	Prod_Descp	BU_Name	FiscalYear	FiscalYear	MethodOfOrder
CustTypeID	ProdTypeID	SupplierAddress1	Quarter	Quarter	MethodOfShipping
Cust_Name	Buid	SupplierAddress2	FiscalQuarter	FiscalQuarter	Department
Cust_Address1	Supplier_Name	SupplierCity	Month	Month	
Cust_Address2	Abbrev	SupplierState	FiscalMonth	FiscalMonth	

Cust_City	Price1	SupplierZipcode	Week	Week	
Cust_State	Price2	Department	FiscalWeek	FiscalWeek	
Cust_Zip	Unit_Cost		Day	Day	
Department	Type_Description		FiscalDay	FiscalDay	
	Department		Date	Date	
				Department	
Measured Facts:	**invoiceID, ShippingCost, SalesAmount, Days to Ship, Discounted, Qty,Department**				

3. Entity Definitions

Table Name	Attribute Name	Description
Customer_Dim	Customer_SK	Surrogate key for customer dimension table
	Cust_id(NK)	The ID of the customer(Natural key)
	CustType	The type of customer
	CustTypeID	This is the ID for the CustType.
	Cust_Name	The name of the customer
	Cust_Address1	The first address line of the customer
	Cust_Address2	The second address line of the customer
	Cust_City	The city of the customer
	Cust_State	The offical USPS state abbreviation
	Cust_Zip	The 5 digit USPS zip code
	Department	Department of the customer

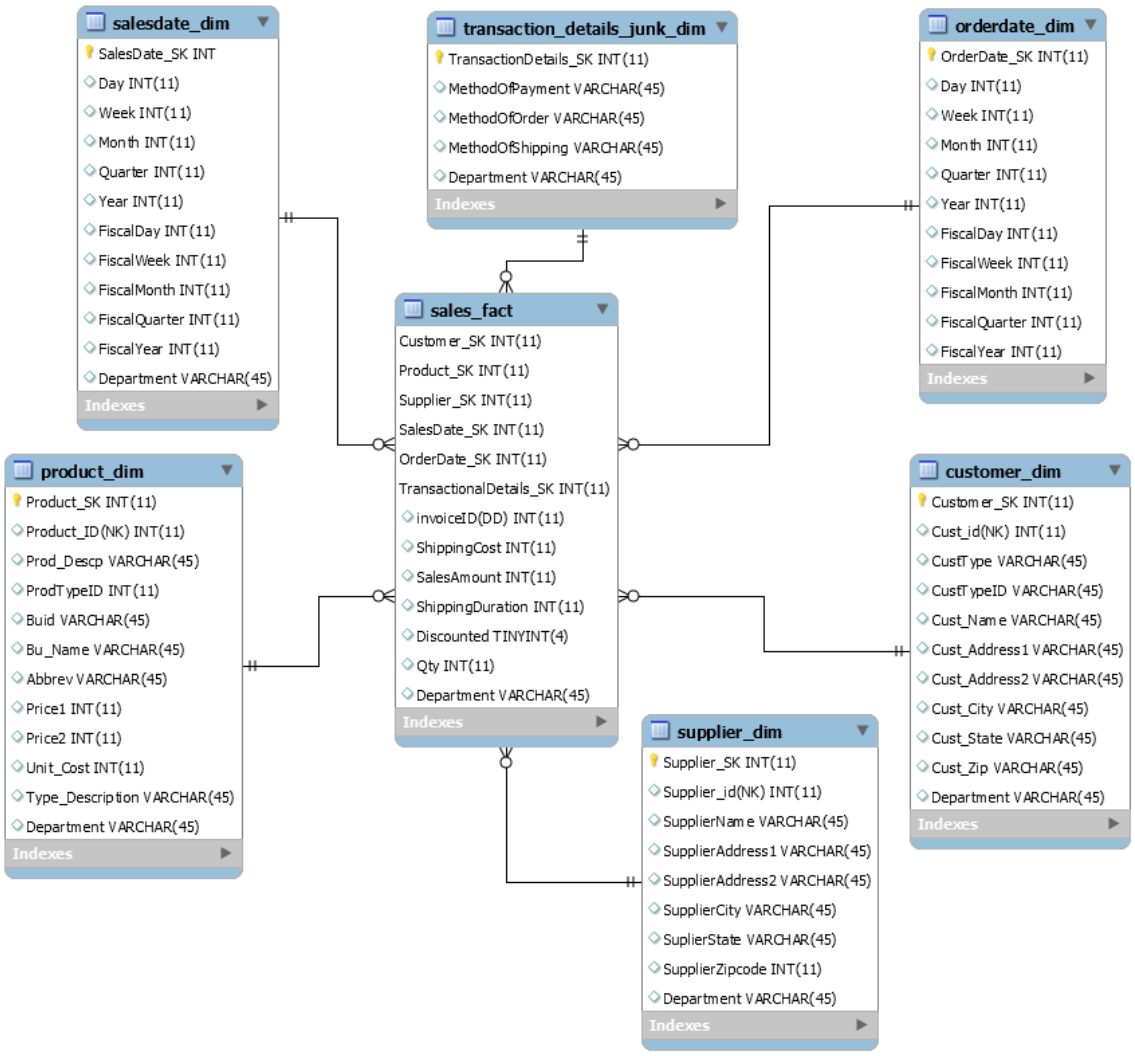
Table Name	Attribute Name	Description
Product_Dim	Product_SK	Surrogate key for product dimension table
	Product_ID(NK)	The ID of the product(Natural key)
	Prod_Descp	The description of the product
	ProdTypeID	The ID for the product type
	Buid	The Business unit ID
	BU_Name	The business unit name
	Abbrev	The business unit abbreviation
	Price1	The standard price of the product per unit
	Price2	The discounted price of the product per unit
	Unit_Cost	The standard cost of the product per unit
	Type_Description	The description for the product type
	Department	Department to which the product associates
Supplier_Dim	Supplier_SK	Surrogate key for supplier dimension table
	Supplier_id(NK)	The ID of the supplier (Natural key)
	SupplierName	The name of the supplier
	SupplierAddress1	The first address line of the supplier
	SupplierAddress2	The second address line of the supplier
	SupplierCity	The city of the supplier

Table Name	Attribute Name	Description
	SupplierState	The official USPS state abbreviation
	SupplierZipcode	The 5 digit USPS zip code
	Department	Department to which the product associates
OrderDate_dim	OrderDate_SK	Surrogate key for OrderDate dimension table
	Year	Year from the date of order
	FiscalYear	Fiscal Year from the date of order
	Quarter	Quarter from the date of order
	FiscalQuarter	Fiscal Quarter from the date of order
	Month	Month from the date of order
	FiscalMonth	Fiscal Month from the date of order
	Week	Week from the date of order
	FiscalWeek	Fiscal Week from the date of order
	Day	Day from the date of order
	FiscalDay	Fiscal Day from the date of order
	Date	For Fact table
SalesDate_dim	SalesDate_SK	Surrogate key for SalesDate dimension table
	Year	Year from the date of sales

Table Name	Attribute Name	Description
	FiscalYear	Fiscal Year from the date of sales
	Quarter	Quarter from the date of sales
	FiscalQuarter	Fiscal Quarter from the date of sales
	Month	Month from the date of sales
	FiscalMonth	Fiscal Month from the date of sales
	Week	Week from the date of sales
	FiscalWeek	Fiscal Week from the date of sales
	Day	Day from the date of sales
	FiscalDay	Fiscal Day from the date of sales
	Department	Department to which the product associates
	Date	For Fact table
Transaction_Details_Junk_dim	TransactionDetails_SK	Surrogate key for Junk dimension table
	MethodOfPayment	Method by which order payment was done
	MethodOfOrder	Method by which order was placed
	MethodOfShipping	Method by which order was shipped to the customer
	Department	Department to which the product associates
sales_fact	Customer_SK	Surrogate key for customer dimension table

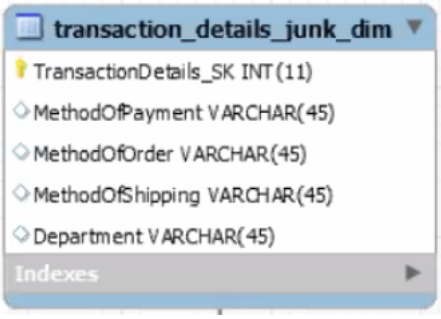
Table Name	Attribute Name	Description
	Product_SK	Surrogate key for product dimension table
	Supplier_SK	Surrogate key for supplier dimension table
	OrderDate_SK	Surrogate key for OrderDate dimension table
	SalesDate_SK	Surrogate key for SalesDate dimension table
	TransactionDetails_SK	Surrogate key for Junk dimension table
	invoiceID	The ID of the invoice
	ShippingCost	Cost of shipping the order
	SalesAmount	The dollar amount of the sale
	Days to Ship	Total days needed to ship the order
	Discounted	Indicate whether the product is discounted or not.
	Qty	The quantity purchased
	Department	Department to which the product associates

II. Dimensional Model

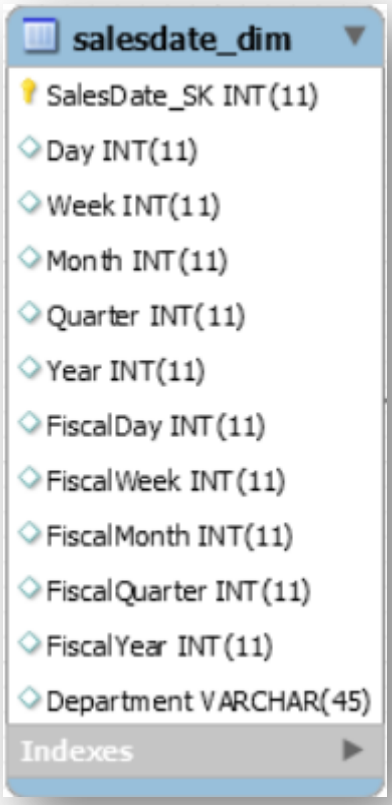


The above data mart of all the divisions consist of 6 dimensions viz. Product_dim, Customer_dim, Supplier_dim, orderdate_dim, salesdate_dim and 1 fact table sales_fact table

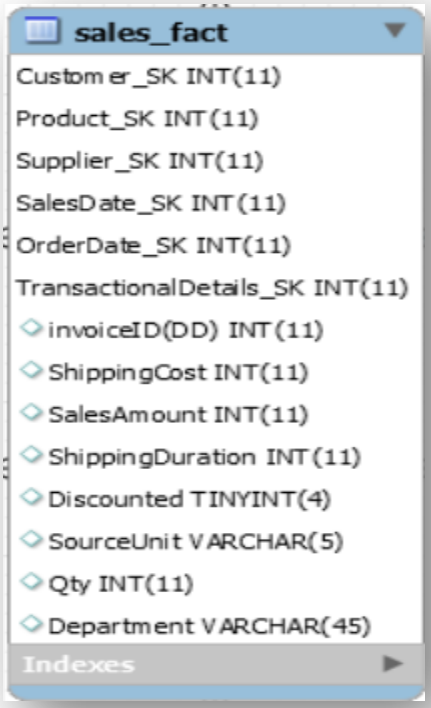
The fact table consist of measurable facts like ShippingCost, ShippingDuration, Discounted, SourceUnit, Quantity, Sales_Amount



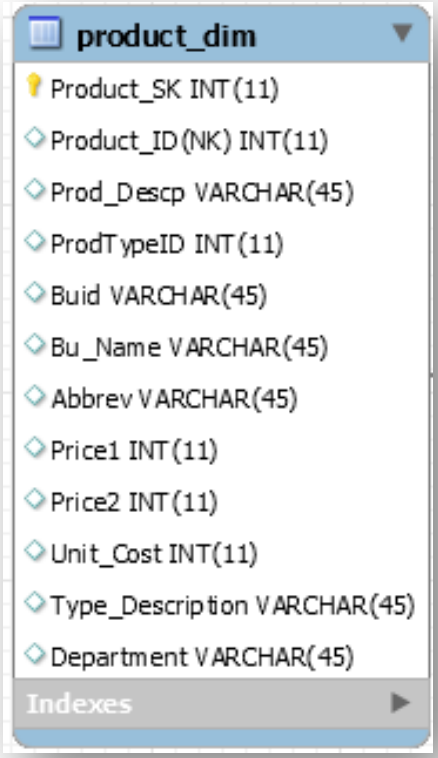
Junk_Dimension



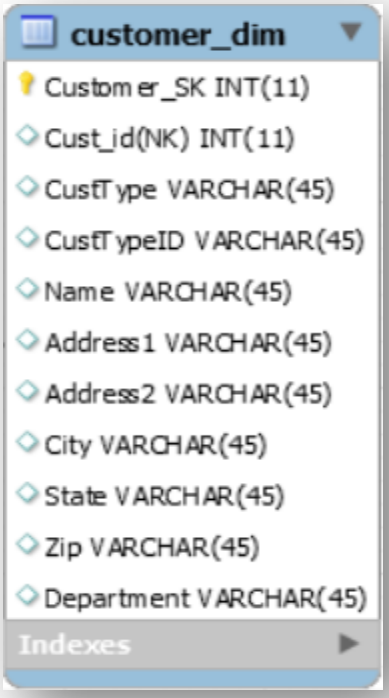
Sales_date Dimension



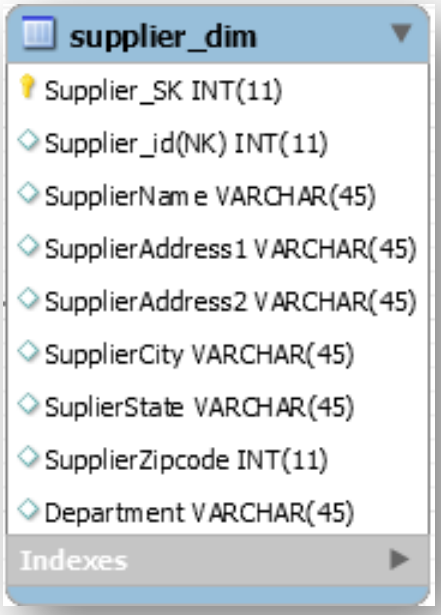
Sales_Fact Dimension table



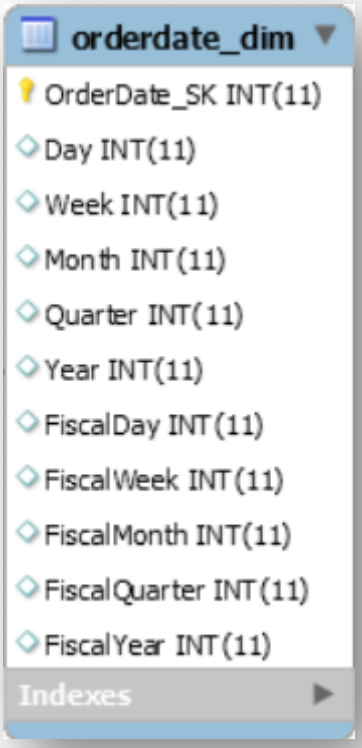
Product_Dimesion



Customer_Dimesion table



Supplier Dimension table



Order_Date Dimension

III. Data Staging: ETL – Data Extract File Definitions

Dimension Table	Department	File	Format
Customer_Dim	TPCE	customer	csv
	TPCE	customer_type	csv
	TPCW	TPCWcustomer	csv
	TPCW	TPCWcustomer_type	csv
	PEC	PECcustomer	csv
	PEC	PECcustomer_type	csv
Product_Dim	TPCE	product	csv
	TPCE	prod_type	csv
	TPCW	TPCWproduct	csv
	TPCW	TPCWproduct_type	csv
	PEC	PECproduct	csv
	PEC	PECproduct_type	csv
Supplier_Dim	TPCE	supplier	csv
	TPCW	TPCWproduct	csv
	PEC	PECproduct	csv
OrderDate_dim	PEC	PECinvoice	csv
SalesDate_dim	TPCE	invoice	csv
	TPCW	TPCWinvoice	csv
	PEC	PECinvoice	csv
Transaction_Details_Junk_dim	PEC	PECinvoice	csv
sales_fact	TPCE	invoice	csv
	TPCE	invoice_details	csv
	TPCW	TPCWinvoice	csv
	PEC	PECinvoice	csv

IV. Data Staging: ETL – Source-to-Target Mappings

Follow the same format as indicated in "The Data Warehouse ETL Toolkit" by Kimball & Caserta, Fig. 3.1 on page 60. This is available on Books 24x7. The table should be in alphabetical order table name and column name.

Target					**Source**				
Table Name	**Column Name**	**Data Type**	**Table Type**	**SCD Type**	**Database Name**	**Table Name**	**Column Name**	**Data Type**	**Transformation**
Customer_Dim	Customer_SK	**INT**	Dimension	**0**				INT	Added in Pentaho
	Cust_id(NK)	**INT**	Dimension	**0**	TPCE TPCW PEC	customer(CSV) TPCWcustomer(CSV) PECcustomer(CSV)	CUSTID custID custID	INT	Changes in attribute name
	CustType	**VARCHAR**	Dimension	**1**			custtype custtype custtype	VARCHAR	Changes in attribute name
	CustTypeID	**VARCHAR**	Dimension	**2**			CUSTTYPEID custtypeID custtypeID	VARCHAR	Changes in attribute name
	Cust_Name	**VARCHAR**	Dimension	**2**			NAME name name	VARCHAR	Changes in attribute name
	Cust_Address1	**VARCHAR**	Dimension	**2**			ADDR 1 address address	VARCHAR	Changes in attribute name Divide : Address 1 and Address 2
	Cust_Address2	**VARCHAR**	Dimension	**2**			ADDR 2 address address	VARCHAR	Changes in attribute name Divide : Address 1 and Address 2
	Cust_City	**VARCHAR**	Dimension	**2**			CITY city city	VARCHAR	No Changes

Target						Source				
	Cust_State	**VARCHAR**	Dimension	**2**			STATE state state	VARCHAR	Changed to Upper Case
	Cust_Zip	**VARCHAR**	Dimension	**2**			ZIP Zip Zip	VARCHAR	No Changes
	Department	**VARCHAR**	Dimension	**2**				VARCHAR	Added in Pentaho
Product_Dim	Product_SK	**INT**	Dimension	**0**	TPCE TPCW PEC			INT	Added in Pentaho
	Product_ID(NK)	**INT**	Dimension	**0**		Product CSV) TPCWproductCSV) PECproduct (CSV)	PRODID No Header prodid	INT	Changes in attribute name
	Prod_Descp	**VARCHAR**	Dimension	**0**			DESCRIPTION No Header prodDescription	VARCHAR	Changes in attribute name
	ProdTypeID	**INT**	Dimension	**2**			productTypeID NO Header productTypeID	INT	Changes in attribute name
	Buid	**VARCHAR**	Dimension	**0**		business_unit (CSV)	BUID Buid Buid	VARCHAR	No Changes
	BU_Name	**VARCHAR**	Dimension	**2**		TPCW business_unit (CSV)	NAME name Name	VARCHAR	Changes in attribute name
	Abbrev	**VARCHAR**	Dimension	**2**		PECbusiness_unit (CSV)	ABBREV Abbrev Abbrev	VARCHAR	Added Miscellaneous
	Price1	**INT**	Dimension	**2**		Product CSV) TPCWproductCSV) PECproduct (CSV)	PRICE1 Price1 Price1	INT	Changed to number
	Price2	**INT**	Dimension	**2**			PRICE2 Price2 Price2	INT	Changed to number
	Unit_Cost	**INT**	Dimension	**2**			UNITCOST Unitcost Unitcost	INT	Changed to number

Target						Source			
	Type_Description	**VARCHAR**	Dimension	**2**		Product_type(CSV) TPCWproduct_type(CSV) PECproduct_type (CSV)	typeDescription NoHeader typeDescription	VARCHAR	Changes in attribute name
	Department	**VARCHAR**	Dimension	**2**				VARCHAR	Added in Pentaho
Supplier_Dim	Supplier_SK	**INT**	Dimension	**0**	TPCE TPCW PEC			INT	Added in Pentaho
	Supplier_id(NK)	**INT**	Dimension	**0**		Supplier(CSV) TPCW product (CSV) PECproduct(CSV)	SUPPLIERID No Header suppiler ID	INT	Changes in attribute name
	SupplierName	**VARCHAR**	Dimension	**2**			NAME No Header Name	VARCHAR	Changes in attribute name
	SupplierAddress1	**VARCHAR**	Dimension	**2**			ADDR 1 No Header Address	VARCHAR	Split address in address1 and 2
	SupplierAddress2	**VARCHAR**	Dimension	**2**			ADDR 2 No Header Address	VARCHAR	Split address in address1 and 2
	SupplierCity	**VARCHAR**	Dimension	**2**			CITY No Header City	VARCHAR	No Changes
	SupplierState	**VARCHAR**	Dimension	**2**			STATE No Header State	VARCHAR	Changed to Upper Case
	SupplierZipcode	**INT**	Dimension	**2**			ZIP No Header Zip	INT	No Changes
	Department	**VARCHAR**	Dimension	**2**				VARCHAR	Added in Pentaho
OrderDate_dim	OrderDate_SK	**INT**	Dimension	**0**	TPCE TPCW PEC			INT	Added in Pentaho

Target						Source			
	Year	INT	Dimension	2		PECinvoice(CSV)	orderDate	INT	Split date
	FiscalYear	INT	Dimension	2				INT	Split date
	Quarter	INT	Dimension	2				INT	Split date
	FiscalQuarter	INT	Dimension	2				INT	Split date
	Month	INT	Dimension	2				INT	Split date
	FiscalMonth	INT	Dimension	2				INT	Split date
	Week	INT	Dimension	2				INT	Split date
	FiscalWeek	INT	Dimension	2				INT	Split date
	Day	INT	Dimension	2				INT	Split date
	FiscalDay	INT	Dimension	2				INT	Split date
SalesDate_dim	SalesDate_SK	INT	Dimension	0	TPCE TPCW PEC	invoice (CSV) TPCWinvioce(CSV) PECinvoice(CSV)	salesDate	INT	Added in Pentaho
	Year	INT	Dimension	2				INT	Split date
	FiscalYear	INT	Dimension	2				INT	Split date
	Quarter	INT	Dimension	2				INT	Split date
	FiscalQuarter	INT	Dimension	2				INT	Split date
	Month	INT	Dimension	2				INT	Split date
	FiscalMonth	INT	Dimension	2				INT	Split date
	Week	INT	Dimension	2				INT	Split date
	FiscalWeek	INT	Dimension	2				INT	Split date

Target					Source				
	Day	INT	Dimension	2				INT	Split date
	FiscalDay	INT	Dimension	2				INT	Split date
Transaction_Details_Junk_dim	TransactionDetails_SK	INT	Dimension	0				INT	Added in Pentaho
	MethodOfPayment	VARCHAR	Dimension	1	PEC TPCE TPCW	PECinvoice(CSV)	shippingMethod	VARCHAR	No Changes
	MethodOfOrder	VARCHAR	Dimension	1			paymentMethod	VARCHAR	No Changes
	MethodOfShipping	VARCHAR	Dimension	1			orderMethod	VARCHAR	No Changes
	Department	VARCHAR	Dimension	1				VARCHAR	Added in Pentaho
sales_fact	Customer_SK	INT	Fact			Customer_Dim		INT	Pulled from Customer_Dim
	Product_SK	INT	Fact			Produce_Dim		INT	Pulled from Product_Dim
	Supplier_SK	INT	Fact			Supplier_Dim		INT	Pulled from Supplier_Dim
	OrderDate_SK	INT	Fact			OrderDate_dim		INT	Pulled from OrderDate_Dim
	SalesDate_SK	INT	Fact			SaleDate_Dim		INT	Pulled from SalesDate_Dim
	TransactionDetails_Junk_SK	INT	Fact			TransactionDetails_dim		INT	Pulled from TransactionDetails_Junk_Dim
	invoiceID	INT	Fact	1	TPCE TPCW PEC	invoice(CSV) TPCWinvoice(CSV) PECinvoice(CSV)	INVOICEID Invoiced Invoiced	INT	

Target					Source				
	ShippingCost	**INT**	Fact	**1**				INT	Calculated in pentaho
	SalesAmount	**INT**	Fact	**1**	TPCE TPCW PEC	invoice (CSV) TPCWinvoice(CSV) PECinvoice(CSV)	AMT Amt Amt	INT	
	Days to Ship	**INT**	Fact	**1**				INT	Calculated in pentaho
	Discounted	**BOOLEAN**	Fact	**1**	TPCE TPCW PEC	invoice_details (CSV) TPCWinvoice(CSV) PECinvoice(CSV)	DISCOUNTED	BOOLEAN	
	Qty	**INT**	Fact	**1**			QTY Qty Qty	INT	
	Department	**VARCHAR**	Fact	**1**				VARCHAR	Added in Pentaho

. SQL Code – Tables & Constraints

```
CREATE DATABASE  IF NOT EXISTS `sales_order` /*!40100 DEFAULT CHARACTER SET utf8mb4
COLLATE utf8mb4_0900_ai_ci */ /*!80016 DEFAULT ENCRYPTION='N' */;
USE `sales_order`;
-- MySQL dump 10.13  Distrib 8.0.21, for macos10.15 (x86_64)
--
-- Host: localhost    Database: sales_order
-- ------------------------------------------------------
-- Server version  8.0.23

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!50503 SET NAMES utf8 */;
/*!40103 SET @OLD_TIME_ZONE=@@TIME_ZONE */;
/*!40103 SET TIME_ZONE='+00:00' */;
/*!40014 SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0 */;
/*!40014 SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS,
FOREIGN_KEY_CHECKS=0 */;
/*!40101 SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='NO_AUTO_VALUE_ON_ZERO'
*/;
/*!40111 SET @OLD_SQL_NOTES=@@SQL_NOTES, SQL_NOTES=0 */;

--
-- Table structure for table `customer_dim`
--

DROP TABLE IF EXISTS `customer_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `customer_dim` (
  `Customer_SK` int NOT NULL AUTO_INCREMENT,
  `Cust_id(NK)` int DEFAULT NULL,
  `CustType` varchar(45) DEFAULT NULL,
  `CustTypeID` varchar(45) DEFAULT NULL,
  `Cust_Name` varchar(45) DEFAULT NULL,
  `Cust_Address1` varchar(45) DEFAULT NULL,
  `Cust_Address2` varchar(45) DEFAULT NULL,
  `Cust_City` varchar(45) DEFAULT NULL,
  `Cust_State` varchar(45) DEFAULT NULL,
  `Cust_Zip` varchar(45) DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Customer_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=122 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `customer_dim`
--

LOCK TABLES `customer_dim` WRITE;
/*!40000 ALTER TABLE `customer_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `customer_dim` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `orderdate_dim`
--
```

```
DROP TABLE IF EXISTS `orderdate_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `orderdate_dim` (
  `OrderDate_SK` int NOT NULL AUTO_INCREMENT,
  `Day` int DEFAULT NULL,
  `Week` int DEFAULT NULL,
  `Month` int DEFAULT NULL,
  `Quarter` int DEFAULT NULL,
  `Year` int DEFAULT NULL,
  `FiscalDay` int DEFAULT NULL,
  `FiscalWeek` int DEFAULT NULL,
  `FiscalMonth` int DEFAULT NULL,
  `FiscalQuarter` int DEFAULT NULL,
  `FiscalYear` int DEFAULT NULL,
  `Date` date DEFAULT NULL,
  PRIMARY KEY (`OrderDate_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=2158 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `orderdate_dim`
--

LOCK TABLES `orderdate_dim` WRITE;
/*!40000 ALTER TABLE `orderdate_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `orderdate_dim` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `product_dim`
--

DROP TABLE IF EXISTS `product_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `product_dim` (
  `Product_SK` int NOT NULL AUTO_INCREMENT,
  `Product_ID(NK)` int DEFAULT NULL,
  `Prod_Descp` varchar(45) DEFAULT NULL,
  `ProdTypeID` int DEFAULT NULL,
  `Buid` varchar(45) DEFAULT NULL,
  `Bu_Name` varchar(45) DEFAULT NULL,
  `Abbrev` varchar(45) DEFAULT NULL,
  `Price1` int DEFAULT NULL,
  `Price2` int DEFAULT NULL,
  `Unit_Cost` int DEFAULT NULL,
  `Type_Description` varchar(45) DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Product_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=109 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `product_dim`
--

LOCK TABLES `product_dim` WRITE;
/*!40000 ALTER TABLE `product_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `product_dim` ENABLE KEYS */;
```

```
UNLOCK TABLES;

--
-- Table structure for table `sales_fact`
--

DROP TABLE IF EXISTS `sales_fact`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `sales_fact` (
  `Customer_SK` int NOT NULL,
  `Product_SK` int NOT NULL,
  `Supplier_SK` int NOT NULL,
  `SalesDate_SK` int NOT NULL,
  `OrderDate_SK` int NOT NULL,
  `TransactionalDetails_SK` int NOT NULL,
  `invoiceID(DD)` int DEFAULT NULL,
  `ShippingCost` int DEFAULT NULL,
  `SalesAmount` int DEFAULT NULL,
  `ShippingDuration` int DEFAULT NULL,
  `Discounted` tinyint DEFAULT NULL,
  `Qty` int DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  PRIMARY KEY
(`TransactionalDetails_SK`,`SalesDate_SK`,`OrderDate_SK`,`Product_SK`,`Customer_SK`,`Supplier_SK`
)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `sales_fact`
--

LOCK TABLES `sales_fact` WRITE;
/*!40000 ALTER TABLE `sales_fact` DISABLE KEYS */;
/*!40000 ALTER TABLE `sales_fact` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `salesdate_dim`
--

DROP TABLE IF EXISTS `salesdate_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `salesdate_dim` (
  `SalesDate_SK` int NOT NULL AUTO_INCREMENT,
  `Day` int DEFAULT NULL,
  `Week` int DEFAULT NULL,
  `Month` int DEFAULT NULL,
  `Quarter` int DEFAULT NULL,
  `Year` int DEFAULT NULL,
  `FiscalDay` int DEFAULT NULL,
  `FiscalWeek` int DEFAULT NULL,
  `FiscalMonth` int DEFAULT NULL,
  `FiscalQuarter` int DEFAULT NULL,
  `FiscalYear` int DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  `Date` date DEFAULT NULL,
  PRIMARY KEY (`SalesDate_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=198459 DEFAULT CHARSET=utf8;
```

```
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `salesdate_dim`
--

LOCK TABLES `salesdate_dim` WRITE;
/*!40000 ALTER TABLE `salesdate_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `salesdate_dim` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `supplier_dim`
--

DROP TABLE IF EXISTS `supplier_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `supplier_dim` (
  `Supplier_SK` int NOT NULL AUTO_INCREMENT,
  `Supplier_id(NK)` int DEFAULT NULL,
  `SupplierName` varchar(45) DEFAULT NULL,
  `SupplierAddress1` varchar(45) DEFAULT NULL,
  `SupplierAddress2` varchar(45) DEFAULT NULL,
  `SupplierCity` varchar(45) DEFAULT NULL,
  `SuplierState` varchar(45) DEFAULT NULL,
  `SupplierZipcode` int DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Supplier_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=14 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `supplier_dim`
--

LOCK TABLES `supplier_dim` WRITE;
/*!40000 ALTER TABLE `supplier_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `supplier_dim` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `transaction_details_junk_dim`
--

DROP TABLE IF EXISTS `transaction_details_junk_dim`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `transaction_details_junk_dim` (
  `TransactionDetails_SK` int NOT NULL AUTO_INCREMENT,
  `MethodOfPayment` varchar(45) DEFAULT NULL,
  `MethodOfOrder` varchar(45) DEFAULT NULL,
  `MethodOfShipping` varchar(45) DEFAULT NULL,
  `Department` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`TransactionDetails_SK`)
) ENGINE=InnoDB AUTO_INCREMENT=48 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `transaction_details_junk_dim`
--
```

```
LOCK TABLES `transaction_details_junk_dim` WRITE;
/*!40000 ALTER TABLE `transaction_details_junk_dim` DISABLE KEYS */;
/*!40000 ALTER TABLE `transaction_details_junk_dim` ENABLE KEYS */;
UNLOCK TABLES;
/*!40103 SET TIME_ZONE=@OLD_TIME_ZONE */;

/*!40101 SET SQL_MODE=@OLD_SQL_MODE */;
/*!40014 SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS */;
/*!40014 SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS */;
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;
/*!40111 SET SQL_NOTES=@OLD_SQL_NOTES */;

-- Dump completed on 2021-05-08 19:10:56
```

VI. Data Staging Activities - ETL

1. Data Cleansing

DM Table	Attribute	Problem	Resolution Strategy (attach code)
PECcustomer	Custtype	The data was enclosed in double-quotes The data was not uniform Ex: SATELOCALGOVT, USGOVT,COMERCIAL	The double quotes where replaced with space Replaced the ununiform data to maintain data uniformity throughout the data mart
PECcustomer	Address	The data was too big in one attribute The data was not uniform Ex: St, Street; Rd. Road; Ave. Avenue	Added Comma delimiter in address field to split into addr1 & addr2 in csv file of customer. Replaced the data where necessary to have consistent data Made Rd. to Road St. to Street Ave. to Avenue
PECcustomer	Name	Partial records had Inc. while some had Incorporation Similarly some data had Co. while some had Company	Standardized the name, Inc to Inc. Co. to Company Corp to Corporation by replace in string method in pentaho.

PECcustomer	Add1 & Addr2	Some records had St, Ave and Rd etc. so standardized them	Standardized the addr1 & addr2, Rd. to Road St. to Street Dr. to Drive Av. To Avenue Ave. to Avenue by using replace in string in pentaho.
TPCWcustomer	All attributes	All attributes- values where enclosed in double quotes	Replaced double quotes & Comma with space in the fields using replace in string in pentaho.
TPCWCcustomer	State	The records where inconsistent with PEC state	Standardized all the state names with their ABBREV using replace in string in pentaho.
TPCWcustomer	CustType	Records where inconsistent some had Gov ,Comm. Edu so	Standardized the data from Govt to US_Govt State to State_Local_Govt Comm to Commercial Edu to Education using replace in string in pentaho.
TPCWcustomer	Address	The address field was not consistent	Added Comma delimiter in address field to split into addr1 & addr2
TPCWcustomer	Name	Records where inconsistent Some had Inc, Corp*, Cor, Chem	Standardized the name, Inc. to Inc, Co. to Company, Corp to Corporation, Comm to Commercial, Chem to chemicals Firstfed America BanCompany to Firstfed America Ban Company, YuliChem to Yuli Chemicals using replace in string in pentaho.

TPCWcustomer	Address	Some records had St, Ave and Rd,Dr., etc. so standardized them	Standardized the addr1 & addr2, Rd. to Road St. to Street Dr. to Drive Av. to Avenue Ave. to Avenue & Replaced all . to space using replace in string in pentaho.
TPEcustomer	Name	Records had inconsistency Co., Corp, Inc,Vew etc	Standardized the data from Inc. to Inc Co. to Company Corp to Corporation Vec to V.e.v And removed " ." and replaced with space by using replace in string in pentaho.
PECbusinessunit.csv	ABBREV	The data was quoted in doubled quotes And Had null values	Added missing value And replaced double quotes(") with space
TPCWbusiness_unit.csv	ABBREV	Null value	Added missing value
TPCEbusinessunit.csv	ABBREV	Null value And Data inconsistency like Supply, chemical	Added missing value & replace Supply with Supplies & replaced Chemical with Chemicals
PECinvoice	All Fields except orderDate	Remove	Select orderDate by using replace in string in pentaho to separate orderDate from the csv file.
PECInvoice	OrderDate	Standardization	Standardization the date by replacing '-' &'.' by '/'.

PECInvoice	OrderDate	Standardization	Standardizaring the date format to mm/dd/yy
PECInvoice	OrderDate	Extract date, month, year	Extract date, week, quarter month, year from given date using calculator function of pentaho
PECInvoice	OrderDate	Calculating Fiscal day, fiscal week, fiscal month, fiscal year	Writing a javaScript to calculate fiscal day, fiscal month, fiscal year, fiscal week, fiscal quarter
PECinvoice	All Fields except saleDate	Remove	Select SalesDate by using replace in string in pentaho to separate SalesaDate from the csv file.
PECinvoice .csv	saleDate	Standardization	Standardization the date by removing '-', '.' and adding '/'.
PECinvoice .csv	All fields	Standardization	Standardization records which includes some random data.
TPCWinvoice.csv	saleDate	Standardization	Standardization the date by removing '-' , '.', ',' by '/'.
TPCWinvoice.csv	saleDate	Standardization	Standardization the date to mm/dd/yyyy
PECinvoice TPCWinvoice TPCEinvoice	CalendarYear CalendarQuarter CalendarMonth CalendarWeek CalendarDay	Extract Year of date A Quarter of date A Month of date A Week of date A Day of month of date A	Splitting SalesDate and using calculator in pentaho into CalendarYear CalendarQuarter CalendarMonth CalendarWeek CalendarDay

PECinvoice TPCWinvoice TPCEinvoice	CalendarYear CalendarQuarter CalendarMonth CalendarWeek CalendarDay	Calculate fiscal Year, fiscal week, fiscal month, fiscal quarter	Calculating fiscal Year, fiscal week, fiscal month, fiscal quarter from date using Modified JavaScript
TPCWinvoice.csv	Custid	Standardization	Negative values were changed to absolute values.
SupplierTPCE	NAME ADDR1 ADDR2 CITY STATE ZIP SUPPLIER ID	Standardization	Standardized these attributes using select values to supplierName supplierAddress1 supplierAddress2 supplierState supplierCity suppplierZipcode supplier_Id
ProductTPCW	Field_000 Field_001 Field_002 Field_003 Field_004 Field_005 Field_006 Field_007 Field_008 Field_009 Field_0010	Standarization	Extracted all the values and renamed them to Prodid ProdDescription Price1 Price2 UnbitCXost SupplierNBame Addr1 Addr2 City Zip ProdType_Id
ProductTPCW	supplierCityState	Combined	Splitted city and state using split fields twith delimiter as',' supplierCity supplierState

ProductTPCW	ProdId State ProdTypeID City	Enclcosed in quotes	Replacing quotes with space
ProductTPCW	supplierName supplierAddr1 supplierAddr2 supplierState supplierCity suppplierZipcode supplierID	Duplicate Rows	Duplicate rows were removed using Unique rows (Hash Set)
ProductTPCW	supplierAddr1	Extra attribute Attn :	Attn: was removed] before all address 1 in supplier were using replace in string.
ProductPEC	prodid prodDescription productTypeID	Quotes	Quotes where replaced and some values where changed like Equip* to Equipment using Replace in string
TPC_W_Product.csv	Product name	Some records had Equip	Hence replaced with Equipment using replace string (Using Regular expression).
TPCWProduct.csv	Unit cost	Values were null	Calculated unit Cost for PEC data using the formula: Total Manufacturing Cost/ Total Quantity, from the manufacturing_cost.csv and pec.invoice.csv file.

TPC_W_Product_Type.csv	All fields	The data contains double quotes””.	Removed quotes using Select Value in pentaho
TPC_W_Business_Unit.csv	ABBREV	Standardizing	Added missing data and formatted exiting data for uniformity like Supply to supplies.
TPC_E_Product.csv	Description	Some of the description contains Equip.	Hence replaced with Equipment using replace string
TPC_E_Product_Type.csv	Type description	Some of the type description contains Equip.	Hence replaced with Equipment using replace string (Using Regular expression).
TPC_E_Product_Type.csv	BUID	Data column contains double quotes ““	Hence removed using string operations.
TPC_E_Business Unit.csv	ABBREV	Supply and Chemical is present.	So, replaced it with Supplies and Chemicals using replace string
PEC_Product.csv	Missing Source column	Missing source column	Created a new source column to recognize the source of every record from which company it has been collected.
PEC_Product_Type.csv	All fields	The data contains double quotes””.	Hence removed using replace in string.
PEC_Product_Type.csv	Type description	Some of the type description contains Equip.	Hence replaced with Equipment using replace string (Using Regular expression).
PEC_Business_Unit.csv	ABBREV	Some of the miscellaneous data is missing.	Hence added using replace string.
Sales_Fact_Table (PEC_Product)	All Fields	The Values were enclosed in double quotes.	Replaced double quotes with space in the fields using replace in string in pentaho.

Sales_Fact_Table (PEC_Invoice)	Missing Source column	Missing source column	Created a new source column to recognize the source of every record from which company it has been collected.
Sales_Fact_Table (TPC_W_Product)	Field_000 Field_001 Field_002 Field_003 Field_004 Field_005 Field_006 Field_007 Field_008 Field_009		The header row names were missing for every column hence replaced the name with the appropriate headed name and renamed them to Product ID Product Name Price 1 Price 2 Unit Cost ProdTypeID supplierName supplierAddr1 supplierAddr2 supplierCityState supplierZipcode
Sales_Fact_Table (TPC_W_Invoice)	All Fields	The Values were enclosed in double quotes.	Replaced double quotes with space in the fields using replace in string in pentaho.
Sales_Fact_Table (TPC_E_Invoice)	Missing Source column	Missing source column	Created a new source column to recognize the source of every record from which company it has been collected.

2. Data Transformation

DM Table	**Image Creation Process** (attach code)
Product_Dimension	Extracted the Product, Product_type and Business_unit for all the sources and manufacturing cost for PEC Calculated unit cost for PEC data using the formula: Total Manufacturing Cost/ Total Quantity, from the manufacturing_cost.csv and pec.invoice.csv file. Appended the calculated unitCost in PEC where the unitCost value is null Add Source (for TPCE TPCW and PEC) Merged product, product_type and BUID for all sources. Add null Record Added surrogate key and send it to output file. **Product_Dimension.ktr** **Refer to the images below for transformation**
Transaction_Profile_ Dimension	Extracted PEC invoice csv file Added DivisionID (1 for TPCE, 2 for TPCW and 3 for PEC) Merge all division Data As TPCW and TPCE does not have any values, added 2 rows with all attributes having initial value 0. One row for TPCW and another row for TPCE. Add null Record Added surrogate key and sent it to output file. **Transaction_Profile_ Dimension.ktr** **Refer to the images below for transformation**

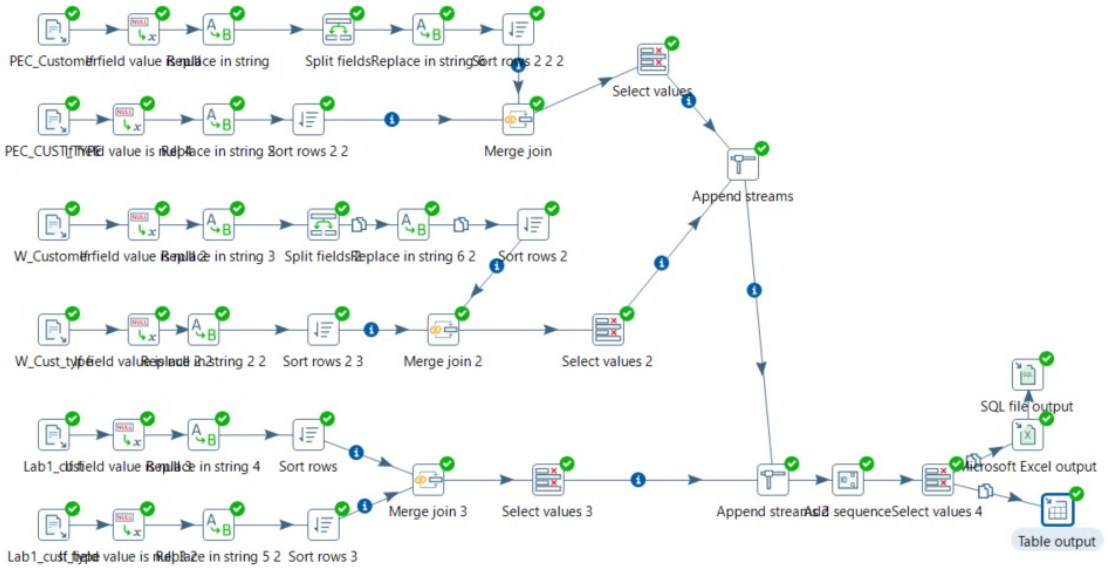
Supplier_Dimension	Extracted Supplier from TPE, Product TPCW and productPEC. Added source to all the files to get track of data from divisions. Standardized the format of data from supplier TPC E. NAME to supplierName ADDR1 to supplierAddr1 ADDR2 to supplierAddr2 CITY to city STATE to state Zip to supplierZip Removed ProductID , ProductName, Price1, Price2, UnitCost,PRODTYPEID Splitted citystate to city and state Updated all mis-spelled address and city and state. Sorted the data from TPC W and TPC E to according to supplierName and merge them. After merging all data removed extra column supplierName_1, supplierAddr1, supplierAddr2, City_1, state_1, supplierZipcode Removed prodid, prodDescription, Price1, price2, UnitCost, productTypeID from productPEC. Removed all null values using filter rows. Added surrogate key using add sequence. Added supplierAddress1, supplierAddress2, city, state and zip using add constants. Appended all the values from TPCE and PEC. Generated supplier name, supplier address, city, state anfd zip and appended those values with previous append. Added surrogate key. Added SCD using add constant. Taking table output using table output. Supplier_Dimension.ktr **Refer to the images below for transformation**

Customer_Dimension	Extract customer and customer_type files for all three divisions PEC, TPCW,TPCE Add ADDR1 & ADDR2 column to TPCW and PEC Map states to its abbreviation in TPCW Replace the abbreviations such as Ave, Rd, Dr in ADDR 2 with complete name in all divisions Add DivisionID (1 for TPCE, 2 for TPCW and 3 for PEC) Merge Customer and Customer_type of each division Merge all the three division tables Add Null value Add surrogate keys and send it to table output file. **Customer_Dimension.ktr** **Refer to the images below for transformation**
Order_Date_Dimension	Extract PECinvoice.csv file Convert orderDate into MM/dd/yyyy format. Removed all other attributes and only kept orderDate Used Unique hashset to take unique records of saleDate Add fiscal year, fiscal month, fiscal week, and fiscal quarter to the orderDate. Add null records Add surrogate keys and send it to table output file. **Order_Date_Dimension.ktr** **Refer to the images below for transformation**
Sales_Date_Dimension	Convert the sale date into MM/dd/yyyy format. Removed all other attributes and only kept orderDate Add fiscal year, fiscal month, fiscal week and fiscal quarter to the salesDate. Merge all the 3 divisions Used Unique hashset to take unique records of saleDate Add null Record Add surrogate keys and send it to table output file. **Sale_Date_Dimension.ktr** **Refer to the images below for transformation**

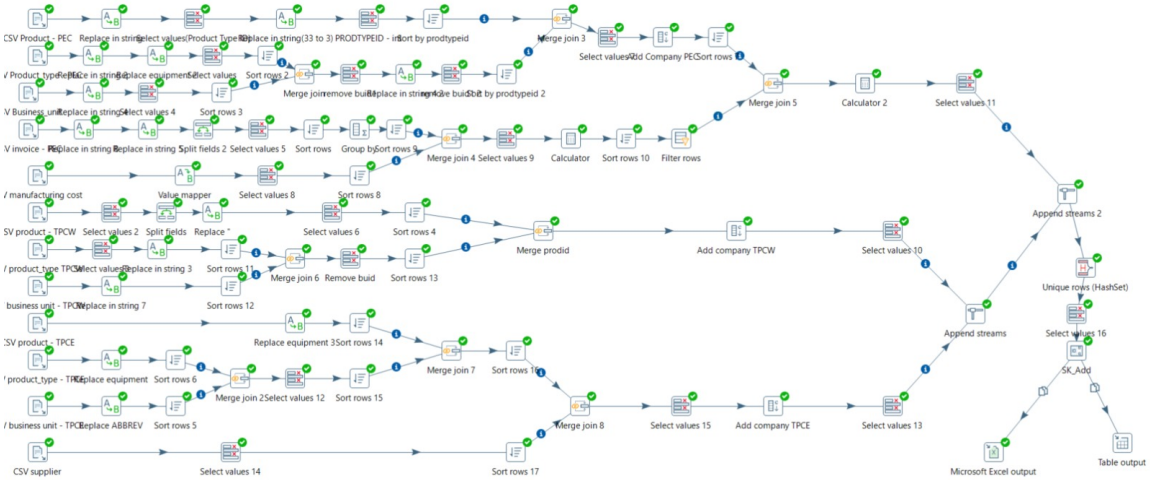
Sales_Fact_Dimension	Extracted the Product, invoice files from the three sources that is PEC, TPC_W and TPC_E Changes the price data type to number. Sorted all the rows with respect to prodID for the sources. Then merged the TPC_W_invoice and TPC_W product on prodID Then merged the TPC_E_invoice and TPC_E product on prodID Then merged the TPC_E_invoice and TPC_E product on prodID Then calculated the number of days to ship by subtracting order date from sales date using the calculator only for PEC data this is because the order date and sales date are present only in this source. Total sales amount was calculated using java script using the formula. If discount = 0 then we multiply price 1 and quantity, if discount is 1 then we multiply price 2 and quantity. Then the invoice was remned to invoiceID. Post this unwanted attribute for this fact table were removed using select value function. Then all the data was appended for the three sources. Finally given to the table output. **Transcation_Profile_Dimension.ktr** **Refer to the images below for transformation**

KTR ScreenShots :

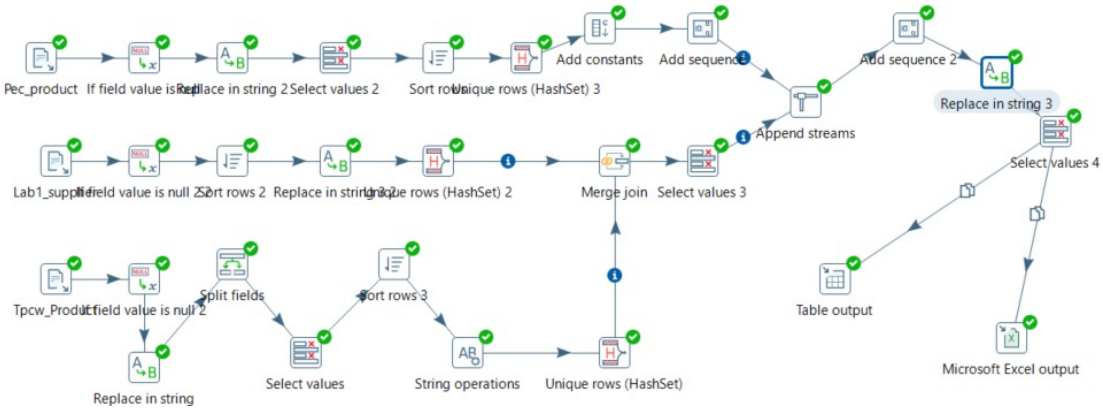
Customer_Dim :



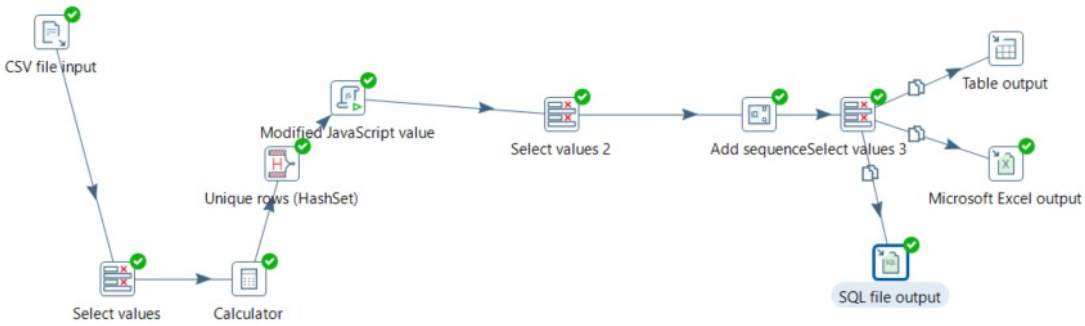
Product_Dim:



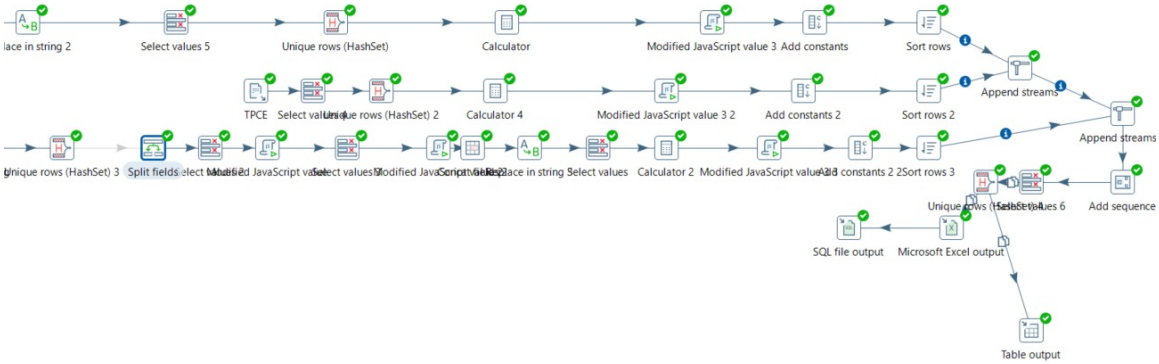
Supplier_Dim :



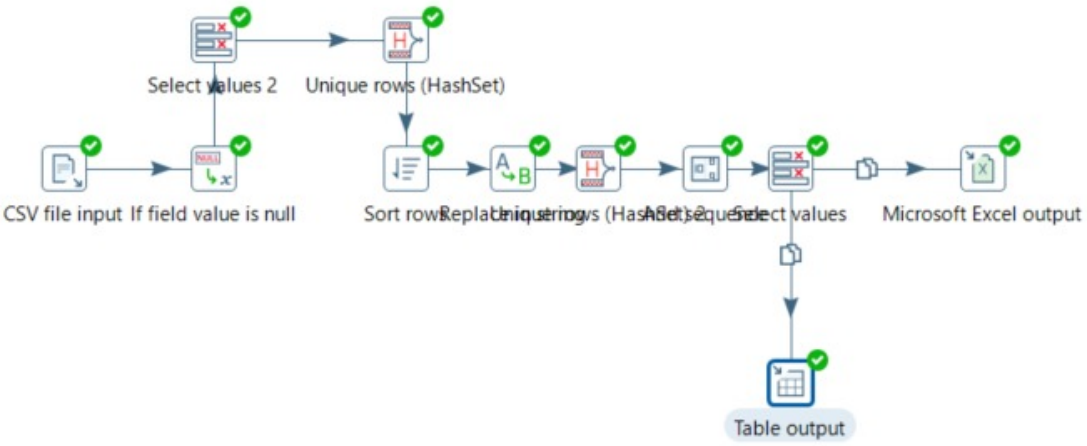
OrderDate_dim:



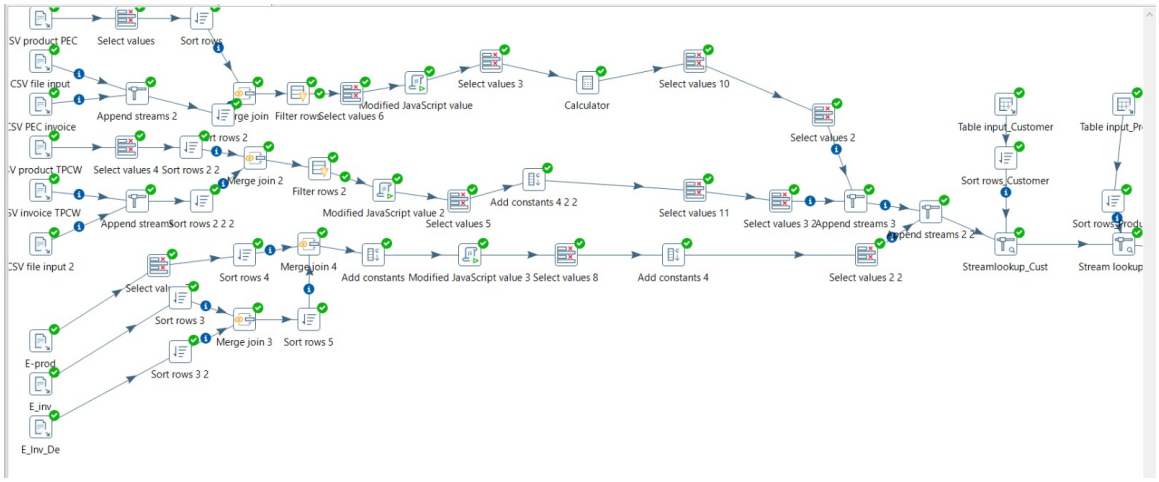
SalesDate_dim:



Transaction_Details_Junk_dim:



sales_fact :



3. Table Population

DM Table	**Table Population Process** (attach code)
Customer dimension	PEC,TPCW,TPCE get Customer & Customer_type Addr1 & Addr2 separated Removes abbreviation Ave., Rd., Dr in Addr Merge Customer & customer_type Sorted the rows with typename Merge all 3 divsions Add surrogate keys & upload data to Table output
Transaction_Junk_Dim	Merge all 3 divsions Add surrogate keys & upload data to Table output Add null Record Added surrogate key and sent it to output file.
Product_Dimension	Extracted the Product, Product_type and Business_unit for all the sources and manufacturing cost for PEC Calculated unit cost using the formula: Total Manufacturing Cost/ Total Quantity, from the manufacturing_cost.csv and pec.invoice.csv file. Add department (for TPCE TPCW and PEC) Add null Record Added surrogate key and send it to output table.
Supplier_Dimension	Extracted data from TPCE,TPCW,PEC Add departments Separated Address to Addr1 & Addr2 Made uniformity in data in State Corrected spells and formats of data Sorted the rows with Name Extracted unique rows only Merged data from TPCE & TPCW Appended values of PEC in the data Select needed data fields and add Surrogate key Uploaded the data to Table output
OrderDate	Excluded the values not needed and only kept order date Separated the date, month, year from given sales dates Selected unique rows Calculate fiscal month, fiscal year, fiscal week, fiscal day Add surrogate key Select needed fields Uploaded data to output table

Sales_date	Imported data from PECInoivce, TPCE & TPCW invoice Select needed field & exclude unwanted fields Select unique rows dates Extract month, day, week, year Calculate fiscal month, fiscal year, fiscal week, fiscal day Add Department Sort rows in according to dates Append all the data and extract only unique dates in unique department Add surrogate key and upload data to table output

1) Sales_Fact

```
1 •  SELECT * FROM sales_order.sales_fact;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Customer_SK	Product_SK	Supplier_SK	SalesDate_SK	OrderDate_SK	TransactionalDetails_SK	invoiceID(DD)	ShippingCost	SalesAmount	ShippingDuration	Discounted
105	1	1	6	7	11	34660	1	119	14	0
119	1	1	10	22	11	10192	1	137	6	0
101	1	1	10	27	32	38637	2	234	27	0
93	1	1	12	11	18	33528	2	101	19	0
103	1	1	22	7	4	36098	14	200	30	0
105	1	1	26	28	21	11317	24	262	24	0
102	1	1	33	46	14	44059	5	228	15	0
109	1	1	35	39	33	11361	30	369	30	0
120	1	1	35	63	14	25041	22	315	5	0
106	1	1	35	45	20	41602	2	155	19	0
83	1	1	35	38	26	49196	9	178	26	0
100	1	1	35	53	19	55317	22	359	13	0
116	1	1	44	43	16	18056	6	199	30	0
109	1	1	44	72	24	31954	26	367	0	0
110	1	1	44	64	17	35448	1	104	9	0
91	1	1	45	55	14	38785	9	227	18	0
87	1	1	45	68	4	55021	13	268	6	0
88	1	1	50	72	10	11290	5	158	6	0
92	1	1	50	75	18	27479	21	264	5	0
119	1	1	50	65	1	49137	3	132	14	0
93	1	1	51	79	14	23360	26	369	2	0
101	1	1	52	76	34	14723	11	127	5	0

2) Junk_Dim

```
1 •  SELECT * FROM sales_order.transaction_details_junk_dim;
```

Result Grid | Filter Rows: | Edit: | Export/Import:

TransactionDetails_SK	MethodOfPayment	MethodOfOrder	MethodOfShipping	Department
1	cash	email	air	PEC
2	cash	internet	air	PEC
3	cash	phone	air	PEC
4	cash	email	n/a	PEC
5	cash	internet	n/a	PEC
6	cash	phone	n/a	PEC
7	cash	email	train	PEC
8	cash	internet	train	PEC
9	cash	phone	train	PEC
10	cash	phone	truck	PEC
11	cash	email	truck	PEC
12	cash	internet	truck	PEC
13	charge	phone	air	PEC
14	charge	email	air	PEC
15	charge	internet	air	PEC
16	charge	email	n/a	PEC
17	charge	internet	n/a	PEC
18	charge	phone	n/a	PEC
19	charge	email	train	PEC
20	charge	internet	train	PEC
21	charge	phone	train	PEC
22	charge	email	truck	PEC
23	charge	internet	truck	PEC

3) Product_Dim

```
1 •  SELECT * FROM sales_order.product_dim;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

SK	Product_ID(NK)	Prod_Descp	ProdTypeID	Buid	Bu_Name	Abbrev	Price1	Price2	Unit_Cost	Type_Description	Department
1		Enumerator Polishing Equipment	3	A	NULL	Equipment	165	143	NULL	Polishing Equipment	PEC
2		Planetesimal Manufacturing Equipment	2	A	NULL	Equipment	347	307	NULL	Manufacturing Equipment	PEC
3		Tailor Jacks	15	D	TPC West	MISC	510	430	400	Jacks	PEC
4		Miniaturizing Manufacturing Equipment	2	A	NULL	Equipment	456	380	NULL	Manufacturing Equipment	PEC
5		Flake Photo Equipment	1	A	NULL	Equipment	232	204	NULL	Photo Equipment	PEC
6		Denigrating Polishing Equipment	3	A	NULL	Equipment	566	454	NULL	Polishing Equipment	PEC
7		Bandage Manufacturing Equipment	2	A	NULL	Equipment	298	267	NULL	Manufacturing Equipment	PEC
8		Chalmers Polishing Equipment	3	A	NULL	Equipment	289	243	NULL	Polishing Equipment	PEC
9		Defeated Tray Supplies	8	B	TPC West	Supplies	560	477	448	Tray Supplies	PEC
10		Chartable Photo Equipment	9	C	NULL	Chemicals	410	347	NULL	Photo Chemicals	PEC
11		Sailors Manufacturing Equipment	2	A	NULL	Equipment	340	290	NULL	Manufacturing Equipment	PEC
12		Chromium Photo Equipment	1	A	NULL	Equipment	566	457	NULL	Photo Equipment	PEC
13		Coward Covers	13	D	TPC East	MISC	430	349	328	Covers	PEC
14		Travelings Photo Chemicals	9	C	TPC West	Chemicals	274	242	225	Photo Chemicals	PEC
15		Syntax Polishing Equipment	3	A	NULL	Equipment	227	191	NULL	Polishing Equipment	PEC
16		Insensitivity Manufacturing Equipment	2	A	NULL	Equipment	201	180	NULL	Manufacturing Equipment	PEC
17		Sortie Covers	13	D	TPC East	MISC	248	213	207	Covers	PEC
18		Allis Polishing Equipment	3	A	NULL	Equipment	202	176	NULL	Polishing Equipment	PEC
19		Habitually Manufacturing Equipment	2	A	NULL	Equipment	509	419	NULL	Manufacturing Equipment	PEC
20		Courthouses Manufacturing Equipment	2	A	NULL	Equipment	286	244	NULL	Manufacturing Equipment	PEC
21		Weeks Polishing Equipment	3	A	NULL	Equipment	199	174	NULL	Polishing Equipment	PEC
22		Extinguisher Manufacturing Equipment	2	A	NULL	Equipment	120	98	NULL	Manufacturing Equipment	PEC

4) SalesDate_Dim

```
1 •  SELECT * FROM sales_order.salesdate_dim;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

SalesDate_SK	Day	Week	Month	Quarter	Year	FiscalDay	FiscalWeek	FiscalMonth	FiscalQuarter	FiscalYear	Department	Date
1	1	14	4	2	2009	1	26	1	1	2010	PEC	2009-04-01
2	15	29	7	3	2005	15	41	4	2	2006	PEC	2005-07-15
3	1	1	1	1	2005	1	41	10	4	2005	PEC	2005-01-01
4	2	2	1	1	2005	2	42	10	4	2005	PEC	2005-01-02
5	3	2	1	1	2005	3	42	10	4	2005	PEC	2005-01-03
6	4	2	1	1	2005	4	42	10	4	2005	PEC	2005-01-04
7	5	2	1	1	2005	5	42	10	4	2005	PEC	2005-01-05
8	6	2	1	1	2005	6	42	10	4	2005	PEC	2005-01-06
9	7	2	1	1	2005	7	42	10	4	2005	PEC	2005-01-07
10	8	2	1	1	2005	8	42	10	4	2005	PEC	2005-01-08
11	9	3	1	1	2005	9	43	10	4	2005	PEC	2005-01-09
12	10	3	1	1	2005	10	43	10	4	2005	PEC	2005-01-10
13	11	3	1	1	2005	11	43	10	4	2005	PEC	2005-01-11
14	12	3	1	1	2005	12	43	10	4	2005	PEC	2005-01-12
15	13	3	1	1	2005	13	43	10	4	2005	PEC	2005-01-13
16	14	3	1	1	2005	14	43	10	4	2005	PEC	2005-01-14
17	15	3	1	1	2005	15	43	10	4	2005	PEC	2005-01-15
18	16	4	1	1	2005	16	44	10	4	2005	PEC	2005-01-16
19	17	4	1	1	2005	17	44	10	4	2005	PEC	2005-01-17
20	18	4	1	1	2005	18	44	10	4	2005	PEC	2005-01-18
21	19	4	1	1	2005	19	44	10	4	2005	PEC	2005-01-19
22	20	4	1	1	2005	20	44	10	4	2005	PEC	2005-01-20
23	21	4	1	1	2005	21	44	10	4	2005	PEC	2005-01-21

5) Customer_Dim

```
1 •  SELECT * FROM sales_order.customer_dim;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Customer_SK	Cust_id(NK)	CustType	CustTypeID	Cust_Name	Cust_Address1	Cust_Address2	Cust_City	Cust_State	Cust_Zip	Depar
1	3	COMMERCIAL	C	Baxter May	NULL	796-2366 Vel Avenue	Meriden	FL	58328	PEC
2	6	COMMERCIAL	C	Raphael Allison	NULL	734-2598 Sed Avenue	La Crosse	HI	58702	PEC
3	9	COMMERCIAL	C	The Product Company (East)	NULL	20021 Barnes Road	Stratford	NY	13470	PEC
4	11	COMMERCIAL	C	Austin Ferrell	NULL	8666 Justo Road	Pullman	PA	55979	PEC
5	14	COMMERCIAL	C	Serrano	NULL	7212 Eleifend Street	Princeton	KY	66785	PEC
6	15	COMMERCIAL	C	Colton Maldonado	NULL	1869 Leo Road	Cedar Rapids	SD	97640	PEC
7	16	COMMERCIAL	C	Googol	NULL	5453 Dictum Avenue	Jenks	LA	96393	PEC
8	20	COMMERCIAL	C	Pewter Gym	NULL	9028 Non Street	Sandy	FL	87202	PEC
9	27	COMMERCIAL	C	Clare Baird	6152 Auctor	Road	Wilson	MS	29338	PEC
10	29	COMMERCIAL	C	Blevins	NULL	2571 Donec Avenue	Winooski	SD	65512	PEC
11	30	COMMERCIAL	C	Knox Reid	500 Eu	Avenue	Pass Christian	WY	35949	PEC
12	32	COMMERCIAL	C	Ferengi Treasures	NULL	716-6837 Vulputate ...	Paramount	NY	35644	PEC
13	33	COMMERCIAL	C	Tallulah	NULL	3376 Aliquet Street	Batavia	NM	50742	PEC
14	34	COMMERCIAL	C	The Product Company (West)	NULL	21 E Bullard Avenue	Fresno	CA	93710	PEC
15	1	EDUCATION	E	Haynes	2920 Auctor	Road	Moraga	AZ	70314	PEC
16	7	EDUCATION	E	Martin Donaldson	NULL	2704 At Road	Gardner	GA	31421	PEC
17	18	EDUCATION	E	Melvin House	NULL	8371 Nullam Road	Isle of Palms	ME	89644	PEC
18	19	EDUCATION	E	Emerson Electric Company	NULL	1792 Squash Drive	South Texar...	TX	75501	PEC
19	21	EDUCATION	E	Hop Adams	NULL	704 Nisl Road	Sun Valley	NH	84458	PEC
20	22	EDUCATION	E	Xavier Harmon	NULL	2285 Ante Street	Farmington	WV	16456	PEC
21	28	EDUCATION	E	Schultz Learning	NULL	8542 Quis Street	Spartanburg	ND	73217	PEC
22	31	EDUCATION	E	Starfleet Academy	NULL	3061 Donec Avenue	Canandaigua	UT	75963	PEC

Result Grid | Form Editor | Field Types | Query Stats | Execution Plan

5) Supplier_Dim

```
1 •  SELECT * FROM sales_order.supplier_dim;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Supplier_SK	Supplier_id(NK)	SupplierName	SupplierAddress1	SupplierAddress2	SupplierCity	SuplierState	SupplierZipcode
1	11	NULL	NULL	NULL	NULL	NULL	NULL
2	12	TPC East	NULL	NULL	NULL	NULL	NULL
3	13	TPC West	NULL	NULL	NULL	NULL	NULL
4	7	Afg Industries Inc.	Dorothy Polk	5377 Cookbook Circle	Vero Beach	FL	32967
5	8	American General Ventures Inc.	Kimberly Yarborough	1616 Goggles Drive	Sterling	KS	67579
6	3	Black Hills Corporation	Edena Gilbert	1618 Greenland Street	Tallahassee	FL	32304
7	5	Dollar General Corporation	Antoinette Hickey	3988 Chanced Avenue	S Middleton	PA	17007
8	9	Fedders Corporation	Virgil Hubert	7844 Carelessly Drive	Overlook	WA	98366
9	6	First Bancshares Inc. Mo	Jensine Lease	11885 Aural Turnpike	Louisville	KY	40253
10	4	Immunex Corporation	Elbert Glad	5841 Pervades Drive	Fairfax	VA	22035
11	10	Saurer Gruppe Holding Ag	Pearce Valdez	3727 Elms Drive	South Sain...	MN	55077
12	2	SInc.lair Broadcast Group Inc.	Carla Shelby	6037 Ecuadorian Street	Clinton	MI	49236
13	1	Specialty Teleconstructors Inc.	Zylina Cox	2161 Chalmers Street	Salinas	CA	93905
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

VII. End User Applications

1. Queries

User Question/Reporting Need
Displaying top sales by customers:
SQL Code
use sales_order; SELECT customer_dim.Cust_Name AS 'Customer_Name', SUM(sales_fact.SalesAmount) AS 'Sales' FROM sales_fact JOIN customer_dim ON customer_dim.Customer_SK = sales_fact.Customer_SK GROUP BY customer_dim.Cust_Name ORDER BY SUM(sales_fact.SalesAmount) DESC; 1 • use sales_order; 2 • SELECT customer_dim.Cust_Name AS 'Customer_Name', SUM(sales_fact.SalesAmount) AS 'Sales' FROM sales_fact 3 JOIN customer_dim ON customer_dim.Customer_SK = sales_fact.Customer_SK 4 GROUP BY customer_dim.Cust_Name 5 ORDER BY SUM(sales_fact.SalesAmount) DESC; Result Grid | Filter Rows: | Export: | Wrap Cell Content:

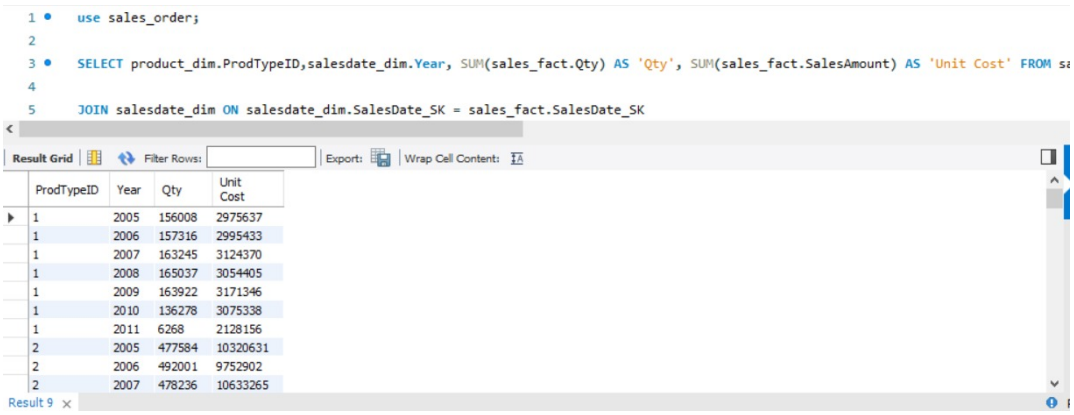
Customer_Name	Sales
Lnr Property Corporation	6652533
Gte Corporation	6528625
Helpmate Robotics Inc.	6457578
Greenman Technologies Inc.	6345620
Publishing Company Of North America Inc.	6325020
Ligand Pharmaceuticals Incorporated	6307797
Sterling Financial Corporation Washington	6299070
Usx-steel Group Inc.	6271407
Railamerica Inc.	6209266
Scp Pool Corporation	6201163
United Healthcare Corporation	6144456

Supporting Index(es)
Customer and sales (fact table)

User Question/Reporting Need
Per year sales of a company by product type
SQL Code

use sales_order;

SELECT product_dim.ProdTypeID,salesdate_dim.Year, SUM(sales_fact.Qty) AS 'Qty', SUM(sales_fact.SalesAmount) AS 'Unit Cost' FROM sales_fact
JOIN salesdate_dim ON salesdate_dim.SalesDate_SK = sales_fact.SalesDate_SK
JOIN product_dim ON product_dim.Product_SK = sales_fact.Product_SK
GROUP BY product_dim.ProdTypeID,salesdate_dim.Year
ORDER BY product_dim.ProdTypeID;



Supporting Index(es)

Product, SalesDate

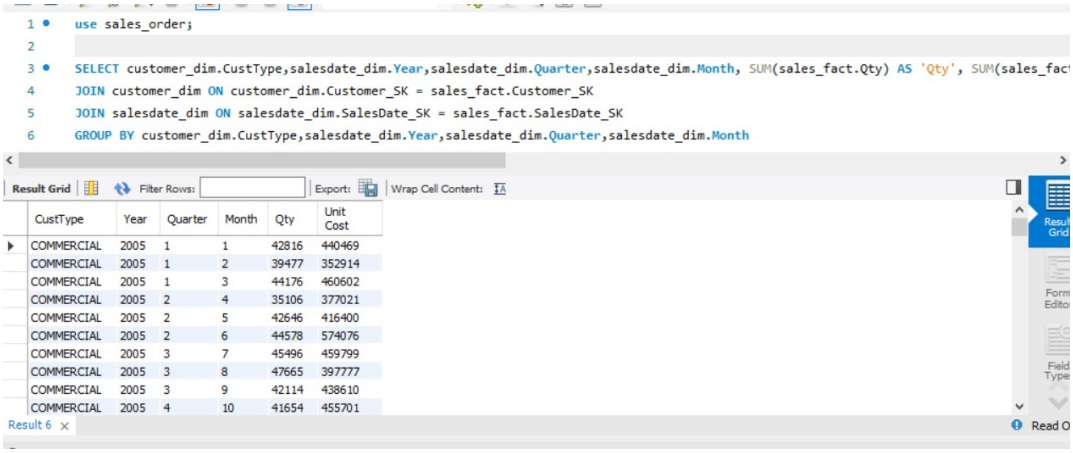
User Question/Reporting Need

Query showing the sales and and quantity of each customer or customer type on an annual, quarterly and monthly basis

SQL Code

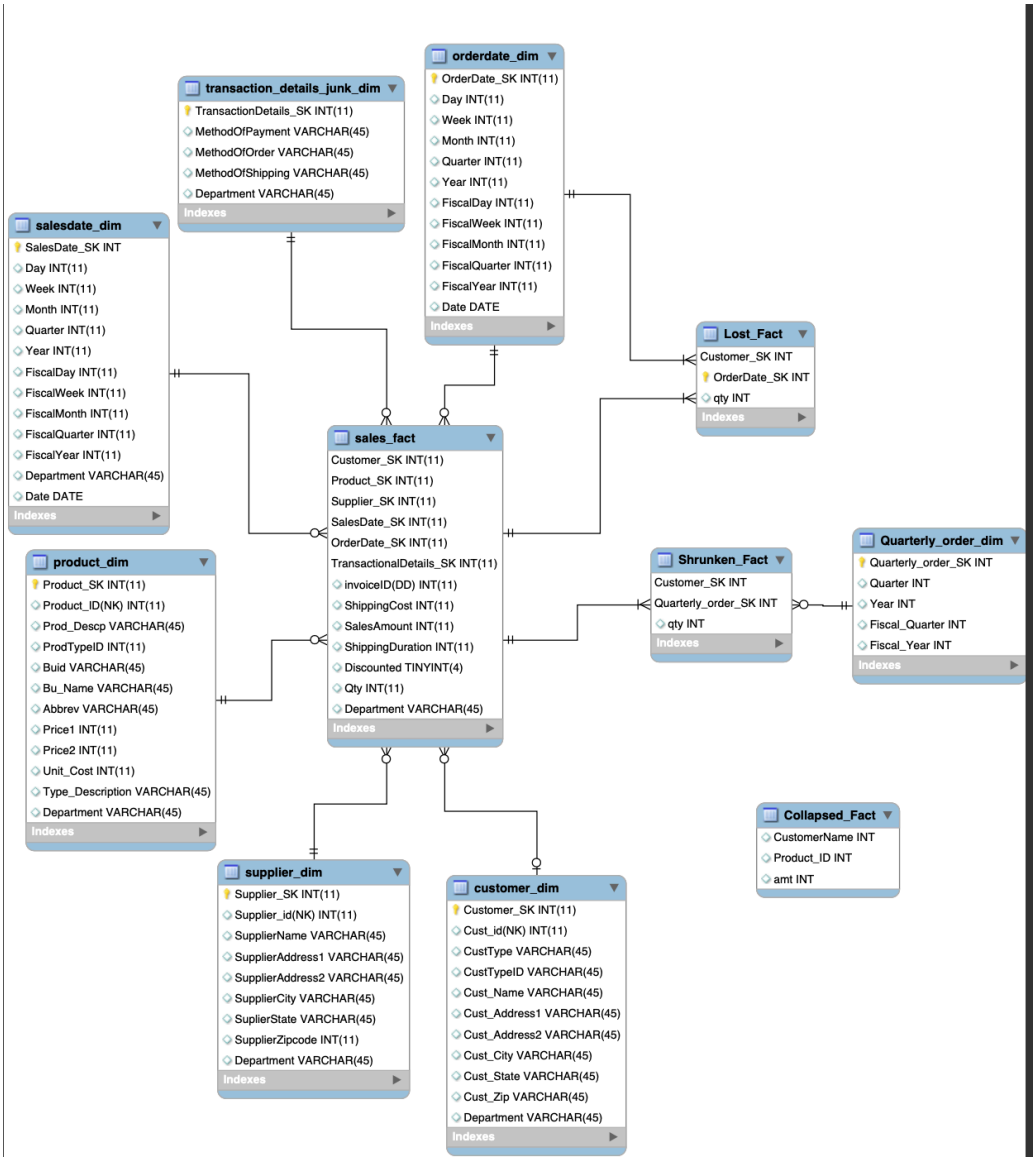
```
use sales_order;

SELECT
customer_dim.CustType,salesdate_dim.Year,salesdate_dim.Quarter,salesdate_dim.
Month, SUM(sales_fact.Qty) AS 'Qty', SUM(sales_fact.SalesAmount) AS 'Unit
Cost' FROM sales_fact
JOIN customer_dim ON customer_dim.Customer_SK = sales_fact.Customer_SK
JOIN salesdate_dim ON salesdate_dim.SalesDate_SK = sales_fact.SalesDate_SK
GROUP BY
customer_dim.CustType,salesdate_dim.Year,salesdate_dim.Quarter,salesdate_dim.
Month
ORDER BY customer_dim.CustType;
```



Supporting Index(es)

3. Aggregated Mata Marts



VIII. Handling Slowly Changing Dimensions (SCD)

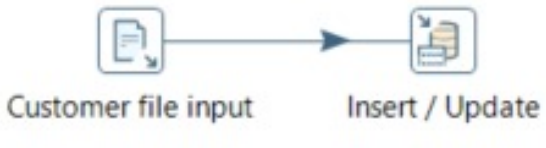
- **What type of slowly changing dimension would you implement? Why? (There should be at least two different SCD types other than Type 0)**

- We are implementing SCD Type 1 because it's very useful when we need to overwrite the data. This is implemented when we don't have to store the historical data. We are also implementing SCD Type 2 which is the most used SCD type. We have chosen these two types so as to show the difference of storing historical data. SCD Type 2 stores the "Effective from" and "Effective to" dates so as to use the recent updated data. It stores the historical data (version control) with the help of a field "Version". This also keeps track of how many times the data is updated.

- **Create a set of sample source data (minimum 25 records) and make a copy of dimension tables to demonstrate your choices of slowly changing dimension types. Don't apply this to your original data mart. Use Pentaho Kettle steps or any other tool that you have used in the project. Describe how the slowly changing dimension types were implemented and include the code. The Pentaho transformation or any other tools should be generic so that you can apply to any other source datasets.**

- Set of sample source data is attached in the zip file under SCD folder.

- **SCD Type 1 :**



This is implemented using the Insert/Update in Pentaho. This will lookup for the comparison attribute and update in the required field. Here, in this case we have changed the names of Cust_IDs : 27, 30, 1, 26, 35.

In the SCD 1 it doesn't stores the historical data and just stores the recent data.

Output is shown below :

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

Customer_SK	Cust_id(NK)	CustType	CustTypeID	Cust_Name	Cust_Address1	Cust_Address2	Cust_City	Cust_State	Cust_Zip	Department
9	27	COMMERCIAL	C	Samruddhi Deshpande	6152 Auctor	Road	Wilson	MS	29338	PEC
11	30	COMMERCIAL	C	Vikram Parmar	500 Eu	Avenue	Pass Christian	WY	35949	PEC
15	1	EDUCATION	E	Sanskruti Deshpande	2920 Auctor	Road	Moraga	AZ	70314	PEC
26	4	STATE_LOCAL_GOVT	S	Maya Brewer	PO Box 825	6836 At	Union City	ND	82051	PEC
47	15	COMMERCIAL	C	Merritt Long	Suite 30	3072 Cursus Avenue	West Haven	WY	83709	TPCW
69	23	STATE_LOCAL_GOVT	S	Chemfix Technologies Inc.	Suite 50	11533 Wonderingly Drive	Pasadena Hills	MI	63121	TPCW
71	26	STATE_LOCAL_GOVT	S	Ninad Ingale	3142 Congue	Street	Chicago	IL	76404	TPCW
72	30	STATE_LOCAL_GOVT	S	Setron	Dept #80	3085 Breakthrough Drive	Hart	MI	64865	TPCW
74	40	STATE_LOCAL_GOVT	S	Gte Corporation	Dept #555	7196 Heaver Lane	Winter Garden	FL	34777	TPCW
80	35	US_GOVT	F	Nihar Ingale	6979 Quis	Street	Nome	DE	58359	TPCW
81	1	COMMERCIAL	C	Room Plus Inc.	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	TPCE
84	16	COMMERCIAL	C	Meridian Resources Corp...	Dept #804	1977 Symbiotic Turnpike	Fall River	MA	2724	TPCE
85	33	COMMERCIAL	C	Polymer Group Inc.	Suite 204	11368 Harder Drive	Maidstone	VT	5905	TPCE
86	39	COMMERCIAL	C	Ajay Deshpande	Dept #5	10947 Eagle Street	Lisbon Falls	ME	4252	TPCE
87	40	COMMERCIAL	C	Processing Equipment Co...	Dept #2	29843 Klingon Road	Naperville	IL	60563	TPCE
90	5	EDUCATION	E	Sterling Financial Corpora...	Dept #505	7258 Meriting Avenue	Portland	OR	97201	TPCE
91	7	EDUCATION	E	Fedders Corporation	Dept #520	3311 Blatantly Circle	Santa Fe Spri...	CA	90670	TPCE
92	8	EDUCATION	E	United Healthcare Corpor...	Dept #855	675 Hordes Turnpike	Malaga	NM	88263	TPCE
94	26	EDUCATION	E	Firstfed America Bancorp ...	Dept #851	5656 Settler Street	Ward Prairie	TX	75840	TPCE
99	9	US_GOVT	F	Greate Bay Casino Corpo...	Dept #242	4255 Depart Lane	Randolph Air ...	TX	78148	TPCE
101	12	US_GOVT	F	Cohesant Technologies Inc.	Dept #690	3334 Eggshell Circle	Oakley	SC	29461	TPCE
103	18	US_GOVT	F	Scp Pool Corporation	Suite 251	755 Leek Street	Enid	OK	73706	TPCE

istomer_dim 1

- **SCD Type 2 :**

SCD Type 2 is implemented using the Dimension lookup and update in Pentaho. This will lookup for the comparison attribute and update in the required field. Here, in this case we have changed the “names” of Cust_IDs : 35, 39, 36

In the SCD 2 it stores the version of the data and keeps the track of historical data.

Output is shown below :

Customer_SK	Cust_id(NK)	CustType	CustTypeID	Cust_Name	Cust_Address1	Cust_Address2	Cust_City	Cust_State	Cust_Zip	Department	version	EffectiveDate	ExpireDate
315	30	STATE_LOCAL_GOVT	S	Setron	Dept #80	3085 Breakthrough Drive	Hart	MI	64865	TPCW	3	2021-05-06 14:52:55	2199-12-31 23:59:59
316	31	STATE_LOCAL_GOVT	S	Setron	Dept #80	3085 Breakthrough Drive	Hart	MO	64865	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
317	7	EDUCATION	E	Fedders Corporation	Dept #520	3311 Blatantly Circle	Santa Fe Spri...	CA	90670	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
318	12	US_GOVT	F	Cohesant Technologies Inc.	Dept #690	3334 Eggshell Circle	Oakley	SC	29461	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
319	9	US_GOVT	F	Greate Bay Casino Corporation...	Dept #242	4255 Depart Lane	Randolph Air ...	TX	78148	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
320	20	US_GOVT	F	Labor Ready Inc.	Dept #759	4618 Days Drive	Waterville	WA	98858	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
321	35	US_GOVT	F	Integra -- A Hotel & Restauran...	Dept #822	5222 Blazed Street	Parsons	TN	38363	TPCE	2	2021-05-06 14:52:55	2021-05-06 14:53:22
322	26	EDUCATION	E	Firstfed America Bancorp Inc.	Dept #851	5656 Settler Street	Ward Prairie	TX	75840	TPCE	2	2021-05-06 14:52:55	2199-12-31 23:59:59
323	1	COMMERCIAL	C	Room Plus Inc.	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	TPCE	2	2021-05-06 14:52:55	2199-12-31 23:59:59
324	8	EDUCATION	E	United Healthcare Corporation	Dept #855	675 Hordes Turnpike	Malaga	NM	88263	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
325	4	STATE_LOCAL_GOVT	S	Maya Brewer	PO Box 825	6836 At	Union City	ND	82051	PEC	1	1900-01-01 00:00:00	2199-12-31 23:59:59
326	40	STATE_LOCAL_GOVT	S	Gte Corporation	Dept #555	7196 Heaver Lane	Winter Garden	FL	34777	TPCW	2	2021-05-06 14:52:55	2199-12-31 23:59:59
327	28	STATE_LOCAL_GOVT	S	Gte Corporation	Dept #532	7196 Heaver Lane	Winter Garden	FL	34777	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
328	5	EDUCATION	E	Sterling Financial Corporation ...	Dept #505	7258 Meriting Avenue	Portland	OR	97201	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
329	18	US_GOVT	F	Scp Pool Corporation	Suite 251	755 Leek Street	Enid	OK	73706	TPCE	1	1900-01-01 00:00:00	2199-12-31 23:59:59
330	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1	NULL	NULL
331	35	US_GOVT	F	SCD change 1	6979 Quis	Street	Nome	DE	58359	TPCW	3	2021-05-06 14:53:22	2199-12-31 23:59:59
332	39	COMMERCIAL	C	SCD change 2	Dept #5	10947 Eagle Street	Lisbon Falls	ME	4252	TPCE	2	2021-05-06 14:53:22	2199-12-31 23:59:59
333	36	US_GOVT	F	SCD change 3	Dept #633	11119 Periphery Turnpike	Simpson	LA	71474	TPCE	2	2021-05-06 14:53:22	2199-12-31 23:59:59
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

IX. Many-to-Many (N-M) Relationship Implementation Option

Many to many relationships indicate connections between elements where a parent record in one element contains a few records in another element and a vice versa. We use snowflake model for this implementation

Strategy 1

In this strategy, we make a different substance and add property IDs to the fact table. In our model, for estimating the work of the supplier, we need to associate it with the shipping_company dimension. Along these lines, by utilizing Method 1, we can make a different relationship entity which is like the acquainted entity in RDBMS, by having association among supplier and shipping_company measurements with the assistance of surrogate keys.

Strategy 2

In this strategy, many to many relationship is changed over into a compound property relationship. In this, we need to regard one property as child of the other and have a compound key for the lower-level trait and furthermore need to add both characteristic IDs. Thus, in our model, by utilizing the Method 2, we can have the supplier table as a join fact table which would allude to the shipping dimension. Subsequently the supplierID would be a quality of the supplier_dimension which looks into the supplier data from the supplier table.

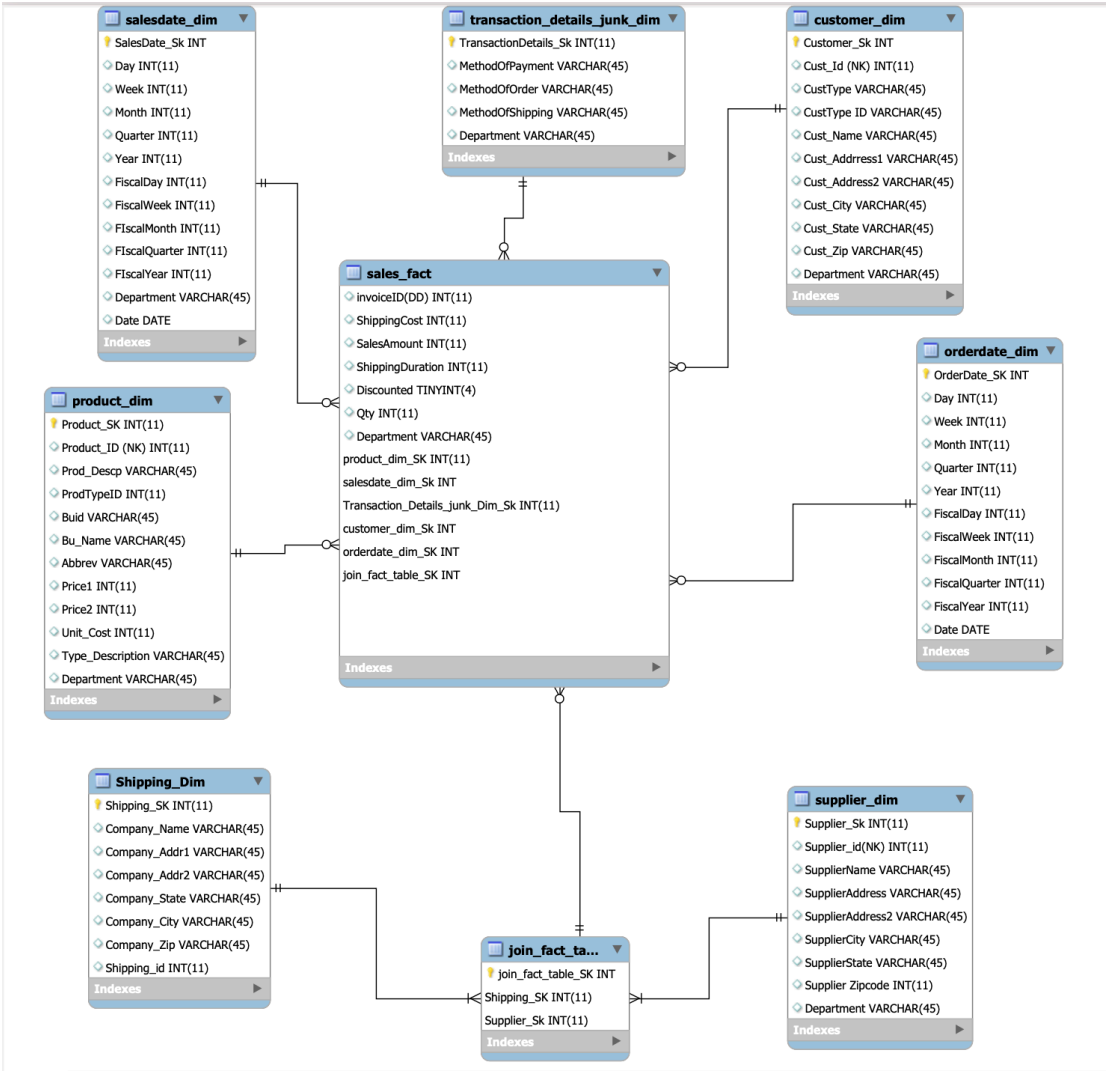
Strategy 3

This strategy, it's a flexible arrangement and have following qualities:

It further improves on the compound trait relationship which is in technique 2 into the simple attribute relationship. Instead of all the columns in the fact table, it only needs one attribute column. Another property is made which is a connection of surrogate keys between its parent attributes, which is lower in level than both of them. Thus, instead of incorporating supplier_sk and Shipping_company_sk we add this SKU attribute.

Many-to-Many (N-M) Relationship Implementation Option used :

In our project, we have proposed join_fact_table where we have made many to many connections among supplier and shipping dimension using a fact table which we have named as join_fact_table. In this strategy, the client adds another segment in the dimension table for every conceivable values that are available.

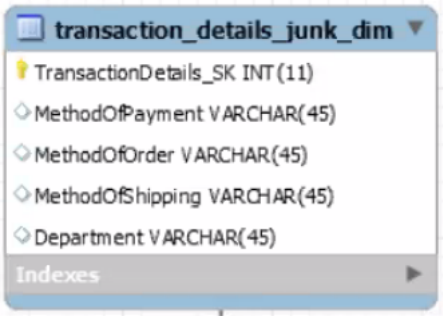

salesdate_dim
SalesDate_Sk INT
Day INT(11)
Week INT(11)
Month INT(11)
Quarter INT(11)
Year INT(11)
FiscalDay INT(11)
FiscalWeek INT(11)
FIscalMonth INT(11)
FIscalQuarter INT(11)
FIscalYear INT(11)
Department VARCHAR(45)
Date DATE
Indexes
transaction_details_junk_dim
TransactionDetails_Sk INT(11)
MethodOfPayment VARCHAR(45)
MethodOfOrder VARCHAR(45)
MethodOfShipping VARCHAR(45)
Department VARCHAR(45)
Indexes
customer_dim
Customer_Sk INT
Cust_Id (NK) INT(11)
CustType VARCHAR(45)
CustType ID VARCHAR(45)
Cust_Name VARCHAR(45)
Cust_Addrress1 VARCHAR(45)
Cust_Address2 VARCHAR(45)
Cust_City VARCHAR(45)
Cust_State VARCHAR(45)
Cust_Zip VARCHAR(45)
Department VARCHAR(45)
Indexes
sales_fact
invoiceID(DD) INT(11)
ShippingCost INT(11)
SalesAmount INT(11)
ShippingDuration INT(11)
Discounted TINYINT(4)
Qty INT(11)
Department VARCHAR(45)
product_dim_SK INT(11)
salesdate_dim_Sk INT
Transaction_Details_junk_Dim_Sk INT(11)
customer_dim_Sk INT
orderdate_dim_SK INT
join_fact_table_SK INT
Indexes
orderdate_dim
OrderDate_SK INT
Day INT(11)
Week INT(11)
Month INT(11)
Quarter INT(11)
Year INT(11)
FiscalDay INT(11)
FiscalWeek INT(11)
FiscalMonth INT(11)
FiscalQuarter INT(11)
FiscalYear INT(11)
Date DATE
Indexes
product_dim
Product_SK INT(11)
Product_ID (NK) INT(11)
Prod_Descp VARCHAR(45)
ProdTypeID INT(11)
Buid VARCHAR(45)
Bu_Name VARCHAR(45)
Abbrev VARCHAR(45)
Price1 INT(11)
Price2 INT(11)
Unit_Cost INT(11)
Type_Description VARCHAR(45)
Department VARCHAR(45)
Indexes
Shipping_Dim
Shipping_SK INT(11)
Company_Name VARCHAR(45)
Company_Addr1 VARCHAR(45)
Company_Addr2 VARCHAR(45)
Company_State VARCHAR(45)
Company_City VARCHAR(45)
Company_Zip VARCHAR(45)
Shipping_id INT(11)
Indexes
supplier_dim
Supplier_Sk INT(11)
Supplier_id(NK) INT(11)
SupplierName VARCHAR(45)
SupplierAddress VARCHAR(45)
SupplierAddress2 VARCHAR(45)
SupplierCity VARCHAR(45)
SupplierState VARCHAR(45)
Supplier Zipcode INT(11)
Department VARCHAR(45)
Indexes
join_fact_ta...
join_fact_table_SK INT
Shipping_SK INT(11)
Supplier_Sk INT(11)
Indexes

X. Appendix (Fix Lab #3 Problems)

Deliverables :
Missing Dump File, so cannot :
There was a mysqldump.exe version mismatch error in the sql workbench, due to which the dump files were not getting generated. The dump files generated were blank (0KB). We have resolved that issue and extracted the new dump files.

Part #1 :
Missing Optional Cardinality for the many side :
The Relation to the Dimension table from fact table is many to many and has an optional cardinality, we have missed the optional cardinality in our dimension table.

Name of Junk Dimension :
Name of the Junk Dimension has been changed to Transaction_Details_Junk_Dim.



Part #2 :
PECcustomer.csv:
Values not well separated :

#	custID	name	Addr1	Addr2	city	state	zip	custtype	Department
1..	33	Tallulah		3376 Aliquet Street	Batavia	NM	50742	COMMERCIAL	PEC
1..	34	The Product Company (West)		21 E Bullard Avenue	Fresno	CA	93710	COMMERCIAL	PEC
1..	1	Haynes	2920 Auctor	Road	Moraga	AZ	70314	EDUCATION	PEC
1..	7	Martin Donaldson		2704 At Road	Gardner	GA	31421	EDUCATION	PEC
1..	18	Melvin House		8371 Nullam Road	Isle of Palms	ME	89644	EDUCATION	PEC
1..	19	Emerson Electric Company		1792 Squash Drive	South Texarkana	TX	75501	EDUCATION	PEC
1..	21	Hop Adams		704 Nisl Road	Sun Valley	NH	84458	EDUCATION	PEC
2..	22	Xavier Harmon		2285 Ante Street	Farmington	WV	16456	EDUCATION	PEC
2..	28	Schultz Learning		8542 Quis Street	Spartanburg	ND	73217	EDUCATION	PEC
2..	31	Starfleet Academy		3061 Donec Avenue	Canandaigua	UT	75963	EDUCATION	PEC
2..	38	Firstfed America Bancorp Inc.		5656 Settler Street	Ward Prairie	TX	75840	EDUCATION	PEC
2..	39	Ronan French		8773 Feugiat Road	Gary	WA	17387	EDUCATION	PEC
2..	2	Rudyard-Knapp		9261 Nulla Street	Sault Ste. Marie	LA	94298	STATE_LOCAL_GOVT	PEC
2..	4	Maya Brewer	PO Box 825	6836 At	Union City	ND	82051	STATE_LOCAL_GOVT	PEC
2..	5	Mullins Inc..		4009 Ornare Road	Birmingham	NE	59509	STATE_LOCAL_GOVT	PEC
2..	10	Hammett Farley		1668 In Street	Yonkers	AK	18651	STATE_LOCAL_GOVT	PEC
2..	12	Cain		1116 Augue Street	Warwick	DC	38450	STATE_LOCAL_GOVT	PEC
3..	25	Atkins		4607 Risus Road	Waukegan	NC	43147	STATE_LOCAL_GOVT	PEC
3..	37	Beverly Equipment		7581 Quisque Street	Bloomington	HI	88440	STATE_LOCAL_GOVT	PEC
3..	40	Zena Machines		1277 HenDriveerit Avenue	San Juan	WA	07066	STATE_LOCAL_GOVT	PEC
3..	8	Gemma Castro		9101 Duis Street	Passaic	VT	40926	US_GOVT	PEC

Address values has been split in Address 1 and Address 2

Part #3 :
Customer dimension
Addresses not well separated :

#	custID	name	Addr1	Addr2	city	state	zip	custtype	Department
1..	33	Tallulah		3376 Aliquet Street	Batavia	NM	50742	COMMERCIAL	PEC
1..	34	The Product Company (West)		21 E Bullard Avenue	Fresno	CA	93710	COMMERCIAL	PEC
1..	1	Haynes	2920 Auctor	Road	Moraga	AZ	70314	EDUCATION	PEC
1..	7	Martin Donaldson		2704 At Road	Gardner	GA	31421	EDUCATION	PEC
1..	18	Melvin House		8371 Nullam Road	Isle of Palms	ME	89644	EDUCATION	PEC
1..	19	Emerson Electric Company		1792 Squash Drive	South Texarkana	TX	75501	EDUCATION	PEC
1..	21	Hop Adams		704 Nisl Road	Sun Valley	NH	84458	EDUCATION	PEC
2..	22	Xavier Harmon		2285 Ante Street	Farmington	WV	16456	EDUCATION	PEC
2..	28	Schultz Learning		8542 Quis Street	Spartanburg	ND	73217	EDUCATION	PEC
2..	31	Starfleet Academy		3061 Donec Avenue	Canandaigua	UT	75963	EDUCATION	PEC
2..	38	Firstfed America Bancorp Inc.		5656 Settler Street	Ward Prairie	TX	75840	EDUCATION	PEC
2..	39	Ronan French		8773 Feugiat Road	Gary	WA	17387	EDUCATION	PEC
2..	2	Rudyard-Knapp		9261 Nulla Street	Sault Ste. Marie	LA	94298	STATE_LOCAL_GOVT	PEC
2..	4	Maya Brewer	PO Box 825	6836 At	Union City	ND	82051	STATE_LOCAL_GOVT	PEC
2..	5	Mullins Inc..		4009 Ornare Road	Birmingham	NE	59509	STATE_LOCAL_GOVT	PEC
2..	10	Hammett Farley		1668 In Street	Yonkers	AK	18651	STATE_LOCAL_GOVT	PEC
2..	12	Cain		1116 Augue Street	Warwick	DC	38450	STATE_LOCAL_GOVT	PEC
3..	25	Atkins		4607 Risus Road	Waukegan	NC	43147	STATE_LOCAL_GOVT	PEC
3..	37	Beverly Equipment		7581 Quisque Street	Bloomington	HI	88440	STATE_LOCAL_GOVT	PEC
3..	40	Zena Machines		1277 HenDriveerit Avenue	San Juan	WA	07066	STATE_LOCAL_GOVT	PEC
3..	8	Gemma Castro		9101 Duis Street	Passaic	VT	40926	US_GOVT	PEC

Address values has been split in Address 1 and Address 2

Product dimension:
First data point has prodID missing :
We have checked if the prodID is null we are dropping the row. Hence the issue is solved.

Not found surrogate key :
Surrogate Key has been added to Product Dimension.

1	Price2	Unit_Cost	Bu_Name	ProdTypeID	Build	TYPEDESCRIPTION	NAME	ABBREV	Department	product_sk
5	142.5	<null>	<null>	003	A	Polishing Equipment	Processing Equipment	Equipment	PEC	1
3	307	<null>	<null>	002	A	Manufacturing Equipment	Processing Equipment	Equipment	PEC	2
8	429.7	399.7	TPC West	015	D	Jacks	Miscellaneous	MISC	PEC	3
2	379.6	<null>	<null>	002	A	Manufacturing Equipment	Processing Equipment	Equipment	PEC	4
3	204.4	<null>	<null>	001	A	Photo Equipment	Processing Equipment	Equipment	PEC	5
6	454.5	<null>	<null>	003	A	Polishing Equipment	Processing Equipment	Equipment	PEC	6
8	267	<null>	<null>	002	A	Manufacturing Equipment	Processing Equipment	Equipment	PEC	7
2	242.6	<null>	<null>	003	A	Polishing Equipment	Processing Equipment	Equipment	PEC	8
7	476.9	448.3	TPC West	008	B	Tray Supplies	Disposable Supplies	Supplies	PEC	9
2	347.4	<null>	<null>	009	C	Photo Chemicals	Chemicals	Chemicals	PEC	10
5	290.4	<null>	<null>	002	A	Manufacturing Equipment	Processing Equipment	Equipment	PEC	11
5	457.5	<null>	<null>	001	A	Photo Equipment	Processing Equipment	Equipment	PEC	12
8	349	328.1	TPC East	013	D	Covers	Miscellaneous	MISC	PEC	13
4	242	225.1	TPC West	009	C	Photo Chemicals	Chemicals	Chemicals	PEC	14
7	190.9	<null>	<null>	003	A	Polishing Equipment	Processing Equipment	Equipment	PEC	15
7	180.4	<null>	<null>	002	A	Manufacturing Equipment	Processing Equipment	Equipment	PEC	16
2	213.2	206.8	TPC East	013	D	Covers	Miscellaneous	MISC	PEC	17
8	176.3	<null>	<null>	003	A	Polishing Equipment	Processing Equipment	Equipment	PEC	18

Supplier :
Supplier dimension not well created.
If Separate supplier dimension present, why Supplier Name in Product Dimension?

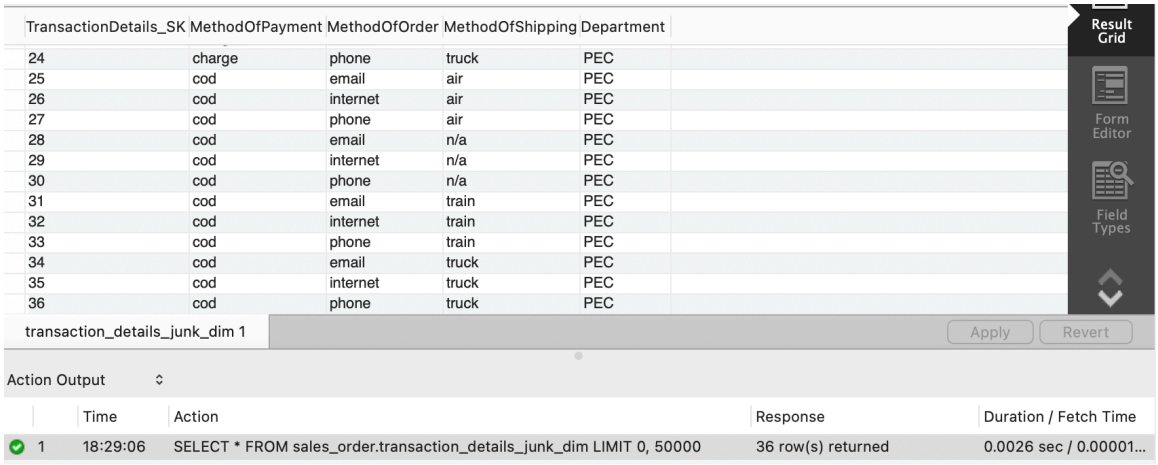
Product Dimension does not have Supplier Name the name field was of BU Name

Sales Date dimension :
Sales Date dimension output file not found

Added Sales Date Dimension output in this submission

Junk dimension :
There should be a total of 36 rows in Junk dimension. Found more

We have removed the duplicate rows and there are 36 rows now in the Junk Dimension.



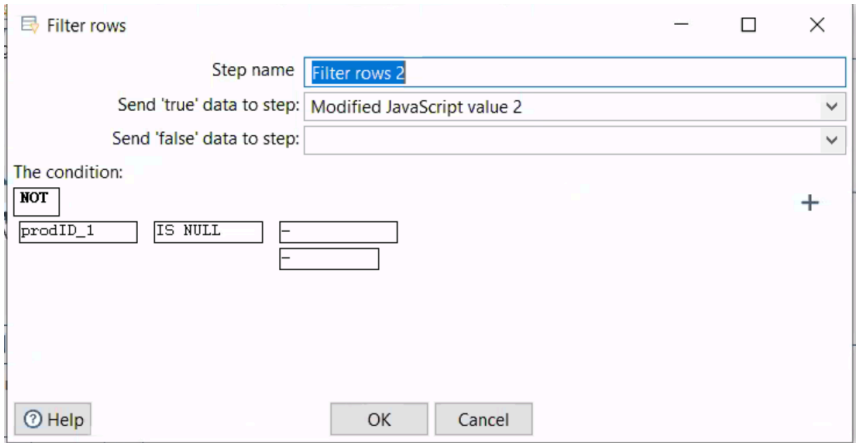
TransactionDetails_SK	MethodOfPayment	MethodOfOrder	MethodOfShipping	Department
24	charge	phone	truck	PEC
25	cod	email	air	PEC
26	cod	internet	air	PEC
27	cod	phone	air	PEC
28	cod	email	n/a	PEC
29	cod	internet	n/a	PEC
30	cod	phone	n/a	PEC
31	cod	email	train	PEC
32	cod	internet	train	PEC
33	cod	phone	train	PEC
34	cod	email	truck	PEC
35	cod	internet	truck	PEC
36	cod	phone	truck	PEC

Null records in dimensions are created appropriately to handle non-match sales records (3pts)

- Since could not find dump file, cannot check for null records

There was a mysqldump.exe version mismatch error in the sql workbench, due to which the dump files were not getting generated. The dump files generated were blank (0KB). We have resolved that issue and extracted the new dump files.

We have removed the null records by using “filter rows” method in Pentaho.



Code steps not shown :

SQL Code and KTR Screenshots has been added.

Part #4
Found only 198458 records in Sales_Fact file. Required 273594 records Missing Dump file, cannot load the data

Part #5

Define three User Queries. For each user query
User queries has been added to the report in section : VII

Sample output (2pts) - No sample outputs

Sample output has been added after every query.