CERN 중심 대용량 실험 데이터센터 인프라 동향

박형우 이형진 여밀연 윤희준

KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

<제목 차례>

1. 서론 ····· **1**

가. 조사의 배경 ····· 1

나. 데이터의 대용량화에 따른 조사 주안점 ····· 1

다. 조사 범위 ····· 2

2. 인프라 중심 WLCG Tier1 센터 동향 ····· **3**

가. 스위스 CERN 데이터 센터 ····· 3

나. 독일 KIT-Gridka 데이터 센터 ····· 13

다. 영국 RAL 데이터 센터 ····· 18

라. 미국 페르미 랩 데이터 센터 ····· 24

마. 일본 KEK 데이터 센터 ····· 29

3. 미국 NSF 및 OECD 활동 동향 ····· **34**

가. 미국 NSF 지역별 빅데이터허브구축사업 (2단계) 동향 ····· 34

나. NSF 빅데이터 인프라 구축 지원 사업 동향 ····· 36

다. OECD 빅데이터 정책 동향 ····· 39

4. 국내동향 ····· **51**

가. 데이터 댐 사업 동향 ····· 51

나. 박데이터 활용 지원 사업 동향 ····· 54

다. 국내 데이터 산업 동향 ····· 58

5. 결언 ····· **62**

가. 데이터 센터 인프라 동향 특징 ····· 62

나. 미국 인프라 지원 사업의 동향 특징 ····· 62

다. 스토리지 산업체의 기술 동향 특징 ····· 63
라. GSDC 인프라 구축 방향에 대한 고려 사항 ····· 63
6. 참고문헌 ····· 65
7. 부록 1: 데이터센터 인프라 기술 약어집(알파벳 순) ····· 67
8. 부록 2: 데이터 스토리지 산업체 기술 동향 ····· 73

<표 차례>

표 1 CERN 인프라 동향 키워드 ····· 3
표 2 독일 KIT 인프라 동향 키워드 ····· 13
표 3 영국 RAL 인프라 동향 키워드 ····· 18
표 4 미국 페르미랩 인프라 동향 키워드 ····· 24
표 5 일본 KEK 인프라 동향 키워드 ····· 29
표 6 미국 NSF 지역별 빅데이터허브구축사업 동향 키워드 ····· 34
표 7 미국 NSF 빅데이터 인프라 구축 지원 사업 동향 키워드 ····· 36
표 8 OECD 빅데이터 정책 동향 키워드 ····· 39
표 9 OECD 고잉 디지털 모듈별 관련 부서 ····· 41
표 10 데이터 댐 사업 동향 키워드 ····· 51
표 11 빅데이터 활용 지원 사업 동향 키워드 ····· 54
표 12 국내 데이터 산업 동향 키워드 ····· 58
부록 1: 데이터센터 인프라 기술 약어집(알파벳 순) ····· 66
표 14 SingleStore사 동향 ····· 73
표 15 빅데이터센터 설계 동향 ····· 74

표 16 씨게이트(Seagate)사 동향 ····· 75
표 17 DataCore Software사 동향 ····· 76
표 18 Elbencho사의 스토리지 벤치 마크 동향 ····· 77
표 19 KIOXIA사 동향 ····· 78
표 20 Spectra Logic 사와 OpenDrives 사 동향 ····· 79
표 21 SCSI 사업자 협회 동향 ····· 80
표 22 Fungible사와 LANL연구소 동향 ····· 81
표 23 삼성전자 동향 ····· 82
표 24 Spectra Logic 사 동향 ····· 83
표 25 VAST Data 사 동향 ····· 84
표 26 SNIA and OpenFabrics Alliance 동향 ····· 85
표 27 스페인 슈퍼컴퓨팅 네트워크 사업의 스토리지 서비스 동향 ····· 86
표 28 IBM사의 스토리지 제품 라인 동향 ····· 87
표 29 Ohio Supercomputer Center의 스토리지 확장동향 ····· 88
표 30 오크릿지 슈퍼컴센터의 Exascale System용 스토리지 규격 발표 89
표 31 스타트업 22dot6사의 동향 ····· 90
표 32 Academic Sinica HPC 스토리지 구축 동향 ····· 91
표 33 미네소타 대학교 슈퍼컴퓨팅 연구소 동향 ····· 92
표 34 Qumulo사 동향 ····· 93
표 35 독일 율리히 슈퍼컴센터 동향 ····· 94
표 36 DDN사와 Bytesnet사 동향 ····· 95
표 37 Pengchen Lab 동향 ····· 96
표 38 이탈리아 Pisa 대학 동향 ····· 97

표 39 EU 프로젝트 ADMIRE 동향 ····· 98
표 40 Supermicro사 동향 ····· 99
표 41 미국 샌디에고 슈퍼컴센터 동향 ····· 100
표 42 호주 Pawsey 슈퍼컴퓨팅 센터 동향 ····· 101
표 43 Panasas사 동향 ····· 102
표 44 미국 HPE 사의 GreenLake 동향 ····· 103
표 45 Spectra Logic사 테이프 저장 장치 동향 ····· 104
표 46 Kioxia 사 동향 ····· 105
표 47 펜실베니아 주립 대학 Huck 연구소 동향 ····· 106
표 48 Adaptive Computing사 동향 ····· 107
표 49 Quantum사 동향 ····· 108
표 50 일본 후지(FUJY)사의 테이프 스토리지 동향 ····· 109
표 51 Exxact사 동향 ····· 110
표 52 Lightbits사 동향 ····· 111

<그림 차례>

그림 1 CASTOR HSM 기능 ····· 9
그림 2 기존 (구) CERN 스토리지 체계 ····· 10
그림 3 EOS 기능 ····· 10
그림 4 Ceph 데이터/스토리지 서비스 ····· 11
그림 5 데이터 & 스토리지 서비스 활용 예시 ····· 11
그림 6 독일 KIT 스토리지 시스템 ····· 16

그림 7 FERIMILAB DUNE 실험 ····· 25
그림 8 2021년 3월 15일 기준 KEK 컴퓨팅 시스템 ····· 31
그림 9 Open Storage Research Infrastructure 구축 사업 ····· 39
그림 10 프로젝트 4개 모듈별 결과물 형태 ····· 41
그림 11 코로나19 연구출판물과 당뇨병, 치매 데이터의 오픈액세스 비교 43
그림 12 국경간 데이터 이동 관련 규제의 연도별 개정 및 누적 건수 ····· 45
그림 13 국경간 데이터 이동 관련 규제 기준 ····· 46
그림 14 개인정보보안 장치를 위한 OECD 주요 원칙 ····· 47
그림 15 10인 이상 기업의 보안사고 경험 비율 년 (2019년) ····· 48
그림 16 데이터 기업들의 시가 총액 평균성장지수 ('85년 ~'20년) ····· 50
그림 17 국가별 빅데이터 벤처캐피탈(VC) 투자 ('19 년 기준) ····· 50
그림 18 국내 데이터산업 직접 매출 시장 규모 ····· 59
그림 19 2020년 데이터산업 인력 및 데이터직무별 인력비중 ····· 60

1. 서론

가. 조사의 배경

- 연구장비의 초대형화에 따른 실험 데에터의 증가율이 급격히 증가하면서 이를 효율적으로 저장, 관리, 분석하기 위한 글로벌 커뮤니티들의 최신 노력을 조사하고, 현장 중심의 버톰업 방식의 정책 수립을 위한 기초 자료 작성을 위하여 필요함.
- 인공지능과 머신 런닝 기술의 시대로 돌입하면서 데이터의 대용량화가 연구 환경에 주류를 이루게 됨. 이에 따른 해외 데이터 센터의 대응 방향 모니터링 필요성 증대
- 이러한 변화를 조사하기 위한 해외 데이터 센터의 인프라 동향을 세계 최대 가속기 실험 장비를 보유한 스위스 CERN 연구소를 중심으로 한 글로벌 협력 Tier1 급 데이터 센터들의 변화해가는 인프라 구축 동향을 조사하고자 함.

나. 데이터의 대용량화에 따른 조사 주안점

- 데이터의 대용량화는 필연적으로 인프라 규모의 확장을 불러오는데 무한적으로 인프라를 확대할 수는 없음. 따라서 선진 데이터 센터에서 선호하는 첨단 인프라 자원의 형태와 인프라 구성 및 서비스 방식의 변화를 파악하는 것이 중요함
- 데이터의 대용화에 따른 인프라 확장은 인프라의 효율적인 공동 활용 효과를 극대화하여야하는 당위성을 설명하여야 함. 이를 위한 해외 유명 데이터 센터들의 선택하고 있는 접근 방식의 분석이 중요함
- GSDC 사업과 연관하여 실제로 기초과학 분야의 초대형 연구장비에서 발생하는 데이터의 급증에 대비한 인프라 규모의 급증을 늦추는 것을 포함한 예산 절감 노하우를 파악하는 것이 중요함

다. 조사 범위

- 스위스 CERN연구소 중심의 상위 급 Tier 1 센터 인프라 동향
- 미국 NSF 사업에서 주관하는 빅데이터 인프라구축 및 지원사업 동향
- OECD에서 추구하는 빅데이터 정책 동향
- 국내에서 추진하고 있는 빅데이터 관련 정부 사업들의 최신 동향
- 국내외 산업체의 스토리지 기술 동향

2. 인프라 중심 WLCG Tier1 센터 동향

가. 스위스 CERN 데이터 센터

표 1 CERN 인프라 동향 키워드

가.1 CERN 개요, **가.2** CERN IT 부서 미션, **가.3** Computing Facilities 측면, **가.4** Linux (OS), Elastic Search(검색), 대화형 서비스, **가.5** 구성 (Configuration), 클라우드 및 메시징, **가.6** 모니터링 및 실험 테스트 프레임워크, **가.7** 컴퓨팅 서비스, **가.8** 데이터베이스 서비스, **가.9** 스토리지 – CEPH, CVMFS, CERNBox, **가.10** 스토리지 – SWAN, FTS, DTO, EOS, **가.11** 스토리지 – Science Mesh 및 컨퍼런스, **가.12** CERN 테이프 아카이브(CTA) 구축, **가.13** CTA 및 테이프 인프라, **가.14** 공유파일시스템 전략 업데이트, **가.15** 네트워킹 및 전화 통신 서비스, **가.16** 협업 서비스, **가.17** 디지털 저장소 (Digital Repositories), **가.18** 디지털 저장소

가.1 CERN 개요

- 1954년 23개국이 참여하여 설립
- 현재 (2020) 기준 세계 최대 입자물리 연구소
- 2500여 스태프 근무, 17000여명 이용자 지원

가.2 CERN IT 부서 미션

- CERN 연구소가 고유 미션을 달성하도록 지원함. 예를 들면.
- Meryrin 사이트에 있는 main data center 운영
- 600 PB 규모의 데이터 저장

가.3 Computing Facilities 측면

- 정량적 규모(2021년 3월 15일 기준): 서버 13000대, 코어 224,200개,

디스크 72,600개 테이프 93대, 라우터 319대

- 장비 조달 진행 현황: 2020년 12월부터 3.6 MHS06 구모의 서버 1대. 300PB 규모의 외장 스토리지 1대 조달 구매 진행 중
- CERN Prévessin Computer Center(PCC): 입찰 완료 및 12월 재무 위원회에서 승인됨. 계약 협상을 시작하기 전에 낙찰자와 최종 확인 상태.
- 기타 : Openstack/Ironic 이미지에서 완전히 사용할 수 있는 새로운 결과물에 대한 벤치마킹. DC 모니터링과 OpenDCIM 서비스 중. 온도 모니터링 서비스. 5개의 복도(랙 150개)를 지속적으로 정비 중. ServiceNow에서 데이터 개인 정보 보호 구현 작업 중

가.4 Linux (OS) , Elastic Search(검색), 대화형 서비스

- 리눅스: 칩 기반 시스템 사용 사례로 ARM64 지원 설정. CentOS 변동 관점에서 리눅스 퓨처스 살펴보기(이번 주 추가)
- Elastic search (과거에는 Elastic Stack이라는 검색 엔진): 버전 5(ES5) 및 버전 6 (ES6) 가 단계적으로 종료를 알리는 캠페인 중이어서 OpenDistro 버전을 평가하기위한 파일럿 인스턴스를 생성 중임. 버전 7.11에서의 변동 (SSPL 1.0)에 따라서 Amazon에서의 OpenDistro 서비스 제공 예정임. 2021년도 3월 말에 관련 타임라인을 발표할 예정
- 대화형 서비스: CentOS 8의 대화형 서비스를 위한 LXPLUS8(Linux Public Login User Service 8)이 제공되고 있고 LXTUNNEL 서비스(원격 근무자에게 인기)도 제공 중 임. PodMan 서비스도 제공 중임. CentOS에 동작하는 Stream 8은 2021년 2분기에 제공될 예정임. LXPLUS 사용자들을 위한 GPU 제공은 현재 옵션을 평가 중임.

가.5 구성(Configuration), 클라우드 및 메시징

- 구성 관리: Puppet 6으로의 마이그레이션이 완료됨. PuppetDB v5 마이그레이션은 성능 문제로 중단됨. (https://tickets.puppetlabs.com/browse/PDB-4830). 컨테이너 보안을 위하여 Pilot

Hashicorp사의 Vault를 연구 중. CentOS Stream 8 및 ARM64 지원
- 클라우드 및 컨테이너: 많은 서비스들이 향상 중에 있음.
- 메시징: 현재 46개의 브로커가 서비스 되면서 사용량이 계속 증가하고 있음. 최신 오픈 소스 버전인 ActiveMQ 5.16.1로 업그레이드됨

가.6 모니터링 및 실험 테스트 프레임워크

- 모니터링: 일부 사용자 use case들을 (WLCG Transfers, ATLAS Job Monitoring 등)를 스파크 집계 및 Elastic 검색 스토리지를 기반으로 하는 새로운 워크플로우로 이동시킴. 모든 WLCG 및 실험에 대해 VO 피드(Feed)에서 CRIC (Computing Resource Information Catalogue)로의 migration topology enrichment 완료. MONIT 인프라와 서비스를 위한 SLI(Service Levels Indicators) 및 SLO(Service Level Objectives) 도입.
- 실험 테스트 프레임워크(ETF: Experiment Test Framework): ETF를 Checkmk 1.6으로 업데이트함 (https://indico.cern.ch/event/970604/ 참조). 새로운 작업 제출 및 작업자 노드 테스트 프레임워크가 현재 운영 중. 현재 DUNE 및 CMS(경량급의 일부 사이트 모니터링)에서 ETF를 시험 중.

가.7 컴퓨팅 서비스

- HTCondor 팜은 현재 28만 개 이상의 코어 제공: 새 자원이 베어 메탈 리소스로 추가 됨. Kubernetes에서 배치 컴퓨팅 테스트는 계속 진행 중.
- 저비용의 서비스를 위한 pre-emptible instance 서비스를 포함한 7000코어 수준의 Azure에서의 실제 수준의 운영 테스트 진행 중
- 이론 과학, HSE(건강 안전 환경) 및 빔 분야의 데이터를 위하여 Infiniband 기반 11000 개의 코어를 HPC SLURM(Simple Linux Utility for Resource Management)으로 제공 중: 예산이 다양한 사용 사례를 지원할 GPU 구입을 가능하게 함으로 2021년에 더 많은 자원

을 제공하고 퍼블릭 클라우드에서의 burst data 처리를 시험함. TPU(Tensor Processing Unit)와 같은 퍼블릭 클라우드 리소스가 포함된 배치, 노트북, 대화형, Gitlab CI(Continuous Integration)도 사용할 수 있음.

- HEPiX IPv6 WG과 함께 IPv6 상의 worker 노드 서비스 검증 중: 일부 CERN 서비스에서는 dual stack(IP v4/v6 동시 서비스)을 지원하지 않음. 외부 서비스들은 테스트 중. FreeIPA가 추가된 CentOS 8 서비스가 Active Directory 및 Token을 대체하기위하여 시험될 것임.

가.8 데이터베이스 서비스

- 인프라 및 스토리지: 인프라, 서버 및 NetApp NAS 스토리지에 대한 주요 업데이트.
- 온디맨드 데이터베이스: 고가용성 지원. 자동 전환 및 MySQL용 ProxySQL을 통해 마스터/복제본 설정을 기반으로 한 솔루션 구축. EOS의 기존 백업에서 데이터베이스 복구에 대한 자동화된 테스트 수행
- 오라클 데이터베이스 서비스: 버전 18c 및 19c에서 대부분의 서비스를 사용할 수 있도록 업그레이드 및 통합
- 자바 웹 서비스: Kubernetes 기반 Apache Tomcat과 Oracle WebLogic에서 서비스의 90%까지 구현.
- Pentaho 서비스 보고: 사용자의 인스턴스 구축 자동화, 스케일아웃 클러스터링 및 액세스 이중화 부문을 향상시킴. CERN SSO와의 통합 구현
- Hadoop, Spark, SWAN: NXCALS(Hadoop 기반)를 이용하여 CERN Accelerator Logging 서비스 제공. 플랫폼 안정화를 위한 많은 노력(특히 HBase) 수행. 현재 베타 버전인 SWAN(CERN 서비스가 통합된 주피터 노트북)을 위한 Kubernetes 구현 개발. 스파크 3는 Hadoop 인프라와 SWAN에서 제공하는 Spark as a service에 구축됨.
- 스트리밍(Kafka): use case를 충족하는 멀티 테넌트(Multi-tenant: 다중 고객) 솔루션 개발. CERN Kafka/스트리밍 서비스는 CERN 근접 측정계에 사용되는 LoRa/IoT(사물 인터넷)의 기본 구성요소임.

가.9 스토리지 - CEPH, CVMFS, CERNBox

- CEPH 클러스터와 스토리지 신규 공급: 용량 위주의 IOPS(최소 500 IOPS + GB당 5 IOPS)를 지원하는 새로운 높은 IOPS와 짧은 지연 시간을 지원하는 블록 스토리지 도입.

- 데이터 분석을 위해 컨테이너 이미지를 작업자 노드에 효율적으로 배포하기 위한 CVMFS(CernVM File System) 프로토타입 지원. Gitlab 및 Harbor 서비스도와의 통합 진행 중

- CERNBox (단순 개방형 문서 형식 및 애플리케이션): 테스트 중인 마크다운 편집기 통합(CodiMD) 중 인데 곧 전체 통합 예정임. 점점 더 많은 애플리케이션을 통합할 예정임. 예를 들면, 온리 오피스(OnlyOffice), 콜라보라(Collabora) 등

가.10 스토리지 - SWAN, FTS, DTO, EOS

- 일부 Tier 2 센터들과 SWAN(Service for Web based Anlysis)을 이용한 데이터 분석 도구 서비스에 대해 논의 중: 스토리지, 컴퓨팅 및 소프트웨어를 포함한 주피터(Jupyter Notebook)를 앞세운 전체 소프트웨어 스택 통합 등

- FTS(File Transfer Service) v3.10.0 릴리스됨
- 거의 모든 운영 인스턴스가 최신 버전으로 업그레이드됨
- DTO(Data Transfer Orchestrator)의 첫 번째 버전을 최근 출시함. : VO가 DAQ(Data AcQuisition)에서 스토리지로 데이터를 안정적으로 전송하는 데 사용함.
- 최신 HTTP-TPC로 활성화된 모든 EOS 물리 인스턴스: EOS는 disk-based, low-latency storage service. Having a highly-scalable hierarchical namespace를 나타냄

가.11 스토리지 - Science Mesh 및 컨퍼런스

- 서비스의 동기화 및 공유를 위한 Science Mesh 연합이 성장하고 있음.:
- 테스트베드에 조기에 참여하고자 하는 사이트는 cs3mesh4eosc.eu에서 자세한 정보를 얻을 수 음.
- 컨퍼런스 및 워크샵: 5차(화상) EOS 워크샵(3월 1-4일) 및 CS3 2021 컨퍼런스(Cloud Storage Synchronization and Sharing conference, 1월 25-28일). Gather를 사용하여 새 대화형 네트워킹 형식을 테스트함. 가능한 데이터 보호 문제를 해결하기 위해 이러한 툴을 폭넓게 채택해야 함.

가.12 CERN 테이프 아카이브(CTA) 구축

- CTA(CERN Tape Architecture) 3개가 새로 구축됨. : CMS 및 LHCb 인스턴스가 기존 Alice 및 Atlas 서비스에 추가됨.
- 소규모 실험과 데이터 보존을 위한 CTA 공공 서비스가 구축됨.: CERN의 소규모 실험들이 CASTO(The CERN Advanced STORage manager)에서 CTA로 마이그레이션되기 시작함. 첫 번째로 데이터 수집이 진행 중 (CAST(The CERN Axion Solar Telescope) 실험. FTS를 사용하여 DAQ에서 데이터 전송). 첫 번째 데이터 보존 사용 사례로 Babar 데이터를 CTA로 가져옴.
- CTA는 이제 250PB 이상의 테이프 데이터를 관리함.

가.13 CTA 및 테이프 인프라

- 테이프 리콜 최적화: CTA는 RAO(Recommended Access Order, 권장 액세스 주문서)라고 하는 테이프 리콜 최적화에 대한 지원을 시연. 기본적으로 지원하지 않는 LTO(Linear Tape Open) 드라이브의 효율성 향상

- 테이프 인프라: LHC 3차 실행을 위한 하드웨어 계획의 진도가 나갔음. 조달은 아직 시작되지 않았으며 특히 LTO-9 대기 중임. 최소 4개의 라이브러리와 약 200개의 드라이브로 실행되며, Enterprise와 LTO로 분할되고 SSD를 통해 액세스됨.

가.14 공유파일시스템 전략 업데이트

- CERN에서 공유 파일 시스템의 미래를 전망할 수 있는 워킹 그룹 구성:
- 전체 보고서는 다음 URL 참조. https://cds.cern.ch/record/2750122
- Working Group의 권장 사항을 IT 관리 부서에 제시: 현재 비슷한 서비스 비용으로 CERN의 요구를 충족할 수 있는 비예약형 AFS(Andrew File System)의 대체품은 없음. Open AFS 커뮤니티는 충분히 활성화되어 있음.
- WG가 추천 내용: AFS 서비스를 lxplus 및 lsbatch용 Linux 홈 디렉토리로 유지 관리하려면 AFS에 대한 의존도를 계속 낮추어야함, CERN은 장기적으로 AFS에서 벗어나야 함.
- 전략 커뮤니티에 참여한 모든 사람들은 3차 실행이 끝날 때까지 새로운 솔루션을 위하여 상당한 노력을 기울여야 함.

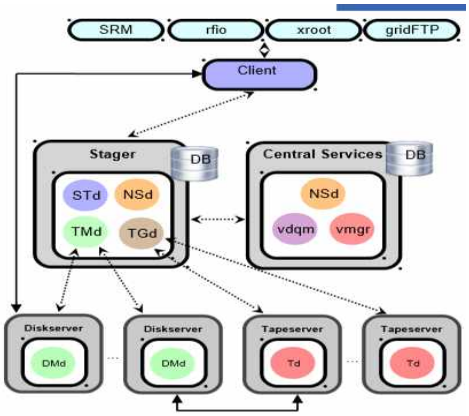

그림 1 CASTOR HSM 기능

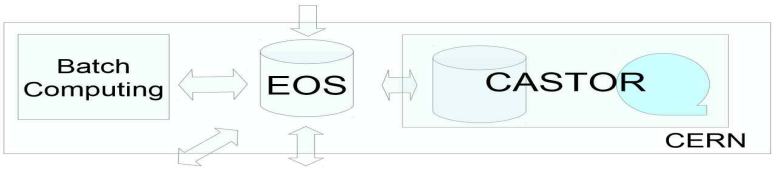

그림 2 기존 (구) CERN 스토리지 체계

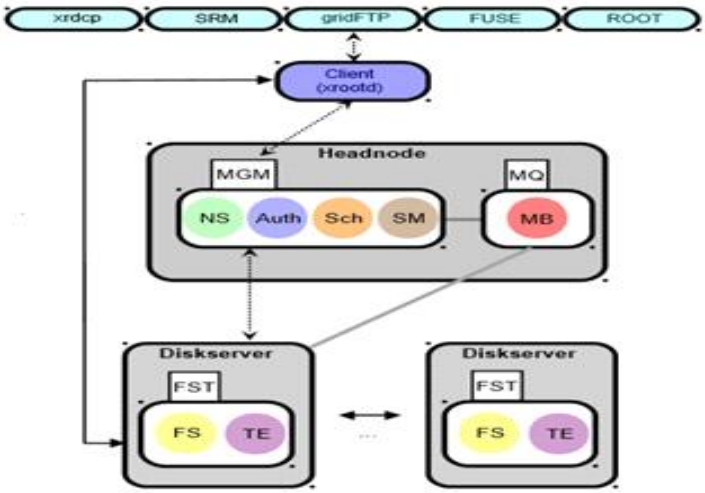

그림 3 EOS 기능

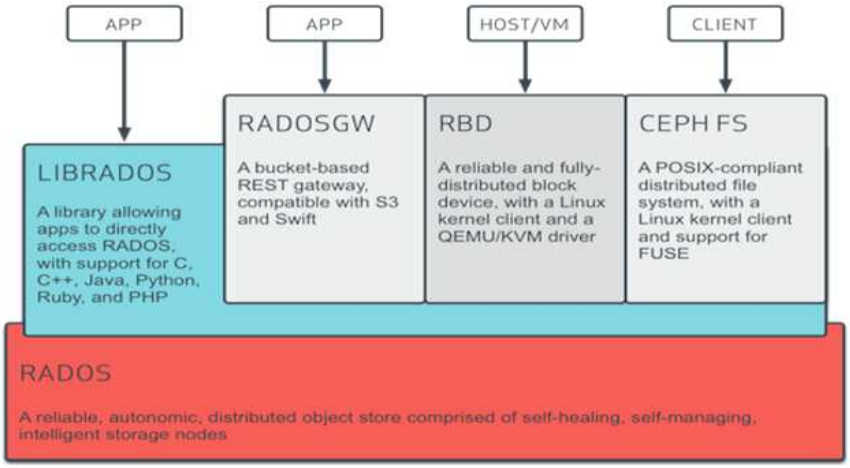

그림 4 Ceph 데이터/스토리지 서비스

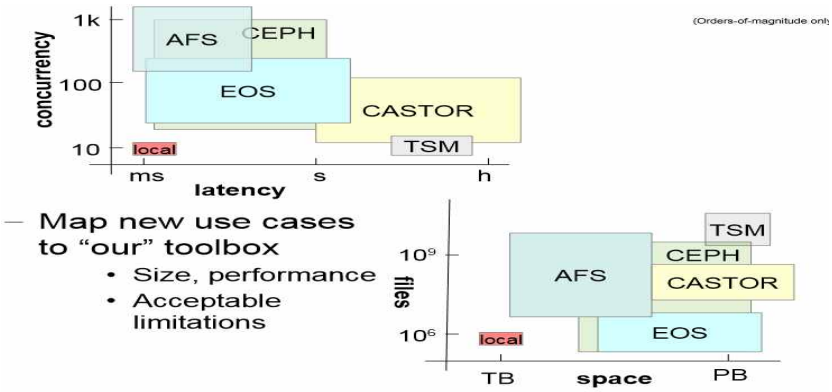

그림 5 데이터 & 스토리지 서비스 활용 예시

가.15 네트워킹 및 전화 통신 서비스

- 네트워킹: 마지막 HEPiX에서 언급한 GEANT 200Gbps가 현재 서비스 중임. LHCONE 200Gbps + 인터넷 200Gbps가 제공됨. DE-KIT에 대한 100Gbps LHCOPN 링크가 추가로 구축되었슴(총 200Gbps). 인터넷 게이트에 새로운 Palo Alto 방화벽 설치 및 운영 준비(4월 예상). 암스테르담 및 프랑크푸르트의 Azure 데이터 센터와 연결 구성.
- 전화 통신: CERN 모바일 전화 서비스를 고려 중임. 데스크톱 스타일의

CERN 전화 클라이언트 개발이 4월에 다시 시작됨 (파일럿 실험용). E-PLMN (Equivalent Public Land Mobile Network)파일럿 과제는 성공하였음. 새로운 SIM(Subscriber Identification Module)의 보급은 4월에 시작될 것으로 예상됨. 다른 스위스산 휴대폰처럼 쉽게 오렌지색 휴대폰으로 전환할 수 있음.

가.16 협업 서비스

- 화상회의: Zoom을 기반으로 한 새로운 서비스 제공. Vidyo 서비스는 퇴출
- 이메일: Dobecot/OpenXChange 기반 파일럿 서비스는 여전히 진행 중
- 도브콧은 주로 보안을 염두에 두고 작성된 Linux/UNIX 유사 시스템용 IMAP 서버. Open-XCange는 MIT 과학자와 CERN이 제공하는 보안 전자 메일 서비스를 통한 오픈 소스 스택 이니셔티브임.
- 웹 호스팅: SLC6에서 최신 컨테이너 기반 호스팅 환경으로 최대 2,500개의 웹 사이트 이동 (2020년 11월). OKD4(2021년 2월)에 기반한 차세대 Platform-as-a-Service 웹 애플리케이션 호스팅 파일럿 시작.
- OKD는 opensource version of Red Hat OpenShift. Kubernetes의 Origin Community Distribution"을 의미

가.17 디지털 저장소 (Digital Repositories)

- InvenioRDM; 전 세계 20개 파트너와 함께 구축한 Research Data Management 저장소 (Zenodo 기반). 리소스, 인덱싱, 검색, 보관, API 등에 대한 액세스 제어가 포함된 v1.0 임.
- CERN 문서 서버: 최신 Invenio 기술 (InvenioILS)을 사용하여 CERN 라이브러리 카탈로그(서적 대출) 개발. 이제 프로덕션 단계로 전환.
- InvenioILS는 Invenio App ILS. Integrated Library System based on Invenio)
- CERN 개방형 데이터: CERNOpen Data-client를 사용하여 CLI를 통

해 데이터셋(Dataset)을 다운로드 가능. CMS 중이온 물리 데이터 제공(210TB). ORCID(The Open Researcher & Contributor ID) 및 ROR(the Research Organization Registry) 식별자 지원.

가.18 디지털 저장소

- REANA(Reusable Analysis: 재사용 분석): 새 릴리스: 0.7.0, 0.7.1, 0.7.2, Helm 차트를 통해 배포, 사용자 리소스 할당량, 하이브리드 워크플로우 지원 (HTCondor, Kubernetes, Slum). 새 블로그: https://blog.reana.io/
- 헬름 차트를 통해 가장 복잡한 Kubernetes 애플리케이션도 정의, 설치 및 업그레이드할 수 있음
- 디지털 메모리: Archivematica 기반 장기 보존 플랫폼 구축. 아날로그 콘텐츠(비디오 6.000개, 이미지 420.000개, 오디오 8.700개)를 디지털화하여 CDS(CERN Document Server)에 부분적으로 업로드.

나. 독일 KIT-Gridka 데이터 센터

표 2 독일 KIT 인프라 동향 키워드

나.1 Tier-1 배치 시스템 및 컴퓨팅 노드 및 워크 노드(CEs & WNs), **나.2** Tape 시스템, **나.3** Network **나.4** 온라인 스토리지 시스템, **나.5**. 스토리지 시스템 하드웨어, **나.6** GPFS Setup, **나.7** GPFS 상의 dCache/xrootd

나.1 Tier-1 배치 시스템 및 컴퓨팅노드 및 워크노드(CEs & WNs)

- 2017년 3월 이후에는 HTCondor 만 서비스 : 활한 운영 중이나 배울

것이 많았음. MC/SC(Multi Core / Single Core) 스케줄링 및 조각 모음의 지속적인 최적화 진행 중. TLAS 파일럿이 비어 있는 문제→ aCT((Advanced Resource Connector Control Tower) 와 함께 확인.

- ARC(Advanced Resource Connector) CE: 2017년 여름까지 많은 안정성과 확장성 문제 있었음. ALICE 및 LHCb 실험을 위해서는 패치가 필요함
- 작업자 노드(WN): 850개 노드, 320kHS06, 그리고 최대 24000 SC(Single Core) 작업 슬롯. 코어당 1.5/1.6 작업 슬롯. 현재 몇 개의 AMD 테스트 시스템을 제외하고는 Intel CPU 만 사용함

나.2 Tape 시스템

- Oracle사 재품 SL8500 2개, IBM 사 제품 TS3500 2개: 20 개의 T10K-D, 17개의 T10K-C, 5개의 TS1140, 5개의 LTO-5. T10K-C 드라이브에 몇 가지 문제가 있는 데 "쓰기 효율성 부족", "롤러 고착", 1000여대의 카트리지가 영향을 받음., 그리고 최신 펌웨어 빌드에서만 모니터링됨.
- Tier-1에서는 TSM(Tivoli Storage Manager)을 사용하고 있고 HPSS 마이그레이션 준비 중
- LSDF(Large Scale Grid Faclity): 재해 복구를 위해 TSM에서 GHI(GPFS-HPSS-Interface)로 전환 중

나.3 Network

- 이중으로 WAN 연결을 위한 2개의 경계 라우터(Cisco Nexus 7010/7710): 2x 10G ~ CERN, 2x 100G ~ LHCONE/OPN/GPI. 2018년 CERN으로 100G(+20G 백업) 연결
- 4개의 내부망 패브릭용 라우터(Nexus 7010/7710): 워커노드(WN)는 10G 지원. 각각의 WN 랙은 패브릭 익스텐더를 통해 40G로 1개의 라우터(백업 없음)에 연결됨

- 최신에 설치한 40G를 지원하는 파일 서버는 Nexus 5696Q(TOR 스위치)에 연결. 이 ToR 스위치는 이중화된 4x100G 업링크를 2대의 Nexus 7710에 연결되어 있음
- 이전부터 사용하였던 구 파일 서버는 1x/2x 10G를 사용하여 하나의 라우터에 연결됨(백업 없음)

나.4. 온라인 스토리지 시스템

- 2016년 말에 새로운 스토리지 시스템을 설치 하였음: GridK에 20PB 할당, LSDF에는 6PB 할당
- 2017년 4월부터 서비스 중. 2018년 여름에 2.7+2 PB를 확장하여 설치. 그리고 11 + 2.2 PB를 확장을 위항여 주문 예정(2018년 가을 기준)
- 시스템 통합사로 NEC 선정: NEC (OEM NetApp) 스토리지 시스템. IBM Spectrum Scale (aka GPFS: General Parallel File System). NEC GPFS NSD(Network Shared Disk) servers. NEC (OEM SuperMicro) 프로토콜 서버들. Mellanox 인피니 밴드(IB) 제품(fabric)

나.5 스토리지 시스템 하드웨어

- Tier-1용으로 23PB 스토리지 시스템 구성
- 13 대의 NEC SNA660(NetApp E5600) 운영: 300개의 8TB HDD 운영. 백업 지원하는 액티브-액티브 컨트롤러 2대. RAID-6 대신 디스크 DDP(Dynamic Disk Pool) 60개 운영
- GPFS 메타데이터용 NetApp E2800 2대 운용: 1.6TB SSD(Solid State Disk) 29개 운영, 백업 기능이 있는 중복 액티브-액티브 컨트롤러 (RAID-1) 2대 운영

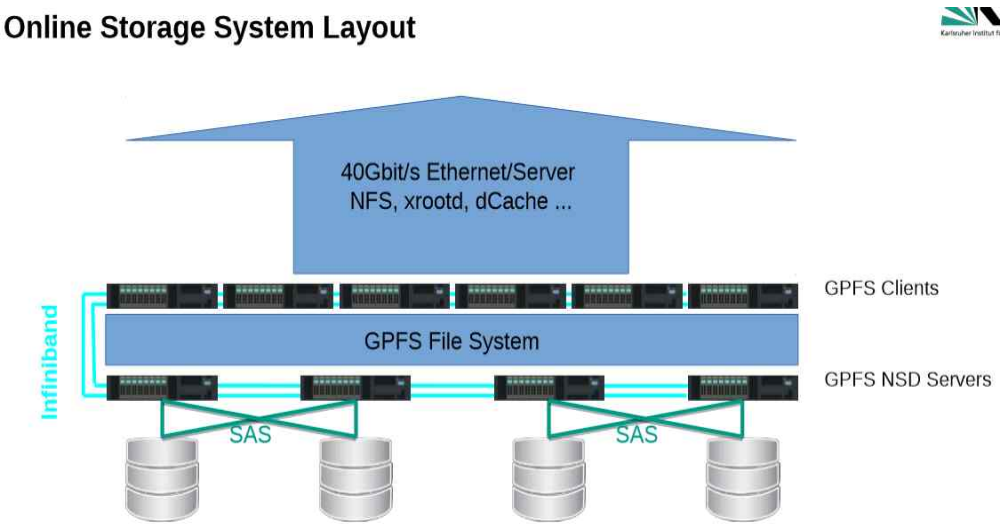

그림 6 독일 KIT 스토리지 시스템

- NSD 서버 16대(20개 코어, 128GB RAM) 운영: 디스크 콘트롤로는 SAS(Serial Attached SCSI), 프로토콜 노드에는 FDR(Fourteen Data Rate: 14Gbps) 인피니밴드 (IB) 사용
- 백업 기능이 있는 2대의 non-blocking 제품으로 Mellanox FDR IB 스위치 12개 사용
- NFS/dCache/xrootd 용으로 44대의 프로토콜 서버(20 코아, 128G 램 지원) 운영. NSD 서버까지는 FDR IB지원, WNs/WAN 따지는 40G 이더넷 지원
- 벤치마크 성능: 70GB/s 속도의 쓰기 + 읽기 지원

나.6 GPFS Setup

- 메타데이터 4 copies 복제 서비스 : GPFS로 2 copies 복사. 메타데이터 스토리지 시스템의 RAID-1으로 2 copies 복제 서비스
- VO당 1개(또는 2개) 파일 시스템 서비스: ATLAS 8.6PB, ALICE 5.9+0.66PB, CMS 5.3PB(LHCb는 다른 스토리지 시스템을 사용함). 과거의 운영 경험에 따라 VO 워크로드 분리 수행
- NSD 서버 클러스터 : 모든 파일 시스템를 호스트함. 16대의 서버를 두

대의 서버를 통해 각각의 NSD 서버에 액세스 가능

- VO당 "프로토콜 서버" 클러스터 1세트 운영. VO 당 8-10대의 서버 구성: GPFS로 VO 파일 시스템을 원격 마운트함. NSD 서버와 dCache/xrootd 서버를 분리하였음

나.7 GPFS 상의 dCache/xrootd

- 서버당 1개의 dCache 풀 구성 / VO당 8-10대의 서버 구성: mover 제한을 최상으로 조정한 것임 (현재 관련 튜닝 작업은 거의 없음). 향후 해결 할 과제로는 dCache 풀(pool)당 ATLAS 데이터 1PB를 60분 소요하는 것을 → 6시간 이내로 dCache 풀(pool)들이 r/w 하게 지원하는 것을 목표로 함

- xrootd SE당 파일 시스템 1개 구성: 개별2222222 네임스페이스 디렉토리(oss.space + localroot)를 삭제할 수 있음

- 데이터 마이그레이션

  - 1단계: ATLAS/CMS/Belle의 dCache를 통해 자동 처리; 최대 6.5PB를 1개월 이내 이동

  - 2단계: ALICE xrootd 데이터의 rsync 작업; 최대 4.5PB 이동. 파일 크기로 인해 매우 천천히 이동

  - 3단계: LHCb의 dCache를 통해 DDN(Data Direct Network) SFA12K(Storage Fusion Architecture 12K; 일종의 블록 스토리지를 나타냄) 기반 스토리지로 자동 이동

-13.5PB의 이전 DDN스토리지가 전환되어 활용될 예정임

다. 영국 RAL 데이터 센터

표 3 영국 RAL 인프라 동향 키워드

다.1 재택근무 (WFH: Work From Home) 및 데이터 센터 운영, **다.2** RAL Tier-1 개요, **다.3** 컴퓨팅 장비 (Batch Farm), **다.4** Castor 장비, **다.5** RAL 스토리지, **다.6** 에코 (Echo) 스토리지, **다.7** SCD (RAL의 Scientific Computing Department) 클라우드, **다.8** 네트워킹, **다.9** 그리드 서비스, **다.10** 기타 활동

다.1 재택근무 (WFH: Work From Home) 및 데이터 센터 운영

- 코로나로 재택 근무가 중요해짐.
    - STFC(the Science and Technology Facilities Council, 영국) 사이트들은 재택근무 환경이 잘 구축 되어 있음 - 네트워크가 양호하게 연결. 사이언스 컴퓨팅 인프라는 자동화 및 원격 관리 능력이 잘 되어있음.
    - 연구소들이 스태프들을 위한 환경 구축을 증진시키고 있음
- 줌(Zoom)이 협업 작업 도구의 국제 표준이 됨
    - UKRI(UK Research and Innovation) 전체에 대하여 면허를 구매함. 줌의 모든 레벨로 매일 사용이 가능함.
- SCD (Scientific Computing Department)에서도 슬랙(SLAC: Stanford Linear Accelerator Center)이 광범위하게 사용됨
    - 특정 주제에 대하여 많은 채널이 '공용' 및 비공개로 제공됨
- 2020년 3월 말 ~ 6월까지 영국이 폐쇄되었음
    - 시스템 운영을 위한 사이트 출입이 매우 제한됨
- 2020년 7월에서 12월
    - 재택근무를 할 수 없는 직원들을 위하여 점진적인 현장 복귀 추진: 작업장, 연구소, 재택근무 환경이 어려운 직원, 전직원의 30%~40% 이내로 근무
- 2021년 1월 21일부터 지금까지 (2012.3. 21 기준) 영국 폐쇄 상태
    - 다시 폐쇄 상태이나 현장 작업은 계속 진행

- 기관 폐쇄:
    - DC(Data Center) 운영 중단. 엔지니어가 방문할 수 있지만 비상시에만 방문 가능

    - Covid-Safe 절차 및 프로토콜, 위험 평가, 방법 진술, PPE (personal protective equipment): 가능한 것에 상당한 제약을 가할 수 있음: DC(Data Center)에서의 Face Cover (얼굴 커버) – 유해한 환경에도 불구하고 강제 공기 흐름이 방울과 에어로졸 확산 증가
    - 정상 활동의 지연. 많은 하드웨어 설치 및 해체 작업이 지연됨: 11월/12월에 대규모로 일이 몰림
- 공기 순환의 상태 지표로서의 CO2 센서
    - CO2 레벨 측정이 공기 혼합 및 공기 흐름을 측정하는 데 효과적인 것으로 확인
    - 개방형 사무실 공간에서 네트워크형 CO2 센서를 사용하여 CO2 레벨 모니터링 시작. 밖이 추울 때도 창문이 열림.
    - DC 전산실에서도 유사한 작업 수행할 계획

다.2 RAL (Rutherford Appleton Laboratory) Tier-1 개요

- 180명의 스태프와 7500명의 사용자: 애플리케이션 개발 및 지원, HPC & HTC 컴퓨팅 및 데이터 시설, 시스템 관리, 데이터 서비스, 수치 분석, 소프트웨어 엔지니어링
- LHC 실험을 위한 Tier1 센터 역할 수행
- JASMIN (RAL 슈퍼컴퓨터 이름): 기후와 지구시스템 사이언스를 위한 데이터 집중형 컴퓨팅 지원
- DAFNI(Data and Analytics Facility for National Infrastructure): 국가 인프라의 분석, 모델링, 그리고 시뮬레이션을 위한 허브
- Daresbury 연구소에서는 계산 과학 및 계산 엔지니어링 수행: 생물학과 생명과학, 이론 및 계산 물리학, 계산 화학, 엔지니어링 및 환경

다.3 컴퓨팅 장비 (Batch Farm)

- ARC CE(Advanced Resource Connector Compute Element: Grid front-end)들로 HTCondor 지원.
    - 2020년에 HTCondor CE를 평가했지만 ARC보다 더 좋은 잇점을 발견하지 못했음.
- 작업들은 도커 컨테이너에서 실행
- 또한 모든 WN(work Node)에서 XCache(컨테이너 내) 실행
- SCD Cloud와 협력하여 CPU 조달 수행
    - 각각 4대의 서버를 포함하는 92 x 2U 섀시(28 Tier-1, 64 클라우드)
    - 듀얼 AMD EPYC 7452(서버 프로세서)
    - 512GB 메모리
    - 3.5 DWPD(Drive Writes Per Day)를 지원하는 3.84TB SSD
- 향상된 내구성의 SSD를 제외하고 이번 조달은 이전과 매우 유사함
    - FY19/20 Tier-1 CPU의 SSD 사용량은 평균 2.14 DWPD임.
- 약 316,000 HS06 성능을 제공하는 800개 이상의 노드에 걸쳐 31,500개의 슬롯 제공.
- HTCondor 지원: 유지 관리 중인 Docker 컨테이너 이미지 내에서 작업 시작.- 기본 OS에서 실행 환경 분리. 작업 실행에 사용할 수 있는 CentOS 8 이미지와 비슷함.
- SCD(Scientific Computing Department) 클라우드에 보급을 자동화하고 정상적으로 작동함. 일반적으로 600개 이상의 코어 그리고 종종 1200~1800개 코어까지의 SCD 보급을 자동화함. 서비스를 위한 GPU 제공에도 활용됨
- Argus(인증서버)를 통한 중앙 서스펜션(보류, 유예 등)이 모든 서비스에 적용됨

다.4 Castor 장비

- 지난 6개월(2019년 상반기) : Ceph 기반 Echo 스토리지를 위한 '디스

크 전용' CASTOR pool 해체 작업 계속. ATLAS, CMS, non-LHC 사용자는 이전 완료. LHCb는 2019년 5월 7일 완료. Alice 데이터는 2019년 4분기 말 종료 예정. 새로운 테이프 캐시 pool이 현재 풀 프로덕션 레벨의 서비스 중. 2019년 5월 초까지 모든 사용자 마이그레이션 하였음.

- 테이프 보관 서비스의 진화에 대한 보다 자세한 내용 참조 (https://indico.cern.ch/event/810635/contributions/3593326/)

다.5 RAL 스토리지

- RAL 스토리지 솔루션을 여러 형태로 지원
    - STFC(the Science and Technology Facilities Council)의 보유 장비로서의 스토리지 지원: 뮤온 실험을 위한 과제 ISIS를 위한 스토리지와 CLF(Central Laser Facility)과제를 위한 스토리지 지원
    - STFC 클라우드 팀 서비스에서 지원하는 스토리지
    - LHC 실험(GridPP)를 위한 Tier1 센터 스토리지

다.6 에코 (Echo) 스토리지

- 영국 RAL과 STFC site에 구축한 a large Ceph 스토리지 cluster
- Ceph 백엔드 스토리지는 믿을 수 없을 정도로 잘 운영되고 있음.
    - 아마도 가장 큰 Erase Code Ceph 인스턴스
- Nautilus에서 안정적임
    - 3/4분기 Octopus 보기를 권장.
- FY 20/21 - 48개의 스토리지 노드 추가 설치
    - 24 x 16TB 드라이브, 듀얼 Intel Xeon 4208, 192GB 메모리 및 25Gb/s NIC.
- 총 구축 용량은 69.9 PB의 원시 데이터.
- Ceph용 XrootD 플러그인이 Vector Reads를 구현하기 위해 개선됨. GridFTP 플러그인은 서서히 철수 할 예정.

- S3 엔드포인트 사용이 증가 (비LHC 실험).
    - 10개 미만의 VO가 340TB와 1억 4천 6백만 개의 객체 (object) 스토리지 사용.

다.7 SCD (RAL의 Scientific Computing Department) 클라우드

- 현재 Scientific Computing 부서 내의 새로운 그룹 내의 클라우드 활동은 사용자에게 동적 인프라 및 클라우드 서비스를 제공하는 데 주력하고 있음
- OpenStack을 기반으로 IaaS 플랫폼을 구축함: 최대 9k vCPU 코어, 160개 이상의 프로젝트(사용자) 지원. 가상화 또는 컴퓨팅을 위한 GPU의 여러 가지 서비스 제공
- STFC(Science and Technology Facilities Council) 및 외부에서 과학자에게 동적 컴퓨팅 리소스 제공: LSST(Large Synoptic Survey Telescope), EUCLID(은하계 연구 과제), AENEAS(Advanced European Network of E-infrastructures for Astronomy with the SKA), UKAEA(United Kingdom Atomic Energy Authority) 지원. IS(neutron source), DLS(Diamond Light Source), CLF(Central Laser Facility), ASTEC(Accelerator Science and Technology Centre), UKATC(UK Astronomy Technology Centre), RALSpace (우주선 연구), Hartree (가상 엔지니어링 센터) 지원
- 시스템 관리자 모집함, alexander(dot)dibbo(at)stfc(dot)ac(dot)uk 참조할 것

다.8 네트워킹

- IPv6 이슈
    - Tier1 방화벽 바이패스 현재 100Gb/s 사이트 경계 라우터에 직접 연결되었지만 40Gb/s로 서비스 됨: IPv6 패킷 손실 문제를 해결하지 못함
    - JANET(2019년 9월)에서 불량 라우터 카드를 발견하고 교체함:

이로 인해 Tier2s의 IPv6 문제 해결

- Tier1 사이트-내부 IPv6 트래픽에 문제가 발생한 후 사이트 코어 스위치 중 하나가 재부팅됨. 이 때문에 T1에서 PPD(Particle Physics Department)-T1 트래픽에 대한 나머지 문제가 해결

- GridPP는 영국의 PerfSonar 인프라를 업데이트하기 위해 새로운 표준 하드웨어 구성을 개발 중이며 커뮤니티의 입력을 요청함. 특히 특정 CPU 아키텍처가 유의미한 차이를 만든다는 증거를 가진 사람이 있는 경우. 상세 내용은 다음 URL 참조 요망.
- https://indico.cern.ch/event/849677/contributions/3570742/attachmen ts/1911907/3159545/GridPP-perfSONAR-refresh-01.pdf

다.9 그리드 서비스

- 직원이 부족함: Catalin은 EGI로 옮겨갔으며, 다른 직원은 SCD 내에서 역할 변경됨.
- 대부분의 그리드서비스가 지원되고 있지만 많은 개발이 진행되지 않음
- 2명의 그리드 시스템 관리자의 구직 공고가 HEPiX(2019년 10월 21일) 이후 월요일에 마감됨. 상세 내용은 다음 URL을 참조 요망. https://www.topcareer.jobs/Vacancy/irc250600_9896.aspx
- 상세 내용은 James Adams에게 물어볼 것

다.10 기타 활동

- WLCG 보안 운영 센터 WG 참여: RAL에서 데이비드 크룩스가 리드. 수요일에 C.f. 데이빗이 발표 예정. 상세 내용은 다음 URL 참조 요망. https://indico.cern.ch/event/810635/contributions/3593242/
- 엔터프라이즈 가상화: Hyper-V 클러스터에서 VMware 클러스터로의 마이그레이션 완료. 독립형 Hyper-V 시스템 마이그레이션은 2019년 11월 중에 완료됨.

라. 미국 페르미 랩 데이터 센터

표 4 미국 페르미랩 인프라 동향 키워드

라.1 페르미 랩 개요, **라.2** 현재 및 단기 실험 내용, **라.3** 향후 대표 실험: DUNE (Deep Underground Neutrino Experimeent), **라.4**. 페르미 연구소 전산 설비, **라.5** 분산 컴퓨팅, **라.6** 컴퓨팅 자원 및 액세스의 발전, **라.7** 페르미랩 HEPCloud, **라.8** 컨테이너 컴퓨팅 이니셔티브, **라.9** Fermilab 스토리지, **라.10** 페르미랩 인증서

라.1 페르미 랩 개요

- 마국에서 유일한 HEP 전용 연구소 : 1967년 창립, 시카고 서쪽 위치, 대표 성과로는 bottom quark(1977), top quark(1995), tau neutrino(2000)
- 현재 사이언스 프로그램을 빛(intensity), 에너지, 그리고 우주 경계로 확장 중: 짧은 기준선(short baseline)과 긴 기준선(long baseline)에서에서 중성미자 실험을 하고 Fermilab 중심으로 정밀 뮤온(Muon) 실험을 수행함. CMS 및 LHC Accelerator 프로그램에서 중요한 역할 및 직접 암흑물질 검출 실험 및 암흑에너지 조사에서의 중요한 역할을 수행함
- 전체 과학 프로그램에서 중요한 컴퓨팅 역할 수행
- 페르미연구소 가속기 복합단지: Main Injector: proton 에너지가 120GeV, Booster: proton 에너지가 8 GeV

라.2 현재 및 단기 실험 내용

- Neutrinos: Long baseline 실험으로는 NovA (최근 끝난 실험: MINERvA, MINOS). short naseline 실험으로는 MicroBooNE, ICARUS (2019), SBND (2020)
- 정밀 뮤온 물리학: 뮤온 g-2, Mu2e (2021)

라.3 향후 대표 실험: DUNE (Deep Underground Neutrino Experimeent)

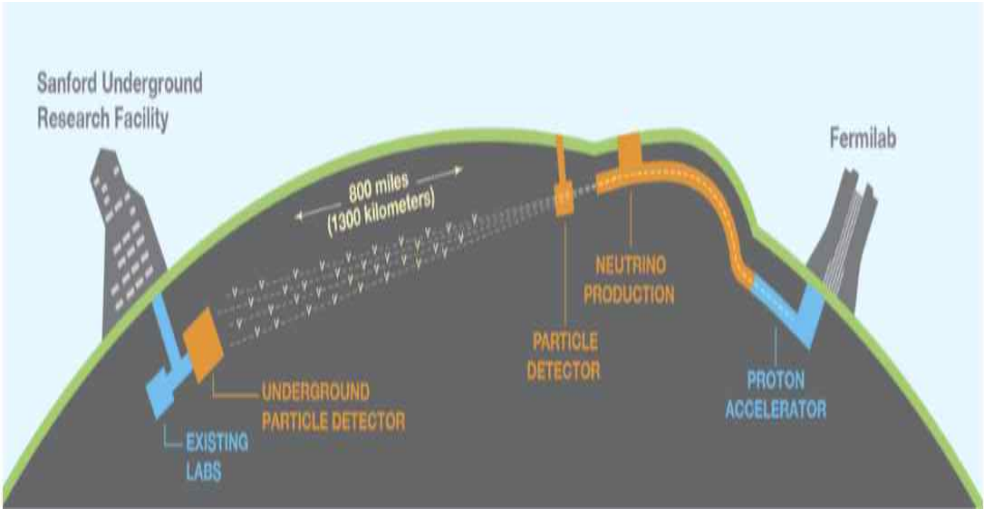

그림 7 FERIMILAB DUNE 실험

라.4 페르미 연구소 전산 설비

- 파인만(Feynman) 컴퓨팅 센터(FCC): 전산실 2개 각각 0.75MW의 공칭 전력과 냉각 지원. 발전기 백업이 포함된 UPS 지원, 전력 의존도가 높은 대부분의 서비스 지원. 중앙 서비스(이메일, 웹), 대부분의 Disk 서버, 비 CMS 실험을 위한 Primary 테이프 스토리지
- 그리드 컴퓨팅 센터(GCC): 3개의 전신실 각각 공칭 0.9 MW 냉각 설비 및 전원을 지원함. UPS를 사용하지만 제너레이터 백업은 없음. 대부분의 컴퓨팅 서버 및 CMS 테이프 스토리지을 운영 중.
- 두 센터 사이를 총 600Gbps 링크로 연동하여 운영함

라.5 분산 컴퓨팅

- 페르미그리드: 모든 빛의 강도 관련 선도 연구 (IF: Intensity Frontier)와 기타 비CMS 실험 지원 (~19k 코어, ~200kHS06 지원)

- CMS Tier-1 및 LPC: 미국 CMS 연구자를 위한 분석 전용 LPC(LHC Physics Center) 클러스터 (~27k 코어, ~285kHS06 지원). 일반적으로 중앙 CMS(Tier-1)용으로 LPC에 5k-7k 코어 할당.
- 래티스 QCD. 미국 LQCD에서 사용하는 HPC 리소스 지원: 최대 19k 코어를 추가하는 다양한 클러스터 지원. 80 테슬라 M2050 GPU 지원
- 실험에 의한 때때로 OSG(Open Science Grid) 자원을 상당량 이용함
- 실험들이 점점 더 많은 HPC 및 상용 클라우드를 활용함
- 테스트/개발을 위하여 대화형으로 자원 지원. 대부분 VM을 테스트 실험에 따라 할당

라.6 컴퓨팅 자원 및 액세스의 발전

- 미래 수요를 대비한다면 전용 형식의 / 로컬에 설치된 리소스는 부족할 것으로 예상됨
- "비전통적" 리소스에 대한 수요 및 가용성 증가 예상 : 예를 들면, HPC, GPU/가속기,
- 물리적 리소스와 할당된 기능 간에 정의되어 있는 현재 링크 방법에서 벗어날 필요가 있음: 예를 들면, 유형/실험이 x인 작업은 항상 리소스 y에 제출되는 것
- 논리적 워크로드 설명으로 전환하고 작업 특성을 충족하는 리소스와 매치시킬 필요가 있음: 그러면 작업 유형 및 보다 이기종 리소스 풀에서 모두 확장 가능
- 사용자를 위한 과학 게이트웨이 제공: 예를 들면 HEPCloud로서 요구사항/제한 사항과 함께 작업이 제출되면 의사결정 엔진으로 리소스 매칭을 수행

라.7 페르미랩 HEPCloud

- 의사 결정 엔진이 작업 요구 사항에 따라 지능적으로 리소스 요청을 생성함
- HPC(NERSC) 리소스 및 클라우드 (아마존, Google) 자원을 공략함

- 새로운 자원을 사용하기 위한 실험들은 조절되어야 함: 운영과 실험이 서로 협응하고 있는 지 확인하기 위해 조정 작업이 필요함. 신규 자원의 예산이 책정되거나 할당되면서부터 주의를 기울여야 함

라.8 컨테이너 컴퓨팅 이니셔티브

- 컨테이너 컴퓨팅은 이미 여러 서비스에서 사용 중: FermiGrid/CMS Tier 1에서 현재 모든 배치 작업이 도커 컨테이너 내부에서 실행됨. 작업들을 서로 격리하고 작업당 자원의 한계를 늘릴 수 있음. HAProxy 소프트웨어 로드 밸런서가 도커화되어 여러 서비스에 사용됨
- 컨테이너를 제작, 테스트 및 저장하는 도구: 기본 이미지에서 빌드할 서버 플랫폼. On-site Docker registry (VMWare Harbor). Harbor(VMWare)는 서버 간에 프로젝트 복제가 내장되어 있으므로 프록시 서버 뒤에 Active-Active 쌍이 있음. Jenkins를 사용하여 다양한 용도로 CI/CD 컨테이너 워크플로우 테스트
- 여러 사용자를 위한 컨테이너 관리(managing)/조정(orchestration): Red Hat OpenShift의 오픈 소스 버전인 OKD를 사용함. 현재 임의 서비스를 위한 소규모 프로덕션 클러스터. 첫 번째 테스트 대상 애플리케이션은 Rucio임 .
- CMS 개발자들이 최대 100개까지의 구형 작업 노드를 가지고 실험하기 위해서 훨씬 더 큰 OKD 클러스터 설치를 시작하려고 함. 잘 알려진 기술을 활용하여 처음부터 무언가를 구축할 필요를 없앰. 멀티 테넌트(여러 사용자 지원/분리). SDN 관리(소프트웨어 정의 네트워크). 많은 워크플로우를 수용할 수 있을 정도로 유연함. 모르는 것을 배울 기회임
- 컨테이너 관련 서비스 및 인프라에 대해 다른 기관과 논의 및 협력하는 데 관심이 있음

라.9 Fermilab 스토리지

- 2018년 6월 이전까지 모든 테이프 라이브러리를 Oracle/스토리지 (테크 SL8500 : 7*10k-슬롯) 사용 : T10KC, T10KD, LTO4 드라이브

조합으로 구성. 125PB 동적 스토리지. Enstore에서 관리

- dCache를 통한 Disk 스토리지를 우선 제공함: 페르미연구소 실험을 위하여 public dCache instance 제공하는 데 테이프 백업 r/w, 영구적(실험에 의해 할당됨) 및 공유 스크래치를 조합하여 제공함. 전용 CMS dCache 지원하는 데 테이프 캐시와 전용 Disk 스토리지 사이를 분리하여 지원함.
- LPC (페르미랩 LHC 물리센터, 미국 CMS) 분석 사용자를 위하여 EOS 인스턴스 지원
- 테이프 시스템을 LTO8/IBM TS4500으로 전환: TS4500 라이브러리 2개 구입(~8.6k 슬롯/ea). Public용(LTO8 드라이브 56개) 및 CMS용(LTO8 드라이브) 도입. 초기 100개의 카트리지 배치와는 별도로 모든 새로운 미디어는 "M8"(LTO7은 9TB 용량으로 포맷됨). 2019 여름 세 번째 TS4500 도입. 개발(엔스토어) 및 운영 통합에 상당한 노력 소요.
- 다음 마이그레이션 주요 단계로는 T2에서 LTO8/M8 미디어로 총 125PB의 데이터 마이그레이션 예상
- 현재 163.5PB의 액티브 테이프 스토리지 지원 : 2018년 8월 이후 LTO8 100개, M8 테이프 2100개 작성. 2018년 11월 전송 기록 : 14PB 전송
- 38PB의 dCache 스토리지: CMS는 최대 24PB의 디스크 인스턴스, 1PB의 테이프 캐시를 지원하며 dCache 2.13 실행. public으로는 r/w, 영구 및 스크래치에 걸쳐 13PB 지원하며 dCache 4.2 실행. 월 18PB 이상 전송
- 7PB EOS(CMS LPC) 스토리지 지원. EOS 4.4.10 실행
- 기타 스토리지: 2PB NAS(인터랙티브 사용), 1PB Lustre(LQCD)

라.10 페르미랩 인증서

- OSG CA는 작년에 종료: Fermilab은 새로운 IGTF 인증 CA를 찾아야 함.
- InCommon은 신뢰 연합(Trust Federation)이며 회원에게 CA 서비스를 제공함. InCommon도 IGTF 인증 CA 서비스 제공. 매우 합리적인 비용

으로 뛰어난 서비스 제공
- 당사의 모든 서비스 인증서를 InCommon IGTF CA로 전환하는 과정: 2019년 4월 전환 시작, 5월 중순 종료
- 다른 CA인 CILogon Basic CA를 통하여 인증을 받기 때문에 개인/사용자 인증서에 영향을 미치지 않음: InCommon 전환이 완료되면 사용자 인증서를 IGTF 인증과 함께 제공되는 CILogon Silver CA로 업그레이드할 계획임

마. 일본 KEK 데이터 센터

표 5 일본 KEK 인프라 동향 키워드

마.1 SuperKEKB: e+e_ intensity frontier, **마.2** J-PARC: Japan Proton Accelerator Research Complex, **마.3** 대규모 컴퓨터 시스템, **마.4** KEK 컴퓨팅 센터 : 조달 및 기타 시스템 현황, **마.5** KEK 컴퓨팅 센터 서비스, **마.6** 국제 망 연결 현황, **마.7** 기타

마.1 SuperKEKB: e^+e^- intensity frontier

- 광의 세기를 8x1035 cm-2s-1 로 설계. 이전의 KEKB 보다 40배 이상 강함
- 첫번째 충돌은 2018년 4월 26일
- 2 단계 실험이 2018년 7월 18일에 종료 되었고. 첫 번째 충돌은 2단계에서 관찰됨.
- 빔 작동은 2019년 3월 11일에 3단계 실험이 실행되면서 전체 벨(Belle) II 검출기(실리콘 VTX)로 다시 시작됨.
- SuperKEKB는 2019년 여름 휴업기간을 거쳐 2019년 10월 15일부터 가동을 재개.
- 누적 통계의 최종 목표는 약 100PB의 원시 데이터에 해당하는 50 ab-1임. 이는 이전의 Belle 실험보다 50배 큰 데이터임

마.2 J-PARC: Japan Proton Accelerator Research Complex

- Fast extraction 과 Slow extraction 동시 지원
- Main Ring: Top Energy 30FeV, FX Design Power 0.75MW, SXPower Expectation 0.1MW 이하
- Rapid Cycle Synchrotron: Energy 3GeV, Repetition 25Hz,
- LINAC(Electron-positron injector, a linear accelerator): 400MeV
- 재료 및 생명과학 설비

마.3 대규모 컴퓨터 시스템

- 리눅스 Cluster + 스토리지 시스템(GPFS/HSM)
- CPU: 15,200개 코어
    - Intel Xeon Gold 6230 2.1GHz
    - CPU/노드 2개, 코어/노드 40개
    - 380개 노드
    - 745 HS06/노드(@2.1GHz(하이퍼스레딩 미포함)
- 메모리: 87TB
    - 코어당 4.8GB(80%) + 코어당 9.6GB(20%)
- 디스크: 25.5PB
    - 17PB: 실험 그룹을 위한 GPFS
    - 8.5PB: HSM 캐시로서의 GPFS-HPSS-Interface(GHI)
- 테이프: 최대 용량으로 100PB 제공
- 그리드 인스턴스가 KEKCC에서 실행 중이며 리소스를 다른 그룹과 공유함

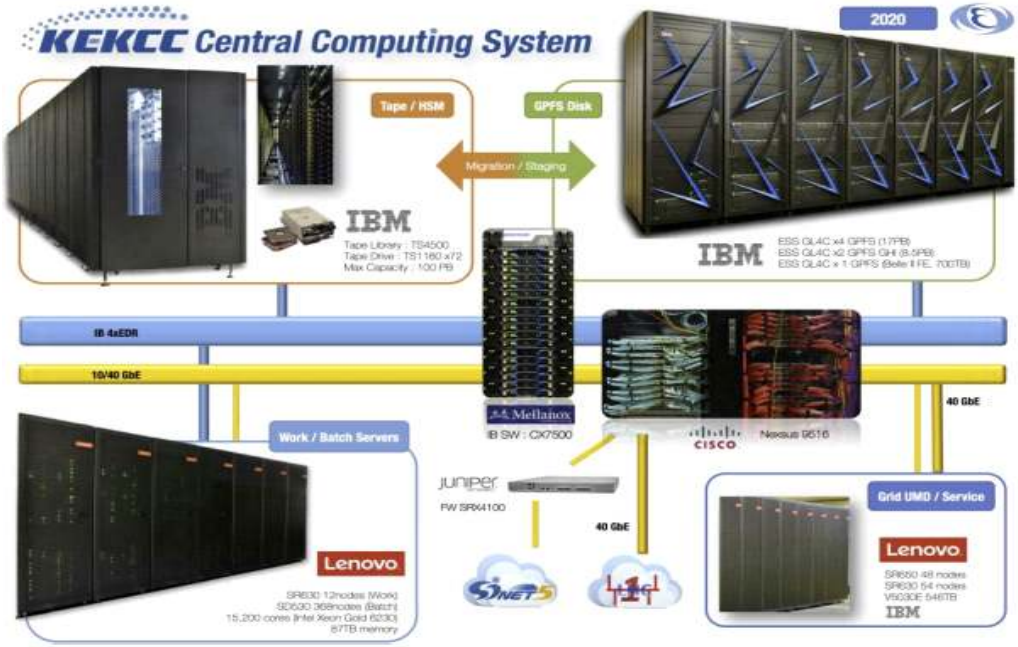

그림 8 2021년 3월 15일 기준 KEK 컴퓨팅 시스템

마.4 KEK 컴퓨팅 센터 : 조달 및 기타 시스템 현황

- 많은 KEK 프로젝트 지원: Belle/Belle2, ILC, J-PARC의 다양한 실험 등
    - KEK는 렌털 시스템으로 자원 확보. 4~5년마다 완전히 교체.
    - 현재 장비는 2020년 9월에 시작되 2024년 8월 이후 종료 예정.
- 데이터 분석 시스템
    - 로그인 서버들과 배치 서버들: Lenovo Think System SD530, Intel Xeon Gold 6230 2.1GHz, 15,200개 코어로 283kHS06 지원(40개 코어 x 380개 노드), Linux Cluster(CentOS 7.7) + LSF(작업 스케줄러) 지원
    - 스토리지 시스템: IBM Elastic 스토리지 시스템(GPFS용 17PB + HSM 캐시 용 8.5PB, 총 25.5PB), HPSS (IBM

TS4500 테이프 라이브러리, 최대 100PB), 테이프 드라이브 (TS1160 x72), 스토리지 인터커넥트 (IB 4xEDR, 100Gbps를 나타냄), 그리드 SE(StoRM:Storage Resource Manager) 및 iRODS(The Integrated Rule-Oriented Data System)의 GHI (Global Horizontal Irradiance) 액세스, 총 처리 성능 (디스크와 GPFS는 100 GB/s 이상, HSM과 GHI는 60 GB/s 이상)

- 그리드 컴퓨팅 시스템: UMD/EGI 및 iRODS/RENCI
- 범용 IT 시스템: 메일, 웹(Indico, Wiki, 문서 아카이브), CA 지원.

마.5 KEK 컴퓨팅 센터 서비스

- 2PB 이상의 Belle II 원시 데이터 아카이브 서비스. 적절하게 예측한다면 2029년에 최대 50 PB이 예상됨
- BelleII를 위한 CVMFS (CernVM File System) 저장소 서비스: Belle.kek.jp
    - Belle II는 원래 belle.cern.ch로 시작.
    - 각 지역에 2개의 복제본(Stratum-1) 구축: 아시아의 IHEP/KEK, EU의 DESY/RAL, 미국의 BNL과 미국의 두 번째 사이트 후보 FNAL. DESY the Deutsches Elektronen-Synchrotron.
    - Dave와 Jakob이 CVMFS 조정 그룹으로서 지원해줌.
    - 클라이언트 설정 파일 배포

- ATLAS CVMFS 저장소를 위한 복제본 호스팅 서비스
    - ICEPP/도쿄-LCG2가 ATLAS를 담당. ICEPP : the International Center for Elementary Particle Physics
    - 같은 국가에 2개 이상의 Stratum-1을 보유하지 않음. Stratum-1: 상위 레벨의 네트워크 타입 서버
- 데이터 관리 혁신 서비스
    - Rucio/BNL로 마이그레이션 완료 (2021년 1월). Rucio: the next-generation Data Management system in ATLAS.

- LFC(LCG File Catalog)는 해체 단계로 이동하여 2021년 여름에 폐기.

마.6 국제 망 연결 현황

- 일본 토쿄에서 미국 로스앤젤레스, 뉴욕, 그리고 유럽 암스테르담 까지 100 Gbps로 연결됨
- 현재 LHCONE 사이트용 IPv4와 IPv6의 경로를 다른 경로를 사용하도록 구성함

마.7 기타

- 우리는 2020년 9월에 전체 시스템 교체를 완료함.
- RHEL6(Red Hat Enterprise Linux 6)에서 실행 중지: 2021년 여름 전에 VOMS(Virtual Organization Membership Service) 및 StoRM 용 OS를 마이그레이션할 계획.
- IPv6 지원 LHCONE: 현재 중앙 네트워크 스위치의 메모리 공간 제약 때문에 LHCONE의 IPv6 경로 경로를 광고하고 있지 않음. 여름에 매년 정전되는 동안 상황이 개선되기를 희망함.

3. 미국 NSF 및 OECD 활동 동향

가. 미국 NSF 지역별 빅데이터허브구축사업 (2단계) 동향

표 6 미국 NSF 지역별 빅데이터허브구축사업 동향 키워드

가.1 미국 빅데이터 동향 요약, **가.2** 지역별 빅데이터 허브(Regional Big Data Hub) 구축 사업 개요, **가.3** South Big Data Hub 구축 사업, **가.4** West Big data Hub 구축사업, **가.5** Midwest Big Data Hub 구축 사업, **가.6** Northeast Big Data Innovation Hub 구축사업, **가.7** 기타 특이 사항

가.1 미국 빅데이터 동향 요약

- 2015년부터 시작한 지역별 빅데이터 허브 구축 사업은 2019년에 1단계 종료하고 2019년부터 2단계 사업(2023년까지) 추진 중
- 빅데이터허브구축사업의 성공적 추진을 위하여 빅데이터 생태계 육성을 위한 관련 H/W, S/W 그리고 활용 활성화 지원 사업을 병행하여 추진함
- 관련 H/W 사업으로는 OSN (Open Storage Network: NSF 연구 데이터 저장 스토리지 인프라 구축 사업) 구축 사업과 NESE (NorthEast Storage Exchange) 구축 사업을(테이프 장치 중심 스토리지 인프라 구축사업) 추진
- 관련 S/W 사업으로는 빅 데이터 Bridge 사업 (이전 DXC : Data eXa Cell) 사업을 추진함(Tapeless 기반(저가의 HDD 중심) 스토리지 인프라 구축 사업)
- 관련 활용 활성화 사업으로는 Big Data Hub spoke 사업을 추진함

가.2 지역별빅데이터허브(Regional Big Data Hub) 구축사업 개요

- 사업 기간 : 2단계 (2019 ~ 2023) 추진 중. (1단계는 2015 ~ 2019)
- 2단계 예산: 허브 당 4년간 400만 달러, 총 16백만 달러 (1단계의 2

배)

- 연장 목적: 데이터 사이언스 커뮤티이 구축, 주요 분야에서의 데이터 기반 연구에서 혁신 주도, Hub 간 협력 수행, 지속적 서비스 등 사회적 성과 확산을 위하여

가.3 South Big Data Hub 구축 사업

- 미국 남부 16개 주 및 D.C. 지역을 포함하며, 조지아 공대(Georgia Institute of Technology)와 노스캐롤라이나 대학(University of North Carolina)이 공동으로 운영 및 조정
    - 분야: 보건의료, 해안재난, 산업 데이터, 재료 및 제조, 생물서식지 데이터
- OSN (Open storage Network) 구축사업으로 NSF로부터 180만 달러 지원 받음 (4대 지역별 빅데이터 허브 센터들이 OSN 사업을 주도함)
- 클라우드 컴퓨팅 사업을 마이크로 소프트 사업의 Azure cloud computing으로부터 지원 받음

가.4 West Big data Hub 구축사업

- 알래스카와 하와이를 포함한 서부 13개 주를 포함하며, UC San Diego, UC Berkeley, 워싱턴 대학(University of Washington)에 의해 공동운영 및 조정
    - 분야: 대용량데이터 기술, 데이터 집약형 과학적 발견, 천연자원, 자연재해, 개인맞춤형 의료
- Jupyter 노트북을 활용하여 빅데이터 서비스 이용 지원

가.5 Midwest Big Data Hub 구축 사업

- 중서부 12개 주를 담당하며, 일리노이대학교(University of Illinois at Urbana-Champaign)에 의해 운영 및 조정

• 분야: 농업, 수자원, 스마트시티

- Big Data Spoke 사업에서 300만 달러 지원 받아 8개 기관에 지원

가.6 Northeast Big Data Innovation Hub 구축사업

- 북동부 9개 주를 담당하며, 콜롬비아 대학(Columbia University)에 의해 운영 및 조정

• 분야: 에너지, 금융, 교육 관련 데이터과학, 기후 및 환경

- Big Data 관련 워크샵 마다 만달러를 지원하고 추가로 Microsoft Azure 클라우드를 25천달러까지 사용 지원

가.7 기타 특이 사항

- 빅데이터 허브 구축사업의 성공적 추진을 위하여 NSF에서는 빅데이터 허브 활용사업 (Big Data Spoke 사업) 빅데이터 인프라 확충 사업 (Big Data Bridge 개발 사업, Open Storage Network 구축사업, Northeast Storage eXchange 구축사업)을 병행하여 추진하고 4대 지역별 빅데이터 허브 센터는 이들 사업에서 중요 역할을 또한 수행함.

나. NSF 빅데이터 인프라 구축 지원 사업 동향

표 7 미국 NSF 빅데이터 인프라 구축 지원 사업 동향 키워드

나.1 Big Data Hub spoke 사업, **나.2** Bridge 사업, **나.3** NESE (NorthEast Storage Exchange) 사업, **나.4** OSN (Open Storage Network) 구축 사업, **나.5** OSIRIS (Open Storage Research Infrastructure) 구축사업

나.1. Big Data Hub spoke 사업

- 2016년도부터 예산 1000만 달러로 10개 빅데이터 응용 분야 지원
- 빅데이터 허브 센터 구축사업과 연계하여 추진(지역별 빅데이터 허브 센터 활용 활성화 사업)

나.2 Bridge 사업

- National Data Service 사업을 위하여 NSF 사업으로 PSC (피츠버그 슈퍼컴 센터)에서 수행함
- Data Supercell 또는 Data eXa Cell (DXC) 사업으로 시작하여 Bridge 사업으로 이어짐
- 예산 규모는 760만 달러(2019 기준). (NSF의 DIBB (Data Infrastructure Building Blocks 프로그램의 일환으로 사업 수행)
- 주된 사업 내용은 테이프 저장 장치 대신 디스크를 사용해서 빅데이터 서비스 기술 개발(Multiple Meta Data server를 사용하여 지원)

나.3 NESE (NorthEast Storage Exchange) 사업

- NSF의 DIBB 사업의 일환으로 추진. Data Centric Infra 구축 목적으로 추진, 하버드 대학에서 수행.
- 기존 Cern 실험 데이터의 ATLAS Tier 2 사업을 하다가 미국에 산재해 있는 Atlas Tier 2 4개, Tier1 1개를 위한 Storage Exchange 사업으로 발전시킴
- Globus 사용하고 Ceph의 OSD(Object Storage Device)를 지원함. Tape 장치를 활용하여 대규모 데이터 서비스 제공

나.4 OSN (Open Storage Network) 구축 사업

- NSF에서 180만 달러, 슈미트 재단에서 100만 달러 지원
- 저가의 스토리지 장비로 200PB 구축 목표
- 주 목적이 NSF 사업에서 생성된 데이터 저장이 목표

- 예상 분야로 LHC, LSST, OOI(해양학: Oceanography), genmics 등
- 2018년부터 지역 빅데이터 허브 센터들이 OSN 사업을 주도하기 시작, 예산 지원을 받아가면서.

나.5. OSIRIS (Open Storage Research Infrastructure) 구축 사업

- OSiRIS는 NSF가 자금을 지원하는 시범 프로젝트로, 주요 미시간 연구 대학을 위한 소프트웨어 정의 스토리지 인프라 평가사업.
- OSiRIS는 여러 혁신적인 개념을 결합하여 분산형 다중 기관 스토리지 인프라를 제공함. 이를 통해 3개 캠퍼스의 연구원들이 컴퓨팅 시설 위치에서 직접 데이터를 읽고, 쓰고, 관리하고, 공유할 수 있음.
- 목표는 모든 캠퍼스에서 잘 연결된 위치에서 동일한 스토리지 인프라에 투명하고 고성능의 액세스를 제공하는 것임. 우리는 네트워크 검색, 모니터링 및 관리 도구의 조합과 CEPH 기능의 창의적인 사용을 통해 이를 가능하게 할 것임.
- 데이터에 대한 컴퓨팅 액세스를 "in-place"로 지원하는 단일 데이터 인프라를 제공함으로써 연구 커뮤니티가 직면한 많은 데이터 집약적 및 협업 문제를 해결함, 이러한 커뮤니티가 자체 대학의 경계를 넘어 쉽게 연구 협업을 수행할 수 있도록 지원함.
- OSiRIS는 상용의 기존 하드웨어를 CEPH 소프트웨어와 결합하여 고성능 소프트웨어 정의 스토리지 시스템을 구축함. 이 시스템은 SSD가 장착된 스토리지 헤드 노드 + 60-디스크 SAS 쉘프, RHEV 클러스터(2-호스트), Globus Connect 서버, PERFSONAR 네트워크 모니터링 노드 및 안정적인 OpenFlow 지원 스위치 등 여러 빌딩 블록 구성요소로 구성됨.
- OSiRIS는 CEPH 스토리지 클러스터를 필요한 스토리지 하드웨어를 구성하는 주요 수단으로 사용하는 소프트웨어 정의 스토리지(예: 스토리지 로직이 소프트웨어 계층으로 추상화된 범용 하드웨어) 서비스를 미시건 대학에 배포할 것임.
- OSiRIS는 National Science Foundation의 지원금으로 지원되며, 수석 조사자는 물리학과 연구 과학자 숀 맥키이고 네트워크 및 스토리지 지원 협업 계산 과학 센터(CNSECCS) 책임자임. CNSECCS 및 OSiRIS는

미시간 계산 과학 발견 및 엔지니어링 연구소(MICDE)의 후원으로 운영됨.

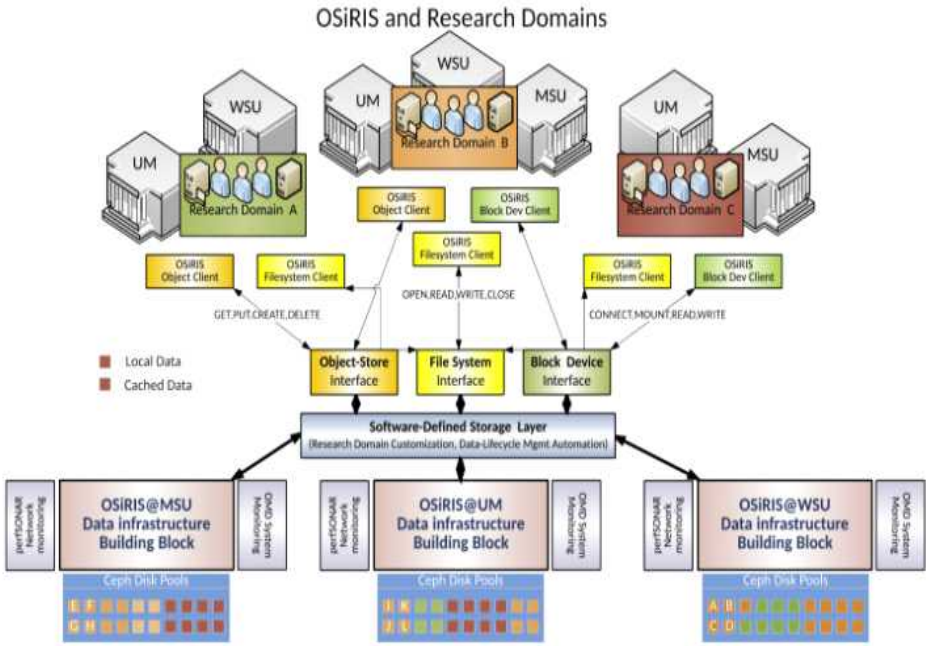

그림 9 Open Storage Research Infrastructure 구축 사업

다. OECD 빅데이터 정책 동향

표 8 OECD 빅데이터 정책 동향 키워드

다.1 OECD 고잉 디지털 (Going Digital) 프로젝트 개요, **다2** 고잉 디지털 3단계 개요, **다.3** 모듈 1: 데이터 관리•접근•공유•통제 (Data stewardship, access, sharing and control), **다.4** 모듈2: 신뢰 기반 국가간 데이터 이동 촉진 (Cross-border data flows with trust), **다.5** 모듈 3: 데이터에 의해 형성되는 기업과 시장 (Data shaping firms and markets), **다.6** 모듈 4: 데이터 및 데이터 이동 측정 (Measurement of data and data flows)

다.1 OECD 고잉 디지털 (Going Digital) 프로젝트 개요

- 디지털변혁 에 따른 경제 사회적 변화를 규명하기 위하여 OECD의 디지털경제정책위원회 를 중심으로 다수의 위원회가 참여하는 최대 프로젝트 중 하나
- 회원국들과 원활한 소통을 위해 각국 대표부 대사단이 참여하는 비공식 회의체인 " Friends of Going Digital" 운영

  - 1 단계 ('17년 ~ '18년) 디지털변혁의 사회 경제 전반의 영향 고찰 (OECD 14개 위원회 참여)
  - 2 단계 ('19년 ~ '20년) 핵심 기술인 인공지능(AI) 과 블록체인 분석 (OECD 20개 위원회 참여)
  - 3 단계 ('21년 ~ ''22년) 성장과 웰빙을 위한 데이터 거버넌스 추진 중 (OECD 16개 위원회 참여)

다.2 고잉 디지털 3단계 사업 개요

- 추진배경
  - 인공지능과 머신러닝 같은 신기술 기반 분석도구는 데이터를 자본 및 노동과 유사한 새로운 생산 요소로 변화시키고, 신비즈니스 모델• 디지털 상품•서비스를 창출하며 기업이 경쟁하는 방식을 변화시킴
  - 코로나 19 대응에서 데이터는 중요한 자산으로 활용되고 있으나 △기술• 정책적 상호운용성, △접근• 공유• 사용, △신뢰• 프라이버시•보안, △측정 등의 도전 과제도 발생하여 해결 방안이 필요

- 주요목적
  - △ 성장과 웰빙의 자원으로서 △ 데이터와 사회·경제적 영향에 대한 이해도 제고, △증거 기반 확보, △각국 정책 프레임워크 개발 지원

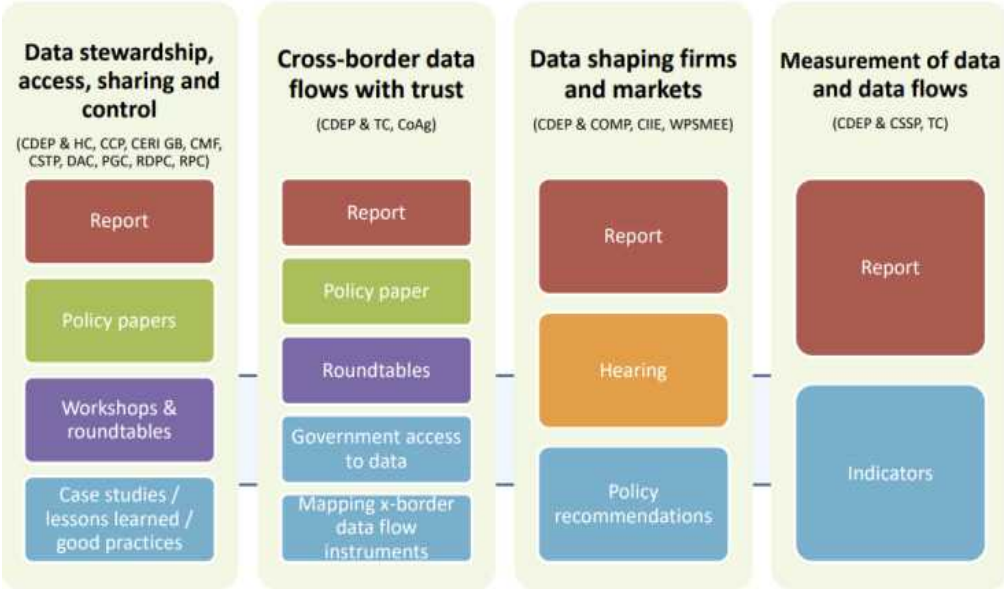

그림 10 프로젝트 4개 모듈별 결과물 형태

표 9 OECD 고잉 디지털 모듈별 관련 부서

모듈	내용	OECD 내 관련 부서
1	데이터 관리•접근•공유 통제에 관한 이슈 정책 및 사례 분석: 기존 데이터 거버넌스 프레임워크, 정책, 이행 현황 등을 규명하고 데이터 정책을 사용하여 경제와 사회 문제 특히 위기에 대응하는 방안에 관한 모범사례를 도출 (보건 위원회(HC) 주도)	CFE, DAF, DCD, EDU, ELS, GOV, STI
2	신뢰 기반 국가간 데이터 이동 촉진 방안 탐색 : 프라이버시보호와 무역 관점 통합 국가간 데이터 이동 관련 토론을 통해 상호 운용 가능한 국가간 데이터 이동 제도 촉진 (무역위원회(TC) 주도)	STI, TAD
3	데이터에 의해 형성되는 기업과 시장 : 기업의 데이터 활용이 생산성과 경쟁에 주는 영향 분석 ((경쟁위원회(TC) 주도)	CFE, DAF, STI
4	데이터 및 데이터 이동 측정에 관한 작업 : 데이터 관련된 향상된 측정 방법과 지표 마련 (통계 정책위원회 (CSSP)주도)	SDD, STI, TAD

- 프로젝트의 4개 모듈과 주요 결과물
    - (결과물) ① 4개 모듈 결과를 통합한 전체 종합보고서, ② 데이터 측정 및 지표 보고서, ③ OECD 권고안•원칙 수립 지원, ④ 고잉디지털 툴킷 (https://goingdigital.oecd.org) 과 과 AI정책 저장소 (https://oecd.ai)에 반영 ⇒ 이들을 통해 "국가의 데이터 거버넌스 평가 프레임워크" 를 마련하고, 디지털경제장관회의 ('22년말 ~'23년초 예상) 를 통해 마무리

다.3 모듈 1: 데이터 관리•접근•공유•통제 (Data stewardship, access, sharing and control)

- 목적: 사회적 문제 해결 등 공공 정책 목적 달성을 위한 측면에서 데이터 공유와 사용의 기회요인과 도전과제를 탐색

- 공공 펀딩 연구 데이터 접근성 권고문 개정: Recommendation of the Council concerning Access to Research Data from Public Fund ('21년 1월 OECD 이사회 채택)

    - 개정 배경: 급속한 디지털화에 따른 연구 환경 변화를 반영하여 연구 데이터의 범위에 디지털 데이터 소프트웨어 알고리듬 플랫폼 등을 포함하고 이러한 연구데이터의 공개 및 접근에 대한 국제표준을 보완
    - 참고> 연구데이터 공유 중요성; 코로나 와 오픈사이언스 19 (Open Science). 코로나19 대응을 위해 과학계는 자원을 전례없이 동원하였으며 '20.1월~11월 동안 7만 5천건 코로나 관련 연구출판물을 발간, 그 중 3/4 이상이 오픈액세스였음 → 오픈된 코로나 19 연구데이터들은 백신과 진단키트 등의 연구개발을 가속화

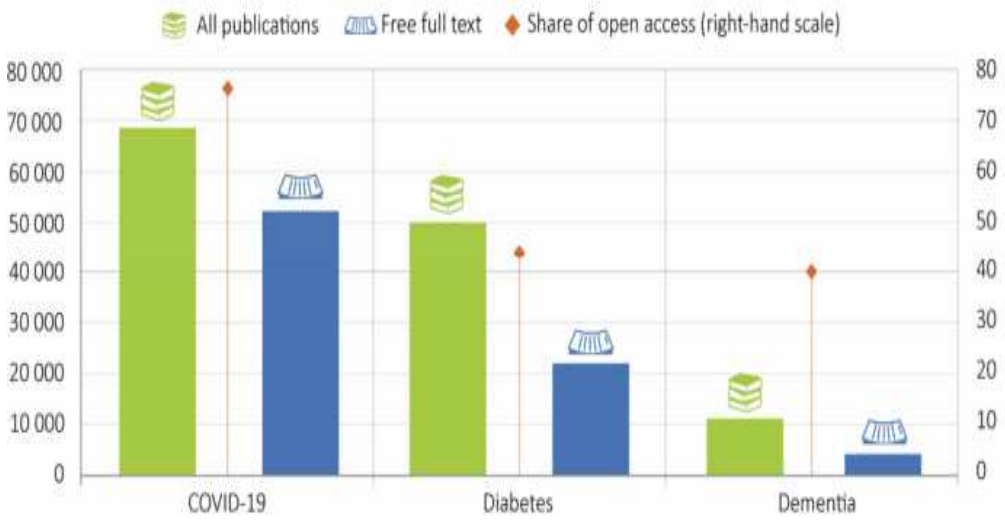

그림 11 코로나19 연구출판물과 당뇨병, 치매 데이터의 오픈액세스 비교
* 출처: OECD STI Outlook 2021 (‘21.1월)

- (구성 및 내용) △ 신뢰를 제고하는 데이터 거버넌스 기술, △ 표준과 실제 사례: 상호호환성 및 재사용 제고 방안 권고, △ 책임• 소유• 관리 연구데이터 관리를 위한 연구자• 펀딩기관• 연구조직• 민-관섹터의 책임과 역할 명확화 △인센티브와 보상, △ 지속가능한 인프라, △ 인적 자본 육성, △ 연구데이터 접근 향상을 위한 국제 협력 강화 등

- 데이터 접근 및 공유 증진(EASD) 권고안 마련중: Recommendation of the Council on Enhancing Access to and Sharing of Data

- **데이터 접근(access)**: 데이터의 잠재적인 이용을 위해 적용 가능한 기술적 재무적 법률적 기관의 요청에 따라 데이터를 요구하거나 찾아오는 행위.
- **데이터 공유(sharing)**: · 다른 사람 기관들의 사용을 위하여 데이터 접근을 제공

- **목적**: 여러 분야와 규제환경에서 데이터 접근•공유를 위한 일관된 데이터 거버넌스 마련과 기존 OECD 권고안 들의 일관성 제고
- **추진현황**: OECD의 디지털경제정책위원회(CDEP), 과학기술정책위원회(CSTP), 공공거버넌스위원회(PGC) 등 개 위원회 주도적 참여 △ '17년 JSG(Joint Steering Group)구성, △ '20.4월 1차, '20.10 월 2차 초안 마련,△ '21.2월 이해관계자 협의 및 OECD내 타 위원회 회람, △ CDEP, CSTP, PGC 승인을 얻어 이사회 제출 예정

- **권고안 구성: 3개 섹션**으로 구성
- **섹션1**: 데이터 생태계 전반의 신뢰 강화: △ 이해관계자에 대한 권한 부여와 적극적 참여 촉진, △ 전략적인 범정부 접근 방식 채택 촉구, △ 개인• 기관의 권리를 보호하고 다른 합법적 이익과 목적들을 고려하면서, 데이터 접근과 공유의 혜택을 최대화
- **섹션2**: 데이터 투자 활성화 및 데이터 접근 공유 장려: △ 지속가능한 비즈니스 모델과 시장개발을 위한 인센티브 제공 및 여건 촉진
- **섹션3**: 사회 전반에 효율적이고 책임성 있는 데이터 접근•공유•사용 촉진: △ 신뢰 기반 국경간 데이터 접근 공유 조건의 개선, △ 공공 민간기관 전반의 데이터 발견가능성•접근성• 호환성•재사용성 촉진, △데이터 가치 주기에서 모든 이해관계자들이 효과적으로 데이터를 책임 있게 사용할 수 있도록 하는 역량 강화 조치

다.4 모듈2: 신뢰 기반 국가간 데이터 이동 촉진 (Cross-border data flows with trust)

- 목적: 정부의 합법적인 공공 정책(법집행•국가안보) 목표를 가능하도록

하면서 국경간 데이터 이동의 혜택을 얻을 수 있는 정책 옵션 탐색

- 국경간 데이터 이동 관련 규제 측면에서의 접근 방법 공통점 조사:*Mapping Commonalities in regulatory approaches to cross-border data transfers (OECD ;21.4월)

  • 이슈: 국경간 데이터 이동 규제는 늘어나고 사법관할권별로 달라 정부의 개인정보보호 등 법집행은 물론 기업의 글로벌 활동도 어렵게 함

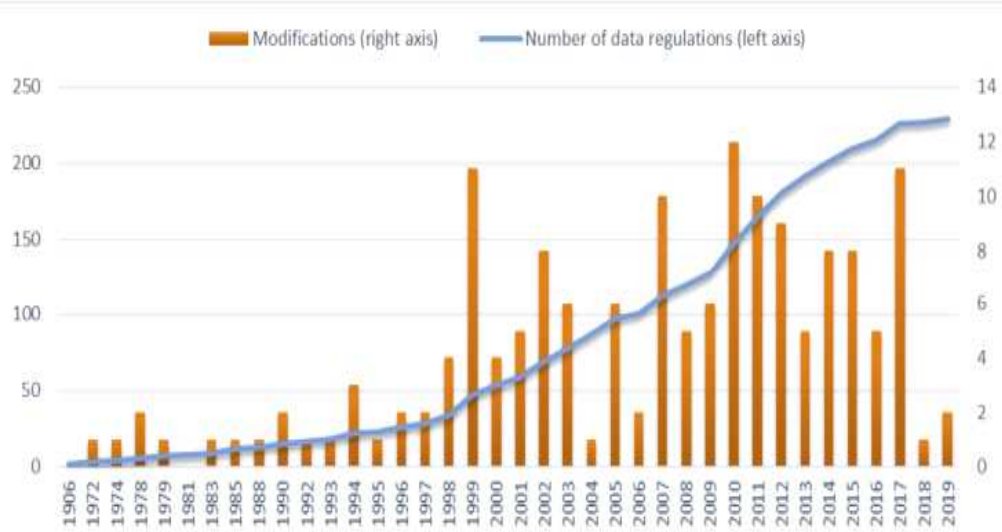

그림 12 국경간 데이터 이동 관련 규제의 연도별 개정 및 누적 건수

  • 조사결과: 단일화된 수단은 없으며 4가지로 분류: ①일방적 조치, ②다원적 합의, ③무역협정 및 파트너십, ④표준과 기술 주도 정책

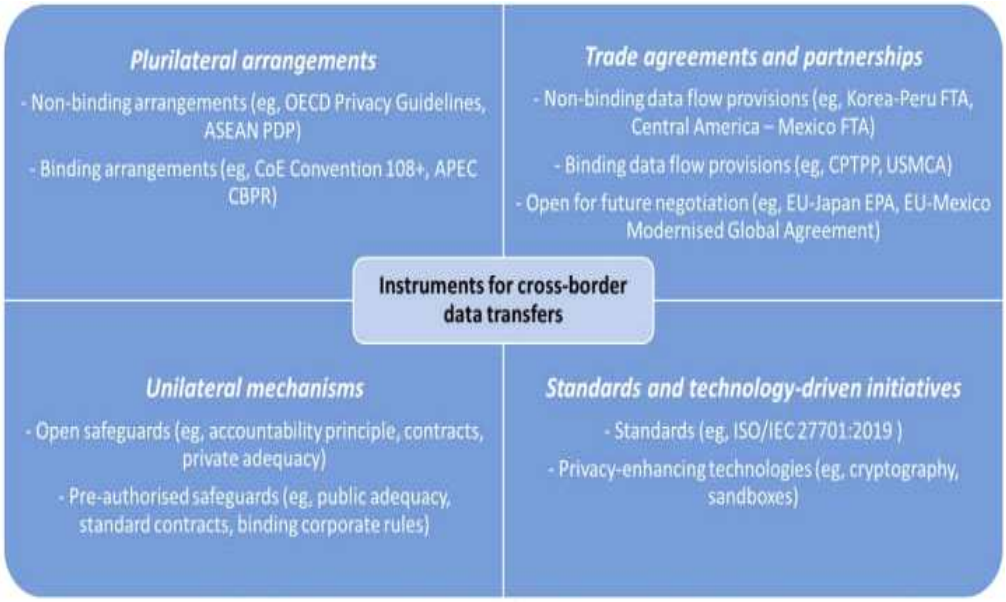

그림 13 국경간 데이터 이동 관련 규제 기준

- 분석결과: △“데이터 보호 안전 조치” 와 “국경간 이동 가능” 2가지 목적을 모두 추구, △ 통합 사례 증가 (예: 무역협정 체결시 개인정보 보호 조치와 함께 국경간 데이터 이동을 포함), △ 수단들간 상호 보완 (예: 무역협정들의 다원적 합의 참조 증가)

- 정부의 민간부문 보유 개인 정보 접근에 관한 상위 원칙 개발: *(가칭) High-level Recommendation on Government Access to Personal Data Held by the Private Sector

- 배경: OECD 디지털경제정책위원회(CDEP)가 “OECD 프라이버시 가이드라인” 을 재검토한 결과, 국경간 자유로운 데이터 흐름에 장애가 되어 해결해야 하는 가장 시급한 과제* 로 “신뢰에 기반한 정부의 민간보유 개인정보 접근” 을 지목하고 상위 원칙 개발을 추진

*2020.7월 유럽 사법 재판소가 미국 정부의 민간 부문 보유 개인 정보 접근 차단을 감시하는 옴부즈맨 제도의 독립성을 문제삼아 EU-미국 간 체결한 프라이버시 쉴드(Privacy Shield)를 무효화한 것을 계기로

본 이슈 부각

- 추진경과 및 향후계획: △ 상위 원칙 초안 마련을 위해 23개 회원국 전문가로 구성된 그룹 결성 (‘21.2월), △ 초안 마련 (’21.4월), △ G7 디지털 기술 장관 선언에 언급 (‘21.4월), △ 이해 관계자 협의 및 수정 (’21.6월), △ OECD 디지털경제정책위원회 (CDEP) 승인 (‘21.7월~8 월중), △ 2021년 10월 각료회의(MCM) 에서 채택을 목표로 진행 중

- 상위 원칙 고려 사항: △개인정보보호와 법집행•국가안보간 균형, △정부의 민간 보유 개인정보 접근에 대한 안전장치 확보, △민주주의 원칙과 법률에 기반, △ 신뢰 기반의 데이터 이동 촉진 등의 측면에서 국가들의 공통점을 강조
- 현재까지 도출된 안전장치 대 주요원칙: △법적 근거, △합법적 목표, △접근 승인, △개인정보처리 제한, △투명성, △독립적 감독, △법적 구제 등으로 구성

그림 14 개인정보보안 장치를 위한 OECD 주요 원칙

다.5 모듈 3: 데이터에 의해 형성되는 기업과 시장 (Data shaping firms and markets)

- 목적: 기업들간 불균형적인 데이터 사용이 비즈니스 모델과 시장 역동성에 미치는 영향을 이해
- 중소기업의 디지털 전환 보고서 발간: Digital Transformation of SMEs(OECD '21.2월)

  • 개요: 디지털화는 중소기업에 성과 개선•혁신•촉진•생산성 향상•대기업과의 공평한 경쟁 등 기회를 제공하지만, 중소기업은 내부 리소스와 인식• 기술• 격차• 재무 문제 등으로 디지털 전환에 뒤처짐
  • 빅데이터 분석 및 클라우드 컴퓨팅: OECD 회원국들의 기업규모별 활용도를 조사한 결과 대중소기업간 격차가 크게 나타남:

  *2017년 기준 OECD 평균: 250인 이상 33.5%, 50인 이상 18.9%, 10인 이상 10.7% (출처: OECD Digital Economy Outlook, '20.11월)

  • 디지털 보안: 일반적으로 중소기업은 디지털 집약도는 낮지만 보안 사고 탐지 능력, 디지털 보안 관행의 미성숙 등 취약성이 높음

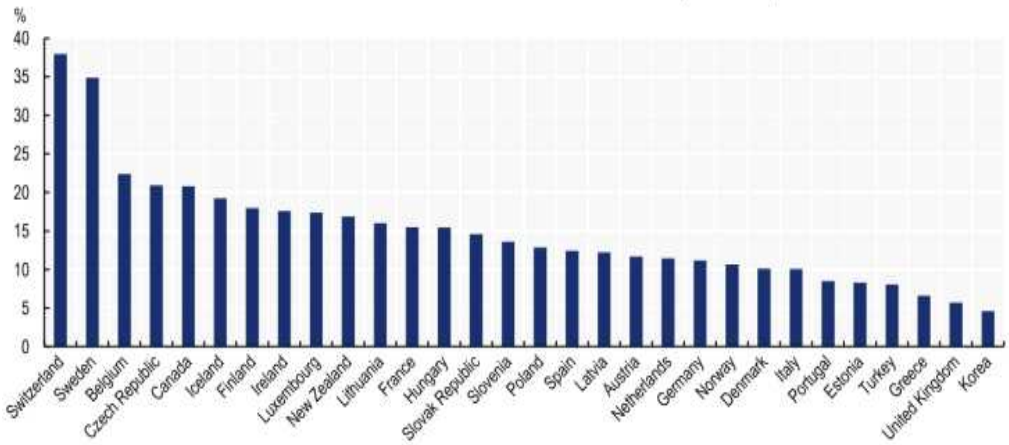

그림 15 10인 이상 기업의 보안사고 경험 비율 년 (2019년)

- 데이터 이동성(Data Portability)과 경쟁: *Data Portability: Analytical Report 초안 (OECD, '21.4월)
    - 데이터 이동성 정의: 자연인• 법인이 자신의 정보를 보유한 데이터 보유자에게 컴퓨터 등 기계가 읽을 수 있는 구조화된 형식으로 특정 제 3 자에게 임시•지속적으로 이전할 것을 요구하는 능력 또는 권리
    - 데이터 이동성 의미: 개인과 기업의 개인정보와 비즈니스 데이터에 대한 통제권을 강화하는 동시에 디지털 서비스 제공자와 플랫폼간 데이터 접근과 공유를 향상시키는 핵심 수단
    - 기회 요인: △시장 경쟁 및 소비자 선택권 증진, △데이터 주도 혁신 촉진, △ 데이터 흐름과 공유 촉진, △ 개인정보 자기결정권 달성
    - 도전 과제: △데이터 이동성 범위에 대한 불확실성, △데이터 디지털 보안과 개인정보보호, △법적 책임성 문제, △표준 및 API 등 상호 운용성, △규정준수 비용, △분야별 법집행 기관간 협력

    *Application Programming Interface: 서버에 데이터를 요구하는 프로토콜

다.6 모듈 4: 데이터 및 데이터 이동 측정 (Measurement of data and data flows)

- 목적: 데이터 및 데이터 이동의 역할• 가치• 경제적 영향 등을 측정할 수 있는 방법과 구체적인 증거 제공
- 현황: 데이터 및 이동과 관련된 더 나은 측정 방법과 지표 마련을 위해 데이터 분류와 다양한 측정 방법을 시도하고 있음

    - 사례 1: 미국 데이터 주도 기업 시가총액 분석, '85년 $88B('20 년 기준 $200B)에서 '20년3 분기 $5T에 육박, 모든 기간에서 이들의 성장 지수가 나스닥과 뉴욕증권거래소의 평균 성장 지수보다 높았음

    * Data-driven Company: 핵심 비지니 대략의 데이터 수집•저장•분석과 연관이 있는 나스닥 44개, 뉴욕증권거래소 32개 기업 (출처: Perspective on the Value of Data and Data Flows, OECD, '20.12)

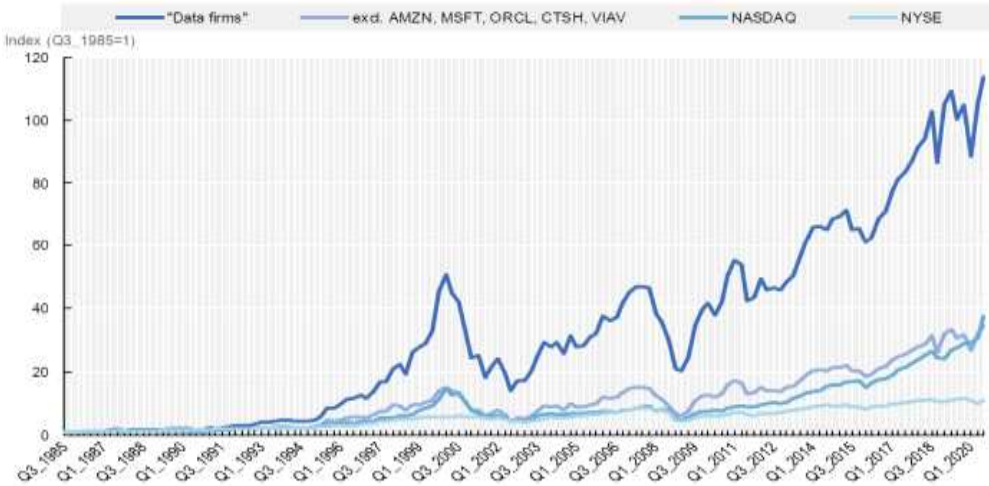

그림 16 데이터 기업들의 시가 총액 평균성장지수 (‘85년 ~’20년)

- 사례2: 벤처캐피탈 투자와 데이터 가치, 규모측면에서는 미국, 중국이 GDP대비 비중으로는 싱가폴 이스라엘 스위스 홍콩 순으로 조사

*Measuring the value of data and data flows: venture capital investments in big data firms (OECD, ‘21.3)

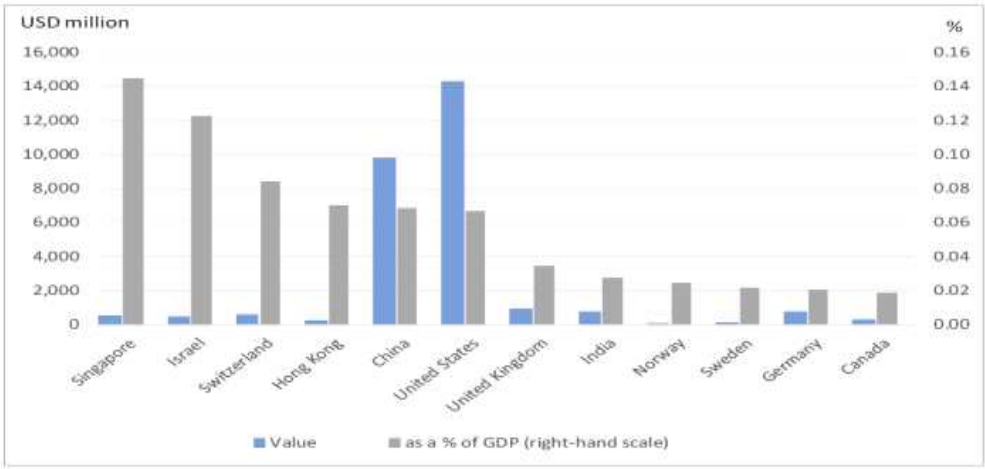

그림 17 국가별 빅데이터 벤처캐피탈(VC) 투자 (‘19 년 기준)

4. 국내동향

가. 데이터 댐 사업 동향

표 10 데이터 댐 사업 동향 키워드

가.1 국내 동향 개요, **가.2** 데이터 댐 사업 개요, **가.3** 빅데이터 플랫폼 구축사업, **가.4** 통합 데이터 지도 서비스 사업, **가.5**. 마이데이터 혁신 사업, **가.6** 국가 바이오 데이터 스테이션 구축 사업

가.1 국내 동향 개요

- 코로나19로 인한 비대면화와 디지털 경제로의 가속화를 위하여 2020년 7월 정부는 관계 부처 합동으로 디지털 뉴딜 사업 발표. 2025년도까지 총 58.2조원 투자. 정부는 우선 데이터댐과 지능형 정부 , 스마트 의료 인프라를 대표 과제로 선정
- 정부는 3개 중점추진 과제를 포함해 디지털 뉴딜 추진 주요 분야로 DNA 생태계 강화, 교육인프라 디지털 전환, 비대면 산업 육성, SOC 디지털화 등을 제시
- DNA 생태계 사업을 강화하기 위해 국민생활과 밀접한 분야에 쓰일 공공 데이터 개방. 1 · 2 · 3차 전 산업으로 5G · AI를 융합하여 확산시키고, 5G · AI 기반의 지능형 정부 및 K-사이버 방역체계 구축. 2025년까지 총사업비 38조5000억원(국비 31조 9000억원)을 투자해 일자리 56만7000개 창출
- DNA 생태계란 데이터(Data), 네트워크(Network), 인공지능(AI)을 결합한 복합 체계.

가.2 데이터 댐 사업 개요

- 디지털 뉴딜 정책의 핵심 사업으로 ‘데이터댐’ 사업 추진. 데이터를 민간에 공유케하여 인공지능(AI) 기술 및 산업 발전을 촉진시킴.
- 데이터 댐이란 ? 광범위한 데이터를 '댐'에 가둬두고 필요한 곳에 사용할 수 있도록 하는 것. 이를 위해 '수로'에 해당되는 네트워크가 필요.

이왕이면 속도가 아주 빠른 5G네트워크가 활용돼야 함. 그리고 데이터를 수집하거나 데이터가 소비되는 끝단에서 최적의 활용을 위해 인공지능(AI)과 융합시킴.

- 데이터 댐 구축은 빅데이터 플랫폼, 데이터 스토어, 공공데이터 포털 등 관련 산업의 부흥케하고 데이터를 매매하고 이를 활용해 새로운 서비스 산업 창출이 가능함. 예를 들자면 개인의 취향을 저격하는 맞춤형 숙박, 의료, 교육, 외식, 모빌리티 서비스들이 '내 데이터'를 기반으로 내 주변에 공기처럼 존재하게 하는 것임.
- 과학기술정보통신부(이하 과기정통부)와 한국지능정보사회진흥원(NIA)은 인공지능(AI) 학습용 데이터 170종, 4억8000만건을 'AI 허브'를 통해 개방.
- 공개하는 AI 학습용 데이터는 △음성·자연어(한국어 방언 등 39종) △헬스케어(암진단 영상 등 32종) △자율주행(도로주행영상 등 21종) △비전(스포츠 동작 영상 등 15종) △국토환경(산림수종 이미지 등 12종) △농축수산(가축행동 영상 등 14종) △안전(노후 시설물 이미지 등 19종) △기타(패션상품 이미지 등 18종) 등.

가.3 빅데이터 플랫폼 구축사업

- 디지털 뉴딜('20.7월 발표) 핵심 프로젝트인 '데이터 댐'의 일환으로 구축한 빅데이터 플랫폼 사업 추진. 빅데이터 플랫폼 구축사업은 과기정통부에서 산업 전반에 필요한 양질의 데이터를 공급하기 위해 데이터 수집·가공·분석·유통 등 전주기를 지원하는 사업.
- '19년도에는 금융, 환경, 문화, 교통, 헬스케어, 유통, 통신, 중소기업, 지역경제, 산림, '20년도에는 농식품, 해양수산, 소방안전, 스마트치안, 라이프로그, 디지털 산업혁신 분야 지원.
- 민간·공공기관의 빅데이터 플랫폼을 통해 협력하여 '20년말 기준 3,246종의 데이터를 축적하였고, 이용자들의 활용실적은 83,336건으로 '19년(2,942건)에 비해 큰 폭의 상승세(28.3배)를 보임.

가.4 통합 데이터 지도 서비스 사업

- 과학기술정보통신부(장관 최기영, 이하 '과기정통부')와 한국지능정보사회진흥원(원장 문용식)은 데이터 댐에 축적된 데이터가 효과적으로 검

색·활용될 수 있도록 기능을 확대·개편한 통합 데이터지도 서비스 제공 (통합 데이터지도 주소 : www.bigdata-map.kr)

- 통합 데이터지도는 과기정통부가 구축한 기존 빅데이터 플랫폼 외에도 국내 다른 데이터 플랫폼과의 연계를 확대하여 더욱 다양한 데이터의 소재 정보를 제공함. 인공지능 학습용 데이터를 축적한 인공지능 허브, 데이터 오픈마켓인 데이터스토어 뿐 아니라 대한무역투자진흥공사(KOTRA)의 무역·투자 빅데이터도 통합 데이터지도와 연계하여, 이용자가 약 29만 건의 데이터정보를 쉽게 검색하고 활용할 수 있을 것임.

가.5. 마이데이터 혁신 사업

- 과기정통부는 `19년부터 금융상품 추천서비스, 응급 상황 시 진료기록 공유서비스, 소상공인 문서·자금 플랫폼 서비스 등 다양한 분야에서 혁신적인 마이데이터 서비스를 발굴·지원하여 전 영역으로 확산하기 위한 마이데이터 실증 사업 추진.
- 현재까지 총 97개 기관이 참여하였으며, 9개 분야*에서 대형병원, 금융사, 통신사 등 데이터 보유기관과 혁신서비스를 개발하는 활용기관을 연계하여 17개 서비스 개발을 지원함. ※ `21년 마이데이터 사업 예산 : 100억원.

※ 마이데이터(MyData)란 정보주체를 중심으로 산재된 개인데이터를 한 곳에 모아 개인이 직접 열람·저장하는 등 통합·관리하고, 이를 활용하여 개인을 위한 맞춤형 서비스를 제공하는 것.

가.6 국가 바이오 데이터 스테이션 구축 사업

- 사업 개요
  - (사업목적) 부처·사업·연구자별로 흩어져 있는 국가 바이오 연구 데이터의 체계적 통합 수집·관리·제공 시스템 구축
  - (사업기간/예산) '21~계속/ '21년 관계 부처 총 442억원

  ※ 과기정통부 410억원, 질병청 25억원, 농진청 7억원(다부처 국가생명연구자원 선진화 사업)

  - (사업범위) 국가 바이오 R&D를 통해 생산·활용되는 모든 데이터*(제3자 제공 동의를 받은 익명화된 임상 데이터 등 포함). 생화학분석, 이

미지(영상), 임상 및 전임상, 분자구조, 표현형 정보, 화학반응 등

- 주요내용
  - (수집·품질관리) 부처·연구자별 데이터의 체계적 수집을 위해 등록 양식 표준화 및 데이터의 정제 등을 통한 품질 관리
  - (임상정보) 바이오 R&D 사업에서 생산되는 임상 연구 데이터는 임상유전체생명정보시스템(CODA)과 데이터 스테이션을 연계하여 수집·제공
  - (분석·활용) 빅데이터 시각화, 분석도구 등 데이터 활용 서비스 제공
  - (인프라 성장지원) 데이터 사이언티스트 등 데이터 활용 인력양성 및 국제협력

- 향후 계획
  - '21.10 국가 바이오 데이터 스테이션 시범 운영(데이터 등록, 검색, 분석)
  - '22.~ 데이터 분석·활용 지원 및 인력양성

나. 박데이터 활용 지원 사업 동향

표 11 빅데이터 활용 지원 사업 동향 키워드

나.1 데이터 활용 지원 사업 개요, **나.2.** 데이터 바우처사업, **나.3** 데이터 플래그십 사업, **나.4.** 중소기업 빅데이터 활용 지원 사업, **나.5** 인공 지능 고성능 컴퓨팅 지원 사업, **나.6.** 인공 지능 중심지 (AIHub) 구축 지원 사업, **나.7.** K-HUB(인공지능 혁신허브) 구축사업 지원

나.1 데이터 활용 지원 사업 개요

- 디지털 뉴딜의 핵심 사업 데이터 댐을 통하여 축적된 양질의 데이터를 활용하는 사업을 지원하는 사업 추진. 데이터 바우처 사업, 데이터 플래그십 사업, 중소기업 빅데이터 활용 사업 추진.

나.2. 데이터 바우처사업

- 2019년부터 추진해온 데이터 바우처 지원사업은 중소기업, 소상공인, 등이 데이터를 활용하여 비즈니스 혁신 및 신제품 서비스 개발시 필요한 데이터상품 또는 가공 서비스 비용을 지원하는 사업. 2021년도에 총 2580개의 바우처 (총 예산 1230억원)을 지원함.
- 2012년도 부터는 다양한 분야에서 데이터 활용을 통한 변화와 성장의 기회를 제공하고자 16개 정부부처 및 소속 전문 기관 뿐만 아니라 지자체와 협력을 확대하여 수요를 발굴하고 우수 사례를 창출하는 한편 그 결과물을 확산함

나.3 데이터 플래그 십 사업

- 데이터 플래그십 사업은 빅데이터와 AI 기술을 활용하여 사회 현안 해결 및 비즈니스 혁신을 위한 제품 및 서비스 개발을 지원하는 사업. 총 12개 과제를 지원. 공공기관, 민간 기업 구분없이 지원 (총 예산 60억원)
- 2013년도부터 시작한 데이터 플래그십 사업의 주요 우수 성과로는 서울시 심야 버스 노선 수립, 실종자 과학수사 지원, 코로나 19 역학 조사지원시스템 고도화 등

나.4. 중소기업 빅데이터 활용 지원 사업

- 총 예산 10억으로 4개 지자체 지원 지자체별 20개기업 지원. 사업 내용은 중소기업의 빅데이터 활용 방법 및 데이터 수집, 가공, 분석 역량 등 컨설팅 교육.

나.5 인공 지능 고성능 컴퓨팅 지원 사업

- 사업 목적: 대용량 컴퓨팅 환경 구성이 어려운 중소 및 벤처기업 등이 막대한 계산량이 필요한 AI 학습에 병렬 연산 장치를 이용하여 신속히 제품 및 서비스를 연구개발 할 수 있도록 인공 지능에 특화된 고성능 컴퓨팅 자원 제공 (aihub.or.kr 참조)
- 2021년에는 고성능 컴퓨팅 사업 공급 업체로 네이버 클라우드 (Naver Cloud)의 GPU 서버를 선정하여 연중 상시 지원 체제 구축

- 2020년의 경우 883개 중소, 벳처기업, 연구소, 대학 등을 지원하여 인공지능 기술 개발기간 평균 3.9개월 단축, 월 평균 5백만원 비용 절감 확인
- 2021년도에는 1000개 기업,기관 이상 지원할 예정이고 지원 용량도 2020년도 월 상시 25.7 PF에서 월 상시 30.6 PF로 확장함.
- 제공 GPU 서버 예시: CPU (서버용 2.1GHz), 메모리 (128GB), 스토리지 (2TB), 회선 속도 (1bps) 로 10TF, 20TF, 40TF, 80TF로 사용자에 따라서 구분하여 지원

나.6. 인공 지능 중심지 (AIHub) 구축 지원 사업

- 사업 개요
  - 과학기술정보통신부(장관 임혜숙, 이하 '과기정통부')와 한국지능정보사회진흥원(원장 문용식, 이하 '지능정보원')은 국내 인공지능 기술 및 산업 발전의 촉진제가 될 인공지능 학습용 데이터 170종(4억 8천만 건)을 인공지능) 중심지(aihub.or.kr)를 통해 6월 18일부터 개방.
  - 과기정통부는 '17년부터 기업, 연구자, 개인 등이 시간 및 비용 문제로 개별 구축하기 어려운 인공지능 학습용 데이터를 구축·개방해왔으며, '20년부터는 디지털 뉴딜 '데이터 댐' 구축 프로젝트의 일환으로 구축 규모를 대폭 확대하여 추진.

- 사업 내용(성과)
  - 그동안('17~'19년) 21종의 인공지능 학습용 데이터를 구축·개방하여 인공지능 중심지 이용자와 데이터 활용이 빠르게 증가하고 있으며, 개방 데이터를 활용한 인공지능 서비스 개발 및 성능 향상 등의 성과 창출.

  ※ (이용자) '19년 4,439명 → '20년 13,092명(누적), (활용) '19년 16,177회 → '20년 49,085회(누적)

  - 인공 지능 학습용 데이터 구축사업과 일자리 창출사업과 연계하기위하여 대중 참여 제작 기업 (크라우드웍스 라벨러, 메타버스에서의 노동자 역할 수행 등) 들과 협력하여 라벨러와같은 대규모참여형 노동자의 처우개선 적정 보수 지급 교육기회 제공 등 협력 추진
  - 인공지능 학습용 데이터(170종) 목록.

나.7. K-HUB(인공지능 혁신허브) 구축사업 지원

- 과학기술정보통신부(장관 임혜숙, 이하 '과기정통부')는 인공지능 혁신허브 사업을 수행할 기관으로 고려대학교 주관의 'K-Hub 그랜드 컨소시엄'을 최종 선정함.
- 인공지능 혁신허브 사업은 국내 대학·기업·연구소 등에서 축적되고 있는 인공지능 연구 역량을 모으고, 개방형 인공지능 연구 환경을 조성하여 세계적 수준의 경쟁력을 갖춘 국가 연구체계를 구축함
- 'K-Hub 그랜드 컨소시엄'에는 국내·외 57개 대학, 119개 기업, 18개 연구소 등 총 215개 기관이 참여함. 대학은 고려대학교를 포함한 12개 대학이 2개의 연구그룹*으로 나뉘어 연구를 주도하고, 강원대, 동국대 등 국내 14개 대학 및 매사추세츠공과대(MIT), 카네기멜론대(CMU) 등 해외 31개 대학이 협력.

  ※ (연구그룹1) 고려대, KAIST, 포항공대, 연세대, 한양대, 경북대, (연구그룹2) 서울대, 이화여대, 경희대, 성균관대, UNIST, 전남대. 아울러, 국내 대·중소·벤처 기업 102개와 구글·페이스북 등 17개 해외 기업, ETRI·KIST·KETI 등 국내 주요 출연연 7개와 프라운호퍼·막스플랑크연구소 등 해외 연구소 11개도 협력할 예정임.

- K-Hub 그랜드 컨소시엄은 향후 최대 5년 간 약 445억원을 지원받아 ①인공지능 분야 고난도·도전형 연구, ②개방형 연구 환경 조성 ③최고 수준의 인재양성 등 국가 인공지능 역량 강화 및 인재양성의 구심점 역할 수행.
- 두 개의 연구 그룹은 핵심원천 기술 확보를 위한 초거대 인공지능 개발과 로봇, 드론, 모바일, 의료, 에너지 등 분야별 특화 인공지능 연구를 주요 방향으로 정하고, 각 그룹이 세부 주제에 대한 연구를 수행.
- 대규모 연구 프로젝트 수행에 필요한 연구 환경을 조성하기 위해 최대 1,000명이 동시에 이용 가능한 규모인 50페타플롭스*(PF) 규모의 컴퓨팅 센터를 구축하고, KT, 네이버 등 기업에서 보유하고 있는 다양한 컴퓨팅 인프라를 연계하여 컨소시엄 참여 대학 및 기업의 연구수행에 제공.

※ 1 페타플롭스는 초당 1천조회 이상의 연산을 수행한다는 의미로, 전 지구의 인구가(약 76억명)이 각각 1초동안 13만회 이상의 계산을 수행하는 것과 유사

다. 국내 데이터 산업 동향

표 12 국내 데이터 산업 동향 키워드

다.1 국내 데이터 산업 현황, **다.2** 데이터 3법 시행

다.1 국내 데이터 산업 현황

- 분류 체계 : 데이터 관련 비즈니스를 대분류 3개, 중분류 10개로 구분해 조사
  - 데이터 처리 및 관리 솔루션 개발·공급업(약어: 데이터 솔루션): 데이터 수집·연계 솔루션 개발·공급업, DBMS 솔루션 개발·공급업, 데이터 분석 솔루션 개발·공급업, 데이터 관리 솔루션 개발·공급업, 데이터 보안 솔루션 개발·공급업, 빅데이터 통합 플랫폼 솔루션 개발·공급업 (6개)
  - 데이터 구축 및 컨설팅 서비스업(약어: 데이터 구축/컨설팅): 데이터 구축·가공 서비스업, 데이터 관련 컨설팅 서비스업 (2개)
  - 데이터 판매 및 제공 서비스업(약어: 데이터 서비스): 데이터 판매·중개 서비스업, 정보제공 서비스업 (2개)

- 국내 데이터 산업 규모 (www.kdata.or.kr 참조)
  - 2020년(E) 전체 시장규모는 19조 2,736억원, 전년 대비 2조 4,154억원 증가(14.3%↑)

  * 전체 시장규모 = 데이터 관련 직접매출 + 간접매출(광고비, 시스템운영관리 매출)

* 최근 3년간('18년~'20년(E)) 연평균 성장률은 11.3%로 높은 성장 예상

- 2020년(E) 직접 매출규모는 11조 4,350억원, 전년 대비 1조 4,602억원 증가(14.6%↑)
- 대분류로 보면, ▲ '데이터 판매 및 제공 서비스업'이 9조 3,752억원으로 가장 큰 매출 비중(48.6%)을 차지했으며, ▲ '데이터 구축·컨설팅 서비스업'이 7조 4,361억원(38.6%), ▲ '데이터 처리·관리 솔루션 개발·공급업'이 2조 4,623억원(12.8%) 순으로 나타남.
- 특히, '분석솔루션(38.0% ↑)', '보안솔루션(26.3% ↑)' 매출 증가 등으로 인해 '데이터 처리·관리 솔루션 개발·공급업(11.8% ↑)'의 '19년 대비 증가율이 가장 높을 것으로 나타남.
- 중분류별로는 ▲ '데이터 판매·중개 서비스업(41.9% ↑)', ▲'데이터 분석 솔루션 개발·공급업(38.0% ↑)', ▲ '데이터 관련 컨설팅 서비스업(31.5% ↑)' 순으로 전년대비 크게 성장함.

그림 18 국내 데이터산업 직접 매출 시장 규모

- 국내 데이터 직무 인력 규모
  - 데이터산업 내 데이터 직무인력은 10만 1,967명으로, 전년대비 12,909명 증가(14.5%↑)

* 직무별 인력 수는 ‘데이터 개발자’ 3만 4,085명(33.4%), ‘데이터 엔지니어’ 1만 6,450명(16.1%)으로, 두개 직무 인력이 전체의 약 50%를 차지

* 향후 5년간(‘21~’25년) 직무별 인력수요는 데이터 과학자(31.4%)가 가장 높고, 데이터 분석가·데이터 개발자, 데이터 컨설턴트 순

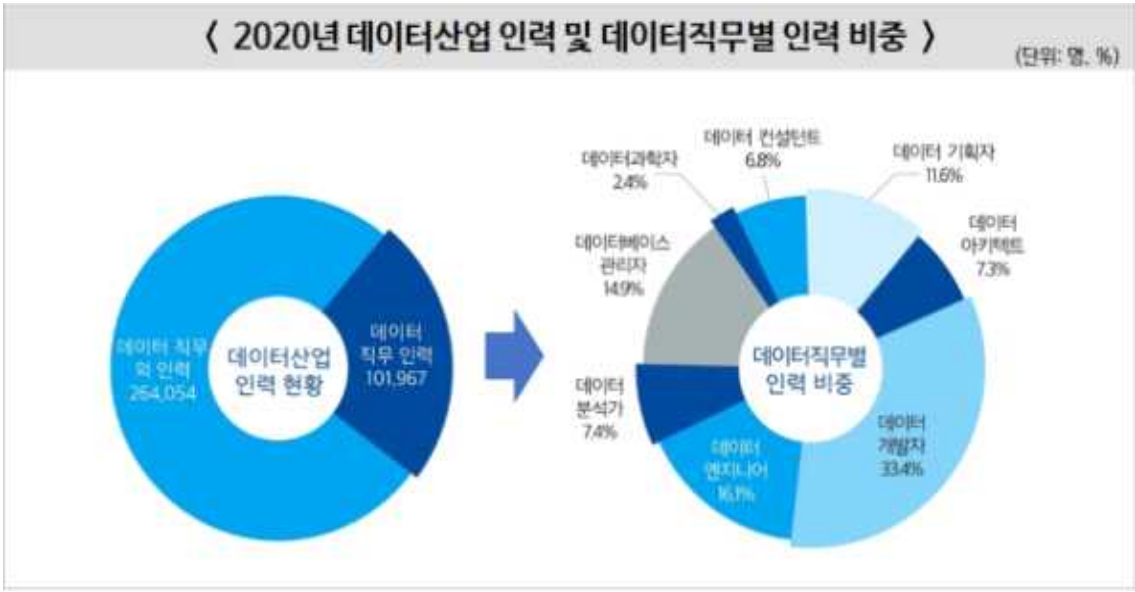

그림 19 2020년 데이터산업 인력 및 데이터직무별 인력비중

다.2 데이터 3법 시행

- 데이터 경제 활성화를 이끌 데이터3법(개인정보보호법·신용정보법·정보통신망법)이 2020년도 8월 5일 시행된 지 1년이 지남. 데이터3법 시행으로 부처급 기관 개인정보보호위원회가 출범하고 가명정보 활용, 마이데이터 등 산업 활성화 기반이 마련됨. 코로나19 상황을 타개하기 위한 정부 디지털 뉴딜 사업이 더해지면서 다양한 데이터 관련 투자와 사업이 시행됨.
- 정부는 ‘데이터를 가장 잘 쓰는 나라’를 목표로 데이터 3법을 개정함, 가명정보 제도를 통해 과거 불가능했던 빅데이터 분석을 활용하여 혁신

적인 제품과 서비스를 창출하고, ‘데이터 경제’ 시대를 앞당김.

- 이를 위하여 복잡한 결합절차, 인프라 부족, 개인정보 유출 우려 등을 가명정보 활용을 주저하게 되는 요인을 줄이는 가명정보제도 추진. 가명정보 제도는 정보주체 동의 없이 과학적 연구 등을 목적으로 식별가능성을 최소화하여 정보처리자가 직접 활용하거나 제3자에게 제공하는 제도임.

  ※ 가명정보 활용 시범사업: ① 암 질병 치료효과 분석, ② 암 환자의 합병증 및 만성질환 예측 연구, ③ 국가보훈대상자신용실태 연구, ④ 노후소득보장 종합연구, ⑤ 불법스팸 실태연구, ⑥ 지역별, 상권별, 상품별 소비패턴 분석, ⑦ 맞춤형 산림치유 프로그램 분석

5. 결언

가. 데이터 센터 인프라 동향 특징

- 대형 연구 장비의 초대형화에 따른 실험 데이터의 급속한 증가 그리고 이런 초대용량 데이터 분석을 위한 인공 지능 및 머신 런닝 기술의 연구 현장 도입은 데이터센터로 하여금 인프라 구축에 있어서 스토리지 확장을 최우선 순위로 추진케하고 있음. 이런 배경으로 해외 유명 데이터 센터들이 경쟁적으로 스토리지 인프라 구축에 2021년도 인프라 예산을 투입하고 있었음.
- 그리고 스토리지 인프라 확장에 있어서 특이점은 일반적인 스토리지 저장 용량의 증가뿐만 아니라 대용량의 데이터를 고속으로 처리하기 위하여 선진 데이터 센터들이 기존의 하드 디스크 중심의 스토리지 인프라가 아닌 반도체 메모리 중심의 스토리 인프라로 확장을 추진하고 있음. 아직도 하드 디스크 방식의 스토리지 보다 고비용임에도 불구하고, 신규 도입 스토리지 장비를 반도체 방식의 스토리지로 구입하여 스토리지 인프라를 구축하고 있음
- 해외 데이터 센터들은 스토리지 인프라 확장 속도가 대형 연구장비 실험 데이터의 증가 속도를 따라잡지 못할 것을 인식하면서 스토리지 인프라를 보다 더 마노은 연구자들이 쉽게 공동 활용 할 수 있도록 스토리지 가상화를 데이터센터 인프라 고도화 방향의 하나로 계획을 수립하고 추진하고 있다는 것을 파악함

나. 미국 인프라 지원 사업의 동향 특징

- 미국 과학재단(NSF)에서는 실험 데이터의 대용량화에 따라 데이터 센터의 스토리지 인프라 확장 지원에 있어서 크게 두가지 정책을 추진하고 있음을 파악할 수 있었음
- 첫째는 상용 솔루션을 이용한 데이터 센터 인프라 확장 및 고도화 사업 지원임. 이 방식은 세계적인 거대 공룡 기업들이 기부 형태로 자사의 장비와 예산을 정부와 공동으로 지원하여 구축하는 사업 방식임. 기존

의 미국의 주요 선도 연구 지원 데이터센터의 인프라 구축 방식임

- 두 번째는 데이터센터들이 정부와 기업의 지원에도 불구하고 데이터의 증가가 매우 크고 빠르기 때문에 이를 충분히 감당할 수 없다고 판단하여 상용 솔루션을 역으로 최소화하고, 그 대신에 오픈 소스와 저가의 장비를 최대한 활용하여, 급증한 실험데이터를 처리할 수 있는 새로운 인프라 구축을 관련 글로벌 연구자 그룹들과 공동으로 플랫폼을 개발하고 구축 운영하려고 하는 것임.

다. 스토리지 산업체의 기술 동향 특징

- 스토리지 산업체에서는 실험 데이터의 급증으로 인하여 향후 실험 데이터 관련 데이터 센터 인프라의 수요 동향이 크게 두가지 방향으로 진행될 것으로 예측하고 이를 준비하고 있음
- 첫 번째는 하드웨어 측면으로서, 반도체 메모리가 점차 하드디스크 스토리지 인프라를 점점 대체할 것으로 예측함. 따라서 이러한 반도체 메모리와 서버 등 IT 연구 장비와 연구자를 고속으로 연동하기 위하여 이들 장비간 인터페이스의 성능을 고성능화하는 기술 및 제품을 개발 출시하고 있음
- 두 번째는 부족한 스토리지 인프라를 모든 사람들이 쉽게 언제든지 활용할 수 있도록 가상화 기술 단계를 더 높은 수준으로 개발 적용하는 것임. 하지만, 아직 이에 관한 표준화 기술이 빠르게 뒤받침을 하고 있지 못해서 선진 데이터 센터들은 특정 기업체의 기술 및 제품에 종속되는 위험을 무릎쓰고 해당 제품을 선택할지 아니면 자체적으로 유지보수 인력 소요가 많이 드는 오픈 소스로의 인프라 방향을 선택할 지는 아직은 케이스 바이 케이스 형태로 보고되고 있어서 대세를 읽기는 어려웠음.

라. GSDC 인프라 구축 방향에 대한 고려 사항

- 앞에서 언급한 선진 데이터 센터에서의 동향과 같이 우리도 여건이 허용되는 범위 내에서 큰 흐름에 동참할 준비를 하여야 한다고 판단함,

빠르게 다가올 제터 바이트 데이터 컴퓨팅시대를 대비하여 지금부터라도 인프라 확장에서 스토리지 인프라 확장에 우선 순위를 현재 보다 조금 더 높여 줄 것을 제안함.

- 이는 단순히 대형장비의 초대형화에 따른 실험 데이터의 폭발적인 증가를 대비하는 것 외에도, 인공 지능 및 머신 런닝으로 인한 대용량 실험 데이터에서 우수한 첨단 연구 결과를 산출할 수 있게 됨에 따른 AI 기반 실험데이터 처리가 많은 연구자들이 선호하는 새로운 연구방식으로 자리매김을 하고 있어서 이에 대한 대비를 하여야 한다고 판단하기 때문임.
- 그리고 단순 저장 방식의 기존 스토리지 인프라 구축에서 진화하여 이제는 대용량 실험 데이터 및 인공지능 및 머신 런닝 기술로 무장한 연구자들의 협업 연구 플랫폼으로 인식될 수 있는 지능형 인프라-플랫폼으로 인식을 전환할 시기가 도래하였다고 판단함.
- 따라서 스토리지 인프라의 고성능화를 가상화와 병행하여 추진하여야 하는데, 추진 방법은 점진적으로 반도체 기반 스토리지 인프라 확충 및 오픈 소스기반 스토리지 인프라의 가상화를 지금부터 조금씩 준비할 것을 제언함.

6. 참고문헌

- CERN Site Report, https://indico.cern.ch/event/995485/contributions/4256456/attachments/2207924/3736138/CERN%20Site%20Report%20-%20HEPiX %20Spring %202021.pdf
- RAL Site Report, https://indico.cern.ch/event/995485/contributions/4263427/attachments/2207923/3736135/2021-03%20-%20HEPiX%20Spring%202021%20-%20RAL%20Site%20Report.pdf
- KEK Site Report, https://indico.cern.ch/event/995485/contributions/4259752/attachments/2207802/3741713/hepix-spring-2021-kek-site-report-iwai-with-script-rev6.pdf
- The Economic and Fiscal Impacts of Fermilab on Illinois, Dec. 6, 2019
- Annual Report 2018, KEK, March, 2019
- SuperKEKB MR Status and Plans, Belle II Physics Week, Oct. 28, 2019
- Grid_Ka KIT Site Report, https://indico.cern.ch/event/637013/contributions/2739322/attachments/1541598/2417809/kit-site-report-hepix-kek-20171017.pdf
- Fermi Lab Site report, https://indico.cern.ch/event/765497/contributions/3366608/attachments/1819221/2974778/Jayatilaka-FNAL-HEPiX.pdf
- CERN Annual Report 2018
- CERNAnnuak Report 2019
- Big Data Regional Innovation Hubs (BD Hubs), https://beta.nsf.gov/funding/opportunities/big-data-regional-innovation-hubs-bd-hubs
- Bi Data Regional Innovation Hubs and Spokes Workshop,

https://www.nsf.gov/cise/bdspokes/index.jsp
- Midwest BigData HUB, https://midwestbigdatahub.org/
- Northeast BigData, https://nebigdatahub.org/
- South Big Data Innovation Hub, https://southbigdatahub.org/
- West Big Data Innovation Hub, https://westbigdatahub.org/
- Entering a new phase in the Nig Data Innovation Ecosystem, MidWest Big data Hub, Oct. 29, 2019
- Open Storage Network, Midwest Big Data Hub All Hands Meeting, Oct. 20-30, 2019
- The Data eXaCell - DXC, Pittsburgh Supercomputing Center, May. 17, 2016
- Data eXaCell, Pittsburgh Supercomputing Center, NSF Award ACI-1261721
- Northeast Storage Exchange, Annual Report to the National Science Foundation, 2020
- Big Data Regional Innovation Hubs (BD Hubs), NSF 18-598
- Directorate for computer and Information and engineering (CISE)
- Buliding the Open Storage Network, Alex Szalay, The Johns Hopkins University
- The Open Storage Network: Distributed Storage Cyberinfrastructure for Data-Driven Science, https://www.openstoragenetwork.org/
- OSN UserGuid
- Ceph at RAL, HEPiXSpring Workshop 2021, STFC UK Research and Innovation, 15-19 March 2021
- OECD Going Digital 3 단계, 주 OECD 대표부, 윤두희 참사관
- HPCwire, https://www.hpcwire.com/
- InsideHPC, https://insidehpc.com/
- 과학기술정보통신부 보도자료, https://www.msit.go.kr/index.do

7. 부록 1: 데이터센터 인프라 기술 약어집(알파벳 순)

표 13 스토리지 인프라 기술 약어집(알파벳 순)

aCT	ARC(Advanced Resource Connector) Control Tower
ActiveMQ	open source ActiveMQ, 아파치 오픈소스 메세징
ADMIRE	Adaptive multi-tier intelligent data manager for Exascale (EU 프로젝트)
AFS	Andrew File System
AOD	Analysis Object Data
A-MQ	Red Hat A-MQ, 유연한 메시지 플랫폼, 실시간 통합지원 및 IoT연결 구현
ARC-CE	Advanced Resource Connector-Compute Element, Grid front-end
CAST	The CERN Axion Solar Telescope
CASTOR	The CERN Advanced STORage manager.계층적 스토리지(i.e. HDD and tape)관리 시스템
CASTOR HSM	HSM (Hierarchical Storage management(ie: disk, tape))
CBPF	Centro Brasileiro de Pesquisas Físicas , The Brazilian Center for Research in Physics
CDS	CERN Document Server
Ceph	free-software storage platform. object storage for end-user applications, block storage for virtual machines.
Ceph Nautilus	Ceph 버전 14.2.8의 이름, Nautilus 는 앵무조개 라는 뜻을 갖고 있음
CERNBox	provides a cloud synchronisation service for all CERN users between personal devices (laptops, smartphone, tablets) and a centrally- managed data storage. CERNBox is built on top of the ownCloud S/W.
CI/CD	GitLab CD/CI 참조
CMDB	Configuration Management Data Base. IT 서비스 운영에 필요한 절차, 문서, 소스, 구성, 토폴로지, 어플리케이션 등을 위한 데이터 저장소
colledtd	*"System and applications metrics collector"*. collectd gathers statistics about the system
CRIC	*Computing Resource Information Catalogue*
CTA	Consolidated Tape Association , oversees the dissemination of real-time trade and quote information in New York Stock Exchange
CVMFS	CernVM File System (CernVM-FS) is a network file system based on HTTP and optimized to deliver experiment software in a fast, scalable, and reliable way
DAQ	Data AcQuisition
DCIM	OpenDCIM: a free, web based Data Center Infrastructure Management application
DDN	Data Direct Network, 출처 KIT GridKa, 기업 명칭
DDP	Dynamic Disk Pool, NetApp사 스토리지 시스템 지원 기능

DESY	the Deutsches Elektronen-Synchrotron
DIMM	Dual In Line Memory Module
Documentation	CERN Authoring
DOMA	Data Organization, Management and Acces
Dovecot	Dovecot is an IMAP server for Linux/UNIX-like systems, written with security primarily in mind
Doodle	the simplest way to schedule meetings to clients, colleagues, or friends. Find the best time for one-to-ones and team
DRAM	Dynamic random access memory
DRYAD	Dryad is a leader in data curation and data publishing.
DTO	Data Transfer Orchestrator
DUNE	Deep Under ground Neutrino Experiment, Detector for neutrino, hosted by FNAL
DWPD	Drive Writes Per Day, 수치로 드라이브 TBW 사양을 사용하여 보증 기간 (SSD 내구성)
Echo	Erasure-Coded High-throughput Object store, 영국 RAL과 STFC site 에 구축한 a large Ceph cluster
EDR	IB에서 Extended Data Rate 이면 64비트를 66비트로 보내는 부호 변환을 나타내며 속도는 4 x EDR (25 Gbps)= 100Gbps 임
E-FAX	E-Fax Service allows CERN users to send and receive faxes as emails (가입 필요), TONE에서는 지원이 않됨
E-PLMN	Equivalent Public Land Mobile Network
EOS	energie ouest suisse (runs some powerlines). EOS is a disk-based,low- latency storage service. Having a highly-scalable hierarchical name- space, and with data accesspossible by the XROOT protocol, it was initially used for physics data storage. Today, EOS provides storage for both physics and user usecases. Instances of EOS include EOS USER, EOS PUBLIC, EOS ATLAS, EOSCMS.
EOS	Earth Observation System
Elasticsearch	루씬 기반의 검색 엔진. HTTP 웹 인터페이스와 스키마에서 자유로운 JSON 문서와 함께 분산 멀티테넌트 지원 전문 검색 엔진을 제공
ETF	Experiment Test Franework
FDR	INfiniBand 에서 사용하는 속도 용어로서 Fourteeen Data Rate로 14Gbps를 나타냄
FreeIPA	Redhat 에서 지원하는 무료 오픈 소스 ID 관리 시스템
FTS	File Transfer Service
EOS FUSE	FUSE client allows to access EOS as a mounted file system.
GDB	(WLCG) Grid Deployment Board
GHI	Global Horizontal Irradiance
GitLab CD/CI	The GitLab built-in Continuous Integration, Continuous Deployment, and Continuous Delivery toolset to build, test, and deploy …
GPFS	General Parallel File System. IBM 사의 고성능 공유 디스크 파일

	시스템
HAProxy	HAPROXY는 기존의 하드웨어 스위치를 대체하는 소트우에어 로드밸런서로 네트워크 스위치에서 제공하는 L4, L7 기능 및 로드밸런서 기능 제공
Harbor	Enterprise Container Registry
Helm Charts	Helm Charts help you define, install, and upgrade even the most complex Kubernetes application
HPSS	IBM high-performance storage saver
HSE	Health & Safety and Environmental Protection
HSM	Hierarchical Storage management
IAM	Identity and Access Management
Indico	powerful event (meetings, conferences, lectures) management system developed at CERN
IB	InfiniBand , 쌍방향 시리얼 버스를 이용하는 통신 방식
ICEPP	the International Center for Elementary Particle Physics
IDM	Identity Management
IPMI	Intelligent Platform Management Interface,표준화된 메시지 기반의 하드웨어 관리 인터페이스
IRONIC	Ironic is an open source project that fully manages bare metal infrastructure, Bare Metal as a Service
iRODS	The Integrated Rule-Oriented Data System
ITIL	IT Infrastructure Library. IT 서비스 관리에 대한 프레임워크 구현을 돕기 위한 문서들의 집합
IMAP	Internet Messaging Access Protocol
InvenioILS	Invenio App ILS. Integrated Library System based on Invenio
Invenio	Invenio는 대규모 디지털 리포지토리를위한 오픈 소스 소프트웨어 프레임 워크. 기관 리포지토리 및 연구 데이터 관리 시스템에서 디지털 자산 관리 도구 제공
IOPS	Input/Output Operations Per Second, HDD, SSD, SAN 같은 컴퓨터 저장 장치를 벤치마크하는 데 사용되는 성능 측정 단위
JBOD	Just a Bunch Of Disks
Jenkins	사용자 보다 더 빠르고 더 강력하게 더 쉽게 빌드(Build), 테스트, 그리고 배포(deployment) 도구 등 체인 전체를 통합할 수 있는 방법 제공
Keycloak	an open source identity and access management solution, 시험 중인 New SSO 서비스. authentication server. REST API 및 clients 구성 지원
KLI	Key Performance Indicator
Kopano	CERN Mail Server, Malt-Mail
KRB5	Kerberos V
Lemon	CERN Legacy Tool, Lemon Web Monitoring, CERN Cluster monitoring Project
LFC	LCG File Catalog
LHC	Large Hadron Collider

LSDF	Large Scale Data Facility (독일 Gridka 에서 사용)
LSF	Load Sharing Facility
LTO	Linear Tape Open. LTO 컨소시엄 기술의 표준 폼 팩터는 Ultrium이라는 이름과 함께하며 8세대 LTO Ultrium은 2018년 출시. 동일 크기의 카트리지에 12 TB 보관.
LXPLUS	Linux Public Login User Service
MAlt	Microsoft Alternatives project, 2018년부터 S/W 라이센스 비용 절감 준비 프로젝트. Microsoft, 등의 유료 S/W에서 open-source solutions로 전환 지원
Markdown	Markdown(file.md)-based documentation
Mattermost	CERN의 Online Chat Service
MC/SC scheduling	Multi Core/Single Core Scheduling
MDS	ceph Meta Data Servers
MLC	Multi level Cell in SSD
multitenency	웹을 통해 D/B 정보를 관리 및 공유 지원. SaaS 플랫폼 구현을 위한 핵심 기능 수행. 하이퍼바이저를 기반으로한 가상 컴퓨터(VM) 기술을 이용하여 H/W 공유
MONIT	The Unified Monitoring Architecture for Monitoring the CERN Data Centers and the WLCG Infrastructure. A free, open-source process supervision tool.
MRML	multi-resource modelling lang
Naemon	Open Source System and Network monitoring Application
Newdle	a collaborative enterprise meeting scheduling tool(2019. 12 기준 개발 중), based on the Open Source application with the same name
NOTED	Network-Optimized Transfer of Experimental Data, Project 이름
NSD	Network Shared Disk. IBM 사의 GPFS 구성 요소
NUC	Next Unit of Computing (인텔사의 small form factor PC with a tiny footprint(작은 공간))
NVMe	Non-Volatile Memory express
NXCALS	Next Accelerator Logging Service
ODDC	On-Demand Data Center, Adapribe Computing 사의 상표 , 퍼블릭 클라우드 서비스를 위하여 특정 벤터에게 종속하지 않는 그리고 확장 가능한 클라우드 시스템 관리 플랫폼을 나타냄
OKD	opensource version of Red Hat OpenShift. Kubernetes의 Origin Community Distribution"을 의미
Open-Xchange	the open source stack initiative, with the ... and secure email service offered by MIT scientists and CERN
OpenDistro	Elasticsearch와 Kibana의 상용화에 따른 새로운 open 버젼
Open JDK	Open Java Development Kit
ORCID	The Open Researcher & Contributor ID
OSDs	object storage demons 또는 Ceph의 Object Storage Device
PCC	Prévessin Computer Center, CERN Tier 0 센터

PCIe	Peripheral Component Interconnect express, 인텔이 주도 직렬 구조의 인터페이스
Pentaho	히타치사의 정형 및 비정형 데이터에 대해 수집부터 변환, 적재, 분석, 그리고 시각화까지 일련의 과정을 모두 제공하는 원스탑 빅데이터 솔루션
PGID	specific placement group ID
PPE	personal protective equipment
PPD	Particle Physics Department in RAL
RADOS	Reliabl, Autonomic, Distributed Object Store comprised of self-healing, self managing, intelligent storage nodes
RAID	Redundant Array of Independent Disks 혹은 Redundant Array of Inexpensive Disks, 복수 배열 독립 디스크
RAL	Rutherford Appleton Laboratory, 영국
RAO	Recommended Access Order
RBD	RADOS Block Device
RDM	Research Data Management
REANA	Reusable Analyses (버전 0.5) reusable and reproducible research data analysis platform., Jupyter 노트북 사용
redfish	Redfish 표준은 서버, 스토리지, 네트워킹 및 컨 버지 드 인프라 관리를위한 RESTful 인터페이스를 제공하는 산업 표준 프로토콜을 제공하는 일련의 사양입니다
RedCLARA	남미국가들을 연결하는 LHCONE
RHEL	Red Hat Enterprise Linux
ROR	ROR(the Research Organization Registry)
Rucio	the next-generation Data Management system in ATLAS
SAS	Serial Attached SCSI,
SATA	Serial Advanced Technology Attachment
S3	Simple Storage Service
SCD	Scientific Computing Department
SCM	Storage Class Memory
SED	Self-Encrypting Drive
ServiceNow	Now Platform service for processing business services
SFA	Storage Fusion Architecture, KIT GridKa 출처
SIM	Subscriber Identification Module
SLA	Service Level Agreement, 서비스 제공자가 이용자에게 약속한 서비스 가용성 및 성능에 관한 계약
SLAC	Stanford Linear Accelerator Center
SLC	Single Level Cell in SSD
SLO	Service Level Objective, 서비스 제공자가 도달하기 원하는 목표

SLI	Service Level Indicator
SLURM	Simple Linux Utility for Resource Management)
SENSE	SDN for end-to-end Networked Sience at the Exascale
SSD	Solid State Drive
STFC	the Science and Technology Facilities Council
StoRM	Storage Resource Manager
Spark	Apache Spark- A big data processing framework for ETL, analytics, machine learning, streaming
SRM	Storage Resource Management (CERN 에서 2008년 대에 사용)
stratum-1	The basic definition of a stratum-1 time server is that it be directly linked (not over a network path) to a reliable source of UTC time such as GPS, WWV
SWAN	Service for Web based Anlysis, 클라우드에서 대화형 데이터 분석을 위한 일종의 플랫폼
TBW	Total Byte Written, SSD에 오류가 발생하기 전에 기록할 수 있는 총 데이터 양
TCO	Total Cost of Ownership
Thruk Monitorimg	Webinterface for Naemon, Icinga, Shinken and Nagios
TLC	Triple Level Cell in SSD
TONE	Telephony Open-source Network Evolution (Fixed Line Phone Service, Mobile Phone Service, Skype for Business
TPC	A Time Projection Chamber, 예: Linear Collider TPC 등
TPU	텐서 처리 장치(Tensor Processing Unit, TPU)는 구글에서 2016년 5월에 발표한 데이터 분석 및 딥러닝용 하드웨어
tungsten fabric	multicloud, multistack sdn open-source, cloud-grade networking and security
UKRI	UK Research and Innovation
vault	토큰, 비밀번호, 인증서 ,암호화 키와 같은 비밀 정보나 민감한 데이터를 UI, CLI, HTTP API를 활용할 때 안전하게 저장하고 제어할 수 있게 해주는 오픈 소스
VO	Virtual Organization
VOMS	Virtual Organization Membership Service
VRF	Virtual Routing and Forwarding
Weblogic	기존 및 클라우드 환경에서 애플리케이션을 개발 및 배포하기위한 기업 환경에 적합한 핵심 클라우드 플랫폼 제공
WIPO	World Intellectual Property Organization
Zenodo	a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN
zero-touch	windows 10 migration

8. 부록 2: 데이터 스토리지 산업체 기술 동향

표 14 SingleStore사 동향

<table>
<tr><td>제목</td><td colspan="3">SingleStore 사는 간편한 클라우드 데이터 통합을 위한 AWS Glue 서비스 제공
(SingleStore Adds AWS Glue for Simpler Cloud Data Integration)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년 1월 6일</td><td>출처</td><td>insideHPC</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://insidehpc.com/2021/01/singlestore-adds-aws-glue-for-simpler-cloud-data-integration/</td></tr>
<tr><td>요약</td><td colspan="3">• 운영 분석 및 최신 애플리케이션을 위한 데이터베이스 기업인 SingleStore 사는 클라우드 데이터 통합을 확장할 수 있도록 하는 AWS Glue의 기본 서비스를 추가 기능 지원을 발표함.
• 개발자, 데이터 엔지니어 및 데이터 과학자는 이를 통해 AWS(Amazon Web Services)에 있는 SingleStore 서비스를 보다 쉽게 구축 활용 할 수 있음.
• AWS Glue는 개발자, 데이터 분석가 및 데이터 과학자가 데이터를 쉽게 추출, 정리, 강화, 표준화 및 로드할 수 있는 서버리스 (serverless) 데이터 준비 서비스임.
• AWS Glude에서 제공하는 ETL(추출, 변환 및 로드) 작업을 위한 완전히 관리되고 확장된 Apache Spark 환경은 SingleStore의 분산 SQL 설계와 잘 연결이 됨.
• 이 기능을 사용하면 다른 환경에서 Single Store 사의 Spark Connector와 동일한 방식으로 AWS에서 병렬 수집을 통해 더 빠르게 처리할 수 있음
• SingleStore 사에서 제공하는 AWS 통합 서비스는 고성능의 긴밀한 통합이어서 고객은 이러한 통합을 통해 추출, 변환 및 로드 프로세스가 가속화되아서 만족도가 더욱 높아질 것임
• SingleStore사에서 제공하는 Apache Spark 통합은 AWS Glue 콘솔에서 액세스할 수 있는 새로운 AWS Glue Custom ETL Connectors 기능을 통해 사용할 수 있음.
• AWS Glue Custom ETL Connectors의 얼리 어답터 기업인 SingleStore는 AWS에서 발표된 AWS Glue Custom ETL Connectors 기능의 출시에 참여하도록 협정된 기업임.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 클라우드 기반 HPC 환경에서도 데이터 통합 처리도 고성능화가 서비스들이 출현하고 있음을 보여줌.</td></tr>
</table>

표 15 빅데이터센터 설계 동향

<table>
<tr><td>제목</td><td colspan="3">데이터 센터를 위한 최신 HPC 및 빅 데이터 설계 전략
(Modern HPC and Big Data Design Strategies for Data Centers)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년 1월 12일</td><td>출처</td><td>insideHPC</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://insidehpc.com/2021/01/modern-hpc-and-big-data-design-strategies-for-data-centers-part-3/</td></tr>
<tr><td>요약</td><td colspan="3">• 오늘날 회사내 또는 클라우드에서의 데이터 처리는 대용량의 데이터를 처리해야하는 경우가 많아짐., 특히 고성능 컴퓨팅(HPC), 인공지능(AI) 그리고 딥러닝(DL) 분야에서 대용량 데이터의 융합 처리가 빈번해짐.
• 따라서 데이터 센터도 HPC 시스템 급 서버 및 이 서버에서 실행되는 대용량의 데이터 처리 및 저장이 가능한 인프라 아키텍처를 필요로 함. 새로운 데이터센터 아키텍쳐는 두 가지 특징을 갖게되는 데 IO-heavy 컴퓨팅과 빅데이터 처리/저장임
• 첫 번째의 IO-heavy 컴퓨팅을 위해서는 응용 프로그램이 데이터 계산 수행 전에 디스크에서 데이터를 가져올 때까지 통상 대기해야 하므로 IO가 바인딩될 수 있음.
• IO 바인딩 병목 해결 방법 중의 하나로 이전 SATA 스토리지 디바이스보다 훨씬 더 높은 IOPS (Input/Output Operations Per Second) 성능을 제공하는 NVMe를 활용하는 것임. 예를 들어 PCIe 3.0x4 와 PCIe 4.0 x 16은 4세대 확장 카드 또는 16레인 구성 슬롯을 나타내며 지원할 수 있는 대역폭 양을 두 배로 확대해 줌.
• 두 번째인 빅 데이터 처리 및 저장 작업은 기존 데이터 처리 애플리케이션 소프트웨어로 처리할 수 없을 정도로 크거나 복잡한 데이터 세트를 분석, 체계적으로 추출 및 저장하는 작업을 포함함. 일반적으로 빅데이터 처리 조직은 예측 분석, 사용자 행동 분석 또는 기타 고급 데이터 분석 방법을 사용하여 데이터에서 가치를 추출한 다음 데이터를 저장함.
• 따라서 빅데이터 처리 조직에서는 대용량 스토리지를 갖춘 시스템이 필요함. 대량의 데이터를 고밀도로 그리고 대용량으로 저장하도록 설계된 시스템을 선택하는 것이 중요함.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• HPC, AI, 그리고 딥러닝을 위한 데이터센터도 HPC 급 서버의 운영은 물론 초고속의 IO-heavy 로컬컴퓨팅 및 대용량의 데이터 저장을 위한 인프라 아키텍쳐로 데이터센터도 진화하여야 함.</td></tr>
</table>

표 16 씨게이트(Seagate)사 동향

<table>
<tr><td>제목</td><td colspan="3">씨게이트, 세계에서 가장 빠른 AI 슈퍼컴퓨터를 설계하는 I4DI 콘소시움에 합류.
(Seagate Joins I4DI Driving Design of World's Fastest AI Supercomputer)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년1월19일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/seagate-joins-i4di-driving-design-of-worlds-fastest-ai-supercomputer/</td></tr>
<tr><td>요약</td><td colspan="3">• 세계적인 데이터 저장장치 전문 업체인 Seagate는 혁신적인 글로벌 기술 리더의 파트너십인 I4DI (Innovations for Digital Infrastructure)의 최신 멤버로 합류함
• I4PI에서는 HPC 분야의 성능, 전력 및 효율성 문제를 해결하기 위해 유럽 기반 슈퍼컴퓨팅 솔루션을 개발하고 Open Compute와 유사한 자유와 개방성을 유럽 HPC 시장에 제공함.
• I4PI 컨소시엄 참여 기관의 범위는 데이터 센터 및 슈퍼컴퓨터 기술을 도입하는 업체로부터 전 세계 AI 및 HPC 솔루션 연구 분야의 혁신적인 리더까지 포함하는 - 스토리지, 네트워킹, 랙 제조, 전력, 소프트웨어, 서비스 및 통신 분야의 - 기업들이 모여서 전 유럽 지역을 변화시킬 수 있도록 지원함.
• Seagate는 혁신적인 하드 드라이브, 솔리드 스테이트 드라이브, 시스템 및 복구 서비스로 폭발적인 데이터 증가를 극복하는 데 기여한 40년 이상의 경험을 I4DI에 제공함.
• EU는 전 세계 슈퍼컴퓨팅 자원의 30%를 소비하지만 최근 EU 보고서에 따르면 EU 실제 보유 자원은 세계 용량의 5%에 불과함. 이러한 불일치로 인해 EU 위원회는 효율성과 자원 활용도를 향상시키면서 슈퍼컴퓨터와 데이터 센터의 확장 구축을 착수함.
• I4DI에서 설계하는 데이터 센터는 InoCloud와 같은 미래 정부 또는 상용 클라우드 이니셔티브을 위한 의 청사진 역할을 할 것임. 상업적으로나 정부/AI 워크로드 모두에서 이중으로 사용하기 때문에 새로운 I4DI 데이터 센터의 효율성이 향상되어 하드웨어 활용률이 현재 40%에서 90% 이상으로 향상될 수 있을 것으로 예상함.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 데이터 저장 장치 전문 생산업체들이 빅데이터와 AI 기술이 융합되면서 HPC기반 AI 슈퍼컴퓨터 시장에서도 주도적인 역할을 수행하기 시작함.</td></tr>
</table>

표 17 DataCore Software사 동향

<table>
<tr><td>제목</td><td colspan="3">소프트웨어 정의 스토리지 전문회사인 DataCore Software사가
객체 스토리지 전문회사인 Caringo사를 인수함
(DataCore Software Acquires Caringo)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년1월26일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/datacore-software-acquires-caringo/</td></tr>
<tr><td>요약</td><td colspan="3">• 소프트웨어 정의 스토리지 전문업체인 DataCore Software사는 오늘 Caringo 사를 인수함. 이는 통합된 한 회사에서 데이터를 블록, 파일,및 객체 등 모든 형태에 대하여 완벽한 동종 최고의 소프트웨어 정의 스토리지 솔루션을 제공하기 위함.
• Caringo사 또한 객체 스토리지 시장의 선구자로, 끊임없는 데이터 증가 문제를 해결하기 위하여 소프트웨어로 스토리지의 경제성을 해결하는 회사임. 주력 제품인 Swarm은 페타바이트 규모의 하이퍼스케일 데이터 스토리지, 액세스 및 분석을 위한 기반을 제공하는 동시에 데이터 무결성을 보장하고 하드웨어 종속성에서 벗어나게함.
• Caringo사 인수로 DataCore사가 추구하는 스토리지로부터 데이터까지 하나의 솔루션으로 제공하는 ONE 비전 계획은 더욱 가속화될 것임. 즉, 소프트웨어 정의 스토리지의 강력한 성능을 실현하여 기존의 사일로 및 하드웨어 종속성을 해소하고 스토리지 산업을 통합하는 것이 더욱 강하게 추진함
• IT 기술로 대용량 데이터 서비스의 유연성, 하드웨어 독립성 및 효율성을 제공한다는 것이 입증됨에 따라 소프트웨어 정의 스토리지 기술의 채택이 가속화됨. 많은 기업들이 대량의 데이터를 관리하기 위한 보다 더 비용 효율적이고, 안전하고, 신뢰할 수 있는 방법을 찾으려하면서 객체 스토리지 기반의 데이터 서비스 채택이 증가함.
• 오늘날의 경제 환경에서 IT 조직은 기존의 서비스 규정 준수, 백업 사용 증가, 비즈니스 연속성 및 지속적인 데이터 증가를 해결하면서도 동시에 비용 절감을 모색해야 함. 이를 위하여 스토리지 서비스와 데이터 서비스를 고성능으로, 그리고 벤더 종속성이 해결된 seamless 통합으로 제공하는 것이 더욱 더 중요해질 것임.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• HPC 분야에서 대용량의 데이터의 처리 수요가 늘어나면서, 이를 경제적으로 해결하고 IT 기술들의 통합 제공을 위하여 관련 기업들의 합종연횡이 활발히 진행되고 있음을 알 수 있음</td></tr>
</table>

표 18 Elbencho사의 스토리지 벤치 마크 동향

<table>
<tr><td>제목</td><td colspan="3">Elbencho: AI를 위한 새로운 스토리지 벤치마크
(Elbencho: A New Storage Benchmark for AI)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년2월3일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/elbencho-a-new-storage-benchmark-for-ai/</td></tr>
<tr><td>요약</td><td colspan="3">• Elbencho는 새로운 오픈 소스 스토리지 벤치마크 툴로서, 고성능을 요구하는 조직이 스토리지 액세스에 GPU를 포함하는 최신 스토리지 시스템의 성능을 쉽게 평가할 수 있도록 지원함.
• 기존에는 스토리지 시스템 벤더가 주로 단순한 대규모 블록 스트리밍 대역폭을 기반으로 수치를 발표함. 이는 회전식 Disk의 메커니즘이 다른 거의 모든 워크로드에 대해 제대로 수행되지 않기 때문임.
• 최근 몇 년간 반도체 메모리의 급격한 가격 하락으로 인해 All-Flash 시스템이 등장함. 이를 통해 데이터 과학자들을 위한 훨씬 더 까다로운 액세스 패턴을 지원하는, 즉 고속 플래시 스토리지 장치에 최적화된 새로운 알고리즘들이 개발되고 있음.
• 이는 동시성이 높은 작은 파일이나 랜덤 읽기를 많이 읽어야 한다는 요구 조건을 갖는 딥러닝이 주요 개발 동기 임. 그러나 고속 NVMe 또는 Optane SSD를 사용하면 스토리지 소프트웨어(즉, 파일 시스템 또는 블록 서비스)가 데이터 경로 설정에서 훨씬 더 중요한 부분이 됨. 이러한 새로운 액세스 패턴에 대한 스토리지 시스템 성능을 예측하기가 어려워짐.
• BeeGFS 병렬 파일 시스템을 만든 Sven Breuner는 위와 같은 사실을 깨닫고 새로운 벤치마크 툴인 Elbencho를 개발함. 다양한 크기의 작은 파일, 액세스 지연 시간, 공유 파일에 대한 랜덤 액세스에 대한 IOPS 등을 테스트하는 등 광범위한 테스트 사례를 지원함.
• 모든 테스트 사례에 대해 통계를 실시간으로 표시하고 데이터 로드 시 시스템이 어떻게 동작하는지 확인할 수 있음. 또한 Elbencho는 엔비디아의 새로운 GDS(GPUDirect Storage) API를 공식적으로 지원하여 GPU에 얼마나 많은 스토리지 성능을 제공할 수 있는지 확인할 수 있는, 최초의 독립적으로 제공하는 툴임.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 초고속 AI GPU 컴퓨팅을 위하여 개발된 신 기술의 스토리지 시스템을 평가할 수 있는 스토리지 벤치마크 도구들이 새롭게 등장하기 시작함</td></tr>
</table>

표 19 KIOXIA사 동향

제목	KIOXIA의 최신 NVMe SSD를 스토리지 플랫폼인 Supermicro PCIe 4.0 서버에서 사용 가능해짐 (Latest NVMe SSDs from KIOXIA Now Available on Supermicro PCIe 4.0 Server, Storage Platforms)		
게시일	2021년2월24일	출처	hpcwire
		유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/latest-nvme-ssds-from-kioxia-now-available-on-supermicro-pcie-4-0-server-storage-platforms/		
요약	• 요즘은 서버와 스토리지 인터페이스가 PCIe 4.0 기술로의 전환이 한창인데, KIOXIA 사는 CM6 및 CD6 시리즈로 PCIe 4.0 기반 NVM Express(NVMe)를 지원하는 엔터프라이즈 용 제품의 라인업을 발표함. 그리고, 데이터 센터 용 솔리드 스테이트 드라이브(SSD)를 Super Micro Computer, Inc.의 PCIe 4.0 기반 플랫폼과의 호환성 승인을 받음. • 또한 Ultra, WIO, BigTwin, FatTwin, SuperBlade, 1U/2U NVMe 등 모든 플래시 어레이, GPU 가속 시스템 및 Super Workstation을 포함하는 다양한 엔터프라이즈급 랙마운트 시스템을 제공함. 엔터프라이즈급 성능, 클라우드 기반 데이터 센터 아키텍처, 그리고 성능 중심 및 지연 시간에 민감한 애플리케이션 지원 등의 사용자 요구사항을 해결하기 위해 NVMe SSD로 기업들이 갈아타고 있음. • 사유는 최신 성능인 PCIe 4.0으로 업그레이드된 NVMe 기반 스토리지에 대한 사용자의 수요가 증가하였기 때문임. • NVMe 서버 및 스토리지 제품군은 기존 SAS 및 SATA 인터커넥트보다 혁신적인 기능을 제공함. ※ 스토리지 플랫폼 인터페이스의 핵심 키워드로 자리잡아가는 'PCI-Exrpess 4.0'(이하 PCIe 4.0)은 x4 레인(lane) 기준으로 대역폭이 최대 64GT/s(초당 기가트랜스퍼)에 달함. 데이터를 1초에 최대 7.88GB(기가바이트) 전송할 수 있는 속도이고 기존 PCIe 3.0보다 두 배나 빠른 수준이어서 최신 그래픽카드와 SSD용 인터페이스로 각광 받음.		
결론 및 시사점	• PCIe 4.0 기술의 가장 큰 특징은 엄청난 대역폭과 데이터 전송 속도인데 이를 그래픽 카드와 SSD 인터페이스에 빠르게 적용되면서 HPC스토리지 기술이 생활 주변까지 확산되고 있음을 알 수 있음.		

표 20 Spectra Logic 사와 OpenDrives 사 동향

제목	Spectra Logic 사와 OpenDrives 사는 단대단 스토리지 관리 기술을 공동 개발함 (Spectra Logic and OpenDrives Partner to Provide End-to-End Data Storage, Management)		
게시일	2021년3월2일	출처	hpcwire
		유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/spectra-logic-and-opendrives-partner-to-provide-end-to-end-data-storage-management/		
요약	• 데이터 스토리지 및 데이터 관리 솔루션의 선두 주자인 Spectra Logic 사와 엔터프라이즈급 초확장 NAS(Network-Attached Storage) 솔루션의 글로벌 공급자인 OpenDrive사는 데이터 스토리지 워크플로우를 개선하고 조직의 민첩성 보장과 비용 절감을 위한 완벽한 통합 데이터 스토리지 및 스토리지 라이프사이클 관리 솔루션을 발표함 • 이 결합된 솔루션은 OpenDrive사의 고성능, 저 지연 시간의 특성을 갖는 기본 스토리지를 Spectra Logic의 StorCycle (Storage Lifecycle Management 소프트웨어)와 통합하여 워크플로우 생산성을 가속화하고 비즈니스 효율성을 향상시킴. • 즉, Primary 스토리지와 Secondary 스토리지의 현대적인 '차세대' 조합과 데이터 및 컨텐츠가 적시에 올바른 위치에 있도록 지원하는 다이내믹 스토리지 라이프사이클 관리 전략을 제공함. OpenDrive사의 초확장 NAS 솔루션은 유연하고 확장 가능하며 빠른 엔터프라이즈급 Primary 스토리지를 제공하여 낮은 전력 소비율과 높은 성능을 제공함. • 이 통합 솔루션은 SMB, NFS 및 LTFS와 같은 개방형 스토리지 표준을 사용하며 단시간 안에 설치될 수 있음. 기본 계층의 데이터를 줄임으로써 조직은 기본 계층 스토리지 용량을 극대화하는 동시에 워크플로우의 요구에 따라 보다 경제적인 영구 계층을 구성함. • 즉, 비활성 데이터와 완료된 프로젝트를 Primary 스토리지에서 클라우드 스토리지, 기존 Disk 및 테이프의 모든 계층을 포함하는 저비용 스토리지 계층으로 신속하게 식별하고 마이그레이션할 수 있는 이더넷 기반 스케일아웃 스토리지를 제공함.		
결론 및 시사점	• HPC 스토리지 관리에서도 데이터의 라이프사이클 관리 기반 현대적이고 비용 효율적인 솔루션들이 출현하고 있음		

표 21 SCSI 사업자 협회 동향

<table>
<tr><td>제목</td><td colspan="3">SCSI 사업자 협회는 스토리지사업을 24기가 SAS 기술로 재편하기위한 2021년도 이사회 구성
(SCSI Trade Association Announces 2021 Board of Directors as Storage Industry Transitions to 24G SAS Technology)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년3월10일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/scsi-trade-association-announces-2021-board-of-directors-as-storage-industry-transitions-to-24g-sas-technology/</td></tr>
<tr><td>요약</td><td colspan="3">• 데이터 스토리지 인터커넥트에 사용되는 SAS(Serial Attached SCSI) 기술 분야에서 세계적으로 인정받는 대표 단체인 SCSI Trade Association(STA)은 오늘 2021년 이사회를 발표함.
• 스토리지 업계를 주도하는 STA 리더들은 전략적인 마케팅 프로그램을 개발하는 업무를 담당함. 주된 업무는 INCITS T10 기술 위원회의 업무를 지원하고, 시장에서 SAS의 가치에 대해 업계 대상 교육을 기획함. 이번에 4명의 담당자와 2명의 위원을 1년 임기로 새로 선출함. STA 대표로 KIOXIA America 사의 기업 마켓팅 이사인 카메론 T 브렛이 3번째 임기를 위해 재선됨. Amphenol Corporation 사의 북미 비즈니스 개발 매니저인 David Einhorn이 재무 담당 이사로 이사회에 합류함. 다른 이사들은 모두 2020년 로스터와 같은 위치에서 참여함.
• Brett 대표는 "업계가 12 Gb/s에서 24 Gb/s SAS로 전환함에 따라 엔터프라이즈 데이터 센터의 요구 사항을 지속적으로 충족하는 동시에 보다 강력한 데이터 안정성과 연결성을 제공하는 노련한 스토리지 베테랑 그룹의 리더십을 환영한다" 그리고 "24G SAS 생태계가 이제 완전히 개발되었습니다. SAS 상호 호환성 (plugfest) 시험이 완료되고 OEM 고객 검증이 끝나면 24G SAS 솔루션을 시장에서 구입할 수 있게 됩니다."라고 발표함
• STA는 회원사의 산업에 SCSI의 이점을 전달하는 데 초점을 맞추기 위해 1996년에 설립됨, STA는 SAS(Serial Attached SCSI) 기술의 이해와 사용을 장려하며, 향후 업계의 요구 사항을 충족하기 위한 SCSI 표준의 발전에 기여함. STA 웹 사이트 (http://www.scsita.org) 참조.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 전송속도가 최대 640 Mb/s인 가장 오래된 스토리지 연결 방식인 SCSI (Small Computer System Interface) 기술까지도 HPC를 지원하기위하여 최대 24 G/s로 (40배 정도로) 성능을 향상시키려 함.</td></tr>
</table>

표 22 Fungible사와 LANL연구소 동향

<table>
<tr><td>제목</td><td colspan="3">Fungible사와 LANL연구소의 EMC3는 복잡한 시뮬레이션 응용을 위하여 확장형 스토리지 개발 협력 발표.
(Fungible and LANL EMC3 Collaborate Explore Scalable Storage for Complex Simulation Applications)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년3월16일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/fungible-and-lanl-emc3-collaborate-explore-scalable-storage-for-complex-simulation-applications/</td></tr>
<tr><td>요약</td><td colspan="3">• Fungible사와 LANL 연구소는 공동으로 패브릭에 대한 컴퓨팅 NVMe의 확장성을 탐색하고, 극단적인 규모의 물리적 시뮬레이션에 의해 생성된 데이터에 대한 분석 기능의 오프로드를 시연할 것임. 기존 분산 스토리지 아키텍처에서는 상기 데이터에 의한 병목 현상이 발생하는 데, 스토리지 시스템의 효율성을 높이기 위해서 스토리지 처리 스택을 더 잘 지원하는 새로운 하드웨어 접근 방식이 필요함
• 제안 방식으로는 프로그래밍이 가능한 네트워크 및 스토리지 장치에 컴퓨팅 작업을 오프로드 하는 것인데, 규모에 따라서 적절하게 실행될 수 있는 애플리케이션과 서비스 류의 성능을 향상시킬 수 있는 방안으로 부각됨. 특히 클라우드 및 HPC 데이터 센터 모두에 매우 중요함.
• LANL 연구소는 가장 복잡한 물리적 시뮬레이션 문제의 처리 시간을 6개월에서 6일로 단축하기위하여 Fungible 사의 DPU를 이용하여 이를 실현코자 함.
• Fungible Storage Cluster는 고성능, 보안, 스케일아웃, 세분화된 데이터 스토리지 플랫폼으로, 2RU 폼 팩터에서 세계 최초로 15M IOPS를 제공하고, 단일 40RU 랙에서 300M IOPS로 선형적으로 확장되며, 더 많은 랙까지 확장 가능함.
• 업계 최고의 성능을 통해 워크로드를 통합하여 스토리지 미디어의 활용률을 높이고 설치 공간과 IOPS당 비용을 기존 소프트웨어 정의 스토리지 솔루션에 비해 최소 3배 이상 향상함.
※ LANL(로스 알라모스 국립 연구소), EMC3(Efficient Mission Centric Computing Consortium), DPU(Data Processing Unit), IOPS(Input/Output Operations Per Second)</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 매우 복잡한 시뮬레이션을 위한 HPC 스토리지 서비스를 효율적으로 제공하기 위해서는 클라우드 또는 HPC 데이터 센터에서 필요에 따라 실시간으로 확장이 가능한 스토리지 서비스가 필요함</td></tr>
</table>

표 23 삼성전자 동향

<table>
<tr><td>제목</td><td colspan="3">삼성전자는 업계 최초로 하이K메탈게이트 기반 DDR5메모리 개발
(Samsung Develops Industry's First High-K Metal Gate-based DDR5 Memory)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년3월24일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/samsung-develops-industrys-first-high-k-metal-gate-based-ddr5-memory/</td></tr>
<tr><td>요약</td><td colspan="3">• 첨단 메모리 기술 분야 세계 선두주자인 삼성전자가 하이케이 메탈게이트(HKMG) 공정 기술을 이용하여 업계 최초로 512GB DDR5 모듈을 개발하여 기존 DDR5 D램 메모리 포트폴리오를 확대함.
※ DDR5 SDRAM(Double Data Rate 5th generation Synchronous Dynamic Random-Access Memory)
• DDR4 성능의 2배 이상인 초당 최대 7,200메가비트(Mbps)로 성능을 보여주는 새로운 DDR5는 슈퍼컴퓨팅, 인공지능(AI) 및 머신러닝(ML)에서 가장 많은 컴퓨팅 성능을 필요로 하는 워크로드와 데이터 분석 애플리케이션에 활용될 수 있음
• 따라서, DDR5를 이용하면 고객들에게 의료 연구, 금융 시장, 자율 주행, 스마트 시티 등에 필요한 컴퓨터에 전력을 더 공급할 수 있는 고성능의 에너지 효율적인 메모리 솔루션을 제공할 수 있음.
• AI/ML 분야에서는 데이터 이동 량, 데이터 저장 또는 처리될 데이터의 규모가 기하급수적으로 증가함에 따라 DDR5로의 전환은 향후 클라우드, 데이터 센터, 네트워크 및 엣지 컴퓨팅 기술 구현에 매우 중요한 전환점을 제공할 것임.
• 삼성 DDR5는 TSV(Through-Silicon-Via) 기술을 활용해 8겹의 16Gb D램 칩을 쌓는 방법을 통하여 최대 512GB 용량을 제공함. TSV 방법은 2015년 삼성이 최대 256GB 용량의 서버 모듈을 선보이면서 D램에 처음 활용되었음.
• 삼성은 현재 DDR5 메모리 제품군의 다양한 종류를 고객에게 샘플링하여 검증하고 궁극적으로는 최첨단 제품으로 인증하여 AI/ML, 엑사스케일 컴퓨팅, 분석, 네트워킹 및 기타 데이터 집약적인 워크로드 확산을 가속화함.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 대용량 데이터 기반 HPC 컴퓨팅을 위하여 시스템의 반도체 메모리 용량 또한 이전의 외장 스토리지 수준까지 빠르게 확장 되고 있음.</td></tr>
</table>

표 24 Spectra Logic 사 동향

<table>
<tr><td>제목</td><td colspan="3">Spectra Logic 사는 2021년 데이터 스토리지 전망 보고서를 발표함
(Spectra Logic Releases 2021 Data Storage Outlook Report)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년3월31일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/spectra-logic-releases-2021-data-storage-outlook-report/</td></tr>
<tr><td>요약</td><td colspan="3">• 플래시 메모리 전망: 플래시 공급업체는 점점 더 많은 계층이 축적된 플래시 칩을 만들기 위해 멀티 레이어 플래시 다이(flash dies)를 함께 연결하는 string-stacking 기술에 의존하게 될 것임. 이로 인해 플래시의 비용 절감 효과가 커질 것임. 시스템 및 클라우드 공급자는 영역 기반 인터페이스(데이터의 성능 요구사항에 맞는 영역에 데이터를 물리적으로 배치할 수 있음)를 활용하여 플래시 메모리를 더 긴 수명, 더 나은 성능 및 더 큰 용량으로 확보할 수 있음.
• 디스크 전망: 디스크 업계의 두 주요 선두 기업은 대용량 드라이브로 전환하는 전략을 크게 달리하고 있음. 하나는 전적으로 전통 방식 기반의 HAMR에 초점을 맞추고, 다른 하나는 eMAMR, MAMR, HAMR, SMR을 활용하여 저장 용량을 20% 더 향상시킴.
• 테이프 전망: LTO-9 테이프의 용량은 카트리지당 18TB로 LTO-8에 비해 50% 증가함. 테이프 시스템 공급업체가 지원할 경우 테이프를 활용할 수 있는 애플리케이션의 수를 크게 늘릴 수 있는 새로운 사실상의 표준 인터페이스가 등장함.
• 하이브리드 영구 스토리지 전망: 새로운 하이브리드 스토리지 시스템을 도입하면 클라우드 및/또는 사내 처리 기능을 활용하는 동시에 해당 처리의 원시 및 정제된 데이터를 장기간 보존할 수 있음. • 영구 메모리 전망: 다양한 용량으로 쉽게 구입할 수 있는 3D XPoint 기술을 기반으로 한 영구 메모리는 DRAM 시장에 큰 영향을 줄 것임.
※ HAMR: Heat-Assisted Magnetic Recording, SMR: Shingled Magnetic Recording, MAMR: Microwave Assisted Magnetic Recording, TDMR: Two-Dimensional Magnetic Recording, LTO-9: Linear Tape-Open 9, 3D XPoint: 3D Cross Point(비휘발성 메모리 기술),</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• HPC 스토리지를 위해서 관련 스토리지 기술들이 공통적으로 저비용, 고효율화 기반 대용량화 기술들을 경쟁적으로 도입하고 있음을 전망함</td></tr>
</table>

표 25 VAST Data 사 동향

<table>
<tr><td>제목</td><td colspan="3">VAST Data 사는 세분화된 스토리지 소비 모델인 Gemini를 발표함.
(VAST Data Announces Gemini Disaggregated Storage Consumption Model)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년4월6일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/vast-data-announces-gemini-disaggregated-storage-consumption-model/</td></tr>
<tr><td>요약</td><td colspan="3">• 스토리지 소프트웨어 회사인 VAST Data가 하이퍼스케일 시대를 위해 엔터프라이즈 스토리지 어플라이언스 모델인 제미니(Gemini)를 공개함. 하드웨어와 소프트웨어의 사업을 세분화한 이 모델은 고객이 원가로 필요한 하드웨어 관리 소프트웨어를 구글, 아마존 등 하이퍼스케일 데이터센터가 구입하는 수준의 저비용으로 구입할 수 있도록 함.
• 이 제품은 유연성과 비용 측면에서 소프트웨어 정의 비즈니스 모델의 모든 이점을 제공함. 제품의 주요 특징은 다음과 같음.
✓ 세분화된 상업 모델: 제미니는 고객에게 하드웨어 비용 투명성과 시스템 성능을 높일 수 있는 섬세한 선택의 자유를 제공함.
✓ 무한 스토리지 라이프 사이클: 업그레이드나 마이그레이션을 위해 오프라인으로 전환하지 않는 최신 클라우드 인프라와 마찬가지로 Universal Storage의 비대칭 아키텍처는 시간이 지남에 따라 데이터 수요에 필요한 엑사바이트 규모의 스토리지 클러스터로 서비스 중단 없이 확장할 수 있음.
✓ VAST Co-Pilot: 전문 운영자 지원 프로그램으로서, 신규 고객에게 시스템 관리, 계획, 확장 및 일상적인 운영의 모든 측면을 처리하는 전문 클라우드 운영자 서비스임. 레벨 3급 엔지니어 지원 서비스를 진행하고 있음.
✓ VAST Data의 무손상 보장: VAST는 저비용 플래시 서비스로 10년 데이터 무손상 보장, 가동 시간과 데이터 내구성 보장 등 모든 Universal Storage 기능을 제공함과 동시에 무조건적인 60일 이내 사용후 반환 권리 제공. 이는 고객이 기존 HDD 기반 아키텍처에서 벗어나려고 할 때 새로운 시스템 아키텍처 채택의 위험을 제거해줌.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 수요기반 실시간 HPC 스토리지 서비스를 하이퍼스케일 데이터센터 수준으로 제공하는 것이 가능하도록 스토리지 관리 S/W 기술이 개발됨.</td></tr>
</table>

표 26 SNIA and OpenFabrics Alliance 동향

<table>
<tr><td>제목</td><td colspan="3">SNIA와 개방형 패브릭 연합은 고성능 데이터센터 인프라의 표준 기반 관리를 개선하기 위한 공식 협약 체결
(SNIA and OpenFabrics Alliance Formalize Agreement to Advance Standards-Based Management of High Performance Data Center Infrastructure)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년4월13일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/snia-and-openfabrics-alliance-formalize-agreement-to-advance-standards-based-management-of-high-performance-data-center-infrastructures/</td></tr>
<tr><td>요약</td><td colspan="3">• SNIA(스토리지 Networking Industry Association)와 OFA(Open Fabrics Alliance)는 데이터 센터 인프라에서 고성능 패브릭과 원격 네트워크 서비스의 표준 기반 오픈 소스 관리를 개선하기로 합의함.
• OFA가 개발하고 있는 새로운 오픈 소스 패브릭 관리 프레임워크 OFMF(Open Fabrics Management Framework)는 NVMe 및 NVMe over Fabrics를 포함한 서버, 스토리지 및 패브릭의 포괄적인 관리를 위해 DMTF Redfish 및 SNIA Sodfish 사양을 활용할 예정임.
• SNIA Swordfish 스토리지 관리 사양은 DMTF (Distributed Management Task Force) Redfish 사양의 확장임. 따라서 OFA와의 공동 작업을 통해 IT 관리자는 서버 및 스토리지에 사용하고 있는 동일한 Redfish 및 SNIA Soldfish 에코시스템에 패브릭 및 네트워크의 원활한 표준 기반 관리를 추가할 수 있음.
• 새로운 대규모 컴퓨팅 설계는 다중 노드이며 고속 네트워크를 통해 서로 연결함. 시간이 흐르면서 많은 다양한 유형의 고성능 네트워킹 장치가 이러한 현대적인 결합 컴퓨팅 플랫폼에 사용되기 시작함. 최적화된 네트워크를 구축, 보호 및 유지하려면 이러한 고속 네트워크가 패브릭 관리 요구 사항을 충족해야 함. 이러한 업계 공동 작업을 통해 고급 스토리지, 분산 컴퓨팅 및 메모리에 대한 원격 및 공유 액세스를 제공할 수 있음.
• OFA와 SNIA가 더 큰 고성능 네트워크 생태계의 보편적 관리를 위해 공동 투자를 하는 것임. SNIA Sodfish와 DMTF Redfish 사양은 OFPA가 확장하여 OFMF 생성을 통해 패브릭을 관리할 수 있는 단일 통합 프레임워크를 제공할 것임</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 대용량 데이터를 지원하기위한 HPC스토리지의 효율적인 구축을 위하여 개방형 인프라 기반 시장 표준화 노력을 알 수 있음</td></tr>
</table>

표 27 스페인 슈퍼컴퓨팅 네트워크 사업의 스토리지 서비스 동향

<table>
<tr><td>제목</td><td colspan="3">스페인 슈퍼컴퓨팅 네트워크 사업에서 처음으로
18개 과제에 7.5페타바이트의 스토리지를 제공하는 과제를 공모함.
(Spanish Supercomputing Network's First Call Grants 7.5 Petabytes of Storage to 18 Projects)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년4월20일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/spanish-supercomputing-networks-first-call-grants-7-5-petabytes-of-storage-to-18-projects/</td></tr>
<tr><td>요약</td><td colspan="3">• 스페인 슈퍼컴퓨팅 네트워크 (RES: Red Espaolaola de Supercomputación)는 대량의 데이터를 사용해야 하고, 과학계의 다른 구성원들과 공유하고자 하는 연구자들에게 처음으로 스토리지 용량을 제공할 계획을 수립함.

• RES는 데이터 서비스에 대한 첫 번째 공모과제에서 유전체학, 분자생물학, 화학, 천체물리학, 기후 분석 및 예측, 인공지능 등 다양한 분야에서 18개 프로젝트에 7.5페타바이트 (1페타바이트는 1,000테라바이트에 해당)의 스토리지 제공하는 공모 사업을 승인함. 이 공모에 제출된 프로젝트의 평가는 그 프로젝트의 과학적 우수성과 데이터의 보존, 개선 및 공유를 위한 계획도 포함함. 데이터 관리 계획 준수에 대한 연례 검토에 따라 3년에서 5년 동안 스토리지가 제공됨. 어떤 경우에는 저장된 데이터의 분석, 개선 및 유지보수가 가능하도록 컴퓨팅 용량이 같이 부여되기도 함.

• 이번 공모를 통해 RES는 2007년부터 연간 약 6억 시간의 컴퓨팅 시간을 과학계에 제공했던 슈퍼컴퓨팅 리소스에 대한 액세스 중심의 기존 서비스를 보완함. 스토리지 추가 제공 목적은 인공지능 또는 대용량 데이터를 새로운 방법으로 과학 연구를 할 수 있게 하는 것임. 18개 프로젝트에 제공되는 데이터는 스페인 슈퍼컴퓨팅 네트워크에 연결된 모든 슈퍼컴퓨터센터가 참여하는 분산 스토리지 인프라에 저장됨.

※ 참여 슈퍼컴 센터: Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), Centro de Supercomputación de Castilla y León (SCAYLE), Consorci de Serveis Universitaris de Catalunya (CSUC), Fundación Pública Galega Centro Tecnolóxico de Supercomputación de Galicia (CESGA), Instituto Astrofísico de Canarias (IAC) and Port d'Informació Científica (PIC),</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• AI 기반 대용량 데이터 사이언스 연구를 효율적으로 지원하기 위하여 범국가적으로 슈퍼컴퓨터들을 통합 활용하려는 노력을 알 수 있음.</td></tr>
</table>

표 28 IBM사의 스토리지 제품 라인 동향

제목	IBM사는 스토리지 제품 라인에 '컨테이너-네이티브' 솔루션인 Spectrum Fusion 제품 추가 계획 발표. (IBM Plans to Add ‘Container-Native’ Spectrum Fusion to Its Storage Line)		
게시일	2021년4월27일	출처	hpcwire
		유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/2021/04/27/ibm-plans-to-add-container-native-spectrum-fusion-to-its-storage-line/		
요약	• IBM은 컨테이너 네이티브 기반 새로운 SDS(소프트웨어 정의 스토리지) 솔루션인 IBM Spectrum Fusion을 2021년 하반기에 출시할 계획임. • IBM Spectrum Fusion의 첫 번째 출시 형태는 컨테이너 네이티브 하이퍼 컨버전스 인프라(HCI) 시스템이고, 이는 컴퓨팅, 스토리지 및 네트워킹을 단일 솔루션으로 통합할 것임. • IBM은 Red Hat OpenShift를 통해 기업에게 가상 머신과 컨테이너 환경을 지원하고 클라우드, 에지 및 컨테이너형 데이터 센터를 위한 소프트웨어 정의 스토리지를 제공함. • 2022년 초에 SDS 전용 버전으로 출시되는 Spectrum Fusion은 IBM의 일반 병렬 파일 시스템 및 데이터 보호 소프트웨어의 완전한 컨테이너형 버전을 통합하여 전사적으로 데이터 검색 및 액세스를 간소화함. • Spectrum Fusion 솔루션을 사용하면 기업은 데이터의 단일 복사본만 관리하고 더 이상 애플리케이션 워크로드를 전사적으로 이동할 때 중복 데이터를 생성할 필요가 없어짐. 따라서 관리 기능이 쉬워지고 분석 및 AI가 간소화됨. • 데이터 복사본을 하나로 유지함으로써 개인정보보호법(GDPR: General Data Protection Regulation) 활동은 강화될 수 있으며, 동시에 여러 복사본이 존재할 때 보다는 보안 노출이 감소될 수 있음. • Spectrum Fusion 제품은 IBM Cloud Satellite와 통합되어 기업이 단일 관리 창으로 에지, 데이터 센터 또는 퍼블릭 클라우드에서 클라우드 서비스를 완벽하게 관리할 수 있도록 지원함.		
결론 및 시사점	• HPC 스토리지 구현에서 기존의 물리적 용량 및 데이터 I/O 속도 확대 중심의 구현에서 이제는 hybrid cloud compute environment를 위하여 가상화 및 콘테이터 기반 Software Defined Storage 기술까지도 중요시되고 있음을 알 수 있음.		

표 29 Ohio Supercomputer Center의 스토리지 확장동향

<table>
<tr><td>제목</td><td colspan="3">Ohio Supercomputer Center는 2020년도 연구보고서에서 COVID 대응, Pitzer 시스템, 데이터 스토리지 확장에 대하여 발표함.
(Ohio Supercomputer Center 2020 Research Report Discusses COVID Response, Pitzer, Data Storage Expansions)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년5월3일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/ohio-supercomputer-center-2020-research-report-discusses-covid-response-pitzer-data-storage-expansions/</td></tr>
<tr><td>요약</td><td colspan="3">• 연구보고서는 OSC (Ohio Supercomputer center)를 사용하는 연구원들이 사무실에서 그리고 클러스터 시스템을 이용한 연구 활동과 COVID-19에 대한 OSC의 대응 연구들에 대한 활동을 담고 있음. 그리고 센터의 시스템 및 서비스에 대한 뉴스도 다루고 있음.
• 개발 활동 뉴스에는 430만 달러 규모의 Pitzer 클러스터 시스템 확장, 8.6 페타바이트의 데이터 스토리지 추가 도입, 그리고 Ruby 클러스터 ㅅ;스템의 해체 활동들을 포함하고 있음.
• OSC에서 수행된 연구는 유체역학에서 정책 과학, 그리고 학술 연구에서 상업 공학에 이르기까지 광범위한 분야에서 조사 연구 활동을 수행하고 있음.
• 연구보고서에서 소개된 이야기들은 OSC를 활용한 연구 작업들을 나타내며, 이 결과물들은 전 세계 주요 학술 컴퓨팅 센터 활동과 유사하게, 수십 개의 과학 분야에서 peer-review 방식의 학술지에 1,990개 이상 인용되고 있음.
• 오하이오 주에서 지식 경제의 리더 역할을 수행하는 오하이오 슈퍼컴퓨터 센터는 교직원과 학생들에게 힘을 실어주고 새로운 연구와 사업 기회를 가능하게 함.
• 바이오사이언스에서 첨단 소재에 이르기까지 OSC의 강력한 고성능 컴퓨팅은 오하이오주를 발전시키는 데 매우 중요한 다양한 분야에 기여함
• 올해는 OSC의 유연한 정책으로 OSC의 수십 명의 연구원들이 COVID-19 위기 연구로 갑자기 연구를 전환함에도 즉시 지원할 수 있어서 오하이오의 전략적 대응을 구체화하는 데 도움을 줌.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 오하이오 슈퍼컴센터에서 Covid 19 연구 등 급한 정책적 연구 등을 위한 시스템 확장에서도 HPC 스토리지 확장이 매우 중요하게 다루고 있음을 알 수 있음.</td></tr>
</table>

표 30 오크릿지 슈퍼컴센터의 Exascale System용 스토리지 규격 발표

제목	미국 오크릿지 슈퍼컴센터의 Frontier Exascale 시스템용 슈퍼컴퓨터를 위한 스토리지 사양 발표. (OLCF Announces Storage Specifications for Frontier Exascale System)		
게시일	2021년5월20일	출처	hpcwire
		유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/olcf-announces-storage-specifications-for-frontier-exascale-system/		
요약	• 컴퓨팅 처리 속도가 초당 최고 속도가 10^{18} 회 또는 10억의 10억배 회에 달할 것으로 예상되는 OLCF (Oak Ridge Leadership Computing Facility)의 Exascale 컴퓨팅의 계산 능력은 Cray Frontier가 2022년부터 사용자에게 서비스를 개방할 예정임. Exascale 컴퓨팅의 I/O 시스템은 시스템 내 스토리지 계층과 중앙 전체 파일 시스템이라는 두 가지 주요 구성 요소로 구성되고, 오리온이라고 불리는 중앙 전체 파일 시스템은 오픈 소스 Lustre와 10^{21} 바이트 규모의 ZFS (Zettabyte file system)기술을 사용함. 오리온은 세계에서 가장 크고 가장 빠른 단일 파일 POSIX 네임스페이스 파일 시스템임. 이 시스템은 대규모의 중앙 집중식 병렬-분산형 파일 시스템을 개발 및 배치한 4세대 파일 시스템임. Orion은 극한의 스케일과 하드 디스크/NVMe 하이브리드 기술을 적용하여 기술적 한계를 뛰어넘음. 주요 특징은 다음과 같음. - 플래시 기반 고성능 스토리지 계층은 5,400개의 NVMe(비휘발성 메모리 익스프레스) 장치로 200만 이상의 랜덤 읽기 IOPS와 10 TBps의 최대 읽기-쓰기 속도로 11.5 페타바이트(PB)의 용량 제공. - 하드 디스크 기반 대용량 저장 계층은 최대 읽기 속도 5.5 TBps에 679 PB의 용량과 최대 쓰기 속도 4.6 TBps에 200만 이상의 랜덤 읽기 IOPS를 제공하는 47,700개의 수직형 자기 기록 장치로 구성됨. 플래시 기반 메타데이터 계층을 구성하여 480 NVMe 디바이스로 10 PB의 용량을 추가 제공함. - 시스템 내 스토리지 계층은 PCIe Gen4 링크를 통해 연결된 컴퓨팅 노드의 로컬 스토리지 디바이스를 사용하여 초당 75 TBps 이상의 최대 읽기 속도, 35 TBps 이상의 최대 쓰기 속도, 초당 150억 개 이상의 랜덤 읽기 입출력 작업(IOPS)을 제공함.		
결론 및 시사점	• 엑사스케일 컴퓨팅에서 제터 바이트급의 HPC 스토리지 구현이 사용자 서비스 단계에 도달하였음을 알 수 있음		

표 31 스타트업 22dot6사의 동향

제목	스타트업 22dot6에서 엔터프라이즈 시장에 초월적 추상형 스토리지 아키텍처 (TASS)를 출시함 (Startup 22dot6 Unveils Transcendent Abstractive Storage Architecture to Enterprise Market)		
게시일	2021년5월25일	출처	hpcwire
		유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/startup-22dot6-unveils-transcendent-abstractive-storage-architecture-to-enterprise-market/		
요약	• 스토리지 가상화 및 H/W 추상화를 혁신적으로 뛰어넘는 TASS (Transcendent Abstracted Storage System)는 기업에게 H/W 제조사, 또는 물리적이든 가상이든 클라우드 상의 스토리지의 위치, 그리고 프로토콜 또는 플랫폼에 관계없이 스토리지 리소스 상의 모든 데이터를 투명하게 액세스하고, 이동시키고, 관리 및 보호함. 따라서 TASS는 사용자, 애플리케이션 및 데이터 서비스를 통합할 수 있게 함. • TASS 아키텍처는 NAS, 객체 및 클라우드 시스템에서도 단일 스토리지 시스템의 모양, 느낌 및 단순성을 제공함으로써 과거, 현재 및 미래 기술로 구성된 기존의 모든 플랫폼들이 고립된 섬과 분리된 사일로처럼 되어가는 현상을 제거함. • 22dot6의 첫 번째 TASS 제품인 소프트웨어 Valence는 기존 인프라의 한계를 뛰어넘는 스케일업/스케일아웃을 지원하며 가상화, 하드웨어 추상화, 데이터 서비스 및 스토리지 접근성 등 모든 서비스 요소들 사이에 브리지를 제공하기 위한 통합된 추상화 계층을 지원함. • 관리자는 Valence를 사용하여 실제 스케일업/스케일아웃 아키텍처 내에서 수십억 개의 파일과 엑사바이트의 데이터를 관리할 수 있으며, 물리적 위치 또는 가상 위치당 최대 60개의 데이터 전송/액세스 노드로 구성된 일반 컴퓨팅 HPC 클러스터를 지원하고 단일 글로벌 네임스페이스 내에서 초당 600-1200GB의 데이터 액세스를 제공함. • 주요 특징으로는 간편한 통합, 마이그레이션, 동화(assimilation), 전환, 고가용성, 고성능, 데이터 보호, Active/Active 다중 사이트, 메타데이터 관리, 통합 데이터 및 시스템 프로파일링, 사용자 정의 목표, 관리 및 보고, 멀티프로토콜 지원, Multi-tenant 지원 등		
결론 및 시사점	• 스토리지 가상화 기술이 크게 진화하여 HPC 환경에서도 범용 수준으로 상용화 단계가 보편화 되어 가고 있음을 알 수 있음.		

표 32 Academic Sinica HPC 스토리지 구축 동향

<table>
<tr><td>제목</td><td colspan="3">Academic Sinica, HPC 애플리케이션용으로 인포트렌드 스토리지를 선택하다
(Academia Sinica Chooses Infortrend Storage for HPC Applicationt)</td></tr>
<tr><td rowspan="2">게시일</td><td rowspan="2">2021년6월1일</td><td>출처</td><td>hpcwire</td></tr>
<tr><td>유관부서</td><td>대용량데이터허브센터</td></tr>
<tr><td>원문보기</td><td colspan="3">https://www.hpcwire.com/off-the-wire/academia-sinica-chooses-infortrend-storage-for-hpc-application/</td></tr>
<tr><td>요약</td><td colspan="3">• 타이완의 저명한 학술 기관인 Academia Sinica는 1928년 수학, 물리, 생명과학 등 다양한 분야의 학술 연구를 수행하기 위해 설립됨. 첨단 과학 프로젝트를 위한 연구 데이터와 HPC 워크로드를 처리하려면 대용량 스토리지 용량과 높은 스토리지 성능이 필요함. 기존 스토리지 환경은 원활한 대용량 데이터 기반 HPC 분석을 하기엔 적합하지 않은 성능을 가지고 있고, 빠르게 성장하는 연구 데이터를 위한 충분한 용량을 제공하지 못했으며, 관리가 쉽지 않았기 때문에 더 이상 연구 기관의 증가하는 수요를 충족시킬 수 없었음.
• 또한 IT 서버실은 새 스토리지 시스템을 설치할 수 있는 랙 공간이 제한되어 있음. 따라서 앞에서 언급한 문제를 성공적으로 해결할 수 있는 스토리지 솔루션을 찾고 있었음.
• Academia Sinica는 수치예측시스템의 연구 수요를 위하여 Infortrend의 유니파이드 SAN/NAS EonStor GS 3000 스토리지를 선택함. 관찰 모델은 현장 센서에서 나온 방대한 양의 새로운 관찰을 기반으로 하는 데 이에 대한 분석은 컴퓨팅 리소스에 크게 의존함. GS 3000 Gen2(60pcs x 12TB NL-SAS가 모두 장착된)는 연구 데이터를 수집하기 위한 4,000MB/s 쓰기 속도와 연구 고성능 컴퓨팅을 수행하기 위한 8,000MB/s 읽기 속도를 제공하여 까다로운 애플리케이션에 완벽하게 적합함. IT 관리자는 제한된 랙 공간에 GS를 쉽게 설치하고 4*16Gb FC 호스트 인터페이스를 통해 네트워크에 연결함. GS는 관리 부담을 없애기 위해 직관적인 브라우저 기반 GUI EonOne을 제공함. 연구자의 데이터에 대한 용량 요구 사항은 매우 빠르게 증가하므로 Infortrend는 확장을 위한 최고의 가격 대비 성능을 제공함.</td></tr>
<tr><td>결론 및 시사점</td><td colspan="3">• 대만의 국가과학연구소에서 연구데이터의 대용량화에 따른 HPC 스토리지 확보전략을 위하여 가성비 위주의 솔루션 확보전략을 추구함</td></tr>
</table>

표 33 미네소타 대학교 슈퍼컴퓨팅 연구소 동향

제목	미네소타 대학교 슈퍼컴퓨팅 연구소에서 10 페타바이트의 Panasas 스토리지를 추가 설치함 (University of Minnesota Supercomputing Institute Adds 10 Petabytes of Panasas Storage)		
게시일	2021년6월8일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/university-of-minnesota-supercomputing-institute-adds-10-petabytes-panasas-storage/		
요약	• (개요) 고성능 컴퓨팅(HPC) 데이터 스토리지 솔루션의 선두 업체인 Panasas*사는 새로운 세대의 병렬 스토리지 시스템을 공급하기 위해 미네소타 대학교의 슈퍼컴퓨팅 연구소(MSI)**와 13년 간의 관계를 확장했다고 발표함. MSI는 연구진이 생물정보학, 인공지능, 머신러닝 분야로 더 나아가기위하여 클러스터 스토리지 시스템의 용량과 성능을 모두 업그레이드할 계획임 * Panasas 사: 전 세계의 산업 및 연구 혁신을 지원하는 고성능 컴퓨팅(HPC) 데이터 스토리지 솔루션을 제공하는 기업. ** MSI (University of Minnesota Supercomputing Institute): HPC 연구에 필요한 소프트웨어, 하드웨어 및 전문가를 지원하는 연구소 • (접근방안 소개) 대학은 이를 위해 기존 Panasas 병렬 스토리지 시스템을 차세대 Panasas Ultra 하드웨어로 업데이트하는 것부터 시작함. 이는 약 10페타바이트(PB)의 새로운 용량을 제공하여 5PB의 순증가를 나타내며, 다양한 데이터 집약적인 연구 이니셔티브를 보다 효과적으로 지원함. 그리고, HPE (Hewlett Packard Enterprise) 하드웨어를 사용하여 구축된 MSI의 ceph*** 기반 Tier 2 스토리지 용량도 5PB까지 증가할 예정임. Panasas사와의 파트너십을 통해 MSI의 'Agate' 클러스터는 점점 더 정교해지는 다양한 사용 사례를 해결하는 데 매우 중요한 고성능 컴퓨팅 기술을 제공할 수 있게 됨. ***ceph: Free 소프트웨어 스토리지. 플랫폼 최종 사용자 응용을 위한 객체 스토리지 또는 가상 시스템을 위한 블럭형 스토리지를 서비스함.		
결론 및 시사점	• 해외 슈퍼컴퓨팅센터 마다 HPC 스토리지 확장을 위한 노력이 활발하게 진행되고 있음을 알 수 있음.		

표 34 Qumulo사 동향

제목	Qumulo사는 HPE사의 *GreenLake 클라우드 서비스에게 비정형 데이터를 위한 Pay-as-You-Go 파일 플랫폼 제공 (Qumulo, HPE GreenLake Cloud Services to Provide Pay-As-You-Go File Platform for Unstructured Data)		
게시일	2021년6월17일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/qumulo-hpe-greenlake-cloud-services-to-provide-pay-as-you-go-file-platform-for-unstructured-data/		
요약	• (개요) 오늘날 생성되고 있는 비정형 데이터의 양은 지난 10년 동안 기하급수적으로 증가했으며, 레거시 아키텍처들은 더 이상 증가하는 수요를 따라갈 수 없음. 의료, 금융 서비스, 미디어 및 엔터테인먼트, 기타 산업에 걸친 기업들은 저장, 관리, 보호 및 분석해야 하는 방대한 양의 비정형 파일 데이터를 생산하고 있음. • (접근방안 소개) **Qumulo사와 ***HPE(Hewlett Packard Enterprise)사는 협업을 통해 퍼블릭 클라우드와 함께 제공되는 비정형 데이터 관련 모든 성능, 확장성 및 비용 효율성을 프라이빗 데이터 센터 내에서 제공케함. 예를들면 HPE Apollo 4200 Gen10 하이브리드 플래시 와 HPE ProLiant DL325 Gen10 Plus 서버 상에 구축된 Qumulo 사의 최신 파일 데이터 플랫폼은 엔터프라이즈 조직에 전례 없는 보안 및 데이터 가시성을 제공함. 이러한 강력한 협업을 통해 고객은 데이터 오버프로비저닝 필요성이 없어짐에 따라 자본 지출을 최대 30% 절감하고, 글로벌 IT 프로젝트 구축 시간을 75% 단축하며, 지원 부하를 줄여 IT 팀의 생산성을 40%까지 높일 수 있음. *GreenLake: 애플리케이션과 데이터의 위치에 상관없이 어디서나 종량제로 클라우드를 누릴 수 있음. HPE 사의 클라우드 서비스 명칭 **Qumulo: 고성능 파일 데이터 플랫폼 제작 회사 ***HPE: 엣지 투 클라우드(Edge to Cloud) 서비스형 플랫폼 회사.		
결론 및 시사점	• HPC 인프라의 진화 중 스토리지 관련하여 어떤 크기의 비정형 데이터도 사용자 원하는 용량만큼 사용하게하고, 사용한 만큼만 요금을 부과하는 매우 거대한 가변적 인프라 형태로 진화하고 있음.		

표 35 독일 율리히 슈퍼컴센터 동향

제목	독일 율리히 슈퍼컴센터는 대용량 데이터 처리를 위하여 JURECA 슈퍼컴퓨터를 업그레이드함. (Jülich Upgrades JURECA Supercomputer for Large Data Volumes)		
게시일	2021년6월23일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/julich-upgrades-jureca-supercomputer-for-large-data-volumes/		
요약	• (개요) Jülich 슈퍼컴센터는 모듈식 슈퍼컴퓨터 JURECA를 upgrade 하여 초당 23.5 x 10^{15}번 연산이 가능함. 특징은 방대한 양의 데이터 처리를 위하여 프랑스 벤더인 Atos가 제공하고 독일 회사인 PartTec의 클러스터 소프트웨어가 운영하는 JURECA - DC(Data Centric) 모듈을 설치한 것임. 이 모듈만 해도 세계 30위 내의 슈퍼컴퓨터 중 하나임. • (접근방안 소개) Jülich의 중앙 스토리지 클러스터인 JUST를 52 PB (10^{15} Byte)로 확장함. 특히 *JUST-IME 플래시 메모리 시스템을 채택해 데이터 분석 및 머신러닝 분석을 가속화했는 데 이 모듈은 Hewlett Packard Enterprise 와 **DDN에서 공급했고 율리히 슈퍼 컴퓨터들인 ***JUWELS, JURECA 및 JUSUF에 연결됨. 또한 이 스토리지들은 DDN의 IME(Infinite Memory Engine) 기술을 통하여 대서양 횡단 케이블 ****TAT-14의 약 두 배에 해당하는 2T B/s (10^{12} Bit per second) 이상의 대역폭으로 미국과 연결됨. 이 스토리지 시스템은 2020년에 설치된 슈퍼컴퓨터 JUSUF와 유사하게 EU 프로젝트 ICEI (Interactive Computing E-Infrastructure)의 범위 내에서 *****Fenix e-Infra 의 일부로 조달됨. * JUST-IME : JUST (The Jülich Storage Cluster) - IME(Infinite Memory Engine) IME는 소프트웨어 정의 플래시 스토리지 플랫폼으로서 애플리케이션 I/O를 위한 데이터 경로를 streamline화함. 따라서 IME는 애플리케이션에 직접 인터페이스하여 입출력을 보호하고 파일 시스템 병목 현상을 제거함 ** DDN: DataDirect Networks (DDN®) Storage *** JUWELS, JURECA 및 JUSUF: 율리히 슈퍼컴센터의 슈퍼컴들 **** TAT-14: 14th consortium transatlantic telecommunications cable system *****Fenix: 스타크래프트 게임의 영웅 캐릭터, 유럽의 HPC 공동 활용 인프라 명칭		
결론 및 시사점	• 유럽에서도 빅데이터와 AI를 위한 HPC 인프라의 진화를 위하여 HPC Storage Infra.를 공동으로 구축하고 운영하려는 것을 알 수 있음.		

표 36 DDN사와 Bytesnet사 동향

제목	DDN사와 Bytesnet사는 협력하여 데이터 집약적인 조직을 위하여 '사용량에 따른 과금' 방식의 스토리지 서비스 제공 (DDN Selected by Bytesnet to Provide 'Pay-Per-Use' Storage for Data-Intensive Organizations)		
게시일	2021년6월30일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/ddn-selected-by-bytesnet-to-provide-pay-per-use-storage-for-data-intensive-organizations/		
요약	• (개요) 네덜란드에서는 데이터 집약적인 컴퓨팅 기능이 필요한 정부 및 연구 조직은, 유연성이 떨어지고 비용이 많이 들며 필요한 리소스를 온디맨드 방식으로 액세스할 수 없는 대학 데이터 센터에 연결되어 있는 경우가 많음. 이를 해결하기 위하여 *DDN과 **Bytesnet은 중앙 집중식 인프라를 통해 보다 비용 효율적인 사용량에 따른 종량제 기반 서비스 방안을 제공함. 그리고 연구자의 필요에 따라 자원의 규모를 확장할 수 있게 함. * DDN(Data Direct @ Network): 데이터 스토리지 기업. ** Bytesnet: 네덜란드 정보 서비스 기업 • (접근방안 소개) DDN의 DataFlow 일체형 소프트웨어를 사용하는 광범위한 솔루션 포트폴리오를 활용하여 Bytesnet은 단일 플랫폼에서 제공하는 것 같이 고성능 데이터 워크플로우를 규모에 맞게 설계, 구현 및 운영할 수 있도록 지원함. 예를 들면, Bytesnet은 데이터 처리를 가속화하는 DDN의 병렬 파일 시스템인 EXAScaler를 활용하여 연구자들에게 필요한 대용량 GPU와 함께 더 큰 용량의 데이터를 활용할 수 있게 함. Bytesnet은 연구자들이 인프라 및 시스템 관리보다는 문제 해결에 집중할 수 있는 서비스를 제공하고 DDN은 이러한 엄격한 요구사항을 충족시키는 데 필요한 성능과 처리량을 제공하는 솔루션을 제공함. 양사는 고객들이 자신의 비즈니스를 확장하고 데이터 센터 이용 비용을 효율적으로 대응할 수 있도록하여 데이터 집약적인 연구에 전념할 수 있도록 하는 등 고객의 데이터 저장 및 배포 방식을 지속적으로 개선하기 위해 협력할 것임.		
결론 및 시사점	• AI 와 빅데이터 등 대용량 데이터 기반 R&D 활동 영역이 증가함에 따라 HPC 스토리지 분야에서도 연구자들이 필요할 때 사용한 만큼만 비용을 부담하는 pay per use 시스템이 도입되고 있음.		

표 37 Pengchen Lab 동향

제목	Pengchen Lab의 MadFS와 Intel의 DAOS가 최신 IO500에서 두각을 보이다 (Pengchen Lab's MadFS and Intel's DAOS Shine in Latest IO500)		
게시일	2021년7월7일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/2021/07/07/pengchen-labs-madfs-and-intels-daos-shine-in-latest-io500/		
요약	• (개요) 1년에 두 번 ISC (Intermediate Supercomputing Conference) 및 SC (Supercomputing Conference)에서 스토리지 시스템 성능을 측정하여 상위 500위까지의 스토리지 시스템들을 발표하는 *IO500 목록을 지난 주에 발표함. 이번 ISC 21에서의 발표에서 **Pengchen Laboratory가 ***MADFS를 사용하여 broad category와 10개 클라이언트 노드 category인 2개 부문에서 최고 영예를 수상함. Intel의 DAOS (Distributed Asynchronous Object Storage)는 두 범주 모두에서 2위를 차지하였는데 10노드 category 에서는 가장 높은 대역폭을 입증하였음. * IO500: A Storage Benchmark for HPC **Peng Cheng Laboratory (PCL)는 2018년 3월에 설립된 새로운 중국의 과학 연구 기관. 연구분야는 커뮤니케이션, 네트워킹 및 인공 지능. ***MADFS: Mobile Agent-based Distributed network File System • (접근방안 소개) IO500 결과를 분석하면 대역폭 및 메타데이터 처리 기능이 다양하게 검증된 스토리지 테스트의 핵심 지표임. 이번 발표는 예년의 성능 측정 결과물이 순위순으로 이미 목록에 포함되어 있었기 때문에 일종의 복합적인 테스트 결과를 보여주었음. IO500에서는 특정 시스템의 최고 점수도 같이 보고되는 데, 예를 들면 10노드 클래스에서는 동일한 작업을 수행하여 측정한 후 단일 시스템의 상위 점수를 모두 포함시켜서 발표하기 때문에 한 시스템이 복수의 점수를 가질 수 있음.		
결론 및 시사점	• 지금까지 저가의 스토리지만 생산하던 중국이 이제는 고품질의 HPC 스토리지 분야에서도 1위를 차지할 정도로 발전한 것을 알 수 있음.		

표 38 이탈리아 Pisa 대학 동향

제목	이탈리아 Pisa 대학의 스토리지 환경 살펴보기 (Unwrapping the storage environment at the University of Pisa)		
게시일	2021년7월12일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/2021/07/12/unwrapping-the-storage-environment-at-the-university-of-pisa/		
요약	• (개요) 오늘날의 대학에서는 연구를 위해 고성능 컴퓨팅 클러스터에서 처리되는 대량의 데이터가 필요한 경우가 많음. 이러한 데이터 집약적인 연구에서의 문제의 초점은 HPC/AI 관리자가 하나의 기술로 모든 요구 사항을 충족하지는 못한다는 것임. 이는 사용하는 스토리지 시스템에도 각별한 주의를 기울여야 한다는 것을 의미함. 왜냐하면 일부 연구 애플리케이션은 All-Flash 미디어를 사용하여, 스토리지 시스템의 I/O 성능이 매우 빠른 반면, 다른 애플리케이션은 회전식 Disk로도 충족할 수 있을 정도로 성능 요구사항이 낮기 때문임. 따라서 애플리케이션의 요구 사항을 성능 및 경제성 측면에서 적합한 스토리지 솔루션으로 충족하는 것도 중요한 관건이 됨. • (접근방안 소개) 피사 대학은 *VMware 환경의 중앙집중식 스토리지를 위해 **Dell EMC사의 ***PowerScale 스토리지를 선택함. 이전에는 대용량의 연구 데이터를 여러 개의 사일로(siloes)에 분산 저장함. 하지만 지금은 가상 HPC 시스템에서 액세스하는 통합 중앙 스토리지 시설로 전환하고 있음. PowerScale 플랫폼은 기존 플랫폼의 3배에 달하는 성능을 제공하면서도 안정성과 안정성이 매우 우수함. 이러한 특성으로 인해 PowerScale은 정밀 연구 시뮬레이션에 적합함. 이 가상 HPC 환경에서 일반적으로 호스팅되는 워크로드에는 엔지니어링 및 화학 시뮬레이션뿐 아니라 최신 AI 및 딥러닝 워크로드가 있음. *VMware: 클라우드 컴퓨팅, 가상화, 솔루션 전문 회사 **Dell EMC: 데이터 스토리지, 서버, 데이터 보안, 네트워킹 전문 회사 ***PowerScale: 테라부터 페타 바이트까지 유연하게 지원하는 기업용 스토리지		
결론 및 시사점	• HPC 스토리지에서도 다양한 응용의 특성에 맞게 지원하면서도 가성비를 높이기위하여 HPC 스토리지 가상화가 보편적으로 적용 및 보급되고 있음을 알 수 있음.		

표 39 EU 프로젝트 ADMIRE 동향

제목	EU 프로젝트 ADMIRE가 지능형 어댑티브 스토리지 시스템 개발을 위해 출범함 (EU Project ADMIRE Launched to Develop Intelligent Adaptive Storage Systems)		
게시일	2021년7월19일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/eu-project-admire-launched-to-develop-intelligent-adaptive-storage-systems/		
요약	• (개요) 인공지능에 필요한 방대한 양의 데이터 처리에 고성능 컴퓨터의 사용이 증가하지만 컴퓨팅 집약적인 작업에 초점을 맞춰 개발된 기존의 컴퓨팅 방식에서는 문제가 되어가고 있음. 즉 대용량의 병렬 파일 시스템을 사용하기에는 기존의 일차원적 플랫(flat)형 스토리지 계층 구조 기반 데이터 저장 방식은 점점 더 불편해지고 있음. SSD (solid-state drive, 반도체 스토리지)와 같이 빠르고 비휘발성인 동시에, 에너지 효율적인 대용량 스토리지를 제공하는 신기술이 데이터 집약적인 애플리케이션 지원에 선택되고 있음. 전통적 방식의 기존 스토리지 자원과 새로운 방식인 SSD 잠재력을 통합하여 제공할 수 있고 고객 맞춤 방식의 파일 시스템의 구축 및 이를 위한 적절한 제어 메커니즘이 현재는 부족한 상태임. • (접근방안 소개) 상기 문제를 해결하기 위하여 EU에서는 *ADMIRE 프로젝트를 추진함. 유럽 6개국 14개 기관이 고성능 컴퓨팅을 위한 새로운 적응형 스토리지 시스템을 개발함. 기상 예측, 분자 역학, 난류 모델링, 지도 제작, 두뇌 연구 및 소프트웨어 카탈로그 작성과 같은 분야에서 응용 프로그램의 실행 시간을 크게 개선하는 것이 목표임. 마드리드 카를로스 3세 대학교(UC3M)가 중심이 되고 유럽 고성능 컴퓨팅 공동 사업(EuroHPC JU: European High-Performance Computing Joint Undertaking)과 참여 국가들로부터 자금을 조달받음. * ADMIRE(Adaptive multi-tier intelligent data manager for Exascale)는 EU Horizon 2020 연구 및 혁신 프로그램에서 지원하는 프로젝트. 2021. 4. 1. 출범. 3년간 790만 유로(약 1억5000만 원)의 예산 지원.		
결론 및 시사점	• AI를 위한 빅데이터 컴퓨팅에 최적화된 통합형 HPC 스토리지를 공동으로 개발하기 위한 EU 프로젝트를 2021년 4월에 출범시킴. 우리도 통합형 국가 HPC 스토리지 서비스 인프라 구축이 필요함 .		

표 40 Supermicro사 동향

제목	Supermicro사는 3세대 Intel Xeon 프로세서를 사용하여 새로운 스토리지 시스템을 출시함 (Supermicro Debuts New Storage Systems with 3rd Gen Intel Xeon Processors)		
게시일	2021년7월27일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	hhttps://www.hpcwire.com/off-the-wire/supermicro-debuts-new-storage-systems-with-3rd-gen-intel-xeon-processors/		
요약	• (개요) 대용량 스토리지 확장, HPC 스토리지 구현 그리고 클라우드 규모의 스토리지 구성을 위하여 소프트웨어 정의 클라우드에 기반한 스토리지 시장 성장이 가속화됨에 따라 데이터 센터 인프라를 신속하게 현대화하고, 기술자가 이를 서비스할 수 있게 하는 유연한 스토리지 구성 필요성이 증가하고 있음. 이를 위한 새로운 스토리지 시스템의 고객 요구조건은 스토리지 시스템을 특정 도구가 필요 없는 모듈식 설계, 리소스 절약 기능, 운영비를 포함한 낮은 TCO(Total Cost of Ownership)을 지원하면서 최고의 용량을 제공하는 것임. • (접근방안 소개) *Super Micro Computer사는 최신 3세대 Intel Xeon 확장형 프로세서와 PCI-E 4.0 **NVMe 드라이브에 최적화된 Simply Double 스토리지 시스템과 함께 60-bay 및 90-bay 시스템을 발표함. 이 시스템은 대용량을 위하여 탑로드 지원, 스케일업 및 스케일아웃 아키텍처 설계, PCI-E 4.0 기반 RAID(Redundant Array of Independent Disks) 또는 IT 모드 SAS(Serial Attached SCSI) 컨트롤러의 구성 지원. 핫 스왑 2.5인치 또는 3.5인치 SAS 3 / SATA 3 베이, 2개의 내부 슬림 ***SATA SSD 슬롯 지원. 최대 1.6 페타 바이트를 저비용 스토리지 (JBOD: Just a Bunch Of Disks)로 지원. 후면 액세스 NVMe를 통해 최대 60TB의 SSD 플래시 지원. 서버 노드당 16개의 ****DIMM 슬롯이 있는 이중 소켓 구성으로 3세대 Intel Xeon 프로세서로 확장이 가능함. *Super Micro Computer: 엔터프라이즈 컴퓨팅, 스토리지, 네트워킹 솔루션 전문 기업. **NVMe: Non-Volatile Memory express, ***SATA: Serial Advanced Technology Attachment, ****DIMM: Dual In Line Memory Module,		
결론 및 시사점	• HPC 스토리지를 저비용으로 구축하기 위한 솔루션들이 경쟁적으로 출시하고 있음을 알 수 있음.		

표 41 미국 샌디에고 슈퍼컴센터 동향

제목	미국 샌디에고 슈퍼컴센터는 개방형 사이언스 체인 구축을 위하여 미국 과학재단으로부터 $500,000를 지원받음 (SDSC's Open Science Chain Awarded $500,000 NSF Grant)		
게시일	2021년8월4일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/sdscs-open-science-chain-awarded-500000-nsf-grant/		
요약	• (개요) *SDSC(San Diego Super Computer Center)는 과학 데이터의 재사용에 필요한 개인 정보 보호를 지원하면서 효율적으로 데이터와 메타데이터를 공유하고 검증할 수 있는 안전한 방법을 제공하기위한 OSC (Open Science Chain) 프로그램에 50만 달러의 연구비를 NSF (National Science Foundation)로부터 승인받음. OSC는 연구 플랫폼 및 허브에서 퍼블릭 및 프라이빗 데이터 자산의 무결성과 검증을 보존하기 위해 **블록 체인 기술을 적용하여 "개방형 사이언스 체인 - 무결성 서비스 (OSC-IS: OSC-Integrated Service)"를 제공코자함. *SDSC (San Diego Super Computer Center): 고성능 및 데이터 집약적 컴퓨팅 분야의 선두적인 미국의 슈퍼컴 센터. **블록 체인: 분산 스토리지 형태로 데이터를 저장하면서 위조나 해킹이 불가능하도록 데이터 활용 이력을 공개하는 형식의 데이터 공유 기술 • (접근방안 소개) OSC-IS는 스마트 계약과 오프체인 스토리지의 조합을 활용하여 공동 연구에서 민간 연구 데이터를 공유할 때 데이터 기밀성과 프라이버시를 보장하도록 설계함. 이 프로젝트는 데이터 중심 연구 플랫폼 및 허브에 대한 API(Application Program Interface) 기반 데이터 무결성 검증 관리 서비스를 구축하여 데이터 정보 손실을 줄이고 다양한 메타데이터 표준 및 액세스 제어를 관리할 수 있도록 함. 이 프로젝트는 기존의 과학 데이터 관리 분야가 나아갈 시범 사례로 확대하여, 학술 연구에 블록체인을 활용하는 동시에 전반적인 교육 분야에 컴퓨터 보안을 접목하는 것을 보여줄 것임. 미국 NSF 국립보건연구소 또는 기타 공공 기관에서 연구비를 지원하는 프로젝트 등에 참여하는 연구자들은 데이터 검증을 통한 안전한 데이터 공유에 관심이 매우 높음. OSC-IS 프로젝트는 이러한 분산 데이터 공유 분야에서의 연구자들의 의문점을 해결할 것임.		
결론 및 시사점	• 대규모의 데이터를 안전하게 공유할 수 있도록 분산형 HPC 스토리지에도 블럭체인 기술을 도입하기 시작함.		

표 42 호주 Pawsey 슈퍼컴퓨팅 센터 동향

제목	Pawsey 슈퍼컴퓨팅 센터는 130 PB 규모의 다중 계층구조의 HPC 스토리지 인프라를 구축 중임. (Pawsey Supercomputing Centre to Deploy 130PB of Multi-Tier Storage)		
게시일	2021년8월16일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/pawsey-supercomputing-centre-to-deploy-130pb-of-multi-tier-storage/		
요약	• (개요) *Pawsey Supercomputing Center는 총 7천만 달러 규모의 HPC 인프라 구축 프로젝트의 일환으로 130PB (PetaByte: 10^{15} Byte) 규모의 세계 최고 수준의 다중 계층의 스토리지를 구축할 예정임. Pawsey는 데이터 가용성, 데이터 전송 속도 및 전체 스토리지 용량을 향상시키기 위한 전체 예산 중 7백만 달러의 예산을 객체형 스토리지 계층을 포함하는 다중 계층형 HPC 스토리지 인프라 구축을 위하여 **Dell사 그리고 ***Xenon사와 스토리지 도입 계약을 체결함. *Pawsey Supercomputing Center: 호주 슈퍼컴센터. **Dell: 컴퓨터 및 소프트웨어 등 미국의 전자제품 제조 및 판매 기업 ***Xenon:: 호주의 슈퍼컴, 스토리지, AI 솔루션 전문 기업 • (접근방안 소개) - 구축 목표인 130 PB 규모의 HPC 스토리지 인프라는 기존의 15 PB 규모의 하드 디스크와 70 PB 규모의 테이프 저장장치에 신규로 60 PB 규모의 Object 스토리지인 Ceph 소프트웨어를 도입하여 구축함. 최종 스토리지 인프라의 모습은 5000개의 16 TB(10^{12} Byte) 디스크로 구성되는 데이터 저장소, 메타데이터 스토리지용으로 400개 이상의 NVME (Non-Volatile Memory express) 디스크 할당. 그리고 인프라 전체 측면에서의 데이터 전송 능력의 총합을 400Gb/s 이상으로 지원함. 오픈 파일 시스템인 Versity 파일 시스템용으로 5PB 규모의 프런트 엔드 캐시 스토리지를 지원함. 그리고 32개의 IBM TS1150 고성능 엔터프라이즈 테이프 라이브러리 드라이브와 10,000개 이상의 미디어 슬롯을 제공하는 2개의 IBM TS1160 차세대 드라이브로 대용량 데이터 처리를 지원함.		
결론 및 시사점	• HPC 스토리지의 고도화를 위하여 스토리지 인프라의 다중 계층화 과정에서, 스토리지 가상화를 담당하는 계층에 객체 스토리지인 Ceph 스토리지 소프트웨어 도입이 전 세계의 흐름임을 알 수 있음		

표 43 Panasas사 동향

제목	Panasas사는 차세대 휴대용 스토리지 플랫폼인 PanFS9을 출시함 (Panasas Delivers PanFS9, Its Next-Generation Portable Storage Platform)		
게시일	2021년8월25일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	hhttps://www.hpcwire.com/off-the-wire/panasas-delivers-panfs9-its-next-generation-portable-storage-platform/		
요약	• (개요) *Panasas사는 고객에게 성능향상, 최적화 및 다중 계층 보안 기반을 제공한 **PanFS를 차세대 휴대용 병렬 파일 시스템으로 진화시킨 버전 PanFS9을 발표함. 이러한 발전은 오늘날 HPC 및 AI/ML 워크로드의 까다로운 요구사항을 충족하는 동시에 중요하고 기밀스러운 데이터를 보호하고 안전한 다자간 사용환경을 제공코자함. *Panasas: HPC 환경을 위한 네트워크 연결형 스토리지 인프라를 구축 솔루션을 제공하는 데이터 스토리지 전문 회사 **PanFS: 객체형 스토리지를 사용하는 병렬 파일 시스템 • (접근방안 소개) - PanFS 버전 9은 기존의 ActiveStor Ultra 스토리지 플랫폼 기능을 확장하여 성능 저하나 비용 영향 없이 유휴 하드웨어 기반 암호화를 비롯한 우수한 보안 계층과 Linux 커널 보안 수정 세트인 SELinux (Security-Enhanced Linux)에 대한 파일 레이블 지정을 지원함. - PanFS 9는 물리적 데이터 계층에서 자체 암호화 드라이브 하드웨어 기반 AES (Advanced Encryption Standard) 256비트 암호화를 사용하여 ***NIST가 승인한 유휴 데이터 보호 기능을 지원하고, 드라이브를 제거하는 즉시 추가 인증없이 자동으로 암호화 삭제를 지원함. 그리고 지능형 HPC 융합 및 AI/ML 워크로드 처리를 위해 하이브리드 HPC 스토리지에 대한 새로운 표준인 DDA(Dynamic Data Acceleration) 기반 스토리지 및 메타데이터 노드의 모든 소프트웨어 스택에 대한 성능 최적화 및 향상된 기능을 지원함. ***NIST: 미국 국립표준기술연구소		
결론 및 시사점	• 인공지능과 머신러닝 기술이 많은 HPC 융합 응용에 적용되면서 이를 위한 HPC 스토리지 플랫폼의 이식성 및 보안 기능이 강화된 스토리지 플랫폼이 출시되고 있음.		

표 44 미국 HPE 사의 GreenLake 동향

제목	HPE는 미국 NSA와 20억 달러 규모의 GreenLake 서비스형 HPC 클라우드 구축 사업을 획득함 (HPE Wins $2B GreenLake HPC-as-a-Service Deal with NSA)		
게시일	2021년9월1일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/2021/09/01/hpe-wins-2b-greenlake-hpc-as-a-service-deal-with-nsa/		
요약	• (개요) 고객들은 대규모 데이터 집약적인 프로젝트에서 HPC 기능을 쉽고 단순하한 사용과 민첩한 관리 결합을 요구함. 미국 *NSA는 안전한 사내 솔루션을 서비스로 제공하는 HPE GreenLake 플랫폼을 사용하여 다양한 복잡한 데이터 요구사항을 유연하게 해결할 수 있는 보안이 강화된 HPC 클라우드 솔루션을 확보하고자 함 *NSA (National Security Agency):미국 국방부 소속의 국가안보국(암호해독 첩보국)으로 해외 통신과 해외 신호 정보의 수집과 분석을 책임짐. . **HPE (Hewlett Packard Enterprise):: 미국의 기업용 클라우드 서비스 관련 제공 회사 ***GreenLake: HPE 사에서 제공하며 위치에 구애받지 않는 애플리케이션 및 데이터를 위한 클라우드 서비스를 제공함. • (접근방안 소개) - 2017년 11월 출시된 ***GreenLake는 다양한 업종에서 수행하는 중요한 워크로드에 사용량에 기반하여 IT 서비스를 제공하기 위한 HPE의 클라우드와 같은 data lake 서비스형 HPC 클라우드 솔루션임. HPE는 GreenLake를 가장 빠르게 성장하는 사업부로 사업 정책을 추진함. 2020년 후반까지 계약 금액은 4.5달러였으며 매년 40%씩 증가하였음 - GreenLake 서비스 포트폴리오는 가상 머신, 컨테이너, 하이브리드 클라우드 통합 및 베어 메탈을 포함하며 계속 확장되고 있슴. 2020년 12월, HPE는 HPC 사전 구성 서비스를 통하여 HPC 프로젝트의 구현 속도를 최대 75%까지 높이고 자본 지출을 최대 40%까지 줄임. ***GreenLake: HPE 사에서 제공하는 위치에 구애받지 않는 애플리케이션 및 데이터를 위한 클라우드 서비스.		
결론 및 시사점	• HPC 서비스를 인공지능과 머신러닝에 쉽게 적용하기위하여 대용량의 데이터를 안전하고 쉽게 처리하도록 DATA LAKE 서비스를 고도화함		

표 45 Spectra Logic사 테이프 저장 장치 동향

제목	Spectra Logic사는 테이프 라이브러리 솔루션 제품군으로 LTO-9 유용성을 발표함 (Spectra Logic Announces LTO-9 Availability for its Family of Tape Library Solutions)		
게시일	2021년9월8일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/spectra-logic-announces-lto-9-availability-for-its-family-of-tape-library-solutions/		
요약	• (개요) 디스크와 클라우드를 이미 보유하고 있어도 데이터가 대규모로 증가를 하면 기관은 테이프 스토리지를 다시 구축하려함. 테이프는 테라바이트당 최저 비용으로 탁월한 안정성과 함께 최고의 스토리지 용량을 제공하기 때문임. LTO-9 (Linear Tape Open-버젼 9)과 같은 테이프 기술은 테이프 미디어를 오프라인 모드로 보관을 하면 랜섬웨어에 대하여 뛰어난 보호 기능을 제공하여 악의적 해커로부터 데이터를 보호할 수 있음 *Spectra Logic: 데이터 관리 및 데이터 스토리지 솔루션 전문 기업 • (접근방안 소개) - 새로운 9세대 LTO(Linear Tape-Open) 테이프 기술은 훨씬 더 높은 데이터 스토리지 밀도와 속도를 제공하므로 테이프 카트리지당 최대 18TB(10^{15} 바이트)의 기본 용량(최대 45개)을 제공함. 이전 버전인 LTO-8 보다 용량이 50% 증가함. LTO-9 드라이브는 비압축 전송 속도를 최대 400MB/초까지 제공하여 현재 하드 디스크 드라이브보다 거의 2배 더 빠름. 단일 테이프 라이브러리 장치에서 3 종류의 테이프 기술(LTO, IBM의 TS 11xx 및 Oracle의 T10000)을 지원함. 테이프 장치 프레임을 3개에서 45개까지 확장 가능함, 56,000개 이상의 테이프미디어 슬롯을 제공하여 대량의 데이터 로드가 가능함. 고성능 및 안정성 향상을 위한 첨단 이중 로봇 공학을 지원함. 타사 소프트웨어 패키지와 함께 사용할 수 있는 최대 호환성 과 고객 브랜드를 보여주는 사용자 지정 전면 라이브러리 패널을 제공함.		
결론 및 시사점	• 과거의 기술 또는 사양 산업이 될 수 있는 테이프 저장장치가 최근 AI 와 머신러닝으로 인한 대규모 데이터의 수요가 늘어남에 따라, 사용자 환경에 맞도록 기능이 개선되면서 재구축되고 있음.		

표 46 Kioxia 사 동향

제목	Kioxia 사는 PCIe 4.0 기반의 스토리지 급 SSD 메모리를 출시함 (Kioxia Introduces PCIe 4.0 Storage Class Memory SSDs)		
게시일	2021년9월14일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/spectra-logic-announces-lto-9-availability-for-its-family-of-tape-library-solutions/		
요약	• (개요) *XL-Flash인 **Kioxi 사의 혁신적인 ***BiCS FLASH 3D 플래시 메모리는 셀당 1비트인 SLC(Single Level Cell in SSD)로서 데이터 센터 및 엔터프라이즈 스토리지에 짧은 대기 시간과 높은 성능을 제공함. DRAM(Dynamic Random Access Memory)과 같은 휘발성 메모리 솔루션은 까다로운 응용에 필요한 액세스 속도를 제공하지만 비용이 많이 듦. 따라서 고밀도의 비휘발성 플래시 메모리인 SCM(Storage Class Memory)을 사용하여 비용 효율 문제를 해결함. *XL-Flash: 일본 도시바의 낸드 플래시 상표 **KIOXIA : 반도체 메모리 기반 스토리지 솔루션 전문 회사 ***BiCS FLASH : 112개층의 적층형 3차원 플래시 메모리; • (접근방안 소개) - Kioxia의 SCM 솔루션인 KIOXIA FL6 시리즈는 이중 포트 및 ****PCIe 4.0을 지원하는 SSD(Solid State Drive)로서 DRAM과 TLC(Triple Level Cell) 기반 드라이브 간의 격차(gap)를 해소하여 캐싱 계층, 계층화 및 쓰기 로깅(logging)과 같은 대기 시간에 민감한 응용 서비스에 적합함. 그리고, 대기열(queue) 깊이가 작은 워크로드에서 우수한 성능을 제공하지만 워크로드가 점점 더 까다로워지고 혼합적일 때 강점이 더 드러남. 이러한 환경에서 KIOXIA FL6 드라이브는 다양한 지연 시간에 민감한 애플리케이션에 높은 신뢰성을 제공함. 60 DWPD(Drive Writes Per Day), 800GB ~ 3.2TB 용량, 250만 시간의 MTBF(Mean Time Between Failure)의 신뢰성과 SED Self-Encrypting Drive) 및 *****FIPS 140-2 보안 옵션을 지원함. ****PCIe 4.0: Peripheral Component Interconnect express 4.0으로서. 컴퓨터 메인보드에서 각종 주변기기를 직렬로 연결하는 규격. 최대 64GB/s 속도 지원. *****FIPS 140-2: 미국 연방정부 정보처리 표준 140조-2항. IT 제품이 기밀은 아니지만 민감한 용도로. 사용될 때 충족해야 할 암호화 및 관련 보안 요건을 규정함		
결론 및 시사점	• AI와 머신러닝을 위한 HPC 스토리지 구축에서 대규모 데이터를 매우 빠르게 처리하기위하여 PCIe 4.0와과 SSD 접목 기술이 대두되고 있음		

표 47 펜실베니아 주립 대학 Huck 연구소 동향

제목	펜실베니아 주립 대학 Huck 연구소는 MemVerge사의 빅 메모리 기술을 활용하여 생명과학 분야의 게놈 분석 프로세스를 가속화함 (MemVerge Big Memory Technology Speeds Genomic Analytic Processes for Penn State Huck Institutes of the Life Sciences)		
게시일	2021년9월21일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/memverge-big-memory-technology-speeds-genomic-analytic-processes/		
요약	• (개요) 메모리는 너무 작고 스토리지는 너무 느림. 이에 대응하여, *MemVerge사에서는 대용량 스토리지에 메모리의 나노초 성능을 동시에 제공하는 새로운 빅 메모리 컴퓨팅 영역을 개척함. 이는 펜 스테이트의 게놈 염기서열 분석 연구를 위하여 메모리보다 더 큰 데이터로 인한 문제를 가성비 문제와 같이 해결하기 위함임. *MemVerge: 세계 최초로 빅 메모리 기술을 DRAM과 영구 메모리로 확장한 기업 • (접근방안 소개) - MemVerge 메모리 장비의 소프트웨어는 서로 다른 유형의 메모리 하드웨어를 소프트웨어 정의 메모리 풀(pool)로 투명하게 가상화함. DRAM의 성능은 동일하지만 용량은 여러 배임. 투명한 메모리 서비스 외에도 메모리 장비는 페타바이트 단위의 용량, 성능, 가용성 및 이동성을 메모리 속도와 클라우드 전반에서 프로비저닝할 수 있는 업계 최초의 데이터 서비스 제품군을 제공함 - 메모리 장비 소프트웨어는 Penn State의 과학자들에게 가상화된 DRAM(Dynamic Random Access Memory) 및 PMEM(Persistent Memory) 메모리와 **ZeroIO 스냅샷에 대한 투명한 액세스를 제공함. 인메모리 스냅샷 기술을 통해 연구자들은 적응형 분석을 위해 워크로드를 특정 시점으로 신속하게 롤백하고 번개처럼 빠른 충돌 복구로 복원력을 보장할 수 있음. 이를 통해 새로운 게놈 테스트를 중단 시간 없이 계속할 수 있음. **ZeroIO snapshot: storage IO 없이 응용 프로그램의 메모리 page를 복구하는 기술		
결론 및 시사점	• 내장형 메모리 수준의 초고속 HPC 스토리지 서비스를 위하여 소프트웨어 정의 메모리 (Software-Defined Memory) 기술을 도입하여 세계 최초로 빅 매모리 장비를 개발한 사례 소개		

표 48 Adaptive Computing사 동향

제목	Adaptive Computing사는 주문형 데이터센터 구축을 지원하는 ODDC 6.0 의 GA 버전 출시 (Adaptive Computing Announces the GA Release of the On-Demand Data Center 6.0 Cloud Enablement Platform)		
게시일	2021년9월29일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/adaptive-computing-announces-the-ga-release-of-the-on-demand-data-center-6-0/		
요약	• (개요) HPC를 이용하기 어려운 중소기업은 클라우드 데이터 센터의 HPC 자원을 의존할 수밖에 없음. HPC 클라우드를 이용하면 사내 시스템이 없어도 AI, 머신러닝, 딥러닝, 데이터 분석, 시뮬레이션 등 강력한 컴퓨팅 집약적 애플리케이션을 실행할 수 있음. 따라서 중소기업은 고가의 HPC 하드웨어를 구입하는 대신 임시로 클라우드 인프라 리소스를 사용하다가 클라우드 수요가 최고조에 이를 때가 되어야 사내 시스템을 확장하거나 컴퓨팅 용량 확장을 퍼블릭 클라우드 공급업체와의 계약을 추가하는 것을 선호함. • (접근방안 소개) - HPC 클라우드는 중소 규모의 고객이 구축하는 데이터 센터가 특정 벤더에 종속되지 않고 주문(ODDC*) 기반으로 HPC 클라우드를 동적으로 확장 및 축소 구성이 가능한 솔루션을 제시해야 함. - Adaptive Computing** 사는 Oracle Cloud Infrastructure(OCI)를 활용하여 고객에게 사내 시스템 수준으로 OCI에서 작업을 설정하고 실행케하는 지능형 클라우드 관리 플랫폼의 GA*** 버젼을 개발 출시함. 이 버젼은 고성능의 베어 메탈(기본형) HPC 컴퓨팅 인스턴스와 클러스터 네트워킹을 제공하여 다른 어떤 퍼블릭 클라우드 및 데이터 센터와도 연결이 가능하면서 통합 성능 또한 크게 향상시켜줌. *ODDC(On-Demand Data Center): Adaptive Computing 사의 주문기반 자원서비스 **Adaptive Computing : 미국의 HPC 설치 및 관리 전문 회사 ***GA(General Availability): 모든 테스트가 끝난 개발의 마지막 버전		
결론 및 시사점	• 중소기업이 AI, 머신러닝 등 HPC 관련 사업을 시작부터 데이터 증가에 따른 성장 과정까지 연속적으로 지원할 수 있도록 벤더 중립성과 단계별 시스템 확장성을 동시에 지원하는 솔루션 출시		

표 49 Quantum사 동향

제목	Quantum사는 새로운 형태의 스토리지 및 완벽한 관리 객체 스토리지 서비스를 위한 신제품군을 발표함 (Quantum Announces New Class of Storage and New Line of Fully Managed Object Storage Services)		
게시일	2021년10월5일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/quantum-announces-new-class-of-storage-and-new-line-of-fully-managed-object-storage-services/		
요약	• (개요) IDC* 사는 전세계적으로 생성되는 비정형 데이터의 양이 매년 30%씩 기하급수적으로 증가함에 따라 세계는 새로운 데이터 시대를 예측함. 따라서, 데이터 증가와 스토리지 예산 간의 격차는 점점 커지는 데, 생성되는 데이터의 60%가 cold data일 것으로 추산됨. cold data 란 법규정 상의 요구사항과 데이터가 갖고 있는 막대한 가치로 인해 사용 빈도는 낮아도 수년 또는 수십 년 동안 저장 및 보호를 해야하고, 사용자 이용은 항상 가능케 사용자 액세스를 열어 놓아야 하는 비활성 데이터를 나타냄. 기존 스토리지 아키텍처 상에서는 대부분의 기업이 이러한 데이터 폭증에 대처할 예산, 설치 공간 또는 리소스를 확보하지 못해 데이터센터 운영 또는 예산을 절충해야 함. *IDC(International Data Corporation): IT 및 통신, 컨수머 테크놀로지 부문 세계 최고의 시장 분석 및 컨설팅 기관 • (접근방안 소개) - 이러한 긴급한 과제를 해결하기 위해 Quantum**사는 콜드 데이터 용으로 설계되고 하이퍼스케일 아키텍처를 기반으로 한 as-a-service 형태의 서비스를 제공하는 새로운 종류의 객체 스토리지인 ActiveScale 콜드 스토리지를 출시함. ActiveScale 콜드 스토리지는 객체 스토리지 소프트웨어와 2차원 삭제 코딩 및 Quantum Scalar 테이프 스토리지를 결합하여 안전하고 내구성이 뛰어나며 매우 저렴한 새로운 차원의 스토리지를 제공함. **Quantum: 데이터 관리 및 보존과, 데이터 분석 솔루션을 제공하는 미국 기업		
결론 및 시사점	• AI와 머신 런닝을 위한 데이터가 급증하면서 대규모 데이터의 장기 관리를 위한 HPC 스토리지 인프라의 새로운 확장 방향의 한 지류를 발견할 수 있음.		

표 50 일본 후지(FUJY)사의 테이프 스토리지 동향

제목	후지(FUJY)사는 지속 가능한 데이터 스토리지 이니셔티브를 착수함 (FUJI Launches Sustainable Data Storage Initiative)		
게시일	2021년10월14일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/fuji-launches-sustainable-data-storage-initiative/		
요약	• (개요) FUJIFILM 사는 테이프 기술을 데이터 스토리지와 관련된 전기 소비량과 CO2 배출량을 줄이는 기술로 재등장 시키기위하여 “지속 가능한 스토리지 전략으로 그린 데이터 센터 진행 가속화” IDC** 백서 발간을 통하여 이니셔티브를 시작함. *FUJIFILM: 후지카메라, 필름제작, 충전지, 사진용품 관련 일본 기업. **IDC: 장비, 응용, 네트워크, 서비스 등 고객의 시장을 조사하는 미국 기업. • (접근방안 소개) - IDC 보고에 따르면 2017년부터 2020년까지 데이터센터의 에너지 소비는 31% 증가했으며 저장 데이터 양은 2025년까지 매년 27%씩 증가할 것으로 예상함. - 데이터 센터 전력 소비의 급격한 증가를 따라잡기 위한 수력, 풍력 및 태양열 발전은 환경에도 자체적인 영향을 미침(예, 풍력 터빈과 태양 전지판의 폐기). - 전력 및 환경 문제를 해결하기 위하여 IDC는 전략적으로 더 많은 데이터 스토리지를 테이프로 마이그레이션하면 2030년까지 이산화탄소 배출량을 43.7%(6억 6천 4백만 톤) 줄일 수 있음을 제시. - 테이프 스토리지는 사이버 범죄자에 대한 불변성, 암호화 기능 및 오프라인 보호와 같은 추가적인 보안 이점을 제공함. - 테이프는 기가 바이트당 비용 기준으로 가장 비용 효율적인 스토리지 미디어이며, 작동 시 최소 전력이 필요하며, 뛰어난 비트 오류율(BER)로 30년이 넘는 기간 동안 데이터를 저장할 수 있음.		
결론 및 시사점	• HPC 스토리지에서 경쟁력을 잃어가던 테이프 장비가 인공지능과 머신런닝 데이터 증가로 인한 늘어난 데이터센터 에너지의 절약과 탄소 문제 해결 수단으로 재등장하는 것을 알 수 있음		

표 51 Exxact사 동향

제목	Exxact사는 SoftIron사와 파트너 관계를 맺고 Ceph 기반 소프트웨어 정의 스토리지 솔루션 제공 (Exxact Partners with SoftIron to Provide Ceph-Based Software Defined Storage Solutions)		
게시일	2021년10월19일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/exxact-partners-with-softiron-to-provide-ceph-based-software-defined-storage-solutions/		
요약	• (개요) 점점 더 복잡해지는 AI와 HPC 데이터센터에 대한 기업의 요구를 충족하기 위해 데이터센터들은 빠르게 변화함. 따라서 미래 데이터센터 경쟁력은 성능, 효율성, 보안 및 사용 편의성의 저하 없이 민첩하고 안정적이며 확장 가능한 운영 능력에 달려 있음. • (접근방안 소개) - SoftIron*사의 대표적인 HyperDrive** 엔터프라이즈 스토리지 솔루션은 내구성, 강력한 복원력 및 무한 확장성이 뛰어난 오픈 소스 SDS(Software Defined Storage) 플랫폼인 Ceph***의 성능을 최적화한 것임. 블록, 파일 및 객체 스토리지를 지원하는 Ceph는 데이터 센터의 미래 스토리지 인프라 솔루션임. 왜냐하면, HyperDrive는 내부 전문가가 필요 없을 정도로 단순하고 간편하게 클러스터 관리 및 운영을 지원함. - Exxact****사는 완벽하게 통합된 데이터 센터 솔루션을 고객에게 제공하고 지원할 수 있는 SoftIron사의 공식 파트너사로서 Ceph 스토리지를 직관적으로 사용하기 쉽고, 확장 가능한 우수한 작업별 어플라이언스를 제공하여 간편한 사용 편의성 및 간소화된 운영을 제공함. *SoftIron사 : HyperDrive를 제작 판매 하는 비상장 미국 기업 **HyperDrive : 기업을 위한 Ceph형 스토리지를 간단하게 구축케하는 제품명 ***Ceph: Software Defined Storage 서비스를 제공하는 오픈 소스 솔루션 중 하나. ****Exxact사 : HPC 분야 인프라 구축 및 SI (System Integration) 전문 미국 기업		
결론 및 시사점	• 기존의 성능 중심의 HPC 스토리지 구축에서 사용자 편의성까지 고려한 소프트웨어 정의형 HPC 스토리지 도입으로 HPC 비전문가도 인공지능 및 머신런닝 데이터를 활용할 수 있게 됨.		

표 52 Lightbits사 동향

제목	Lightbits 연구소는 인텔의 혁신기술인 인프라 프로세싱 유닛 상에서 소프트웨어 정의 NVMe 스토리지의 장점을 입증함 (Lightbits Labs to Demonstrate Advantages of Software-Defined NVMe* Storage on Infrastructure Processing Unit at Intel Innovation**)		
게시일	2021년10월27일	출처	HPCwire
구분	인프라구축동향	유관부서	대용량데이터허브센터
원문보기	https://www.hpcwire.com/off-the-wire/lightbits-labs-to-demonstrate-advantages-of-software-defined-nvme-storage/		
요약	• (개요) 데이터 센터에서 기존의 스토리지 아키텍처를 사용하여 서버 팜을 관리할 때 고객의 예측 불가능한 수요에 따라 로컬 스토리지를 과도하게 제공하는 것이 일반적인데, 이는 데이터센터의 운영 비용을 많이 들게 함. • (접근방안 소개) - 이러한 비효율성은 단순하고 확장 가능한 소프트웨어 정의 기반 블록 스토리지 솔루션을 사용하여 해결이 가능한데 Lightbits*** 사는 가벼운 운영체제를 활용하여 인텔사의 IPU**** 기반 인프라를 구축한 클라우드 사업자가 인프라 작업에 할당된 CPU를 IPU로 이동시켜서 서버의 활용률을 높일 수 있게 함. 즉, 수익을 창출하지 않는 서버를 IPU로 이동하는 방법을 통하여 클라우드 사업자가 보유한 서버 CPU의 100%를 수익화하고, 고객의 필수 작업 처리에 성능을 집중시키고, 그리고 데이터 보안을 강화하는 것을 용이하게 함. Lightbits 사의 데모는 AI 가속기가 내장된 3세대 Intel Xeon Scalable 프로세서, Intel Optane 영구 메모리(PMEM) 및 100Gb Intel 이더넷 800 시리즈 네트워크 어댑터에서 수행되었음. *NVMe: Non-Volatile Memory express, 비휘발성의 반도체 메모리 **Intel Innovation: 가상화, 실시간 컴퓨팅, AI를 위하여 인텔사에서 제공하는 엣지, 클라우드 그리고 오픈 소스 친화적인 혁신 도구 또는 관련 발표 행사를 통칭함. *** Lightbits Labs: 프라이빗 및 에지 클라우드용 NVMe 기반, 확장 가능한 프트웨어정의 기반 EBS (아마존의 Elastic Block Storage) 전문의 미국 기업 ****IPU: Infrastructure Processing Unit: 인프라 자원의 효율성과 관리 용이성을 위하여 인텔사에서 제공하는 프로그램 가능한 컴퓨팅과 네트워크 자원 서비스.		
결론 및 시사점	• 인공지능 및 머신 런닝으로 인한 데이터의 증가로 데이터 센터의 스토리지 인프라 규모가 증대하면서 스토리지 자원을 요구 기반으로 할당하기위한 기술들이 혁신 기술로 소개 되고 있음.		

발 행 처 한국과학기술정보연구원
발 행 일 2021년 11월
주 소 (34141) 대전광역시 유성구 대학로 245
대표전화 042) 869-1004, 1234
팩 스 042) 869-1091
홈페이지 https ://www.kisti.re.kr

• 본 보고서의 내용은 비상업적 이용만 가능하며 변형 등 2차적 저작물의 작성을 금지합니다.
• 본 보고서의 활용은 KISTI 저작권 정책
(www.kisti.re.kr/pageView/521?t=1568788952964)을 준수하시기 바랍니다.

한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

본원
[34141] 대전광역시 유성구 대학로 245
TEL. 042) 869-1004,1234
FAX. 042) 869-1091

분원
[02456] 서울특별시 동대문구 회기로 66
TEL. 02) 3299-6114
FAX. 02) 3299-6244

OPEN 공공누리
출처표시 상업용금지 변경금지
공공저작물 자유이용허락

비매품/무료
93560
9 788929 412043
ISBN 978-89-294-1204-3