A brief review of the conference Computing in High Energy and nuclear Physics - CHEP 2013

Amsterdam, 14-18 October 2013

http://www.chep2013.org/

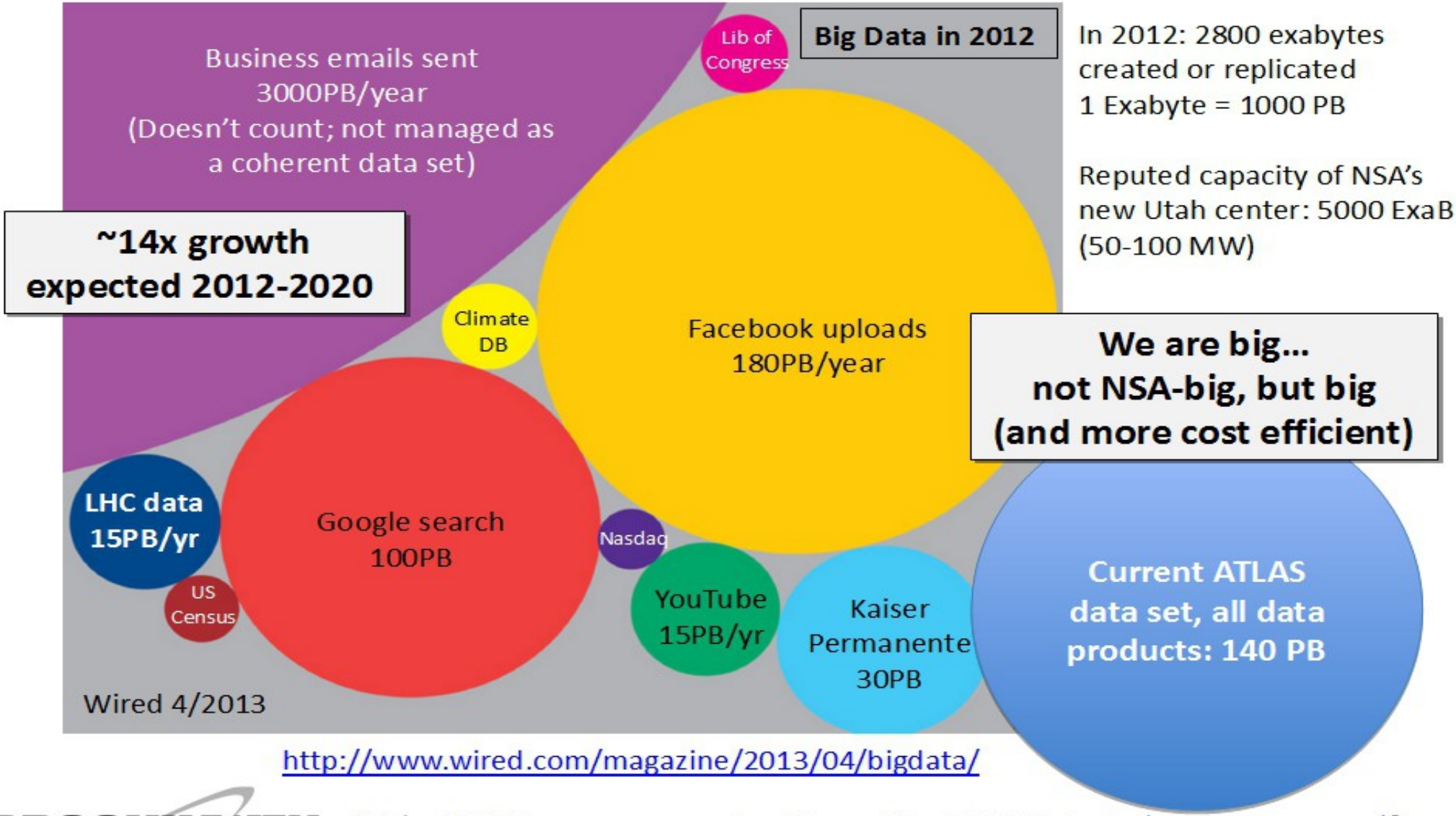
~500 participants have been registered

Andrey Y Shevel

1. Intro on Big Data
2. Data acquisition, trigger and controls.
3. Event processing, simulation and analysis.
4. Distributed processing and data handling.
    1. 3A and 3B
5. Data stores, data bases and storage systems.
6. Software engineering, parallelism and multi-core programming.
7. Facilities, production infrastructures, networking and collaborative tools.
8. CHEP2013 trends
9. Spare slides

Data Management

Where is LHC in Big Data Terms?



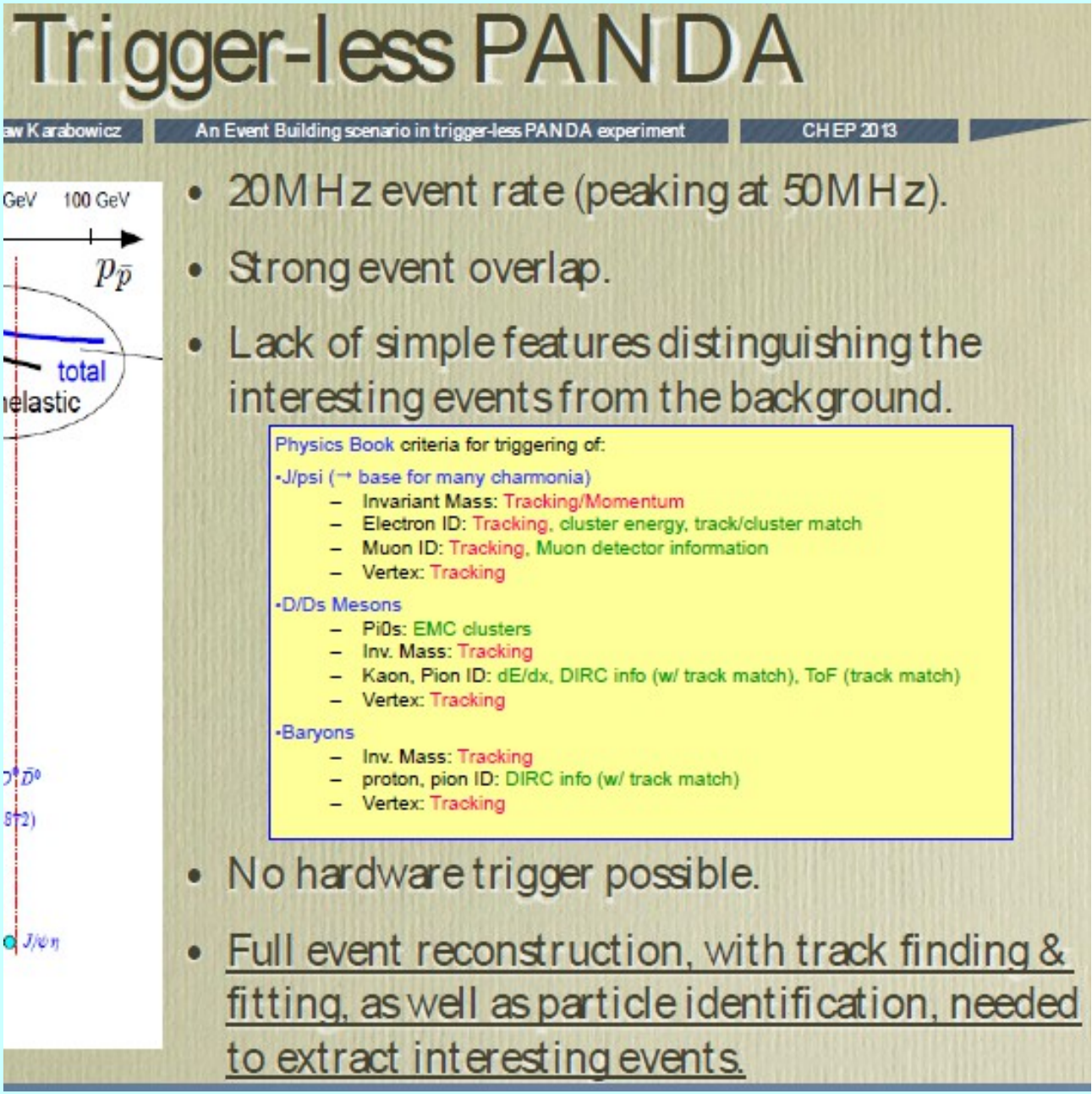
http://www.wired.com/magazine/2013/04/bigdata/

Data Volumes

Decimal		
Value		**Metric**
1000^1	KB	kilobyte
1000^2	MB	megabyte
1000^3	GB	gigabyte
1000^4	TB	terabyte
1000^5	PB	petabyte
1000^6	EB	exabyte
1000^7	ZB	**zettabyte**
1000^8	YB	yottabyte

Track1: Data acquisition, trigger and controls

- 53 submissions
  - 25 posters



DAQ around the World

Belle - http://en.wikipedia.org/wiki/Belle_experiment

NOVA - http://www-nova.fnal.gov/

MICE - http://mice.iit.edu/

PANDA - http://www-panda.gsi.de/

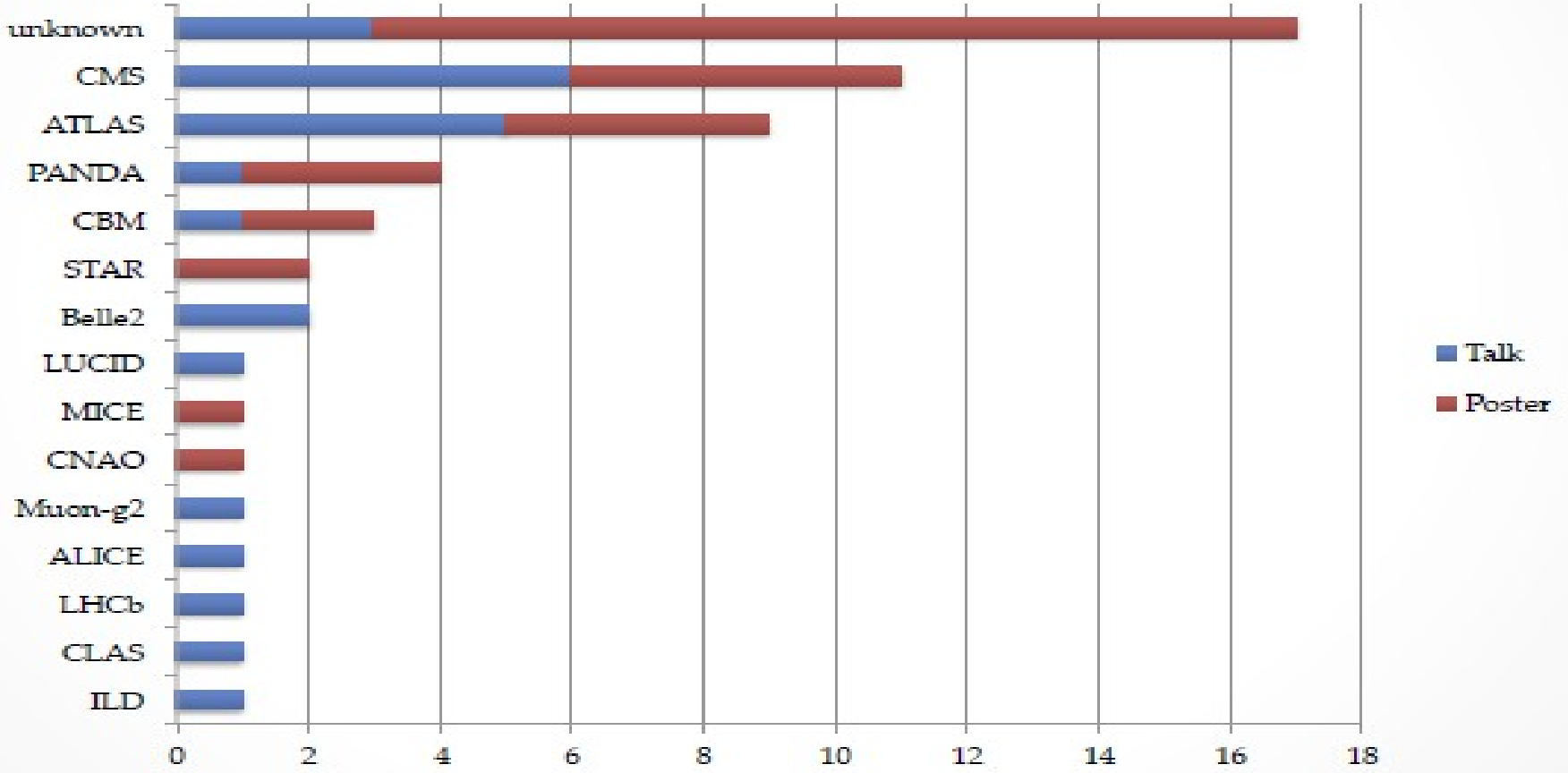
CTA - http://www.cta-observatory.org/

H.E.S.S. - http://www.mpi-hd.mpg.de/hfm/HESS/pages/collaboration/

Icecube - http://icecube.wisc.edu/

Track2: Event processing, simulation and analysis

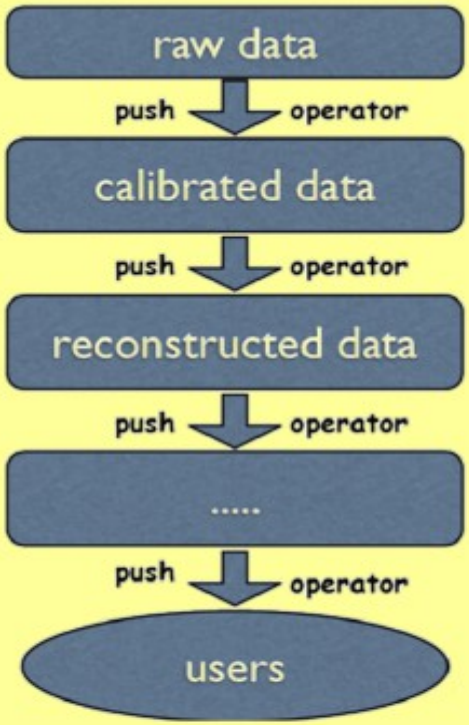
Which experiments are represented?



Alternative ways for data processing and mining?

not only HEP experiments deal
with larga amount of data

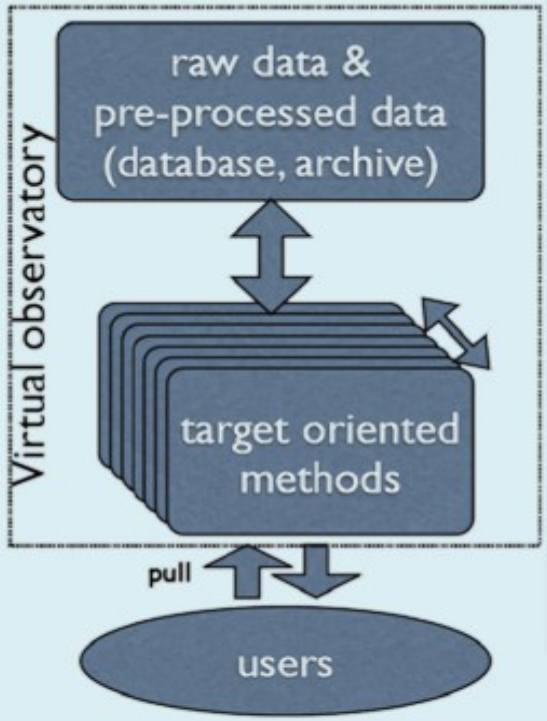
The *Waterfall* model



Forward chaining
"Tier" architecture
Driven by raw data
Process in pipeline
Operators push data
Results in release
Static archive
Raw data is obsolete

Standard in HEP

The *Target* model



Backward chaining
"Target" architecture
Driven by user query
Process on-the-fly
Users pull data
Information system
Dynamic archive
Raw data is sacred

Used in Astronomy

Edwin Valentijn et al. (Astro-WISE, target)

Could it work also for HEP?

Track 3: distributed processing and data handling

- 142 accepted submissions
- uCERNVM - bootloader with client CVMFS which loaded later on whole OS
- ROOT@HADOOP (Hadoop is a free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation).
- Rare re-use of the developments from one collaboration in another one

Track3: Data handling summary

- Data handling and processing are becoming more and more challenging with increasing complexity of scientific experiments.
- Interest is growing among other big data scientific communities to use HEPdriven workload management systems. Efforts on making these systems generic (DIRAC, PanDA) for any user community.
- LHC experiments are facing the challenges of the upcoming run. Computing models are evolving to meet the requirements; new data and workload management systems are being developed. Adaptations to computing models to support use of multicore jobs, federated storage infrastructure, cloud computing and opportunistics resources.

Track4: Data Storages, Data Bases, and Storage Systems

- Filesystems
- Data preservation
- Storage systems
- Data management systems
- Databases
- Metadata services
- IO and Data structures.

Own cloud storage

- http://www.et-js.org/ - elastic transfer
  - Create your Personal Storage Cloud
  - Directly, transfer your files from your workstation to an other PC
  - Third-party Data transfer
  - Flexible data & storage sharing

Track4: summary

- Many areas of evolution: filesystems, storage systems, data management services and federations, database technologies and approaches.
- New, and established solutions, are being built/developed incorporating standards and flexibility to change - Good news !
- Many studies ongoing to optimise our storage and data systems and the performance is good ... but is it good enough for I/O challenges to come?

Track5: Software engineering, parallelism & multi-core programming

- 56 submissions (26 are oral)
  - Main topics
    - Beyond x86 servers
    - Vectorization
    - Concurrency
    - C++11
    - Software Engineering

Software engineering, parallelism & multi-core programming - continuation

- To improve performance per watt
- Alternative platforms & co-processors
  - Porting or Reimplementing ?
    - ARM
    - Xeon Phi
    - GPU (many examples — track fitting, even GEANT)

Concurrency

Adapt applications and application frameworks to many-core systems to exploit different sources of parallelism.

- Multiple events
- Within an event
- Within an algorithm

18/

New standard C++11

For Experiments

- Don't convert to a new language standard, but prepare for a continuous standard and compiler delivery process
- Benefit from safer C++
- Benefit from better compilers

For Physicists

- There is nothing to be done *by them* - they should not need to act
- ROOT, Geant, frameworks should demonstrate the advantage of simple code clear ownership, improved standard library

```
TH1::AddFunction(std::unique_ptr<TF1>)
```

- C++11 and after brings us closer to the ultimate goal:
  - Write correct code and analyses easily!
  - From data taking to physics result quickly!

Software engineering

LHC experiments are using the Long Shutdown to optimize code and infrastructure.

– Optimizing

– Profiling

– Quality Assurance

– Integrated tools.

- Particular attention to release and build management.
- LHCb, ATLAS & CMS (again to move to OpenSource codes, e.g. github, etc)

Track6: Facilities, production infrastructures, networking and collaborative tools

82 submission were accepted (28 are oral)

- Facilities — 6 (monitoring, comparison, etc)
- Production infrastructures — 50
- Networking — 15
- Collaborative tools - 10

Cost of production infrastructure

- Cost of computing/core at dedicated data (comparison BNL vs EC2) centers compare favorably with cloud costs
  - $0.04/hr (RACF) vs. $0.12/hr (EC2)
  - Near-term trends for costs
    - Hardware (goes down)
    - Infrastructure (goes up)
    - Staff (stable)
    - Data duplication (goes up)
- Data duplication requirements will raise costs and complexity – not a free ride

Additional CERN cluster at Hungary (they won open tender)

- First tests are promising
- Large scale deployment 2014-2015
  - But work still required to finalise operational procedures

Production infrastructure

- Cloud, cloud,... e.g. ATLAS HLT cluster is now used as cloud for period when it is not used in data taking
- Public and hybrid clouds (e.g. Amazon EC2 and Google Computer Engine)
- Agile (at CERN) based on many OpenSource components (OpenStack, puppet, git, etc)
- Hardware is growing at CERN by pace +100nodes/week

Virtual infrastructure

- Now you can create the VM (or group of them)
- Configure
- Test and run
- Cleanup and destroy after you do not need it anymore

Networking

- PerfSONAR
- PerfSONAR-PS (both testing frameworks for testing the WAN every 6 hours intra-region, every 12 hours T2-T1 inter-region, and once a week elsewere)
  - And other tests for WLCG network.
- Mass testing for 100 Gbit lines in many participants.
- Many labs do test IPv6

Collaborative tools

- Indico 1.0+ indico.cern.ch

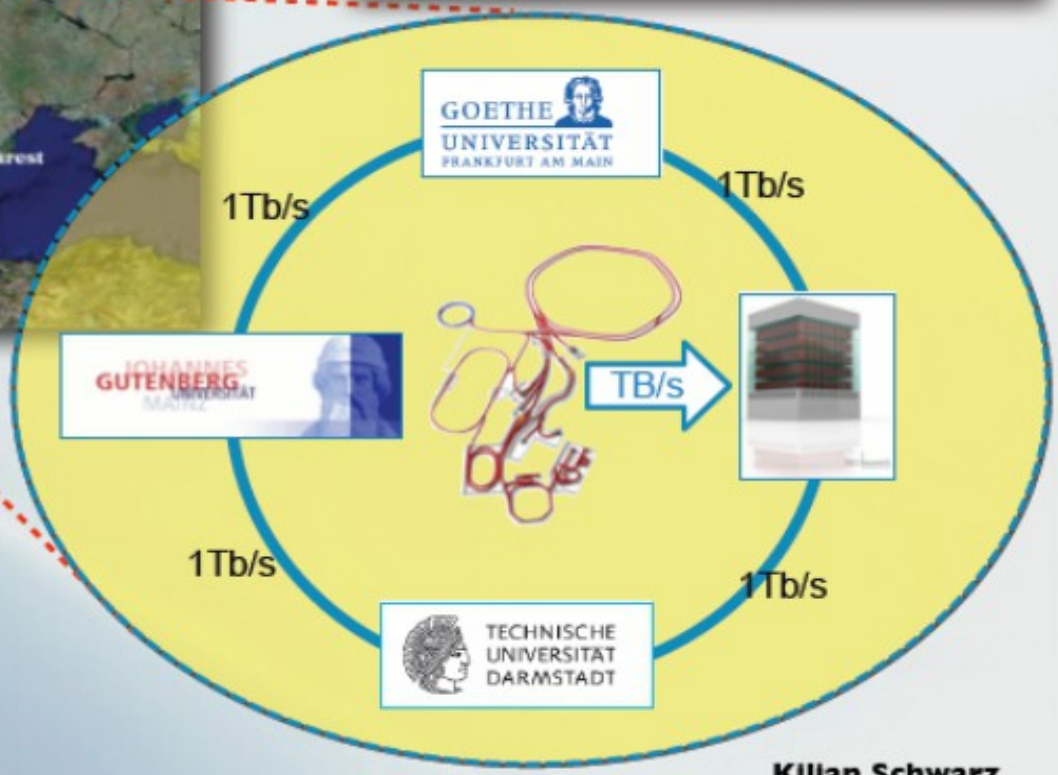
CHEP 2013 Trends

- More special computing devices: GPU, FPGA.
- Clouds, clouds, … and Open Source
- More attention for experimental data preservation and computing/analysis results be re-produced (even EU committe - Kostas GLINOS).
- Plans for separate channels for data and for common Internet activity (FNAL, CERN).
- Plans for private dropbox-like storage facilities.
- 100 Gbit lines are becoming standard (regional network ring at GSI is 1 Tbit).
- More attention to COTS products

Distributed T0/T1 centre embedded in Grid/Cloud

APPA
CBM
LQCD
NuStar
Panda (66k cores, 12PB disks, 12PB/y tape)

in 2018
300000 cores + grid
40 PB disk
40PB/y archive

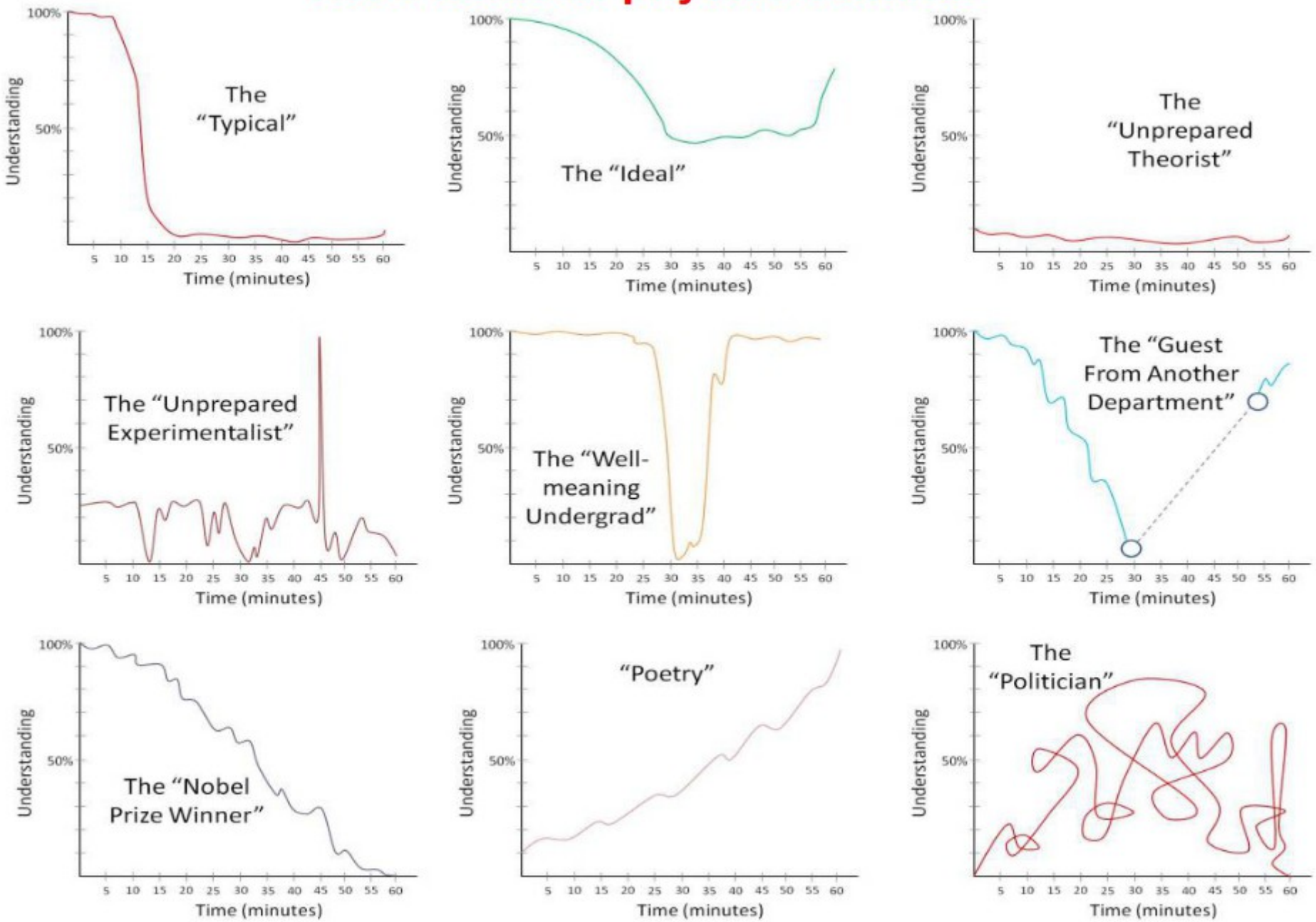


computing center for high-energy, hadron, nuclear, and atomic physics

serve >20 collaborations

Spare slides

The 9 kinds of physics seminar



http://manyworldstheory.com/2013/10/03/the-9-kinds-of-physics-seminar/