Module 1

OLAP Operations

OLAP stands for *Online Analytical Processing* Server. It is a software technology that allows users to analyze information from multiple database systems at the same time. It is based on multidimensional data model and allows the user to query on multi-dimensional data (eg. Delhi -> 2018 -> Sales data). OLAP databases are divided into one or more cubes and these cubes are known as *Hyper-cubes*.

Basic analytical operations of OLAP

Four types of analytical OLAP operations are:

1. Roll-up
2. Drill-down
3. Slice and dice
4. Pivot (rotate)

1) Roll-up:

Roll-up is also known as "consolidation" or "aggregation." The Roll-up operation can be performed in 2 ways

1. Reducing dimensions
2. Climbing up concept hierarchy. Concept hierarchy is a system of grouping things based on their order or level.
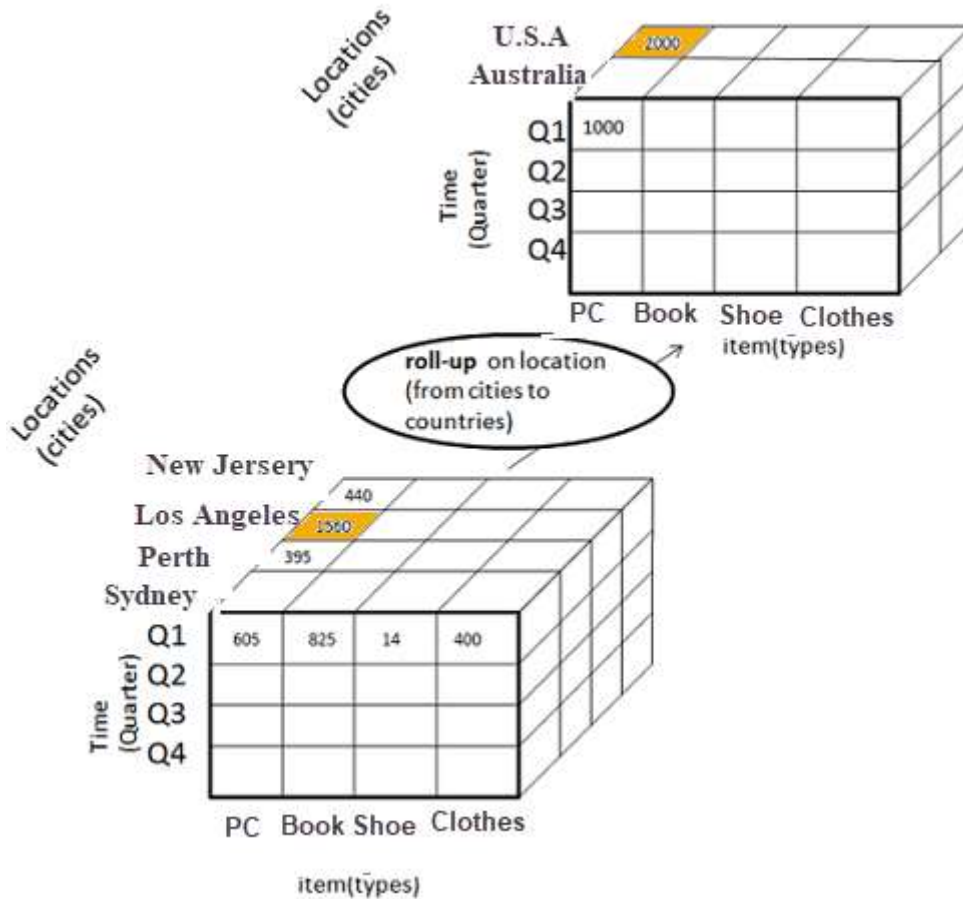
Consider the following diagram

Roll-up operation in OLAP

- In this example, cities New jersey and Lost Angles and rolled up into country USA
- The sales figure of New Jersey and Los Angeles are 440 and 1560 respectively. They become 2000 after roll-up
- In this aggregation process, data is location hierarchy moves up from city to the country.
- In the roll-up process at least one or more dimensions need to be removed. In this example, Cities dimension is removed.

2) Drill-down

In drill-down data is fragmented into smaller parts. It is the opposite of the rollup process. It can be done via
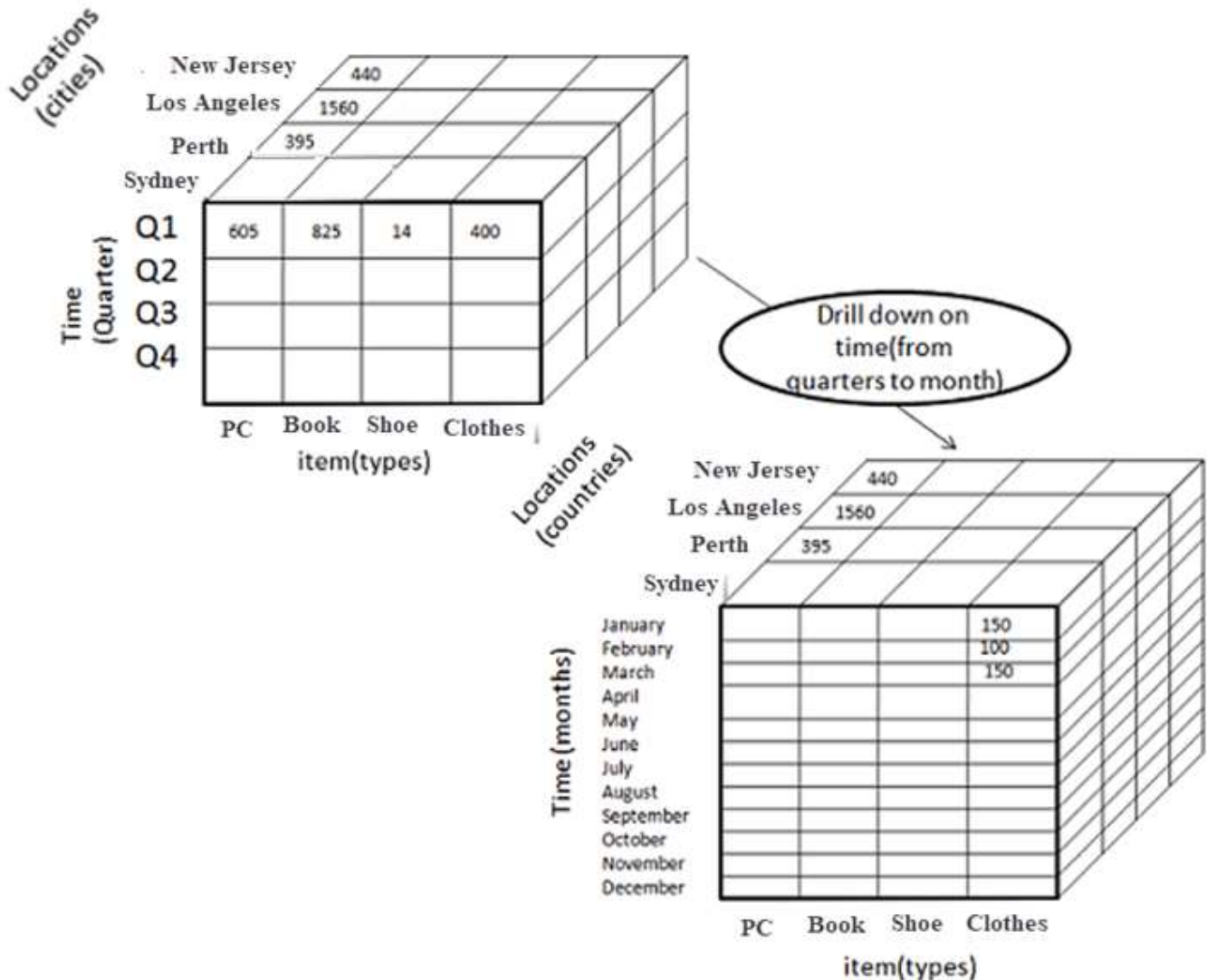
- Moving down the concept hierarchy
- Increasing a dimension

Drill-down operation in OLAP

Consider the diagram above

- Quater Q1 is drilled down to months January, February, and March. Corresponding sales are also registers.
- In this example, dimension months are added.

3) Slice:

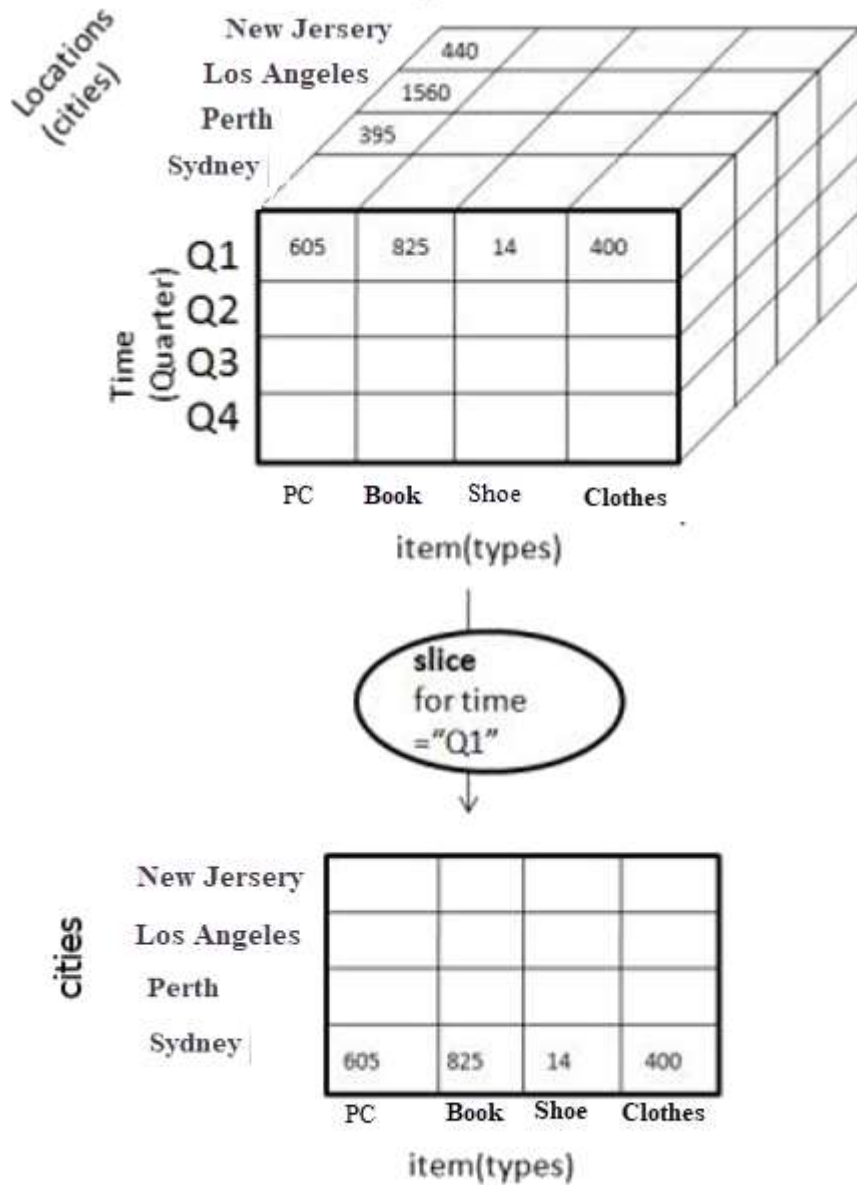Here, one dimension is selected, and a new sub-cube is created.

Following diagram explain how slice operation performed:

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : V   Subject :DWM   Academic Year: 2023 - 2024

Slice operation in OLAP

- Dimension Time is Sliced with Q1 as the filter.
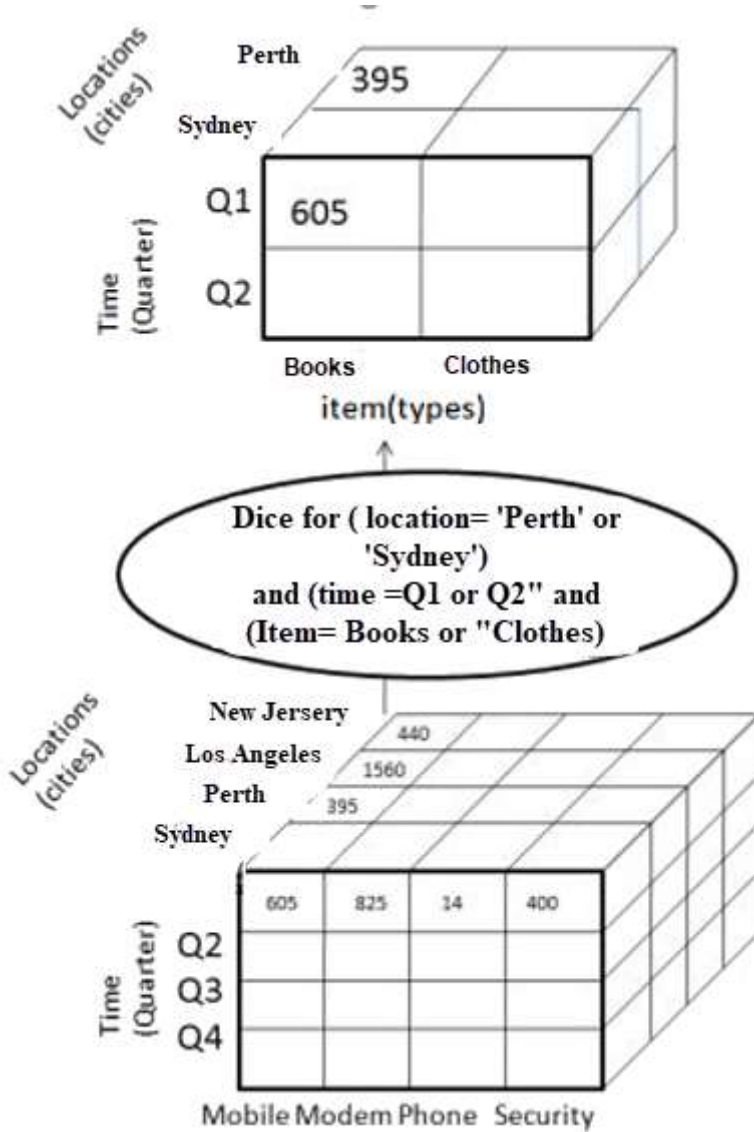- A new cube is created altogether.

Dice:

This operation is similar to a slice. The difference in dice is you select 2 or more dimensions that result in the creation of a sub-cube.

Dice operation in OLAP

4) Pivot

In Pivot, you rotate the data axes to provide a substitute presentation of data.

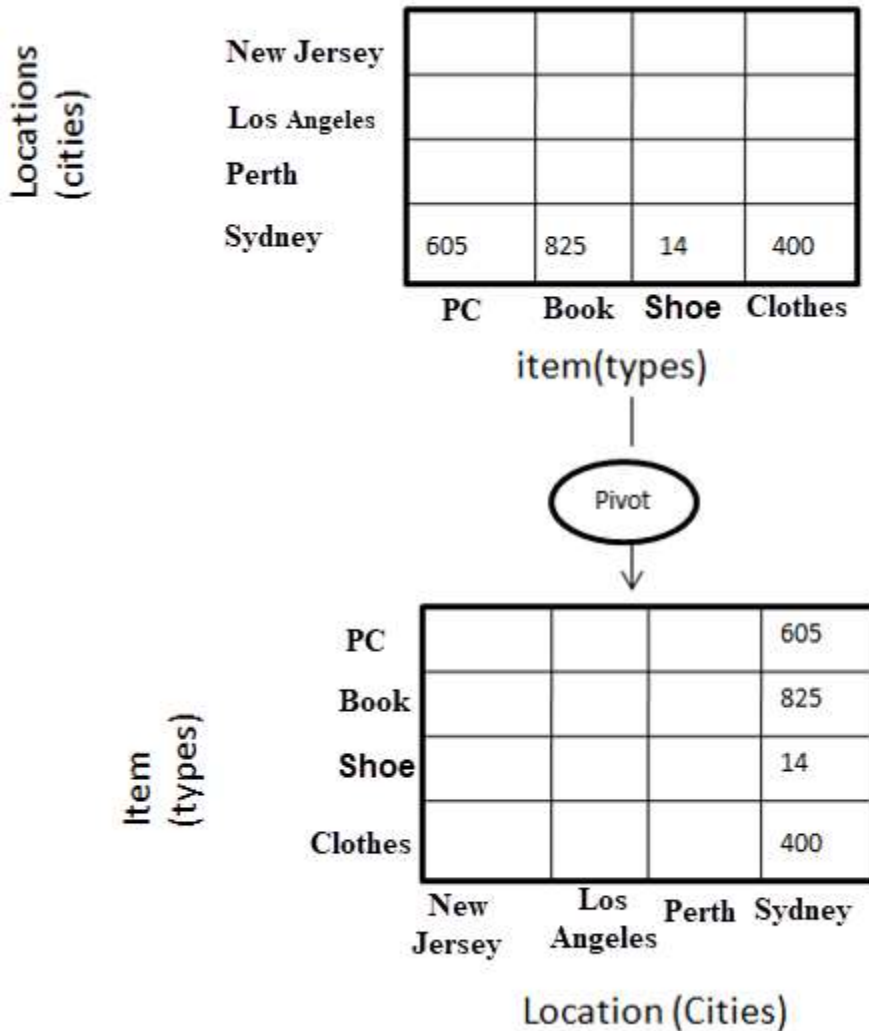In the following example, the pivot is based on item types.

Pivot operation in OLAP

What are the similarities between OLAP and OLTP?

Both online analytical processing (OLAP) and online transaction processing (OLTP) are database management systems for storing and processing data in large volumes. They require efficient and reliable IT infrastructure to run smoothly. You can use them both to query existing data or store new data. Both support data-driven decision-making in an organization.

Most companies use OLTP and OLAP systems together to meet their business intelligence requirements. However, the approach to and purpose of data management differ significantly between OLAP and OLTP.

Key differences: OLAP vs. OLTP

The primary purpose of online analytical processing (OLAP) is to analyze aggregated data, while the primary purpose of online transaction processing (OLTP) is to process database transactions.

You use OLAP systems to generate reports, perform complex data analysis, and identify trends. In contrast, you use OLTP systems to process orders, update inventory, and manage customer accounts.

Other major differences include data formatting, data architecture, performance, and requirements. We'll also discuss an example of when an organization might use OLAP or OLTP.

Data formatting

OLAP systems use multidimensional data models, so you can view the same data from different angles. OLAP databases store data in a cube format, where each dimension represents a different data attribute. Each cell in the cube represents a value or measure for the intersection of the dimensions.

In contrast, OLTP systems are unidimensional and focus on one data aspect. They use a relational database to organize data into tables. Each row in the table represents an entity instance, and each column represents an entity attribute.

Data architecture

OLAP database architecture prioritizes *data read* over *data write* operations. You can quickly and efficiently perform complex queries on large volumes of data. Availability is a low-priority concern as the primary use case is analytics.

On the other hand, OLTP database architecture prioritizes *data write* operations. It's optimized for write-heavy workloads and can update high-frequency, high-volume transactional data without compromising data integrity.

For instance, if two customers purchase the same item at the same time, the OLTP system can adjust stock levels accurately. And the system will prioritize the chronological first customer if the item is the last one in stock. Availability is a high priority and is typically achieved through multiple data backups.

Performance

OLAP processing times can vary from minutes to hours depending on the type and volume of data being analyzed. To update an OLAP database, you periodically process data in large batches then upload the batch to the system all at once. Data update frequency also varies between systems, from daily to weekly or even monthly.

In contrast, you measure OLTP processing times in milliseconds or less. OLTP databases manage database updates in real time. Updates are fast, short, and triggered by you or your users. Stream processing is often used over batch processing.

OLAP systems act like a centralized data store and pull in data from multiple data warehouses, relational databases, and other systems. Storage requirements measure from terabytes (TB) to petabytes (PB). Data reads can also be compute-intensive, requiring high-performing servers.

On the other hand, you can measure OLTP storage requirements in gigabytes (GB). OLTP databases may also be cleared once the data is loaded into a related OLAP data warehouse or data lake. However, compute requirements for OLTP are also high.

Example of OLAP vs. OLTP

Let's consider a large retail company that operates hundreds of stores across the country. The company has a massive database that tracks sales, inventory, customer data, and other key metrics.

The company uses OLTP to process transactions in real time, update inventory levels, and manage customer accounts. Each store is connected to the central database, which updates the inventory levels in real time as products are sold. The company also uses OLTP to manage customer accounts—for example, to track loyalty points, manage payment information, and process returns.

In addition, the company uses OLAP to analyze the data collected by OLTP. The company's business analysts can use OLAP to generate reports on sales trends, inventory levels, customer demographics, and other key metrics. They perform complex queries on large volumes of historical data to identify patterns and trends that can inform business decisions. They identify popular products in a given time period and use the information to optimize inventory budgets.

When to use OLAP vs. OLTP

Online analytical processing (OLAP) and online transaction processing (OLTP) are two different data processing systems designed for different purposes. OLAP is optimized for complex data analysis and reporting, while OLTP is optimized for transactional processing and real-time updates.
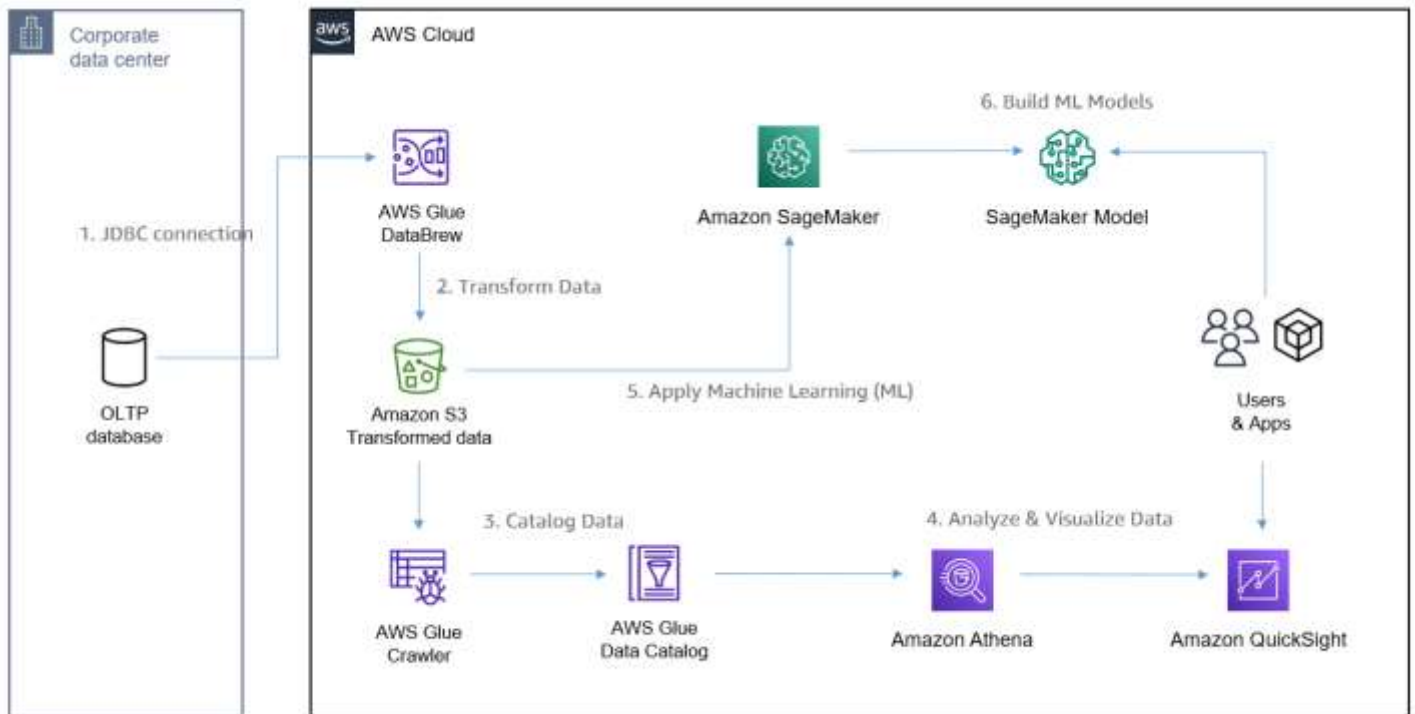
Understanding the differences between these systems can help you make informed decisions about which system meets your needs better. In many cases, a combination of both OLAP and OLTP systems may be the best solution for businesses that require both transaction processing and data analysis. Ultimately, choosing the right system depends on the specific needs of your business, including data volume, query complexity, response time, scalability, and cost.

PARSHWANATH CHARITABLE TRUST'S

A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

| Semester : V | Subject :DWM | Academic Year: 2023 - 2024 |

Summary of differences: OLAP vs. OLTP

Criteria	OLAP	OLTP
Purpose	OLAP helps you analyze large volumes of data to support decision-making.	OLTP helps you manage and process real-time transactions.
Data source	OLAP uses historical and aggregated data from multiple sources.	OLTP uses real-time and transactional data from a single source.
Data structure	OLAP uses multidimensional (cubes) or relational databases.	OLTP uses relational databases.
Data model	OLAP uses star schema, snowflake schema, or other analytical models.	OLTP uses normalized or denormalized models.
Volume of data	OLAP has large storage requirements. Think terabytes (TB) and petabytes (PB).	OLTP has comparatively smaller storage requirements. Think gigabytes (GB).
Response time	OLAP has longer response times, typically in seconds or minutes.	OLTP has shorter response times, typically in milliseconds

| Example applications | OLAP is good for analyzing trends, predicting customer behavior, and identifying profitability. | OLTP is good for processing payments, customer data management, and order processing. |