Prerequisites

Remarks:

- Parallel means multithreading (not multiprocessing) with 50 threads
- WRITE profile is a batch write (multiple sequential write) considering no data were previously in datastore
- READ profile is a batch read (multiple sequential read) considering no data were previously in datastore
- WRITE/READ profile is a random read/write considering that all data were written previously in datastore
- The time is the arithmetic mean between 3 trials and is in seconds

Requirements:

- Python 3.5
- Ubuntu 18.04 with mc (minio command line)
- MinIO RELEASE.2020-03-14T02-21-58Z
- Python lib h5py: 2.10.2
- Python lib google-cloud-storage: 1.26.0
- Python lib minio: 5.0.8

HDF5 vs POSIX sequential vs parallel

	Sequential Posix	Sequential HDF5	Parallel POSIX	Parallel HDF5
WRITE				
1000 tiny files	0.05	0.5	0.06	0.4
10 000 tiny files	0.6	7.2	0.73	4.2
100 000 tiny files	7.2	70	10.46	57
READ				
1000 tiny files	0.03	0.10	0.05	0.13
10 000 tiny files	0.35	1.2	0.6	1.5
100 000 tiny files	33	11	20	19
READ/WRITE				
1000 tiny files	0.09	0.45	0.16	0.30
10 000 tiny files	0.99	5.3	1.5	4.05
100 000 tiny files	23	54	25	40

MinIO vs GCS

Prerequisites:

- To upload file directly to GCS, we used the gcs python blob.upload
  - ➔ Not tested with gsutil rsync (directory) or gsutil cp (file)
- To download file directly to GCS, we used gsutil rsync with recsurivity (directory)
  - ➔ Not tested with blob.download
- To upload file through MinIO, we used fput_object
  - ➔ Not used mc mirror (directorty)
- To download file through MinIO, we used get_object
  - ➔ Not used mc cp recursive or mc mirror

A tiny one is 1 MB

	Sequential GCS	Parallel GCS	Sequential MinIO GCS	Parallel MinIO GCS
WRITE				
1 tiny file	0.12		1.23	
10 tiny files	1.05	0.28	9.93	1.10
100 tiny files	11,03	1.62	93.06	3.85
READ				
1 tiny file	2,91		1.01	
10 tiny files	27.61	3.80	9.96	1.02
100 tiny files	126,45	93.20	94.17	2.12
READ/WRITE				
10 tiny files	8.3	2.49	9.73	1.0
100 tiny files	74.38	44.80	94.12	2.30

A medium one is 10 MB

	Sequential GCS	Parallel GCS	Sequential MinIO GCS	Parallel MinIO GCS
WRITE				
1 medium file	0.58		0.23	
10 medium files	4.71	0.65	4.26	0.84
100 med files	43.26	4.85	24.28	4.49
READ				
1 medium file	2,05		1.41	
10 medium files	13.90	5.25	12.9	4.94
100 med files	137.80	430.76	133.62	306.527
READ/WRITE				
10 medium files	9.07	2.61	8.22	3.07
100 med files	88.84	156.94	86.11	146.07

A big one is 100 MB

	Sequential GCS	Parallel GCS	Sequential MinIO GCS	Parallel MinIO GCS
WRITE				
1 medium file	0.98		18.82	
5 medium files	6.51	4.27	86.44	17.62
READ				
1 medium file	2,99		1.06	
5 medium files	8.92	7.85	4.87	1.07
READ/WRITE				
1 medium files				
5 med files	9.23	4.07	53.04	16.64