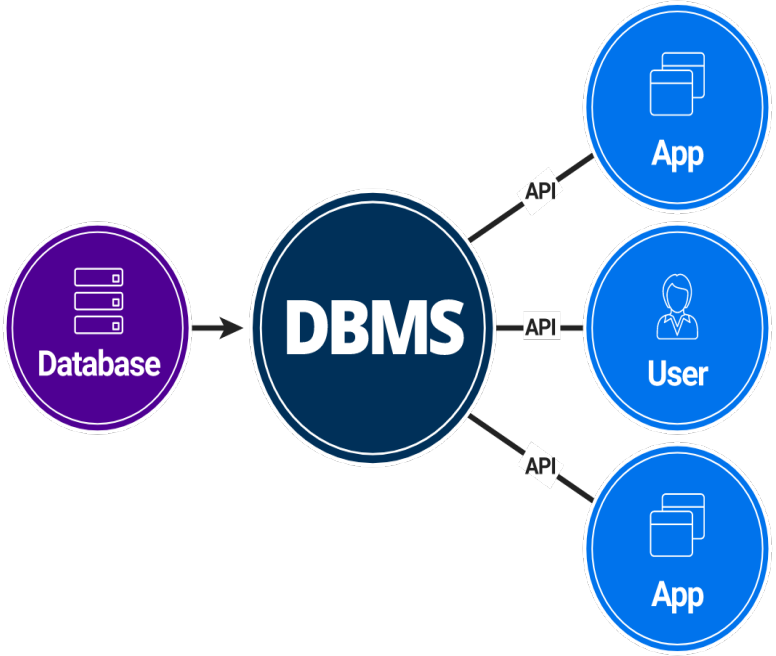
Module_1

Introduction to Database Systems & Structure Query Language

CS 3410

KENNESAW STATE UNIVERSITY

1.1 Introduction and Background



- 1960s were when databases were used from a computer-based format.
- Databases are utilized to store the necessary data for usage later over time
- Data is structured in rows and columns featuring different fields for queries and then later stored in multiple tables to showcase the relationship between them.
- A database management system, or DBMS for short, is a form of software that allows an organization to access and manipulate data that will be showcased in a form that is unable to be changed by other applications and users.

KENNESAW STATE UNIVERSITY

1.2 Limitations and Conventional File Processing

- Files are used to store specific data for future use and recollection.
- When computers first became mainstream, files were stored like paper, in the form of flat files. This information was collected in notepads separated by spaces, commas, semicolons or other symbols.
- In order to organize these files, they were based on their categories consisting of only related information with specific names.
- The only downside to this is that you were unable to open the files without using a specific coding language to edit it.
- While it appeared convenient at the time, it is easy to identify the many disadvantages to using this system.

Data Redundancy and Inconsistency

- One of the major problems with this system was **Data redundancy and inconsistency.**
- Most of the information is also constantly duplicated due to how tedious it would be to access others code and double check information. For example, if a customer of a bank has two accounts, the data accompanied by these accounts would be stored in two separate files in order to satisfy both accounts as they are made. This leads to **Data redundancy.**
- The countless copies of this data could also have **discrepancies** making it impossible to know which information is accurate.
- Whenever a new value needs to be entered into the database, every single file with this data has to be updated to prevent this. This would lead to tedious work that wasn't 100% accurate in the end due to never knowing exactly which files and the specific information you wanted the change without opening them

Integrity and Organization

- If a specific set of information is needed to be organized in a new way, unless it was anticipated prior to the initially being created, it was nearly impossible to achieve this.
- The application needed to display the information in the requested way would not have existed.
- This system doesn’t allow data to be retrieved in a convenient manner leading to different systems to be created down the line.
- Integrity problems were also created due to the data values in a database needing to satisfy certain types of consistency constraints.
- The involvement of multiple users may result in inconsistent data which is normally prevented using supervision. However, in a file processing system this supervision is lackluster due to the several applications and various languages.

Security

- Security is also a major issue in this system. In a database, every user in the database system shouldn’t be able to access all the data.
- In a file processing system since different programmers add their own application, there is either a universal password or so many passwords that the information is scrambled and the people requiring it can’t access it.
- Since every new file is only added when needed, it is difficult to constantly change the permission for each individual file in order to ensure security standards.
- **These disadvantages would lead many to convert to a Database Approach rather than file system. This Database corrected many of these errors reducing the development time and increasing the data integrity of every file.**

1.3 Advantages of Databases

- Data is prevalent in all aspects of both the consumption and production of business and services in today's world.
- Due to such, it is essential that data be stored and recorded in an efficient and accurate manner.
- This is where databases come into play, they offer several benefits and advantages as follows:
- Efficiency
- Versatility
- Categorization and Organization
- Multi-Access

Efficiency

- Businesses benefit the most from efficiency, as they save costs on wasted time/labor.
- The safety of having an efficient DBMS allows businesses to operate on a larger scale and handle large amounts of data.
- Businesses can safely utilize their data when they have trust in the accuracy and efficiency of their data storing method.

Versatility

- Versatility is important on both a consumer and producer level.
- It allows producers to access data from many different platforms whether it be by desktop, laptop, tablet or even smart device.
- For consumers this is especially useful, as most consumers in today's world access accounts via non-desktop device, so being able to manipulate and record their own data from their smart device is extremely useful and powerful.

Categorization and Organization

- The ability to categorize and organize data is very valuable to businesses.
- DBMS allow for structuring data in a way that is easily understood as well as accessed.
- Some DBMS allow for self-defined relationships between 'entities' in order to further simplify the way data is stored, and allow for more utilization

Multi-Access

- The ability to access data in a multitude of ways from multiple different users is known as multi-access.
- Multi-access is useful because it allows multiple authorized users to access the same data.
- Example:
- A Manager of Human Resources will have access to the same data of potential hires as a General Manager of a desired location.

1.4 Cost and Risks of Database Approach

- Databases off their fair share of great benefits and advantages, however, there is always give and take and there are costs and risks associated with running a DBMS.
- Some costs and risks are as follows:
- Implementation
- Installation and Maintenance
- Conversion costs
- Need for Explicit Backup/Recovery
- Organizational Conflict

Implementation

- The need for personnel who will implement the database system will be costly
- There is also a need to train onboarding employees on how to utilize your database system, which will not be free.
- The cost of hiring a single employee to work on your system may exceed $100,000 annually.

KENNESAW STATE UNIVERSITY

Installation and Maintenance

- DBMS require personnel to install and operate the newly created system based on the needs and demands of the business.
- As time passes though, these systems require constant maintenance and updating.
- As a business expands, there is a need to possibly upgrade the system as well as the hardware that the system runs on, which is definitely not cheap.

Conversion Costs

- A significant cost is also taken into account when one has to transfer data from a previous system onto a newly created database.
- This sounds simple, however it is extremely difficult to implement and takes an extraordinary amount of time, thus it is fairly expensive to perform.
- One wouldn't want to go cheap for Conversion, as it is very important that you transfer ALL of your data without loss.

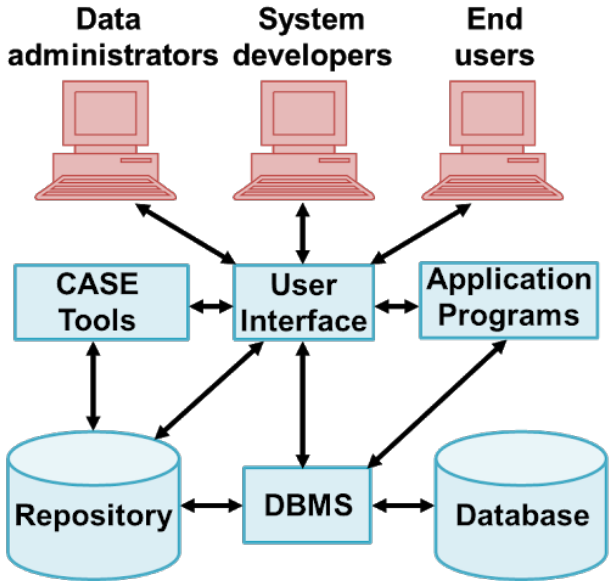
Need for Explicit Backup/Recovery

- A business must absolutely operate with backup data/ recovery data.
- In any business/system there is always a possibility for data loss via a multitude of ways, EX: software ,hardware, or human error.
- Thus to avoid losing data COMPLETELY, business must store backup data.
- This is data nonetheless which is not cheap to store, so it raises the cost of running a DBMS.

Organizational Conflict

- When implementing a large database system it is common for opposing views to clash within an organization.
- This can lead to conflict as well as inconsistency within a system
- The cost here lies in hiring strong and defined leadership to maintain consistency and resolve conflict.

1.5 Components of a Database Environment



- 5 major components of a Database Environment
  - People
  - Hardware
  - Software
  - Data
  - Procedures

KENNESAW STATE UNIVERSITY

People

Various important Roles

Necessary for building the database environment

Roles Included:

- System and Database Administrators – Manages the current server, physical database, and procedures required for securing the database
- Database designers - Programs all queries, relationships, and data
- Programmers – Same role as a database designer
- Analysts - Reviews all the data the programmer/database designer has implemented
- End Users - Utilizes the database management system and make the system more usable for other users

Hardware & Software

Software

- Software Includes:
  - Operating System
  - Programming tools
  - Server Access

Hardware

- Hardware Includes:
  - Actual computer
  - Networking components
  - Flash Disks needed for saving files and data

KENNESAW STATE UNIVERSITY

Data & Procedures

- Data Includes:
  - Actual database
  - Necessary business procedures and/or rules that manage the system.

- Procedures are implemented as a way to structure the overall design on how the database should work and regulate all the data that should be going in and coming out of the database

1.6 Database Systems Development Life Cycle

- The Database Life Cycle contains six phases.
- These phases include…
- **Database initial study (application design)**
- **Database Design (prototyping)**
- **Implementation and loading**
- **Testing and Evaluation**
- **Operation**
- **Maintenance and evolution**.

The Database Life Cycle (DBLC)

Database Initial Study

- The purpose of this study is to analyze the company situation, define problems and constraints, define objectives, and define the scope and boundaries. Each section can be broken down in order to further understand the usefulness behind creating this study.

· **Analyzing the company situation** pertains to defining the general conditions within a company including its organization structure and its mission. In order to correctly do this, the designer must discover what the company's operation components are, the way they function, and how they interact.

· **Defining Problems and Constraints** pertains to the discovery of the issues within the company formally and informally. These problems may appear unstructured, however overall, problems are usually connected allowing the designer to overcome them by the end of the process.

· **Defining Objectives** is a part of the new proposed database system showing that it is designed to solve the major problems identified previously.

· **Defining the Scope and Boundaries** pertains to the designer recognizing the existence of his limits: scope and boundaries. The system's scope shows the extent of the design according to the requirements. The system also is connected to limits known as boundaries which are external. These boundaries are set by the accompanying hardware and software.

KENNESAW STATE UNIVERSITY

Database Design

- Database design is the second phase focusing on the design of the database that supports the company operations and objectives in the future. This can be viewed as the most critical DBLC phase due to the insurance it brings in making sure that the final product meets the guidelines.

Implementation and Loading

- Implementation and Loading pertain to as series of instructions when dealing with the creation of tables, attributes, etc. in the domain. In this phase, the design specifications are installed creating the exact database required by the parent company. This can be done in 3 phases.

· **Install the DBMS:** Installing a new instance of a DBMS in the system taking place on an existing or new server.

· **Creating the DBMS**: Creates the table spaces and file groups accompanied by the database.

· **Loading and Converting Data:** After the database is created, the data must enter the new tables. This requires them to be merged and imported from other databases or the ones previously used in order to ensure the same data is relayed into the newer, better system.

KENNESAW STATE UNIVERSITY

Testing and Evaluation

- Testing and evaluation pertain to the decision made to ensure integrity, security, performance, and recoverability of the database. Following the plans laid out previously, this fine-tunes the database to ensure that performs as expected.

· **Test the database:** During this step, the database is tested to ensure it has the integrity and security required by the company. This is enforced through the proper use of primary and foreign key rules.

· **Fine-Tune the database:** This is the editing of the database with the results of the previous step in mind. If no fine-tuning is required, this step can be skipped.

· **Evaluating the Database:** The database must be reviewed thoroughly to ensure that the data contained is protected against loss- promoting the use of a backup.

KENNESAW STATE UNIVERSITY

Operation and Maintenance and Evolution

- Operation is the second to last step identifying that the database is fully functional. At this point, the database us complete and the new system has space to evolve as needed by the developers.
- The final step is maintenance and evolution. This step is directed by the database administrator allowing them to perform routine maintenance activities regarding the database. Some of these activities include **Backup, Corrected Maintenance**, **Adaptive Maintenance** and the **Assignment of access permissions** to welcome new user and edit old user.
- **All together these steps make up the Database Life cycle and ensure that a fully functional database is created by the end goal allowing for around the clock maintenance within the company promoting a highly efficient system meeting the guidelines presented at the beginning of the process.**

SQL (Structured Query Language)

- Databases offer businesses a smoother operating work situation. The implementation of a database management language such as SQL (Structured Query Language) allows businesses to access and modify data that is stored in a relational database.

Primary Key

- **primary key**, are used to create the relationships between the tables. For example, in the STUDENT table StudentNumber serves as the primary key. Each value of StudentNumber is unique and identifies a particular student.
- If the numbers used in primary key columns such as StudentNumber and ClassNumber are repeatedly created and assigned in the database itself, then the key is also called a **surrogate key** *(Kroenke, Auer, Vandenberg, Yoder, 2018).*

STUDENT

Field Name	Data Type	Description (Optional)
StudentNumber	AutoNumber	Surrogate key for STUDENT
LastName	Short Text	Student's last name
FirstName	Short Text	Student's first name
EmailAddress	Short Text	Student's email address

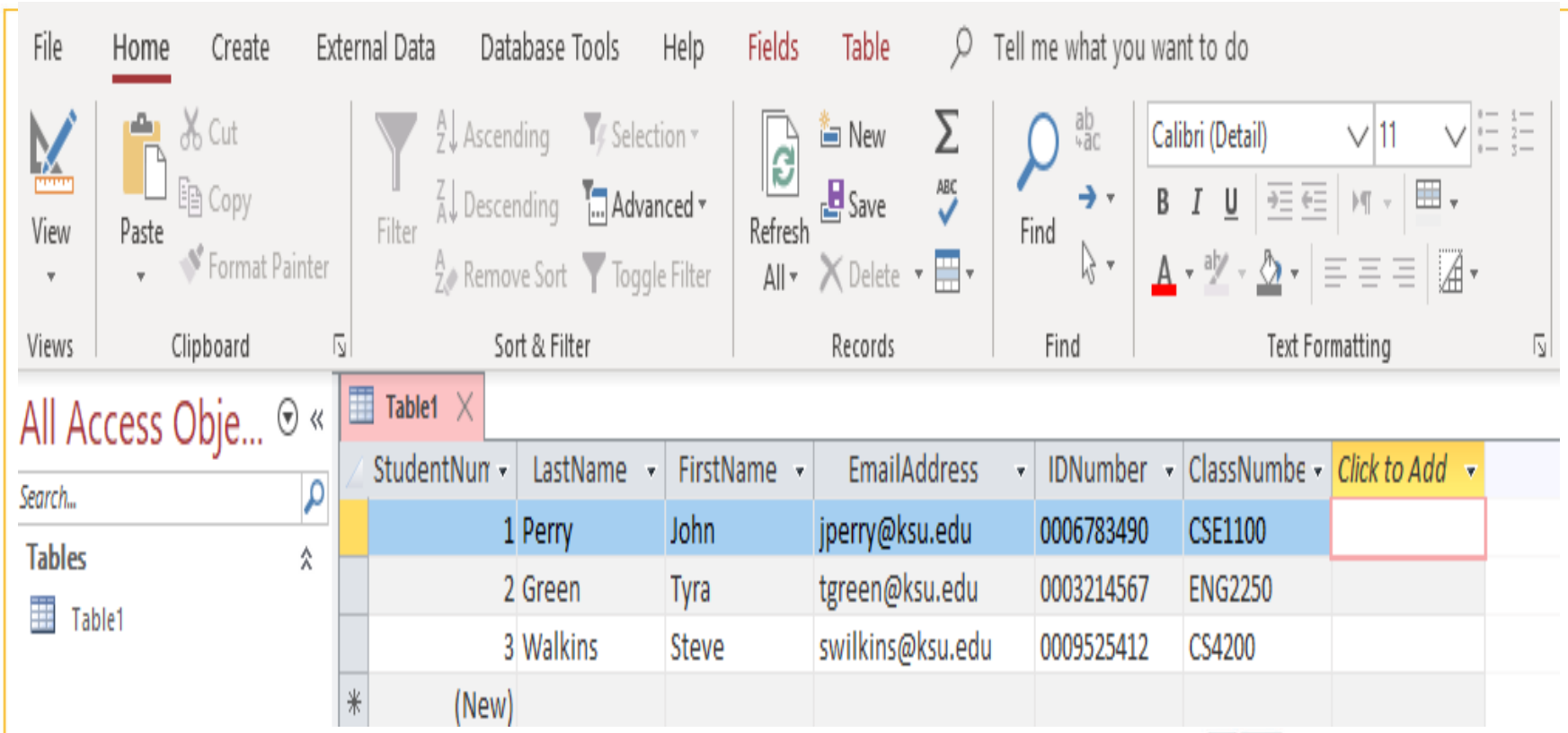
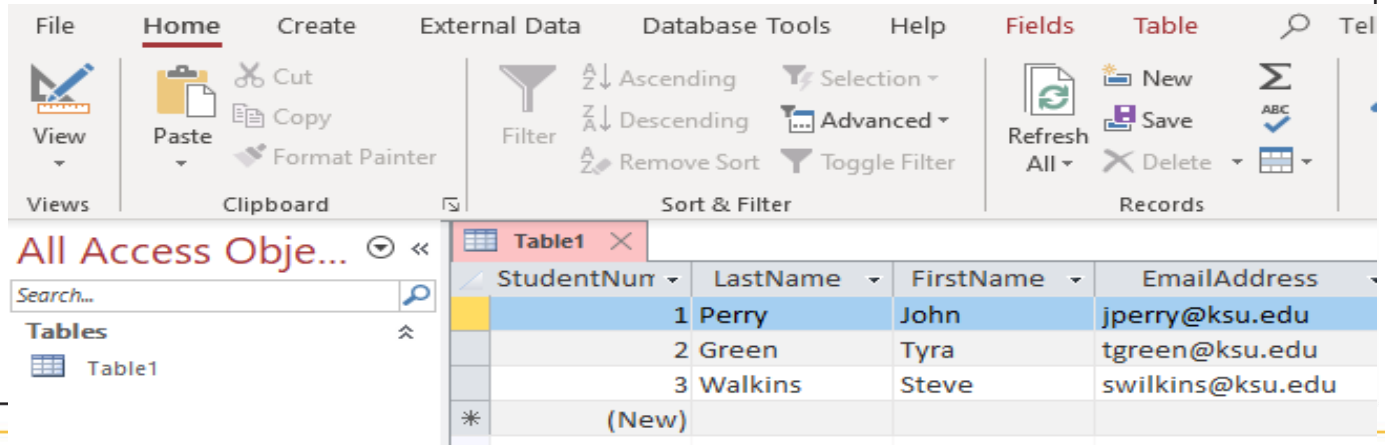


Figure 1-2 The Primary key and Surrogate key

KENNESAW STATE UNIVERSITY

- In the table below shows when more than one column in a table are merged to form of the primary key, is known as a **composite key**.
- A **foreign key** provides a **relationship** or link between two tables.
- *In the GRADE column, StudentNumber and ClassNumber each now serve as a **foreign key**.*

Figure 1-3 shows a **Microsoft Access 2016** point of view of the tables and their relationships.



- Figure 1-4 shows the greater database application, part of a **customer relationship management (CRM)** system, which manages customers and their contacts, purchases, support requests, and so forth.
- The CRM system uses software to support a larger company, which may include anywhere from 500 rows to 10 million or more.

- An **enterprise resources planning (ERP)** system is an information system that affects every department in a company, including sales, inventory, planning purchasing and other business purposes.
- **SAP** (System, Applications & Products in Data Processing) is the vendor used with ERP applications for large companies.

Application	User	Number of Users	Standard Size	Comments
Customer appointment (Doctor dentist)	Manger	20-15	500 rows	Marketing software
Customer relationship Management (CRM)	Senior Manger	10-15	10 million rows	Vendors applications such as Oracle
Data mining	Business Analysts	1-5	1000 to million rows	Data extracted and use by statistical data mining tools.

What is Microsoft Access?

- **Microsoft Access** is not just a database management system (DBMS) but is also a **personal database system.**
- Microsoft Access is a combination of the relational Microsoft Jet Database engine with a **graphical user interface (GUI)** and software-development tools.
- **Microsoft Access** is one of the office suites that is intended for individuals and small works groups such as interact with application through data entry process forms, generate reports, run the queries.

Conclusion

With the implementation of various types of databases and database management systems, data is now easier to store and update as time goes on.

Data can be structured in various ways.

Data also goes through many phases in its life cycle.

Database management may not be the most cost efficient based on organizational needs and funding.

KENNESAW STATE UNIVERSITY