Working with Spark in GCP Dataproc

[How to copy file from HDFS to the local file system](https://stackoverflow.com/questions/17837871/how-to-copy-file-from-hdfs-to-the-local-file-system)

[Ask Question](https://stackoverflow.com/questions/ask)

bin/hadoop fs -copyToLocal /hdfs/source/path /localfs/destination/path

bin/hadoop fs -get /hdfs/source/path /localfs/destination/path

# [How to copy a file from a GCS bucket in Dataproc to HDFS using google cloud?](https://stackoverflow.com/questions/54429642/how-to-copy-a-file-from-a-gcs-bucket-in-dataproc-to-hdfs-using-google-cloud)

[Ask Question](https://stackoverflow.com/questions/ask)

For the case of a single “small” file:

hdfs dfs -cp gs://<bucket>/<object> <hdfs path>

hdfs dfs -cp gs://<bucket>/<object> hdfs://<master node>/<hdfs path>

For the case case of “large” file or large directory of files:

hadoop distcp gs://<bucket>/<directory> <HDFS target directory>

Or you can leave your data on GCS

<https://cloud.google.com/solutions/migration/hadoop/hadoop-gcp-migration-data>