UNIVERSITY OF NOTRE DAME

connect

USCMS Tier-3 meeting

Kenyi Hurtado, Kevin Lannon, James Letts

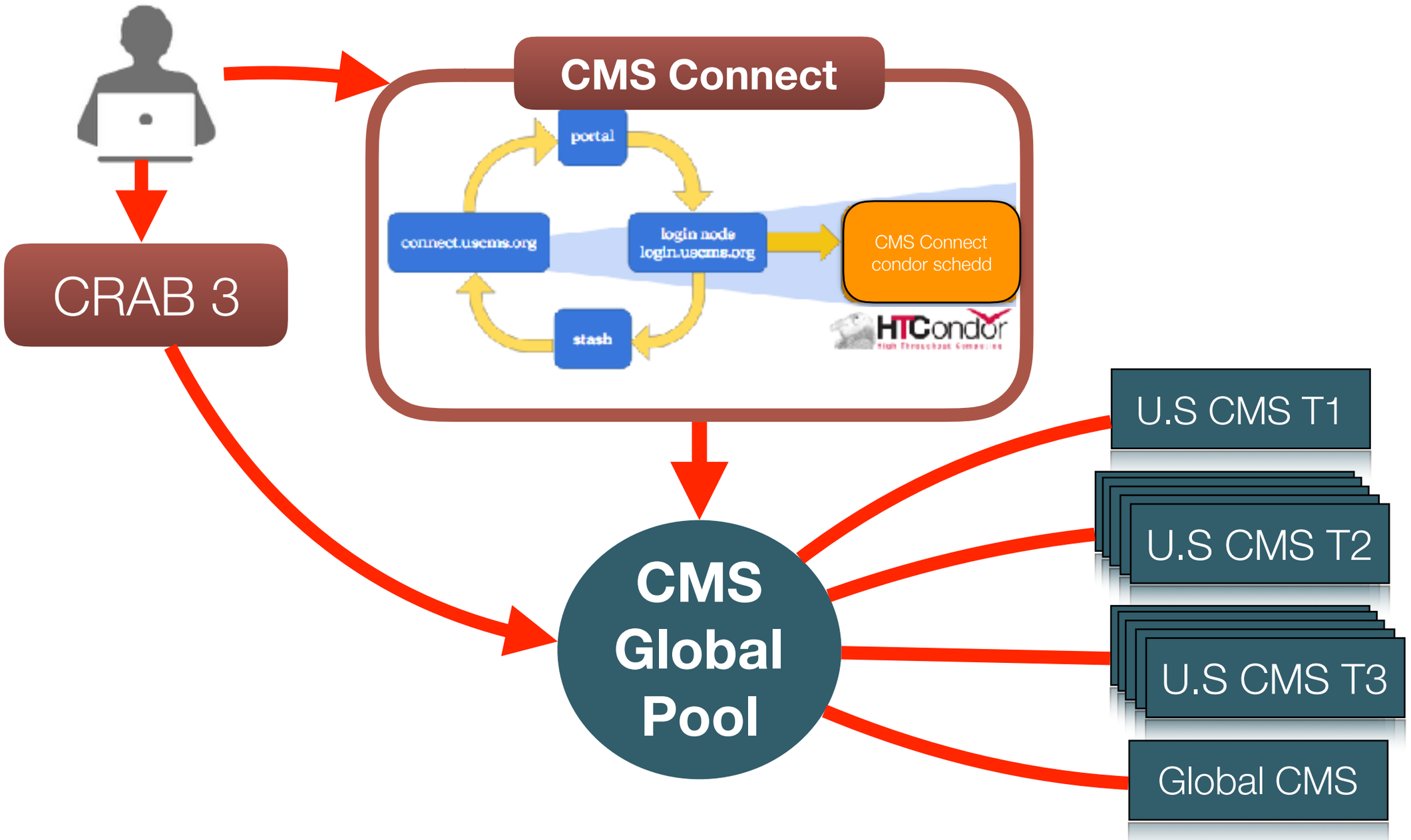
May 12, 2017

Covered in this talk

- **What is CMS Connect?**
  - CMS Connect and CRAB (differences and similarities)
- **Service Platform details**
- **How to use the service in 5 steps**

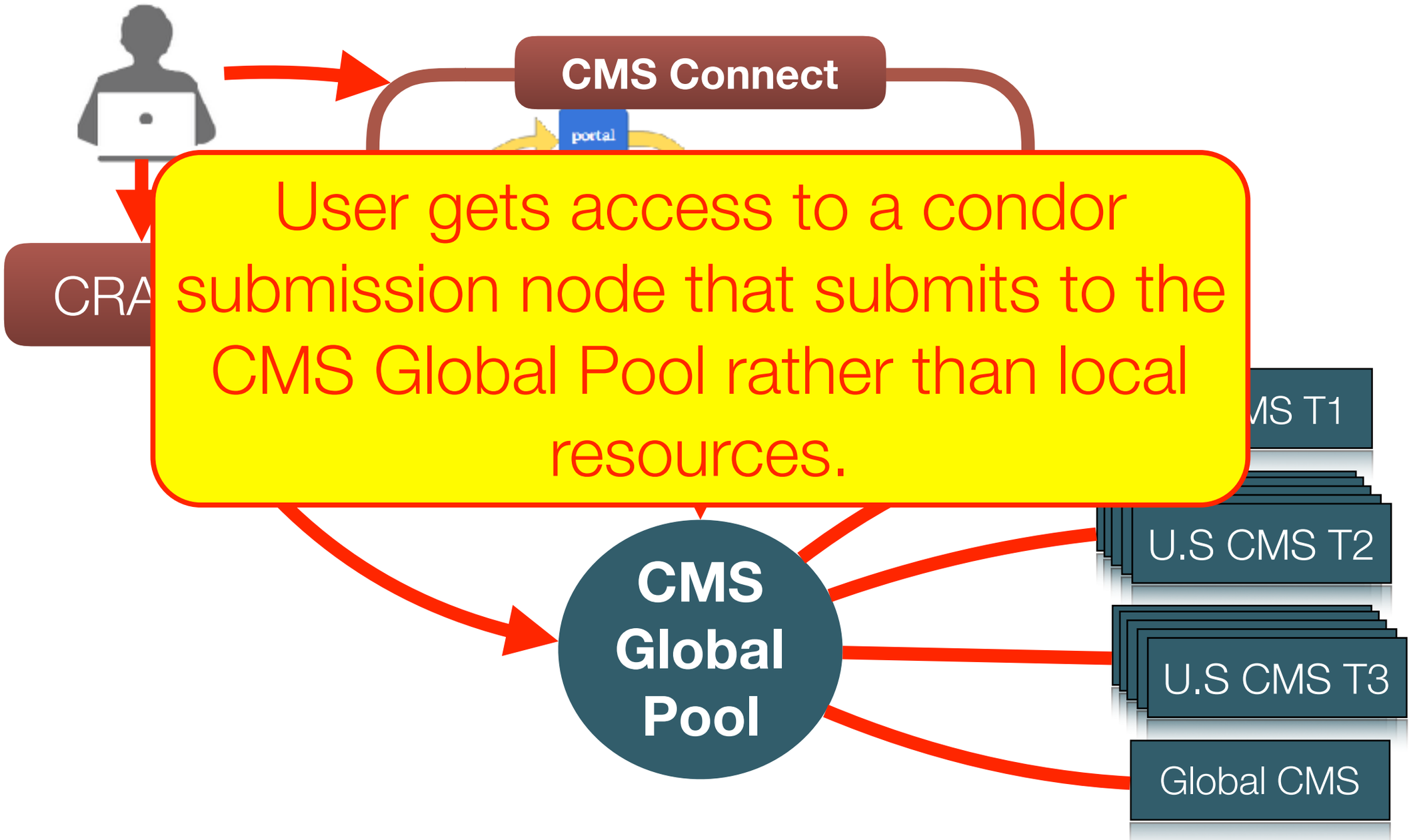
What is CMS Connect?

A **service** to provide Tier 3-Like capability to Institutions **for non-CRAB** workflow **jobs**.



What is CMS Connect?

A **service** to provide Tier 3-Like capability to Institutions **for non-CRAB** workflow **jobs**.



What are non-CRAB workflow jobs exactly?

To answer this, let's start dividing jobs in two categories:

- **Production**: *Organized processing of data performed by a central **operations team***
- **(Offline) Data Processing jobs**
  - **Analysis**: *Data analysis performed by **individual users***

What are non-CRAB workflow jobs exactly?

Analysis jobs could also be divided in two different workflow sub-categories

Workflows using **cmsRun**: the CMSSW executable for e.g *event processing*

Analysis

*Data analysis performed by **individual users***

Late-stage analysis workflows: Regular condor-like jobs for e.g making histograms, plots, analyzing trees, etc.

What are non-CRAB workflow jobs exactly?

Analysis jobs could a
sub-categories

CRAB 3 handles this (CMS Remote Analysis Builder)

Workflows using **cmsRun**: the CMSSW executable for e.g *event processing*

Analysis

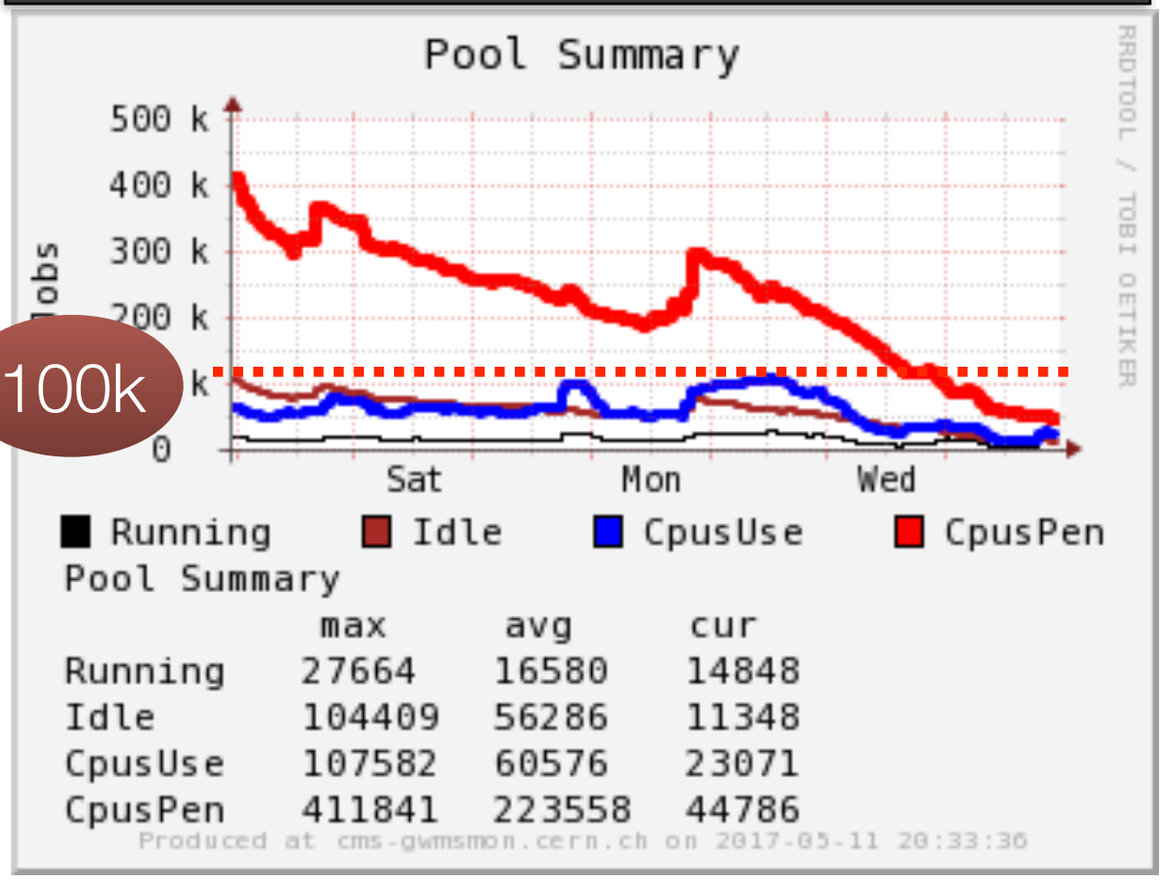
*Data analysis performed by **individual users***

CMS-Connect focuses on this category

Late-stage analysis workflows: Regular condor-like jobs for e.g making histograms, plots, analyzing trees, etc.

CMS Global Pool Scale

Production Summary



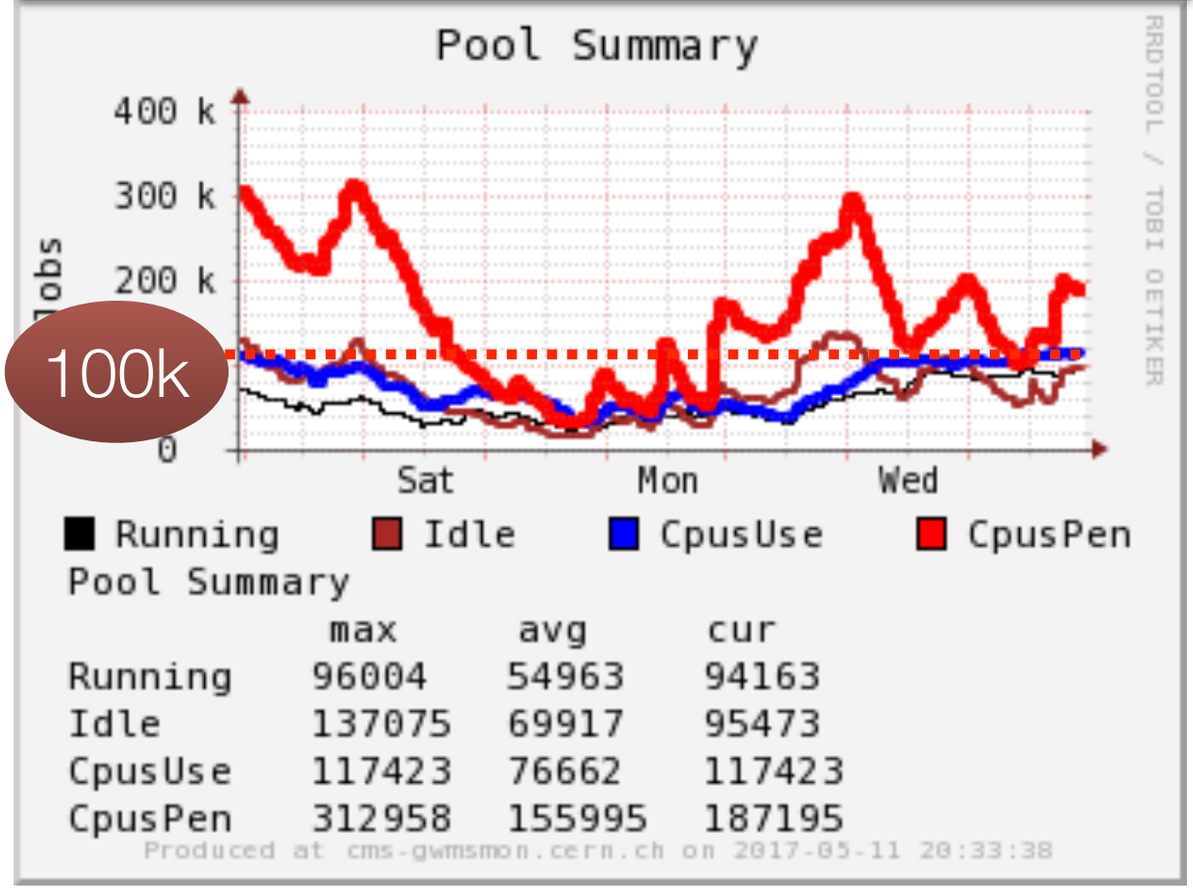
100k

Pool Summary

	max	avg	cur
Running	27664	16580	14848
Idle	104409	56286	11348
CpusUse	107582	60576	23071
CpusPen	411841	223558	44786

Produced at cms-gwmsmon.cern.ch on 2017-05-11 20:33:36

Analysis Summary



100k

Pool Summary

	max	avg	cur
Running	96004	54963	94163
Idle	137075	69917	95473
CpusUse	117423	76662	117423
CpusPen	312958	155995	187195

Produced at cms-gwmsmon.cern.ch on 2017-05-11 20:33:38

Over 200K cores globally distributed available for these two categories (reached ~250K recently).

CMS Connect and CRAB

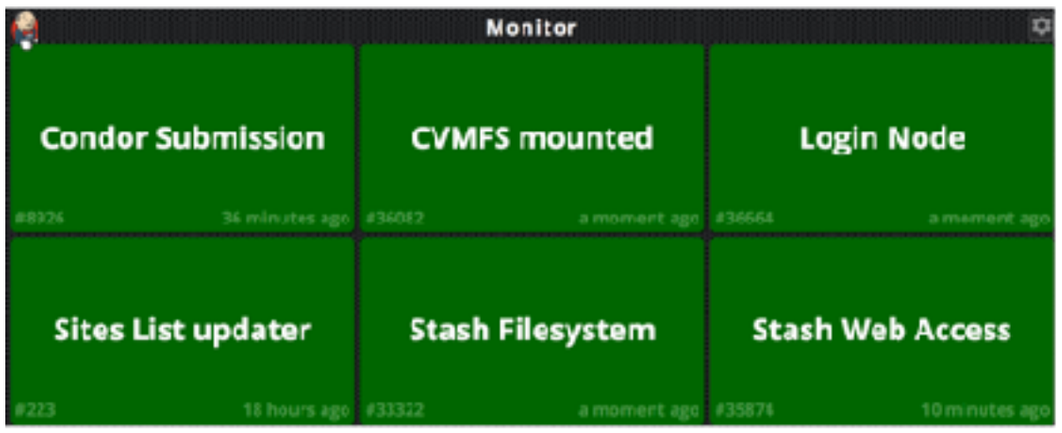
- **Differences**
  - Different workflows: CMS Connect is a complementary service to CRAB for a different type of workflows.
  - CRAB is focused on cmsRun jobs, while CMS Connect focuses on condor-like jobs.
  - CRAB can handle file transfers, site resources, framework setup, etc… in CMS Connect case, these depend on the user, as in any regular HTCondor job (that's why we recommend to use CRAB if what you need to run is a cmsRun job).

CMS Connect and CRAB

- **Similarities**
  - Both:
    - Require your VO CMS user proxy certificate.
    - Can access the CMS Global Pool as analysis jobs.
    - Report to the CMS Dashboard.

Platform overview

- CMS Connect uses the CI-Connect platform
  - OSG supported platform used by OSG Connect (and other experiments like ATLAS).
- Servers are hosted by OSG & U. Chicago and are connected on a DMZ network.
- Service components monitored via jenkins



Quotas

Per user:

- Home area: 100 Gb
  - Can be checked via “quota" command on the submission node. Meant for long-term files storage (source code, scripts, etc).
- Stash storage area: No quota
  - Shared with OSG Connect, about 3.2 PB total, currently ~1.6 PB free. Meant for short-medium term files (should be transferred to CERN/FNAL EOS or US T2 user storage allocations eventually).
- Maximum of 100K submitted jobs per user, 20K jobs per submission.

Global details:

- Currently, maximum of 10K running jobs in the system.
- Fair-share handled by the Global Pool, same analysis accounting group than CRAB

References

- CMS Connect links
  - Website: http://connect.uscms.org
  - Documentation: http://docs.uscms.org
- CHEP 2016
  - Poster: https://goo.gl/Nvavaz
  - Paper: http://stash.osgconnect.net/~khurtado/cms-connect-paper-v4.pdf

Extra Slides
(User-oriented mini-tutorial)

How to use CMS Connect (in 5 steps)

The following slides show the main steps to get users started on the service.

A quick start guide can be found here:
http://docs.uscms.org/CMS+Connect+Quickstart

An extended tutorial can be found below:
https://indico.cern.ch/event/533066/

1. Signing up to the service

http://connect.uscms.org/signup

CMS Connect uses Globus Nexus for user authentication and accounting.

Users need to sign up on the link above in order to get and account into the submission node.

Using your same CERN username for the globus username is not necessary, but strongly recommended.

Sign-up will ask you to create a Globus ID

2. After approval, users can login into the submission node

$ ssh <username>@login.uscms.org

… the approval usually takes up to a business day.

Note: Users will soon need to generate SSH-keys in order to login.

Link for instructions

3. Transferring CMS grid certificates

After logging in login.uscms.org:

$ copy_certificates

We are using the CMS Global Pool, so we need to provide a valid user CMS VO grid proxy, just like with CRAB or WMCore.

More details here: http://docs.uscms.org/CMS+Connect+Quickstart

4. Start submitting via condor

```
#[example.jdl]
Universe = vanilla
Executable = short.sh
Arguments = 5 # to sleep 5
seconds
Error = log/job.err.$
(Cluster)-$(Process)
Output = log/job.out.$
(Cluster)-$(Process)
Log = log/job.log.$(Cluster)
Queue 10
```

```
$ condor_submit example.jdl
Submitting
job(s).....................................
.....................................
.................
10 job(s) submitted to cluster 938.
```

More details here:
https://docs.uscms.org/Quick+Condor+Tutorial

5. (Optional) Selecting the CMS Sites:

Add the list of Sites in your condor submit files. E.g:

+DESIRED_Sites="T2_US_UCSD,T3_US_NotreDame"

Or export the list using the environment variable:
CONDOR_DEFAULT_DESIRED_Sites
Additional commands:

```
$ get_condor_sites
$ set_condor_sites
```

If not specified, all CMS resources are used by default.

More details here: http://docs.uscms.org/Selecting+Sites

Acknowledgements

- U. Chicago OSG Team:
  - Rob Gardner
  - Lincoln Bryant
  - Suchandra Thapa
  - Balamurugan Desinghu
  - David Champion

connect

CILogon

Open Science Grid

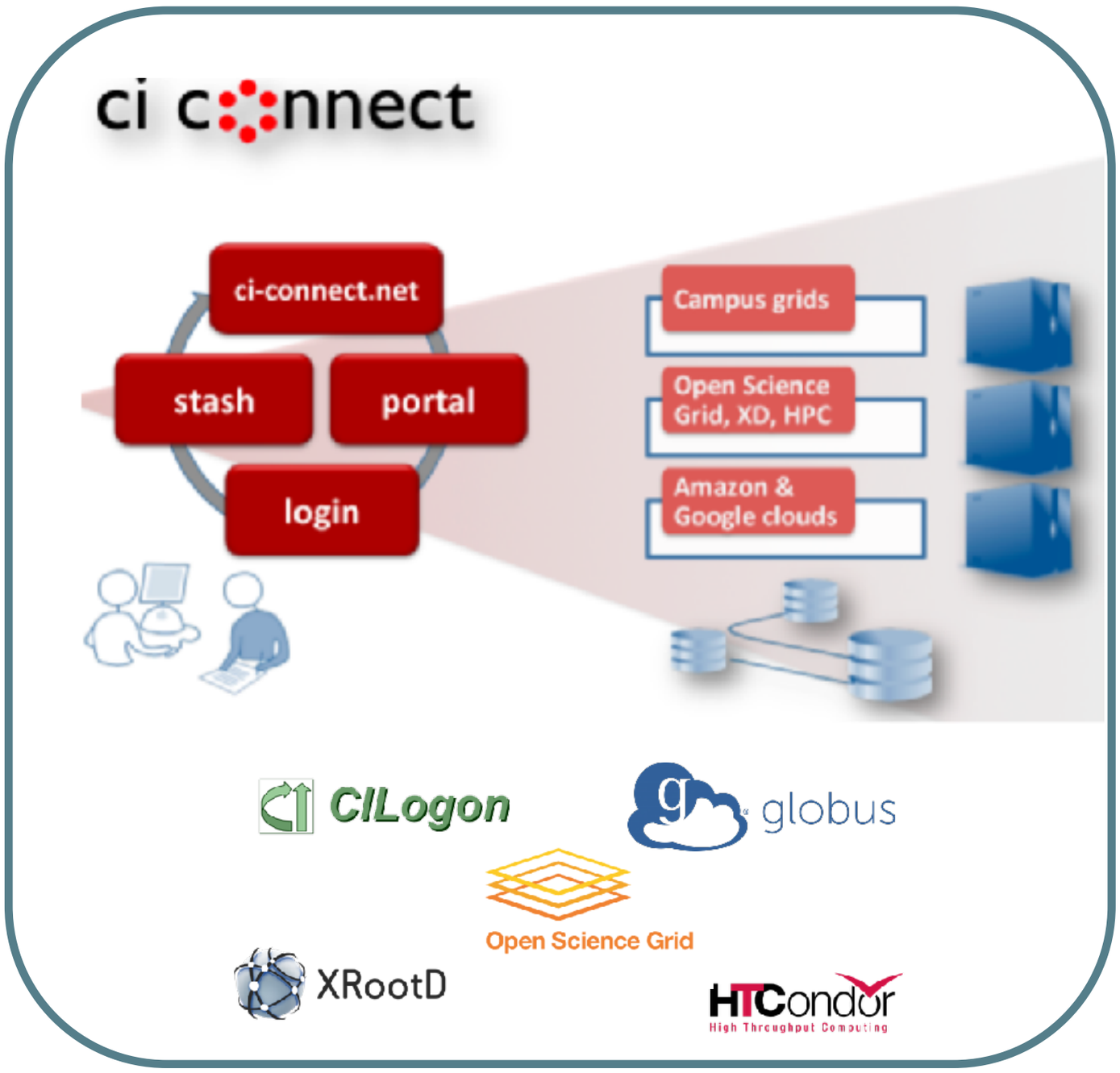
XRootD

HTCondor
High Throughput Computing

Backup Slides

Technology behind CMS-Connect

- Based on CI Connect Platform

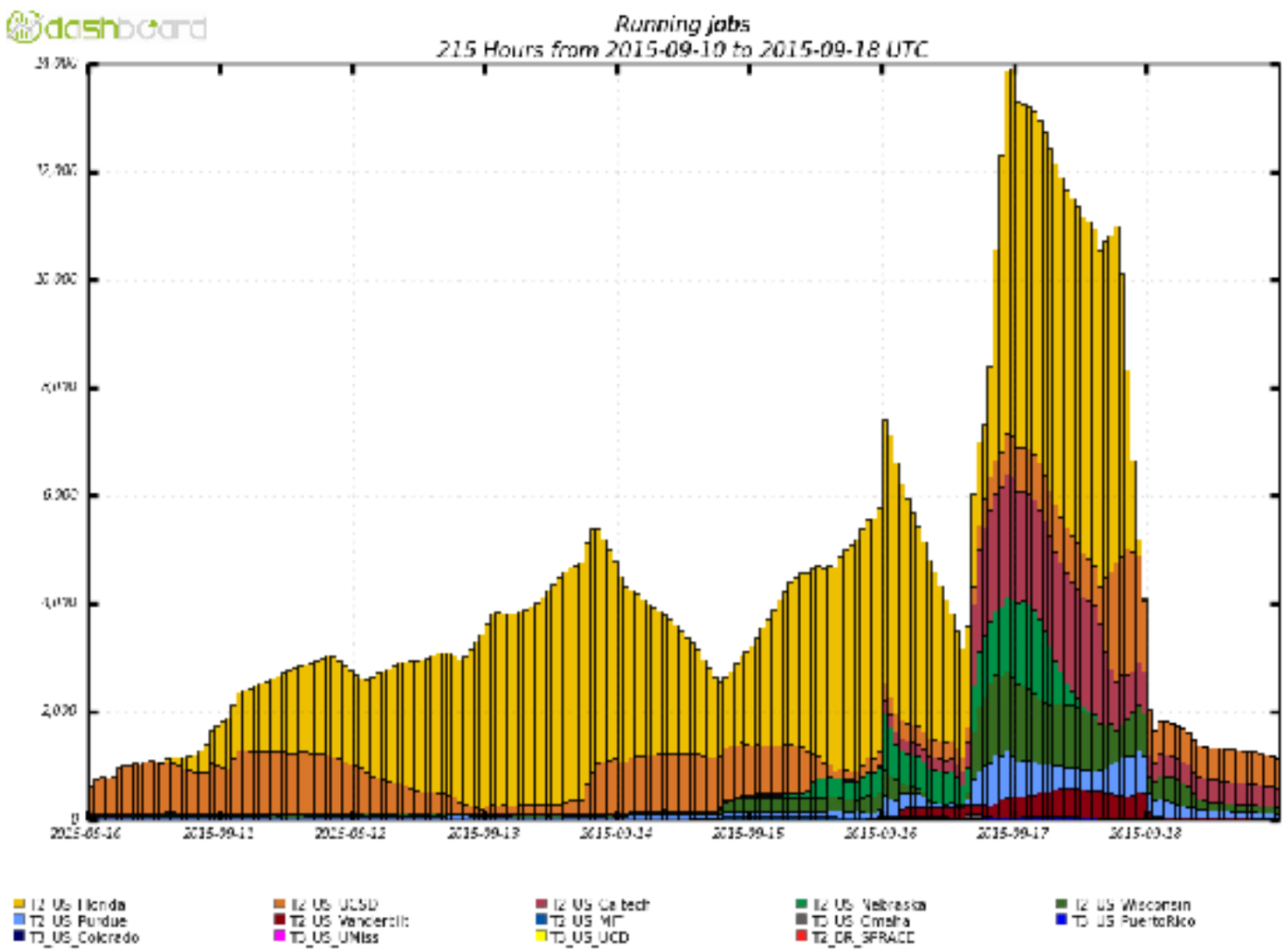


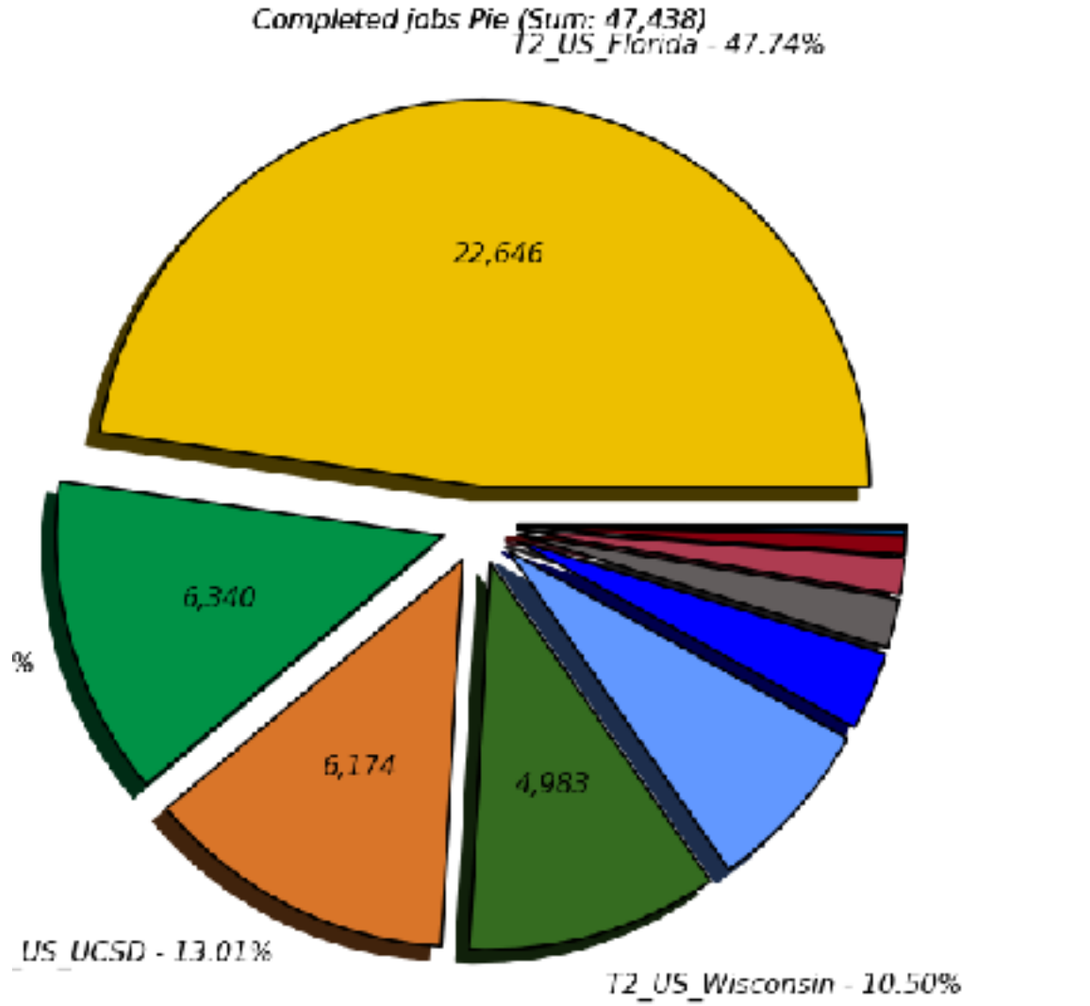
- Globus Platform
  [CILogon + InCommon + X509]
  - Identity Management.
  - Groups, Projects.
- Login Host
  - Auto provisioning of user accounts.
- Connecting CPU resources
  - HTCondor.
- Distributed Data Access
  - XRootD, Globus access, http.
- Distributed Software
  - cvmfs

https://goo.gl/J7VQtJ

CMS Dashboard Reporting

T2 US Sites historical view example





Integrated CMS Dashboard monitoring with condor job submission.