SOFTWARE GUILD

Intro to Databases

.NET Cohort

Coding Bootcamp

SOFTWARE GUILD

Lesson Goals

- Learn the basic terminology around relational databases

About SQL

- SQL stands for "Structured Query Language" and it is the most used database language today. SQL is an ANSI standard, so while different vendors databases implement some special commands, the basic SQL commands work across all databases that are compliant.

- SQL Server is the database environment and the SQL language is the tool set.

What is a Database?

A database is a collection of related information

- It can be mundane things:
  - An address book
  - A telephone book
  - A report card
  - A check book
  - A family tree
  - A diary

- It can be complex things:
  - An order tracking system
  - A web store
  - Every phone call and email you've ever made (thanks NSA!)

SOFTWARE GUILD

Let's start simple though

ItemID	ItemDescription	Price
1	Milk	$1.00
2	Cereal	$3.50
3	Bread	$0.99
4	Mountain Dew	$4.99
5	Chips	$0.75
6	Hamburgers	$7.99

- Which is the most expensive item? Which is cheapest? What is the average price?
- In a simple example like this, we can quickly do it by hand, but with SQL we can process millions of records with sub-second response times!

SOFTWARE GUILD

Some basic terminology

ItemID	ItemDescription	Price
1	Milk	$1.00
2	Cereal	$3.50
3	Bread	$0.99
4	Mountain Dew	$4.99
5	Chips	$0.75
6	Hamburgers	$7.99

- The horizontal values are organized into *rows* (or *records*), the vertical values are organized into *columns* (or *fields*).
- The collection of rows and columns is called a *table*.
- We could say that this table is populated with six records.
- An individual cell is called a *value.*
- When we return data in this format, it's called a *result set.*

SOFTWARE GUILD

About Nulls

AddressID	Address1	Address2
1	100 Oak St	Apt 305
2	404 Elm St	NULL
3	123 Main St	NULL
4	526 Main St	Suite 714B

- Sometimes data in a database field is optional. In this case it can be NULL.
  - NULL is a special keyword that means "unknown" or "undefined."
  - We will visit this in detail later, but the upshot is that because a null is unknown, it can never be used as a comparison or in a calculation.

Database Management Systems (DBMS)

- Database Management Systems (like MySQL) are responsible for protecting the data they contain.

- They manage the backup and recovery process as well as enforcing the rules and constraints of the system as defined by the Database Administrators (DBAs)

- Most databases have many tables which often relate to each other. For this reason, we also refer to them as Relational Database Management Systems (RDBMS)

Database Engine Duties

- The Database Engine is responsible for the following:
  1. Provide reliable storage for data
  2. Provide a means to rapidly retrieve the data
  3. Provide consistent access to the data
  4. Maintain data Integrity
  5. Control access to data with security

Reliable Storage

- Reliable storage starts at the hardware level. So, a Database Administrator (DBA) often needs to understand how to structure their hardware.

- A transaction log makes a record of every change that is made to a database. **It is critical that the transaction log be backed up as well as the main database file**.

Reliable Storage: Illustrated

1. A SQL statement is sent to the engine.
2. Data changes are written to the transaction log.
3. The changes are made to the data file.
4. If the changes succeed, a commit is written to the log. If they fail, a rollback occurs.
5. The engine reports status or returns data back to the caller.

| SQL Statement | → ← | Database Engine | → ← | Transaction Log | → ← | Database File |
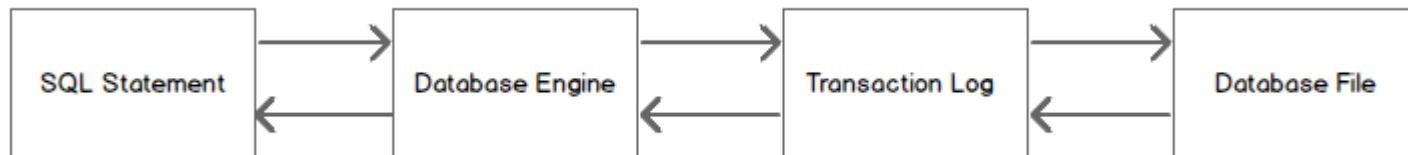
Rapid Data Access

- Databases allow for the creation of *indexes*, enabling fast data access.

- When you query a table with no indexes, your database has to read each and every page of data looking at every row to see if it satisfies your search. This is called a *table scan* and for large data sets, it is bad.

SOFTWARE GUILD

Index Structures

- Indexes are stored separately from the actual data pages — they only contain pointers to data pages or rows. Indexes in SQL Server are *balanced trees* (B-Trees).

Level 2 (Root)

Level 1 (Intermediate)

Level 0 (Leaf)
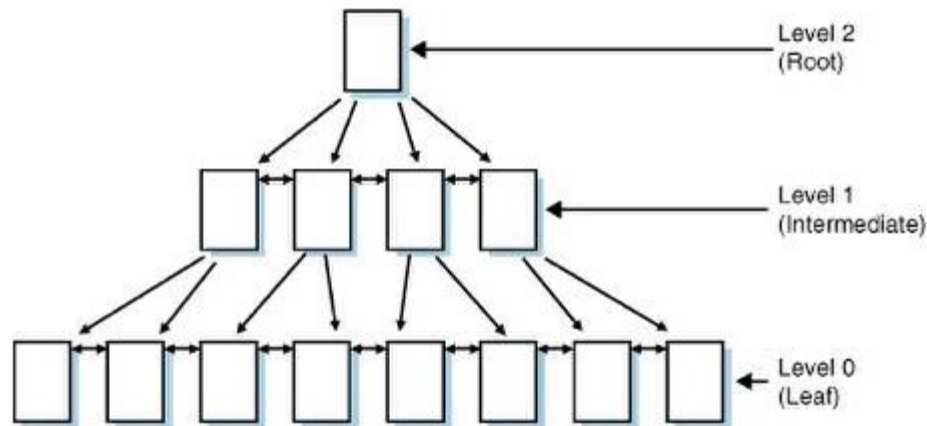
B-Trees?

- Say we want to index the LastName of a person record.

- Think of the levels of the tree as file drawers, each in alphabetical order. Each level points to the next index.

- By starting at the top of the tree and working our way down alphabetically, we can avoid searching all of the records, greatly increasing our efficiency.

Simplified Index
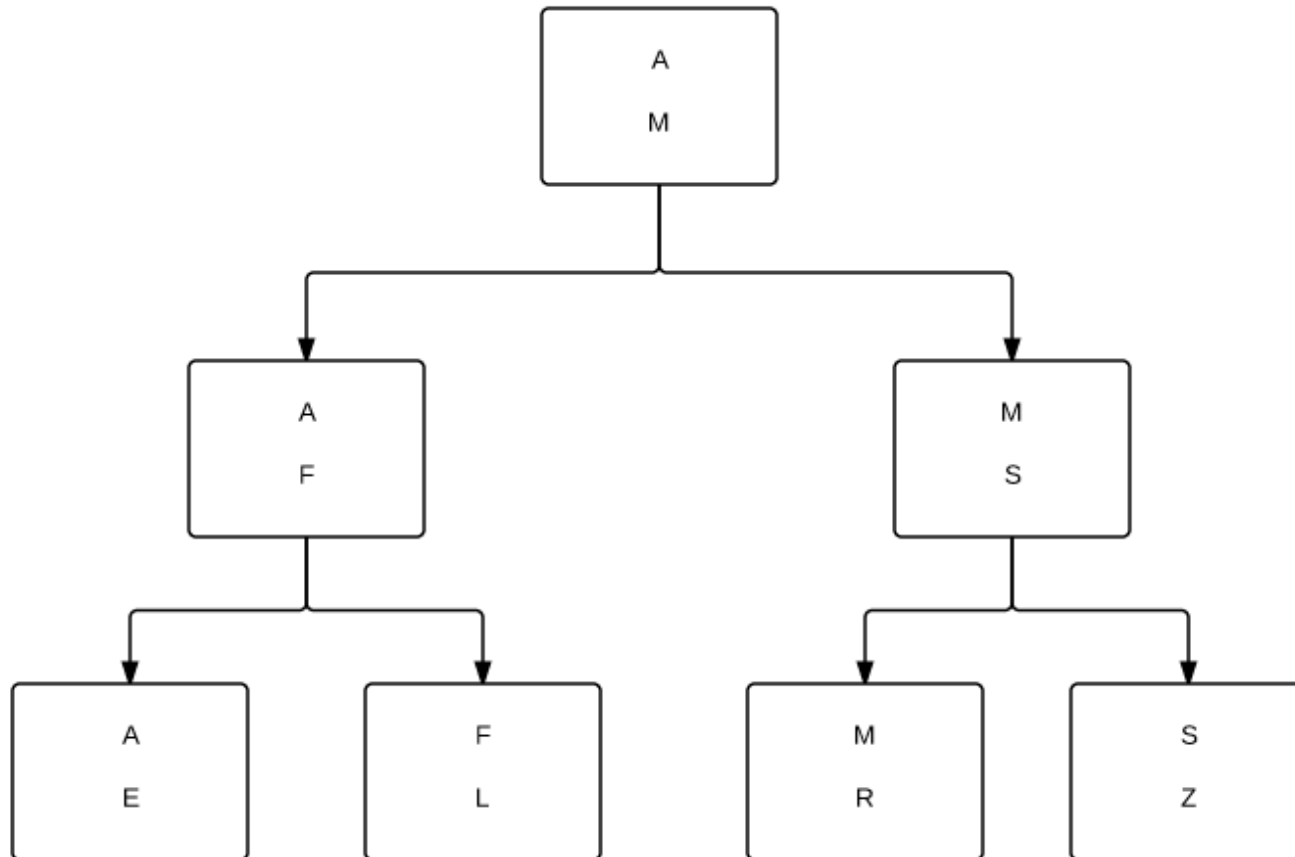
Consistent Data Access

- Getting to data fast means nothing if the data is not accurate. The general idea of consistent data access is to only allow one client at a time to change data **and to prevent others from reading data while it is being changed**.

- This is called *locking.* When two processes try to change the same data concurrently in a way that prevents each from completing, it is called a *deadlock.*

SOFTWARE GUILD

Data Integrity

- When we set up table definitions in a database, we can provide standards that the database will enforce.

- The simplest example is making certain columns of data required, and enforcing types on those columns (int, character, etc.).

- The relational example is if we have a table with customer data, and another table with addresses for those customers, we can set rules that prevent a customer from being deleted unless all addresses associated with them are also deleted.

Security

- Most major databases allow for the creation of login/password information to access data (or plug into your corporate network's security to use your network credentials).

- Normally, a DBA will create *roles* which are granted privileges to database objects (read, write, delete) and procedures.

- Users are added and removed from roles by the DBAs and it prevents DBAs from having to assign each individual user security privileges. Adding users to roles gives them buckets of privileges.

SQL Server Runs as a Service

- By default, SQL Server starts up when the machine boots and runs as a service (we can see it in the services utility, it actually has several services).

- These services run constantly in the background so they are always ready to handle data requests.

Terms You Should Know

- **Column**: A field in a database table.   (AKA field)
- **Database**: A collection of objects to store and retrieve data.
- **Database Management System**: A set of software components and programs that creates, maintains, and controls access to a database.
- **Data type**: An attribute of a field that tells SQL Server what kind of data it may accept. Examples include integers, dates and characters.
- **NULL**: Unknown or unspecified values.

Terms You Should Know (con't)

- **Query**: A question asked to get information from data in a database.
- **RDBMS**: Relational Database Management System is a tool that allows for safe storage of data and quick retrieval of information.
- **Record**: A row of data in a table.
- **Record Set**: The set of data returned as an answer to a query. (AKA result set)
- **Service**: A process, much like an application, that runs in the background of your system.
- **SQL**: Structured Query Language.

SOFTWARE GUILD

Fin

- Next up: Basic Queries!