

# Google Cloud - Serverless Data Processing with Dataflow

---

<b>Code:</b>	GCP-SRVLS-DATA
<b>Length:</b>	3 days
<b>URL:</b>	<a href="#">View Online</a>

---

This training is intended for big data practitioners who want to further their understanding of Dataflow in order to advance their data processing applications. Beginning with foundations, this training explains how Apache Beam and Dataflow work together to meet your data processing needs without the risk of vendor lock-in. The section on developing pipelines covers how you convert your business logic into data processing applications that can run on Dataflow. This training culminates with a focus on operations, which reviews the most important lessons for operating a data application on Dataflow, including monitoring, troubleshooting, testing, and reliability.

## Skills Gained

- Demonstrate how Apache Beam and Dataflow work together to fulfill your organization's data processing needs.  
Summarize the benefits of the Beam Portability Framework and enable it for your Dataflow pipelines.
- Enable Shuffle and Streaming Engine, for batch and streaming pipelines respectively, for maximum performance.  
Enable Flexible Resource Scheduling for more cost-efficient performance.
- Select the right combination of IAM permissions for your Dataflow job.
- Implement best practices for a secure data processing environment.
- Select and tune the I/O of your choice for your Dataflow pipeline.
- Use schemas to simplify your Beam code and improve the performance of your pipeline.
- Develop a Beam pipeline using SQL and DataFrames.
- Perform monitoring, troubleshooting, testing and CI/CD on Dataflow pipelines.

## Who Can Benefit

- Data engineer.
- Data analysts and data scientists aspiring to develop data engineering skills

## Prerequisites

To get the most out of this course, participants should have completed the following courses:

- "Building Batch Data Pipelines"

- “Building Resilient Streaming Analytics Systems”

## Course Details

### Course Outline

- Module 1: Introduction
- Module 2: Beam Portability
- Module 3: Separating Compute and Storage with Dataflow
- Module 4: IAM, Quotas, and Permissions
- Module 5: Security
- Module 6: Beam Concepts Review
- Module 7: Windows, Watermarks, Triggers
- Module 8: Sources and Sinks
- Module 9: Schemas
- Module 10: State and Timers
- Module 11: Best Practices
- Module 12: Dataflow SQL and DataFrames
- Module 13: Beam Notebooks
- Module 14: Monitoring
- Module 15: Logging and Error Reporting
- Module 16: Troubleshooting and Debug
- Module 17: Performance
- Module 18: Testing and CI/CD
- Module 19: Reliability
- Module 20: Flex Templates
- Module 21: Summary

---

## Schedule (as of 4 )

Date	Location
------	----------

---

Refer a friend or colleague and get up to \$100 Amazon gift card\* — when they book training!

[Learn More](#)

