

Tutorial 1.A

High Performance Machine Learning, Deep Learning, and Data Science: Principle and Practice

Dhabaleswar K. (DK) Panda, Hari Subramoni, Aamir Shafi, and Nawras Alnaasan
Department of Computer Science and Engineering, The Ohio State University

Abstract

Recent advances in Machine and Deep Learning (ML/DL) have led to many exciting challenges and opportunities. Modern ML/DL and Data Science frameworks including TensorFlow, PyTorch, and Dask have emerged that offer high-performance training and deployment for various types of ML models and Deep Neural Networks (DNNs). This tutorial provides an overview of recent trends in ML/DL and the role of cutting-edge hardware architectures and interconnects in moving the field forward. We will also present an overview of different DNN architectures and ML/DL frameworks with special focus on parallelization strategies for model training. We highlight new challenges and opportunities for communication runtimes to exploit high-performance CPU/GPU architectures to efficiently support large-scale distributed training. We also highlight some of our co-design efforts to utilize MPI for large-scale DNN training on cutting-edge CPU/GPU architectures available on modern HPC clusters. The tutorial covers training traditional ML models including— K-Means, linear regression, nearest neighbours—using the cuML framework accelerated using MVAPICH2-GDR. Also, the tutorial presents accelerating GPU-based Data Science applications using MPI4Dask, which is an MPI-based backend for Dask. Throughout the tutorial, we include hands-on exercises to enable attendees to gain first-hand experience of running distributed ML/DL training and Dask on a modern GPU cluster.

Bios

Dhabaleswar K. (DK) Panda is a Professor of Computer Science and Engineering and University Distinguished Scholar at the Ohio State University. His research interests include parallel computer architecture, high-performance networking, Exascale computing, Big Data, Deep Learning, programming models, accelerators, high-performance file systems and storage, virtualization, and cloud computing. He has published over 500 papers in major journals and international conferences related to these research areas. Dr. Panda and his research group members have been doing extensive research on modern networking technologies including InfiniBand, High-Speed Ethernet, RDMA over Converged Enhanced Ethernet (RoCE), Omni-Path, and EFA. Dr. Panda and his team have been actively working on high-performance MPI and PGAS libraries (<http://mvapich.cse.ohio-state.edu>), Deep Learning libraries (<http://hidl.cse.ohio-state.edu>) and Big Data libraries (<http://hibd.cse.ohio-state.edu>). Dr. Panda has served (or serving) as Program Chair/Co-Chair/Vice-Chair of many international conferences. He is an IEEE Fellow and a member of ACM. More details are available at <http://www.cse.ohio-state.edu/~panda>.

Dr. Hari Subramoni is an assistant professor in the Department of Computer Science and Engineering at the Ohio State University. His current research interests include high performance interconnects and protocols, parallel computer architecture, network-based computing, exascale computing, network topology aware computing, QoS, power-aware LAN-WAN communication, fault tolerance, virtualization, big data, deep learning and cloud computing. He has published over 100 papers in international journals and conferences related to these research areas. He has been actively involved in various professional activities in academic journals and conferences. Dr. Subramoni is doing research on the design and

development of MVAPICH2 (High Performance MPI over InfiniBand, iWARP and RoCE) and MVAPICH2-X (Hybrid MPI and PGAS (OpenSHMEM, UPC and CAF)) software packages. He is a member of IEEE & ACM.

Dr. Aamir Shafi is currently a Research Scientist in the Department of Computer Science & Engineering at the Ohio State University where he is involved in the High Performance Big Data project led by Dr. Dhabaleswar K. Panda. Dr. Shafi was a Fulbright Visiting Scholar at the Massachusetts Institute of Technology (MIT) in the 2010-2011 academic year where he worked with Prof. Charles Leiserson on the award-winning Cilk technology. Dr. Shafi received his PhD in Computer Science from the University of Portsmouth, UK in 2006. He got his Bachelors in Software Engineering degree from NUST, Pakistan in 2003. Dr. Shafi's current research interests include architecting robust libraries and tools for Big Data computation with emphasis on Machine and Deep Learning applications. Dr. Shafi co-designed and co-developed a Java-based MPI-like library called MPJ Express. More details about Dr. Shafi are available from <https://people.engineering.osu.edu/people/shafi.16>.

Nawras Alnaasan is a Graduate Research Associate at the Network-Based Computing Laboratory, Columbus, OH, USA. He is currently pursuing a Ph.D. degree in computer science and engineering at The Ohio State University. His research interests lie at the intersection of deep learning and high-performance computing. He works on advanced parallelization techniques to accelerate the training of Deep Neural Networks and exploit underutilized HPC resources covering a wide range of DL applications including supervised learning, semi-supervised learning, and hyperparameter optimization. He is actively involved in several research projects including HiDL (High-performance Deep Learning) and ICICLE (Intelligent Cyberinfrastructure with Computational Learning in the Environment). Alnaasan received his B.S. degree in computer science and engineering from The Ohio State University. Contact him at alnaasan.1@osu.edu.

Tutorial 1.B

Telemetry Tracing, Profiling, Analysis & Performance of workload offloads on Intel Scalable Processors with Integrated Accelerators

Mrittika Ganguli, Karthik Raman, and Kalyan Jee
Intel

Abstract

This tutorial provides a comprehensive overview of the Intel 4th Generation Xeon Processors (Code name: Sapphire Rapids), highlighting their advanced features, performance improvements, and integration with accelerators. We delve into the role of accelerators in enhancing computational capabilities and discuss the importance of telemetry in optimizing system performance and resource allocation. By leveraging accelerators and telemetry, the Intel 4th Gen Xeon Processors enable businesses to achieve unprecedented levels of efficiency, scalability, and real-time insights.

Bios

Mrittika Ganguli is a Principal Engineer and Director, Cloud Native Pathfinding in Intel's Network and Edge Architecture team (NEX OCTO). She has 25+ years of experience in cloud hardware and software management, network and storage processing control and data plane, cloud orchestration, telemetry QOS and scheduling Architecture. She is active in CNCF and Open Infra opensource initiatives and initiated a Service Mesh Performance (SMP) index called Meshmark. She has a MS in CS and 50+ patents and multiple IEEE papers in this area. Mrittika will present the problem statement in cloud and edge.

Karthik Raman is an accomplished technical expert with more than 20 years of expertise, 5 Years at Intel. in network architecture and cloud platforms. Karthik is leading solutions for telemetry architecture, accelerated computing, and SmartNIC in his present position as Cloud Solutions Architect He has developed complex systems for the optimization of workload performance and has created cutting-edge network architectures and strategies for next-generation datacenter and cloud infrastructure. He has a strong understanding of cloud infrastructures and has designed and deployed scalable and resilient platforms for several of Intel's Top CSP customers. Karthik will discuss Telemetry tracing across the Xeon platform and integrated accelerators and walk through a live demonstration of workload offload showcasing performance baselining and improvements on Intel 4th Gen Scalable processors and QAT offload.

Kalyan Jee is an accomplished technical expert with more than 20 years of expertise in cloud platform architecture. Kalyan is leading solutions for accelerated computing in his present position as Cloud Solutions Architect - 10 Years at Intel. He has developed complex systems using data accelerators for the optimization of workload performance for next-generation data centers and cloud infrastructure. He has a strong understanding of cloud infrastructures and has designed and deployed scalable and resilient platforms for several of Intel's Top CSP customers. Kalyan will present deep dive technical architecture on Xeon data accelerators, QAT, DSA, and IAA on Intel 4th Gen Scalable processors (also known as Sapphire Rapids).

Networking Technologies for High-Performance Computing: Principles and Solutions

Dhabaleswar K. (DK) Panda and Hari Subramoni

Department of Computer Science and Engineering, The Ohio State University

Abstract

InfiniBand (IB), High-speed Ethernet (HSE), RoCE, Omni-Path, EFA, Tofu, and Slingshot technologies are generating a lot of excitement towards building next generation High-End Computing (HEC) systems including clusters, datacenters, file systems, storage, cloud computing and Big Data (Hadoop, Spark, HBase and Memcached) environments. This tutorial will provide an overview of these emerging technologies, their offered architectural features, their current market standing, and their suitability for designing HEC systems. It will start with a brief overview of IB, HSE, RoCE, Omni-Path, EFA, Tofu, and Slingshot. In-depth overview of the architectural features of IB, HSE (including iWARP and RoCE), and Omni-Path, their similarities and differences, and the associated protocols will be presented. An overview of the emerging NVLink, NVLink2, NVSwitch, AMD Infinity Fabric, Slingshot, and Tofu architectures will also be given. Next, an overview of the OpenFabrics stack which encapsulates IB, HSE, and RoCE (v1/v2) in a unified manner will be presented. An overview of libfabrics stack will also be provided. Hardware/software solutions and the market trends behind these networking technologies will be highlighted. Sample performance numbers of these technologies and protocols for different environments will be presented. Finally, hands-on exercises will be carried out for the attendees to gain first-hand experience of running experiments with high-performance networks.

Bios

Dhabaleswar K (DK) Panda is a Professor and University Distinguished Scholar of Computer Science and Engineering at the Ohio State University. He is serving as the Director of the ICICLE NSF-AI Institute (<https://icicle.ai>). He has published over 500 papers. The MVAPICH2 MPI libraries, designed and developed by his research group (<http://mvapich.cse.ohio-state.edu>), are currently being used by more than 3,300 organizations worldwide (in 90 countries). More than 1.69 million downloads of this software have taken place from the project's site. This software is empowering many clusters in the TOP500 list. High-performance and scalable solutions for Deep Learning frameworks and Machine Learning applications from his group are available from <https://hidl.cse.ohio-state.edu>. Similarly, scalable and high-performance solutions for Big Data and Data science frameworks are available from <https://hibd.cse.ohio-state.edu>. Prof. Panda is an IEEE Fellow and recipient of the 2022 IEEE Charles Babbage Award. More details about Prof. Panda are available at <http://www.cse.ohio-state.edu/~panda>.

Dr. Hari Subramoni is an assistant professor in the Department of Computer Science and Engineering at the Ohio State University. His current research interests include high performance interconnects and protocols, parallel computer architecture, network-based computing, exascale computing, network topology aware computing, QoS, power-aware LAN-WAN communication, fault tolerance, virtualization, big data, deep learning and cloud computing. He has published over 100 papers in international journals and conferences related to these research areas. He has been actively involved in various professional activities in academic journals and conferences. Dr. Subramoni is doing research on the design and development of MVAPICH2 (High Performance MPI over InfiniBand, iWARP and RoCE) and

MVAPICH2-X (Hybrid MPI and PGAS (OpenSHMEM, UPC and CAF)) software packages. He is a member of IEEE & ACM.

Tutorial 1.D

Open Fabrics Interfaces Libfabric

Rajalaxmi Angadi and Juee Himalbhai Desai
Intel

Abstract

OpenFabrics Interfaces (OFI) is a framework focused on exporting fabric communication services to applications. OFI is best described as a collection of libraries and applications used to export fabric services. The key components of OFI are application interfaces, provider libraries, kernel services, daemons, and test applications. Libfabric is a core component of OFI. It is the library that defines and exports the user-space API of OFI and is typically the only software that applications deal with directly. It works in conjunction with provider libraries, which are often integrated directly into libfabric. In this tutorial we would introduce OFI libfabric and briefly explain important libfabric APIs. To explain how these APIs can be used we would do a walk-through of a simple application that exchanges hello messages using (Reliable Datagram) RDM. This will then be proceeded to show more advanced aspects of libfabric such as tagged messages, remote memory access and memory registration modes. The tutorial will also show how libfabric stands out by having software interfaces codesigned with fabric hardware providers and leveraging and expanding existing RDMA open-source community.

Bios

Rajalaxmi Angadi is a Middleware Engineering Manager at Intel Corporation, leading Libfabrics software productization team. She has 17+ years of Industry experience in various capacities ranging from design/development to management. She has been heavily involved in development of schemas and mockups for SNIA swordfish in the Scalable Storage Management Technical Working group. She has co-authored multiple technical proposals and contributed to the NVM Express consortium. She is a key contributor of distributed endpoint manager, an open-source tool for managing NVMe-oF resources. She is also a core team member of Technical Women at Intel Network cross-site committee focusing on EMEA region. Prior to Intel, Rajalaxmi has worked with a stealth mode startup, Symantec Software, Computational Research Lab, Hewlett Packard & Flextronics. She holds Masters in Information Technology and Management, and a Bachelors in Computer Science and Engineering.

Juee Himalbhai Desai is a Middleware Development Engineer at Intel Corporation working under Rajalaxmi Angadi to support libfabrics software productization. She has 2.5 years of industry experience working in computer networking. She has been with Intel developing and testing libfabric components for the past 10 months. Prior to Intel, she was with Puribus Networks working on physical layer technologies. She holds an MS in Electrical Engineering from USC and a BE in Electronics and Communication.

Tutorial 2.A

Electro-optical and die-to-die Interconnect Technologies in Datacenters: Significance, Advantages, and Future Implications

Priyank Shukla
Synopsys

Abstract

This tutorial presents a comprehensive exploration of interconnect technologies in datacenters, aiming to provide attendees with a deep understanding of their significance and implications. It introduces interconnect technologies in datacenters such as Ethernet, CXL, PCIe, and InfiniBand, and provides insights into the motivations, advantages, and use cases of these technologies. The tutorial highlights the importance of emerging electro-optical interfaces, die-to-die interconnects and 800GbE/1.6TbE, in shaping the future of datacenters. This tutorial equips attendees with the knowledge and insights necessary to navigate the dynamic landscape of interconnect technologies in datacenters enabling them to make informed decisions regarding system-level power and performance trade-offs.

Bio

With over 15 years of experience in the semiconductor industry, Priyank Shukla is an ASIC IP Product Manager who has managed and driven complex product development cycles from concept to release. Currently, he serves as Sr. Staff Product Manager of Interface IPs at Synopsys Inc. In this role, he oversees a portfolio of products that includes 112G/224G high-speed SerDes, 400G/800G/1.6TbE Ethernet, and PCIe5-7. I am also an IEEE 802.3 Working Group Voter contributing to IEEE802.3dj/df task forces.

Tutorial 2.B

Understanding How Microarchitecture Impacts Microservices

Ashraf Mahgoub, Harshad Sane, and Kshitij Doshi
Intel

Abstract

Microservices offer numerous benefits, including scalability, modularity, and scheduling flexibility. Accordingly, organizations that adapt microservices architecture in their applications can handle higher traffic loads scalably and allocate resources effectively. However, the underlying microarchitecture plays a crucial role in microservices performance. The interaction between the microservices behavior, the underlying microarchitecture, and placement decisions impose various trade-offs. For instance, two microservices that have a high inter-communication traffic are better collocated on the same host to minimize communication latency. Obviously, this comes with reduced placement flexibility. On the other hand, if the two microservices contend for local resources (such as last level cache), spreading them on different hosts minimizes resource contention.

In this tutorial, we describe a methodical approach for addressing this issue. Specifically, we recap ‘map-and-zoom’ performance methodology in the light of microservices. The methodology pursues multiple objectives such as debugging performance regressions and identifying scaling bottlenecks. We show how a family of three tools (developed by Intel) can be used jointly to achieve agile bottlenecks detection as well as fast reaction times. Finally, we show how performance metrics counters can be collected (in a lightweight manner) for each process/container on a given host, analyzed in real-time, and depicted in time-line view.

Bios

Ashraf Mahgoub is a cloud software engineer in the Cloud Engineering and Solutions Group. His technical expertise is in designing and implementing automated systems for resource management and performance optimization for cloud-native applications. He obtained his B.S and M.S degrees in Computer Engineering from Cairo University (2011 and 2015 respectively). He obtained his Ph.D. degrees in Computer Science from Purdue University (2022). His research interests span cloud-native systems, automated performance optimization, and self-aware & Adaptive Computing.

Harshad Sane is a performance engineer in the Data Center and AI group with a deep technical expertise in system software, memory, and CPU architectures. He specializes in performance monitoring, software optimization, and tool development with focus in the cloud domain. Harshad joined Intel as a College Graduate (RCG) in 2008 after completing his M.S. in Electrical and Computer Engineering from UC Boulder.

Kshitij Doshi works at Intel Corporation in the Data Center and AI group, where he focuses on performance optimization of workloads and cloud instances. He obtained his undergraduate degree in Electrical Engineering from IIT Mumbai (1982) and his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Rice University (1985 and 1989, respectively). His research interests span distributed systems, memory and storage architectures, and resource management.

High Performing and Predictable Java Network Programming

Dinesh Kumar, Harshad Sane, Rodolfo De Vega, and Kshitij Doshi
Intel

Abstract

Modern applications are increasingly structured as loosely coupled collaborations, working in networked interactions while producing and retrieving data from massive datasets through random I/Os. Solution developers, once used to the simplicity of doing I/O through synchronous calls, have to increasingly architect for asynchronous-streaming interactions among loosely coupled peers and data sources/sinks. It is common for I/O operations to not only be multiplexed at platform software level, but also to pass through multiple layers of protocol stacks and runtimes, and a variety of proxying-caching-transformation and filtering operations, before being presented to the core logic of a service. Transitions through these intermediary functionalities need to preserve non-blocking I/O interactions that are designed in at the lowest plane, e.g., at a user-kernel/application-runtime boundary. For programmers of core business logic in an object oriented or functional language such as Java, Scala, Kotlin, these benefits are realized through the Netty framework which lets programmers follow a keep-working-instead-of-waiting style in an efficient event-driven model for I/O. In this tutorial we will describe Netty programming at high level and then transition to showing where other inefficiencies can still arise and how they can be addressed by pairing Netty with IO_uring and with hardware supported network steering capabilities.

Bios

Dinesh Kumar is a Network Software Engineer in the Networking and Edge Group, working on Ethernet product development and programable Network solutions. as well om Network Technologies enablement and optimizations for cloud and data center applications. He joined Intel in 1999 after completing his M.S in Computer Science from IET University, India.

Harshad Sane is a performance engineer in the Data Center and AI group with a deep technical expertise in system software, memory, and CPU architectures. He specializes in performance monitoring, software optimization, and tool development with focus in the cloud domain. Harshad joined Intel as a College Graduate in 2008 after completing his M.S. in Electrical and Computer Engineering from UC Boulder.

Rodolfo De Vega is a cloud software engineer in the Data Center and AI group, supporting customers and cloud service providers in optimizing their workloads for performance, both on-premises and in the public cloud. He joined Intel in 2000 after completing his B.S. in Computer Science from Saint Martin's University.

Kshitij Doshi works at Intel Corporation in the Data Center and AI group, where he focuses on performance optimization of workloads and cloud instances. He obtained his undergraduate degree in Electrical Engineering from IIT Mumbai (1982) and his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Rice University (1985 and 1989, respectively). His research interests span distributed systems, memory and storage architectures, and resource management.

Tutorial 2.D

Leveraging SmartNICs for HPC and Data Center Applications

Jeffrey Young, *Georgia Institute of Technology*

Rich Graham, *NVIDIA*

Oscar Hernandez, *Oak Ridge National Laboratory*

Richard Vuduc, *Georgia Institute of Technology*

Abstract

The past few years have witnessed a surge in the number of advanced network adapters, known as “SmartNICs”, that offer additional functionalities beyond standard packet processing capabilities. These devices often feature programmable lightweight processing cores, FPGAs, and even CPU- and GPU-based platforms capable of running separate operating systems. Though primarily aimed at data center operations, such as infrastructure management, packet filtering, and I/O acceleration, SmartNICs are increasingly being explored for high-performance computing (HPC) application acceleration.

This half-day tutorial offers an in-depth exploration of the state-of-the-art for SmartNICs and the emerging software ecosystems supporting them. Attendees will engage in hands-on exercises to better understand how to use SmartNICs for HPC application acceleration, including MPI collective operation offloading, OpenMP remote offload, and algorithmic modifications to maximize on-board processing power. Participants will have the opportunity to execute these exercises using cutting-edge SmartNICs like NVIDIA's BlueField-3 Data Processing Unit (DPU). The tutorial presenters will further discuss optimizing applications to harness SmartNICs as communication accelerators in HPC systems.

Bios

Jeffrey Young is a senior research scientist in Georgia Tech's School of Computer Science. With a background in computer architecture, his main research interests have focused on the intersection of high-performance computing and novel accelerators including GPUs, Xeon Phi, FPGAs, and Arm SVE processors. He is the director of a novel architecture testbed, the CRNCH Rogues Gallery, that aims to simplify and democratize access to novel post-Moore accelerators in the neuromorphic, reversible, and novel networking spaces. He received his PhD in computer engineering in 2013 from Georgia Tech's ECE department.

Dr. Richard Graham is the Senior Director of HPC Technology at NVIDIA's Networking Business unit. His main area of expertise revolves around HPC network software and hardware capabilities for present and upcoming HPC technologies. Before joining Mellanox/NVIDIA, Dr. Graham accumulated thirteen years of experience at Los Alamos National Laboratory and Oak Ridge National Laboratory, where he held technical and administrative positions in computer science. His technical focus encompassed communication libraries and application analysis tools. Additionally, Dr. Graham played a significant role as a co-founder of the Open MPI collaboration and served as the chairman of the MPI 3.0 standardization efforts.

Oscar Hernandez holds a PhD in Computer Science and currently works at Oak Ridge National Laboratory (ORNL). Dr. Hernandez conducts research on programming models, compilers, and tools deployed at supercomputers such as Summit and Frontier at the Leadership Computing Facility (OLCF). At ORNL, he has contributed to the standardization of parallel languages and APIs for accelerated nodes, including OpenACC/OpenMP, as well as communication libraries and frameworks like OpenSHMEM and UCX. Furthermore, he has been involved with the Exascale Computing Project, leading various initiatives to implement these technologies on Exascale systems. Oscar has also collaborated closely with application teams, including the CAAR, INCITE, and ALCC projects, as well as numerous projects funded by DOE, DoD, NSF, and Industrial Partners within the Oil & Gas industry. Additionally, he has extensive experience delivering tutorials at various events, including Supercomputing, ISC, Exascale Computing Annual Meeting, and for NSF.

Richard (Rich) Vuduc is an Associate Professor at the Georgia Institute of Technology in the School of Computational Science and Engineering, a department devoted to the study of computer-based modeling and simulation of natural and engineered systems. His research lab, The HPC Garage is interested in high-performance computing, with an emphasis on algorithms, performance analysis, and performance engineering. He is a recipient of a DARPA Computer Science Study Group grant; an NSF CAREER award; a collaborative Gordon Bell Prize in 2010; Lockheed-Martin Aeronautics Company Dean's Award for Teaching Excellence (2013); and Best Paper Awards at the SIAM Conference on Data Mining (SDM, 2012) and the IEEE Parallel and Distributed Processing Symposium (IPDPS, 2015), among others. Most recently, Dr. Vuduc has led an effort to map high-performance applications like LAMMPS and MueLu to Data Processing Unit (DPUs) with early results resulting in an IPDPS 2022 Best Paper nominee.