ASW-III

ETL Process

William Huang

What is ETL Process

ETL (Extract, Transform and Load) is a process in data warehousing responsible for pulling data out of the source systems and placing it into a data warehouse. ETL involves the following tasks:

- extracting the data from source systems data from different source systems is converted into one consolidated data warehouse format which is ready for transformation processing.

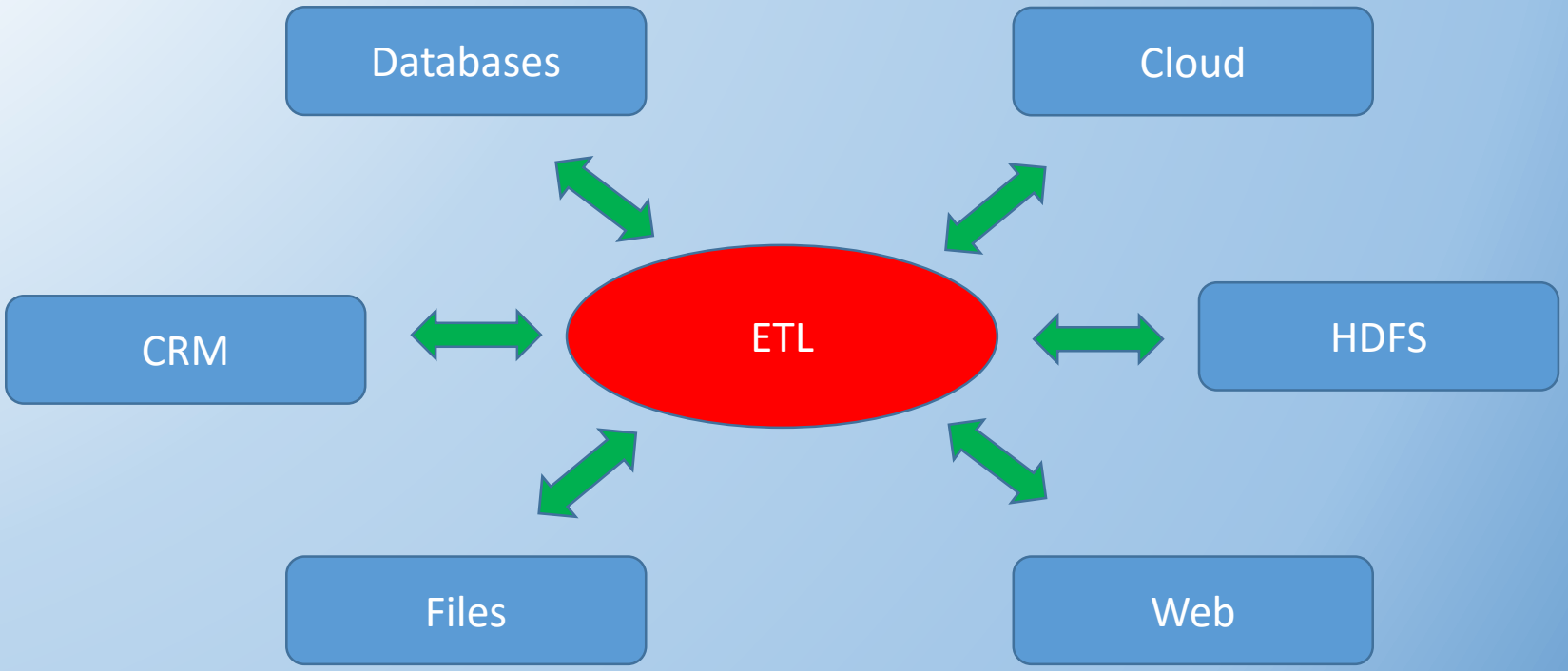
- transforming the data may involve the following tasks:

  - applying business rules (so-called derivations, e.g., calculating new measures and dimensions),
  - cleaning (e.g., mapping NULL to 0 or "Male" to "M" and "Female" to "F" etc.),
  - filtering (e.g., selecting only certain columns to load),
  - splitting a column into multiple columns and vice versa,
  - joining together data from multiple sources (e.g., lookup, merge),
  - transposing rows and columns,
  - applying any kind of simple or complex data validation (e.g., if the first 3 columns in a row are empty then reject the row from processing)

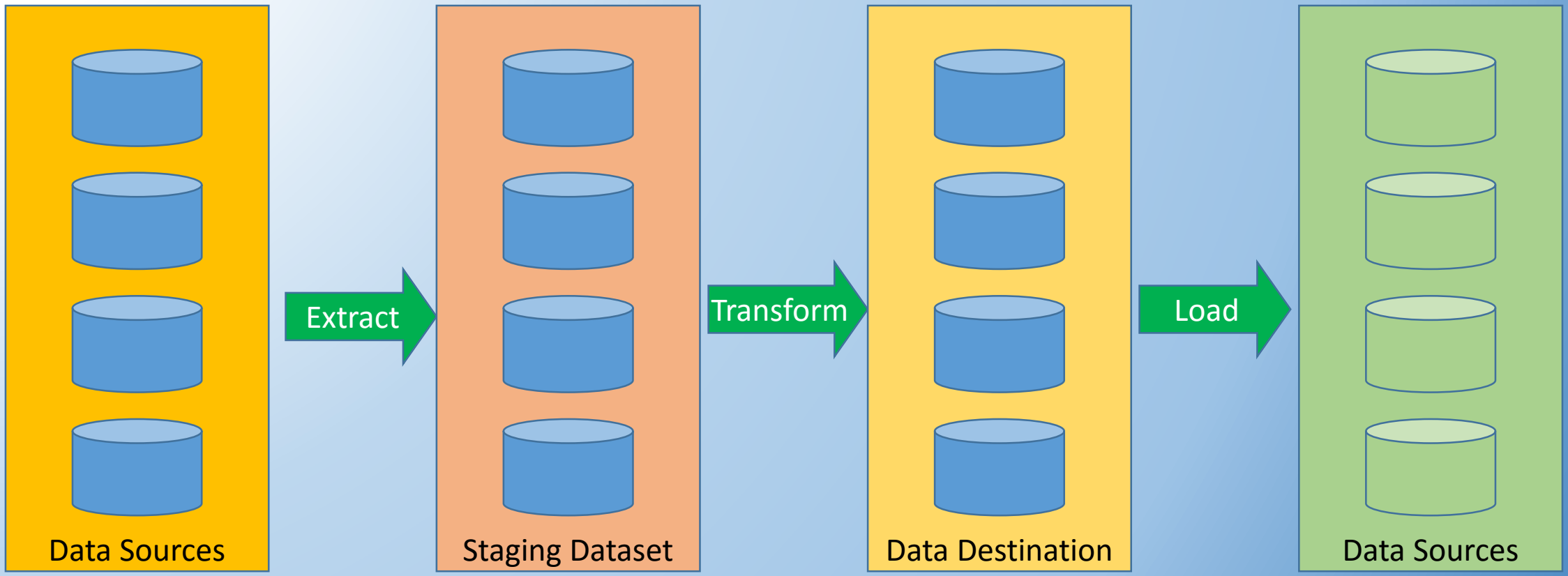
- loading the data into a data warehouse or data repository other reporting applications

ETL Process

Cross Platform Data Process



ETL Process Stages



Most Popular ETL Tools in 2019

- Informatica – PowerCenter
- IBM Infosphere Information Server
- Oracle Data Integrator
- Microsoft SQL Server Integrated Services (SSIS)
- Ab Initio
- Talend Open Studio for Data Integration
- CloverDX Data Integration Software
- Pentaho Data Integration
- Apache Nifi
- Jasper
- SAS Data Integration Studio
- SAP BusinessObjects Data Integrator
- Oracle Warehouse Builder
- Sybase ETL
- DBSoftlab

ETL Tools used in SAW-III

Data: MS Sql Server, HIVE, Text file, Web data, …

ETL tools: SSIS, SQL, Linux script, Apache Nifi, …

Work Place: Local computer, Standfort Lab, MS Azure.

ETL Development Steps

1. Understand the business requirements and the project objectives.
2. Understand data model
3. Design data mapping
4. Implement data mapping
5. Integrate Test for data mapping.

Q & A