Workshop on Data Fabric for Hybrid Clouds (WDFHC)

held in conjunction with
29th IEEE INTERNATIONAL CONFERENCE ON
HIGH PERFORMANCE COMPUTING, DATA, & ANALYTICS (HiPC 2022)

December 18, 2022

A number of organizations have adopted the hybrid-cloud paradigm to optimize business processes. Hybrid clouds span public and private clouds, different public cloud providers as well as edge and cloud resources. A hybrid cloud architecture enables enterprises to scale computing resources, optimize cost and use the best in the class tools. However, it also poses a set of challenges around data discovery, access, governance, usage, and quality. This is especially the case for data-rich machine learning (ML) and artificial intelligence (AI) workloads that are gaining dominance within enterprises.

Data fabric (or a data mesh) is a data management design to make data curation and access transparent in a distributed environment, and alleviate these challenges. A data fabric enables access to the right data to authorized users in an interoperable format while respecting the regulatory and governance constructs. It continuously links enterprise-wide data through their active and passive metadata from different databases, data lakes, and warehouses; orchestrates data placement and integration; and uses semantic inferencing and knowledge graphs to support decision-making by ML and real-time applications.

The Workshop on Data Fabric for Hybrid Clouds (WDFHC) is a half a day event co-located at HiPC 2022 with the aim of getting academia and industry researchers together to share their perspectives on challenges to data management in a hybrid cloud environment, and the role of data fabrics in addressing these challenges. The workshop explores questions such as: *(1) What are the requirements from Data Fabrics to operate in a multi-cloud ecosystem? (2) What are the design principles for Data Fabrics to meet their objectives? (3) What are the core components needed for Data Fabric to support AI/ML applications?*

The workshop has accepted three papers for publications in the proceedings that examine several of these problems. The article on compliance-aware distributed deep learning by Maurya, et al. examines the challenges of machine learning on hybrid clouds while Demicco, et al. offer their perspectives on security analysis when performing labeling-based integrity schemes. We also have an overview of verifiable compliance in the context of various data privacy laws presented by Awasthi. These, combined with two invited talks by renowned researchers and practitioners, Prof. DK Panda (Ohio State University) and Sanil Kumar (SODA Foundation) in the area will make for an engaging and fruitful interaction at the workshop, and advance our collective understanding of this emerging discipline.

The workshop chairs thank the authors for considering our forum for submitting their works, and the reviewers for providing timely and insightful feedback to the authors to enhance their contributions. We also thank the HiPC conference for providing a platform to host this workshop.

Workshop Organization

General Chairs
Yogesh Simmhan,
Indian Institute of Science, Bangalore
Sameep Mehta
IBM Research India, Bangalore

Program Chairs
Kalapriya Kannan
Hewlett Packard Enterprise (HPE), Bangalore
Subhajit Sidhanta
Indian Institute of Technology (IIT), Bhilai

Program Committee
Soumajit Pramanik, IIT Bhilai
Sushanta Karmakar, IIT Guwahati
Nirnay Ghosh, IIEST Shibpur
Madhumita Mallick, Huawei Technologies India
U Das, Rubrik, Inc.
Preeti Syal, Amazon Web Services
Pranay Lohia, Microsoft
Chiranjib Sur, Shell
Bivas Mitra, IIT Kharagpur
Arghya Kusum Das, University of Alaska Fairbanks
Sayan Goswami, Ahmedabad University
Vishwesh Jatala, IIT Bhilai