SS2 Data Processing Lesson Note Second Term

WEEK : 2

TOPIC: Relational Database

CONTENT

Relational database

A relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. The relational database was invented by E. F. Codd at IBM in 1970.

The standard user and application program interface to a relational database is the structured query language (SQL).

A relational database is a set of tables containing data fitted into predefined categories. Each table (which is sometimes called a relation) contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. For example, a typical business order entry database would include a table that described a customer with columns for name, address, phone number, and so forth. Another table would describe an order: product, customer, date, sales price, and so forth.

A user of the database could obtain a view of the database that fitted the user's needs. For example, a branch office manager might like a view or report on all customers that had bought products after a certain date. A financial services manager in the same company could, from the same tables, obtain a report on accounts that needed to be paid.

Creating and modifying relations using SQL

SQL Statements

Most of the actions you need to perform on a database are done with SQL statements.

The following SQL statement selects all the records in the "Customers" table:

Example

SELECT * FROM Customers;

Some of The Most Important SQL Commands

- SELECT - extracts data from a database
- UPDATE - updates data in a database
- DELETE - deletes data from a database
- INSERT INTO - inserts new data into a database
- CREATE DATABASE - creates a new database
- ALTER DATABASE - modifies a database
- CREATE TABLE - creates a new table
- ALTER TABLE - modifies a table
- DROP TABLE - deletes a table
- CREATE INDEX - creates an index (search key)
- DROP INDEX - deletes an index

The SELECT statement is used to select data from a database.

The SQL SELECT Statement

The SELECT statement is used to select data from a database.

The result is stored in a result table, called the result-set.

SQL SELECT Syntax

SELECT column_name,column_name

FROM table_name;

and

SELECT * FROM table_name;

WEEK : 3-4

TOPIC: Relational Database- Integrity Constraints

CONTENT

RELATIONAL MODEL -- Integrity Constraints

Integrity Constraints over Relations

An integrity constraint (IC) is a condition specified on a database schema and restricts the data that can be stored in an instance of the database. If a database instance satisfies all the integrity constraints specifies on the database schema, it is a legal instance. A DBMS permits only legal instances to be stored in the database.

Many kinds of integrity constraints can be specified in the relational model:

Database integrity refers to the validity and consistency of stored data. Integrity is usually expressed in terms of constraints, which are consistency rules that the database is not permitted to violate.

Constraints may apply to each attribute or they may apply to relationships between tables.

Integrity constraints ensure that changes (update deletion, insertion) made to the database by authorized users do not result in a loss of data consistency. Thus, integrity constraints guard against accidental damage to the database.

EXAMPLE- A brood group must be 'A' or 'B' or 'AB' or 'O' only (can not be any other values else).

TYPES OF INTEGRITY CONSTRAINTS

Various types of integrity constraints are-

1.Domain Integrity

2.Entity Integrity Constraint

3.Referential Integrity Constraint

4.Key Constraints

1. Domain Integrity- Domain integrity means the definition of a valid set of values for an attribute. You define data type, length or size, is null value allowed , is the value unique or not for an attribute ,the default value, the range (values in between) and/or specific values for the attribute.

1. Entity Integrity Constraint- This rule states that in any database relation value of attribute of a primary key can't be null.

EXAMPLE- Consider a relation "STUDENT" Where "Stu_id" is a primary key and it must not contain any null value whereas other attributes may contain null value e.g "Branch" in the following relation contains one null value.

Stu_id Name Branch

11255234 Aman CSE

11255369 Kapil ECE

11255324 Ajay

11255237 Raman CSE

11255678 Aastha ECE

3.Referential Integrity Constraint-It states that if a foreign key exists in a relation then either the foreign key value must match a primary key value of some tuple in its home relation or the foreign key value must be null.

4.Key Constraints- A Key Constraint is a statement/condition that a certain minimal subset of the fields of a relation is a unique identifier for a tuple.

There are 4 types of key constraints-

1.Candidate key.

2.Super key

3.Primary key

4.Foreign key

WEEK : 5

TOPIC: Querying Relational Database

CONTENT

SELECT Query Basics

Print

Why MS-Access?

A number of RDBMS vendors provide a GUI to aid their users in developing queries. These can be particularly helpful to novice users as it enables them to learn the overarching concepts involved in query development without getting bogged down in syntax details. For this reason, we will start the course with Microsoft Access, which provides perhaps the most user-friendly interface.

1. Download an Access database and review its tables

Throughout this lesson, we'll use a database of baseball statistics to help demonstrate the basics of SELECT queries.

1.Open the database in MS-Access.

One part of the Access interface that you'll use frequently is the "Navigation Pane," which is situated on the left side of the application window. The top of the Navigation Pane is just beneath the "Ribbon" (the strip of controls that runs horizontally along the top of the window).

The Navigation Pane provides access to the objects stored in the database, such as tables, queries, forms and reports. When you first open the baseball_stats.accdb database, the Navigation Pane should appear with the word Tables at the top, indicating that it is listing the tables stored in the database (PLAYERS, STATS and TEAMS).

2.Double-click on a table's name in the Navigation Pane to open it. Open all three tables and review the content. Note that the STATS table contains an ID for each player rather than his name. The names associated with the IDs are stored in the PLAYERS table.

1. Write a simple SELECT query

With our first query we'll retrieve data from selected fields in the STATS table.

1.Click on the Create tab near the top of the application window.

2.Next, click on the Query Design button (found on the left side of the Create Ribbon in the group of commands labeled as Queries).

3.In the Show Table dialog, double-click on the STATS table to add it to the query and click Close, (when you do this in Access 2010 the ribbon switches to the Design ribbon).

4.Double-click on PLAYER_ID in the list of fields in the STATS table to add that field to the design grid below.

5.Repeat this step to add the YEAR and RBI fields.

6.At any time, you can view the SQL that's created by your GUI settings by accessing the View drop-down list on the far-left side of Design Ribbon, (it is also available when you have the Home tab selected, as shown below).

As you go through the next steps look at the SQL that corresponds to queries you are building.

1. Restrict the returned records to a desired subset

1.From the same View drop-down list, select Design View to return to the query design GUI.

2.In the design grid, set the Criteria value for the RBI field to >99.

3.Test the query by clicking on the red exclamation point, (it should return 103 records).

1. Sort the returned records

1.Return to Design View.

2.In the design grid click in the Sort cell under the RBI column and select Descending from the drop-down list. This will sort the records from highest RBI total to lowest.

3.Test the query.

1. Add additional criteria to the selection

1.Return to Design View and set the Criteria value for the YEAR field to >1989. This will limit the results to seasons of over 100 RBI since 1990.

2.Test the query, (it should return 53 records).

3.Return to Design View and modify the Criteria value for the YEAR field to >1989 And<2000, which will further limit the results to just the 1990s.

4.Test the query, (it should return 34 records).

5.Return to Design View and change the Criteria value for the YEAR field back to >1989, beneath that cell (in the :or cell) add <1960.

As you should be able to guess, I'm asking you to write a query that identifies 100-RBI seasonssince 1989 OR prior to 1960. However, the query as written at this point doesn't quite yield that result, look at the WHERE line in the SQL view. Instead, it would return 100-RBI seasons since 1989 and all seasons prior to 1960 (not just the 100-RBI ones). To produce the desired result you need to repeat the >99 criterion in the RBI field's :or cell, check the SQL view to see the change.

6.Test the query, (it should return 74 records).

You've probably recognized by now that the output from these queries is not particularly human friendly. In the next part of the lesson we'll see how to use a join between the two tables to add the names of the players to the query output.

WEEK : 6-7

TOPIC: File Organization

CONTENT

FILE

A collection of data or information that has a name, called thefilename. Almost all information stored in a computer must be in a file. There are many different types of files: data files, text files ,program files, directory files, and so on. Different types of files store different types of information. For example, program files store programs, whereas text files store text.

Computer File

A computer file is a resource for storing information, which is available to a computer program and is usually based on some kind of durable storage. A file is "durable" in the sense that it remains available for other programs to use after the program that created it has finished executing. Computer files can be considered as the modern counterpart of paper documents which traditionally are kept in office and library files, and this is the source of the term.

File organization

File organization" refers to the logical relationships among the various records that constitute the file, particularly with respect to the means of identification and access to any specific record. "File structure" refers to the format of the label and data blocks and of any logical record control information

It is used to determine an efficient file organization for each base relation. For example, if we want to retrieve student records in alphabetical order of name, sorting the file by student name is a good file organization. However, if we want to retrieve all students whose marks is in a certain range, a file ordered by student name would not be a good file organization. Some file organizations are efficient for bulk loading data into the database but inefficient for retrieve and other activities.

The objective of this selection is to choose an optimal file organization for each relation.

Types of File Organization

In order to make effective selection of file organizations and indexes, here we present the details different types of file Organization. These are:

- Heap File Organization
- Hash File Organization
- Indexed Sequential Access Methods (ISAM) File Organization

Heap (unordered) File Organization

An unordered file, sometimes called a heap file, is the simplest type of file organization.

Records are placed in file in the same order as they are inserted. A new record is inserted in the last page of the file; if there is insufficient space in the last page, a new page is added to the file. This makes insertion very efficient. However, as a heap file has no particular ordering with respect to field values, a linear search must be performed to access a record. A linear search involves reading pages from the file until the required is found. This makes retrievals from heap files that have more than a few pages relatively slow, unless the retrieval involves a large proportion of the records in the file.

Pros of Heap storage

Heap is a good storage structure in the following situations:

When data is being bulk-loaded into the relation.

The relation is only a few pages long. In this case, the time to locate any tuple is Short, even if the entire relation has been searched serially.

When every tuple in the relation has to be retrieved (in any order) every time the relation is accessed. For example, retrieve the name of all the students.

Cons of Heap storage

Heap files are inappropriate when only selected tuples of a relation are to be accessed.

Hash File Organization

In a hash file, records are not stored sequentially in a file instead a hash function is used to calculate the address of the page in which the record is to be stored.

The field on which hash function is calculated is called as Hash field and if that field acts as the key of the relation then it is called as Hash key. Records are randomly distributed in the file so it is also called as Random or Direct files. Commonly some arithmetic function is applied to the hash field so that records will be evenly distributed throughout the file.

WEEK : 8-9

: Internet

CONTENT

Internet

The Internet is the global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link billions of devices worldwide. It is a network of networks that consists of millions of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries an extensive range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web (WWW), electronic mail,telephony, and peer-to-peer networks for file sharing.

Web browser

A web browser (commonly referred to as a browser) is a software application for retrieving, presenting, and traversing information resources on theWorld Wide Web. An information resource is identified by a Uniform Resource Identifier (URI/URL)

and may be a web page, image, video or other piece of content.[1] Hyperlinks present in resources enable users easily to navigate their browsers to related resources.

Although browsers are primarily intended to use the World Wide Web, they can also be used to access information provided by web servers in private networks or files in file systems.

The major web browsers are Firefox, Internet Explorer/Microsoft Edge,[2][3][4] Google Chrome, Opera, and Safari.

- Anchor

Synonymous with hyperlinks, anchor refers to non-linear links among documents. Or more simply put, it's the word or phrase that can be clicked to connect to another page or resource.

- ASCII

(pronounced "Ask-ee") An acronym for American Standard Code for Information Exchange, ASCII is an international standard in which numbers, letters, punctuation marks, symbols and control codes are assigned numbers from 0 to 27. Easily transferred over networks, ASCII is a plain, unadorned text without style or font specifications.

- Avatar

This term refers to an interactive representation of a human in a virtual reality environment.

- Bandwidth

The range of transmission frequencies a network can use. The greater the bandwidth the more information that can be transferred over that network at one time.

- Baseband

A transmission method in which a network uses its entire transmission range to send a single signal.

- Broadband

A transmission method in which the networks range of transmission frequencies is divided into separate channels and each channel is used to send a different signal. Broadband is often used to send different types of signals simultaneously.

- Browser

A type of software that allows you to navigate information databases.

- Client

A computer that has access to services over a computer network. The computer providing the services is a server.

- DNS

An acronym for Domain Name Server, DNS refers to a database of Internet names and addresses which translates the names to the official Internet Protocol numbers and vice versa.

- Document

When used in reference to the World Wide Web, a document is any file containing text, media or hyperlinks that can be transferred from an HTTP server to a client program.

- Document Window

This is the Mosaic program's scrollable window in which HTML documents can be viewed.

- Download

To transfer to your computer a copy of a file that resides on another computer.

- FAQ

This is the acronym for Frequently Asked Questions. A common feature on the Internet, FAQs are files of answers to commonly asked questions.

- Firewall

This term refers to security measures designed to protect a networked system from unauthorized or unwelcome access.

- FTP

File Transfer Protocol is a protocol that allows the transfer of files from one computer to another. FTP is also the verb used to describe the act of transferring files from one computer to another.

- Gopher

A menu-oriented tool used to locate online resources.

- Gopherspace

A term used to describe the entire gopher network.

- GUI

An acronym for Graphical User Interface, this term refers to a software front-end meant to provide an attractive and easy to use interface between a computer user and application.

- Home Page

The document displayed when you first open Mosaic. Home Page also refers to the first document you come to at a Web site.

- Hotlists

Lists of frequently used Web locations and URLs (Uniform Resource Locators).

- Host

A computer acting as an information or communications server.

- HTML

An acronym for HyperText Markup Language, HTML is the language used to tag various parts of a Web document so browsing software will know how to display that document's links, text, graphics and attached media.

- HTML Document

A document written in HyperText Markup Language.

- HTTP

The abbreviation for Hypertext Transfer Protocol, HTTP is used to link and transfer hypertext documents.

- Hypermedia

The hypertext concept extended to include linked multiple media.

- Hypertext

This term describes the system that allows documents to be cross- linked in such a way that the reader can explore related documents by clicking on a highlighted word or symbol.

- IP

The abbreviation for Internet Protocol, IP refers to the set of communication standards that control communications activity on the Internet. An IP address is the number assigned to any Internet-connected computer.

- ISDN

The abbreviation for Integrated Services Digital Network, ISDN is a telecommunications standard that uses digital transmission technology to support voice, video and data communications applications over regular telephone lines.

- JPEG

The acronym for Joint Photographic Experts Group, JPEG is an image compression format used to transfer color photographs and images over computer networks.

Links

These are the hypertext connections between Web pages. This is a synonym for hotlinks or hyperlinks.

- POP

An acronym for Point of Presence, POP is a service provider's location for connecting to users.

- Router

A communications device designed to transmit signals via the most efficient route possible.

- Search Engine

This term refers to a program that helps users find information in text-oriented databases.

- Server

A computer system that manages and delivers information for client computers.

- TCP-IP

The basic protocols controlling applications on the Internet.

- URL

This is the abbreviation for Uniform Resource Locator, The addressing system used in the World Wide Web and other Internet resources.

- Web Browser

This is the software that allows a user to access and view HTML documents. Examples of Web browsers include Mosaic, Cello and Lynx.

- Web Document

An HTML document that is browsable on the Web.

- Web Page

An HTML document that is accessible on the Web.

Uses of the internet

Internet is today one of the most important part of our daily life. There are large numbers of things that can be done using the internet and so it is very important. The following are uses of the internet.

1. Communication

At the moment the easiest thing that can be done using the internet is that we can communicate with the people living far away from us with extreme ease. Now people can not only chat but can also do the video conferencing. Email, social networking sites are some of the prime example of it.

1. Research

Now the point that has been placed next is research. In order to do research you need to go through hundreds of books as well as the references and that was one of the most difficult jobs to do earlier. Since the internet came into life, everything is available just a click away. You just have to search for the concerned topic and you will get hundreds of references that may be beneficial for your research. And since internet is here to make your research public, you can then benefit a large amount of people from the research work that you have done. Research is one such thing which has got lots of benefit from this evolution of internet. Research process has now got wings and has gained the most due to the internet.

1. Education

The next point that we have in this list is education. Yes you read it right. Education is one of the best things that the internet can provide. There are a number of books, reference books, online help centres, expert's views and other study oriented material on the internet that can make the learning process very easier as well as a fun learning experience. There are lots and lots of websites which are related to different topic. You can visit them and can gain endless amount of knowledge that you wish to have. With the use of internet for education, you are non-longer dependent on some other person to come and teach you. There are various number of tutorials available over the internet using which you can learn so many thing very easily. There can't be any excellent use of the internet other than education as it is the key to achieve everything in life.

1. Financial Transaction

The next use mentioned here is financial transaction. Financial transaction is the term which is used when there is exchange of money. With the use of internet in the financial transaction, your work has become a lot easier. Now you don't need to stand in the queue at the branch of your particular bank rather you can just log in on to the bank website with the credential that has been provided to you by the bank and then can do any transaction related to finance at your will. With the ability to do the financial transaction easily over the internet you can purchase or sell items so easily. Financial transaction can be considered as one of the best uses of resource in the right direction.

1. Real Time Updates

Real time updates have been placed at the number fifth position here. This has been mentioned here in regards to the news and other happenings that may be on-going in different parts of the world but with the use of internet we come to know about it very easily and without any difficulty. There are various websites on the internet which provides you with the real time updates in every field be it in business, sports, finance, politics, entertainment and others. Many a time the decisions are taken on the real time updates that are happening in various parts of the world and this is where internet is very essential and helpful.

1. Email:

By using internet now we can communicate in a fraction of seconds with a person who is sitting in the other part of the world. Today for better communication, we can avail the facilities of Email. We can chat for hours with our loved ones. There are plenty messenger services and email services offering this service for free. With help of such services, it has become very easy to establish a kind of global friendship where you can share your thoughts, can explore other cultures of different ethnicity.

1. Information:

The biggest advantage that internet offering is information. The internet and the World Wide Web has made it easy for anyone to access information and it can be of any type, as the internet is flooded with information. The internet and the World Wide Web has made it easy for anyone to access information and it can be of any type. Any kind of information on any topic is available on the Internet.

1. Business:

World trade has seen a big boom with the help of the internet, as it has become easier for buyers and sellers to communicate and also to advertise their sites. Now a day's most of the people are using online classified sites to buy or sell or advertising their

products or services. Classified sites save a lot of money and time so this is chosen as medium by most of people to advertise their products. We have many classified sites on the web like craigslist, Adsglobe.com, Kijiji etc.

1. Social Networking:

Today social networking sites have become an important part of the online community. Almost all users are members use it for personal and business purposes. It's an awesome place to network with many entrepreneurs who come here to begin building their own personal and business brand.

1. Shopping:

In today's busy life most of us are interested to shop online. Now a day's almost anything can be bought with the use of the internet. In countries like US most of consumers prefer to shop from home. We have many shopping sites on internet like amazon.com, Dealsglobe.com etc. People also use the internet to auction goods. There are many auction sites online, where anything can be sold.

11 Entertainment:

On internet we can find all forms of entertainment from watching films to playing games online. Almost anyone can find the right kind of entertainment for themselves. When people surf the Web, there are numerous things that can be found. Music, hobbies, news and more can be found and shared on the Internet. There are numerous games that may be downloaded from the Internet for free.

12 E-Commerce:

Ecommerce is the concept used for any type of commercial maneuvering, or business deals that involves the transfer of information across the globe via internet. It has become a phenomenon associated with any kind of shopping, almost anything. It has got a real amazing and range of products from household needs, technology to entertainment.

13 Services:

Many services are now provided on the internet such as online banking, job seeking, purchasing tickets for your favorite movies, and guidance services on array of topics in the every aspect of life, and hotel reservations and bills paying. Often these services are not available off-line and can cost you more.

14 Job Search:

Internet makes life easy for both employers and job seekers as there are plenty of job sites which connects employers and job seekers.

15 Connect: People are connecting with others though internet.

INTERNET SECURITY

Internet security is a branch of computer security specifically related to the Internet, often involving browser security but also network security on a more general level as it applies to other applications or operating systems on a whole. Its objective is to establish rules and measures to use against attacks over the Internet.[1] The Internet represents an insecure channel for exchanging information leading to a high risk of intrusion or fraud, such as phishing.[2] Different methods have been used to protect the transfer of data, including encryption and from-the-ground-up engineering.[3]

WEEK : 10

TOPIC: Internet Abuse

CONTENT

Internet abuse

Internet abuse refers to improper use of the internet and may include:

Internet has so many advantages, if it is used positively without caring any false means and attractive charms of internet. Internet is producing false trend specially among the students which easily become a victim of its colorful and charming aspects. Some disadvantages of internet may be described as following.

- The students waste their precious hours on sitting on internet without taking any positive and constructive benefits.
- Most of the people are using internet to satisfy their sexual desires.
- Nude pictures and sexual material is easily be available through internet which destroy the moral values of young boys and girls.
- Computer hiking is very common by the use of internet. Some extreme minded people can digest the money through the use of credit card and others.
- The student waste their time in useless talking with each others.
- Several hours ion internet without any purpose produce wrong effects on a person's health.

Types of Abuse

Spamvertized

A type of unsolicited e-mail (spam) that specifically advertises a product.

Copyright Infringement

The unauthorised use of material in a manner that violates copyright law.

Hacking

The unauthorised use of computer and network resources.

Open Relay

An SMTP server that allows third party relay of email messages. An open relay makes it possible for an unscrupulous sender to route large volumes of email messages.

Open Proxy

A proxy server that is configured so that any internet user can use it. Open proxy servers are widely used by spammers due to the fact that it hides the spammer's IP Address from recipients.

Port Scan

Port scanning software is designed to search a network host for open ports which can be exploited.

Virus/Worm

A virus is a program or piece of code loaded onto your computer without your knowledge and runs against your wishes. They can replicate themselves and spread over a network. A worm spreads from computer to computer but unlike a virus has the capability to travel without any human action.

Phishing

The fraudulent process of attempting to acquire sensitive information such as usernames, passwords or credit card details by masquerading as a trustworthy entity in electronic communication such as a website.

Blacklisting

Domains/ email addresses/ IP Addresses/ IP Ranges are denied access to a service because of a "bad name".

DoS Attack

A denial of service attack that attempts to make a computer resource unavailable to its intended users.

Dictionary Attack

A method of breaking into a password protected computer or server by systematically entering every word in a dictionary as a password.

Probes

A probe is an attempt to gain access to a computer and its files through a known or probable weak point in the computer system for the purpose of monitoring or collecting data about network activity.

Unsolicited Bounce

An unsolicited bounce is an email "bounce message" that is sent to a user whose name was forged into the headers of the email that provoked the bounce. It is also referred to as a misdirected bounce since it is sent to a user who should not have received it.

419 Scam

A 419 Scam is an attempt to steal money from people based on the promise that a person will gain extra money if they advance a small amount.

Auto Responder

An auto-responder is a computer program that automatically answers email sent to it. Users may consider these as spam.

Fraudulent Activity

Fraudulent activity refers to any attempt to unlawfully obtain money, information or resources from unsuspecting users within a network.