# POSIX access to remote storage via OpenID Connect

*Federico* Fornari[1,*], *Ahmad* Alkhansa[1], *Alessandro* Costantini[1], *Carmelo* Pellegrino[1], and *Davide* Salomoni[1]

[1]INFN-CNAF, viale Berti Pichat 6/2, Bologna, Italy

**Abstract.** INFN-CNAF is one of the Worldwide Large Hadron Collider Computing Grid (WLCG) Tier-1 data centers, providing support in terms of computing, networking, storage resources and services also to a wide variety of scientific collaborations, ranging from physics to bioinformatics and industrial engineering [1]. Recently, several collaborations working with our data center have developed computing and data management workflows that require access to S3 storage services and/or the integration with POSIX capabilities. Nevertheless, the access to the data must be regulated by federated authentication and authorization mechanisms, such as OpenID Connect (OIDC), which is largely adopted by communities like WLCG [2] and within the European Open Science Cloud (EOSC) [3]. In the present work, the possibility to regulate POSIX access by integrating JSON Web Token (JWT) [4] authentication, provided by INDIGO-IAM as Identity Provider [5], with solutions based on S3 (for object storage) and WebDAV (for hierarchical storage) protocols has been evaluated an related results have been reported. In such respect, a comparison between the performance yielded by S3 and WebDAV protocols has been carried out within the same distributed environment with the aim to better identify the solution most suitable for the different use cases.

## 1 Introduction

The INFN-CNAF Tier-1 data center hosted in Bologna, Italy, provides computational assets to numerous research groups spanning High-Energy Physics (such as the LHC experiments at CERN), Astroparticle Physics, Gravitational Waves, Nuclear Physics, and related domains. INFN-CNAF data center serves also as a hub for various scientific endeavors ranging from chemistry to biology, to health science and so on. Within these communities, an increasing number of use cases is expressing the desire to access cloud storage resources in a POSIX-like manner. This reflects the evolving landscape of data storage and management, with cloud resources becoming increasingly integrated into research workflows. It is worth mentioning here that other use cases with similar requirements are foreseen in the near future.
At the time being, the INFN Cloud infrastructure, where CNAF is one of the major contributors, provides access to cloud object-storage resources by exposing S3 buckets via MinIO Gateway [6], whereas Tier-1 hierarchical storage resources are exposed through StoRM WebDAV [7] via WebDAV protocol. These are non-POSIX protocols, and the authentication/authorization mechanism is managed in both cases with OpenID Connect (OIDC) via

*e-mail: federico.fornari@cnaf.infn.it

the INDIGO-IAM application. Moreover, CNAF is involved in the replacement of MinIO, and a good candidate seems to be Ceph [8] RADOS Gateway (RGW) which has been included in the integration tests presented in the following sections.

## 2 Materials and Methods

As already pointed out in the previous section, there is an increasing number of use cases which require to access cloud storage resources in a POSIX-like manner. To do that, different technologies have been selected and integrated. Anyway, it is worth to highlight that these technologies should be used carefully when dealing with global data namespaces (like scratch buckets or experiments' storage areas) since they do not provide native features to isolate client environments. Access Control Lists or applications like Open Policy Agent (OPA) [9] are valuable tools to implement the desired level of security management.

### 2.1 Mount of S3 bucket via s3fs-fuse

MinIO and Ceph RADOS Gateway (RGW) have been integrated with s3fs-fuse libraries for Secure Token Service (STS) support via OpenID Connect for S3 data access. Both comply with STS, providing temporary S3 credentials via JSON Web Tokens (JWT). MinIO's native STS API lacks INDIGO-IAM JWT profile support, addressed by delegating STS to HashiCorp Vault [10]. The need to use custom shared libraries for authentication and authorization management, both for MinIO and Ceph RGW, brought us to the adoption of s3fs-fuse, a widely used open-source command-line tool designed for the management of object storage files in a POSIX-like way.
In the case of MinIO, a C++ plugin named s3fs-ovm-lib [11] responsible of handling credential processing for s3fs-fuse has been developed. It accomplishes this purpose by exploiting the following components:

- oidc-agent [12] C++ API, used to retrieve an access token from INDIGO-IAM;
- HashiCorp Vault [13] C++ API, employed to acquire temporary S3 credentials from MinIO.

s3fs-ovm-lib can update expired S3 credentials through a new INDIGO-IAM token request (see Figure 1 for more details). The user does not need to worry about the validity period of a single token (which is usually 1 hour), since the application handles new token requests automatically after credentials expiration.
In the case of Ceph RGW, an additional C++ credential plugin, named s3fs-rgw-iam-lib [14], has been developed, as well, to automatically handle INDIGO-IAM authentication by retrieving INDIGO-IAM access tokens, which are then provided to Ceph RGW when an S3 operation is requested. During this process, RGW validates the authenticity of the INDIGO-IAM token. To authorize the S3 operations and make them compliant with the role of the user registered in INDIGO-IAM, Open Policy Agent (OPA) has been installed and integrated with Ceph RGW. To ensure OPA remains up-to-date with the latest user information registered in INDIGO-IAM, an IAM-Ceph-OPA Adapter application [15] has been implemented. Upon receiving an affirmative response from OPA, Ceph RGW provides s3fs-fuse with temporary S3 credentials, allowing the successful mount of the specified Ceph RGW bucket (see Figure 2 for details).

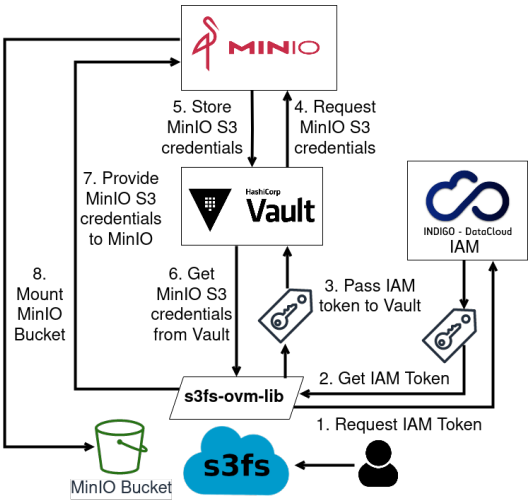


**Figure 1.** The architecture and related workflow followed by s3fs-fuse to mount a MinIO bucket in user space.

### 2.2 Mount of WebDAV storage area via Rclone

INFN-CNAF offers WebDAV as the predominant data transfer technology, by providing a non-POSIX storage solution based on StoRM WebDAV and INDIGO-IAM services for Tier-1 resources. The WebDAV data access has been managed by using Rclone [16] to mount a storage area via INDIGO-IAM token authentication. In this case the storage area is exposed by the StoRM WebDAV application, but the solution is general enough to be used with other WebDAV data management servers (e.g. Apache, NGINX).
Rclone is a versatile tool in the realm of data management, and provides the capability to mount a remote storage area via WebDAV protocol, thereby extending POSIX access to StoRM WebDAV storage system. By leveraging oidc-agent application, Rclone guarantees the automatic renewal of INDIGO-IAM tokens, facilitating authentication and authorization processes when interacting with StoRM WebDAV. The setup adopted for the present work has been made up by coupling StoRM WebDAV to a CephFS POSIX file system (see Figure 3 for more details).

## 3 Scalability Tests

The solutions presented in Sect. 2.1 and Sect. 2.2 have been evaluated within a common distributed environment consisting of a 8-node Ceph (v17.2.5) cluster. This robust infrastructure encompasses a cumulative processing power of 320 CPUs and 1.5TB of RAM, with 4 nodes designated as clients and 4 nodes acting as servers. The connection among the Ceph cluster's nodes is made through the implementation of a 2×10 Gbit/s bonding channel configuration. Each of the 4 server nodes is equipped with 30×8 TB rotational disks connected via a Serial Attached SCSI interface from 2 separate JBODs, each embedding an array of 60 rotational disks. The setup collectively contributes to a comprehensive capacity of 120 disks, with a total raw storage space of 960TB. The system relies on distinct storage pools for Ceph RGW and for CephFS, both with the standard replica 3 policy implemented.

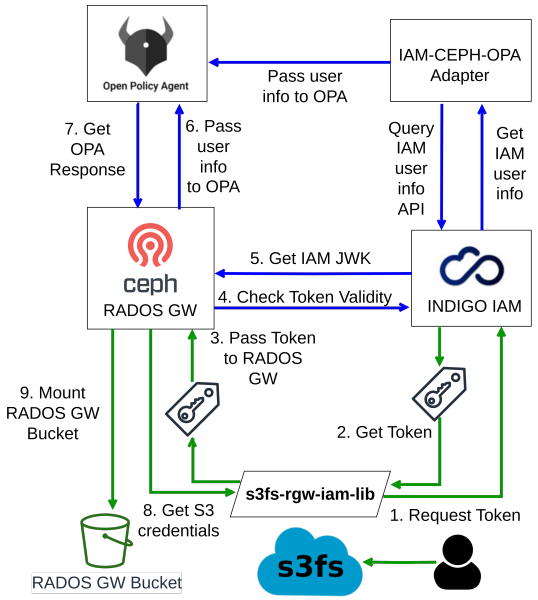


**Figure 2.** The architecture and related workflow followed by s3fs-fuse to mount a Ceph RGW bucket in user space. Green arrows highlight user-side interactions, while blue arrows indicate intra-services interactions.

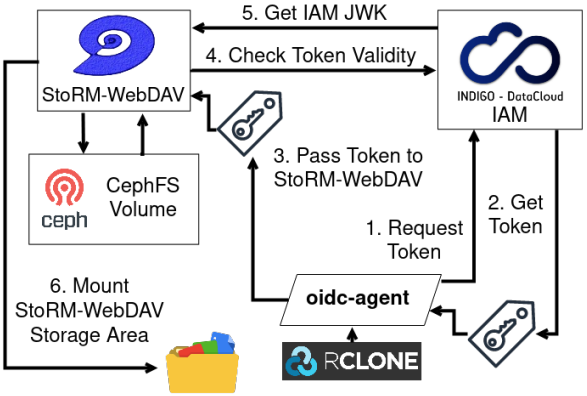


**Figure 3.** The architecture and related workflow followed by Rclone to mount a StoRM WebDAV storage area in user space.

Three out of four Ceph client nodes are dedicated to hosting gateway services for Ceph RGW, MinIO Gateway (over CephFS) and StoRM WebDAV.

Each gateway has been tested separately by running client containers on the complementary Ceph client nodes. These containers are equipped with s3fs-fuse and Rclone, which enable the possibility of mounting buckets and storage areas respectively, and also include the Fio [17] tool used to carry out the comprehensive performance tests (see Figure 4 for more details).

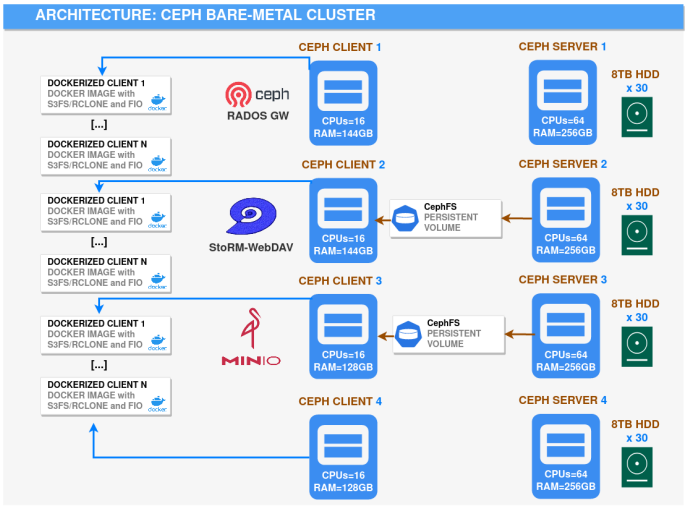


**Figure 4.** Schema of the testbed used to compare the performance of the considered storage solutions.

### 3.1 Results

The performance indicators considered for the tests are IOPS (input/output operations per second) and throughput expressed in MB per second. For each indicator, sequential and random reads/writes have been performed by running Fio and involving a single 1GB-sized file per container. While client side results are directly reported by Fio application, server side scores are measured through Prometheus service deployed to monitor the whole Ceph cluster. IOPS results are reported in Figure 5, while throughput results are grouped in Figure 6.

Comparing the different storage solutions, the following interesting observations arise:

- client vs. server results: as the number of parallel clients increase, server-side metrics rise, contrasting with declining client-side metrics; server-side measurements aggregate across all clients, while client-side metrics are client-specific; with a growing number of containerized clients, fixed CPU and memory resources on each of the four nodes lead to each client accessing a diminishing share of computing power despite the increased overall amount of data operations;
- server-side read operations are hindered by abundant RAM acting as cache, consistently restricting disk access and leading to persistently low read throughput despite a high IOPS rate; a validation test on MinIO with CephFS showed that flushing RAM brought read throughput closer to write throughput, although these results are not displayed here; on client side, s3fs-fuse (cache-enabled) outperforms Rclone, demonstrating its caching mechanism's effectiveness with 3-4 times better read throughput results;
- MinIO performace: MinIO over CephFS exhibits slightly superior performance on client side compared to Ceph RGW: this can be due to fundamental architectural differences, since the interaction between RGW and the RADOS system is mediated by librados, a component that enables an interface to Ceph monitors and OSDs, while CephFS does not rely on intermediate layers to contact RADOS daemons;
- Rclone performance: when evaluating performance using Rclone in conjunction with StoRM WebDAV, it becomes evident that the results are generally less favorable compared to using s3fs-fuse.

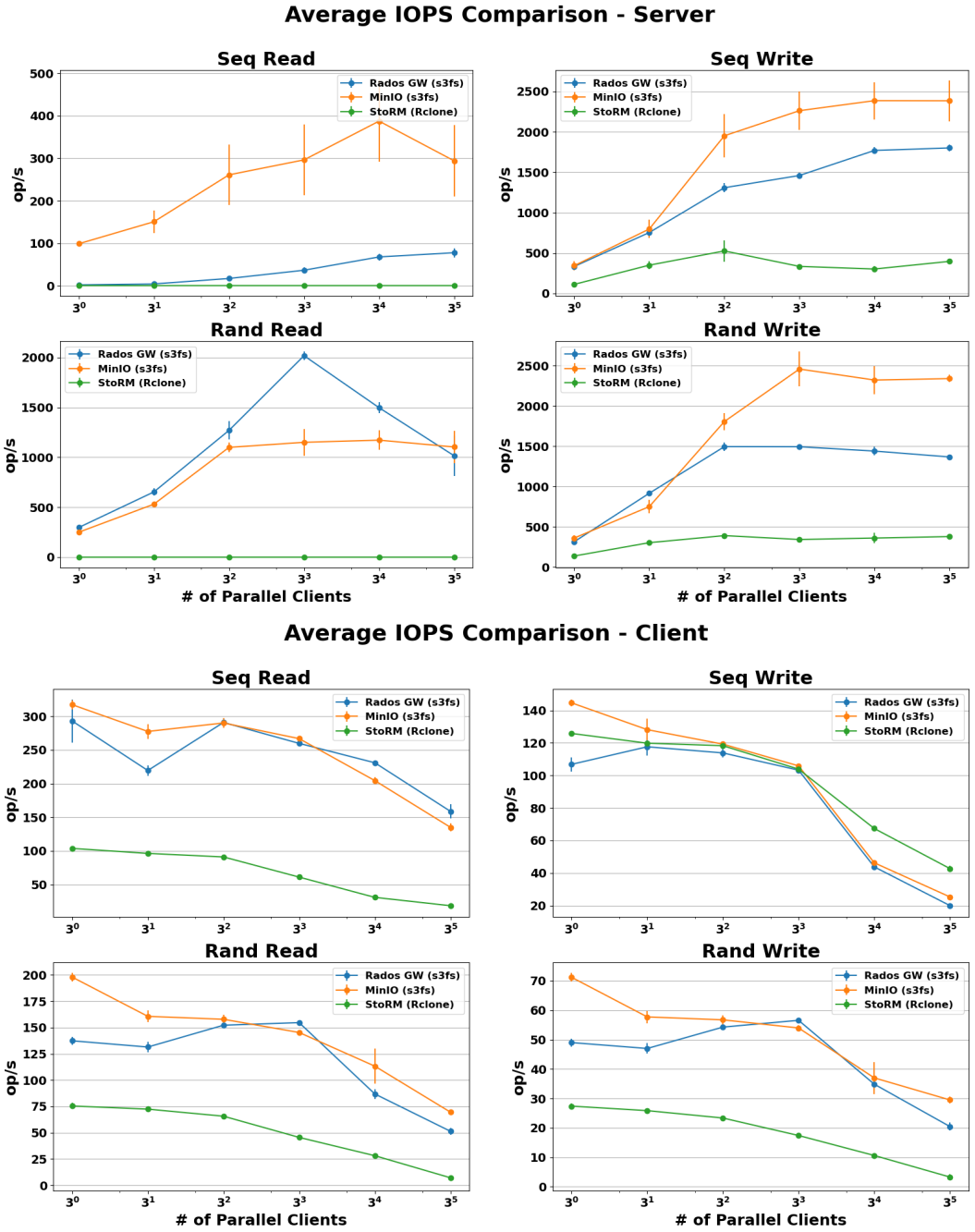


**Figure 5.** IOPS results on server (top) and client (bottom) side from the performance tests.

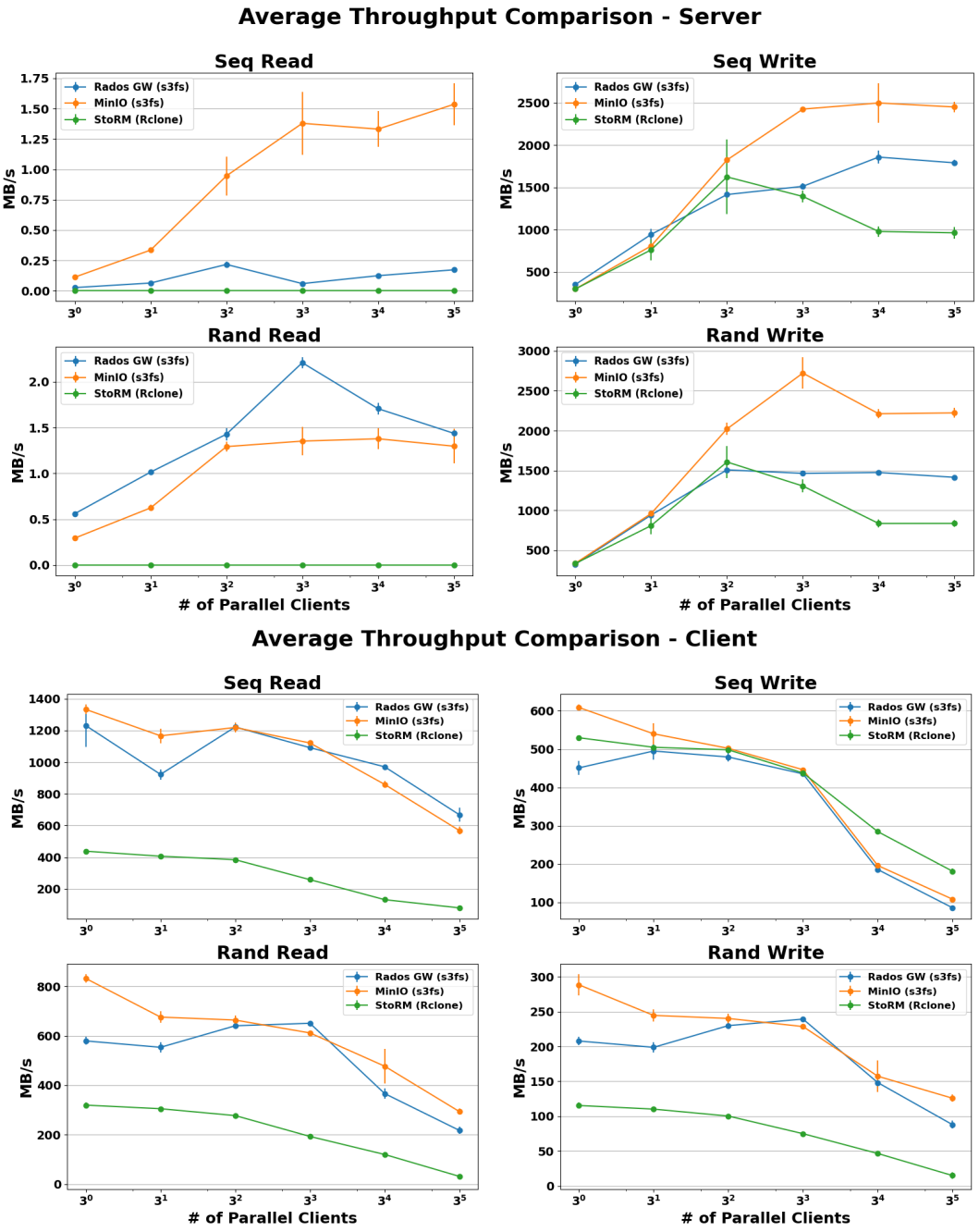


**Figure 6.** Throughput results on server (top) and client (bottom) side from the performance tests.

## 4 Conclusions

During the exploration of the presented storage solutions and their performance characteristics, several noteworthy findings have emerged. Firstly, the s3fs-fuse application is a promising candidate, showing its potential to facilitate remote storage local mounts while leveraging the OpenID Connect authentication/authorization mechanism. On the other hand, while Rclone offers flexibility through tunable parameters, it is important to note here that its out-of-the-box performance falls short when compared to the capabilities of s3fs-fuse. This implies that additional effort may be required to optimize the use of Rclone. When examining combinations of storage components, the interaction between MinIO and CephFS emerges as a contender, demonstrating support for slightly higher throughput in comparison to the Ceph RGW counterpart.
In the next future, we can explore further by increasing the number of client nodes for scalability insights, identifying potential bottlenecks. We may also evaluate alternative WebDAV storage services like ownCloud and Nextcloud to better understand Rclone's capabilities and limitations in different contexts.

## References


[1] D. Cesini, L. dell'Agnello, B. Martelli, G. Maron, L. Morganti, D. Salomoni, *CNAF - Towards the challenges of the coming decade* (PRINGO srl, Roma, 2021)

[2] B. Bockelman, A. Ceccanti, T. Dack et al., EPJ Web Conf. **251**, 02028 (2021)

[3] K. Wierenga, L. Johansson, C. Kanellopoulos et al., *EOSC Authentication and Authorization Infrastructure (AAI): Report from the EOSC Executive Board Working Group (WG) Architecture AAI Task Force (TF)* (European Commission (EU), 2021), ISBN 978-92-76-28113-9, 46.21.02; LK 01

[4] M.B. Jones, J. Bradley, N. Sakimura, *JSON Web Token (JWT)*, RFC 7519 (2015), https://www.rfc-editor.org/info/rfc7519

[5] A. Ceccanti, E. Vianello et al., *indigo-iam/iam: INDIGO Identity and Access Management Service v1.8.0* (2022), https://doi.org/10.5281/zenodo.7065682

[6] Harshavardhana et al., *MinIO* (2022), https://github.com/minio/minio

[7] A. Ceccanti, E. Vianello, M. Caberletti, *italiangrid/storm-webdav: Storm webdav v1.4.1* (2021), https://doi.org/10.5281/zenodo.4890906

[8] S. Weil et al., *Ceph* (2022), https://github.com/ceph/ceph

[9] T. Sandall et al., *OPA* (2022), https://github.com/open-policy-agent/opa

[10] W. Hearn, Z. Seguin, *vault-plugin-secrets-minio* (2022), https://github.com/StatCan/vault-plugin-secrets-minio

[11] F. Fornari, *s3fs-oidc-vault-minio-lib* (2023), https://github.com/ffornari90/s3fs-oidc-vault-minio-lib

[12] G. Zachmann, D. Dykstra, L. Marschke et al., *indigo-dc/oidc-agent: oidc-agent 4.2.6* (2022), https://doi.org/10.5281/zenodo.5846932

[13] J. Mitchell et al., *Vault* (2022), https://github.com/hashicorp/vault

[14] F. Fornari, *s3fs-rgw-iam-lib* (2023), https://github.com/ffornari90/s3fs-rgw-iam-lib

[15] A. Alkhansa, *rgw-opa-iam-adaptor* (2023), https://github.com/ahmadalkhansa/IAM-Ceph-OPA-Adaptor

[16] N. Craig-Wood et al., *Rclone* (2022), https://github.com/rclone/rclone

[17] J. Axboe, *Flexible I/O Tester* (2022), https://github.com/axboe/fio