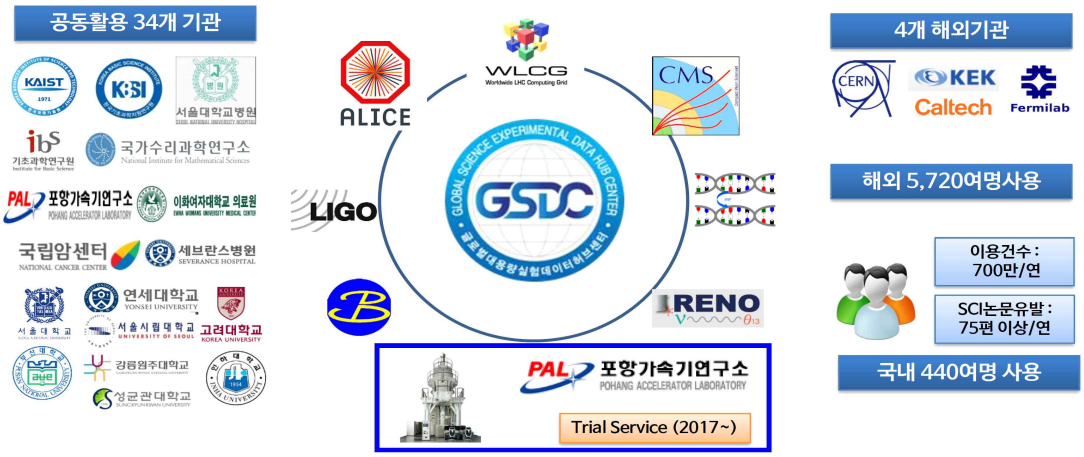
GSDC 구조생물학분야 TEM 분석 팜 소개

('20.2. 한국과학기술정보연구원 국가슈퍼컴퓨팅본부 대용량데이터허브센터)

□ 글로벌 대용량데이터허브센터(GSDC) 소개

- 데이터집약형 기초과학 실험지원을 위한 한국형 통합 데이터센터 역할
- 약 10,000+ 코어, 10+ PB 스토리지 및 첨단연구망 연계 기초과학 데이터 통합 인프라 구축·운영
- 세계 주요 가속기 및 대형 검출기 등에서 생성된 대용량의 실험데이터에 대한 저장·공유·분석 환경 구축 및 연구자 맞춤형 서비스를 위한 연구개발 수행

국내 7개 연구 커뮤니티 공동활용
데이터집약형 다양한 연구도메인으로 확대

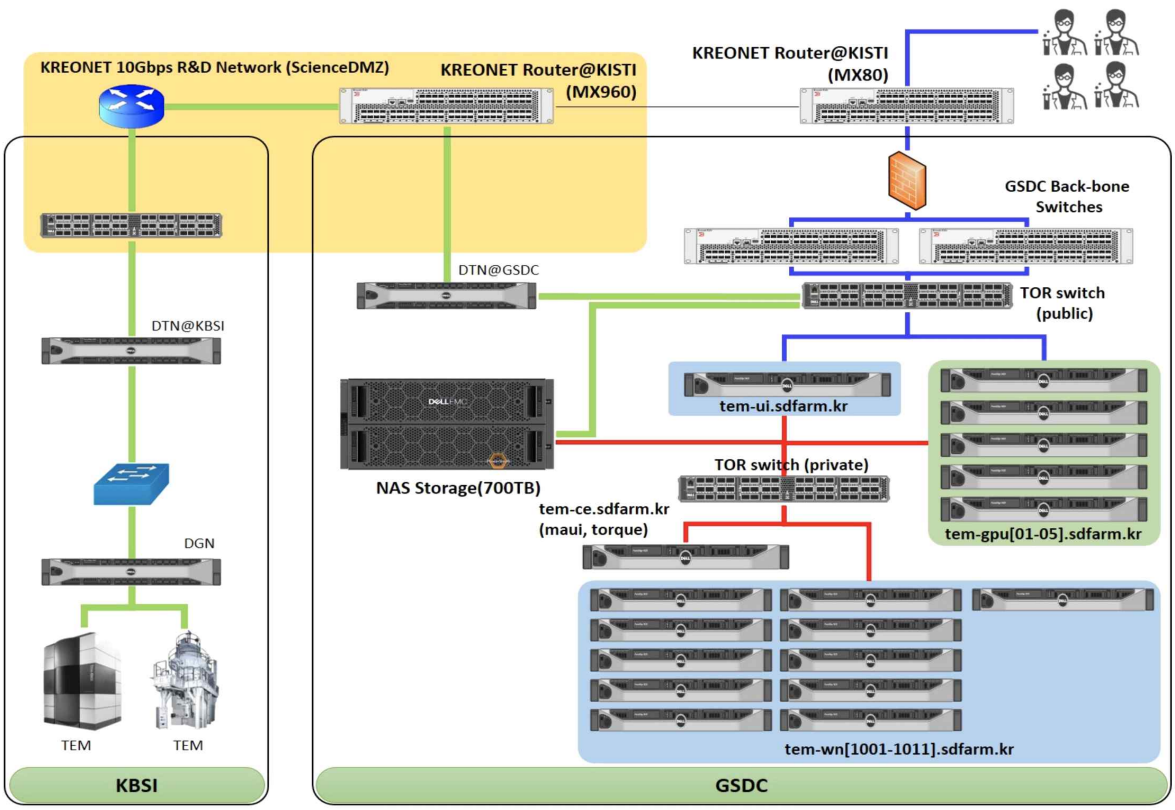


□ 국내 대형연구장비 연계 데이터 집약형 연구지원 확대

- ALICE, CMS, BELLE, LIGO, 유전체, RENO 실험 이외에 국내 대형연구장비 (전자현미경, 포항방사광가속기 등)와 연계 데이터 집약형 연구지원 확대
- 오창 KBSI에 있는 초극저온전자현미경과 연계한 TEM 데이터 분석 팜을 구축·운영 시범 서비스 (2017년~) 및 공식 서비스 (2019년 09월 ~)
- 포항가속기연구소 4세대 방사광가속기와 연계한 데이터 분석 환경을 구축·운영 시범 서비스 (2018년~)

□ GSDC 구조생물학분야 TEM 데이터 분석 팜 현황

- KBSI ↔ GSDC 간 전용의 10Gbps 고속데이터 전송망 제공 (KREONET, KISTI 첨단연구망 ScienceDMZ 활용)
- 408 CPU 코어, 10 GPGPUs, 800 TB 스토리지 제공
- Relion, cisTEM, cryoSPARC 등 다양한 데이터 분석 도구 활용 환경 제공



□ GSDC TEM 분석 팜 컴퓨팅/스토리지 자원 (2020년 2월 기준)

- TEM 분석 팜 구조 및 시스템 명세

Category	Name	Specification	Resources size
Login	tem-ui.sdfarm.kr	• CPU : Intel(R) Xeon(R) CPU E5-2697v3 @ 2.60GHz 14Core * 2 CPUs • RAM : DDR4 8GB * 24 (192GB) • HDD : 12G SAS HDD 1.2TB * 2EA (RAID-1)	28 cores
Computing (master)	tem-ce.sdfarm.kr	• CPU : Intel(R) Xeon(R) CPU E5-2697v3 @ 2.60GHz 14Core * 2 CPUs • RAM : DDR4 8GB * 24 (192GB) • HDD : 12G SAS HDD 1.2TB * 2EA (RAID-1)	28 cores
Computing (workers)	tem-wn[1001-1011].sdfarm.kr	• CPU : Intel(R) Xeon(R) CPU E5-2697v3 @ 2.60GHz 14Core * 2 CPUs • RAM : DDR4 8GB * 24 (192GB) • HDD : 12G SAS HDD 1.2TB * 2EA (RAID-1)	308 cores
	tem-gpu[01-05].sdfarm.kr	• CPU : Intel® Xeon® CPU E5-2690v4 @ 2.60GHz 14Core * 2 CPUs • RAM : DDR4 16GB * 24 (384GB) • SSD : 6G SATA SSD 800GB * 2EA (RAID-1) • GPU : NVIDIA P100 * 2ea (tem-gpu[01-03]) • GPU : NVIDIA P40 * 2ea (tem-gpu[04-05])	140 cores
Storage	Dell EMC Isilon NAS	Network attached storage 700 TB	
Total		504 CPU cores, 10 GPGPUs, 700TB Storage	

※ 2020년 100 CPU 코어, 100TB 스토리지 증설 예정 (총 408 CPU 계산 코어, 10 GPGPU, 800TB 스토리지)

- TEM 분석 팜 관리 소프트웨어 (배치 시스템, OpenMPI, Nvidia CUDA 등)

Category	Name	Description	Version (module path)
OS	Scientific Linux	Operating system	6.x
System M/W	Environment module	• Module environment • https://modules.readthedocs.io/en/latest	v3.2.10
	OpenPBS(torque)	• Cluster resources management • http://www.adaptivecomputing.com/products/torque	v6.1.2
	OpenMPI	• Messaging Pass Interface(MPI) • Reference implementation for MPI standard • https://www.open-mpi.org	v1.8.8 (mpi/gcc/openmpi/1.8.8)
	cuda	• Compute Unified Device Architecture(CUDA) • NVIDIA CUDA Runtime & Toolkit • https://developer.nvidia.com/cuda-toolkit	9.1 (cuda/9.1)
	python	• Python runtime	v2.6.6

- 데이터셋 분석 도구

Category	Name	Description	Version (module path)
Tools	**Relion**	A stand-alone computer program that employs an empirical Bayesian approach to refinement of (multiple) 3D reconstructions or 2D class averages in electron cryo-microscopy (cryo-EM). • https://www3.mrc-lmb.cam.ac.uk/relion/index.php	v3.0.7 (apps/gcc/4.4.7/relion/cpu/3.0.7) (apps/gcc/4.4.7/relion/gpu/3.0.7)
	cisTEM	User-friendly software to process cryo-EM images of macromolecular complexes and obtain high-resolution 3D reconstructions. • https://cistem.org	v1.0.0 (apps/gcc/4.4.7/cistem/1.0.0)
	CryoSPARC	CryoSPARC is the state-of-the-art platform used globally for obtaining 3D structural information from single particle cryo-EM data. • https://cryosparc.com	Not deployed yet (TBD)

※ 2020년 웹 기반 CryoSPARC 시범 서비스 예정. Relion, cisTEM, cryoSPARC 등 지속적인 업데이트

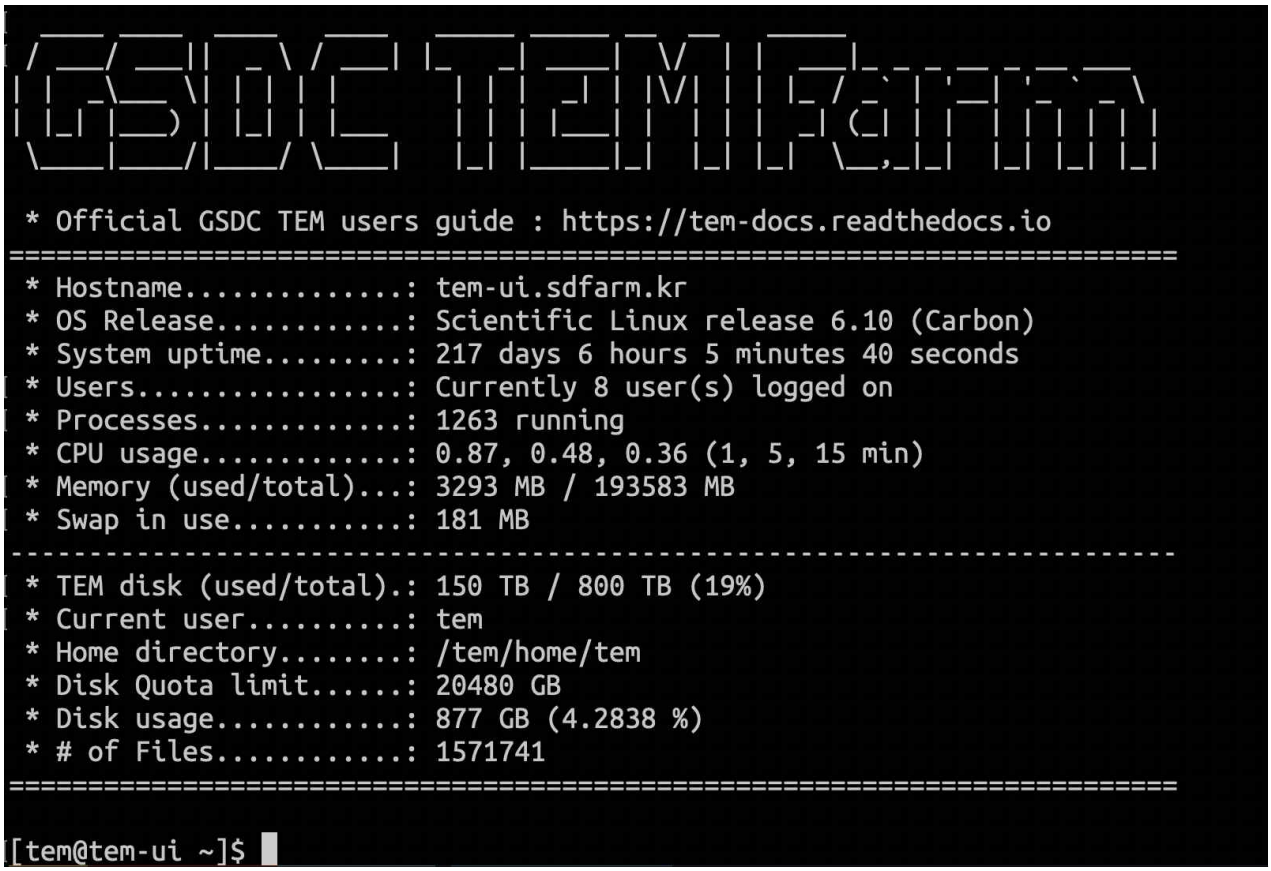
☐ GSDC TEM 분석 팜 사용자 계정 (account) 별 신청 절차

1. 첨부의 TEM 계정신청서 작성 후, 이메일로 계정신청서 제출 (연구자 → GSDC TEM 서비스 담당자)
2. 계정 및 접속 정보 (GSDC TEM 서비스 담당자 → 연구자)

☐ GSDC TEM 분석 팜 사용자 계정 (account) 별 지원 현황

- tem-ui.sdfarm.kr(로그인 서버)을 통한 SSH 터미널 접속 지원
- 데이터 분석 작업 실행을 위한 CPU 계산자원 공용 큐, GPU 계산자원 공용 큐
- 원본 데이터셋 저장 및 분석용 스토리지 20TB
- 데이터셋 분석 도구 실행을 위한 사용자 기술 지원

□ GSDC 구조생물학분야 TEM 서비스 사용자 매뉴얼



https://tem-docs.readthedocs.io

- GSDC TEM 분석 팜 개요 및 소개
- TEM 분석 팜 접속 방법 (리눅스/맥, 윈도우 환경)
- 데이터 분석 도구 활용을 위한 모듈 환경 (module environment)
- 배치시스템 작업 스크립트 작성 방법 및 예시
- 배치시스템 작업 공용 큐 (Queue) 소개 (CPU Queue, GPU Queue)
- Relion : CPU/GPU 계산자원을 활용한 데이터 분석 방법
- cisTEM : CPU 계산자원을 활용한 데이터 분석 방법 등

□ 연락처 및 문의

- 유정록 (junglok.yu@kisti.re.kr, 042-869-0622 KISTI GSDC)
- 여일연 (ilyeon9@kisti.re.kr, 042-869-0658, KISTI GSDC)

□ 추가 안내사항

- KBSI가 아닌 연구그룹 연구실에서의 원활한 실험데이터셋 전송 (GSDC ↔연구실)를 위해서는 첨단연구망 사용 신청 필요
- 매년 초 신청, 평가 진행됨

붙임 1 첨단연구망 (KREONET) 연계 실험실 네트워크 구성 (예)

포항 공대 자체적으로 구축되어야 할 네트워크 구간

1. 연구실험실-연구동 전산실-학교전산실 간 네트워크 연결 코어여부 및 코어 종류(케이블 싱글 or 멀티 모드)확인 (포항공대 전산실)
   1. 코어 존재시 활용 가능 여부 확인
   2. 여유 코어가 없으시 추가 포설 여부 확인
2. 코어 종류에 맞는 실험실 내부 L2 스위치 필요

연구망에 의해 구축될 네트워크 구간

1. 연구망 활용을 위한 신청서 작성 후 선정 과정 필요
   1. 선정 여부에 따라 연구망과 협의 진행
   2. 선정 관련 공고 및 방법은 추후 전달 예정