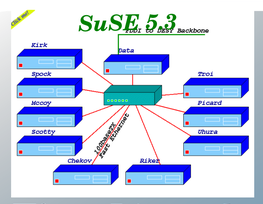
History of Computing and Data Preservation at hermes

SGI Challenge XL

Interactive analysis (28x200MHz R4400)

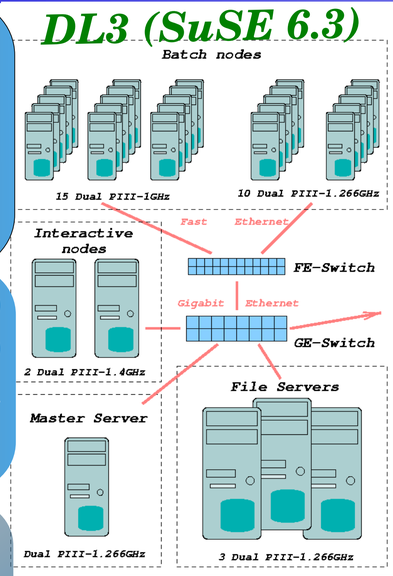
File storage (2TB disc space)

6xDEC (OSF)
9xAlpha (Ultrix)
For DAQ and Slow/fastcontrol



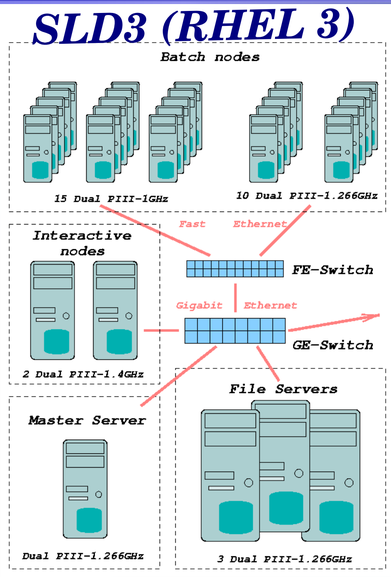
First Linux data production farm in HEP (1996)
10x200MHz Dual-PPro

DL3 (SuSE 6.3)



Interactive analysis
Batch (PBS) analysis
Data/MC production
NFS File service,Webserver
Mailing lists, Database

SLD3 (RHEL 3)



Interactive analysis
Batch (PBS) analysis
Data/MC production
NFS File service,Webserver
Mailing lists, Database

Data Preservation Phase

DESY IT

dCache

OpenAFS

Storage

Computing

BIRD

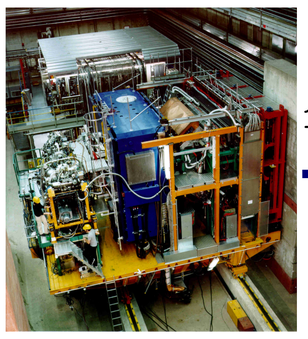
GRID

EGI

1994 — 2001 — 2007 — 2012 — 2018+

Data taking — Analyses, publishing, archiving

Data Storage



12 years data

~500 TB!

Experimental data collected and simulated through 12 years has been stored in formats of various complexity. During active years data have been stored on local NFS fileservers as well as duplicated on tapes. In the Data Preservation phase all local resources are preferably transferred to lab-supported platforms. For each level a corresponding storage type is being developed by DESY-IT to assure prompt access and safe storage based upon the usage type:

Type	Size	Availability	Storage	Value
uDST recent	7TB	Full (24x7)	NFS/dCache	Low (backup available)
uDST all	25TB	Slow (week/month)	Tape(current)	High (hard to reproduce)
MC recent	30TB	Full (24x7)	NFS/dCache	Low (easy to reproduce)
MC old	25TB	Slow (week/month)	Tape(current)	Average
RC data recent	40TB	Slow (week/month)	Tape(current)	High (hard to reproduce)
RC data old	130TB	Slow (month+)	Tape(current)	Average
Raw data	150TB	Slow (month++)	Tape(current)	High (not reproducible)
User data	2TB	Fast (24x7)	AFS/NFS	High (nightly backup)
Group data	5TB	Fast (24x7)	AFS/NFS	High (nightly backup)
Ext.Software	100GB	Fast (24x7)	AFS/NFS	High (nightly backup)

Data access via standard protocols (dCache, AFS, in future NFSv4.1) allow flexible classification of resources requiring different levels of availability and access speeds.

Software validation

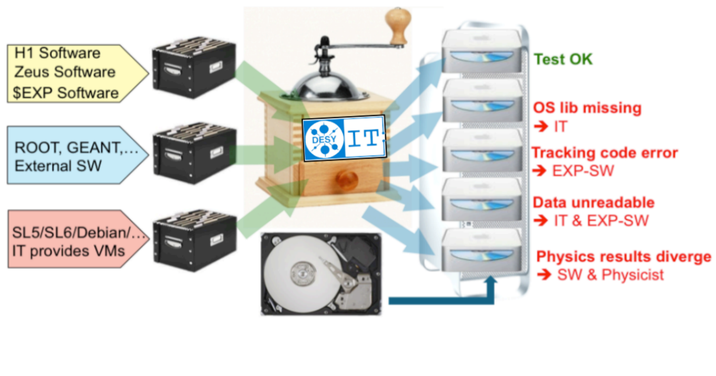
20 years of software development made the further porting to modern OSes/compilers rather complicated. Newer hardware prefers newer OSes (in the future only 64bit!). Several scenarios considered within the DPHEP community:

Experiments		Preservation Model	Use Case
	1	Provide additional documentation	Publication related info search
	2	Preserve the data in a simplified format	Outreach, simple training analyses
	3	Preserve the analysis level software and data format	Full scientific analysis, based on the existing reconstruction
	4	Preserve the reconstruction and simulation software as well as the basic level data	Retain the full potential of the experimental data

Cost, complexity and benefits

The task of level 4 software preservation is complicated due to incompatibilities between old software and new OSes/compilers. The porting to modern versions is greatly simplified by a Software Preservation project under the guidance of DESY-IT, allowing easy testing and validation of selected physics analyses under various OS/compiler combinations.

Older OSes:
SGI (IRIX)
DEC (Ultrix)
Alpha (OSF)
Intel (Linux SuSE, RHEL)
External dependencies:
ADAMO
CERNLIB
GEANT
ROOT
Dialects:
F77,C/C++,Perl, Bash, Tcl/Tk



Documentation

Analyses in the archival phase rely heavily on accurate documentation. The paper documentation of HERMES has been indexed and archived in a designated area of the DESY Library, together with other HERA experiments' materials.

Digital documentation has migrated from a local webserver to DESY-IT-central (virtual) server with support for static html and wiki.

INSPIRE comes as a new and expandable service to replace SPIRES, also allowing to host internal documentation in a protected area and link those to publications as additional resource.

A lot of important information is exchanged through mailing lists - the archived information is made available online and linked from corresponding pages on web- and wiki-pages.

In the archival period restructuring of the collaboration is inevitable. A new Collaboration Board is created to take care of responsibilities of decision-taking, effectively taking over the duties of the Council, Editorial Board and the Management.