



IISWC 2020

Inspur Workshop Digital

About IISWC 2020

Date: October 27 – October 29, 2020

Host: IEEE Computer Society & Technical Committee on Computer Architecture

Sponsors: Inspur, Huawei, ARM, Facebook, etc.

Focused Topics: characterizing and understanding emerging applications in consumer, commercial and scientific computing

- Characterization of applications in domains including Machine learning, data analytics, data mining, Quantum computing, High performance computing, Cloud and edge computing, Blockchain services
- Emerging workloads and architectures, such as Quantum computations and communication, Artificial intelligence and transactional memory workloads
- Characterization of OS, Virtual Machine, middleware and library behavior
- Implications of workloads in system design
- Benchmark methodologies and suites
- Measurement tools and techniques

IISWC 2020 Website: <http://www.iiswc.org/iiswc2020/index.html>

Practice and Research of Commercial Systems and Application Domains

Tuesday, October, 27 (EDT, UTC-5)

No.	Topic	Speaker
1	Frontiers in Multimodal AI	Yaqian Zhao <i>Deputy General Manager of Architecture Research Department, Inspur Information</i>
2	Federated Intelligence Accelerates AI Landing	Jianzong Wang <i>Deputy Chief Engineer, PingAn Technology</i>
3	Proactive Disk Failure Prediction System towards Reliable Data Centers	Tuanjie Wang <i>Senior Software Development Engineer of Storage R&D Department, Inspur Information</i>
4	Key Design Issues of Massive Optical Archive Storage Systems	Jie Yao <i>Professor of Wuhan National Laboratory for Optoelectronics Huazhong University of Science and Technology</i>
5	Large-Scale Optimization Strategies for Typical HPC Workloads	Qian Wang <i>Senior application engineer of AI&HPC Department, Inspur Information</i>
6	Deep Neural Network Compression and Model Deployment	Tao Li <i>Professor of Intelligent Computing System Lab Nankai University</i>
7	Analysis and Practice of SPC-1 (Storage Benchmarking)	Hao Sun, <i>Senior Software Development Engineer of Storage R&D Department, Inspur Information</i>
8	The Past, Present and Future of Blockchain Storage	Peter Ye <i>Senior Architect of Storage Product Line Department, Inspur Information</i>



Yaqian Zhao

Deputy General Manager of Architecture Research Department

Inspur Information

Frontiers in Multimodal AI



Recently, the exploration of multimodal AI is prospective since our experience of the world is multimodal. The idea to solve the difficult multimodal tasks is firstly pre-training a deep learning model with large amounts of data and then fine-tuning it to adapt the downstream tasks. Breakthroughs in single modal tasks of CV and NLP have brought a golden opportunity for Visual & Language Learning. The presentation will cover some mature pre-training multimodal models, such as VisualBert. With the usage in typical multimodal downstream tasks, this talk will explain the succeed of these models.

Jianzong Wang

Deputy Chief Engineer

PingAn Technology



Federated Intelligence Accelerates AI Landing

In the new development of artificial intelligence, as a core asset and under the constraints of laws and regulations and enterprise privacy protection, the data has formed information isolated island that hinders the implementation of AI. The core technology to break this data dilemma is federal intelligence, which is also the key to our breakthrough and application. Federal intelligence includes federated learning, federated data tribes, federated inference, and federated incentive systems throughout the framework system. Based on the theoretical framework of federated intelligence, the hive platform can be applied to smart finance, smart cities, intelligent medical, and etc. In this way, each participant can establish a joint learning relationship with others, thus not only achieving the goal of reducing costs and increasing efficiency, but also making the best use of the value of the industry.

Tuanjie Wang

Senior Software Development Engineer of Storage R&D

Department, Inspur Information



Proactive Disk Failure Prediction System towards Reliable Data Centers

In modern datacenter, hard disk drive has the highest failure rate. Current storage system has data protection feature to avoid data loss caused by disk failure. However, data reconstruction process always slows down system services. If disk failures can be predicted accurately, data protection mechanism can be performed in advance. Disk failure prediction dramatically improve the reliability and availability of storage system. We analyze disk SMART data in detail, design effective feature engineering and two-stage normalization preprocessing method. And we optimize XGBoost model through hyperparameter tuning and cost-sensitive learning. Finally, we apply ensemble learning to further improve the accuracy of prediction. The experimental results of Alibaba dataset show that our system predict disk failures within 30 days effectively and the F1-score achieves 39.98.

Jie Yao

Professor of Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology

Key Design Issues of Massive Optical Archive Storage Systems



- Why massive optical archive storage system is promising for long-term preservation of data?
- What are the key design issues?
- Solutions on above issues
- Conclusions

Qian Wang

Senior Application Engineer of AI&HPC Department
Department Inspur Information

Large-Scale Optimization Strategies for Typical HPC Workloads

Ensuring performance of applications running on large-scale clusters is one of the primary focuses in HPC research. In this talk, we will show our strategies on performance analysis and optimization for applications in different fields of research using large-scale HPC clusters. Our strategies are designed to comprehensively analyze runtime features of applications, parallel mode of the physical model, algorithm implementation and other technical details. This three levels of strategy covers platform optimization, technological innovation, and model innovation, and targeted optimization based on these features. State-of-the-art CPU instructions, network communication and other modules, and innovative parallel mode of some applications have been optimized. After optimization, these applications will perform even better when running up to 8000 cores, with obvious increase in performance compared to the non-optimized ones.



Tao Li

Professor of Intelligent Computing System Lab

Nankai University

Deep Neural Network Compression and Model Deployment

Deep neural networks (DNNs) have achieved great success in many fields. However, DNN models often have a large number of computations and many parameters, which hinder their deployment in low memory resource devices or applications with strict latency requirements, such as mobile phones and embedded devices. DNN compression is the key technology to meet this challenge. This talk will discuss DNN compression and heterogeneous acceleration methods on several FPGA platforms. Finally, we will introduce our recent works on DNN quantization and model deployment for the applications such as intelligent fundus image computing and person ReID.



Hao Sun

Senior Software Development Engineer of Storage R&D Department,
Inspur Information



Analysis and Practice of SPC-1

In today's society, with the rapid growth of data volume and the multiplication of data business volume, storage products are playing an increasingly important role in the business. The performance of storage products is an important technical index of storage products. However, due to the complexity of business scenarios, it is difficult to measure the performance of storage devices. How to comprehensively measure storage performance is an urgent problem to be solved. This paper introduces the following contents on storage performance:

- How does SPC-1 test storage performance
- Inspur storage performance optimization practice

Peter Ye

Senior Architect of Storage Product Line Department

Inspur Information

The Past, Present and Future of Blockchain Storage

- Why Blockchain (New point of view)
- Distributed Cloud - Blockchain IT Infrastructure (New IT model with checks and balance)
- Two kinds of Blockchain Storage
- Why Blockchain Storage?
- Zero Knowledge Proof
- Inspur Storage and Blockchain Storage





| Thank You!