LECTURE 1

INTRODUCTION TO DATABASES & SQL

Ankita Pai

Aegis

SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

Agenda

- Definition of DBMS

- Characteristics of Databases

- What is SQL?

- SQL Sublanguages

- Components of Database System

- Users of database system

- Role of Data Scientists in SQL

- Data Models ,Types of Models, Benefits

- The Entity-Relationship (ER) Model, Notations

- Relational Terminology

LEARNING OUTCOME

- **To understand**

  – Definitions related to database

  – Data Scientists & SQL

  – SQL and its Sublanguages

- **Learn To Draw**

  – ER Diagrams

  – Mapping ER Model to Relational Model

MySQL WORKBENCH

MySQL and MySQL Workbench

- Open source tool

- Popular choice for working with databases for use in web applications.

- Client server system => clients communicate to server through SQL.

- Free Graphical tool

- Makes it easier to work with MySQL databases.

Data

- **Data** are simply facts or figures
    - Bits of **information**, but not **information** itself.
- **Data** processed, interpreted, organized, structured or presented => meaningful or useful data called **information**.

Database

- A database is a
  - collection of data
  - with some inherent meaning
  - representing some aspect of real world
  - designed, built and populated with data for a specific purpose.

Characteristics of Databases

- Metadata and Self describing nature

- Persistent Data

- Support of multiple views of the data

- Sharing of data and multiuser transaction processing

Example of Metadata

Filename:	Tadzik.jpg
Author:	Piotr Kononow
Date:	August 15, 2016 6:40:10PM
File:	5,312 × 2,988 JPEG 15.9 megapixels 3,393,448 bytes (3.2 megabytes)
Camera:	Samsung SM-G920F 4.3 mm
Lens:	Max aperture f/1.9 (shot wide open) Auto exposure Program AE
Exposure:	1/402 sec f/1.9 ISO 40
Flash:	none

DBMS

- Collection of programs that enables user to create and maintain a database =>
  - general-purpose software
  - provides the users with the processes of defining, constructing and manipulating the database for various applications.

Components of Database System

- Database
- DBMS software
- Application Programs
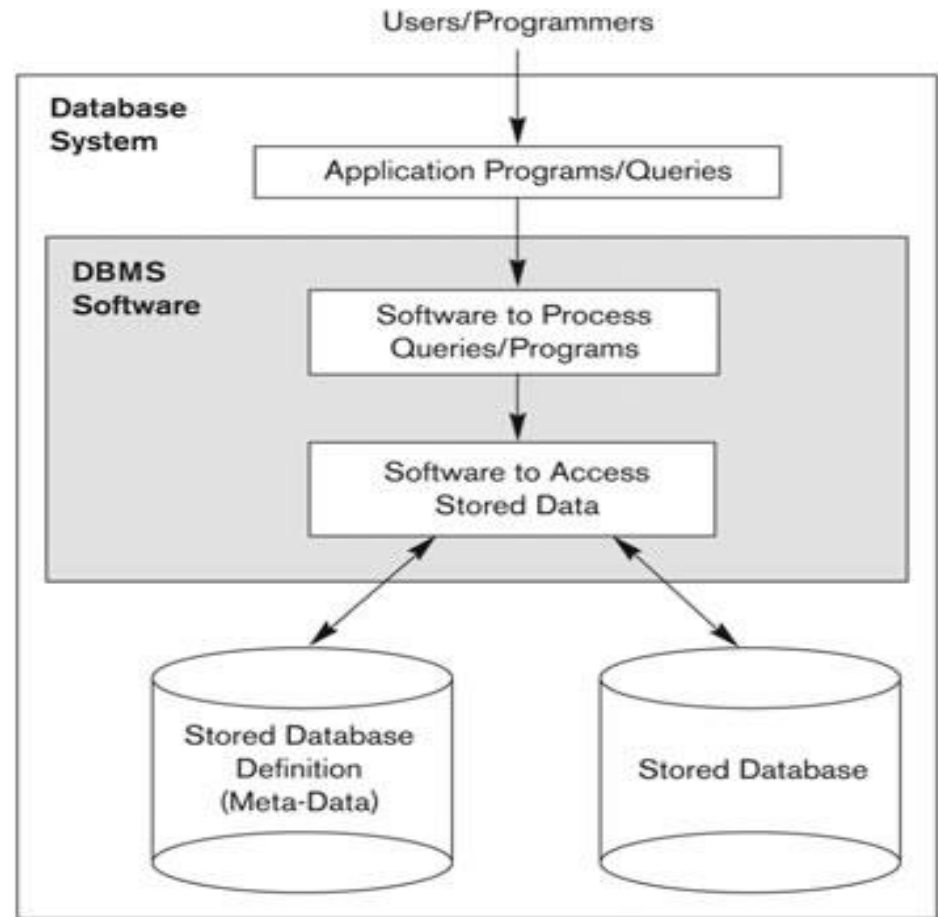
Types of Databases

- Traditional database

- Multimedia databases

- Geographic information systems

- Data warehouses

Advantages of using databases

- Redundancies and Inconsistencies can be reduced.

- Better services can be provided to the users.

- Cost of developing and maintaining system is lower.[data and application program are independent]

- Security and Integrity can be improved.[data used by many users at a time]

TASK

1. What is internal to a DBMS software?
   a) Application programs and queries
   b) Software to process queries/programs
   c) Software to access stored data
   d) Meta-data
2. DBMS provides user with processes to define data in the database. Justify
3. Database is collection of data. True/False
4. Give an example of a real time database.
5. What do you understand by the self describing nature of the database?

SQL: STRUCTURED QUERY LANGUAGE

What is SQL?

- Standard computer language for relational database management and data manipulation

- Used to query, insert, update and modify data

- Used to communicate with databases

- SQL is a non procedural language

- SQL cannot write complete applications

- Simple, but powerful

SQL Sublanguages

DDL	DML	DRL	DCL	TCL
CREATE ALTER DROP RENAME	INSERT UPDATE DELETE	SELECT	GRANT REVOKE	COMMIT ROLLBACK SAVEPOINT

Users of SQL

- Backend developer
- QA Engineer
- Database Administrator(DBA)
- Data Analyst
- System Administrator
- Data Architect

- ETL Developer (Extract, Transform, Load)
- Systems Engineer
- Data Scientist

DATA SCIENTISTS

&

SQL

How do Data Scientists use SQL?

- Retrieve data

- May create their own table or test environment

- Combine multiple sources together

- Writes complex queries for analysis

- Forecasting sales/revenue/stock prices etc.

- For building recommendation systems.

SQL and Database Management Systems

- How you write syntax will depend on what DBMS you are using

- Each DBMS has its own syntax

- You will tweak based on the syntax your DBMS uses

- Some commonly used RDBMS are:
  - IBM DB2 Oracle
  - PostgreSQL
  - MySQL
  - Microsoft SQL Server
  - Apache Open Office Base
  - SQLite

Aegis

SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

TASK

1. Which person is responsible for overall activities for database?
   a) Database designer
   b) Database analyst
   c) Database administrator
   d) Database manager
2. SQL can be used to copy, read, write and create data. True/False
3. What is the main purpose for which data scientists use SQL?
   a) Data Retrieval
   b) Making predictions based on data
   c) Building models
   d) Writing complex queries for analysis
4. Differentiate and compare between DBA and data scientists.

DATA MODELS

What is a data model?

- A database model shows the logical structure of a database

  - Includes the relationships and constraints that determine how data can be stored and accessed.

- Most data models can be represented by an accompanying database diagram.

Types of Database Models

- Hierarchical database model

- Relational model

- Network model

- Object-oriented database model
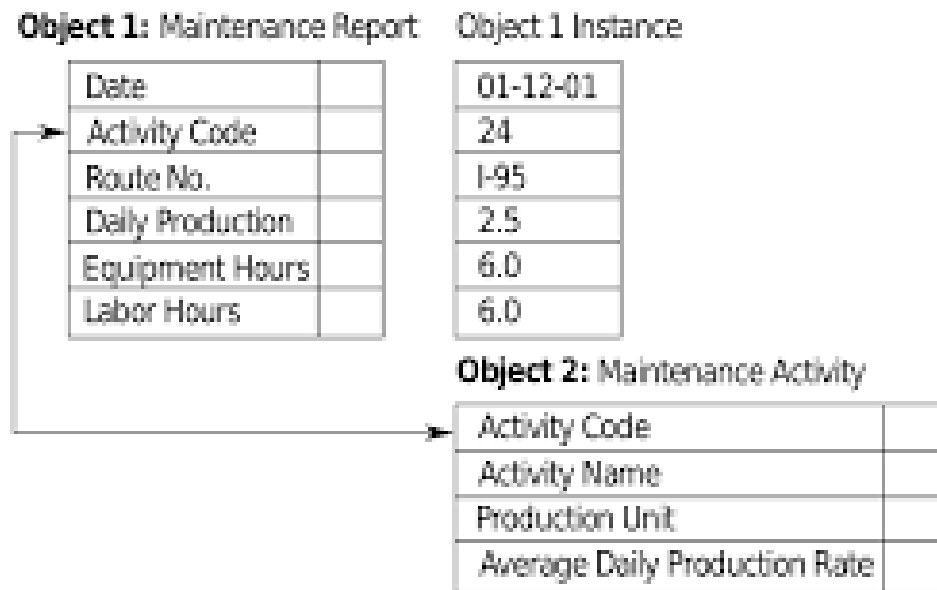
Benefits of data modelling

- Focusing on essentials

- Ease of communication and understanding

- Product or process improvement

- Exploring Alternatives[what-if]

Building Blocks of a Data Model

- **Entity:** Real or abstract object. Person, place thing or event;

- **Attribute:** Characteristics of an entity.

- **Relationship:** Describes association among entities.
  - One-to-one
  - One-to-many
  - Many-to-many

What is ER model?

- Composed of entity types

- Specifies relationships that can exist between instances of those entity types

- Helps to understand and represent a business process

- Represented visually

- Shows links between tables (Primary keys and Foreign keys)
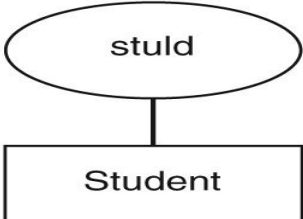
Primary Key and Foreign Key
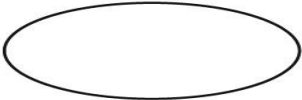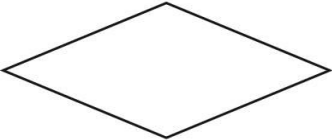
- **Primary key:**
  - A column or set of columns whose values uniquely identify every row in a table

- **Foreign key:**
  - One or more columns of a table that provides a link between two tables.
  - The values in the foreign key are the unique key values of another table
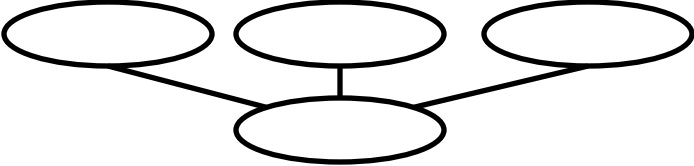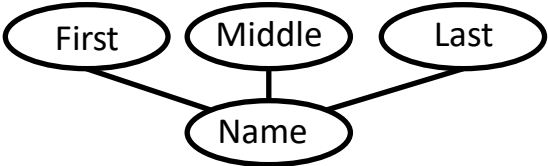
TASK
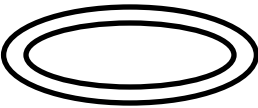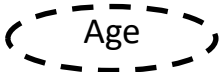
1. Which of the following are benefits of a Relational Data Model?
   a) It leaves data unstructured.
   b) It allows you to easily write queries against it
   c) It simplifies the connections between the data
   d) It is the least popular data model.

2. Data models represent the _____ structure of a database.

3. Match the following:

   i. Hierarchical model   a) Many to many relationship

   ii. Relational model   b) Multiple reusable software elements

   iii. Object Oriented model   c) Tables

   iv. Network model   d) Tree like structure

4. Which of the following could be possible primary attributes for an entity?
   a) Height
   b) Contact number
   c) Zip code
   d) PAN ID number

5. Give one example of each:
   – One-to-one
   – One-to-many
   – Many-to-many

Symbols used in ER Diagrams

SYMBOL	NAME	MEANING	EXAMPLE
▭	Rectangle	Entity Set	Student
⬭	Oval	Attribute	stuId
◇	Diamond	Relationship	Enroll
—	Line	Links: Attribute to Entity	stuId — Student

Types of Attributes

Symbol	Description	Meaning	Example
(oval)	Oval	Simple Attributes	Price
(multiple ovals connected with lines)	Multiple ovals connected with lines	Composite Attributes	First — Middle — Last — Name
(oval)	Oval	Single-valued Attributes /Stored Attributes	Height
(concentric ovals)	Concentric Ovals	Multi-valued Attributes	MobileNo
(dashed oval)	Dashed oval	Derived Attributes	Age
(underlined attribute oval)	Underlined attribute	Primary Attribute	Employee_ID

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

For example
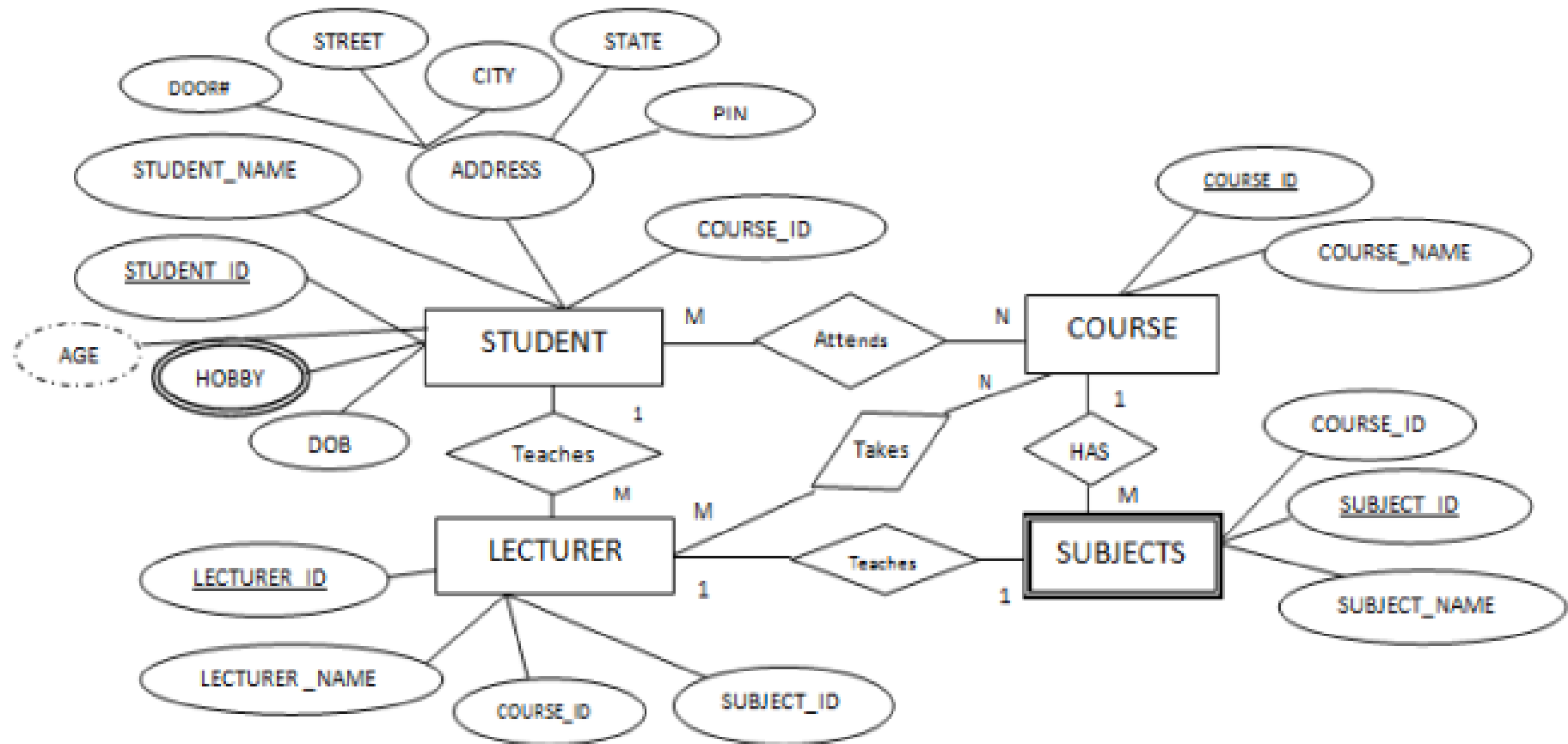
Database Schema

- Skeleton structure that represents the logical view of the entire **database**.

- Defines how the data is organized .

- Ex: **Cricket Database**
  - Match(MatchID, Team1,Team2, Ground, Date)
  - Player(PlayerID, Lname, Country, Yborn, Bplace)
  - Batting(MatchID, PID, Fours, Sixes)

Mapping ER Model to Relational Model

- Mapping Strong Entities

- Mapping Weak Entities

- Mapping Relationship
  - One to One Binary Relationship

  - One to Many Relationship

  - Many to Many Relationship

TASK

1. Draw the ER diagram for:
    a) A banking database
    b) A hospital
2. _____ is used as the primary key of a weak entity set.
3. Give two examples each for:
    a) Derived attribute
    b) Multi-valued attribute
    c) Composite attribute

TASK

1. Draw the relationship between the following entity sets using the ER diagram.
   a) Passenger entity and aircraft entity
   b) Book entity and member entity (Library database)

2. Write the database schema for:
   a) Courier service
   b) Book store

RELATIONAL DATABASE

Concepts

- **DATABASE:** A container (usually a file or set of files) to store organized data; a set of related information

- **TABLE:** A structured list of data

- **COLUMN:** A single field in a table; all tables are made up of one or more columns

- **ROW**: record in a table

	A	B	C	D	E	F	G	H	I	J	K
1	sku	name	brand	Price	item_height	item_length	item_width	item_unit_o	item_weight	item_unit_of_weight	
2	PFI-F0314BPY	Pfister F-031-4	Pfister	129.99	3.8	20.6	12.7	inches	6.5	pounds	
3	PFI-F042HAK0	Pfister Amher	Pfister	79.99	7.69	4.81	6.56	inches	3.42	pounds	
4	PFI-FWK1340S	Pfister F-WK1	Pfister	119.99	2.5	18.5	10	inches	5.12	pounds	
5	PFI-GT343TCC	Pfister GT34-3	Pfister	149.99	2.5	18.5	10	inches	6.2	pounds	
6	PFI-MP8LNKK	Pfister Langsto	Pfister	99.99	7.67	14.22	7.67	inches	3.2	pounds	
7	B00B4QEP0U	Pfister GT529	Pfister	109.99	2.5	18.5	10	inches	3.17	pounds	
8	PFI-GT529DCC	Pfister GT529	Pfister	179.99	2.5	24.5	10.5	inches	3.17	pounds	
9	PFI-GT529DSS	Pfister GT529	Pfister	114.99	2.5	24.5	10.5	inches	3.17	pounds	
10	B007LEP02Q	Pfister F-031-4	Pfister	249.99	4.1	20.6	12.8	inches	6.5	pounds	
11	PFI-F0314BPS	Pfister F-031-4	Pfister	144.99	4.1	20.6	12.8	inches	6.5	pounds	
12											

Terminology

- Relation => Table
- Attribute => Fields => Columns
- Tuple => Record => Row
- Degree of a Relation => No. of columns
- Cardinality of a Relation => No. of rows

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

TASK

1.   Which of the following is true of a *Relational* Database?

     a)   Shows the relationships between the different tables

     b)   Information is not stored in a way that's conducive to querying and analysis.

     c)   It's an operational database

     d)   Used to optimize querying data, making it easy and intuitive to access the data

2.   Relational database is transactional in nature. True/False

3.   What is the degree of the relation with the schema: Vehicle(Chassis_no, Model_no, Engine_displacement, Manufacturer, Price, Type, Mileage, Transmission, No_of_seats, Fuel)

4.   Tuples can otherwise be referred to as fields. True/False

References

- Textbook:
  - Fundamentals of Database Systems by Ramez Elmasri, Shamkant Navathe
  - Database Management Systems by Raghu Ramakrishnan, Johannes Gehrke
  - Database System Concepts by Abraham Silberschatz, Henry Forth, Sudarshan
- Web material:
  - https://www.1keydata.com/sql/sql-data-types.html
  - https://www.tutorialspoint.com/mysql/mysql-data-types.htm
  - https://dev.mysql.com/doc/refman/8.0/en/

Topics to be covered in next Lecture

- Data Types

- DDL Commands

- Adding Comments in SQL

Reference for these topics above:

Book: Elmasri-Navathe:"Fundamentals of database system"

Chapter: 8   **Page No:** 205-236

We will have MCQ on this Lecture.