1. DataCleansing

File (.csv & others if applicable)	Attribute	Problem	Resolution Strategy (attach code)

PECbusiness_unit/TPCWbusiness_unit/TPCEbusiness_unit — BUID — spaces in BUID — Replace String (in Pentaho)

PECbusiness_unit/TPCWbusiness_unit/TPCEbusiness_unit — ABBREV — Null Value in Miscellaneous column — Replace String (in Pentaho)

PECcustomer_type/TPCW/TPCE — TypeName — Gov and Govt — Replace String (in Pentaho)

PECcustomer/TPCW/TPCE — address — Special Characters (.,) — Replace String (in Pentaho)

PECcustomer/TPCW/TPCE — City — Special Characters (.,) — Replace String (in Pentaho)

PECcustomer/TPCW/TPCE — custType — Special Characters (.,) — Replace String (in Pentaho)

PECcustomer/TPCW/TPCE — custType — Gov and Govt — Replace String (in Pentaho)

PECinvoice/TPCW/TPCE — salesDate — Date format — Convert to string and With Java Script (in Pentaho)

PECinvoice/TPCW/TPCE — orderDate — Date format — Convert to string and With Java Script (in Pentaho)

PECmanufacturingCosts — Year — Change YY to YYYY — Replace string (in Pentaho)

PECproduct_type/TPCW/TPCE — typeDescription — Replace Equip with equipment — Replace string (in Pentaho) using Regex

PECproduct_type/TPCW/TPCE — BUID — Extra spaces — Replace string (in Pentaho)

PECproduct/TPCW/TPCE — prodDescription — Replace Equip with equipment — Replacesttring (in Pentaho) using Regex
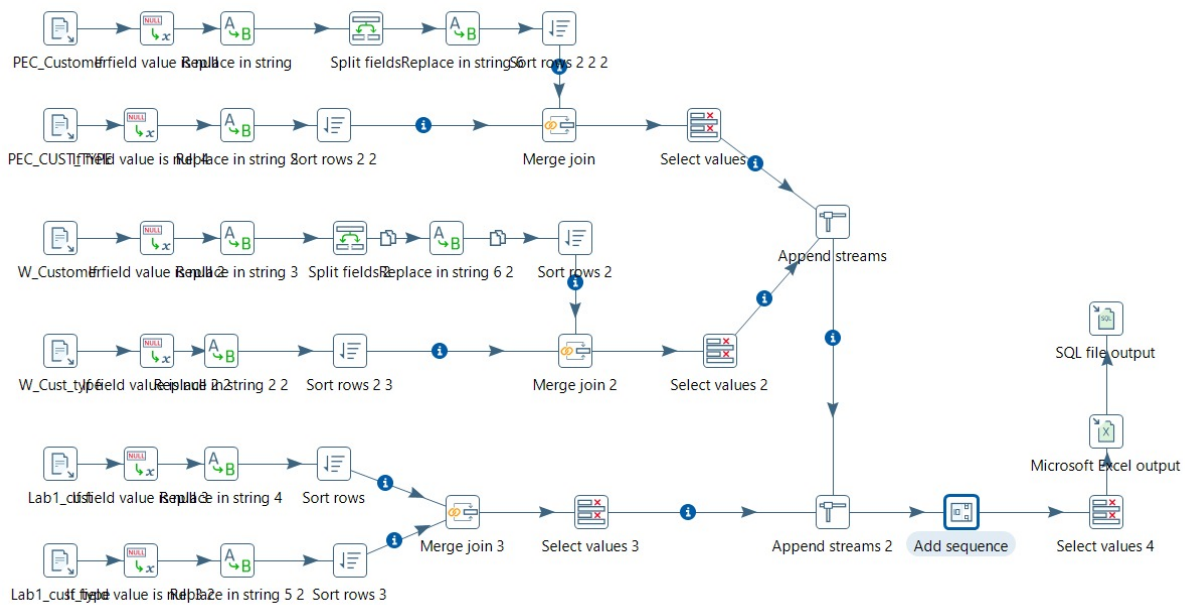
PECproduct/TPCW/TPCE — unitCost — calculate — using manufacturingcost and month (in Pentaho)

Attached KTR files

3. TablePopulation

DM Table	Table Population Process (attach code)

Customer_Dim - creating through KTR files for cleaning with surrogate keys

MySQL Workbench

xampp - Warning - not suppor...

File   Edit   View   Query   Database   Server   Tools   Scripting   Help

Query 1   product   product   product   customer   customer_dim

```
1 •  SELECT * FROM deshpande_205.customer_dim;
```

Navigator

SCHEMAS

Filter objects

- deshpande_205
  - Tables
    - customer_dim
    - junk_dim
    - orderdate_dim
    - product_dim
    - sales_fact
    - salesdate_dim
    - supplier_dim
      - Columns
        - Supplier_SK
        - Supplier_id
        - SupplierName
        - SupplierAddress1
        - SupplierAddress2
        - SupplierCity
        - SupplierState
        - SupplierZipcode
      - Indexes
      - Foreign Keys
      - Triggers
    - Views
    - Stored Procedures
    - Functions
  - new_schema
  - pe_kettle
    - Tables
    - Views
    - Stored Procedures
    - Functions

Administration   Schemas

Information

Table: junk_dim

Columns:

TransactionDetails_SK	int(11) AI PK
MethodOfPayment	varchar(
MethodOfOrder	varchar(
MethodOfShipping	varchar(

Result Grid

Customer_SK	Cust_id(NK)	CustType	CustTypeID	Name	Address1	Address2	City	State	Zip	Department
1	3	COMMERCIAL	NULL	Baxter May	796-2366 Vel Avenue	NULL	Meriden	FL	58328	PEC
2	6	COMMERCIAL	NULL	Raphael Allison	734-2598 Sed Avenue	NULL	La Crosse	HI	58702	PEC
3	9	COMMERCIAL	NULL	The Product Company (East)	20021 Barnes Road		Stratford	NY	13470	PEC
4	11	COMMERCIAL	NULL	Austin Ferrell	8666 Justo Road		Pullman	PA	55979	PEC
5	14	COMMERCIAL	NULL	Serrano	7212 Eleifend Street		Princeton	KY	66785	PEC
6	15	COMMERCIAL	NULL	Colton Maldonado	1869 Leo Road		Cedar Rapids	SD	97640	PEC
7	16	COMMERCIAL	NULL	Googol	5453 Dictum Avenue		Jenks	LA	96393	PEC
8	20	COMMERCIAL	NULL	Pewter Gym	9028 Non Street		Sandy	FL	87202	PEC
9	27	COMMERCIAL	NULL	Clare Baird	6152 Auctor	Road	Wilson	MS	29338	PEC
10	29	COMMERCIAL	NULL	Blevins	2571 Donec Avenue		Winooski	SD	65512	PEC
11	30	COMMERCIAL	NULL	Knox Reid	500 Eu	Avenue	Pass Christian	WY	35949	PEC
12	32	COMMERCIAL	NULL	Ferengi Treasures	716-6837 Vulputate ...		Paramount	NY	35644	PEC
13	33	COMMERCIAL	NULL	Tallulah	3376 Aliquet Street		Batavia	NM	50742	PEC
14	34	COMMERCIAL	NULL	The Product Company (West)	21 E Bullard Ave		Fresno	CA	93710	PEC
15	1	EDUCATION	NULL	Haynes	2920 Auctor	Road	Moraga	AZ	70314	PEC
16	7	EDUCATION	NULL	Martin Donaldson	2704 Ad Road		Gardner	GA	31471	PEC

Output

Action Output

#	Time	Action	Message
1	23:37:53	SELECT * FROM deshpande_205.customer_dim LIMIT 0, 1000	121 row(s) returned

CSV file input   If field value is null   Select values 2   Sort rows   Add sequence   Select values   Microsoft Excel output

Unique rows (HashSet)

SQL file output

CSV file input

Modified JavaScript value

Unique rows (HashSet)

Select values 2   Add sequence Select values 3

Select values   Calculator

Microsoft Excel output   SQL file output
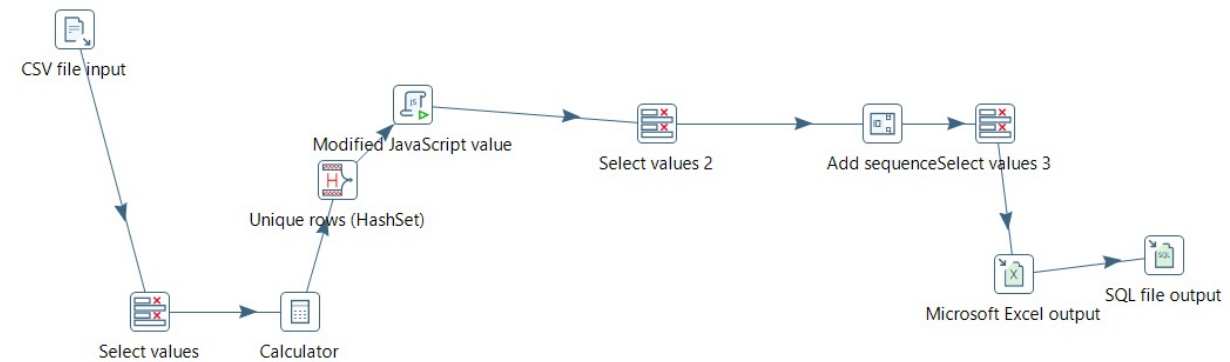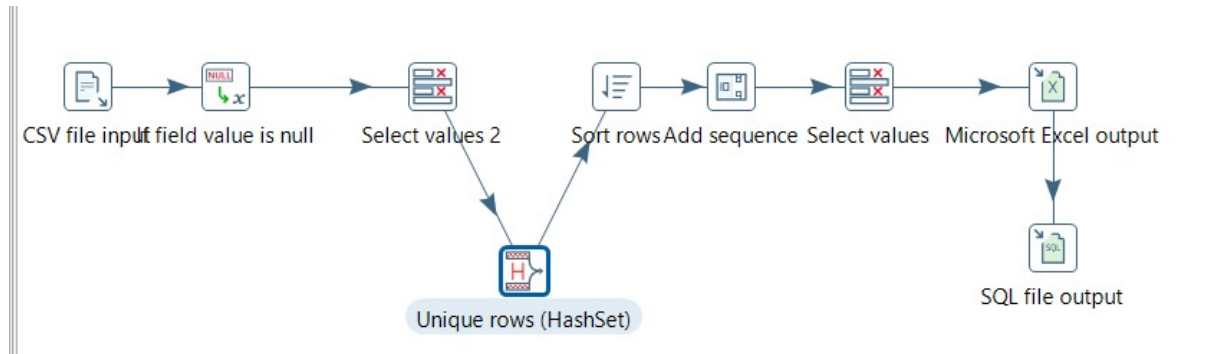
Junk_Dim - creating through KTR files with for cleaning surrogate keys

OrderDate_Dim - creating through KTR files for cleaning with surrogate keys

Product_Dim - creating through KTR files for cleaning with surrogate keys
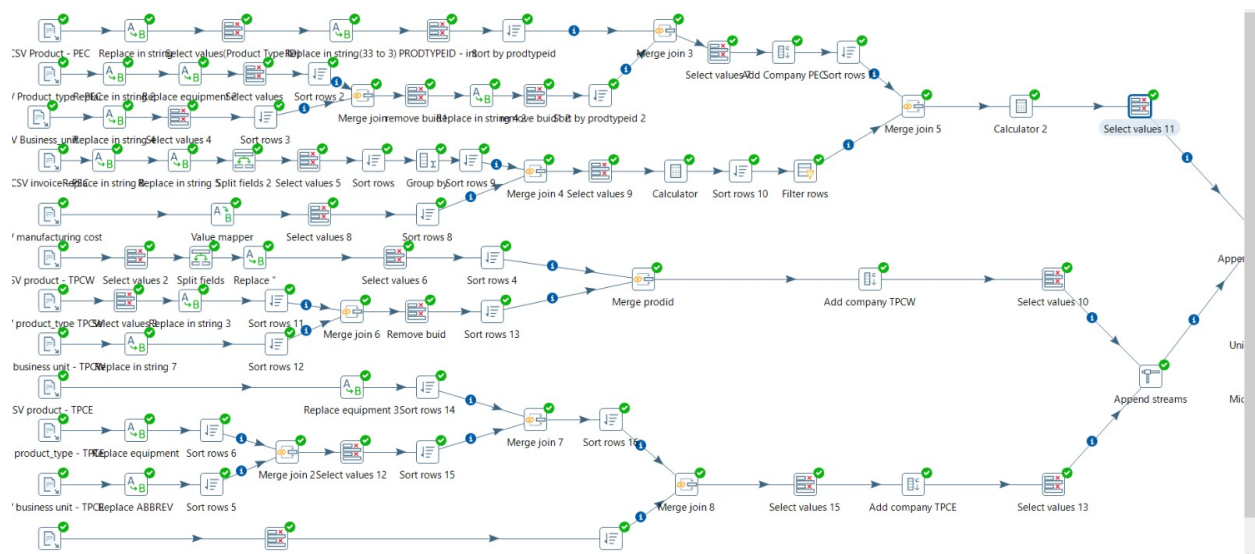
SalesDate_Dim - creating through KTR files for cleaning with surrogate keys
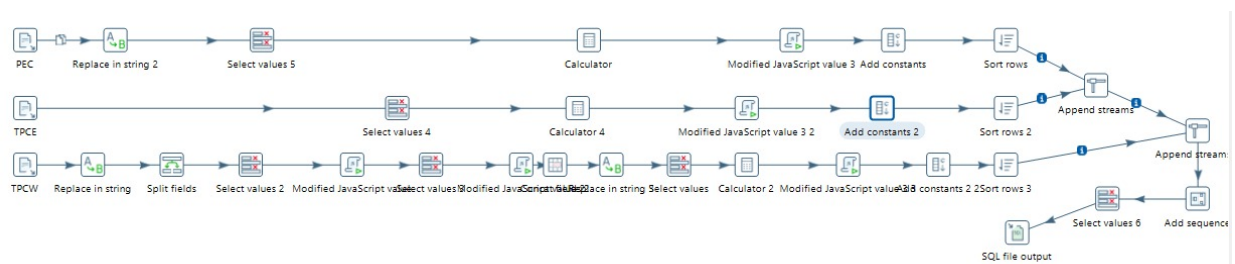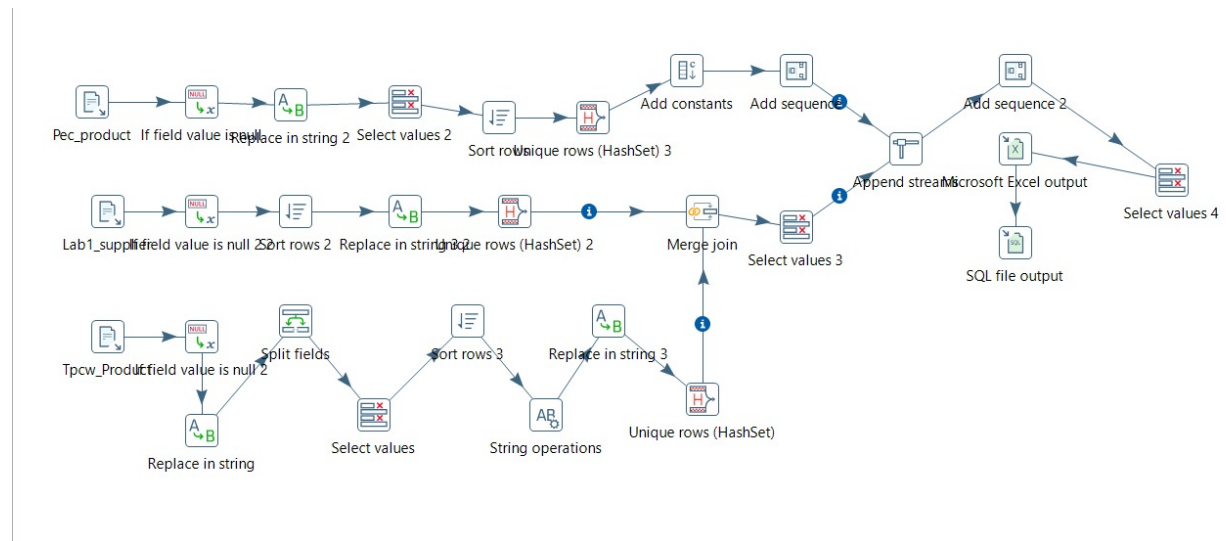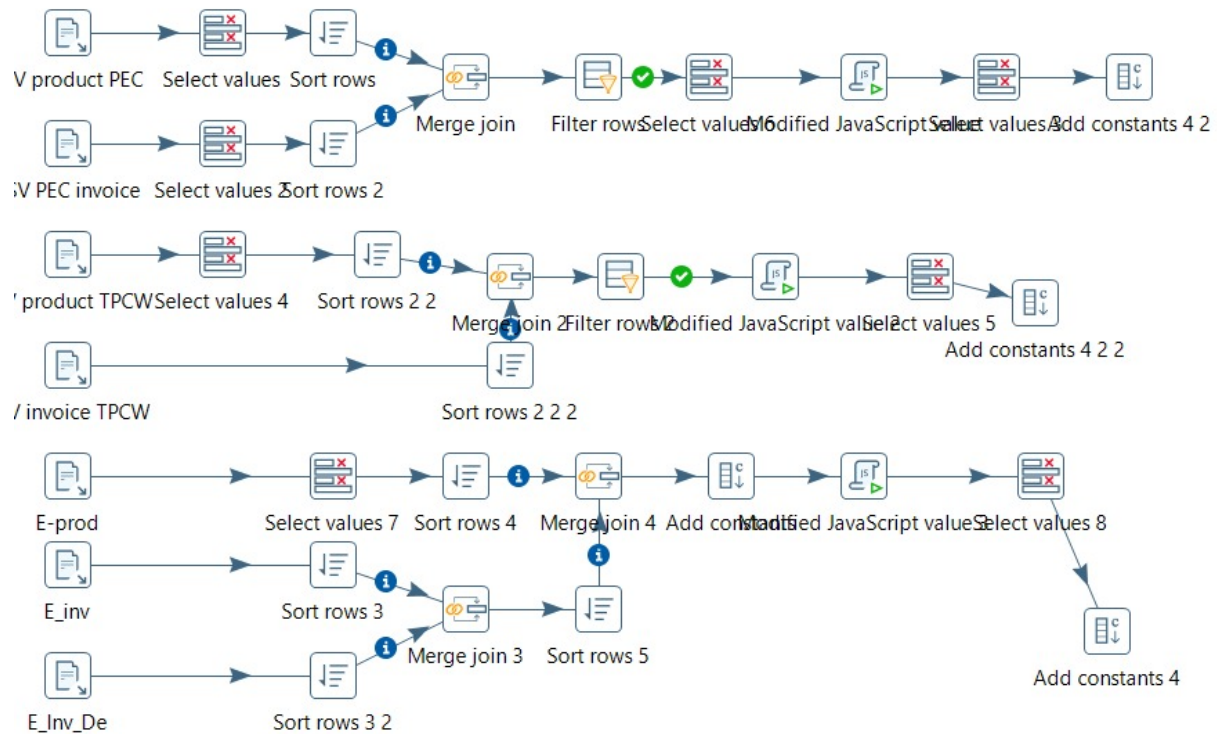
Supplier_Dim - creating through KTR files for cleaning with surrogate keys

Sales_Fact - creating through KTR files

2. DataTransformation

DM Table	Image Creation Process (attach code)

Dimension tables and Fact table created in SQL Workbench attached Dimension model

junk_dim
- TransactionDetails_SK INT(11)
- MethodOfPayment VARCHAR(45)
- MethodOfOrder VARCHAR(45)
- MethodOfShipping VARCHAR(45)
- Indexes

orderdate_dim
- OrderDate_SK INT(11)
- Year INT(11)
- Quarter INT(11)
- Month INT(11)
- FiscalMonth INT(11)
- FiscalYear INT(11)
- FiscalQuarter INT(11)
- FiscalWeek INT(11)
- FiscalDay INT(11)
- Day INT(11)
- Week INT(11)
- Indexes

salesdate_dim
- SalesDate_SK INT(11)
- Year INT(11)
- Quarter INT(11)
- Month INT(11)
- Day INT(11)
- Week INT(11)
- FiscalDay INT(11)
- FiscalWeek INT(11)
- FiscalMonth INT(11)
- FiscalQuarter INT(11)
- FiscalYear INT(11)
- Department VARCHAR(45)
- Indexes

sales_fact
- Customer_SK INT(11)
- Product_SK INT(11)
- Supplier_SK INT(11)
- SalesDate_SK INT(11)
- OrderDate_SK INT(11)
- TransactionalDetails_SK INT(11)
- invoiceID(DD) INT(11)
- ShippingCost INT(11)
- SalesAmount INT(11)
- ShippingDuration INT(11)
- Discounted TINYINT(4)
- SourceUnit VARCHAR(5)
- Qty INT(11)
- Department VARCHAR(45)
- Indexes

product_dim
- Product_SK INT(11)
- Product_ID(NK) INT(11)
- Prod_Descp VARCHAR(45)
- ProdTypeID INT(11)
- Buid VARCHAR(45)
- Name VARCHAR(45)
- Abbrev VARCHAR(45)
- Price1 INT(11)
- Price2 INT(11)
- Unit_Cost INT(11)
- Type_Description VARCHAR(45)
- Department VARCHAR(45)
- Indexes

supplier_dim
- Supplier_SK INT(11)
- Supplier_id INT(11)
- SupplierName VARCHAR(45)
- SupplierAddress1 VARCHAR(45)
- SupplierAddress2 VARCHAR(45)
- SupplierCity VARCHAR(45)
- SuplierState VARCHAR(45)
- SupplierZipcode INT(11)
- Indexes

customer_dim
- Customer_SK INT(11)
- Cust_id(NK) INT(11)
- CustType VARCHAR(45)
- CustTypeID VARCHAR(45)
- Name VARCHAR(45)
- Address1 VARCHAR(45)
- Address2 VARCHAR(45)
- City VARCHAR(45)
- State VARCHAR(45)
- Zip VARCHAR(45)
- Department VARCHAR(45)
- Indexes