

국외 출장 결과보고서

결	담당자	팀장	실장	사업단장
	6월 17일	6월 20일	6월 20일	6월 21일
재	김동휘	김동일	김정한	권인혁

1. 출장개요

출장목적	ISC 2024 참석		
출장일정	2024. 5. 11.~2024. 5. 18. (6박 8일)		
성명		소속	직급(직위)
김동휘		전산관리팀	원급연구원
방문기관 (또는 회의장소)	독일 함부르크 콩그레스 센터(Congress Center Hamburg, Germany)		
출발지	인천(Incheon)	도착지	함부르크(Hamburg)

2. 업무수행내용

가. 출장 배경 및 목적

o 배경

- ISC 2024 High Performance Conference는 HPC, AI, 양자 컴퓨팅 등 신기술들 간의 상호 작용 방식과 관련 기술들을 소개하는 학회로써 슈퍼컴퓨팅 관련 넓은 지식을 갱출 수 있음
- 또한, 반도체 제조 비용의 상승으로 인해 HPC의 중요성이 부각되면서 계속해서 증가하는 성능 요구 사항을 충족시키기 위한 강연으로 갖추어져 있음
- 학회 세션 주제로는 HPC와 AI를 위한 스마트 네트워킹 기술, 엑사스케일 컴퓨팅, 시스템과 어플리케이션 시뮬레이션 및 모델링, HPC, AI, Cloud에서의 러스터, 퀀텀 컴퓨팅, 과학에서의 생성형 AI, HPC 워크로드 분석을 위한 AI 및 ML, 유럽의 HPC와 AI 등으로 구성되어 있음

o 목적

- 사업단에서 사용중인 연구개발 클러스터 장비들의 성능을 올리기 위한 세션들에 참석하여 HPC, AI 및 여러 분산시스템, 프로세서 기술, 메모리 시스템에 대한 지식을 습득
- 슈퍼컴퓨팅, HPC, AI 기술들이 유럽에서 실제 어떻게 활용하고 있는지 파악하여 사업단 수치예보모델개발 업무에 활용할 수 있도록 함
- GPU 가속화, 엑사스케일을 활용한 GPU 프로그래밍, 데이터센터 냉각 솔루션, AI 및 HPC 간 최적화 프로세싱 기술 등을 활용하여 사업단의 안정적인 전산 인프라 환경에 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 튜토리얼

- High-Performance and Smart Networking Technologies for HPC and AI

고성능 네트워킹 기술은 GPU, 가속기, 데이터센터 처리 장치(DPU) 및 다양한 애플리케이션 워크로드를 갖춘 HPC 및 AI용 차세대 하이엔드 컴퓨팅(HEC) 시스템을 구축하는 데 많은 관심을 불러일으키고 있습니다. 이 튜토리얼에서는 이러한 최신 기술, 아키텍처 기능, 현재 시장 지위 및 HEC 시스템 설계에 대한 적합성에 대한 개요를 제공합니다. IB, HSE, RoCE 및 Omni-Path 상호 연결에 대한 간략한 개요부터 시작합니다. 이러한 상호 연결의 아키텍처 기능에 대한 심층적인 개요는 관련 실습을 통해 제공됩니다. 이어서 새롭게 떠오르는 NVLink, NVLink2, NVSwitch, EFA 및 Slingshot 아키텍처에 대한 개요가 이어집니다. 그런 다음 성능과 확장성을 지원하는 상용 고성능 네트워크의 고급 기능을 소개합니다. 그런 다음 DPU/IPU(스마트 NIC)와 같은 향상된 오프로드 가능 네트워크 어댑터와 해당 기능에 대한 개요를 제공합니다. 다음으로 Open Fabrics Verbs, LibFabrics, UCX와 같은 고성능 네트워크용 소프트웨어 스택의 개요를 살펴보고 이러한 스택의 성능을 비교합니다. 다음으로, 이러한 상호 연결, 솔루션 및 샘플 성능 수치를 위한 MPI 라이브러리를 설계할 때 기술들이 소개됩니다.

- Performance Engineering for Linear Solvers

CPU 및 GPU 노드의 선형 솔루션에 대한 코드 분석, 성능 모델링 및 최적화를 다룹니다. 성능 엔지니어링은 성능 모델에 대한 우수한 예로 간단한 루프를 사용하고 최적화를 안내하는 방법을 가르치는 경우가 많습니다. 그러나 사전 조정된 전체 선형 솔루션은 수렴할 때까지 실행되는 반복 체계에 포함된 여러 연속 루프로 구성됩니다. 결과적으로 "최적 성능"이라는 개념은 하드웨어 리소스 효율성과 반복 솔루션 수렴을 모두 고려해야 합니다. 우리는 선형 반복 솔루션에 맞춰진 성능 엔지니어링 프로세스를 전달합니다. 조밀하고 희박한 데이터 구조를 위한 하드웨어 구성 및 스토리지의 기본 개념을 소개한 후 Roofline 성능 모델이 예측 및 진단 방식으로 이러한 솔루션에 어떻게 적용될 수 있는지, 그리고 솔루션의 하드웨어 효율성을 평가하는 데 어떻게 사용할 수 있는지 보여줍니다. 순수 메모리 경계와 같은 중요한 코너 케이스. 그런 다음 CG(Conjugate Gradient Method) 알고리즘을 주요 예로 사용하여 사전 조건이 지정된 솔버의 구조로 진행합니다. 전체 솔루션의 핫스팟과 병목 현상이 식별되고 선조건자 및 캐시 차단과 같은 고급 성능 최적화 기술이 도입됩니다.

○ 세션 참석

- Conference Keynote: Beyond Exascale Computing

1세대 엑사스케일 컴퓨팅 시스템은 강력하고 새로운 애플리케이션 기능과 시스템 소프트웨어를 갖추며 온라인 상태입니다. 동시에 더욱 강력한 시뮬레이션, 기계 학습 방법의 채택, 새로운 기기 및 점점 더 유비쿼터스되는 데이터 수집 장치에서 발생하는 대규모 데이터 분석 문제를 위해 고성능 컴퓨팅에 대한 요구가 계속해서 증가하고 있습니다. 가장 넓은 의미에서 계산과학 연구는 물리과학과 생명과학을 넘어 사회과학, 공공정책, 심지어 인문학까지 확대되고 있습니다. 칩 기술이 스케일링 한계에 직면하고 약한 스케일링의 이점이 줄어들면서 이러한 새로운 요구를 충족하는 것이 점점 더 어려워질 것입니다. 공급망 제한, 시스템 통합업체 수 감소, 클라우드 제공업체의 영향력 증가 등을 포함한 컴퓨팅 시장의 혼란으로 인해 미래의 슈퍼컴퓨터를 확보하고 배포하는 방법에 대한 기본 가정이 바뀌고 있습니다. 동시에 클라우드 컴퓨팅, 특수 하드웨어, 과학 기술이자 컴퓨팅 하드웨어 및 소프트웨어의 동인으로서의 AI의 역할에 대한 논의가 진행되고 있습니다. 이 강연에서 엑사스케일 컴퓨팅의 미래에 관한 미국 국립 아카데미 합의 보고서의 결과 중 일부를 제시할 것입니다. 또한 연구가 직면한 몇 가지 과제와 기회에 대한 여러 관점에 대해 이야기합니다.

- High Performance Software Foundation BOF

과학 컴퓨팅과 디지털 엔지니어링에서 HPC의 사용이 보편화되고 AI 사용 사례가 증가함에 따라 점점 더 많은 데이터 센터에서 GPU 및 기타 가속기를 배포합니다. 오픈 소스 소프트웨어 프로젝트는 이러한 혁신을 가능하게 하는 데 매우 중요했지만 개발 및 사용자 기반은 지속되어야 합니다. 2023년 11월, Linux 재단은 HPC 소프트웨어 프로젝트를 위한 중립적 기반을 제공하고 지속적인 개발을 관리하는 데 도움을 주기 위해 HPSF(고성능 소프트웨어 재단)를 설립하겠다는 의도를 발표했습니다. HPSF는 기술 프로젝트, 이벤트, 교육, 홍보를 통해 HPC 소프트웨어 생태계를 유지할 것이며, 실무 그룹이 HPC 프로젝트가 생태계 전체에 도움이 되는 공통 목표를 위해 협력할 수 있도록 지원할 것입니다. BOF는 ISC 2024에서 HPSF의 공식 출시를 시작하며 ISC 참석자들에게 HPSF 창립자, 초기 회원 조직 및 소프트웨어 프로젝트와 상호 작용할 수 있는 기회를 제공하는 것을 목표로 합니다. 해당 강연은 HPSF의 사명, 거버넌스 구조 및 향후 계획에 대해 소개하였습니다.

- Using HPC in the cloud to speed up your time to solution?

AWS에서 HPC 워크로드를 실행하여 결과 도출 시간을 단축할 수 있는 방법에 대해 설명을 합니다. R&D 팀이 필요할 때마다 언제든지 손쉽게 사용할 수 있는 고급 기술을 최대한 활용하는 데 필요한 친숙한 스케줄러, 라이브러리 및 개발자 도구가 모두 포함된 작업 준비 HPC 환경을 시작하는 방법을 배우게 됩니다. AWS를 사용하면 과학자와 엔지니어는 친숙한 인터페이스와 온디맨드 리소스를 사용하고 사실상 모든 규모에서 시뮬레이션을 실행할 수 있습니다. 결과 도출 시간을

단축하는 동시에 낭비를 방지하기 위해 비용을 엄격하게 관리할 수 의견을 공유할 것입니다.

- Lustre in HPC, AI and the Cloud

HPC용 오픈 소스 및 개방형 병렬 파일 시스템인 Lustre는 역사적으로 주로 미국 국립 연구소에서 주도했지만 기존의 거점을 넘어 고급 AI/ML 및 익스트림 클라우드 인프라를 포함하도록 범위를 확대했습니다. 전 세계의 기여자들과 함께 커뮤니티에서 개발한 기술입니다. Lustre 클라이언트는 x86, Power 및 ARM과 같이 광범위하게 배포된 명령어 세트 아키텍처에 사용할 수 있습니다. ISC 2024 BoF에서는 개발자, 관리자, 사용자 및 솔루션 제공업체가 모여 새로운 기능, 혁신적인 개발 및 배포는 물론 Lustre를 지속적으로 발전시키는 새로운 과제와 커뮤니티의 현재 상태에 대해 논의할 것입니다. 향후 전 세계 Lustre 배포를 어떻게 형성할 것인지 논의할 것입니다. 이제 완전한 IPv6 및 RHEL9 서버 지원을 제공하는 새로운 Lustre 2.16 릴리스의 이점, 경험 및 문제를 다룰 것이며 향후 릴리스 2.17 및 2.18의 기능에 대해서도 논의할 것입니다. 세션은 몇 가지 참조 사이트 업데이트와 알려진 사용자 요구 사항 및 문제에 특별히 초점을 맞춘 현재 릴리스부터 2026년 상반기까지 Lustre 로드맵에 대한 공유를 합니다.

- Quantum Computing

퀀텀은 2세대 QCCD(양자 전하 결합 소자) 양자 프로세서를 구축했습니다. 이러한 프로세서는 큐비트와 교감 냉각 및 셔틀 작업을 위해 트랩된 이온을 사용하여 개별 큐비트와 큐비트 쌍 사이에서 충실도가 높은 게이팅 작업을 달성하여 완전히 연결되도록 합니다. 이번 강연에서는 이들 기계의 성능을 엄격하게 벤치마킹하기 위한 우리의 노력을 검토하고 이들 기계의 강점과 약점을 강조합니다. 또한 단기적인 양자 이점과 양자 오류 수정을 위한 우리의 노력에 대해 간략하게 설명합니다. 마지막으로 이러한 프로세서 확장을 목표로 하는 몇 가지 기술 개발 노력과 향후 장치에 미치는 영향에 대해 간략히 설명합니다.

- Generative AI for Science

생성 AI 및 기반 모델은 텍스트 및 이미지 처리(예: GPT4, DALL-E3, SORA)에서 눈부신 성공을 거두었습니다. 우리는 현대 AI가 과학에서 비슷한 역할을 할 수 있고 물리학, 화학, 생물학의 오랜 문제를 해결할 수 있다고 믿습니다. 텍스트와 이미지에 비해 과학 분야에서는 고품질 데이터가 상대적으로 적으므로 모델 설계, 귀납적 편향 및 내장된 물리학이 더 큰 역할을 합니다. 이 강연에서는 분자 시뮬레이션의 오랜 문제를 소개하고 이 분야에서 생성 AI와 기반 모델을 사용하는 최근 접근 방식에 대해 논의합니다.

- Artificial Intelligence and Machine Learning for HPC Workload Analysis

HPC 시스템은 이미 낮은 수준의 하드웨어 원격 측정 및 오류 보고 시스템에서 생성된 데이터부터 하드웨어 성능 카운터, 작업 예약 및 시스템 로그에 이르기까지 매일 테라바이트 규모의

모니터링, 사용량 및 성능 데이터를 생성하고 있으며 관리자 문제 해결의 자연어 텍스트도 포함하고 있습니다. 티켓과 메모. 미래의 시스템은 훨씬 더 크고 복잡해질 것입니다. 이는 이러한 시스템에서 사용자 행동을 모니터링하고 특성화하는 과제를 증가시킵니다. 한편, 기계 학습 및 인공지능 기술은 이미 크고 복잡한 데이터 세트에서 지식을 특성화하고 추출하는 효과를 입증하기 시작했지만 이러한 노력은 이제 막 다양한 영역에서 잠재력의 전체 가치를 실현하기 시작했습니다. 이러한 이유로 우리는 HPC 워크로드 분석을 위한 인공지능 및 기계 학습에 관한 첫 번째 BoF를 제안합니다. 이 BoF는 최첨단 연구 아이디어를 논의할 뿐만 아니라 데이터 과학, 기계 학습, 통계, 응용 수학, 시스템 설계, 시스템 모니터링, 시스템 탄력성 및 기타 분야에 걸쳐 작업하는 연구자들 간 의견을 공유할 수 있는 세션입니다.

- Sustainability, Carbon-neutrality and Procurement for HPC

우리는 HPC 시스템의 운영 효율성을 크게 개선했습니다. 이제 우리는 시스템의 수명주기 전반에 걸쳐 더 넓은 범위의 환경 영향을 고려해야 합니다. 여기에는 제품의 설계 및 제조 방법, 운송 방법, 작동 방법, 더 이상 유용하지 않은 제품을 분해하고 재사용 및 재활용하는 방법이 포함됩니다. 이러한 고려 사항은 명확하지 않을 수 있습니다. 예를 들어, 제조 비용은 시스템의 수명 주기 탄소 발자국을 지배하며 이러한 추세는 증가하고 있습니다. 시스템의 전체 수명 주기에 걸쳐 탄소 배출량을 어떻게 고려하기 시작할 수 있습니까? 우리는 시스템의 성능, 성능, 에너지를 이해할 수 있는 많은 역량을 갖추고 있지만 탄소 배출량에 대해서는 그렇지 않다고 말할 수 없습니다. 탄소 배출량을 1차 최적화 목표로 삼아야 할까요? 이 BoF는 HPC 조달의 지속 가능성 요구 사항을 해결하기 위한 에너지 효율적인 HPC 실무 그룹의 제안 초안을 검토 및 논의하기 위해 제시할 것입니다. 기본 요구사항은 제품 온실가스(GHG) 배출량 추정치입니다. 보다 발전된 요구 사항은 제품(GHG) 배출 보고를 위한 표준화된 방법론을 사용하는 것입니다. 혁신적인 요구 사항은 물 소비, 환경 독성 및 오염을 포함한 GHG 배출 요구 사항을 확장합니다. 이외에 온실가스 감축을 위해 많은 의견을 공유하고자 합니다.

- Powering European AI with EuroHPC Supercomputers

해당 세션의 목표는 EuroHPC JU 슈퍼컴퓨터가 AI 애플리케이션에 어떻게 사용될 수 있는지, 그리고 이러한 슈퍼컴퓨터에 대한 현재 액세스 가능성은 무엇인지 제시하는 것입니다. EuroCC 국가 역량 센터를 통해 제공되는 지원 가능성도 소개됩니다. 워크숍은 또한 AI 커뮤니티의 현재와 미래의 HPC 요구 사항을 더 잘 이해하기 위해 청중과 상호 작용하고 토론할 수 있는 기회가 되어야 합니다. 생성적 AI와 기반 모델은 전례 없는 속도로 발전하고 있으며 기술과 사회의 미래를 형성하는 데 중추적인 역할을 하게 될 것입니다. AI는 이제 유럽 디지털 전략의 핵심 정책 영역이 되었습니다. HPC와 결합된 AI는 보다 혁신적이고 효율적이며 지속 가능하고 경쟁력 있는 경제에

기여하는 동시에 시민을 위한 안전, 교육 및 의료를 개선할 수 있습니다. 이번 강연은 EuroHPC JU의 임무를 확장하는 맥락에서 개최될 예정이며, 추가로 유럽의 AI 스타트업 및 연구원을 지원하기 위한 AI 전용 슈퍼컴퓨팅 인프라 구축도 포함됩니다.

- The Future of Benchmarks in Supercomputing

슈퍼컴퓨팅이 엑사스케일 시대의 시뮬레이션, 데이터 과학 및 인공 지능의 새로운 워크플로우를 환영함에 따라, 이 워크숍의 목표는 "슈퍼컴퓨팅 커뮤니티가 성능 벤치마크를 어떻게 발전시켜야 합니까?"라는 질문을 제기하고 참여하고 토론하고 해결하는 것입니다. 워크숍은 Top500, HPCG, MLPerf, TeraSort 등의 활동적인 구성원과 업계, 학계, 정부의 주요 인사를 초대하여 진화에 대한 가치, 필요성 및 바람을 논의하는 청중 참여가 포함된 프레젠테이션 및 패널 토론으로 구성됩니다. 미래의 슈퍼컴퓨팅 시스템 설계 및 아키텍처를 안내하기 위해 새로운 애플리케이션을 포괄적이고 수용하는 벤치마크 제품군입니다. 이 워크숍은 (i) 벤치마크 설계(대표성, 비용 고려 등), (ii) 측정항목 선택(에너지 효율성, I/O 포함 등), (iii) 발전을 위한 민첩성(문제 크기, 최신 커널, 미니 앱, 컨테이너화된 하니스 등), (iv) 목적 명시(순위 목록, 공동 설계, 마케팅)와 관련된 주제에 대한 커뮤니티 참여를 촉진하기 위해 프레젠테이션을 구성하였습니다.

3. 출장성과·시사점 및 향후 업무 활용 계획

가. 출장성과·시사점

○ 안정적인 전산 인프라 환경 구축을 위한 방안

- CPU 및 GPU, 서버, 시스템 등 운영 효율성을 개선할 수 있는 많은 방법론에 대해 알 수 있었습니다. HPC 고성능 컴퓨팅의 발전에 따라 반도체, 하드웨어 또한 많은 발전을 이루었다는 것을 알 수 있었습니다. 실제로 많은 발열과 전기를 사용하는 슈퍼컴퓨팅 환경에서 이를 해결할 수 있는 수냉식 냉각시스템, HPC 워크로드 시간 단축 방안 등 다양한 이론들을 알 수 있었습니다. 이를 사업단에 적용할 수 있다면 안정적인 인프라 환경을 구축할 수 있을 것으로 보입니다.

○ 엑사스케일 컴퓨팅 시대의 도래

- 세계 최초의 엑사스케일 컴퓨팅 시스템인 미국의 Frontier의 등장, 내년 독일 EU 최초의 엑사스케일 슈퍼컴퓨터가 출시됨에 따라 앞으로의 HPC 시장에서는 엑사스케일의 슈퍼컴퓨터가 대거 등장할 것으로 예상된다. 이에 따라 차후에 사업단에 엑사스케일 컴퓨팅의 도입에 대비하여 EDR/HDR/NDR 등 Infiniband Networking 시스템과 Lustre 등 스토리지 시스템 그리고 관련된 하드웨어와 소프트웨어 기술 등에 대해 알아야할 필요가 있다. 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.

나. 향후 업무 활용 계획

○ 친환경을 위한 그린 컴퓨팅 환경 조성

- 사업단에서 운영중인 GAON2를 포함한 여러 연구개발 클러스터, 정보시스템 등으로 인해 발생하는 에너지를 줄일 수 있도록 해야합니다. CPU, GPU, 노드, 데이터센터 규모에 이르기까지 모든 영역에서 에너지 효율성을 분석하고 개선할 필요가 있습니다. 현재도 지구 온난화는 끊임없이 진행 중이며 연간 약 200테라와트시를 사용하는 데이터센터는 전체 에너지 사용량의 작은 부분 (약 1%)을 차지하지만 주의가 필요합니다. 따라서 차세대 시스템 도입을 함에 있어 친환경 제품을 구매해야 하며 가능한 데이터센터는 차후에 수냉 시스템을 사용하여 친환경적으로 변화해야 합니다. ISC 2024 Conference에서도 AI, 스토리지 등 기술들은 눈에 띄게 발전하였지만 이에 맞는 환경 대비책은 부족하여 이에 많은 노력이 필요하다고 강조하였습니다.