

CernVM Users Workshop

Thursday, 5 March 2015 - Friday, 6 March 2015

CERN

Book of Abstracts

Contents

Welcome and Introduction	1
Welcome and Introduction	1
Status and Roadmap of CernVM	1
Status and Roadmap of CernVM-FS	1
CernVM-FS Interface to S3	1
CernVM Web-API	1
Feedback from ATLAS	1
CernVM for CMS open data	2
Feedback from LHCb	2
Feedback from OSG	2
Feedback from Asia	2
Feedback from EGI	2
Integrating VMs into WLCG	2
Docker Containers and Their Emerging Roles in Distributed Applications	2
Personalized Opportunistic Computing for CMS at Large Scale	3
Volunteer Computing Projects at CERN	3
Long-Term Data Preservation	4
Status and Roadmap of Vac and Vcycle	4
Distributed Systems: The Hard Problems	4
Hybrid cloud environments and networking on AWS	4
Big Data in the Cloud - Processing and Performance	4
Wrap-up	5
LHCb VM-Dirac (TBC)	5

The Data Bridge: Spanning trusted and untrusted environments	5
CVMFS as a High Speed Filesystem for Auxiliary Data	5
WLCG Stratum 0	5
CernVM elastic clusters and feedback from ALICE	6
Feedback from CMS	6
CernVM: Notes from the GridPP new user engagement programme	6

0

Welcome and Introduction

Corresponding Author: gerardo.ganis@cern.ch

Status of the CernVM Ecosystem / 1

Welcome and Introduction

Corresponding Author: gerardo.ganis@cern.ch

Status of the CernVM Ecosystem / 2

Status and Roadmap of CernVM

Corresponding Author: jakob.blomer@cern.ch

Status of the CernVM Ecosystem / 3

Status and Roadmap of CernVM-FS

Corresponding Author: rene.meusel@cern.ch

Status of the CernVM Ecosystem / 4

CernVM-FS Interface to S3

Corresponding Author: seppo.heikkila@cern.ch

Status of the CernVM Ecosystem / 5

CernVM Web-API

Corresponding Author: ioannis.charalampidis@cern.ch

Feedback from experiments and users / 7

Feedback from ATLAS

Corresponding Author: douglas.benjamin@cern.ch

Feedback from experiments and users / 8

CernVM for CMS open data

Corresponding Author: katri.lassila-perini@cern.ch

Feedback from experiments and users / 9

Feedback from LHCb

Corresponding Author: ben.couturier@cern.ch

Feedback from experiments and users / 10

Feedback from OSG

Corresponding Author: jose.caballero@cern.ch

Feedback from experiments and users / 11

Feedback from Asia

Corresponding Author: felix@twgrid.org

Feedback from experiments and users / 12

Feedback from EGI

Corresponding Author: catalin.condurache@cern.ch

Feedback from experiments and users / 13

Integrating VMs into WLCG

Corresponding Author: andrew.mcnab@cern.ch

Technological trends / 14

Docker Containers and Their Emerging Roles in Distributed Applications

Docker has stormed into the IT landscape recently as a new abstraction for linux containers. Its ease of use is set to revolutionize the way we package applications, share and deploy them on compute clusters. In this presentation we will discuss Docker principles, illustrate the new application life cycle that Docker offers and give an overview of the booming Docker ecosystem. We will then introduce several container management tools aimed at supporting Docker based applications in clusters. Specifically we will present Kubernetes and show how Mesos is playing its part. With this new IT landscape emerging, we will present some ideas on how these new technologies might impact LHC experiments and their applications.

About the author.

Sebastien Goasguen built his first compute cluster in the late 90's when they were still called Beowulf clusters while working on his PhD; he has been working on making computing a utility since then. He has done research in grid computing and high performance computing and with the advent of virtualization moved to cloud computing in the mid 2000s. While at Purdue and Clemson he was involved in the OSG and TeraGrid project and worked on CERN LXCLLOUD as a scientific associate in 2009 and 2010. He is currently a Senior Open Source Solutions Architect at Citrix, where he works primarily on the Apache CloudStack project, helping develop the CloudStack ecosystem. Sebastien is a project management committee member (PMC) of CloudStack and Apache libcloud and a member of the Apache Software Foundation, he focuses on the cloud ecosystem and has contributed to Knife-cloudstack, Eutester and Ansible. He is the author of O'Reilly 60 recipes for CloudStack and is currently writing the O'Reilly Docker cookbook.

Technological trends / 15

Personalized Opportunistic Computing for CMS at Large Scale

Corresponding Author: dthain@cse.nd.edu

In recent years, the WLCG has evolved toward a model of centralized scheduling of largely dedicated clusters provided by each member institution, missing some opportunities to harness opportunistically available resources. This talk will present our recent experience in building Lobster, an opportunistic workload management system for the elements of the CMS collaboration at Notre Dame, where our prototype system scavenging cycles from a Tier-3 center runs on O(10K) cores, which is on the scale of the largest Tier-2 centers. I will discuss the challenges inherent in this environment, explain how Parrot, CVMFS, and Work Queue have evolved to meet these needs, and suggest ways to exploit emerging technologies such as containers within the context of scientific workflows.

About the author.

Douglas Thain is an Associate Professor of Computer Science and Engineering at the University of Notre Dame, where he designs large scale distributed computing systems to power the needs of advanced science and engineering research in fields such as bioinformatics, high energy physics, and molecular dynamics. His team publishes the Cooperating Computing Tools, a collection of open source software that provides application frameworks, workflow systems, and data access tools for clusters, clouds, and grids. Prof. Thain received the Ph.D. from the University of Wisconsin, where he contributed to the HTCondor high-throughput computing system.

CernVM related activities / 16

Volunteer Computing Projects at CERN

Corresponding Author: nils.hoimyr@cern.ch

CernVM related activities / 17

Long-Term Data Preservation

Corresponding Author: frank.berghaus@cern.ch

Status of the CernVM Ecosystem / 18

Status and Roadmap of Vac and Vcycle

Corresponding Author: andrew.mcnab@cern.ch

Technological trends / 20

Distributed Systems: The Hard Problems

The focus of this presentation will be around some interesting issues that arise in distributed systems, and how Riak and applications that use it go about solving them.

About the author.

Nicholas Bellerophon works as a client services engineer at Basho Technologies, helping customers setup and run distributed systems at scale in the wild. He has also worked in massively multiplayer games, and recently completed a live scalable simulation engine. He is an avid TED-watcher with interests in many areas of the arts, science, and engineering, including of course high-energy physics.

Technological trends / 21

Hybrid cloud environments and networking on AWS

Author: Giulio SORO¹

¹ *Amazon Web Services*

The presentation will be around AWS networking services and solutions to connect on premises data centers with the AWS Cloud, as well as infrastructure by code capabilities and strategies.

About the author:

Giulio Soro is a Solutions Architect within the EMEA Public Sector Team of Amazon Web Services. Previously he worked as Technical Project Manager in the Sardinian Public Sector, supervising regional datacenter management and private cloud computing projects. Giulio holds a PhD in Electronics Engineering and a Master of Science in Telecommunications Engineering from University of Cagliari.

Technological trends / 22

Big Data in the Cloud - Processing and Performance

This talk will cover how big data and big computation is being handled in the Cloud. It gives practical examples using GCP including tools that apply like Google BigQuery, DataFlow, Container Engine/Kubernetes, and ways to evaluate their performance using Google's open source benchmarking tool PerfKitBenchmarker.

About the author.

Anthony F. Voellm is currently leading the Google Cloud Performance Team and has a wide range of experience from kernel and database engines to graphics and automated image and map extraction from satellite images. Anthony is an avid inventor with 7 technology patents issued. In his current role at Google Anthony is focused on delivering Price Performance to existing products like Google Compute Engine and Google Cloud Storage while also innovating new offerings. Anthony holds a Master of Science from George Washington University, BA in Physics and a BS in Computer Science and Mathematics from the University of Vermont.

Discussion / 23

Wrap-up

Corresponding Author: jakob.blomer@cern.ch

24

LHCb VM-Dirac (TBC)

Corresponding Author: cinzia.luzzi@cern.ch

CernVM related activities / 25

The Data Bridge: Spanning trusted and untrusted environments

Corresponding Author: laurence.field@cern.ch

CernVM related activities / 26

CVMFS as a High Speed Filesystem for Auxiliary Data

Corresponding Author: dwd@fnal.gov

Feedback from experiments and users / 27

WLCG Stratum 0

Corresponding Author: steve.traylen@cern.ch

Status of the CernVM Ecosystem / 28

CernVM elastic clusters and feedback from ALICE

Corresponding Author: dario.berzano@cern.ch

Feedback from experiments and users / 29

Feedback from CMS

Corresponding Author: andrew.david.lahiff@cern.ch

Feedback from experiments and users / 30

CernVM: Notes from the GridPP new user engagement programme

Corresponding Author: tom.whyntie@cern.ch