

# Transferring Data to Object Storage from Other Cloud Providers or Local File Systems

ORACLE WHITE PAPER | JULY 2018





## Revision History

The following revisions have been made to this white paper since its initial publication:

---

<b>Date</b>	<b>Revision</b>
July 16, 2018	Updated the instructions for installing rclone system wide to include installation on Oracle Linux 7.

---

You can find the most recent versions of the Oracle Cloud Infrastructure white papers at <https://cloud.oracle.com/iaas/technical-resources>.



## Table of Contents

Overview	4
Target Audience and Requirements	4
Costs and Fees	5
System Preparation	5
Install Rclone	5
Configure Object Storage	6
Test the Connection to Object Storage	7
Configure Your Source	8
Test the Connection to Your Existing Cloud Provider	9
Transfer Data	10
Verify Transferred Files (Optional)	11
Resources	11



## Overview

The Oracle Cloud Infrastructure Object Storage service provides unlimited scale for large amounts of unstructured data like videos, backups, and logs. Data in Object Storage is automatically replicated across multiple fault domains.

One of the first steps in using Object Storage is to transfer your existing data. You can choose from a variety of different methods, tools, and services, and your choices depend on your specific use cases, the amount of data that needs to be transferred, and available network bandwidth.

This paper focuses on a primary use case, transferring data to Object Storage from other cloud providers. It provides specific instructions and examples for transferring data from Oracle Cloud Infrastructure Object Storage Classic, Amazon S3, and Microsoft Azure Blob Storage. It also provides instructions on a secondary use case, transferring data to Object Storage from a local file system.

---

**Note:** Although not addressed specifically, the same methodology can be used to transfer data from cloud providers that are not discussed here, including Google Cloud Storage, IBM Cloud Object Storage, Ceph, DigitalOcean Spaces, Rackspace Cloud Files, or any object storage that provides support for OpenStack Swift.

---


## Target Audience and Requirements

This document is intended for technical audiences, including systems and network administrators, site reliability engineers, solutions architects, and engineers.

To perform the actions outlined in this paper, you must meet the following knowledge and access:

- Basic familiarity with Linux and command line (CLI) tools
- Non-administrator-level access to a workstation or server running Linux with uninterrupted and fast connectivity to the internet
- Valid credentials for Object Storage

The methods outlined in this paper are suitable only if the amount of transferred data is less than 100 terabytes (TiB). The maximum amount of data that can be transferred depends on available internet bandwidth and, to a lesser degree, the technical specifications of the workstation or server that is used. If the amount of data is higher than 100 TiB, this method might take a prohibitively long time.



We recommend using an Oracle Cloud Infrastructure Compute instance when you transfer data from other cloud providers. All examples in this document have been tested on a VM.DenseIO1.4 Compute instance. For a list of available Compute instances, see the [Compute pricing page](#).

## Costs and Fees

Oracle does not charge additional fees for transferring data from other cloud providers. Data transferred to Object Storage is priced based on [standard Object Storage pricing](#).

If you use an Oracle Cloud Infrastructure Compute instance for transferring data, you are charged separately for that usage. For pricing of Compute instances, see the [Compute pricing page](#).

---

**Important:** If you transfer data from your existing cloud provider, they might charge you for transferring data *out* of their object storage. This cost is independent of what Oracle charges. Refer to the pricing information of your cloud provider.

---

## System Preparation

Several steps are required to prepare a system before transferring data. This section provides information about those steps.

### Install Rclone

[Rclone](#) is an open-source command-line utility that synchronizes files and directories between a local file system and a variety of cloud backends. The methods provided in this paper use rclone, so you must first install and configure it for Oracle Cloud Infrastructure Object Storage.

You can install rclone system-wide or only for your own use. Installing system-wide requires administrative privileges.

### System-Wide Installation

On Oracle Linux 7, you can install rclone by using yum:

```
$ sudo yum install -y rclone
```

On other Linux distributions, you can use the installation script provided by rclone. Download and run that script to install the latest version of rclone system-wide:

```
$ curl https://rclone.org/install.sh | sudo bash
```

## Installation for Individual Use

If you don't have administrative privileges or don't want to install rclone system-wide, you can download the latest binary and use it directly. The following instructions download rclone, create a `/bin` directory in your home folder, and add the `/bin` directory to your PATH environment variable:

```
$ mkdir ~/Temp
$ cd ~/Temp/
$ curl -O https://downloads.rclone.org/rclone-current-linux-amd64.zip
$ unzip rclone-current-linux-amd64.zip
$ cd rclone-*-linux-amd64
$ mkdir ~/bin
$ cp rclone ~/bin/
$ export PATH=$PATH:~/bin
$ echo 'export PATH=$PATH:~/bin' >> ~/.bashrc # Optional, only if you want to add it to
your path at every login
```

---

**Note:** PATH is reset when you log out unless you put it to your `.bashrc`.

---

## Custom Installation

Rclone supports methods and OSs that are outside the scope of this document. If you need a custom installation, see the [rclone installation documentation](#).

## Configure Object Storage

To configure rclone for Oracle Cloud Infrastructure Object Storage, you must use the Amazon S3 Compatibility API. Use the following instructions.

### Obtain the Necessary Information

You must obtain the secret key, access key, region API endpoint, and namespace for configuration.

To create a secret key and access key pair, follow the instructions in [Managing User Credentials](#) in the IAM service documentation.

The API endpoint is in the following form:

```
https://<your_namespace>.compat.objectstorage.<your_region_identifier>.oraclecloud.com
```

Replace `<your_namespace>` with your namespace, and replace `<your_region_identifier>` with the region in which your account was created.

To identify your namespace and get more information about namespaces, see [Understanding Object Storage Namespaces](#) in the Object Storage service documentation.

For more information about regions and a list of available regions, see [Regions and Availability Domains](#) in the service documentation.

## Set Environment Variables

After you have your secret key, access key, API endpoint, and region identifier, configure rclone by setting the environment variables in the following way:

```
$ export RCLONE_CONFIG_OCI_TYPE=s3
$ export RCLONE_CONFIG_OCI_ACCESS_KEY_ID=<your_access_key>
$ export RCLONE_CONFIG_OCI_SECRET_ACCESS_KEY=<your_secret_key>
$ export RCLONE_CONFIG_OCI_REGION=<your_region_identifier>
$ export RCLONE_CONFIG_OCI_ENDPOINT=
https://<your_namespace>.compat.objectstorage.<your_region_identifier>.oraclecloud.com
```

---

**Note:** You can configure rclone by using environment variables or putting the configuration into an rclone configuration file. In this paper, configuration examples use environment variables. For information about using a configuration file or about configurations in general, see the [rclone documentation](#).

---

Verify whether rclone has recognized the configuration via the environment variables:

```
$ rclone listremotes
oci:
$
```

---

**Note:** The rclone endpoint `oci` is now available.

---

## Test the Connection to Object Storage

Test the configuration as follows:

```
$ rclone lsdir oci: # List all buckets on OCI
-1 2018-01-31 19:46:35 -1 ArchiveBucket
-1 2018-01-17 20:27:18 -1 FileShare
-1 2017-10-05 05:02:05 -1 Public
-1 2018-02-21 18:54:35 -1 Temp
$ rclone mkdir oci:test_bucket # Optionally create a new
bucket
$ rclone lsdir oci:
-1 2018-01-31 19:46:35 -1 ArchiveBucket
-1 2018-01-17 20:27:18 -1 FileShare
-1 2017-10-05 05:02:05 -1 Public
-1 2018-02-21 18:54:35 -1 Temp
-1 2018-03-01 00:25:24 -1 test_bucket # Note the new test_bucket
that was just created
$ rclone rmdir oci:test_bucket # Remove test_bucket
```

If you get an error, verify your configuration.

## Configure Your Source

With rclone, you can define your source as a local file system or a different cloud provider. This section provides specific configuration examples for a local file system and several cloud providers, but you can also configure others.

### Local File System

No specific configuration is required for using a local file system with rclone. You can simply choose a local directory as your source. To do so, export your source path as an environment variable:

```
$ export SOURCE=/path/to/source
```

### Oracle Cloud Infrastructure Object Storage Classic

To configure rclone for Oracle Cloud Infrastructure Object Storage Classic, you need your username, password, service instance ID, and the authentication URL.

To find the authentication URL, follow the instructions in the [REST API documentation for Object Storage Classic](#). A valid authentication URL looks as follows:

```
https://foo.storage.oraclecloud.com/auth/v1.0
```

To find the service instance ID, use the service REST API endpoint URL. For example, if your REST API endpoint URL is `https://foo.storage.oraclecloud.com/v1/my-service-bar`, then your service instance ID is `my-service-bar`. For more information about finding the service instance ID, see the [REST API documentation](#).

After you have all the necessary information, configure rclone by setting the environment variables in the following way:

```
$ export RCLONE_CONFIG_OCIC_TYPE=swift
$ export RCLONE_CONFIG_OCIC_USER=<your_service_instance_ID>:<your_username>
$ export RCLONE_CONFIG_OCIC_KEY=<your_password>
$ export RCLONE_CONFIG_OCIC_AUTH=<your_authentication_URL>
$ export SOURCE=ocic:<your_source_container> # This step sets your source as OCI-C
```

---

**Note:** To use Object Storage Classic or any other Swift-compatible source, you need to use rclone version 1.4 or later. Version 1.3.9 has a bug that results in not refreshing authentication tokens, causing authorization failures.

---



## Amazon S3

To configure rclone for Amazon S3, you need the access key, the secret key, and the region of your bucket (if applicable). After you have them, you can configure rclone by setting the environment variables in the following way:

```
$ export RCLONE_CONFIG_S3_TYPE=s3
$ export RCLONE_CONFIG_S3_ACCESS_KEY_ID=<your_access_key>
$ export RCLONE_CONFIG_S3_SECRET_ACCESS_KEY=<your_secret_key>
$ export RCLONE_CONFIG_S3_REGION=<region_of_your_bucket>
$ export SOURCE=s3:<your_source_bucket> # This step sets your source as S3
```

---

**Note:** Setting `RCLONE_CONFIG_S3_REGION` is required only if your bucket is in a region other than `us-east-1`.

---

## Microsoft Azure Blob Storage

To configure rclone for Microsoft Azure Blob Storage, you need the storage account name and key. After you have them, you can configure rclone by setting the environment variables in the following way:

```
$ export RCLONE_CONFIG_AZURE_TYPE=azureblob
$ export RCLONE_CONFIG_AZURE_ACCOUNT=<your_storage_account_name>
$ export RCLONE_CONFIG_AZURE_KEY=<your_key>
$ export SOURCE=azure:<your_source_bucket> # This step sets your source as
Azure
```

## Other Cloud Providers

You can configure other Cloud Providers if they are supported by rclone. For a full list of supported cloud providers and information about their configuration parameters, see the [rclone website](#).

## Test the Connection to Your Existing Cloud Provider


After configuring your existing cloud provider, verify whether rclone recognizes both source and destination configurations:

```
$ rclone listremotes
oci:
s3:
$
```

---

**Note:** The rclone endpoint `oci` is available, as are the endpoints that you configured (`oci`, `s3`, or `azure`). If you are using a local file system as your source, only `oci` is listed as a remote.

---



You can now test the connection by invoking the following command to list all the files or objects on your source:

```
$ rclone ls $SOURCE
10737418240 disk.img
   77011 log.txt
 6270216 rclone-current-linux-amd64.zip
$
```

If you don't get an error and can see the contents of the source that you want to transfer, you can proceed. If you get an error, verify your configuration.

## Transfer Data

After your configuration is complete, you can start the transfer. Depending on the amount of data and the connection speed, this transfer can take a very long time (days or weeks).

To monitor progress, we recommend starting the transfer in verbose mode from GNU Screen, so you can run the process in the background and watch its progress as needed. GNU Screen is preinstalled with most Linux distributions. For more information about using it, see the [GNU Screen Documentation](#).

Following are the basic commands to use GNU Screen and start it in verbose mode. Replace `<destination_bucket>` as needed; a bucket will be created automatically if one doesn't already exist.

```
$ screen      # Start GNU Screen, press Enter on the next screen to get a new shell

$ rclone --verbose --cache-workers 64 --transfers 64 --retries 32 copy $SOURCE
oci:<destination_bucket>
[....]
[....] # While the transfer is ongoing, you can detach from GNU Screen by pressing Ctrl+a d
[....]
$
$ screen -r   # This command will get you back to GNU Screen, you can detach again by
pressing Ctrl+a d
```

---

**Note:** The optimal number of cache-workers and transfers depends on your specific use case, available bandwidth, resources available on the workstation or server, and throttling limits applied by your cloud provider. In general, the higher the number of transfers and cache-workers, the faster the transfer will complete. But beyond a certain number, you might start getting errors. We recommend that you start with 64 concurrent transfers and cache-workers and adjust the number as necessary.

---

When the transfer is complete, verify that no errors occurred. If you see errors, try decreasing the number of cache-workers and transfers and try again.

## Verify Transferred Files (Optional)

Independent of how data is transferred to Object Storage (whether from a local file system or another cloud provider, or by some other means), you can use rclone to verify whether the data copied to Object Storage is identical to its source.

You can verify the data quickly or more thoroughly.

Use the following method to perform a relatively quick verification:

```
$ rclone --verbose --cache-workers 64 --transfers 64 --retries 32 check $SOURCE  
oci:<destination_bucket>
```

This method detects differences in the source and destination buckets by verifying file sizes and MD5 hashes. However, MD5 hashes can't be used for objects that are uploaded by using multipart uploads. As a result, large objects are verified only by comparing their file sizes, which provides only a cursory verification.

To perform a more thorough verification, you can instruct rclone to download files and compare them:

```
$ rclone --verbose --cache-workers 64 --transfers 64 --retries 32 check --download $SOURCE  
oci:<destination_bucket>
```

Although this method performs a more thorough comparison, it requires downloading copies from both object stores.

---

**Important:** When you perform a verification with downloads, depending on where your job is run, you might incur additional data transfer fees from your existing cloud provider, Oracle, or both.

---

## Resources

For more information, see the following resources:

- [Overview of the Object Storage service](#)
- [Oracle Cloud Infrastructure Amazon S3 Compatibility API](#)
- [Rclone](#)







**Oracle Corporation, World Headquarters**

500 Oracle Parkway  
Redwood Shores, CA 94065, USA

**Worldwide Inquiries**

Phone: +1.650.506.7000  
Fax: +1.650.506.7200

CONNECT WITH US

-  [blogs.oracle.com/oracle](https://blogs.oracle.com/oracle)
-  [facebook.com/oracle](https://facebook.com/oracle)
-  [twitter.com/oracle](https://twitter.com/oracle)
-  [oracle.com](https://oracle.com)

**Integrated Cloud Applications & Platform Services**

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0718

Transferring Data to Object Storage from Other Cloud Providers or Local File Systems  
July 2018  
Author: Ilker Temir

