

**SEN-209: Designing And Implementing Databases**  
**Lab 9: Exploring Modern Data Systems: OLTP, OLAP, Distributed, Cloud, Warehouse, and Data Lake**

**Name:** Thanawin Pattanaphol

**Date:** 01324096

**Part A – Conceptual Review**

<b>Terms</b>	<b>Short Description</b>
OLTP	Shortened of “Online Transaction Processing”; a system designed for managing or processing many real-time transactions.
OLAP	Shortened of “Online Analytical Processing”; a system optimized for performing complex queries or multi-dimensional analytical queries.
Distributed Database	A single database that is distributed across multiple servers, instances or locations.
Cloud Database	A database system provided by cloud provider.
Data Warehouse	A central storage system that keeps structured integrated data from multiple sources.
Data Lake	A central storage system that keeps raw data from many sources.

**Part B – Research Task: Real-World Systems**

Choose one real-world company or service. Investigate their data management architecture based on public available sources.

Chosen service: Facebook

<b>System Type</b>	<b>Example Technology Used</b>	<b>Purpose</b>	<b>Example Data Managed</b>
OLTP	MyRocks (of MySQL)	Process large amount of queries related to user interactions quickly.	Likes, comments, and shares.
OLAP	Scuba, Scribe	Process large amount of logs and turning them into analytical reports.	Logs of events
Distributed Database	MyRocks / MySQL	Storing various types of data; user interactions, profiles, posts, and so on.	User profiles, images, videos, text messages, etc.
Data Warehouse	Hive	Creating reports, analytics, or	Users, ad clicks, sales reports.

**SEN-209: Designing And Implementing Databases**  
**Lab 9: Exploring Modern Data Systems: OLTP, OLAP, Distributed, Cloud, Warehouse, and Data Lake**

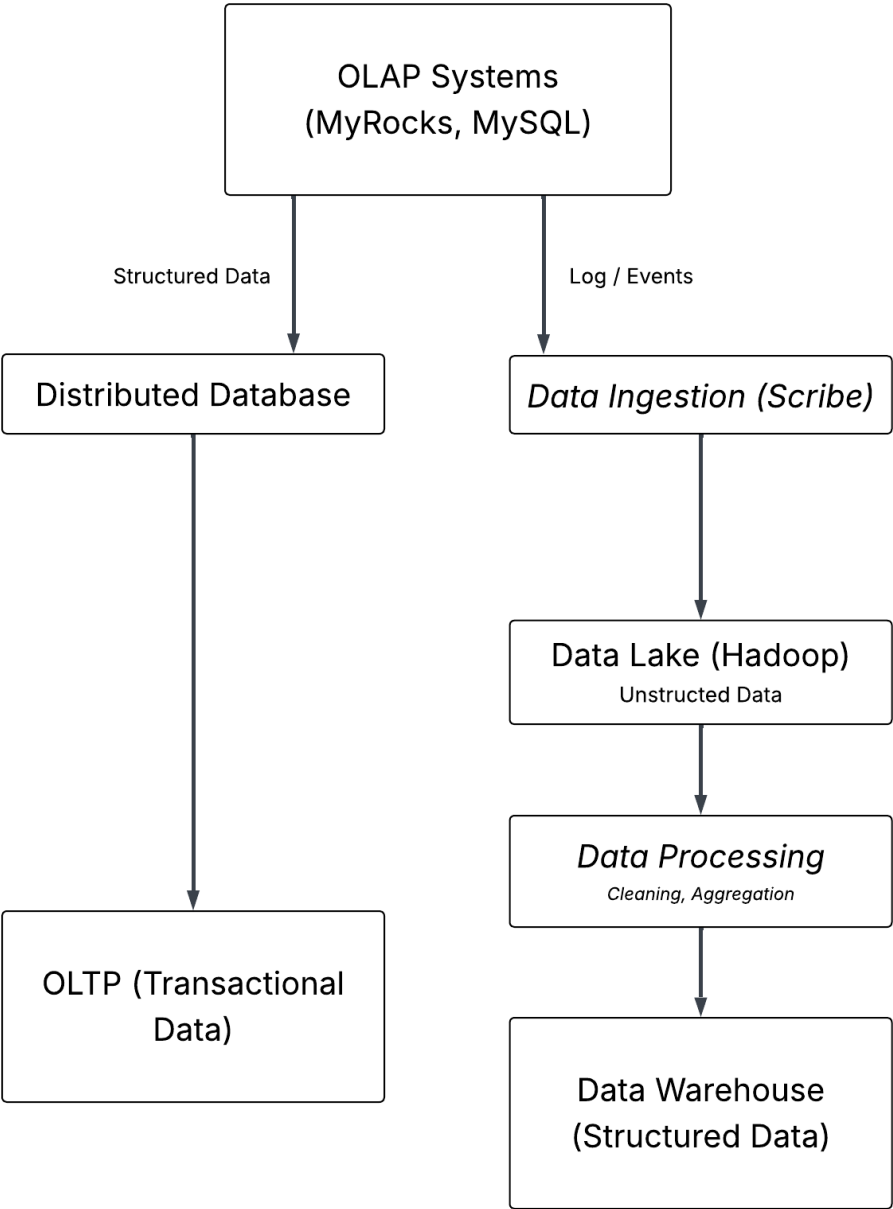
**Name: Thanawin Pattanaphol**

**Date: 01324096**

		dashboards.	
Data Lake	Hadoop	Storage of raw and curated data along with machine learning and data exploration.	Raw logs, JSON events, images, videos, posts.

**Part C – Concept Integration Task**

Chosen organization: Facebook



**SEN-209: Designing And Implementing Databases**  
**Lab 9: Exploring Modern Data Systems: OLTP, OLAP, Distributed, Cloud, Warehouse, and Data Lake**

**Name:** Thanawin Pattanaphol

**Date:** 01324096

**Part D – Case-Based Application**

Chosen business scenario: Huge book store

1. What data would be stored in the OLTP system?

The data that would be stored in this supposed system are such as, book purchases, customer requests, reviews, etc.

2. What analytical questions would be answered by the OLAP or Data Warehouse?

In this context, questions such as, “What is the most popular book during June 2025?”, “What is the most popular genre”, “How are the overall reviews of this specific book?”, “Overall customer satisfaction?”

These questions are ones that are related to the statistics / analytics of data and they can be answered using OLAP systems or Data Warehouses.

3. What kind of data (raw logs, media files, etc.) would go into a Data Lake?

The raw data lake would be storing files such as, book cover images, profile pictures, catalogs of books, PDFs of e-Books, raw JSON book description or format files, and so on.

4. Which parts of your design could benefit from cloud or distributed databases?

Cloud or distributed databases would greatly benefit the Data Lake and the Data Warehouse, since these will most likely be built on top existing relational database solutions along with how said databases will be needed to scale as the system enlarges.

**Part E – Reflection**

1. What are the main differences between OLTP and OLAP systems in terms of design and performance?

OLTP and OLAP systems differentiate on their purposes, in which, OLTP systems are more focused towards quicker, shorter and simpler processes and are more appropriate for systems that require real-time, high volume transactional operations e.g. banking systems, online shopping services. On the other hand, OLAP systems are focused towards complex, historical, and analytical processes, and is therefore more appropriate towards queries that require complex data analysis, data aggregations and other queries that require a significant amount of historical data.

**SEN-209: Designing And Implementing Databases**  
**Lab 9: Exploring Modern Data Systems: OLTP, OLAP, Distributed, Cloud, Warehouse, and Data Lake**

**Name: Thanawin Pattanaphol**

**Date: 01324096**

2. Why are Data Warehouses still used when Data Lakes exist?

It might seem that Data Warehouses and Data Lakes are similar to each other, however, there are several differences between two of these things. As the names suggest, Data Lakes are similar to real-life “lakes”, in which, there is no clear rigid structure of the data stored in it, they are essentially a place where raw data is stored; fit for the purpose of a system in need for a large and raw data. However, what if one needs to analyze said data? This is when Data Warehouses come in; as the name suggests, a Data “Warehouse” is similar to a “warehouse” in which the data is organized in a structured way, additionally, Data Warehouses also keep historical data; appropriate for data analysis thanks to its organized structure.

3. How does distribution and cloud deployment improve database reliability and scalability?

Database distributions and cloud deployments essentially attempts to eliminate the “single-point-of-failure”, known in this context as a single centralized database instance, by “distributing” said database instance into across multiple physical servers or locations. This practice of distributions, along with additional cloud deployments of the database, increases the system’s “fault tolerance” as it is less likely for the entire system to fail when there are multiple of the same database instances running simultaneously on different systems and in different physical locations along with enabling the ability for the administrators to create more instances of the database, therefore, increasing scalability.

4. If you were designing your own startup’s system, which combination would you choose — and why?

This depends on the startup system that would be created. If the startup is requires that of real-time, speedy, and quick transactions (or queries) such as, e-commerce website, event reservation systems, or convenience stores, then it would be the most appropriate to use OLTP systems along with Data Warehouses as they provide the needed characteristics. However, if the startup is rather related to that of analytics, e.g. dashboards, then OLTP systems with Data Lakes would be the most applicable. Although, it must be noted that most startup systems would most likely include all of the elements, in the case of the book store, OLTP system and Data Warehouses would be used for handling quick and large user requests and OLAP and Data Lakes for analytical or trend analysis tasks.