What is loading?

Loading is the ultimate step in the ETL process. In this step, the extracted data and the transformed data are loaded into the target database. To make the data load efficient, it is necessary to index the database and disable the constraints before loading the data. All three steps in the ETL process can be run parallel. Data extraction takes time and therefore the second phase of the transformation process is executed simultaneously. This prepared the data for the third stage of loading. As soon as some data is ready, it is loaded without waiting for the previous steps to be completed.

The loading process is the physical movement of the data from the computer systems storing the source database(s) to that which will store the data warehouse database. The entire process of transferring data to a data warehouse repository is referred to in the following ways:

1.Initial Load: For the very first time loading all the data warehouse tables.

2.Incremental Load: Periodically applying ongoing changes as per the requirement. After the data is loaded into the data warehouse database, verify the referential integrity between the dimensions and the fact tables to ensure that all records belong to the appropriate records in the other tables. The DBA must verify that each record in the fact table is related to one record in each dimension table that will be used in combination with that fact table.

3.Full Refresh: Deleting the contents of a table and reloading it with fresh data.

Refresh versus Update
After the initial load, the data warehouse needs to be maintained and updated and this can be done by the following two methods:

•Update-application of incremental changes in the data sources.

•Refresh-complete reloads at specified intervals.

Data Loading-
Data is physically moved to the data warehouse. The loading takes place within a "load window. The tendency is close to real-time updates for data warehouses as warehouses are growing used for operational applications.

Loading the Dimension Tables
Procedure for maintaining the dimension tables includes two functions, initial loading of the tables and thereafter applying the changes on an ongoing basis System geared keys are used in a data warehouse. The reeds in the source system have their own keys. Therefore, before an initial load or an ongoing load, the production keys must be

co to system-generated keys in the data warehouse, Another issue is related to the application of *Type 1*, Type 2, and *Type 3* changes to the data warehouse. *Fig*. shows how to handle it.

Loading changes to a dimension table
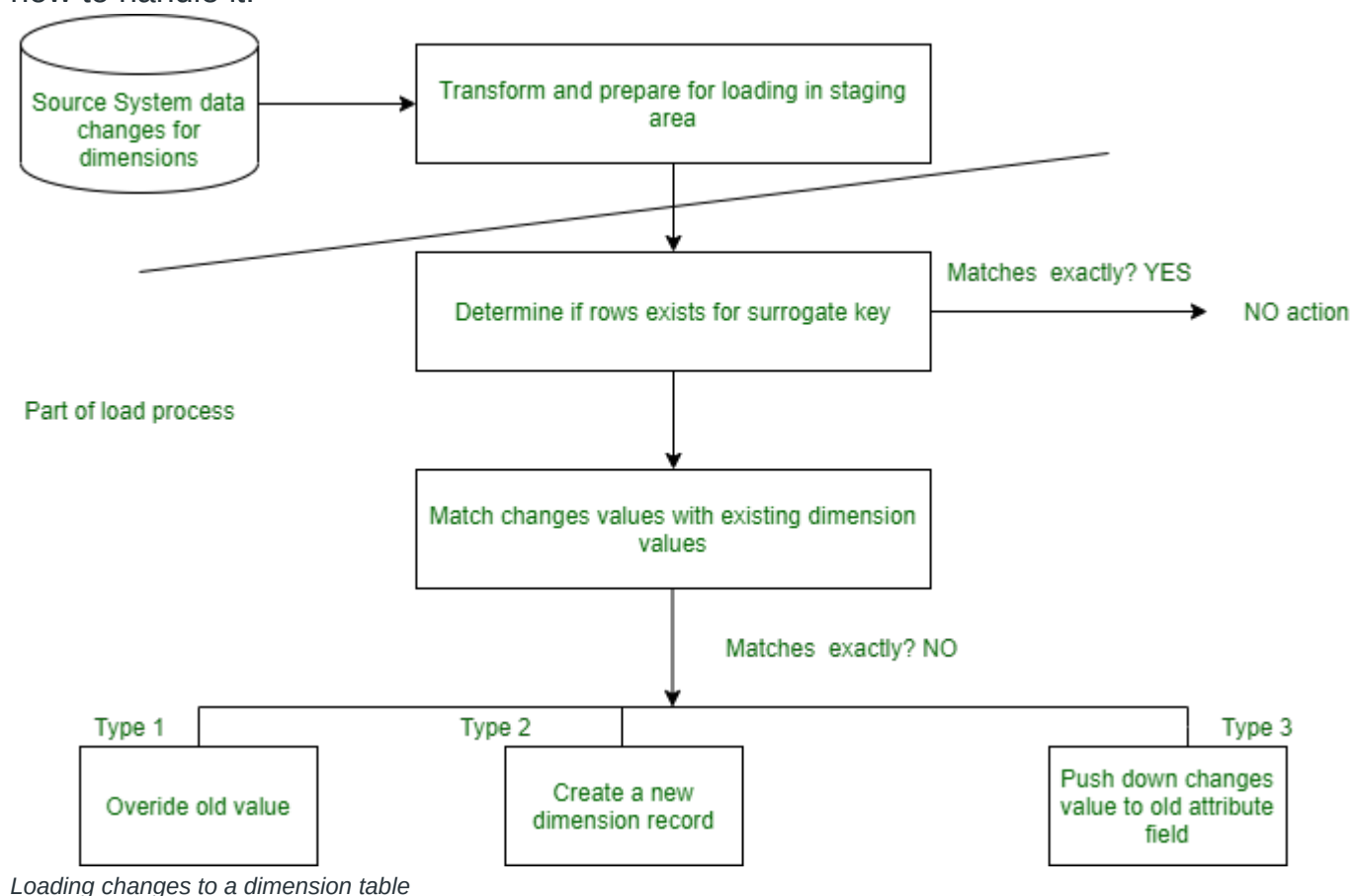
Loading the Fact tables: History and Incremental Loads

• The key in the fact table is the concatenation of keys from the dimension tables.

• So for this reason amplitude records are loaded first.

• A concatenated key is created from the keys of the corresponding dimension tables.

Methods for data loading

• Cloud-based

• Batch processing

• Open-source

ETL Tools

In the present-day market, ETL equipment is of great value, and it is very important to recognize the classified method of extraction, transformation, and loading method.

- Skyvia
- IRI Voracity
- Xtract.io
- Sprinkle
- DBConvert Studio By SLOTIX s.r.o.
- Informatica – PowerCenter
- IBM – Infosphere Information Server
- Oracle Data Integrator
- Microsoft – SQL Server Integrated Services (SSIS)
- Ab Initio

Data loading challenges

Numerous ETL solutions are cloud-based, which is responsible for their speed and scalability. But large enterprises with traditional, on-premises infrastructure and data management processes often use custom-built scripts to collect and load their data into storage systems through customized configurations.

Slow down analysis: Every time a data source is added or changed, the system has to be reconfigured, which is time-consuming and hinders the ability to make quick decisions..

Increase the likelihood of errors: Changes and reconfigurations open the door to human error, duplicate or missing data, and other problems.

Require specialized knowledge: In-house IT teams often lack the necessary skills (and bandwidth) to code and monitor ETL tasks.

Require costly equipment: In addition to investing in the right human resources, organizations have to procure, house and maintain the hardware and other equipment to drive the process on site.