CHEP'15 Program Committee: defining the Tracks

I. Ueda, A. Lyon, D. Bonacorsi

chep2015-pc-chair@icepp.s.u-tokyo.ac.jp

21 Nov 2014

CHEP2015
OKINAWA, japan

21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 – 17, 2015

First of all..

Great work done so far. Thanks everyone.

One note.

✦ Discussion in several threads not unexpectedly went into the content and quality of the abstracts. Let's focus on the definition of tracks now. Once this step is DONE, we move to populate the tracks with abstracts and assess the abstracts themselves, negotiate among the different track conveners, etc. Starting to look at the abstract already at this step will help to speed up the next step, I presume, which is good.

Reminder of CHEP'13 tracks

Data Acquisition, Trigger and Controls

✦ event building and farm networks; compute farms for high-level triggering; configuration and run control; describing and managing configuration data and conditions databases; online software frameworks and tools; online calibration procedures; remote access to and control of data acquisition systems and experimental facilities.

Event Processing, Simulation and Analysis

✦ event generation, simulation and reconstruction; detector geometries, physics analysis; tools and techniques for data classification and parameter fitting; event visualization and data presentation; frameworks for event processing; toolkits for simulation, reconstruction and analysis; event data models.

Distributed Processing and Data Handling ← This was split in 2 tracks by the PC

✦ grid computing; virtualization; infrastructure as a service; clouds; distributed data processing; data management; distributed analysis; distributed processing experience, including experience with grids and clouds; experience with production and data challenges; experience with analysis using distributed resources; interactive analysis using distributed resources; solutions for coping with a heterogeneous environment; mobile computing; monitoring of user jobs and data; grid and cloud software and monitoring tools; global usage and management of resources; middleware reliability, interoperability and security; experiment specific middleware applications.

Data Stores, Data Bases, and Storage Systems

✦ storage management; local I/O and data access; mass storage systems; object dictionaries; event stores; metadata and supporting infrastructure; databases; access patterns and caching strategies; data preservation; data curation and long-term data reproducibility.

Software Engineering, Parallelism & Multi-Core

✦ CPU/GPU architectures; tightly-coupled systems; GPGPU; concurrency; vectorization and parallelization; mathematical libraries; foundation and utility libraries; programming techniques and tools; software testing and quality assurance; configuration management; software build, release and distribution tools; documentation.

Facilities, Production Infrastructures, Networking and Collaborative Tools

✦ basic hardware, benchmarks and experience; fabric virtualization; fabric management and administration; local and wide-area networking; private networks; collaborative systems: progress in technologies and applications; tele-presence and teleconferencing systems; experience in the use of teleconferencing tools.

A (part 1)

A01: ~45 abstracts

- ✦ Ryosuke noted close correlation with A02 and T01. Same observed by other KCs..

A02: ~42 abstracts

- ✦ Andrea found >75% relevant to this keyword. Correlation with A01, less strong with A04, T05, T18. Suggesting to merge A01+A02 to reach ~6 final CHEP tracks. , by moving abstracts more related to physics performance to a "reconstracution" track, or a track in which this would be a subtopic

A03: ~48 abstracts

- ✦ Hisaya noted that about 75% of the total seems to be indeed in this category. Strong correlations with T02 and A04. Could be considered together?

A04: ~46 abstracts

- ✦ Ivan noted that a traditional merging A03+A04+A05 could be done. On the other hand, with growing parallelism, the participants of this track track would be interesting to visit the hardware related tracks as well. Maybe parallelise with care?

A05: ~85 abstracts

- ✦ Andrew noted that ~50% only were indeed ok with this keyword. If this becomes a track: 40ish abstracts. If not, most sensible merges would be: A05+T02/frameworks (77 abstracts), A05+T03/algorithms (63 abstracts), A05+A04/Reco (73 abstracts). Or even merging down further: A05+T03+T12 (74), A05+A04+A03 (91), A05+A11+T02 (100). Q (Daniele): are we paying here the price that the "data analysis" keyword was too generic? To be discussed.

A07: ~76 abstracts

- ✦ Takanori found ~34 well described by this keyword. It covers a wide range of topics. Correlations with T04 and T07, depending on the abstract. Also, A11 ius used sometimes as same meaning of A07. Suggestion is A07+A11+T04+T07. Q: is this doable, it will become a very crowded track. To be discussed.

A08: ~85 abstracts

- ✦ >70% relevant. Reshuffling somehow with A07, A13, T07 would work. To be discussed.

A09: ~23 abstracts

- ✦ >90% properly classified. Eric suggested to have a (smaller) separate track. Also social dynamics aspects, uncovered elsewhere. Should not be in parallel with T15+T16 though. Alternatively, deal with them into T15+T16, but it already crowded. To be discussed.

A10:

- ✦ due to the nature of the keyword, all abstracts tend to belong to some other keyword also. Hence, no need for a specific track. Tadashi offered a list of possible topical areas which each abstract should be absorbed to.

A11: 49 abstracts

- ✦ Simone observed that tracks should be defined based on A keywords, and T keywords should be used mainly to organise the track internally
- ✦ Simone proposed a set of tracks, which is included in the summary at the end
  - so, no detailed summary here is needed

A12: 25 abstracts

- ✦ Mike tagged <50% to be fully relevant to Data/Software Preservation. As discussed already, it is a good candidate for a small, separate track in itself (with e.g. 1 afternoon of parallels only, or so).

A13: 57 abstracts

- ✦ detailed classification, and suggestion about which areas they should belong to. While many cover monitoring topics, shouldn't they be better addressed in specific topical areas? To be discussed.

T (part 1)

T02: ~75 abstracts

- ✦ two scenarios suggested. 1. group is by the function that is being performed by the framework (reco/sim or analysis). 2. group by the area that they run in (offline versus online environment). Preference for 1. Full description of subtopics provided for each scenario.

T04: ~90 abstracts

- ✦ very crowded as a keyword. Do we need a track, or may this fall into different areas and what they mean in their context by "data handling"? <u>To be discussed</u>.

T05: ~39 abstracts

- ✦ Gancho found ~50% of relevant abstracts. Maybe a "Databases, data structures, data handling/access and analysis" would work? <u>To be discussed</u>.

T08: ~60 abstracts

- ✦ Tony found about 2/3 of the abstracts indeed relevant to the keyword. Probably a track to be hosting also abstracts from neighbouring ones?

T09: ~31 abstracts

- ✦ 50% well classified. Maybe merge into a track with e.g. T11..

T10: ~31 abstracts

- ✦ Marco got the impression the keyword was misinterpreted in many cases. Probably a merge with other keywords is not complicated here as well.

T (part 2)

T11: ~15 abstracts

- ✦ ~4 of them seem to be very T11-specific, others could go into areas like A08 or T07. No specific track needed.

T13: ~42 abstracts

- ✦ (roughly) 50% on networks indeed and 50% on neighbouring topics. Focus on the former and drop the latter to more natural tracks. Correlation with T07 is observed. Options: 1. a network specific tracks to relieve others, a "smaller" session in Okinawa; 2. collapse this into a "facilities" track and gain from a larger audience. To be discussed.

T15+**T16**: roughly 60 abstracts (or so)

- ✦ we had no reply from the selected KC for T15, Andrew did a great job in T16 though and it makes sense to propose a joined track
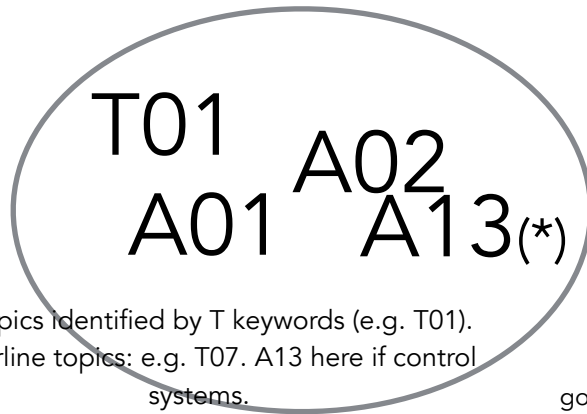
T18+**T17**+**T12**: ~32 abstracts (as from Niko's count)

- ✦ we have separate spreadsheets by Niko/Amitoj/Chris, with abstracts "graded" in terms of relevance
- ✦ the feeling is that T18 would worth a track, possibly joined with T17. No need for a track in itself for T12 so this could well be joined as well
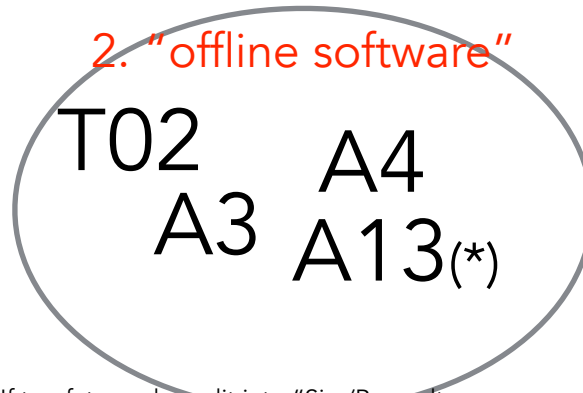
Working slide for the meeting
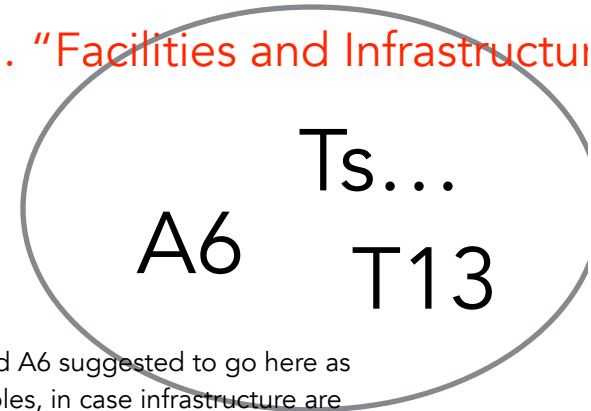
1. "online computing"

T01
A02
A01
A13(*)

Subtopics identified by T keywords (e.g. T01). Borderline topics: e.g. T07. A13 here if control systems.

2. "offline software"

T02
A4
A3
A13(*)

If too fat, maybe split into "Sim/Reco algorithms" and "Sw engineering/techniques". A13 here if it is e.g. mem usage monitoring for multiprocess applications.

5. "Facilities and Infrastructur

Ts…
A6
T13

T13 and A6 suggested to go here as examples, in case infrastructure are discusses, otherwise in "Middleware.."

3. "Middleware, experiment frameworks and tools for (distributed) computing"

A7
A8
A13(*)
A9(*)
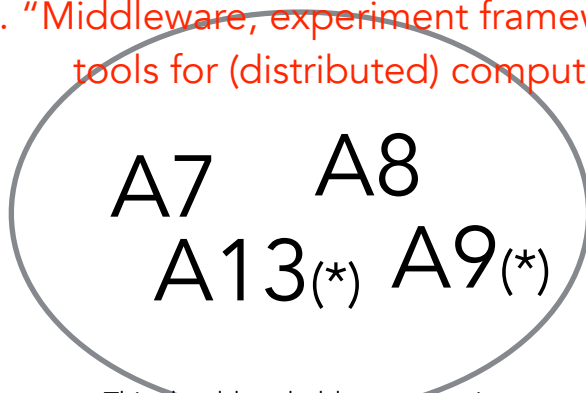
This should probably not contain any operational experience. A9 goes here as "tools". A13 here if it is e.g. monitoring a distributed tool.

4. "Computing activities and computing models"

A9(*)
A5
A10
(A12)
A13(*)
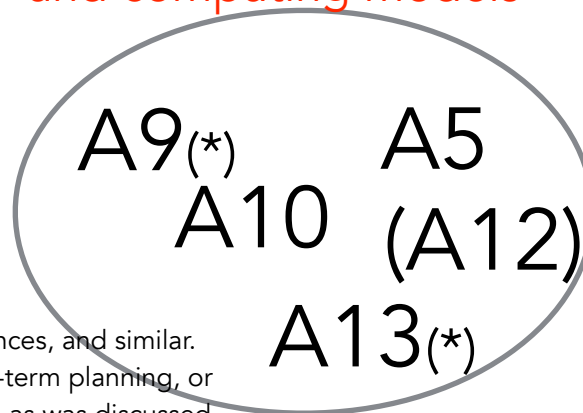
E.g. LHC Run-1 experiences, and similar. A12 may go here as long-term planning, or be a separate small track, as was discussed last time. A9 with volunteer computing may go here. A13 here if it is monitoring for site commissioning (e.g. HammerCloud).

6. Cloud and Virtualisation

T16
T15
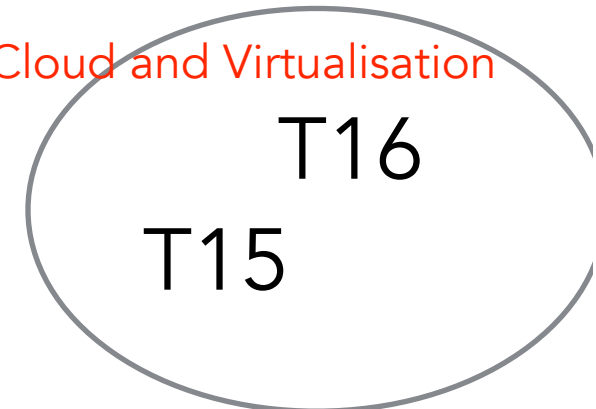
7. "Performance increase using architecture features and co-processors" or "Optimizing for Specialized Hardware"

T18
T12
T17

draft tracks

Online

Offline software

Facilities, Infrastructure, Network

Cloud, virtualisation

Data store and access

✦ A06 + T04 + T06

Middleware, experiment frameworks and tools for (distributed) computing

Computing activities and Computing models

Exploiting hardware features

This slide was added as "working material" during the meeting

Back up material

I. Ueda, A. Lyon, D. Bonacorsi

Application keywords conveners

Application keywords	name	region	affiliation	exp
A01 DAQ	Ryosuke Itoh	AP	KEK	Belle2
A02 Trigger	Andrea Bocci	EU	CERN	CMS
A03 Simulation	Hisaya Kurashige	AP	U. Kobe	ATLAS
A04 Reconstruction	Ivan Kisel	EU	Frankfurt/GSI	none
A05 Data analysis	Andrew Norman	US	FNAL	NOvA
A06 Data stores	Latchezar Betev	EU	CERN	ALICE
A07 Experiment frameworks for WAN distributed computing	Takanori Hara	AP	KEK	Belle2
A08 Middleware and services for production-quality infrastructures	Nicolo Magini	EU	CERN	CMS
A09 Outreach	Eric Yen	AP	ASGC	none
A10 Multi-discipline / multi-experiment topic	Tadashi Maeno	US	BNL	ATLAS
A11 Computing models	Simone Campana	EU	CERN	ATLAS
A12 Data Preservation	Mike Hildreth	US	U. Notre-Dame	CMS
A13 Monitoring	Pepe Flix	EU	PIC	CMS

Technology keywords conveners

Technologies Keywords	name	region	affiliation	exp
T01 Control systems	Jim Patrick	US	FNAL	none
T02 Event processing frameworks	Jim Kowalkowski	US	FNAL	art framework
T03 Data structures and algorithms	James Catmore	EU	U. Oslo	ATLAS
T04 Data handling/access	Christopher Pinkenburg	US	BNL	PHENIX
T05 Databases	Gancho Dimitrov	EU	CERN	none
T06 Storage systems	Alastair Dewhurst	EU	RAL	ATLAS
T07 Computing facilities and infrastructures	Kihyeon Cho	AP	KISTI	Belle2/ALICE
T08 Software design	Tony Wildish	US	U. Princeton	CMS
T09 Software development process	Vincent Garonne	EU	CERN	ATLAS
T10 Performance and validation tools	Marco Clemencic	EU	CERN	LHCb
T11 Continuous integration systems	Brett Viren	US	BNL	none
T12 Parallel programming	Chris Jones	US	FNAL	CMS
T13 Networking	Phil Demar	US	FNAL	none
T14 Collaborative tools	Peter Hristov	EU	CERN	ALICE
T15 Cloud computing	Glenn Moloney	US	U. Melbourne	none
T16 Virtualisation	Andrew McNab	EU	U. Manchester	LHCb/ATLAS
T17 High performance computing	Amitoj Singh	US	FNAL	lattice-QCD
T18 CPU architectures, GPU, FPGA	Niko Neufeld	EU	CERN	LHCb