**Motor Inc ETL Data warehouse design overview**

Overview:

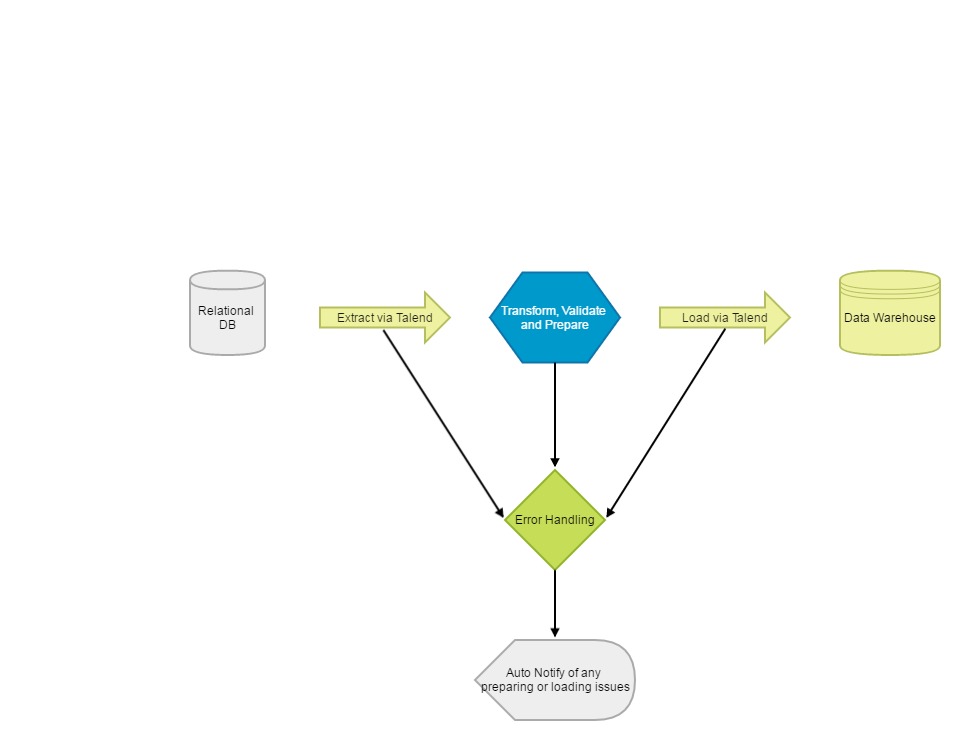
The data will be:

1. Extracted from the 3 relational database postgres tables using Talend for Big Data Integration (talend has inbuilt postgres components to pull data with sql commands from the tables)
2. Be transformed and validated (e.g. POLICY\_STATUS field must be either “ACITVE” or “NOT-ACTIVE”, date fields can be validated for correct formats etc etc.)
3. Loaded into the postgres data warehouse designed (motor\_inc\_test\_create\_ddl.sql)

The talend jobs will run in a unix script on a unix server at fixed and frequent times e.g. daily or every 5 minutes or 30 minutes etc (depending on the business need). Any records in the relational DB that were updated or added since the previous extraction will be extracted and transformed, validated and loaded into the data warehouses.

Any new values in the relational tables that need to be added to the dimension tables will be validated and added. Then the main “Line\_Fact” table values will be prepared e.g. VALID\_LINE will be populated with “TRUE” if LINE.LINE\_STATUS = 0 and LINE.LINE\_ERROR\_CODE = 0. CLAIM\_COSTS will be populated if VALID\_LINE = “TRUE” and the POLICY\_STATUS = “Active”. These aggregated values and other values will be loaded into the “Line\_Fact” table.

Any values that do not load will be collected and logged and email alerts created alerting for example that 5 records did not load – link to the log locations etc.



Talend postgres data integration flows

Job to extract

Job to validate

Job to transform

Job to load dimension tables

Job to load fact table

The data will be pulled from the relational DB and transformed and loaded directly into the data warehouse. The talend data integration job below is just an example and shows csv files being loaded into the data warehouse Line fact table.

