Overview and Progress of the RED ONION project

Susumu Date, Ph. D
The Joint Research Laboratory for Integrated Infrastructure of High Performance Computing and Data Analysis/ Applied Information Systems Research Division
Cybermedia Center, Osaka University, Japan

Cybermedia Center, Osaka University

CMC main building

IT core as data center

- Supercomputing center at Osaka University
  - has a responsibility for providing a powerful high-performance computing environment for university researchers across Japan as a national joint-use facility.

SQUID since May 2021

Supercomputer for Quest to Unsolved Interdisciplinary Datascience

Cybermedia Center, Osaka University

- **Cloud-linked High Performance Computing and High Performance Data Analysis Supercomputer System (Supercomputer for Quest to Unsolved Interdisciplinary Datascience)**
  - **Peak Performance 16.591 PFlops**

SQUID システム構成

CPU nodes	
1520 nodes x peak perf. 5.837 TFlops 8.871 PFLOPS	
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38コア) 2 基
主記憶容量	256 GB

GPU nodes	
42 nodes x peak perf. 161.836 TFlops 6.797 PFLOPS	
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38 コア) 2 基
主記憶容量	512 GB
GPU	NVIDIA HGX A100 8 GPU ボード (Delta)

Vector nodes	
36 nodes x peak perf. 25.611 TFlops 0.922 PFLOPS	
プロセッサ	AMD EPYC 7402P (2.8 GHz 24コア) 1 基
主記憶容量	128 GB
Vector Engine	NEC SX-Aurora TSUBASA Type 20A 8 基

Interconnect	
ノード間接続	Mellanox InfiniBand HDR (200 Gbps)

ONION data aggregation Infra.

S3-compatible Parallel File System 21.2PB	
ファイルシステム	DDN EXAScaler (Lustre)
HDD	20.0 PB
SSD	1.2 PB

S3-compatible Object Storage 500TB	
オブジェクトストレージ	CLOUDIAN HyperStore
HDD	500 TB

JHPCN (Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructure)

- Joint research collaboration framework in Japan
  - https://jhpcn-kyoten.itc.u-tokyo.ac.jp/en/

- International Joint Research Projects

International joint research projects are interdisciplinary joint research conducted in conjunction with researchers outside Japan to address challenging issues that may not be possible to resolve or clarify only with the help of researchers within Japan. For such research projects, there will be subsidy paid to cover travel expenses necessary for holding meetings with foreign joint researchers and so on. For details of the subsidy, please contact our office after your research project has been accepted.

What is ONION?

SQUID since May 2021

Supercomputer for Quest to Unsolved Interdisciplinary Datascience

Cybermedia Center, Osaka University

- **Cloud-linked High Performance Computing and High Performance Data Analysis Supercomputer System (Supercomputer for Quest to Unsolved Interdisciplinary Datascience)**
  - **Peak Performance 16.591 PFlops**

SQUID システム構成

CPU nodes	
1520 nodes x peak perf. 5.837 TFlops	8.871 PFLOPS
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38コア) 2 基
主記憶容量	256 GB

GPU nodes	
42 nodes x peak perf. 161.836 TFlops	6.797 PFLOPS
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38 コア) 2 基
主記憶容量	512 GB
GPU	NVIDIA HGX A100 8 GPU ボード (Delta)

Vector nodes	
36 nodes x peak perf. 25.611 TFlops	0.922 PFLOPS
プロセッサ	AMD EPYC 7402P (2.8 GHz 24コア) 1 基
主記憶容量	128 GB
Vector Engine	NEC SX-Aurora TSUBASA Type 20A 8 基

Interconnect	
ノード間接続	Mellanox InfiniBand HDR (200 Gbps)

ONION data aggregation Infra.

S3-compatible Parallel File System 21.2PB	
ファイルシステム	DDN EXAScaler (Lustre)
HDD	20.0 PB
SSD	1.2 PB

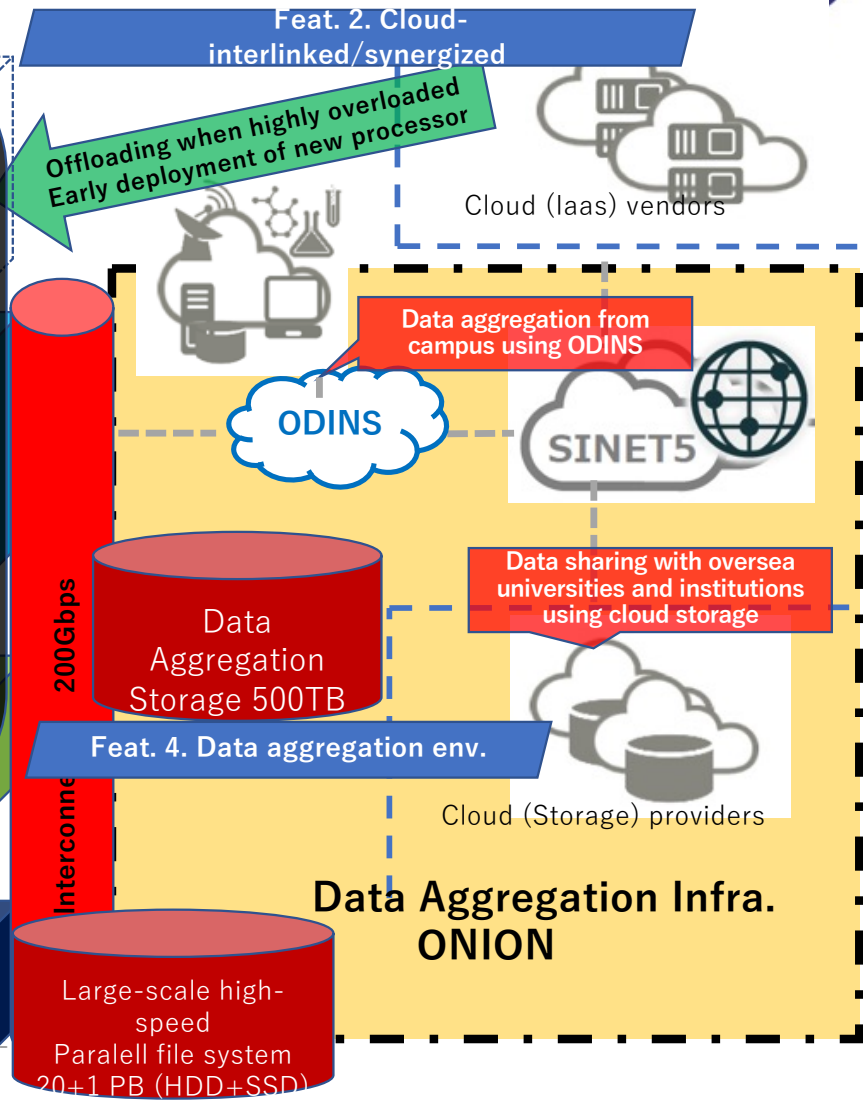
S3-compatible Object Storage 500TB	
オブジェクトストレージ	CLOUDIAN HyperStore
HDD	500 TB

ONION concept

ONION, data aggregation infrastructure

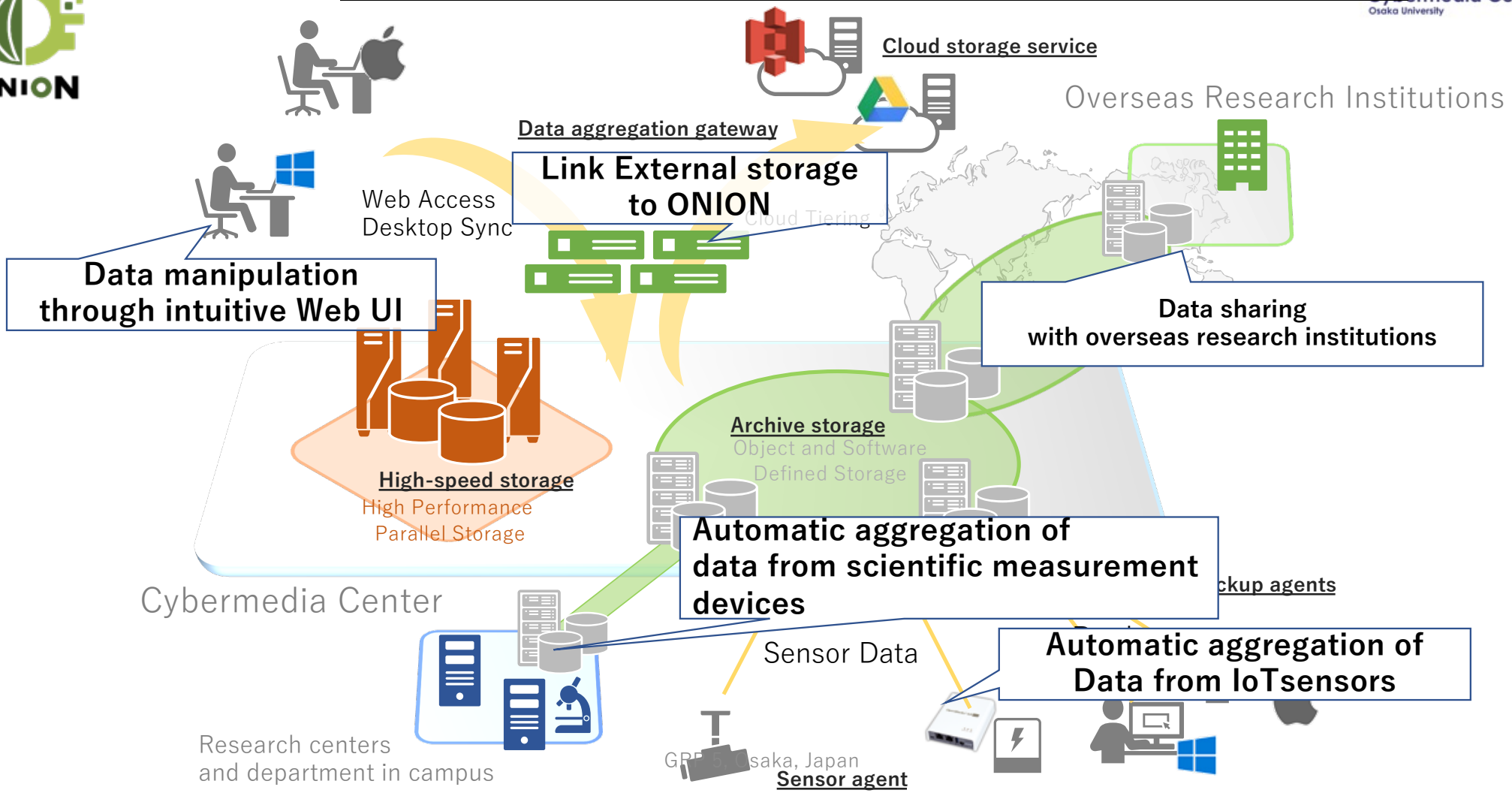
- Osaka university Next-generation Infrastructure for Open research and open innovatioN
- Data aggregation infrastructure that not only enables the sustainable handling of ”super big data” generated in Osaka University in a responsible manner while ensuring the sustainability of such data into the future but also facilitates utilization of research data for “co-creation between academia and industries” and “international research collaboration”.
- introduce as PoC (Proof-of Concept) implementation in the procurement of SQUID on a trial basis.

(1) The primary purpose of procurement is supercomputing system, not for data storage.
(2) Our designed ONION might not be useful and thus not be used.

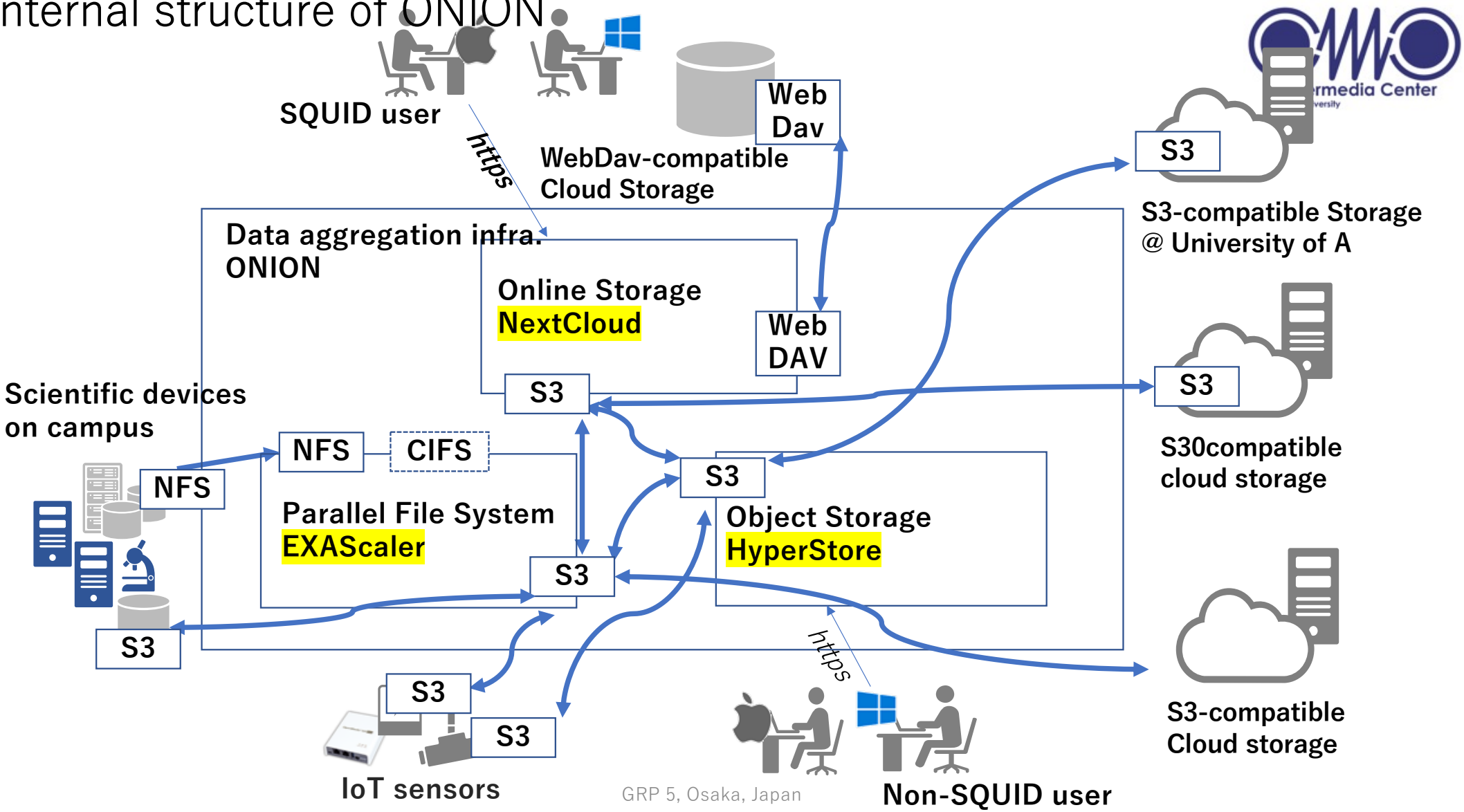


Overview of ONION

3 storage solutions are seamlessly combined in a synergic manner



Internal structure of ONION



Efficient utilization and aggregation of scientific data from scientific measurement devices.

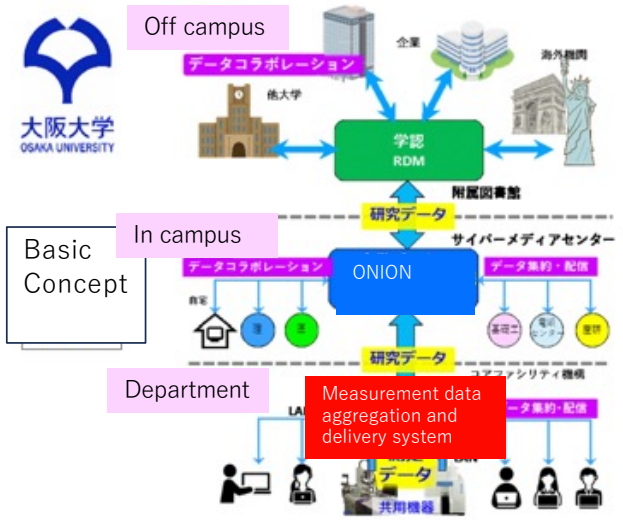
Joint use of scientific devices on campus

Every scientists in university can use scientific devices.

Approximately 300 scientific devices is used.
80,000 measurements /year.
Generates a lot of measurement (research) data

Core Facility center is in charge of these scientific devices.

Concept: Realization of Data utilization infrastructure that aggregates and deliver the data daily generated from scientific devices to collaborators



Tight collaboration among three departments

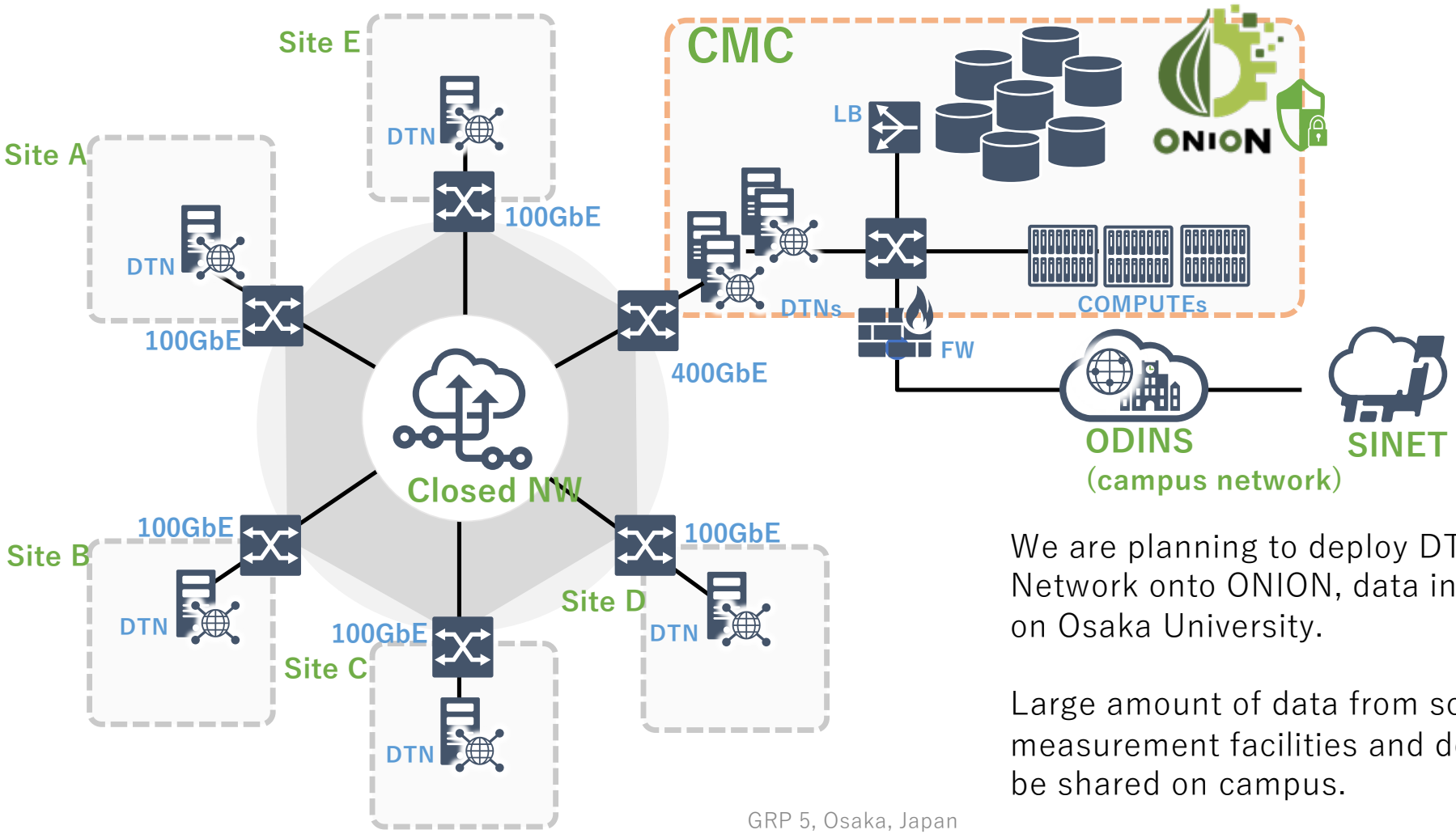
Library — Promotion of "Open Data"

Cybermedia Center — system

Core facility — Date generation

Courtesy of 古谷浩志 (大阪大学 コアファシリティ機構), "大阪大学コアファシリティ機構における測定データ流通・集約基盤の構築と阪奈機器共用ネットワークへの展開", 第1回北陸地区学術データ基盤セミナー ～コアファシリティ連携から研究データエコシステム構築を目指して～, 金沢, 2024年2月.

RED (Research-EnhanceD) ONION (towards Science DMZ)



We are planning to deploy DTN and high-speed Network onto ONION, data infrastructure on Osaka University.

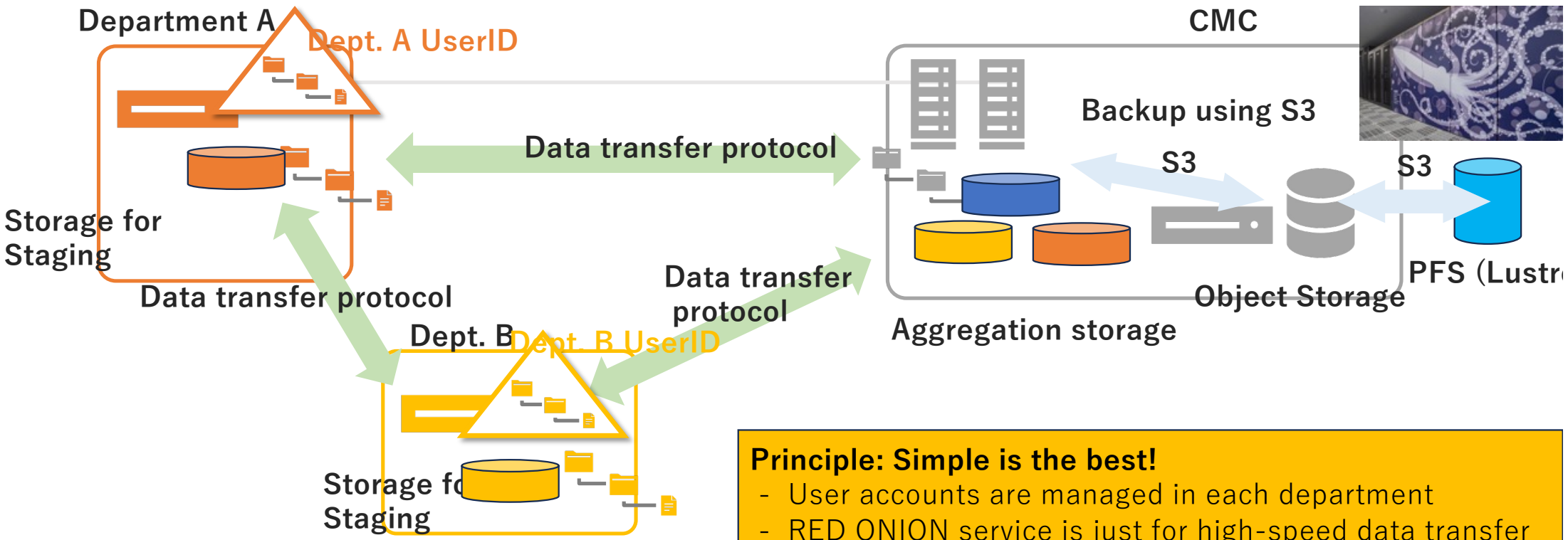
Large amount of data from scientific measurement facilities and devices can be shared on campus.

Towards RED ONION

1. Designing RED ONION from actual operation perspective

2. Performance Evaluation of DTN solutions in LAN/SC2023

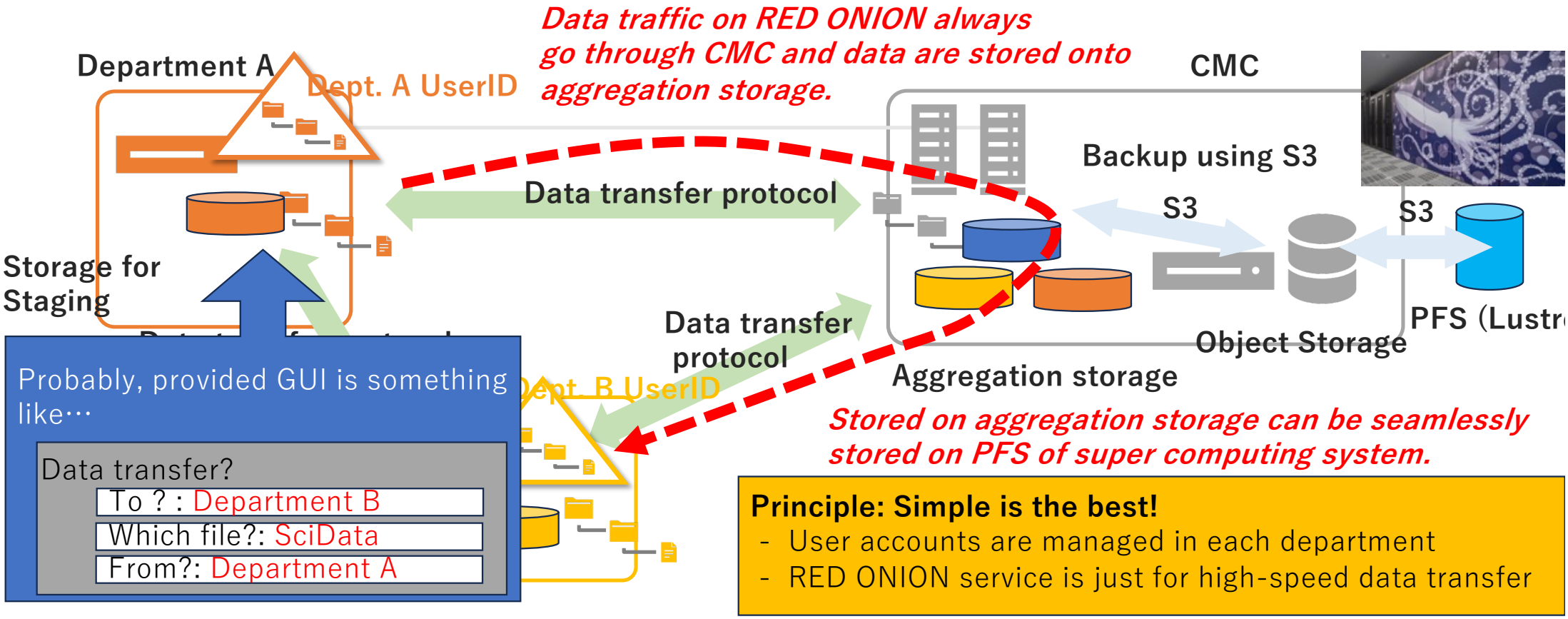
1. Designing RED ONION from actual operation perspective



Principle: Simple is the best!

- User accounts are managed in each department
- RED ONION service is just for high-speed data transfer

1. Designing RED ONION from actual operation perspective



2. Performance Evaluation of DTN solutions in LAN

- Performance of DTN is the most important. The Second importance is easiness in administration, operation and maintenance of DTN.
    - In our case, university does not have operation team enough to manage and administer RED ONION service in addition to supercomputing systems. Therefore, we currently need technical support from IT companies for daily operation.

Performance investigation and evaluation in SC2023

- We worked together with NICT (National Institute of Information and Communications Technology), Japan for performance evaluation / demonstration of RED ONION project in SC2023 Denver.
  - Before SC2023, we pre-tested it between Osaka U and NICT in Japan.

SC23 Network Research Exhibition: Demonstration Preliminary Abstract

[Performance Evaluation of DTNs Towards Research-EnhanceD ONION (RED ONION)]

Susumu DATE, Cybermedia Center, Osaka University, date@cmc.osaka-u.ac.jp

Abstract

The Cybermedia Center (CMC) is a supercomputing center in Osaka University, Japan. The CMC is in charge of providing a high-performance computing environment for Japanese researchers in both academia and industries. From the perspectives of the rising demand on high performance data analysis characterized AI (artificial intelligence), ML (machine learning) and DL (deep learning) and the necessity of research reproducibility, the CMC has been working on data aggregation infrastructure named ONION (Osaka university Next-generation Infrastructure for Open research and open innovatioN) in campus, so that researchers can perform HPC and HPDA immediately after obtaining scientific data from scientific measurement devices and research data including computation results are managed in a proper way. In the future vision, we are exploring the development of RED ONION, which allows research institutions and departments in campus to transfer large amount of research data through the use of Data Transfer Nodes on high-speed campus networks. For the purpose, we are planning to use this SC2023 demonstration opportunity to learn the performance characteristics of candidate DTN technologies over a wide-area network between US and Japan for our RED ONION concept.

Goals

Also, we need 100Gbps network between Osaka University and NICT booth in SC2023. We would like to have a L2 connection between them. Also, it would be greatly helpful if we can use the network even in a time-shared manner. This is because we plan to have direct connection from the CMC to each department or research institute on Campus on the basis of 100G ethernet. Also, we would like to verify the performance of DTNs on a wide-area network because we personally would like to apply this data transfer solution to the network with our partner universities.

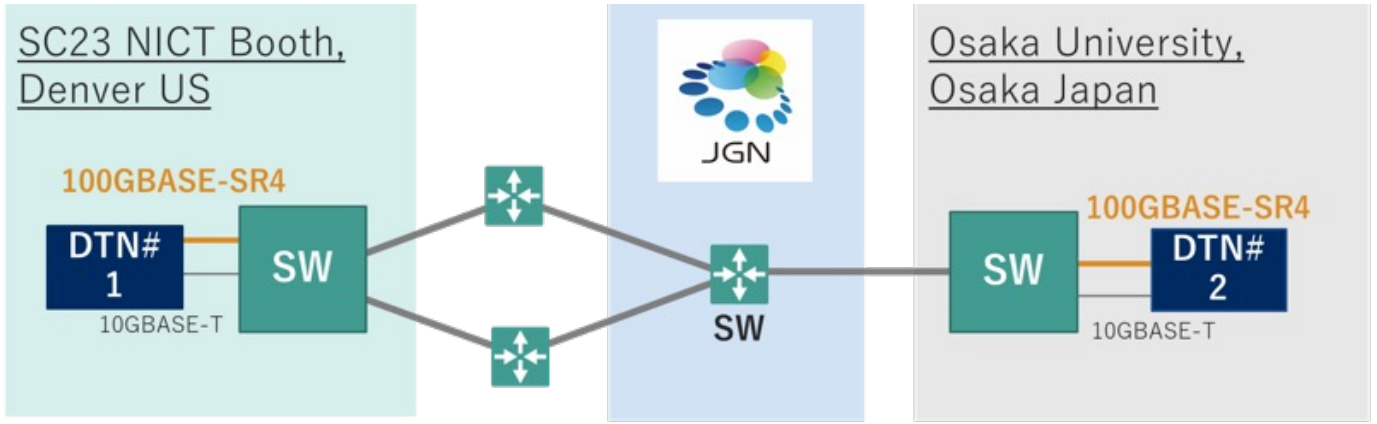
Involved Parties

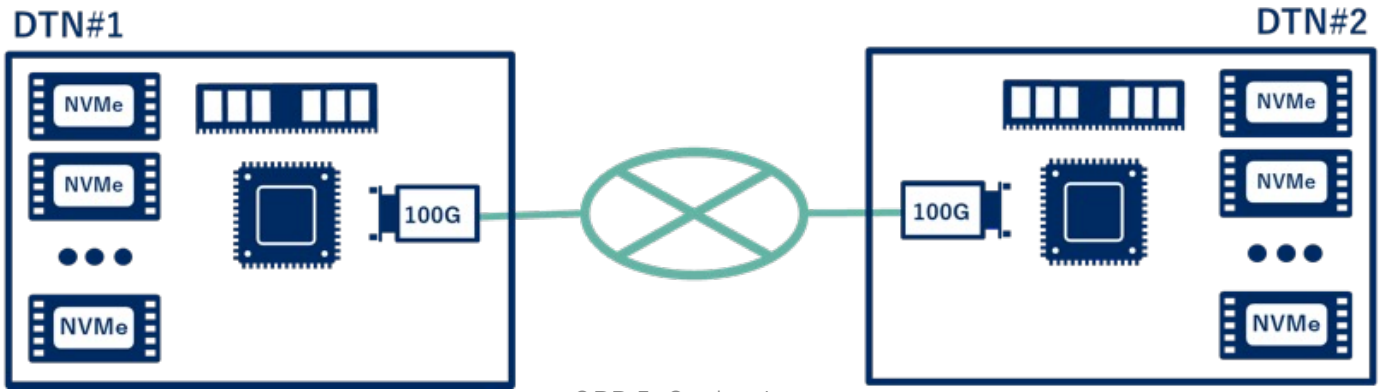
[List of other institutions, researchers and entities involved in the planning and execution of this demonstration. This should include names and contact information]

- Susumu Date, Osaka University, date@cmc.osaka-u.ac.jp
- Kenji Ohira, Osaka University, ohira@cmc.cosaka-u.ac.jp
- Kodai Fukuda, Osaka University, fukuda.kodai@ais.cmc.osaka-u.ac.jp
- Hideyuki Tanushi, Osaka University, h-tanushi.cmc@osaka-u.ac.jp

For high-speed data transfer

- Experiment at SC2023





Memory-to-Memory (tmpfs-to-tmpfs) transfer

Data Transfer Tools	Multiple File Transfer	Single File Transfer
ftp	99Gbps @ 16 files	7.5Gbps
GridFTP	98Gbps @ 16 files	17Gbps
XRootD	41Gbps @ 16 files	1.8Gbps
xxxxx	99Gbps @ 4 files	42Gbps 75Gbps (/dev/null)

- ✓ **Multiple File Transfer**: Many tools can achieve 100Gbps.
- ✓ **Single File Transfer**: Many tools can't achieve 100Gbps. It seems using tmpfs causes performance loss.

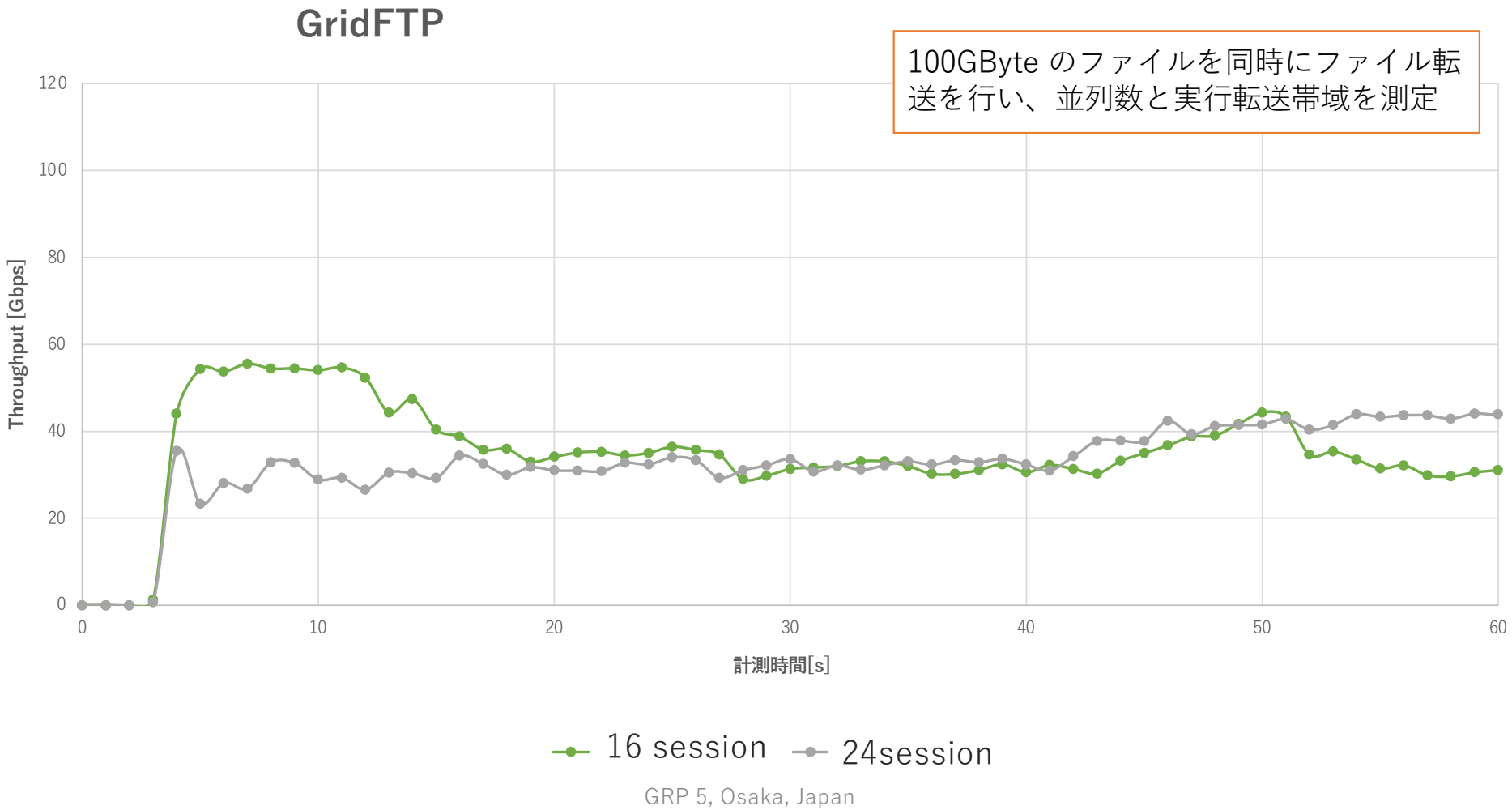
Memory-to-Disk transfer (tmpfs-to-xfs)

Cybermedia Center
Osaka University

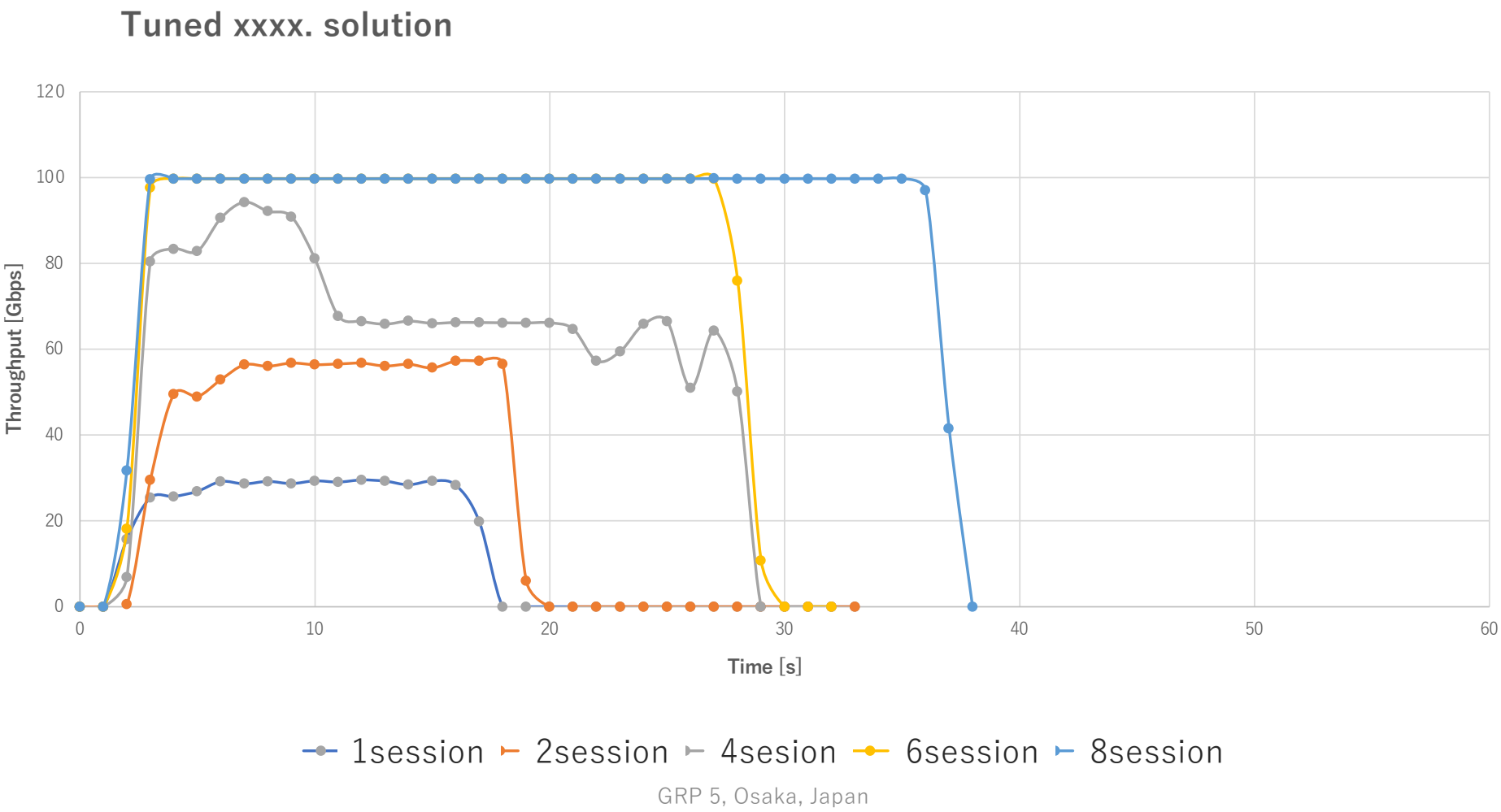
Data Transfer Tools	Multiple File Transfer	Single File Transfer
ftp	50Gbps @ 16 files	10Gbps
GridFTP	36Gbps @ 8 files	8Gbps
XRootD	41Gbps @ 32 files	1.9Gbps
xxxxx	70Gbps @ 4 files	34Gbps

- ✓ **Multiple File Transfer**: Less than 100Gbps. Possible issues on file I/O (Buffer, Sync, Random vs. Direct, Async, Sequential).
- ✓ **Single File Transfer**: Archaea shows good performance, but performance is lower than that of multiple file transfer.

Performance of parallel data transfer

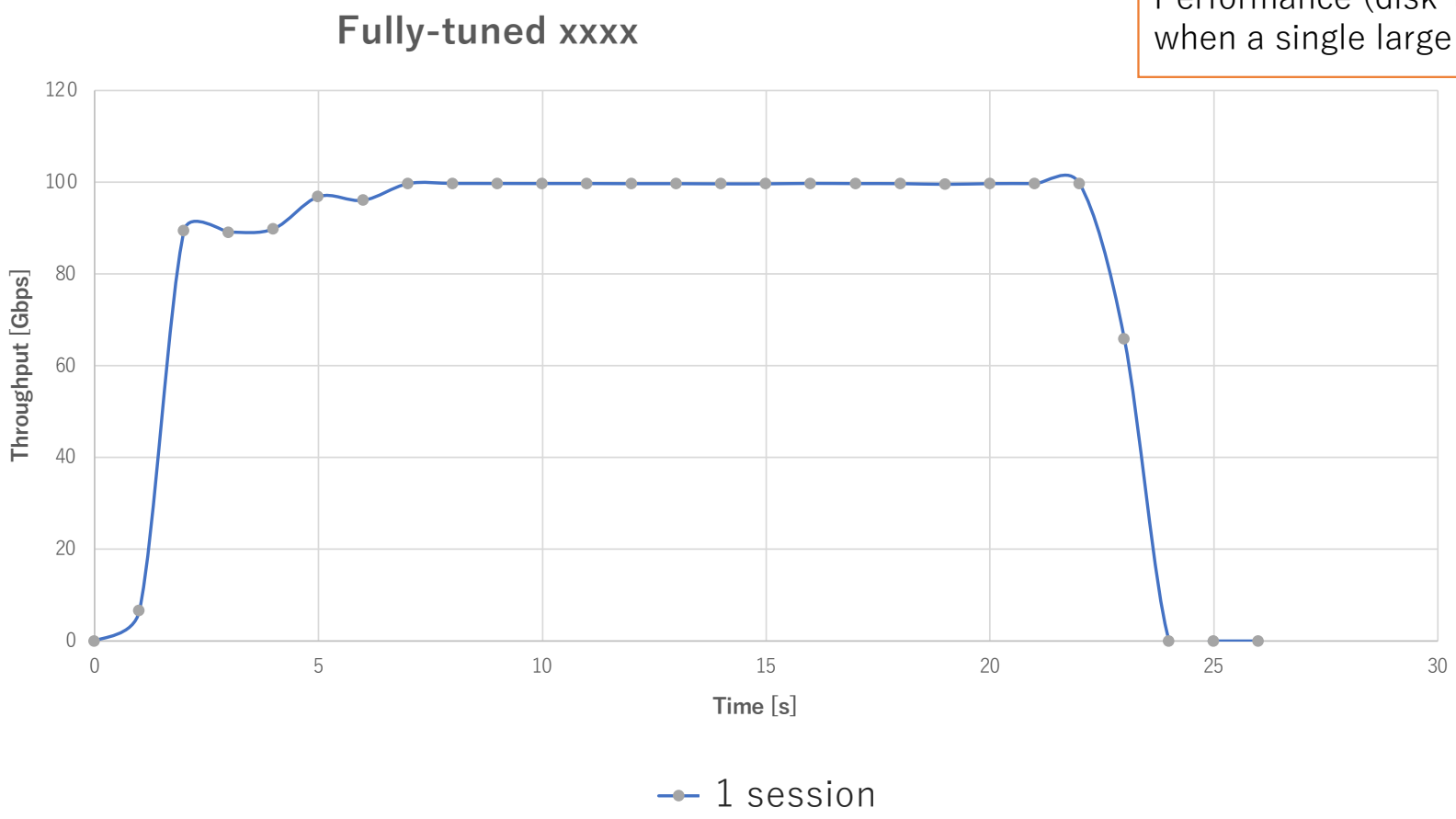


Performance of Tuned xxxx DTN



Performance of Tuned xxxx DTN. As of July 9 in 2024

Performance (disk-to-disk)
when a single large file was sent ,



Summary

Cybermedia Center
Osaka University

- Towards the realization of RED ONION environment that allows our departments in campus to easily transfer large amount of data with each other, we are now designing RED ONION in detail.

- For the purpose, the performance of DTN solutions is the most important. Also, for stable operation and administration of RED ONION, the easiness is also being evaluated.