

# CS313: Lab Assignment 1

B Siddharth Prabhu

200010003@iitdh.ac.in

16 August 2022

## 1 Answers to Question 1

### 1.1 E.F. Codd and Michael Stonebraker

Edgar Frank Codd was an English computer scientist, whose most celebrated achievement is the invention of the relational model for database management. In 1948, he moved to New York to work for IBM as a Mathematical Programmer, and he was appointed IBM Fellow in 1976. In the 1960s and 1970s, he worked out his theories of data arrangement, and issued his paper “A Relational Model of Data for Large Shared Data Banks” in 1970, at IBM. To his disappointment, IBM proved slow to make use of his suggestions until commercial rivals started implementing them. He received the Turing Award in 1981 for his fundamental contributions to the theory and practice of DBMS. He coined the term ‘OnLine Analytical Processing’ (OLAP) in 1993, and also wrote the “Twelve laws of online analytical processing”. Outside of DBMS, he also worked with self-replication in cellular automata, and he showed that a set of 8 states was sufficient for universal computation and construction.

Reference 1: [Wikipedia: E.F. Codd](#)

Reference 2: [Wikipedia: Codd's Cellular Automaton](#)

Michael Ralph Stonebraker is an American computer scientist specializing in database systems. He earned his M.S. and Ph.D. from the University of Michigan in 1967 and 1971 respectively. He joined the University of California, Berkeley as an assistant professor in 1971, and he taught in the Computer Science department for 29 years. That was also where he did his early pioneering work on relational databases. In 1973, Stonebraker and his colleague E. Wong started researching relational database systems, after reading a series of papers published by E.F. Codd on the relational data model. He received a Turing Award in 2014 for inventing several concepts that are central to DBMS. Stonebraker is currently a Professor Emeritus at UC Berkeley and an adjunct professor at MIT. He has founded numerous companies that are successfully commercializing his pioneering database technology work.

Reference 1: [Wikipedia: Michael Stonebraker](#)

Reference 2: [MIT News: Michael Stonebraker wins Turing Award](#)

## 1.2 Data Models and Database Applications

A data model is a conceptual model that organizes elements of data, while standardizing the relations among the data and those between the data and properties of real-world entities. Data models explicitly determine the structure of data. They are generally specified by data specialists, data librarians, or digital humanities scholars using some kind of data modeling notation. Data models can sometimes be referred to as data structures, especially in the context of programming languages. The main functions of data models involve describing the structure, manipulation, and integrity aspects of structured and unstructured data in data management systems. A 'database model' is a type of data model that determines the logical structure of a database. Types of data models include: Conceptual Data Model, Logical Data Model, and Physical Data Model. The Entity-Relationship Model (which is of interest to us in the CS303 Theory Course) is a Logical data model for databases.

Reference 1: [Wikipedia: Data Model](#)

Reference 2: [Wikipedia: Database model](#)

A database application is a computer program whose main purpose is retrieval of information from a computerized database. This information can be inserted, updated or deleted, and can then be reflected in the database. Modern database applications facilitate simultaneous updates and queries from multiple users. Examples of database applications include: Wikipedia, Gmail, Amazon, Youtube, SAP (Systems, Applications & Products in Data Processing), MySQL, etc. In the Mid-1990s, it became more common to build database applications with a Web interface. The advantage of this is that it can be used on devices of different sizes, with different hardware, and with different OS. Other uses of database applications may include transaction processing, or various machine learning calculations. Database applications may provide a way for data to be consumed either by end users or other higher-level applications.

Reference 1: [Wikipedia: Database Application](#)

Reference 2: [MongoDB: What Is A Database Application?](#)

## 2 Answer to Question 2

Here are some large digital applications (such as Google Pay, Amazon) that have a huge database size and large number of transactions:

1. Finance & Payments: PayTM, MobiKwik, PhonePe
2. Retail & IT: Tata Consultancy Services, DMart, Flipkart
3. Ticket Booking: MakeMyTrip, Goibibo, Yatra, ClearTrip
4. Others: Wipro, Tech Mahindra

Reference 1: [FirstPost: Indian retailers embracing Big data analytics to gauge consumer preferences](#)

Reference 2: [WalkThroughIndia: Top 10 Best Flight Booking Websites in India](#)

### 3 Answer to Question 3

To begin, let us have an overview of each of the terms: OLTP and OLAP.

OLTP (OnLine Transaction Processing) is a class of software programs that is capable of supporting applications that are transaction-oriented. Two important characteristics of OLTP are concurrency and atomicity. Concurrency prevents multiple users from altering the same data at the same time, while atomicity guarantees data integrity. What this means is, if one step of a transaction is incomplete, or it fails, then the process will not continue. OLTP Applications must be highly available, because downtime in such an application can have negative consequences, such as lost sales.

OLAP (OnLine Analytical Processing) is a class of software programs that is capable of performing multidimensional analysis at high speeds on large volumes of data from a data warehouse, data mart, or some other unified, centralized data store. Most datasets in business have multiple dimensions, i.e., multiple categories into which the data is broken down for different uses such as presentation, tracking, or analysis. However, in a data warehouse, data sets are organized as tables, each of which can organize data into just two of these dimensions at a time. So, OLAP extracts data from multiple relational datasets and reorganizes it into a multidimensional format that enables quick processing and insightful data analysis.

Now let us compare and contrast the above-described classes of software:

| Category            | OLTP   | OLAP   |
|---------------------|--|--|
| Main Objective      | The main objective is data processing; not data analysis.  | The main objective is data analysis; not data processing.  |
| Characteristic      | It is characterized by large numbers of short, simple, standardized queries.                     | It is characterized by a large volume of data, with complex, aggregative queries.                      |
| Table Normalization | Database Tables are normalized to eliminate data redundancy and enhance data integrity.          | Database Tables are not normalized.  |
| Operations          | It allows both read and write operations. It is mainly based on INSERT, UPDATE, DELETE commands. | Mostly just read is used here, data is rarely ever written. It is mainly based on SELECT commands.     |
| Productivity        | It helps to increase the end-user's self-service and productivity.                               | It helps to increase productivity of business managers, data analysts, and other data knowledge users. |
| Response Time       | Its response time is in the milliseconds.  | Its response time is in seconds to minutes.  |

Table 1: Comparison of OLTP and OLAP

Reference 1: [TechTarget: OLTP \(online transaction processing\)](#)

Reference 2: [IBM: What is OLAP?](#)

Reference 3: [Guru99: OLTP vs OLAP](#)

Reference 4: [Stitch: OLTP and OLAP](#)