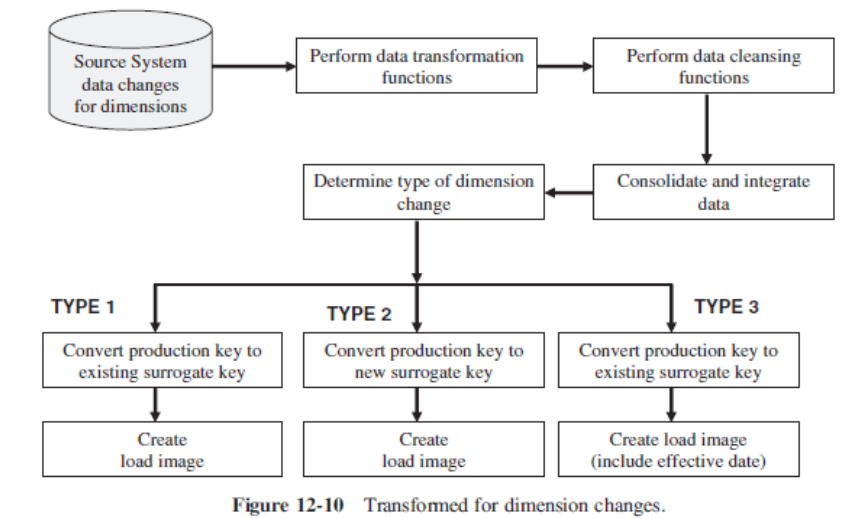
**Data Transformation** involves several basic tasks to prepare raw data for storage and analysis in a data warehouse:

1. *Selection:* Initial step where whole records or parts of records are chosen from source systems. May occur during extraction or as part of transformation if source structure is complex.
2. *Splitting / Joining:* Manipulates selected data parts, including splitting or joining them. Splitting is rare while joining multiple parts from different sources is common in DW.
3. *Conversion:* Crucial for standardizing data across disparate systems & making it user friendly. Involves converting fields to a common format and ensuring their usability.
4. *Summarization:* Aggregates data to a higher level of detail, suitable for analysis. Useful when storing granular transaction data isn't necessary for analysis.
5. *Enrichment:* Rearranges and simplifies fields to enhance their usefulness in the DW. Combines fields from single or multiple records to create a clearer view for analysis.



**Major Transformation Types in Data Transformation:**

1. Format Adjustments: Making changes to data formats for consistency.
2. Decoding Data: Converting encoded data into readable format.
3. Calculations: Creating new values based on existing ones.
4. Splitting Data: Breaking one field into multiple parts.
5. Merging Information: Combining data from different sources.
6. Character Conversion: Changing character encoding schemes.
7. Unit Conversion: Converting data from one measurement unit to another.
8. Date/Time Adjustments: Standardizing date and time formats.
9. Summarizing Data: Condensing detailed information into summaries.
10. Key Restructuring: Modifying keys to meet database requirements.
11. Deduplication: Removing duplicate entries for accuracy.

**Data loading** is transferring data from source systems into a DW or DB. It involves extracting data from various sources, transforming it into a suitable format, and then loading it into the target data storage system. This process ensures that the data is structured, cleansed, and organized for analysis, reporting, or other business purposes. Methods:

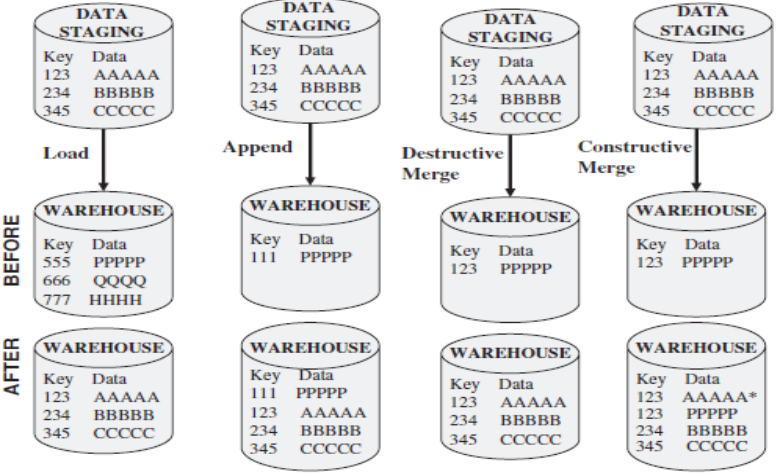
1. *Initial Load:* Populating all data warehouse tables for the first time.
2. *Incremental Load:* Applying ongoing changes periodically.
3. *Full Refresh:* Erasing table contents and reloading with fresh data.

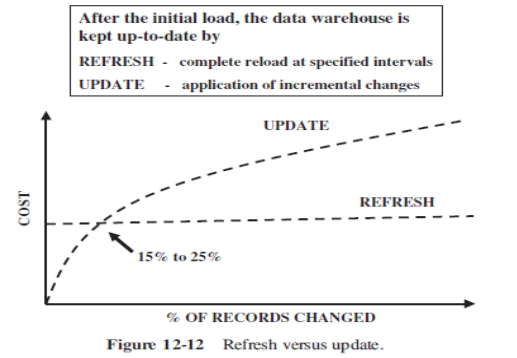
**Data Loading Considerations:**

1. *Time Concerns:* Loading can take a long time, so scheduling is crucial.
2. *Parallel Loading:* Breaking the process into smaller chunks for faster execution.
3. *Continuous Operation:* Parts of the DW can remain active while others are loaded.
4. *Load Challenges:* Addressing issues like unsuccessful record applications & data transport.
5. *Load Utilities:* Using database-provided tools or custom programs for loading.

**Modes of Applying Data:**

1. *Load:* Overwrites existing data with incoming data, or simply adds if table is empty.
2. *Append:* Adds incoming data to existing data, preserving the original records.
3. *Destructive Merge:* Updates existing records with incoming data & adds new records.
4. *Constructive Merge:* Adds new records while preserving existing ones, based on primary key matches.

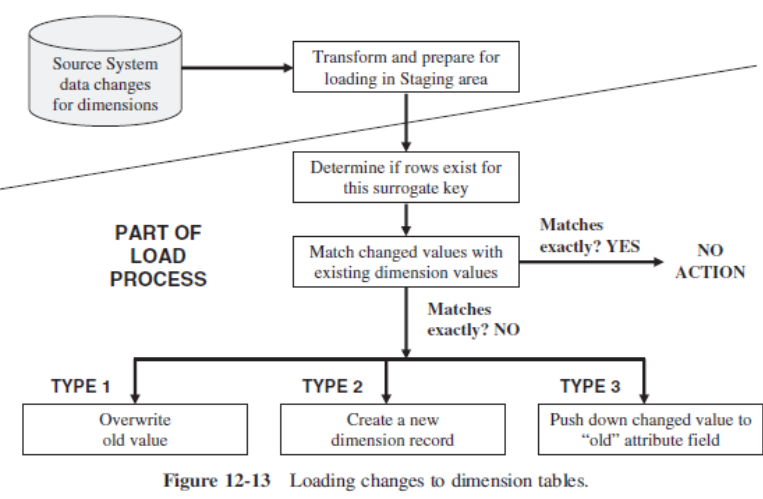




**Data refresh and update** are used to maintain the data in a data warehouse after the initial load:

*Update:* Involves applying incremental changes to the DW based on updates received from the source systems. These changes can include additions, modifications, or deletions to existing data.

*Refresh:* Involves completely reloading the DW with fresh data from source systems at specified intervals. This process replaces existing data in data warehouse with latest data available from sources.



**Loading dimension tables** involves 2 main tasks: initial loading & ongoing updates. The initial loading process brings data from source systems into the DW, while ongoing updates apply changes over time.

Source data often uses different keys than those used in the DW. Converting source keys to system-generated keys in the warehouse is essential before loading data. This key conversion can be integrated into the transformation process or handled separately.

Another issue in loading dimension tables involves managing changes such as type 1, type 2, and type 3 dimension changes. These changes need to be appropriately applied to the DW to ensure data integrity and accuracy.

**When loading fact tables**, the key is formed by concatenating keys from dimension tables. Therefore, dimension records must be loaded before fact table loading. Fact table loads can be historical or incremental. For historical loads, it's crucial to identify and extract relevant historical data, refine business rules, capture audit statistics, perform surrogate key look-ups, and restructure data as needed before preparing load files. Incremental loads involve extracting new or updated transactions since the last load. Using database transaction logs for data capture, loading as frequently as feasible, and applying parallel processing techniques can optimize incremental fact table loads.

**Data quality** refers to the accuracy, completeness, consistency, and reliability of data. Ensuring high data quality is crucial for making informed decisions and deriving valuable insights. However, various issues can arise during data cleansing, which is the process of identifying and rectifying these problems. Some common issues in data cleansing include:

1. *Missing Values:* Data may have missing values due to errors or incomplete entries.
2. *Inaccurate Data:* Errors during data entry or outdated info can lead to inaccuracies.
3. *Duplicate Records:* Identifying and removing duplicate entries ensures data integrity.
4. *Inconsistent Formats:* Standardizing data formats ensures uniformity.
5. *Spelling Errors:* Correcting typos and syntax errors improves data quality.
6. *Irrelevant Data:* Removing outdated or irrelevant data maintains relevancy.
7. *Data Integrity Issues:* Ensuring data relationships are consistent prevents contradictions.