T 12: Computing

Zeit: Montag 11:00–12:30 Raum: P15

T 12.1 Mo 11:00 P15

High Performance Computing usage for ATLAS — •Rodney Walker — Ludwig-Maximilians-Universitaet, Muenchen

Current ATLAS computing is largely provided by a distributed federation of High Throughput Computing(HTC) clusters, based at universities and labs around the world. The model was well grounded and critical to the success of LHC Run 1 data analysis, but budgetary constraints mean that it is unlikely to be sufficient for Run 2 and beyond.

In the meantime, High Performance Computing(HPC) has evolved to become dominated by large x86 linux clusters which, despite the specialized hardware, are comparable in cost to the smaller university-based HTC linux clusters. Indeed, economies of scale, together with novel energy saving features, can make the cost of ownership significantly cheaper. Such HPC clusters are well suited for standard ATLAS production work, although there are a number of hurdles to overcome - technical solutions and working examples will be described.

We also discuss specific ATLAS workflows which can take advantage of the HPC capabilities, such as fast interconnect and high performance files systems. Allocated cpu-time on such general scientific resources, together with the large potential for opportunistic use, by means of backfill, should provide a valuable contribution to ATLAS's future computing needs.

T 12.2 Mo 11:15 P15

An Analysis Framework for the modern HEP — •Raphael Friese, Thomas Müller, Roger Wolf, Joram Berger, Dominik Haitz, Fred Stober, Thomas Hauth, and Günter Quast — Institut für experimentelle Kernphysik, KIT

The analysis of data in high energy physics experiments is characterized by a number of iterations, during which the knowledge and assessment of the physics objects, under study, increase incrementally. Especially in a distributed environment, as for the LHC experiments this can be challenging. Analyses can significantly benefit from a carefully designed analysis software framework. The talk outlines the concepts of such a software framework, used on top of the analysis and reconstruction software CMSSW, of the CMS experiment. This framework is used by the KIT physics analysis groups. Data formats, the envisaged workflows and the advantages in terms of flexibility, reliability and turnaround time will be discussed. The framework is foreseen to serve a large variety of analyses in the areas of jet calibration, QCD multijets measurements, Higgs measurements and SUSY/BSM searches.

T 12.3 Mo 11:30 P15

Hadoop for Parallel Root Data Analysis — •Sebastian Lehrack and Guenter Duckeck — LMU Muenchen

The Apache Hadoop software is a Java based framework for distributed processing of large data sets across clusters of computers using the Hadoop file system (HDFS) for data storage and backup and MapReduce as a processing platform. Hadoop is primarily designed for processing large textual data sets which can be processed in arbitrary chunks, and must be adapted to the use case of processing binary data files which can not be split automatically. However, Hadoop offers attractive features in terms of fault tolerance, task supervision and controlling, multi-user functionality and job management. For this reason, we have evaluated Apache Hadoop as an alternative approach to PROOF for ROOT data analysis. Two alternatives in distributing analysis data are discussed: Either the data is stored in HDFS and processed with MapReduce, or the data is accessed via a standard Grid storage system (dCache Tier-2) and MapReduce was used only as execution back-end. The focus in the measurements are on the one hand to safely store analysis data on HDFS with reasonable data rates and on the other hand to process data fast and reliably with MapReduce. For evaluation of MapReduce, realistic ROOT analyses have been used and event rates were compared to PROOF. We also investigated the data locality on our workstation cluster.

T 12.4 Mo 11:45 P15

VISPA: A new Approach for Porting Data Analysis Workflows to the Web — •Marcel Rieger, Martin Erdmann, Robert Fischer, Christian Glaser, Gero Müller, Thorben Quast, Martin Urban, Daniel van Asseldonk, and Tobias Winchen — III. Physikalisches Institut A, RWTH Aachen

We present the most recent developments of the web browser based Visual Physics Analysis (VISPA) project, which has been successfully used in various high-energy and astroparticle physics data analyses (http://vispa.physik.rwth-aachen.de). The VISPA platform delivers a modern user interface and a framework for creating extensions written in Python and Javascript to provide dedicated functionality for individual needs. Common examples for extensions are a file browser, a code editor, or a terminal. Another key feature is the use of workspaces, i.e., machines connected via SSH providing computing power and data access using remote procedure calls. Every user-accessible machine running Python can act as a workspace, regardless of whether it is a mobile phone, a desktop machine, or even a computing cluster. The benefits of this approach can be applied to data analyses in scientific workflows. With the combination of workspaces, e.g. computing resources, and individual extensions, e.g. experiment software, VISPA represents a new approach for porting data analysis workflows to the web.

T 12.5 Mo 12:00 P15

GENFIT - a Generic Track-Fitting Toolkit — •Johannes Rauch[1] and Tobias Schlüter[2] — [1]Technische Universität München — [2]Ludwig-Maximilians-Universität München

GENFIT is an experiment-independent track-fitting toolkit, which combines fitting algorithms, track representations, and measurement geometries into a modular framework. We report on a significantly improved version of GENFIT, based on experience gained in the Belle II, PANDA, and FOPI experiments. Improvements concern the implementation of additional track-fitting algorithms, enhanced implementations of Kalman fitters, enhanced visualization capabilities, and additional implementations of measurement types suited for various kinds of tracking detectors. The data model has been revised, allowing for efficient track merging, smoothing, residual calculation and alignment.

T 12.6 Mo 12:15 P15

Next-generation Software Framework of the NA61/SHINE Experiment at CERN — •Marek Szuba[1], Andras Laszlo[2], Antoni Marcinek[3], Tom Paul[4], Roland Sipos[2], Michael Unger[1], Darko Veberic[4], and Oskar Wyszynski[3] for the NA61/SHINE-Collaboration — [1]Karlsruhe Institute of Technology, Karlsruhe, Germany — [2]Wigner Research Center for Physics, Budapest, Hungary — [3]Jagiellonian University, Cracow, Poland — [4]University of Nova Gorica, Nova Gorica, Slovenia

NA61/SHINE is an experiment at the CERN Super Proton Synchrotron, studying hadron production in hadron-hadron, hadron-nucleus and nucleus-nucleus collisions to provide valuable contributions to a number of subjects, from neutrino through cosmic-ray to heavy-ion physics. Inaugurated in 2011, its software-upgrade project "Shine Offline" has aimed at providing a modern, extensible and more maintainable replacement for the legacy reconstruction, simulation, calibration and analysis software inherited by NA61/SHINE from its predecessor NA49 while at the same time providing continued support for legacy data format and mission-critical software components. This contribution presents an overview of design considerations, fundamental properties and architecture of Shine Offline, status of its components and results of comparative tests between the new framework and its legacy counterparts. A complementary topic of long-term preservation of NA61/SHINE software and data through the use of cloud computing shall be discussed as well.