

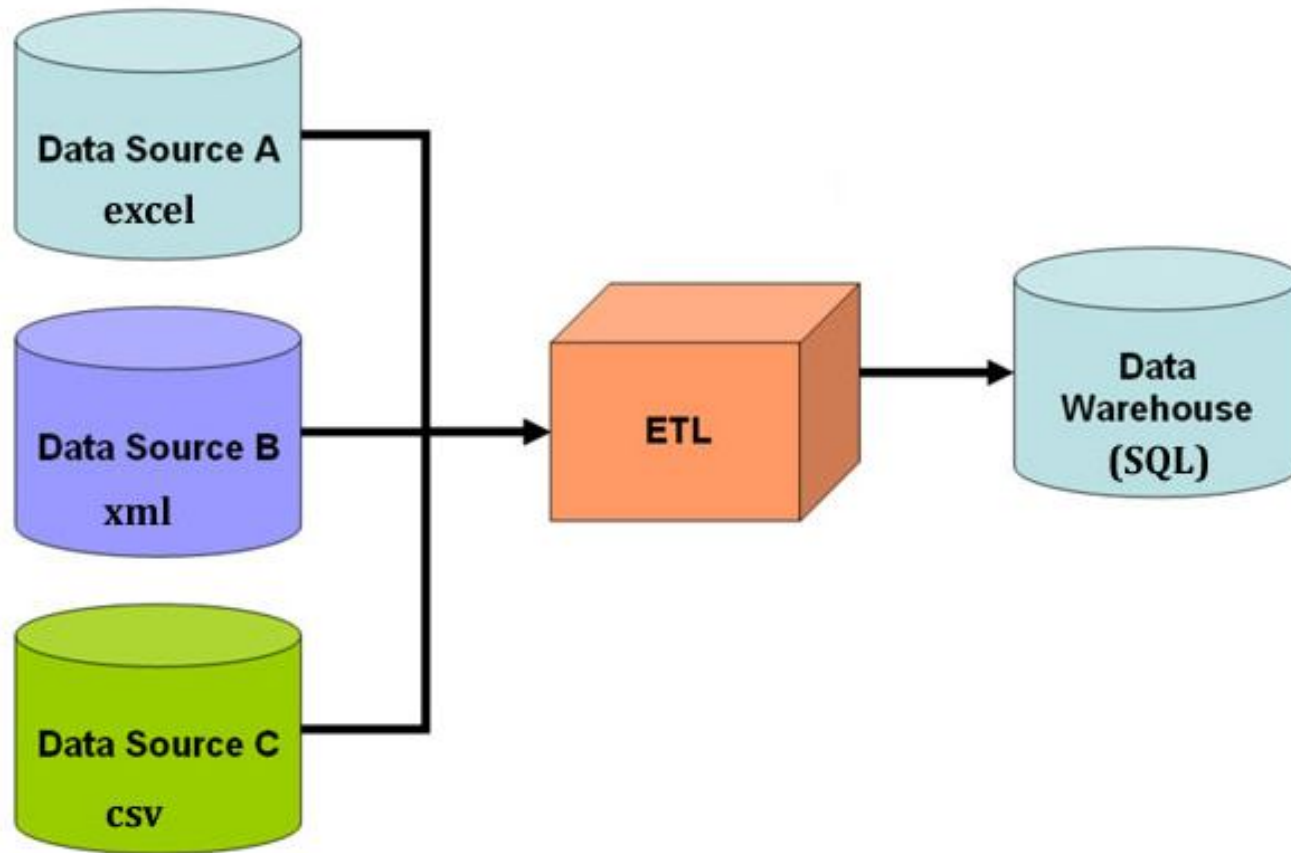


Data Extraction, Cleanup and Transformation Tools

V.. Vaishnavee

AP-AI-DS

SNSCE





Tool Requirements

- **Data transformation** from one format to another based on differences between source and target platforms
- **Data transformation & calculation** based on application of business rules that force certain transformations
- **Data consolidation & integration** – combining several source records into single record to be loaded in DW
- **Metadata synchronization & management** – included storing, updating metadata definitions about source data files



Criteria that affects tools ability to transform, consolidate, integrate and repair data

- Ability to identify data in data source environment
- Support for flat files, indexed files
- Capability to merge data from multiple data stores
- Ability to read information from data dictionaries
- Capability to create summarization, aggregation and derivation records and fields



Vendor Approaches

- Task of capturing data from source data system, cleaning and transforming it & then loading into target data system carried out either by separate products or by single integrated solution.
 - Code generators
 - Database data replication
 - Rule-driven dynamic transformation engines or data mart builders



Access to legacy data

- To manage the interaction between the new applications & growing DW -> organizations use middleware solutions
- Apertus Corporation -> developed Enterprise/Access

Provides 3 tier architecture -> defines how applications are partitioned to meet integration & migration objectives

- Designed for scalability and manageability in DW
- Development tool -> enables rapid development of transparent, production quality client/server interfaces to legacy applications



Access to legacy data – 3 tier architecture

Data Layer -> provides data access & transformation services for management of corporate data assets

Manages data & enforces the business rules for data integrity

Process Layer -> provides services to **manage automation** & support for center business processes

User Layer -> manages user interaction with process and data layer services



Two virtual database models for Enterprise/Access are applicable to the development of DW

1. Enterprise/Access acts as a virtual DB
 - Host interfaces & business transactions are defined entirely in Enterprise/Access
 - This model is used when **building a generalized architecture** & modernizing applications
2. Enterprise/Access & open Gateway use SQL server as a virtual DB
 - Host interfaces are located in Enterprise/Access while business logic is implemented via stored procedures
 - This model used when **migrating legacy system** functionality to relational DB like SYBASE



Vendor Solutions

- Prism Solutions
- SAS Institute
- Carleton corporation's PASSPORT & Metacenter
- Vality corporation
- Evolutionary Technologies – ETI – Extract Tool Suite
- EDA (External Databases Access)/SQL from Information Builders





Carleton corporation's PASSPORT & Metacenter

- PASSPORT & Metacenter provides solution for migrating data to new DW
- PASSPORT:
 - Consists of 2 components

Mainframe based

Workstation based

Collects file, record, table layouts for required inputs & outputs & converts them to PDL

Used to create meta data directory from which it builds COBOL programs to actually create the extracts