

DATA SCIENCE

ADVANCED

4 DAYS

SQL for Data Science

Build a strong foundation in SQL for data analysis by mastering core concepts, advanced querying techniques, and performance optimization. Learn how to use SQL to analyze data, integrate with data science tools, apply governance principles, and solve real world business problems through data driven insights.

Why SQL for Data Science?

Efficient Data Analysis

SQL enables data scientists to efficiently retrieve and manipulate large volumes of structured data.

Seamless Data Integration

SQL provides the ability to integrate and combine data from various sources, allowing for a holistic view of the data.

Standardisation and Collaboration

SQL is a standardised query language used across different database platforms. By training your data science team in SQL, you ensure a common language and skillset for data analysis and manipulation.

What you'll be able to do

- ✓ Gain a solid understanding of SQL concepts, syntax, and best practices.
- ✓ Learn how to use SQL to manipulate and transform data effectively.
- ✓ Learn how to apply advanced SQL techniques, including joining tables, working with subqueries, using window functions, and optimising query performance.
- ✓ Gain the ability to apply SQL for data analysis purposes.
- ✓ Equip participants with the skills to integrate SQL with other data science tools and programming languages.
- ✓ Learn how to apply SQL to solve real-world data science problems commonly encountered in corporate environments.
- ✓ Learn about data governance principles and how to implement them within the SQL context.
- ✓ Learn how to leverage SQL to make data-driven decisions.

Who this is for

- Data Scientists
- Data Analysts
- Business Analysts
- Database Developers/Administrators
- Data Engineers

Curriculum

01 SQL Foundations for Data Science

- Position SQL in analytics and ML data prep workflows
- Query relational tables with SELECT fundamentals
- Read schemas, keys, and relationships confidently
- Write readable queries with consistent style

02 Filtering, Sorting, and Aggregation

- Filter rows with precise WHERE conditions
- Sort and page results for analysis
- Aggregate with GROUP BY and summary functions
- **Hands-on:** Build KPI queries from raw tables

03 Functions and Data Transformation

- Apply string, numeric, date, and conversion functions
- Shape columns for feature and reporting needs
- Handle nulls during transformations
- Compose reusable expression patterns

04 Joins and Subqueries

- Combine tables with inner, left, and advanced joins
- Avoid row explosion and ambiguous keys
- Use subqueries and derived tables for complex logic
- Validate join correctness with row counts and samples

05 Data Manipulation and Constraints

- Insert, update, and delete data safely
- Use transactions for controlled changes
- Apply primary/foreign keys and constraints
- Protect data integrity in shared warehouses

06 Views and Indexing

- Create views for curated analytical layers
- Use indexes to speed common access paths
- Balance write cost vs read performance
- Document views as contracts for analysts

07 Advanced SQL (Window Functions and CTEs)

- Use window functions for rankings and running metrics
- Structure complex logic with CTEs
- Build incremental and analytical query patterns
- Refactor messy SQL into maintainable pipelines

08 Stored Procedures and Functions

- Encapsulate repeated logic in procedures/functions
- Parameterize routines for reuse across teams
- Understand when SQL routines vs app code is better
- Test and version database routines carefully

09 SQL with Python and R

- Connect notebooks and scripts to SQL engines
- Pull query results into pandas/R data frames
- Push transformed data back with controlled loads
- Keep SQL and code boundaries clear for maintainability

10 Dates, Missing Data, and Performance Tuning

- Handle dates/times and late-arriving windows
- Impute or filter missing values deliberately
- Read execution plans and remove hotspots
- Tune queries for warehouse cost and latency

11 Visualizing Query Results

- Shape SQL outputs for charts and dashboards
- Integrate SQL with common visualization workflows
- Design queries that match visual grain
- Validate numbers before publishing visuals