<u>DataStage</u>

Introduction to DataStage

- Overview of DataStage: History, purpose, and capabilities
- ETL Concepts: Extract, Transform, Load processes
- Data Integration Challenges and Solutions

DataStage Architecture and Components

- DataStage Architecture: Client-server architecture, parallel processing
- DataStage Components: Designer, Director, Administrator, Manager
- DataStage Parallel Job Design: Stages, links, containers, and job properties

DataStage Installation and Setup

- Installing IBM InfoSphere Information Server
- Configuring DataStage: DataStage Administrator tasks
- Connectivity to Data Sources: Setting up database connections

DataStage Job Development Basics

- Introduction to DataStage Designer: GUI overview and navigation
- Creating DataStage Jobs: Sequential vs. Parallel Jobs
- Importing and Exporting Metadata

DataStage Stages and Transformations

- Overview of DataStage Stages: Input, output, and processing stages
- Transformations: Filter, Join, Lookup, Sort, Aggregator
- Advanced Transformations: Pivot, Unpivot, Modify, Column Export

DataStage Job Design Techniques

- Job Design Best Practices: Reusability, modularity, and efficiency
- Job Parameters and Variables: Using parameters and environment variables
- Job Sequencing: Creating sequences and job dependencies

DataStage Parallel Processing and Performance Tuning

- Partitioning Techniques: Hash, range, round-robin partitioning
- Performance Optimization: Parallel job design considerations
- Monitoring and Debugging Jobs: Director and logs

Advanced DataStage Features

- Job Control: Job sequences, job control language (JCL)
- DataStage Macros and Shared Containers
- Error Handling and Recovery Strategies

DataStage Connectivity and Integration

- DataStage and Enterprise Data Warehousing (EDW) Integration
- Connecting to Different Data Sources: Relational databases, flat files, APIs
- Real-time Data Integration with DataStage

DataStage Administration and Security

- DataStage Administrator Tasks: User management, security settings
- DataStage Repository: Backup, restore, and maintenance
- Managing Metadata and Data Lineage

DataStage Job Deployment and Automation

- Packaging and Deploying DataStage Jobs
- DataStage Job Scheduling: IBM Scheduler, external schedulers
- Automating Job Execution: Command line utilities and scripting

DataStage Performance Monitoring and Troubleshooting

- Performance Monitoring Tools: DataStage Manager, Operations Console
- Troubleshooting Common Issues: Job failures, performance bottlenecks
- Enhancing Job Performance: Benchmarking and optimization techniques

DataStage Project Work and Case Studies

- Real-world DataStage Project: Design and implementation
- Case Studies: Industry-specific use cases and solutions
- Presentation and Documentation of DataStage Projects