

Team Project 1

Load aw2017neu_dw from several data sources. These sources were extracted and manipulated from Adventureworks2017 so you can use schema documentation to understand data and relationships.

Databases:

Table 1: Source Databases

Database	DBMS	Backup/Export	Schema
aw2017neu_hr	Oracle	aw2017neu_hr_oracle.zip	HumanResources
aw2017neu_person	Oracle	aw2017neu_person_oracle.zip	Person
aw2017neu_sales	PostgreSQL	aw2017neu_sales_postgresql.backup	Sales
aw2017neu_production	SQL Server	aw2017neu_production_sqlserver.bak	Production
aw2017neu_purchasing	MySQL	aw2017neu_purchasing_mysql.sql	Purchasing
Input seed for DimDate	csv	DimDate_input.csv	

Target db: aw2017neu_DW_mysql.sql – MySQL script

Table 2: DW Target Tables

table_schema	table_name	comment
aw2017neu_dw	dim_rejectcodes	Track Rejection Codes, Note: Team manually inputs
aw2017neu_dw	dim_sorsystem	Track SOR_ID, Note: Team manually inputs
aw2017neu_dw	dimcurrency	
aw2017neu_dw	dimcustomerperson	
aw2017neu_dw	dimdate	
aw2017neu_dw	dimemployee	
aw2017neu_dw	dimgeography	
aw2017neu_dw	dimlocation	
aw2017neu_dw	dimpayhistory	SCD 2
aw2017neu_dw	dimproductcosthistory	SCD 2
aw2017neu_dw	dimproductpricehistory	SCD 2
aw2017neu_dw	dimproducts_purchased	only products purchased from vendors
aw2017neu_dw	dimproductsall	
aw2017neu_dw	dimproductvendor	
aw2017neu_dw	dimpromotion	
aw2017neu_dw	dimsalesterritory	
aw2017neu_dw	dimscrapreason	
aw2017neu_dw	dimshipmethod	
aw2017neu_dw	dimstore	
aw2017neu_dw	dimvendorcontacts	
aw2017neu_dw	dimvendors	

table_schema	table_name	comment
aw2017neu_dw	fact_workorder	
aw2017neu_dw	fact_workorder_rejects	
aw2017neu_dw	fact_workorderrouting	
aw2017neu_dw	fact_workorderrouting_rejects	
aw2017neu_dw	factinternetsales	
aw2017neu_dw	factinternetsales_rejects	
aw2017neu_dw	factpurchases	Purchase Orders
aw2017neu_dw	factpurchases_rejects	
aw2017neu_dw	factstoresales	
aw2017neu_dw	factstoresales_rejects	

Deliverables:

Load target DW using the following:

Database	DBMS	Data Integration Software
aw2017neu_hr	Oracle	Talend
aw2017neu_person	Oracle	Talend
aw2017neu_sales	PostgreSQL	Talend
aw2017neu_production	SQL Server	SSIS
aw2017neu_purchasing	MySQL	SSIS
DimDate	csv	Alteryx
List Price & Cost SCD tables	SQL Server	Alteryx, SSIS, Talend

Bonus: Besides completing above also complete a versions of the project using SSIS & Talend entirely.

Submit:

- Schema, List of Tables, Row Counts – in spreadsheet or table
- Screen capture of success run of Talend & SSIS jobs with row counts
- Upload Talend, SSIS & Alteryx jobs

Grading criteria:

- Loading all tables
- Handling errors in loading Fact tables
- Correct target table row counts – total rows, valid rows and invalid rows.
- Split of invalid rows by type and reason for error.
- Structure of data integration code
- Completing data integration