

Hands-On Exercise: Storing Data in Amazon S3

In this exercise, you will copy data from AWS S3 to HDFS.

IMPORTANT: This exercise depends on [Hands-On Exercise: Working with HDFS](#). Be sure to complete that exercise before continuing.

All steps in this exercise that use a terminal window should be run on cmhost.

View Your Environment's AWS Credentials

1. When your exercise environment was created, you were assigned two AWS keys that will give you access to the course S3 bucket. You will find these values in the README.txt file in ~/training_materials/admin/.

Copy Data from S3 to HDFS Using the `hadoop distcp` Command

2. Copy `shakespeare.txt` file from S3 into HDFS using the `hadoop distcp` utility. Enter the command below, which uses the AWS credentials provided in the README.txt file. Enter the correct value in place of `$AWS_ACCESS_KEY` and `$AWS_SECRET_KEY`.

```
$ hadoop distcp \  
-Dfs.s3a.access.key=$AWS_ACCESS_KEY \  
-Dfs.s3a.secret.key=$AWS_SECRET_KEY \  
s3a://admin-public/shakespeare.txt /tmp/
```

3. Confirm that the data is now in HDFS:

```
$ hdfs dfs -tail /tmp/shakespeare.txt
```

This is the end of the exercise.