SQL Course Curriculum for Data Engineering

Module 1: SQL for Data Engineering Introduction

Introduction to SQL and Data Engineering

- What is SQL and why it's essential for Data Engineering
- SQL in the modern data ecosystem
- Career opportunities with SQL skills
- Course roadmap and learning objectives

How to use SQL as a Data Engineer

- Real-world SQL applications in data pipelines
- SQL for ETL/ELT processes
- Database management and optimization
- Collaboration with data scientists and analysts
- Case studies from industry projects

Database History and Overview

- Evolution of database systems
- Types of databases (Relational vs NoSQL)
- SQL standards and variants (MySQL, PostgreSQL, SQL Server, etc.)
- Client-Server architecture overview
- Cloud databases introduction (BigQuery, Redshift, Snowflake)

Installation and Setup

- Installing PostgreSQL/MySQL
- Setting up database GUI tools (pgAdmin, DBeaver, MySQL Workbench)
- Cloud database setup (free tier options)
- Basic configuration and security
- Connecting to databases from different clients

Module 2: SQL Building Blocks

Important SQL Keywords/Functions Overview

- DML vs DDL vs DCL vs TCL
- Essential SQL clauses (SELECT, FROM, WHERE, etc.)
- Function categories (Aggregate, Scalar, Window)
- Reserved words and naming conventions

Data Types and Creating Tables

- Numeric types (INT, DECIMAL, FLOAT)
- Character types (VARCHAR, CHAR, TEXT)
- Date/Time types (DATE, TIMESTAMP, INTERVAL)
- Boolean and special data types
- CREATE TABLE syntax with examples
- Choosing appropriate data types

SQL for Data Engineering & Different Types of Queries

- Analytical queries vs transactional queries
- Batch processing queries
- Real-time query patterns
- Performance considerations

Creating Database in SQL

- CREATE DATABASE command
- Database configuration options
- Schema design principles
- Setting up development vs production databases

Module 3: Database and SQL Concepts

Operators in SQL

- Arithmetic operators (+, -, *, /, %)
- Comparison operators (=, <>, <, >, BETWEEN, IN)
- Logical operators (AND, OR, NOT)
- Pattern matching (LIKE, ILIKE, wildcards)
- Set operators (UNION, INTERSECT, EXCEPT)

ALTER Query

- Adding, modifying, dropping columns
- Changing data types and constraints
- Table renaming and schema changes
- Best practices for schema evolution

Database Constraints

- NOT NULL constraints
- UNIQUE constraints
- CHECK constraints
- DEFAULT values

● Constraint naming and management

Primary Key, Foreign Key & Other Keys

● Primary Key concepts and implementation
● Foreign Key relationships and referential integrity
● Composite keys
● Surrogate keys vs natural keys
● Indexes and their role in keys

Understand ACID Properties

● Atomicity in transactions
● Consistency guarantees
● Isolation levels
● Durability of transactions
● Real-world implications for data engineering

Understand Normalization

● First Normal Form (1NF)
● Second Normal Form (2NF)
● Third Normal Form (3NF)
● Boyce-Codd Normal Form (BCNF)
● DE normalization for performance

INSERT, UPDATE, DELETE Statements

● INSERT with VALUES and SELECT
● UPDATE with JOINs
● DELETE with conditions
● TRUNCATE vs DELETE
● Bulk operations best practices

JOINS (INNER, LEFT, RIGHT, FULL, CROSS)

● INNER JOIN deep dive
● LEFT/RIGHT JOIN scenarios
● FULL OUTER JOIN use cases
● CROSS JOIN applications
● Join performance optimization

SELF JOIN

- Hierarchical data queries
- Employee-manager relationships
- Finding pairs and sequences
- Practical self-join patterns

☐ ORDER BY, GROUP BY & HAVING CLAUSE

- Sorting with ORDER BY
- GROUP BY with multiple columns
- HAVING vs WHERE differences
- Grouping sets and rollups

☐ Aggregation Functions

- COUNT, SUM, AVG, MIN, MAX
- Statistical functions (STDDEV, VARIANCE)
- Conditional aggregation
- Grouping and filtering aggregated data

☐ Module 4: Advanced SQL Topics

☐ Understand Subqueries

- Correlated vs non-correlated subqueries
- Subqueries in SELECT, FROM, WHERE
- EXISTS and NOT EXISTS
- Performance implications

☐ Understand CTE (Common Table Expression)

- CTE syntax and structure
- Recursive CTEs for hierarchical data
- Multiple CTEs in single query
- CTEs for query organization

☐ Window Function Basics

- OVER() clause fundamentals
- Partitioning data with PARTITION BY
- Ordering within windows with ORDER BY
- Frame clauses (ROWS vs RANGE)

☐ Analytical Functions (RANK, DENSE_RANK, ROW_NUM, LAG, LEAD)

- Ranking functions comparison
- LAG/LEAD for time series analysis
- Running totals and moving averages
- Percentile calculations

☐ Views

- Creating and managing views
- Materialized views vs regular views
- View security and permissions
- Performance considerations

☐ CTAS (Create Table As Select)

- CTAS syntax and use cases
- Table cloning and data archiving
- Performance benefits
- Temporary tables with CTAS

☐ TEMP Table

- Local vs global temp tables
- Temp tables in stored procedures
- Performance optimization with temp tables
- Cleanup and management

☐ Casting and Data Type Conversion

- CAST and CONVERT functions
- Implicit vs explicit conversion
- Handling conversion errors
- Date and string formatting

☐ Working with Date time

- Date extraction functions
- Date arithmetic and intervals
- Time zone handling
- Date formatting and parsing

☐ CASE Statements

- Simple CASE vs searched CASE
- Conditional aggregation with CASE
- Pivoting data with CASE
- Nested CASE statements

Stored Procedures

- Creating and executing procedures
- Parameters and return values
- Control structures (IF, WHILE, LOOP)
- Error handling in procedures

Module 5: Data Modeling

Building ER Model

- Entity Relationship Diagram fundamentals
- Identifying entities and attributes
- Relationship types (one-to-one, one-to-many, many-to-many)
- Cardinality and modality
- Tools for ER modeling

Building Data Model

- Conceptual vs logical vs physical models
- Dimensional modeling (star schema, snowflake)
- Fact and dimension tables
- Slowly Changing Dimensions (SCD)
- Data vault modeling introduction

Module 6: Projects

Project 1: Sales Data Exploration using SELECT

- Business requirements analysis
- Database setup and data loading
- Exploratory data analysis queries
- Sales performance reporting
- Customer segmentation analysis
- Revenue trend analysis

Project 2: E-Commerce Data Modeling and Analysis (SQL + Python)

- Complete database design for e-commerce
- Implementing the schema in SQL
- Advanced analytical queries
- Integrating SQL with Python (pandas, SQLAlchemy)
- Building data pipelines
- Creating analytical dashboard