

CHEP'01 Review

H. Vogt, P. Wegner

Contributions: T. Schmidt, A. Gellrich, T. Naumann

History

Introduction

Plenary Program

Parallel Tracks

DESY Contributions

LHC / Grid

Commodity Hardware & Software

Data Handling & Storage

Data Acquisition

Simulation, Data Analysis, Visualization

Conclusions

History

International Conference on Computing in High Energy (and Nuclear) Physics

18-month period (Europe, USA, World), history:

September 2001: Beijing / P.R. China

February 2000: Padova / Italy

October 1998: Chicago / USA

April 1997: Berlin / Germany

September 1995: Rio de Janeiro / Brazil

April 1994: San Francisco / USA

September 1992: Annecy / France

March 1991: Tsukuba / Japan

April 1990: Santa Fe / USA

April 1989: Oxford / Great Britain

Introduction



CHEP'01
3.-7. September 2001
Friendship Hotel
Beijing / P.R. China

Organized by IHEP of the Tsinghua University Beijing



CHEP'01
Beijing

Plenary Program

Monday (Future):

<i>Status of Preparation for LHC Computing</i>	<i>(M. Delfino / CERN)</i>
<i>Geant4 toolkit: status & utilization</i>	<i>(J. Apostolakis / CERN)</i>
<i>Software Frameworks for HEP Data Analysis</i>	<i>(V. Innocente / CERN)</i>
<i>The LHC Experiments' Joint Controls Project (JCOP)</i>	<i>(W. Salter / CERN)</i>

Tuesday (Running Systems):

<i>The BARBAR Database: Challenges, Trends and Projections</i>	<i>(J. Becla / SLAC)</i>
<i>New data acquisition and data analysis system for the Belle experiment</i>	<i>(R. Itoh / KEK)</i>
<i>The CDF Computing and Analysis System: First Experience</i>	<i>(S. Lammel / FNAL)</i>
<i>The D0 Data Handling System</i>	<i>(V. White / FNAL)</i>



CHEP'01
Beijing

Plenary Program (cont.)

Wednesday (Grid):

<i>Grid Technologies & Applications: Architecture & Achievements</i>	(I. Foster / ANL)
<i>Review of the EU-DataGrid project</i>	(F. Gagliardi / CERN)
<i>US Grid Projects: PPDG nad iVDGL</i>	(R. P. Mount / SLAC)
<i>The Asia Pacific Grid (ApGRID) Project and HEP Application</i>	(S. Sekiguchi / AIST)
<i>Grid Computing with Grid Engine Juxta, and Jini</i>	(S. See / Sun Microsystems)

Friday (Grid/Networks/Clusters/other fields):

<i>Grid Computing at IBM</i>	(G. Wang / IBM)
<i>Present and Future Networks for HEPN</i>	(H. Newman / Caltec)
<i>From HEP Computing to Bio-Medical Research and Vice Versa: Technology Transfer and Application Results</i>	(M. G. Pia / INFN)
<i>Large Scale Cluster Computing Workshop</i>	(Alan Silverman / CERN)
<i>Virtual network computing environment - challenge to high performance computing</i>	GAO Wen (China academy of sciences)
<i>Conference Summaries</i>	(Session convenors)



CHEP'01
Beijing

Parallel Tracks

1. Commodity Hardware & Software (10)
2. Control Systems (9)
3. Data Analysis & Visualization (28)
4. Data Handling & Storage (27)
5. Simulation (8)
6. Information Systems & Multimedia (5)
7. Local & Wide Area Networking (10)
8. Software Methodologies & Tools (26)
9. Triggering & Data Acquisition (28)
10. Grid Computing (26)

Total (177)

Large number of talks especially from US speakers where canceled or held by other authors



CHEP'01
Beijing

DESY Contributions

28 participants, 19 contributions, 14 talks, 5 posters

DESY Hamburg:

IT: C. Beyer (Printing), P. Fuhrmann (dCache), A. Gellrich (Linux)
K. Woller (Linux), K. Ohrenberg, B. Hellwig

IPP: J. Bürger (EDMS), L. Hagge, J. Kreuzkamp (AMS)

H1: U. Berthon (OO), G. Eckerlin (Control), R. Gerhards (OO Framework)
F. Niebergall

ZEUS: U. Behrens (DAQ), U. Fricke (Database), R. Mankel,
K. Wrona (Farms)

HERA-B: V. Amaral (Database)

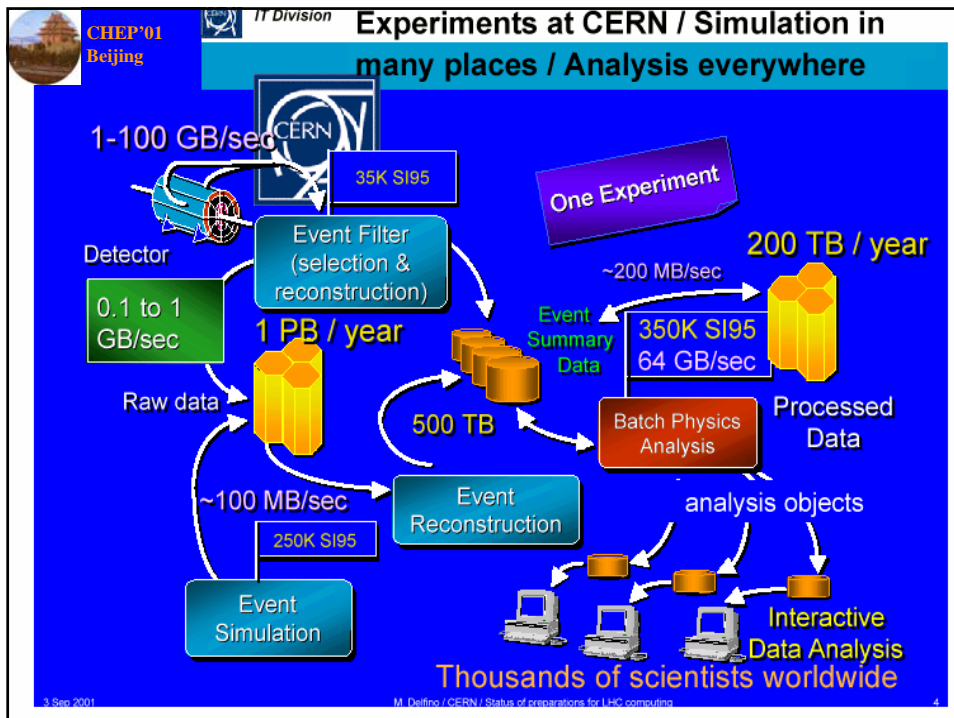
TESLA: H. von der Schmitt (Tesla Computing)

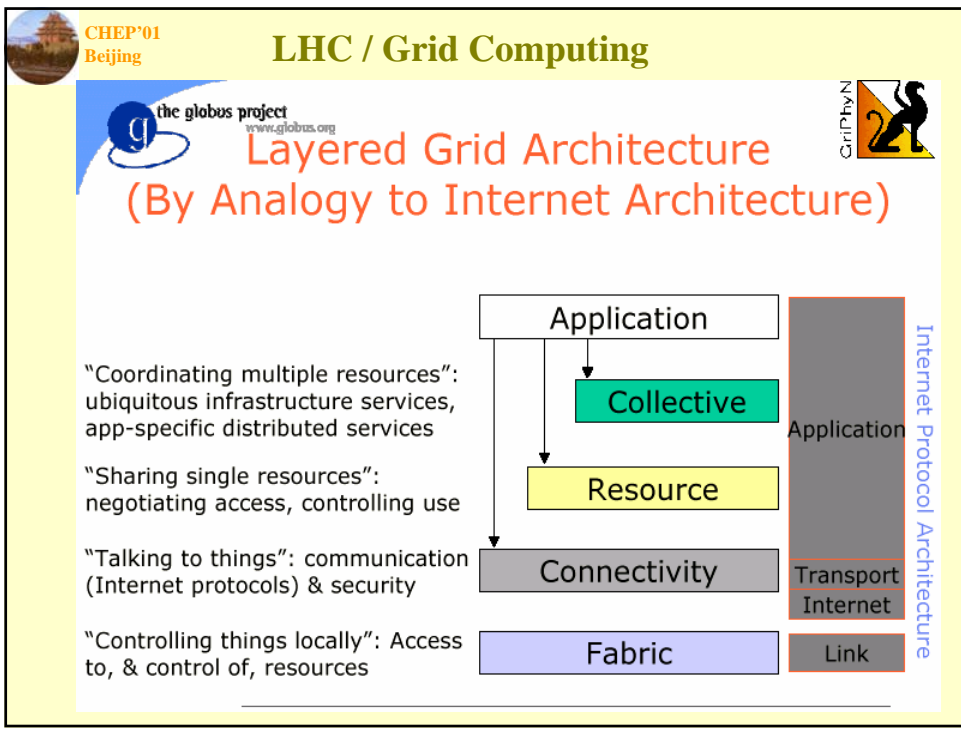
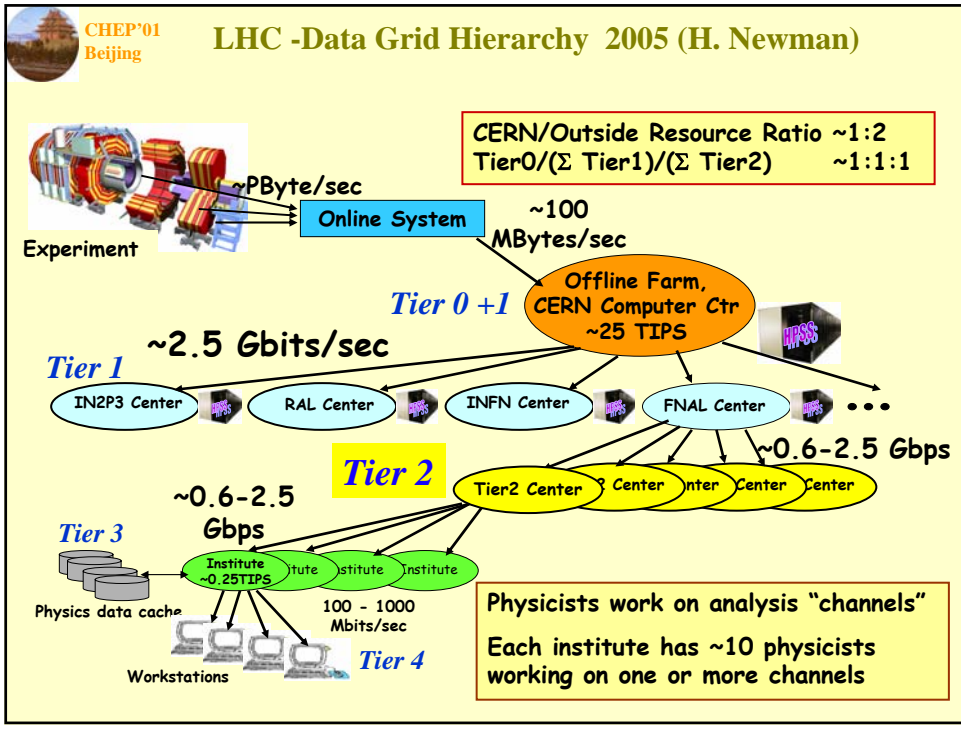
DESY Zeuthen:

H. Leich (PITZ-Interlock), K.-H. Sulanke, H. Vogt,

P. Wegner (APEmille), T. Naumann (H1-Alignment),

J. Hernandez (Hera-B Farms), A. Spiridonov (Hera-B Reconstruction)



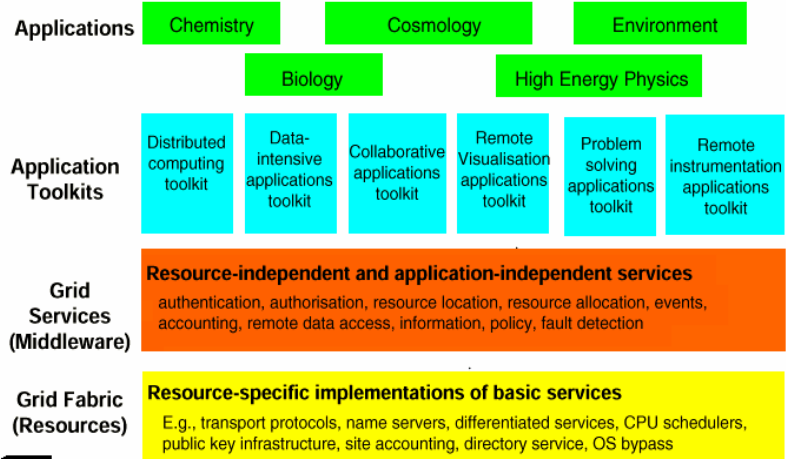




CHEP'01
Beijing

LHC / Grid Computing (B. Segal at CERN HNF meeting)

GRID from a services view



Ben Segal CERN IT/PDP

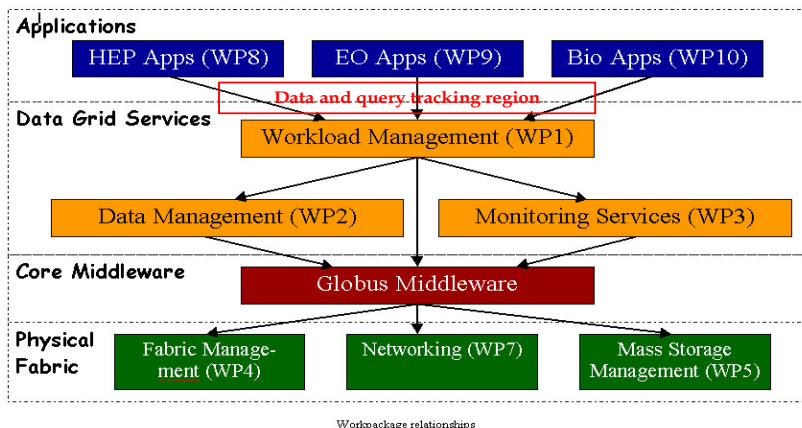
21



CHEP'01
Beijing

LHC / Grid Computing

10-005 Querying Large Physics Data sets over an information Grid (R. McClatchey)



EU-DataGrid

07-09-01

M.Mazzucato – Grid-Beijing

17



CHEP'01
Beijing

LHC / Grid Computing

Goal:

... Grid is (also) the attempt to develop a new world-wide “standard engine” to provide transparent access to resources (computing, storage, network...) for coordinated problem solving by dynamical Virtual Organizations.

Implementation:

Globus, Avaki (commercialized former Legion project)

Projects:

EU Data Grid (CERN + others, 9.8 M Euros EU funding over 3 years)

Grid Physics Network (GriPhyN)

Particle Physics Data Grid (PPDG, \$ 11 M, US DOE, over 4 years)

International Virtual Data Grid Laboratory (iVDGL, \$ 15 M, US NSF, over 5 years)

... and many many others



CHEP'01
Beijing

Cluster Computing

HEP wide: More and more (Linux) PC clusters are replacing large SMP and MPP ‘mainframes’ for offline simulation and analysis

Challenge: automatic management and support, no common tools, side dependent solutions

CERN: Batch farm, about 800 nodes, dual PII/PIII up to 800 MHz, dedicated and shared subfarms controlled managed via LSF (Load Sharing Facility),

Goal : One big shared batch farm with optimal resource scheduling

Only DESY is officially supporting Linux on desktops.

Red Hat – basic distribution (except DESY – SuSE)

Many online farms in HEP experiments:

CDF, D0, RHIC, H1, ZEUS, HERA-B, HERMES, BaBar

LHC: CMS, ATLAS, ALICE – Testfarms O(100),

O(1000) planned from 2005/2006



CHEP'01
Beijing

Cluster Computing

Report on FNAL Cluster Workshop (by invitation, May 2000) :

Review of tools with regard to PC clusters

“Administration is far from simple and poses increasing problems as cluster sizes scale” (A. Silverman, CERN)

Quote of the week - “a cluster is a great error amplifier”
(Chuck Boehm, SLAC)

DESY underrepresented, usage of cfengine, grid engine ... not mentioned

Lattice QCD farm at FNAL:

2000: 80 PCs, Dual PIII 600 MHz, Myrinet
(Replacement for ACPMAPS cluster)

2002: 256 nodes P4(XEON) 2 GHz(?), Myrinet2000

2004: 1000 nodes, 1 Tflop expected, extension to 10 Tflops planned
(deployment of about 200 nodes/year)



CHEP'01
Beijing

Mass Storage

Who's Using What

“Experiment”	Event DB	Metadata DB	Mass Storage System
Alice	Root	MySQL	Castor
AMS	Root	Oracle	
Atlas	Objectivity	MySQL	Castor
BaBar	Objectivity	Objectivity	HPSS
BES	Root	Oracle	----
CDF	Root	Oracle	DIM
CMS	Objectivity	Objectivity	Enstore, HPSS → Castor
COMPASS	CDR	Objectivity	Castor
D0	Flat Files	Oracle	SAM+Enstore, HPSS, etc
JLAB-located	various	various	JASMine
KEK-located	various	various	HPSS
KLOE	YBOS	YBOS+DB2	ADSM+Local
LHCb	Root	In Progress	Castor
Star	Root	MySQL	HPSS
ZEUS	Objectivity	Objectivity	OSM → EuroStore

CHEP 2001

Data Handling & Storage Summary

9



CHEP'01
Beijing

Mass Storage

RAIT (Reliable Array of Inexpensive Tapes)

Parity tapes to data stripes

US DOE/ASCI project, talk by Storage Technology Corporation (STK)

Test setup: 80 MB/sec sustained,

1 FiberChannel Interface striped out to 8 drives

Goal: 8 GB/sec by striping 12 RAIT systems together

dCache

Joint project: DESY-IT, FERMI CD_INTEGRATED SYSTEMS)

“Generic tool for caching, storing and easily accessing huge amounts of data, distributed among a large set of heterogeneous caching nodes“

Single name space (pnfs)

Hot Spot Spreading

OSM interface

Disk storage is the biggest cost risk for LHC computing



CHEP'01
Beijing

DAQ

More than 31 contributions ...

- 3x ATLAS, LHC, CERN
- 3x CMS, LHC, CERN
- 2x ALICE, LHC, CERN
- 1x nTOF, PS, CERN
- 3x DZERO, Tevatron, Fermilab
- 2x CDF, Tevatron, Fermilab
- 1x BeTeV, Tevatron, Fermilab
- 1x H1, HERA, DESY
- 1x HERA-B, HERA, DESY
- 1x Zeus, HERA, DESY
- 1x AMANDA, (DESY)
- 1x ANTARES, (CEA)
- 1x ARGO, YBJ-HACRL
- 1x BaBar, PEP-II, SLAC
- 1x BES III, BEPC-II, IHEP Beijing
- 1x CLEO, CESR
- 1x FLNP, IBR-2, JINR
- 1x I STRA+, IHEP Protvino
- 1x PHENIX, RHIC, BNL
- 1x SND2000, VEPP-2000, IHEP Novosibirsk

9x CERN

6x Fermilab

4x DESY

...



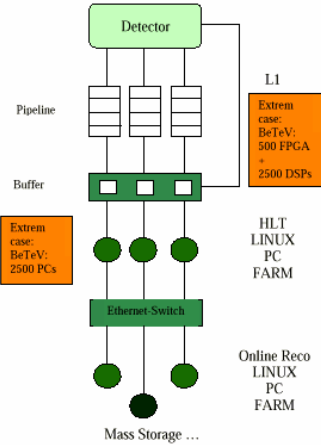
CHEP'01
Beijing

DAQ

Main topic: High Rate Systems ...

Typical solution: pipelined readout + hardware filter + software filter (PC farm)

Examples:



Existing:

HERA-B
500000 Channels
10 MHz, 90 GB/s

Hardware Filter:
10 MHz, 10 GB/s

240 PCs
L2 Software Filter (RoI):
50 KHz, 250 MB/s

L3 Software Filter (Full Data)
500 Hz, 250 MB/s

L4 Software Online Reco
50 Hz, 6MB/s

2.4 MB/s to Mass Storage

Future:

ATLAS
100000000 Channels
40 MHz x 25 Events

Hardware Filter:
75 KHz, 0(200) GB/s

400 PCs:
L2 Software Filter:
0(1 KHz), 0(4 GB/s)

200 PCs
Software Event Filter:
0(100 Hz), 0(4 GB/s)

200 MB/s to Mass Storage



CHEP'01
Beijing

DAQ

Crucial component: Farm communication using Gigabit Ethernet, Myrinet ...

Example:

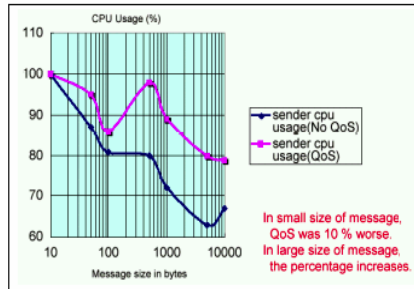
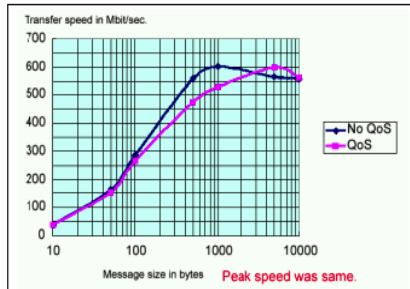
Quality of Service on Linux for the Atlas TDAQ Event Building Network:

Testsetup: 4 Linux PCs with Gigabit Ethernet plus Gigabit Switch

→ QoS can eliminate packet loss on UDP/IP multicast transfer (upto 60% without QoS)

→ Transfer speed high inspite of QoS

→ CPU usage of QoS is small on the transfer





CHEP'01
Beijing

DAQ

Readout/Run Control/Online Software...

Frequently used technologies:

OS	low level software	high level software	client-server/remote actions	GUI, histogramming, etc.	configuration	scripting (run control)
<ul style="list-style-type: none"> Vxworks (realtime) LINUX (dominating, since realtime typically not needed) Solaris 	C(++)	<ul style="list-style-type: none"> C++ Java 	CORBA <ul style="list-style-type: none"> ILU, omniORB, MICO, ... (C++) jdk1.3 CORBA (Java) 	ROOT (C++) AWT and JAS (Java)	<ul style="list-style-type: none"> databases (Objectivity, MySQL) via JDBC XML 	(J)Python

Interesting: Tendency to use many different technologies in parallel:

Examples:

- A Dataflow Meta-Computing Framework for Event Processing in H1
 - Linux PCs + C++ + Java + omniORB + jdk1.3 CORBA + ROOT + JAS + Python
- On the Way to Maturity - The CLEO III Data Acquisition and Control System
 - Vxworks/Solaris/Windows NT CPUs + C++ + Java + VisiBroker CORBA + Objectivity



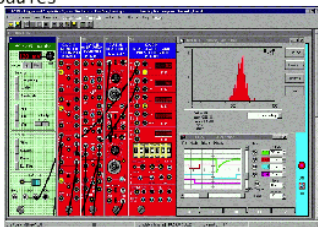
CHEP'01
Beijing

DAQ

Pleasant change to relax: Small systems and simple tools like ...

TASS: Trigger and Acquisition System Simulator

- TASS reproduces in a realistic way commercial NIM, CAMAC and VME modules
- Access to any hardware and software characteristics like with real modules



→ to help physicists developing trigger systems and students learning the fundamentals

TASS aims to be the 'bridge' joining the existing software packages in HEP
Detector/GEANT → DAQ/TASS → Analysis/PAW

RTLinux and ROOT for Data Acquisition in Small Experiments

Hardware:

- PC with ISA-to-CAMAC controller
- plus CAMAC crate etc.
- Interrupt requests are sent to PC using IRQ9

Software:

- RTLinux for hard real-time DAQ (interrupt latency < 15 μs)
- ROOT for GUI control



Offers:

- High DAQ speed
- A reliable and stable DAQ system
- Easy implementation of software design
- Low cost



CHEP'01
Beijing

CHEP'01 URLs

<u>CHEP01 home page:</u>	http://www.ihep.ac.cn/~chep01/
<u>EU Data Grid:</u>	http://www.EU-DataGrid.org
<u>Grid anatomy:</u>	
http://www.globus.org/research/papers/anatomy.pdf	
<u>IEEE task force on cluster computing:</u>	http://www.ieeetfcc.org
<u>Fermilab Large Cluster Workshop:</u>	http://conferences.fnal.gov/lccws/
<u>IGUANA Toolkit:</u>	http://iguana.cern.ch
<u>Java Analysis Studio:</u>	http://www-sldnt.slac.stanford.edu/nld
<u>Tass DAQ simulator</u>	http://tass.roma1.infn.it/