Software Tools and Information System

Track Summary

CHEP 06 -- Mumbai, India

Andreas Pfeiffer
CERN, PH/SFT
andreas.pfeiffer@cern.ch

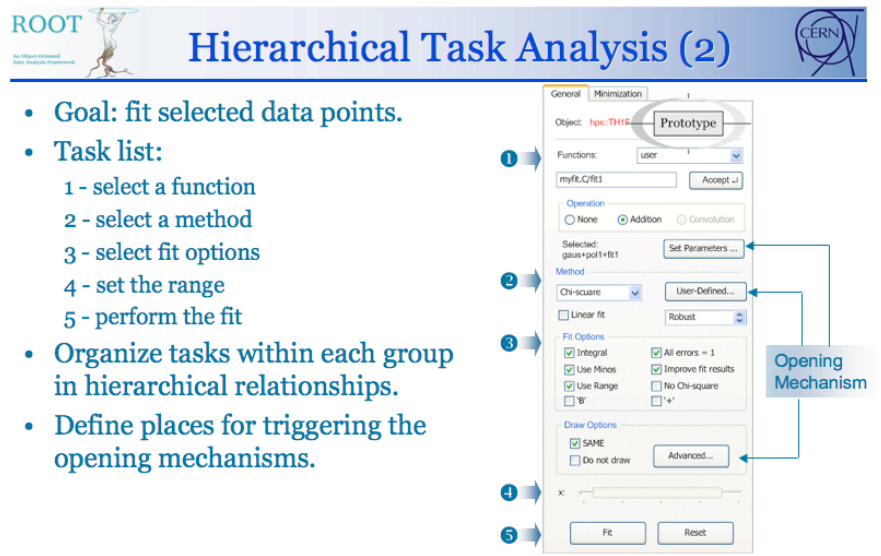
Track summary

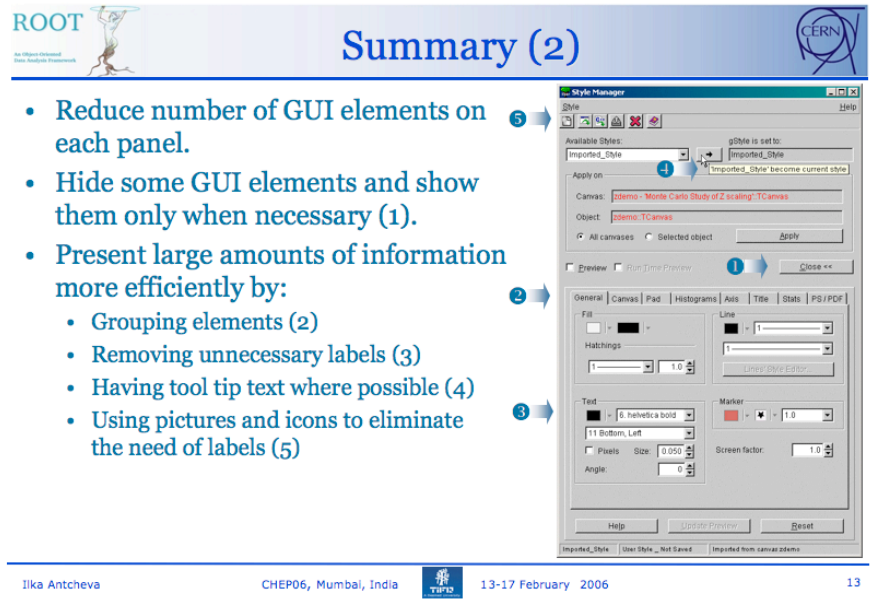
- Received 38 abstracts
  - 33 oral presentations, 5 posters
    - 1 oral arrived as poster, 2 didn't arrive at all
- Four main themes (with "fuzzy" borders)
  - GUI and generic application design
  - Release and distribution management, validation, job configuration and workflow
  - Computing studies: compilers, parallel, performance optimization
  - Communication and information management
- As time is short, I had to make some (subjective) selection
  - Apologies for this …

Andreas.Pfeiffer@cern.ch

GUI and generic application design

- Guidelines for GUI application design
  - Design needs to start with UseCase analysis
  - Three click rule
    - *Miller's law of 7*
  - Use structure to hide complexity
    - Tabs, sub-menus

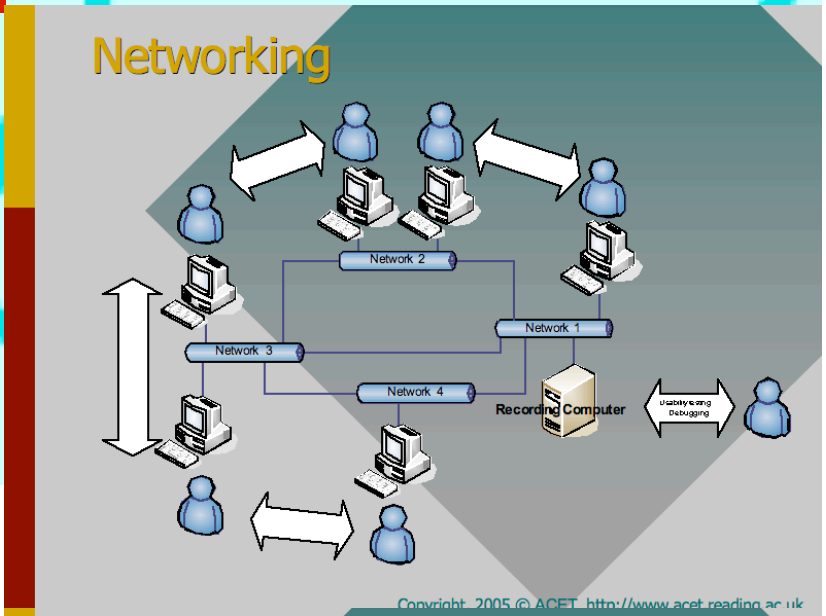


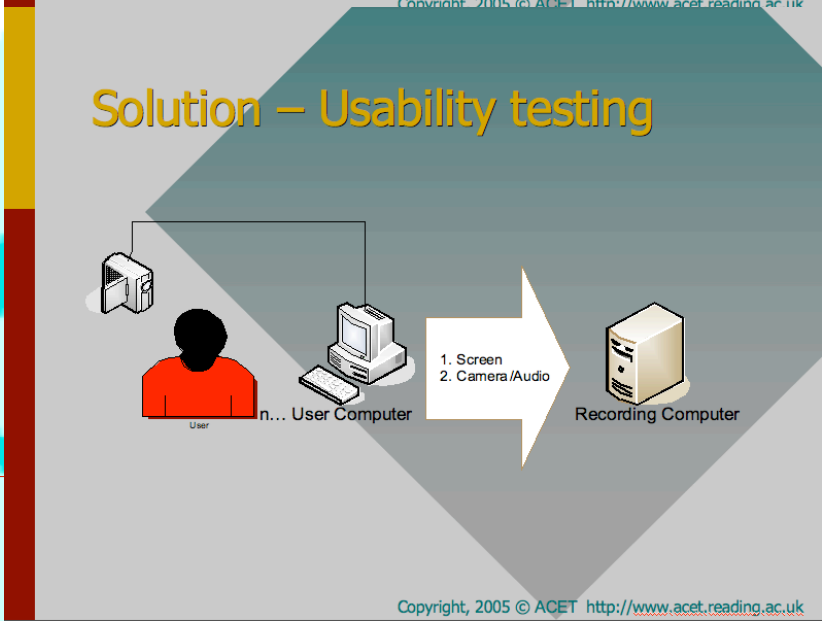


Fons Rademakers, ID 188 cern.ch

GUI and generic application design

- Testing of distributed GUIs/applications
  - UnitTests not enough
  - Usability tests need to capture information as user(s) use it
    - Capture screen
    - Capture video/audio
    - From all inputs in distributed environment
    - Also useful for debugging at a later time
  - Proof of concept done





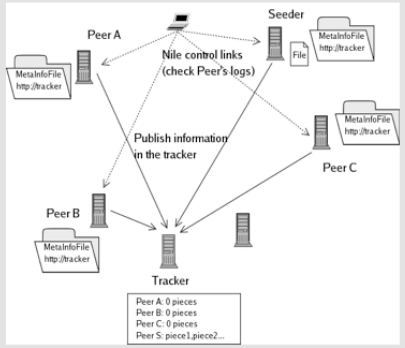
Andreas.Pfeiffer@cern.ch

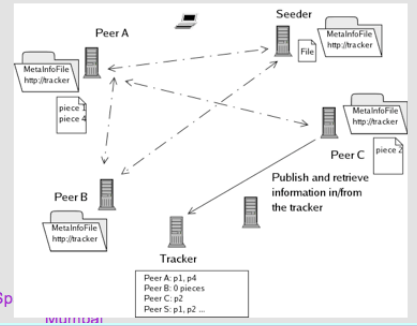
Release and distribution management, validation, job configuration and workflow

- Worm and P2P Distribution in Atlas Trigger/DAQ
  - 6 GB of s/w per release, 600 nodes in various locations
  - Distribution tools not always available outside CERN
  - Using “hacker technology” to distribute:
    - Worm: Nile
      - Distribute tools and commands
    - Peer-to-Peer: BitTorrent
      - Distribute the s/w

Hegoi Garitaonandia, ID 200

A P2P for File Distribution: BitTorrent (2)

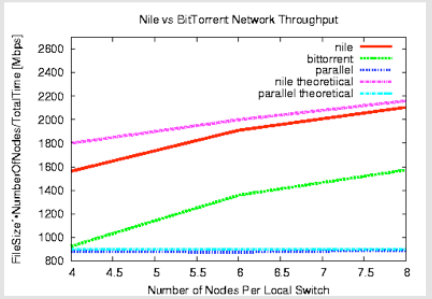




H.Garitaonandia, hegoi@ifae.es, IFAE, Barcelona, Sp

SW Distribution in Pre-Series with Known Network Topology

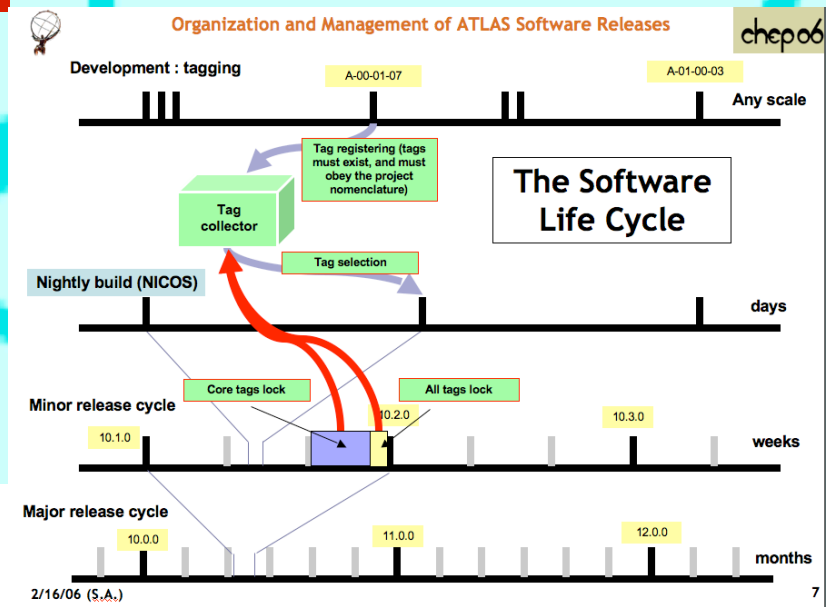
- **Nile 2.0.2 configured in two stages:**
  - **Performed the best of all three.**
  - **Throughput was close to the expected value.**
- The parallel copy can be understood as the **simplest** Quattor configuration:
  - Only one SW repository, HTTP, no Squid caches

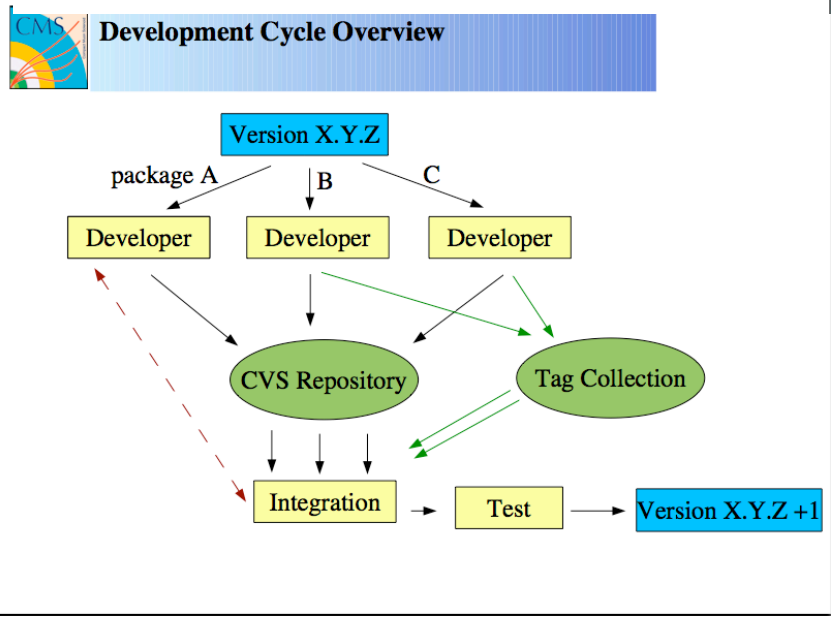


H.Garitaonandia, hegoi@ifae.es, IFAE, Barcelona, Spain 02/14/2006, CHEP06, TIFR, Mumbai

Release and distribution management, validation, job configuration and workflow

- Release Process very similar in Atlas and CMS
  - Main problem: large number of developers and geographical diversity
  - Use different tools for configuration mgt and build
  - Quite some commonality in process (and (some) tools)
    - Nightlies (nicos)
    - collecting/controlling tags





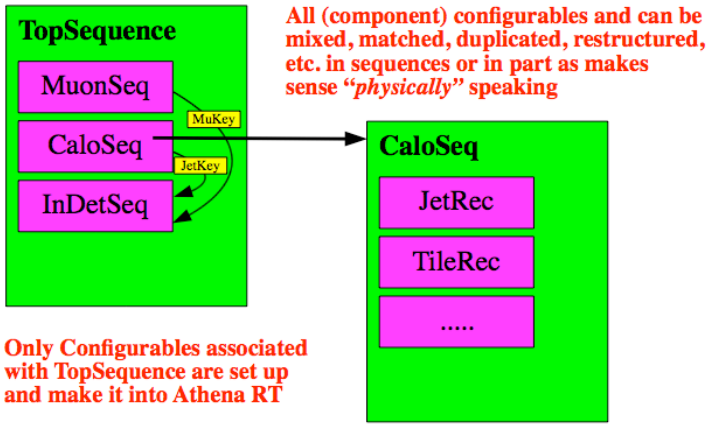
Solveig Albrand (Atlas), ID 71
Stefano Argiro (CMS), ID 246

, India

Release and distribution management, validation, job configuration and workflow

- Physics-level job configuration
  - Targeting end-user physicists
  - Defining configurables
    - Smart low-level building blocks (auto-generated)
    - Structuring possible
    - Build higher level structures needed by physics community

2) Structure support



Summary & Outlook

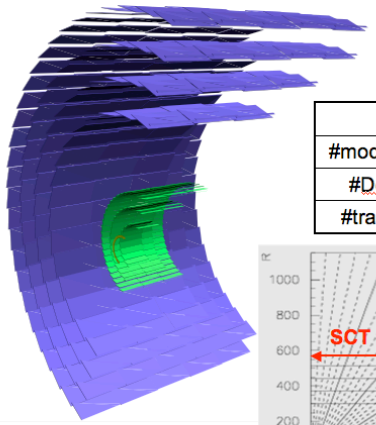
- **Atlas software is changing:**
  - development focus => analysis focus
- **Configuration building blocks provided**
  - Auto-generated, checkable, independent
- **Layered structures now possible**
  - With layered builders (functions/classes)
  - End-user modifiable, exploration-safe
- **Opens up possibilities for new tools**
  - Browsers, validators, code generators, etc.

Computing studies: compilers, parallel, performance optimization

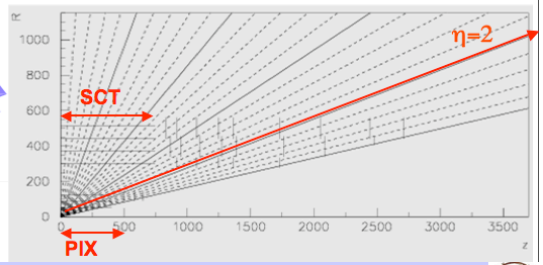
- Parallel computing for the Atlas Si tracker
  - ca 35000 Degrees of Freedom in the tracker
  - Alignment with a global χ^2 function
    - Limited by memory size, precision and execution time
  - Using AMD64 Beowulf cluster at RAL
    - Carefully checked correctness and quality
  - Gained factor of about 30

Müge Karagöz Ünel, ID 100

Solving Various Size Problems for Barrel Track



	cone	SCT	PIX+SCT
#modules	1030	2112	3568
#DoF	6180	12672	21408
#tracks	~250k	~550k	~800k



Comparison of Platforms for Global χ^2 Solution

- Parallel-computing allowed for large problem solutions in a very efficient way.
- The time performance looks very advantageous!
- Number of modules in parantheses show the modules not taken into account for alignment.

Nmodule	NDoF	File size(GB) NxN(triang)	Processing time*		
			SCARF NCPU=16 / 4 (2.2GHz, MB=256MB)	Intel PentiumIV	SCARF single-CPU
1030 (0)	6180	0.3 (0.15)	1m53s / 3m	35m (2.8GHz)	39m
2112 (1)	12672	1.2 (0.6)	10m30s / 26m58s	7h34m (2GHz)	5h28m
3568 (37)	21408	3.4 (1.7)	43m43s/ 121m51s	---	**>20 hrs**

* Measurements include the binary file I/O.

- Also attempted invert these singular matrices. Matrices either found completely singular (as expected) (RCOND=0.0) or RCOND< ε (DoF=6180), and no solution is provided.

Computing studies: compilers, parallel, performance optimization

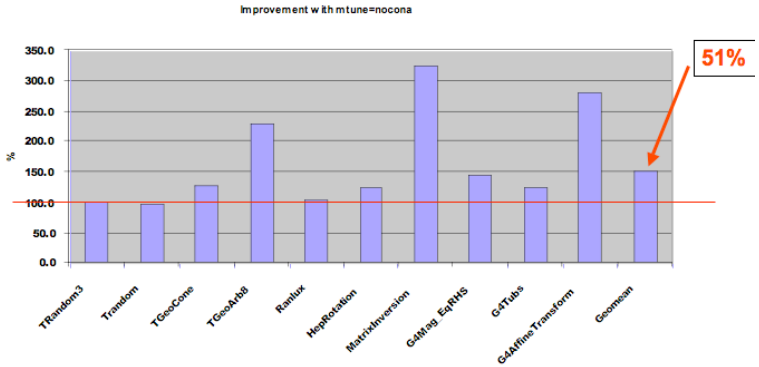
- Software kernels - performance studies using small "snippets" of HEP code
  - 10 snippets from HEP codes selected and analyzed so far
    - More are planned
  - Talking to compiler writers to improve the compilers
    - Mainly commercial and gcc
    - Good feedback
  - Need to care about potential pitfalls
    - Inconveniences are outweighed by the advantages

ndreas.Pfeiffer@cern.ch

Sverre Jarp, ID 31

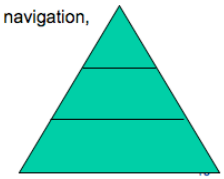
Used in comparisons (1)

- gcc 4.0.2 (with O2) on a 3.6 GHz Xeon (64-bit mode)
  - With/without mtune=nocona



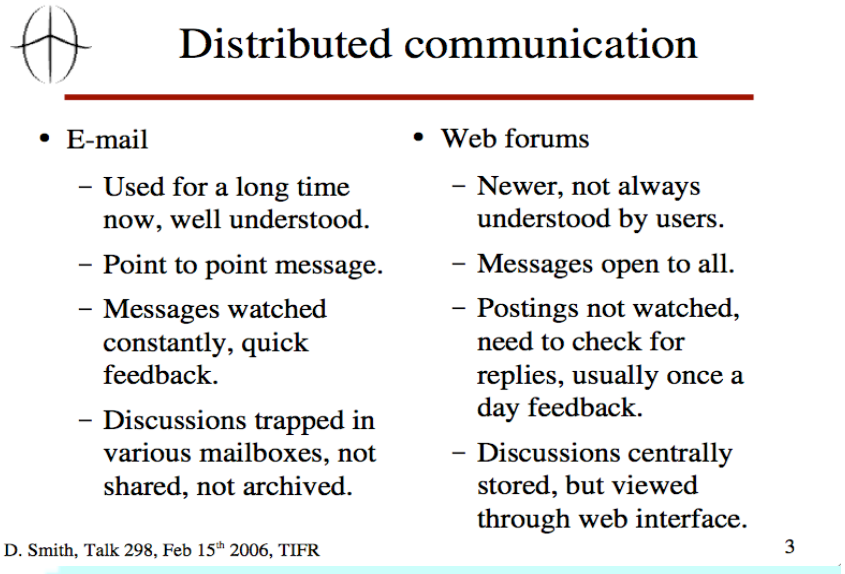
Conclusions

- **Benchmarking and optimization are still important:**
  - LHC physicists will have huge CPU demands
- **But, we have to tread carefully!**
  - "You only test what you REALLY test"
- **As we have seen:**
  - Snippets: Great for testing single compiler features
    - **Mandatory** in discussions with compiler writers
  - ROOTmarks (from *stress* testing)
    - Need to know our domain (file input/output, geometrical navigation, Linear Algebra, STL, etc.)
  - The full-blown LHC applications
    - Best – but extremely complex to port



Communication and Information Management

- HyperNews - managing discussions in HEP
  - First used in BaBar then STAR
    - This year: Atlas, CMS
  - Combining the positive aspects of e-mail with web-based bulletin boards
  - Rich set of features
    - Most recent postings
    - Search engine
    - Central (un)subscription to fora



Distributed communication

- E-mail
  - Used for a long time now, well understood.
  - Point to point message.
  - Messages watched constantly, quick feedback.
  - Discussions trapped in various mailboxes, not shared, not archived.
- Web forums
  - Newer, not always understood by users.
  - Messages open to all.
  - Postings not watched, need to check for replies, usually once a day feedback.
  - Discussions centrally stored, but viewed through web interface.

D. Smith, Talk 298, Feb 15th 2006, TIFR 3

Coordinated discussions in forums

- Discussions centrally managed, divided into forums

SLAC Forum List by Category

Forums by Category | Recent Postings | Member Info | Help
Forums by Time Order | Search in Forums | Members List | Contact Admin
Request a New Forum | Subscribe to Forums | New Member

Categories
Discussions on these forums, announcements and setup
Hypernews Use and Development

Discussions on these forums, announcements and setup
Hypernews Announcements
Hypernews Test
Suggestions for changes to setup of these forums
Hypernews Use and Development
Hypernews Development
Hypernews Feedback/Problems

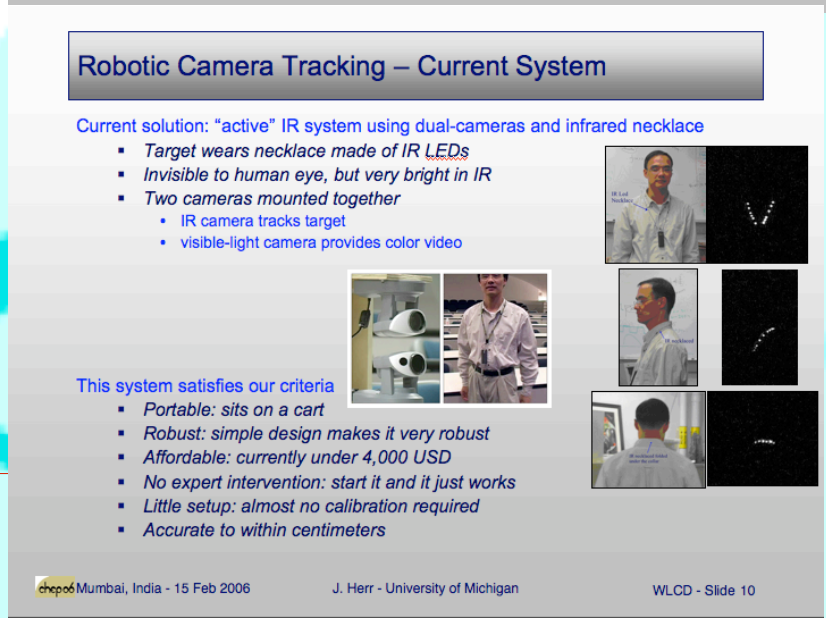
This site runs SLAC HyperNews version 1.11-slac-31, derived from the original HyperNews

D. Smith, Talk 298, Feb 15th 2006, TIFR 8

Douglas Smith, ID 298

Communication and Information Management

- Capturing and archiving lectures for HEP
  - Web Lecture Archive Project
    - Started with pilot in 1999
    - Working on a standard for "Lecture Object"
      - XML extension to SMIL
      - "Seminar" vs. "Tutorial" types with different requirements
    - Automation using robotic cameras which track speaker
      - Developed "active IR tracking"
    - Portable, aim at cost < 10 K$



Steven Goldfarb, ID 344,345

Communication and information management

- HepForge and HEP software development
  - Emphasizing the importance of small well-designed tools
    - Modularity and interfaces
    - A la unix-tools
  - HepForge created as part of CEDAR
    - Full development environment (web, cvs/svn, wiki, tracker, mail-lists, download)
  - In operation since Jan 06
    - After beta test of 6 months

Andy Buckley, ID 120 Pfeiffer@cern.ch

CEDAR: Collaborative e-Science Data Analysis Resource

HepForge is a component of CEDAR:

- JetWeb: global tuning of Monte Carlo generator parameters
- HepData: archival of published experimental data
- HepForge: development environment for HEP software
- HepCode: centralised repository of pheno code/programs
- HepML: set of XML data formats for data sets and MC config

www.cedar.ac.uk

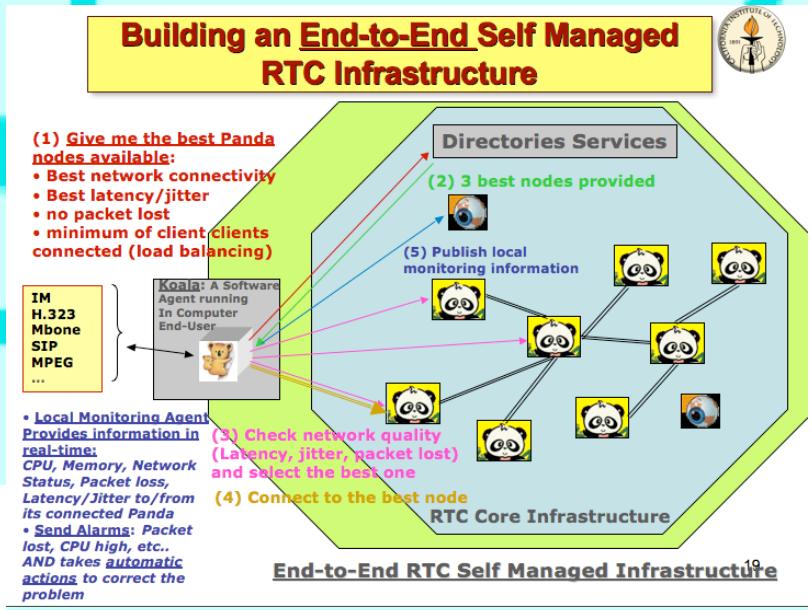
HepForge exists (officially) to implement HepCode

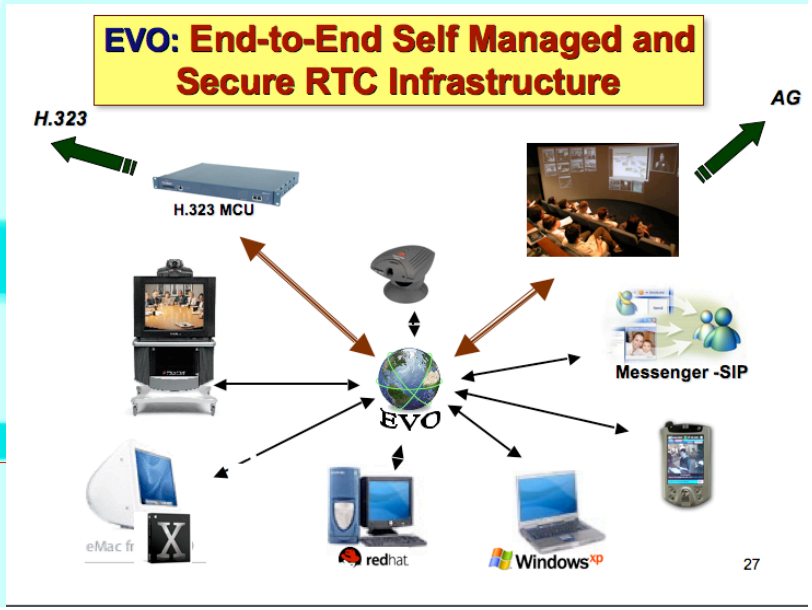
Summary

- HEP *needs* to consider modularity and interfaces more
- HepForge is a small spin-off from CEDAR's main thrust (building and operating a MC generator tuning system)
- HepForge will be used to implement the HepCode system
- HepForge is available for HEP software development now!
- Feedback has all been very positive: system is powerful but very easy to use
- If you have written / are going to write a small–medium re-useable HEP application, please consider HepForge!

Communication and information management

- From VRVS to EVO
  - VRVS well established and well used, will stay
    - Next-generation tools will avoid some of the problems found
  - EVO: adaptive, self-managed infrastructure
    - Decentralized agents ("Panda")
    - Distributed w/o single point of failure
    - "Koala"s at client side contact "nearest" Panda
    - Multi-platform interoperable





Philippe Galvez, ID 154

Conclusions

- CHEP 06
  - Very “LHC” heavy - but the few talks on other topics were interesting!
    - Phenix (RHIC): data taking with up to 600 MB/sec !
  - Grids are being used now
    - … and there are still several of them
  - LHC experiments go from “development” to “deployment” mode
- STIS track
  - Interesting track, spanning a wide field
    - GUI and generic application design
    - Release and distribution management, validation, job configuration and workflow
    - Computing studies: compilers, parallel, performance optimization
    - Communication and information management
  - Interesting and stimulating discussions


SOFT BEVERAGES
MINT COOLER