Building a Data Warehouse

The Product Company

Monish Chandrashekar
Anup Aryal
Dhruvika Jeryal
Shraddha Sarkhot

Objective

- Integrate the sales data from three divisions of the company
- Optimize financial performance by monitoring sales, quantity sold, costs and profits
- Analyze data at different levels.
- Business decision making

Data source and Extraction

1. TPC-E: OLTP

2. TPE-W

   a. Sales data feed

   b. CSV files

3. PEC

   a. Sales data feed

   b. Manufacturing data feed.

   c. CSV files

Cleaning

- Invalid customer ids.

- Invalid dates.

- Extra whitespace on BUID field.

- Missing header (TPCW product).

- Garbage rows.

- Uniformity across divisions. (Customer Type)
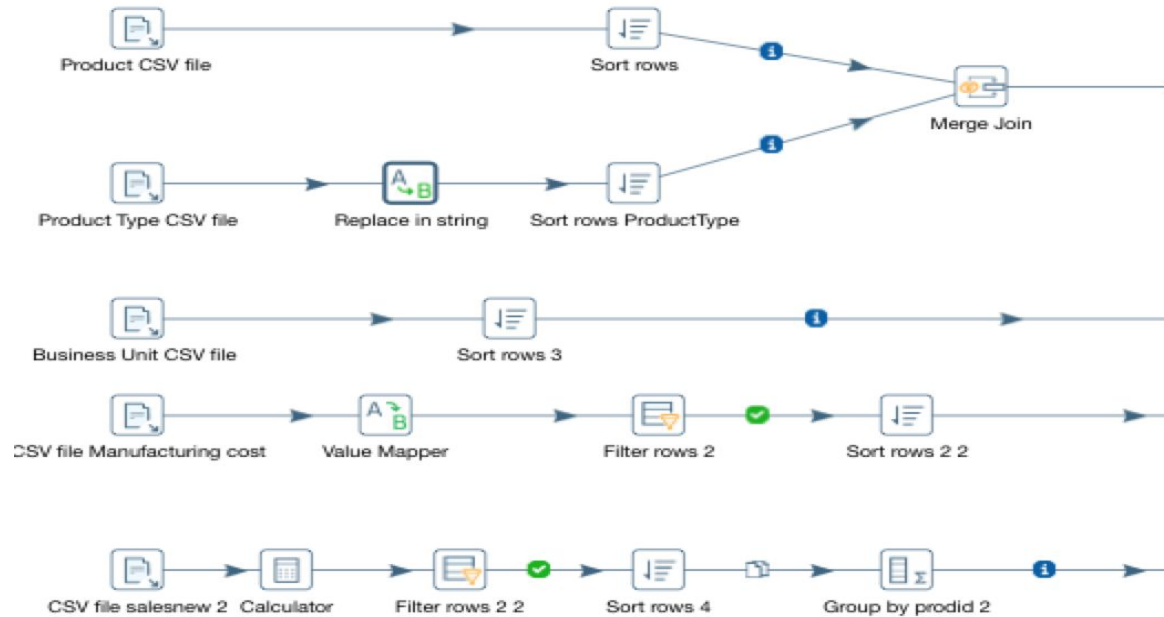
TPC-E Transformation
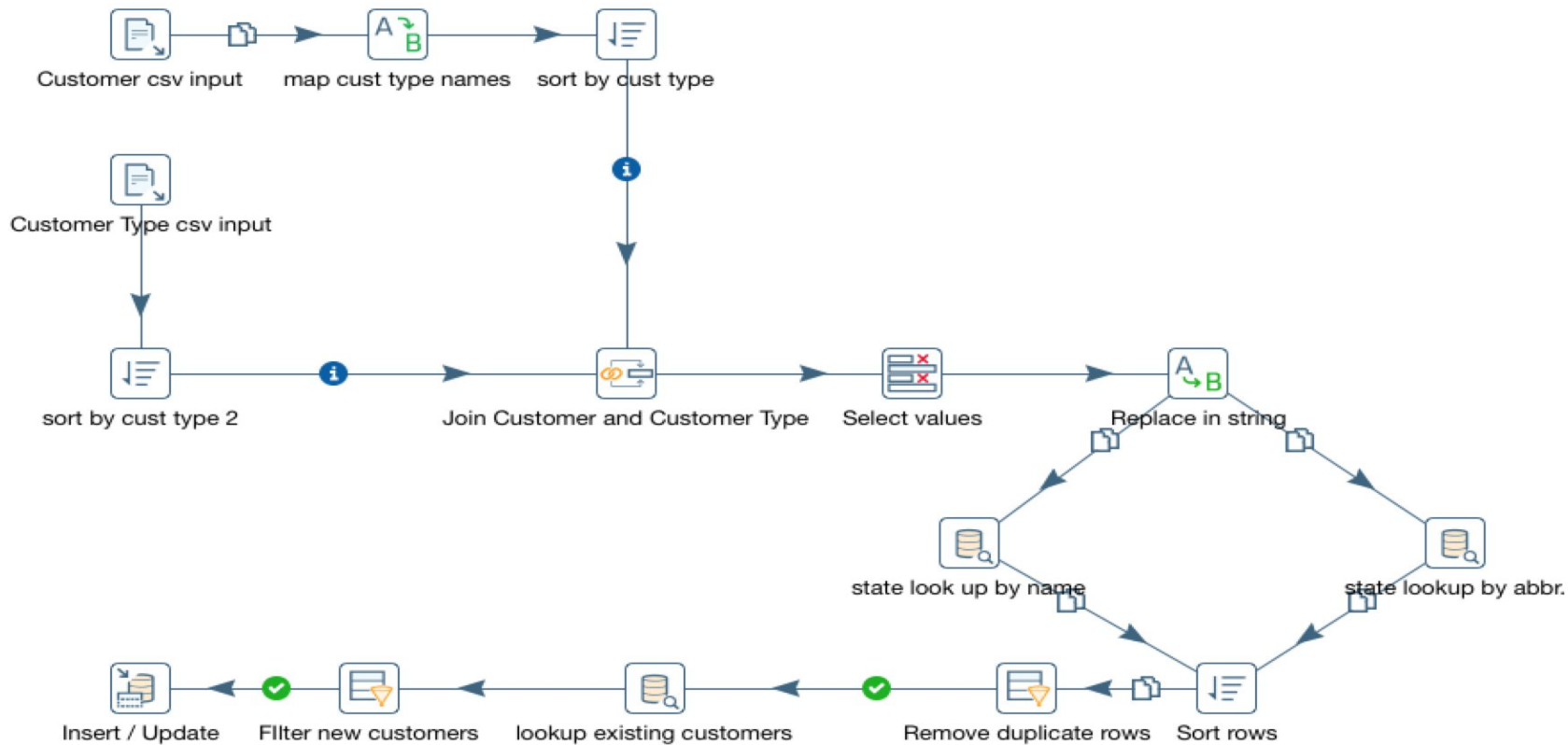
Staging Table

```sql
/*Create Staging table:*/

CREATE TABLE `salesorderstagingtable` (
  `BUID` char(1) DEFAULT NULL,
  `BUName` varchar(25) DEFAULT NULL,
  `BUAbbrev` varchar(25) DEFAULT NULL,
  `customerID` int(11) DEFAULT NULL,
  `customerName` varchar(50) DEFAULT NULL,
  `customerAddr1` varchar(50) DEFAULT NULL,
  `customerAddr2` varchar(50) DEFAULT NULL,
  `customerCity` varchar(50) DEFAULT NULL,
  `customerState` char(2) DEFAULT NULL,
  `customerZip` varchar(5) DEFAULT NULL,
  `customerTypeID` char(1) DEFAULT NULL,
  `customerTypeName` varchar(20) DEFAULT NULL,
  `invoiceID` int(11) DEFAULT NULL COMMENT 'The ID of the invoice',
  `productID` int(11) DEFAULT NULL COMMENT 'The ID of the product',
  `invoiceDtlAmt` decimal(10,2) DEFAULT NULL COMMENT 'The dollar amount of the sale',
  `invoiceDtlQty` decimal(8,2) DEFAULT NULL COMMENT 'The quantity purchased',
  `invoiceHdrSaleDate` date DEFAULT NULL COMMENT 'The date invoice was produced',
  `prodTypeID` int(11) DEFAULT NULL COMMENT 'The ID of the product',
  `prodtypeDescription` varchar(25) DEFAULT NULL COMMENT 'The!description!got!the!profuct!type',
  `productDescription` varchar(50) DEFAULT NULL,
  `productPrice1` decimal(10,2) DEFAULT NULL COMMENT 'The description of the product',
  `productPrice2` decimal(10,2) DEFAULT NULL COMMENT 'The discounted price of the product per unit ',
  `productUnitCost` decimal(10,2) DEFAULT NULL,
  `supplierID` int(11) DEFAULT NULL,
  `supplierName` varchar(50) DEFAULT NULL COMMENT 'The name of the supplier',
  `supplierAddr1` varchar(50) DEFAULT NULL COMMENT 'The first address line of the supplier.',
  `supplierAddr2` varchar(50) DEFAULT NULL COMMENT 'The second address line of the supplier.',
  `supplierCity` varchar(20) DEFAULT NULL COMMENT 'The city of the customer',
  `supplierState` char(2) DEFAULT NULL,
  `supplierZip` char(5) DEFAULT NULL COMMENT 'The five digit USPS zip  code'
);
```

PEC & TPC-W Transformation

Using pentaho DI, Kettle
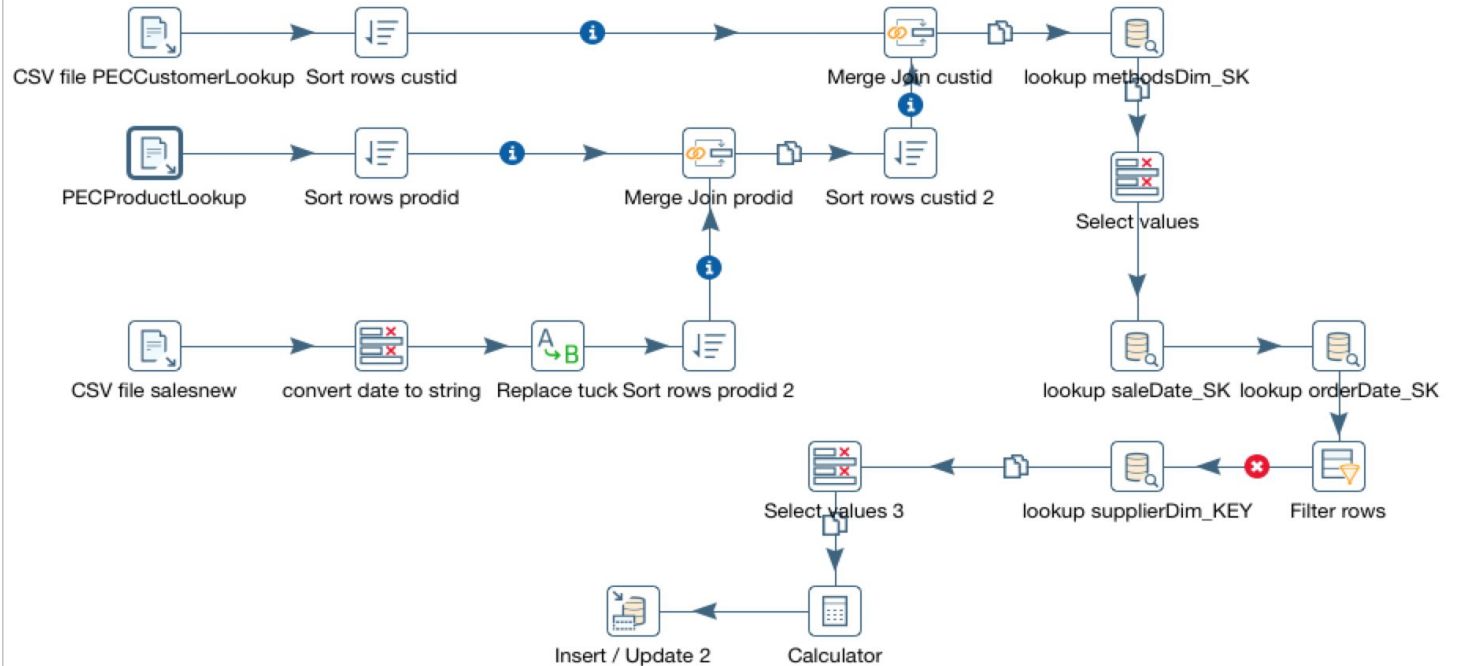

PEC Product transformation

PEC customer dimension

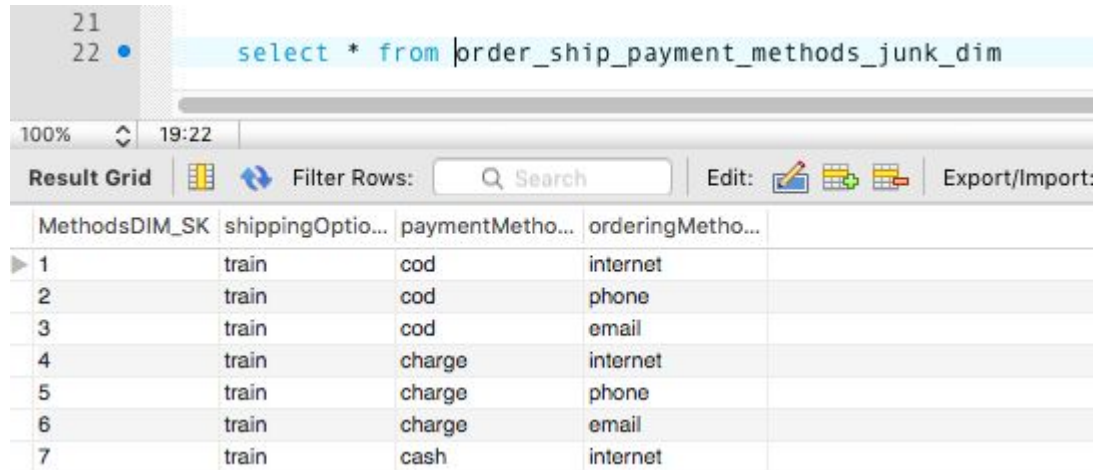# Loading

Using kettle

-Dimensions

-Facts

PEC fact

Loading of Date and Junk Dimension

- Stored Procedure for Date dimension
- Combination of shipping , payment and order methods

End User Applications (Internal Sales Report)

```
 5
 6 ●   select InvoiceNo, d.full_date as sales_date, p.productDescription as product,
 7      p.prodTypeDescription as product_type, customerName, supplierName, sales_amt, Sales_qty
 8          from Sales_Fact sf
 9          join customer_dim c
10              on sf.customerDim_SK = c.customerDim_KEY
11          join supplier_dim s
12              on sf.supplierDi_SK = s.supplierDim_KEY
13          join date_dim d
14              on sf.salesDateDim_SK = d.date_KEY
15          join product_dim p
16              on sf.productDim_SK = p.productDim_SK
17          where c.internalCustomer = 1
18          order by sales_date desc;
```

100% ◇   1:6

Result Grid | ⊞ | ↩ | Filter Rows: | Q Search | Export: 🖫 | Fetch rows: ⊞ ⊞

InvoiceNo	sales_date	product	product_type	customerName	supplierName	sales_amt	Sales_qty
▶ 28389	2010-10-27	Automobiles Fillers	Fillers	The Product Company (West)	TPC East	204.00	98.00
28389	2010-10-27	Palm Polishing Equipment	Polishing Equipment	The Product Company (West)	PEC	246.00	30.00
32414	2010-10-24	Bomber Photo Equipment	Photo Equipment	The Product Company (West)	PEC	153.00	128.00
28392	2010-10-23	Courthouses Manufacturing Equip	Manufacturing Equip	The Product Company (West)	PEC	102.00	66.00
11876	2010-10-19	Chalmers Polishing Equipment	Polishing Equipment	The Product Company (West)	PEC	362.00	175.00
11876	2010-10-19	Denigrating Polishing Equipment	Polishing Equipment	The Product Company (West)	PEC	236.00	47.00
11876	2010-10-19	Escape Manufacturing Equip	Manufacturing Equip	The Product Company (West)	PEC	121.00	109.00
11876	2010-10-19	Weeks Polishing Equipment	Polishing Equipment	The Product Company (West)	PEC	337.00	149.00
11877	2010-10-18	Sailors Manufacturing Equip	Manufacturing Equip	The Product Company (West)	PEC	220.00	107.00
11873	2010-10-16	Defeated Tray Supplies	Tray Supplies	The Product Company (West)	TPC West	209.00	23.00

SCDs

SCD Type 1:

A	B	C	D	E	F	G	H	I	J	K
productDim_SK	productID	productDescription	productPrice1	productPrice2	prodTypeID	prodTypeDescription	productUnitCost	BUID	BUName	BUAbbrev
95	12	Travelings Photo Chemicals	274.39	242.01	12	Fillers	225.07	D	Miscellaneous	NULL

Change prodTypeDescription → photo chemicals

A	B	C	D	E	F	G	H	I	J	K
productDim_SK	productID	productDescription	productPrice1	productPrice2	prodTypeID	prodTypeDescription	productUnitCost	BUID	BUName	BUAbbrev
95	12	Travelings Photo Chemicals	274.39	242.01	9	Photo Chemicals	225.07	C	Chemicals	Chemicals

SCDs

SCD Type 2:

	A	B	C	D	E	F	G	H	I	J	K
	customerDim_KEY	customerID	customerName	customerAddr1	customerAddr2	customerCity	customerState	customerZip	customerTypeID	customerTypeName	internalCustomer
	52	18	Cross	NULL	6605 Rhoncus. Avenue	Hopkinsville	NH	25620	F	US Govt	0
	92	23	Cross	NULL	112 In Road	Opelousas	GA	43882	F	US Govt	0

Add new column to say which address is current.

	A	B	C	D	E	F	G	H	I	J	K	L
	customerDim_KEY	customerID	customerName	customerAddr1	customerAddr2	customerCity	customerState	customerZip	customerTypeID	customerTypeName	internalCustomer	current_flag
	52	18	Cross	NULL	6605 Rhoncus. Avenue	Hopkinsville	NH	25620	F	US Govt	0	N
	92	23	Cross	NULL	112 In Road	Opelousas	GA	43882	F	US Govt	0	Y

SCD Type 6:

Original Row in Customer Dimension

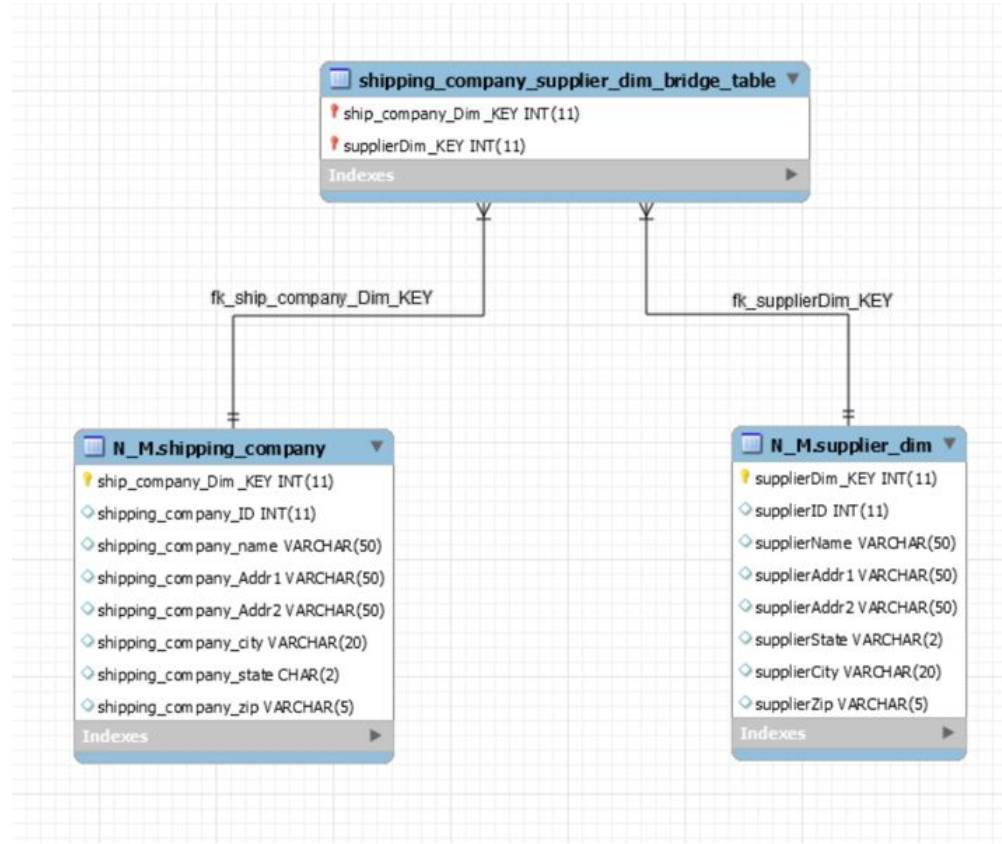| customerDim_KEY	customerID	HistoricalcustomerName	CurrentcustomerName	customerAddr1	customerAddr2	customerCity	customerState	customerZip	customerTypeID	customerTypeName	internalCustomer	CurrentRowIndicator
1	1	Room Plus Inc	Room Plus Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Current

Rows in Customer Dimension following first customer name reassignment

customerDim_KEY	customerID	HistoricalcustomerName	CurrentcustomerName	customerAddr1	customerAddr2	customerCity	customerState	customerZip	customerTypeID	customerTypeName	internalCustomer	CurrentRowIndicator
1	1	Room Plus Inc	Blue Room Plus Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Expired
1	1	Blue Room Plus Inc	Blue Room Plus Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Current

Rows in Customer Dimension following second customer name reassignment

customerDim_KEY	customerID	HistoricalcustomerName	CurrentcustomerName	customerAddr1	customerAddr2	customerCity	customerState	customerZip	customerTypeID	customerTypeName	internalCustomer	CurrentRowIndicator
1	1	Room Plus Inc	Blue Room Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Expired
1	1	Blue Room Plus Inc	Blue Room Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Expired
1	1	Blue Room Inc	Blue Room Inc	Suite 836	6037 Ecuadorian Street	Clinton	MI	49236	C	Commercial	0	Current

N:M relationships

Thank YOU!