End-user data analysis at the LHC*

Nils Faltermann Manuel Giffels Günter Quast

Karlsruhe Institute of Technology (KIT)

Abstract

The Large Hadron Collider (LHC), located at CERN in Geneva, stands as one of the most monumental scientific experiments in human history. This remarkable machine facilitates approximately 40 million particle collisions every second, generating an astronomical amount of data. Even with rigorous filtering of collision events, the data retained for subsequent analysis remains staggering in scale. In addition to the recorded data, conducting a successful physics analysis demands an extensive set of simulations that can be compared to the recorded events. In this presentation, we will delve into our approach to incorporating external resources, such as the NEMO cluster in Freiburg, into our local batch system. This integration greatly enhances accessibility for the complex workflows required for physics data analyses.

1 Introduction

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It is located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. The LHC is a multinational scientific endeavor involving thousands of physicists and engineers from over 100 countries around the world. The LHC accelerates beams of protons or heavy ions to extremely high energies. These particles are guided around the ring by powerful magnetic fields generated by the superconducting magnets. The beams travel in opposite directions and are made to collide at four main detector sites (ATLAS, CMS, ALICE, and LHCb), where the outcome of the particle collisions is recorded. The primary purpose of the LHC is to explore the fundamental properties of particles and the forces that govern them. By colliding high-energy particles together at nearly the speed of light, scientists aim to recreate conditions similar to those just nanoseconds after the Big Bang, providing insights into the origins of the Universe.

2 Experimental Setup

Particle detectors are designed to measure various properties of particles produced in high-energy collisions. These detectors are complex instruments consisting of multiple subsystems designed to track the paths of charged particles, measure their energies, and identify different types of particles. They collect vast amounts of data from collision events, which include information such as particle trajectories, energy deposits, and timing signals. The data collected by the detectors provide insights into the behavior of fundamental particles and the interactions between them.

The CMS (Compact Muon Solenoid) detector is one of the two large general-purpose detectors located at the LHC. It is a cylindrical apparatus with a mass of 14,000 tonnes, about 15 meters in diameter and 21 meters in length, designed to observe a wide range of particles and phenomena produced by high-energy proton-proton collisions. It consists of multiple layers of sub-detectors arranged concentrically around the collision point.

The LHC operates proton-proton collisions at very high rates. The nominal collision rate at the LHC is around 40 million proton-proton collisions per second per experiment. With around 55 million readout channels of the CMS detector, this results in approximately a data rate of $1\,\mathrm{PB/s}$, too much for any storage system to cope with. For this reason, the first step of the data acquisition process is the trigger system, which is designed

*Proceedings of the 9th bwHPC-Symposium 2023. Mannheim Conference Series (MaConf). 2025.

to select collision events of interest for further analysis and drop the rest. The CMS trigger system consists of two levels: the Level-1 trigger (L1) and the High-Level Trigger (HLT). The L1 trigger is implemented in hardware and makes rapid decisions based on information from the calorimeters and muon detectors. It reduces the event rate from the LHC's collision rate of 40 million events per second to around 100,000 events per second. The events passing the L1 trigger are further processed by the HLT, which is implemented in software running on a large farm of computers. The HLT performs more sophisticated event reconstruction and selects events for storage based on specific physics criteria. It further reduces the event rate to a manageable level for offline analysis, around 2,000 events per second. Despite the significant reduction in event rates, the total amount of data in a year that needs to be recorded and processed amounts to at least $20\,\mathrm{PB}$ for each of the detectors at the LHC.

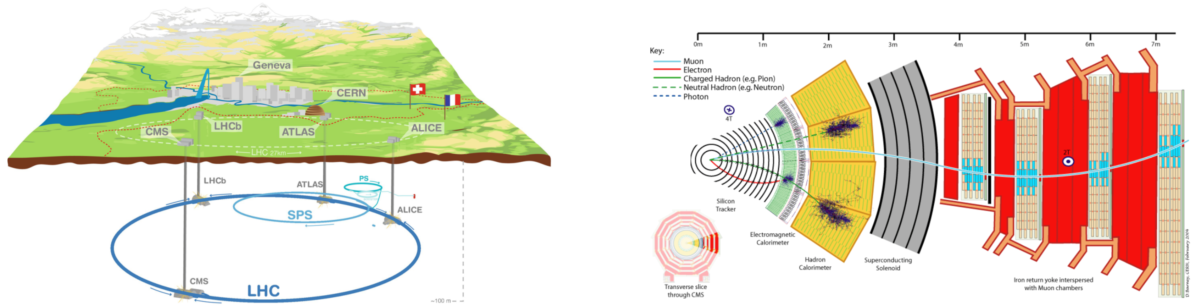

Figure 1: Illustration of the LHC tunnel and its surrounding area with the location of the main particle detectors (left) [1] and a vertical slice of the CMS detector (right) [2], demonstrating the potential path of particles produced in the proton-proton collisions and their signatures.

3 Computing Model

The Worldwide LHC Computing Grid (WLCG) is a global computing infrastructure established to handle the massive amounts of data produced by the different LHC experiments. The WLCG represents a collaboration among thousands of scientists, engineers, and computing experts from research institutions and universities around the world. It involves more than 170 computing centers distributed across more than 40 countries, forming a global network of computing resources for storing, processing, and analyzing data.

The WLCG is organized in a hierarchical, tiered structure: The Tier-0 center is located at CERN itself. It serves as the initial data collection point where raw data from the LHC experiments is stored and distributed to Tier-1 centers for further processing and analysis. Tier-1 centers are major computing facilities located around the world, such as GridKa in Karlsruhe. They are responsible for storing and processing a significant portion of the LHC data, providing computing resources, storage capacity, and networking infrastructure to support the needs of the global particle physics community. Tier-2 centers are regional computing facilities located at universities and research institutions. They serve as hubs for data analysis and simulation tasks, providing computing resources and storage capacity to physicists and researchers within their respective regions. The final physics data analysis is usually carried out at so-called unofficial Tier-3 centers, which consist of local institute clusters.

While there is a constant workload in terms of processing and simulation jobs on higher-level tiers due to centrally organized workflow management systems, the situation at Tier-3 clusters is usually different: The end-user working on physics analyses has different and diverse requirements:

- Processing of data and simulation, deriving new quantities and condensing the information into histograms. Medium to high CPU usage, but usually very high I/O requirements. Typical use case for data-intensive computing.
- Tasks that require processing data multiple times, for example statistical inference and training of machine learning methods. Very high CPU usage, but moderate I/O requirements. Referred to as

CPU-intensive computing.

- The demand for resources is not constant over time. Because of the limited amount of users and workflow of analyses, i.e., first developing and implementing algorithms that are then applied to data, resources are usually requested in bursts. This results in a situation where either not enough resources are available for the demand or resources are idling.

The infrastructure hosted at institute clusters is typically built homogeneously, similar to a higher-tier site in the WLCG, with a tendency towards CPU-intensive computing to support also many other applications outside of high-energy physics. On top of that it cannot be justified to spend the already sparse funding in research for the purchase of additional resources that would not be utilized to the full extent. Given the user requirements and the reality at local institutes, this situation is a prime example of a use case where so-called opportunistic resources can provide the solution. Opportunistic resources are resources that are not permanently dedicated to but temporarily available for a specific task, user or group. With the integration of opportunistic resources, the resource pool can also be made more heterogeneous, thus providing better-suited resources for individual demand, e.g., external GPUs for specialized machine learning applications. As the name already implies, opportunistic resources are also only allocated when there is demand for them, which for instance can also be realized by booking resources from cloud providers.

4 Dynamic Resource Integration

To run user jobs on WLCG resources, placeholder jobs, called pilots, are employed, which interact with an overlay batch system (OBS) to dynamically register available resources to a resource pool. The user only interacts with this OBS and is therefore transparent to the underlying site. This concept works fine in the homogeneous environment of the WLCG sites, where the hardware and software are tailored to the needs of the community. Naturally, the situation for external resources is different. For this reason, the pilot concept is extended to the so-called drone. Depending on the resource, the drone can run natively on the resource as a batch job, or can provide an additional layer of abstraction in terms of a virtual machine or a container to provide the required software environment. The demand and request for additional resources at an external site is managed by COBalD - the Opportunistic Balancing Daemon (COBalD) [3, 4] by aggregating similar drones into an abstract pool. The decision to increase or decrease the resource demand in a pool involves also the utilization of the current resources in this pool. Drones themselves are managed via the Transparent Adaptive Resource Dynamic Integration System (TARDIS) [5, 6], which directly interacts with COBalD. Each site can run an individual TARDIS instance to facilitate specialized configurations and react independently on site-specific circumstances. The TARDIS workflow is illustrated in Fig. 2.

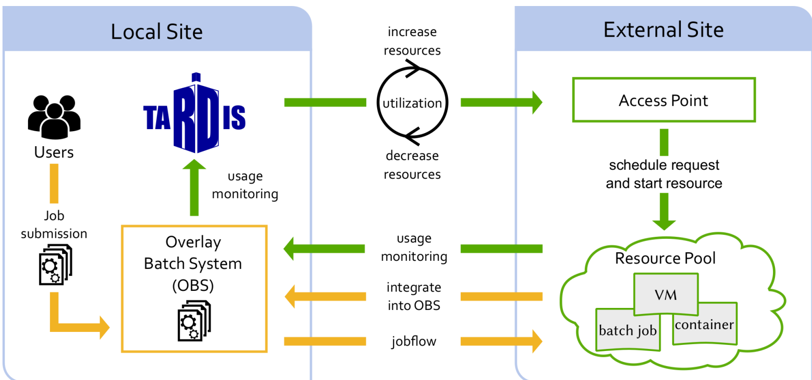

Figure 2: Illustration of *The Transparent Adaptive Resource Dynamic Integration System* (TARDIS) and its working principle to incorporate opportunistic resources.

TARDIS is successfully deployed to the OBS instance (HTCondor) at the Institute of Experimental Particle Physics (ETP), which is part of the Karlsruhe Institute of Technology (KIT). Resources that have been integrated so far consist of research clusters such as the bwForCluster (NEMO), TOpAS (*Throughput Optimized Analysis System*, GridKa), ForHLRII (KIT), HoreKa (KIT), as well as well as commercial cloud providers such as Open Telekom Cloud, Exoscale and 1&1. In Fig. 3 the batch system utilization at the ETP is shown over the time of a week. Local resources were always used in this case but only made up a small fraction of all used resources. During this particular time, many jobs were submitted to the batch system and thus opportunistic resources were allocated to serve this demand, as can be seen by the rising amount of opportunistic resources whenever new jobs were queued in the batch system. Notably at some point during the week, the utilization exceeded the total share of the CMS experiment at the Tier-1 site GridKa.

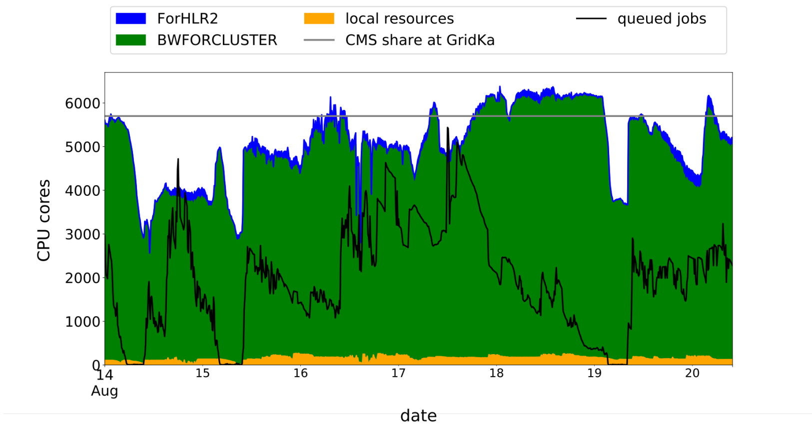

Figure 3: Example usage of the local batch system at the Institute of Experimental Particle Physics (ETP). Shown are the static local resources in yellow and the dynamically integrated opportunistic resources in blue (ForHLR II) and green (bwForCluster NEMO).

5 Scientific Results

During recent years the integration of opportunistic resources enhanced the computing capabilities of the ETP institute cluster significantly. In total, there are more than 20 journal publications up to this day that directly benefited from the additional computing power. The most recent ones involving the CMS experiment are outlined shortly in the following.

In light of the currently ongoing Run 3 of the LHC at $\sqrt{s} = 13.6\,\mathrm{TeV}$ it is important to first look into well-known physics processes to understand and calibrate the detector, particularly after a longer shutdown and maintenance period. This also needs to happen quickly to identify potential problems with the data-taking to resolve them as soon as possible. One of these so-called standard candles is the production of a Z boson that decays into a pair of muons. This process has been measured successfully with high precision [7], paving the way for future Run 3 precision measurements. The result of the measurement is also highlighted in Fig. 4 (left). Especially in the context of an early measurement, where frequent reprocessing of data happens, the addition of opportunistic resources helps for a fast turnaround.

In contrast to the previous example, the measurement of ttH and tH production with H $\rightarrow$ bb decays [8] is targeting an already completed data-taking campaign to probe the coupling of the Higgs boson to the top quark. With the complex final state consisting of many jets, fully reconstructing events becomes a combinatorial task, for which dedicated machine learning algorithms are employed. To extract the results elaborate and computing-intensive statistical calculations are necessary, e.g., maximum-likelihood estimations and multidimensional likelihood scans for various parameters of interest. One of these results is exemplary shown in Fig. 4 (right), where the coupling of the Higgs boson to the top quark and vector bosons is probed simultaneously to derive exclusion limits.

The search for additional Higgs bosons and leptoquarks in $\tau\tau$ final states [9] directly searches for possible extensions to the standard model of physics. To model the already-known background process of a Z boson decaying into a pair of τ leptons, the analysis utilizes a refined method called τ-embedding. This method is a data-simulation hybrid approach, where data events with a Z boson decaying into a pair of muons are cleaned of the muon signature, which is then replaced with simulated τ decays and their interaction with the detector material. The latter part is a highly computationally intensive task and the bottleneck for simulated events in general, but the use of opportunistic resources significantly speeds up the task.

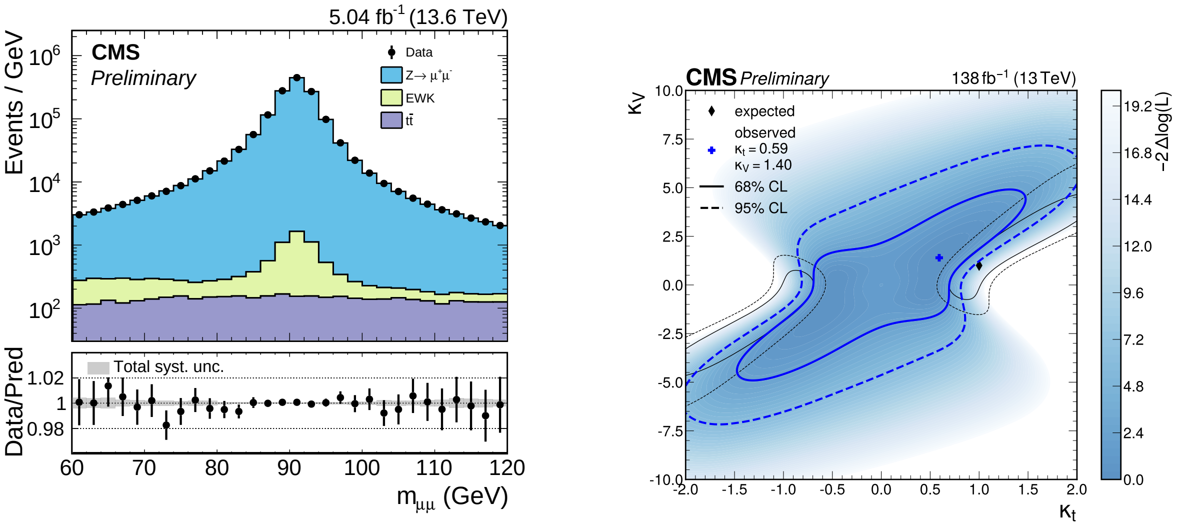

Figure 4: Exemplary scientific results that have been obtained with the help of opportunistic resources. Shown is the invariant mass distribution of the dimuon system in the measurement of the Z boson cross section at $\sqrt{s} = 13.6\,\mathrm{TeV}$ [7] (left) and the two-dimensional likelihood scan of the coupling modifiers for the top quark κ_t and for vector bosons κ_V in the combined measurement of ttH and tH production at $\sqrt{s} = 13\,\mathrm{TeV}$ [8] (right).

6 Conclusions

The enormous data recorded at the Large Hadron Collider (LHC) is a computational challenge on many fronts. While data processing and simulation are mostly automated on dedicated sites, the physics analysis as an end-user is mostly done on conventional computing clusters. To provide additional resources that are transparent to use the COBalD/TARDIS system was developed and successfully deployed to our local batch system at the Institute of Experimental Particle Physics (ETP). The demand for computing resources at the LHC will increase greatly during the next decade when the LHC will be upgraded to the High-Luminosity LHC, where each collision event will contain significantly more particles to analyze. The integration of opportunistic resources can be one of the cornerstones to overcome this upcoming obstacle.

7 Acknowledgements

The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG (bwForCluster NEMO).

References

[1] Overall view of the LHC, Service graphique, CERN, OPEN-PHO-CHART-2014-006

[2] Interactive Slice of the CMS detector, Siona Ruth Davis, CMS Collaboration, CMS-OUTREACH-2016-027

[3] Lightweight dynamic integration of opportunistic resources, M. Fischer et al., EPJ Web Conf. 245 (2020) 07040

[4] MatterMiners/cobald: v0.14.0, M. Fischer et al., Zenodo, doi: 10.5281/zenodo.8199049

[5] Effective Dynamic Integration and Utilization of Heterogenous Compute Resources, M. Fischer et al., EPJ Web Conf. 245 (2020) 07038

[6] MatterMiners/tardis: 0.8.1, M. Giffels et al., Zenodo, doi: 10.5281/zenodo.10411024

[7] Measurement of the inclusive cross section of Z boson production in pp collisions at $\sqrt{s} = 13.6\,\mathrm{TeV}$, CMS Collaboration, CMS-PAS-SMP-22-017, to be submitted to PLB

[8] Measurement of the ttH and tH production rates in the H $\rightarrow$ bb decay channel with $138\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$, CMS Collaboration, CMS-PAS-HIG-19-011, to be submitted to JHEP

[9] Searches for additional Higgs bosons and for vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$, CMS Collaboration, JHEP 07 (2023) 073