

국외 출장 결과보고서

결 재	담당자	팀장	실장	사업단장
	6월 30일	6월 30일	6월 30일	7월 1일
	김군태	김동일	김정한	이우진

1. 출장개요

출장목적	ISC 2022 참석		
출장일정	2022. 5. 28. ~ 2022. 6. 4. (6박 8일)		
성명		소속	직급(직위)
김군태		전산관리팀	선임급연구원
방문기관 (또는 회의장소)	독일 함부르크 콩그레스 센터 (Congress Center Hamburg, Germany)		
출발지	인천	도착지	함부르크

2. 업무수행내용

가. 출장 배경 및 목적

- 슈퍼컴퓨팅 관련 이슈와 신기술, 기술발전 등 세계적인 연구동향과 흐름 파악
 - 2022 ISC High Performance Conference는 슈퍼컴퓨팅 관련 세계각지의 많은 전문가들이 참여하여 발표와 토론을 진행하는 학회로써, 슈퍼컴퓨팅과 관련된 이슈와 새로운 기술들, 기술발전과 그 방향 등 최신의 연구동향을 파악하는데 유리함
 - 사업단에서 운영중인 GAON2 클러스터와 다양한 정보시스템들에 ISC 2022에서 확인할 수 있는 새로운 프로세스와 메모리 기술 등을 적용하여, 보다 효율적으로 사업단 전산시스템을 운영할 수 있는 계기 마련
- 사업단 연구개발 클러스터인 GAON2에 적용할 수 있는 병렬화기법 및 시스템 최적화 방법에 대한 지식 습득
 - 새로운 슈퍼컴퓨팅 기술과 연구에 대한 지식을 습득함으로써 사업단에서 운영중인 GAON2 클러스터에 새로운 병렬화기법 등의 기술을 적용하여 클러스터 생산성 향상에 기여
 - 노후화된 하드웨어와 제한된 리소스를 사용하는 GAON2 클러스터의 클러스터링 최적 성능값을 찾아 적용함으로써 수치예보모델 개발업무 최적화에 기여

나. 주요 업무 수행 사항

※ 학술회의에서의 연구개발성과 발표 내용, 최신 연구개발 기술동향 습득 내용, 외부 전문가와의 토의 내용 등 기술, 기관 방문 시 회의내용에 대한 기술

o Tutorial

- Efficient Distributed GPU Programming for Exascale

GPU는 지난 몇 년동안 세계적으로 HPC의 한 축을 담당하는 장치였고, 최근에는 슈퍼컴퓨터 성능의 대부분을 제공하는 핵심 컴퓨팅 장치로 인식되고 있다. 수만 개의 장치가 있는 미래 GPU 가속시스템을 활용하기 위해선 분산 GPU 응용프로그램을 이해하고 관리 및 최적화를 할 수 있는 기술이 필요하다. 이러한 기술들을 습득하기 위해 MPI를 활용한 GPU 프로그래밍 방법을 실습하고, NCCL과 NVSHMEM 같은 보완 프로그래밍 모델에 대한 강의를 진행했다.

- GPU code optimization and auto-tuning made easy with Kernel Tuner: a hands-on, bring your own code tutorial

현재 진행중인 1세대 엑사스케일 HPC 시스템 환경에서 GPU는 주요 플랫폼으로 사용될 예정이며, GPU 애플리케이션 성능은 소프트웨어가 하드웨어에 어떻게 최적화 되었는지에 따라 달라진다. 소프트웨어에서 사용되는 매개변수는 다양하게 구현되고 코딩되므로, 여기에 사용되는 성능 자동튜닝 기술은 매우 중요하다. 해당 튜토리얼에서는 Python 스크립트를 사용하여 GPU 코드를 자동으로 조정하기 위한 툴인 커널튜너(Kernel Tuner)의 사용방법을 실습하고, OpenCL, CUDA, C++, Fortran을 지원하는 커널튜너에 대한 간략한 강의를 진행했다.

- Advanced OpenMP: Performance and 5.2 Feature

OpenMP(Open Multi-Processing)는 공유 메모리 다중 처리 프로그램 API이며 리눅스와 유닉스, 윈도우 등 여러 플랫폼을 지원한다. OpenMP에 대한 간략한 소개를 시작으로 OpenMP를 고성능으로 활용하기 위한 메모리엑세스, 벡터화(Vectorization) 방법과 가속컴퓨팅을 위한 OpenMP 활용법 등을 실습했다.

o Conference

- Accelerating Arm for Zettascale

ZettaScale Computing은 초당 1,000 ExaFlops 또는 6조에 달하는 부동소수점 연산을 수행하는 컴퓨팅 시스템을 말한다. 2022년 현재까진 미국의 Frontier-System이 유일한 ExaScale 시스템이라 할 수 있으나, ARM에서도 ZettaScale Computing에 대한 준비가 진행되고 있다. 일본의 Fugaku, 산디아의 Astra, 영국의 GW4/EP SRC와 유럽의 Mont-Blanc 프로젝트를 포함한 ARM 기반의 HPC 프로젝트가 증명한 바와 같이 ARM은 HPC에 큰 영향을 미친 주요 업체로, 지난 2년 ARM이 향후 대규모 클러스터 구축의 옵션이 될 것을 증명했다. ARM에서도 Zettascale-System을 준비하고 있으며, 이와 관련된 클러스터의 성장에 대해 논의한다.

- Highlights of the 59th TOP500 List including Awarding

일본 고베에 위치한 R-CCS(Riken Computational Science Center)의 Fugaku-System이 약 2

년의 폴-포지션을 유지하다 미국 테네시주 오크리지 국립연구소(ORNL)의 Frontier-System에게 자리를 내어주었다. 프론티어 시스템(Frontier System)은 1.6ExaFlops/s의 최고 성능을 가지고 있으며, 지금까지 1.102Eflops/s의 HPL 벤치마크 점수를 달성했다. 혼합 정밀도 계산을 위한 성능을 측정하는 HPL-AI 벤치마크에서도 6.86 Exaflops를 달성하여 1위를 차지했다. 현존하는 유일한 ExaFlops 시스템이라 할 수 있으며, 2위로 밀려난 일본의 Fugaku의 442PetaFlops/s와 비교하여 약 4배 정도 계산능력이 높은, 압도적인 성능이라 말할 수 있다. TOP500의 상위 5개 시스템 순위는 아래와 같다.

Rank	System	Core	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	Frontier – HPE Cray EX235a, AMD Optimized 3 rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	Fugaku – Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	LUMI – HPE Cray EX235a, AMD Optimized 3 rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	1,110,144	151.90	214.35	2,942
4	Summit – IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096

5	Sierra – IBM Power System AC922, IBM POWER9 22C 3.1GHz, NV IDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM/NVIDIA/Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438
---	---	-----------	-------	--------	-------

– The Message Passing Interface (MPI): On the Path to MPI 5.0

MPI-4에서는 다음과 같은 주요내용이 추가되었다. 1) Partitioned Communication, 2) New tool interface for events, 3) Solution for “Big Count” operations, 4) Persistent Collectives, 5) New init options via MPI Sessions, 6) Topology Solutions, 7) and much more, ... 본 세션에서는 최근 출시된 MPI-4의 버전과 그 기능 등에 대한 논의를 하고, MPI-4의 다음 단계인 MPI-5로 초점을 바꾸어 새로운 아이디어를 논의한다.

– InfiniBand In-Network Computing Technology and Roadmap

과거 스마트 인터커넥트 개발의 초점은 CPU에서 네트워크로 네트워크 패킷을 전달하는 것이었으나, 차세대 스마트 인터커넥트는 네트워크 시스템 내에서 전송되는 데이터를 상호 연결함으로써 컴퓨팅 기능에도 참여하는 기술로 발전하고 있으며, 이를 In-Network Computing이라 한다. 이러한 기술들은 Infiniband HDR(200Gb/) In-Network Computing을 기반으로 TOP500 슈퍼컴퓨터 플랫폼을 포함한 최근 전 세계의 대형 슈퍼컴퓨터 중 일부에서 사용되고 있다. 현재는 Infiniband NDR(400Gb/s) In-Network Computing 기술이 시장에 출시되고 있으며, MPI의 다대다통신(ALL-To-ALL) 엔진 등과 같은 새로운 컴퓨팅 기능과 응용프로그램 별 알고리즘을 프로그래밍하는 엔진도 추가되어 있다.

– LUSTRE in HPC, AI and Big Data: Widening Scope, New Features and Roadmap

Lustre 파일시스템은 HPC를 위한 오픈소스 파일시스템으로, 지난 2년 간 세계 슈퍼컴퓨터 순위 TOP500에서 1위를 유지한 일본의 Fugaku를 포함하여 전 세계 TOP100 슈퍼컴퓨터의 약 2/3에서 사용되고 있다. 현재 TOP500 중 1위로 선정된 미국의 Frontier도 Lustre 파일시스템으로 구성되어 있다. Lustre는 역사적으로 미국 국립연구소(U.S. National Labs)에서 주도적으로 개발해왔으며, 이제는 AI/ML 인프라 및 클라우드 기반 아키텍처를 포함하도록 범위를 확장하고 있다. 최근 Lustre에서 기능개발에 초점을 두고 있는 부분은 1) 네트워크 선택 정책, 2) OST 풀 할당량, 3) DNE(Distribute NameSpace) 자동 균형 조정, 4) 사용자 데이터 및 디렉토리 암호화, 5) FLR Erasure-Coding 기술(Data Availability Proof에 사용되는 수학적 기법 중 하나) 등이 있다.

- The Green 500: Trends in Energy-Efficient SuperComputing

슈퍼컴퓨팅 시스템은 연산처리 능력에 초점을 맞추어 성능을 높이는 방향으로 성장해왔으나, 2000년대 중반부터 환경에 대한 중요성이 인식되면서 슈퍼컴퓨팅 시스템의 전력 효율에 대한 부분도 중요하게 다뤄지고 있다. 이로 인하여 2010년부터는 슈퍼컴퓨팅 시스템의 전력소비효율을 평가하는 Green500 프로젝트도 진행되고 있으며, 현재는 슈퍼컴퓨팅 시스템을 평가하는 주요 지표로 활용되고 있다. 2022년도 Green500에서 1위를 차지한 시스템은 미국의 Frontier-System으로, 2위로 밀려난 일본의 Fugaku-System과 비교하여 약 32% 향상된 에너지 효율을 가지고 있으며, 그 효율은 와트(Watt) 당 52.23GigaFlops의 계산성능에 달한다. 상단의 TOP500 상위 5위 시스템 리스트를 참조하면 2위인 Fugaku 시스템보다 약 9,000kW 적은 전력을 소모한다.

- The Road to the Hybrid Quantum

양자컴퓨팅 기술은 약물 발견에서 일기예보까지 잠재력을 발전시키면서 HPC의 미래에 큰 역할을 할 것으로 기대되고 있다. 이미 수십개의 양자조직이 NVIDIA cuQuantum 소프트웨어 개발 키트를 사용하여 GPU 양자회로 시뮬레이션을 가속화하고 있으며, 최근 AWS에서는 Braket에서 cuQuantum을 사용할 수 있다고 발표했고, Braket에서 cuQuantum이 양자 머신러닝 워크로드에서 최대 900배 속도를 제공하는 방법을 시연했다. 양자컴퓨팅 기술이 진화함에 따라 고전적인 시스템과 양자 시스템을 하이브리드 양자 컴퓨터로의 결합이 요구되고, 이는 크게 두 가지 요소로 구분된다. 하나는 GPU와 QPU(Quantum Processor Unit) 간 빠른 연결이며, 다른 하나는 효율적이고 사용하기 편리한 통합 프로그래밍 모델이다. 이 두 가지 요소는 하이브리드 시스템의 회로 최적화, 보정 및 오류 수정과 같은 작업에 활용되나, 통합 프로그래밍 모델과 컴파일러 도구 체인이 부족하여 개선과정이 필요하다. 또한 양자 컴퓨터가 작업을 효율적으로 가속화할 수 있는 방법을 찾기 위해 QPU의 이식과 이를 위한 높은 수준의 컴파일러 성능 등이 요구되고 있다.

- Applications of HPC in Industry

HPC는 수십 년 동안 학술 연구 및 산업 혁신에서 중요한 부분을 차지해오고 있다. HPC는 엔지니어, 데이터 사이언티스트, 설계자 및 기타 연구자들이 기존 컴퓨팅보다 훨씬 적은 시간과 비용으로 크고 복잡한 문제를 해결할 수 있도록 도와주고 있다. 거의 모든 산업 분야의 Fortune 1000대 기업에서 HPC를 사용하고 있으며, Hyperion Research에 따르면 2022년 기준 전 세계 HPC 시장의 규모가 약 440억 달러에 달하는 것으로 보여진다. 다양한 산업분야에서 HPC는 아래와 같은 역할을 수행하고 있다.

1) 항공우주 : 비행기 날개 위의 기류와 같은 복잡한 시뮬레이션 생성

- 2) 제조 : 신제품의 설계, 제조 및 테스트를 지원하여 더욱 안전하고 효율적인 프로세스 및 혁신을 실현
- 3) 금융(핀테크) : 복잡한 위험분석, 빈번한 거래, 금융모델링 및 사기탐지 등 수행
- 4) 유전체학 : 조상 연구를 지원하기 위한 DNA 시퀀싱, 약물 상호 작용 분석, 단백질 분석 실험 등 수행
- 5) 석유 및 가스 : 공간 분석을 수행하고 저수지 모델을 테스트하여 석유 및 가스 자원의 위치를 예측하고, 유체 흐름 및 지진 처리와 같은 시뮬레이션을 수행

o Workshop

- Communication, I/O, and Storage at Scale on Next-Generation Platforms

차세대 HPC 플랫폼은 하위 시스템에서 증가하는 이질성을 처리할 수 있어야 한다. 이러한 서버 시스템에는 노드 간 통신을 위한 내부 고속 패브릭, 소프트웨어 정의 네트워크를 지원하기 위한 데이터 처리 장치(DPU) 및 인프라 처리 장치(IPU)와 스토리지 시스템 인프라 등이 포함된다. 워크숍에서는 커뮤니케이션, I/O 및 규모에 맞는 스토리지에 중점을 둔 차세대 플랫폼의 전체적인 개요를 설명하고, 다양한 HPC 패브릭을 사용한 애플리케이션 중심의 성능 분석을 소개한다. 또한 관련된 모범사례를 공유하여 규모에 맞는 애플리케이션 성능 및 당면 과제를 개선하기 위한 포럼을 진행했다.

다. 기타

※ 본인의 담당업무 이외 사업단 홍보 및 대외 소통을 위한 노력, 사업단 채용을 위한 인재 발굴에 관한 사항 등이 있을 경우에 한하여 작성

o 함부르크 독일기후전산센터 방문 및 소통 (DKRZ, Deutsches Klimarechenzentrum)

- 방문일시 : 2022년 06월01일 수요일 15시 (GMT+2)
- 독일기후전산센터는 1987년 기후변화와 지구시스템 연구를 전문적으로 지원하기 위해 막스플랑크연구협회(MPG)와 함부르크대학(시정부), 알프레드-베게너극지해양연구소(AWI), 게슈타흐트연구센터(GKSS)가 협력하여 함부르크에 설립한 슈퍼컴퓨터센터이다.
- 현재 운영중인 슈퍼컴퓨터인 Levante는 AMD 7763 프로세서 2개를 탑재한 Computing-Node 2,832개와 AMD 7713 프로세서 2개와 NVIDIA A100 GPU 4개를 탑재한 GPU-Node 60개로 구성되어 있으며, 네트워크 시스템은 NVIDIA Mellanox Infiniband HDR 100G, 스토리지는 Lustre로 설계한 DDN의 130PB로 구성되어 있다.
- Levante는 수냉식 냉각장치로 구성되어 있다. 30도의 냉각수로 시스템을 냉각하면 약 40도 정도로 배출되는데, 배출된 냉각수는 DKRZ 건물 옥상으로 보내 외부공기와 접촉시켜 다시 냉

각하는 방식을 사용하고 있다. 함부르크는 따뜻한 계절에도 평균고온이 약 22도이고 평균저온이 약 13도로 측정되는 도시이기 때문에 가능한 친환경적인 열전달 방식이라 볼 수 있다.

3. 출장성과 · 시사점 및 향후 업무 활용 계획

가. 출장성과 · 시사점

○ HPC 기술 동향 파악 및 발전 방향

- 세계 최초의 엑사스케일 컴퓨팅 시스템인 미국의 Frontier가 등장함에 따라 앞으로의 HPC 시장에서는 엑사스케일의 슈퍼컴퓨터가 대거 등장할 것으로 예상된다. 이에 따라 엑사스케일 환경에서의 EDR/HDR/NDR 등 Infiniband Networking 시스템과 Lustre 등 스토리지 시스템 그리고 관련된 하드웨어와 소프트웨어 기술 등과 관련된 긴밀한 협업이 요구된다. 앞으로의 HPC 발전방향을 참고하여 사업단에서도 현재 연구하고 있는 모델들을 보다 고차원적인 시스템 성능에 맞추어 최적화함으로써 엑사스케일 슈퍼컴퓨터 등 최신의 HPC 시스템에서도 적합하게 활용할 수 있도록 개선하는 작업이 필요하다고 할 수 있다.

○ 친환경 시스템 개발 및 운영

- 환경과 관련된 이슈가 꾸준히 제기됨에 따라 세계의 HPC 시장에서도 이와 관련된 이슈를 끊임없이 제기하고 논의하고 있다. 주요 화두는 전력을 적게 사용하면서도 고성능의 연산능력을 갖추는 HPC의 개발로, 최근 TOP500에서 1위를 차지한 미국의 Frontier 시스템이 대표적이다. 미국의 Frontier HPC는 2위인 일본의 Fugaku HPC보다 약 4배의 연산능력을 가지지만 약 30% 정도의 전력을 소모하여, 친환경 슈퍼컴퓨터 지표인 Green500에서도 1위를 차지했다.
- 독일 함부르크의 독일기후전산센터(DKRZ)는 수냉식 냉각시스템을 도입하여 친환경적으로 시스템을 운영하고 있다. 함부르크 지역 여름철 평균고온이 22도인 점을 활용하여 HPC 시스템에 사용한 냉각수의 온도를 낮추고, 다시 순환하여 사용하는 방식을 취하고 있다. 이처럼 주변의 자연환경을 활용하여 시스템에 접목시키는 등 친환경적으로 시스템을 운영하기 위한 노력이 필요하다.

○ In-Networking Computing 및 MPI

- EDR(100GbE) Infiniband 네트워킹 시스템이 보편화되고, HDR(200GbE), NDR(400GbE)의 시스템이 출시되고 계획됨에 따라, 네트워킹 시스템은 단순 데이터패킷 전송이 아닌 시스템 연산 성능의 한 축을 담당하는 부분으로 자리잡고 있다.
- 이러한 네트워크 시스템을 기반으로 MPI 등 네트워크 통신을 기반으로 하는 분산 및 병렬처리 시스템의 주요 엔진들의 패치가 진행되고 있으며, 최근에는 MPI-4 작업이 진행중이며 MPI-5에 논의도 진행되고 있다.

나. 향후 업무 활용 계획

○ 친환경적인 사업단 정보시스템 운영 논의

- 사업단 IDC에서 운영중인 정보시스템과 GAON2 연구개발 클러스터 등에 적용 가능한 저전력 고성능의 친환경적인 운영방식을 논의할 필요가 있다. 현재 운영되는 시스템의 불필요한 전력 소모를 줄이는 방법을 논의하여 적용하고, 추후 도입할 장비들도 불필요한 하드웨어는 줄이면서 성능을 극대화 할 수 있는 방법을 협의하여 적용할 계획이다.

○ GAON2 연구개발 클러스터 최신 라이브러리 설치 및 최적화

- 세계적인 추세에 따라 사업단에서 연구개발 중인 데이터, 모델 등이 엑사스케일의 HPC와 같은 최신의 시스템에서도 동작할 수 있도록 사업단의 GAON2 연구개발 클러스터를 업데이트할 필요가 있다. 물론 하드웨어 시스템 상으로 불가능할 수 있지만, 적용 가능한 최신의 MPI와 최신의 라이브러리 등을 설치하여 운영함으로써, 사업단에서 생산하는 데이터와 모델 등이 다른 최신의 HPC에서도 동작할 수 있도록 설계를 추진할 계획이다.