Data Warehouse - Introduction

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Agenda

In this session, we will discuss the following topics:

- What is a Data Warehouse?
- Applications of Data Warehouse
- Evolution of Data Warehousing
- Data Warehouses vs. Data Marts
- Operational Data Stores
- Warehouse Components
- Data Warehouse Architectures
- Data Staging

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

What is a Data Warehouse?

- Data Warehouse is a technology used to manage large amounts of data in a **centralized repository**.
- They are designed for querying and analysis and are optimized for read-only access.
- They typically contain historical data that has been extracted from various sources and transformed to fit in a consistent data model.
- They provide a **single source of truth** for an organization’s data.
- They help business users to **make data-driven decisions**.
- A well-designed data warehouse helps organizations identify opportunities and challenges and guide strategic planning and resource allocation.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Applications of a Data Warehouse

- Business Intelligence: A single, integrated view of data for decision making.
- Trend Analysis: Detecting patterns and trends in historical data for forecasting.
- Customer Relationship Management (CRM): 360-degree view of customers, providing a better understanding of customer behavior and preferences.
- Risk Management: Identify and monitor risks, including operational risks, credit risks, and market risks.
- Regulatory Compliance: Monitor and report on regulatory compliance, such as financial reporting, auditing, and data privacy regulations.
- Supply Chain Management: Manage and track inventory and supplier performance, and optimize logistics operations.

This file is meant for personal use by rumpe.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Evolution of a Data Warehouse

- 1970s – Edgar Codd proposed a set of rules called the "relational model," which gave rise to relational databases making it easier to store and manage a large amount of data.
- Mid-1980s – One of the earliest examples of a data warehouse is built by IBM for a supermarket chain.
- Early 1990s – Traditional operational and transactional databases did not satisfy the requirements for data analysis, as they were designed and optimized to support daily business operations with a primary focus on concurrency, recovery, and consistency. The concept of data warehousing has become more widely recognized by businesses.
- Mid 1990s - Inmon and Kimball published their respective proposals on Data Warehouse Design. To date, a Data Warehouse is based on design suggestions by Inmon or Kimball or a hybrid of the two.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Evolution of a Data Warehouse

- Late 1990s: The emergence of data mining and online analytical processing (OLAP) allows businesses to extract insights from their data more easily.
- Early 2000s: The rise of the internet and e-commerce led to the development of web-based data warehousing solutions.
- Mid-2000s: The advent of big data and the rise of cloud computing leads to new challenges and opportunities for storing and analyzing a large amount of data.
- 2010s: The rise of machine learning and artificial intelligence leads to the development of new data warehousing tools and techniques for processing and analyzing large data sets.
- Present: Data warehousing continues to evolve. New technologies and approaches are emerging to help business store and analyze their data more effectively.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Data Warehouse vs Data Marts

Data Warehouse

- Contains **all the data** of the business organization (all the business units) in one single centralized repository
- Large-scale, enterprise-wide
- Optimised for complex queries and analysis
- In the long run, having a data warehouse can help ensure the consistency and accuracy of data

Data Marts

- Contains a **subset** of data (typically data warehouse) in separate repositories relevant to a specific business unit
- Smaller-scale
- Optimised for fast access and quick decision-making
- Data marts can exist without a data warehouse.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Operational Data Stores (ODS)

- Definition: An ODS is a type of centralized data repository that is used to support “operational” business processes to make real-time data-driven decisions for business operations.
- Examples of Operational processes are order/inventory/customer management.
- It is designed to integrate data, which might not be fully cleansed or transformed, from multiple sources such as transactional databases, messaging systems, and external data feeds.
- Organisations can have multiple ODS, and it is possible for an ODS to have two or more business operations, either via data integration or data virtualization.
- RDBMS is a commonly used technology to build ODS.
- ODS are updated frequently compared to the Data Warehouse.
- ODS is often used as a staging area for data that will be transformed and loaded into the Data Warehouse.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Warehouse Components

- Source Systems: Operational systems that collect and store transactional data.
- ETL: Process of **e**xtracting data from source systems, **t**ransforming it into a suitable format for the data warehouse, and **l**oading it into the data warehouse.
- Datawarehouse Database: Central repository that holds historical and aggregated data.
- Metadata: Provides information about the data stored inside the data warehouse (data definitions, data lineage, data quality).
- Access tools: Software tools used to access, query, analyze and report on the data inside the data warehouse.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Data Warehouse Architecture

- A data warehouse architecture refers to the overall **design** and **structure** of a data warehouse system.
- An effective data warehouse architecture provides a framework for organizing and integrating data from disparate sources into a single, consistent, and easily accessible repository for business analysis and decision making.
- It includes the underlying data models, data storage, data integration processes, and data access methods required to support the analytics and reporting needs of the organization.

This file is meant for personal use by rumpe.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Data Warehouse Architecture - Layers

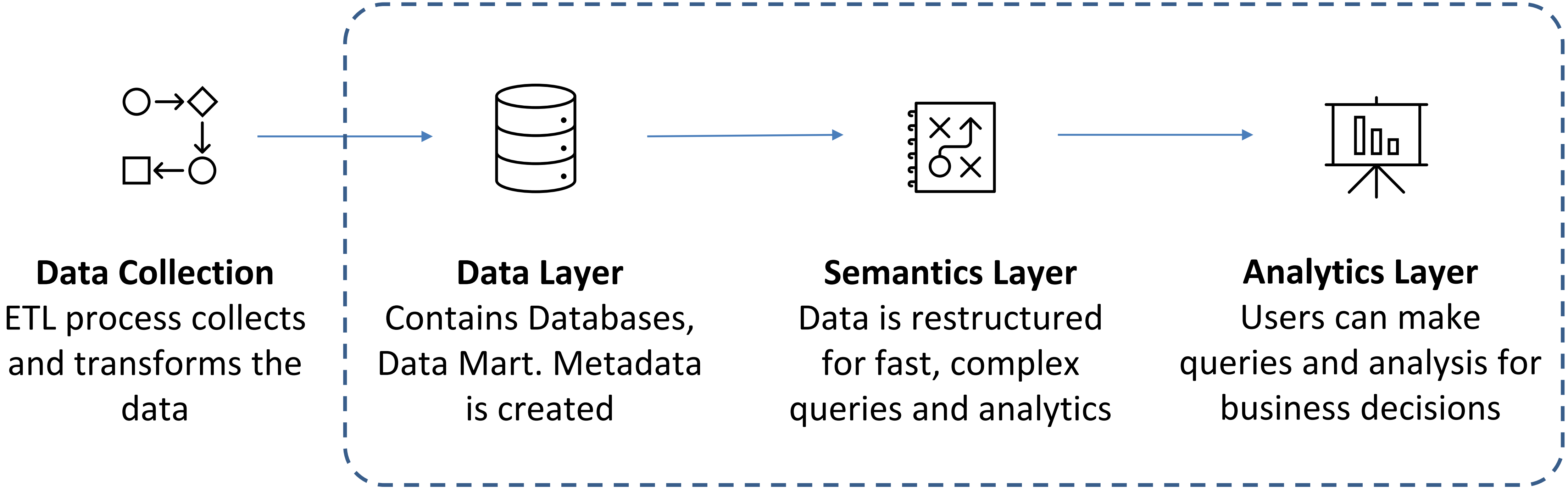
- Data Layer: Data is extracted from various sources, transformed into a suitable format, and loaded into the Data Layer.
- Semantic Layer: It is the middle tier, where online analytical processing (OLAP) and online transactional processing (OLTP) servers restructure the data for quicker execution of complex queries.
- Analytics layer: The top tier is client facing, holds the data warehouse access tools that let users interact with data, create dashboards, KPIs monitoring and reporting, data mining, and more.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Sharing or publishing the contents in part or full is liable for legal action.

Data Warehouse Architecture - Layers



Data Collection
ETL process collects and transforms the data

Data Layer
Contains Databases, Data Mart. Metadata is created

Semantics Layer
Data is restructured for fast, complex queries and analytics

Analytics Layer
Users can make queries and analysis for business decisions

Data Warehouse Architecture

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Data Staging

- Data staging involves moving data from its original source into a “staging area”.
- Here, data can be extracted, transformed, cleansed, and organized into staging (temporary) tables.
- The staging area acts as a buffer between the source systems and the data warehouse.
- This approach allows data to be properly prepared before being loaded into the data warehouse.
- The process helps maintain the data integrity and minimize errors in the data warehouse.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Summary

- Data warehousing has evolved from simple data storage to an important business intelligence tool that helps organizations make data-driven decisions.
- Data marts are small, departmental data warehouses designed for specific business units, while data warehouses are centralized repositories that store all enterprise data.
- An operational data store (ODS) is a real-time database that serves as a staging area for operational systems data before it is loaded into the data warehouse.
- The components of a data warehouse include the source system, ETL process, data warehouse database, and BI tools.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited
Sharing or publishing the contents in part or full is liable for legal action.

Summary

- The data warehouse architecture includes various layers, such as staging, integration, storage, and presentation layers, and various technologies, such as RDBMS, OLAP, and data mining tools.
- Data staging is the process of collecting and preparing data from various sources to be loaded into the data warehouse and includes tasks such as data extraction, data transformation, and data loading.

This file is meant for personal use by rumne.10731101@ltimindtree.com only.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Sharing or publishing the contents in part or full is liable for legal action.