Long term data preservation and virtualization

- Why preserving data? What to do with that old data?
- Possible usages of virtualization
- Some work done

DPHEP

Study Group for Data Preservation and Long Term Analysis in High Energy Physics

Yves Kemp, DESY IT

2nd Workshop on adapting applications and computing services to multi-core and virtualization
CERN, 22.6.2010

HELMHOLTZ ASSOCIATION

DESY

Inter-experiment Study Group on Data Preservation (DPHEP)

DPHEP-2009-001
July 30, 2009

www.dphep.org

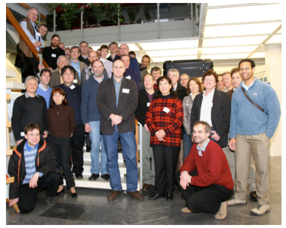
Data Preservation in High-Energy Physics

Study Group for Data Preservation and Long Term Analysis in High Energy Physics

http://dphep.org

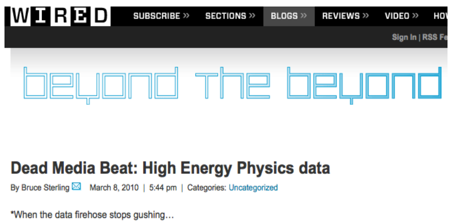
Abstract

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. At the same time, HEP has no coherent strategy for data preservation and re-use. An inter-experimental Study Group on HEP data preservation and long-term analysis was convened at the end of 2008 and held two workshops, at DESY (January 2009) and SLAC (May 2009). This document is an intermediate report to the International Committee for Future Accelerators (ICFA) of the reflections of this Study Group.



- **ICFA subgroup since 08/2009**
  - Presented at HEPAP (DOE/NSF), FALC, ICFA
  - Intermediate document released in November 2009
- **Next Steps:**
  - Next workshop KEK, July 8-10, 2010
  - New labs (JLAB), extension to other experiments (neutrino etc.)
  - Investigate concrete models, technical proposal 2010
  - Defined Resources, progress towards a common infrastructure

3 Workshops in 2009



The 2010 HEP Landscape (Colliders)

"LEP is scheduled to be dismantled soon so that its 27 km tunnel can become the home for the ambitious LHC proton collider, which is due to come into operation in 2005."
[CERN Courier, Dec. 1st, 2000]

- e^+e^-: LEP ended in 2000
  - No follow-up decided (ILC?) - after 2020
- $e^{\pm}p$: HERA end of collisions at HERA in 2007
  - No follow-up decided (LHeC?) - after 2020
- B-factories: BaBar ended in 2008, Belle → Belle II
  - Next generation in a few years (2013-2017)
- pp: Tevatron ends soon (in 2011?)
  - The majority of the physics program will be taken over at the LHC
  - However: p-pbar is unique, no follow-up foreseen

Data taking at HEP experiments takes 15-20 years, and some data are unique

- What is the fate of the collected data?

(where "data" means the full experimental information..)

Data is needed:

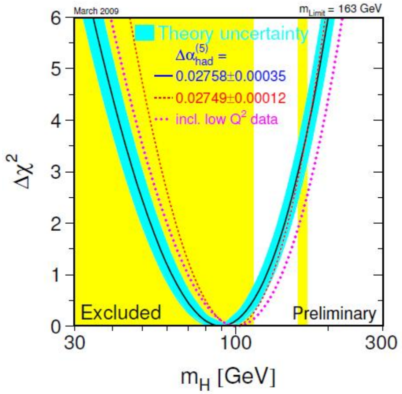
- My personal story: Around 2008, a retired professor asked the DESY data management group:
- "Around 1975*, we had tapes from a bubble chamber experiment at the Computing Center. Are these still available, maybe copied to other media? I got a request from CERN concerning these tapes."
- (we did not have them…)

- Honestly, I thought, no one would need these data anymore
  - New experiments, higher energy, better resolution, …
  - No one able to read / understand the (scientific content of the) data

- **First lesson learned here: Preserve data, and preserve the ability to perform some kind of meaningful operation on it.**

What is “meaningful operation”?

- **>** Long-term completion and extension of scientific programs
  - Allow “late analyses” to be done: +5-10% more publications
  - “Late analyses” benefit from full statistics, best understanding of systematics
- **>** Cross-collaboration analyses
  - Often performed at the end of lifetime of collaborations
  - Even among generations of experiments



- **>** Re-use of the data
  - Re-analyze the data with new theoretical models, new analysis techniques, …
- **>** Education, training and outreach
  - E.g. analysis by students without restrictions (like collaboration membership…)

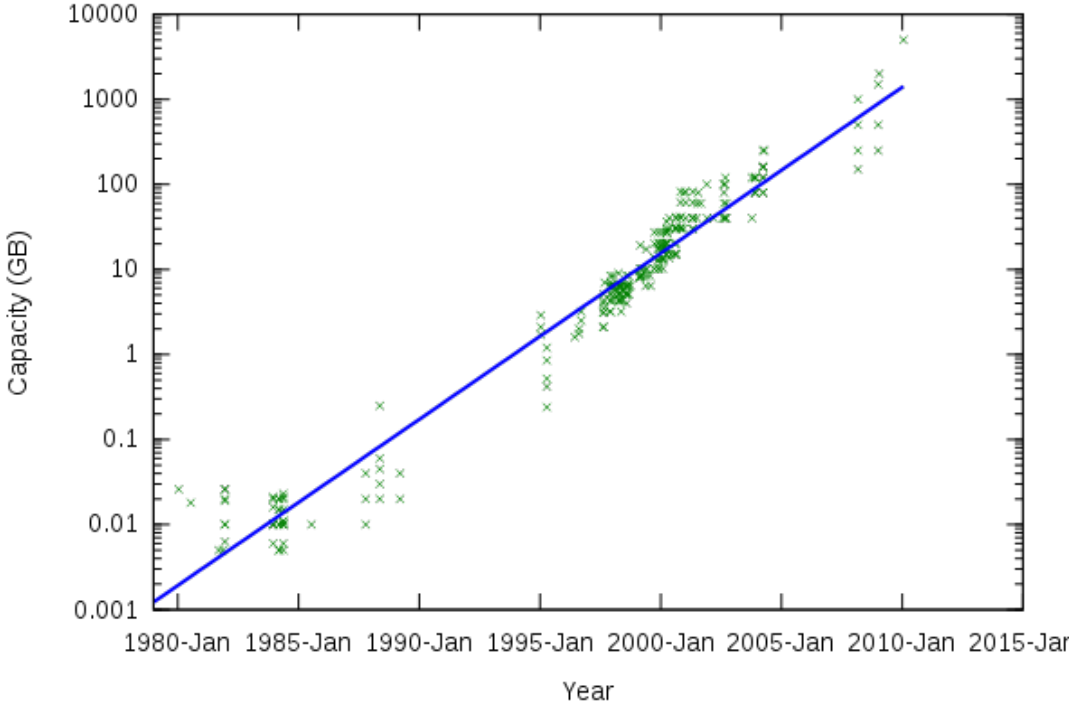
- **>** Different goals – different solutions
  - Both for the “data archival” and the “long term analysis” part

My first and very naïve ansatz

- OK, why don't we just put everything an a virtual machine?
  - Data archival is done elsewhere, just need "to plug that into the VM"
  - Your VM contains everything you need to develop and run code and analysis
- The problem would then be reduced to maintain virtual images, and maintain their ability to run. In the Cloud era, seems like a trivial task
- Problems: Everything in IT is a moving target:
  - Will your network always be the same?
  - Will your access protocol always be the same?
  - Are you sure you do not need new software (e.g. MC generators) that require a new OS?
  - Are you sure your i386/SL4 VM will produce the same results when emulated on a quantum computer in NN years?
  - What about services you need, like CondDB,…
- Naïve virtualization will not work… but still, virtualization can help

The Data Access Problem

- Data access for running or very recently finished experiments is huge
  - You know the LHC numbers. H1 e.g. (finished 2007) has ~1 PB of data
  - Need complicated systems to store and access data
- This problem will remain until you can "put all data in one machine"
  - For 1 PB of data, this could happen in 2025: All data fits in one hard drive (if you believe in the plot below)



Hard drive capacity over time
Wikimedia Commons

Some complex dependencies

Personal Analysis Code

↓

Experiment Framework

↓

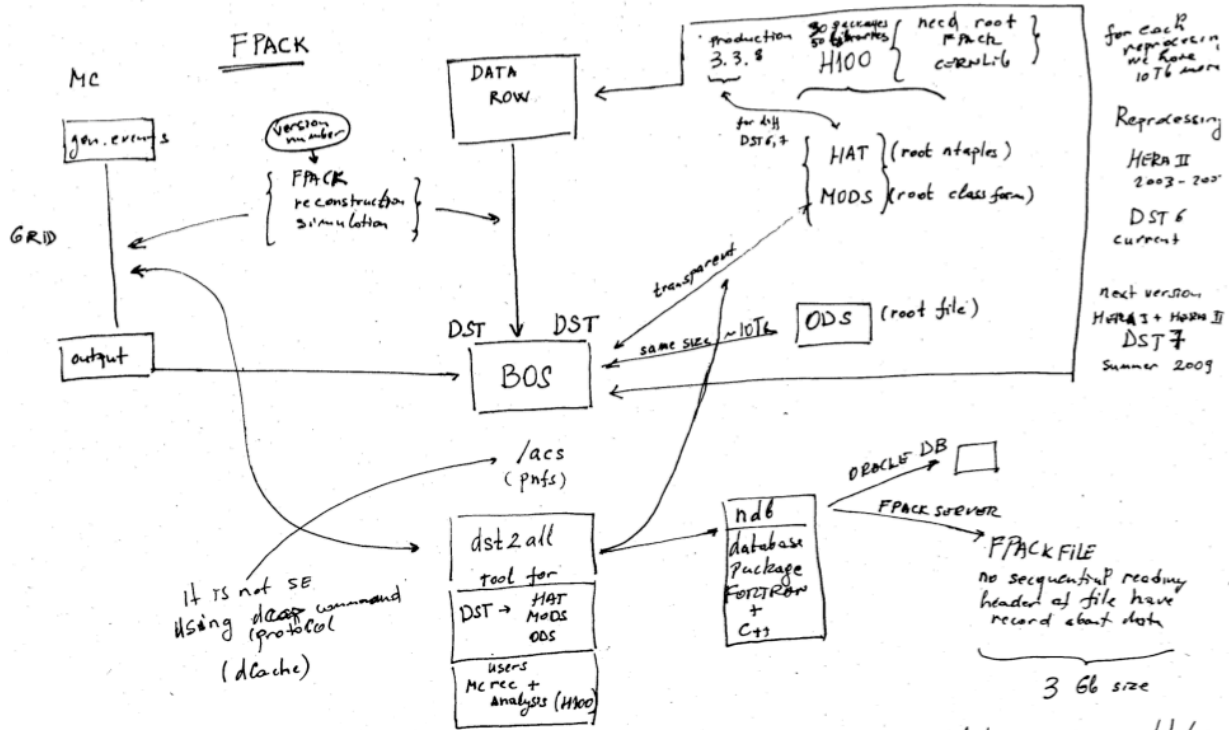
External Dependencies

- ROOT - Geant, Pythia, …
- Compiler - DB (Oracle)
- Grid / Storage Middleware
- OS & Libs ….

↓

Hardware Architecture

- x86 - x86-64
- Power
- Quantum Computers
- …



H1 Data Analysis Model
D. Ozerov, M. Mudrinic

> Will we be able to reproduce the same results after many years of not looking at the data?

- Two hypothetical antipodes involving Virtualization

Scenario 1: "Freezing"

- At the end of the experiment:
  - Datasets closed, final reprocessing done
  - Software framework stable
- Virtual image of the OS with software is done
  - Important: Use a standardized format, like OVF
- Necessary services like Cond DB.:
  - Either integrated into images
  - Or also frozen into another image
- Data access:
  - Either maintain the old protocol/interface
  - Or use high-level protocols
- Running analysis in 20NN (with NN >> 10):
  - Start the whole ensemble of VMs

Scenario 2: Continuous test-driven migration

- Start during running experiment
  - Or even before, when designing software framework
- Define tests
  - In the beginning on MC data, later real data
  - Certain code, running on certain data, yields certain result (e.g. M_{top}=172.4 GeV/c^2)
- Have an automated machinery, which regularly compiles code for different OS / architectures, and runs the tests
- If test fails (e.g. compilation or execution fails, or result divergent)
  - Manual intervention: understand (and fix) problem

→ Such automated tests are usually performed using virtualization techniques and workflows

M.C. Escher's "Waterfall"

Discussion “Freezing” / “test driven migration”

Pro Freezing

- One-time effort, very small maintenance outside of analysis phase
- Also allows software w/o code (but might fail with DRM / licensing issues)

Pro Test-driven migration

- Usability and correctness of code is guaranteed at every moment
- Data accessibility and integrity can be checked as well
- Fast reaction to standard/protocol changes
- General code quality can improve, as designed for portability and migration

Cons Freezing

- Rely on certain standards and protocols that may evolve
- Potential performance problems

Cons Test-driven migration

- Needs long-time intervention, more man-power and resources needed
- Some knowledge of the frameworks must be passed to maintainers

The BaBar Archival System

Lifetime:

It will take ~8 years for SuperB to be approved by the governments and funding agencies, constructed, commissioned and obtain a dataset more significant than BaBar's.

- **2012 to 2018 is the <u>minimum</u> required existance of the BaBar archival system**

____

There may also be a need to validate initial results against those of BaBar and Belle

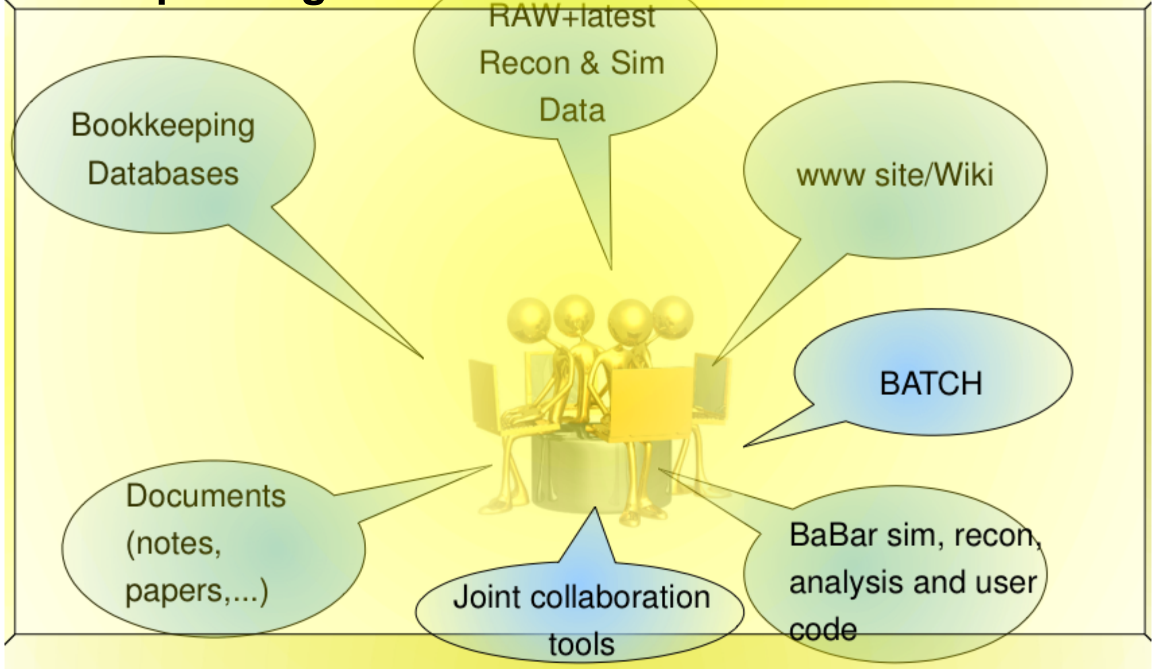
Status:

- Currently working on selecting prototype archival system hardware

32-cores, enough storage for micro data needed by most analyses

- Set it up and get testers to identify faults

Concept using Virtualization Containers:

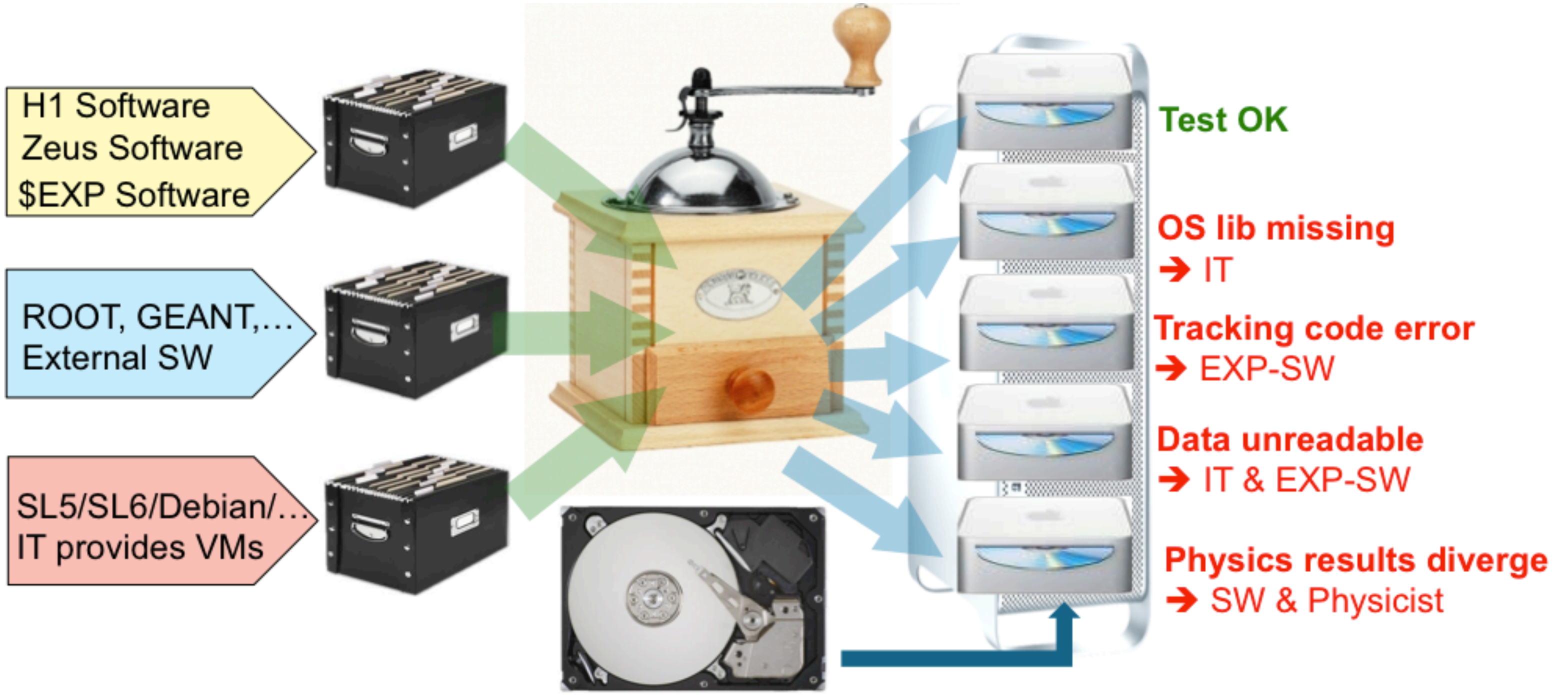


Slides from Homer Neal, @CERN 8.12.2009

Remark: BaBar has migrated to SL5 Before “going virtual”

DESY: H1, Zeus & IT development

- Automatically test SW & data
- Migration helper (as automatic as possible)
- Status: 5% mockup ready for evaluating man-power and needs
- Generic effort: Open to other experiments



Clear separation between providers of input.

Automated VM image generator provided centrally.

Tests defined by $EXP.
Test data store provided by IT.

Different VMs run SW and tests. Depending on results, different action needed.

Combination of the two models

- **Possible scenario:**
- Start with migrating software to most up-to-date OS. Best start this already during running experiment

① Use an automated machinery for testing and migrating your software

  - … as long as your data does not fit into one machine

② If data fits in one machine: Freeze everything using the most up-to-date OS.

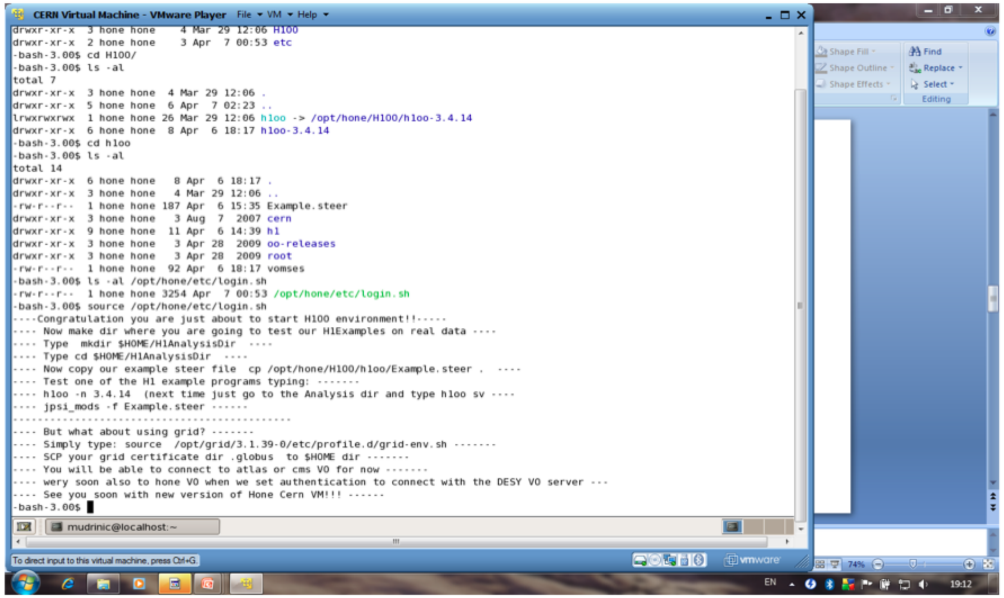
  - And preserve this VM

- What about analysis?
  - During phase 1: Easy: You always have a living system, can add current SW, might need some manpower to do the large scale analysis, but success is guaranteed
  - During phase 2: New code: Might be difficult to incorporate in an (then old) VM
  - During phase 2: Reproduce results: Easy: Problem has become "small", and is reduced to running an ensemble of VM

Acceptance problem of virtualization

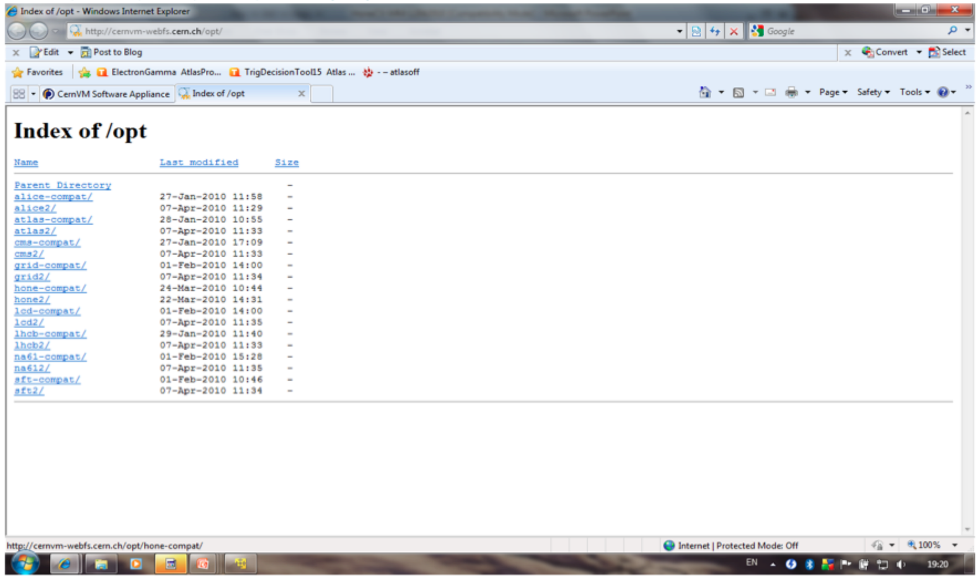
- Physicists are used to hardware. Virtualization is new, kind of a black box.
- Virtualization as a means of preserving ability to analyze precious data might have acceptance problems
- CernVM already now make VMs a reality for physicists
- At DESY, H1 (Mihajlo Mudrinic) has started some efforts together with Predrag:

At the beginning user only need to type:: source /opt/hone/etc/login.sh



Software Repository on the File Server:: http://cernvm-webfs.cern.ch

ATLAS,ALICE,CMS,LHCb,HONE(H1)



- ZEUS are producing MC on VMs, have validated them against PhysM

Summary and outlook

- Data preservation and long term analysis is an important project for HEP and an interesting field for computing science
  - HEP is not alone, but sheer amount of data makes it outstanding
  - This talk was far from complete, many more aspects to cover
- Data preservation alone is worthless, if one does not preserve the ability to perform analysis
- Different scenarios can be envisaged, two were presented that involve virtualization techniques
- Two projects are in a prototype stage, will learn from them
- … more to come
- Slides thanks to Dmitry Ozerov, David South, Mihajlo Mudrini, Cristinel Diacono and Homer Neal

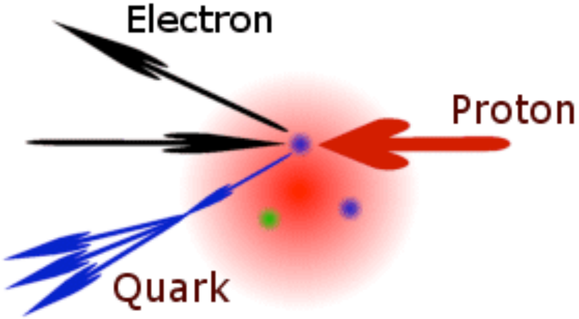
Backup slides

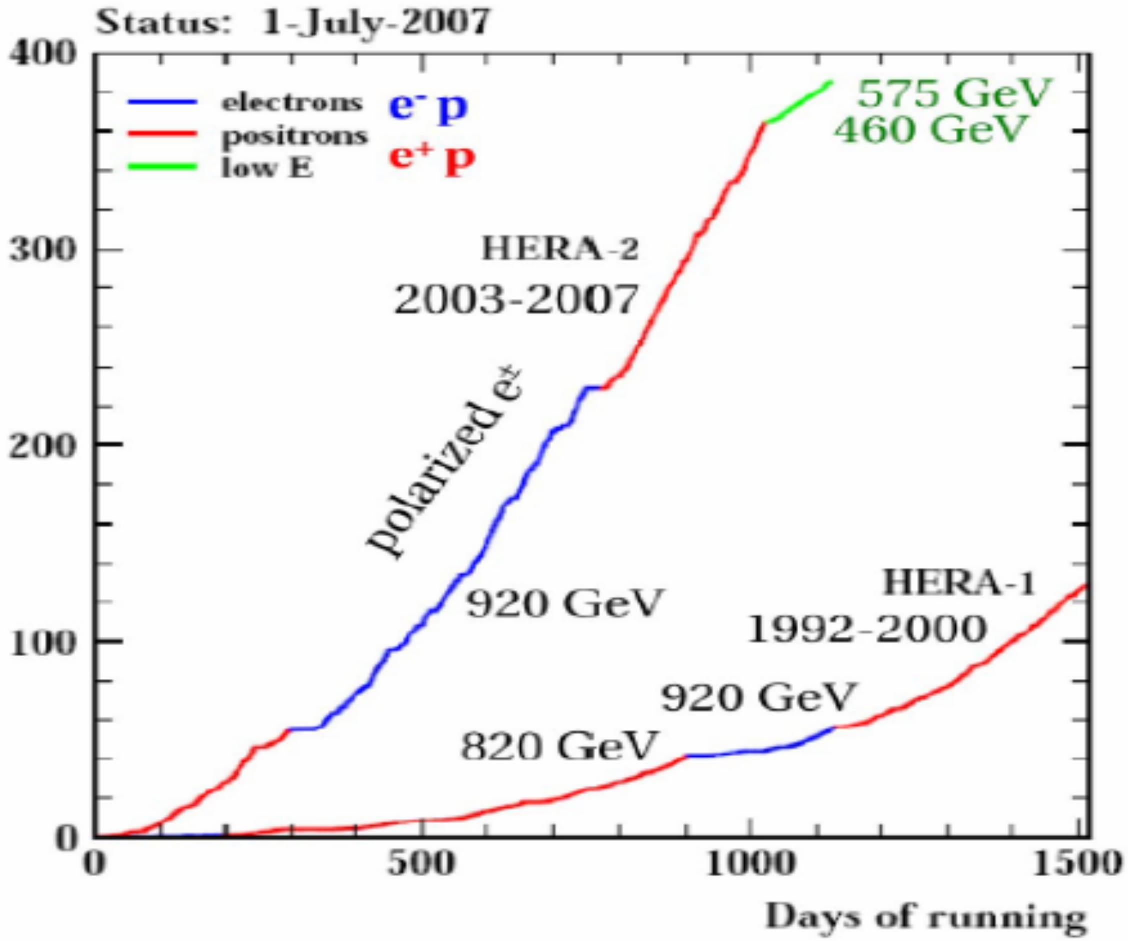
Some backup slides with

- H1 data model
- More on H1 CernVM usage

Hadron-Elektron-Ring-Anlage (HERA) at DESY

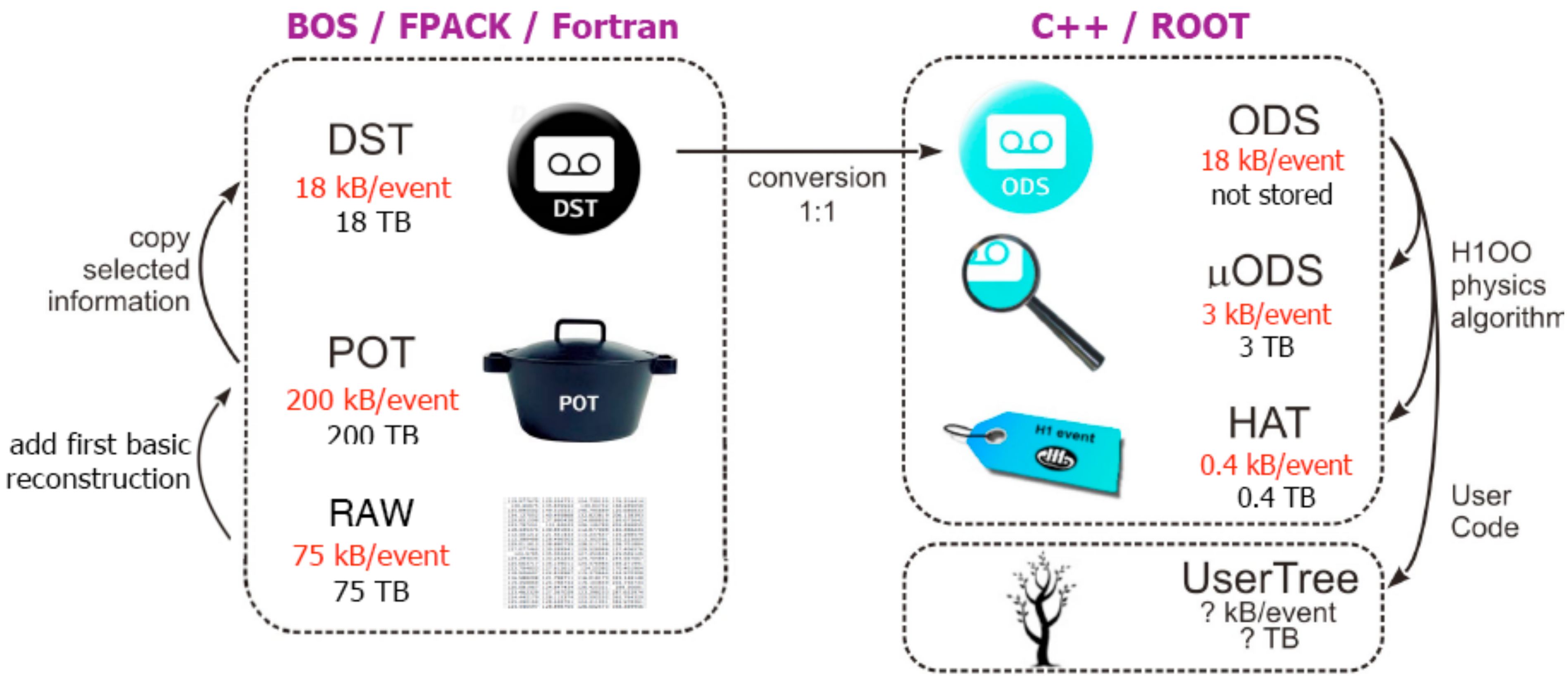
$e^{\pm}$ p collider , HERA at DESY, Hamburg, Germany.
Luminosity collected: **$0.5 fb^{-1}$ per experiment**
Ee= 27.6 GeV Ep= 920 GeV ($\sqrt{s} \approx 320$ GeV)





- Operated 1992-2007
- p: 460-920 GeV, 110 mA
- e: 27.6 GeV, 45 mA

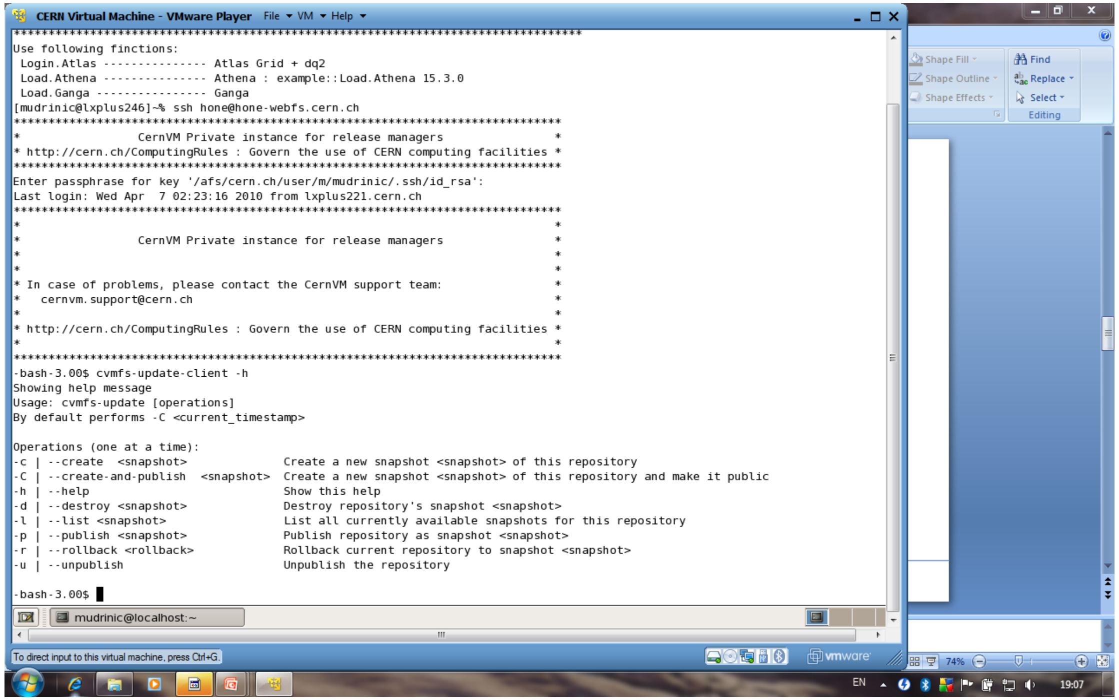
H1 Data Event Model : Present and Past



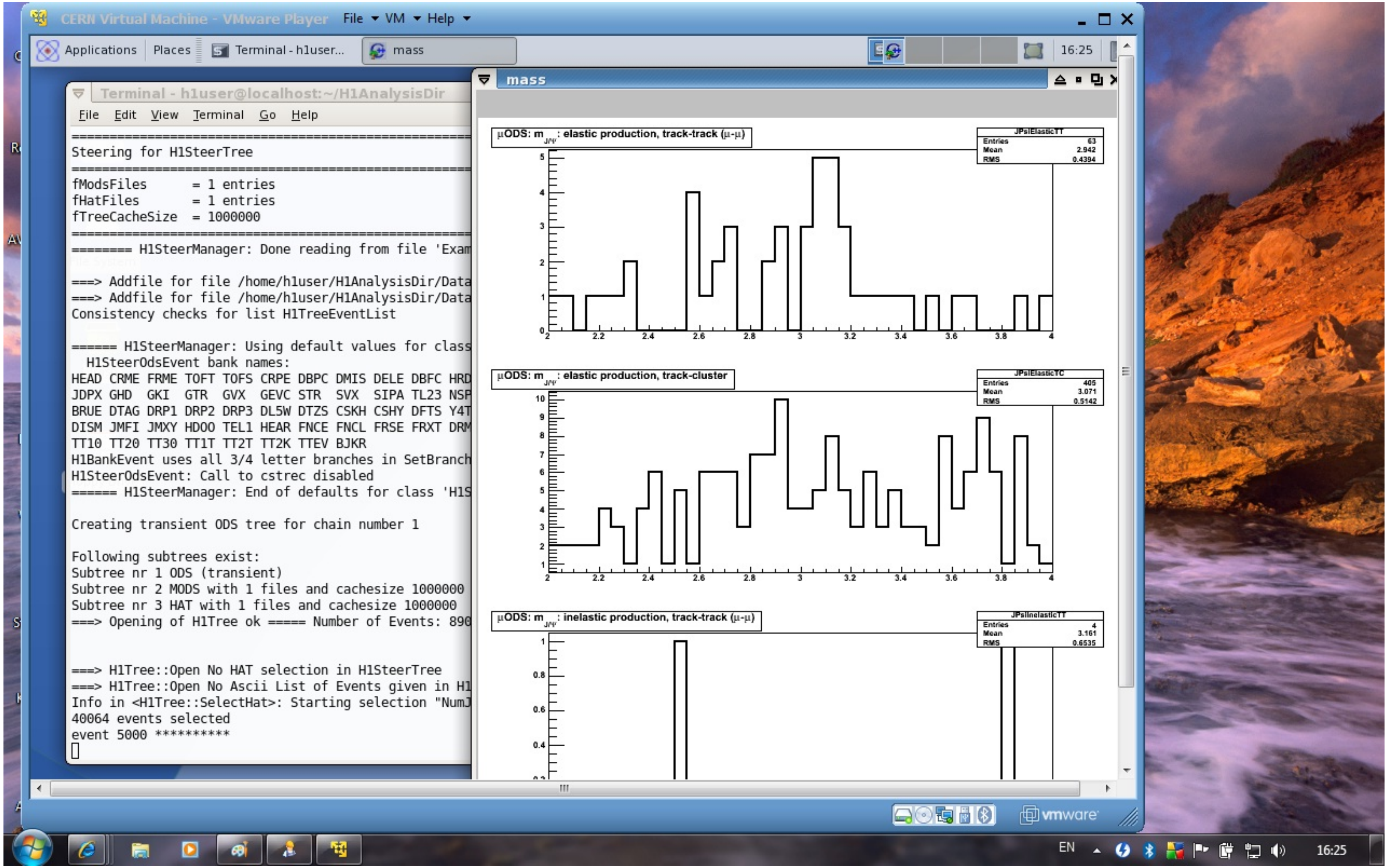
Example: HERA II data
~ 100Tb raw data => 13Tb DST => 1.3 Tb Mods/Hat (root ntuple)

H1OO Setup and Maintenance

- People from CernVM R&D project provided VM on CERN domain dedicated to H1 Experiment hone@hone-webfs.cern.ch
- hone@hone-webfs.cern.ch is accessible through lxplus.cern.ch with ssh key, and is used for testing and publishing H1OO software on CernVM File System.
- After finishing with developmant, whole image of software can be publish using command cvmfs-update-slient
- Many instances of image snapshot could be kept /deleted and publish/unpublish.



Running H1OO Analysis Framework on CernVM



Phase I: Members of the H1Collaboration (user) should gain experience and provide complaints, suggestions and ideas to H1&CernVM experts.

H1OO Virtualization Using CernVM Software Appliance. How to get H1 Collaboration OO analysis fr - Windows Internet Explorer

http://www.desy.de/~mudrim/H1OOCernVm.html

H1OO Virtualization Using CernVM Software Appliance.

How to get H1 Collaboration OO analysis framework and the grid User Interface on your laptop.

Step-by-step Instruction

- Install VMware Player or VirtualBox
- Download Latest CernVM
- Untar the file and open it with your VM Software.
- Play your CernVM image and wait until the end of boot process.
- Read out your IP Address.Fig1
- Open a web browser on your computer, and point to the IP address.Fig2
- Type user: admin password: password.
- Change the admin password.Fig3
- Setup an local user (Group Must be hone!!).Fig4
- Click on preference, "VO set to hone", open advance option and choose "enable grid user interface".Fig5
- Wait until CernVM rebootsFig6
- Login and type: source /opt/hone/etc/login.Fig7
- Good Luck!Fig8

Special Note for VirtualBox Software

Our suggestion is to use "Bridged Networking". The guest will obtain its IP address in the same way that the host does.

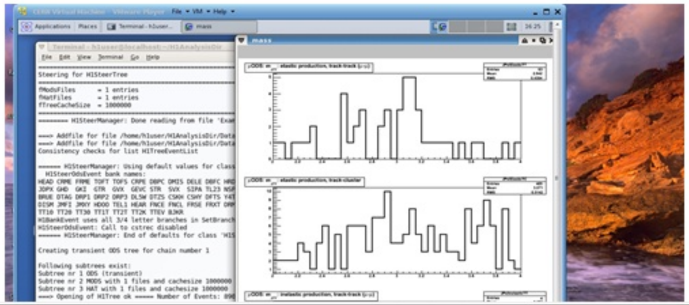
You can find on CERN wiki page special instructions for

Introduction

Virtual machine software (VMware, VitrualBox, ...) is software allowing one to run two operating systems simultaniously on a single machine (you laptop). To promote the idea of using HEP data in scientific training, education and outreach we build "virtual" H1OO Linux image using CernVM Software Applience. CernVM is a CERN based R&D project which delivers a thin Virtual Software Appliance already used by LHC experiments (ATLAS,ALICE,CMS,LHCb). We would also like to thank the members of CernVM R&D project for dedicating one VM on the CERN domain to The H1 Collaboration on which we can build and publish "virtual" Linux images with the preinstal H1OO analysis framework (release 3.4.14).

System minimum requirements

- **Windows users** : desktop or laptop PC running Windows with the VMware Player/VirtualBox software installed (free software)
- **Macintosh users** : desktop or laptop Macintosh running Mac OS with the WMware Fusion/VirtualBox software installed (shareware/free software)
- at least a **1 GHz** processor
- at least **1 Gb** of RAM for the PC or Macintosh
- **2 Gb** of free disk space available



Internet | Protected Mode: Off