Serverless Data Processing with Dataflow

Category: AI, Data & Analytics | **Vendor:** Google Cloud

Duration: 24.00 hours (3 days) **19.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

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.

About This Course

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.

Who Should Attend

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

Prerequisites & Entry Requirements

General 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

Learning Outcomes

Upon successful completion of this course, participants will be able to:

- 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.

Detailed Course Outline

Module 1: Introduction

- Introduce the course objectives.
- Demonstrate how Apache Beam and Dataflow work together to fulfill your organization’s data processing needs.

Module 2: Beam Portability

- Summarize the benefits of the Beam Portability Framework.
- Customize the data processing environment of your pipeline using custom containers.
- Review use cases for cross-language transformations.
- Enable the Portability framework for your Dataflow pipelines.

Module 3: Separating Compute and Storage with Dataflow

- Enable Shuffle and Streaming Engine, for batch and streaming pipelines respectively, for maximum performance.
- Enable Flexible Resource Scheduling for more cost-efficient performance.

Module 4: IAM, Quotas, and Permissions

- Select the right combination of IAM permissions for your Dataflow job.
- Determine your capacity needs by inspecting the relevant quotas for your Dataflow jobs.

Module 5: Security

- Select your zonal data processing strategy using Dataflow, depending on your data locality needs.
- Implement best practices for a secure data processing environment.

Module 6: Beam Concepts Review

- Review main Apache Beam concepts (Pipeline, PCollections, PTransforms, Runner, reading/writing, Utility PTransforms, side inputs), bundles and DoFn Lifecycle.

Module 7: Windows, Watermarks, Triggers

- Implement logic to handle your late data.
- Review different types of triggers.
- Review core streaming concepts (unbounded PCollections, windows).

Module 8: Sources and Sinks

- Write the I/O of your choice for your Dataflow pipeline.
- Tune your source/sink transformation for maximum performance.

- Create custom sources and sinks using SDF.

Module 9: Schemas

- Introduce schemas, which give developers a way to express structured data in their Beam pipelines.
- Use schemas to simplify your Beam code and improve the performance of your pipeline.

Module 10: State and Timers

- Identify use cases for state and timer API implementations.
- Select the right type of state and timers for your pipeline.

Module 11: Best Practices

- Implement best practices for Dataflow pipelines.

Module 12: Dataflow SQL and DataFrames

- Develop a Beam pipeline using SQL and DataFrames.

Module 13: Beam Notebooks

- Prototype your pipeline in Python using Beam notebooks.
- Launch a job to Dataflow from a notebook.

Module 14: Monitoring

- Navigate the Dataflow Job Details UI.
- Interpret Job Metrics charts to diagnose pipeline regressions.
- Set alerts on Dataflow jobs using Cloud Monitoring.

Module 15: Logging and Error Reporting

- Use the Dataflow logs and diagnostics widgets to troubleshoot pipeline issues.

Module 16: Troubleshooting and Debug

- Use a structured approach to debug your Dataflow pipelines.
- Examine common causes for pipeline failures.

Module 17: Performance

- Understand performance considerations for pipelines.
- Consider how the shape of your data can affect pipeline performance.

Module 18: Testing and CI/CD

- Testing approaches for your Dataflow pipeline.
- Review frameworks and features available to streamline your CI/CD workflow for Dataflow pipelines.

Module 19: Reliability

- Implement reliability best practices for your Dataflow pipelines.

Module 20: Flex Templates

- Using flex templates to standardize and reuse Dataflow pipeline code.

Module 21: Summary

- Summary., Introduction, Beam Portability, Separating Compute and Storage with Dataflow, IAM, and Permissions, Security, Beam Concepts Review, Windows, Triggers, Sources and Sinks, Schemas, State and Timers, Best Practices, Dataflow SQL and DataFrames, Beam Notebooks, Monitoring, Logging and Error Reporting, Troubleshooting and Debug, Performance, Testing and CI/CD, Reliability, Flex Templates

Skills You Will Learn:

- 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.

Frequently Asked Questions

Q: What delivery options are available for Serverless Data Processing with Dataflow?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 3-day Serverless Data Processing with Dataflow course provides up to 19.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Serverless Data Processing with Dataflow training?

The training takes place over 3 day(s), with each day lasting approximately 24.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Serverless Data Processing with Dataflow?

Yes, we provide corporate training, dedicated training, and closed classes for Serverless Data Processing with Dataflow. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: What related terms do people search for?

Popular related searches include: Serverless Data Processing with Dataflow, Data Analysis, SDPF, Google Cloud, vILT

Q: Why choose Nexus Human for Serverless Data Processing with Dataflow?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Serverless Data Processing with Dataflow training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 19/10/2026.