

# Google Cloud Dataflow Training

*COURSE CONTENT*

## GET IN TOUCH



Multisoft Systems  
B - 125, Sector - 2, Noida



(+91) 9810-306-956



info@multisoftsystems.com



www.multisoftsystems.com

## About Multisoft

Train yourself with the best and develop valuable in-demand skills with Multisoft Systems. A leading certification training provider, Multisoft collaborates with top technologies to bring world-class one-on-one and certification trainings. With the goal to empower professionals and business across the globe, we offer more than 1500 training courses, which are delivered by Multisoft's global subject matter experts. We offer tailored corporate training; project Based Training, comprehensive learning solution with lifetime e-learning access, after training support and globally recognized training certificates.

## About Course

Google Cloud Dataflow Training by Multisoft Systems is designed to help professionals master scalable data processing and real-time analytics using Google Cloud Platform (GCP). This industry-oriented training provides comprehensive knowledge of Google Cloud Dataflow, Apache Beam, batch processing, streaming pipelines, ETL workflows, and cloud-native data engineering practices.

## Module 1: Introduction to Google Cloud Platform (GCP)

- ✓ Overview of Cloud Computing
- ✓ Introduction to Google Cloud Platform
- ✓ GCP Core Services and Architecture
- ✓ Understanding Cloud Data Engineering
- ✓ Introduction to Google Cloud Console
- ✓ IAM and Resource Management Basics

## Module 2: Introduction to Google Cloud Dataflow

- ✓ Overview of Google Cloud Dataflow
- ✓ Features and Benefits of Dataflow
- ✓ Batch vs Streaming Data Processing
- ✓ Dataflow Architecture and Components
- ✓ Managed Service Concepts
- ✓ Use Cases and Industry Applications

## Module 3: Apache Beam Fundamentals

- ✓ Introduction to Apache Beam
- ✓ Beam Programming Model
- ✓ Pipelines, PCollections, and Transforms
- ✓ Core Beam SDK Concepts
- ✓ Pipeline Execution Lifecycle
- ✓ Beam Runners and Dataflow Runner

## Module 4: Setting Up the Environment

- ✓ Configuring GCP Environment
- ✓ Creating and Managing Projects
- ✓ Installing SDKs and Tools
- ✓ Setting Up Python and Java SDKs

- ✓ Authentication and Permissions
- ✓ Creating First Dataflow Pipeline

## **Module 5: Building Batch Data Pipelines**

- ✓ Batch Processing Fundamentals
- ✓ Reading Data from Different Sources
- ✓ Data Transformation Techniques
- ✓ Writing Output to Storage Systems
- ✓ Pipeline Optimization Techniques
- ✓ Error Handling in Batch Jobs

## **Module 6: Stream Processing with Dataflow**

- ✓ Streaming Data Fundamentals
- ✓ Real-Time Data Processing Concepts
- ✓ Working with Streaming Pipelines
- ✓ Event Time and Processing Time
- ✓ Windowing and Triggers
- ✓ Watermarks and Late Data Handling

## **Module 7: Data Sources and Integrations**

- ✓ Integrating with BigQuery
- ✓ Working with Cloud Pub/Sub
- ✓ Cloud Storage Integration
- ✓ Reading from Databases and APIs
- ✓ Data Ingestion Strategies
- ✓ Multi-Source Data Processing

## **Module 8: Data Transformation and Processing**

- ✓ ParDo Transformations

- ✓ GroupByKey and Combine Operations
- ✓ Filtering and Aggregations
- ✓ Side Inputs and Side Outputs
- ✓ Complex Event Processing
- ✓ Schema Management Techniques

## **Module 9: Monitoring and Debugging**

- ✓ Monitoring Pipelines in GCP
- ✓ Dataflow Job Management
- ✓ Logging and Error Tracking
- ✓ Debugging Pipeline Failures
- ✓ Performance Monitoring
- ✓ Troubleshooting Best Practices

## **Module 10: Performance Optimization**

- ✓ Autoscaling Concepts
- ✓ Resource Optimization Strategies
- ✓ Pipeline Parallelism
- ✓ Cost Optimization Techniques
- ✓ Throughput and Latency Tuning
- ✓ Efficient Resource Utilization

## **Module 11: Security and Governance**

- ✓ GCP Security Fundamentals
- ✓ IAM Roles and Permissions
- ✓ Data Security Best Practices
- ✓ Secure Pipeline Development
- ✓ Compliance and Governance Concepts
- ✓ Encryption and Access Controls

## Module 12: Advanced Dataflow Concepts

- ✓ Stateful Processing
- ✓ Timers and Advanced Windowing
- ✓ Dynamic Work Rebalancing
- ✓ Custom Transform Development
- ✓ Template-Based Pipeline Deployment
- ✓ Flex Templates and Reusable Pipelines

## Module 13: Real-World Projects and Use Cases

- ✓ Real-Time Analytics Pipeline
- ✓ ETL Pipeline Development
- ✓ IoT Data Streaming Scenario
- ✓ Log Processing and Monitoring
- ✓ Data Warehouse Integration
- ✓ Enterprise Data Engineering Use Cases

## Module 14: DevOps and Deployment

- ✓ CI/CD for Data Pipelines
- ✓ Version Control Integration
- ✓ Automated Pipeline Deployment
- ✓ Infrastructure as Code Concepts
- ✓ Managing Production Pipelines
- ✓ Deployment Best Practices

## Module 15: Capstone Project and Certification Preparation

- ✓ End-to-End Dataflow Project
- ✓ Project Review and Optimization
- ✓ Best Practices for Enterprise Implementation