Track1 Summary

ACAT 2017 Seattle

Shih-Chieh Hsu
University of Washington Seattle

on behalf of Track1 Coordinators
Niko Neufeld, Maria Girone, Graeme Stewart

Track1 Coordinations

- Thanks for all contributors for exciting and high quality presentations
  - 82 abstracts submitted
  - 34 talks accepted
  - 38 posters accepted out of which 32 were actually shown
  - 1 round-table discussion
- Many thanks for Track1 Coordinators, and session convenor - Stefan Roiser.
- Disclaimer: being the local coordinator in Track1
  - Special thanks for Niko to help me preparing for this talk.
  - Selected highlight based on common features
  - Apologize for “Not mention” your contributions due to my bias ...

Common Trends

- GRID to Cloud-federation
- Standardization of Tools
- On-line DAQ: harmonization interface
- Heterogeneous resources
- New Paradigm - From Clever to Revolution

Virtualization of the PNNL Computing Infrastructure

- Large services deployed from bare-metal over virtualization to containers.
- Use and adapt existing industry standard solutions.

- Individual machines
- Automated provisioning
- Virtual machines
- OpenStack Cloud
- Repo Mirrors
- Containers
- Kubernetes

Kubernetes + Docker Engine

OpenStack + KVM

Ceph

GitLab

Lustre

LoadBalancing/HA

PerfSonar

Prometheus

Grafana

CheckMK

ElasticSearch

389-DS

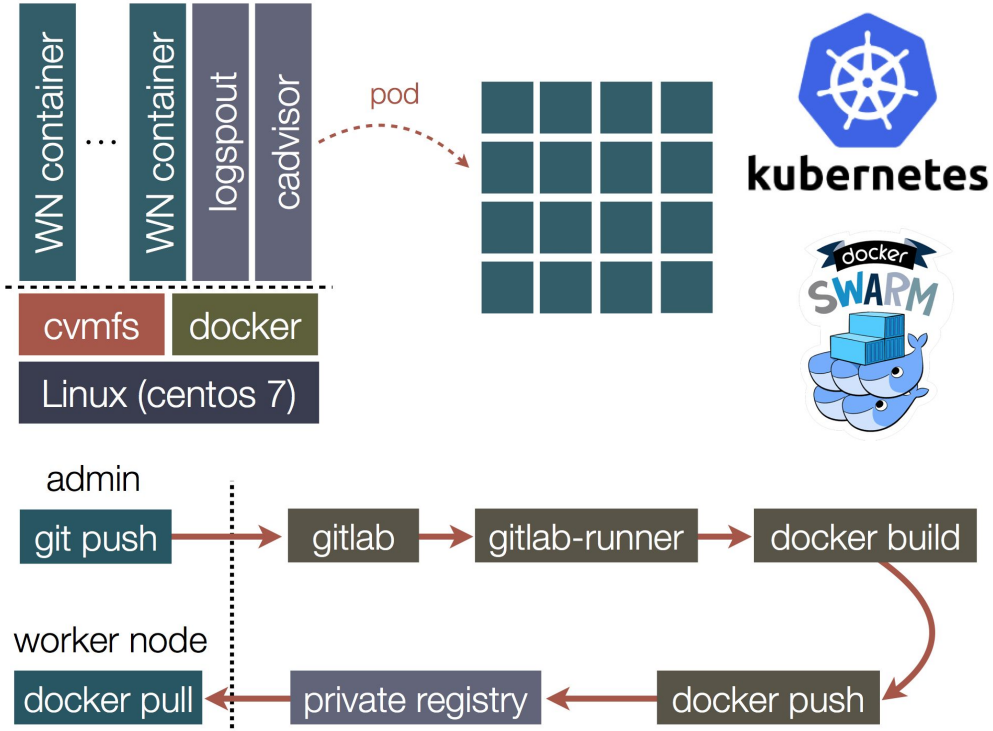
Cobbler

NFS

- **DIRAC**
  - Distributed Data Management System
  - Gatekeeper Services
  - Many development and testing services
- **Condor CE's**
  - DIRAC SiteDirector
  - HTCondor cluster
  - Squid Cache
- **Leadership Class Facility CE's**
  - DIRAC SiteDirector
  - HPC Cluster
- **SE's**
  - BestMan2
  - Gridftp
  - Backed by Lustre
- **Belle2DB**
  - REST Service
  - UI Service
  - Payload Service
  - Squid Cache
  - Postgresql Relational Database
- **FTS3**
- **CVMFS Stratum**
  - Zero
  - One
- **Authorization**
  - Gums
  - VOMS Server with multiple VO's

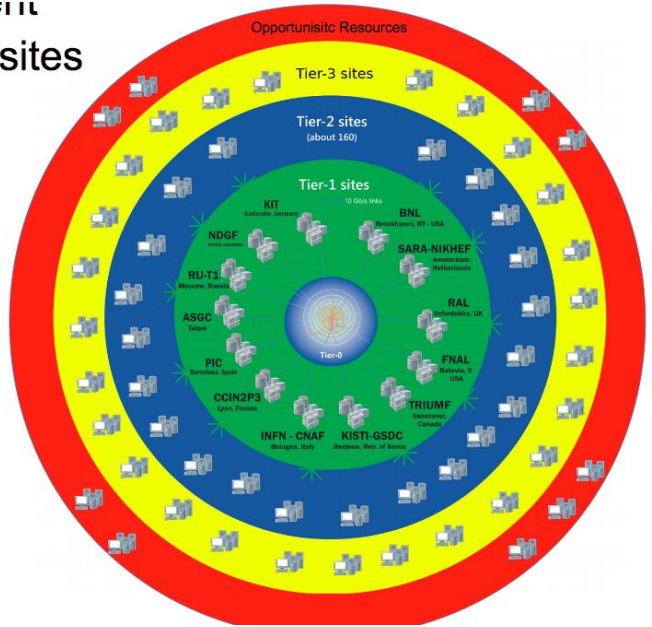
Container model for resource provision at a WLCG

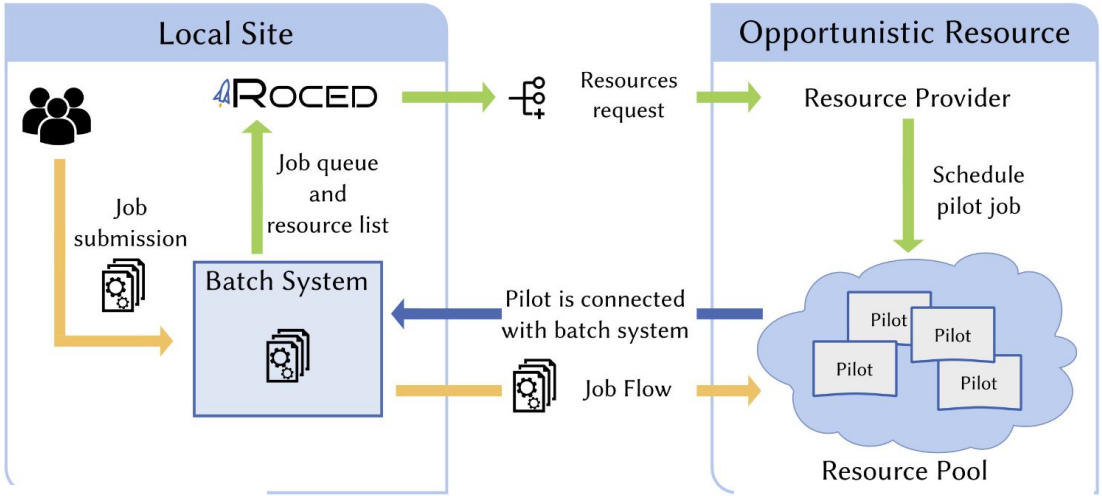
- Reimplement VAC idea using containers.
- Establish PoC by running production ATLAS multi-core payloads
- Integrated monitoring, logging and constructed a CI/CD pipeline



Mastering Opportunistic Computing Resources for HEP

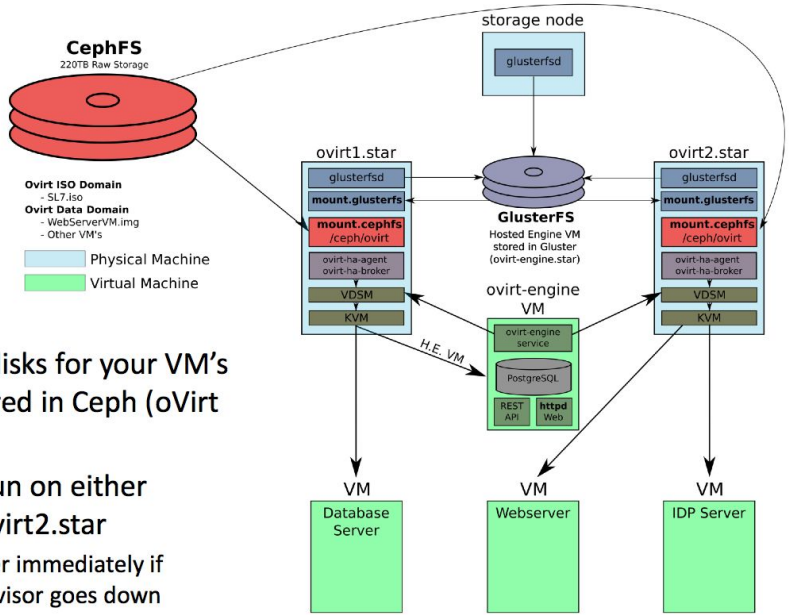
- Use Docker image, on-demand provisioning of VMs, usage of shared HPC





oVirt Virtualization and CephFS in STAR

- Built from Open Source tools instead of using expensive commercial solution
- Self-Hosted Engine feature is a true High Availability setup with oVirt to ensure VMs are always up and remove single point of failure



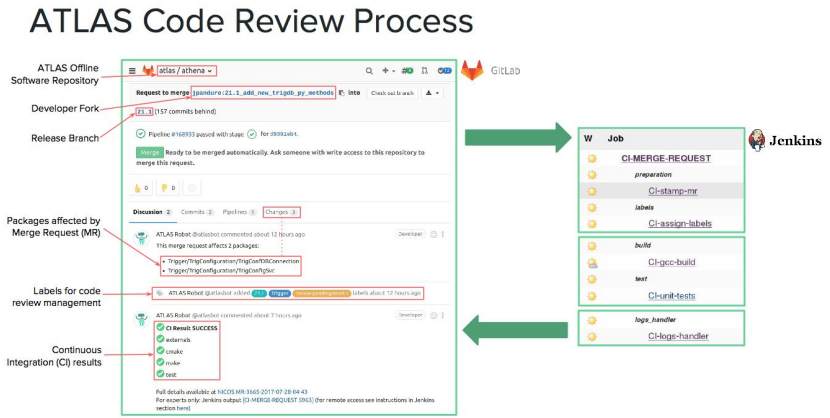
- Create Virtual disks for your VM's that will be stored in Ceph (oVirt Data Domain)
- The VM's can run on either ovirt1.star or ovirt2.star
  - VM will failover immediately if running hypervisor goes down
  - VM's can be migrated live between hypervisors

Using standardized Tools

- A trend to revisit and to use standardized tools to customize what we need

Modernising ATLAS Software Build Infrastructure

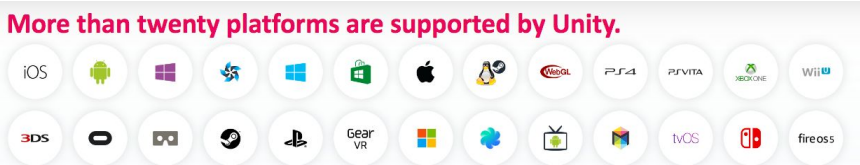
- Moved much of our software infrastructure from custom tools to well established third-party tools
- Established code reviews and continuous integration tests as part of the standard workflow, public documentation, tutorial



CMake git GitLab Jenkins

Visualization in JUNO

- New development using game engine Unity for event display



ROOT(SERENA)

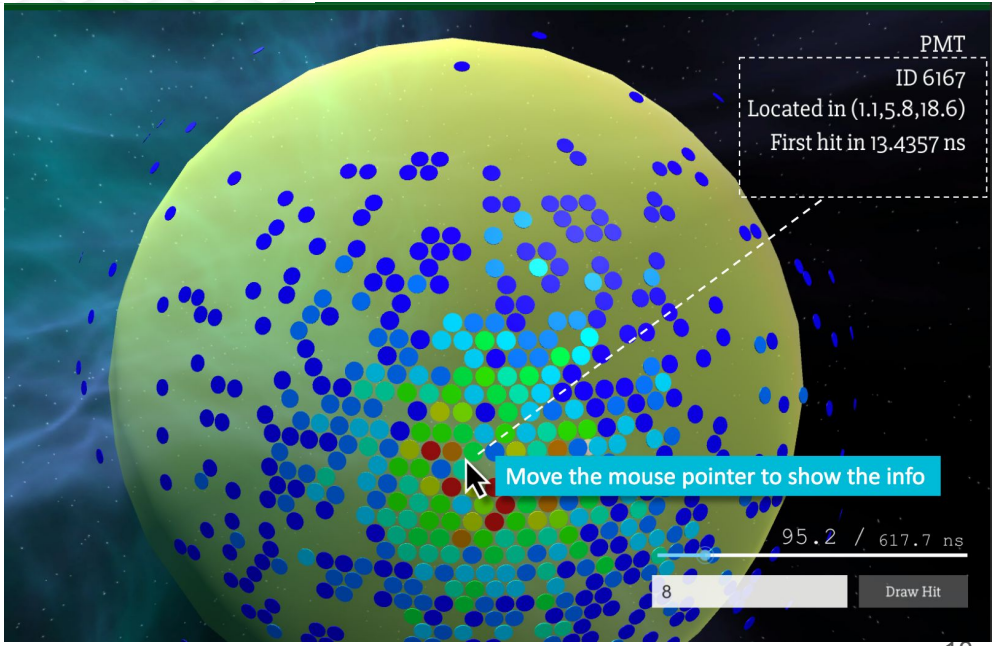
Easy to input and output the data in root file format.
Well developed for now. (Geometry, Event, Mc truth and Reconstruction)

Integrated in JUNO offline as a part of JUNO offline.
Plugin is needed if users want to display remotely.
Visual effect is limited by ROOT.

Unity

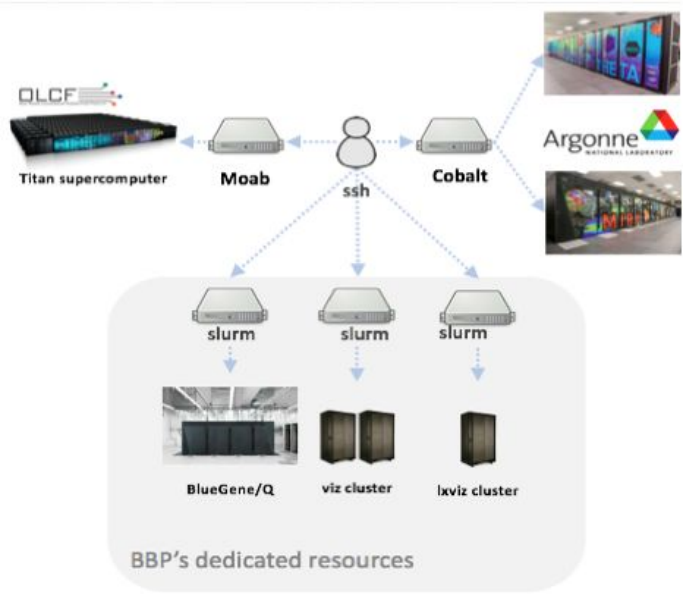
Easy to transplant to other platform like windows, Linux, mac, web.
Built as a client, which can be run in user's own PC without JUNO offline.
Fancy visual effect is available as a game engine.

Need the data conversion when loading a root file.

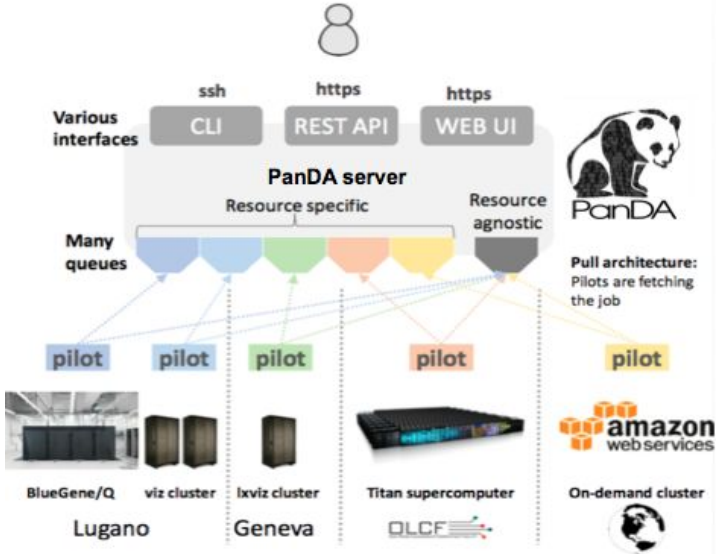


BigPanda and Big Brain Project

- Supercomputers, Clouds and Grids powered by BigPanDA for Brain studies
- ATLAS pioneer to disseminate tools outside HEP.
- Phase I successfully finished - proof-of-principle LHC software works in BBP.



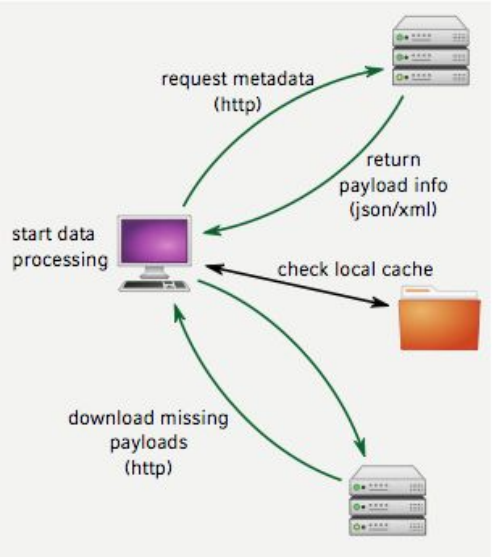
Machine Specific Setup



Common Functional Building Blocks

Conditional Database

- Required for data taking, worldwide for gird/cloud processing, vastly different lifetime/size per sub-detector
- Belle II: Server (PNLL)/Client (LMU)



Martin Ritter/Lynn Wood (P)/Martin Rigger (P)

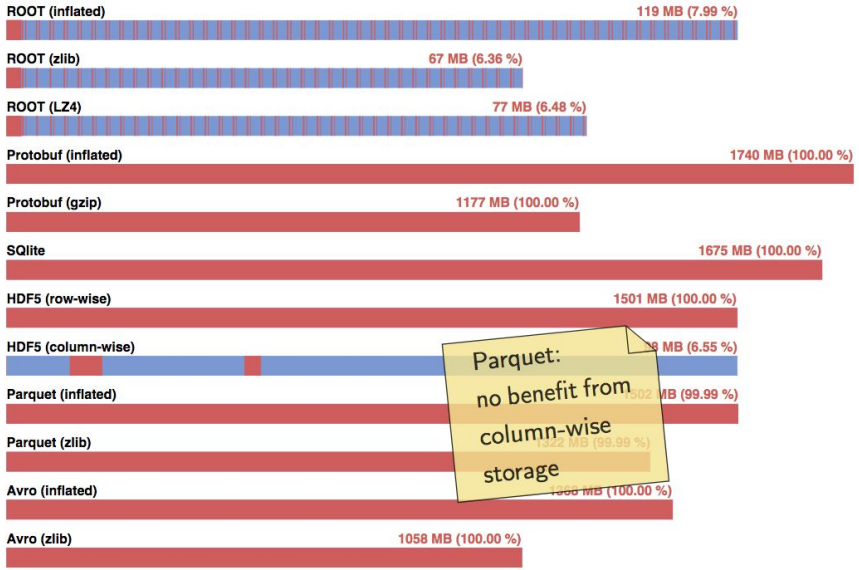
- Crest: a cross-experiment initiative, ATLAS/CMS/NA62/ Belle2, LHCb,
- Access through REST services using standard Open API specifications, implemented in Swagger

Crest Pro-cache

Paul Laycock [P]

Customized Tools specialized to HEP

- Benchmark with 8.5 million LHC run 1 events B → KKK, Check BitFlips, Encoding efficiency, Throughput (Memory/SSD/Reading/Plotting), Cost, I/O Pattern



Data Formats

- Large performance differences by different data formats and libraries
- Protobuf is a good benchmark for serialization performance
- HDF5 is most suitable for HPC-style processing of multi-dimensional arrays
- Parquet is an interesting format but C++ libraries are not yet ready

ROOT I/O

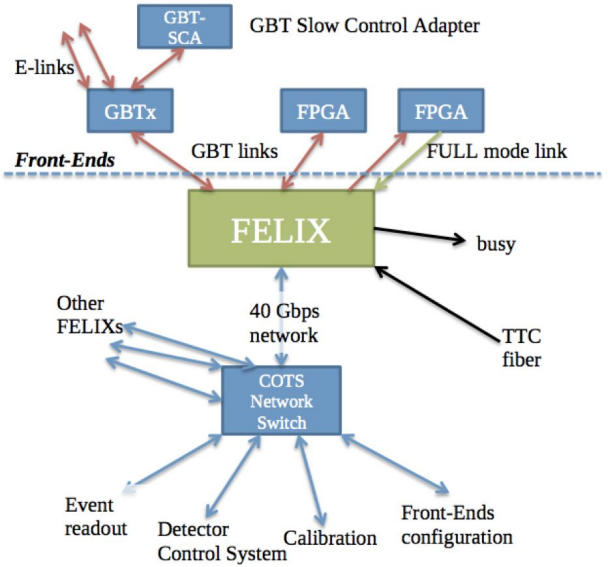
- In general best performance in these tests
- Several issues found & being addressed
- LZ4 turns out to be a good trade-off for analysis data sets
- Optionally checksummed ROOT files might be desirable

On-line DAQ: Harmonization Interface

- A trend to Standardize, homogenize, to more streamline, to have more common system.
- The challenging requirement makes it difficult to maintain development as individual.
- A trend to reduce number of customized electronics

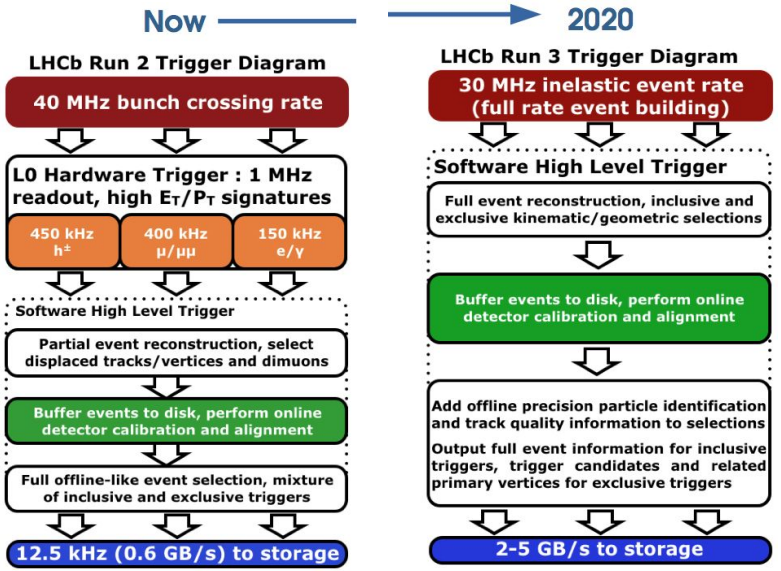
FELIX: ATLAS Readout Upgrade for Run3 and Run4

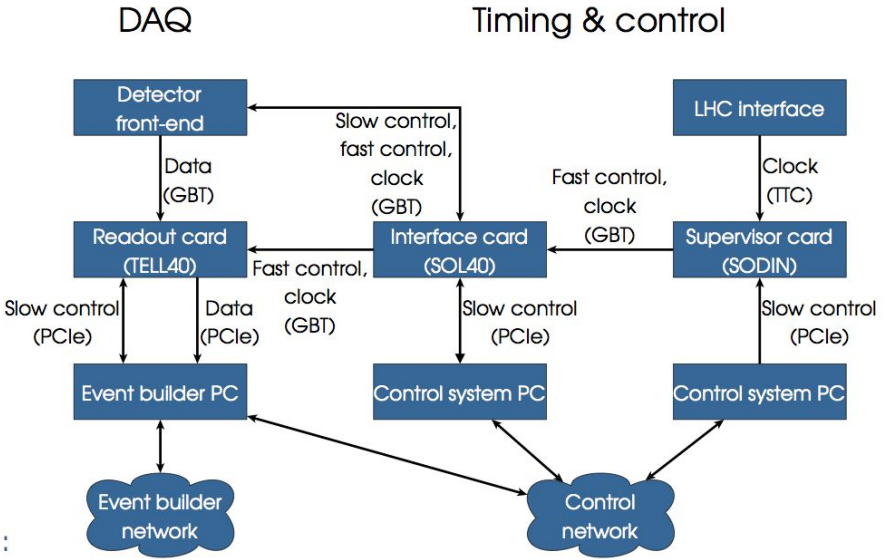
- Front-End LInk eXchange (FELIX) can service distinct paths for many bi-directional purposes, support use of Commercial Off-The-Shelf (COTS), Scalability, Detector Agnostic, use new GBT
- FPGA based router for GBT, based on FPGA PCIe card



LHCb Trigger in Run3

- No more hardware trigger
- Optimize cost with high usage factors in network and use single versatile PCIe card
- Fuse Offline into Online

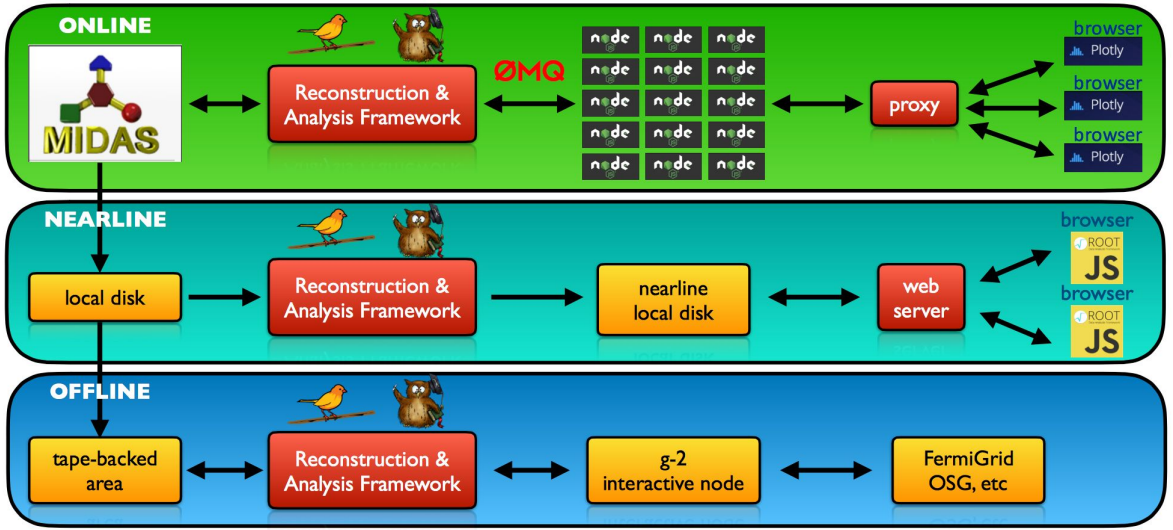




Tommaso Colombo/Rosen Matev/Stefan Roiser

g-2 on-line and off-line reconstruction

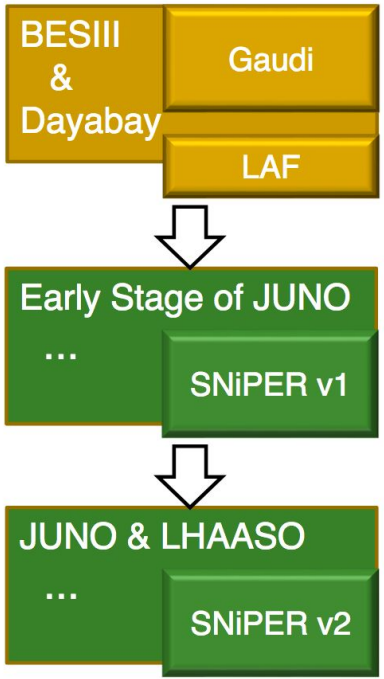
- g-2 implements highly-integrated system from on-line to offline using modern standarded tools.

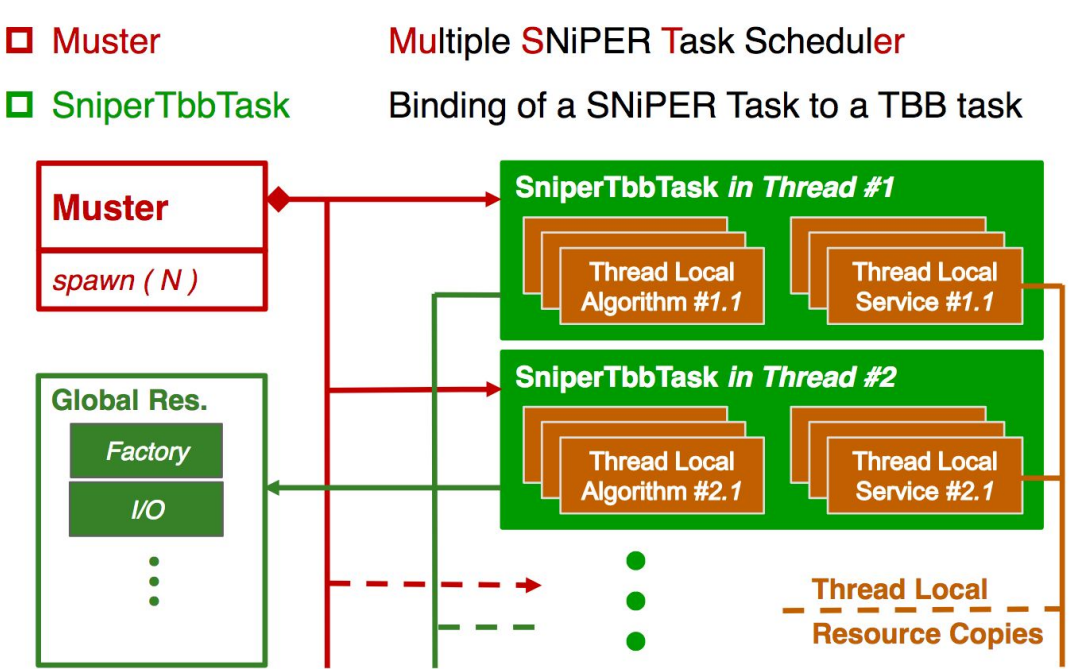


Intel TBB Eigen ØMQ ZeroMQ plotly Plotly

SNiPER on Intel TBB

- SNiPER : Software for Non-collider Physics Experiment
  - a general purpose offline software framework
  - lightweight and simple to use
- SNiPER Muster is implemented based on Intel TBB for multi-threading task



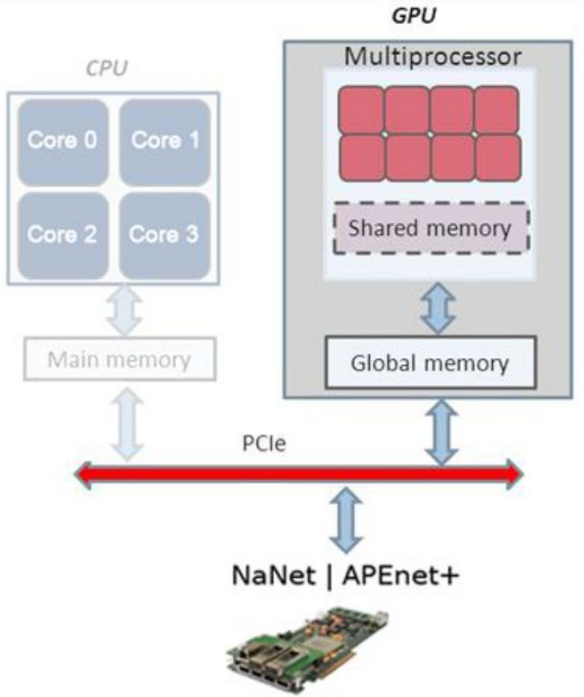


Heterogeneous Resource

Successful application of GPU and FPGA to on-line
Reference: Andrew Putnam’s talk from MicroSoft Catapult Project

Using GPU in Trigger in NA62

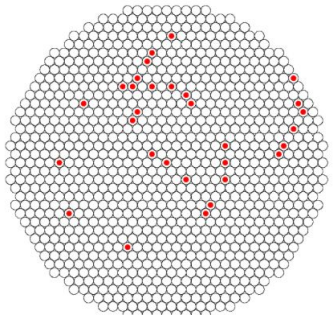
FPGA based NIC is an enabler of Heterogeneous Low Latency stream computing.
First successful GPU integration in a production HEP hardware trigger



L0 RICH trigger algorithm

Requirements for an on-line RICH reconstruction algorithm:

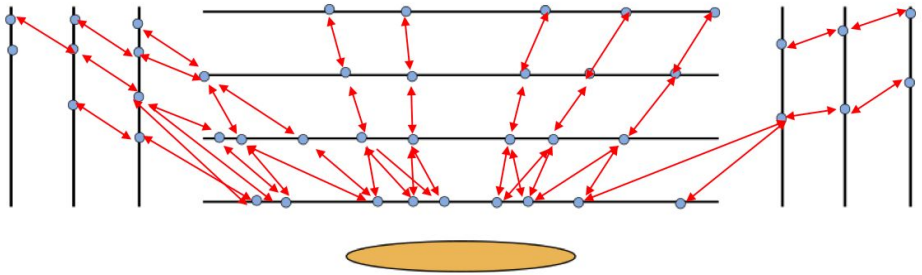
- **Trackless**
  - No information from the tracker
  - Difficult to merge information from many detectors at L0
- **Multi-rings**
  - Many-body decays in the RICH acceptance
- **Fast**
  - Events rate at ~10 MHz
- **Low latency**
  - Online (synchronous) trigger
- **Accurate**
  - Offline resolution required



Use GPU for Particle ID!

Using GPU in Trigger in CMS

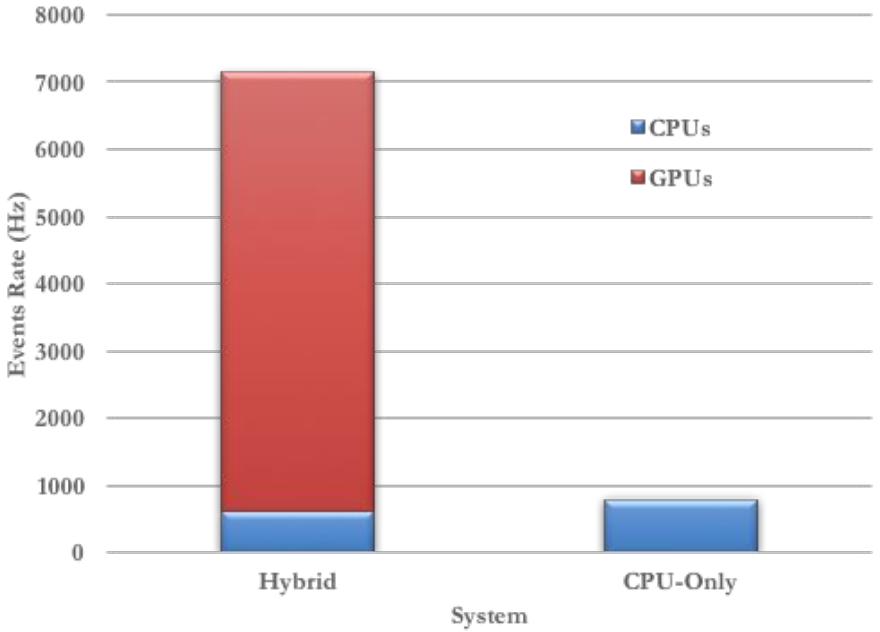
Pixel Track Seeding algorithm have been redesigned with high-throughput parallel architectures with Cellular Automaton (CA)



To reach 100kHz:

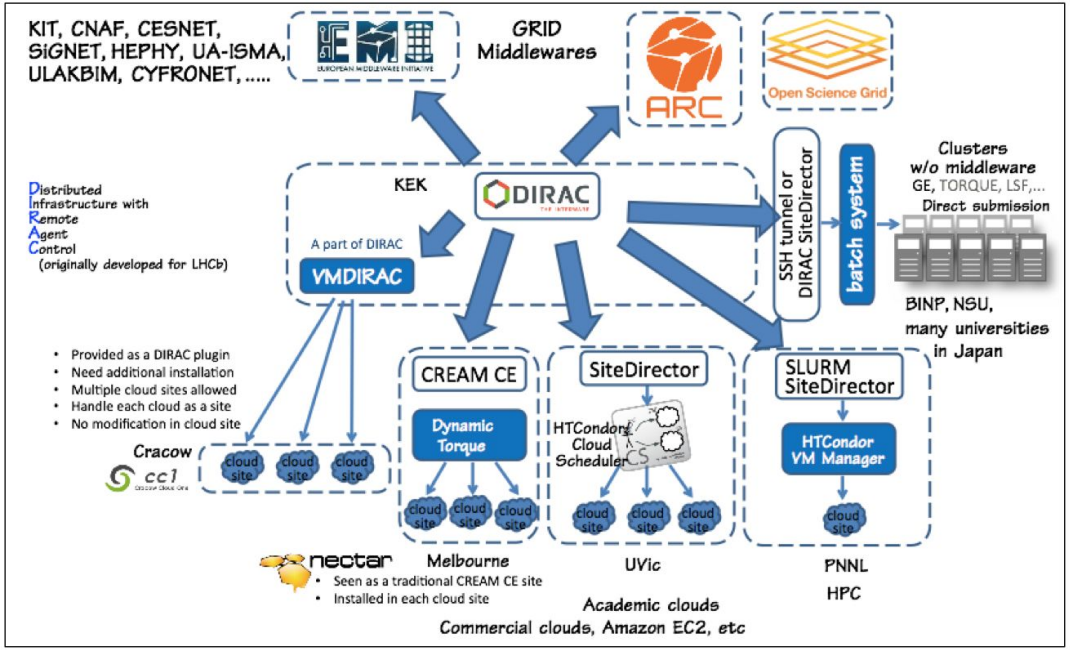
#nodes ~14 Hybrid vs 128 CPU

Price 70x 🍌 320x 🍌



Heterogeneous resources in Belle2

- Nicely use DIRAC (Distributed Infrastructure with Remote Agent Control), existing standard tool within the community



	CY21
CPU [kHEPSpecs]	82.97
Storage [PB]	9.28
Networking In/Out [Gbps]	1.89/0.83

	CY21
CPU @ PNNL [kHEPSpecs]	41.58
CPU @ LCF [kHEPSpecs]	41.39

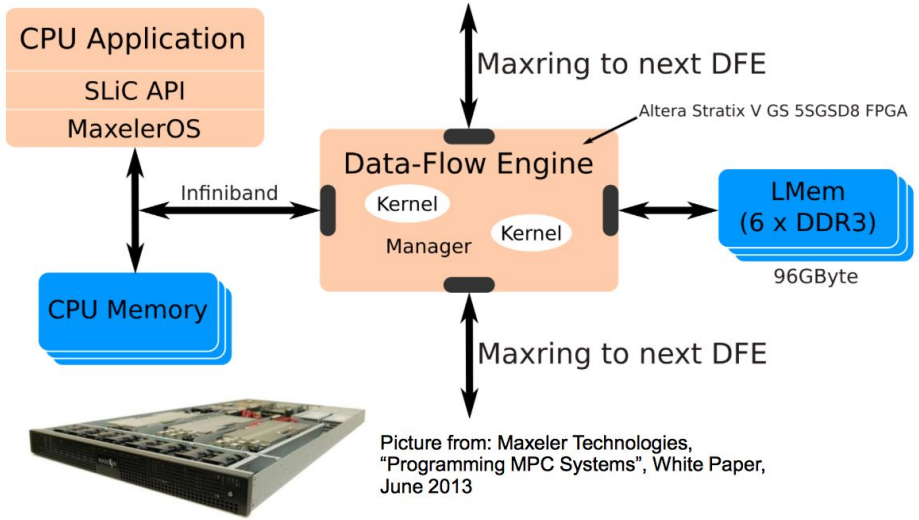
Lattice QCD Computing with FPGA

- Dslash operator can be implemented in a highly parallel and efficient way on an FPGA using the data-flow approach

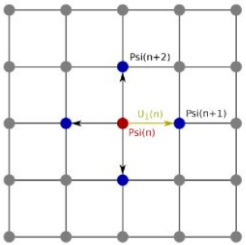
We compute: $\Psi' = (1 - \kappa H)\Psi$

H collects all neighbor terms

$$H = \left(\sum_{\mu=1}^{4} (1 - \gamma_\mu) U_\mu(n)\, \psi(n + \mu) + (1 + \gamma_\mu) U_\mu^T(n - \mu) \psi(n - \mu) \right)$$



Picture from: Maxeler Technologies, "Programming MPC Systems", White Paper, June 2013



Platform	Memory BW	Performance	Perf./BW ratio
OpenSPL FPGA (32 bit fixed-point)	23.64 GByte/s	66 GFLOP/s	2.79
AMD Radeon HD 5870 (DP) [10]	100 GByte/s	60 GFLOP/s	1.2
AMD Radeon HD 7970 (DP)[11]	200GByte/s	120 GFLOP/s	1.2

Round-Table: Using heterogeneous resources for HEP

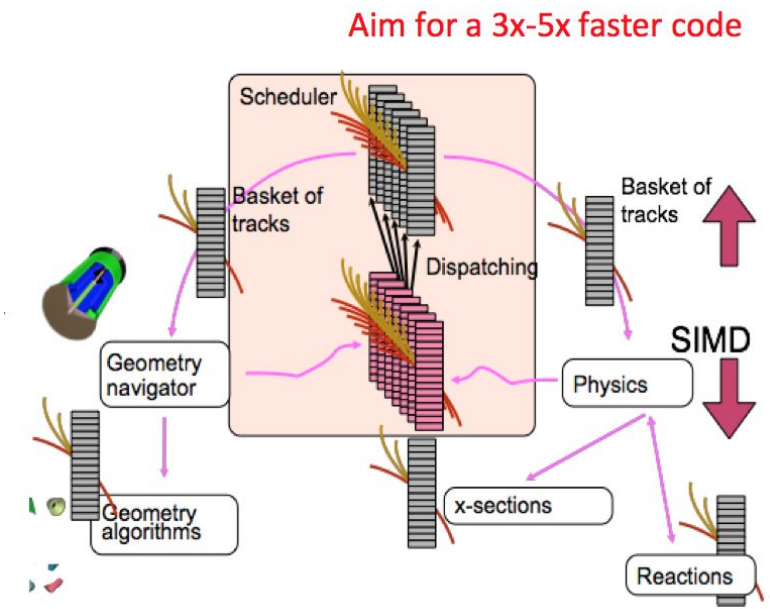
- AI/DL engines are essentially super powerful processors geared at sparse linear algebra
- Should be trying to do what everybody does, because this is where the volume and the low prices are.
- Trend is clearly to standard languages (C++ instead of #pragmas)

Ian Fisk (FNAL) - Chair
Gareth Roy (Glasgow)
Daniel Lo (MicroSoft)
David Lange (Princeton)
Tom Gibbs (NVIDIA Co.)
Jeff Hammond (Intel)

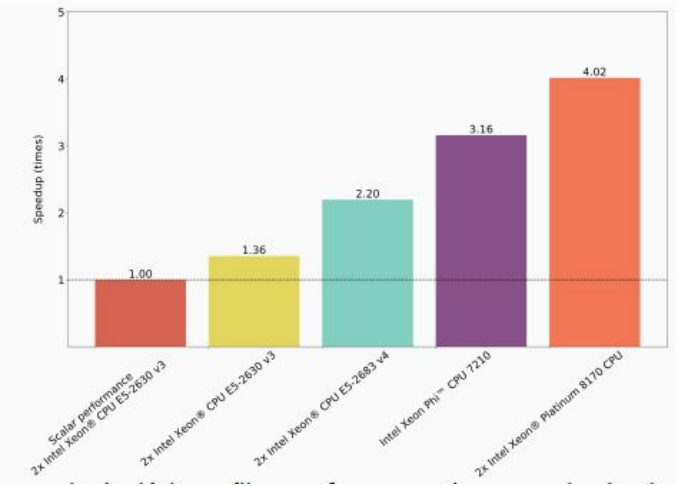
Thanks for enlightening panel discussions!

Clever Optimization: Vectorization

- Faster full simulation & more fast simulation using more efficiently CPU resources.



- Develop cross-architecture Kalman filter for LHCb in modern SIMD and evaluate in 4 variations.

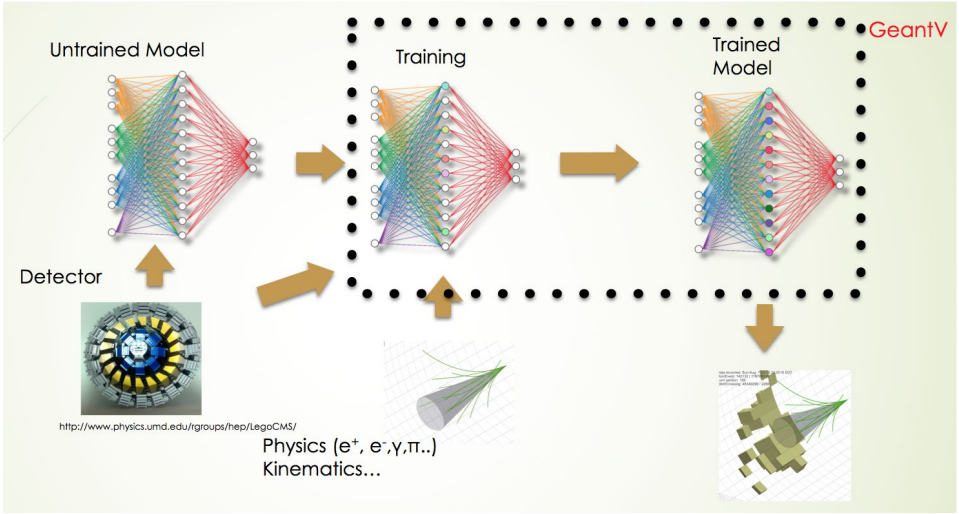


LHCb THCp intel CERN IT Department

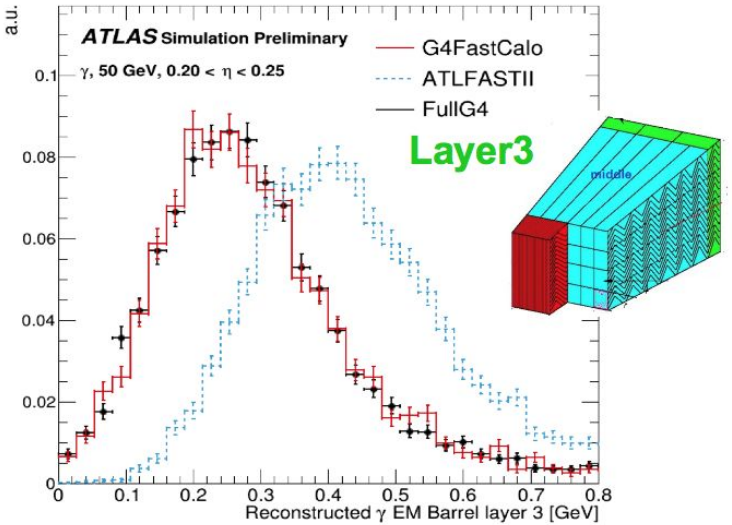
High Throughput Computing Collaboration (HTCC)

Revolutional Optimization: Machine Learning

- Integrate Machine Learning based simulation to GeantV. Fast, modular, and fully configurable.



- ATLAS FastCaloSim based on single particle parametrization using ML techniques.
- 85% simulation spent on CaloSim. Speed up x20



Conclusion

- GRID to Cloud-federation
  Container, Scheuler, to make use of resources of Cloud
- General trend to use more standard, less home-grown stuff,
  cross-experiments, global collaboration with industry
- Using FPGA I/O engine on-line
  a trend to use single customer, small transition layer between frontend and offline DAQ
- Heterogeneous for computing
  Successful application of GPU in on-line trigger and FPGA QCD calculation
- Optimization from clever to revolution
  - Revolutional computing needs new implementation than Virtualization

ACAT Spirit in Action

ACAT LOGO - Aztec feather shield:
how to get out out of this maze (mess) without reinventing the wheel.

- proposed by Denis Perret-Gallix

“We are in even deeper this time!”

- Bruce Denby