ATTENDEES

Conference Program

Preliminary schedule for IEEE CCGrid 2026, 18-21 May 2026, Sydney, Australia. This page will be updated as the detailed technical program, room assignments, and speaker information are finalized.

Monday, 18 May 2026

Tutorials and workshops day

Time	Conf Room 1	Conf Room 2	Conf Room 3	Conf Room 4
08:00 - 17:00	Registration: Foyer			
09:00 - 10:30	**Tutorial 1**	**CIC2**	**FL-RCE**	**Continuum-RI**
10:30 - 11:00	Morning Tea			
11:00 - 12:30	**Tutorial 1**	**CIC2**	**FL-RCE**	**Continuum-RI**
12:30 - 14:00	Lunch			
14:00 - 15:30	**Tutorial 2**	**CIC2**	**HEAD**	
15:30 - 16:00	Afternoon Tea			
16:00 - 17:30	**Mentorship Session**			
18:30 - 20:30	Welcome Reception: Foyer			

Session Details

Tuesday, 19 May 2026

Main conference opening and technical sessions

Time	Conf Room 1	Conf Room 2	Conf Room 3	Conf Room 4
08:00 - 17:00	Registration: Foyer			
09:00 - 09:30	Opening Ceremony			
09:30 - 10:30	Keynote: Prof Guo			
10:30 - 11:00	Morning Tea			
11:00 - 12:30	**GPU-1**	**LLM-1**	**WORKFLOW**	**Doctoral Symposium**
12:30 - 14:00	Lunch			
14:00 - 15:30	**GPU-2**	**LLM-2**	**AI-SEC**	**Posters**
15:30 - 16:00	Afternoon Tea			
16:00 - 17:30	**HPC-SCH**	**AI-SYS**	**GREEN**	
17:30 - 18:30	Panel Discussion			

Session Details

11:00 - 12:30

GPU-1) GPU Communication and HPC Acceleration (

Kernel-Initiated One-Sided Networking for GPU-Accelerated AI Workloads

Khaled Hamidouche, John Bachan, Pak Markthub, Peter-Jan Gootzen, Elena Agostini, Sylvain Jeaugey, Aamir Shafi, Georgios Theodorakis and Manjunath Gorentla Venkata

Full paper

Mammoth: Macro-Level MPI Offloading to Off-path Accelerator in DPU

Jong-Bin Lee, Pu-Rum Seo, Ki-Moon Jeong and Hyun-Wook Jin

Full paper

Scalable, Topology- and Multi-HCA-Aware Hierarchical GPU Allgather Using Parallel Rings

Amirreza Barati Sedeh, Ryan Grant and Ahmad Afsahi

Full paper

WORKFLOW) Distributed Workflow Management and Optimization (

ElastiFlow: Elastic Resource Management for Iterative Scientific Workflows in Hybrid HPC-Cloud Infrastructure

Srishti Dasgupta, Kavitha Subramaniam and Michael Gerndt

Full paper

Efficiently Reproducing Distributed Workflows in Notebook-based Systems

Talha Azaz, Raza Ahmad, Saiful Islam, Douglas Thain and Tanu Malik

Full paper

Compass: Optimizing Compound AI Workflows for Dynamic Adaptation

Milos Gravara, Juan Luis Herrera and Stefan Nastic

Full paper

LLM-1) Efficient LLM Inference Systems - 1 (

Characterizing LLM Inference Energy-Performance Tradeoffs across Workloads and GPU Scaling

Paul Joe Maliakel, Shashikant Ilager and Ivona Brandic

Full paper

Quicktopia: Iteration-Level GPU Frequency Control for Energy-Latency Co-Optimization in LLM Inference

Soyang Baek, Bodon Jeong, Hongsu Byun and Sungyong Park

Full paper

ARKV: Adaptive Resource-Efficient KV Cache Management for Long Context LLM Inference under Memory Constraints

Jianlong Lei and Shashikant Ilager

Full paper

14:00 - 15:30

GPU-2) GPU Resource Management and Runtime Optimization (

ADVICE: Automatic Identification of Variables to Checkpoint through Compiler Augmentation

Xin Huang, Luanzheng Guo, Nathan Tallent and Kento Sato

Full paper

Enhanced SVM for Improving Application Performance under GPU Memory Oversubscription

Bennett Cooper, Thomas Scogland and Rong Ge

Full paper

Priority-Aware GPU Co-Scheduling for High Performance Computing

Naman Kulshreshtha, Tapasya Patki, Aniruddha Marathe, Tom Scogland and Rong Ge

Full paper

LLM-2) Efficient LLM Inference Systems - 2 (

EmbdC: Error-Bounded Lossy Video Embedding Compression For On-Device LLM Inference

Bo Jiang, Taolue Yang, Youyuan Liu, Sheng Di and Sian Jin

Full paper

LLM-Pilot: SLO-Aware and Cost-Efficient LLM Serving on Public Cloud VM Clusters via Offloading

Jinwoo Kim, Kihyun Kim, Hyunsun Chung, Jihoon Yang, James J. Kim, Dong Li and Youngjae Kim

Full paper

EAFAL: An Edge-Based Agentic Framework for Adaptive Selection between SLMs and LLMs

Chamara Madarasingha, Prajyot Singh, Redowan Mahmud, Mahbuba Afrin, Aneesh Krishna and Salil Kanhere

Full paper

AI-SEC) Security and Reliability in AI Systems (

Exploring Silent Data Corruption as a Reliability Challenge in LLM Training

Anton Altenbernd, Philipp Wiesner and Odej Kao

Full paper

Noise-Aware Misclassification Attack Detection in Collaborative DNN Inference

Shima Yousefi and Saptarshi Debroy

Full paper

SeCI: A Framework for Self-Certified Identity for Autonomous AI Agents

Nehal Al-Otaiby, Mohammad Hammoudeh and Jameleddine Hassine

Full paper

16:00 - 17:30

HPC-SCH) HPC Scheduling and Execution Optimization (

GASched: Goal-Adaptive Hierarchical Reinforcement Learning for Multi-Objective HPC Job Scheduling

Minsol Choo and Sangyoon Oh

Full paper

Reducing Backfill Failures from Workload Drift with Lightweight Uncertainty Buffers in HPC Job Scheduling

Jiheon Choi and Sangyoon Oh

Full paper

Automated Configuration of Power-Management Knobs for Optimal HPC Job Executions

Francesco Antici, Andrea Proia, Ryoma Ohara, Toshihiro Hanawa, Zeynep Kiziltan, Andrea Bartolini and Jens Domke

Full paper

AI-SYS) LLM / AI Training & Inference Systems (

Reuse-Aware Min-Cost Flow Scheduling for Distributed LLM Training in Hybrid Optical-Electrical Networks

Yifan Yang, Gongming Zhao, Hongli Xu and Huihui Tang

Full paper

Tula: Optimizing Time, Cost, and Generalization in Distributed Large-Batch Training

Sahil Tyagi and Feiyi Wang

Full paper

VEX: Scaling HNSW-Based Vector Search with DPU Memory and Parallelism

Kihwan Kim, Hyungsun Yoo, Woojung Kim, Donghyun Min, Myungcheol Lee, Jihoon Yang, Weikuan Yu and Youngjae Kim

Full paper

GREEN) Energy, Sustainability & Carbon-Aware Computing (

Green or Fast? Learning to Balance Cold Starts and Idle Carbon in Serverless Computing

Bowen Sun, Christos Antonopoulos, Evgenia Smirni, Bin Ren, Nikolaos Bellas and Spyros Lalis

Full paper

STEAM: Realistic Modeling and Systematic Exploration of Composable Techniques for Sustainable Datacenter

Dante Niewenhuis, Sacheendra Talluri, Alexandru Iosup and Tiziano De Matteis

Full paper

Spanergy: Energy-aware Distributed Tracing for Microservices

Cesar Perdigao Batista, Denis Conan and Sophie Chabridon

Full paper

Wednesday, 20 May 2026

Keynote, best-paper sessions, and community events

Time	Conf Room 1	Conf Room 2	Conf Room 3	Conf Room 4
08:00 - 17:00	Registration: Foyer			
09:30 - 10:30	Keynote: Prof Keahey			
10:30 - 11:00	Morning Tea			
11:00 - 12:30	**EDGE-CLOUD**	**FL-1**	**OPT-SYS**	**SCALE Challenge**
12:30 - 14:00	Lunch (Steering Committee)			
14:00 - 15:30	**EDGE-AI**	**FL-2**	**DATA-SYS-1**	**ICFEC 1**
15:30 - 16:00	Afternoon Tea			
16:00 - 17:30	**EDGE-IoT**	**FL-3**	**DATA-SYS-2**	**ICFEC 2**
19:30 - 22:30	Conference Dinner			

Session Details

11:00 - 12:30

EDGE-CLOUD) Edge-Cloud Resource Management for AI Workloads (

Multi-Objective Load Balancing for Heterogeneous Edge-Based Object Detection Systems

Daghash Alqahtani, Maria Rodriguez, Muhammad Aamir Cheema and Adel N. Toosi

Full paper

Deep Reinforcement Learning-driven Edge Offloading for Latency-constrained XR pipelines

Sourya Saha and Saptarshi Debroy

Full paper

Decentralized Resource Sharing in Edge-Cloud Federations via Multi-Agent Hierarchical Reinforcement Learning

Panagiotis Kokkinakis, Polyzois Soumplis and Emmanouel Varvarigos

Full paper

FL-1) Federated / Decentralized Learning - 1 (

A Multi-Armed Bandit-Based Participant Selection Method for Federated Recommendation Systems

Jintao Liu, Mohammad Goudarzi and Adel N. Toosi

Full paper

FedMO: Mobility-Aware Client Selection in Federated Learning for Drone Delivery Systems

Jiang Yuan, Xiao Liu, Jia Xu, Aiting Yao, Frank Jiang and Xuejun Li

Full paper

Heterogeneous-Resource-Aware Federated Learning with Intelligent LoRA Allocation and Aggregation

Youye Xie, Yao Lian, Kevin Chen, Abdul Latif, Lingzhi Zhao and Reza Farivar

Full paper

OPT-SYS) Scheduling & Systems Optimization (

AdaSched: A Performance-Driven Cluster Scheduler for Deep Learning Workloads using Deep Reinforcement Learning

Han Yin, Jialun Li, Xuan Mo and Weigang Wu

Full paper

Congestion-Aware Pricing for Fast and Efficient Edge-Cloud Computing

Polyzois Soumplis and Emmanouel Varvarigos

Full paper

SD-MoE: Scenario-Driven MoE Forecasting for Intelligent Elastic Scaling in Cloud Clusters

Xianzhao Guo, Weipeng Cao, Minxian Xu, Dachuan Li, Chuanfei Xu, Xi Tao and Zhong Ming

Full paper

14:00 - 15:30

EDGE-AI) Edge Systems for AI Agents (

Orchestrating WASM-based MCP Tool Runtimes for AI Agents across Edge-Cloud Continuum

Moohyun Song, Hayoung Kim, Kyoohyun Lee, Jae Gi Son and Kyungyong Lee

Full paper

SwiftBot: A Decentralized Platform for LLM-Powered Federated Robotic Task Execution

Hailu Xu, Simon Zhang, Zhengxiong Li, Shuai Xu, Xiaokun Yang and Fangtian Zhong

Full paper

DriveCache: On-Board Compute Caching for Scalable Vehicular Edge Computing Networks

Suvarthi Sarkar, Aditya Gupta, Salil Kashyap and Aryabartta Sahu

Full paper

FL-2) Federated / Decentralized Learning - 2 (

Dual-Distilled Heterogeneous Federated Learning with Adaptive Margins for Trainable Global Prototypes

Fatema Siddika, Md Anwar Hossen, Wensheng Zhang, Anuj Sharma, Juan Pablo Munoz and Ali Jannesari

Full paper

RDA-CAFL: Reputation-Dynamic and Distillation-based Asynchronous Conflict-Aware Federated Learning for UAV Networks

Jie Li, Pengyang Li, Liming Sun, Xingwei Wang and Yihang Zhang

Full paper

Evaluating Federated Learning Beyond Simulation: A Deployment-Aware Methodology

Mathis Valli, Cedric Tedeschi, Alexandru Costan, Loic Cudennec and Gabriel Antoniu

Full paper

DATA-SYS-1) Hardware-Aware Storage & Memory Systems (

Making Variable-Size I/O Practical in ZNS SSDs

Sijie Lan, Abutalib Aghayev, Mahmut Kandemir and Umesh Maheshwari

Full paper

NIO-Cache: Device-Affinitive Page Cache Placement Mechanism for NUMA Systems

Jiazheng Zhang, Xiaoyang Wang and Jiwu Shu

Full paper

S-MSHR: A Scalable MSHR Architecture Using Cache Tag Data-Ready Bits and Index Queues

Shuang Wu, Xu Zhang, Yibin Xu, Yangyang Zhao, Tianyue Lu and Mingyu Chen

Full paper

ICFEC 1) Monitoring and Management in the Computing Continuum (

K-Sense: A Non-Invasive eBPF Framework for QoS Inference

Abdullah Muslim, Ali Beiti Aydenlou and Stephan Recker

Cost-Effective Processing of IoT Data in the Computing Continuum

Vasileios Karagiannis, Drazen Ignjatovic, Antonios Iosifidis and Stefan Schulte

LLMEdger: Phase-Aware Model Parallelism Scheduler for LLM Inference on Edge

Xinyang Shen and Lena Mashayekhy

16:00 - 17:30

EDGE-IoT) IoT & Edge Systems Performance (

Benchmarking Message Brokers for IoT Edge Computing: A Comprehensive Performance Study

Tapajit Chandra Paul, Pawissanutt Lertpongrujikorn, Hai Nguyen and Mohsen Amini Salehi

Full paper

Volatility-Aware Adaptive Context Caching for Real-Time Context-Aware IoT Applications

Ashish Manchanda, Prem Prakash Jayaraman, Abhik Banerjee and Arkady Zaslavsky

Short paper

Hierarchical Semi-Supervised Federated Learning for UAV-Enabled Fire Monitoring

Mark Adrian Gambito, Bahman Javadi, Lorenzo Carnevale and Massimo Villari

Short paper

Synthetic Data Generation for Storage Failure Prediction in Large-Scale Systems

Chandranil Chakraborttii, Ana Solorzano and Devesh Tiwari

Short paper

FL-3) Federated / Decentralized Learning - 3 (

FedOort: A Fair and Efficient Optimization Method for Federated Learning

Zhaohua Zheng, Junhui Du, Qiquan Chen, Xu Han, Qijun Huang and Jin Zhang

Full paper

Evidential Trust-Aware Model Personalization in Decentralized Federated Learning for Wearable IoT

Murtaza Rangwala, Richard Sinnott and Rajkumar Buyya

Full paper

CroSatFL: Energy-Efficient Federated Learning with Cross-Aggregation for Satellite Edge Computing

Nan Yang, Bahman Javadi, Rodrigo Neves Calheiros, David Boland and Philip Leong

Full paper

DATA-SYS-2) Data Systems & Storage Engines (

ARC: Adaptive Resource Coordination for Write-Stall Mitigation in LSM-Tree

Guangxun Zhao, Yongjie Zhu, Suhwan Shin, Seehwan Yoo and Jongmoo Choi

Full paper

Altocumulus: Enabling Efficient Erasure Coding in IPFS

Mohammad Rizk, Shadi Ibrahim and Thomas Lambert

Full paper

BucketLSM: What, After All, Is the I/O Bottleneck Behind L0 Key-Range Overlap in LSM-Based KV Stores?

Jaewan Park, Kyungwook Min, Sungjin Byeon, Taewan Noh, Hyungi Park, Xubin He, Hongyeon Kim and Youngjae Kim

Full paper

ICFEC 2) Efficient On-Device Learning and Model Compression (

Layer-Wise Weight Sharing for Efficient Transformer on SoCs

Saeed Khalilian Gourtani, Hang Xu, Nirvana Meratnia and Anuj Pathania

CORAL: Covariance-Guided Resource Adaptive Learning for Efficient Edge Inference

Ahmad Nabhaan, Zaki Sukma, Rakandhiya Rachmanto, Muhammad Santriaji, Byungjin Cho, Arief Setyanto and In Kee Kim

Order-Aware Compression for RF-DETR on Edge Devices: Overcoming Graph Fragmentation and Quantization Instability

Farhan Mahmood, Michalis Karamousadakis, Antonis Porichis, Vishwanathan Mohan and Panagiotis Chatzakos

Thursday, 21 May 2026

Final technical sessions and closing

Time	Conf Room 1	Conf Room 2	Conf Room 3	Conf Room 4
08:00 - 17:00	Registration: Foyer			
09:30 - 10:30	Keynote: Prof Usman			
10:30 - 11:00	Morning Tea			
11:00 - 12:30	**CLOUD**	**ML-SYS-1**	**EMERGE-SYS**	**ICFEC 3**
12:30 - 14:00	Lunch			
14:00 - 15:30	**SERVERLESS-SEC**	**ML-SYS-2**		**ICFEC 4**
15:30 - 16:00	Afternoon Tea			

Session Details

11:00 - 12:30

CLOUD) Cloud Infrastructure and Orchestration (

Larger Cloud Servers, Fewer Hosts? On the Evolution of VM Sizes in IaaS Platforms

Pierre Jacquet, Camille Coti and Marcos Dias De Assuncao

Full paper

TelePod: Live Migration for Stateful Containers

Mingjie Yan, Atharva Ranade, Xin Zhang and Kartik Gopalan

Full paper

Cost-Justified Multi-type Resource Fair Scheduling for Kubernetes

Bo Yan, Sujoy Sikdar and Madhusudhan Govindaraju

Short paper

XCAgent: Automating Multi-Cloud Deployment of Agentic Workflows on FaaS Platforms

Varad Kulkarni, Vaibhav Jha, Nikhil Reddy, Anand Eswaran, Praveen Jayachandran and Yogesh Simmhan

Short paper

ML-SYS-1) Learning-Based Performance Modeling (

CloudFormer: An Attention-based Performance Prediction for Public Clouds with Unknown Workload

Amirhossein Shahbazinia, Darong Huang, Luis Costero and David Atienza

Full paper

PM2Lat: Highly Accurate and Generalized Prediction of DNN Execution Latency on GPUs

Truong-Thanh Le, Hoang-Loc La, Amir Taherkordi, Frank Eliassen, Phuong Hoai Ha and Peiyuan Guan

Full paper

Towards Proactive AIOps: Transfer Learning for Unsupervised Anomaly Detection via Bi-LSTM

Autoencoder in the Computing Continuum

Danny De Novi, Lorenzo Carnevale and Massimo Villari

Full paper

EMERGE-SYS) Advances in Emerging Systems and Security (

Efficient Concurrent GHZ State Distribution in Quantum Networks

Charles Cao, Weisheng Si, Jong Choi and Sajal Das

Full paper

QRAP: Adaptive Quantum-Safe Risk-Aware Prioritization for Cloud Applications

Surabhi Garg, Delton Myalil, Nidhi Singh, Shiv Shankar, Meena Singh Dilip Thakur and Rajan M A

Short paper

BCRLSecureLink: A Blockchain, Cryptography, and Reinforcement Learning-Based Defense Against Link Discovery Attacks in SDVN

Patikiri Arachchige Don Shehan Nilmantha Wijesekara, Harsha Sandaruwan Gardiyawasam Pussewalage, Kalupahana Liyanage Kushan Sudheera and Geeth Priyankara Wijesiri

Short paper

Lifting to tensors when compiling scientific computing workloads for AI Engines

Nick Brown and Gabriel Rodriguez-Canal

Short paper

ICFEC 3) Intelligent Infrastructure and Multi-Tier Orchestration (

Multi-Provider Caching in Multi-Tier Fog Networks

Ferdous Sharifi, Young Choon Lee and Shaahin Hessabi

NL-CPS: Reinforcement Learning-Based Kubernetes Control Plane Placement in Multi-Region Clusters

Sajid Alam, Amjad Ullah and Ze Wang

UAV-Enabled Integrated Sensing, Semantic Communication, and Computation: Disaster-Oriented Edge Computing and Sensing

Yaxi Liu, Wencan Mao, Xulong Li, Yu Xiao, Wei Huangfu and Keping Long

14:00 - 15:30

SERVERLESS-SEC) Security and Analysis of Serverless Systems (

Kumo: A Security-Focused Serverless Cloud Simulator

Wei Shao, Chongzhou Fang, Khaled Khasawneh, Setareh Rafatirad and Houman Homayoun

Full paper

LogLearners: Identifying Compromised AI Functions in Serverless Systems

Adil Bin Bhutto, Erik Elmroth and Monowar Bhuyan

Full paper

PoliFlow: Inferring Control-Flow Policies from Serverless Workflows

Pedro Escaleira, Vitor A. Cunha, Joao P. Barraca, Diogo Gomes and Rui L. Aguiar

Full paper

ML-SYS-2) ML Systems & Optimization (

ACNNS: A Multi-interest Recommendation Model with Capsule Network

Yan Zhang, Xiaotong Cui, Nan Wang and Yingli Zhong

Full paper

InverseTune: Inverse Synthetic Fine-Tuning for Reliable Structured Output in Small Language Models

Markus Goetz and Hojjat Baghban

Short paper

SOtrain: Efficient LLM Fine-Tuning on a Consumer GPU via User-space I/O and Scheduling Optimizations

Zhengguo Liu, Hao Lan and Jiwu Shu

Short paper

RapidGNN: Communication-Efficient Distributed Training on Large-Scale Graph Neural Networks

Arefin Niam, Tevfik Kosar and M. S. Q. Zulkar Nine

Short paper

ICFEC 4) Security and Coordination in Federated Learning (

A Federated LLM-based Framework for DDoS Defense in Mobile Edge Computing

Shuo Zhang, Kousuke Mori and Toshio Hirotsu

Atlas Synchronization in the Hierarchical Federated Learning Continuum

Antonios Iosifidis, Vasileios Karagiannis and Stefan Schulte