Computing Technology for Physics Research

Summary Track 1, ACAT 2013 / Beijing

Axel Naumann, CERN
Jerome Lauret, BNL
Gang Chen, IHEP
David Britton, University of Glasgow

Topics

- Programming Languages, software quality, IDE and UI
- Distributed and parallel computing
- New architectures, many-cores
- Virtualization
- Online monitoring and control, HLT

Contributions

- 6 plenaries
- 23 parallels
- 3 posters
- 14 hours by 27 presenters

Coverage

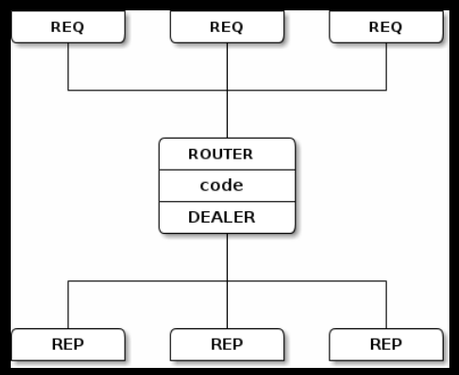
Prog langs, software quality, IDE and UI

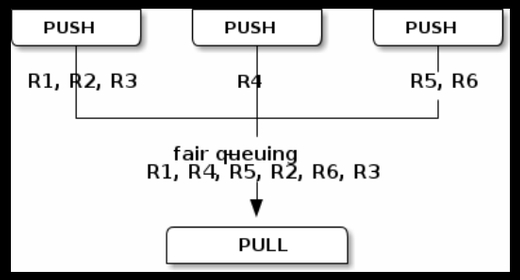
- Rieger: web-based collaborative data analysis
- Tkachov: Oberon
- This topic is consistently underrepresented

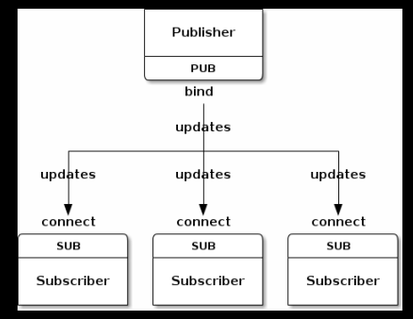
Distributed and parallel computing - Data

- al-Turany: Flexible data transport for online reconstruction @ FAIR

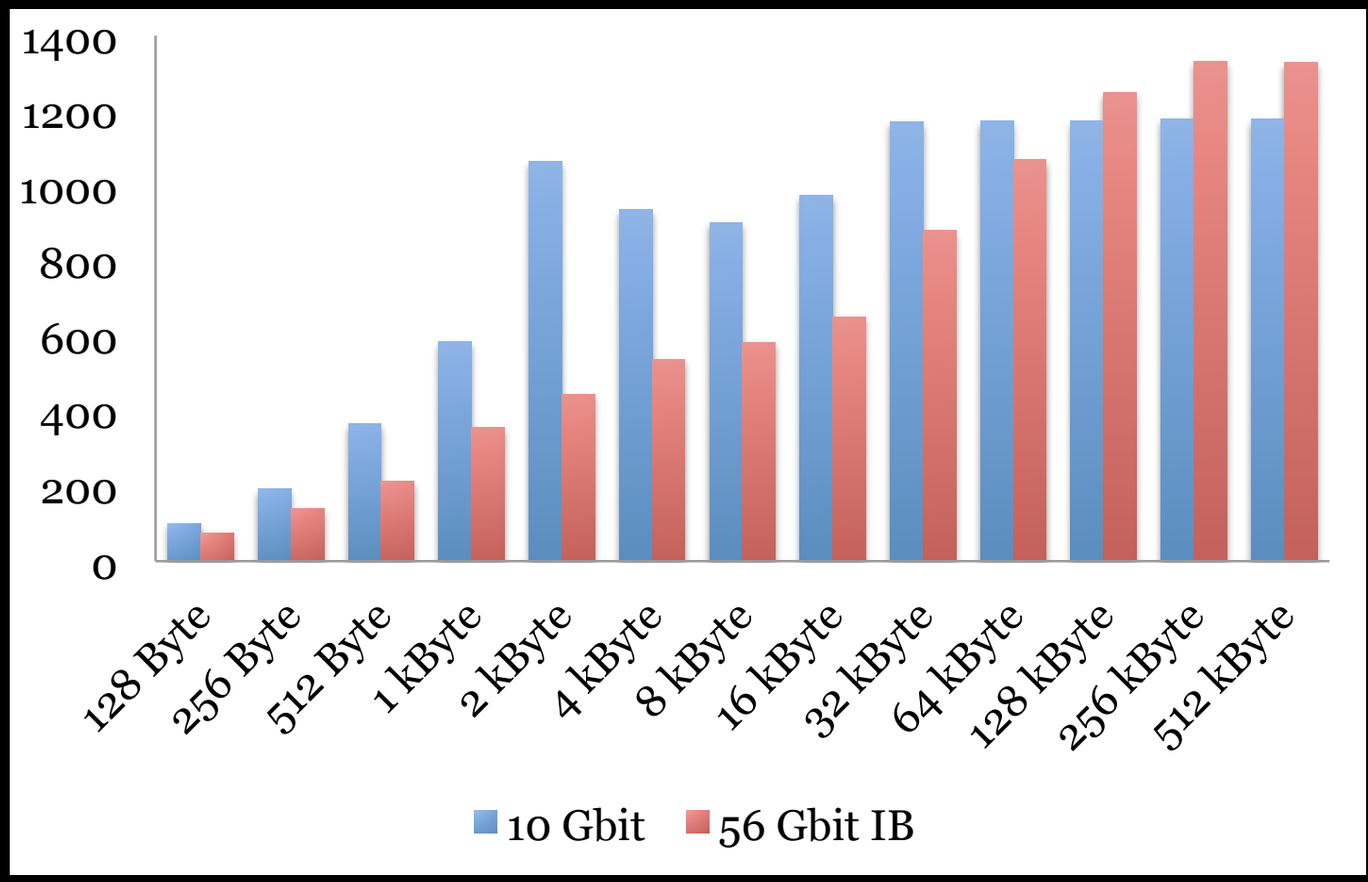
0MQ







- Request/Reply
- Publish/Subscribe
- Pipeline
- Exclusive Pair

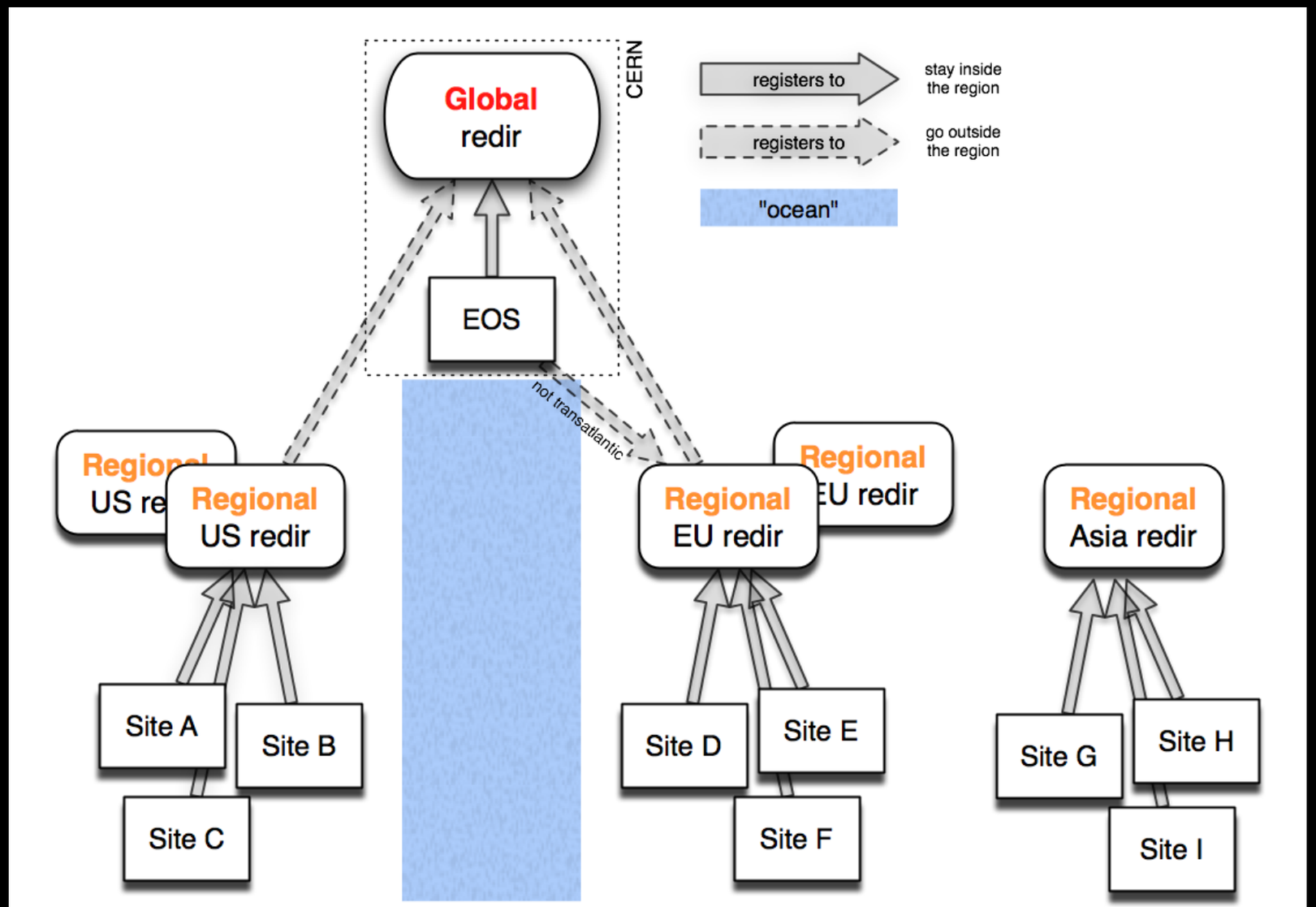


Distributed and parallel computing - Data

- al-Turany: Flexible data transport for online reconstruction @ FAIR
- Elmer: CMS use of a data federation

Adding I/O Paths

- Fallback, interactive, remote CPU overflow



Distributed and parallel computing - Data

- al-Turany: Flexible data transport for online reconstruction @ FAIR
- Elmer: CMS use of a data federation
- Gheata: Performance optimization for distributed analysis at ALICE
- Horsky: Influence of distributing data storage on physics analysis

Distributed and parallel computing - Data (2)

- Makatun: Cache performance in distributed environment for data processing

Cache Eviction

What caching algorithm is the best?

Average improvement over FIFO

Algorithm	cache hits	cache data hits
MS	116 %	-20 %
LRU	8 %	5 %
ARC	13%	11%
LVCT	86 %	2 %

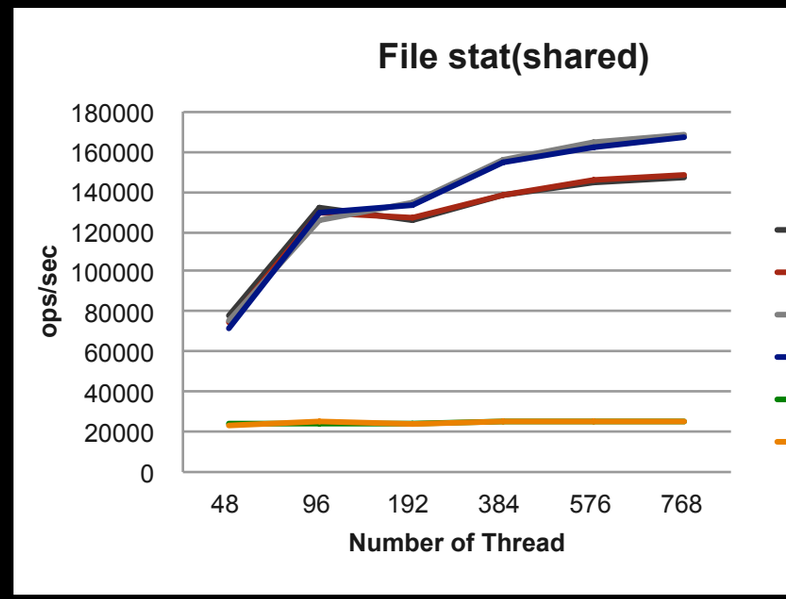
For studied access patterns

- MS has the best cache hits performance but the worst cache data hits
- ARC has the highest cache data hits
- LVCT balances between cache hits and cache data hits

Distributed and parallel computing - Data (2)

- Makatun: Cache performance in distributed environment for data processing
- Liu: HPC at BESIII (tag)
- Sangaline: Experience, use, performance of Hadoop FS in nuclear physics analysis
- Triendl (DataDirect Networks): Storage architectures for large data-analysis systems

Distributed Filesystem



- Forget NFS
- Much improved Lustre 2.3/2.4
  - I/O hierarchies; checkpointing writes versus global update - POSIX kills us
  - Look out for Fast Forward for Exascale I/O

Distributed and parallel computing

- Grigoras: JAliEn - interface between AliEn jobs and central services
- Nairz: ATLAS distributed computing

New architectures, many-cores

- Elmer: Initial explorations of ARM processors for scientific computing
- Gheata: Concurrent vector-based steering framework for particle transport
- Kanzaki: Event generation using GPU
  - more in cooperation with track 3?
- Meng: TianHe-1A, HPC in China

Tian-He, HPC in China

- Long tradition Chinese supercomputing
- Culminated in Tian-He 1A
  - 14K Xeons
  - 7K Teslas
  - 2K SPARC-based CPUs
  - 260GB RAM

New architectures, many-cores

- Nowak: Opportunities and choice in a new vector era

Inhomogeneous Everything

What should next-generation code look like?

Balance of programmability and performance	• Can't entirely sacrifice one for the other • Can't ignore either
Using universal standards, languages and compilers	• Sustainability, interoperability
Scaling in multiple dimensions	• Parallelization, vectorization, multi-process domain, message passing readiness, "asynchronicity", lack of coherency • Good domain decomposition a prerequisite
Power efficient	• Optimal balance of data movement and compute

New architectures, many-cores

- Nowak: Opportunities and choice in a new vector era
- Oyanagi: Lessons learned from Kei

Kei and the World

- History of HPC incl. Japan
- K brought Japan back with Parallel computing
- HPC and being number one
- Applications, applications
- Memory hierarchy makes programming on Exascale very hard

	0206	0211	0306	0311	0406	0411	0506	0511	0606	0611	0706	0711	0806	0811	0906	0911	1006	1011	1106	1111	1206	1211
1	ES	ES	ES	ES	ES														K	K		
2																					K	
3						ES																K
4							ES											T IT				
5																			T IT	T IT		
6			NAL					ES														

New architectures, many-cores

- Nowak: Opportunities and choice in a new vector era
- Oyanagi: Lessons learned from Kei
- Vicini: GPU for real-time processing in HEP trigger systems
- Vicini: Host and GPU interface in 3D torus network

Virtualization

- Bagnasco: Tier-2 with a private cloud
- Bird: Grid and cloud evolution from now onward

Grid + Cloud

- Move towards Cloud is happening
- Integration into grid still needed, e.g. for federations
- Private cloud is baseline; buying cloud hours costly + needs standards
- HEP Cloud will benefit from being integrated into other science fields, too

Virtualization

- Bagnasco: Tier-2 with a private cloud
- Bird: Grid and cloud evolution from now onward
- Colling: CMS HLT farm as a cloud resource

Online monitoring and control, HLT

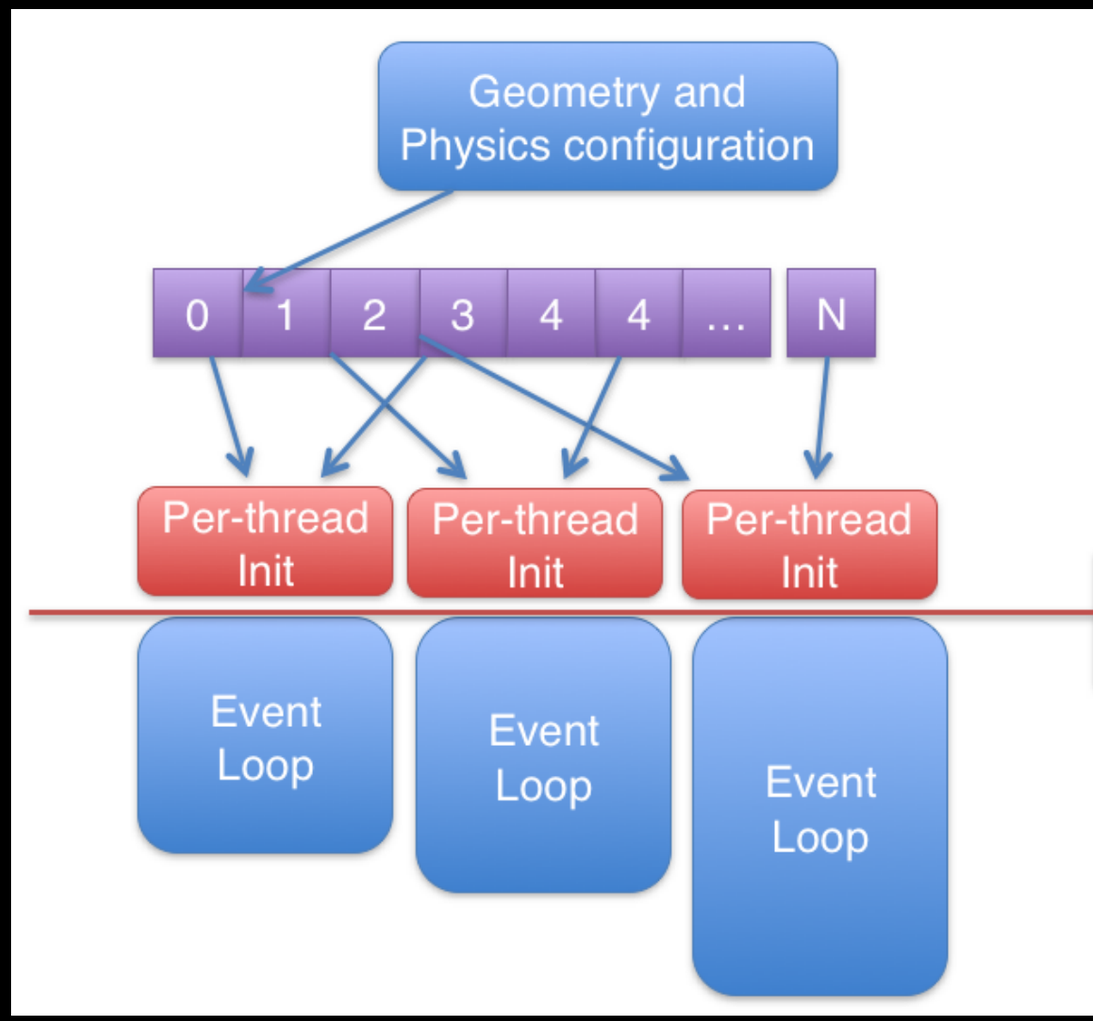
- Dell'Asta: ATLAS muon and tau trigger
- Ionita: ALICE expert system
- Krasznahorkay: Trigger and DAQ in ATLAS
- Krasznahorkay: Data quality monitoring for ATLAS trigger
- Liu: Performance quality monitoring for the Daya Bay reactor neutrino experiment

[This drawer has no name]

- Al-Turany: FairRoot framework
- Elmer: Checkpoint-Restart for HEP software

Checkpoint / Restore

- Pause button for programs
  - e.g. debugging
- Problem is resuming
  - best in controlled environment
- Reduce startup



[This drawer has no name]

- Al-Turany: FairRoot framework
- Elmer: Checkpoint-Restart for HEP software
- Lopez Villarejo: Documenting through activity diagrams
- Luzzi: TGeoCad interface between ROOT and CAD systems

Current Main Topics

Get Data From A to B

- Solid engineering
  - Where is B?
  - How much data
  - In what format (file system,...)
- Measurements!
- Not new, but now without dreaming

Get Data From A to B (2)

- Data structures: optimizing for I/O?
- Data transfer: more connections, more heterogeneity, network as a resource
- Data cache: multi-level, complex
  - inhomogeneous elements (+file system!)
  - file level versus partial read

Groud

- Virtualization as a tool is understood
- Grid is becoming cloudy
  - Grid concept likely not replaced
  - looks like a grid, managed as clouds
- We are just another entity with lots of data
  - Re-use, encourage standards (!= EC2?)

O those CPUs

- They don't do what we want. Thus:
  - data-driven parallelism
  - ever growing toolbox
  - vectorization: language? features? architecture? cost? benefit? user tool?

O those CPUs (2)

- We don't master Intel CPUs. Let's look at others (that we also won't master):
  - GPGPUs, MIC: likely niche but obligatory
  - Heterogeneous System Architecture (!)
  - ARM

Conclusion

Thank you, IHEP!

- Ideal conference infrastructure
- Perfect organization
- Wonderful local team

Thank you, presenters!

- Many **excellent** presentations
- Good discussions during the sessions
- Reflections of computer scientists, experimental and theoretical physicists
- Some new seeds were planted - let's water them and inspect them at ACAT 2014!