

T 12: Data, AI, Computing, Electronics II (Data Management, Workflow)

Time: Monday 16:45–18:15

Location: VG 2.102

T 12.1 Mon 16:45 VG 2.102

Adaptation of the HammerCloud visualization to state-of-the-art tools - online and offline — ●LEA KUTTLER, MICHAEL BÖHLER, and MARKUS SCHUMACHER — Institute of Physics, Albert-Ludwigs-University Freiburg, Freiburg, Germany

HammerCloud (HC) is a framework for automated testing designed to monitor the resources of the Worldwide LHC Computing Grid (WLCG). It uses test jobs that mimic realistic physics analysis or production jobs to identify potential issues within the infrastructure of WLCG computing sites.

HC's web interface provides site administrators with detailed insights and visual summaries of test results. After almost two decades of successful operation, these visualizations, which previously relied on Google Image Charts, were upgraded to interactive Highcharts, offering a more dynamic, and user-friendly experience.

Complementing the detailed insights available on the HC website, the Tier-2-Report provides a broader overview by summarizing important test metrics over weekly or monthly periods to analyze overall performance trends. The generation of these Tier-2-Reports is fully automated, using a tool initially developed to present and discuss HC statistics for the GridKa cloud. A recent update has made this tool applicable for generating similar reports for any other cloud. Additionally, this update introduced a redesigned report layout and enhanced representations of test metrics.

This contribution presents the updated methods for visualizing and discussing test results within the HC framework.

T 12.2 Mon 17:00 VG 2.102

A new document server and publication process tool for Belle II — ●DAVID KOCH and THOMAS KUHR — Ludwig-Maximilians-Universität

An integral part of working in science and in large collaborations in particular is the documentation of work in progress and results, be it in the form of slides, internal notes and reports or papers targeted for publication. Especially the latter however involves much more than just a single document. The process starts from an analysis and includes writing multiple versions of a draft that has to go through many stages and internal approvals until it can be submitted to a journal. The Belle II experiment recently launched its own in-house developed document server, PubDB, that is also a tool to follow the entire process of bringing an analysis to publication. In this talk we show how Belle II uses PubDB to implement its publication procedure policy in a uniform, streamlined and traceable manner. We share ideas and lessons learned that are valuable to the community as a whole.

T 12.3 Mon 17:15 VG 2.102

Orchestrated columnar-based analysis with columnflow — ●MATHIS FRAHM, JOHANNES HALLER, PHILIP KEICHER, NATHAN PROVOUST, MARCEL RIEGER, DANIEL SAVOIU, PETER SCHLEPER, and MATTHIAS SCHRÖDER — Institut für Experimentalphysik, Universität Hamburg

The large datasets and increasing complexity of modern physics analysis in high energy collider physics pose a major challenge to the analysis workflows. Systems are required that can efficiently process large amounts of data, while keeping the execution of the complete analysis manageable. In this talk, we present Columnflow, a tool for columnar-based data analysis. Columnflow provides an orchestrated, yet flexible workflow that automatically handles the bookkeeping of results and dependencies. Typical analysis tasks such as propagation of systematic uncertainties, machine learning applications, and statistical inference are transparently integrated into the workflow. The implemented workflow allows the use of distributed computing resources and is fully configurable, yet accessible to newcomers.

T 12.4 Mon 17:30 VG 2.102

Pre-cache tests with the WLCG Tier-2 centre GoeGrid

and the NHR HPC cluster Emmy using workflows of the ATLAS collaboration at the LHC — ●INGA ŁAKOMIEC, SAIDEV POLISETTY, ARNULF QUADT, and SEBASTIAN WOZNIEWSKI — II Physikalisches Institut, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen

The GoeGrid centre in Goettingen is one of the WLCG Tier-2 sites and contributes to the ATLAS job processing and data storage. The HPC cluster Emmy by the National High Performance Computing (NHR) has been successfully connected with GoeGrid and ATLAS jobs can be run on its resources. However, there is no large local mass storage at Emmy for the WLCG operations. Therefore, data for jobs that are processed at Emmy is currently provided by the GoeGrid storage. A transition of storage and computing resources to the Helmholtz Centres and NHR sites respectively from the university based Tier-2 centres is planned in Germany in the next years. Since some NHR clusters will serve as a big computing centres without a large local mass storage, there is a need to prepare proper caching solutions and validate them.

Small local storage can be available at Emmy for the WLCG tasks after the storage centralisation. The current ATLAS workflow management has been tested for Emmy computing resources together with the small local storage (pre-cache) instance at GoeGrid. Results will be presented for the different number of CPU cores used by heavier workloads exclusively or a mix of production jobs. Then, the transfer and deletion of data in terms of a small disc size will be shown.

T 12.5 Mon 17:45 VG 2.102

Research Data Management at HZDR with HELIPORT — ●STEFAN E. MÜLLER¹, THOMAS GRUBER¹, OLIVER KNODEL¹, MANI LOKAMANI¹, DAVID PAPE¹, MARTIN VOIGT^{1,2}, and GUIDO JUCKELAND¹ — ¹Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany — ²Technische Universität Dresden, Dresden, Germany

The researchers at the Helmholtz-Zentrum Dresden-Rossendorf rely on a large variety of tools and systems when it comes to administer research data. The project planning phase (proposal submission to a beamtime proposal management system, creation of data management plans and data policies), the documentation during experiments or simulation campaigns (electronic laboratory notebooks, wiki pages), backup- and archival systems as well as the final journal and data publications (using collaborative authoring tools, meta-data catalogs, software and data repositories, publication systems) are all processes which involve research data management. Also, modern research projects often require to interact with a variety of software stacks and workflow management systems to allow reproducibility on the underlying IT infrastructure. The "HELMholtz Scientific Project Workflow PlaTform" (HELIPORT), which is currently developed by researchers at HZDR and their collaborators, facilitates the management of research data and metadata by providing an overarching guidance system which combines all the information by interfacing the underlying processes. It also includes a workflow engine which can be used to automate processes like automated data publication or data analysis.

T 12.6 Mon 18:00 VG 2.102

Status and Plans for the CMS Grid at Aachen — MANUEL GIFFELS¹, ●ALEXANDER JUNG², THOMAS KRESS³, MARTIN LIPINSKI⁴, ANDREAS NOWACK³, VALENTINA SARKISOVI², ALEXANDER SCHMIDT², and SHAWN ZALESKI² — ¹Institut für Experimentelle Teilchenphysik, KIT — ²III. Physikalisches Institut A, RWTH Aachen — ³III. Physikalisches Institut B, RWTH Aachen — ⁴I. Physikalisches Institut B, RWTH Aachen

From 2025 onwards, the German CMS Grid Tier-2 model will evolve to incorporate external storage from Helmholtz centers and CPU resources provided by the NHR consortium.

In this presentation, we will provide an overview of the current status and outline our plans for utilizing NHR resources for this purpose, with a focus on the Aachen CMS Grid.