**Enterprise Data Warehouse Data Reconciliation Methodology**

An enterprise [data warehouse](https://dwbi.org/data-modelling/dw-design/5-what-is-data-warehousing.html) often fetches records from several disparate systems and store them centrally in an enterprise-wide warehouse. But what is the guarantee that the quality of data will not degrade in the process of centralization?

**Data Reconciliation**

Many of the data warehouses are built on n-tier architecture with multiple data extraction and data insertion jobs between two consecutive tiers. As it happens, the nature of the data changes as it passes from one tier to the next tier. Data reconciliation is the method of reconciling or tie-up the data between any two consecutive tiers (layers).

**Why Reconciliation is required?**

In the process of extracting data from one source and then transforming the data and loading it to the next layer, the whole nature of the data can change considerably. It might also happen that some information is lost while transforming the data. A reconciliation process helps to identify such loss of information.

One of the major reasons of information loss is loading failures or errors during loading. Such errors can occur due to several reasons e.g.

* Inconsistent or non coherent data from source
* Non-integrating data among different sources
* Unclean / non-profiled data
* Un-handled exceptions
* Constraint violations
* Logical issues / Inherent flaws in program
* Technical failures like loss of connectivity, loss over network, space issue etc.

Failure due to any such issue can result into potential information loss leading to unreliable data quality for business process decision making.

Further more, if such issues are not rectified at the earliest, this becomes even more costly to “patch” later. Therefore this is highly suggested that a proper data reconciliation process must be in place in any data Extraction-Transformation-Load (ETL) process.

**Scope of Data Reconciliation**

Data reconciliation is often confused with the process of data quality testing. Even worse, sometimes data reconciliation process is used to investigate and pin point the data issues.

While data reconciliation may be a part of data quality assurance, these two things are not necessarily same.

Scope of data reconciliation should be limited to identify, if at all, there is any issue in the data or not. The scope should not be extended to automate the process of data investigation and pin pointing the issues.

A successful reconciliation process should only indicate whether or not the data is correct. It will not indicate why the data is not correct. Reconciliation process answers “what” part of the question, not “why” part of the question.

**Methods of Data Reconciliation**

**Master Data Reconciliation**

Master data reconciliation is the method of reconciling only the master data between source and target. Master data are generally unchanging or slowly changing in nature and no aggregation operation is done on the dataset. That is - the granularity of the data remains same in both source and target. That is why master data reconciliation is often relatively easy and quicker to implement.

In one business process, “customer”, “products”, “employee” etc. are some good example of master data. Ensuring the total number of customer in the source systems match exactly with the total number of customers in the target system is an example of customer master data reconciliation.

Some of the common examples of master data reconciliation can be the following measures,

1. Total count of rows, example
   * Total Customer in source and target
   * Total number of Products in source and target etc.
2. Total count of rows based on a condition, example
   * Total number of active customers
   * Total number of inactive customers etc.

**Transactional Data Reconciliation**

Sales quantity, revenue, tax amount, service usage etc. are examples of transactional data. Transactional data make the very base of BI reports so any mismatch in transactional data can cause direct impact on the reliability of the report and the whole BI system in general. That is why reconciliation mechanism must be in-place in order to detect such a discrepancy before hand (meaning, before the data reach to the final business users)

Transactional data reconciliation is always done in terms of total sum. This prevents any mismatch otherwise caused due to varying granularity of qualifying dimensions. Also this total sum can be done on either full data or only on incremental data set.

Some examples measures used for transactional data reconciliation can be

1. Sum of total revenue calculated from source and target
2. Sum of total product sold calculated from source and target etc.

**Automated Data Reconciliation**

For large warehouse systems, it is often convenient to automate the data reconciliation process by making this an integral part of data loading. This can be done by maintaining separate loading metadata tables and populating those tables with reconciliation queries. The existing reporting architecture of the warehouse can be then used to generate and publish reconciliation reports at the end of the loading. Such automated reconciliation will keep all the stake holders informed about the trustworthiness of the reports.