Team Project 01 (Version 1.1)

Databases:

- Source database:
    - AdventureWorks2014_SOR
- Source files:
    - EmployeePayHistory.csv
    - ProductList_CostHistory.csv
    - ProductList_PriceHistory.csv
    - DimDate.csv
- Target databases: AdventureWorks2014_Project (student creates)
    - Target database for Talend – either MySQL or Oracle as your team can choose.
    - Target database for SSIS – SQL Server

Data Integration Tools to be used:

- Talend
- SSIS

Objectives:

- Create schema & Load tables in target database that will contain data on:
    - All product purchases made by AdventureWorks from vendors, i.e. purchase orders
    - A "rejects" table for product purchases that have orders that do not meet referential integrity rules (enforced by data integration processes NOT using database constraints or triggers)
    - List of products that might be purchased from Vendors, in other words, all the products that are not made by AdventureWorks
        - Need to include product attributes such as list price, standard cost, model name, subcategory name and category name.
        - Note: if hierarchy applicable then this needs to be a star and NOT a snowflake design
    - A slowly changing dimension (SCD) for list prices for products above
        - SCD 2 or higher
    - A slowly changing dimension (SCD) for standards costs for products above
        - SCD 2 or higher
    - Vendors that AdventureWorks might purchase products from
    - Products that AdventureWorks might purchase from vendors and associated with the vendors that sell them.
    - Vendors' employees who interact with AdventureWorks and include attributes about those people.
    - AdventureWorks' employees who have bought products from vendors and include attributes about those people.
    - A slowly changing dimension (SCD) for AdventureWorks employee pay (only employees listed above)
        - SCD 2 or higher
    - AdventureWorks' sales territories

- o   AdventureWorks' ship methods
- o   A geography dimension that AdventureWorks could use as outrigger for dimensions above.
- o   A calendar dimension that AdventureWorks could use for facts and dimensions
- o   AdventureWorks product inventory for purchase products and locations stored in

- The tables created and data in each is based on your design. There may need to be additional dimensional tables based on how you design the target database

Rules of Project:
- Cannot use any views, stored procedures or functions from AdventureWorks2014_SOR
- Need to follow data integration standards (separate documents)

Submit:
- Schema, List of Tables, Row Counts – in spreadsheet or table
- Screen capture of target data model (only one db schema is required)
- Screen capture of success run of Talend & SSIS jobs with row counts
- Performance statistics of Talend load
- Queries for validations (SQL queries)
    - o   Data integration rejection records
        - ▪   Number of selected & rejected rows
        - ▪   Number of rows by rejection reason
    - o   Ranked order of Vendors by purchase amount $
    - o   Ranked order of products purchased by amount $
        - ▪   By category
        - ▪   By subcategory
        - ▪   By product model (top 20)
        - ▪   By product (top 20)
    - o   List of employees who purchased products with phone, email & address
    - o   List of employees who purchased products with pay rate & raises (SCD)
    - o   List of purchasing vendor contacts with vendor name, phone, email & address
    - o   List of product prices by product order by product and SCD effective ascending
    - o   List of standard costs by product order by product and SCD effective ascending