ITTcert
BY ORSYS

Course : Serverless Data Processing with Dataflow

Official course, preparation for Google Cloud certification exams

Practical course - 3d - 21h00 - Ref. SDD
Price : 2890 € E.T.

With this training course, you'll deepen your mastery of Dataflow to take your data processing applications to the next level. You'll discover how Apache Beam and Dataflow work together without the risk of vendor lock-in. You'll learn how to transform your business logic into Dataflow pipelines, then master the essential operations: supervision, troubleshooting, testing and reliability.

Teaching objectives

At the end of the training, the participant will be able to:

- Explain how Beam and Dataflow can be used together to process data efficiently
- Activate Portability, Shuffle/Streaming Engine and Flexible Scheduling to optimize costs and performance
- Choosing the right IAM access and applying good pipeline security practices
- Configure and optimize I/O, schemas, SQL/DataFrames to simplify and accelerate the pipeline
- Monitoring, testing, troubleshooting and CI/CD of Dataflow pipelines

Intended audience
Data engineers, data analysts and data scientists aspiring to develop data engineering skills.

Prerequisites
Have taken the course "Data Engineering on Google Cloud Platform" Ref DGC or have equivalent knowledge.

Certification
We recommend you take this course if you want to prepare for certification as a "Google Cloud Professional Data Engineer".
Comment passer votre examen ?

Practical details

Teaching methods

Training in French. Official course material in digital format and in English. Good understanding of written English.

Course schedule

1 Beam portability

- Beam portability.
- Runner v2.
- Container environments.
- Inter-language transformations.

2 Separate calculation and storage with Dataflow

- Dataflow Streaming Engine.
- Flexible resource planning.
- Dataflow.
- Dataflow Shuffle service.

3 IAM, quotas and permissions

- IAM.
- Quotas.

4 Security

- Data localization.
- Shared VPC.
- private IP.
- CMEK.

5 Review of Beam concepts

- Beam bases.
- Utility transformations.
- DoFn life cycle.

6 Windows, watermarks, triggers

- Windows.
- Watermarks.
- Triggers.

7 Sources and Sinks

- Sources and Sinks.
- Text IO and File IO.
- BigQuery IO.
- Pub/Sub IO.
- Kafka IO.
- Bigtable IO.
- Avro IO.
- Splittable DoFn.

8 Schematics

- Beam diagrams.
- Code examples.

9 Status and timers

- API State.
- API Timer.

10 Best practices

- Schematics.
- Management of unprocessable data.
- Error handling.
- AutoValue code generator.
- JSON data management.
- Use the DoFn lifecycle.
- Pipeline optimization.

11 Dataflow SQL and DataFrames

- Dataflow and Beam SQL.
- SQL windowing.
- Beam DataFrames.

12 Notebooks Beam

- Notebooks Beam.

13 Monitoring

- List of jobs.
- Job information.
- Job graph.
- Job metrics.
- Metrics Explorer.

14 Error logging and reporting

- Logging.
- Error Reporting.

15 Troubleshooting and debugging

- Troubleshooting process.
- Types of problems.

16 Performance

- Pipeline design.
- Data structure.
- Sources, Sinks and external systems.
- Shuffle and Streaming Engine.

(17) Testing and CI/CD

- Overview of testing and CI/CD.
- Unit testing.
- Integration testing.
- Building artifacts.
- Deployment.

(18) Reliability

- Introduction to reliability.
- Supervision.
- Geolocation.
- Disaster recovery.
- High availability.

(19) Flex Templates

- Classic templates.
- Flex Templates.
- Using Flex Templates.
- Templates provided by Google.

Dates and locations

REMOTE CLASS
2026 : 16 June, 29 Sep., 1 Dec.

PARIS LA DÉFENSE
2026 : 16 June, 29 Sep., 1 Dec.