WLCG Environmental Sustainability Workshop

Report of Contributions

https://indico.cern.ch/e/wlcg-sustainability-2024

Contribution ID: **3** Type: **not specified**

Approach to Sustainability at the CERN T0 and CERN DC

Wednesday 11 December 2024 15:55 (15 minutes)

CERN has been tracking its energy consumption and carbon footprint at the level of the organisation and department for several years. We are compliant with the ISO 50001 energy management certification. However, this gives only a rough overview of the impact. CERN IT has been working on improving the monitoring of efficiency and power consumption as well as understanding the total impact of their activities. In this contribution we will give an overview of CERN-ITs activities and future strategy and direction.

Primary author: SCHULZ, Markus (CERN)

Presenter: SCHULZ, Markus (CERN)

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **4** Type: **not specified**

Life Cycle Analysis of Compute Nodes

Thursday 12 December 2024 16:20 (15 minutes)

A model for computing the total life cycle emissions for compute nodes, based on providing a certain level of scientific computing over long time.

This takes real-world data from data centers, electricity generation, and vendor server information and simulates different replacement scenarios (from replacing it as soon as possible to keeping it a long time) to be able to optimize for minimal total emissions.

This work has largely been presented at ISGC and HEPiX, but since then we have a publicly available simulation tool as well as some minor improvements to graphs and conclusions.

The model is available at: https://codeberg.org/wimvanderbauwhede/low-carbon-computing/src/branch/master/LCA-model-equations

Primary author: WADENSTEIN, Mattias (University of Umeå (SE))

Presenter: WADENSTEIN, Mattias (University of Umeå (SE))

Session Classification: Hardware and Fabrics

Track Classification: All contributions

Contribution ID: **6** Type: **not specified**

ADVANCING DATA CENTER SUSTAINABILITY: Empirical Measurements of AI Training Power Demand on a GPU-Accelerated Node

Data center sustainability, a phenomenon that has grown in focus due to the continuing evolution of Artificial intelligence (AI)/High Performance Computing (HPC) and High Throughput Computing (HTC) systems; furthermore, the rampant increase in carbon emissions resulted due to unprecedented rise in Thermal Design Power (TDP) of the computer chips. With the exponential increase of demand towards the usage of such systems, major challenges have surfaced in terms of productivity, Power Usage Effectiveness (PUE), and thermal/scheduling management. Deploying AI/HPC infrastructure in data centers will require substantial capital investment.

This study at the ATLAS Tier-1 site, Scientific Data and Computing Center (SDCC), Brookhaven National Laboratory (BNL) quantified the energy footprint of this infrastructure by developing models based on the power demands of AI hardware during training. We measured the instantaneous power draw of an 8-GPU NVIDIA H100 HGX node while training open-source models, including the ResNet image classifier and the Llama2-13b large language model. The peak power draw observed was about 18% below the manufacturer's rated TDP, even with GPUs near full utilization. For the image classifier, increasing the batch size from 512 to 4096 images reduced total training energy consumption by a factor of four when model architecture remained constant. These insights can aid the scientific data center facilities such as CERN identify the 'stranded power' within existing facilities and assist in capacity planning and provide researchers with energy use estimates. Future studies will explore the effects of liquid cooling technologies and carbon-aware scheduling on AI workload energy consumption. These results can help ATLAS in the development of the ATLAS software or operational model which may significantly reduce the carbon footprint at their data centers and identify opportunities for heat reuse.

Primary author: LATIF, Imran (Brookhaven National Laboratory)

Presenter: LATIF, Imran (Brookhaven National Laboratory)

Session Classification: Hardware and Fabrics

Track Classification: All contributions

Contribution ID: **7** Type: **not specified**

The Role of the HEP Benchmark Suite in Measuring Performance to Support Sustainability in High-Energy Physics Computing

Thursday 12 December 2024 11:50 (15 minutes)

The HEP Benchmark Suite provides a comprehensive framework for evaluating the performance and power consumption of computing resources within the Worldwide LHC Computing Grid (WLCG), supporting its sustainability objectives. This presentation will outline the toolkit's capabilities and present examples where the suite has successfully identified opportunities to improve both energy efficiency and overall computational efficiency within the WLCG. Finally, we will discuss potential future enhancements to expand the suite's capacity to further support WLCG's sustainability initiatives.

Primary authors: GIORDANO, Domenico (CERN); SZCZEPANEK, Natalia Diana (CERN)

Presenter: GIORDANO, Domenico (CERN)

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **8** Type: **not specified**

Aware Computing in ATLAS: Estimating Carbon Footprint for Workloads with PanDA

Thursday 12 December 2024 11:10 (15 minutes)

The ATLAS experiment relies on the PanDA workload management system to handle its analysis and production tasks across the Worldwide LHC Computing Grid (WLCG). In response to a request from the ATLAS Sustainability Forum, we've introduced a new feature: simple, informative estimates of the carbon emissions generated by each job.

To achieve this, we calculate CO_2 emissions by retrieving time-dependent, regional electricity grid emission intensities and using estimated core power consumption values. Emission totals are calculated in grams of CO_2, differentiated by region or a global average, and presented to users upon task completion. These estimates are also available in monitoring tools, distinguishing emissions by successful and failed jobs.

This feature is not meant to deter essential work but to raise awareness of the environmental impact of computational tasks, encouraging users to consider optimizing their workloads for a greener future.

Primary authors: ALEKSEEV, Aleksandr (The University of Texas at Arlington (UTA)); BRUERS, Ben (Deutsches Elektronen-Synchrotron (DE)); LIN, Fa-Hui (University of Texas at Arlington (US)); BARREIRO MEGINO, Fernando Harald (University of Texas at Arlington); BORODIN, Misha (University of Iowa (US)); WALKER, Rodney (Ludwig Maximilians Universitat (DE)); MAENO, Tadashi (Brookhaven National Laboratory (US)); KORCHUGANOVA, Tatiana (University of Pittsburgh (US))

Presenter: BARREIRO MEGINO, Fernando Harald (University of Texas at Arlington)

Session Classification: Environmental Sustainability: The Experiments

Track Classification: All contributions

Contribution ID: **10** Type: **not specified**

Sustainable computing with RF2.0 at DESY

Wednesday 11 December 2024 15:05 (15 minutes)

Sustainability is becoming an ever-increasing part in the planning, designing and operation of large-scale infrastructures. Due to the international nature of the collaborators, these large-scale infrastructures end up being both joint-ventures building large machines, and the pooling of smaller resources to provide services for these machines. Research Facility 2.0 is an EU-funded project with a vision of developing a more sustainable path for the future of research. Its remit covers both of these areas by researching: the design and use of technologies for use at future accelerators; and the approaches we can take to manage energy at support infrastructures. Both have unique challenges in terms of environmental impact, but with the latter being of more relevance to the WLCG. DESY is the only institution on this project looking to develop strategies for the energy management of "green"data-centres. This talk shall show the results of some initial investigations in implementing energy management strategies at the DESY computing centre.

Primary author: Dr SPITERI, Dwayne (DESY)

Co-author: Mr GASTHUBER, Martin (DESY)

Presenter: Dr SPITERI, Dwayne (DESY)

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **13** Type: **not specified**

Carbon costs of storage & carbon reporting: a UK perspective.

Thursday 12 December 2024 14:50 (20 minutes)

In order to achieve the higher performance year on year required by the 2030s for future LHC upgrades at a sustainable carbon cost to the environment, it is essential to start with accurate measurements of the state of play. Whilst there have been a number of studies of the carbon cost of compute for WLCG workloads published, rather less has been said on the topic of storage, both nearline and archival.
We present a study of the embedded and ongoing carbon costs of storage in multiple configurations, from Tape farms through to SSDs, within the UK Tier-1 and Tier-2s and discuss how this directs future policy.

Primary authors: Dr DEWHURST, Alastair (STFC); PACKER, Alison (STFC); HAYS, Jonathan (University of London (GB)); OWEN, Richard Alexander; SKIPSEY, Samuel Cadellin

Presenters: OWEN, Richard Alexander; SKIPSEY, Samuel Cadellin

Session Classification: Hardware and Fabrics

Track Classification: All contributions

Contribution ID: **14** Type: **not specified**

Raising awareness on IT's Environmental Impact: Experiences from a Collaborative Workshop

Wednesday 11 December 2024 17:40 (10 minutes)

Collaborative approaches to raise awareness about climate change or ICT impact in the environment have become very popular in the past years. Examples like Climate Fresk or Digital Collage, where collective intelligence is used to understand and identify mitigation strategies, have found a tremendous success among general public and are happening worldwide. Getting inspiration from these initiatives, a collaborative workshop was designed with a technical focus on IT services relevant for the IT Compute & Devices group at CERN. The goal of the workshop was to organise a team building event, where participants could share expertise and knowledge, understand the impact of IT technologies in the environment and why it is important to mitigate them. The aim of the presentation is to share our experience designing and running the workshop, including lessons learned and next steps.

Primary authors: ALANDES PRADILLO, Maria (CERN); BHAT, Varsha

Presenters: ALANDES PRADILLO, Maria (CERN); BHAT, Varsha

Session Classification: Projects and Reports

Track Classification: All contributions

Contribution ID: **15** Type: **not specified**

The 3 most important factors in datacenter planning: Location, location, location

Wednesday 11 December 2024 13:45 (15 minutes)

The single largest factor in determining the CO2 footprint of a compute resource is the CO2 intensity of the electricity used to power it. The carbon intensity ranges from 20 to 800gCO2/kWh, depending on the energy mix of the region. Due to the interconnectivity of the European electricity grid, it is not sufficient to look only at the regional electricity carbon intensity, but also at that of interconnected regions that could use that same energy. In a very real sense, a MW lower consumption in France can replace a MW of coal generated electricity in Germany.

On the other hand, there is very limited interconnectivity between the low carbon intensity Nordic grids and central Europe. An unused Nordic kWh cannot be transferred to reduce German generation from coal. It follows that locating our compute resource in Sweden, rather than central Europe, reduces its carbon footprint from around 1000gCO2/kWh to 20gCO2/kWh. One can question the assumptions in reaching the factor 50 improvement, but this will anyway make any O(10) percent improved efficiency of server, cpu, cooling, software, etc rather negligible.

We describe the power grids, interconnects and bottlenecks which lead to the objective conclusion that we should locate our computing where low carbon power is abundant and cheap but not transmittable. We also explore the artificial constraints which make minimizing CO2 emissions in this way a frustratingly futile endeavour.

Primary authors: WADENSTEIN, Mattias (University of Umeå (SE)); WALKER, Rodney (Ludwig Maximilians Universitat (DE))

Presenter: WALKER, Rodney (Ludwig Maximilians Universitat (DE))

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **16** Type: **not specified**

Natural job drainage and power reduction in PIC Tier-1 using HTCondor

Wednesday 11 December 2024 16:15 (15 minutes)

This study presents preliminary analyses of natural job drainage and power reduction patterns in the PIC Tier-1 data center, which uses HTCondor for workload scheduling. By examining historical HTCondor logs, we simulate natural job drainage behaviors, in order to understand natural job drainage patterns: when jobs naturally conclude without external intervention. These findings provide insights into the center's capability to modulate resource usage according to external factors like green energy availability cycles.

To further validate and extend these observations, simulations were conducted under various load conditions to evaluate the influence of job types and VO-specific durations on drainage cycles. An analysis of power consumption pre- and post-drainage, facilitated by ipmitool, allows for estimating potential power and carbon emission reductions in drainage scenarios. Building on these insights, machine learning models are being developed to predict optimal power scaling adjustments.

We propose a conceptual feedback loop to HTCondor that could enable real-time power adjustments based on fluctuations in green energy availability. By exploring these ideas, this research aims to contribute to a more sustainable data center model, offering a framework for adapting workload management to dynamic environmental factors.

Primary author: FLIX MOLINA, Jose (CIEMAT - Centro de Investigaciones Energéticas Medioambientales y Tec. (ES))

Co-authors: ACOSTA SILVA, Carles (PIC); MERINO, Gonzalo (IFAE - Institute for High Energy Physics); CASALS HERNANDEZ, Jordi (Port d'Informació Científica (PIC)); FABREGA, Kevin (UAB)

Presenter: FLIX MOLINA, Jose (CIEMAT - Centro de Investigaciones Energéticas Medioambientales y Tec. (ES))

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **17** Type: **not specified**

Advancing a more energy-efficient and sustainable SDCC at BNL

Wednesday 11 December 2024 16:35 (20 minutes)

The Scientific Data and Computing Center (SDCC) at Brookhaven National Laboratory provides data services (storage, transfer and management), computing resources and collaborative tools to our worldwide scientific communities. Growing needs from our various programs coupled with data centre floor space and power constraints led us to the construction and occupancy of a new power-efficient data center and on-going studies of new hardware architectures that yield higher performance per unit of power (kWh). This presentation describes BNL's sustainability activities in support of current and future needs of core programs such as RHIC, EIC, HL-LHC, Belle-II, LQCD and NSLS-II.

Data center sustainability is a phenomenon that has grown in focus due to the continuing evolution of AI, HPC and HTC systems, leading to the rampant increase in carbon emissions due to the unprecedented rise in TDP of current computer chips. With an exponential increase of demand towards the usage of such systems, major challenges have surfaced in terms of productivity, PUE and thermal/scheduling management. Deploying AI/HPC infrastructure in data centers will require substantial capital investment.

This study at the SDCC quantifies the energy footprint of this infrastructure by developing models based on the power demands of AI hardware during training. We measured the instantaneous power draw of an 8-GPU NVIDIA H100 HGX node while training open-source models, including the ResNet image classifier and the Llama2-13b large language model. The peak power draw observed was about 18% below the manufacturer's rated TDP, even with GPUs near full utilization. For the image classifier, increasing the batch size from 512 to 4096 images reduced total training energy consumption by a factor of four when model architecture remained constant. These insights can aid scientific data center facilities to identify the 'stranded power'within existing facilities, assist in capacity planning and provide staff with energy use estimates. Future studies will explore the effects of liquid cooling technologies and carbon-aware scheduling on AI workload energy consumption. These results can help in the development of software or operational models, which may significantly reduce the carbon footprint at the data centers and identify opportunities for heat reuse.

Primary authors: CARAMARCU, Costin; LATIF, Imran (Brookhaven National Laboratory); CASELLA, Kevin; Mr COWAN, Matthew (Brookhaven National Laboratory); SMITH, Thomas; WONG, Tony

Presenter: LATIF, Imran (Brookhaven National Laboratory)

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **18** Type: **not specified**

Environmental efficiency of tape storage at CERN

Thursday 12 December 2024 15:15 (15 minutes)

The current capacity of the CERN's data storage tape infrastructure is around 1 EB.

The talk will first explain the 3 building blocks of this technology (automation, tape drives and media), their lifecycle and different resource requirements.

By using data compiled by IBM and Spectra Logic we will then outline the environmental and sustainability impacts of CERN's configuration in terms CO2 emissions and power consumption.

In the 3rd section we will explore how different usage patterns have different requirements on the overall infrastructure.

Primary author: BAHYL, Vladimir (CERN)

Presenter: BAHYL, Vladimir (CERN)

Session Classification: Hardware and Fabrics

Track Classification: All contributions

Contribution ID: **19** Type: **not specified**

Environmental Sustainability in Digital Research Infrastructures - the GreenDIGIT project

Wednesday 11 December 2024 17:00 (15 minutes)

GreenDIGIT (https://greendigit-project.eu/) is a TECH-01-01 Horizon Europe project that started in March 2024 to pursue environmental sustainability within digital services and service ecosystems that research infrastructures (RIs) rely on. GreenDIGIT brings together institutes from 4 digital RIs: EGI, SLICES, EBRAINS, and SoBigData to address the pressing need for sustainable practices in digital service provisioning.

Central to GreenDIGIT's mission is the identification of good practices in lemphasized textow-impact computing, which involves evaluating the broader computing landscape and identifying opportunities for improvements across various stakeholder groups both within, and beyond the RIs represented in the project consortium. The project will establish reference architectures and design principles that facilitate environmental impact considerations throughout the entire RI lifecycle, ensuring that sustainability is embedded within the entire design-implementation-operation-termination phases of RIs at the level of nodes, services and components.

Moreover, GreenDIGIT will develop and validate innovative technologies and methodologies that empower digital service providers to reduce energy consumption and/or overall operational environmental impact. By supplying technical tools for both providers and researchers, the project promotes the design and execution of environmentally conscious digital applications, aligning with principles of Open Science and FAIR data management.

Through education and support initiatives GreenDIGIT will foster a culture of sustainability among researchers and service providers, equipping them with best practices for lifecycle management and operation. By mapping the landscape of environmentally friendly computing, and by offering policy and technical recommendations, the project not only enhances the sustainability of research infrastructures but also contributes to the broader goal of mitigating climate change.
The first tangible outcome of the project is a landscape analysis that was conducted within the 4 participating RIs, and within other RIs that operate significant digital infrastructures. The presentation will share the first findings from this landscape report, which included responses from 15 EGI sites, many being present in WLCG too.

In the second phase of the project EGI members will extend the EGI-WLCG service configuration database (GOCDB) with environmental metrics to fuel "green workload optimisation strategies" in HTC and Cloud infrastructures. Based on the metrics the DIRAC workload manager (CNRS), the AI4OS AI/ML framework (CSIC), the Terraform DevOps tool (SZTAKI) and a Data Center Energy scheduler (CESNET) will be extended to optimise job/VM/task execution according to green strategies.

In summary, the GreenDIGIT project is a transformative initiative that champions environmental sustainability in digital research services, offering various outcomes and collaboration opportunities to the WLCG community.

Primary authors: CONDURACHE, Catalin (EGI Foundation); SIPOS, Gergely

Co-author: DEMCHENKO, Yuri (SNE Group, University of Amsterdam)

Presenter: CONDURACHE, Catalin (EGI Foundation)

Session Classification: Projects and Reports

Track Classification: All contributions

Contribution ID: **20** Type: **not specified**

Summary of the LOCO2024 Workshop on Low Carbon Computing

Wednesday 11 December 2024 17:20 (15 minutes)

LOCO2024, the 1st International Workshop on Low Carbon Computing was held in Glasgow on 3 Dec. It was an initiative of the Scottish Informatics and Computer Science Alliance that aimed to bring together researchers and practitioners from all over the world working on low carbon and sustainable computing.

The workshop provided a forum for presenting and discussion of new ideas, ongoing work and early results.

In this talk I will provide a summary of the relevant work presented and discussion points raised, on topics such as carbon awareness and energy efficiency, frugality and sufficiency, sustainable software engineering etc.

Primary author: VANDERBAUWHEDE, Wim (University of Glasgow)

Presenter: VANDERBAUWHEDE, Wim (University of Glasgow)

Session Classification: Projects and Reports

Track Classification: All contributions

Contribution ID: **21** Type: **not specified**

Calorimetric simulation with quantum computers: the energy perspective

Thursday 12 December 2024 12:10 (15 minutes)

During the HL-HLC era the WLCG, according to the current estimates, will be requested to provide millions of CPU years of computation annually. The largest single computing loads are particle tracking, hard scatter event generation and calorimetry simulation. Improving the computational performance of these loads is a very active research area that includes creation of GPU based codes and algorithms, applications of AI and to some extent the application of quantum computing (QC). There have been some recent advancements in the area of application of QC to simulation of particle showers in calorimeters. We conduct a first examination of energy cost of conducting calorimetry simulations using traditional approaches, AI-based approaches deployed on GPUs and QC-based approaches.

Primary author: FEDORKO, Wojtek (TRIUMF)

Presenter: FEDORKO, Wojtek (TRIUMF)

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **22** Type: **not specified**

Watt Counts: ARM Compute & Energy Accounting in the WLCG

Thursday 12 December 2024 15:35 (15 minutes)

In the push for sustainable solutions in High Energy Physics (HEP) computing, our WLCG Tier2 site at ScotGrid Glasgow has adopted ARM-based servers, achieving full production integration with increasingly positive results. Today we run significant workloads on ARM, including ATLAS production tasks, and our findings indicate a measurable reduction in energy consumption compared to traditional x86 servers. This talk will present the energy savings achieved so far and an outlook for the near future.

We will also outline recent efforts to standardize power-efficiency metrics across HEP centers by developing a unified HEP-Score/Watt output format in collaboration with the HEPiX Benchmark Working Group, to facilitate consistent data aggregation and reporting on energy efficiency.

Additionally, drawing inspiration from ATLAS PanDA CO2 reporting, we provide an example of power usage accounting at our site. Leveraging runtime data from Prometheus and power statistics, we calculate the energy consumption per Virtual Organization (VO), showing a potential approach to site-level CO2 accounting.

Primary author: SIMILI, Emanuele

Co-authors: BORBELY, Albert Gyorgy (University of Glasgow (GB)); BRITTON, David (University of Glasgow (GB)); STEWART, Gordon; SKIPSEY, Samuel Cadellin

Presenter: SIMILI, Emanuele

Session Classification: Hardware and Fabrics

Track Classification: All contributions

Contribution ID: **23** Type: **not specified**

Power efficiency gains from GPU optimised workloads with Geant4 based detector simulations

Thursday 12 December 2024 14:30 (15 minutes)

Improvements in algorithmic efficiency are, in principle, the 'free-est'sources of energy-efficiency gains, as reducing CPU time at a constant power clearly reduces total energy. However, it is not clear a priori that the same applies to porting parts of a code to special purpose accelerators such as GPUs.
We investigated this case for the Celeritas project's GPU offloading for Geant4's EM calorimeter simulation –this particularly relevant as ATLAS currently spends ˜40% of its CPU budge in simulation, and ˜20% on Geant4 'full-sim'.
We show that, for all cases, both run-time and energy consumption are reduced by moving to GPU. In the best cases, energy use is reduced by 42% and run-time by 54%.
We will discuss how these observations extend to strategic considerations for applications of accelerators to NetZero targets in the HL-LHC regime.

Primary author: Dr BORBELY, Albert Gyorgy (University of Glasgow (GB))

Co-authors: BRITTON, David (University of Glasgow (GB)); SIMILI, Emanuele (University of Glasgow (GB)); STEWART, Gordon; SKIPSEY, Samuel Cadellin

Presenter: Dr BORBELY, Albert Gyorgy (University of Glasgow (GB))

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **25** Type: **not specified**

Estimating the environmental cost of HEP software and computing hardware at the University of Manchester

Wednesday 11 December 2024 14:25 (15 minutes)

In this talk, we will describe the studies undertaken at the University of Manchester to estimate and improve the energy efficiency of computing hardware and software used by students and researchers.

The goal of these studies is to build an understanding of the environmental impact of paticle physics research focusing on two fronts:
1) the carbon cost of the hardware uses for high power computing hardware and the local computing cluster
2) the energy efficiency of data analysis software and machine learning models in "big data"-related scientific fields including as high-energy particle physics.

The focus of this contribution will be the energy efficiency of scientific software algorithms and MC generation packages, taking FastJet and Herwig as examples, as well as the energetic cost of top tagging algorithms. We will then describe our plans towards a lifecycle analysis for computing hardware, as well as the policies that could be put in place for hardware re-use; this may benefit also the Tier2 where we plan to do testing in a second phase. We will also describe our plans for integration into teaching materials for software development courses.

Primary authors: FORTI, Alessandra (The University of Manchester (GB)); DOGLIONI, Caterina (The University of Manchester (GB)); FITSCHEN, Tobias (The University of Manchester (GB))

Presenter: FITSCHEN, Tobias (The University of Manchester (GB))

Session Classification: Infrastructure

Contribution ID: **27** Type: **not specified**

Sustainability for Einstein Telescope Computing

Thursday 12 December 2024 10:50 (15 minutes)

Einstein Telescope (ET) is the proposed third generation ground-based gravitational wave observatory to be built in Europe in the mid-2030s. Sustainability is a key concern for Research Infrastructures and, thanks to its significant computing needs, ET has been charged with incorporating Sustainability criteria in its Computing Model already at the design phase. Although new, dedicated ET computing facilities will be limited in size, they will still consume significant energy. Distributed computing, using shared facilities, is expected to represent the majority of computing resources and thereby energy consumption. This presentation will summarise the current understanding of sustainability issues for Einstein Computing.

Primary authors: STAHL, Achim (Rheinisch Westfaelische Tech. Hoch. (DE)); TONELLO, Nadia (Barcelona Supercomputing Center); VERDIER, Patrice (IN2P3); LAYCOCK, Paul James (Universite de Geneve (CH)); Dr BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare, Torino)

Presenter: LAYCOCK, Paul James (Universite de Geneve (CH))

Session Classification: Environmental Sustainability: The Experiments

Track Classification: All contributions

Contribution ID: **29** Type: **not specified**

Sustainability Efforts at the KIT Scientific Computing Center

Wednesday 11 December 2024 14:05 (15 minutes)

KIT operates not only the GridKa WLCG Tier-1 Center but also extensive HPC facilities, large-scale storage services, and the fundamental KIT IT infrastructure.
Across these domains, we are actively working on various initiatives to optimize energy consumption and maximize waste heat reuse. We'll present our multifaceted approach which encompasses the following key areas:

1. Component-Level Energy Profiling: We meticulously measure the energy consumption of individual components within our data centers to identify potential hardware optimizations and enhance data center efficiency.
2. Energy-Efficient Hardware Procurement: When acquiring hardware for large-scale computing installations, we prioritize energy efficiency over the projected lifespan. Additionally, our involvement in the WLCG benchmarking working group enables us to rigorously test and select the most energy-efficient hardware.
3. Optimized Data Center Cooling: We implement different data center cooling strategies to significantly improve the Power Usage Effectiveness (PUE) of our facilities.
4. Energy-Efficient Software and Resource Allocation: We develop software solutions that leverage distributed and opportunistic resources, which are especially energy-efficient. Furthermore, this approach can be adapted to dynamically adjust the allocation of local computing resources based on external factors such as real-time CO2 emissions associated with the energy mix.
5. Energy-Conscious Data Center Planning: The planning and location of new data centers are primarily driven by the potential for waste heat reuse, ensuring sustainable and efficient operations.

Primary author: PETZOLD, Andreas (KIT - Karlsruhe Institute of Technology (DE))

Presenter: PETZOLD, Andreas (KIT - Karlsruhe Institute of Technology (DE))

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **30** Type: **not specified**

Development and implementation of an environmentally friendly Workload Management System in DiracX

Thursday 12 December 2024 11:30 (15 minutes)

The goal of this thesis is to develop an optimized task allocation algorithm for the Workload Management System (WMS) in DiracX, aimed at reducing energy consumption and maximizing the utilization of renewable energy sources. To achieve this, a combination of advanced techniques—including numerical optimization, clustering, and artificial intelligence (e.g., reinforcement learning)—will be explored. The proposed algorithm will take into account a variety of factors, such as (but not limited to):
* The energy mix at the country level (including forecasts, if available),
* Local resource center optimizations (e.g., waste heat recovery, cooling methods, dynamic downclocking),
* Global and local policy considerations (e.g., resource and climate pledges, energy regulations),
* The efficiency of high-energy physics (HEP) jobs at resource centers (based on factors like hardware performance, job failure rates, outages, etc.).

As this project is still in its early stages, the specific algorithm and the results of subsequent simulations and real-world tests have not yet been developed. However, this presentation will outline potential strategies for integrating environmental sustainability aspects into job allocation mechanisms on the grid and discuss methods for validating the effectiveness of the proposed solutions.

Primary author: GIEMZA, Henryk (National Centre for Nuclear Research (PL))

Presenter: GIEMZA, Henryk (National Centre for Nuclear Research (PL))

Session Classification: Environmental Sustainability: The Experiments

Track Classification: All contributions

Contribution ID: **39** Type: **not specified**

ALICE

Primary author: PIANO, Stefano (INFN (IT))

Presenter: PIANO, Stefano (INFN (IT))

Session Classification: Environmental Sustainability: The Experiments

Contribution ID: **40** Type: **not specified**

ATLAS

Thursday 12 December 2024 09:20 (15 minutes)

Primary author: MARSHALL, Zach (Lawrence Berkeley National Lab. (US))

Presenter: MARSHALL, Zach (Lawrence Berkeley National Lab. (US))

Session Classification: Environmental Sustainability: The Experiments

Contribution ID: **41** Type: **not specified**

CMS

Thursday 12 December 2024 09:40 (15 minutes)

The rapid evolution of computing technologies, including AI and high-performance computing (HPC), has significantly impacted scientific advancements, particularly in high-energy physics (HEP). However, this progress comes at an environmental cost, as the energy and carbon footprints of computing tasks continue to grow. This presentation explores the energy consumption and carbon footprint of CMS simulation workflows, focusing on GENSIM jobs at the Large Hadron Collider (LHC).

We introduce a monitoring framework designed to measure energy usage and assess the environmental impact of HEP workloads. Preliminary results from the INFN-CNAF nodes demonstrate that optimizing resource allocation can significantly reduce energy consumption without compromising output. Metrics are normalized to provide actionable insights for sustainable computing practices.

Future developments aim to refine error estimation, build a structured database for footprint data collection, and establish automated analysis pipelines. These efforts align with the broader goal of mitigating the environmental impact of computational physics experiments.

Primary author: PICCINELLI, Andrea (University of Notre Dame (US))

Presenter: PICCINELLI, Andrea (University of Notre Dame (US))

Session Classification: Environmental Sustainability: The Experiments

Contribution ID: **42** Type: **not specified**

LHCb

Thursday 12 December 2024 10:00 (15 minutes)

Primary author: GIEMZA, Henryk (National Centre for Nuclear Research (PL))

Presenter: GIEMZA, Henryk (National Centre for Nuclear Research (PL))

Session Classification: Environmental Sustainability: The Experiments

Contribution ID: **44** Type: **not specified**

Carbon Reporting in GridPP

Presenter: OWEN, Richard Alexander

Contribution ID: **46** Type: **not specified**

Lightning Talk: Estimating the environmental cost of HEP software and computing hardware at the University of Manchester

Presenter: ALANDES PRADILLO, Maria (CERN)

Contribution ID: **47** Type: **not specified**

Introduction to the Workshop, Structure and Goals

Wednesday 11 December 2024 13:30 (10 minutes)

Presenters: BRITTON, David (University of Glasgow (GB)); CAMPANA, Simone (CERN)

Contribution ID: **68** Type: **not specified**

Sustainable AI

Pending..

Primary author: Dr VALLECORSA, Sofia (CERN)

Presenter: Dr VALLECORSA, Sofia (CERN)

Session Classification: Workloads and Metrics

Contribution ID: **71** Type: **not specified**

FlashSim at CMS: performance and resources of deep learning based simulation for HEP

Thursday 12 December 2024 13:50 (15 minutes)

The CMS FlashSim simulation framework is an end-to-end ML based simulation that can speed up the time for production of analysis samples of several orders of magnitude with a limited loss of accuracy. Detailed event simulation at the LHC is taking a large fraction of computing budget. As the CMS experiment is adopting a common analysis level format, the NANOAOD, for a larger number of analyses, such an event representation is used as the target of this ultra fast simulation that we call FlashSim. Generator level events, from PYTHIA or other generators, are directly translated into NANOAOD events at several hundred Hz rate with FlashSim. CMS showed how training FlashSim on a limited number of full simulation events is sufficient to achieve very good accuracy on larger datasets for processes not seen at training time. In this work, we focus on presenting the usage of compute resources for this new framework, for both training and inference, and try to establish a comparison with the full simulation of the CMS experiment.

Primary author: VASELLI, Francesco (Scuola Normale Superiore & INFN Pisa (IT))

Presenter: VASELLI, Francesco (Scuola Normale Superiore & INFN Pisa (IT))

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **72** Type: **not specified**

Holistic cost analysis of running a computing center

Wednesday 11 December 2024 14:45 (15 minutes)

The Tier-2 computing center at MIT has been established in 2006 and has been a prominent contributor to the CMS computing ever since. Over the years hardware has aged and we are faced with a major re-design which affects our principal design ideas that have driven our purchases over the years. In this context we have done a more holistic evaluation of the cost which for the first time includes power consumption based on our own measurements under realistic loads.
We compare cost of running an aging computing center and propose hardware retirement schedules that are more cost effective when considering power consumption and rack space.

Primary author: WANG, Zhangqier (Massachusetts Inst. of Technology (US))

Co-authors: PAUS, Christoph (Massachusetts Inst. of Technology (US)); GONCHAROV, Maxim (Massachusetts Inst. of Technology (US))

Presenter: WANG, Zhangqier (Massachusetts Inst. of Technology (US))

Session Classification: Infrastructure

Track Classification: All contributions

Contribution ID: **74** Type: **not specified**

Summary and Outlook

Friday 13 December 2024 12:00 (20 minutes)

Co-authors: DI GIROLAMO, Alessandro (CERN); BRITTON, David (University of Glasgow (GB)); LETTS, James (Univ. of California San Diego (US)); RADEMACKER, Jonas (University of Bristol (GB)); SCHULZ, Markus (CERN); WADENSTEIN, Mattias (University of Umeå (SE)); CAMPANA, Simone (CERN); MARSHALL, Zach (Lawrence Berkeley National Lab. (US))

Presenters: DI GIROLAMO, Alessandro (CERN); BRITTON, David (University of Glasgow (GB)); LETTS, James (Univ. of California San Diego (US)); RADEMACKER, Jonas (University of Bristol (GB)); SCHULZ, Markus (CERN); WADENSTEIN, Mattias (University of Umeå (SE)); CAMPANA, Simone (CERN); MARSHALL, Zach (Lawrence Berkeley National Lab. (US))

Session Classification: WLCG: Ideas for a plan

Contribution ID: **75** Type: **not specified**

Introduction to the discussion topics

Friday 13 December 2024 09:00 (20 minutes)

Session Classification: WLCG: Ideas for a plan

Contribution ID: **76** Type: **not specified**

Discussion A

Friday 13 December 2024 09:20 (1h 20m)

Session Classification: WLCG: Ideas for a plan

Contribution ID: **77** Type: **not specified**

Discussion B

Friday 13 December 2024 11:00 (1 hour)

Session Classification: WLCG: Ideas for a plan

Contribution ID: **78** Type: **not specified**

Accelerating Monte-Carlo event generation on CPUs and GPUs

For HL-LHC, recent studies have shown that around 20% of the total computing resources will be used for Monte-Carlo simulations. Without efficiency improvements the computing needs of the experiments will not be met by the available resources. Motivating research avenues in the realm of GPU acceleration.
In this talk, we will present the current status of the GPU and vectorised version of the widely used event generator MadGraph5_aMC@NLO.
We will discuss the performance improvements achieved so far, comparing the results with the original CPU version, and giving insights on the power consumption of these new approaches.

Presenter: MASSARO, Daniele (CERN)

Contribution ID: **79** Type: **not specified**

Accelerating Monte-Carlo event generation on CPUs and GPUs

Thursday 12 December 2024 14:10 (15 minutes)

For HL-LHC, recent studies have shown that around 20% of the total computing resources will be used for Monte-Carlo simulations. Without efficiency improvements the computing needs of the experiments will not be met by the available resources. Motivating research avenues in the realm of GPU acceleration.
In this talk, we will present the current status of the GPU and vectorised version of the widely used event generator MadGraph5_aMC@NLO.
We will discuss the performance improvements achieved so far, comparing the results with the original CPU version, and giving insights on the power consumption of these new approaches.

Primary author: MASSARO, Daniele (CERN)

Presenter: MASSARO, Daniele (CERN)

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **80** Type: **not specified**

Energy Efficient Computing with ALICE O2

Thursday 12 December 2024 09:00 (15 minutes)

ALICE upgraded its computing infrastructure and software to handle the increased data rates and enhanced detector performance in Run 3. This upgrade significantly reduced energy consumption and CO2 emissions through:
- Optimized resource usage: Maximizing GPU utilization, parallelizing and vectorizing CPU code, and organizing analyses into trains.
- Energy-efficient techniques: Employing low PUE centers for tuning and validation of analysis, reusing generated events in MC simulation, and integrating support for ARM architectures.
- EPN infrastructure: Cost savings and performance improvements through energy-efficient cooling systems and extensive use of GPUs for real time data compression.
These strategies will be discussed in detail during the talk.

Primary author: PIANO, Stefano (INFN (IT))

Presenter: PIANO, Stefano (INFN (IT))

Session Classification: Environmental Sustainability: The Experiments

Contribution ID: **81** Type: **not specified**

Energy Sustainability in Artificial Intelligence

Thursday 12 December 2024 13:30 (15 minutes)

The rapid advancement of artificial intelligence (AI) technologies is driving transformative changes across various sectors, including High Energy Physics (HEP). While the energy consumption associated with AI systems poses a challenge to sustainability, the integration of AI within HEP offers significant benefits in terms of efficiency and modernization of the HEP computing model. This talk addresses the critical issue of energy sustainability in AI, focusing on the factors contributing to its high energy usage, such as the computational intensity of deep learning algorithms. In response to these challenges, several strategies can be identified to reduce AI's energy footprint. These include the design of optimal deep learning architectures, the implementation of advanced model training techniques and the use of energy-efficient hardware.

Primary author: Dr VALLECORSA, Sofia (CERN)

Presenter: Dr VALLECORSA, Sofia (CERN)

Session Classification: Workloads and Metrics

Track Classification: All contributions

Contribution ID: **82** Type: **not specified**

Raising Awareness on IT's Environmental Impact

Primary authors: ALANDES PRADILLO, Maria (CERN); BHAT, Varsha

Presenter: ALANDES PRADILLO, Maria (CERN)

Session Classification: Projects and Reports

Track Classification: All contributions