Scientific computing and data preservation at MPP

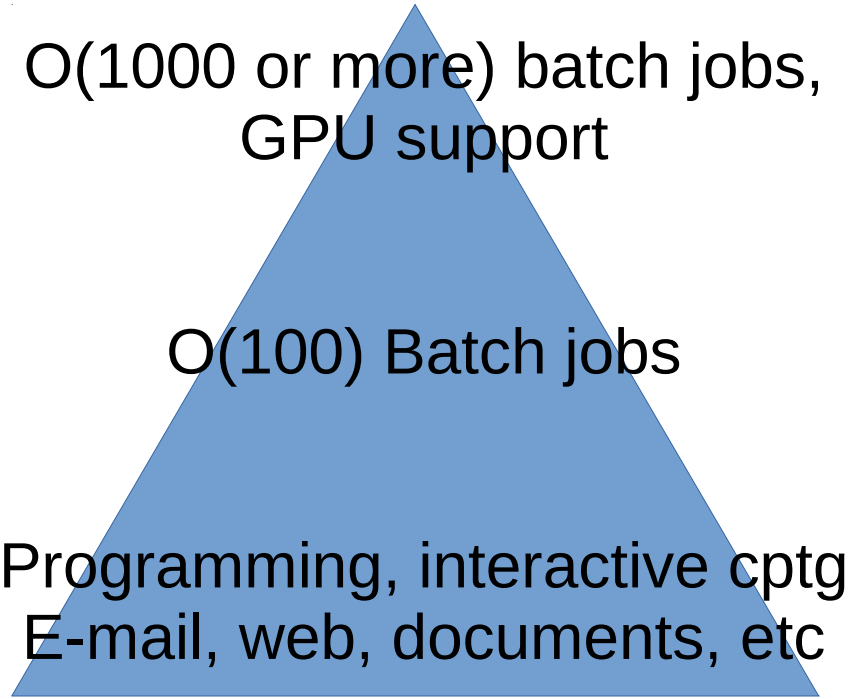
MPP Project review 2020
Stefan Kluth
14.12.2020

People

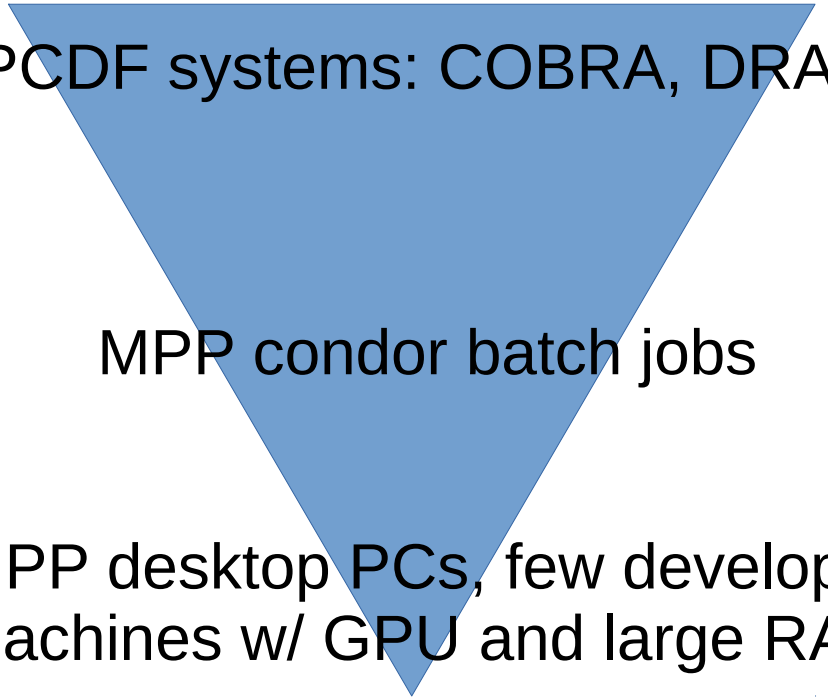
- MPP IT Fachabteilung (T. Hahn)
  - Manuel Krämer, Konstantinos Kiriakidis, Alfred Kriesel, Katrin Krebs, Uwe Leupold, Yaozhi Pan
- MPP at MPCDF (SK)
  - Cesare Delle Fratte, Sergio Tafula
- MPP Computing Commission
  - I. Abt, S. Bethke, T. Hahn, SK (chair), D. Paneque, F. Simon, O. Schulz, S. Stonjek
  - Meetings (generally) public

Overview

Activity



Compute systems



MPCDF

COBRA: 3424 compute nodes, 136,960 cores, 128 Tesla V100-32 GPUs, 240 Quadro RTX 5000 GPUs, 529 TB RAM, 7.9 TB HBM2, since 2018

DRACO: >900 compute nodes, 30.688 cores, 212 GTX980 GPUs, 128 TB RAM, since 2016

MPCDF MPP cluster

- 130 (+special) nodes, >3500 cores, 4 GB/core
  - Node groups at, bt, ft, kt, login nodes mppui1, 2, 3
- Large RAM nodes zt1, zt2 (Henn group):
  - Zt1: 6 TB, 192 cores, zt2: 3 TB, 36 cores
- More than 5 PB (soon) storage
  - /ptmp/mpp (gpfs), dCache
- CentOS7, Slurm batch, singularity, /cvmfs

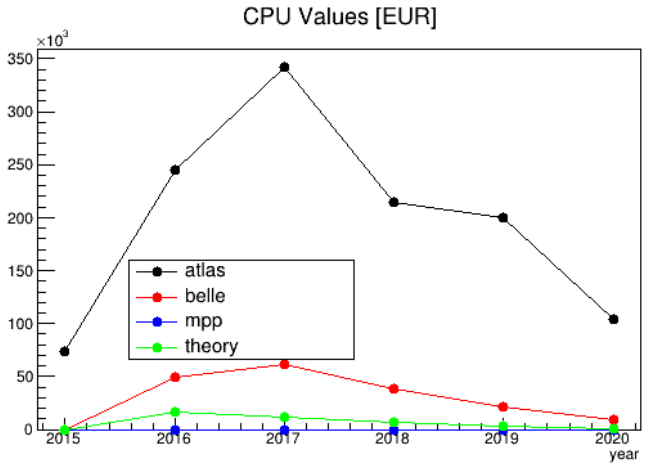
MPP desktop PCs

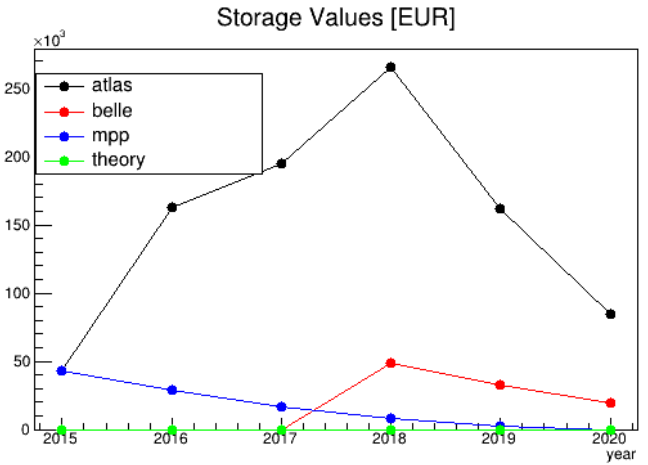
- Theory groups
  - > 80 PCs ≥8 core, 2-4 GB/core, ssd
  - opensuse tumbleweed, condor, /remote/ceph, /cvmfs
- Expt. groups
  - > 100 PCs ≥8 core, 2-4 GB/core, ssd
  - Ubuntu 18 LTS, condor, /remote/ceph, /cvmfs
- Common
  - CEPH storage > 2 PB
  - Few servers with 512-1024 GB RAM, Nvidia GPUs, R&D, local jobs

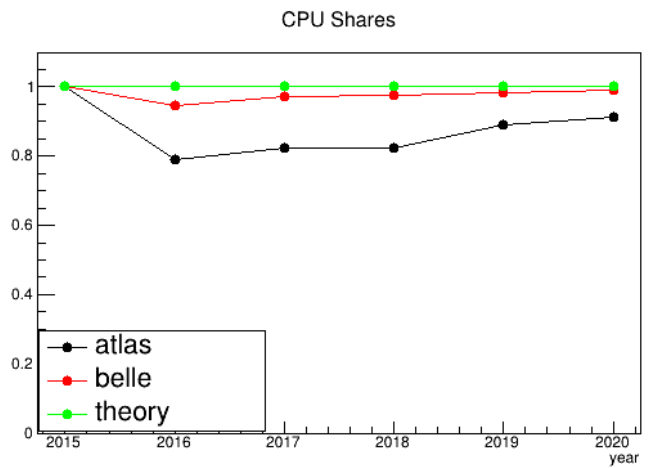
MPP 1 Linux

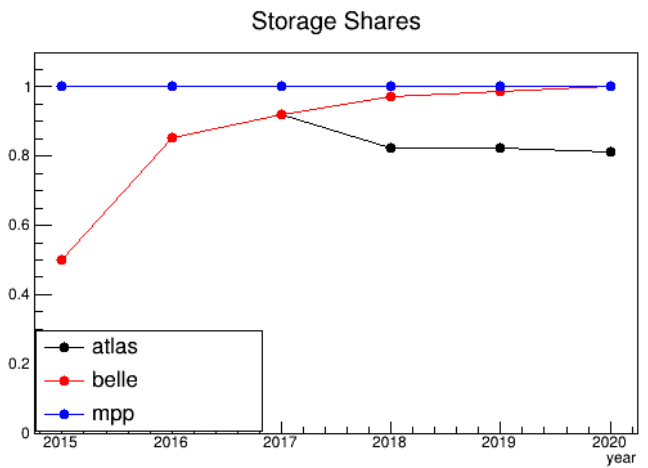
- Plan to move all desktop PCs to common (one) Linux
  - Rolling release (opensuse tumbleweed)
  - Desktops, Singularity, /cvmfs, /remote/ceph, condor
- Test machines available
  - Please test your (local) workflows
- Rolling release Linux
  - Guarantee support of modern hardware
  - Decent and recent desktop software
  - Scientific workflows in containers, data in /remote/ceph, batch jobs with condor

MPCDF MPP cluster









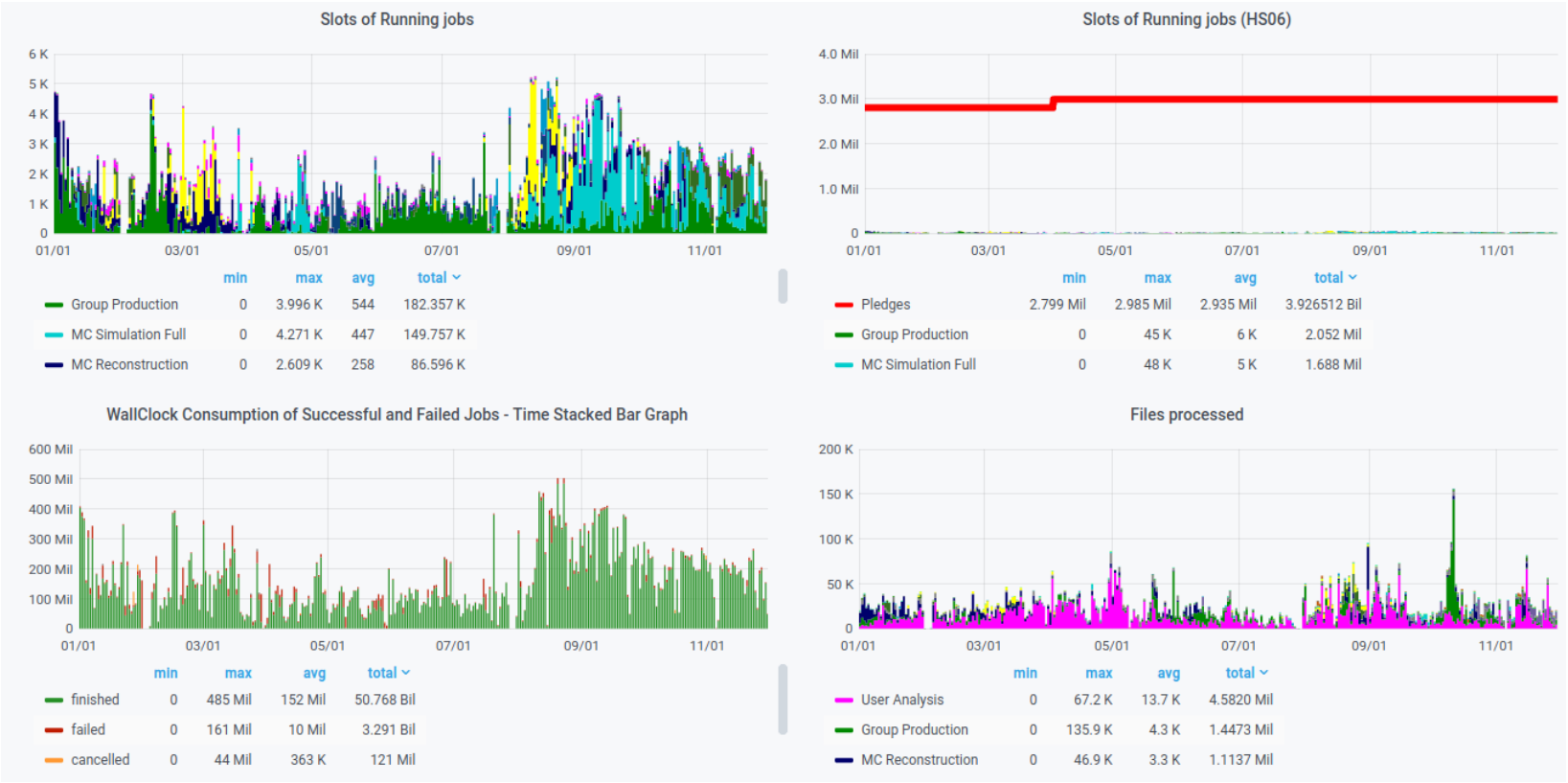
Status Sep 2020

New procurements coming:

CPU server Henn and Zanderighi groups, 96 dual AMD EPYC 32 core

New storage (dCache)

WLCG on MPP cluster @ MPCDF

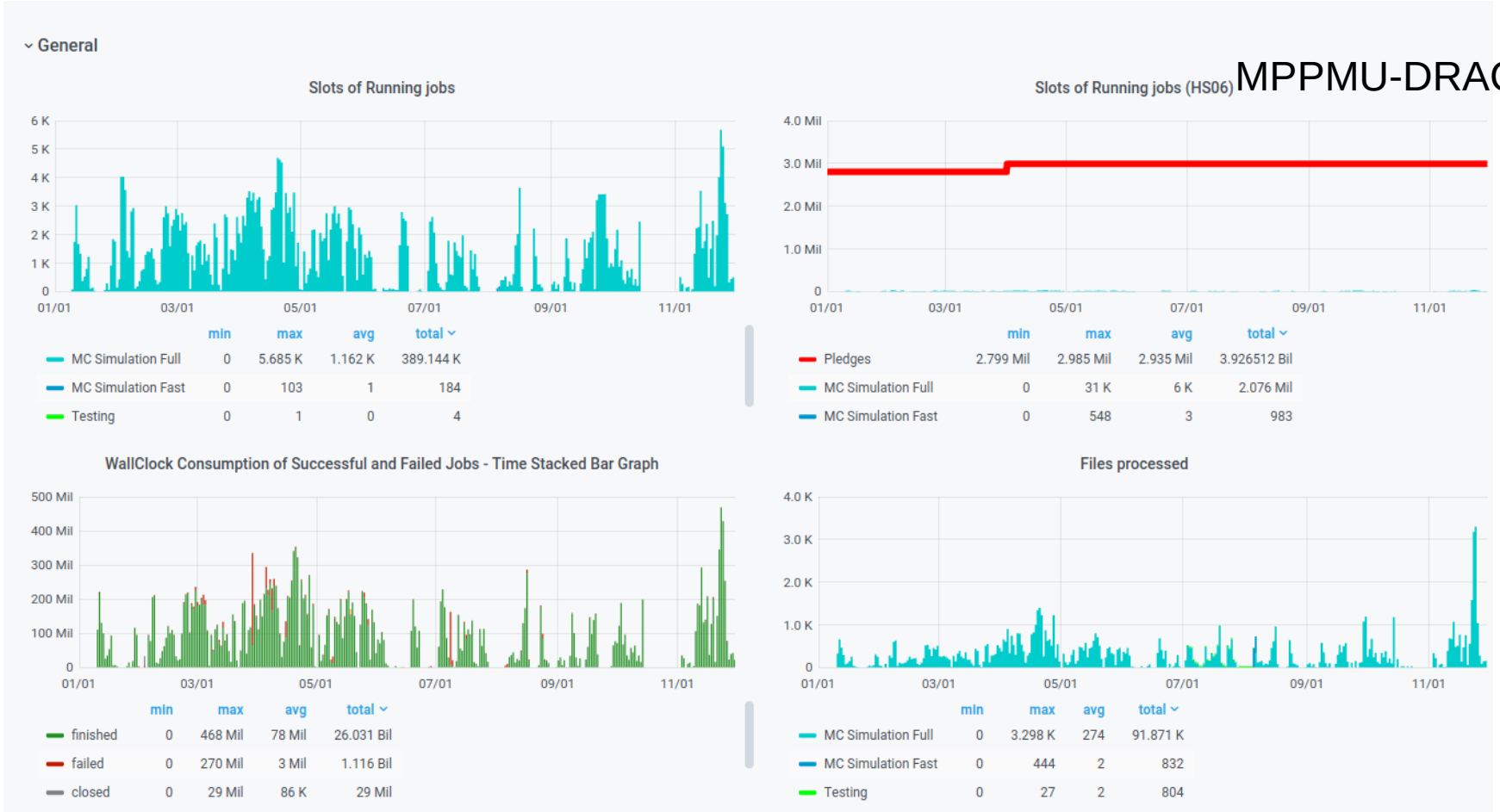


Panda queues MPPMU and ANALY_MPPMU

Pledges in HS06 wrong by factor 1000 …

Almost there due to Low WLCG share In summer

WLCG on MPP cluster @ MPCDF



MPPMU-DRACO_MCORE

Examples from Theory group

Three-photon NNLO: MPCDF MPP cluster (20 x 5000 x 24h jobs) arxiv:2006.04133

Higgs, Drell-Yan MiNNLO$_{PS}$: MPP condor and MPCDF MPP cluster (ca. 50 x 2000 x 10h jobs, many tests+debug), arxiv:2010.04681

Z γ MiNNLO$_{PS}$: COBRA 100 x 1000 x 24h jobs, arxiv:2010.10478

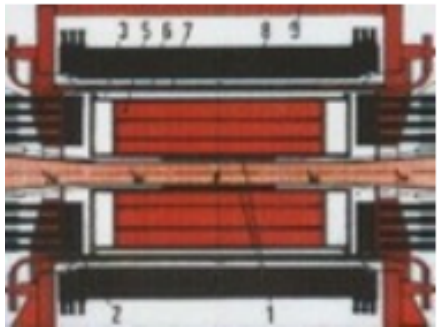
WW MiNNLO$_{PS}$: DRACO project ongoing, ca. 20 x 1000 x 12h jobs

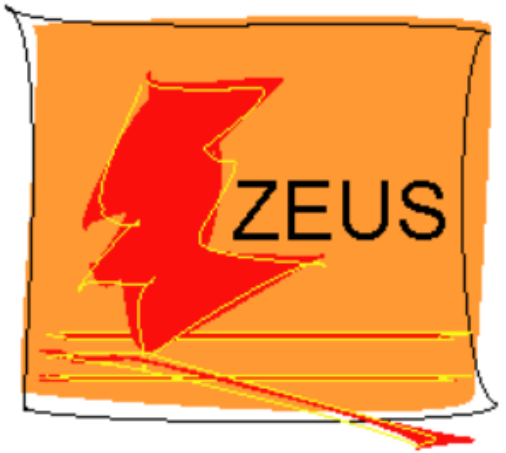
ttbar MiNNLO$_{PS}$: MPP interner Condor Cluster (project ongoing ca. 30 x 500 x 10h jobs, Mazzitelli, Monni, Nason, Re, MW, Zanderighi)

Reduction to master integrals (FIRE6): zt 10 x 1-2 weeks jobs, 2-3 TB RAM, 60 cores, arXiv:2007.04851

Evaluation of millions of polylogs (GiNaC) to 15k digits: zt ca. 10 x 24h jobs, < 1 TB RAM, 60 cores, arXiv:2003.03120

Data preservation





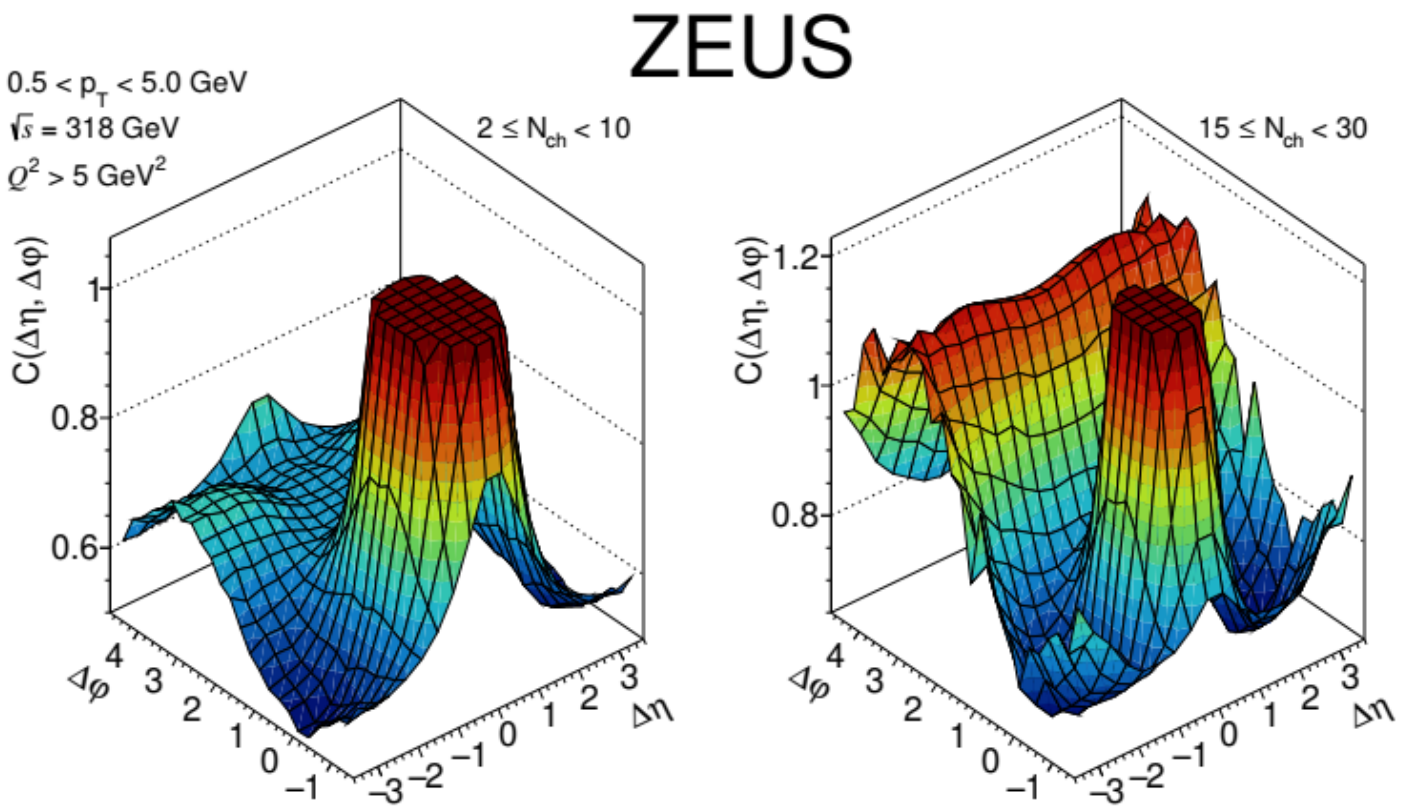
e^+e^- annihilation: JADE (PETRA 1979-1985), OPAL (LEP 1989-2000)

ep scattering: H1, ZEUS (HERA 1990-2007)

I. Abt[H], S. Bethke[J,O], D. Britzger[H], A. Caldwell[Z], V. Chekelian[H],
G. Grindhammer[H], J. Hessler[H], C. Kiesling[H], SK[J,O],
A. Verbytskyi[O,Z] (H. Abramowicz[Z], A. Levy[Z], H v.d.Schmitt[J,O])

ZEUS two-particle correlations

Ridge effect in ep? $C(\Delta\eta,\Delta\varphi) = N^{pair}_{same}(\Delta\eta,\Delta\varphi) / N^{pair}_{mixed}(\Delta\eta,\Delta\varphi)$



Near-side peak ($\Delta\varphi \approx 0$)
Particles in same jet

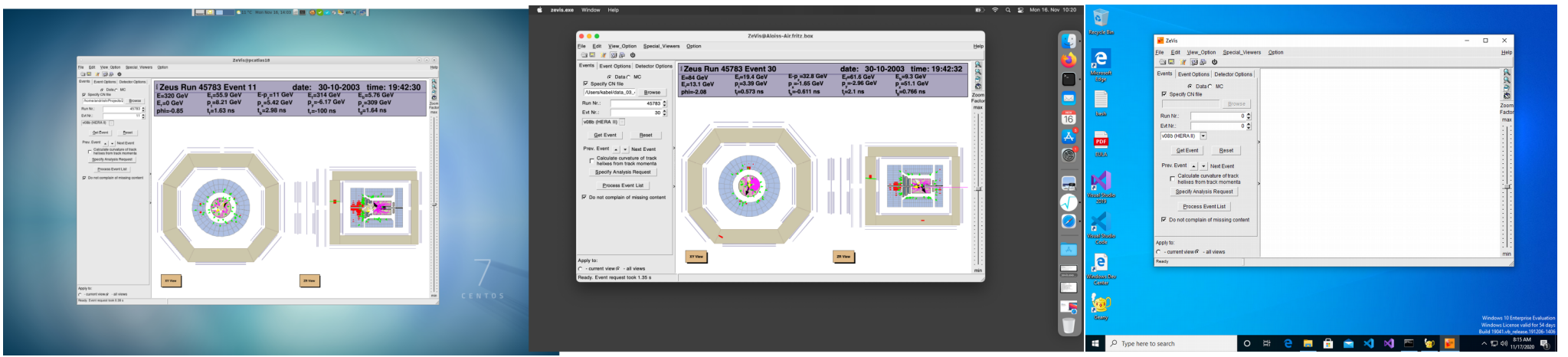
Away-side peak ($\Delta\varphi \approx \pi$)
Particles in other jet

No evidence for extra "hydrodynamic" two-particle correlations in ep (present in pp, pPb)

[JHEP 2004 (2020) 070]

DP machine room

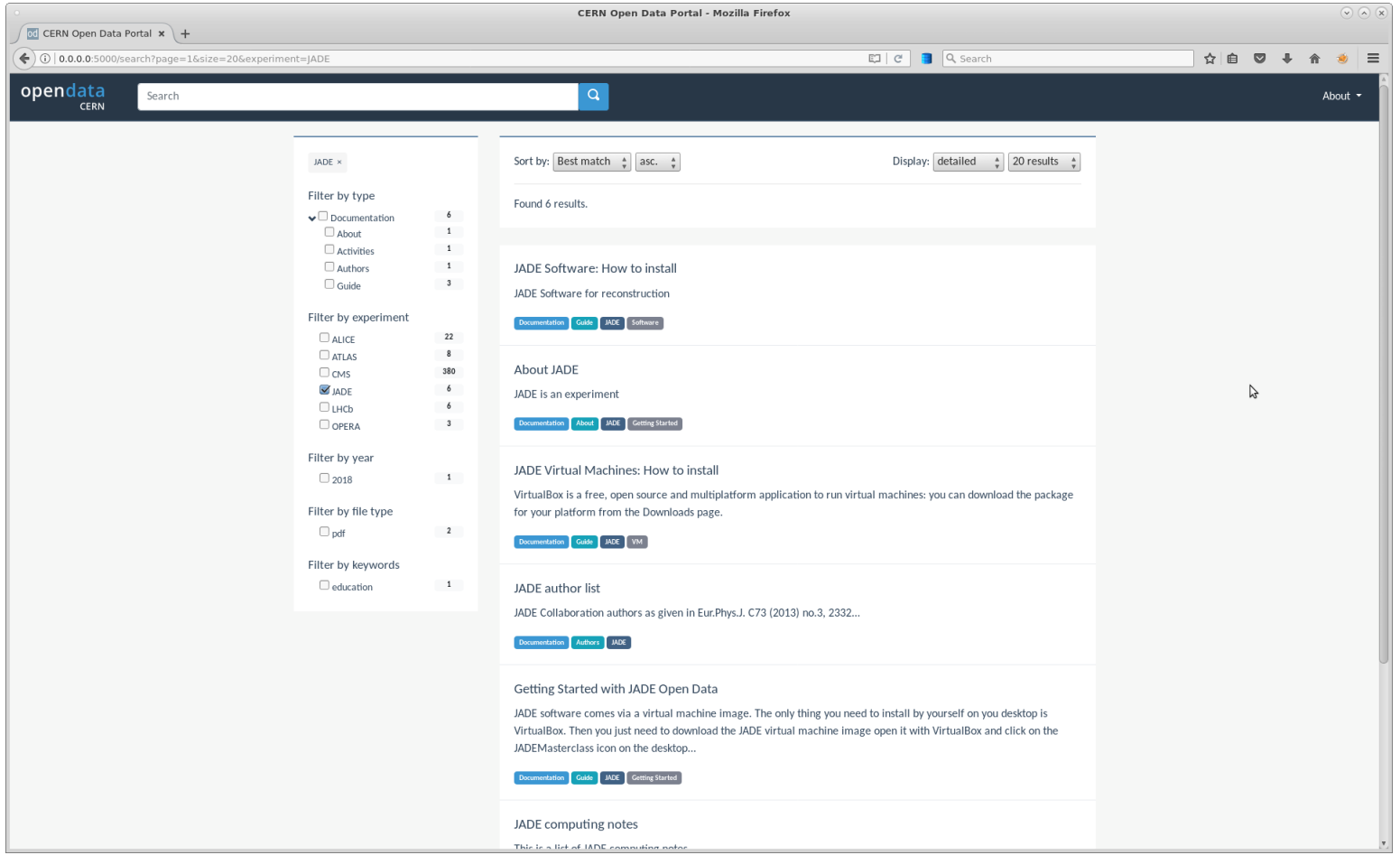
- ZEUS software
  - CNINFO (ZEUS DB): CentOS7/8, MacOS, W10
  - FORMOZA (MC / HepMC3): CentOS8
  - ZeVis (event display): C++ / Root6, cmake build, CentOS7/8, MacOS



JADE Software

- Fortran IV or 77 plus “extensions”
  - Late 70ies to 80ies
  - Now with gcc gfortran, or commercial Fortran compilers, cmake, Github CI, on CentOS7/8, MacOS
- Components
  - Ancient MCs (Jetset, Herwig 5.x, …)
  - Detector simulation (w/o Geant!)
  - Reconstruction, event display, data reduction
  - Some more recent analysis codes

JADE, OPAL at CERN opendata

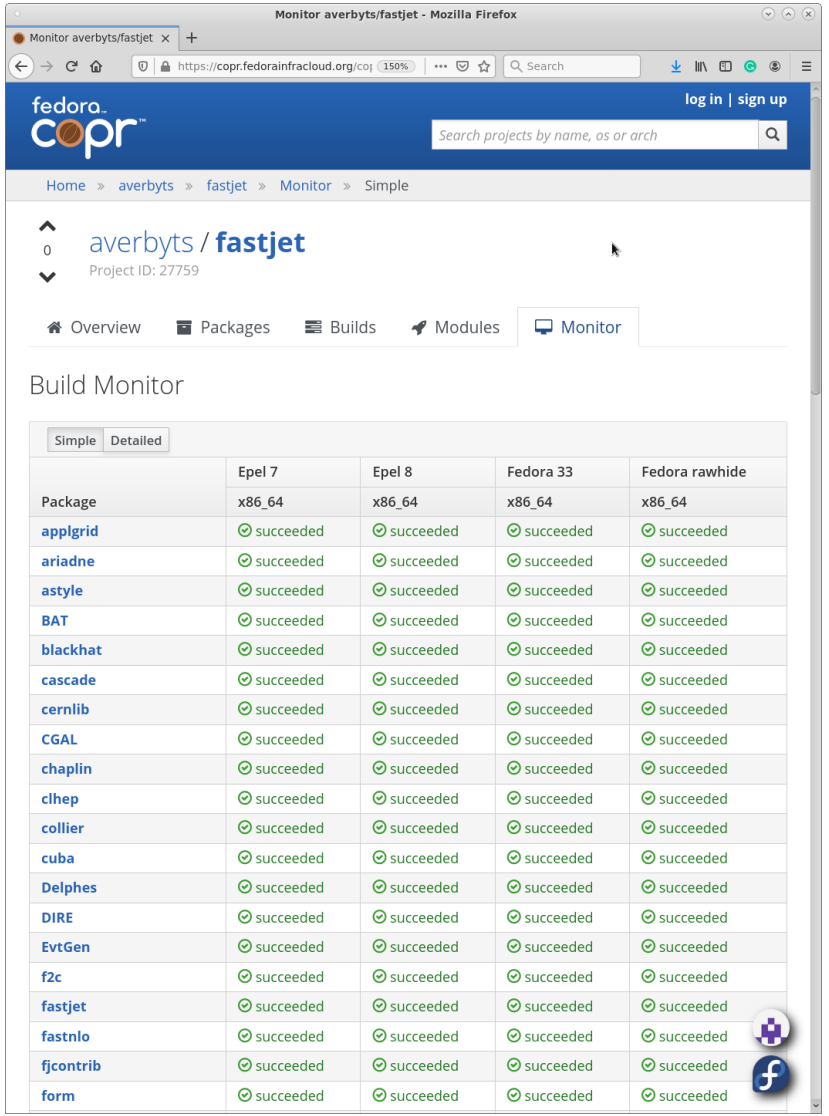


Data, software, Documentation

JADE not fully “public”, need agreement of collaboration

Could host in similar way at MPP (data already on own-cloud)

SW deployment: COPR



Package	Epel 7 x86_64	Epel 8 x86_64	Fedora 33 x86_64	Fedora rawhide x86_64
applgrid	succeeded	succeeded	succeeded	succeeded
ariadne	succeeded	succeeded	succeeded	succeeded
astyle	succeeded	succeeded	succeeded	succeeded
BAT	succeeded	succeeded	succeeded	succeeded
blackhat	succeeded	succeeded	succeeded	succeeded
cascade	succeeded	succeeded	succeeded	succeeded
cernlib	succeeded	succeeded	succeeded	succeeded
CGAL	succeeded	succeeded	succeeded	succeeded
chaplin	succeeded	succeeded	succeeded	succeeded
clhep	succeeded	succeeded	succeeded	succeeded
collier	succeeded	succeeded	succeeded	succeeded
cuba	succeeded	succeeded	succeeded	succeeded
Delphes	succeeded	succeeded	succeeded	succeeded
DIRE	succeeded	succeeded	succeeded	succeeded
EvtGen	succeeded	succeeded	succeeded	succeeded
f2c	succeeded	succeeded	succeeded	succeeded
fastjet	succeeded	succeeded	succeeded	succeeded
fastnlo	succeeded	succeeded	succeeded	succeeded
fjcontrib	succeeded	succeeded	succeeded	succeeded
form	succeeded	succeeded	succeeded	succeeded

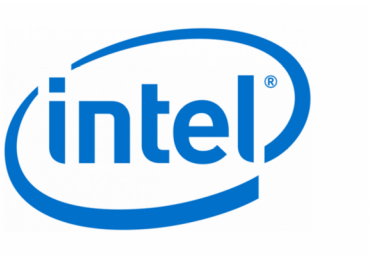
COPR: ”Community projects”

copr.fedorainfracloud.org/coprs/averbyts/fastjet

Supports CentOS7/8, Fedora, Suse, ...

Basis for stand-alone Singularity/Docker containers w/o LCG over /cvmfs, or for Github/Gitlab CI

Software preservation



vs.

Virtual machines, Linux Containers

Archive software and environment for later unmodified running. Risks: compatible hardware or VM / linux container environment disappear

Just an example, Apple Si very competitive, don’t expect every server to be on Apple Si anytime soon, but rapid transition to ARM possible

Need proper build on new platforms: cmake, CI tests, packaging (COPR), automatic deployment

Summary

- Scientific computing at MPP and MPCDF
  - Central for many theory and exptl results
  - Increasing demand for CPU and storage
  - MPP: unify Linux, /remote/ceph (> 2PB) service, condor
  - MPCDF: new CPU server, > 5 PB storage
- Data preservation
  - Many physics results (only one example shown)
  - Rely on software (and documentation) preservation