Sprint 3

# Summary

Here we are!!! After assessing the data quality of our source and generating a report to quickly provide value added to our data (and our jobs). It’s time to load the data into an actual star schema that will be used to analyze bus movements and activity throughout the day and the regions

# Requirements

The analyst provided you with a mapping document that he built with the architect. The mapping document contains target specifications (the data model that you’ll have to pour data into) as well as source and transformations. It’s basically your roadmap to build an ETL process to load specified columns from the source into a star schema.

In addition to the usual gtfs\_stm.zip file containing data about bus schedules and stops, you’ll have to ingest a geo.zip file, containing postal code, district and city information about each bus stop.

# Provided documentation and artifacts

* A mapping document (mapping.xls) detailing the target schema, the source and the transformations you’ll have to make. The file is located @ ./Sprint3/Starter
* A data folder containing all the data needed to perform the load into data warehouse. The data folder is located @ ./Data

# Expected deliverables

* A database named “STM\_DW” with all the tables specified in the target section of the mapping document. The database should contain 4 dimensions and 1 fact table. The tables should all have primary key and foreign keys when needed.
* A SSIS solution that loads the files into the data warehouse, the following features are required:
  + Being able to do incremental loads: if I load the file twice, I’m expecting to see only the new and modified records.
  + Some kind of error management. Don’t want to see the package fail because of data quality
  + Calculating the delta between current and previous load

# Advice and hints

* Read chapter 6 and 7 of the first book. Everything is in there.
* Experiment as much as you can, you’ll face tough challenges that can be solved by thinking in SSIS terms
* Google extensively,