

POSIX-like access via HTTP: OIDC AuthN/AuthZ solutions provided for research communities

Ahmad Alkhansa (ahmad.alkhansa@cnaifnfn.it)

Diego Ciangottini (ciangottini@pg.infn.it)

Alessandro Costantini (alessandro.costantini@cnaifnfn.it)

Federico Fornari (federico.fornari@cnaifnfn.it)

Jacopo Gasparetto (jacopo.gasparetto@cnaifnfn.it)

Diego Michelotto (diego.michelotto@cnaifnfn.it)

Carmelo Pellegrino (carmelo.pellegrino@cnaifnfn.it)

Massimo Sgaravatto (massimo.sgaravatto@pd.infn.it)

Daniele Spiga (daniele.spiga@pg.infn.it)

The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this license or copyright law is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this license.



Context

- Several **emerging use cases** of experiments/collaborations asking for **local POSIX access** to storage:
 - WLCG experiments would like to **access cloud storage** resources in a **POSIX-like** way
 - Multiple solutions are available (**Ceph, S3, CVMFS, CernBOX**), but which is the most suitable?
 - Internal and external projects where INFN is involved
- Need to test proper technologies/services
- **POSIX-like access via HTTP**
 - Take into account not only solutions working for S3 (i.e. WebDAV)
- Provide a GP WebApp acting as a GUI

Tested Solutions

Server side



INDIGO-IAM used as
AuthN/AuthZ service



Client side



- <https://github.com/s3fs-fuse/s3fs-fuse>
- operate files in S3 bucket like a local file system using FUSE



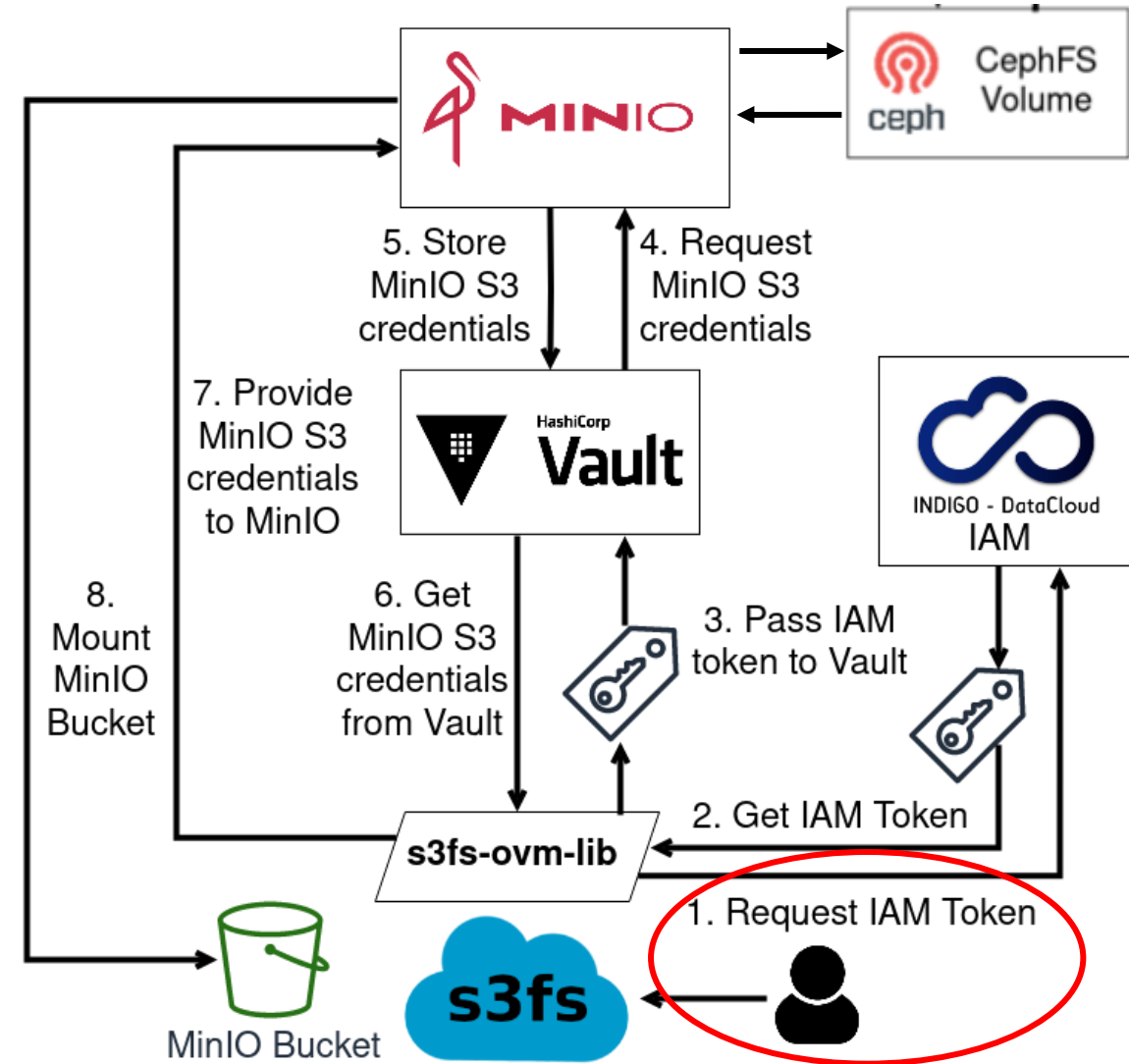
Rclone and s3fs-fuse

Limitations on tested solutions

- **Rclone**
 - **S3 credentials valid for 1h** using Rclone as it is
 - Considering the use of a client application to **automatically refresh temporary S3 credentials**
- **S3fs-fuse**
 - Does not support WebDAV protocol

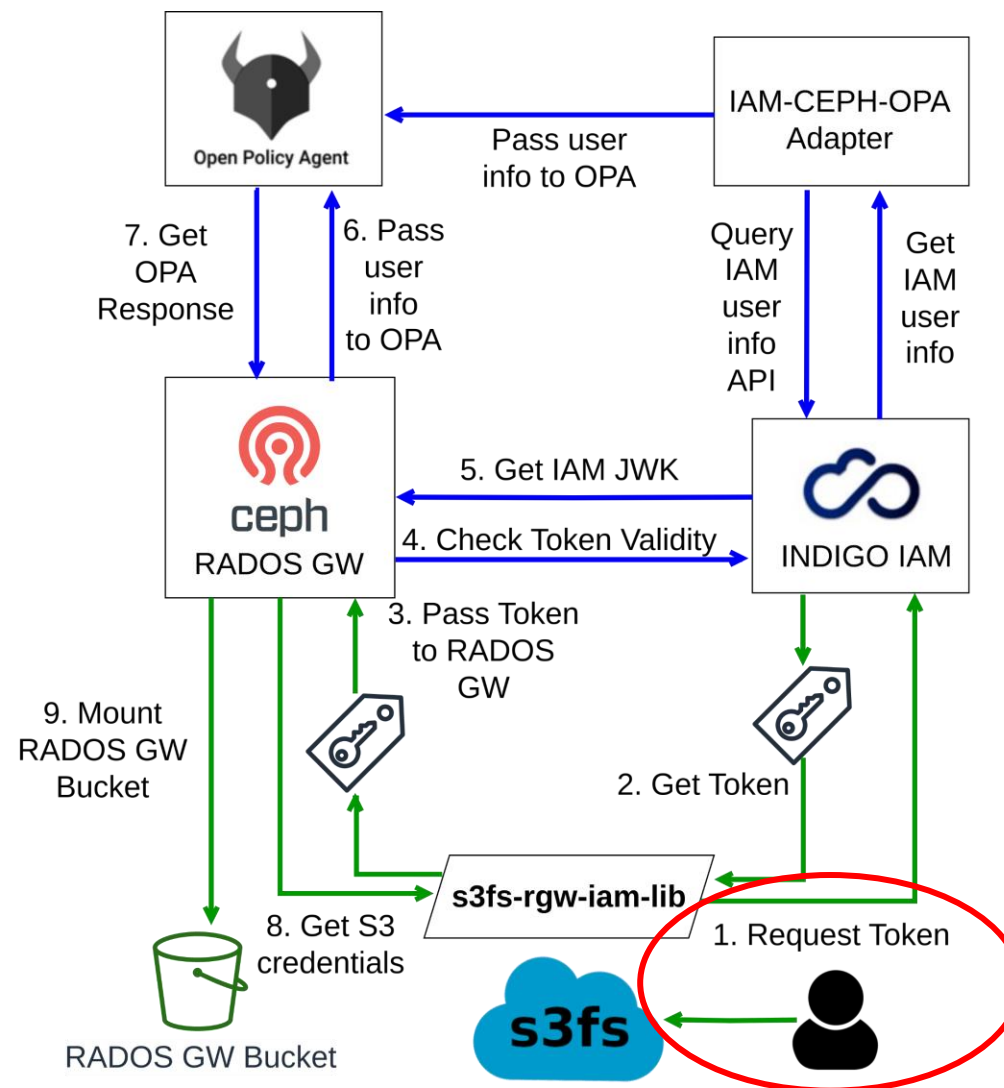
MinIO

- **Hashicorp Vault** interact with **MinIO** to get temporary **S3 credentials**
 - can be configured to be accessed through **OpenID Connect provider**
 - can **supply Secure Token Service (STS)** functionality for MinIO
- **s3fs** with an INFN plugin for
 - **oidc-agent** C++ API to get an **access token** from Indigo **IAM**
 - **Vault** C++ API to obtain **S3 temporary credentials** from **MinIO**
- A **policy** must be defined in **MinIO** and is **linked** to a **Vault role** to perform operations on **buckets** based on IAM token **groups** claim value



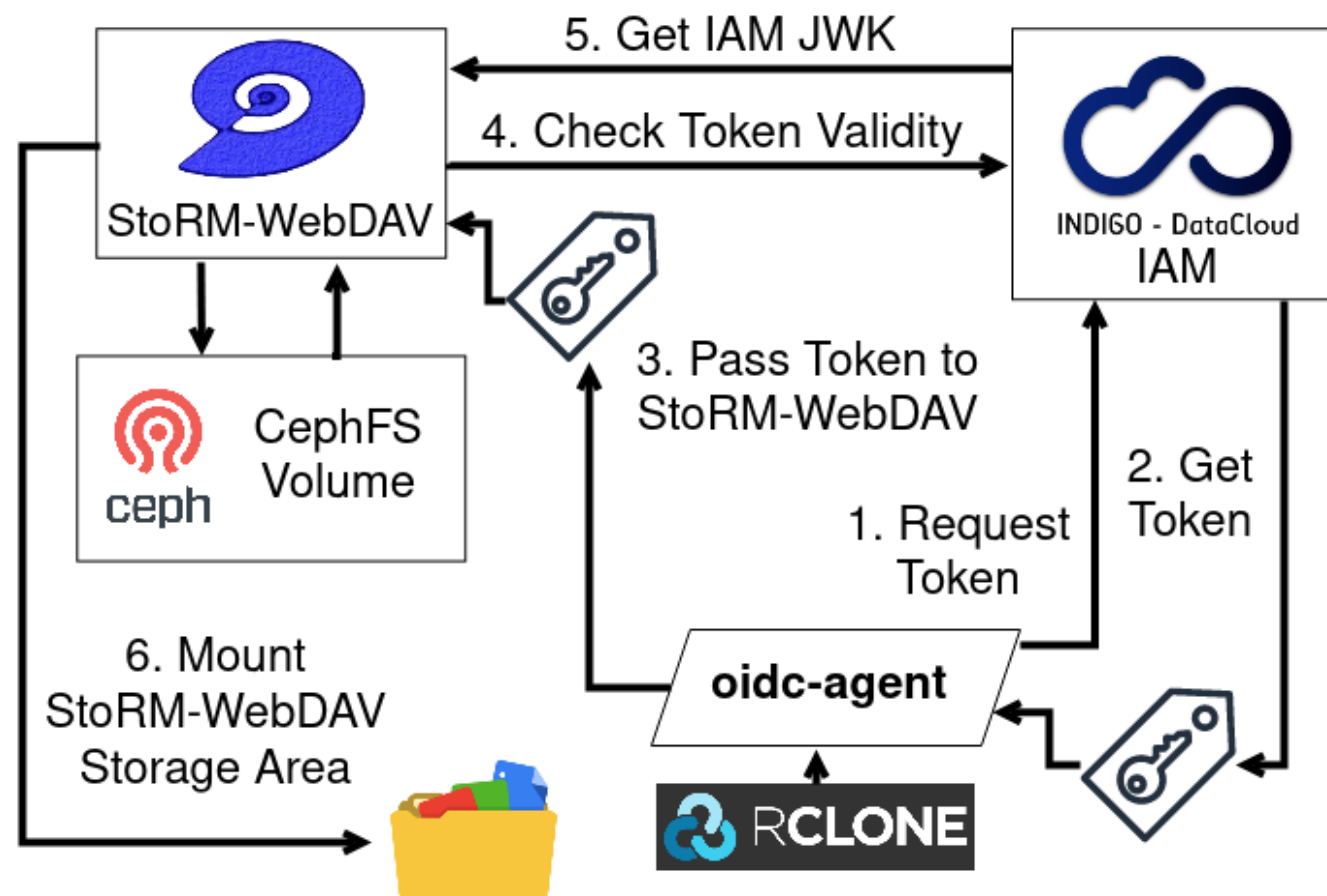
CEPH RGW

- **s3fs** with a INFN plugin developed for IAM **AuthN/AuthZ** with Rados Gateway.
 - The library retrieves IAM access token for performing **STS with RGW**.
- **RGW** validates the token with IAM then sends an **authorization request to OPA**.
- **OPA's response** depends on the content of **RGW input**, existing **policies** and stored **user-related** information received from the **adapter**.



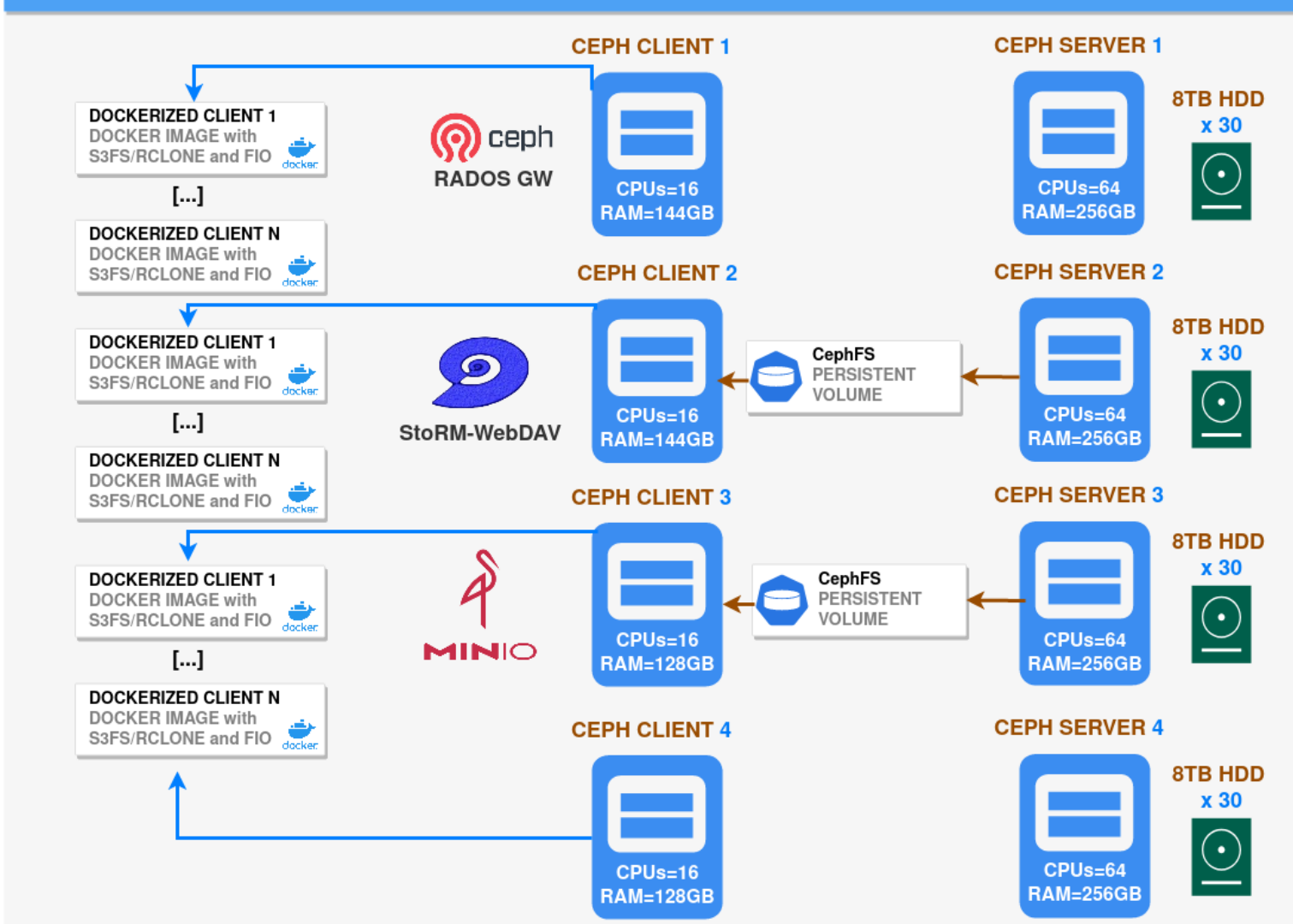
StoRM-WebDAV

- **Rclone** can mount a **StoRM-WebDAV** storage area (SA) providing **POSIX** access
 - For **WebDAV** remote storage, **Rclone** allows the user to provide a command (**oidc-agent**) for the application to **automatically** renew tokens
- **StoRM-WebDAV** exports data from **POSIX** file system (**CephFS**)
- **no object storage**



Scalability Tests – Testbed Setup

ARCHITECTURE: CEPH BARE-METAL CLUSTER

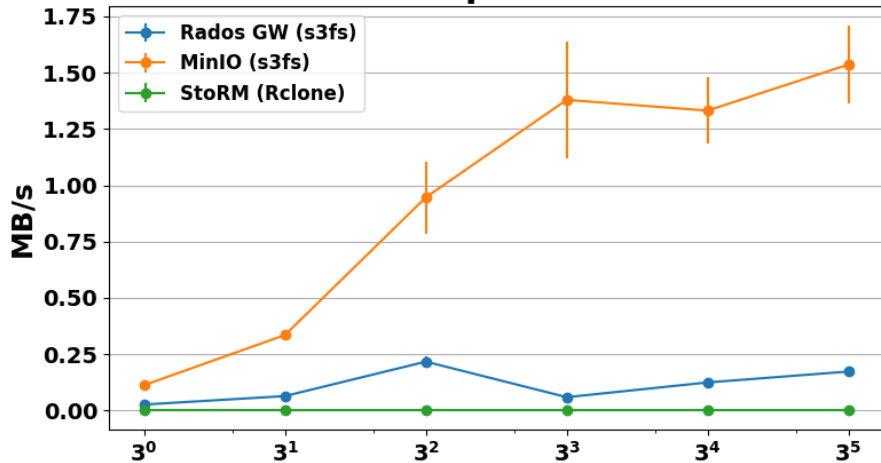


- Ceph testbed:
 - 4 server nodes
 - 4 client nodes
 - 2x10 Gbit NIC per node
 - 120 8TB HDD
- 3 Ceph client nodes host gateway services:
 - Rados GW
 - MinIO
 - StoRM-WebDAV
- 4 Ceph client nodes host client containers with s3fs/Rclone to mount personal buckets/storage areas and with fio to perform tests

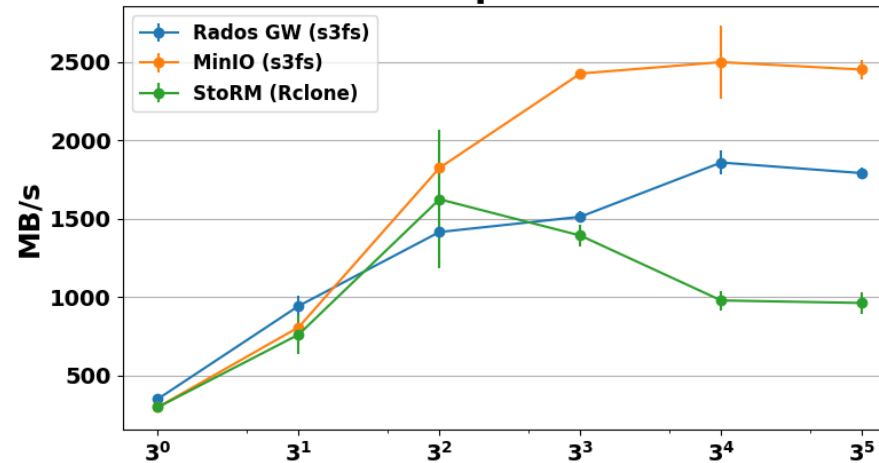
Scalability Tests – Server Side Results

Average Throughput Comparison - Server

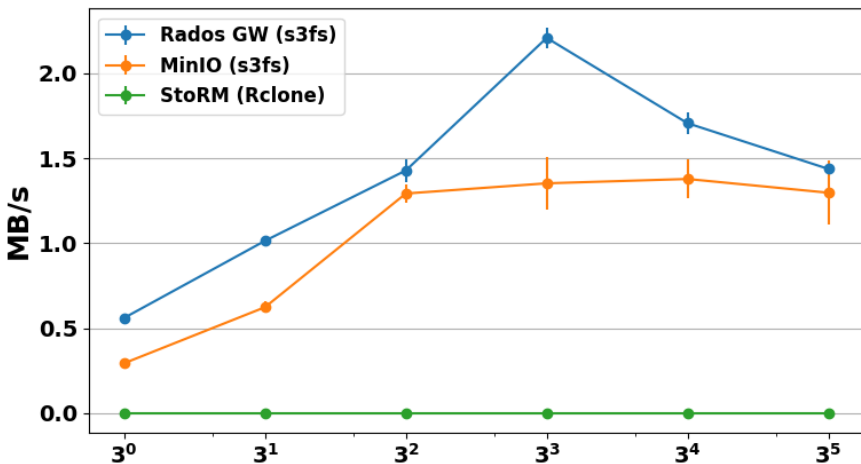
Seq Read



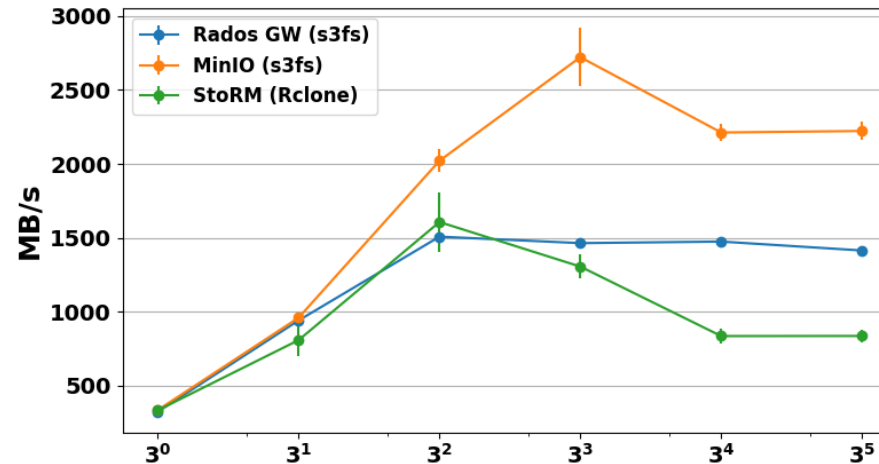
Seq Write



Rand Read



Rand Write

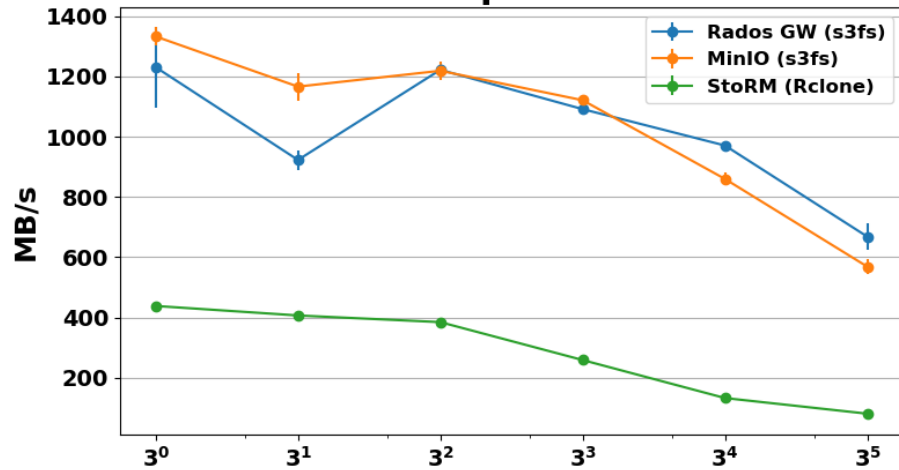


- Each **point** in the plots consists of the **mean** and relative **error of 5 runs**
- Each **run** is a **fiio** sequential/random write/read of a single **O(GB) file per client**
- **Throughput** from the server side

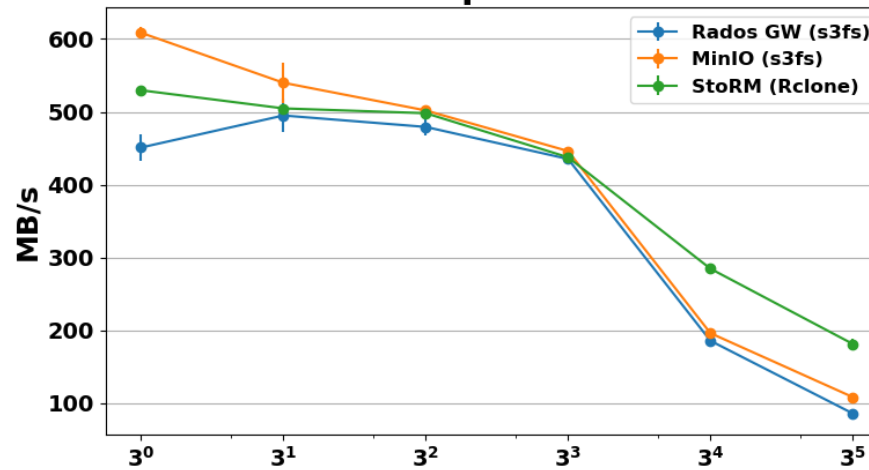
Scalability Tests – Client Side Results

Average Throughput Comparison - Client

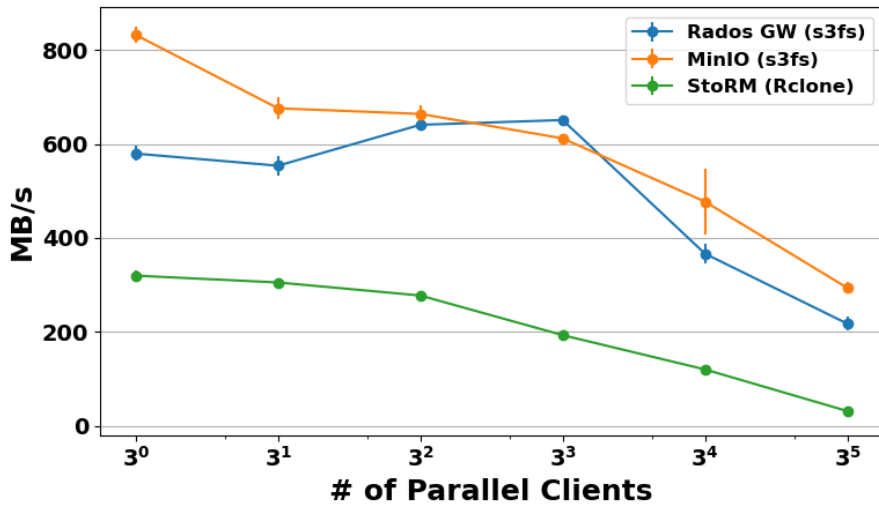
Seq Read



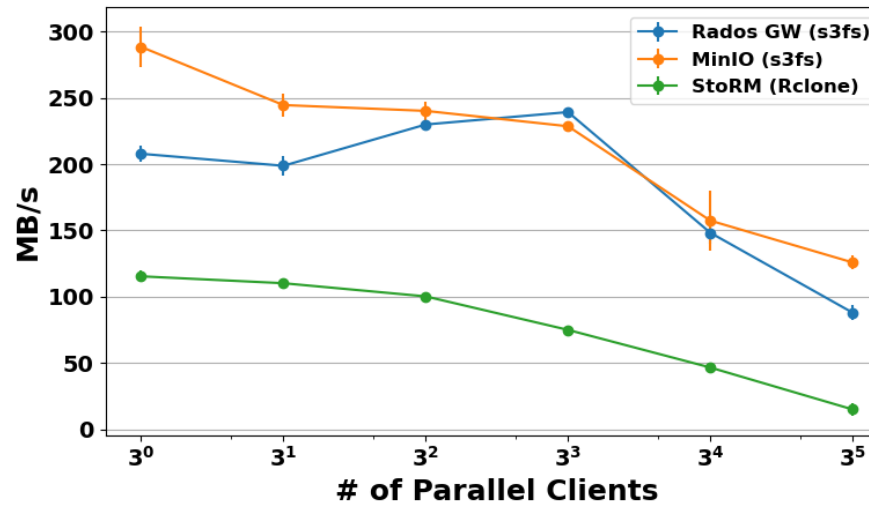
Seq Write



Rand Read



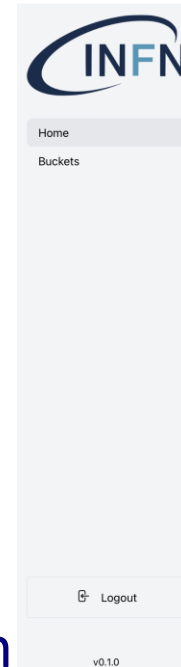
Rand Write



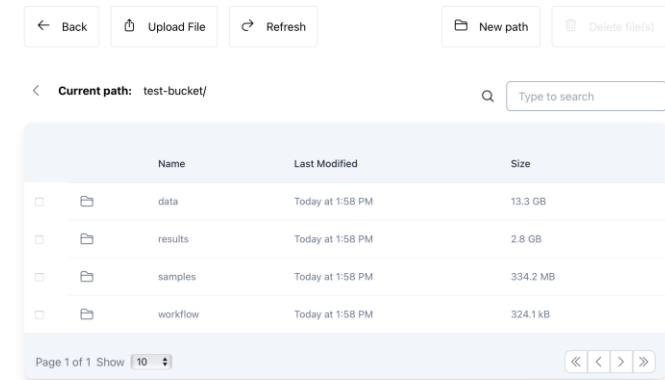
- Throughput seen by **fiio** during the same tests
- In general, **s3fs** (cache-enabled) yields **better results** w.r.t. Rclone
- **MinIO and CEPH RGW** generally shows comparable performance
- **Rclone + StoRM-WebDAV** shows lower values

Web Application

- Need of a GP Web Application able to
 - Use API to interact (at present) with S3 (CEPH RGW, can be extended)
 - Support OpenID Connect and OAuth2.0
 - Simple to deploy and to modify
- Used technologies
 - **React** library for web user interfaces
 - **FastAPI** framework for building APIs with Python
 - **AWS SDK for JavaScript** to create, configure, and manage buckets.

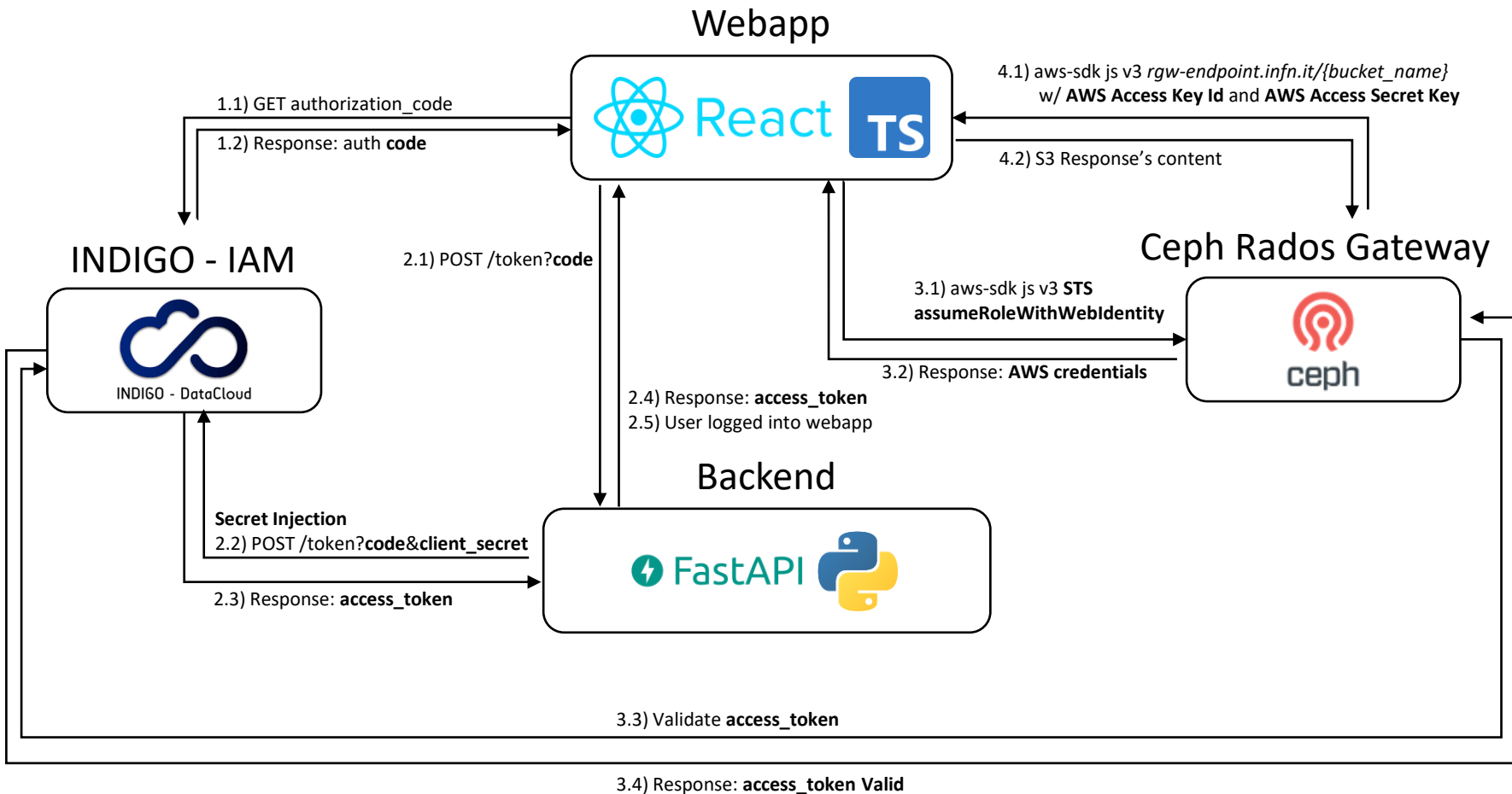


test-bucket



| | Name | Last Modified | Size |
|--------------------------|----------|------------------|----------|
| <input type="checkbox"/> | data | Today at 1:58 PM | 13.3 GB |
| <input type="checkbox"/> | results | Today at 1:58 PM | 2.8 GB |
| <input type="checkbox"/> | samples | Today at 1:58 PM | 334.2 MB |
| <input type="checkbox"/> | workflow | Today at 1:58 PM | 324.1 kB |

Web Application



- **S3 operations** performed with **official AWS SDK (v3)** for JavaScript
- **Direct access** with Access Key Id and Access Secret Key
- **Authentication** via Indigo IAM using STS AssumeRoleWithWebId entity

Conclusions

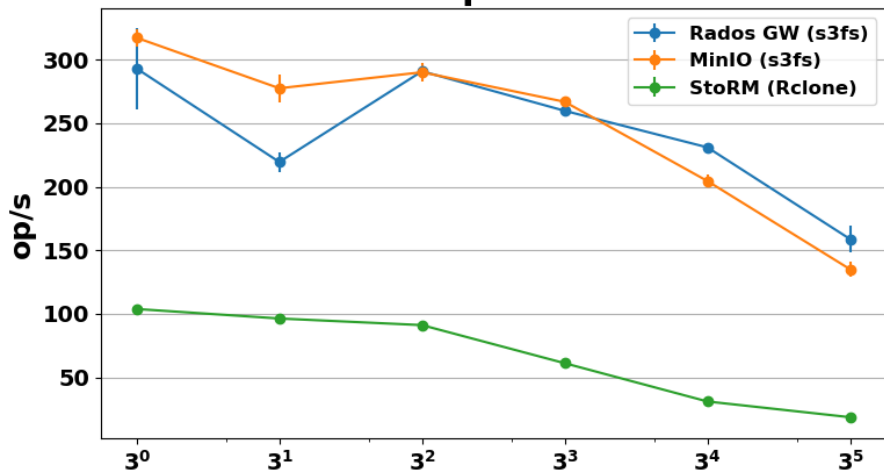
- **s3fs-fuse** seems to be a **promising** application to **support** the **remote** storage local **mount** with OpenID Connect AuthN/AuthZ mechanism
- **Rclone** shows **not exciting** performance **out of the box** with respect to s3fs-fuse
- **MinIO** and **CEPH RGW** seems to have comparable performances
 - The internal expertise on CEPH propend to adopt CEPH as S3 solution
- A WebApp (under development) will act as a **GUI** to interact with S3
- **Future**
 - Increase quality and quantity of the tests
 - Properly tune Rclone and/or involve **alternative WebDAV** storage services for **Rclone** (e.g. **ownCloud**)
 - Extend the WebApp functionalities to support other storage services

THANK YOU!

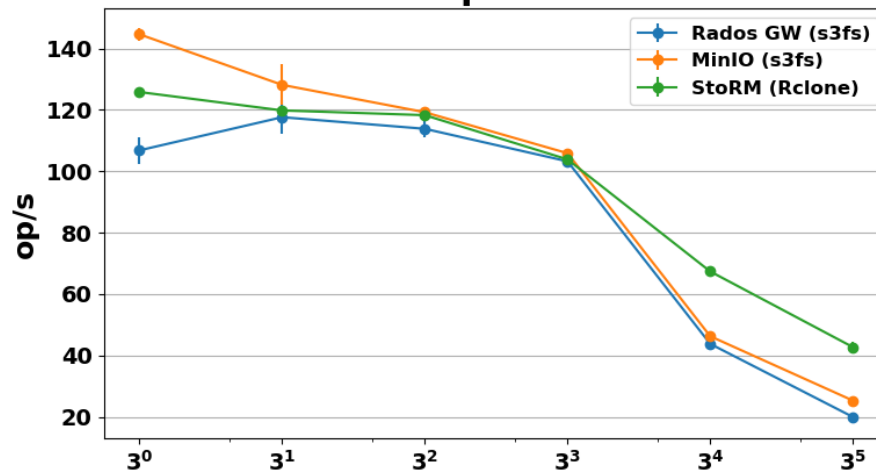
Scalability Tests – Client Results

Average IOPS Comparison - Client

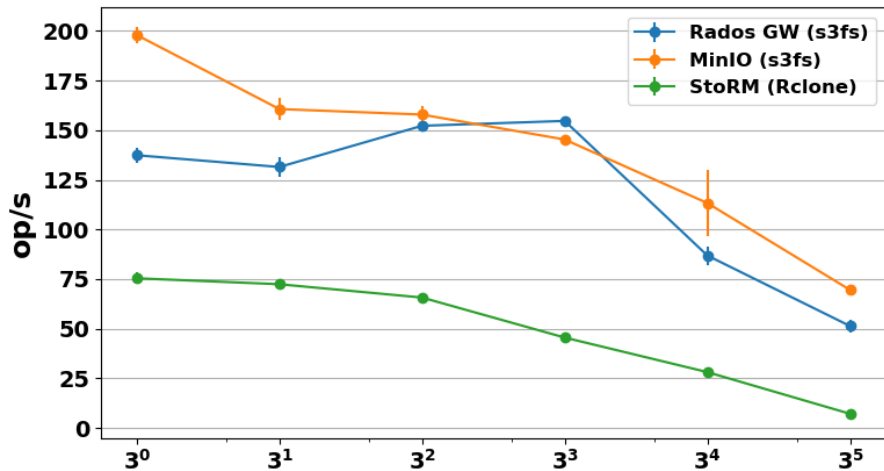
Seq Read



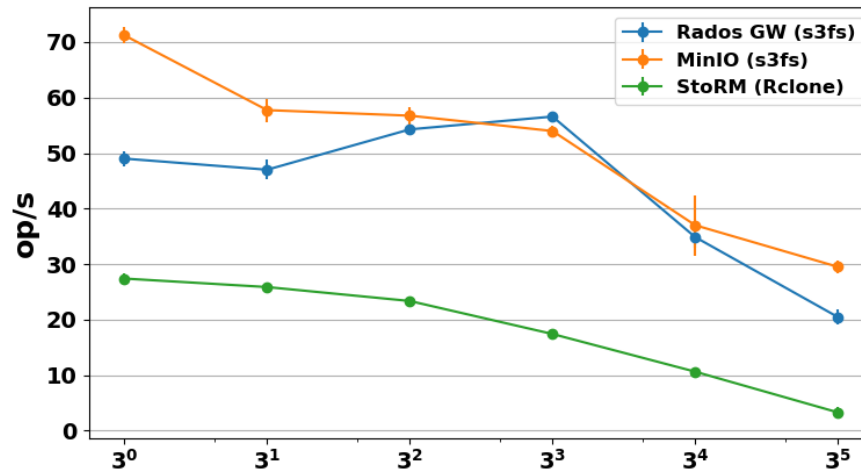
Seq Write



Rand Read



Rand Write



of Parallel Clients

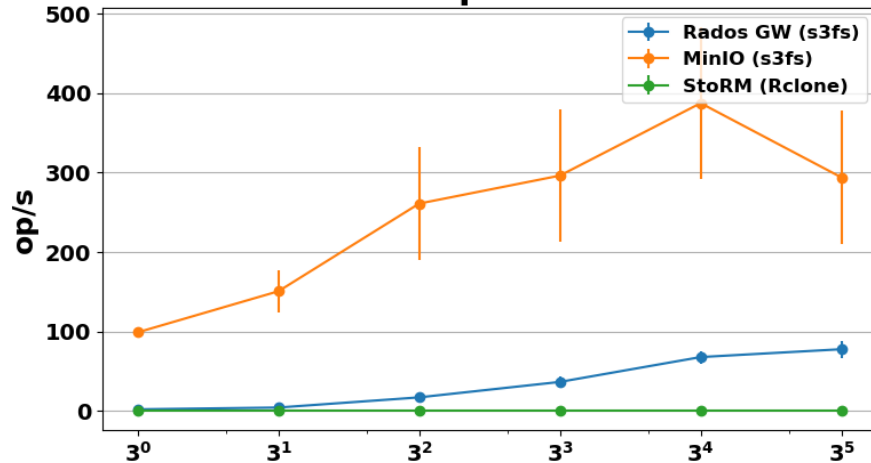
of Parallel Clients

- Each **point** in the plots consists of **mean** and **relative error of 5 runs**
- Each **run** is a **fiio** sequential/random write/read of a single **O(GB)** file **per client**
- These are the **IOPS** seen by **fiio** during the performance tests

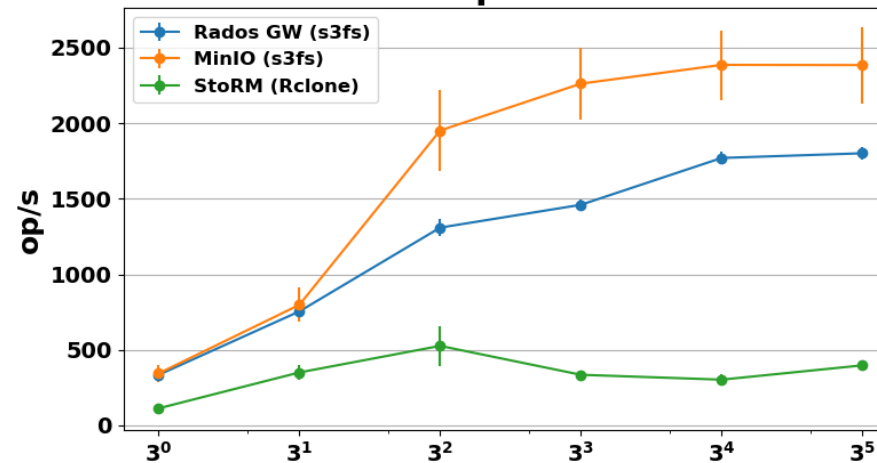
Scalability Tests – Server Results

Average IOPS Comparison - Server

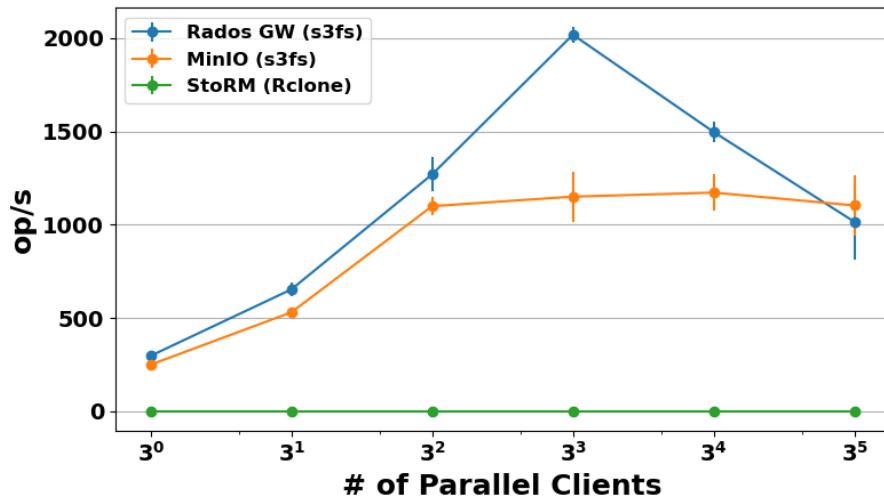
Seq Read



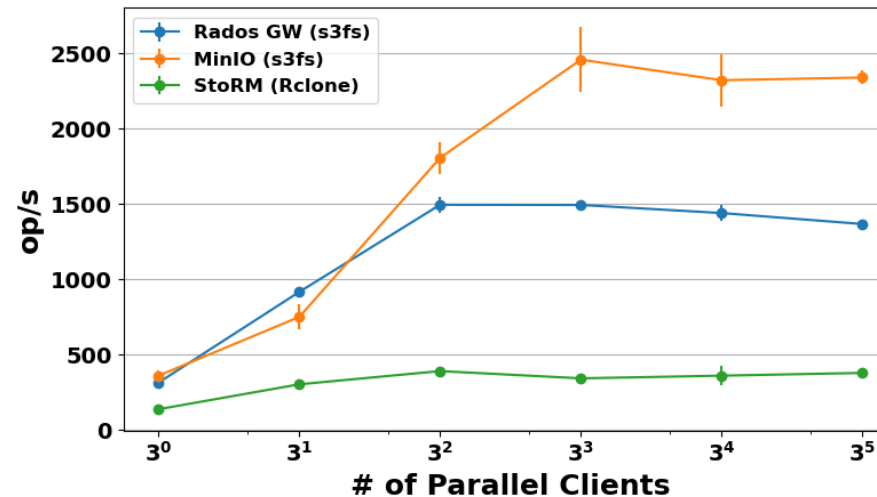
Seq Write



Rand Read



Rand Write



- Each **point** in the plots consists of **mean** and **relative error of 5 runs**
- Each **run** is a **fiio** sequential/random write/read of a single **O(GB)** file **per client**
- These are the **IOPS** seen by Ceph cluster during the tests for the interested **Ceph pool**