Contribution ID: **141**   Type: **Oral Presentation**

Supercomputing, Databases and Data Ecosystems: The Best of Three Worlds for Efficient Data-Driven Workflows

Tuesday, 17 March 2026 16:00 (25 minutes)

As supercomputing facilities increasingly run data analytics and artificial intelligence (AI) workloads, an efficient handling of external data sources, of storage and of data flows becomes paramount. However, often this is just addressed via optimization of parallel file systems, network connectivity and I/O libraries. In industry, it is common practice to support data-driven applications with optimized database-management systems, object storage, and data-streaming and caching techniques. The respective backend technologies are also increasingly important for data exchange and sharing via standardised methods (e.g., the IDSA Dataspace protocol) which give a response to data availability challenges in AI.

We thus argue that the IT behind data-driven workflows must be evolved at supercomputing centres – which is one core idea of our Extreme Data Analytics project "EXA4MIND". On the computing-infrastructure side, the centres have shifted from exclusively running High-Performance-System (HPC) clusters to offering a mixed landscape of HPC, Cloud-Computing, GPU-computing, quantum-computing and other systems.

EXA4MIND makes best usage of such infrastructure and provides flexibly deployable modules for boosting and managing data-driven workflows. For piloting, it involves IT4Innovations (CZ) and LRZ (DE) as HPC centres and treats application cases in molecular dynamics (MD), autonomous driving, smart vineyards and public-health data. Within this scope, the project produces the following modules for Extreme Data processing: (i) storage submodules for the management of temporarily-instantiated or longer-term data stores optimised for each application case (object storage middleware, SQL and noSQL databases including vector databases); (ii) the Advanced Query and Indexing System (AQIS), where Extreme Data and AI workflows are orchestrated and data queries across backends can be executed via natural-language query and through extensions to the LEXIS Platform 2 enabling efficient orchestration across computing systems; (iii) toolboxes for data transfer, validation, preprocessing and analytics; (iv) connectivity modules enabling FAIR research data management and connectivity to European or international data ecosystems such as Data Spaces according to the IDSA Dataspace protocol..

In this contribution, we present the EXA4MIND approach and architecture, giving users from science, SMEs and industry a direction for realising Extreme Data workflows.

Primary authors: Dr HACHINGER, Stephan (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities); Dr GOLASOWSKI, Martin (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Mr HAYEK, Mohamad (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities); Prof. KARAGOZ, Pinar (Department Of Computer Engineering, Middle East Technical University); Prof. TOROSLU, Ismail Hakki (Department Of Computer Engineering, Middle East Technical University); Prof. OTYEPKA, Michal (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Mr CIZ, David (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Dr HURYCH, David (valeo.ai); Dr HARSH, Piyush (Terraview GmbH); Mrs CULELL, Maria Ignacia (AUSTRALO); Dr SLANINOVA, Katerina (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Mr FURMANEK, Radek (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Mr SWIATKOWSKI, Jan (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava); Dr MARTINOVIC, Jan (IT4Innovations National Supercomputing Center (IT4I), VŠB - Technical University of Ostrava)

Presenter: Dr HACHINGER, Stephan (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities)

Session Classification: Converging High PerformanComputing Infrastructures: Supercomputers, clouds, accelerators

Track Classification: Track 9: Converging High Performance Computing Infrastructures: Supercomputers, clouds, accelerators