

CS313, Lab - 1

Rishit Saiya, 180010027

September 5, 2020

1

1.1

Michael Ralph Stonebraker is a computer scientist who specializes in database research. Currently he is a Professor Emeritus at UC Berkeley and an adjunct professor at MIT. He is an editor for the book Readings in Database Systems. He won Turing Award in 2014. His research and products are central to many relational database systems. He is also the founder of many database companies, including Ingres Corporation, Paradigm4, etc.

Source: [Wiki - Micheal Stonebraker](#)

Edgar Frank Codd was an English Computer Scientist who, while working for IBM, invented the relational model for database management, the theoretical basis for relational databases and relational database management systems. He coined the term Online analytical processing (OLAP) and wrote the "twelve laws of online analytical processing". He received the Turing Award in 1981, and in 1994 he was inducted as a Fellow of the Association for Computing Machinery. His relational model, a very influential general theory of data management, remains his most mentioned, analyzed and celebrated accomplishment.

Source: [Wiki - Edgar Frank Codd](#)

1.2

A database application is a computer technology whose main purpose is entering and retrieving information from a computer database. Modern database applications facilitate simultaneous updates and queries from multiple users. Some of the database applications are Amazon, Oracle relational database, eBay, Facebook, Google, etc. Back in mid-1990s, it became more common to build database applications with a Web interfaces. They had an advantage that it could be used in devices of various sizes, with different hardware components, and with different operating systems.

Source: [Wiki - Database Application](#)

A data model organizes different elements of data and standardizes on the basis of their relation to one another and to the properties of real-world entities. Data models are mainly specified by a data specialists, data librarians, or digital humanities scholars in a data modeling notation. These notations are represented in graphical form. A data model main function remains to determine the structure of data. In the context of programming languages, a data model can be sometimes referred as a data structure.

For Example: Some of the data models are Flat model, Hierarchical model, Network model, Relational model, etc.

Source: [Wiki - Data Model](#)

2

Some large database applications that have a huge database size and large number of transactions are as follows:

- Bookings: GoIbibo, RedBus, Oyo
- Payments: Paytm, PhonePe
- Others: Dream11, Flexcube, Zerodha, UpStox

Source: [Large Scale DB Apps](#), [UPI - India based](#)

3

OLTP:

Online Transactional Processing or OLTP is collection/class of software programs which are capable of supporting the applications which are Transaction-Oriented. These systems are designed to support on-line transaction and process query quickly on the Internet. If a transaction fails, built-in system logic ensures data integrity.

For Example: Telemarketers entering telephone survey results, POS (point of sale) system of any supermarket, Sending a text message, etc.

Source: [OLTP Reference](#)

OLAP:

On-Line Analytical Processing or OLAP is a classification of software programs which authorizes different professionals like analysts, executives and managers to get an overview into information through fast, consistent, interactive channel access in a wide variety of possible views of data. Since most businesses data have multiple dimensions i.e., multiple categories into which the data are broken down for presentation, tracking, or analysis, OLAP is the most convenient way to process data at multiple dimensions.

| | OLTP | OLAP |
|---------------------|--|--|
| Characteristics | Handles a large number of small transactions | Handles large volumes of data with complex queries |
| Inserts and Updates | Short and fast inserts and updates initiated by end users | Periodic long-running batch jobs refresh the data |
| Purpose | Control and run essential business operations in real time | Plan, solve problems, support decisions, discover hidden insights |
| Productivity | Increases productivity of end users | Increases productivity of business managers, data analysts, and executives |
| Data view | Lists day-to-day business transactions | Multi-dimensional view of enterprise data |

Table 1: OLTP vs. OLAP

For Example: Business reporting for sales, Marketing, Management Reporting, Business Process Management (BPM), Budgeting and Forecasting, Financial Reporting, etc.

Source: [OLAP Reference](#)

Table 1 shows the differences between OLTP & OLAP over different parameters.

Source: [OLTP vs. OLAP Reference](#)

4

After some searching, I got the below link given in the source. It mentions that the speed is 360 MB/sec.

Our given Bank data is 25 TB, so we calculate as follows:

$$360 \text{ MB} \rightarrow 1 \text{ sec}$$

$$2.5 \times 10^7 \text{ MB} \rightarrow ?$$

$$T = \frac{2.5 \times 10^7 \text{ MB} \times 1 \text{ sec}}{360 \text{ MB}}$$

$$T \approx 69,445 \text{ sec}$$

This 69,445 sec compute to approximately 19.3 hours. The average environment is considered in terms of disk speed only.

Source: [Bank Reference](#), [LTO-8](#)