

# KI Cloud R&D 플랫폼을 위한 데이터 플랫폼 설계 및 구현

조혜영<sup>°</sup>, 박준영, 손아영, 정기문  
한국과학기술정보연구원 슈퍼컴퓨팅본부  
{chohy<sup>°</sup>, jypark, ayson, kmjeong}@kisti.re.kr

## Design and Implementation of the Data Platform for KI Cloud

Hyeyoung Cho<sup>°</sup>, Junyoung Park, A-young Son, Gi-Mun Jeong  
Division of National Supercomputing  
Korea Institute of Science and Technology Information

### 요약

최근 인공지능(AI) 및 빅데이터 기술의 비약적인 발전에 따라, 고성능컴퓨팅(High Performance Computing, HPC) 환경에 대한 요구는 점차 고도화되고 다변화되고 있다. 이에 따라 기존의 베어메탈 기반 인프라에서 벗어나, 사용자 맞춤형 자원 할당이 가능한 클라우드 기반 HPC 환경의 필요성이 증가하고 있다. 한국과학기술정보연구원(KISTI)은 이러한 요구를 반영하여 HPC 사용자 대상 R&D 플랫폼인 KI Cloud를 개발하였다. 본 논문에서는 KI Cloud 상에서 다양한 HPC 워크로드를 효과적으로 지원하기 위한 데이터 플랫폼의 설계 및 구현 방안을 제안한다. 특히, 멀티 클러스터 환경의 스토리지 유형을 고려한 통합 데이터 처리 구조를 중심으로 데이터 플랫폼 아키텍처를 설계하였다.

### 1. 서론

최근 인공지능(AI)과 빅데이터 기술의 급격한 발전으로 고성능 컴퓨팅(HPC) 수요가 다양해지면서, 기존의 전통적인 베어메탈 시스템 및 스케줄러 기반 인프라 구조에서 벗어나 사용자 맞춤형 클라우드 환경에 대한 필요성이 대두되고 있다. 특히 HPC 사용자들이 단일 소프트웨어 환경을 공유하는 방식의 한계를 극복하고, 유연한 인프라 자원 할당과 다양한 서비스 유형을 요구하면서, HPC 영역에서도 클라우드에 대한 관심이 높아지고 있다. 이에 한국과학기술정보연구원(KISTI)에서는 고성능 컴퓨팅(HPC) 사용자를 위한 R&D 플랫폼인 KI Cloud를 개발하여 서비스하고 있다[1,2].

이러한 통합 플랫폼에서 고성능 HPC를 성공적 지원하기 위해서는 멀티 클러스터 환경에서 다양한 스토리지 유형을 지원하는 데이터 플랫폼의 설계가 핵심 기술이다. 본 논문에서는 HPC R&D 플랫폼인 KI Cloud를 위한 데이터 플랫폼 설계하고, 구현 결과를 소개한다. 본 논문의 구성은 다음과 같다. 2장에서는 HPC 클라우드 플랫폼 KI Cloud에 대해 기술한다. 3장에서는 KI Cloud R&D 플랫폼을 위한

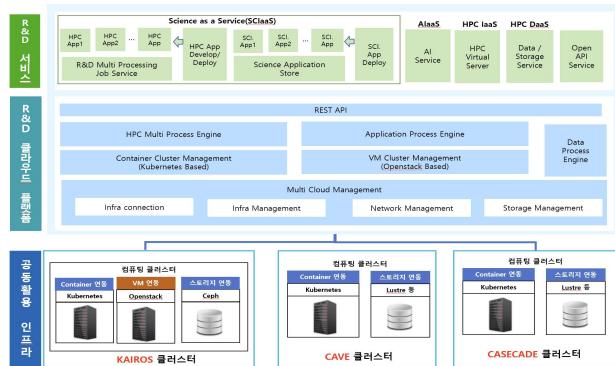
데이터 플랫폼 설계 및 구현을 기술하고, 4장에서는 결론 및 향후 계획을 기술한다.

### 2. KI Cloud 플랫폼

KI Cloud는 KISTI가 과학기술 분야의 고성능컴퓨팅(HPC) 및 인공지능 관련 연구를 지원하기 위해 자체적으로 구축한 클라우드 플랫폼이다[1]. KISTI의 슈퍼컴퓨터 사용자뿐만 아니라 국내 연구자들에게 HPC 기반의 IaaS(Infrastructure as a Service)와 PaaS 서비스를 제공하고 있다. 본 플랫폼은 웹 기반 포털을 통해 클라우드 환경에서 기계학습 및 데이터 분석 등의 연구가 원활하게 수행될 수 있는 인프라를 제공한다.

KI Cloud는 오픈소스 기반의 Openstack[3]과 Kubernetes[4] 기술을 핵심 구성요소로 설계하고 개발하였다. Openstack을 통해 사용자는 맞춤형 가상서버, 스토리지 등 IaaS 서비스를 이용할 수 있으며, Kubernetes 기반의 컨테이너 환경에서는 Jupyter Notebook과 같은 대화형 프로그래밍 툴뿐 아니라 분산형 프로그래밍이 가능한 PaaS 기능도 지원된다. 플랫폼의 전체 아키텍처 및 서비스 구성은 각각 [그림 1]에 나타내었다. KI Cloud는 멀티 클러스터 자

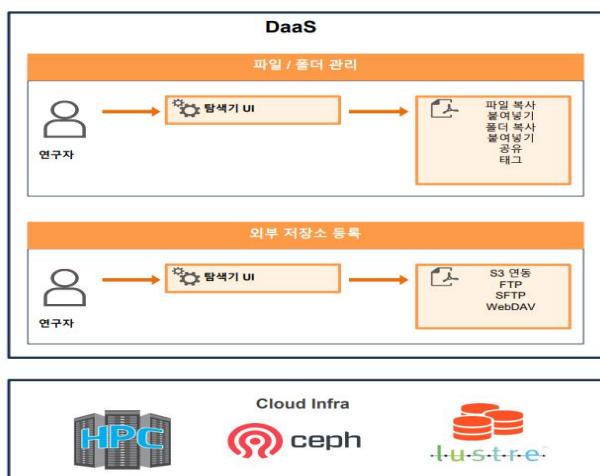
원을 제공한다. KISIT 내 자원인 KAIROS, CAVE, CASECADE 클러스터들을 연동 및 관리하는 모듈을 구성하였다. Kubernetes 기반의 멀티 클러스터 자원 관리가 가능하다.



[그림 1] KI Cloud 플랫폼 구조

### 3. 데이터 플랫폼 설계 및 구현

DaaS(Data as a Service) 서비스는 KI Cloud를 위한 데이터 서비스로, 원격 서버나 클라우드 스토리지의 파일을 접근하고, 다운로드 할 수 있는 사용자 웹 인터페이스를 제공한다. 또한 여러 외부 스토리지와 연동하는 기능을 제공한다. [그림 2]에 DaaS 구조도를 나타내었고 [표 1]에 외부저장소 연결 정보를 표시하였다.



[그림 2] DaaS 구조도

DaaS 주요 기능은 다음과 같다.

- 파일 / 폴더 관리
- 파일이나 폴더 관리 기능 제공
- 일반적인 파일 및 폴더의 복사, 붙여넣기 기능과 즐겨찾기, 태그, 공유 기능 제공
- SCIAaaS에서 활용하는 장소의 사용자 인터페이스

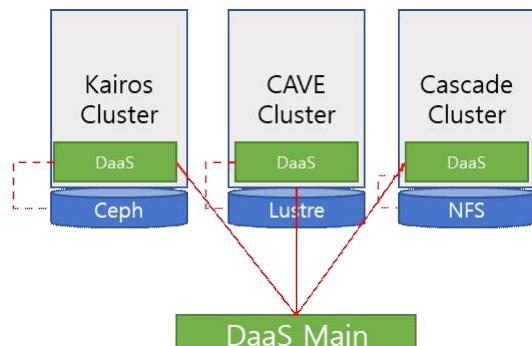
를 제공

- 멀티 클러스터들의 모두 별개의 저장소이며, 자원에 배포되는 클러스터의 저장소에 PV를 사용
- DaaS를 통해 통합 관리 기능 제공
- 외부 저장소 등록
  - 외부 저장소를 등록(마운트) 하기 위하여 S3, FTP, SFTP, WebDAV(https) 연동 인터페이스 방식을 제공

[표 1] 외부 저장소 연결 설정 정보

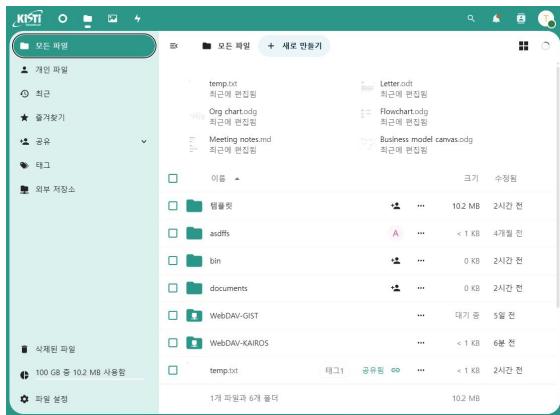
외부 저장소	연결 설정 정보
S3	<ul style="list-style-type: none"> <li>• 버킷</li> <li>• 호스트 명(EndPoint)</li> <li>• 접근키: Access Key ID</li> <li>• 비밀키: Secret Key ID</li> </ul>
FTP / SFTP	<ul style="list-style-type: none"> <li>• 호스트 IP, Port</li> <li>• 사용자 접속 정보</li> <li>• 키파일</li> </ul>
WebDAV	<ul style="list-style-type: none"> <li>• Https</li> </ul>

KI Cloud에는 멀티 클러스터 환경을 제공하며 스토리지 네트워크 연동하여 구성하였다. [그림 3]에 멀티 클러스터에서 DaaS를 통한 데이터 관리 구조를 나타내었다. KI Cloud에는 Kubernetes기반 PaaS 서비스인 SCIAaaS(SCience as a Service)를 통해, 과학분야에서 활용되는 소프트웨어 중심으로 클라우드 기술을 활용한 R&D 서비스를 제공한다. SCIAaaS에서 활용하는 KAIROS, CAVE, CASECADE 클러스터의 스토리지는 모두 별개의 저장소이며 KAIROS는 Ceph[5], CAVE는 Lustre[6,7], CASECADE는 NFS를 사용한다. SCIAaaS에서는 자원에 배포되는 클러스터의 저장소에 PV(Persistent Volumes)를 사용한다. 이러한 스토리지들은 DaaS를 통해 통합 관리 가능하다.



[그림 3] 멀티 클러스터에서 DaaS를 통한 데이터 관리 구조도

[그림 4]에 KI Cloud 데이터 플랫폼인 DaaS의 구현화면을 나타내었다. 그림의 디렉토리 중 WebDAV-KAIROS는 KAIROS 클러스터의 PV를 연결을 나타낸다.



[그림 4] DaaS 구현 화면

#### 4. 결론 및 향후 계획

KI Cloud는 과학기술 분야의 고성능컴퓨팅(HPC) 및 데이터 분석 연구를 지원하기 위해 KISTI가 개발한 R&D 플랫폼이다. 본 논문에서는 KI Cloud 플랫폼에서 다양한 HPC 워크로드를 효과적으로 지원하기 위한 데이터 플랫폼의 설계 및 구현 방안을 제안한다. 특히, 멀티 클러스터 환경에서 다양한 스토리지 유형을 효율적으로 통합·관리하기 위한 데이터 플랫폼 설계 및 구현 방안을 제시하였다. DaaS는 S3, FTP, SFTP, WebDAV 등을 활용하여 다양한 외부 저장소와의 연동을 지원하고, KAIROS, CAVE, CASCADE의 멀티 클러스터 환경에서도 통합된 데이터 관리 인터페이스를 제공함으로써 클라우드 환경에서의 데이터 통합 서비스 운용 가능성을 입증하였다.

향후 연구에서는 DaaS의 데이터 접근 성능 향상과 보안성 강화를 위한 기능 고도화를 진행하고, 다양한 과학기술 분야의 응용 요구에 대응할 수 있도록 서비스 확장성을 중심으로 한 플랫폼 고도화 전략을 마련할 계획이다.

#### 참고문헌

- [1] KI Cloud (KISTI Intelligent Cloud), <https://home.kicloud.kisti.re.kr>
- [2] Hyeyoung Cho, et al, "Design and Development of KI Cloud Platform for High Performance Computing," KIPS fall conference, 2020.
- [3] Openstack, <https://www.openstack.org>
- [4] kubernetes, <https://kubernetes.io/docs/home>
- [5] ceph, <https://ceph.io>
- [6] lustre, <https://www.lustre.org/>
- [7] Braam, P. J., & Zahir, R. (2002). Lustre: A scalable, high performance file system. Cluster File Systems, Inc.

#### ACKNOWLEDGMENT

본 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구임.(과제번호: K25L1M2C2-01, 수요자 맞춤형 연구환경 제공을 위한 초고성능컴퓨팅 기술개발)