HEPiX-2010-Spring workshop brief review

http://www.lip.pt/hepixspring2010

http://indico.cern.ch/conferenceTimeTable.pyconfId=73181#20100419

and

where do we go

The presentation overview

- ➔About HEPiX workshop series
- ➔The HEPiX agenda
- ➔Storage
- ➔Virtualization
- ➔Linux distributions
- ➔Benchmarking
- ➔Disasters
- ➔Common things
- ➔HEPD

HEPiX aims

- The HEPiX forum unites IT system support staff, including system administrators, system engineers, and managers from the High Energy Physics (HEP) and Nuclear Physics laboratories and institutes, ...
- The HEPiX meetings are an excellent source of information for IT specialists in scientific computing.
- Members of HEPiX are responsible computing persons from many HEP laboratories around the World.
- HEPiX main site: *https://www.hepix.org/*

HEPIX 2010 Spring Agenda

- Site Reports (11)
- Storage and File Systems (8)
- Monitoring & Infrastructure Tools (4)
- Virtualization (7)
- Grid and WLCG (3)
- Operating Systems and Applications (3)
- Miscellaneous (1)
- Benmarking (2)
- + keynote speeches and closing remarks (6)
  - In total ~45 presentations (~20 local and ~25 remote over EVO); ~110 registered persons.

Site Reports

- LIP and Grid in Portugal *(by Goncalo BORGES)*
- RAL Site Report *(by Martin BLY)*
- BNL RHIC/ATLAS Computing Facility Site Report *(by Christopher HOLLOWELL)*
- CERN site report *(by Helge MEINHARD)*
- DESY site report *(by Wolfgang FRIEBEL)*
- Petersburg Nuclear Physics Institute (PNPI) status report *(by Andrey Shevel)*
- SLAC Site Report *(by Randy MELEN)*
- Fermilab Site Report *(by Chadwick KEITH)*
- INFN Tier1 site report *(by Vladimir SAPUNENKO)*
- Site report from PDSF *(by Jay SRINIVASAN)*
- Jefferson Lab Site Report *(by Sandy PHILPOTT)*

Storage and File systems

- Progress Report 2010 for HEPiX Storage Working Group *(by Andrei MASLENNIKOV)*
- Evaluation of NFS v4.1 (pNFS) with dCache *(by Patrick FUHRMANN)*
- Building up a high performance data centre with commodity hardware *(by Andreas HAUPT)*
- CERN Lustre evaluation and storage outlook *(by Tim BELL)*
- LCLS Data Analysis Facility *(by Alf WACHSMANN)*
- GEMSS: Grid Enabled Mass Storage System for LHC experiments *(by Vladimir SAPUNENKO)*
- OpenAFS Performance Improvements: Linux Cache Manager and Rx RPC Library *(by Jeffrey ALTMAN)*
- Lustre-HSM binding *(by Thomas LEIBOVICI)*

Monitoring & Infrastructure Tools

- ➔Lavoisier : a way to integrate heteregeneous monitoring systems *(by Cyril L'ORPHELIN)*
- ➔Scientific Computing: first quantitative methodologies for a production environment *(by Alberto CIAMPA)*
- ➔RAL Tier1 Quattor experience and Quattor outlook *(by Ian Peter COLLIER)*
- ➔Spacewalk and Koji at Fermilab *(by Troy DAWSON)*

Virtualization

- Update on HEPiX Working Group on Virtualisation *(by Tony Cass)*
- Virtualization at CERN: a status report *(by Ulrich SCHWICKERATH)*
- Virtual machines over PBS *(by Marc RODRIGUEZ ESPADAMALA)*
- An Adaptive Batch Environment for Clouds *(by Ian GABLE)*
- Virtualization in the gLite Grid Middleware software process. Use-cases, technologies and future plans *(by Lorenzo DINI)*
- Virtual Network and Web Services (An Update) *(by Thomas FINNERN)*
- Virtualisation for Oracle databases and application servers *(by Carlos GARCIA FERNANDEZ)*

Grid and WLCG

- ➔CESGA Experience with the Grid Engine batch system *(by Esteban FREIRE GARCIA)*
- ➔CERN Grid Data Management Middleware plan for 2010 *(by Oliver KEEBLE)*
- ➔EGEE Site Deployment: The UMinho-CP case study *(by Tiago Sá)*

Benchmarking

- ➔Preliminary Measurements of Hep-Spec06 on the new multicore processor *(by Michele MICHELOTTO)*
- ➔Hyperthreading influence on CPU performance *(by Joao MARTTINS)*

Operating Systems and Applications

- ➔Scientific Linux Status Report and Plenary Discussion *(by Troy Dawson)*
- ➔Windows 7 Deployment at Cern *(by Michal BUDZOWSKI)*
- ➔TWiki at CERN: Past Present and Future *(by Pete JONES)*

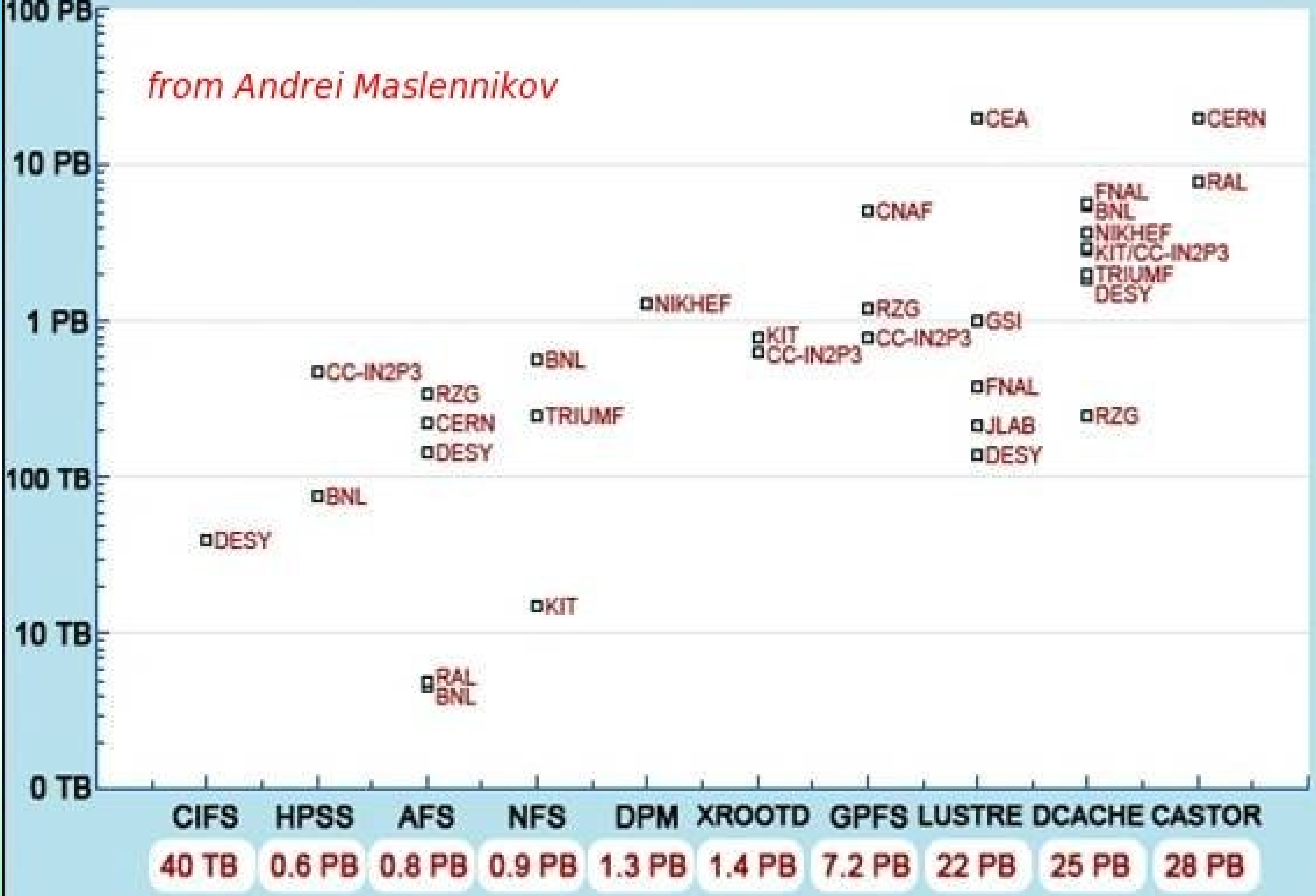
Miscellaneous

➔Lessons Learned from a Site-Wide Power Outage
(by John BARTELT)

Scale of computing facilities

- ➔DESY/Zeuten (GPU;2.9K cores) - Andreas Haupt
- ➔INFN-PISA (1.9K cores; ~350TB disks) — Alberto Ciampa
- ➔RAL (2.8K cores; ~1.2 PB disks) — Martin Bly
- ➔BNL (10K cores; ~6 PB disks; ~15 PB tapes) — Christopher Hollowell
- ➔CERN (added ~16K cores) — Helge Mainhard
- ➔DESY/HH (4.9K cores) — Wolfganf Friebel
- ➔SLAC (GPU; 8.2K cores; ~3.5 PB disks;~6.7 PB tapes) - Randy MELEN
- ➔FNAL (Grid cluster ~3K servers) - Chadwick KEITH
- ➔INFN (added 2.2K cores; ~6.8 PB disks; 10 PB tapes) - Vladimir SAPUNENKO
- ➔PDSF (LBNL) (GPU) - Jay SRINIVASAN
- ➔JLAB (GPU; 5.7K cores) - Sandy PHILPOTT

Terabytes on disk per type of the shared area



Largest Russian HEP computing resource is JINR computing facility

- ➔150+ servers
- ➔1K+ cores
- ➔0.5+ PB disks
- ➔20 Gbit line JINR — Moscow
- ➔All required software and services as expected for Tier2 (cite from http://lit.jinr.ru/Inf_Bul_5/IB_LIT_5(46)_2010_color.pdf)

Storage access featurs

- ➔NFSv4.1 is near far from public - Patric Furman
- ➔Dcache, GPFS, Lustre are in wide use;
- ➔Lustre is not acceptable (yet) as a file system for Tier0 — Tim Bell
- ➔No ideal file system for all scenario of data handling - *parameters tuning is required if you need max througput.*

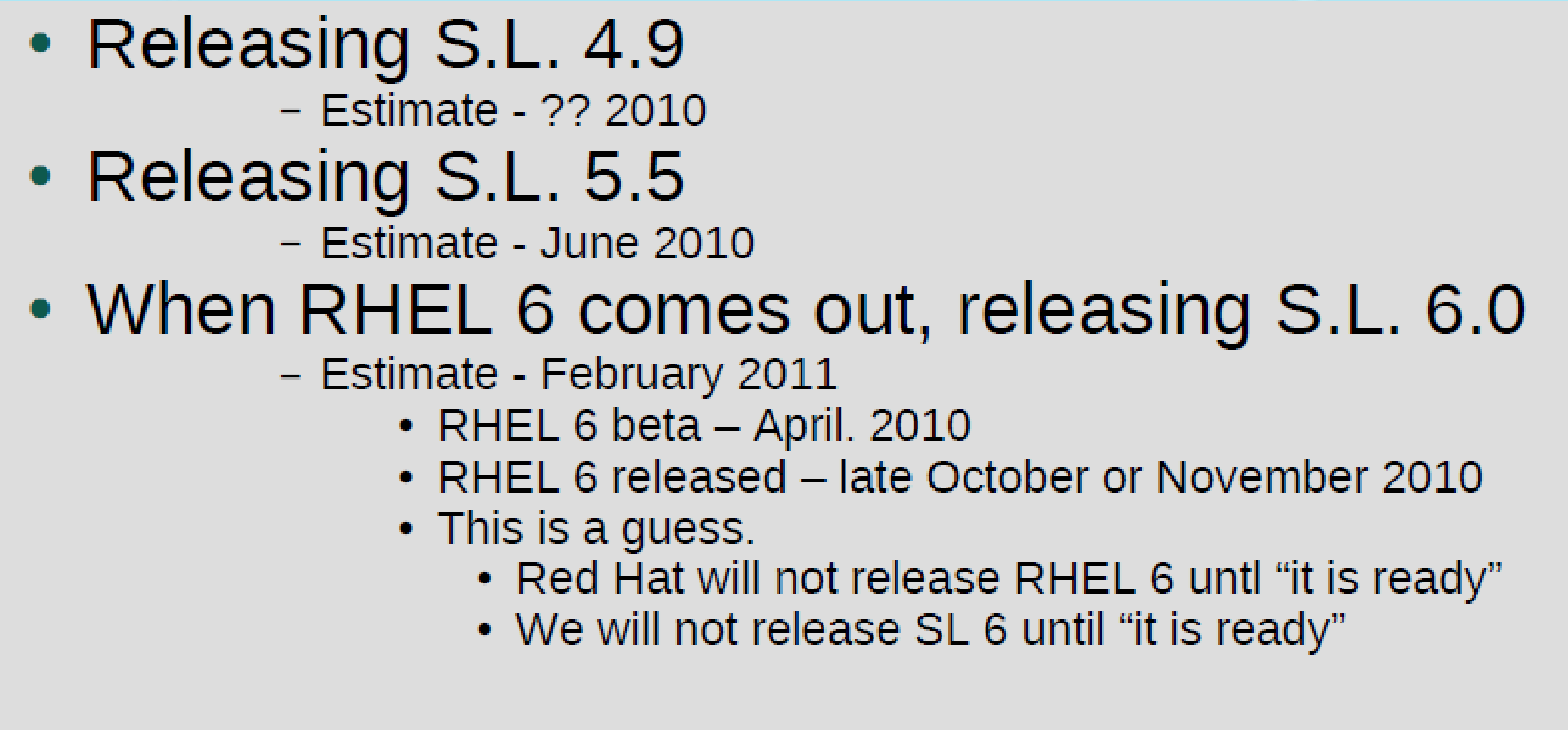
Virtualization, Clouds

- Many developments to handle the virtual images:
  - To create
  - To revoke
  - To send
  - To spread around servers
  - To balance the load

Linux distributions

- Scientific Linux
  - *http://www.scientificlinux.org/*
- CERN Scientific Linux
  - *http://linux.web.cern.ch/linux/*
- NauLinux (derived from Scientific Linux)
  - *http://www.naulinux.ru/*

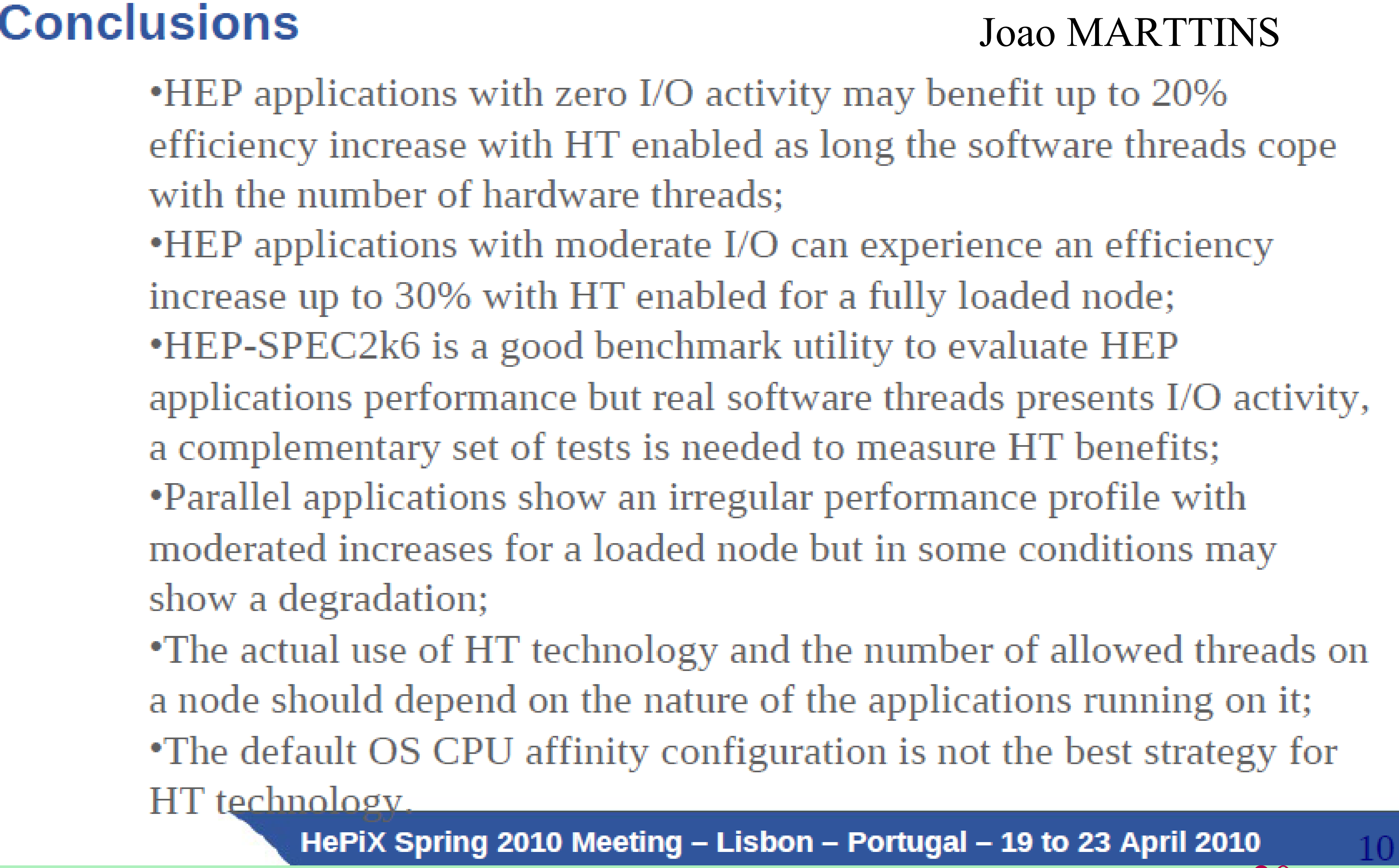
Nearest future for Scientific Linux (cite from Troy Dawson)



- Releasing S.L. 4.9
  - Estimate - ?? 2010
- Releasing S.L. 5.5
  - Estimate - June 2010
- When RHEL 6 comes out, releasing S.L. 6.0
  - Estimate - February 2011
    - RHEL 6 beta – April. 2010
    - RHEL 6 released – late October or November 2010
    - This is a guess.
      - Red Hat will not release RHEL 6 untl "it is ready"
      - We will not release SL 6 until "it is ready"

Benchmarking

Joao MARTTINS

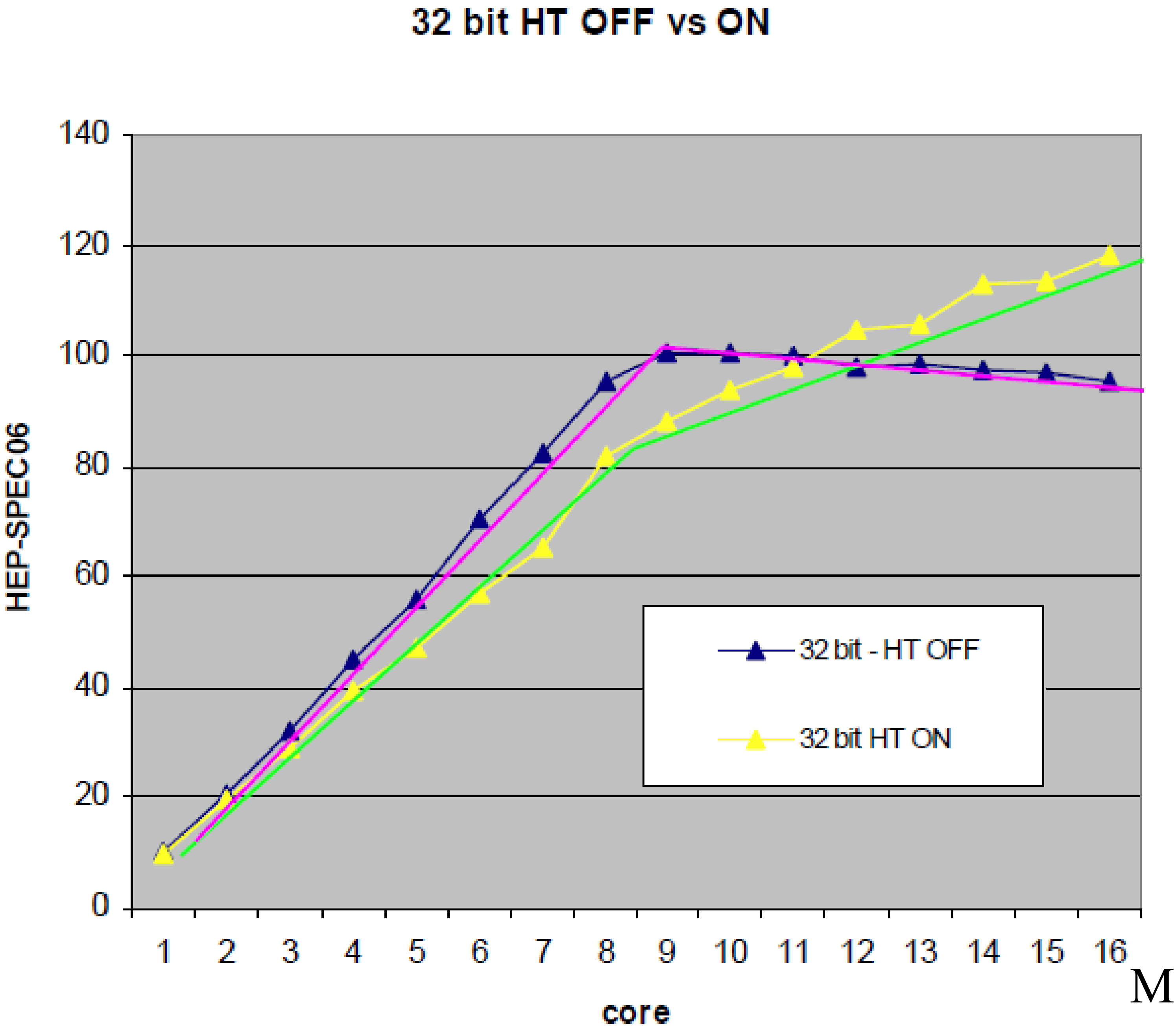


Conclusions

- HEP applications with zero I/O activity may benefit up to 20% efficiency increase with HT enabled as long the software threads cope with the number of hardware threads;
- HEP applications with moderate I/O can experience an efficiency increase up to 30% with HT enabled for a fully loaded node;
- HEP-SPEC2k6 is a good benchmark utility to evaluate HEP applications performance but real software threads presents I/O activity, a complementary set of tests is needed to measure HT benefits;
- Parallel applications show an irregular performance profile with moderated increases for a loaded node but in some conditions may show a degradation;
- The actual use of HT technology and the number of allowed threads on a node should depend on the nature of the applications running on it;
- The default OS CPU affinity configuration is not the best strategy for HT technology.

HePiX Spring 2010 Meeting – Lisbon – Portugal – 19 to 23 April 2010 10

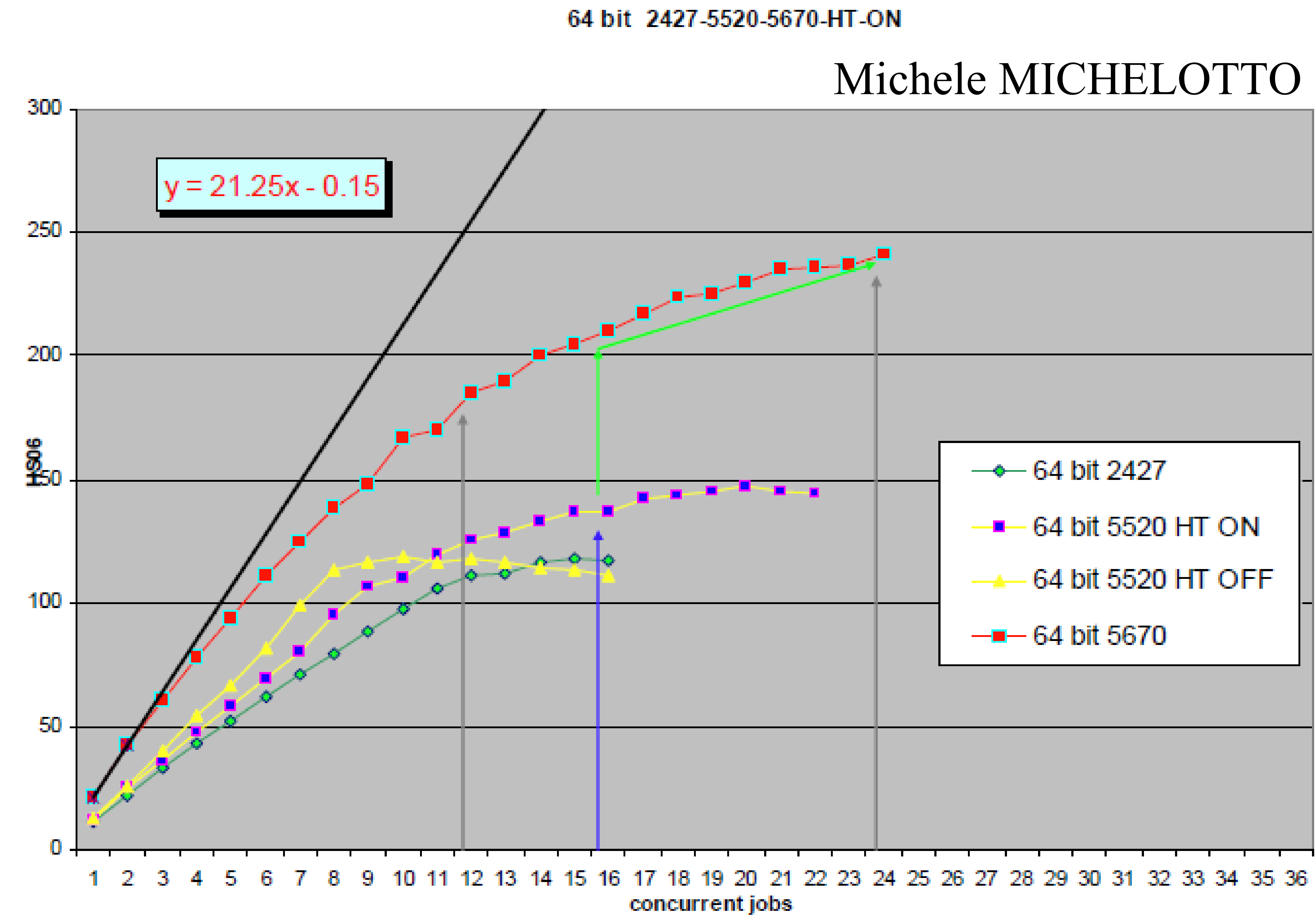
Benchmarking: HT on vs HT off



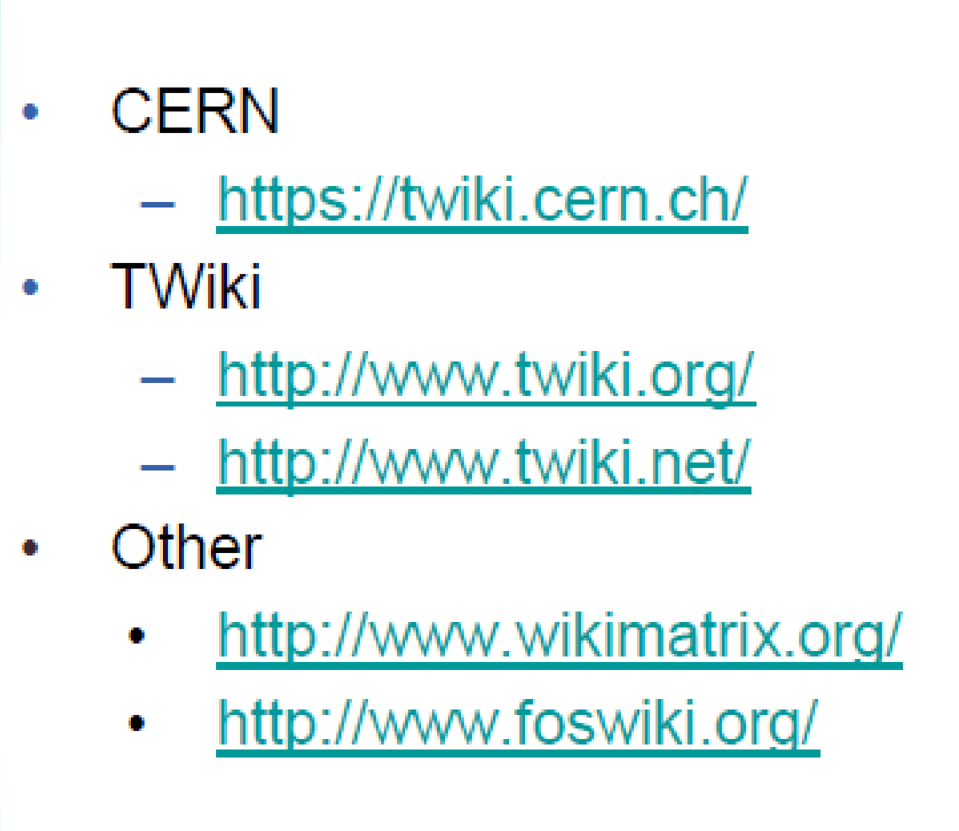
HT ON:81.81
HT OFF: 95.96
HT OFF is better up to 11concurrent run

Michele MICHELOTTO

Benchmarking: a range of new CPU



Twiki (cite from Peter Jones)



- CERN
  - https://twiki.cern.ch/
- TWiki
  - http://www.twiki.org/
  - http://www.twiki.net/
- Other
  - http://www.wikimatrix.org/
  - http://www.foswiki.org/

Failures and desasters

- Two reports about serious problems:
  - FNAL — one of UPSs were out of order due to outdated breaker — *(by Chadwick KEITH)*
    - Part of clusters were switched off due to temperature reasons because of air conditioner became off power.
  - SLAC — site wide power cut for two days (!) because of storm - *(by John BARTELT)*
    - Mails, web sites, procurement hosts were switched off.

Common things

- External connectivity: 10 Gbit and more
- Disk space around min 1 PB and more *(around 20 PB at CERN)*
- Electrical Power from 100s KW to 1s MW
- Almost all run Scientific Linux 5 (Berkeley — CentOS)
- Popular micro processor ~ Intel Xeon 5520/5570
- 3 GB per core; 1 job per core
- Popular number of cores per server 8-16 (i.e. main memory 24-48 GB per server)
- GPU in many sites (mainly for Lattice QCD)
- DESY, SLAC change/extend area of scientific research (changing in personalities)
- Developing new and integration existing complex components.
- Power/Cooling

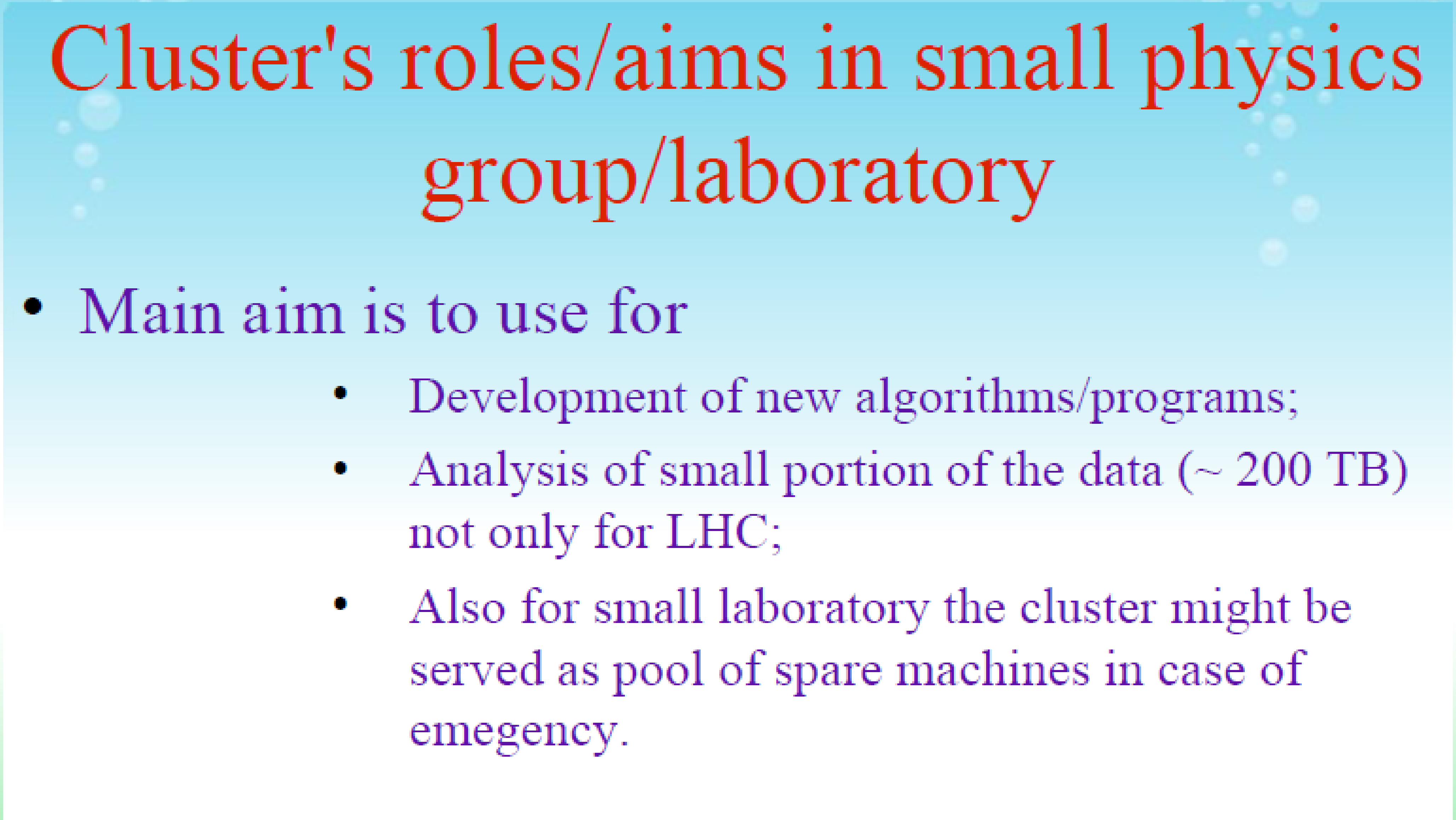
- Remote control rooms established for ATLAS and CMS

ATLAS remote control room in Zeuthen

CMS remote control room in Hamburg

And HEPD …

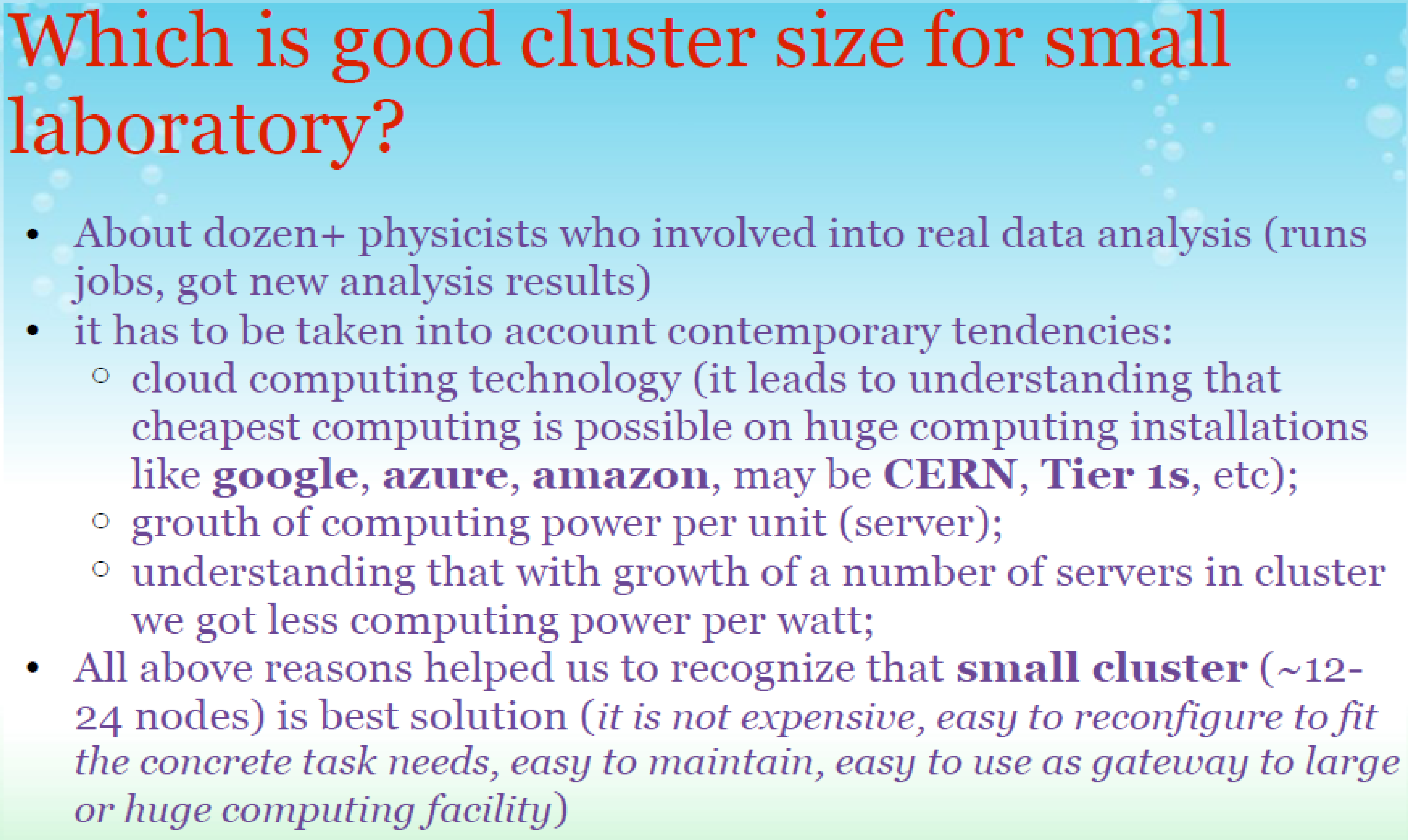
Cite 1 from Shevel's report on HEPIX



Cluster's roles/aims in small physics group/laboratory

- Main aim is to use for
  - Development of new algorithms/programs;
  - Analysis of small portion of the data (~ 200 TB) not only for LHC;
  - Also for small laboratory the cluster might be served as pool of spare machines in case of emegency.

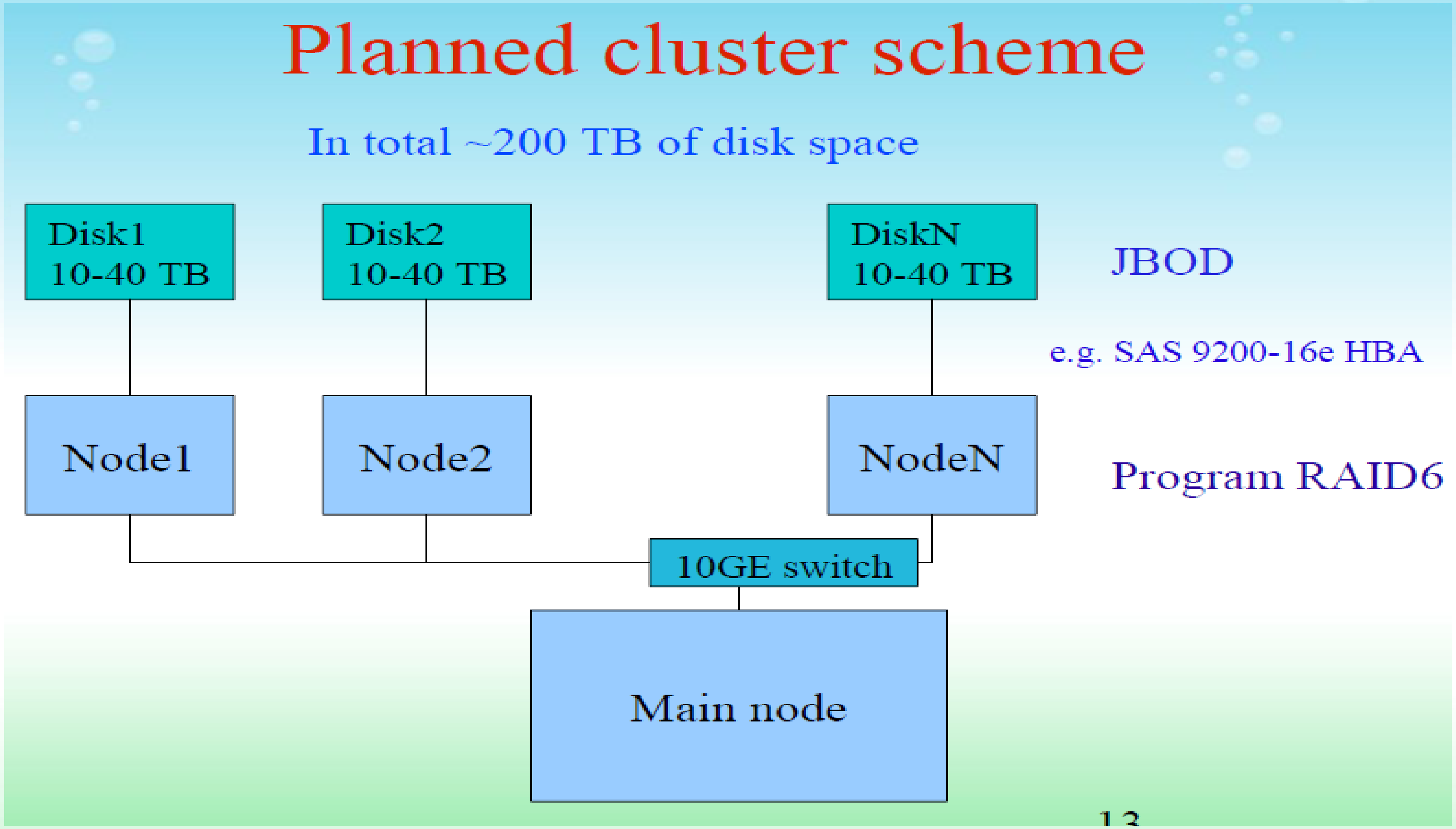
Second Shevel's cite



Which is good cluster size for small laboratory?

- About dozen+ physicists who involved into real data analysis (runs jobs, got new analysis results)
- it has to be taken into account contemporary tendencies:
  - cloud computing technology (it leads to understanding that cheapest computing is possible on huge computing installations like **google**, **azure**, **amazon**, may be **CERN**, **Tier 1s**, etc);
  - grouth of computing power per unit (server);
  - understanding that with growth of a number of servers in cluster we got less computing power per watt;
- All above reasons helped us to recognize that **small cluster** (~12-24 nodes) is best solution (*it is not expensive, easy to reconfigure to fit the concrete task needs, easy to maintain, easy to use as gateway to large or huge computing facility*)

Future HEPD Cluster



Planned cluster scheme

In total ~200 TB of disk space

Disk1 10-40 TB

Disk2 10-40 TB

DiskN 10-40 TB

JBOD

e.g. SAS 9200-16e HBA

Node1

Node2

NodeN

Program RAID6

10GE switch

Main node

Thank you !   Questions?