Whitepaper

High Performance Computing

Einfacher Zugriff für die gesamte Organisation



https://opencloud.eu

OpenCloud

Inhaltsverzeichnis

OpenCloud: Einfacher Zugang zu HPC-Daten für die ganze Organisation 3

Zugang zu Cluster-Daten: zwei Notlösungen 4

Das HPC-Add-on als Zugriffsschicht 5

Zugänge schaffen, Kontrolle behalten 6

Vorteile im Betrieb 7

Das HPC-Add-on im Überblick 8

Souveränität durch offene Standards 9

OpenCloud: Einfacher Zugang zu HPC-Daten für die ganze Organisation

Bei Simulationen auf Rechenclustern, an Sequenzierern und Elektronenmikroskopen oder anderen Messgeräten fallen erhebliche Datenmengen an, meist mehrere Petabyte. Hochschulen, Forschungseinrichtungen und wissenschaftliche Institute legen die entstehenden Daten oft auf großen, parallelen Dateisystemen ab, wie sie für HPC-Umgebungen (High Performance Computing) typisch sind.

Wer darauf zugreifen möchte, benötigt normalerweise ein eigenes Benutzerkonto. Alle anderen haben keinen Zugriff: vom Sekretariat über Partnerinstitute bis hin zu externen Kolleg*innen. Um die Daten zugänglich zu machen, kopieren viele Organisationen sie in ein zweites, browserbasiertes System oder richten VPN- und Shell-Zugänge ein. Beide Wege haben jedoch handfeste Nachteile.

Genau hier setzt das **HPC-Add-on für OpenCloud** an: Es legt sich als Zugriffsschicht direkt auf das vorhandene parallele Dateisystem und nutzt es als Speicher-Backend. Neu geschriebene Dateien erscheinen unmittelbar in OpenCloud und sind über Browser, Desktop-Clients, mobile Apps und Freigabe-Links erreichbar – ohne VPN und ohne Kommandozeile. So arbeiten alle an einem einzigen Datenbestand, von den Forschenden am Cluster bis zu den externen Partner*innen.

OpenCloud basiert auf einem Fork der Open Source-Software „ownCloud Infinite Scale“ (OCIS), dessen Komponenten u.a. von Entwickler*innen der Science-Organisation CERN sowie anderen Aktiven mitentwickelt wurden. OpenCloud wird nun von der Heinlein Gruppe mit neuen Ideen und einem klaren Fokus auf Datenschutz, Interoperabilität und nachhaltige Digitalisierung weiterentwickelt.

Zugang zu Cluster-Daten: zwei Notlösungen

In der Praxis haben sich zwei Wege etabliert. Beide lösen das Zugangsproblem nur scheinbar:

Cluster-Dateisystem · Petabytes
CephFS · GPFS · Leil FS · nur aus dem Cluster-Netz

1 Daten auf ein zweites System kopieren



Doppelter Speicherbedarf
Die Kopie belegt denselben Platz wie das Original, inklusive Backup.



Verzögerungen
Während des Kopiervorgangs kann niemand produktiv mit den Daten arbeiten.



Versionschaos
Original und Kopie laufen mit der Zeit unweigerlich auseinander.

2 VPN- und Shell-Zugänge verteilen



Hohe Einstiegshürde
Wer mit der Kommandozeile nicht vertraut ist, bleibt außen vor.



Sicherheitsrisiko
Externe Partnerinstitute benötigen einen VPN-Zugang ins interne Netzwerk.



Schatten-IT
Ist der Weg zu kompliziert, weichen viele auf USB-Datenträger oder private Cloud-Konten aus.

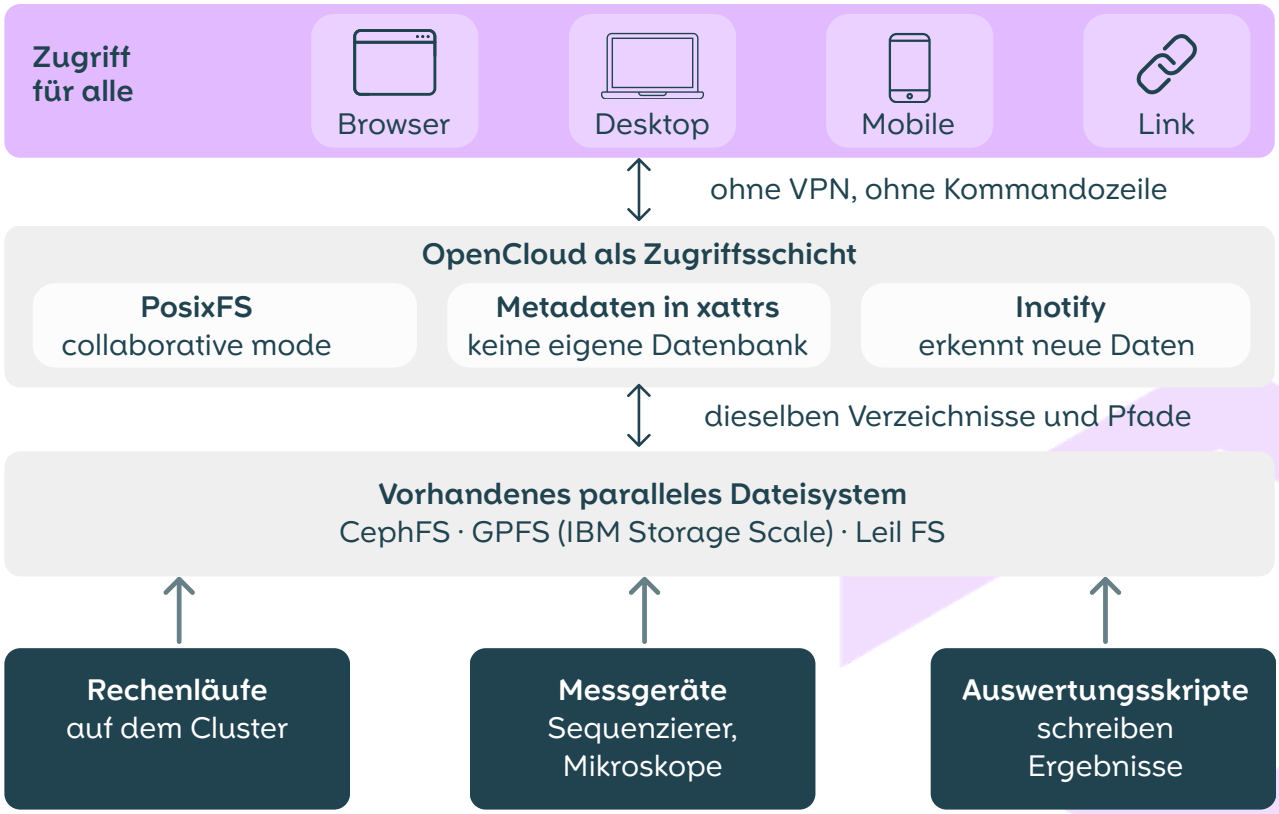


Das Ergebnis: Unkontrollierte Datenwege
Beide Herangehensweisen verschieben das eigentliche Problem, statt es zu lösen.

Das HPC-Add-on als Zugriffsschicht

Das Add-on arbeitet direkt auf dem Dateisystem und nutzt dafür den Speichertreiber PosixFS im Collaborative Mode. Dieser Modus erlaubt es, dass neben OpenCloud auch andere Prozesse denselben Verzeichnisbaum beschreiben. OpenCloud greift also auf exakt dieselben Verzeichnisse und Pfade zu wie die Rechenläufe, Messgeräte und Auswertungsskripte. Eine Kopie oder Migration der Daten erübrigt sich damit.

Auch die Metadaten bleiben auf dem Dateisystem: OpenCloud legt sie in den erweiterten Attributen (xattrs) der jeweiligen Dateien ab. Neu geschriebene und modifizierte Dateien erkennt OpenCloud über Inotify, eine Schnittstelle des Linux-Kernels, die Änderungen im Dateisystem überwacht und Programme ohne messbare Verzögerung informiert. Ergebnisse aus Rechenläufen und Messungen sind daher sofort im Browser oder in anderen grafischen Clients verfügbar, ohne dass OpenCloud das Dateisystem durchsuchen muss.



Zugänge schaffen, Kontrolle behalten

Damit sind die Daten sichtbar – es bleibt allerdings zu klären, wer Zugriff darauf erhält. Anwender*innen sollen mit den Daten arbeiten können, ohne dass die Organisation ihr Netzwerk öffnet. OpenCloud trennt daher Zugangswege, Freigaben und Rechte sauber voneinander.

1 Zugriff über vertraute Wege

Für die Arbeit mit den Cluster-Daten braucht es weder einen Shell-Zugang noch Kenntnisse der Kommandozeile. Benutzer*innen arbeiten über die Weboberfläche oder eigene Desktop-Clients für Windows, Linux und macOS, dazu kommen Apps für iOS und Android. Darunter liegt WebDAV, das Dateizugriff und Synchronisierung übernimmt. Für die Office-Integration kommt WOPI hinzu. Wer die Daten weiterverarbeiten möchte, bindet sie per WebDAV direkt ins eigene Betriebssystem ein.

Freigaben innerhalb und außerhalb der Organisation 2

Innerhalb der Organisation teilen Anwender*innen Daten mit anderen registrierten Benutzerkonten. Für alle anderen erzeugen sie Links, über die Personen ohne eigenes Konto auf Dateien zugreifen können. Die Einrichtung von VPN- und SSH-Zugängen ist damit nicht mehr erforderlich. Link-Freigaben lassen sich bei Bedarf abschalten.
Partnerinstitute mit eigenen OpenCloud-Instanzen oder kompatiblen Plattformen wie Nextcloud oder ownCloud profitieren vom OCM-Protokoll (Open Cloud Mesh). So geben sie Dateien standortübergreifend frei, auch über Organisationsgrenzen hinweg. Vor der ersten Freigabe steht ein Einladungsprozess, der auf vertrauenswürdige Instanzen beschränkt ist. Welche das sind, legt die Organisation selbst in der Konfiguration fest.

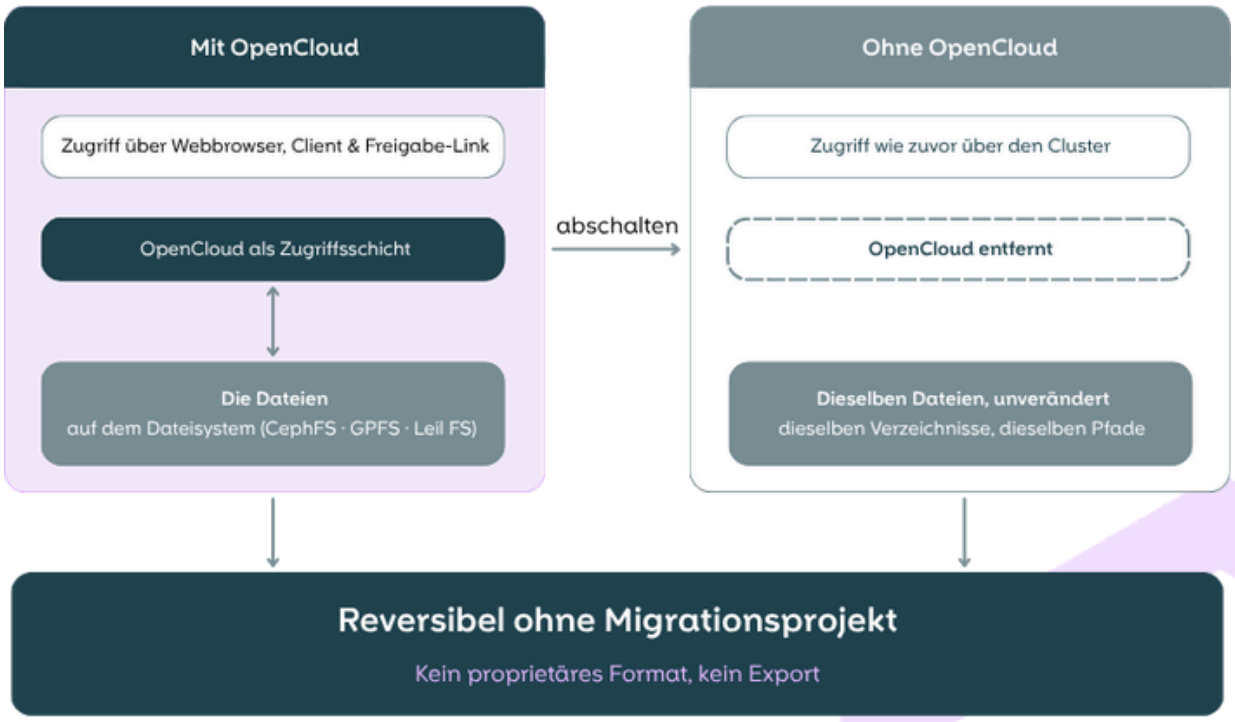
3 Anmeldung und Rechteverwaltung

Die Anmeldung erfolgt über OIDC und OAuth2, mit den Konten und Gruppen aus der vorhandenen Identitätsverwaltung. Alternativ lässt sich die eingebaute Nutzerverwaltung von OpenCloud verwenden.
Die OpenCloud-Rechteverwaltung kennt drei verschiedene Rollen. Nutzerrollen gelten systemweit und regeln die Verwaltung von Konten, Gruppen und Spaces. Innerhalb eines Spaces greifen die Space-Rollen. Davon unabhängig bestimmen Freigabe-Rollen, was Empfänger*innen mit geteilten Dateien und Ordnern anfangen dürfen, ohne dass sich ihre Rolle im System oder in einem Space ändert.

Vorteile im Betrieb

Die OpenCloud-Architektur vereinfacht den laufenden Betrieb:

- **Ein einziger Datenbestand:** Keine Kopie, kein zusätzliches Backup, keine Wartezeit durch Kopiervorgänge. Externe Forschende erhalten schneller Zugang.
- **Weniger Schatten-IT:** Ein Zugang ohne Shell oder VPN, dafür keine USB-Datenträger und private Clouds. Daten bleiben im kontrollierten Bereich der IT.
- **Reversibel ohne Migration:** Nach dem Abschalten liegen die Daten weiterhin als normale Dateien auf dem Dateisystem, in denselben Verzeichnissen und mit denselben Pfaden. Kein proprietäres Format, kein Export.



Unterstützte Dateisysteme

Das HPC-Add-on unterstützt drei parallele Dateisysteme, die in Rechenzentren etabliert sind. CephFS ist im Hochschul- und Forschungsumfeld weit verbreitet und in vielen Organisationen als Speicherbackend vorhanden. In größeren HPC-Umgebungen dominiert GPFS, das IBM seit einigen Jahren unter dem Namen Storage Scale führt. Hinzu kommt Leil FS (ehemals SaunaFS), ein verteiltes, POSIX-kompatibles Dateisystem, das unter der GPL-3.0 steht. Für Lustre und BeeGFS gibt es derzeit keine Anbindung.

Das HPC-Add-on im Überblick

Vorteil	Beschreibung
Direkter Dateisystem-Zugriff	OpenCloud arbeitet als Zugriffsschicht auf denselben Verzeichnissen und Pfaden wie Rechenläufe, Messgeräte und Auswertungsskripte.
Kein zweiter Datenbestand	Keine Kopie, kein zusätzliches Backup, keine Wartezeit durch Kopiervorgänge.
Sofort sichtbar	Neu geschriebene Dateien erkennt OpenCloud über Inotify und zeigt sie ohne separaten Suchlauf an.
Zugriff ohne Shell und VPN	Weboberfläche, Desktop-Clients für Windows, Linux und macOS, Apps für iOS und Android, WebDAV.
Freigaben nach außen	Links ohne eigenes Konto der Empfänger*innen, standortübergreifend über OCM mit Partnerinstituten.
Anmeldung und Rechte	OIDC und OAuth2 mit den Konten aus der vorhandenen Identitätsverwaltung, getrennte Nutzer-, Space- und Freigabe-Rollen.
Etablierte Dateisysteme	CephFS, GPFS (IBM Storage Scale) und Leil FS.
Kein Lock-in	Nach dem Abschalten bleiben die Daten normale Dateien auf dem Dateisystem, kein proprietäres Format, kein Export.

Souveränität durch offene Standards

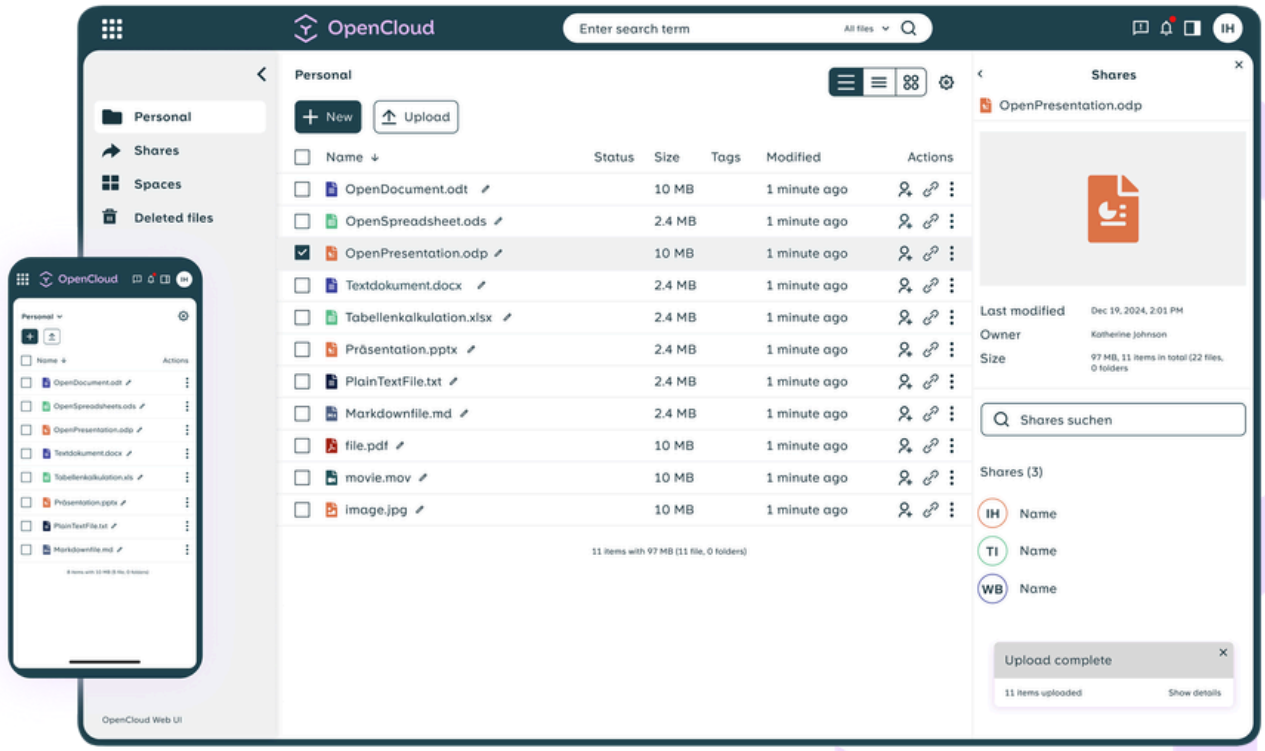
Mit dem HPC-Add-on bleiben die Daten, wo sie sind. OpenCloud verändert weder ihre Struktur noch ihr Format, sondern legt sich als Zugriffsschicht darüber – unabhängig davon, welche Dateisysteme laufen. Wer Zugriff auf die Cluster-Daten bekommt, entscheidet die Organisation.

Offene Standards machen das möglich: POSIX-kompatible Dateisysteme, WebDAV für den Dateizugriff, OIDC und OAuth2 für die Anmeldung, OCM für Freigaben über Organisationsgrenzen hinweg. OpenCloud ist Open Source, alle Datenwege sind nachvollziehbar.

Auch der Weg zurück ist jederzeit möglich: Wird OpenCloud abgeschaltet, bleiben einfach normale Dateien auf dem eigenen Dateisystem zurück. Eine Abhängigkeit, die ein Migrationsprojekt zum Auflösen bräuchte, entsteht gar nicht erst.

Möchten Sie Ihre HPC-Daten für die ganze Organisation zugänglich machen? Wir unterstützen Sie bei Architektur, Planung und Umsetzung.

Kontaktieren Sie uns gern unter sales@opencloud.eu.
Wir freuen uns auf Ihre Anfrage.



OpenCloud

OpenCloud ist die leistungsstarke Open-Source-Lösung für Filemanagement, Filesharing und Content Collaboration, die höchste Ansprüche an Nutzerfreundlichkeit, Performance, Stabilität und Skalierbarkeit erfüllt. Die Software kann flexibel On-Premise selbst betrieben oder sicher über Partner bezogen werden. Mit einer intuitiven Bedienung, nahtloser Integration, dem personenunabhängigen File-Management („Spaces“) und sicherer Versionierung ist OpenCloud ideal für den öffentlichen Sektor, Bildungseinrichtungen, Forschung sowie große Unternehmen und Provider geeignet.

Gemeinsam mit OpenTalk und mailbox bildet OpenCloud eine der drei starken Säulen der Heinlein Gruppe, die sichere und digital souveräne IT-Infrastrukturen für Europa bereitstellt.



https://opencloud.eu