Database Management - C175

Tuesday, April 20, 2021 8:22 PM

Operational Databases are Summerized and Stored

Analitical Databases are olap online

Olap access is by business intelligance tools

Unstructed data is raw state data

Structured data is data formatted for storage

Semistructured data - data that has been processed to a certain extent, has relevant meaning

data is accessed with SQL commands.

Record is one row of data

a File is a collection of related records

Data redundency is bad and can cause poor security and data inconsistency

Uncontrolled data redundency is a big issue

Data anomoly is inconsistent data, usually caused by uncontrolled data redundency

Data Model - ERD, First step in designing a database / database blueprint

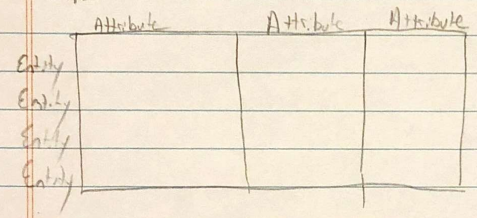
Attribute - Column / Field "Customer name"

One to many - Painter creates many paintings
Many to many - employee learns skills and skills can be learned by anyone
One to one - Store manager manages one store
Constraints - placed on data to ensure data integrity (a rule)
Schema - all of the objects grouped together
Subschema - part of the schema seen by the program
DML commands - Select, insert, update, delete
Commit and rollback
DDL Data definition Language - Create, alter, drop
Each row in a relation is called a tuple
RDBMS - Relational Database Management system
UML - Unified Modeling language
ADT - Abstract data type

Primary Keys
- Simple - Consists of one column
- Composite - Consists of multiple columns
  - (First name + Last name + Email) (Composite Key) — Natural
- Compound - multiple columns that are keys themselves

Basic Database operations
Create
Read
Update
Delete

3 Types of List Anomalies
Deletion
update
Insertion



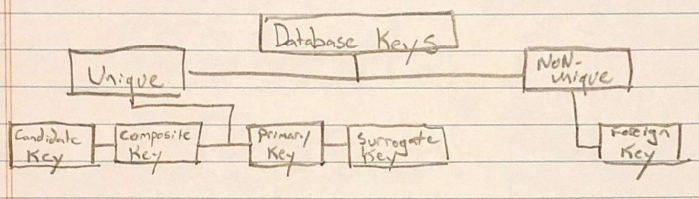
SQL - Structured query Language
ETL - Extract, transform, Load

4 Components of a database system are:
- Users
- Database Applications
- Database management System (DBMS)
- Database

- A Database is a self-describing collection of related records
- Tables within a database are related

The DBMS (Database management system) - serves as an intermediary between database applications and the database

A relation is a two-dimensional table
columns are attributes
intersection between row and colum are called cells
No two rows within a table can be identical



- A composite key is a key that is composed of two or more attributes
- A candidate key is called "candidate" because it has the potential to become the primary key. It is a unique key
- A Primary key is a candidate key chosen to be the main key for the relation
- A surrogate key is a unique, numeric value that is added to a relation to serve as the primary key. They have no meaning to users
- A Foreign key is a primary key from one table that is placed into another table
- A Null value means that no data exists
- A Functional Dependency is a relationship between attributes in which one attribute determines the value of another attribute in the same table

Data Normalization Video 2 of 8 around 35 mins in

A process of analyzing a relation to ensure that it is well formed i.e absent of the anomaly (Deletion, update or insertion)

As a general rule, a well-formed relation will not encompass more than one business concept.

Table with multivalued Attributes

↓

Remove multivalued Attributes

↓

First Normal Form

↓

Remove partial dependencies

↓

Second Normal Form

↓

Remove transitive dependencies

↓

Third Normal Form

SQL Data Definition

Create - Creates database objects

Alter - Modify the structure and/or characteristics of existing database objects

Drop - to delete existing database objects

Count - Counts the number of rows that match the specified criteria

Min - Finds the minimum value for a specific column matching specified criteria

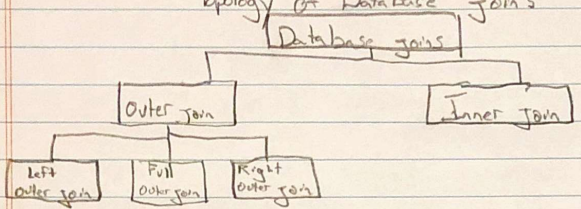
Max - Finds the maximum value for a specific column matching specified criteria

SUM - Calculates the sum (total) for a specific column for those rows matching the Criteria

AVG - Calculates the numerical average (mean) of a specific column for those rows matching the Criteria

STDEV - Calculates the Standard deviation of the values in a numeric Column whose rows match the Criteria

Topology of Database joins



one-and-only-one —||

one-to-Many —|<

zero-to-One —O|

zero-to-Many —O<

- Functional dependency - The relationship (within the relation) that describes how the value of one attribute may be used to find the value of another attribute
- Determinant - An Attribute that can be used to find the value of another attribute in the relation

The Three most critical database administration functions are:

1) Concurrency control
2) Security
3) Backup and Recovery

- Concurrency - people or applications can access the same information at the same time

Dirty read - Transaction reads a modified record that has not yet been committed to the database
Transient read - Transaction re-reads a dataset and finds that the data has changed
Phantom read - Transaction re-reads dataset and finds new record has been added

Implicit locks are issued automatically by the DBMS based on an activity

Explicit locks are issued by users requesting exclusive rights to specified data

- table locks
- Row Locks
- Column Locks
- Cell Locks

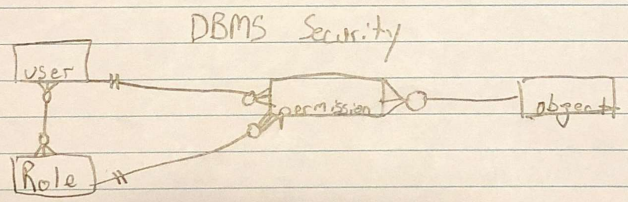
Dead lock - when two processes lock resources at the same time

Consistent transactions are often referred to by the acronym ACID

Atomic
Consistent
Isolated
Durable

A cursor is a point into a set of rows that is the result set from a SQL Select statement

Processing Rights define:
- Who is permitted to perform certain actions
- When certain actions are allowed to be performed



BI systems fall into two broad Categories:

- Reporting systems that Sort, Filter, group and make elementary calculations on operational data.
- data mining applications that perform sophisticated analyses on data; analyses that usually involve complex statistical and mathematical processing

Operational Database

- Used for structured transaction data processing
- Current data are used
- Data are inserted, updated and deleted by users

Dimensional Database

- Used for unstructured analytical data processing
- Current and historical data are used
- Data are loaded and updated systematically, not by users

Data Mining Techniques

- Cluster analysis - Identifies groups of entities that have similar characteristics.
- Decision tree analysis - Classifies entities into groups based on past History.
- Regression - Produces mathematical equations that can be used to predict Future events based on past observations
- Neural Networks - Use training data to learn how to create accurate predictions / estimations
- Market Basket Analysis - Determines patterns of associated buying behavior

- Most data that can be encountered are best classified as semistructured
- To reveal meaning, information requires context
- The organization of data within Folders in a manual File system is determined by its expected use
- An XML database supports the storage and management of semistructured XML Data
- A workgroup database supports a relatively small number of users or a specific department within an organization.
- Knowledge is the body of information and facts about a specific subject.
- Data constitutes the building blocks of information
- Unstructured data exist in the format in which it was collected,
- Data inconsistency exists when different and conflicting versions of the same data appear in different places
- A data dictionary contains at least all of the attribute names and characteristics for each table in the system.
- Information is the result of processing raw data to reveal its meaning
- Performance tuning refers to activities that make a database operate more efficiently in terms of storage and access speed
- XML, Extensible Markup Language is a special language used to represent and manipulate data elements in a textual format
- An application might be written by a programmer or it might be created through a DBMS utility program.
- Data is said to be verifiable if the data always yields consistent results.
- Accurate, relevant, and timely information is the key to good decision making.

One disadvantage of a database system over previous data management approaches is increased cost

End-user data is raw facts of interest to the end user

Analytical databases focus primarily on storing data used to generate information required to make strategic decisions.

A query is a specific request issued to the DBMS for data manipulation

The term islands of information refers to scattered locations storing the same basic data.

Metadata provide a description of the data characteristics and the set of relationships that link the data found within a database.

An ad hoc query is a spur-of-the-moment question.

A file is a collection of related records

DDL Commands

- Drop
- Alter
- Rename
- Create
- Truncate

Ternary Relationship involves three different entity types

Normalization - the process of organizing the fields and tables of a relational database to minimize redundancy and dependency

DML Commands

- Update
- Delete
- Insert
- Merge
- Select

- Weak Entity is also called a dependent entity
- Aggregation - is used for populating summaries that can be used at the staging area of ETL

Data warehouse
- Centralized repository of information
- Organized around relevant subject areas
- Provides platform for queries
- Used for analysis and not transactional processing
- Data is non-volatile
- Target location for integrating data from multiple sources

Every non-Key must provide a fact about the Key, the whole Key, and nothing but the Key
- "the Key" Ensures 1NF
- "the whole Key" Ensures 2NF
- "Nothing but the Key" Ensures 3NF