BlackTECH Data Analytics School

BlackTECH

Structured Query Language in Data Analytics

Introduction to Structured Query Language

What We Hope To Cover

1. My Learning Journey
2. Learning Outcomes
3. Before we Begin
4. Databases: The Big Picture
5. Relational Databases (RDBMS)
6. Data Exploration using SQL Server (T-SQL)
7. Writing SQL Queries with SQL Server Management Studio (SSMS)
8. Links to Resources
9. Questions

My Learning Journey (facilitator)

A proactive Business Intelligence analyst with over 15 years MS SQL Database Development / analyst role in financial and telecommunication space, T-SQL programming, Relational and Multi-Dimensional Data Modeling, ETL Development , Data Visualization and Process Optimization

EDUCATION & CERTIFICATIONS

- B.Sc. Biochemistry
- B.Sc. Computer science (In View)
- Microsoft Certified Professional (MCP)
- Microsoft Certified Solution Developer (MCSD)

SOME PAST PROJECTS I HAVE INVOLVED IN

- Absa bank International Banking MIS
- Absa Bank Retail bank Transactional Highlander Project ETL, Mart , Reporting Development
- MTN Prepaid retention Report and Dashboard
- Absa Customer Value management reporting (Naxian Project)

EXPERIENCE

- 5 Years IT Training: Damelin, Boston College Johannesburg
- 2 Years SQL Developer, DBA, Data Warehouse Admin Mayibuye Group
- 2 Years MIS Analyst: MTN SA, Johannesburg
- 12 Years Data Analyst (SAS, SQL): Absa Bank (Former Barclays Bank Africa)
- Currently Data Analyst (Santander Consumer Edmonton Canada)

My Learning Journey (Co-facilitator)

A lifelong learner and Software Engineer with bias to the .NET Stack, I have had the opportunity to work on projects leveraging Several Databases including Microsoft SQL Server Database in the last 9 years.

EDUCATION & CERTIFICATIONS

- Certificate in Big Data Analytics @ York University
- M.Sc. Information Technology
- B.Eng. Electrical & Electronics Engineering
- Oracle 10g Database Administrator Certified Professional (OCP)
- Microsoft Certified Professional (MCP): Querying Microsoft SQL Server 2012/2014
- Microsoft Certified Azure Developer Associate

EXPERIENCE

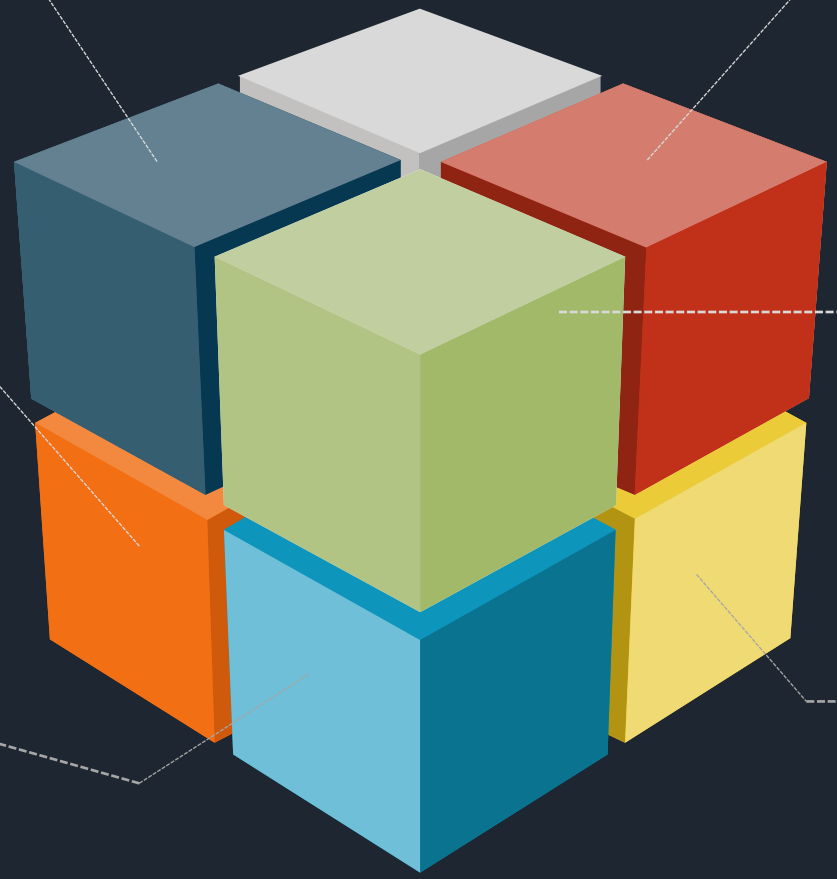
- 2 Years IT Training: FirstLogic, AfriHUB
- 6 Years .NET Developer and DBA: Federal Inland Revenue Service (FIRS)
- 2 Years .NET Developer: Rian Consulting, Toronto
- 2 Years .NET Developer: LexisNexis Canada

SOME PAST PROJECTS I HAVE INVOLVED IN

- FIRS Tax Collection dashboard using SQL Server Reporting Services (SSRS)
- FIRS 2016 recruitment Portal
- Integration Projects with Nigerian Customs Service (NCS), Treasury Single Account (TSA), NIBSS, Remita, Interswitch, Autoswift, e-Tranzact
- Enterprise Logistics Management System

Learning Objectives

- Have a high-level overview Databases
- Understand the Different types of Databases
- Understand the concept of Relational Databases
- Understand Structured Query Language (SQL)
- Manipulate Data using SQL
- Learn advanced SQL Concepts on your own



Before We Begin

Nice to Know / Have

- English Language
- Open Mind

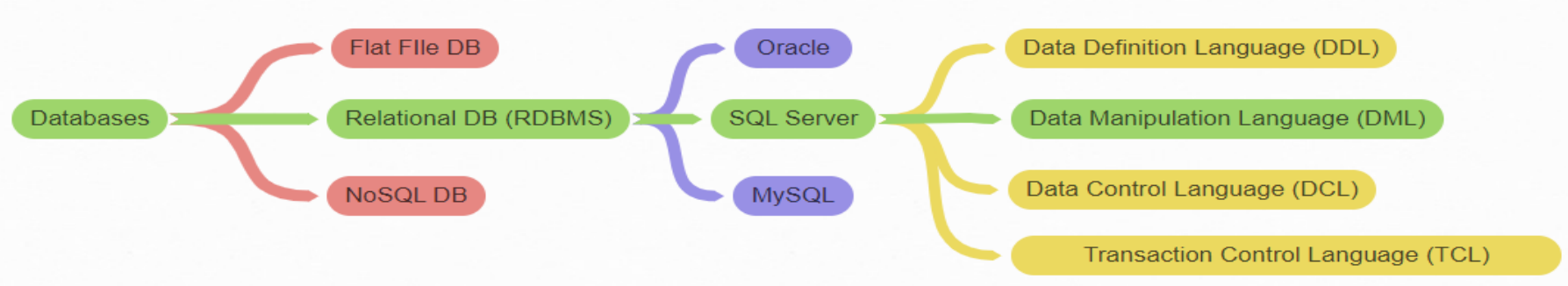
My Approach to Learning a New IT Skill

- Learn the basics
- Focus more on understanding the capabilities and limitations than knowing everything, that's why you have documentation, google & Stack overflow
- Get your hands dirty

Databases: The Big Picture

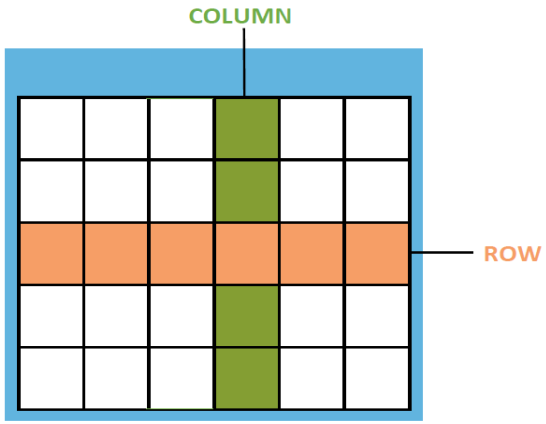
A **database** is an organized collection of data, generally stored and accessed electronically from a computer system (Wikipedia)

Database Type	Example
Flat file Databases	Excel, CSV
Relational Database Management Systems (RDBMS)	MySQL, SQL Server, Oracle
No SQL Databases	Mongo DB, Cosmos DB, Neo4J



Relational Databases (RDBMS)

- RDBMS Stores data in Tables which can be related to each other
- Tables are database objects and consist of rows, columns and relationships
- Columns hold Specific Data types: (String, Numeric, Date and Time, Others)
- While Rows/Records hold information about an observation
- Indexing is a data structure technique which allows you to quickly retrieve records from a database
- Contains other Database Objects (Stored Procedures, Views, Triggers, Synonyms)
- Information about Data stored in the DB is called Metadata
- Database Schema is a blueprint of how the DB is organized



Relational Databases (RDBMS)

Tables also have constraints which are used to specify rules for data in a table, constraints could be at table level or column level

Database Type	Example
Primary Key (PK)	Uniquely identifies every row in a table and cannot be empty (Unique + NOT NULL)
Foreign Key (FK)	References the Primary Key column in another table, relationship can be *1- to-1, 1-to-Many*
Unique Key	Cannot have duplicate values in a column
Composite key	A table level constraint that makes a combination of 2 columns the primary key
NOT NULL	Column cannot have an empty value
Check	used to limit the value range that can be placed in a column

Overview of Data Normalization

There are three basic steps in data normalization namely

- First Normal form 1NF
- Second Normal form 2NF
- Third Normal form 3NF

1. First Normal Form (1NF)

1NF ensures the data must be atomic, no repeating data in a group. To be considered 1NF, each entry must have only one single value for each cell and each record must be unique.

For example, you are recording the name, address, gender of a person, and if they bought cookies.

2. Second Normal Form (2NF)

The data must be in 1NF, working to ensure no repeating entries. Following that, data must have only one primary key. To separate data to only have one primary key, all subsets of data that can be placed in multiple rows should be placed in separate tables. Then, relationships can be created through new foreign key labels.

For example, you are recording the name, address, gender of a person, if they bought cookies, as well as the cookie types. The cookie types are placed into a different table with a corresponding foreign key to each person's name.

3. Third Normal Form (3NF)

For data to be in 2NF requirements. Following that, data in a table must only be dependent on the primary key. If the primary key is changed, all data that is impacted must be put into a new table.

For example, you are recording the name, address, and gender of a person but go back and change the name of a person. When you do this, the gender may then change as well. To avoid this, in 3NF gender is given a foreign key and a new table to store gender.

Data normalization is an iterative process forms, the rules will become clearer while separating your data into tables and levels will become effortless. These tables will then make it simple for anyone within an organization to gather information and ensure they collect correct data that is not duplicated.

Benefits of data normalization

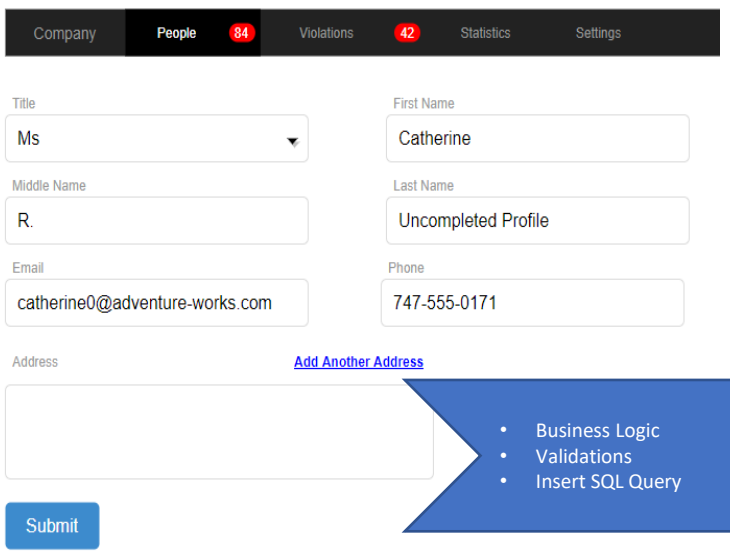
- More space as duplicates are eliminated
- Faster question answering
- Better Data segmentation
- Enhances Data Integrity

Some Drawbacks

- Can get overly complex
- Redundancy When storing information with different requirements

Relational Databases (RDBMS)

Data can be added to the Database through Application forms, APIs, direct SQL queries and bulk insert from Excel and other flat file sources among others



- Business Logic
- Validations
- Insert SQL Query

Customer Table

100 % | Results | Messages

	CustomerID	Title	FirstName	MiddleName	LastName	EmailAddress	Phone
1	29485	Ms.	Catherine	R.	Abel	catherine0@adventure-works.com	747-555-0171
2	29486	Ms.	Kim	NULL	Abercrombie	kim2@adventure-works.com	334-555-0137
3	29489	Ms.	Frances	B.	Adams	frances0@adventure-works.com	991-555-0183
4	29490	Ms.	Margaret	J.	Smith	margaret0@adventure-works.com	959-555-0151
5	29492	Mr.	Jay	NULL	Adams	jay1@adventure-works.com	158-555-0142
6	29494	Mr.	Samuel	N.	Agcaoili	samuel0@adventure-works.com	554-555-0110
7	29496	Mr.	Robert	E.	Ahlering	robert1@adventure-works.com	678-555-0175
8	29497	Mr.	François	NULL	Ferrier	françois1@adventure-works.com	571-555-0128
9	29499	Ms.	Amy	E.	Alberts	amy1@adventure-works.com	727-555-0115

CustomerAddress Table

100 % | Results | Messages

	CustomerID	AddressID	AddressType
1	29485	1086	Main Office

Address Table

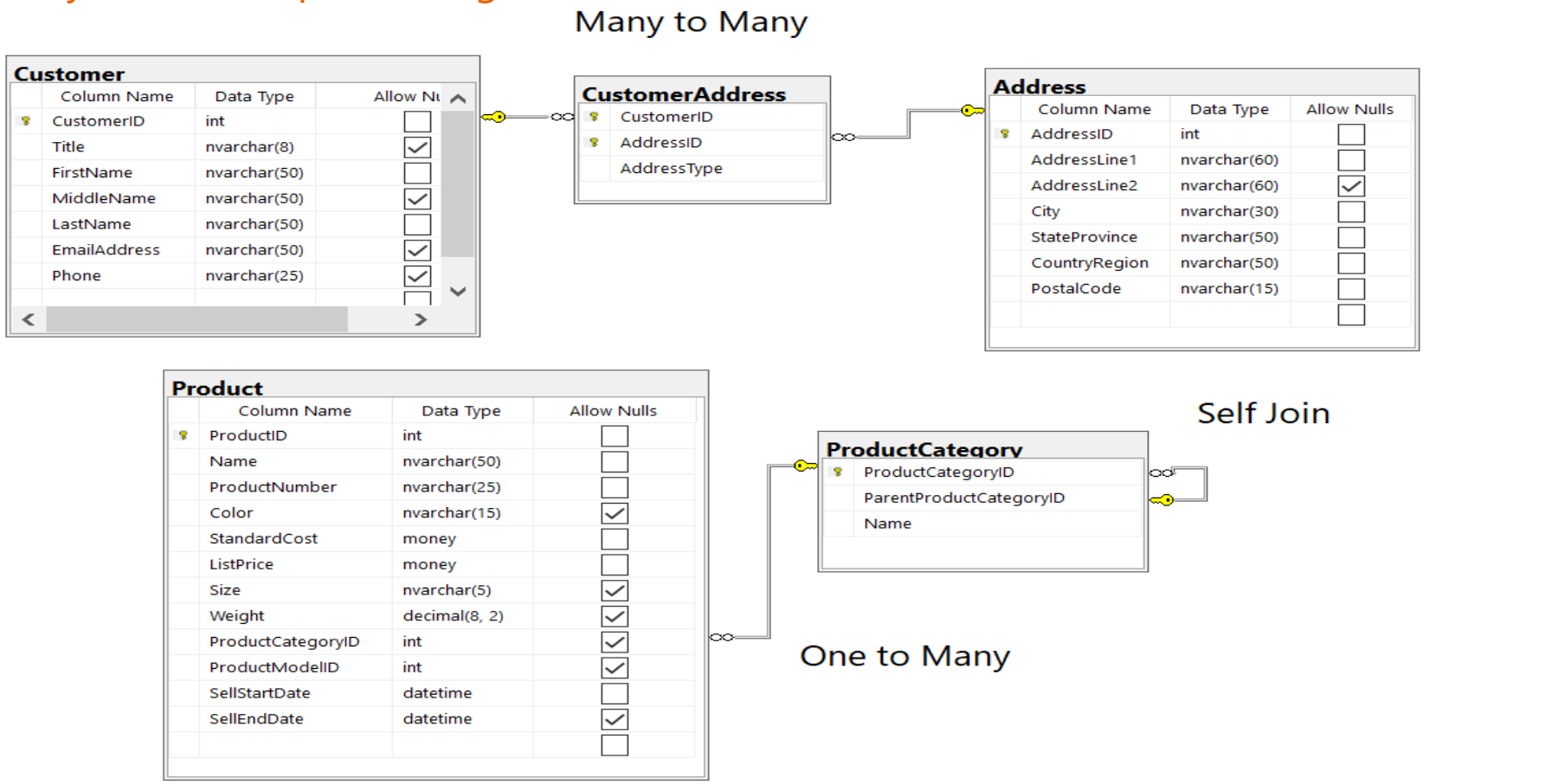
100 % | Results | Messages

	AddressID	AddressLine1	AddressLine2	City	StateProvince	CountryRegion	PostalCode
1	1086	57251 Serene Blvd	NULL	Van Nuys	California	United States	91411

Relational Databases (RDBMS)

A Diagram showing how the tables relate is called and

Entity Relationship (ER) Diagram



Data Exploration Using SQL Server

Data Exploration and Preparation takes up more than 50% of the workload in a Machine Learning Project

Microsoft SQL Server uses a flavour of SQL called Transact SQL (T-SQL)
SQL Server has 4 main components

1. Database Engine
2. SQL Server Integration Services (SSIS)
3. SQL Server Reporting Services (SSRS)
4. SQL Server Analysis Services (SSAS)

SQL Server Management Studio (SSMS) is Microsoft's IDE for managing SQL Server Databases

SQL Server Editions and Versions

SQL Server Versions are Tagged by release year:
e.g. SQL Server 2012, SQL Server 2014, SQL Server 2016, SQL Server 2019

SQL Server Editions are packaged based on Cost and Functionality

1. Enterprise
2. Standard
3. Web
4. Express

SQL Server can also be installed on Linux

IDEs and Software Installation

1. Ms SQL Server Express
2. SQL Server Management Studio
3. AdventureWorks Sample Database
4. Azure Data Studio

Introduction to SQL

Used to write queries against Relational Databases

SQL statements can either be Data definition language (DDL) statements or Data Manipulation language (DML) Statements.

DDL statements are used to create/alter database objects while DML statements manipulate data stored in the database

Language Type	Command
DCL	GRANT, REVOKE
DDL	CREATE, ALTER, DROP
DML	SELECT, INSERT, UPDATE, DELETE
TCL	BEGIN TRAN, COMMIT TRAN, ROLLBACK TRAN

Introduction to SQL

Sample DDL Statement

```
CREATE TABLE Department (
DepartmentId INT PRIMARY KEY,
DepartmentName NVARCHAR(250))

CREATE TABLE Employees (
EmployeeId INT PRIMARY KEY,
LastName NVARCHAR(250),
DepartmentId INT FOREIGN KEY REFERENCES Department (Dep
artmentId)
ON DELETE CASCADE);
```

Data Exploration Using SQL Server

Some important SQL Commands

Language Type	Command
WHERE	Used to filter rows based on a value
CASE, Searched CASE	Create a new column based on certain conditions
HAVING	Filters rows with aggregate functions
ORDER BY	Used for sorting
GROUP BY	Used with an aggregate function for grouping rows
AS	An alias for giving alternative names columns
INNER JOIN, OUTER JOIN(LEFT, RIGHT, FULL), CROSS JOIN	Used to join tables that have relationships

Data Exploration Using SQL Server

Operators

Comparison	Set	Conditional	Logical	Math
=	UNION	BETWEEN	AND	+
>	UNION ALL	IS NULL	OR	-
<	INTERSECT	LIKE		*
>=	EXCEPT	IN		/
<=				%
<>				

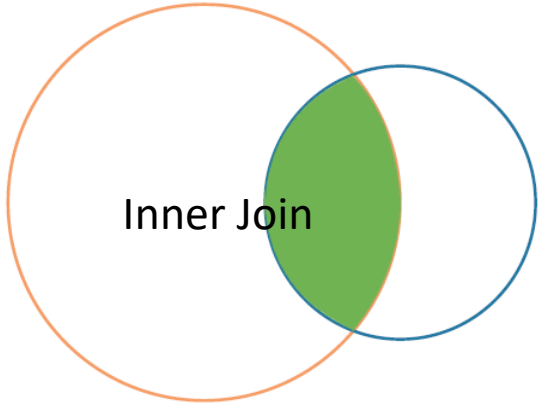
Data Exploration Using SQL Server

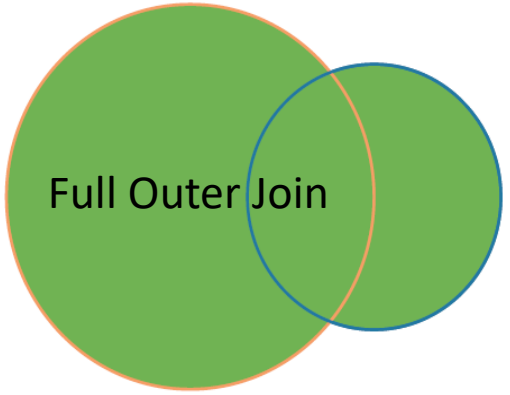
Wild Cards

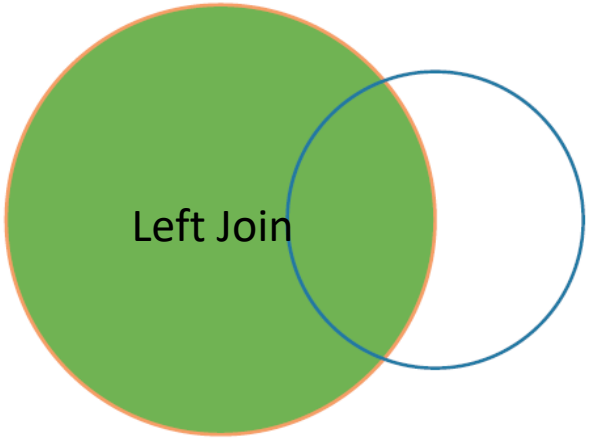
Symbol	Description	Example
%	Represents zero or more characters	bl% finds bl, black, blue, and blob
_	Represents a single character	h_t finds hot, hat, and hit
[ ]	Represents any single character within the brackets	h[oa]t finds hot and hat, but not hit
^	Represents any character not in the brackets	h[^oa]t finds hit, but not hot and hat
-	Represents a range of characters	c[a-b]t finds cat and cbt
*	Returns all columns	

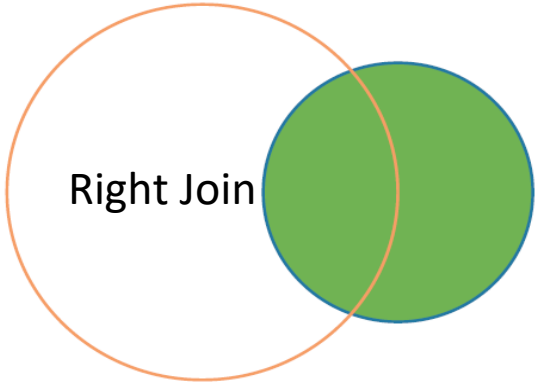
Data Exploration Using SQL Server

Understanding Joins









Data Exploration Using SQL Server

Some in-built Functions

String	Conversion	Date & Time	Logical	Aggregate	Window
LEN	CAST/ TRY_CAST	GETDATE	IIF	AVG	ROW_NUMBER
LTRIM/RTRIM	CONVERT/TRY_CONVE	DATENAME		SUM	RANK
CONCAT/ +	RT	DATEPART		MAX	DENSE_RANK
REPLACE	PARSE	DATEDIFF		MIN	LAG
UPPER/LOWER	ISNULL	DATEADD		COUNT	LEAD
SUBSTRING				DISTINCT	FIRST_VALUE
					LAST_VALUE

Data Exploration Using SQL Server

SQL Server Management Studio (SSMS) is one of Microsoft's IDE for working with the SQL Server Database

- Exploring SSMS
- Importing Data into SSMS
- Writing DML statements in SSMS
- Exporting Data from SSMS

Links to Resources

- Google it
- Stack Overflow
- YouTube
- https://www.sololearn.com/home
- https://www.w3schools.com/sql/default.asp
- https://www.tutorialsteacher.com/sql

Course outline

Week 1

- **Module 1: SQL Intro & Setup**
- Installing MSSQL and Azure Data Studio
- Restoring the AdventureWorks Database
- Understanding the Different Types of Databases
- Relational Database Management Systems
- Database Normalization
- Components, Versions & Editions of MSSQL Server
- Overview of Azure Data Studio
- **Practice Session**
- Using SELECT Statement
- Top Clauses

Week 2

- **Module 2: Single Table Analysis**
- **Module 1 Review**
- Introduction to writing SQL statements
- Types of SQL Statements (DML, DDL, DCL, TCL)
- Using SELECT Statement to Filter by Column
- **Practice Session**
- The FROM Clause - Tell the Server Which Table Your SQL Query Will Pull Data From
- **Assignment**

Week 3

- **Module 3: Analysis of Multiple Tables with Joins**
- **Module 2 Review**
- Row Filtering using the WHERE Clause
- Common Filtering Operators (AND, OR, IN, LIKE)
- Using Wildcards
- **Practice Session**
- Aggregating data using GROUP BY and HAVING Clause
- Sorting Data using the ORDER BY Clause

Course outline

Week 4

- **Module 4: Analysis of Multiple Tables with Joins**
- **Module 3 Review**
- Introduction to SQL JOINS and Querying Multiple Tables
- Writing Subqueries
- The UNION Operator
- The Union All
- Intersect
- Except

Week 5

- **Module 5: Analysis of Multiple Tables with Joins**
- **Module 4 Review**
- Introduction to Insert statement
- Introduction to Update statements
- Writing Subqueries
- Common Type expressions CTE

Week 6

- **Module 6: Analysis of Multiple Tables with Joins**
- **Module 5 Review**
- Updates with Subqueries
- Pivot Queries
- General Reviews

Connect With Us

www.blacktechhub.org

+1-647-927-0903

info@blacktechhub.org

https://www.linkedin.com/company/blacktechhub/

https://instagram.com/blacktech_hub

https://twitter.com/blacktech_hub


BlackTECH
Realize Your Vision
Technology
Entrepreneurship
Career
Hub...