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

Kenyi Hurtado Anampa, Ben Tovar, Douglas Thain, Kevin Lannon

University of Notre Dame (US)

Although different computing resources are often available at universities, meeting all requirements to deploy a valid Tier 3 able to run CMS workflows on the grid is not an easy task to achieve without root-level access to them. For instance, providing the right operating system (OS) with access to the CERNVM File System (CVMFS) on the worker nodes or installing and maintaining a Compute Element (CE) on the submit host is not always allowed or possible due to lack of system administrator privileges, network policies, etc. This work describes the use of VC3 (a service for automating the deployment of virtual cluster infrastructures) and the OSG Hosted CE service in order to provide the grid environment and components needed to build a CMS Tier 3 using Notre Dame opportunistic campus resources without root privileges.

Notre Dame Computing Resources

The Notre Dame CMS group operates a Tier 3 with about 1,300 cores for local and grid analysis jobs. 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, but these resources lack the software components and environment needed by CMS analysis workflows.

Deploying these needs the way it is done at the Tier 3 is not an option due to the lack of root access level to the nodes. To address this, Notre Dame uses Lobster[1] (a workflow management tool designed to harness these non-dedicated resources) to provide these CMS requirements locally, but this does not deal with CMS grid jobs with other standard submission tools in CMS like CRAB3.

[1] http://lobster.crc.nd.edu
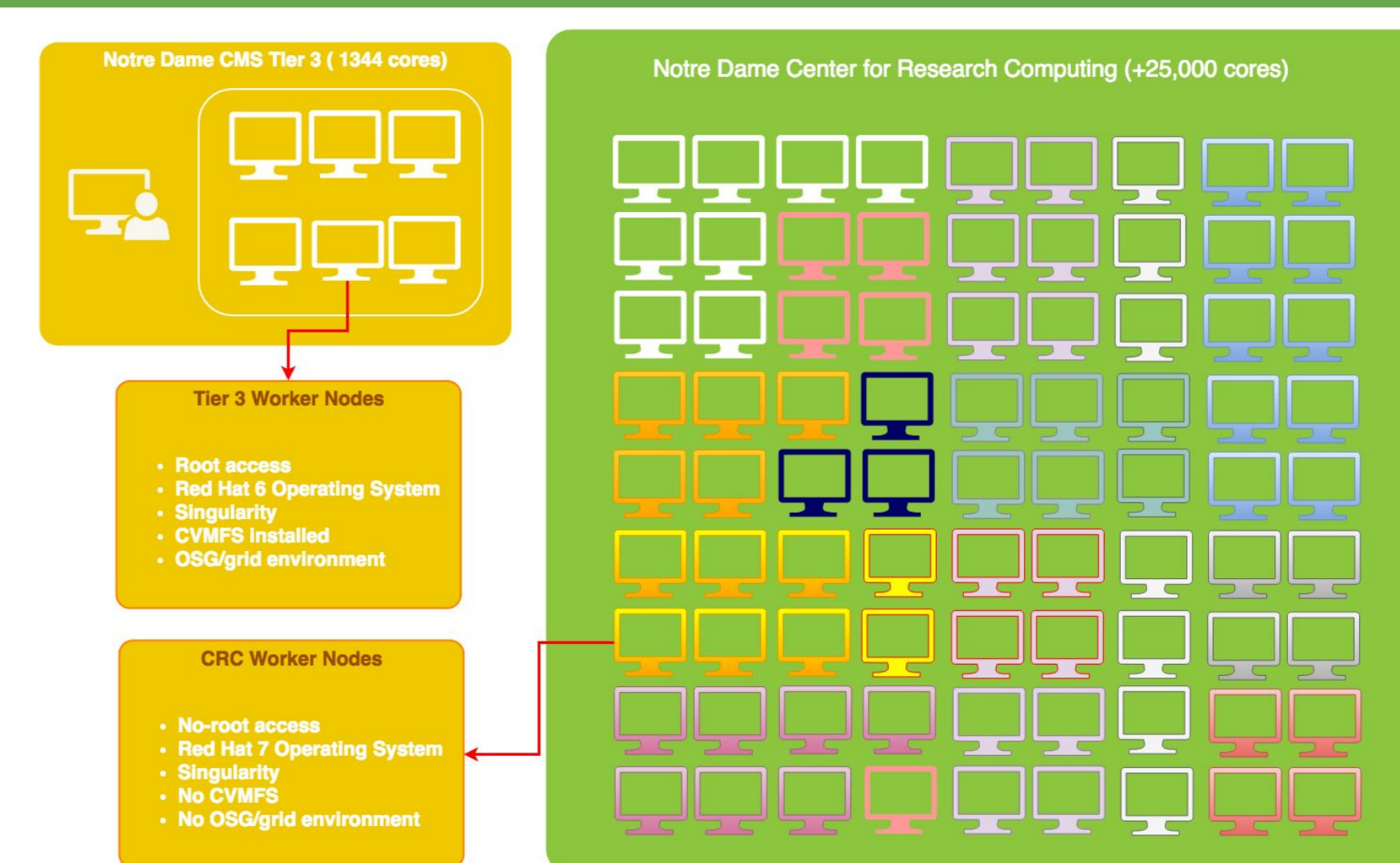

VC3: Virtual Clusters for Community Computation

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 DOE-funded VC3 project[2,3] 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.

In this work, we use VC3 to build a virtual cluster on top of the Notre Dame CRC campus cluster with all the software stack components and environment required by CMS analysis workflows to work through the grid without requiring root access level to these nodes.

[2] http://www.virtualclusters.org/
[3] VC3 is supported by the Department of Energy NGNS program via grant DE-SC0015711

VC3 Architecture
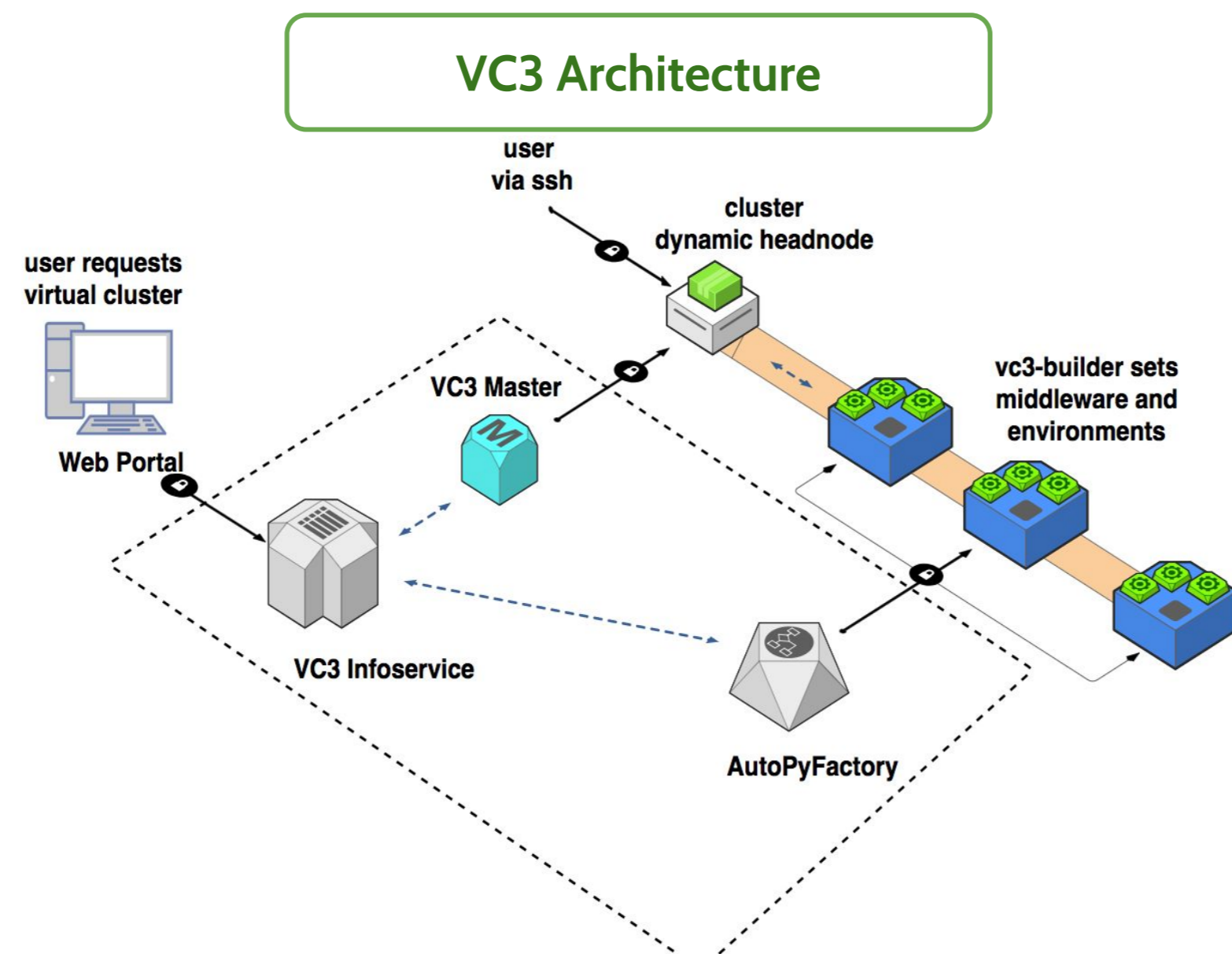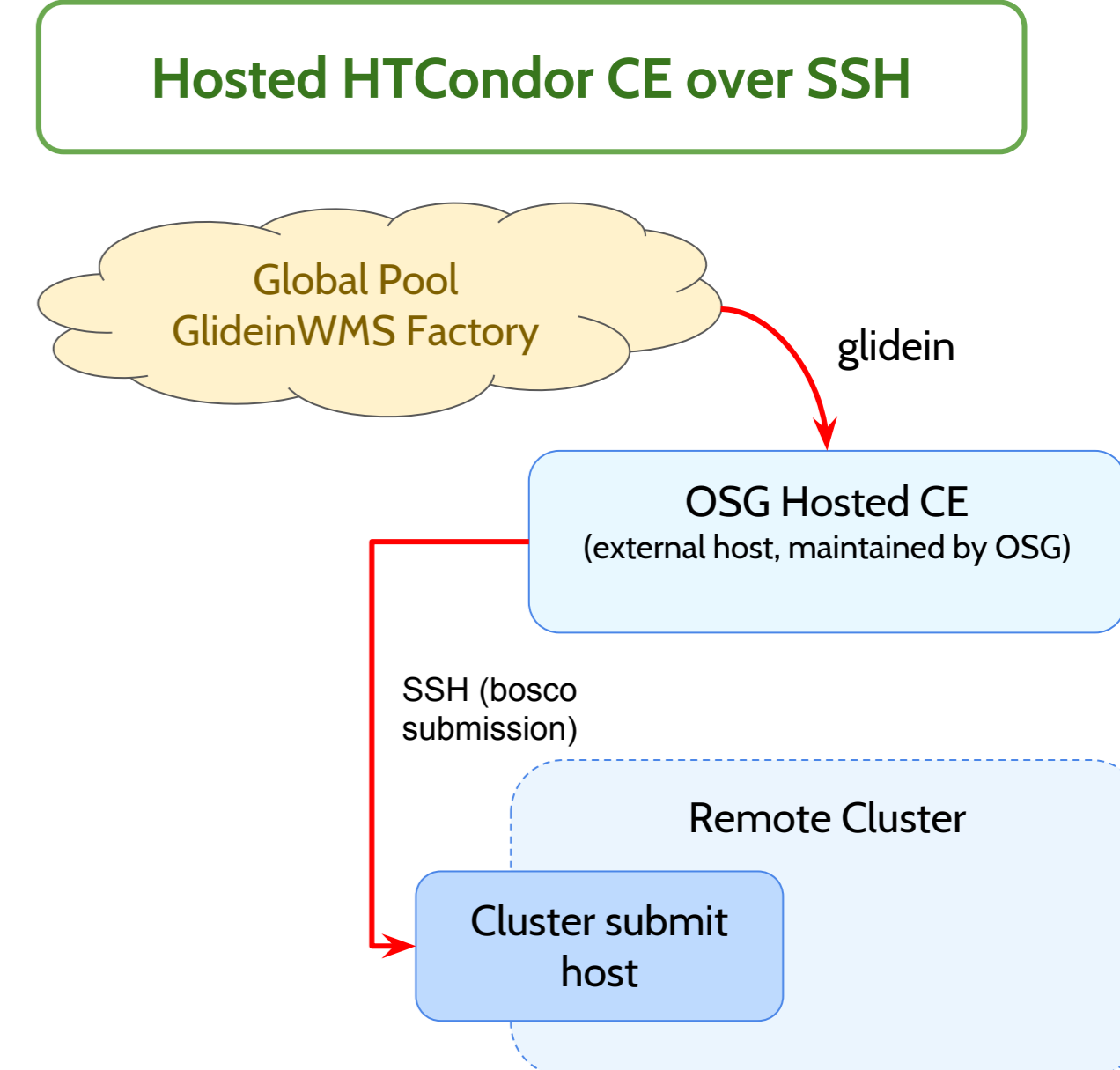
The OSG Hosted CE service

The Open Science Grid[4] Hosted CE service is an HTCondor-CE gateway working over SSH via Bosco[5].

This service provides and maintains an HTCondor-CE instance outside the remote cluster infrastructure.

Host certificates and the CE installation and configuration is handled by OSG and the only requirement to connect this to a remote cluster is having SSH access (through an authentication key) to a normal account in the cluster submit host.

[4] http://opensciencegrid.org/docs
[5] https://osg-bosco.github.io/docs/

Hosted HTCondor CE over SSH

Building a Tier 3 on top of the CRC resources with VC3 and the OSG Hosted CE service

A Grid Compute Element able to run CMS analysis workflows requires the following on the worker nodes:
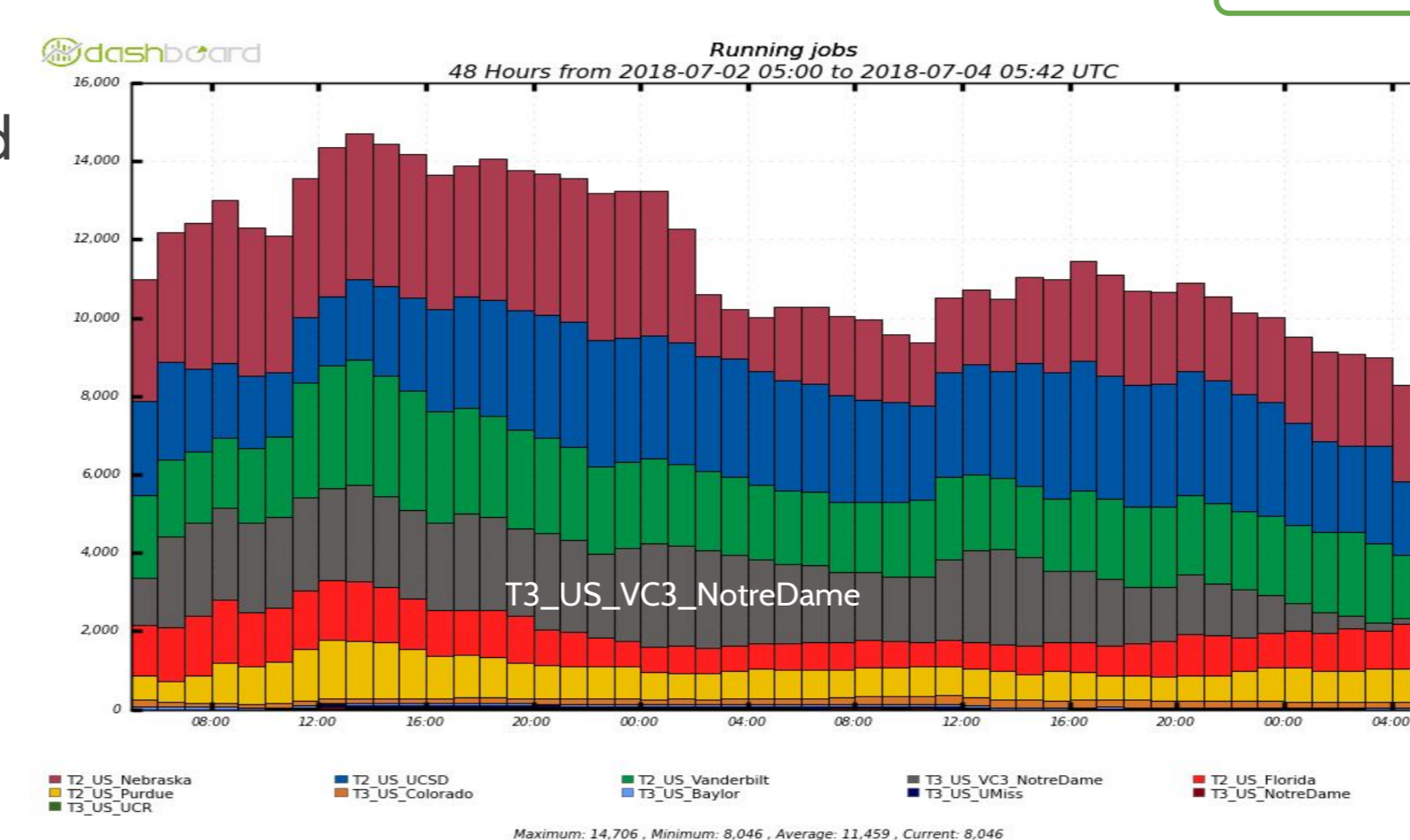
- The CernVM File System (CVMFS)
- Virtual Organization Membership Service components for authentication
- Grid transfer tools (gfal, xrootd)
- CMS information for the Site (PhEDEx, squid location, Trivial File Catalog) properly defined and linked to /cvmfs/cms.cern.ch/SITECONF/local
- Red Hat 6 or 7 with (optionally) singularity.
  - While the CMS framework works with RHEL 7, physics results validation is still ongoing and RHEL 6 is still the main Operating System used in CMS analysis workflows.

While most of these software components need root level privileges for installation and configuration, this can also be achieved with an extra thin layer deployed with VC3 to provision everything in user space.
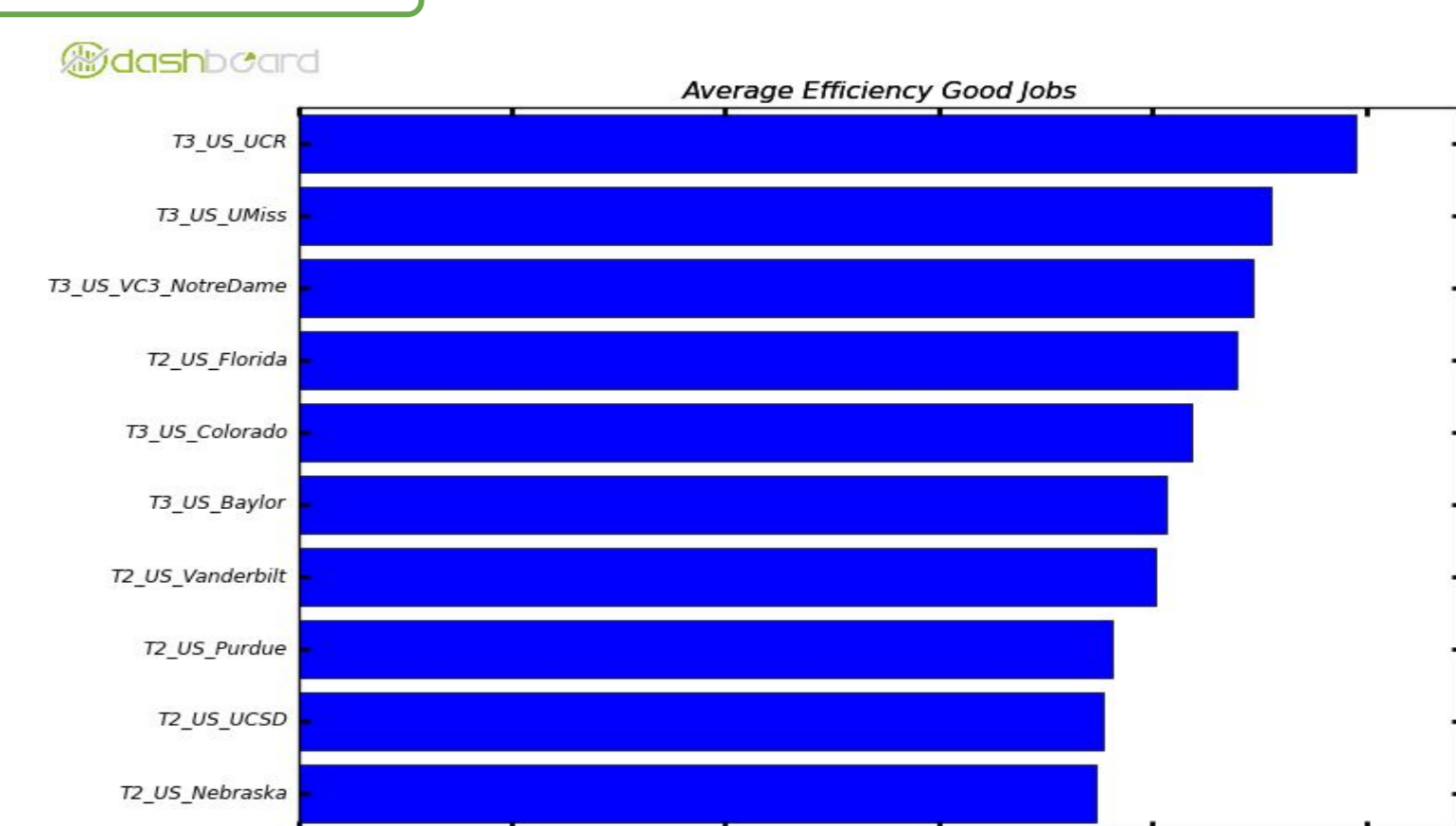
Additionally, a job gateway (HTCondor–CE) with the proper host certificates for authentication installed and maintained on a machine that can submit jobs into the local batch system is also needed in order to add the resource to the CMS Global Pool. To overcome this at the user-level, the OSG Hosted CE service is used instead.

Traditional Grid Compute Element scheme

Grid Compute Element provisioned with VC3 + OSG Hosted CE
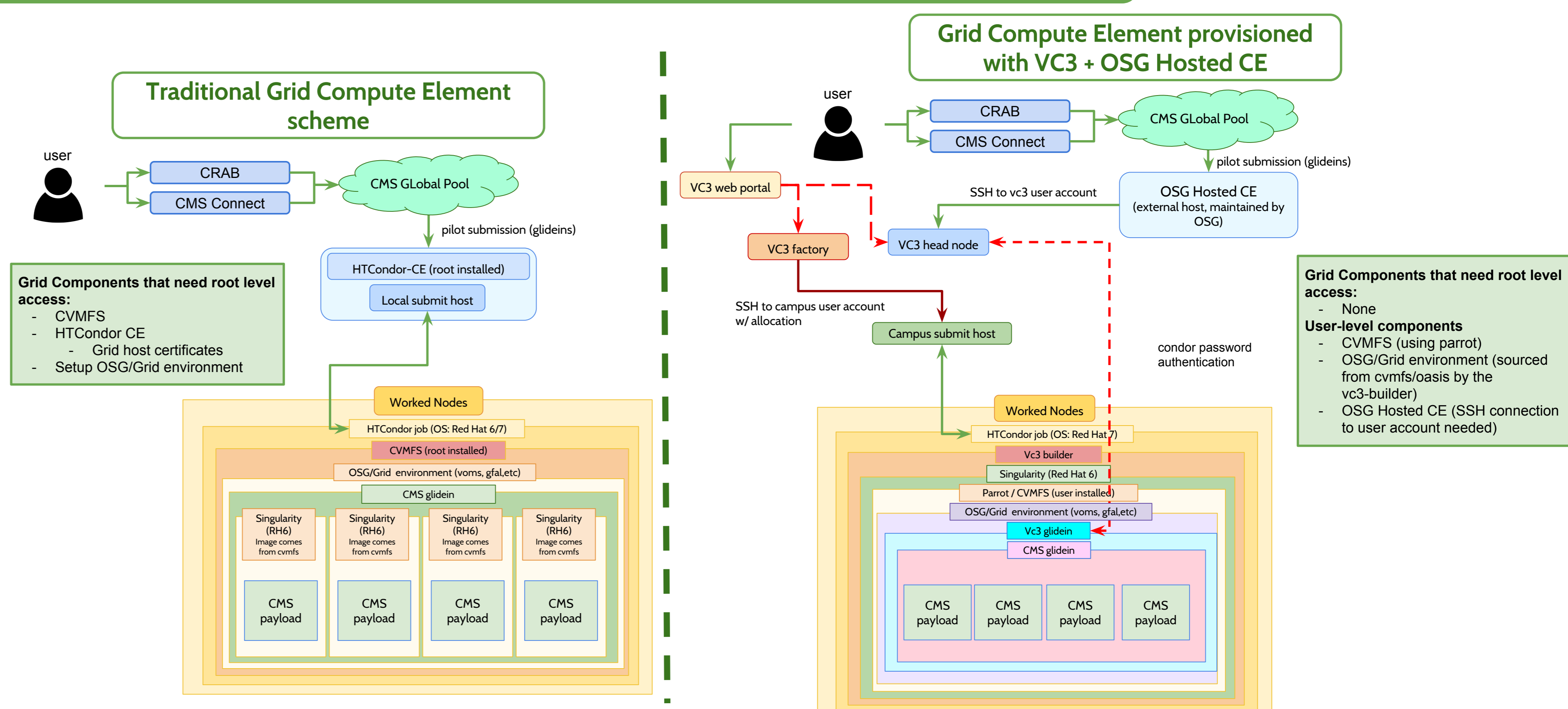
Running CRAB jobs

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 CRAB3 (or CMS Connect) to match to this resource and submit to it.

The plots in this section 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.

CMS Dashboard