Data Integration with Cloud Data Fusion Training

Course Outline

Module 1: Introduction to data integration and Cloud Data Fusion

- Data integration: what, why, challenges

- Data integration tools used in industry

- User personas

- Introduction to Cloud Data Fusion

- Data integration critical capabilities

- Cloud Data Fusion UI components

Module 2: Building pipelines

- Cloud Data Fusion architecture

- Core concepts

- Data pipelines and directed acyclic graphs (DAG)

- Pipeline Lifecycle

- Designing pipelines in Pipeline Studio

Module 3: Designing complex pipelines

- Branching, Merging and Joining

- Actions and Notifications

- Error handling and Macros

- Pipeline Configurations, Scheduling, Import and Export

Module 4: Pipeline execution environment

- Schedules and triggers

- Execution environment: Compute profile and provisions

- Monitoring pipelines

Module 5: Building Transformations and Preparing Data with Wrangler

- Wrangler

- Directives

- User-defined directives

Module 6: Connectors and streaming pipelines

- Understand the data integration architecture

- List various connectors

- Use the Cloud Data Loss Prevention (DLP) API

- Understand the reference architecture of streaming pipelines

- Build and execute a streaming pipeline

Module 7: Metadata and data lineage

- Metadata

- Data lineage