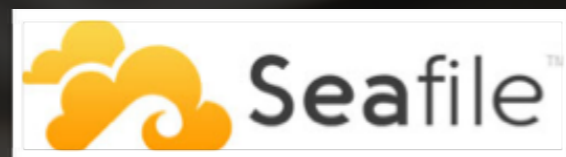


# NATIONAL DATA STORAGE SYNERGY WITH SYNC&SHARE SERVICES



Maciej Brzeźniak,  
Krzysztof Wadówka,  
Eugeniusz Pokora,

Norbert Meyer  
Paweł Wolniewicz  
Radosław Januszewski  
Mirosław Kupczyk  
Rafał Mikołajczak

PSNC, HPC Department



# AGENDA

- PSNC, National Data Storage & sister projects
- Why sync & share as-an-interface?
- Synergies on non-technical level



## CONTEXT - WHO WE ARE?



- Poznań Supercomputing and Networking Centre is one of the largest **HPC centres** in Poland
- Academic IT services provider: cloud computing, storage, backup/archival, processing; NREN

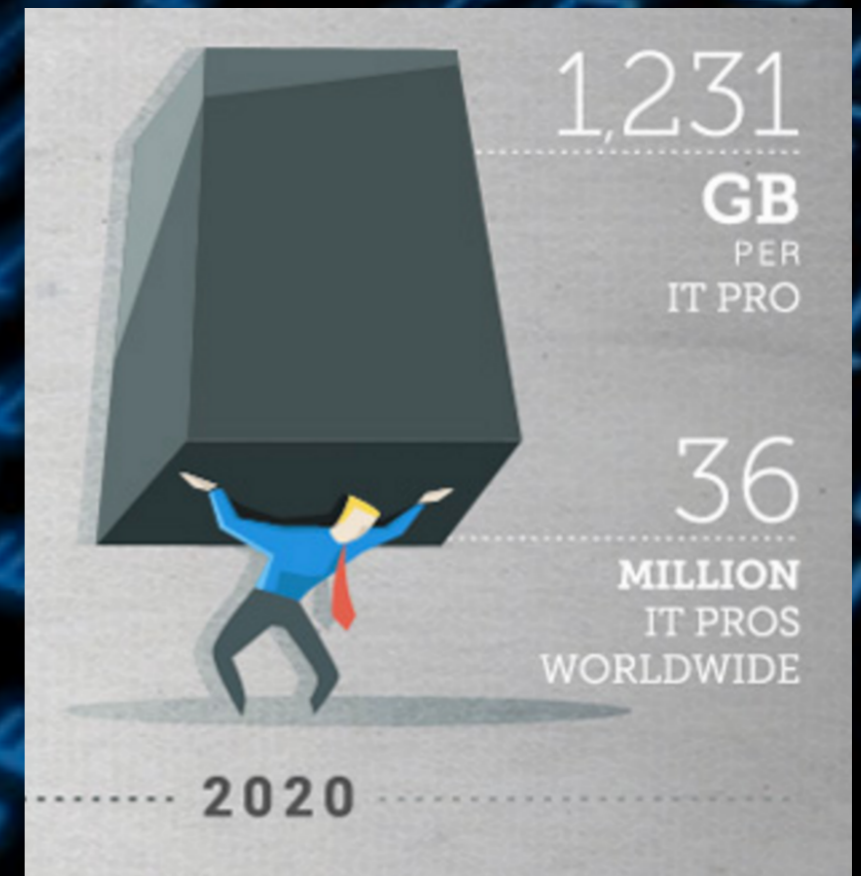


- **sync & share service provided country-wide by PSNC**
- aimed at large user base (10s thousands); currently 3k
- in production since 2015, 3k users, 180 TB data now
- based on Seafile Pro, MariaDB, GPFS cluster and HAproxy
- OnlyOffice integrated since 2020
- why Seafile? see our previous talks @CS3'17 and @TNC'18



# A SLIDE FROM 2018 TALK @CS3

- **DATA VOLUMES:**
  - PetaBytes...
  - 10-100s **GB/s**, 100s k **IOPS**
- **HPC storage:**
  - fast but hermetic
  - not easy to access
- **Dilemma: performance vs usability:**
  - SFTP, GridFTP, NFS, SSH - it's 80's...
  - HPC storage especially hermetic
- **Users:**
  - want **performance of Lustre**...
  - with the Dropbox' **ease of use**



Source: IDC





# NDS AND OTHER PSNC PROJECTS

- **National Data Storage** project (2021-2024)
  - „Data Lake” project
  - services:
    - long term storage based on object storage and tape storage (plus geo-replication)
    - fast data storage & access, data access acceleration based on SSD/NVMe pools
    - data exploration, analysis, processing through integrated services / interfaces
    - data preservation and archival using embedded digital repository and data protection software
  - sync & share - a convenient interface to data (lake)
- **Sister projects:**
  - **PRACE-LAB** (2019-2021) - **HPC** (CPU) & **cloud** computing,  
storage: auxiliary to compute services/infrastructure; developments: services / resources provisioning
  - **PRACE-LAB2** (2020-2023) - focused on **HPC** (CPU+**GPU**), BigData and AI  
storage: dedicated BigData cluster for demanding use-cases, extended Lustre systems for selected sites



# HPC / CLOUD / STORAGE INFRASTRUCTURE & PROJECTS



## PROJECTS: Future ICT & PRACE-LAB

SYSTEM	EAGLE	ALTAIR	E+A
YEAR	2015	2020	2021
PFLOPS	1,3	5,9	7,2
CORES	33K	63K	96K
RAM	303TB	250TB	550TB
NETWORK	IB FDR	IB EDR	FDR/ EDR





# HPC / CLOUD / **STORAGE**

## INFRASTRUCTURE & PROJECTS



### PROJECTS: Future ICT & PRACE-LAB, National Data Storage

SYSTEM / PROJECT	FutureICT	PRACE-LAB	NDS
<i>YEAR</i>	2015	2020	2022
<b>SSD ARRAYS</b>	0,08	0,15	0,15
<b>HPC STORAGE</b>	6,7	<b>15</b>	<b>15</b>
<b>FILE SERVERS</b>	0,5	0,5	0,5
<b>BLOCK ARRAYS</b>	7	7	7
<b>DISK SERVERS</b>	25,15	<b>45</b>	<b>150</b>
<b>SSD/NVMe SERVERS</b>	0,2	<b>0,4</b>	<b>3,6</b>
<b>TAPES</b>	28,50	28,50	<b>130 *)</b>

\*) We will come back this at the end

Systems' capacity in PetaBytes



# SYNERGIES



- Great... but what this means for sync & share?
  - Lots of **disk** (HDD) and flash (SSD/NVMe) **storage space**  
(goodbye limits, goodbye delays)
  - **New sync & deployment architecture**  
(goodbye GPFS, welcome to Ceph)
  - New **users** and **applications**  
(goodbye boredom)
  - New **integrations** and **interfaces** needed  
(goodday ease of use/mgmt & integration)

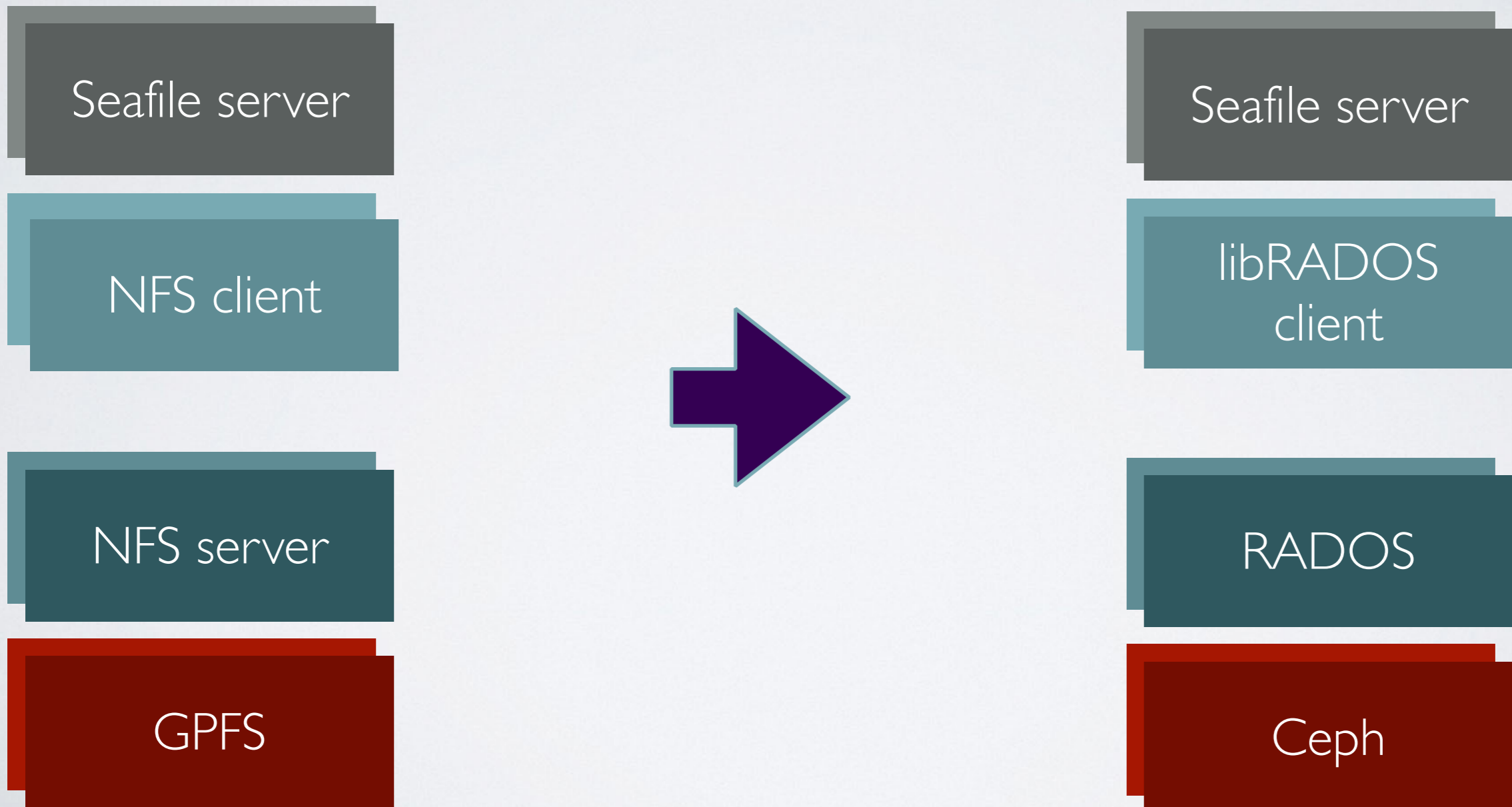




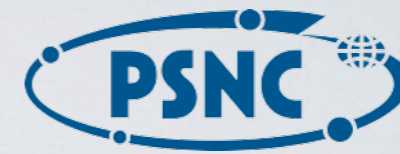
# NEW DEPLOYMENT ARCHITECTURE

GPFS is great but it costs licenses & dedicated maintenance effort

**Better use Ceph** for better scalability



# DID WE SAY LIMITS?

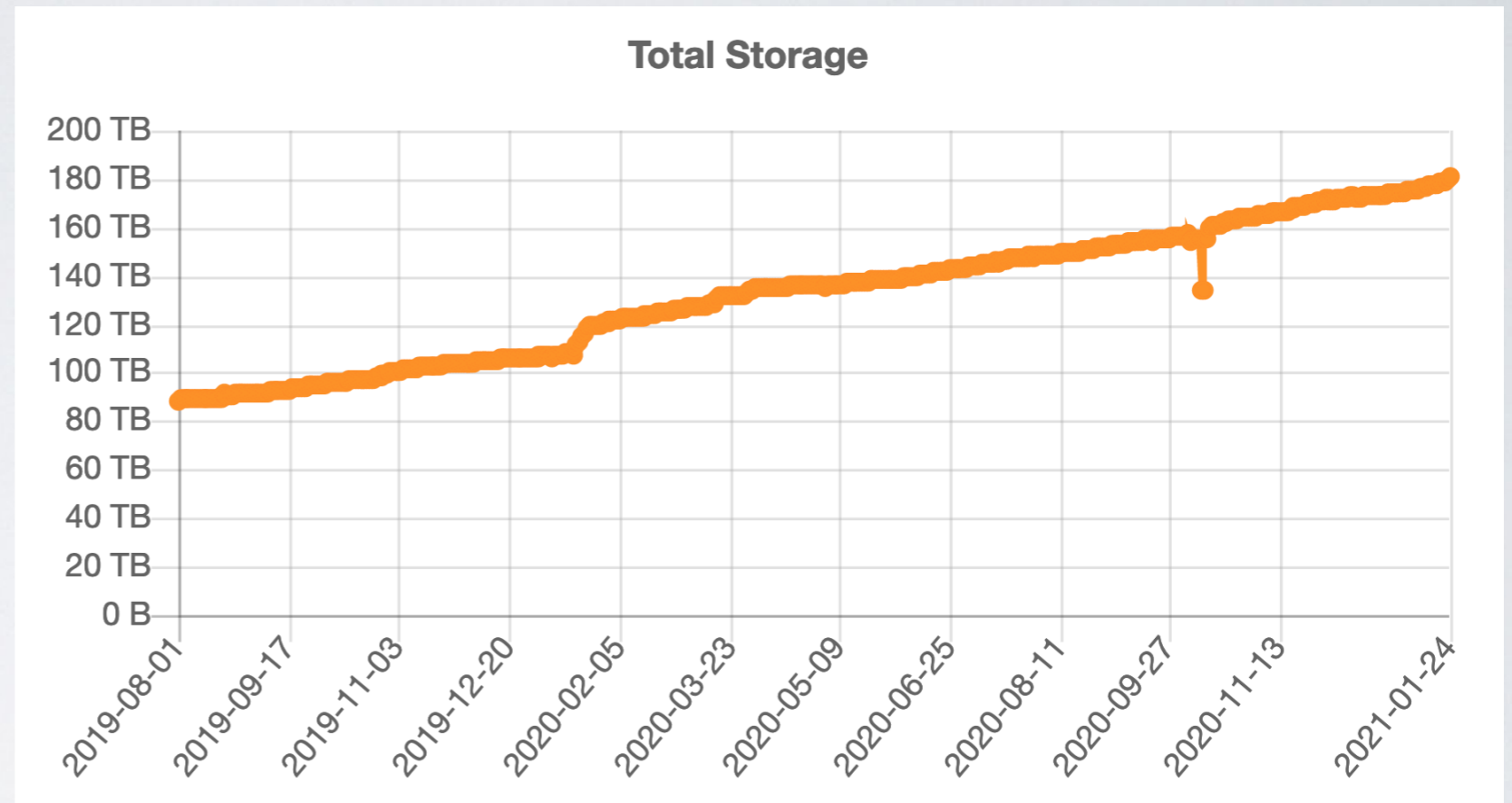


- Typical quota (2020):
  - 100GB / person,
  - 1-10TB / project

- Total space used:
  - **180 TB** logical
  - 212 TB physical (versioning overhead)

- Users demand more!:
  - 10+TB / project
  - 1+ TB / person

- Data mgmt complexity increases for growing # of users & groups

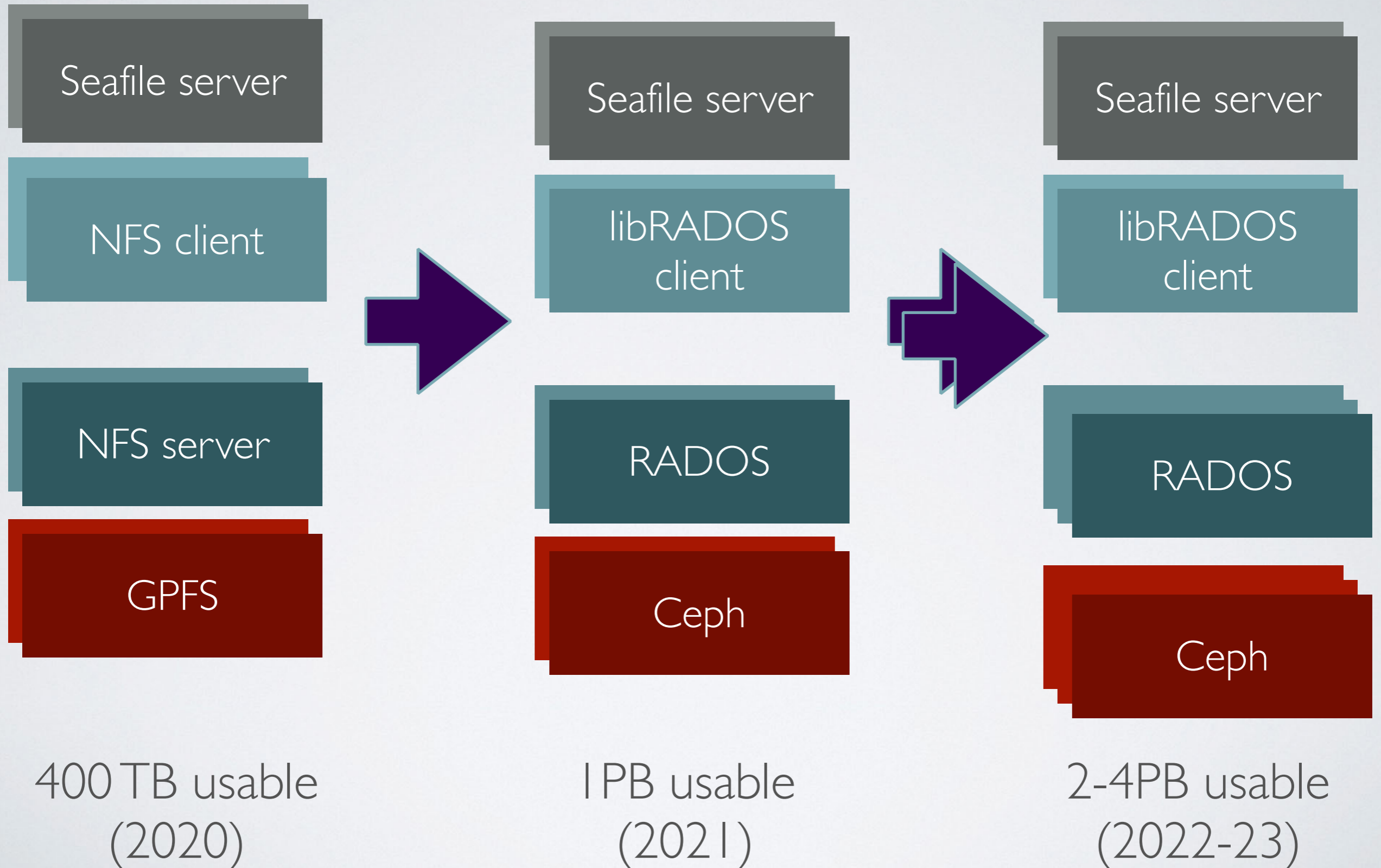


Project / User	Space usage [TB]
EU project on multimedia	59,97
PL project on biology	42,72
PR dept @PSNC	8,25
Lab @ PSNC	4,69
User in technical university	4,33
User @PSNC	3,07

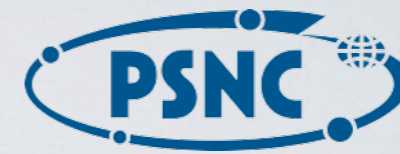




# NEW DEPLOYMENT - MORE SCALABLE



# PERFORMANCE?



- Our tests (in 2017) - see results on the next slide:
  - Small files:
    - GPFS: serves I/O faster, smaller latencies, predictable performance
    - Ceph: higher latency, less consistent performance, capable for caching (RAM!)
  - Large files:
    - GPFS and Ceph perform similarly
    - Ceph: upload: 3x replication penalty, downloads: ~10% difference
- => **NO SERIOUS RISKS FOR OVERALL PERFORMANCE**
- Tests to be repeated with:
  - new infrastructure (faster servers, more spins) and Seafile
  - need the fine-tuning of the configuration

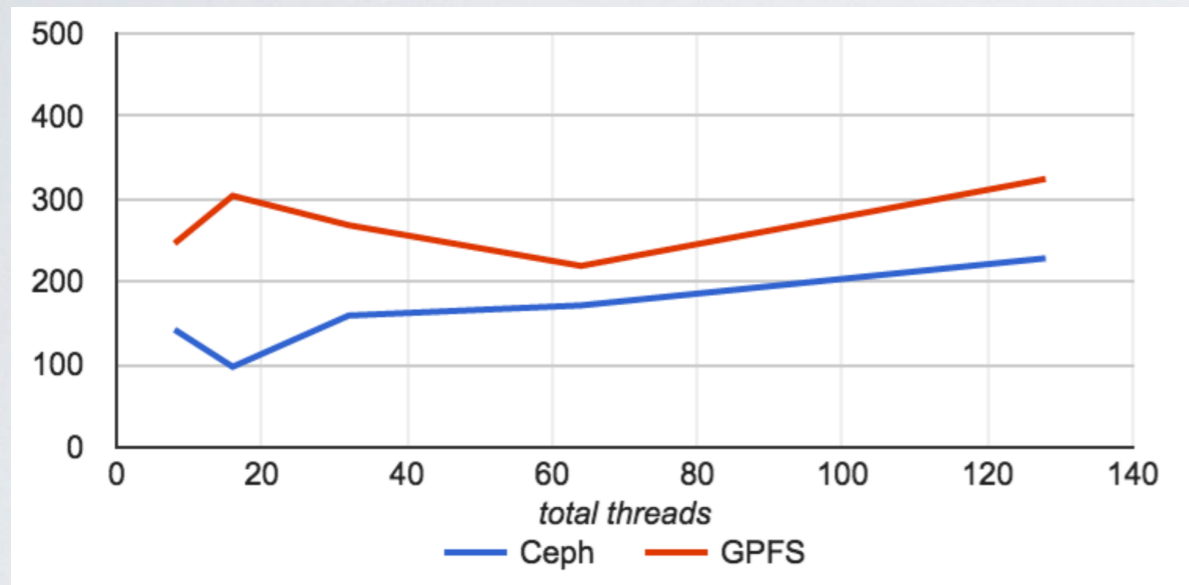


# SEAFILE PERFORMANCE

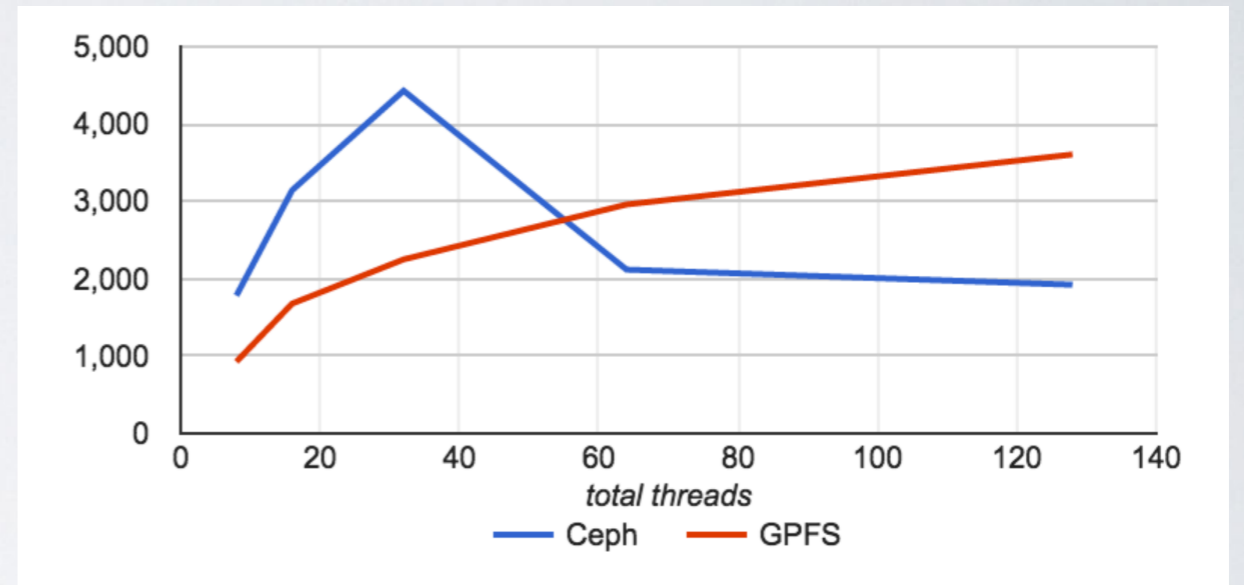
WITH CEPH AND GPFS BACKENDS (2017)

**SMALL FILES** TEST: 45K X 100KB FILES [FILES/S]

**UPLOAD**

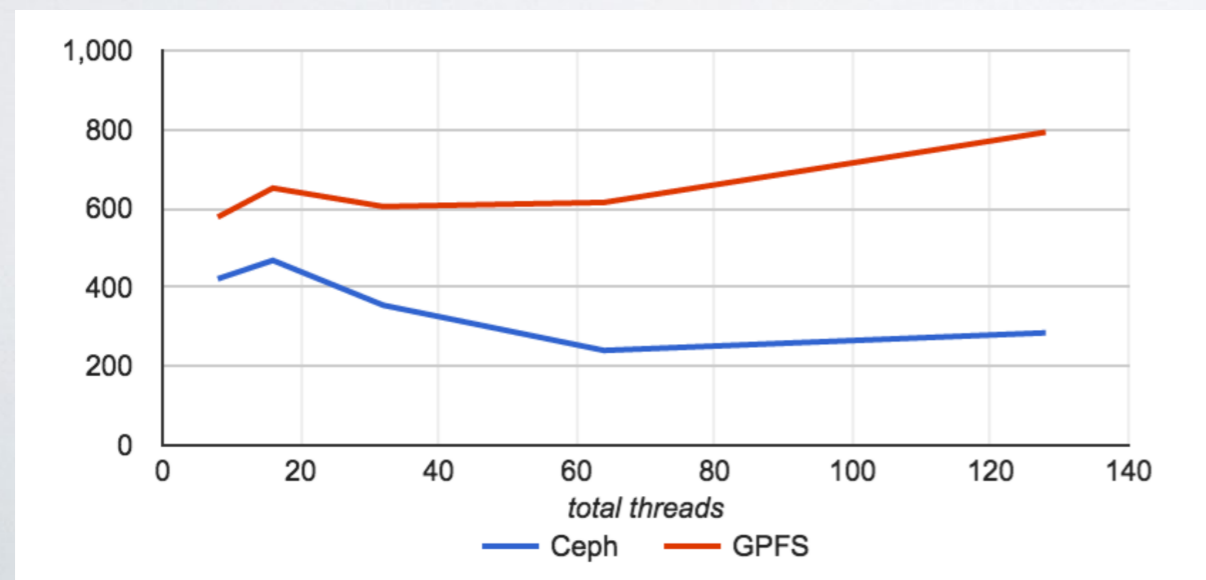


**DOWNLOAD**

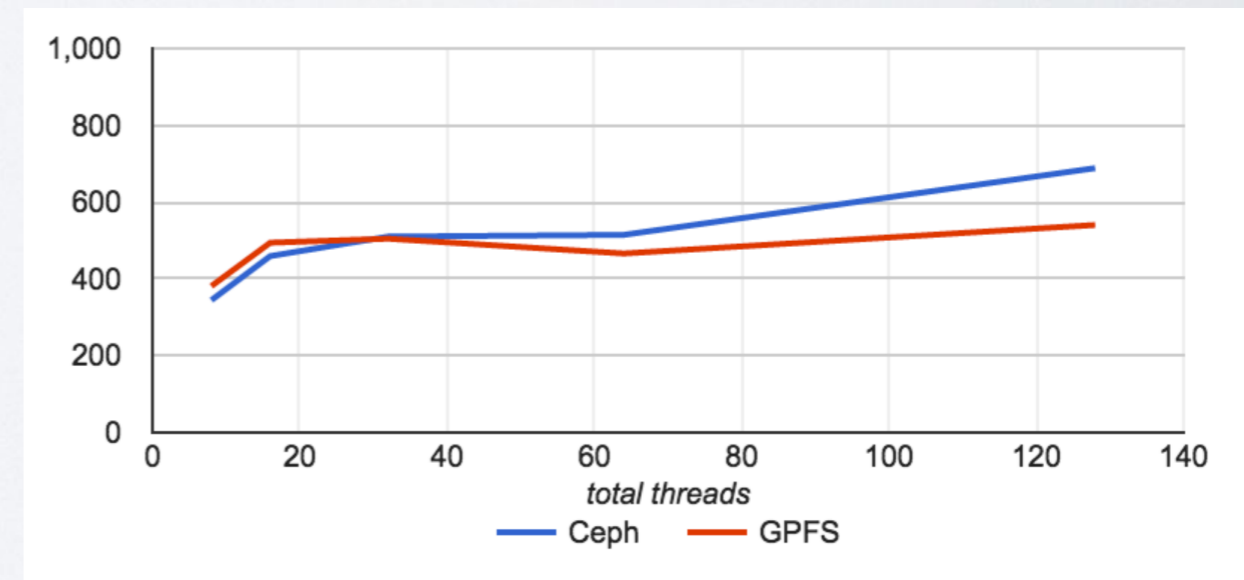


**LARGE FILES** TEST: 4,4 GB FILES: [MB/S]

**UPLOAD**

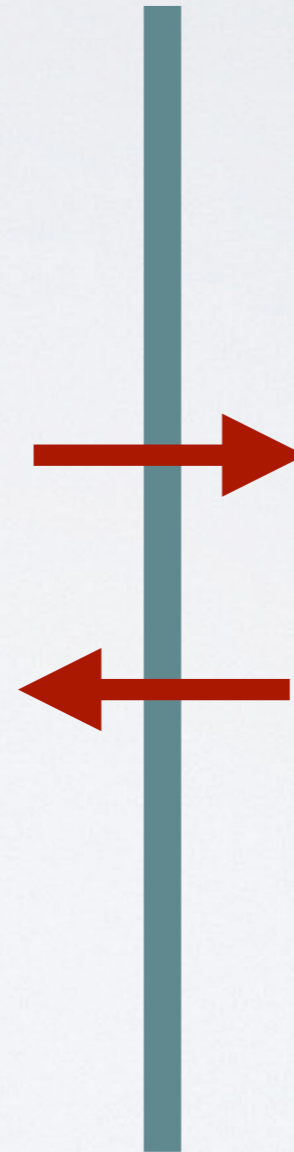


**DOWNLOAD**



# A SLIDE FROM 2018 TALK @CS3

## TYPICAL DATA FLOW \*) IN HPC



\*) They call it: „hassle”

User & data

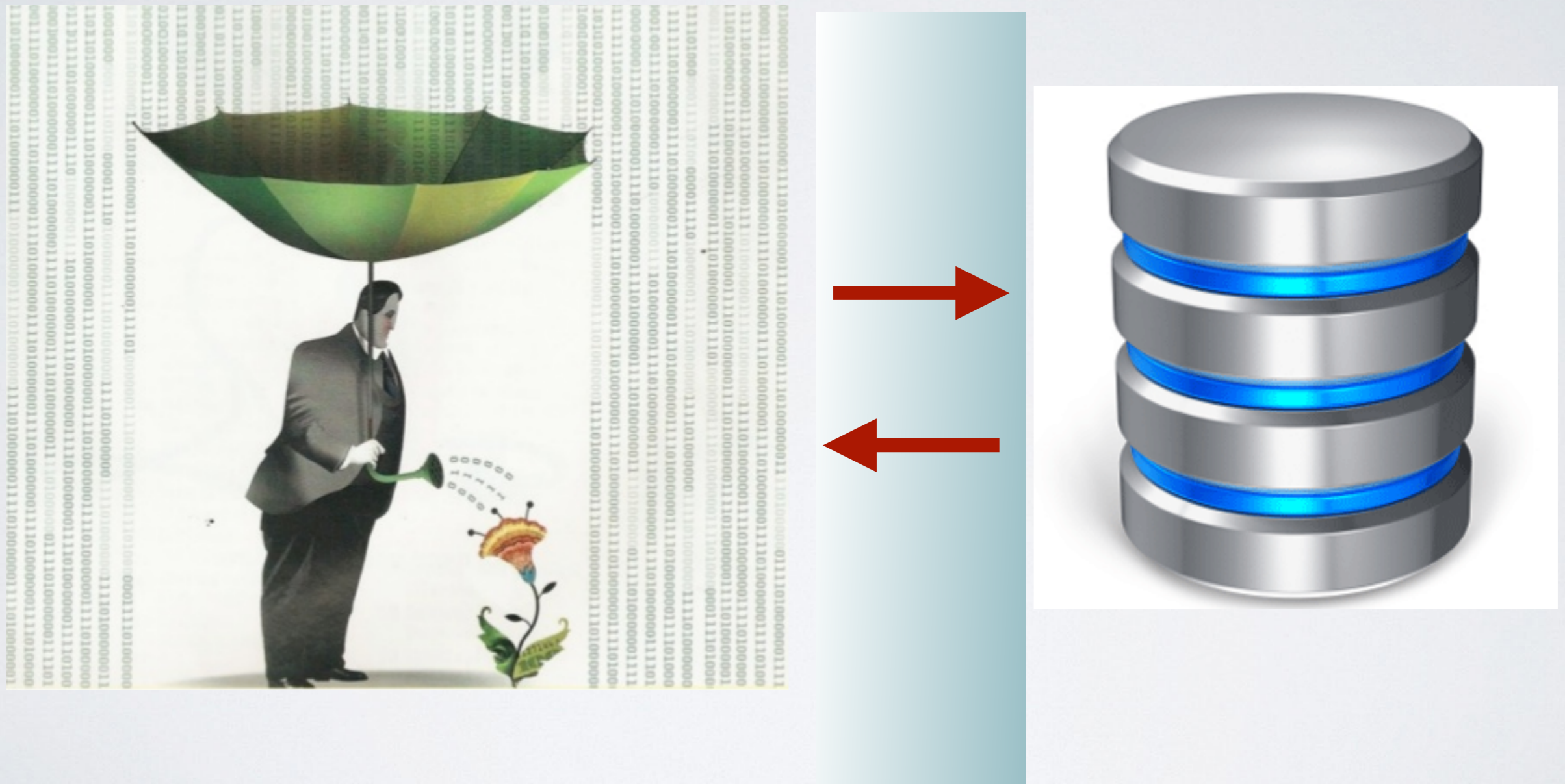
**Barrier**

Storage



# A SLIDE FROM 2018 TALK @CS3

## HOW TO MAKE DATA FLOW?



User & data

**Interface**

Storage

# PREVIOUS WORK



- **PRACE-LAB** - a HPC/Cloud-focused project:
  - sync & share as the **data exchange interface** for HPC/cloud users:
    - users store & access data in sync & share:
      - convenience, ease of use
      - all basic sync & share features plus collaboration tools
    - data are accessible for in-HPC and in-cloud applications:
      - synchronised (using sync clients) - space overhead
      - accessed (using virtual drive) - **no** space overhead



# Sync&share - **data exchange interface** for cloud/HPC



**User's system:**

- 10-100s MB/s
- 0.1-1TB

Cloud

**Cloud storage:**

- 0.1-1 GB/s
- 10s TB



Sync & share

Input data:



Output data:



**Sync&share (Seafile):**

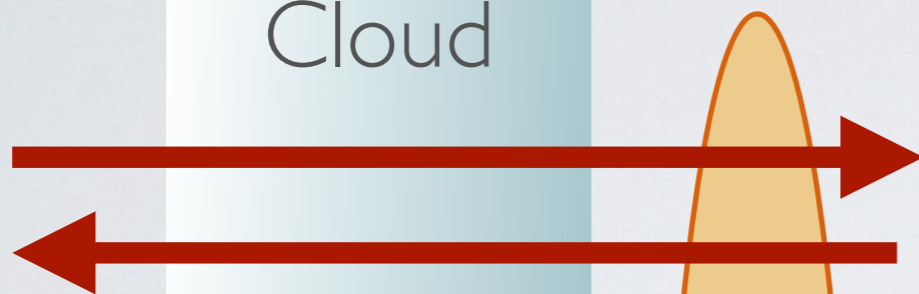
- 1-10 GB/s
- ~1 PB



HPC

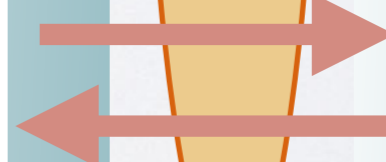
**HPC scratch and/or home**

- 10-100 GB/s
- 1-10s PB



SYNC & SHARE

SYNC & SHARE



# Sync&share - **data exchange interface** for cloud/HPC

- **analysis:**

- a synchronisation **delay is acceptable** in many use-cases (see our presentation at CS3'18 in Kraków)
- **ease of use comes at the cost** of extra space usage (cloud, HPC)
- using proper access tools eliminates overhead:
  - Seafile provides a **SeaDrive** client (virtual drive based on FUSE)
  - **only the active/ used / modified data are sync'd** down to the Cloud storage and/or HPC storage
- issues?:
  - sync&share **load** on cloud storage / scratch filesystem
  - users and **identity** management - and credentials „delegation”?
    - workaround is to create robotic user and dedicated „organisation”
    - more advanced solution in progress (including Keycloak...)

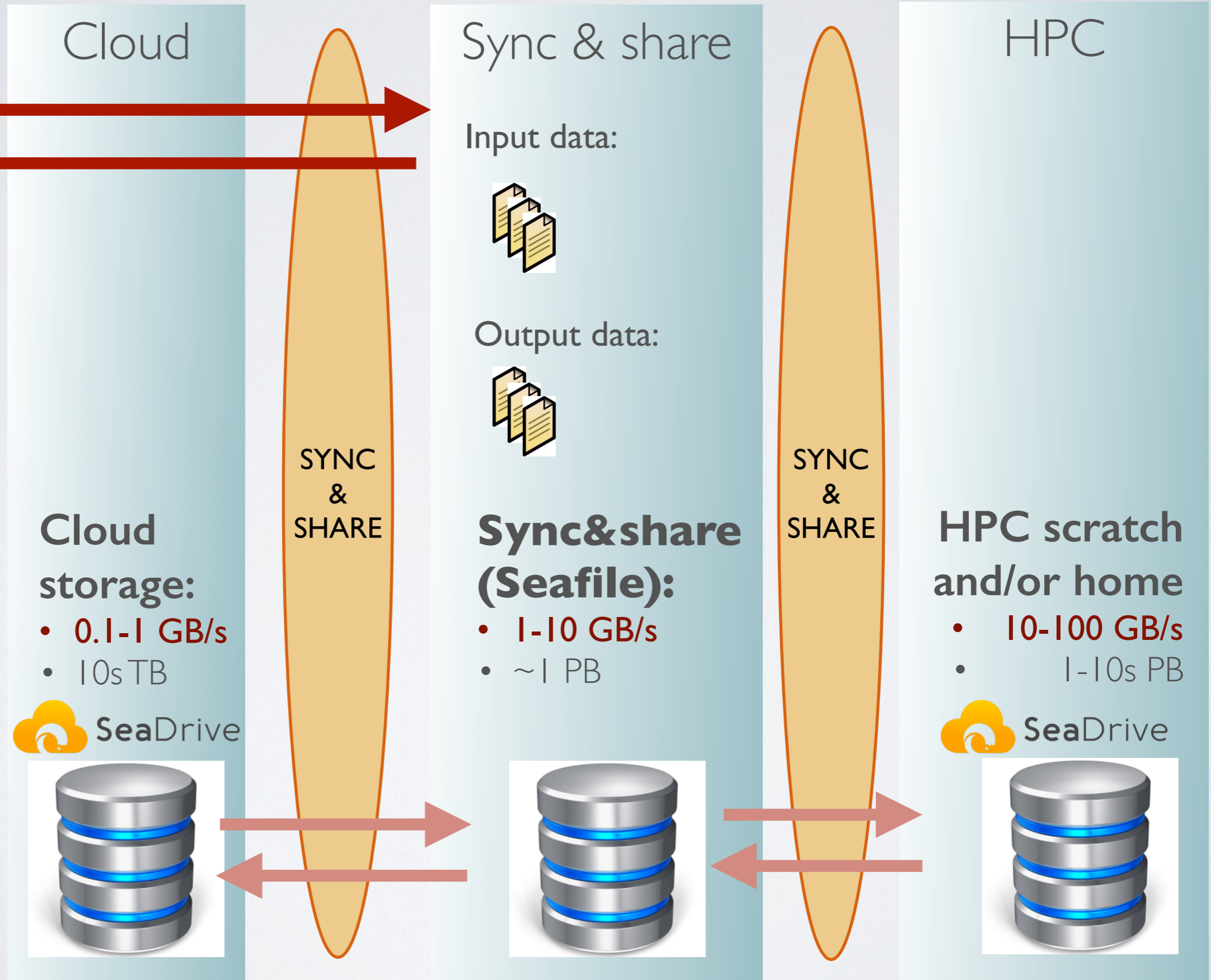


# Sync&share - **data exchange interface** for cloud/HPC



**User's system:**

- 10-100s MB/s
- 0.1-1TB



Cloud

Sync & share

HPC

Input data:



Output data:



SYNC  
&  
SHARE

SYNC  
&  
SHARE

Cloud  
storage:

- 0.1-1 GB/s
- 10s TB



Sync&share  
(Seafile):

- 1-10 GB/s
- ~1 PB



HPC scratch  
and/or home

- 10-100 GB/s
- 1-10s PB



# Sync&share - **data exchange interface** for cloud/HPC



**User's system:**

- 10-100s MB/s
- 0.1-1TB



Cloud



SYNC  
&  
SHARE

**Cloud storage:**

- 0.1-1 GB/s
- 10s TB



Sync & share

Input data:



Output data:



**Sync&share (Seafile):**

- 1-10 GB/s
- ~1 PB

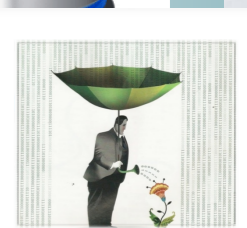
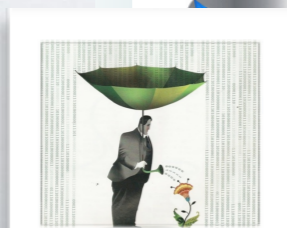


HPC

SYNC  
&  
SHARE

**HPC scratch and/or home**

- 10-100 GB/s
- 1-10s PB



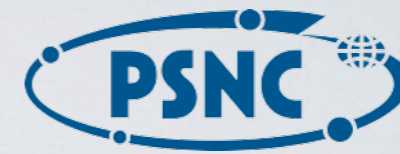


# ONGOING WORK



- **PRACE-LAB** - automatic **provisioning of sync & share space** to cloud/HPC
  - users / groups need the **dedicated ,collaboration' space**
    - **HPC: grant users** collaborate on datasets related to their research
    - **Cloud: tenants** may need shared storage (Manila style)
  - Implementation: **,robotic user'** and **,organisation'** (Seafile feature) created, users signed to **,organisation'**; quota defined for **,organisation'**
    - **HPC: grant users space** / data **,mounted'** in the scratch FS and syncd using sync client or accessed on-demand through Seadrive
    - **Cloud: tenant-dedicated space** **,mounted'** as **,share'** (à la OpenStack Manila-managed shares); sync client or Seadrive for data access;

# ONGOING WORK



- **PRACE-LAB:** automatic **provisioning of sync & share space** to cloud/HPC
  - **analysis:**
    - automation makes life of **users** easier:
      - **no need for manual setup** of shares in cloud and HPC
      - integrated management of compute and storage services
    - automation helps **admins** managing storage spaces and datasets for cloud and HPC; this enables:
      - **removing manual work** on storage spaces configuration for users;
      - binding data sets/group dirs in sync&share to HPC grants / cloud tenants (no orphan data sets residing in sync&share after computing grants expire)
      - associating data sets/group in sync&share to cloud tenants (benefits as in HPC case)
      - managing user/group accounts and data sets **retention** in an **integrated way across various types of services** - that we (historically) managed separately

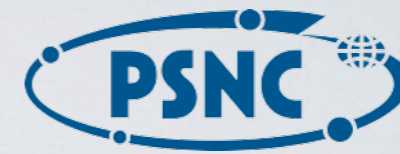


# NEW USE-CASES (I)



- **National Data Storage** project:
  - **basic data management services:**
    - local data storage and access  
(disk pools, SSD/NVMe, tape; automatic tiering)
    - safe data storage  
(geo-replication based on Ceph/S3 for object storage)
  - **sync & share** as the data storage/access interface
    - complementary to S3 and NFS
    - implementation to be worked out:
      - possibly through a gateway
      - doable thanks to **extensive API of Seafire**

# NEW USE-CASES (2)



- **National Data Storage** project:
  - long-term storage and preservation **services integrated:**
    - data **repositories embedded** in (mostly S3-based) storage infrastructure provided by NDS project: Dataverse, Zenodo, dLibra (PSNC) and possibly others
    - sync & share -> **depositing data to** -> **repositories** (based on CS3MESH4EOSC results ported to NDS infrastructure)
    - sync & share -> **long-term storage** (Ceph/S3, tape etc.) (partially based on CS3MESH4EOSC results ported to NDS infrastructure, possibly a separate, custom-made solution)



# NEW USE-CASES (3)



- **National Data Storage** project:
  - PSNC's sync & share service **federations**:
    - integration with 'the MESH'
      - based on the results of **CS3MESH4EOSC** project:
        - CS3 APIs as the basis
        - Seafile-side implementation of the MESH APIs  
(approach to be worked out, hopefully supported by Seafile)
        - PSNC is the CS3MESH4EOSC partner
  - federation of sync & share systems of the NDS project partners  
in order to achieve (at least ) the national level federation



# SUMMARY



- There are and will be **synergies** among projects and sync & share services; in both directions
  - Projects build & deliver the infrastructure and basic services for:
    - increasing scale & reach of sync & share (capacity, performance, technology: e.g. GPFS->Ceph)
    - improved reliability (better data protection: replication / Erasure Coding, geo-replication)
  - Sync & share **as the front-end:**
    - makes life of users easier (applies to those *already convinced* to use our infrastructure)
    - attracts / brings new users (first they use simple services, then become ,pro' users)
  - Sync & share **extensions / integrations::**
    - make sync & share system more functional and addressing new use-cases
    - enable integrated management of data services and data sets



# NATIONAL DATA STORAGE SYNERGY WITH SYNC&SHARE & SERVICES

THANK YOU!  
QUESTIONS?

Contact:

[maciekb@man.poznan.pl](mailto:maciekb@man.poznan.pl)

[box@pionier.net.pl](mailto:box@pionier.net.pl)

[www.psnc.pl](http://www.psnc.pl)

[box.pionier.net.pl](http://box.pionier.net.pl)

