Activity & Plan:
Collaboration with ALICE Computing

School of Computer Science
Chungbuk National University

January 7, 2022

Seo-Young Noh

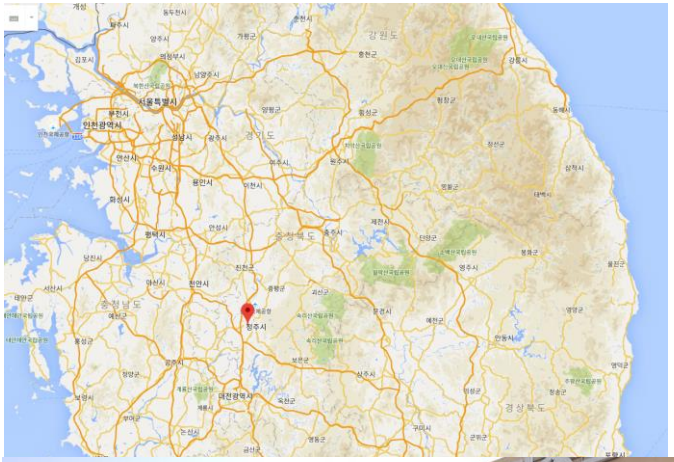
Contents

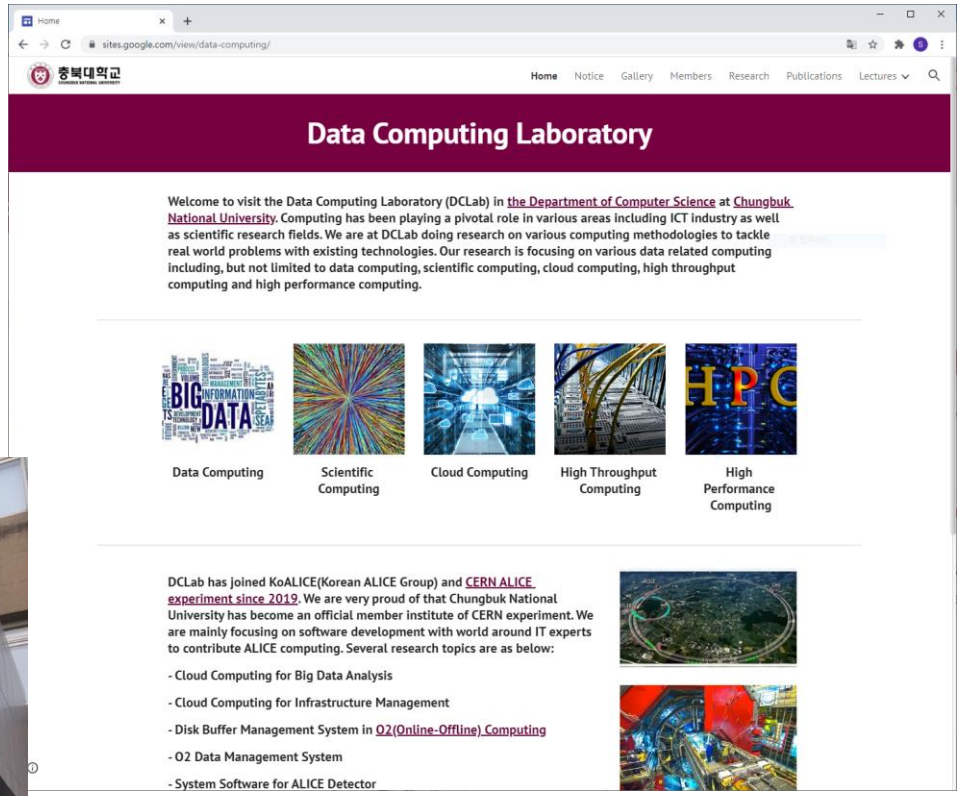
1. Team Introduction
2. Collaboration with ALICE Computing
3. Activities
4. Plan

Team Introduction

Data Computing Laboratory

- https://dclab.cbnu.ac.kr





Members

Current Students (MS)

Moon-Hyun Kim(20)

Jun-Yeong Lee(20)

Graduate in Feb, 2022

Kyung-Jun Kim(21)

NAVER Cloud

New Ph.D Student

Jae-Hyuck Shin (Part Time)
10-year Career at Samsung Electronics

New M.S Student

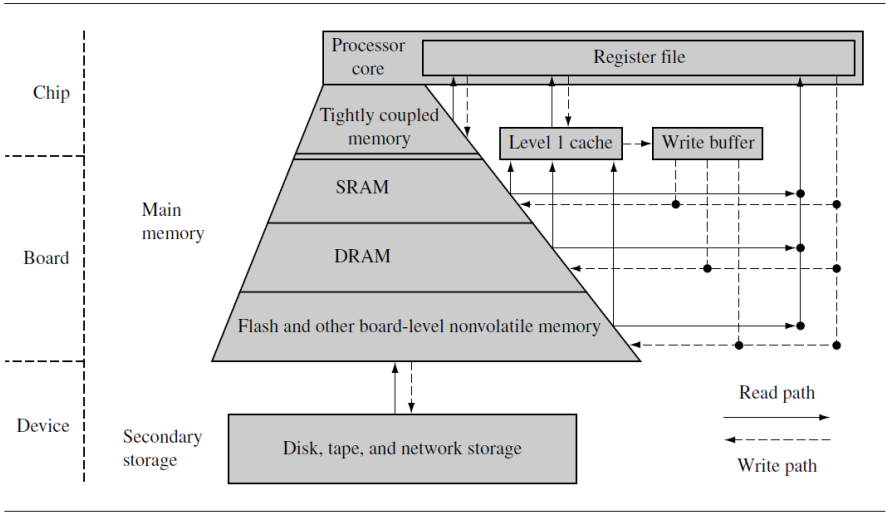
Seong-Min Kim

Collaboration with ALICE Computing

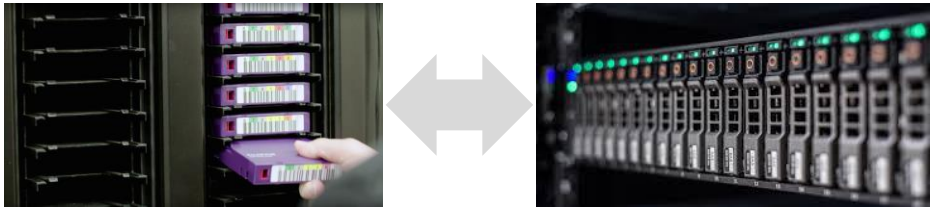
Collaboration Topics

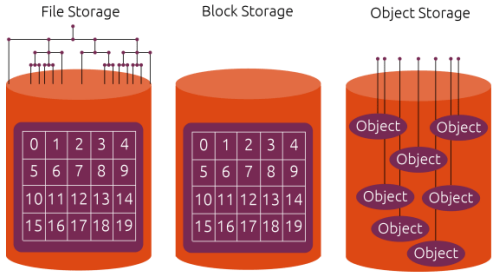
#1 Data Handling SW Development

➲ Disk Buffer Management System

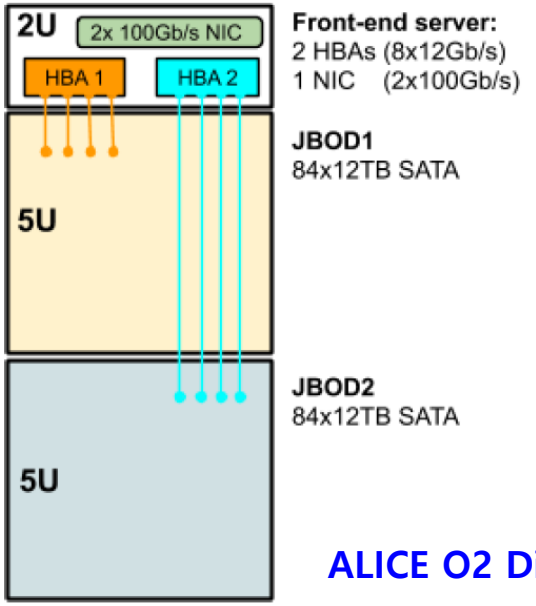


Different Storage Media





Different Data Storing Mechanism



Front-end server:
2 HBAs (8x12Gb/s)
1 NIC (2x100Gb/s)

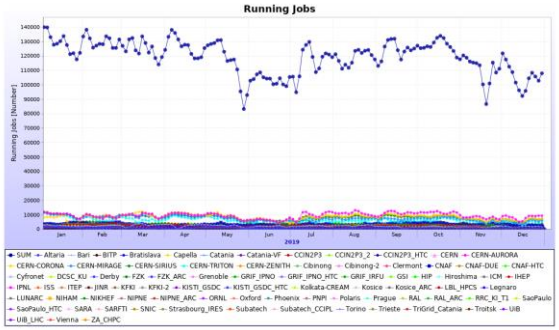
JBOD1
84x12TB SATA

JBOD2
84x12TB SATA

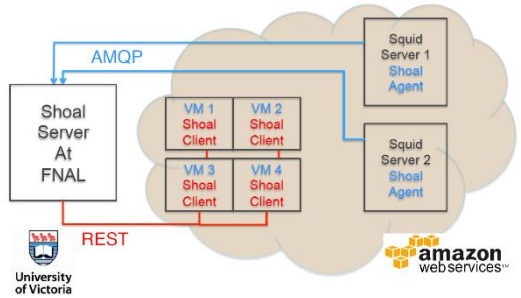
ALICE O2 Disk Buffer

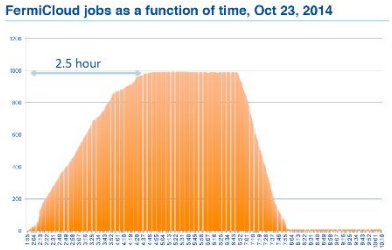
Collaboration Topics

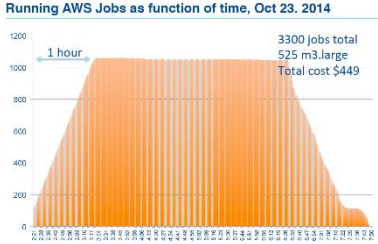
#2 Control & Monitoring SW



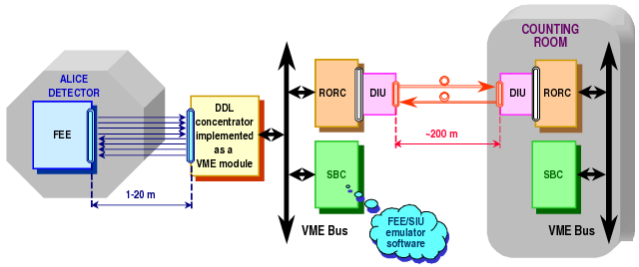
#3 Cloud Computing for Data Processing



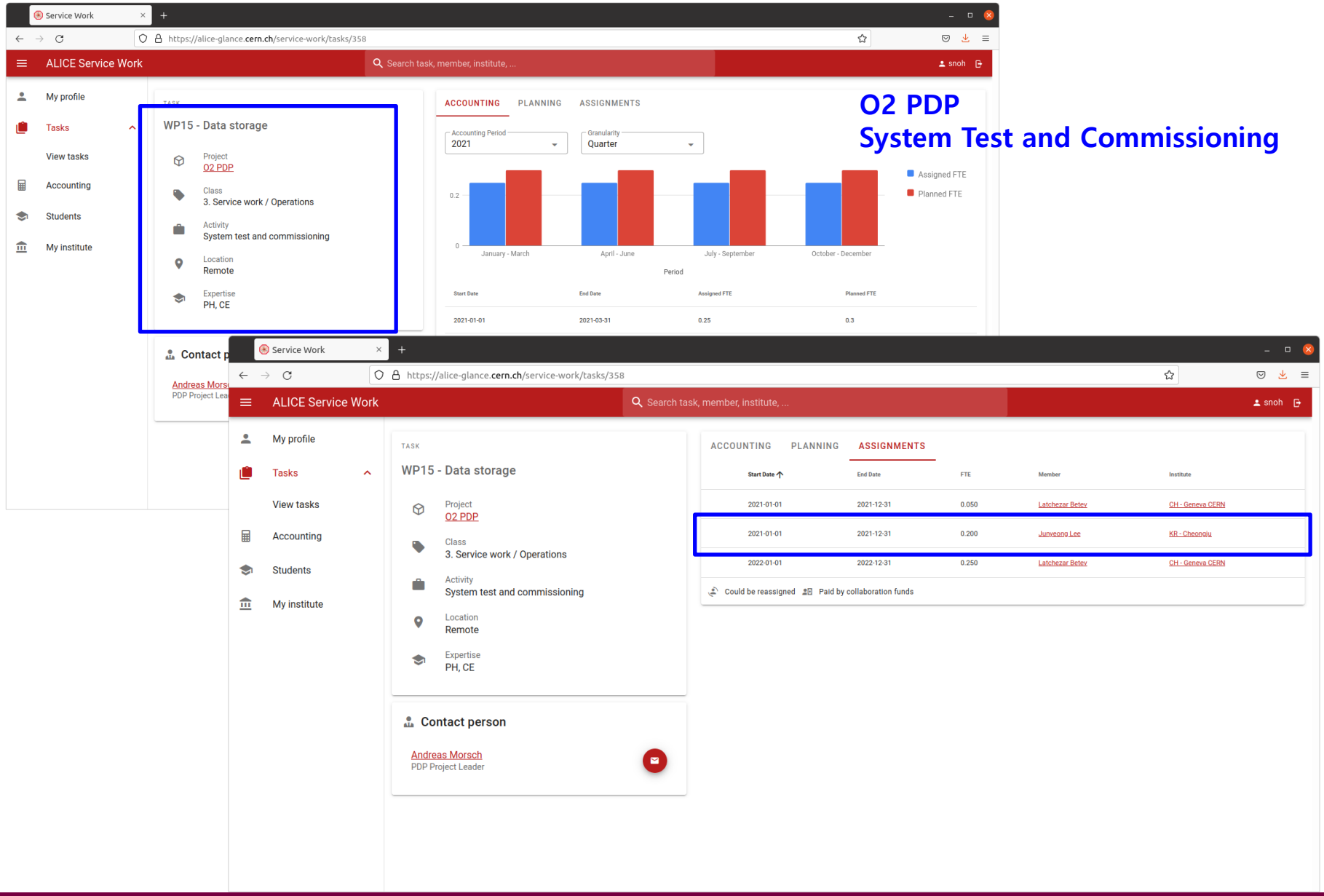




#4 CERN ALICE Detector System SW



#1 Data Handling – O2 PDP WP15 Data Storage

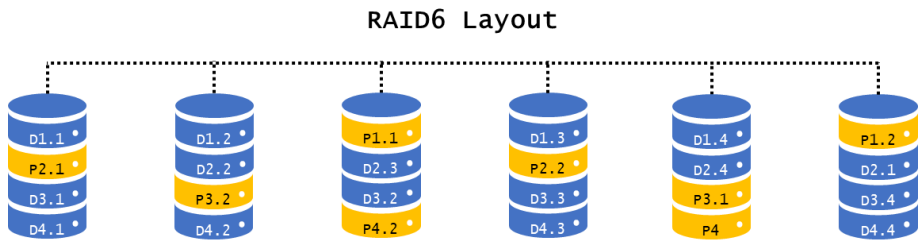


O2 PDP
System Test and Commissioning

#1 Data Handling – O2 PDP WP15 Data Storage

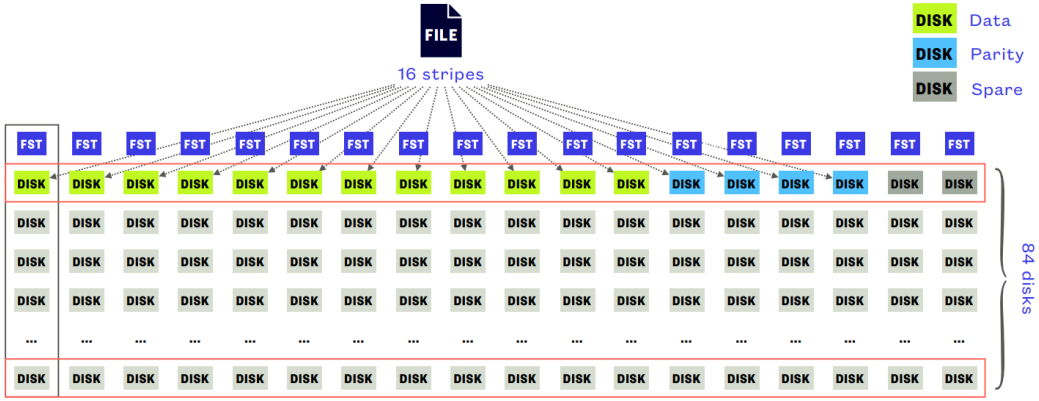
ALICE O2

- 10+2(RAID6 RAIN) Configuration
- 10 Data Disks
- 2 Parity Disks
- 16% Space Overhead



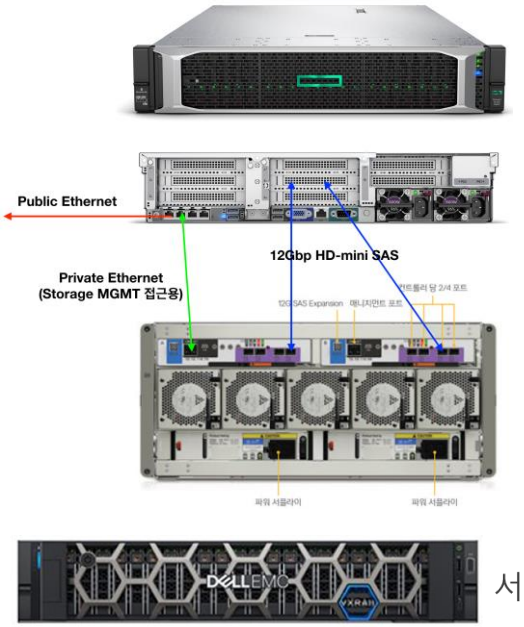
KISTI CDS (Custodial Disk Storage)

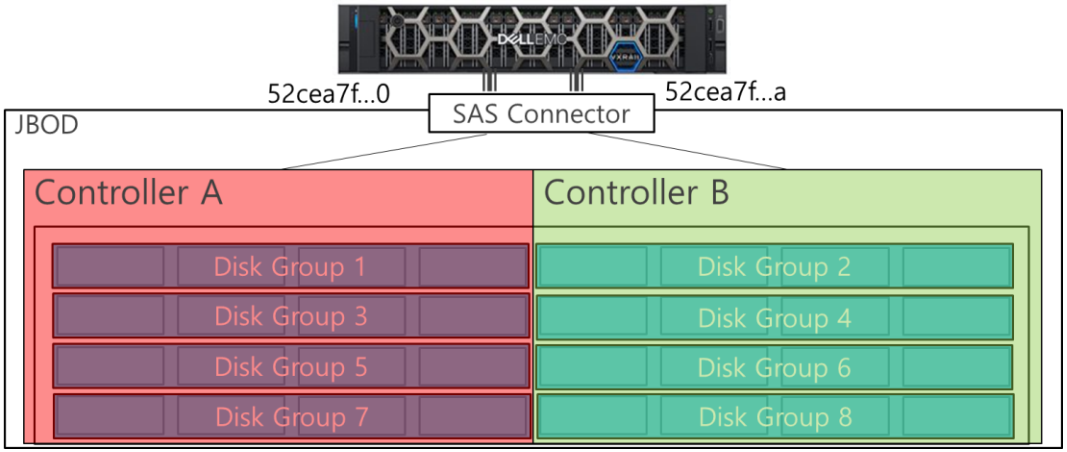
- 12+4(QRAIN) Configuration
- 12 Data Disks
- 4 Parity Disks
- Additional 2 Spare Disks
- 33% Space Overhead

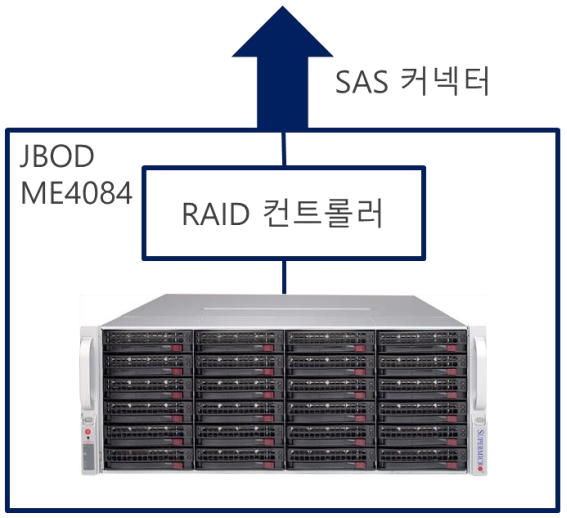


#1 Data Handling – O2 PDP WP15 Data Storage

- **Server – HP ProLiant DL560 G10**
  - Xeon Gold 6230 @ 20Core * 4 (80 Core)
  - DDR4 16GB * 48 (768GB)
  - 480GB SSD (Booting)
  - 12Gbps HBA
- **JBOD – Dell PowerVault ME4084**
  - 12TB * 70EA HDD
  - 12Gbps SAS 4 Port







#1 Data Handling – O2 PDP WP15 Data Storage

EOS Storage Setup and Evaluation

```
EOS Console [root://localhost] |/eos/> fs ls
```

host	port	id	path	schedgroup	geotag	boot	configstatus	drain	active
fst1.eos.docker	1095	1	/jbod1	default.0	docker::test	booted	rw	nodrain	online
fst2.eos.docker	1095	2	/jbod2	default.1	docker::test	booted	rw	nodrain	online
fst3.eos.docker	1095	3	/jbod3	default.2	docker::test	booted	rw	nodrain	online
fst4.eos.docker	1095	4	/jbod4	default.3	docker::test	booted	rw	nodrain	online
fst5.eos.docker	1095	5	/jbod5	default.4	docker::test	booted	rw	nodrain	online
fst6.eos.docker	1095	6	/jbod6	default.5	docker::test	booted	rw	nodrain	online
fst7.eos.docker	1095	7	/jbod7	default.6	docker::test	booted	rw	nodrain	online
fst8.eos.docker	1095	8	/jbod8	default.7	docker::test	booted	rw	nodrain	online

```
EOS Console [root://localhost] |/eos/> node ls
```

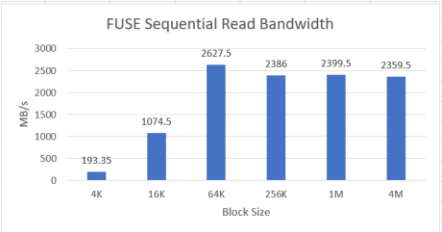
type	hostport	geotag	status	activated	txgw	gw-queued	gw-ntx	gw-rate	heartbeatdelta	nofs
nodesview	fst1.eos.docker:1095	docker::test	online	on	off	0	10	120	0	1
nodesview	fst2.eos.docker:1095	docker::test	online	on	off	0	10	120	0	1
nodesview	fst3.eos.docker:1095	docker::test	online	on	off	0	10	120	1	1
nodesview	fst4.eos.docker:1095	docker::test	online	on	off	0	10	120	0	1
nodesview	fst5.eos.docker:1095	docker::test	online	on	off	0	10	120	1	1
nodesview	fst6.eos.docker:1095	docker::test	online	on	off	0	10	120	0	1
nodesview	fst7.eos.docker:1095	docker::test	online	on	off	0	10	120	1	1
nodesview	fst8.eos.docker:1095	docker::test	online	on	off	0	10	120	1	1

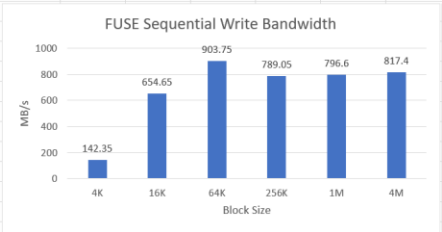
Performance Evaluations on JBOD based EOS System

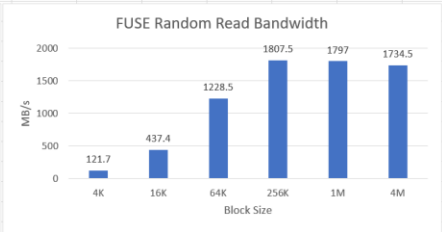
(ALICE O2 like) and KISTI CDS based EOS System

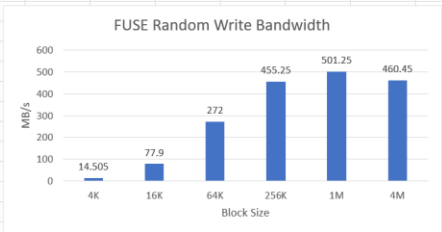
#1 Data Handling – O2 PDP WP15 Data Storage

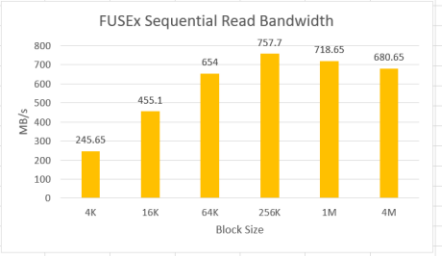
Performance Evaluations

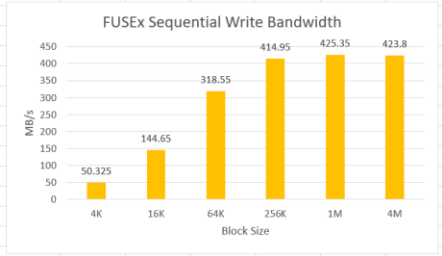


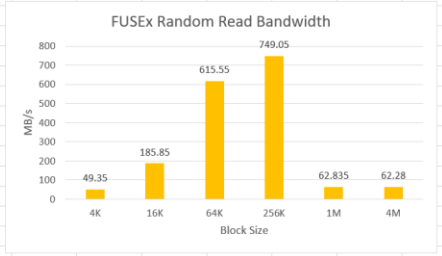


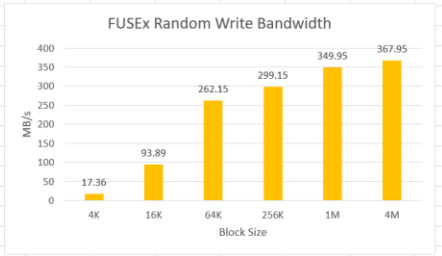




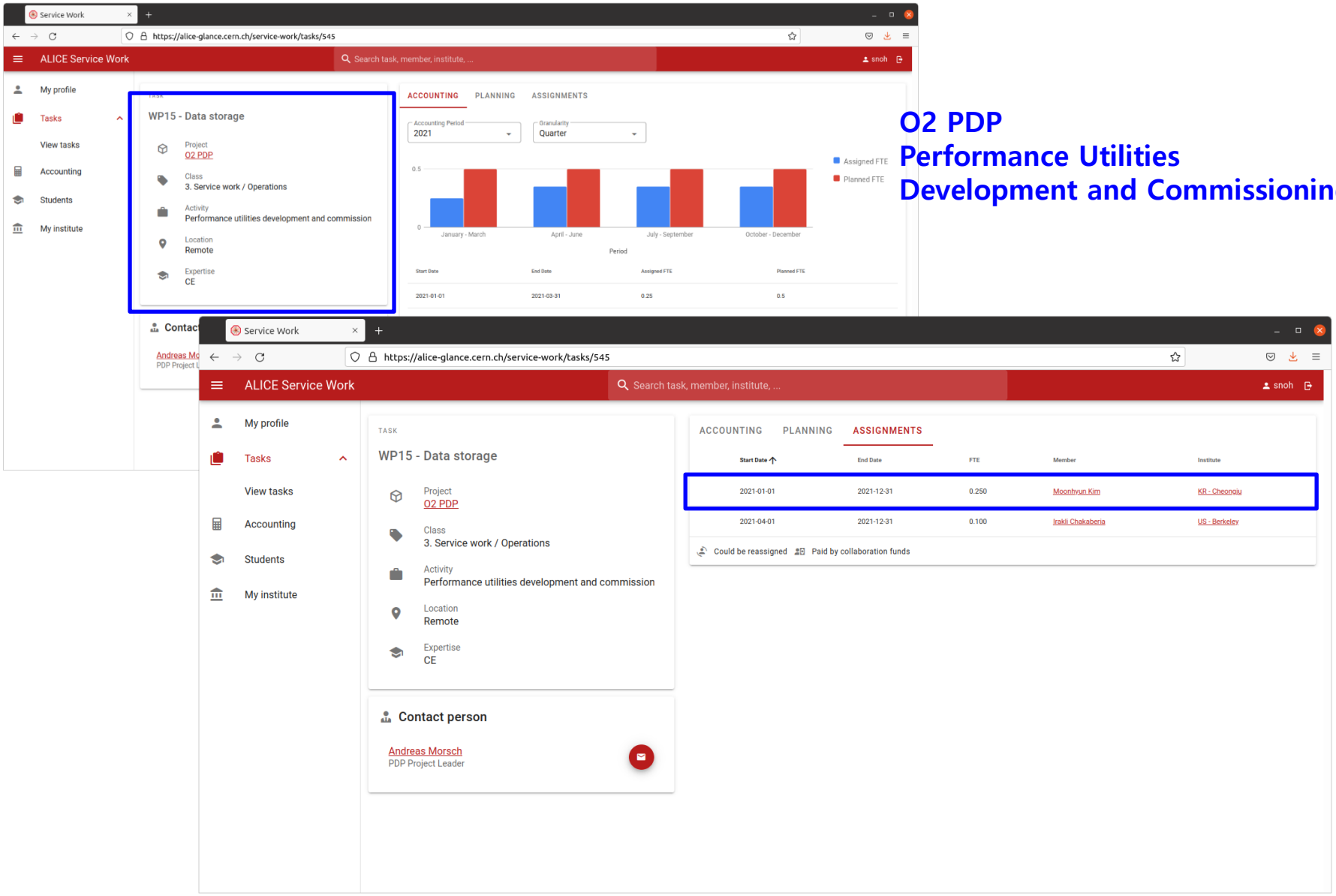








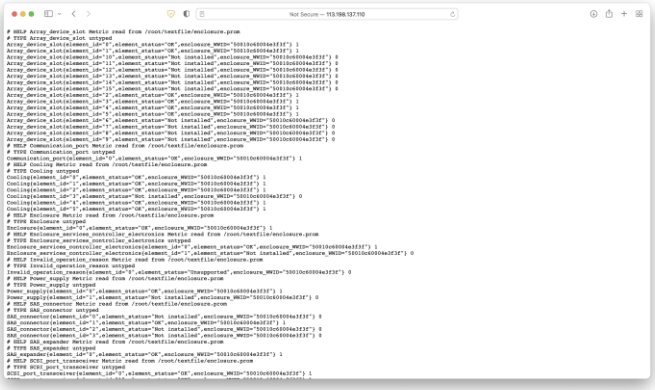
#2 Monitoring – O2 PDP WP15 Data Storage



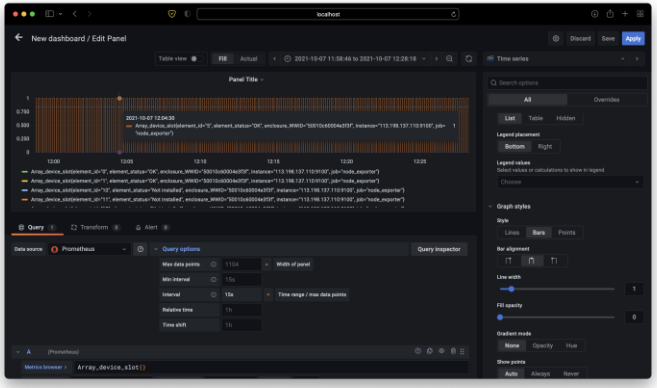
O2 PDP
Performance Utilities
Development and Commissioning

#2 Monitoring – O2 PDP WP15 Data Storage

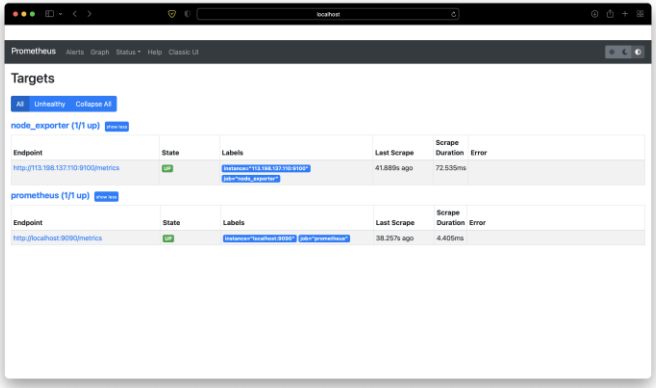
- **Developments**
  - Enclosure Error Reporting
  - Disk Error Reporting
  - EOS I/O Error Reporting
  - Enclosure Multipath Error Reporting



Node_exporter



Grafana

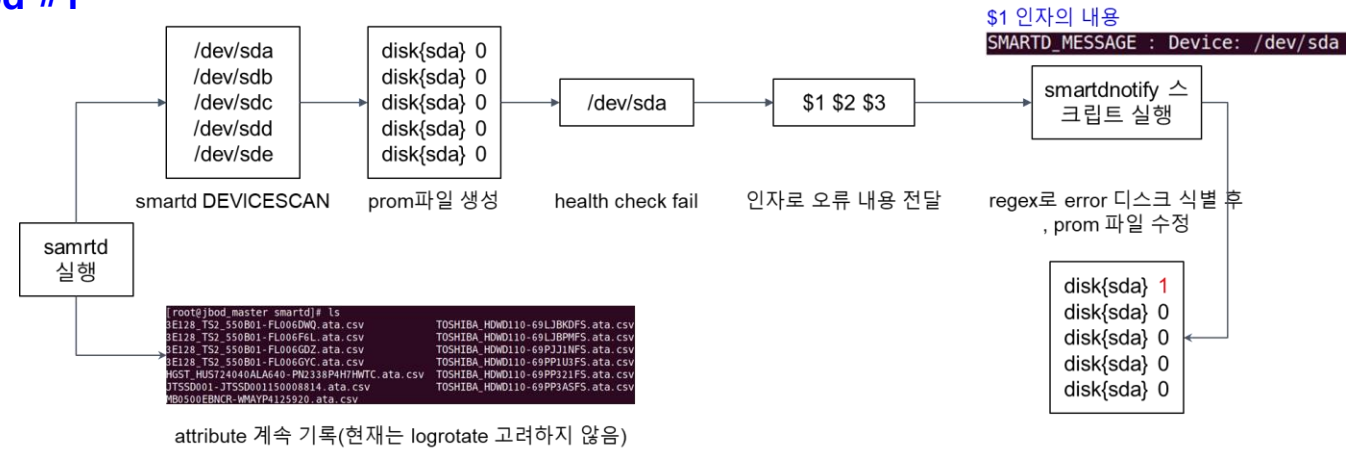


Prometheus

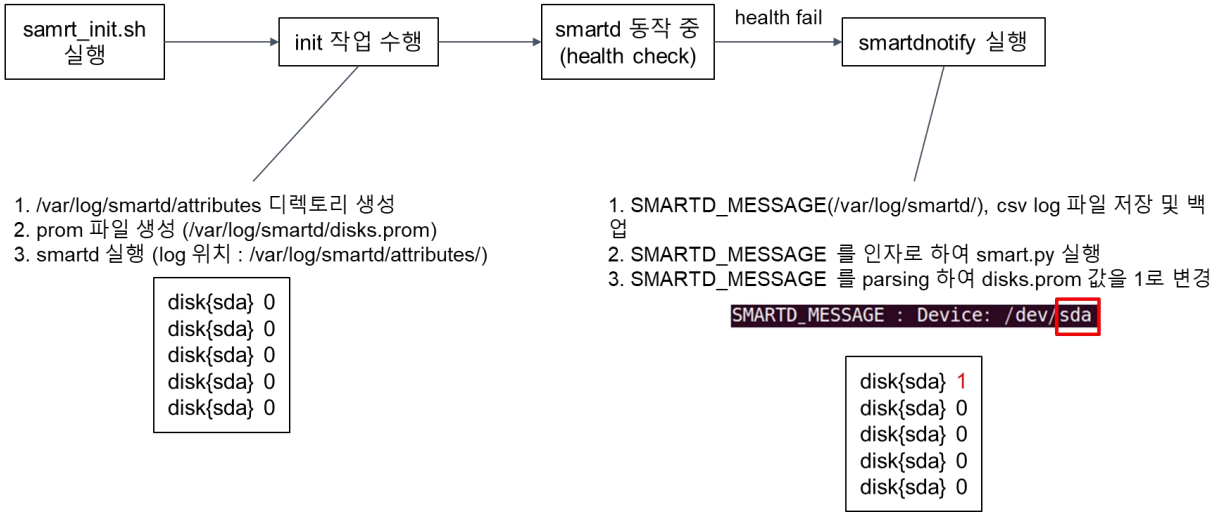
#2 Monitoring – O2 PDP WP15 Data Storage

Monitoring Workflow

Method #1



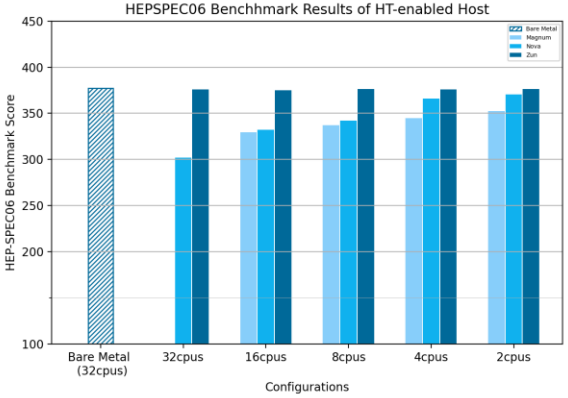
Method #2

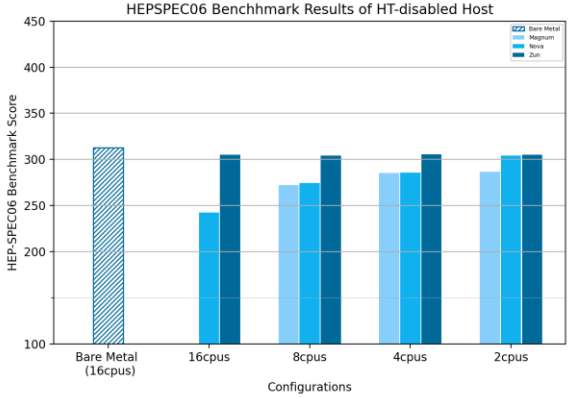


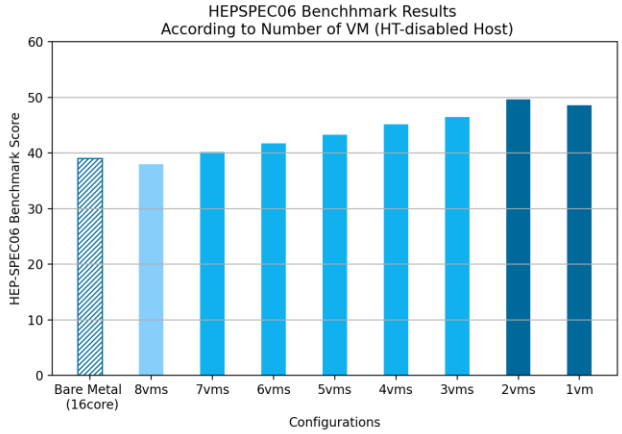
#3 Cloud Computing for Data Processing

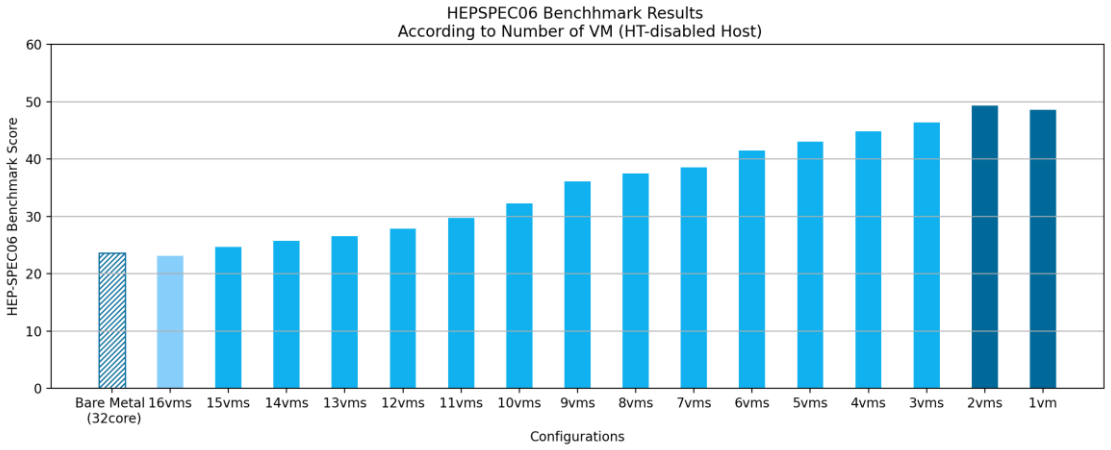
Efficient ALICE Workload Processing in Cloud Infrastructure

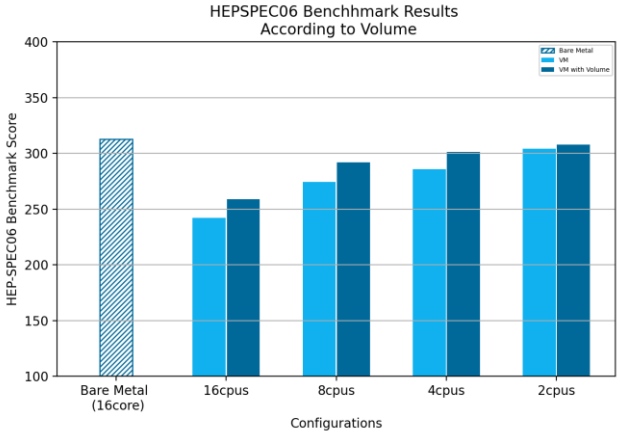
➲ HEPSPEC06 based Performance Benchmarks











Activities

Papers – Journal

electronics

MDPI

Article

Performance Evaluations of Distributed File Systems for Scientific Big Data in FUSE Environment

Jun-Yeong Lee [1], Moon-Hyun Kim [1], Syed Asif Raza Shah [2], Sang-Un Ahn [3], Heejun Yoon [3] and Seo-Young Noh [1,*]

[1] Department of Computer Science, Chungbuk National University, Cheongju-si 28644, Korea; lee1238234@cbnu.ac.kr (J.-Y.L.); moonhyunkim@cbnu.ac.kr (M.-H.K.) rsyoung@cbnu.ac.kr (S.-Y.N)
[2] Department of Computer Science and CRAIB, Sukkur IBA University (SIBAU), Sukkur 65200, Pakistan; asif.shah@iba-suk.edu.pk
[3] Global Science Experimental Data Hub Center, Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea; sahn@kisti.re.kr (S.-U.A.); k2@kisti.re.kr (H.Y.)
* Correspondence: rsyoung@cbnu.ac.kr

Abstract: Data are important and ever growing in data-intensive scientific environments. Such research data growth requires data storage systems that play pivotal roles in data management and analysis for scientific discoveries. Redundant Array of Independent Disks (RAID), a well-known storage technology combining multiple disks into a single large logical volume, has been widely used for the purpose of data redundancy and performance improvement. However, this requires RAID-capable hardware or software to build up a RAID-enabled disk array. In addition, it is difficult to scale up the RAID-based storage. In order to mitigate such a problem, many distributed file systems have been developed and are being actively used in various environments, especially in data-intensive computing facilities, where a tremendous amount of data have to be handled. In this study, we investigated and benchmarked various distributed file systems, such as Ceph, GlusterFS, Lustre and EOS for data-intensive environments. In our experiment, we configured the distributed file systems under a Reliable Array of Independent Nodes (RAIN) structure and a Filesystem in Userspace (FUSE) environment. Our results identify the characteristics of each file system that affect the read and write performance depending on the features of data, which have to be considered in data-intensive computing environments.

Keywords: data-intensive computing; distributed file system; RAIN; FUSE; Ceph; EOS; GlusterFS; Lustre

check for updates

Citation: Lee, J.-Y.; Kim, M.-H.; Raza Shah, S.A.; Ahn, S.-U.; Yoon, H.; Noh, S.-Y. Performance Evaluations of Distributed File Systems for Scientific Big Data in FUSE Environment. *Electronics* **2021**, *10*, 1471. https://doi.org/10.3390/electronics10121471

Academic Editor: Antonio F. Diaz

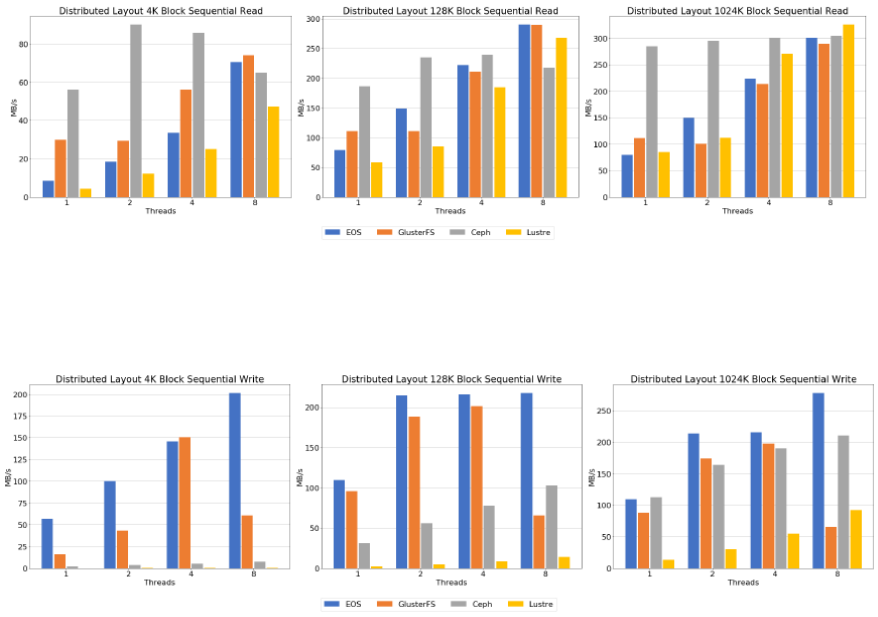
Received: 14 May 2021
Accepted: 15 June 2021
Published: 18 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

As the amount of computing data increases, the importance of data storage is emerging. Research from IDC and Seagate predicted that the size of the global data sphere was only a few ZB in 2010, but it would increase to 175 ZB by 2025 [1]. CERN, one of the largest physics research groups in the world, produces 125 petabytes of data per year from LHC experiments [2]. Due to the tremendous amount of experimental data produced, data storage is one of key factors in scientific computing. In such a computing environment, the capacity and stability of storage systems are important because the speed of data generation is high, and it is almost impossible to reproduce the data. Although there are many approaches to handling such big data, RAID has been commonly used to store large amounts of data because of its reliability and safety. However, RAID requires specific hardware and software to configure or modify storage systems. Moreover, it is difficult to expand with additional storage capacity. RAID is likely to affect the stability of the l loss. To overcome these drawbacks, many d and deployed at many computing facilitie distributed file system provides horizontal scalability compared to RAID, which uses

Published in Electronics
June 18, 2022 (SCIE)



Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2008-00458).

Papers – Conference

Korea Computer Congress 2021

대규모 클라우드 컴퓨팅 환경의 자원 효율성 증대를 위한 머신러닝 기반의 다중 관점 가상머신 배치 기법

김문현° 이준영 김경준 노서영
충북대학교 컴퓨터과학과
moonhyun.kim@cern.ch, junyeong.lee@cern.ch, rudwns273@chungbuk.ac.kr, rsyoung@chungbuk.ac.kr

Machine Learning-based Multi-Perspective Virtual Machine Placement for Large Cloud Computing Environments

Moon-Hyun Kim° Jun-Young Lee Kyeong-Jun Kim Seo-Young Noh
Dept of Computer Science, Chungbuk National University

요 약

데이터 센터에서 컴퓨팅 자원의 효율성을 높이는 것은 매우 중요한 문제이다. 대규모 클라우드 컴퓨팅 환경을 위한 데이터 센터에서는 가상화 기술을 통해 서비스를 제공하고, 실시간 스케줄링을 통해 가상머신을 스케줄링하여 컴퓨팅 자원의 효율성과 운영 효율성을 높이고 있다. 기존의 가상머신 스케줄링 방법들은 가상머신이 배치될 호스트의 CPU 사용량, RAM 사용량, 네트워크 대역폭 등 여러가지 지표들을 고려하고자 했다. 여러 지표들을 모두 고려하기 위해 각 리소스를 벡터로 수치화 하여 가상머신이 배치될 적합한 호스트를 찾는 기법들이 연구된 바 있다. 하지만 이러한 방법에서는 고려해야할 지표가 많을 수록 벡터의 차원이 증가하게 되어 가상머신 스케줄링에 필요한 연산양이 증가하게 된다. 이는 곧 가상머신의 배치 속도에 영향을 미치게 되어 전반적인 데이터 센터의 자원 효율성에 악영향을 미칠수 있다는 한계가 존재한다. 따라서 기존의 제시된 방법보다 보다 효율적인 가상머신 배치 방법이 필요하다. 본 논문에서는 가상머신 스케줄링 과정에서 다차원의 지표를 고려하면서도 연산속도를 증가시킬 수 있는 머신러닝 기반의 다중관점 가상머신 배치 방법을 제안한다. 본 논문에서 제시하는 연구 방법은 주성분 분석을 통해 고려해야 할 요소들의 차원을 효과적으로 줄이고, 머신러닝 기반의 클러스터링 기법을 활용한 가상머신 상태 판단 모델을 사용하여 가상머신의 배치 속도 또한 증가 시킬 수 있을 것으로 예상한다. 또한 이를 통해 궁극적으로 대규모 클라우드 컴퓨팅 환경에서의 자원 효율성을 극대화 할 수 있을 것으로 기대한다.

1 서론

데이터의 중요성이 증대되는 만큼 데이터센터의 역할은 커져가고 있다. 데이터센터는 컴퓨팅 자원의 활용성을 증대할 수 있는 능력, 여러 서비스를 격리하는 능력 등이 필요한데, 이러한 주요 이슈를 해결하기 위해 가상화를 기반으로 한 클라우드 데이터 센터가 주목 받고 있다. 가상화 기반의 데이터 센터는 리소스 활용성을 증대시키고, 가상머신을 독립적으로 관리하여 문제가 있는 가상 머신이 데이터 센터 전체에 영향을 주는 것을 막을 수 있다. 이러한 장점들 때문에 데이터 집약형 연구를 수행하는 CERN[1], FNAL[2] 등의 연구소에서도 클라우드 기반의 데이터 센터를 구축하여 연구에 활용하고 있다. 데이터 센터 내에서는 호스트 머신의 가용성을 최대한 높여 데이터 센터의 전반적인 컴퓨팅 자원의 운영 효율성을 높이는 것이 중요하다. 인프라를 관리하는 관점에서 초기의 가상머신 배치 기법은 데이터 센터의 전력 사용량에 가장 큰 영향을 미치는 호스트의 CPU 사용량을 기준으로 배치하는 방법들이 연구되었다[3]. 하지만 수행되는 작업의 특성에 따라 CPU 뿐만 아니라 RAM 사용량, 네트워크 대역폭 사용량, 디스크 사용량 등 고려해야 할 요인들이 상이했기 때문에, 여러 요인들을 복합적으로 고려하는 방법들 또한 기존에 활발하게 연구된 바 있다. 가상 머신을 배치하는데 있어 고려하는 요인들의 수가 증가 할수록 고차원의 연산을 필요하게 되는데, 이는 곧 처리해야할 연산양의 증가로 가상 머신 배치 속도에 영향을 미치게 된다. 이는 실시간으로 가상 머신의 배치 및 재배치를 수행하는 데이터 센터 내 스케줄링 과정에 있어서 연산의 비용을 증가시킬 뿐만 아니라, 작업을 수행하는 전체 호스트의 성능 저하를 일으킬 수 있다. 또한 대규모의 클라우드 컴퓨팅 환경일수록 가상 머신 배치에 있어서 고차원의 요인들을 고려할 때 연산량과 비용이 급증하게 된다. 따라서 본 논문에서는 기존의 고차원의 요인들을 고려하는 가상 머신 배치 방법에서 나아가 연산속도를 증가 시킬 수 있으며, 데이터 센터 내 호스트의 가용성을 증가시킬 수 있는 머신러닝 기반의 가상 머신 배치 기법을 제안한다. 이를 위해 먼저 고차원의 벡터를 저차원으로 줄일 수 있는 방안과 가상 머신의 상태를 빠르게 판단할 수 있는 기법을 제안한다. 본 논문에서 제시하는 방법을 통해 가상 머신 배치 과정에서 고차원의 요인들을 고려하면서도 연산 속도 또한 줄임으로써, 데이터 센터 내에서 가용성과 운영 효율성을 증가 시킬 수 있을 것으로 기대한다.

2 관련 연구

인프라를 관리하는 관점에서 가상머신을 배치할 경우 CPU 사용량이나 RAM 사용량, 네트워크 대역폭 등 호스트의 리소스 사용량을 기준으로 하는 연구가 기존에 활발하게 진행되었다. 호스트의 리소스 사용량을 가능한 한 최대화하기 위해 가상 머신을 배치하

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2008 -00458).

CERN EOS 분산 파일 시스템의 배포 환경에 따른 I/O 성능 비교

이준영°, 김문현, 김경준, 노서영
충북대학교 컴퓨터과학과
junyeong.lee@cern.ch, moonhyun.kim@cern.ch, rudwns273@cbnu.ac.kr, rsyoung@cbnu.ac.kr

I/O Performance Comparisons of CERN EOS Distribute File System under Bare-metal and KVM Deployment Environments

JunYeong Lee°, Moonhyun Kim, KyeongJun Kim, Seo-Young Noh
Department of Computer Science, Chungbuk National University

요 약

대용량의 데이터를 저장하기 위한 분산 파일 시스템은 대형 연구 시설과 대규모의 데이터센터와 같은 많은 양의 데이터를 다루는 곳에서 널리 사용되고 있다. 기존의 분산 파일 시스템은 베어메탈 방식으로 서버의 운영체제에 직접 분산 파일 시스템 소프트웨어를 설치하여 구성하였다. 하지만 최근에는 컨테이너 기반의 Docker, Kubernetes나 OpenStack과 같은 플랫폼을 기반으로 분산 파일 시스템을 가상화 방식을 통해 구성하여 제공하는 사례가 많아지고 있다. 본 논문에서는 세계 최대 규모의 대용량 실험데이터를 생산하는 CERN에서 사용되는 EOS 분산 파일 시스템을 베어메탈 형식과 Linux의 하이퍼바이저인 KVM으로 구현한 스토리지를 벤치마크하여 각 방식에 따른 성능을 측정하였다. 측정 결과를 토대로 구성 간의 성능 비교 및 특징에 대해 분석하였고, 이를 통해 분산 파일 시스템의 가상화에 따른 성능의 영향과 장점에 대하여 제시한다.

1. 서론

대규모의 데이터센터와 같이 거대한 용량의 데이터를 저장하기 위해 사용되는 분산 파일 시스템은 대용량의 데이터를 처리해야 하는 과학 연구 분야에서도 활발히 사용되고 있다. 고성능 컴퓨팅 시설과 연구 시설에서 많이 사용되는 Lustre 파일 시스템의 경우 Oak Ridge 대학이나 Los Alamos의 Cielo 슈퍼컴퓨팅 시설에서 초창기부터 사용되었다[1]. 유럽 입자 물리 연구소(CERN)에서는 LHC(Large Hadron Collider) 입자가속기에서 생산되는 대용량의 데이터를 저장하기 위해 EOS 분산 파일 시스템을 2010년부터 개발하여 현재까지 약 350페타바이트의 용량으로 구성되어 사용되고 있다[2]. 이와 같은 분산 파일 시스템을 구성하기 위해서 기존에는 하드웨어 상에 직접 소프트웨어를 설치하여 구성하는 베어메탈 방식을 사용하였지만, 최근에는 Docker나 OpenStack과 같은 플랫폼을 통해 분산 파일 시스템을 가상화하여 사용자에게 제공하고 있다. 본 논문에서는 CERN의 EOS 분산 파일 시스템을 작은 클러스터를 사용하여 베어메탈과 KVM을 이용한 가상화를 통해 소거 코딩(Erasure Coding)을 사용하는 파일 시스템을 구성하였다. 구성된 파일 시스템을 간단한 벤치마크를 통해 성능을 측정하고 각 구성에 따른 성능의 차이를 분석하였다. 결과를 통해 분산 파일 시스템을 가상화하였을 때의 장점과 가상화로 인해 발생하는 성능의 영향을 분석하였다.

2. 배경

2.1. EOS

CERN에서 2010년에 개발을 시작한 EOS는 LHC(Large Hadron Collider) 입자가속기에서 생산되는 대량의 데이터와 사용자가 생산하는 데이터를 저장하기 위해 개발된 오픈소스 분산 스토리지 솔루션이다[3]. 기존 스토리지에서 널리 사용되는 RAID(Redundant Array of Independant

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2008 -00458).


제21-1853호

최우수논문상

제 목 : CERN EOS 분산파일 시스템의 배포환경에 따른 I/O 성능비교

저 자 : 이준영 김문현 김경준 노서영(충북대학교)

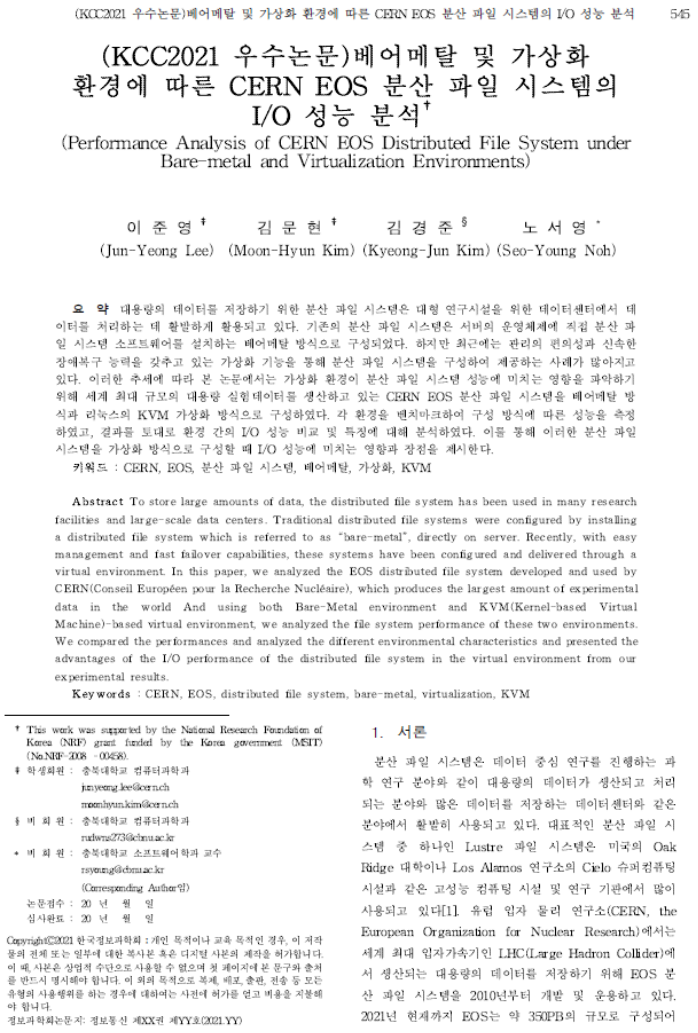
위 논문은 2021 한국컴퓨터종합학술대회에서 컴퓨터시스템 분야의 최우수논문으로 선정되었기에 이 상을 수여함.

2021년 6월 24일

한국정보과학회 회장 나 연 묵

748 papers published, only 2% selected as best papers

Papers – Under Preparation



(KCC2021 우수논문)베어메탈 및 가상화 환경에 따른 CERN EOS 분산 파일 시스템의 I/O 성능 분석 545

(KCC2021 우수논문)베어메탈 및 가상화 환경에 따른 CERN EOS 분산 파일 시스템의 I/O 성능 분석†

(Performance Analysis of CERN EOS Distributed File System under Bare-metal and Virtualization Environments)

이 준 영 † 김 문 현 † 김 경 준 § 노 서 영 *

(Jun-Yeong Lee) (Moon-Hyun Kim) (Kyeong-Jun Kim) (Seo-Young Noh)

요 약 대용량의 데이터를 저장하기 위한 분산 파일 시스템은 대형 연구시설을 위한 데이터센터에서 데이터를 처리하는 데 활발하게 활용되고 있다. 기존의 분산 파일 시스템은 서버의 운영체제에 직접 분산 파일 시스템 소프트웨어를 설치하는 베어메탈 방식으로 구성되었다. 하지만 최근에는 관리의 편의성과 신속한 장애복구 능력을 갖추고 있는 가상화 기능을 통해 분산 파일 시스템을 구성하여 제공하는 사례가 많아지고 있다. 이러한 추세에 따라 본 논문에서는 가상화 환경이 분산 파일 시스템 성능에 미치는 영향을 파악하기 위해 세계 최대 규모의 대용량 실험데이터를 생산하고 있는 CERN EOS 분산 파일 시스템을 베어메탈 방식과 리눅스의 KVM 가상화 방식으로 구성하였다. 각 환경을 벤치마크하여 구성 방식에 따른 성능을 측정하였고, 결과를 토대로 환경 간의 I/O 성능 비교 및 특징에 대해 분석하였다. 이를 통해 이러한 분산 파일 시스템을 가상화 방식으로 구성할 때 I/O 성능에 미치는 영향과 장점을 제시한다.

키워드 : CERN, EOS, 분산 파일 시스템, 베어메탈, 가상화, KVM

Abstract To store large amounts of data, the distributed file system has been used in many research facilities and large-scale data centers. Traditional distributed file systems were configured by installing a distributed file system which is referred to as "bare-metal", directly on server. Recently, with easy management and fast failover capabilities, these systems have been configured and delivered through a virtual environment. In this paper, we analyzed the EOS distributed file system developed and used by CERN(Conseil Européen pour la Recherche Nucléaire), which produces the largest amount of experimental data in the world And using both Bare-Metal environment and KVM(Kernel-based Virtual Machine)-based virtual environment, we analyzed the file system performance of these two environments. We compared the performances and analyzed the different environmental characteristics and presented the advantages of the I/O performance of the distributed file system in the virtual environment from our experimental results.

Keywords : CERN, EOS, distributed file system, bare-metal, virtualization, KVM

† This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2008 -00458).

학생회원 : 충북대학교 컴퓨터과학과
jun.yeong.lee@cern.ch
moonhyun.kim@cern.ch

§ 비 회 원 : 충북대학교 컴퓨터과학과
rudwns273@chnu.ac.kr

* 비 회 원 : 충북대학교 소프트웨어학과 교수
rsyoung@chnu.ac.kr
(Corresponding Author임)

논문접수 : 20 년 월 일
심사완료 : 20 년 월 일

정보과학회논문지: 정보통신 제XX권 제YY호(2021.YY)

1. 서론

분산 파일 시스템은 데이터 중심 연구를 진행하는 과학 연구 분야와 같이 대용량의 데이터가 생산되고 처리되는 분야와 많은 데이터를 저장하는 데이터센터와 같은 분야에서 활발히 사용되고 있다. 대표적인 분산 파일 시스템 중 하나인 Lustre 파일 시스템은 미국의 Oak Ridge 대학이나 Los Alamos 연구소의 Cielo 슈퍼컴퓨팅 시설과 같은 고성능 컴퓨팅 시설 및 연구 기관에서 많이 사용되고 있다[1]. 유럽 입자 물리 연구소(CERN, the European Organization for Nuclear Research)에서는 세계 최대 입자가속기인 LHC(Large Hadron Collider)에서 생산되는 대용량의 데이터를 저장하기 위해 EOS 분산 파일 시스템을 2010년부터 개발 및 운용하고 있다. 2021년 현재까지 EOS는 약 350PB의 규모로 구성되어

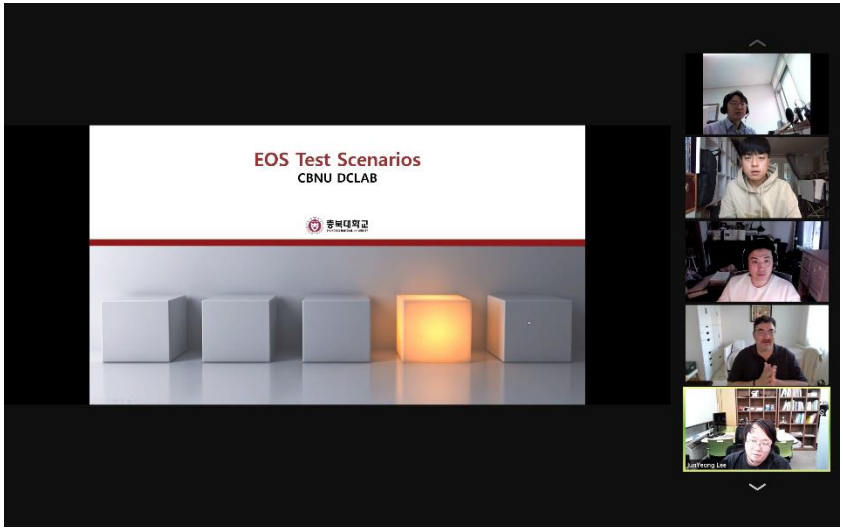
- **Invited Paper**
  will be published in March

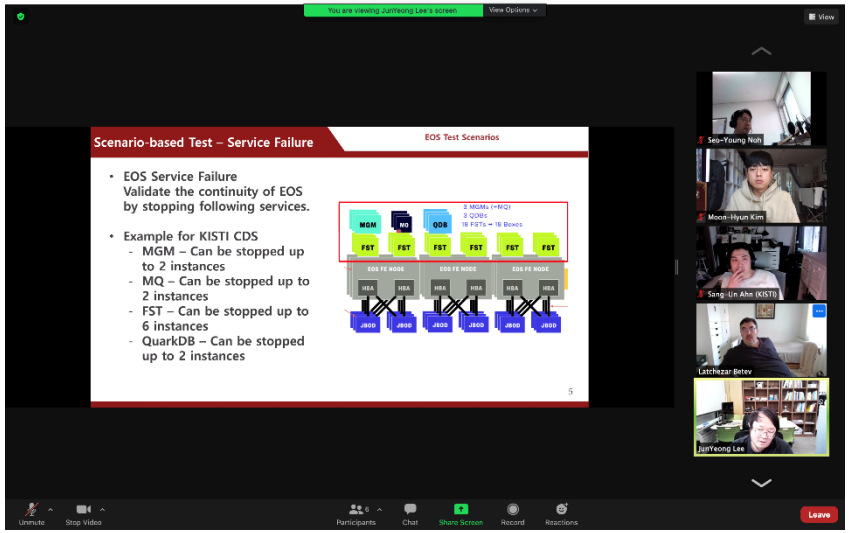
- **SCIE Journal #1 (Jun-Yeong Lee)**
  Analysis of CERN EOS Storage under Physical and Virtualization System

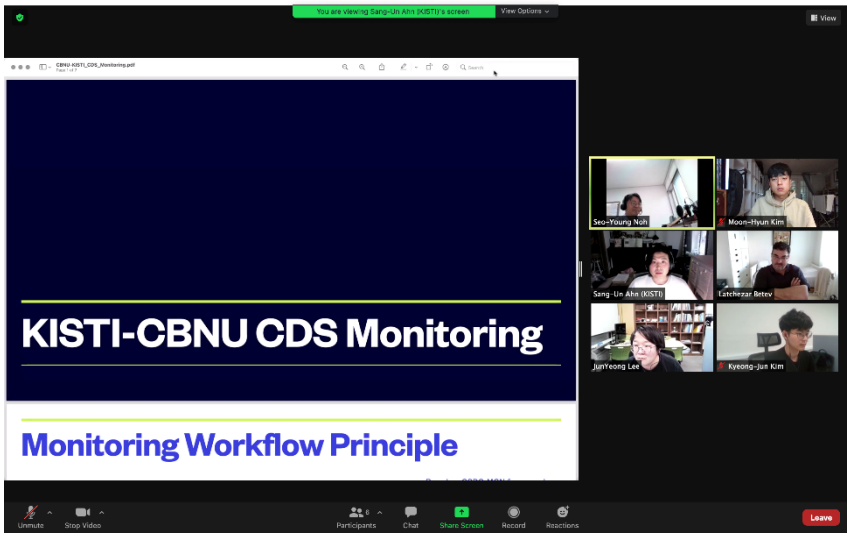
- **SCIE Journal #2 (Moon-Hyun Kim)**
  Benchmarking and Performance Evaluations in OpenStack for Cloud-based Scientific Workloads

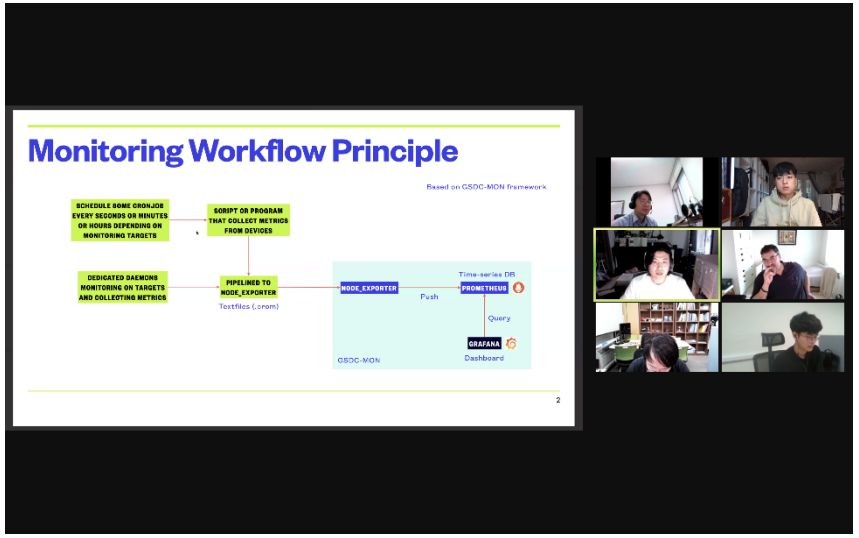
Meetings

CERN-KISTI-CBNU Meeting



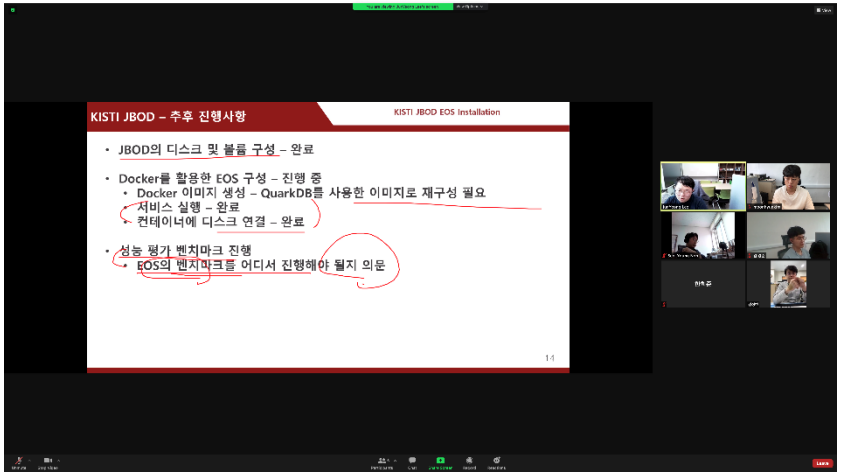


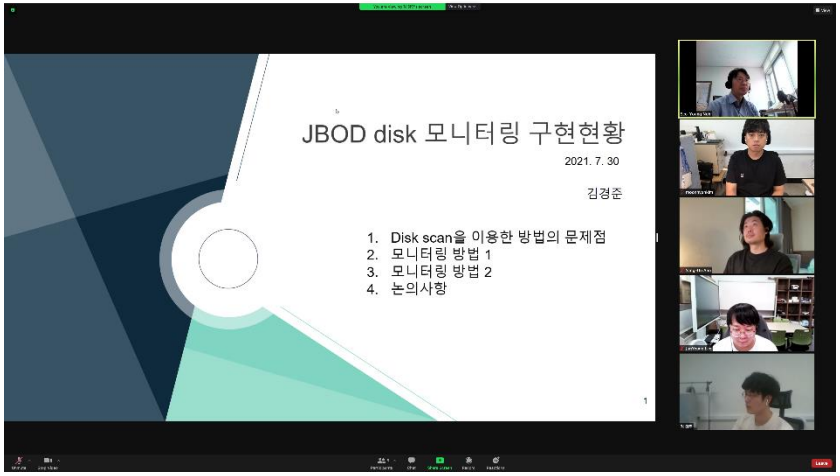


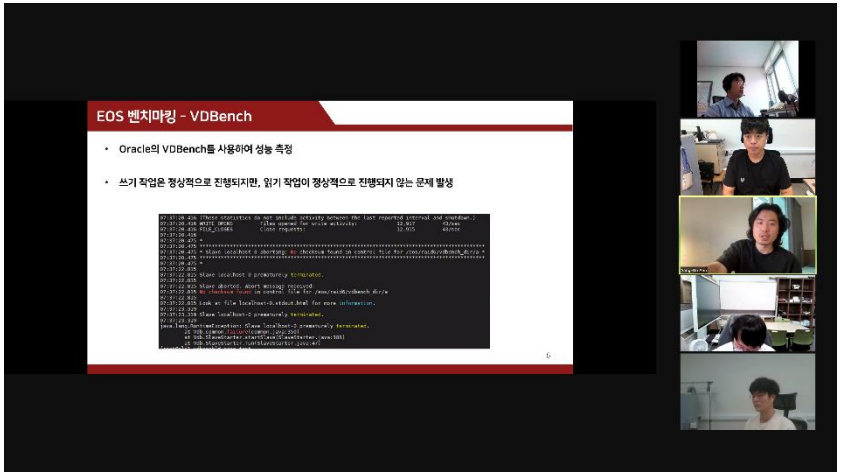


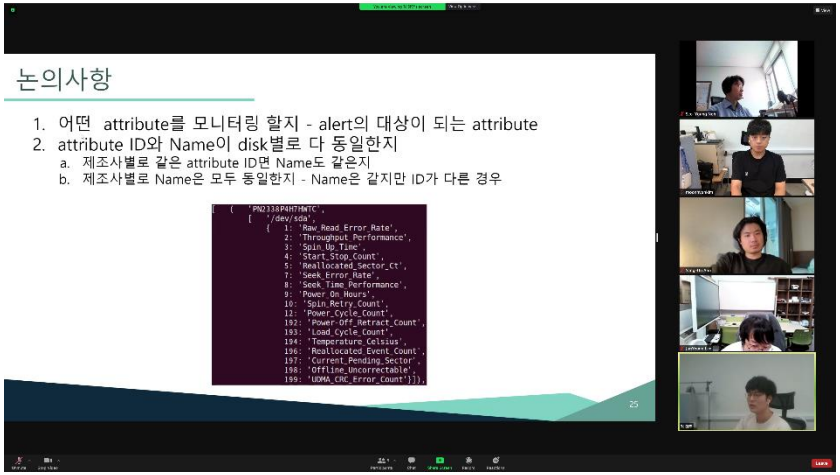
Meetings

KISTI-CBNU Meeting









Plan

Plan

- **Keep Focusing on Current Projects**
  - Data Storage Systems: O2 PDP WP15
  - Monitoring Utilities Development
  - Computing Resource Optimization
- **Involving in Core Software Developments**
  - ITS3 Detector System Software
  - EOS Storage System Software
- **Keep Collaboration with KISTI**

Thank You.