KAIA

Learn. Transform. Get Placed

+91-9035192617

kalaalytics@gmail.com

kalaanalyticstraining.com

Batch starts on

30th March 2024

Online & Classroom Teaching

TECHNOLOGIES AND TOOLS WE TEACH

SQL

snowflake

aws

Power BI

COURSE SYLLABUS

SQL Using Snowflake Cloud

MODULE 1

Module 1: Structured Query Language SQL with Snowflake (Introduction

- Introduction to Database
- Database Management System
- Relational Database Management System

MODULE 2

Module 2: DDL/ DML/ DQL

- Creating databases and database objects
- Dropping and using database
- SQL commands for DDL, DML, DCL and DQL
- SQL CREATE TABLE and data types in designing tables along with constraints
- Table objects and alteration of tables by adding/dropping/renaming columns

MODULE 3

Module 2: Importing Data / Filtering/ Grouping/ Functions

- Data importing and exporting using wizard
- Dropping/Deleting and truncating tables
- Select Queries, SQL clauses, SQL summary query
- Advanced SQL queries using case and when, pivot and unpivot, update with CASE-WHEN and subqueries
- Functions

MODULE 4

Module 4: Joins and sets

- Vertical Join- Appending
- Horizontal Join- Merging(Set)
- Keys and Relationships

MODULE 5

Module 5: Sub Queries

- SQL data model designing using fact and dimension table concept for start schema and snowflakes
- SQL sub-queries
- Sub-queries vs partition by vs with clause

MODULE 6

Module 6: window functions

- Rank () over, Dense_rank () over, Row() over etc
- Aggregated and value window function
- SQL Functions - Character, Numeric and Datetime
- SQL Views

MODULE 7

Module 7: Advanced Topics

- Optimizing using CTE
- Automation using stored procedure
- User defined functions
- Clustered and non clustered index

COURSE SYLLABUS

Data Warehousing & Snowflake

MODULE 1

Introduction to Data Warehousing & Snowflake

- Data warehousing concepts and evolution
- Traditional vs Cloud data warehousing
- Overview of Snowflake and its advantages
- Use cases and industry applications

MODULE 2

Getting Started with Snowflake And Architecture

- Snowflake account setup
- Web Ul overview
- Basic SQL operations
- Navigating databases, schemas, and warehouses
- Multi-cluster shared data architecture
- Database storage layer
- Compute layer (Virtual Warehouses)
- Cloud services layer

MODULE 3

Databases, Schemas & Tables

- Creating and managing databases and schemas
- Table types: Permanent, Transient, Temporary
- Views and materialised views (overview)
- Data types and constraints

MODULE 4

Data Loading & Staging

- Internal and external stages
- COPY INTO command
- File formats (CSV, JSON, Parquet)
- Best practices for efficient data loading

MODULE 5

Cloud Integrations (AWS)

- Working with JSON, Parquet
- Querying semi-structured data
- Auto-ingest vs manual Snowpipe
- Event-based data loading

MODULE 6

Databases, Schemas & Tables

- Creating and managing databases and schemas
- Table types: Permanent, Transient, Temporary
- Views and materialised views (overview)
- Data types and constraints

MODULE 7

Performance Optimization

- Virtual warehouse sizing
- Query profiling and optimization
- Clustering keys
- Materialized views

MODULE 8

Time Travel & Fail-safe

- Data retention concepts
- Querying historical data
- Restoring dropped objects
- Cloning using Time Travel
- Fail-safe overview

MODULE 9

Tasks & Automation And Access Management

- Creating and managing databases and schemas
- Table types: Permanent, Transient, Temporary
- Views and materialised views (overview)
- Data types and constraints
- Role-Based Access Control (RBAC)
- Users, roles, and privileges

Real Time Project

Identifing Potential Customer using RFM value



Power BI

COURSE SYLLABUS

Reporting and Visualization

MODULE 1

Introduction to BI & Power BI

- Power BI Components
- Architecture
- Import vs Direct Query vs Live
- Licensing Overview
- Hands-on: Install Desktop and connect Excel

MODULE 2

Data Connectivity & Power Query

- Connect to Excel, CSV, SQL Server, Folder, Web
- Power Query transformations
- Merge & Append
- Hands-on: Clean sales dataset

MODULE 3

Data Modeling

- Fact & Dimension Tables
- Relationships
- Cardinality
- Date Table
- Hands-on: Build star schema

MODULE 4

DAX Fundamentals (Measures vs Calculated Columns)

- SUM, COUNT, AVERAGE
- IF, SWITCH
- CALCULATE, FILTER, ALL
- YTD, SAMEPERIODLASTYEAR
- RANKX
- Hands-on DAX measures

MODULE 5

Data Visualization

- Charts
- Cards
- Matrix
- Maps
- Slicers
- Drill Down/Through
- Bookmarks
- Buttons
- Hands-on Executive Dashboard

MODULE 6

Power BI Service

- Workspace
- Dashboards
- Refresh
- Sharing
- Hands-on publish report

Real Time Project

Sales Dashboard

Problem Statement

Statement

Understand the problem and the use of solving it

Time Intelligence

Analyzing Charts and Creating Visualizations

Publish to Server

Drill Through, Filters Page actions.

COURSE SYLLABUS

Data Engineering on AWS

MODULE 1

Cloud & IAM foundations for data

- Regions, AZs, shared responsibility
- IAM users, groups, roles, policies
- Least privilege for data services
- Free plan, Budgets, billing alarms
- AWS CLI and boto3 setup

MODULE 2

Data lake storage on Amazon S3

- Buckets, prefixes, storage classes
- Raw → clean → curated zones
- CSV vs JSON vs Parquet
- Partitioning by date
- Encryption and lifecycle rules

MODULE 3

Data catalog & SQL with Glue and Athena

- Databases, tables, schemas
- Glue crawlers and classifiers
- Athena SQL directly on S3
- Partitions and cost per data scanned
- CTAS: convert CSV to Parquet

MODULE 4

Batch ETL with AWS Glue

- Glue jobs, workers, PySpark basics
- DynamicFrame vs DataFrame
- Glue Studio visual ETL
- Joins, dedupe, data-quality checks
- Job bookmarks, incremental loads

MODULE 5

Event-driven processing with Lambda

- Lambda runtime, limits, layers
- S3 event notifications
- Validating and routing files
- CloudWatch Logs and metrics

Performance Optimization

MODULE 6

Data warehousing with Redshift

- OLTP (RDS) vs OLAP (Redshift)
- Star schema: facts and dimensions
- COPY from S3; sort and dist keys
- Spectrum: query the lake in place
- Serverless base capacity and cost

MODULE 7

Streaming ingestion with Kinesis

- Batch vs streaming use cases
- Data Streams: shards, producers, consumers
- Firehose delivery to S3, buffering
- Python event producer

MODULE 8

Orchestration, monitoring & governance

- Step Functions state machines
- EventBridge schedules and rules
- CloudWatch alarms on failed jobs
- Lake Formation permissions overview
- Tagging and teardown scripts

MODULE 9

Review, certification & careers

- Recap: reference lakehouse on AWS
- AWS Certified Data Engineer – Associate roadmap
- Common interview questions
- Free paths: Skill Builder, AWS Educate

Real Time Project

Retail sales analytics pipeline, end to end

