EPJ Web of Conferences **214**, 03035 (2019) https://doi.org/10.1051/epjconf/201921403035
CHEP 2018

Deploying and extending CMS Tier 3s using VC3 and the OSG Hosted CE service

Kenyi Hurtado Anampa[1]*, *Kevin* Lannon[1], *Douglas* Thain[1], and *Ben* Tovar[1]

[1]University of Notre Dame, Notre Dame, IN, USA

Abstract. CMS Tier 3 centers, frequently located at universities, play an important role in the physics analysis of CMS data. Although different computing resources are often available at universities, meeting all requirements to deploy a valid Tier 3 able to run CMS workflows can be challenging in certain scenarios. For instance, providing the right operating system (OS) with access to the CERNVM File System (CVMFS) on the worker nodes or having a Compute Element (CE) on the submit host is not always allowed or possible due to e.g: lack of root access to the nodes, TCP port network policies, maintenance of a CE, etc. The Notre Dame group operates a CMS Tier 3 with 1K cores. In addition to this, researchers have access to an opportunistic pool with +25K cores that are used via lobster for CMS jobs, but cannot be used with other standard CMS submission tools on the grid like CRAB, as these resources are not part of the Tier 3 due to its opportunistic nature. This work describes the use of VC3, a service for automating the deployment of virtual cluster infrastructures, in order to provide the environment (user-space CVMFS access and customized OS via singularity containers) needed for CMS workflows to work. Also, we describe its integration with the OSG Hosted CE service, to add these resources to CMS as part of our existing Tier 3 in a seamless way.

1 Introduction

The Worldwide LHC Computing Grid (WLCG) [1] is composed of 4 layers or "tiers" that provide a specific set of services. Local computing resources used to perform the final stages of data analysis by individual university groups are defined as Tier 3s in this infrastructure.

The University of Notre Dame high energy physics group operates a CMS Tier 3 with about 1,300 cores for analyzing CMS [2] data submitted locally or through the grid. In addition to this, the Center for Research Computing (CRC) provides an opportunistic campus cluster with over 25,000 cores of computing power researchers have access to (as shown in Figure 1), but these resources lack the software components and environment needed by CMS analysis workflows.

This work describes the use of VC3 [3], a service for automating the deployment of virtual cluster infrastructures and the OSG Hosted CE service [4], in order to provide the grid environment and components needed to build a CMS Tier 3 using Notre Dame opportunistic campus resources at the user level.

---

*e-mail: khurtado@nd.edu

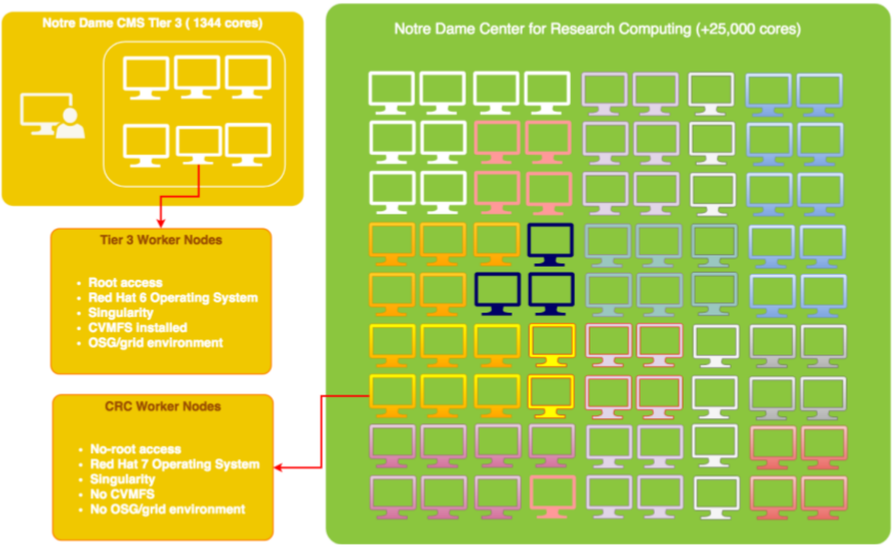


Figure 1. Notre Dame CMS Tier 3 and CRC computing resources

2 VC3: Virtual Clusters for Community Computation

Traditional HPC and campus computing facilities provide considerable amounts of computing power in a fixed environment designed to address local needs. However, this makes the deployment of complex applications that can span multiple sites and require specific software components very difficult.

The VC3 project allows users with one or more site resource allocations to address these issues, facilitating the aggregation of different resources and installation of custom software and cluster middleware needs in user space. The user sees a virtual cluster with a uniform environment with access to a VC3 head node to submit jobs.

Figure 2 shows the architecture of VC3. Users interact with a web portal in order to request the creation of a virtual cluster with one or more resource allocation accounts listed on it. This information is propagated to the system with an info service. A master process creates a head node for the user and a factory instance submits pilot jobs to the resources in the virtual cluster. A public SSH key is also provided by the user through the portal, which is used to give the user access to the VC3 dynamic head node.

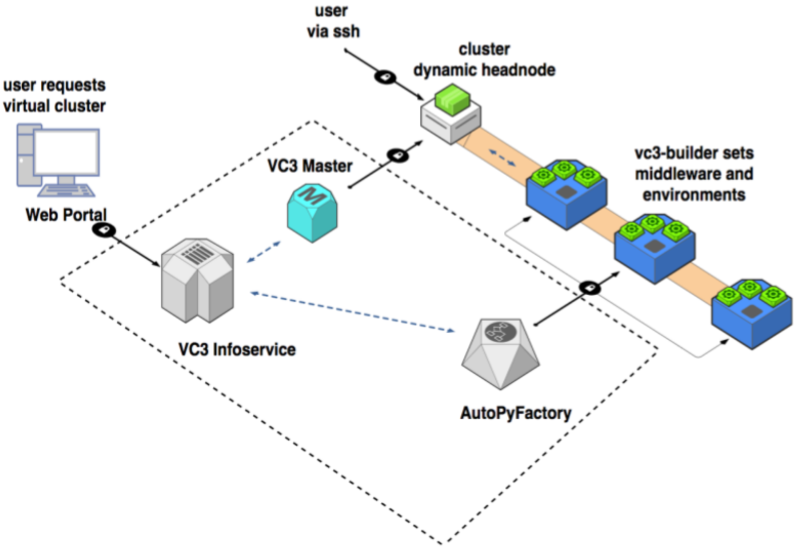


Figure 2. VC3 architecture

3 The OSG Hosted CE service

The Open Science Grid (OSG) [5] is a distributed computing infrastucture designed to facilitate access to high throughput computing resources, located mostly at universities and national laboratories within the US, for large-scale scientific research. Requests to a resource in this infrastucture are materialized as *pilot jobs* running within the resource batch system using the GlideinWMS system [6].

The entry point used for such pilots jobs to reach a resource is HTCondor-CE [7], a job gateway based on HTCondor [8] and specially configured to transform and submit these jobs to the local batch system.

The OSG Hosted CE service is an HTCondor-CE gateway working over SSH. Pilot Jobs are submitted to the remote cluster batch system through an external host running HTCondor-CE and submitting via Bosco [9], as shown in Figure 3.

The Compute Element installation, configuration and management is hanlded directly by OSG. The CE authenticates to an unprivileged user account in the remote cluster submit host using SSH private/public key pairs.

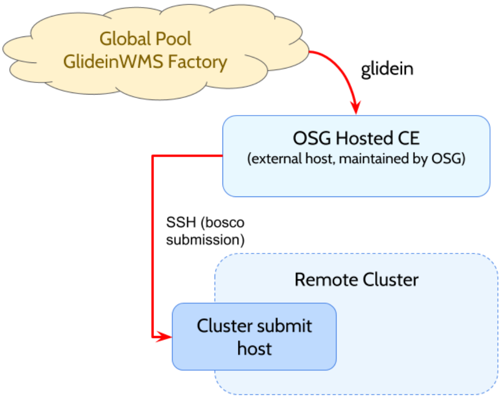


Figure 3. OSG Hosted CE diagram, showing remote submission of pilot (glidein) jobs to a remote cluster submit host over SSH.

4 Deploying a CMS Tier 3 with VC3 and the OSG Hosted CE service

Several components in the worker node environment are needed in order for CMS analysis workflows to work properly in the computing resource. A list of these components is detailed below and shown in Figure 4.

- **The CernVM File System (CVMFS)** [10]: The different CMS software releases are provided using this mechanism.
- **Grid transfer tools**: Utility tools like GFAL or the XRootD client are used to transfer data remotely .
- **Virtual Organization Membership Service components**: The VOMS client is needed to generate grid user proxies for authentication in order to access CMS data.
- **CMS information for the Site**: A set of files containing specific information for the site need to be defined and stored via CVMFS and soft-linked to `/cvmfs/cms.cern.ch/SITECONF/local`.
- **Red Hat 6 or 7 with (optionally) singularity**: CMS software releases are compiled for RHEL 6 and 7 operating systems, which need to be supported either natively or through containers.
- **HTCondor-CE**: This is needed in order for the CMS Global Pool factory to submit pilot jobs to the local resource.

Most of the software components above cannot be directly installed and configured on the Notre Dame CRC campus cluster due to lack of system administrator privileges to the resource. While Lobster [11], a workflow management tool developed at the University of Notre Dame and designed to harness these opportunistic computing resources, can be used to provide the proper CMS environment locally, this does not deal with CMS grid jobs with other standard submission tools in CMS like CRAB [12].

In order to address this issue, an extra thin layer deployed with VC3 is added in order to provision all required CMS components in user space. Additionally, the OSG Hosted CE

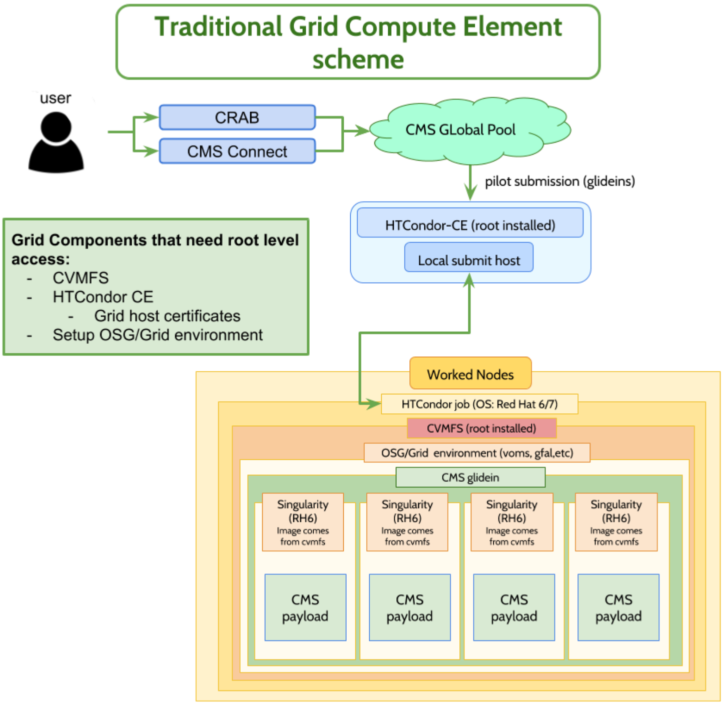

Traditional Grid Compute Element scheme
user
CRAB
CMS Connect
CMS GLobal Pool
pilot submission (glideins)
HTCondor-CE (root installed)
Local submit host
Grid Components that need root level access:
- CVMFS
- HTCondor CE
- Grid host certificates
- Setup OSG/Grid environment
Worked Nodes
HTCondor job (OS: Red Hat 6/7)
CVMFS (root installed)
OSG/Grid environment (voms, gfal,etc)
CMS glidein
Singularity (RH6) Image comes from cvmfs
CMS payload

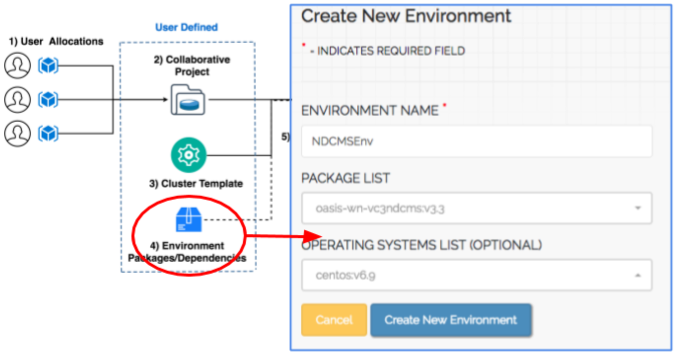
Figure 4. Tier 3 components - Traditional version

service is used in order to provide the job gateway component without requiring anything other than a user account in the Notre Dame CRC submit host.

Figure 5 shows the CMS environment creation for the virtual cluster through the web portal, defining CentOS v6.9 via Singularity and the OSG Worker Node environment. CMVFS is provided via parrot by the VC3 builder and links the site configuration to `/cvmfs/cms.cern.ch/SITECONF/local`, as required by CMS. Figure 6 shows a VC3 head node created from the request, connected to the CRC submit host via SSH. CMS pilot jobs are submitted to the VC3 head node through the OSG Hosted CE component, which later eventually run on CRC computing resources with the proper environment needed by CMS.

Once the CMS software and proper environment in the virtual cluster is in place and a CE for the cluster has been deployed, standard procedure to add this site to the CMS Global Pool factory can be followed. This allows submission services in the experiment like CRAB to match to this resource and submit to it.

The monitor plots in Figure 7 show the CRAB usage for the Notre Dame non-dedicated resources as T3_US_VC3_NotreDame in the CMS dashboard, with scales and efficiencies similar to that of the Tier 2s.



Create New Environment

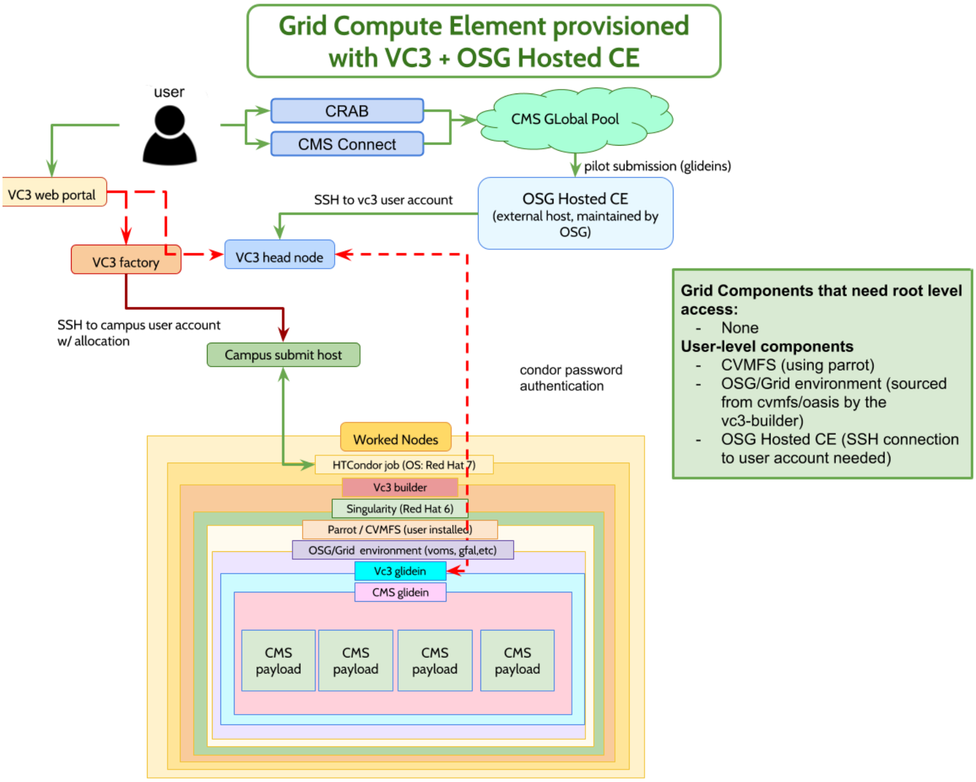
* - INDICATES REQUIRED FIELD

ENVIRONMENT NAME *: NDCMSEnv

PACKAGE LIST: oasis-wn-vc3ndcms:v3.3

OPERATING SYSTEMS LIST (OPTIONAL): centos:v6.9

Figure 5. Creating a Virtual Cluster environment


Grid Compute Element provisioned with VC3 + OSG Hosted CE
user
CRAB
CMS Connect
CMS GLobal Pool
pilot submission (glideins)
VC3 web portal
SSH to vc3 user account
OSG Hosted CE (external host, maintained by OSG)
VC3 factory
VC3 head node
SSH to campus user account w/ allocation
Campus submit host
condor password authentication
Worked Nodes
HTCondor job (OS: Red Hat 7)
Vc3 builder
Singularity (Red Hat 6)
Parrot / CVMFS (user installed)
OSG/Grid environment (voms, gfal,etc)
Vc3 glidein
CMS glidein
CMS payload
CMS payload
CMS payload
CMS payload

Grid Components that need root level access:
- None

User-level components
- CVMFS (using parrot)
- OSG/Grid environment (sourced from cvmfs/oasis by the vc3-builder)
- OSG Hosted CE (SSH connection to user account needed)

Figure 6. Tier 3 components - VC3 plus OSG Hosted CE version

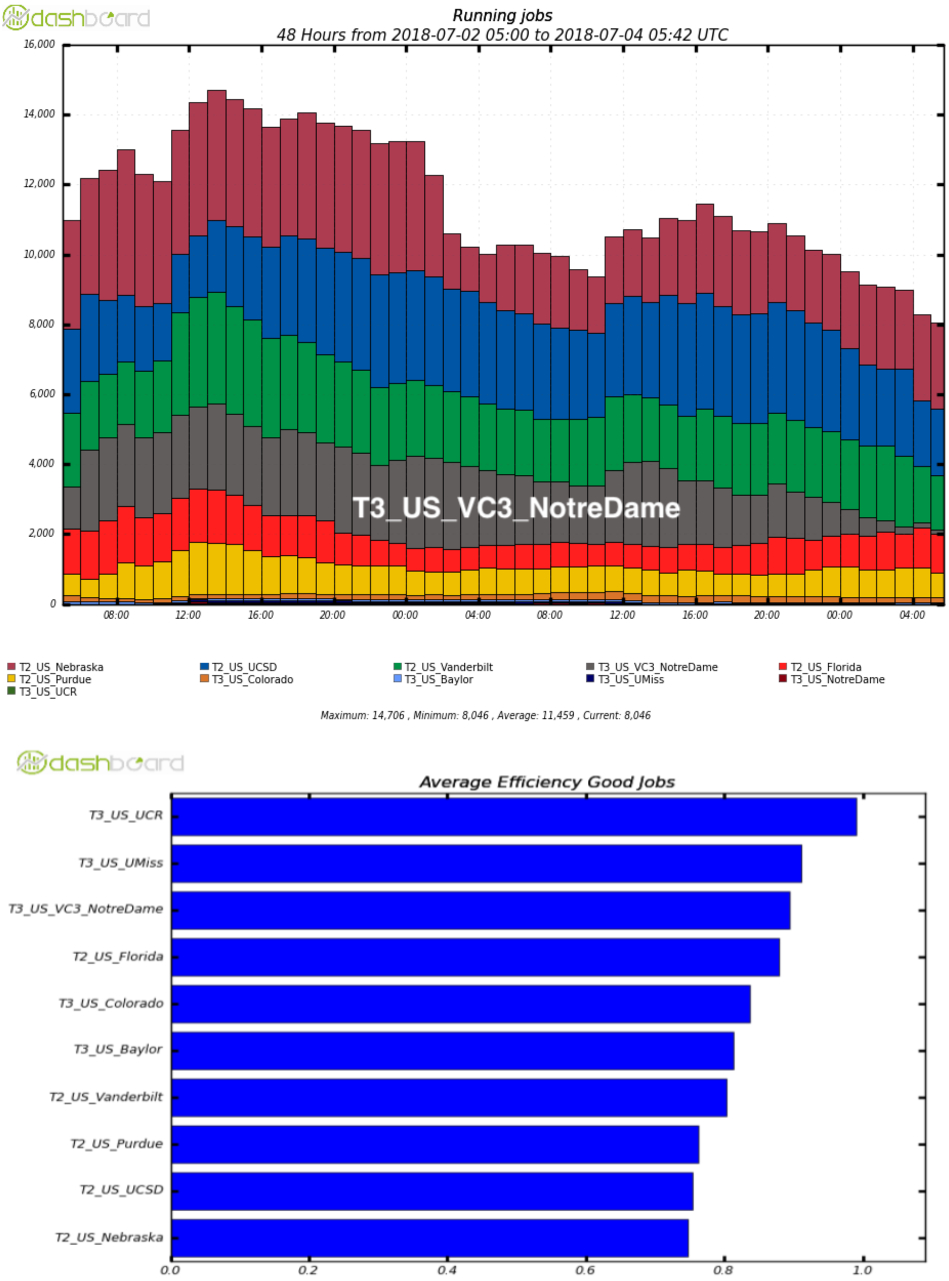


Figure 7. CRAB analysis activity in the CMS Dashboard

5 Conclusions

This work has shown the needs and challenges present in deploying a CMS Tier 3 on opportunistic resources researchers might have access to, but with limited privileges, as well as a solution to it using VC3 and the OSG Hosted CE service. The use of VC3 allowed us to provide several software components needed by CMS analysis workflows in user space, while the OSG Hosted CE service provided a simple way to connect this virtual cluster to the grid (CMS Global Pool) by only requiring SSH access to the opportunistic resource.

VC3 is supported by the Department of Energy NGNS program via grant DE-SC0015711. In addition to VC3, we are grateful to the OSG team for assisting us with the OSG Hosted CE service, without whose support this work would not have been possible.

References

[1] http://wlcg-public.web.cern.ch

[2] S Chatrchyan *et al.* (The CMS Collaboration), The CMS experiment at the CERN LHC, JINST **3**, S08004 (2008)

[3] L Bryant, J Van, B Riedel, R Gardner, J Caballero Bejar, J Hover, B Tovar, K Hurtado and D Thain, VC3: A Virtual Cluster Service for Community Computation, PEARC, **30**, 1-8 (2018)

[4] https://opensciencegrid.org/docs/compute-element/htcondor-ce-overview/#hosted-htcondor-ce-over-ssh

[5] R Pordes *et al.*, The Open Science Grid, J. Phys. Conf. Ser., **78**, 012057 (2007)

[6] I Sfiligoi, GlideinWMS: a generic pilot-based workload management system, J. Phys. Conf. Ser. , **119** 062044 (2008)

[7] B Bockelman, T Cartwright, J Frey, E M Fajardo, B Lin, M Selmeci, T Tannenbaum and M Zvada, Commissioning the HTCondor-CE for the Open Science Grid, J. Phys. Conf. Ser., **664** 062003 (2015)

[8] D Thain, T Tannenbaum and M Livny, Condor and the Grid, *Grid Computing: Making The Global Infrastructure a Reality*, ISBN: 0-470-85319-0 (2003)

[9] D Weitzel, I Sfiligoi, B Bockelman, J Frey, F Wuerthwein, D Fraser and D Swanson, Accessing opportunistic resources with bosco, J. Phys. Conf. Ser. , **513** 032105 (2014)

[10] J Blomer, C Aguado-Sánchez, P Buncic and A Harutyunyan, Distributing LHC application software and conditions databases using the CernVM file system, J. Phys. Conf. Ser., **331** 042003 (2011)

[11] A Woodard, M Wolf, C Mueller, B Tovar, P Donnelly, K Hurtado Anampa, P Brenner, K Lannon, M Hildreth and D Thain, Exploiting volatile opportunistic computing resources with Lobster, J. Phys. Conf. Ser. **664** 032035 (2015)

[12] M Cinquilli, D Spiga, C Grandi, J M Hernandez, P Konstantinov, M Mascheroni, H Riahi and E Vaandering, CRAB: Establishing a new generation of services for distributed analysis at CMS, J. Phys. Conf. Ser., **396(3)** 032026 (2012)