ETL

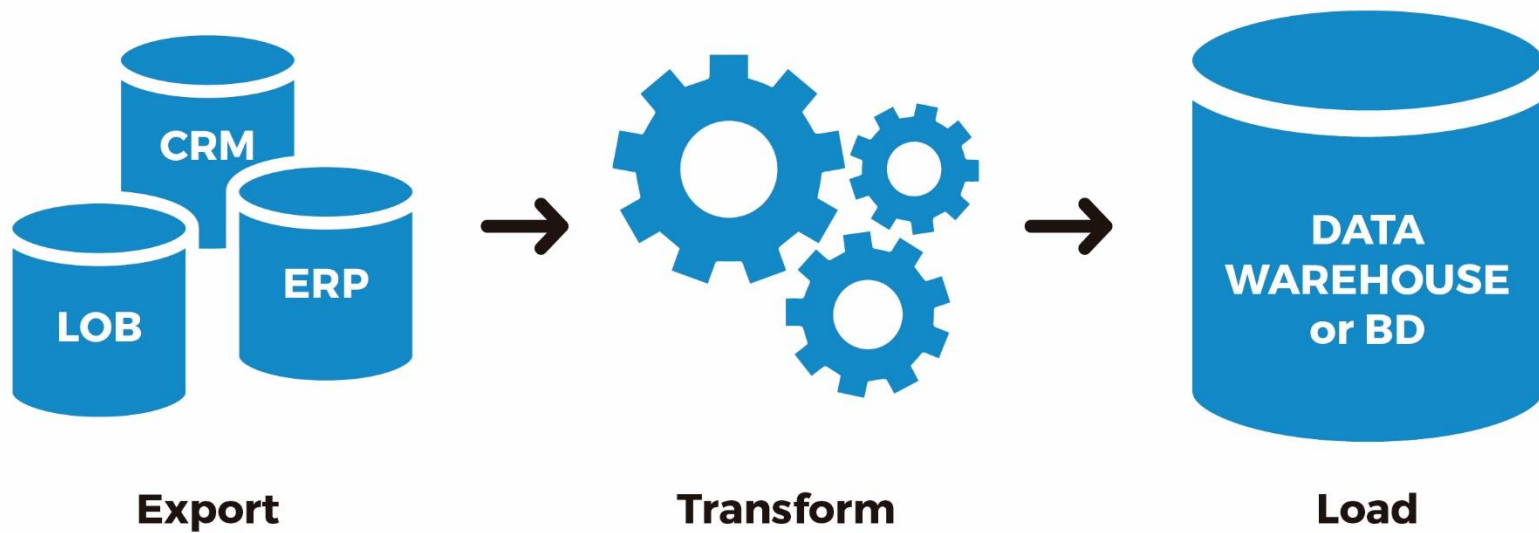**Export**   **Transform**   **Load**

CRM
LOB
ERP

DATA
WAREHOUSE
or BD

ETL

Why ETL?

- Companies need a way to analyze their data for critical business decisions

- Transactional Database can't answer complex business questions

- A data warehouse provides a common data repository

- ETL provide a method of moving the data from various source into a data warehouse.
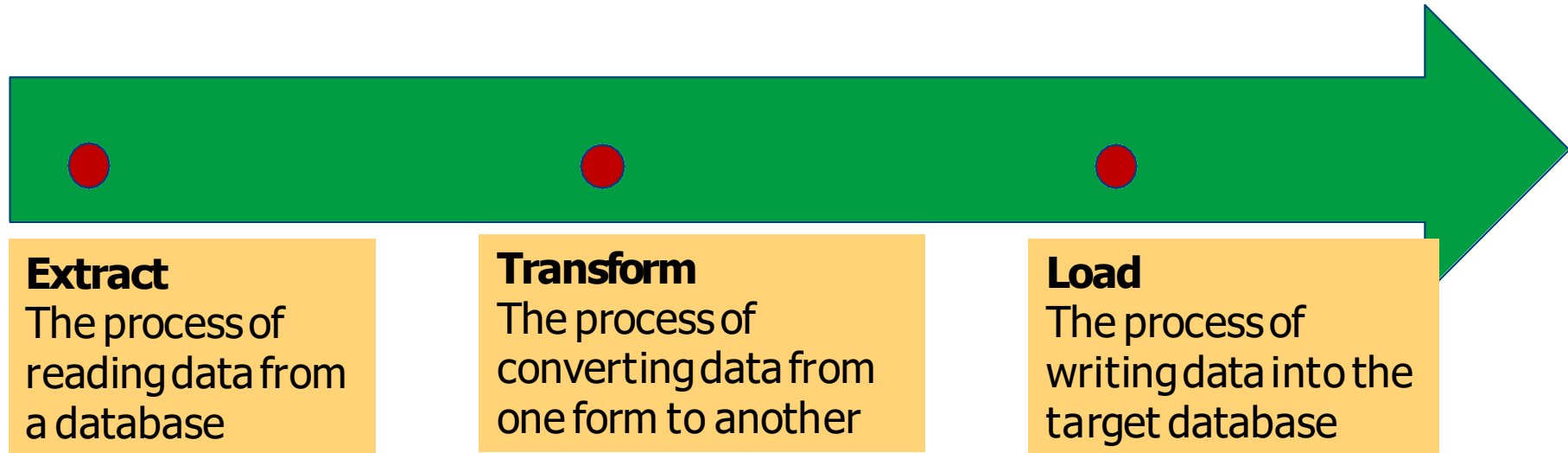
ETL Concept

A Company data may be scattered in different locations and in different formats.

ETL allows you to:

- Migrate the data into a data warehouse.

- Converts the various formats and types to adhere to one consistent system.

ETL is a predefined process for access and manipulate source data and loading it into a

target database.

The ETL Process

Extract
The process of reading data from a database

Transform
The process of converting data from one form to another

Load
The process of writing data into the target database

Extract

Gathering the data

- Raw data that was written directly into the disk

- Data written to flat files or relational tables from structured source systems.

- Data can be read multiple times, if needed

Cleansing the data

- Eliminate duplicates or fragmented data

- Exclude unwanted / unneeded information
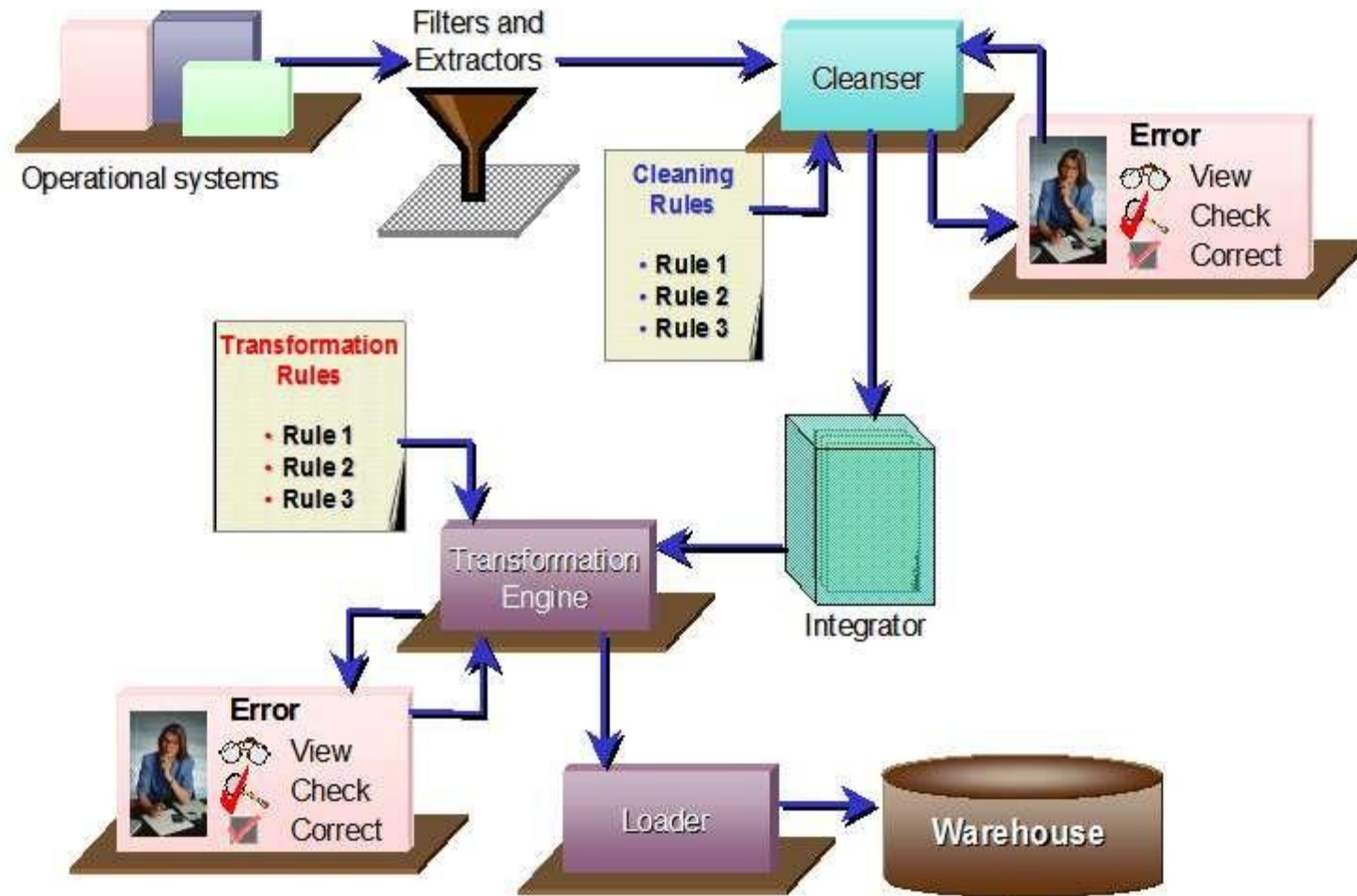
Transform

- Preparing the data to be housed in the data warehouse

- Converting the extracted data

    - Using rules and lookup tables

    - Combining data

    - Verification/Validity checks

    - Standardization

Load

Storing the transformed data in the data warehouse.

Batch/Real-time processing

Can follow star schema and snowflake schema.

ETL – Extract Transform Load

Critical Components

Handling Multiple Source systems – Should have enabled to extract from multiple sources

Captures Changed records (CDC) – Picks up only changed records to enable incremental loads

Auditing & Logging – Need a detailed logging of different process so that loaded data could be audited and errors are debugged.

Fault Tolerance – System needs to gracefully recover from failures during loads.

Notifications- The system needs to send out notifications for failure as well as completion

Scalability – Should be scalable to handle increase in data volume

Accuracy - Data cannot be dropped or changed in a way that corrupts its meaning. Every data point should be auditable at every stage in your process.

Advantage of ETL Tool

- Simple, faster and cheaper development

- Most ETL tools provide a metadata repository, synchronizing metadata from various sources.

- Most ETL tools deliver good performance even for very large dataset

- Most ETL tools provide impact analysis tools for any proposed schema changes.

- Most ETL tools have built in connectors for all the major RDBMS systems.

- Most ETL tools allow reuse of the existing complex programs.

- Several ETL tools offers visual Development Environment.

- Most ETL tools offers built in Scheduler sequencers and documentation.

- Several ETL tools offer various performance optimization options such as parallel processing, complex load balancing etc.

Thank you

Innovative Services

Passionate Employees

Delighted Customers