Data Performance Evaluation of Cloud Storage Providers

Aleksandar Dimov [1] and Stanimir Kirov [1]

[1] *Faculty of Mathematics and Informatics, Sofia University, 5 James Bourchier Blvd., Sofia, 1164, Bulgatia*

Abstract

Many of the current software systems are data-intensive which presents many new challenges not only to IT and to software professionals but also to business and individual users. Some of these challenges are related to decisions on how to store the data that data-intensive systems work with. One common solution is to use cloud storage, which most often is offered by third party. This paper presents a methodology for evaluation of cloud storage providers in the realm of data-intensive systems, based on the fundamental operations that are provided by their services. Further, it also makes a performance comparison of some of the popular cloud storage services in terms of the operations execution times.

Keywords

Data performance comparison; cloud storage providers; data-intensive systems

1. Introduction

An important concern in the realm of data-intensive systems is how users and businesses are going to store their data. Both regular and businesses users are increasingly credulous on cloud-based storage solutions instead of on-premises local storage hardware. Most significant reasons for this include security, availability, scalability and cost-effectiveness. More and more recognizable nowadays is the tendency to migrate data to the cloud or to take seriously the ability to base on the cloud when developing new solutions. In this sense, software engineers and IT professionals are interested to have means for well-informed selection of specific solutions, based on quality of service.

Additionally, most of the contemporary systems are data-intensive [1], [2], which means that they heavily rely on data storage and quality characteristics of such storage. Such systems also often perform data analysis and analytical processing which may be required to happen in real time. In these terms, it becomes especially significant to optimize performance of such systems.

However, in current environment, it may become difficult to select appropriate cloud storage provider, as there exist a lot such services. Users need means to select the best option in a straightforward way. One of the first things someone should do when choosing between cloud services is compare storage options, features, and costs. Next, it is the dependence on a single vendor for so many critical needs. If your data is in the hands of one service provider, the dependence on your provider is huge. To avoid this, users may implement multi-cloud architecture. By using multi-cloud storage connection tool, one can easily switch between cloud service providers that are supported by the tool.

The goal of this paper is to provide a methodological framework for testing of cloud storage providers and show particular results on some of the most popular free storage services. The research question employed by this study is "What are the main factors that users should employ to evaluate cloud storage solutions and how to pick provider that is right for their needs?".

The rest of the paper is structured as follows: Section 2 makes an overview of the related work in the area; Section 3 presents the methodological framework of our approach to testing of cloud storage providers; Section 4 describes the specifics of the testing environment and experiments, we have made; Section 5 presents and analyse the results and finally, section 6 concludes the paper.

2. Related work

There exist a number of research works that directly relate to our and aim at performance comparison of cloud storage providers.

Like [5], where a comparison between Google Cloud Service and iCloud is made by exploration of the features of these two cloud storage services.

In [6], the authors have tested performance of several cloud storage providers including Google Drive and Dropbox and have analysed their applicability in healthcare services by using medical image files for testing and comparison. Like what is shown in this paper, a comparison was based on time duration of several commands, including upload, download, and file deletion.

Another comparison of some popular cloud storage services is provided in [7]. The authors aim to help users to choose the right cloud service for storage by making a comparison on 10 different factors, including performance. It is evaluated by upload and download of files of two different sizes.

There also exist several non-academic surveys [3], [4] that try to rate cloud storage providers, however they do not focus on methodological approach to testing but rather just compare the properties of different plans that cloud storage providers offer.

Another direction of research that have some relation to our work concern performance testing of various cloud services. Like in [8], where High-Perfor-

mance Computing (HPC) is evaluated in terms of performance comparison of Google services Cloud Functions (Function-as-a-service) and Compute Engine (Infrastructure-as-a-Service).

In conclusion, there exist a lot of work in comparison of cloud services and cloud storage in particular. However, in this paper we are trying to fill the gap in relation to the cloud evaluation with respect to data-intensive systems. For this purpose, in next section we present our methodology for testing performance, which is specifically targeted at storage service operations.

3. Comparison methodology

This section will explain the methodological approach for comparison between different cloud storage providers.

The test environment should be fully isolated from other applications, in order to prevent data interference. An additional application is also needed to provide a bridge between the test environment and cloud providers under test. It will also serve as a wrapper that will allow access to different cloud providers and provide the same and fair conditions for all of them.

We will perform the test following three main phases:

1. pre-test phase – a share is created, which is going to be used in the test phase to check the performance of cloud data storage providers

2. test execution phase – this phase consists of execution of 9 operations common for each operating system and execution time is measured for each of them. These operations are the following:

   a) Create share – this operation is used to create a location for storing files;

   b) List share – this operation is used to show files in the share\directory listed;

   c) Move share – this operation is used to move a directory and subdirectories (if available) and files within the share;

   d) Copy share – this operation is used to copy a directory and subdirectories (if available) and files within the share;

   e) Delete share – this operation is used to delete a directory and subdirectories (if available) and files within the share;

   f) Upload file – this operation is used to transfer data from source (computer\PC) to destination (cloud share in this project case);

   g) Download file – this operation is used to transfer data from source (cloud share) to destination (computer\PC);

   h) Copy file – this operation is used to copy files;

   i) Delete file – This operation is used to remove a file from the file system in Create share.

3. post-test phase – this phase has the duty to prepare for the next iteration of the test execution phase. It includes cleaning the test file that that was created during the previous phase. This is needed since free accounts are used that have limited storage space.
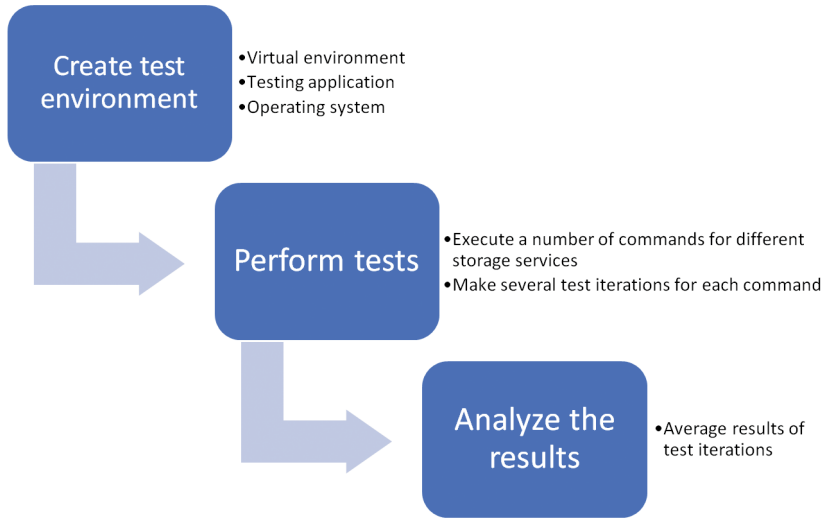
Create test environment
- Virtual environment
- Testing application
- Operating system

Perform tests
- Execute a number of commands for different storage services
- Make several test iterations for each command

Analyze the results
- Average results of test iterations

Figure 1: Cloud providers testing methodology

Testing of a cloud storage provider should be performed while treating it as a black box. Normally, one should not be able to get any kind of internal information for cloud architecture infrastructure as this is considered as security breach and if that happens the cloud infrastructure could be classified as highly unreliable. This way, we are going to use opaque testing technique. With this technique, only the fundamental aspects of the system are being explored. In that way, more data may be collected and conclusions can be very accurate regarding different cloud storage vendor's behaviour and response according to our setup.

In order to perform the test, we should ensure the following requirements that are supposed to the fairest test conditions:

1. Single platform or application should be used to access different cloud storage providers.

2. Virtualization should be used, which is limited to a single virtual machine. This will provide an isolated environment and is a safe, efficient, cheap and flexible way to test applications – one can test everything from server configurations to resource allocation and most importantly for us – storage.

3. The operating system should be less demanding and have good handling of resources so it can have less interference with the application and the test results to be believed as accurate as they can.

4. It should be considered that cloud storage had different characteristics for different uses (different end users or companies could make use of the service in different ways). For this reason, we focused only on file-system based operations and we will use a single application to access different cloud solutions for storage service offered by vendors.

4. Building the testing environment

We are going to use Rclone[2] command-line tool as an intermediary application between a client and cloud provider service. This way the integration is provided between them. Rclone, is an instrument written in Go programming language which is used to download\upload data from computer to a cloud hosted data storage centre. It can connect to various cloud storage centres. This way, the requirement for a single platform to have access to different cloud storage services offered by vendors is going to be fulfilled.

Another objective of using Rclone command-line tool is to produce multiservice cloud delivery model. By developing and implementing it, we can compare supported storage services from a performance perspective. The architecture of the test environment built is shown on Figure 2.
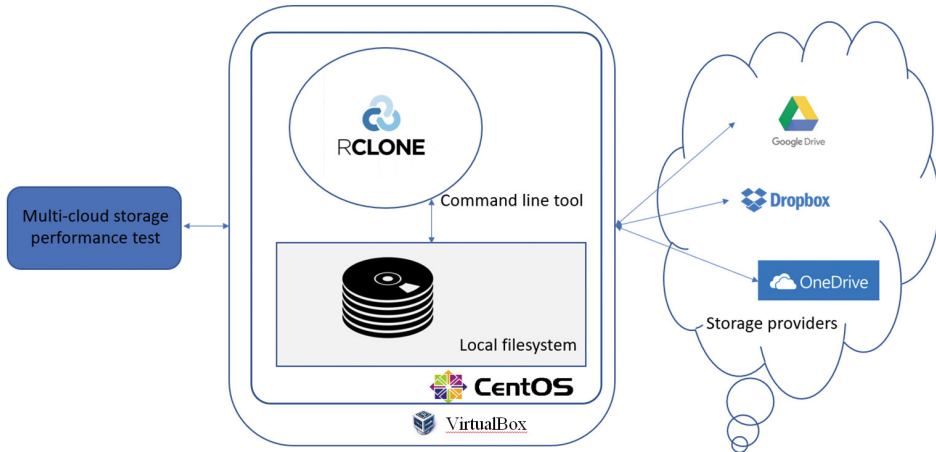

Figure 2: Architecture of a multi-cloud storage performance test

To provide virtualization, Oracle VirtualBox is used. It is a deceptively simple, but powerful and free to use cross-platform virtualization application for x86 hardware, targeted at server, desktop and embedded use [5].

---

[2] https://rclone.org

As an operating system the CentOS Linux distribution was used, as it is a stable, predictable, manageable, and reproducible platform derived from the sources of Red Hat Enterprise Linux [6], [7]. It is available free of charge and technical support is primarily provided by the community via official mailing lists, web forums, and chat rooms. Other reasons for it to be chosen for our work is that it has good documentation; it is highly customizable and is supported by Virtualbox.

As defined in the methodology description in Section 3, we have to implement the operations that are most used on storage. In the list below each operation is shown together with the specific Rclone command that was used to execute it:

- Create share
```
rclone mkdir [Provider]:Testdir
```
- List share
```
rclone lsf [Provider]:Test
```
- Move share
```
rclone move [Provider]:Test [Provider]:Testdir
```
- Copy share
```
rclone copy [Provider]:Testdir [Provider]:Test
```
- Delete share
```
rclone purge [Provider]:Testdir
```
- Upload file
```
rclone copy /home/user/documents/1GB.txt
[Provider]:Test
```
- Download file
```
rclone sync [Provider]:Test /home/user/downloads
```
- Copy file
```
rclone copy [Provider]:Test/1GB.txt
[Provider]:1GB.txt
```
- Delete file
```
rclone delete [Provider]:Test/1GB.txt
```

5. Experiment results and analysis

This section presents the results of performance comparison of cloud service providers. After presenting particular and average times for execution of each command listed in previous section, we also make some analysis of the different providers based also on their pricing plans. To perform the test a single 1GB file with randomly generated contents is used.

All times shown in the tables and figures in this section, as well as in the appendixare in the format minutes:seconds (mm:ss).

The experiment described in this section was undertaken under two important assumptions:

• We are going to test only free services delivered by cloud storage providers. This is an important assumption, because given cloud service provider may limit the resources available to their free tier services, while increasing or removing the said limit for the paid plans.

• Analysis of pricing plans of cloud storage providers has been made only about per month plans of each provider and for personal users only. It is important, because many providers may offer additional services on top of storage, which may influence the price of storage. Cloud providers also offer additional subscriptions, like annual ones, family plans, business, and enterprise plans, etc., which may vary significantly in terms of pricing.

5.1. Test Results

Results of the tests performed given the environment and methodology, described in section 4 are shown in Table 1.

Table 1

Performance results of Cloud storage providers

Operation	Average Times (mm:ss)		
	Google Drive	OneDrive	DropBox
Create share	00:01.7	00:02.4	00:01.3
List share	00:00.8	00:01.4	00:00.9
Move share	00:01.3	00:02.5	00:02.0
copy share	00:01.4	00:02.8	00:01.4
delete share	00:01.2	00:01.8	00:01.4
Upload	03:57.4	04:36.2	04:32.7
Download	05:41.3	06:06.8	02:30.1
Copy	00:04.7	00:06.7	00:02.9
Delete	00:02.2	00:03.1	00:01.7
Duration of all tests	09:52.0	11:03.8	07:14.3

As seen from Table 1, the performance of the three compared cloud storage providers is similar, with slight underperformance of OneDrive in Share/directory operations (Figure 3). However, performance of all three providers in upload/download is similar (Figure 4). More detailed table with test results is presented in Appendix 1.
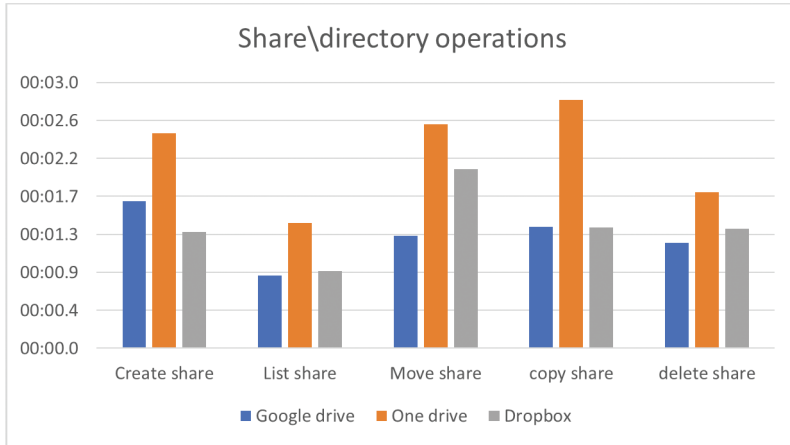
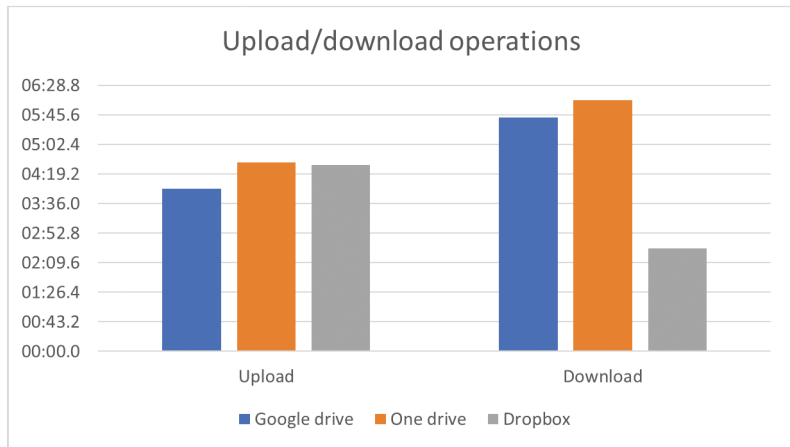**Figure 3:** Average time (mm:ss) of share and directory operations

Figure 4: Average time (mm:ss) of upload and download file operations (file size is 1 GB)

5.2. Pricing plans evaluation

All cloud storage providers have consumer storage plans and support different storage plans for business. Here we are going to focus on consumer storage plans. Please note that all prices refer to individual accounts and they are not options for businesses. Also depending on the plan every provider gives you bonus features that are not part of our research.

This research shows that, the pricing plans of tested cloud storage providers are almost the same. However, it should be noted that Google Drive offers the

largest storage space in their free plan. It is also one of the most generous cloud storage providers with their plans even if the free plan of the storage is shared between different services that they offer.

Table 2
Cloud storage providers pricing comparison

Google drive		OneDrive		Dropbox	
Storage	Price per month	Storage	Price per month	Storage	Price per month
15 GB	Free	5 GB	Free	2 GB	Free
30 GB	6$	100 GB	1.99$	2 TB	9.99$ starting
2 TB	12$	1 TB	6.99$	3 TB	16.58$ starting
5 TB	18$	6 TB	9.99$		

At first glance, Dropbox probably have the best pricing offers for bigger storage needs and offer the best price per space ratio. However, it should also be noted that most providers, together with the storage offer users also a large number of other services as well. This requires a more complex methodology and criteria for pricing comparison of cloud storage providers.

6. Conclusion

In terms of data-intensive systems, it is worth to be able to evaluate different storage options available for small business and individual users. In contemporary systems, most data is stored over the cloud using the services of different cloud storage providers. This paper presents a methodological framework to evaluate cloud storage providers in terms of their performance parameters. It also presents details on specific testing environment and results from testing the performance of three popular cloud providers that also offer free storage options. Additionally, a comparison of the pricing plans of these providers is performed; however, it is difficult to assess them in this respect, as most subscriptions include other service, besides storage.

It should be noted that a certain drawback of cloud solutions is represented by bandwidth limitations and the end user network is very important part of the cloud service. If the network is slow and unstable it may trouble accessing or sharing files and even, make impossible to work on this kind of environment. However, investigation on how end user network affect performance of cloud storage providers is part of our further research.

Directions for future research include:
• Increasing the comparison with more service providers
• Development of methodology for comparison of other quality characteristics of cloud storage providers like reliability, availability, security and

cost-effectiveness. It may also appear beneficial to define a compound measure for cloud storage quality of service, by combining the results of the various tests of such characteristics.

• The experiment may be expanded to include more diverse tests, for example with various file sizes, single transaction with large number of files (both small and large ones), and etc.

7. Acknowledgements

8. References

[1] Kleppmann, M. (2017). Designing Data-Intensive Applications. O'Reilly. Beijing.

[2] Hey, T., Tansley, S., Tolle, K. M. (2009). Jim Gray on eScience: a transformed scientific method. The Fourth Paradigm.

[3] Best cloud storage providers 2021, available at: https://cloudstorageinfo. org/top-10-cloud-storage-providers.

[4] Cloud Storage Comparison – Table Chart: https://www.goodcloudstorage. net/cloud-storage-comparison.

[5] H. Arif, H. Hajjdiab, F. A. Harbi and M. Ghazal, "A Comparison between Google Cloud Service and iCloud," *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 337-340, doi: 10.1109/CCOMS.2019.8821744.

[6] Roy, M., Singh, M. (2022). Performance Evaluation and Comparison of Various Personal Cloud Storage Services for Healthcare Images. In: Tavares, J.M.R.S., Dutta, P., Dutta, S., Samanta, D. (eds) Cyber Intelligence and Information Retrieval. Lecture Notes in Networks and Systems, vol 291. Springer, Singapore.

[7] Zenuni, Xhemal & Ajdari, Jaumin & Ismaili, Florie & Raufi, Bujar. (2014). Cloud storage providers: A comparison review and evaluation. 883. 272-277. 10.1145/2659532.2659609.

[8] Malla, S., & Christensen, K. (2020). HPC in the cloud: Performance comparison of function as a service (FaaS) vs infrastructure as a service (IaaS). Internet Technology Letters, 3(1), e137.

Appendix: Detailed results of performance comparison

Storage service	Operation	Command	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Avg Time
Google drive	Create share	rclone mkdir google-drive:Testdir	00:02.0	00:01.5	00:01.7	00:01.4	00:02.3	00:01.7	00:01.9	00:01.5	00:01.5	00:01.5	00:01.7
	List share	rclone lsf google-drive:Test	00:00.9	00:00.7	00:00.8	00:00.8	00:00.9	00:00.8	00:00.9	00:01.0	00:00.8	00:00.8	00:00.8
	Move share	rclone move google-drive:Test google-drive:Testdir	00:02.0	00:01.0	00:01.1	00:01.2	00:01.4	00:01.5	00:01.2	00:01.2	00:01.1	00:01.2	00:01.3
	copy share	rclone copy google-drive:Testdir google-drive:Test	00:02.8	00:01.0	00:01.1	00:01.1	00:01.4	00:01.5	00:01.3	00:01.2	00:01.2	00:01.2	00:01.4
	delete share	rclone purge google-drive:Testdir	00:01.1	00:01.0	00:01.1	00:01.0	00:01.4	00:01.2	00:01.1	00:01.6	00:01.3	00:01.1	00:01.2
	Upload	rclone copy /home/user/documents/1GB.txt google-drive:Test	03:44.6	03:43.9	03:38.0	04:17.5	03:43.3	03:36.1	03:54.9	04:53.3	04:18.5	03:43.8	03:57.4
	Download	rclone sync google-drive:Test /home/user/downloads	05:08.7	05:47.4	05:25.2	05:27.7	06:36.1	06:48.3	05:01.0	06:30.1	04:49.5	05:19.2	05:41.3
	Copy	rclone copy google-drive:Test/1GB.txt google-drive:1GB.txt	00:04.8	00:03.9	00:04.3	00:04.3	00:05.3	00:05.6	00:05.1	00:04.7	00:04.7	00:04.4	00:04.7
	Delete	rclone delete google-drive:Test/1GB.txt	00:02.0	00:02.1	00:02.1	00:02.3	00:02.1	00:02.5	00:02.4	00:02.2	00:02.4	00:02.3	00:02.2
	Duration of all tests		09:08.9	09:42.4	09:15.3	09:57.3	10:34.0	10:39.2	09:09.6	11:36.7	09:21.0	09:15.6	09:52.0
OneDrive	Create share	rclone mkdir one-drive:Testdir	00:00.8	00:02.6	00:02.1	00:02.0	00:03.6	00:02.5	00:02.2	00:02.5	00:03.3	00:02.8	00:02.4
	List share	rclone lsf one-drive:Test	00:01.1	00:01.3	00:01.9	00:01.3	00:01.4	00:01.2	00:01.6	00:01.1	00:01.7	00:01.5	00:01.4
	Move share	rclone move one-drive:Test one-drive:Testdir	00:02.9	00:02.3	00:02.2	00:02.9	00:03.3	00:02.5	00:02.1	00:02.8	00:02.5	00:02.0	00:02.5
	copy share	rclone copy one-drive:Testdir one-drive:Test	00:05.1	00:02.5	00:02.3	00:02.6	00:02.4	00:02.4	00:03.0	00:02.2	00:02.7	00:03.1	00:02.8
	delete share	rclone purge one-drive:Testdir	00:02.4	00:01.7	00:01.3	00:01.3	00:01.8	00:01.4	00:01.2	00:01.4	00:02.0	00:03.2	00:01.8
	Upload	rclone copy /home/user/documents/1GB.txt one-drive:Test	03:58.1	04:16.5	03:59.1	03:41.1	05:09.1	08:10.3	04:12.4	04:09.4	04:28.6	03:57.4	04:36.2
	Download	rclone sync one-drive:Test /home/user/downloads	02:37.4	04:48.9	04:36.6	05:27.5	21:14.6	04:56.6	05:01.0	04:25.0	05:07.0	02:53.6	06:06.8
	Copy	rclone copy one-drive:Test/1GB.txt one-drive:1GB.txt	00:08.4	00:05.4	00:07.0	00:05.9	00:06.7	00:07.0	00:06.0	00:06.0	00:06.6	00:07.7	00:06.7
	Delete	rclone delete one-drive:Test/1GB.txt	00:04.3	00:02.8	00:02.6	00:04.8	00:02.6	00:02.3	00:02.8	00:02.9	00:02.9	00:03.0	00:03.1
	Duration of all tests		07:00.5	09:24.0	08:55.1	09:29.4	26:45.3	13:26.2	09:32.4	08:53.2	09:57.4	07:14.4	11:03.8
Dropbox	Create share	rclone mkdir dropbox:Testdir	0m1.522s	00:01.5	00:01.6	00:01.7	00:01.5	00:01.4	00:01.3	00:01.5	00:01.3	00:01.4	00:01.3
	List share	rclone lsf dropbox:Test	00:00.8	00:00.8	00:01.0	00:01.0	00:00.8	00:00.9	00:00.8	00:01.0	00:00.9	00:00.9	00:00.9
	Move share	rclone move dropbox:Test dropbox:Testdir	00:01.9	00:01.8	00:02.0	00:04.6	00:01.6	00:01.5	00:01.7	00:01.6	00:01.7	00:01.9	00:02.0
	copy share	rclone copy dropbox:Testdir dropbox:Test	00:01.4	00:01.6	00:01.4	00:01.4	00:01.3	00:01.4	00:01.4	00:01.2	00:01.2	00:01.3	00:01.4
	delete share	rclone purge dropbox:Testdir	00:01.4	00:01.2	00:01.3	00:01.2	00:01.4	00:01.2	00:01.2	00:01.4	00:02.0	00:01.3	00:01.4
	Upload	rclone copy /home/user/documents/1GB.txt dropbox:Test	04:46.8	04:19.6	04:36.9	04:22.1	04:26.5	05:01.9	04:24.8	04:30.2	04:23.7	04:34.2	04:32.7
	Download	rclone sync dropbox:Test /home/user/downloads	02:12.3	02:31.1	02:23.6	02:19.2	02:15.7	03:18.0	02:20.1	02:38.4	02:42.2	02:20.1	02:30.1
	Copy	rclone copy dropbox:Test/1GB.txt dropbox:1GB.txt	00:02.7	00:03.2	00:03.0	00:03.2	00:03.0	00:02.6	00:02.7	00:02.9	00:02.7	00:02.6	00:02.9
	Delete	rclone delete dropbox:Test/1GB.txt	00:01.9	00:01.7	00:01.9	00:01.9	00:01.6	00:01.8	00:01.7	00:01.8	00:01.6	00:01.5	00:01.7
	Duration of all tests		07:09.2	07:02.6	07:12.8	06:56.3	06:53.6	08:30.6	06:55.9	07:20.0	07:17.2	07:05.1	07:14.3