Create a cluster

You can create a Dataproc cluster via a Dataproc API clusters.create (/dataproc/docs/reference/rest/v1/projects.regions.clusters/create) HTTP or programmatic request, using the Cloud SDK gcloud (/sdk/gcloud/reference/dataproc/clusters/create) command-line tool in a local terminal window or in Cloud Shell (https://console.cloud.google.com/?cloudshell=true), or from the Google Cloud Console (https://console.cloud.google.com/dataproc/clustersAdd) opened in a local browser. You can also create clusters programmatically using Cloud Client Libraries, or via API requests.

Cluster Region: You can specify a `global` region or a specific region for your cluster. The global region is a special multi-region endpoint that is capable of deploying instances into any user-specified Compute Engine zone. You can also specify distinct regions, such as `us-east1` or `europe-west1`, to isolate resources (including VM instances and Cloud Storage) and metadata storage locations utilized by Dataproc within the user-specified region. See Regional endpoints (/dataproc/docs/concepts/regional-endpoints) to learn more about the difference between global and regional endpoints. See Available regions & zones (/compute/docs/regions-zones/regions-zones#available) for information on selecting a region. You can also run the `gcloud compute regions list` command to see a listing of available regions.

Compute Engine Virtual Machine instances (/compute/docs/instances) (VMs) in a Dataproc cluster, consisting of master and worker VMs, require full internal IP networking access to each other. The `default` network available (and normally used) to create a cluster helps ensure this access. For information on creating your own network for your Dataproc cluster, see Dataproc Cluster Network Configuration (/dataproc/docs/concepts/network).

Creating a Dataproc cluster

gcloud commandREST & CMD LINE (#rest--cmd... Console (#console)Go (#go)Java (#java)Node.js (#node

To create a Dataproc cluster on the command line, run the Cloud SDK gcloud dataproc clusters create (/sdk/gcloud/reference/dataproc/clusters/create) command locally in a terminal window or in Cloud Shell (https://console.cloud.google.com/?cloudshell=true).

```
$ gcloud dataproc clusters create cluster-name \
    --region=region
```

The above command creates a cluster with default Dataproc service settings for your master and worker virtual machine instances, disk sizes and types, network type, region and zone where your cluster is deployed, and other cluster settings. See the gcloud dataproc clusters create (/sdk/gcloud/reference/dataproc/clusters/create) command for information on using command line flags to customize cluster settings.

Create a cluster with a YAML file

1. Run the following `gcloud` command to export the configuration of an existing Dataproc cluster into a YAML file.

```
$ gcloud dataproc clusters export my-existing-cluster --destination clust
```

2. Create a new cluster by importing the YAML file configuration.

```
$ gcloud dataproc clusters import my-new-cluster --source cluster.yaml
```

Note: During the export operation, cluster-specific fields (such as cluster name), output-only fields, and automatically applied labels are filtered. These fields are disallowed in the imported YAML file used to create a cluster.

Note: You can click the **Equivalent REST or command line** links at the bottom of the Dataproc Cloud Console Create a cluster (https://console.cloud.google.com/dataproc/clustersAdd) page to have the Console construct an equivalent API REST request or **`gcloud`** tool command to use in your code or from the command line to create a cluster.

Last updated 2020-08-17 UTC.