

# Real-Time Data Lake / Modern Data Warehouse

---

March 14, 2022

# Data Warehouse Purpose

## Original Goals of Data Warehousing

- Transform/refine data into information
- Integrate/harmonize data from multiple sources into a single enterprise view
- Deliver fast query performance
- Support end-user ease of use
- Provide seamless integration with dashboard and BI/reporting tools
- Maintain history

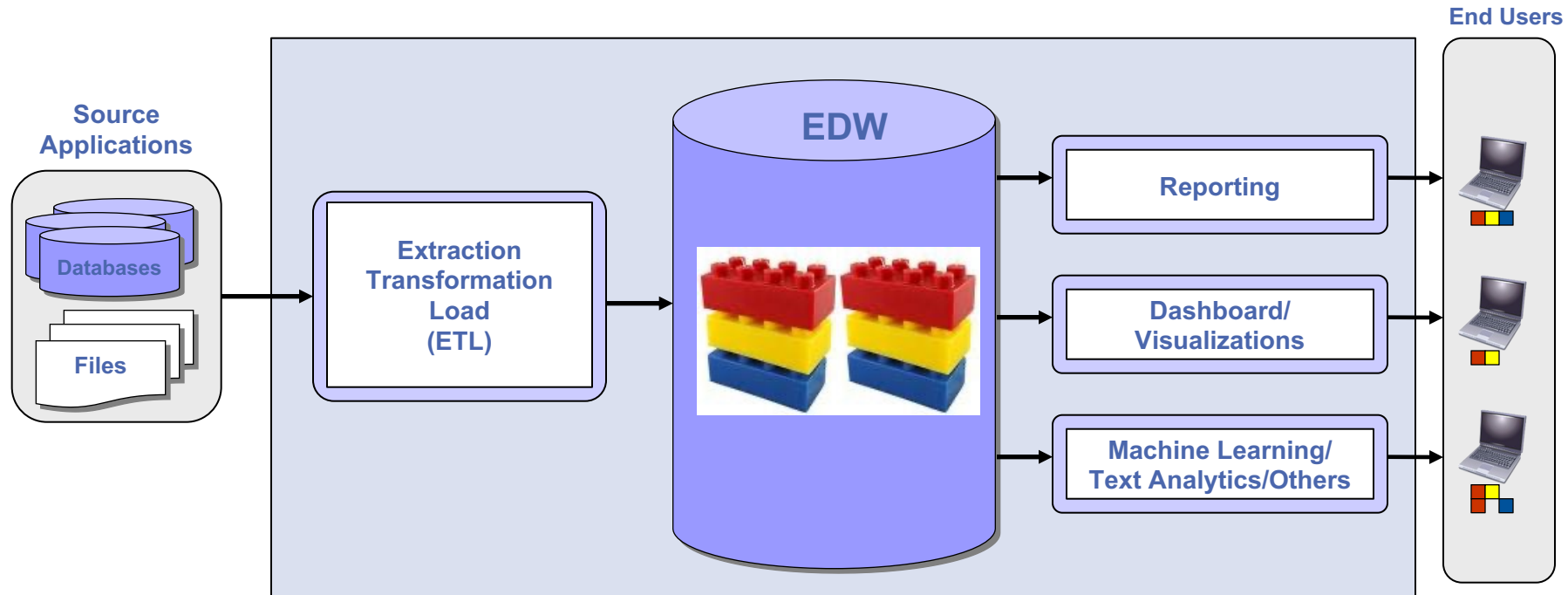
## Limitations

- Did not support real-time data loads
- The data engineering/ingestion (ETL) process was brittle and very complex. Changes frequently required full data reprocessing and reloading
- Maintaining history was complex

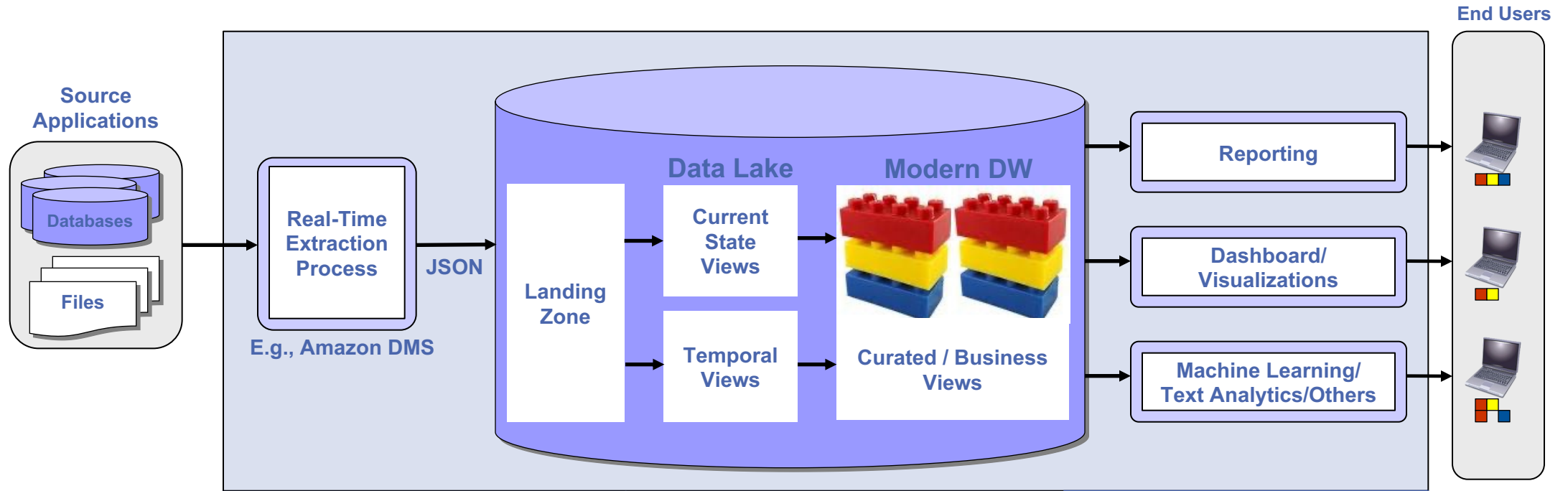
# Data Warehouse Purpose

| Characteristics   | Traditional Data Warehouse | Modern Data Warehouse |
|---|----------------------------|-----------------------|
| Data is transformed/refined into information  | ●                          | ●                     |
| Data from multiple sources is integrated/harmonized into a single enterprise view                                     | ●                          | ●                     |
| Fast query performance  | ●                          | ◐                     |
| End-user ease of use  | ●                          | ●                     |
| Strong dashboard and BI/reporting tool integration  | ●                          | ●                     |
| Close to real-time updates  | ○                          | ●                     |
| Dynamic, “forgiving” data engineering/ingestion process   | ○                          | ●                     |
| Easy to make data engineering changes (e.g., business rules, source changes, etc.).<br>No data reprocessing/reloading | ○                          | ●                     |
| Easy to maintain all of history   | ○                          | ●                     |
| Strongly supports dimensional conformance (master data management)  | ●                          | ◐                     |

# Traditional Data Warehouse (ETL)



# Modern Data Warehouse (ELT)



## Landing Zone

- One-column, vertically-oriented table
- Stores JSON messages
- JSON message contains:
  - Meta data (table name, column name, data types)
  - Data values
  - Update function (insert, update, delete)
- Query performance is surprisingly good given vertical data orientation

## Real-Time Data Lake

### *Current State Views / Temporal Views*

- The views horizontally pivot the vertical JSON messages into mirror images of source system table structures
- Current State Views – Latest data structures and data values
- Temporal Views – Historic data structures and data values
- Forgiving: Automatically stays in sync with source data structure and value changes
- Standardized scripts are run to establish the data lake in 1 or 2 days.

## Modern Data Warehouse

### *Curated / Business Views*

- This is the Transformation in ELT
- Dimensional views, Fact views, and Denormalized business views are created for ease of use and persisting business rules for end users
- Strongly supports agile, evolutionary development
- Business changes do not require reprocess or reloading.

# Modern Data Warehouse Benefits

- Supports all of traditional data warehouse benefits
- Near end-to-end, real-time performance (1 to 3 minute lag from source changes)
- Forgiving – absorbs source data structure and data values changes automatically
- Data lake (current views, and temporal views) is “out of the box” in 1-2 days
- Strongly supports quick, agile, evolutionary data warehouse development
- Curated view changes do not require data reprocessing and reloading
- Maintains full history