ETL Key Performance Indicators for data testing

1)  **Source reject** – Assume that developer created proper metadata for source. But the source contains data which differs with the metadata. In this case, the data will be outright rejected from the source. It will not come into the mainstream for ETL processing. Nobody will come to know that it was rejected from the source. *It should be zero.*

2) **Target reject** – Same as above, developer created a metadata for target. But the processed data contains different data which does not match with the target metadata. The data is rejected before inserting in to the target. If you don’t capture, it is lost.*It should be zero.*

3) **Source is having higher data** – When you compare source data and target data after completion of ETL, you may see that source is having higher records than target, reason can be anything. *It should be zero when compared with the target.*

4) **Target is having higher data** – This scenario exists when source data is deleted but still it exists in the target. This needs target data updating at the time of regular ETL. *It should be zero when compared with the source.*

5) **Source data mismatches with the Target** – This happens during ETL, some transformation is applied on source or during migration from one data source to another, it would automatically convert the data and insert into the target. In short, source is different than target. *Data mismatch need to be verified and this count should be zero.*

6) **Target data mismatches with the Source** – This happens when the data update takes place on source, but not changed in the target. *Data mismatch need to be verified and this count should be zero.*

7) **Timestamp issue** – This is one of the most promising issues during ETL process. When data comes from different time zones, collected at central place and then brought to the staging for Warehousing. Usually timestamp is used for incremental load, this timestamp creates a problem which skips data for specific time zone

1. Another issue in time zone is timestamp – All of the databases do not support the entire time stamp format. While converting data from one database to another database, this time conversion needs to be taken care of.

8) **Time lag during ETL** – Many times ETL job runs for few hours on every day basis. Need to have proper signature for ETL completion. If the signature of ETL completion is not proper and if some data insertion or update happens during the period of ETL, then certainly at consecutive execution the job is going to skip few records. *Need special method to track the source and target data.*

9) **Null handling** **and metadata conversion** – Every time when ETL is executed, proper care for null handling and metadata conversion is to be taken care. *Need schema compliance check.*

10) **Monitoring** – After completion of ETL it is always required that proper auditing and logging is implemented. This is required for ensuring that everything is working as expected. All the above points and other items need to be logged during ETL project development. *Need reporting bursting on the audited and logged data.*

In addition to this there are many activities and practices to follow to ensure that the data movement from source to target is as expected. If it is not ensured then, it directly reflects in the data loss which is a revenue loss for a company. Because, if the source data for the dashboard is wrong, it would be shown as it is in the Dashboard or report. This is false and partial data and not corrects figures. Many times, this is not identified unless it is compared with some other data. This comparison is a far remote chance in today’s fast food and high speed data requirement. In many cases error is not generated during ETL process, which gives false notice that the ETL is working fine.

**Are you doing all of these during your ETL project development**? If not, ensure that you do all these next time when you start with your ETL project. Whether you are a client or a developer or business owner, if you are moving data from one location to other location, try to ensure all of above points are checked or implemented during ETL activity.