# Requirements: V3

Data is recorded on student absences, misbehavior and misbehavior consequences (or resolutions). Samples of reports will be available. With the source data (below) complete the following:

* Create a dimensional data model for a Schools DW
* Load data from the data sources provided into your Schools DWs
  + ETL tool:
    - Talend Enterprise Data Integration
    - SSIS
  + DW database:
    - Schools\_DW – MySQL - loaded using Talend
    - Schools\_DW - Microsoft SQL Server 2014 - loaded using SSIS
* Separate “reject” rows, i.e. rows that have invalid foreign keys (FKs) in regards to dimensional data provided in data sources. Also put a rejection reason in the Rejects table using column such as:
  + DW\_Reject\_Reason nchar(80) NULL
* Each table needs to record:
  + ETL process ID
  + Create Date
  + Modification Date
* Create reports on student absences, misbehavior and behavior resolutions listing number of students, number of instances, types of instances, types of resolution and location of instances. Note: See below.
  + BI tool:
    - Schools DW (MySQL) – Tableau, PowerBI
    - Schools DW (SQL Server) – Qlik Sense, PowerPivot
    - DW\_CTL – Microsoft BI

# Data Sources:

Table : List of Data Source Table with Rows Counts

| **Table Name** | **Row Count** |
| --- | --- |
| Absence\_Reason | 48 |
| Behavior\_Event\_Type | 53 |
| Behavior\_Location\_Type | 14 |
| Behavior\_Resolution\_Type | 22 |
| Behavior\_Role | 5 |
| Dim\_Date | 6,941 |
| Grade\_Level | 24 |
| Schools | 18 |
| Student\_Absence\_AL | 93,216 |
| Student\_Absence\_NE | 1,271,964 |
| Student\_Absence\_NW | 823,689 |
| Student\_Absence\_SE | 1,098,260 |
| Student\_Absence\_SW | 1,002,152 |
| Student\_Behavior\_AL | 1,287 |
| Student\_Behavior\_NE | 30,926 |
| Student\_Behavior\_NW | 7,723 |
| Student\_Behavior\_Resolution\_AL | 969 |
| Student\_Behavior\_Resolution\_NE | 28,595 |
| Student\_Behavior\_Resolution\_NW | 8,832 |
| Student\_Behavior\_Resolution\_SE | 20,285 |
| Student\_Behavior\_Resolution\_SW | 8,770 |
| Student\_Behavior\_SE | 19,413 |
| Student\_Behavior\_SW | 9,527 |
| Students | 363,466 |

Student Absences, Behavior & Resolution data is split by the five school districts (AL, NE, NW, SE & SW) with the data being stored in the following:

* NE - SQL Server
* NW - PostgreSQL
* SE & SW – MySQL
* AL – csv files

The remaining data sources are stored as either tab delimited or xml files.

# Hints:

After data quality checking the results may be:

|  |  |
| --- | --- |
| **Table** | **Row Counts** |
| Student\_Absence | 4,085,018 |
| Student\_Behavior | 62,039 |
| Student\_Behavior\_Resolution | 61,413 |
| **Table** | **Row Counts** |
| Student\_Absence\_Rejects | 204,263 |
| Student\_Behavior\_Rejects | 6,837 |
| Student\_Behavior\_Resolution\_Rejects | 6,038 |

# Deliverables:

* Data Model (PDF)
* ETL jobs – job & workflow diagrams (PDF) – generate Talend documentation
* Dashboard(s)
  + Screenshots
  + Source files for each BI tool zipped
* Database backup of Schools DW (localhost db, user: neu & password info7290)
* Report on ETL process metadata showing job runs
* Class Presentation & Demo of ETL
  + Need to demonstrate each job works with TAs

# Bonus: - data has to correct to apply

* Top 4 Fastest DW Loads (total time to load both DWs) - 2X bonus points
* Top 4 Visually appealing dashboards - 1X bonus points

# Types of analysis:

* Absence
  + Reason type, reason, gender, language, grade level, class affiliation
  + amount of absences (#, periods, minutes), # of tardy
* Behavior
  + Event type, event, role, event location, school, school type, district, grade level, city, gender
  + Number of occurrences
* Behavior Resolution
  + Resolution(s), event type, event, role, event location, school, school type, district, grade level, city, gender
  + Number of occurrences