ONEDATA

Global Distributed Data Access System

Michal Orzechowski

ACK CYFRONET AGH, Krakow, Poland

Onedata in a nutshell

Globally Distributed Data Access System

One solution for:

- **Unified** data access
- **Distributed** data management
- High **throughput** data **migration** between sites
- High speed **remote** data **access** on the fly
- Flexible authorization and access control based on: **tokens**, OpenID, X.509 Certs and others
- **Sharing** data between users
- **Open data** management (in preparation)

Continued Support of

INDIGO - DataCloud

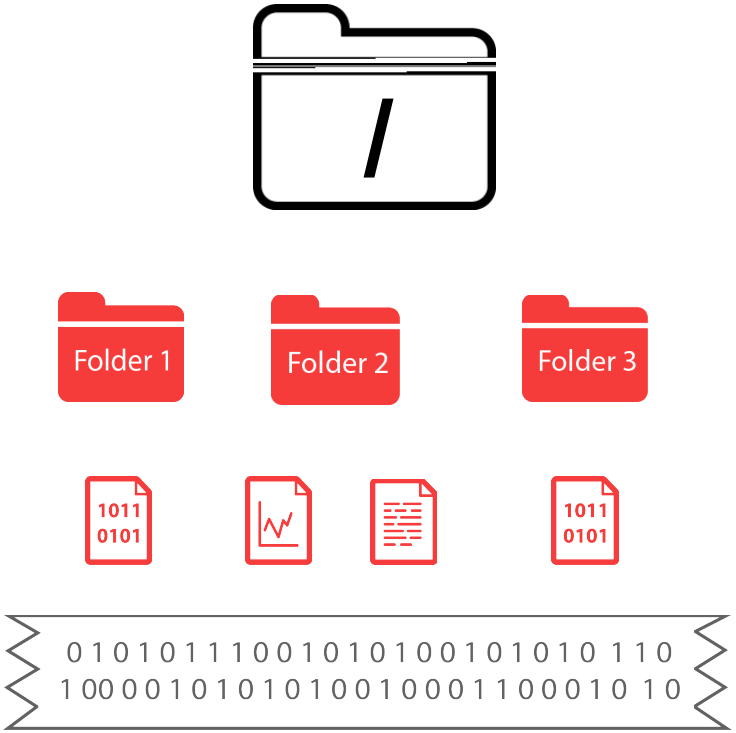
EGI-Engage

PL GRID CORE

INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY

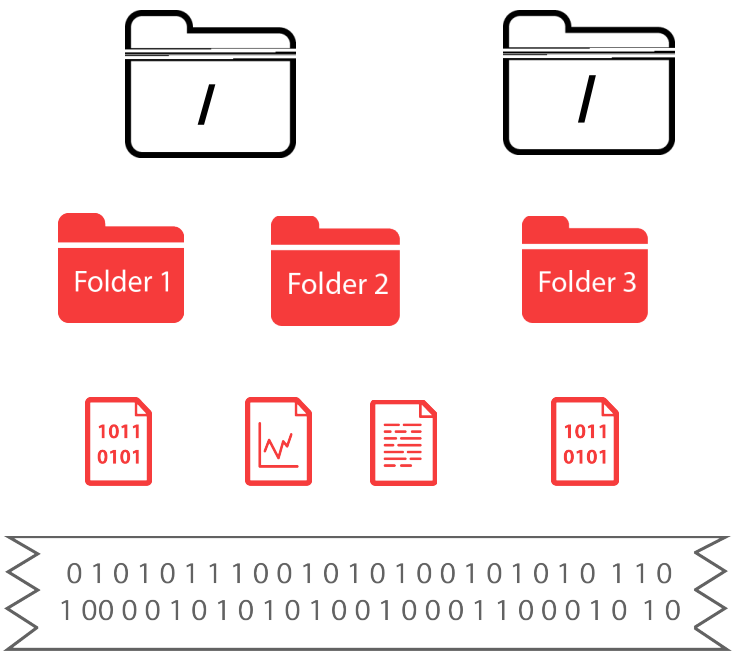
CYFRONET

Onedata Spaces



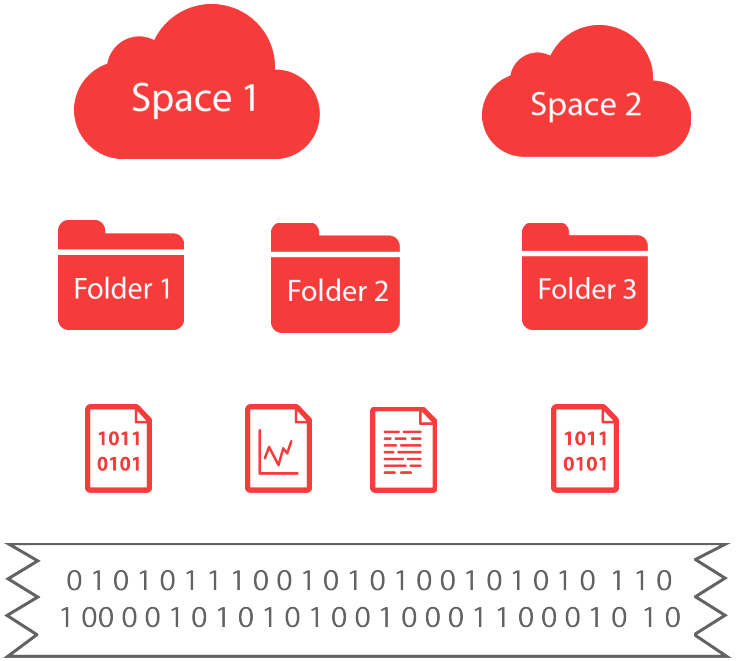
- kind of virtual directory
- **root** for filesystem tree
- unlimited number of spaces
- access via Onedata

Onedata Spaces



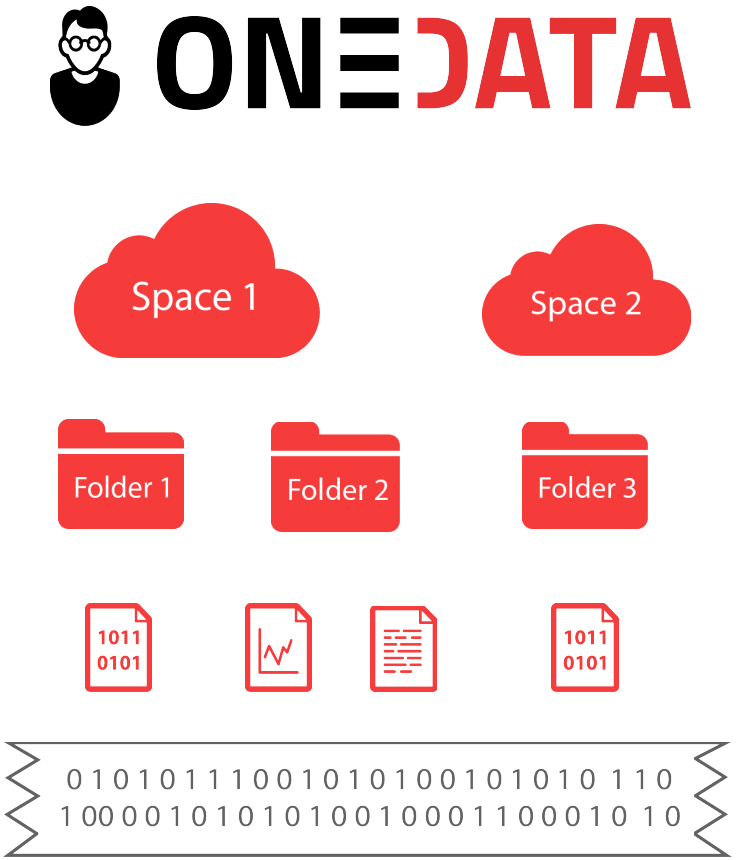
- kind of virtual directory
- **root** for filesystem tree
- unlimited number of spaces
- access via Onedata

Onedata Spaces



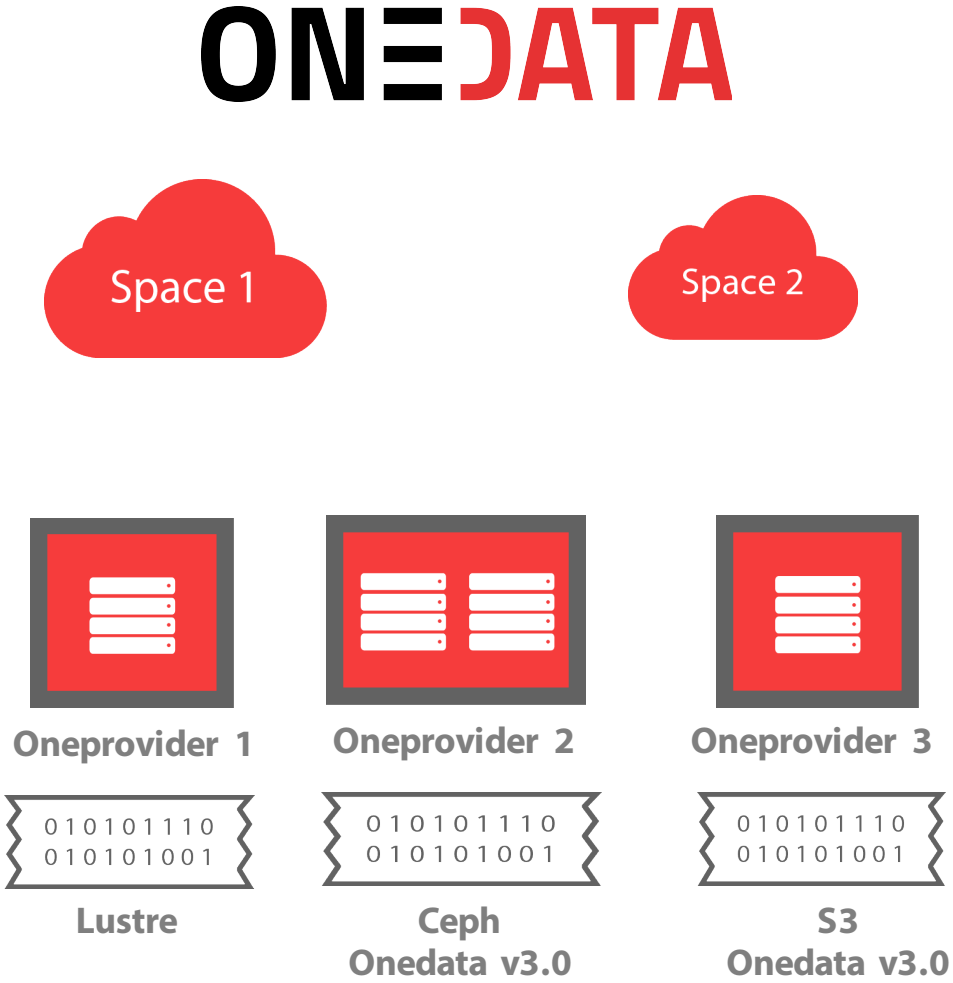
- kind of virtual directory
- **root** for filesystem tree
- unlimited number of spaces
- access via Onedata

Onedata Spaces

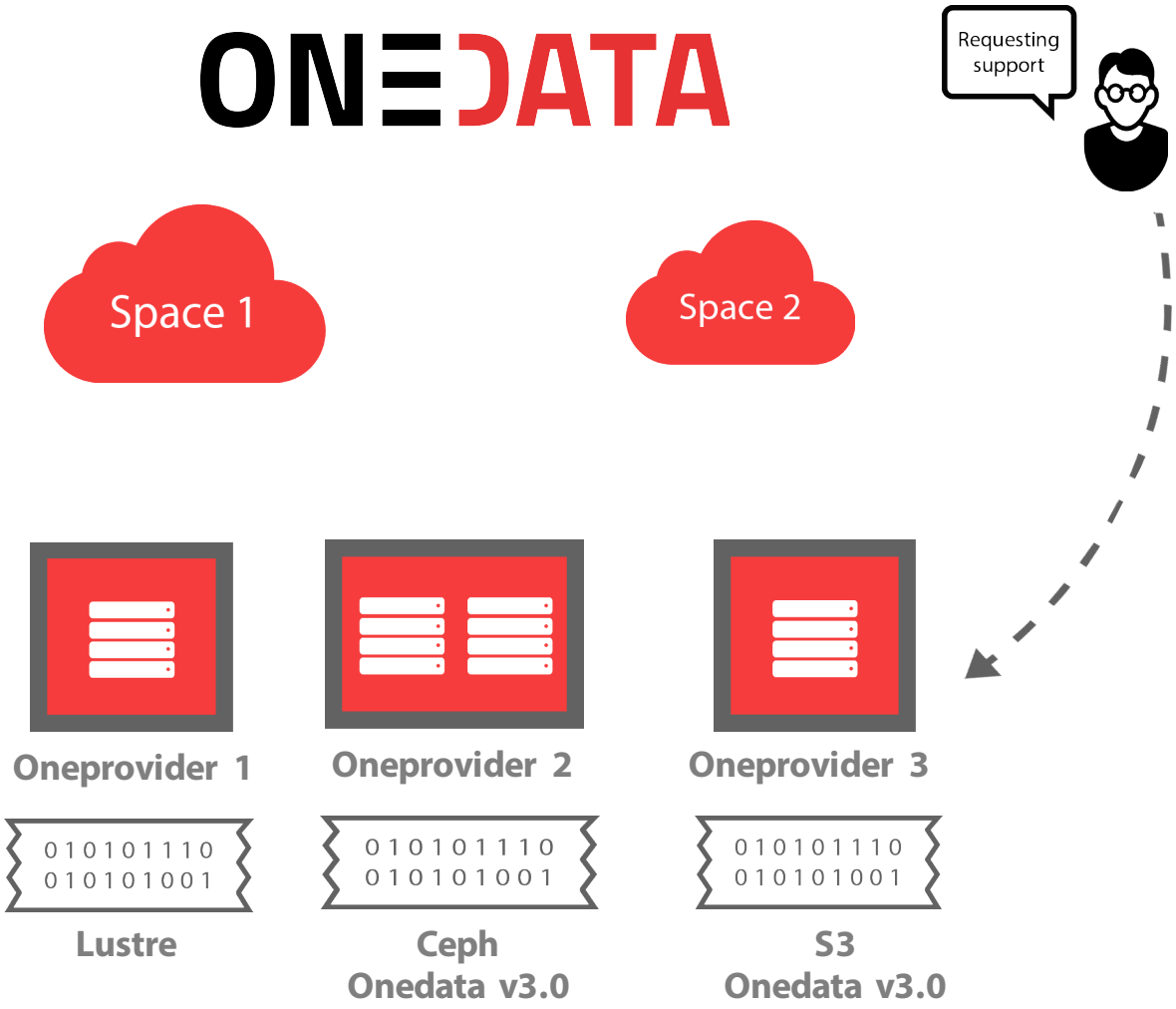


- kind of virtual directory
- **root** for filesystem tree
- unlimited number of spaces
- access via Onedata

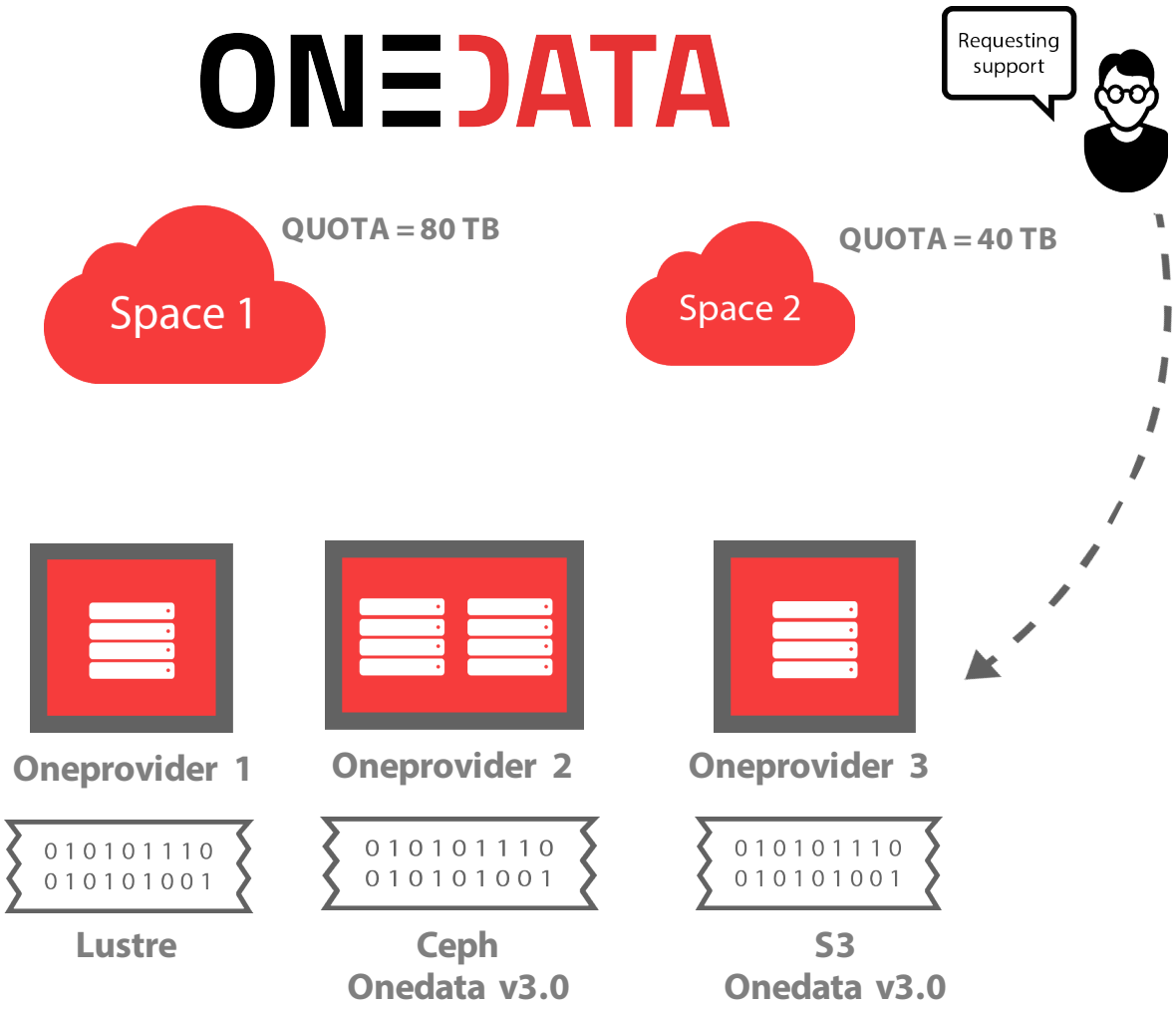
Distributed Data Access System



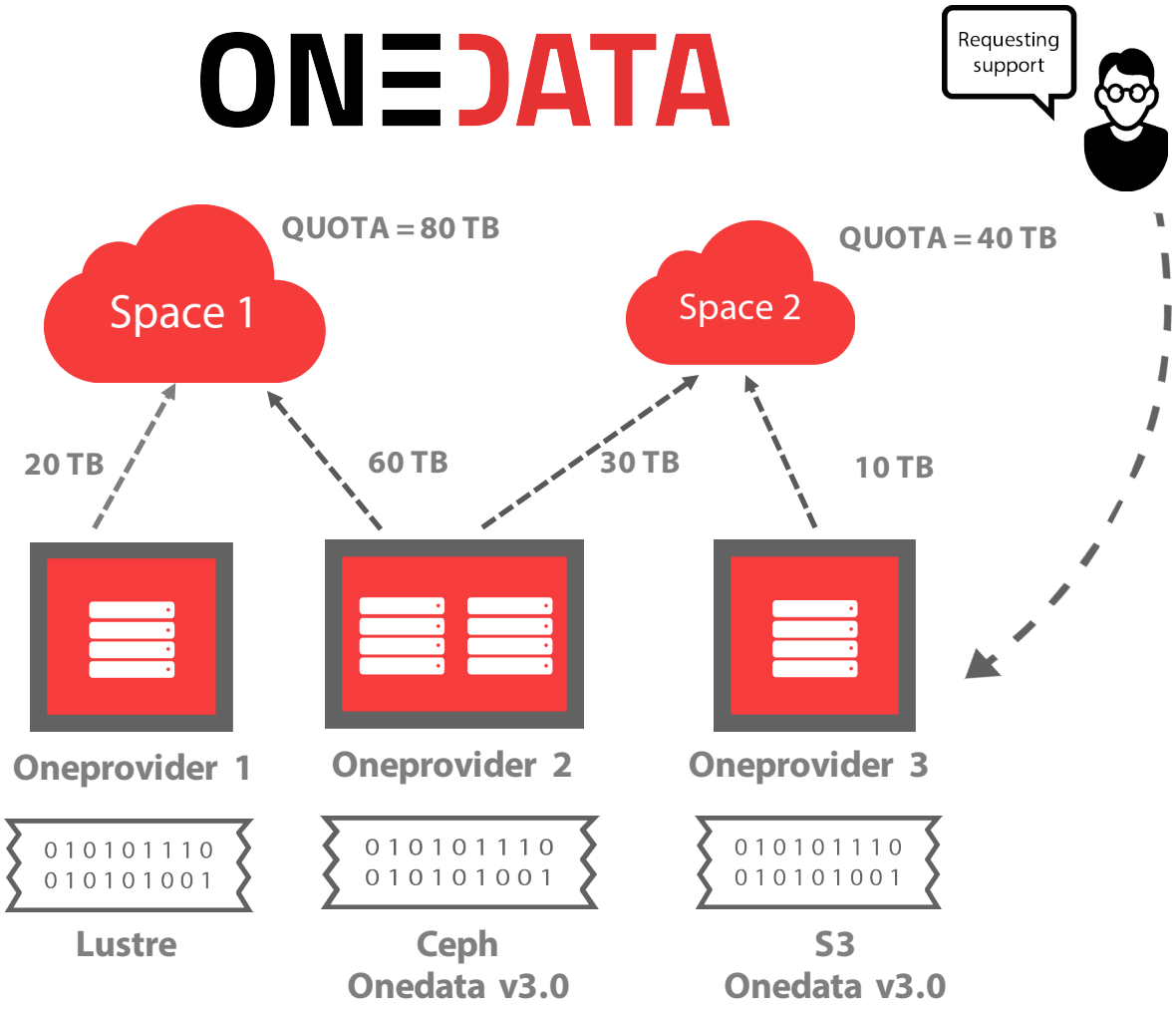
Distributed Data Access System



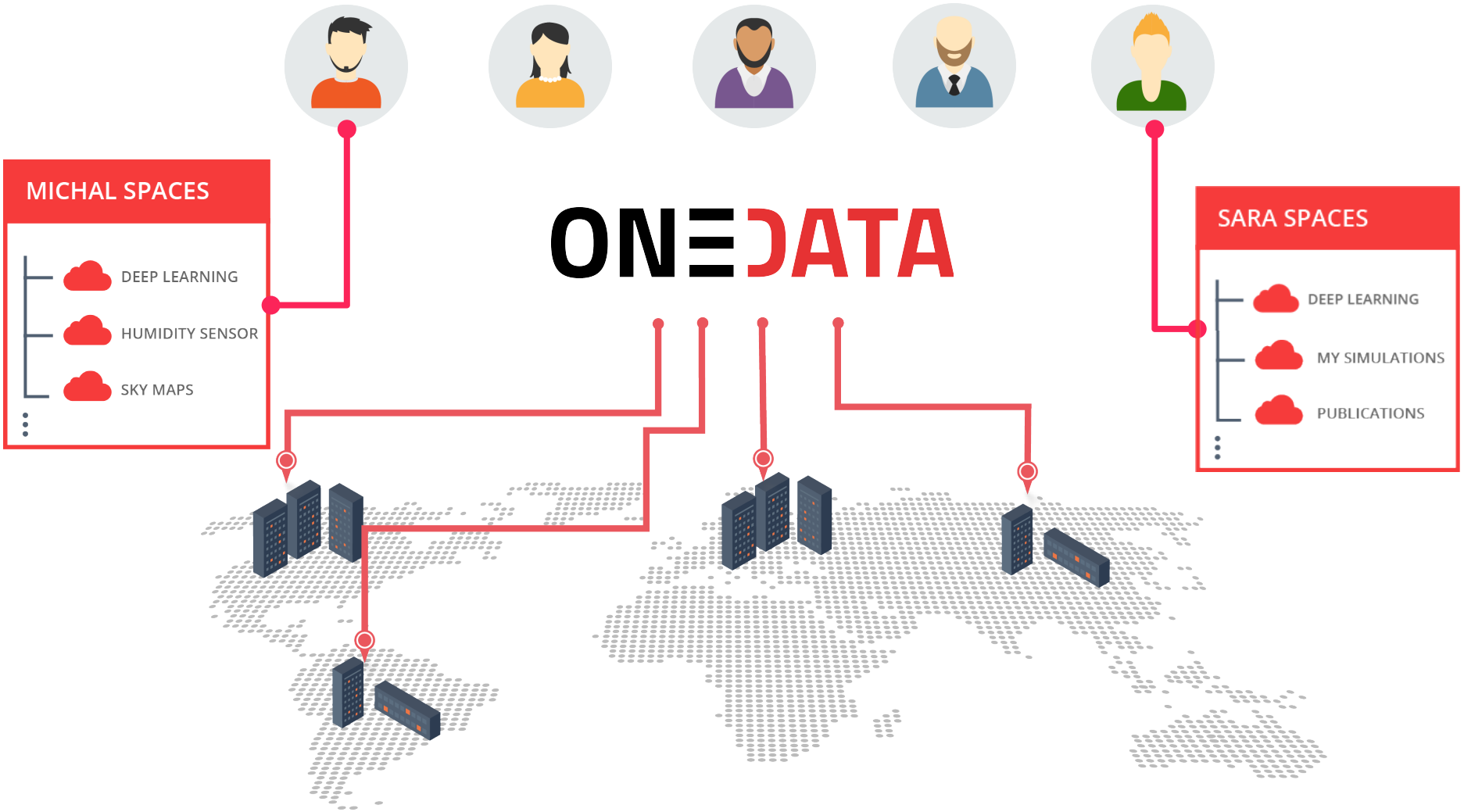
Distributed Data Access System



Distributed Data Access System



Globally Distributed Data Access System

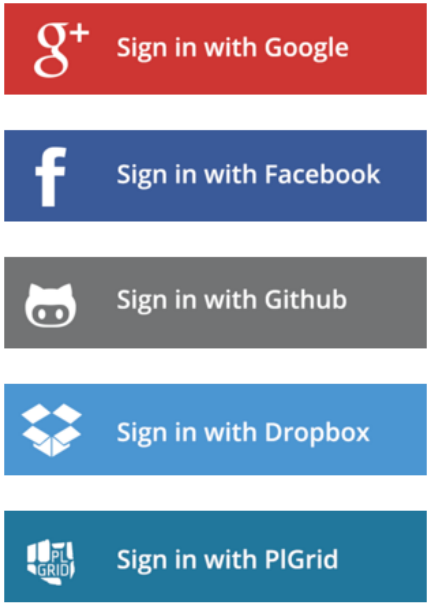


First Login

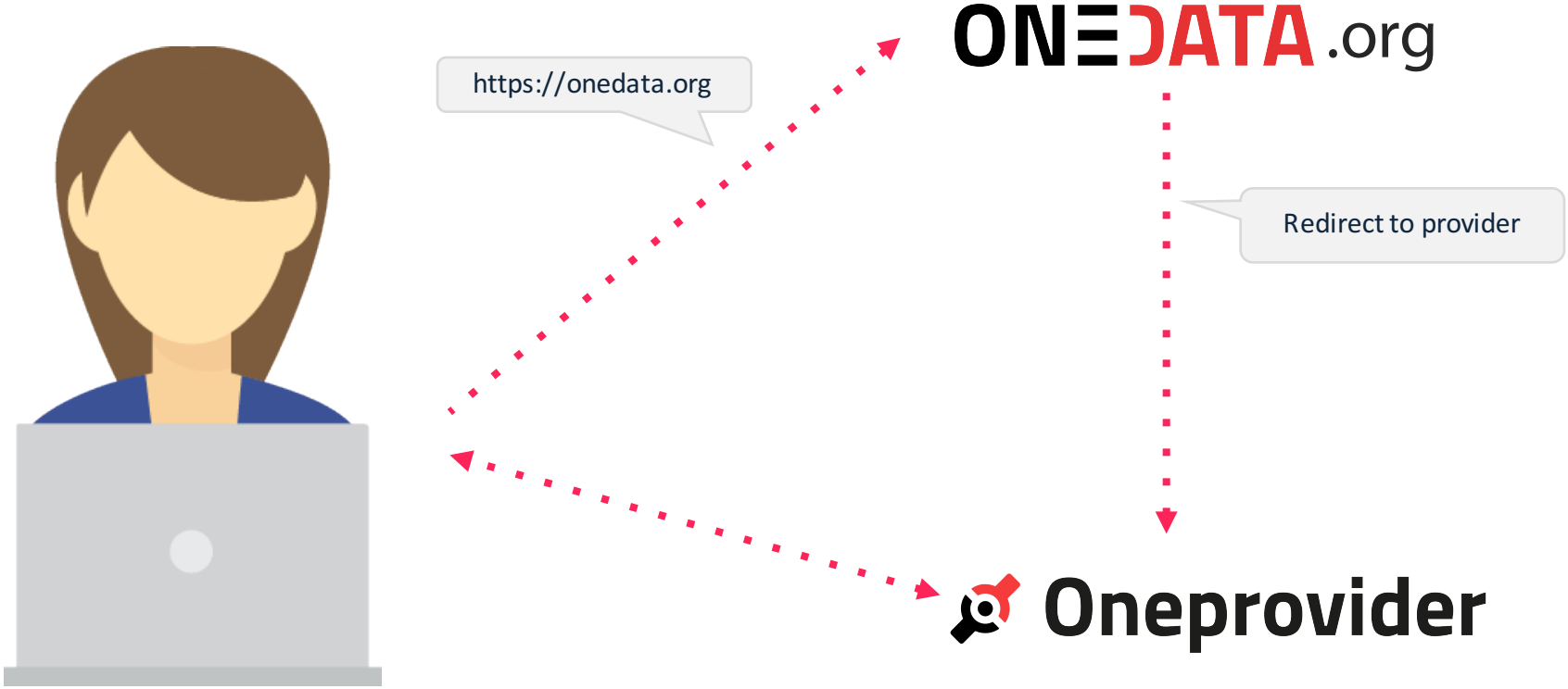
https://onedata.org

ONEDATA.org

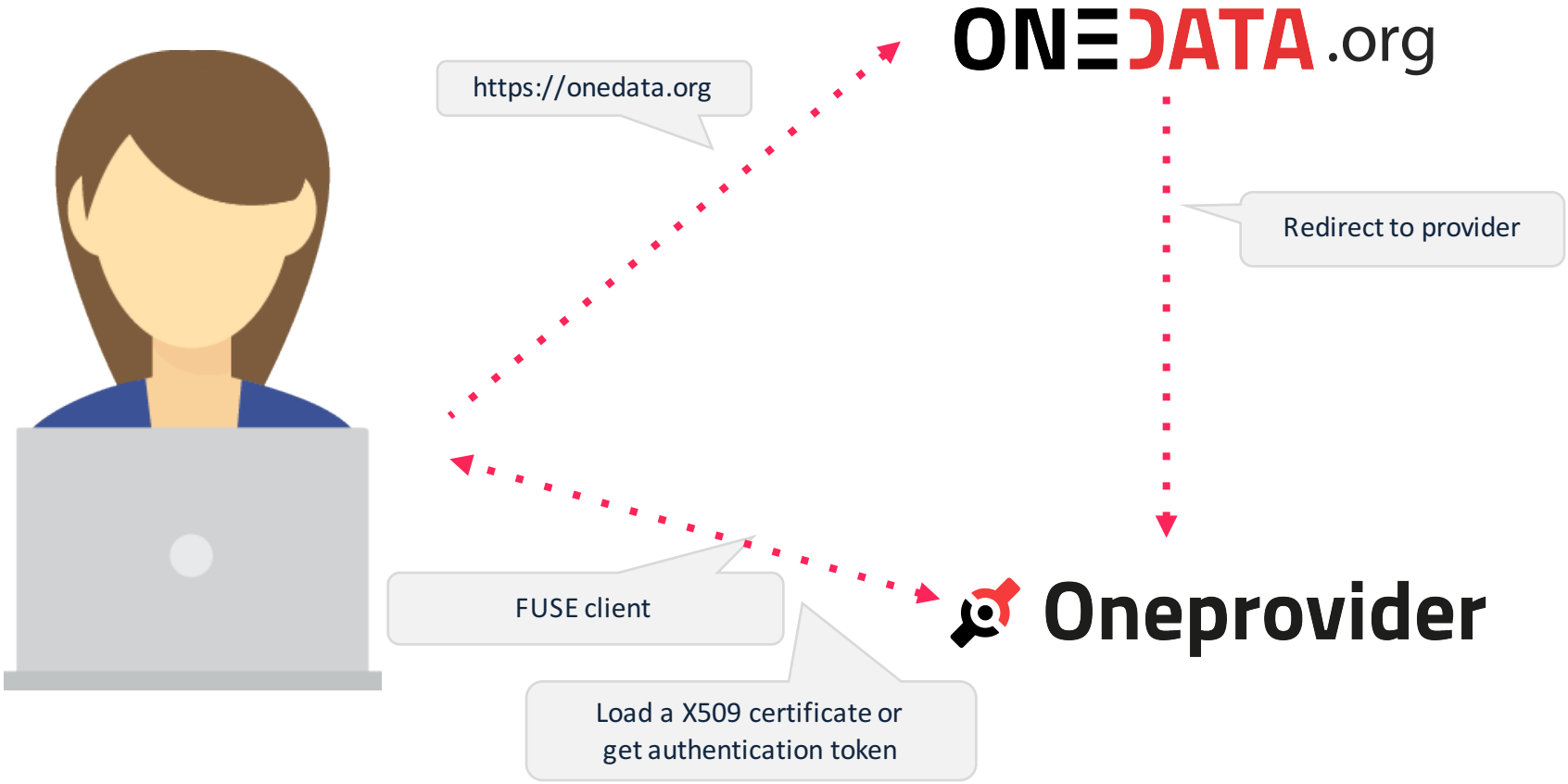
OAuth2
OpenID



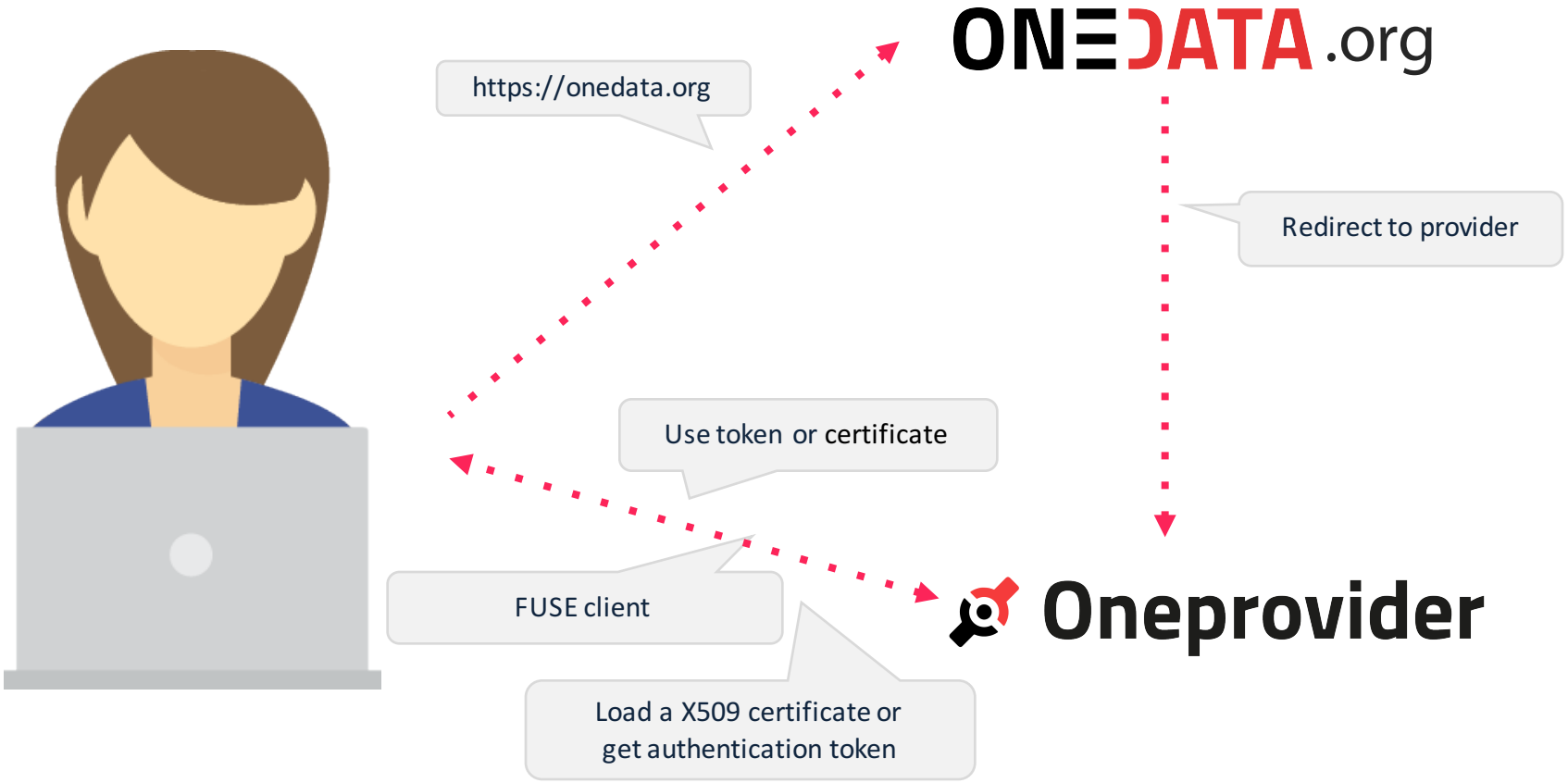
First Login


https://onedata.org
ONEDATA.org
Redirect to provider
Oneprovider

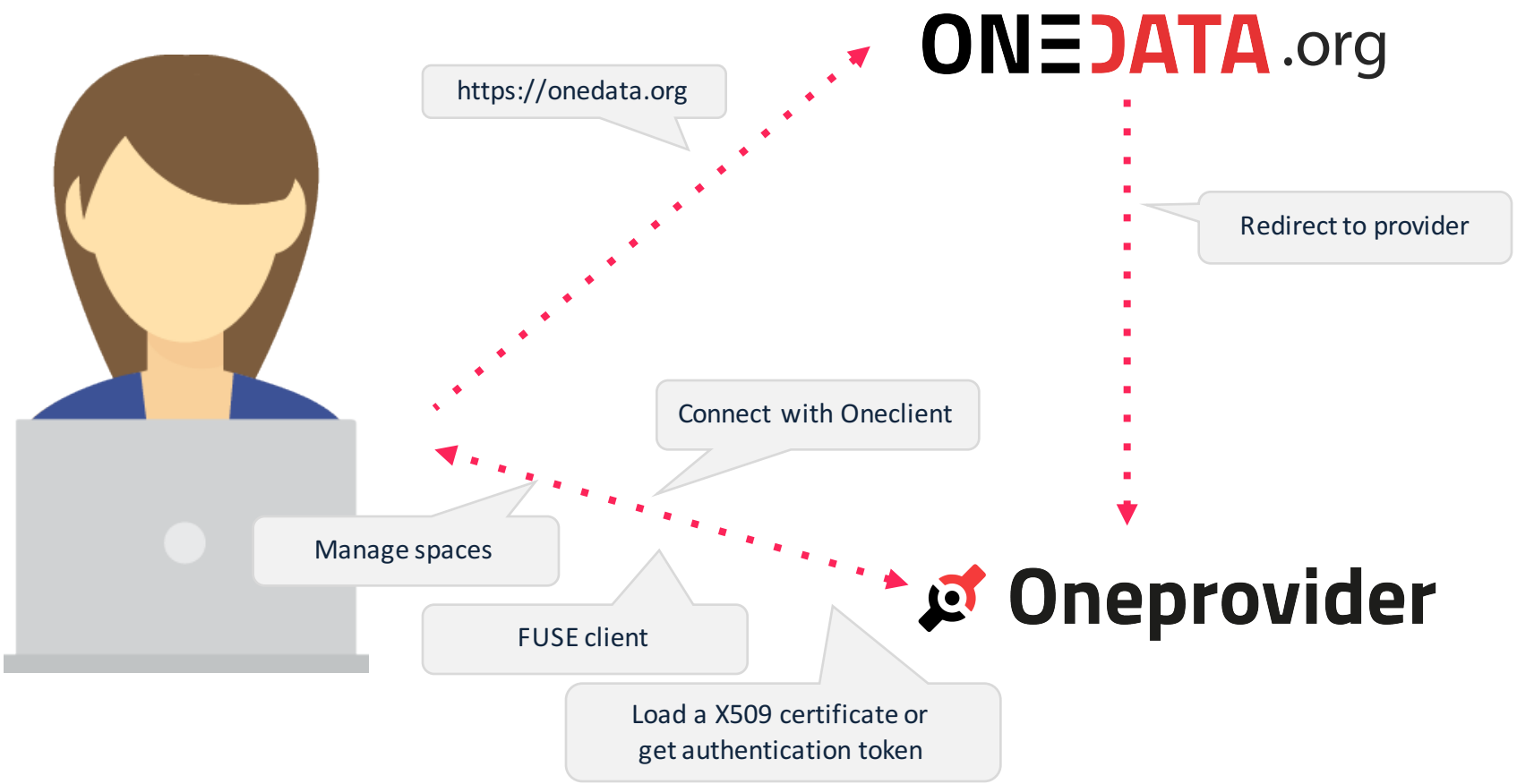
First Login


ONEDATA.org
https://onedata.org
Redirect to provider
FUSE client
Oneprovider
Load a X509 certificate or
get authentication token

First Login


ONEDATA.org
https://onedata.org
Redirect to provider
Use token or certificate
FUSE client
Oneprovider
Load a X509 certificate or
get authentication token

First Login


ONEDATA.org
https://onedata.org
Redirect to provider
Connect with Oneclient
Manage spaces
FUSE client
Load a X509 certificate or
get authentication token
Oneprovider

Spaces accessible via web UI

onedata Data Spaces Groups Tokens Michal Orzechowski

~ / spaces

Search

Sort

Name	Permissions	Size	Access	Modification
..			2016-01-08 07:41:59	2016-01-08 07:48:03
Deep Learning			2016-01-08 07:43:46	2016-01-08 07:48:16
Humidity Sensor			2016-01-15 14:32:14	2016-01-15 14:32:14
Michal's space			2016-01-08 07:41:59	2016-01-08 07:48:03

Access Control Lists

onedata Data Spaces Groups Tokens Michal Orzechowski

Search

Sort

Name	Permissions	Size	Access	Modification
spaces	r w x r - x r - x [755]		2016-01-08 00:38:44	2016-01-08 00:38:44
file1.txt	r w - r w - - - - [660]	612.00 KB	2016-01-08 07:48:04	2016-01-08 07:48:04

Permissions type: POSIX ACL

	read	write	execute
user	✓	✓	
group	✓	✓	
other			

octal form: 660

recursive Ok Cancel

Access Control Lists

onedata Data Spaces Groups Tokens Michal Orzechowski

Search

Sort

Name	Permissions	Size	Access	Modification
spaces	r w x r - x r - x [755]		2016-01-08 00:38:44	2016-01-08 00:38:44
file1.txt	r w - r w - - - - [660]	612.00 KB	2016-01-08 07:48:04	2016-01-08 07:48:04

Permissions type: POSIX ACL

proccessing order

Michal Orzechowski R W X

New ACL entry...

Identifier Michal Orzechowski

Type allow

Perms read write execute

Save Discard

recursive Ok Cancel

Spaces accessible via web UI

https://107.178.216.168/file_manager

onedata Data Spaces Groups Tokens Michal Orzechowski

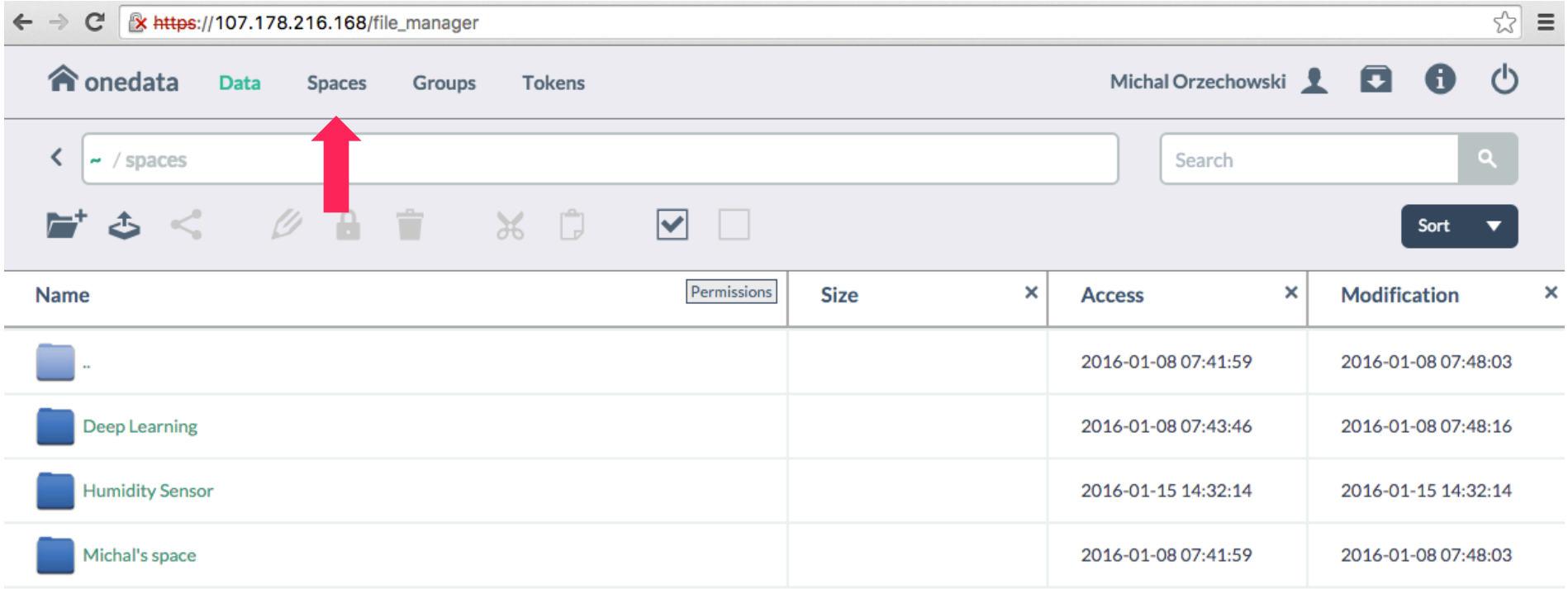
~ / spaces

Search

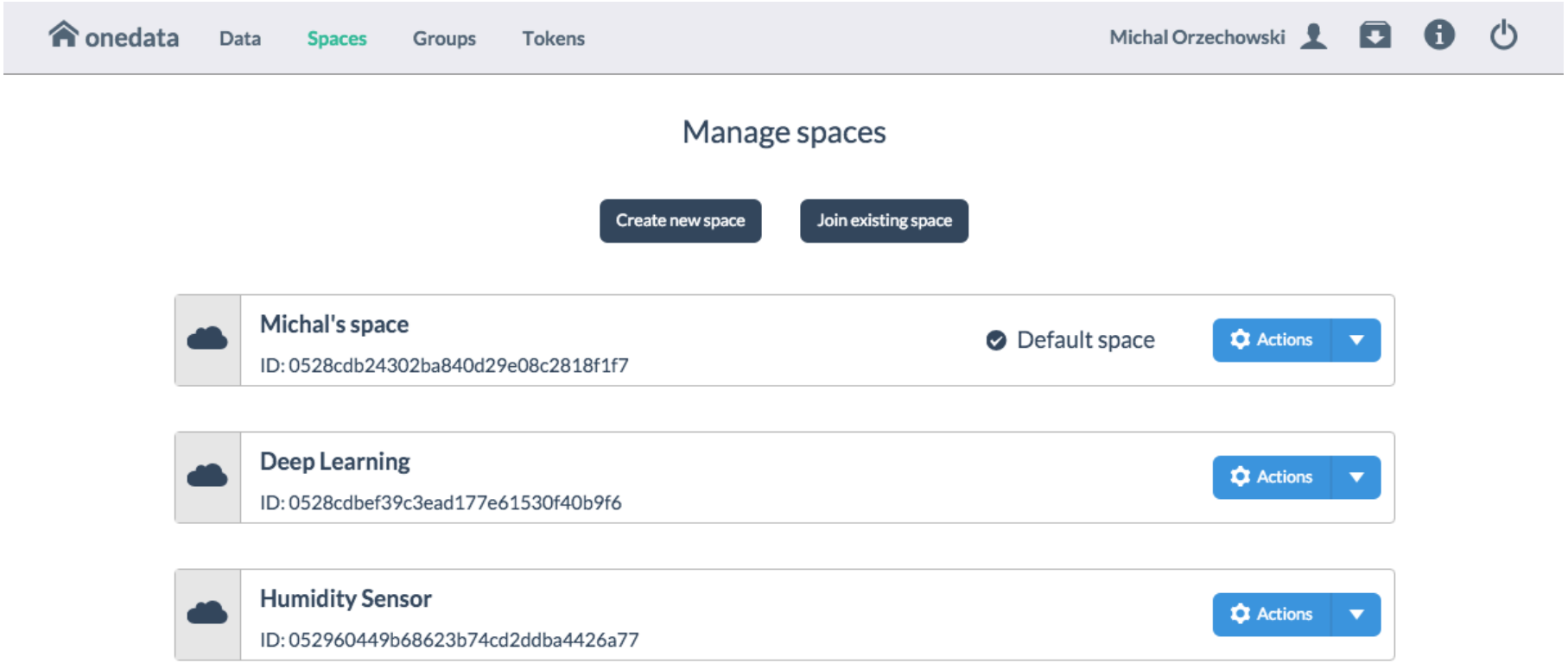
Sort

Name	Permissions	Size	Access	Modification
..			2016-01-08 07:41:59	2016-01-08 07:48:03
Deep Learning			2016-01-08 07:43:46	2016-01-08 07:48:16
Humidity Sensor			2016-01-15 14:32:14	2016-01-15 14:32:14
Michal's space			2016-01-08 07:41:59	2016-01-08 07:48:03

Spaces accessible via web UI



Space Management



Space Management

Deep Learning

ID: 0528cdbef39c3ead177e61530f40b9f6

Actions

USERS & RIGHTS

User	View Space	Modify Space	Remove Space	Invite user	Remove user	Invite group	Remove group	Invite provider	Remove provider	Set privileges
Michal Orzechowski	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sara	✓									

GROUPS & RIGHTS

Group	View Space	Modify Space	Remove Space	Invite user	Remove user	Invite group	Remove group	Invite provider	Remove provider	Set privileges
Machine Learning	✓									

PROVIDERS

Provider name	Provider ID	Provided space
provider5	0528c7cd3f993e0448d32ba3723d8b51	1024.00 MB

- **Sharing** spaces between users
- Rich **collaboration** capabilities
- Fine-grained **access control lists** (ACLs)

Space Management

onedata | Data | Spaces | Groups | Tokens | Michal Orzechowski

Manage groups

Machine Learning

ID: 0529605724abfd1857fe37a1d49f3515

Actions

USERS & RIGHTS

User	View group	Modify group	Remove group	Invite user	Remove user	Create Space	Join Space	Leave Space	Invite provider	Set privilege
Michal Orzechowski	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

SPACES

Space name	Space ID
Deep Learning	0528cdbef39c3ead177e61530f40b9f6

- **Group** Spaces for Communities

Other Ways to Access Spaces

- **POSIX mount with CLI client**

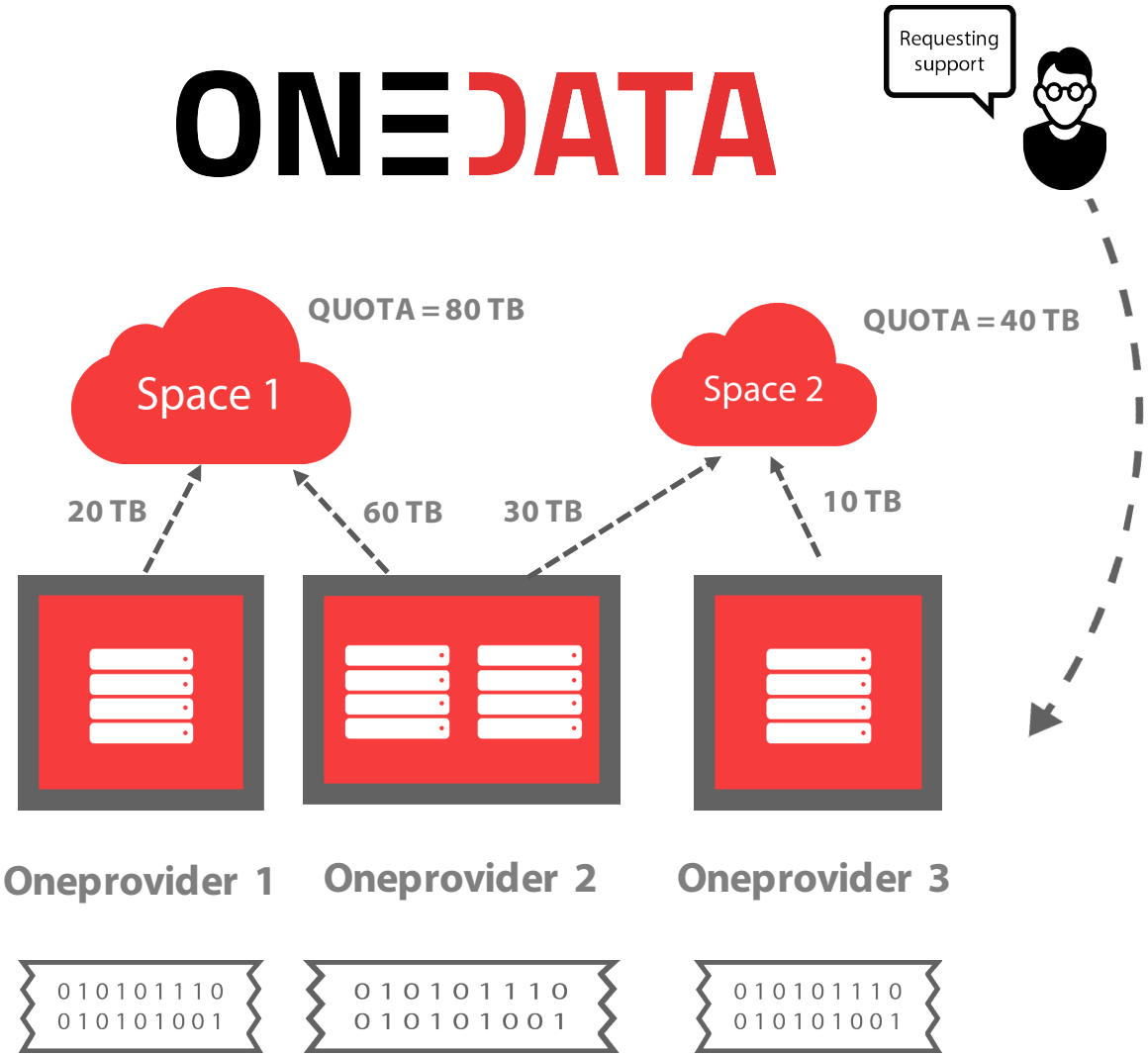
```
>oneclient ~/experiment/mnt/   #Virtual POSIX Filesystem
```

- **REST**

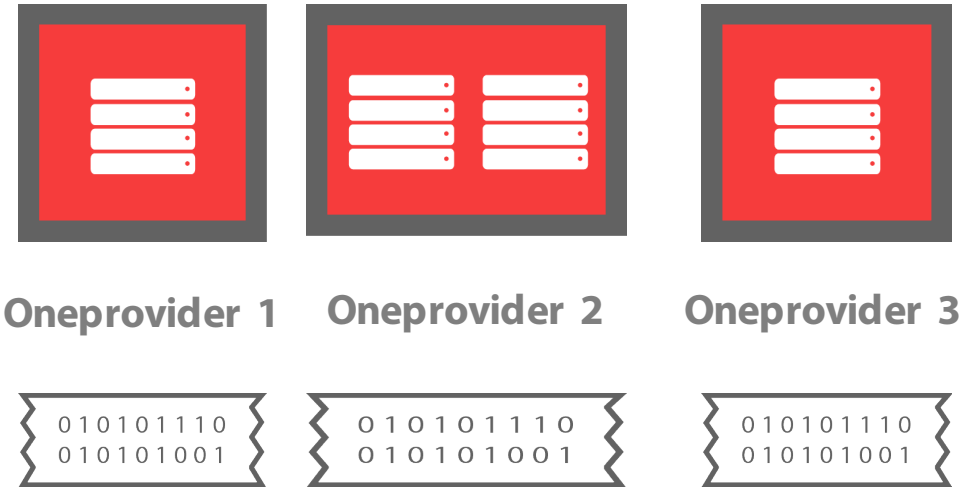
```
>curl -k -H "$TOKEN_HEADER" -X PUT -T ./binary_file
$ENDPOINT/binary_file
```

- **CDMI**

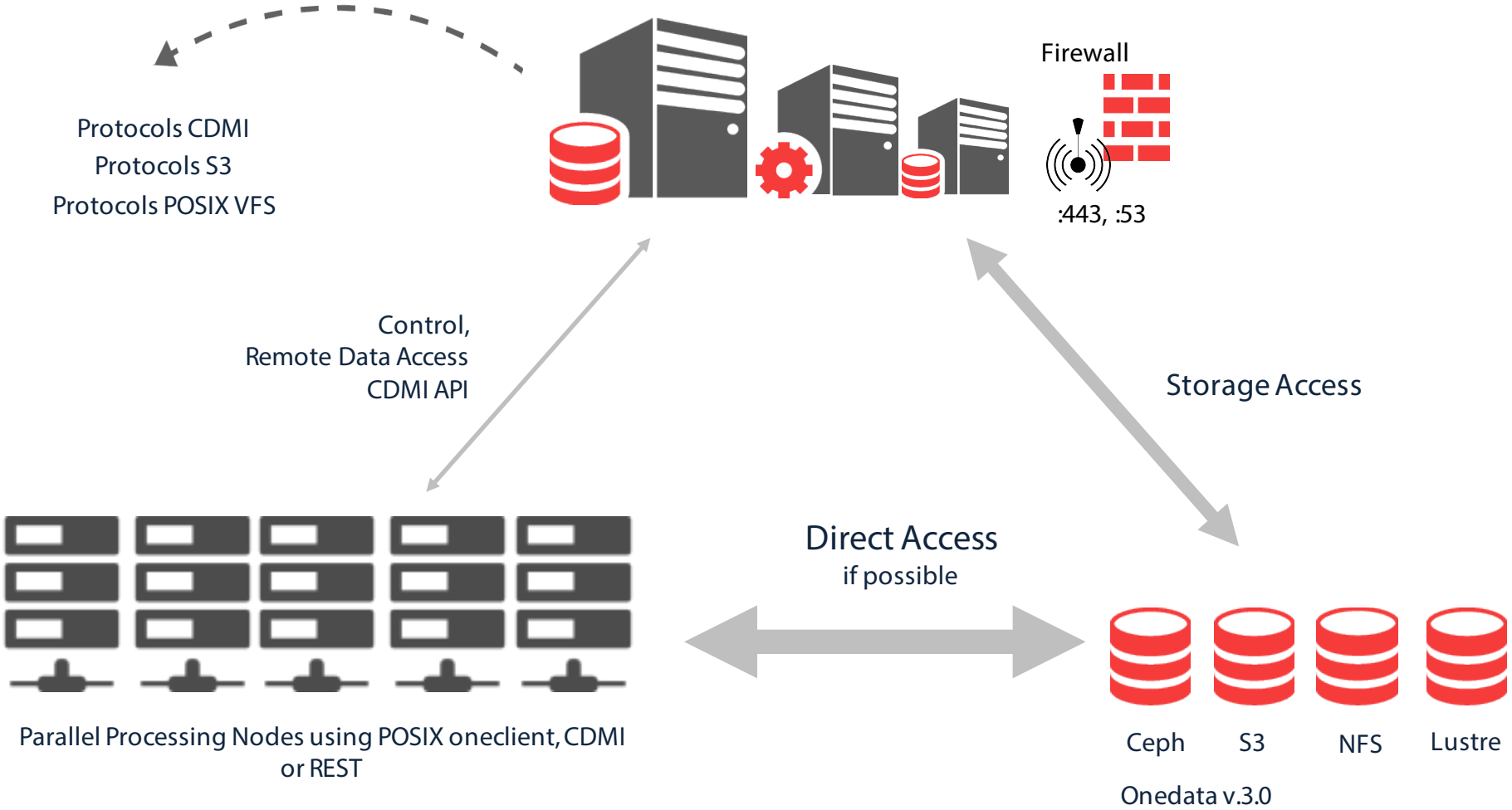
```
>curl -k -H "$TOKEN_HEADER" -H "$CDMI_VSN_HEADER" -
H 'Content-Type: application/cdmi-object'  -X PUT -
d '{"value":"file content"}' $ENDPOINT/file.txt
```


ONEDATA
Requesting support
QUOTA = 80 TB
Space 1
QUOTA = 40 TB
Space 2
20 TB
60 TB
30 TB
10 TB
Oneprovider 1
Oneprovider 2
Oneprovider 3
0 1 0 1 0 1 1 1 0
0 1 0 1 0 1 0 0 1
0 1 0 1 0 1 1 1 0
0 1 0 1 0 1 0 0 1
0 1 0 1 0 1 1 1 0
0 1 0 1 0 1 0 0 1

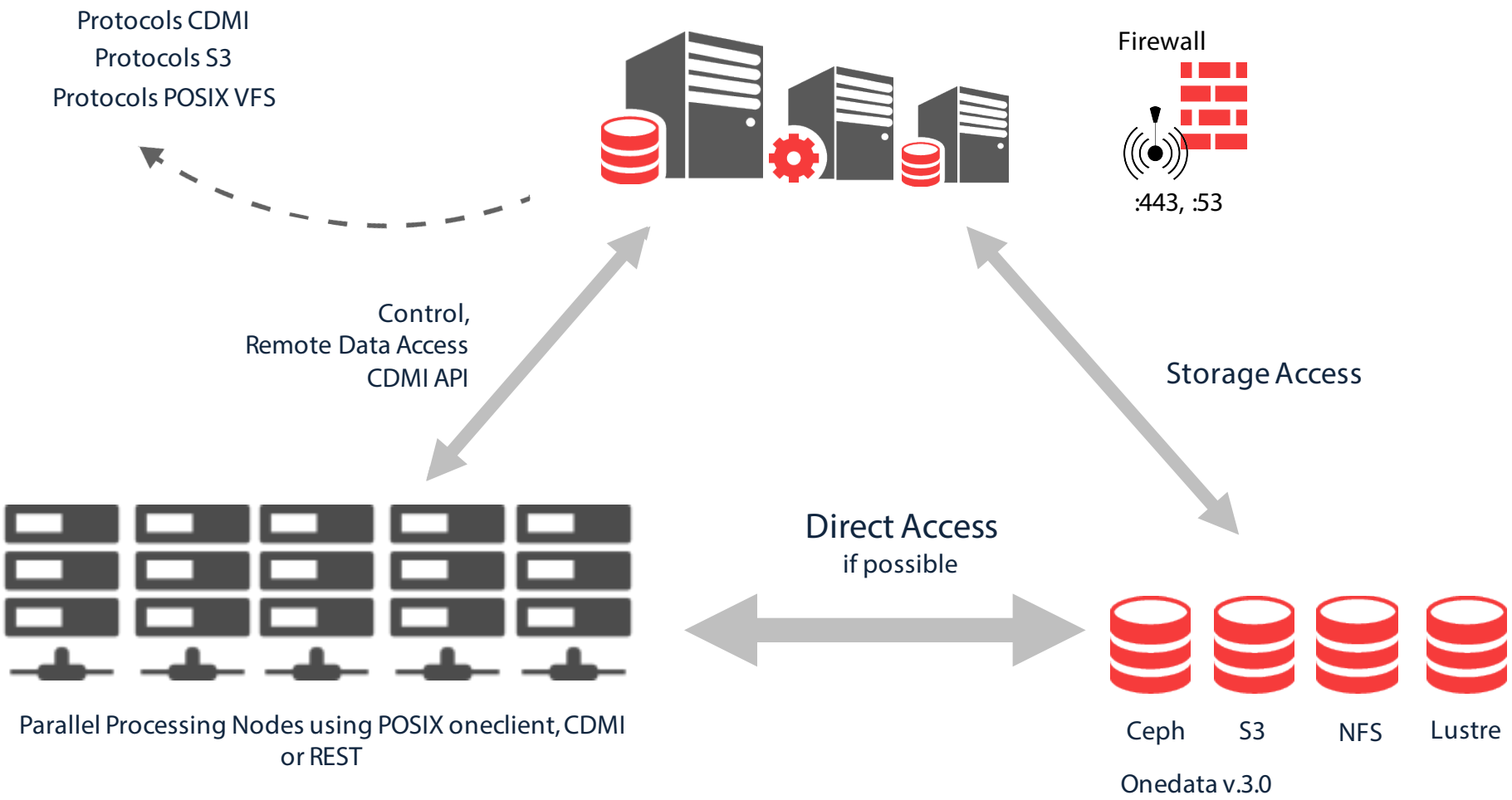
Oneprovider



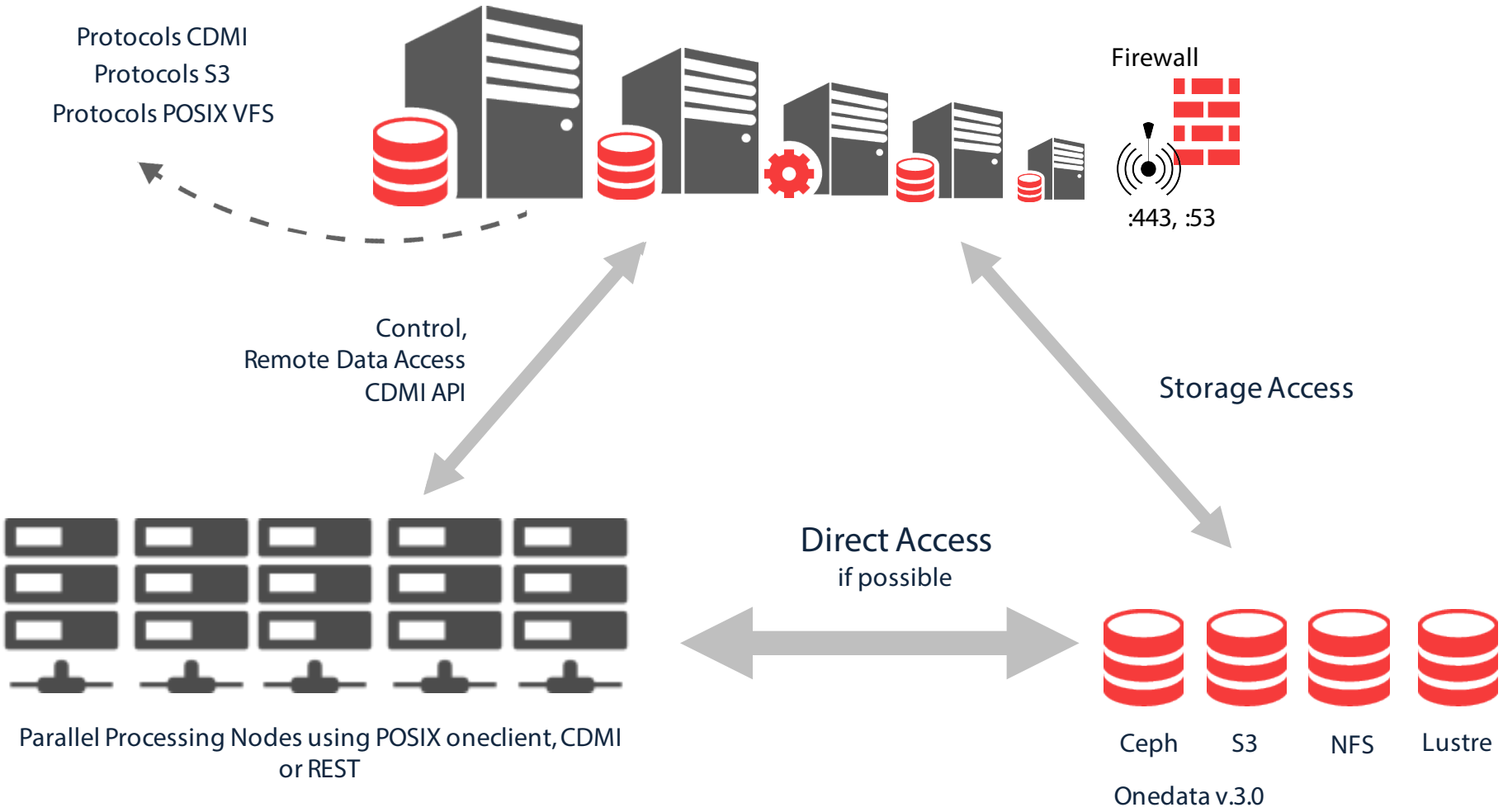
Oneprovider



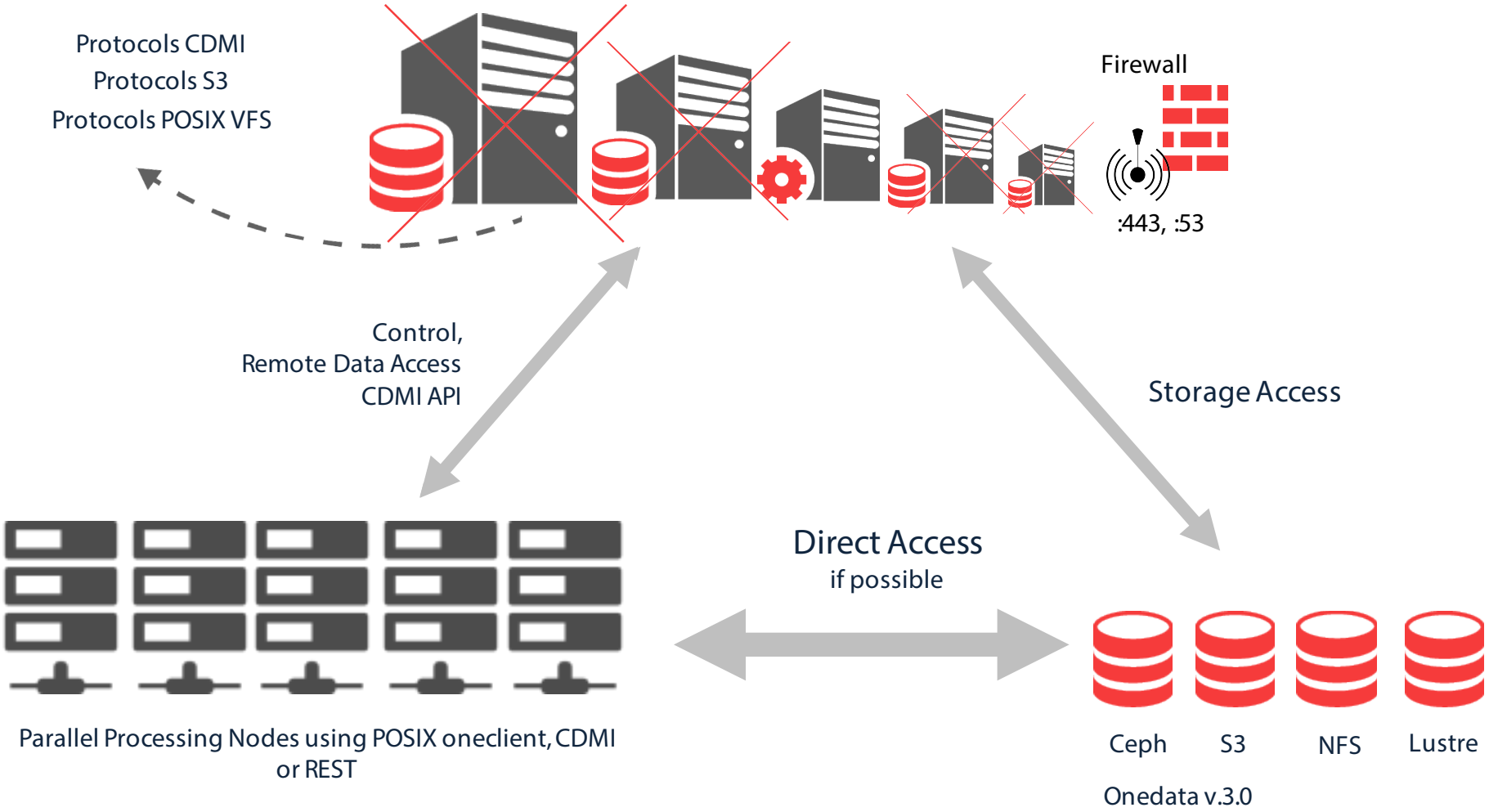
Oneprovider



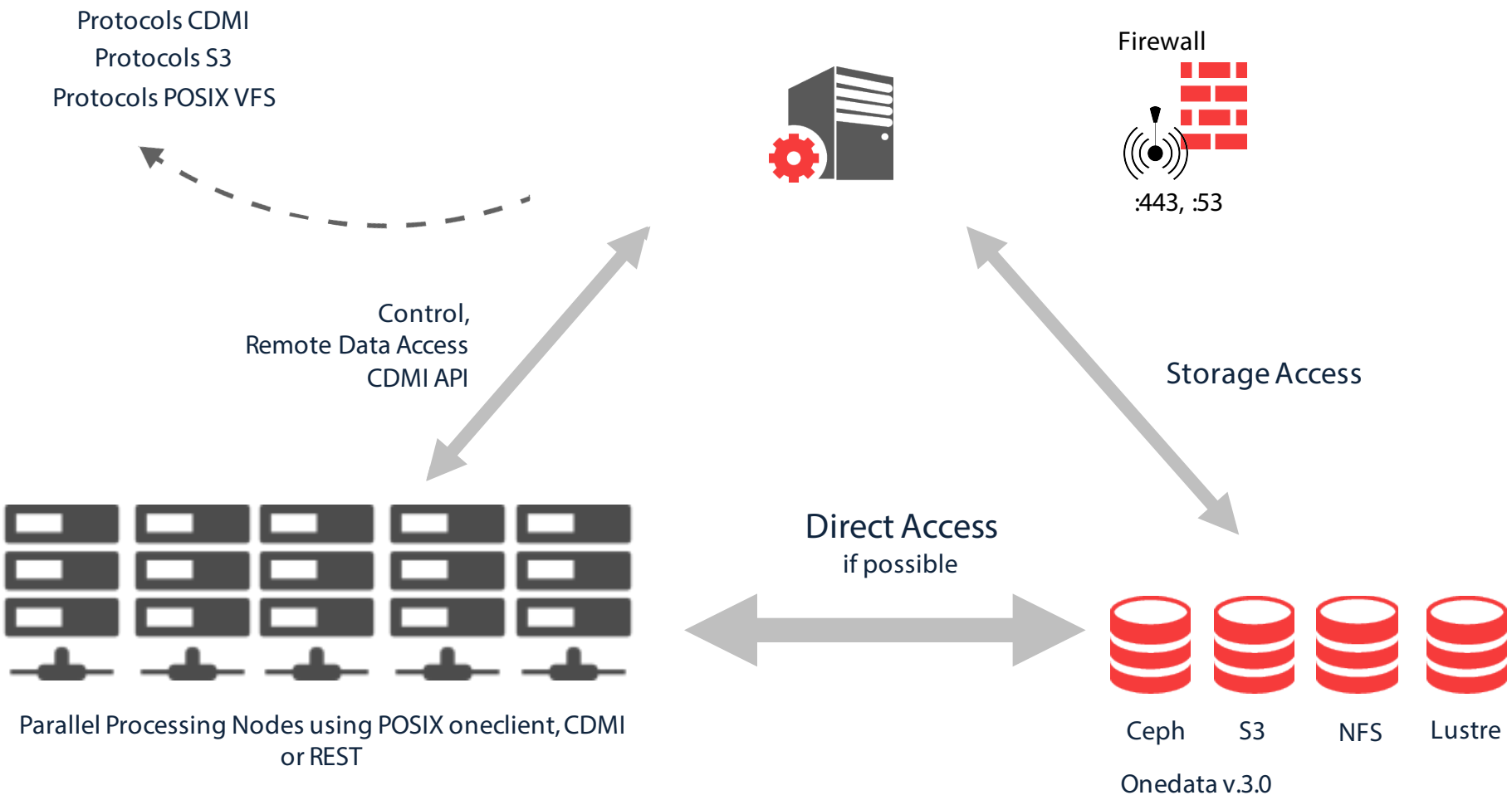
Oneprovider



Oneprovider

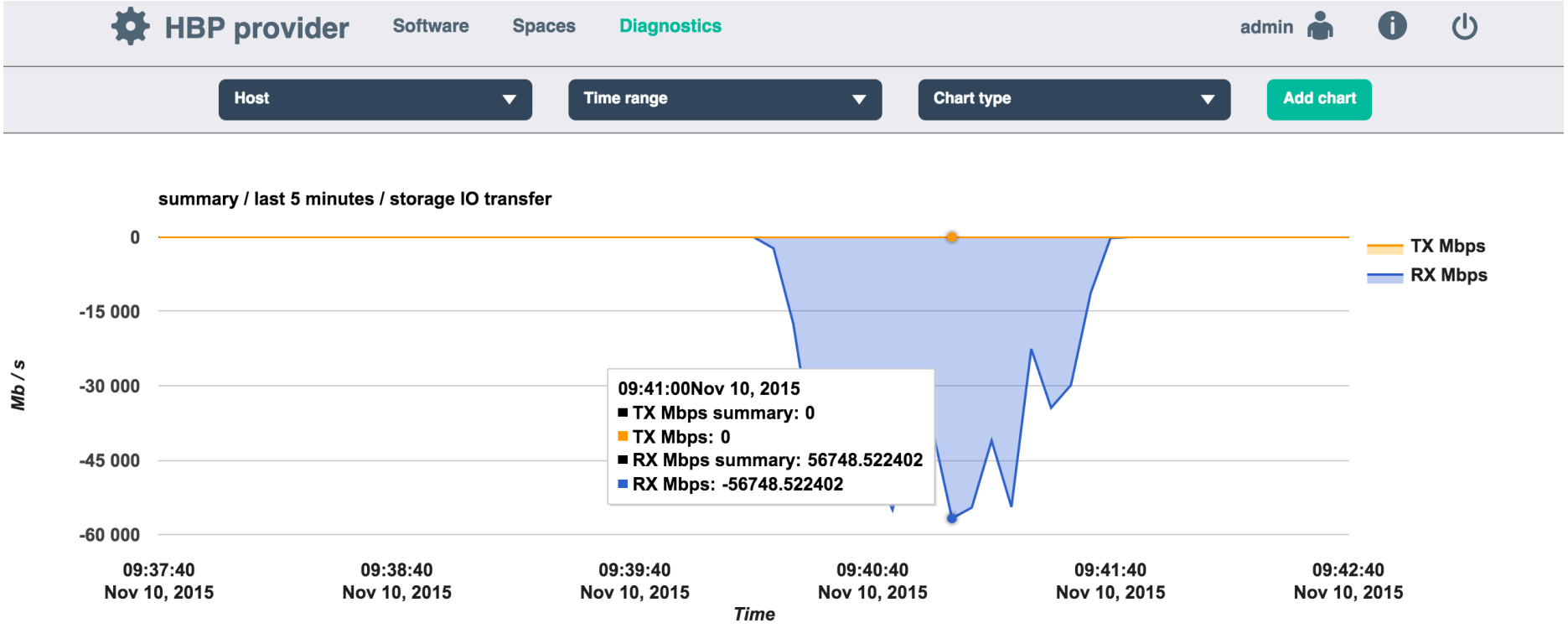


Oneprovider



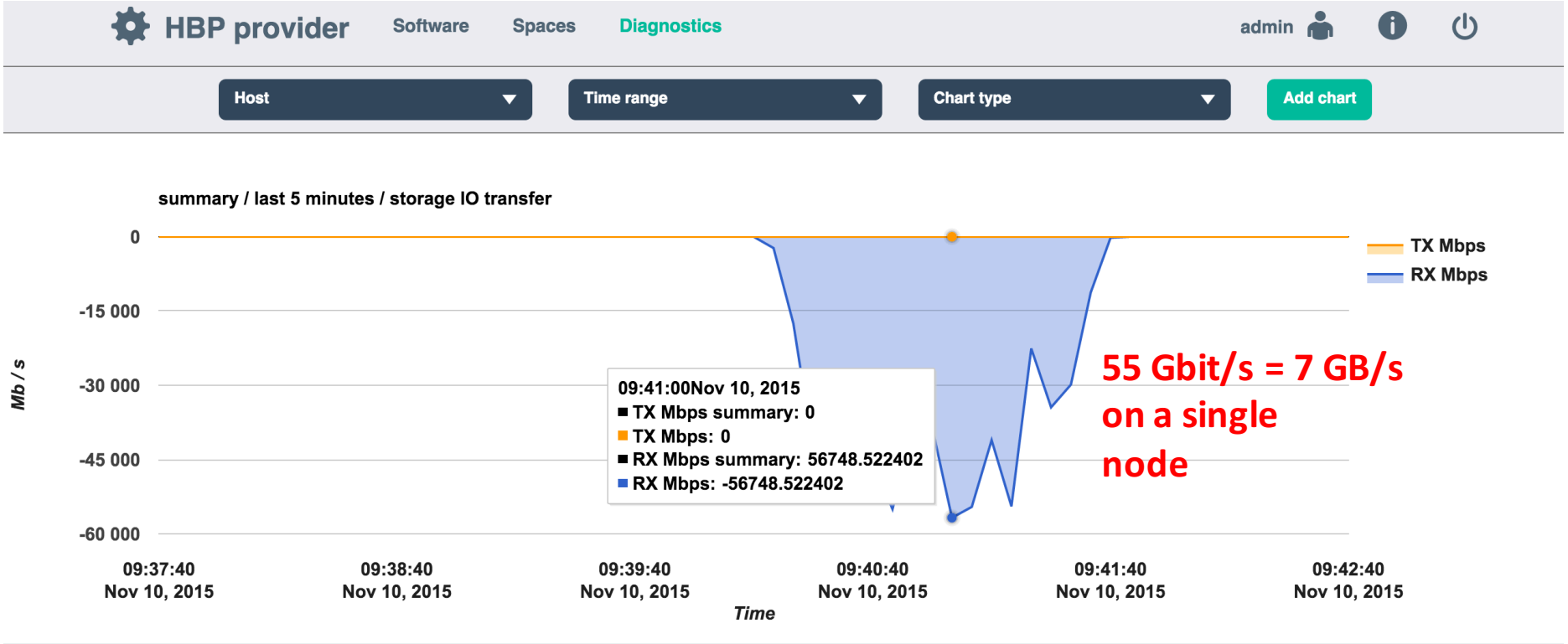
ONEDATA

High Performance Access



ONEDATA

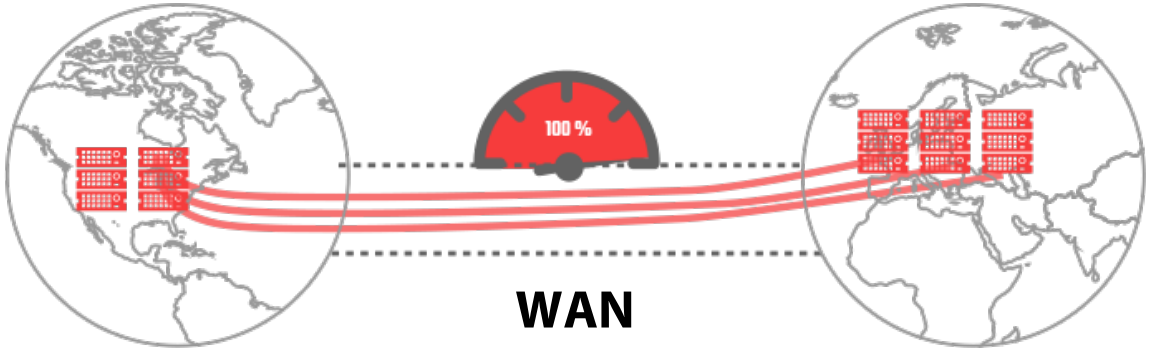
High Performance Access



ONEDATA

Remote Parallel Transfer

Distributed Priority Queue



Transfer started by:

- Access to Rmt. Data
- User in GUI
- API-s
- Policy

Block-based transfer:

- Remote Data Access on the fly
- Pre-staging
- Data Migration
- Data Replication

What's Next?

Onedata 3.0 release (Q1'16)

- New storage drivers: **Ceph** and **S3 Storage**
- Support for **S3 API**
- New **WEB** graphical user **interface**
- Significant **performance improvements** for desktop access (faster fuse)
- Dropbox-like desktop **synchronization client**

Open Data Support

- Publish Space as **open datasets**
- Integration with **DOI/PID** assignment services
- **Metadata** support for Spaces, files and folders

Summary

- Distributed multi-provider storage
- Flexible access control
- Intra-federations scenarios for sharing data
- Works with tokens or X.509
- POSIX client for mounting user's space
- Scalable from Single NAS to large datacentre
- Can be deployed on top of high-performance parallel storage solutions with very small overhead < 5%.
- Support for open data scenarios in preparation
- Onedata is currently supported by:
  PLGrid, EGI-Engage, INDIGO DataCloud

Questions?

ONEDATA.org

Michał Orzechowski
michal.orzechowski@cyfronet.pl

ACK CYFRONET AGH