CSC 365 Vocab Notes

Database: A collection of information, often stored in tables and the tables' rows and columns.

Column: Defines the type of data that can be stored.

Row: An entry in the database. Comprised of multiple fields defined by the table's columns.

Structured Query Language (SQL): The language used to communicate with a database. An implementation of the theoretical Relational Algebra (see below). Defined and in use since 1970s.

SQL Injection: A method of attacking software in which valid SQL statements are included as input into a program and then executed on the database. Ex. Bobby Tables.

Data Definition Language (DDL): Statements that define the rows, columns, and tables in a database.

Data Manipulation Language (DML): Statements the modify data existing within the tables of a database.

Database Management System (DBMS): Software for creating and managing databases.

B-Tree: Data structure used by databases to store large quantities of data within a tree structure with relatively short height.

Query Optimizer: Program that employs relational algebra to speed up query operations.

Key / Index. Used by DBMS to quickly navigate the B-Tree data structure and find requested data.

Clustered Index: Each table may have more than one index, but only one clustered index. Clustered indices are stored 'ideally' sequentially on disk and thus faster than regular indices.

Primary Key: The column of table that is given the speed boosting clustered index.

Foreign Key: A column within a table that acts as a pointer to data existing within another table for the purpose of enabling joins.

Constra

Cartesian Product: The combining of all possible combinations between two sets.

Join (Inner): The operation of combining two tables into a single larger table through a cartesian product. Results are often narrowed down through one or more conditions, primarily the pairing of foreign keys.

Subquery: Queries contained within other queries. This can be useful for generating inline views, generating scalars for use in an update clause.

Inline View: The use of a subquery to generate a table that is joined with the tables from the outer query. Must be aliased.

Outer Join: Three variation **(Left, Right, Full)**. Similar to a inner join, except that depending on which of the three variations, guarantees that certain matching will occur, even if data must be fabricated.

Left Join: Most common. Protects the left table and creates null values in the right table.

Right Join: Protects the right table and creates null values in the left table.

Full Join: Protects neither table, generating null values in either table as necessary.

Epoch: A moment in time used as a reference point. For computers, this is commonly accepted as January 1, 1970 00:00:00:00 UTC and calculated as the number of milliseconds since that moment.

Entity Relationship Diagram (ERD): Diagram that demonstrations the structure of a database and the relationships between the tables.
Junction Table: A method to illustrate a many-to-many relationship. (N-M relationship). Consists of unique combinations of foreign keys. An example would be a table called AuthorXBook, in which an author may have multiple books, and a book may have multiple authors.
Collation Order: The cultural specific manner in which characters are ordered. Not everyone sings their abc's the same way.
Don't Repeat Yourself (DRY): Common acronym for reducing repetitive code. "This project needs DRYing out."
Normalization: The removal of redundancies within a database.
Transaction: The method of combining SQL statements into an atomic block, in which either all statements are executed on the live database, or none are. Either committed or rolled back.
Commit Transaction: Execute all the statements contained within a transaction. Difficult to reverse without restoring a backup or reconstructing the database.
Rollback Transaction: Revert the database to its state before the execution of the SQL statements contained within the transaction. SQL's only equivalent of undo or CTRL-Z.
DataTypes within MySQL

Integer: Standard 32 bit integer. Has variations of Tinyint, smallint, mediumint, bigint (1,2,3,8 bytes)

Unsigned: modifier that prevents numerical values from being negative. Useful for reducing human error.

Decimal: Numerical value required by law to be used in the monetary actions that performs mathematical operations (addition, subtraction, multiplication, division, etc) as humans do: in base ten. Required for its use of precision.

Float: Numerical value to represent numbers with values have the decimal point.

Double: Numerical value to represent numbers with values have the decimal point with greater precision than a float. Useful for operations where precision is important. Ex. scientific experiments.

char(n): string of characters of a strict length of n. If string is less than n long, the remaining space will be padded with zeros.

varchar(n): string of characters up to a maximum length of n. The value of n is stored along with the string and the size of n is dependant upon the number of bytes needed to store n (less than 255 will use one byte). Any unused space is not padded with zeros.

Binary Large Object (BLOB): Long string of binary values.

Character Large Object (CLOB): Long string of character values. Different than a blob because characters can include Unicode.

Tinytext, text, mediumtext, long text: Character strings of different lengths, basically CLOBs

Datetime: 'yyyy-mm-dd hh:mm:ss' represents local date and time.

Timestamp: date time under a specific timezone locale.

Java Database Connectivity Driver (JDBC Driver): Interface standard that defines necessary functions for connecting to databases and performing database operations through a Java application. There are four different types of implementations.

Type 1: A general driver. Designed to work with multiple DBMSs. Will be unable to take advantage of specific features of each DBMS.

Type 2 - 3: Intermediates between 1 and 4.

Type 4: A driver designed specifically for a single DBMS. Capable of using the specific features of that DBMS.

Connecting to MySQL: Pass the following arguments into DriverManager.getConnection(). jdbc:mysql://<ip address. Use localhost if connecting locally>:<port. default:3306/<database name>?useSSL=false <database username> <database password>

Connection Pool: The maintenance of multiple open connections by a database server in order to reduce the functional overhead of opening and closing connections for each individual session. When a session is done with a connection, the session performs clean up operations and then the connection returns to the pool of available connections for the next session to make use of.

Statement - Java Class used to make SQL queries. Vulnerable to SQL Injection.

PreparedStatement - Java Class used to make SQL queries. Sanitizes all input and is thus immune to SQL Injection.

ResultSet - Java Class that represents the result of a SQL query.

Autoboxing - Java's automatic conversion between primitive data types and their Object equivalents. Enables ease of operations like Integer(3) == int 3;

ACID: Atomic Consistent Isolated Durable

Atomic: An action is either or all or nothing

Consistent: All transactions leave the database in a consistent state: Ex. money transfers between accounts in a bank system leave the total amount of money in the same as before the transfers

Isolated: How processes that are accessing/modifying the same data behave

Read Uncommitted: Uncommitted modifications made by one process are visible to other processes. Creates 'dirty reads'. Cheap and fast, but unreliable. Last to commit wins

Read Committed: Only committed modifications made by one process are visible to other processes for reading. No row locking. Whichever process commits last wins. Prevents dirty reads, but still permits processes significant interference over each other.

Read Repeatable: All data read during a process is promised to remain consistent from the start of the transaction until commit. MySQL cheats by giving a working copy of the data unless 'for update' is specified. If 'for update' or an update statement is used, the targeted rows are locked and other processes cannot modify the live data. Problem of Phantom Reads, Deadlock. This is MySQL's default level of isolation, without the for update specification.

Serializable: Two types of row locking: Shared lock (read access) and exclusive lock (write access). More than one process may take advantage of a shared lock and read data under the locked rows. However, if one process aims to modify the data, an exclusive lock is needed and can only be used by one process. If an exclusive lock is sought on rows under shared lock among more than one process, the exclusive lock will block until other processes no

longer need read access. When more than one process seeks an exclusive lock on the same rows, deadlock will occur.
Durable: All committed transactions are permanent and will remain a part of the data set even under the events of system failure (from loss of power or otherwise).
Dirty Read: Data that has not yet been written to disc (permanent storage) and is currently held in volatile memory.
Phantom Reads: When rows that were previously absent under read repeatable are added to the data set.
Relational Algebra (RA): Branch of mathematics that organizes data within relational databases and the methods for querying the contained data.
Relation: An ordered collection of data represented in tables. Similar to the definition of a set of tuples for Discrete mathematics. No repetitive data and no enforced order.
Attribute: The set of possible values for a column.
Tuple: A set of values, one for each attribute. Equivalent to a row.
Projection: Removal of one or more attributes (columns) from a relation. Taking a higher dimensional space and mapping into fewer dimensions. If an object was three dimensional, a projection would be its shadow, represented by (x,y) coordinates, without the z coordinate.Symbol (lowercase pi)
Selection: Takes a subset of tuples, equivalent to a WHERE or HAVING clause, depending on where it occurs in relation to a group by. Symbol (lowercase sigma).
Unconstrained Join: A cartesian product. Represented by 'x'. Ex. R x S
Theta Join: Called a constrained join, equivalent to an inner join. Represented by the bowtie symbol and conditions.
Natural Join: Constrained join on all same-named attributes between two relations. Just a bowtie.
Outer Joins
    Left Join: Bowtie L or bowtie with extra pointing left.
    Right Join Bowtie R or bowtie with extra pointing right.
    Full Join: Bowtie with extra point both left and right.
Renaming Operation: Creates an alias. Represented by rho. Ex. rho <alias> (attribute[s]) (Relation)
Group By Operation: Aggregation. Symbol is lowercase gamma.
Order By Operation: Converts the relation into an ordered list. Must be the last operation. Represented by lower case tau.
Set Operations: Union, Set subtraction works too.
Key: a subset of attributes in which all other attributes are functionally dependant. If you know the values for the key, you can determine the values for the attributes.
Minimal: A condition of a key in which it requires all attributes to determine the values for other attributes. Requires no more attributes and can get by with no less.
Superkey: A set of values that uniquely identifies a tuple, but does not necessarily follow the minimal condition. All keys are superkeys but not all super keys are keys. A simple super key is the use of all of the attributes.

Candidate key: A largely redundant term, but a key that is likely to be considered to be used as a primary key and thus receive a clustered index.
Expression Tree: Technique for mapping operations so that the order of operations is in no way ambiguous. Easily viewed in regular numerical operations, but also extends to relational algebra operations. Useful for generating operations that return equivalent relations but in a quicker manner.
Selection Push: Technique of moving a selection operation further down an expression tree so that the operation reduces the number of tuples early on and thus reduces future computation. Big benefit.
Project Push: Technique of moving a projection operation further down an expression tree so that the operation removes attributes early on and reduces future computation. Marginal benefit.
Join Reassociation: Reducing the total number of joins necessary to complete an operation by reorganizing the order in which they happen.
Binary Operator: Operation that requires two operands. Ex: Joins
Unary Operator: Operation that requires one operand. Ex: Projection, selection, rename, group by, order by.