How-to Guides

Core and common tasks

Create a cluster
Create a cluster using the Google Cloud Console or the gcloud command-line tool.

(/dataproc/docs/guides/create-cluster)

Create a custom image
Learn how to create a custom image and install it on a Dataproc cluster.

(/dataproc/docs/guides/dataproc-images)

Create and manage labels
Create and manage Dataproc user labels.

(/dataproc/docs/guides/creating-managing-labels)

Manage Java and Scala dependencies for Spark
Learn how to manage Java & Scala dependencies and resolve conflicts for Apache Spark applications.

(/dataproc/docs/guides/manage-spark-dependencies)

Manage a cluster
Update or shut down a cluster.

(/dataproc/docs/guides/manage-cluster)

Run jobs on Google Kubernetes Engine[Beta]
Run Dataproc jobs on a GKE cluster.

(/dataproc/docs/concepts/jobs/dataproc-gke)

Set up a project

Set up a new project to use Dataproc.

(/dataproc/docs/guides/setup-project)

Submit a job

Submit different job types using the Google Cloud Console, the gcloud command-line tool, or by SSHing into a cluster instance.

(/dataproc/docs/guides/submit-job)

Logging and monitoring

Job driver output

Use the console, gcloud command-line tool, or Cloud Storage to view Dataproc job driver output.

(/dataproc/docs/guides/driver-output)

Cloud Logging

Use Cloud Logging to view Dataproc cluster and job logs.

(/dataproc/docs/guides/logging)

Cloud Monitoring

Use Cloud Monitoring to view Dataproc cluster metrics.

(/dataproc/docs/guides/monitoring)

Cloud Profiler

Use Cloud Profiler to profile Spark and Hadoop job CPU usage and memory-allocation.

(/dataproc/docs/guides/profiling)

Migrating to Dataproc

Migrating HDFS data
Migrate data from a Hadoop filesystem (HDFS) to Cloud Storage for use with Dataproc and other services.

 (/solutions/migration/hadoop/hadoop-gcp-migration-data)

Migrating clusters
Migrate Apache Spark and Apache Hadoop clusters to Dataproc.

 (/solutions/migration/hadoop/hadoop-gcp-migration-overview)

Migrating jobs
Move Spark, Hadoop, Hive, Pig and other job types to Dataproc.

 (/solutions/migration/hadoop/hadoop-gcp-migration-jobs)

   \