Exploitation of HPC Resources for data intensive sciences

vCHEP 2021 - software track

David Southwick (CERN)
Viktor Khristenko (CERN)
Maria Girone (CERN)
Miguel F. Medeiros (CERN)
Domenico Giordano (CERN)
Ingvild Brevik Høgstøyl (Norwegian University of Science and Technology)
Luca Atzori (CERN)

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

Introduction

Exascale HPC machines will provide processing capacities similar to or greater than the entire compute grid. Common challenges* continue to drive HPC adoption:

Exascale HPC machines are going to be based on heterogeneous hardware architectures

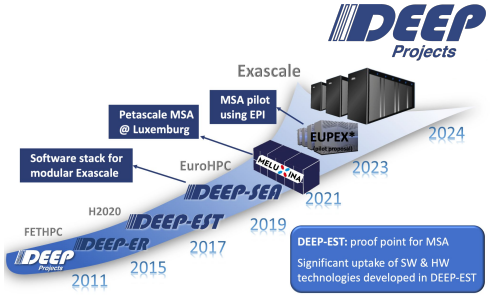
- Need to leverage compute accelerators

Need to understand how efficiently and how much of these resources are used

- Need to develop benchmarking & accounting tools for HPC

HEP must bring data to (and from) Exascale HPC machines to make use of them

- Evaluate various shared storage systems and Data Access mechanisms



DEEP roadmap to Exascale

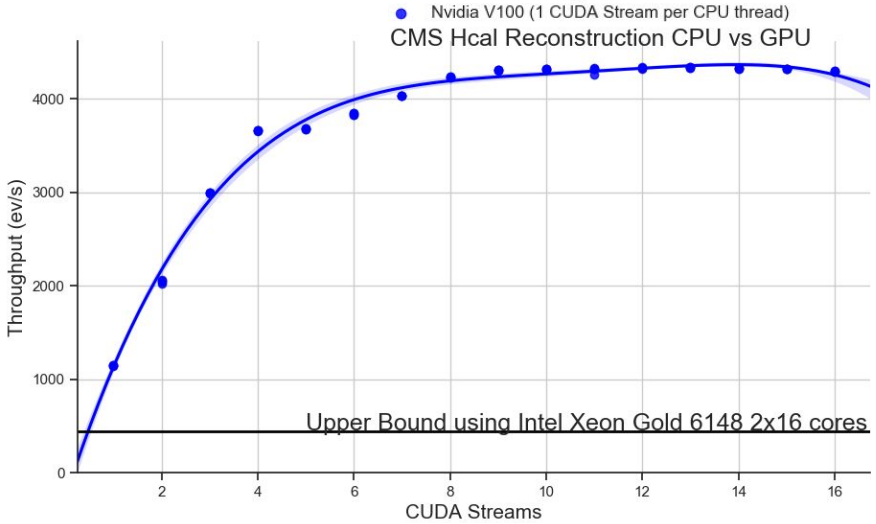
*Common challenges for HPC integration into LHC computing, *M. Girone*

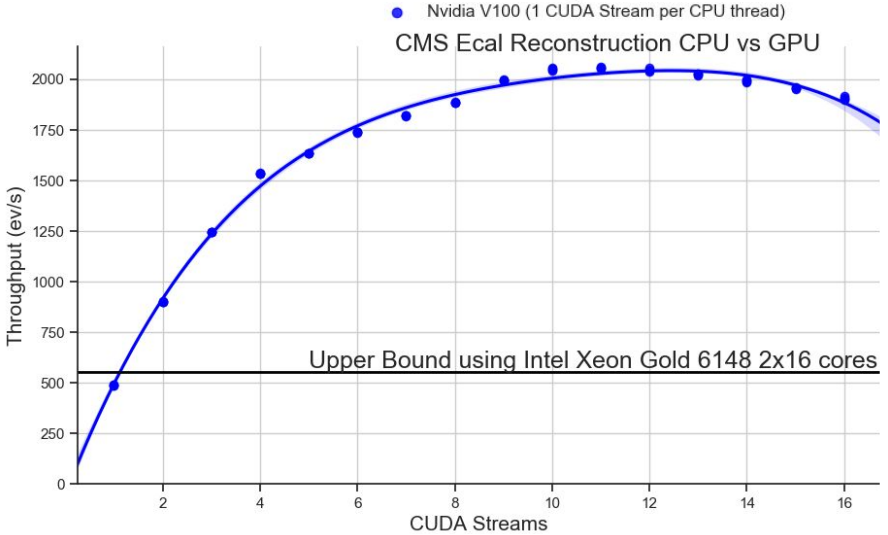
Porting experiment workloads to GPUs
CMS Hcal/Ecal only

HPC Centers exploit heterogeneous computing to reach Exascale:

Using Nvidia V100 GPU for Patatrack (tested http://opendata.cern.ch/record/12303)

- Hcal -> speed of **7-8x**
- Ecal -> speed of **3-4x**



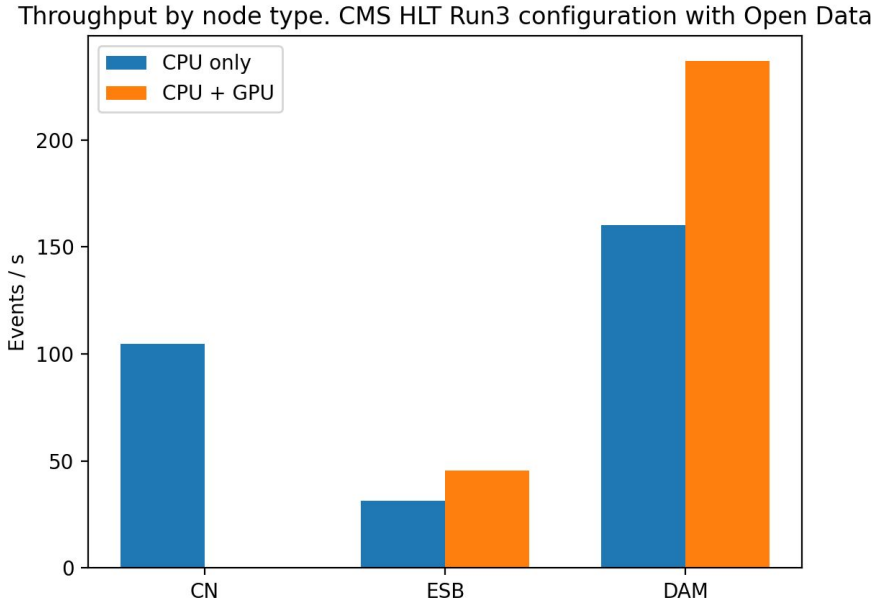


Porting experiment workloads to GPUs
CMS HLT-like Run3

CMS HLT-like Run3 configuration, including Patatrack GPU developments

Open Data source: http://opendata.cern.ch/record/12303

50% more throughput out of nodes with Nvidia GPUs (V100 here)



HEP Benchmark Suite

Extended for HPC

Benchmarking and accounting of heterogeneous compute resources remains on the critical path to HPC adoption. Collaboration with HEPiX Benchmarking Group to refactor & re-tool for HPC execution at scale:

- New unprivileged & modular python3 interface
- Workloads now Singularity by default; Docker/OCI-compatible supported
- Multi-Arch, Multi-GPU containers: enables comparison across heterogeneous architectures
- Easily extendable to other areas of science!

See vCHEP 2021 *HEPiX Benchmarking* plenary from M. Medeiros (this morning, 9:30)

```
# HEP Benchmark Suite requires singularity 3.5.3+, python3.
module load singularity python3
python3 -m pip install --user git+https://gitlab.cern.ch/hep-benchmarks/hep-benchmark-suite.git

echo "Running HEP Benchmark Suite on $SLURM_CPUS_ON_NODE Cores"
srun bmkrun --config default
```

gitlab.cern.ch/hep-benchmarks/hep-benchmark-suite

Benchmarking on HPC

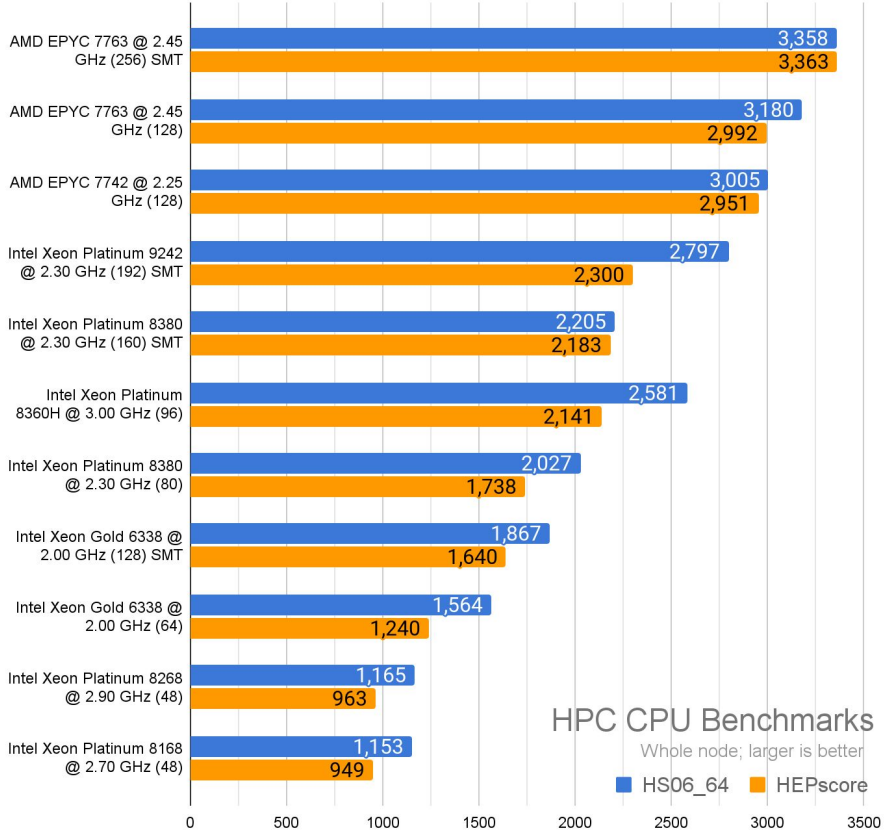
Results

Already deployed across several HPC sites:

- **2,316+ HPC nodes benchmarked**
- **155k+ cores**
- **6.7M+ HEPscore seen (~7M HS06+)**
- **Heterogeneous hardware (AMD/Intel/ARM + Nvidia GPUs)**
- **Automated reporting of all results**

Enabling resource accounting at unprivileged computing sites

Better information for procurement on heterogeneous accelerators



Example results comparing HS06 and HEPscore across recent HPC CPUs

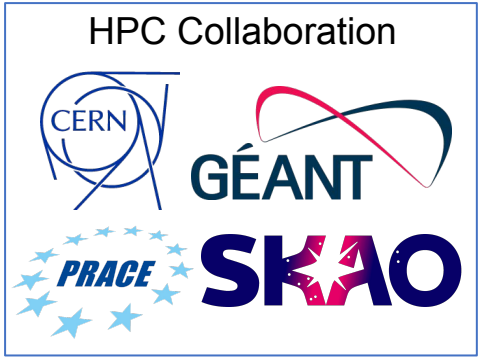
Thank you to supporting HPC sites!

SDSC SAN DIEGO SUPERCOMPUTER CENTER

FLATIRON INSTITUTE

Data Access / HPC

Exascale challenge



Lots of moving parts! Break down challenge into three areas:

1. **Efficient usage of storage systems on site**
2. **Data ingress/egress from HPC centers**
3. **Dynamic scaling interaction between (1) and (2)**

Started tackling (1) - We look at this from processing site perspective only:

- **Performed first tests @SDSC with CMS production-like workflow & ROOT I/O framework**
- **Evaluating other storage systems/confs (e.g. BeeGFS) as well**
  - **via PRACE Summer Of HPC project and CoE RAISE**
- **Developing applications to simulate I/O traffic to/from Shared Storage Systems**
  - **Easier to package and prepare as a benchmark**

Data Access

Efficient usage of storage systems on site

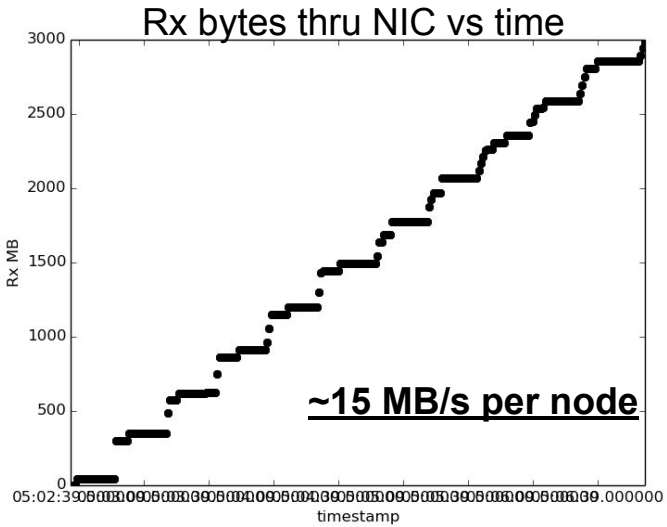
First tests with CMS production-like workloads to evaluate the needs/impact on the Shared Storage Systems

CMS MINIAOD2NANO workflow; scaling to 200 nodes

- ~3000 MB/s aggregate from Ceph (SDSC)
- This is on average (HEP i/o is bursty)
- flat performance -> **289.22 Evs/s/node**

Hypothetically, Exascale HPC O(1M) cores total -> O(10K) nodes -> O(100-150GB/s) aggregate

- Although we might never use from a single site



Summary & Next Steps

To meet a looming computational resource gap, CERN must evolve its computing platform to leverage heterogeneous computing and HPC systems

The DEEP-EST project proved to be an invaluable platform for

- Collaboration with HPC experts from other sciences/centers to perform HEP tests and development towards the usage of Exascale HPCs

Developing benchmarking on HPC Next Steps:

- We continue to work with run3 heterogeneous workloads as they become available (ARM/Power/GPU etc)

Data Access / HPC

- In the coming months, we will leverage on the HPC collaboration with PRACE to demonstrate scale, as well as focus on data ingress/egress

Thank you!

Contact: **egi-ace-po@mailman.egi.eu**
Website: www.egi.eu/projects/egi-ace

EGI Foundation

@EGI_eInfra

HPC Collaboration

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

DEEP Projects

The DEEP projects have received funding from the European Union's Seventh Framework Programme (FP7) for research, technological development and demonstration and the Horizon2020 (H2020) funding framework under grant agreement no. FP7-ICT-287530 (DEEP), FP7-ICT-610476 (DEEP-ER) and H2020-FETHPC-754304 (DEEP-EST).