Getting started with Data Warehousing

Course Description:

This course provides an in-depth understanding of data warehousing concepts, architecture, data modeling, ETL processes, implementation, and optimization techniques. It is designed for professionals aiming to build, manage, and optimize data warehouses.

Course Duration: 24 hours

Prerequisites:

- ✓ Basic understanding of databases and SQL.

Table of Contents :

Module 1: Introduction to Data Warehousing

- What is a Data Warehouse?
- Key Characteristics of a Data Warehouse
- Benefits of Using a Data Warehouse
- Differences Between OLTP and OLAP Systems
- Evolution of Data Warehousing

Module 2: Data Warehouse Architecture

- Three-Tier Architecture
  - Bottom Tier (Data Storage)
  - Middle Tier (OLAP Server)
  - Top Tier (Front-End Tools)
- Components of a Data Warehouse
  - Data Sources
  - ETL (Extract, Transform, Load)
  - Data Storage

- Metadata
- Query and Reporting Tools
- Types of Data Warehouse Architectures
  - Enterprise Data Warehouse (EDW)
  - Operational Data Store (ODS)
  - Data Marts

Module 3: Data Modeling for Data Warehouses

- Dimensional Modeling
  - Star Schema
  - Snowflake Schema
  - Galaxy Schema (Fact Constellation)
- Fact Tables
  - Types of Facts (Additive, Semi-Additive, Non-Additive)
- Dimension Tables
  - Types of Dimensions (Conformed, Junk, Degenerate)
- Slowly Changing Dimensions (SCDs)
  - Type 1, Type 2, Type 3, and Hybrid Approaches

Module 4: ETL (Extract, Transform, Load) Process

- Overview of ETL
- Extraction Techniques
- Transformation Techniques
  - Data Cleaning
  - Data Integration
  - Data Aggregation
- Loading Techniques
  - Full Load vs Incremental Load
- ETL Tools Overview (e.g., Informatica, Talend, SSIS)

Module 5: Data Warehouse Implementation

- Planning and Requirements Gathering

- Data Warehouse Design

- Data Integration and Migration

- Testing and Validation

- Deployment and Maintenance

Module 6: Data Warehouse Optimization

- Indexing Strategies

- Partitioning and Clustering

- Query Optimization Techniques

- Caching and Materialized Views

- Performance Monitoring and Tuning