

Using ssh as portal

The CMS CRAB over glideinWMS experience

by

I Sfiligoi¹, S Belforte²,

J Letts¹, T Martin¹, M D Saiz Santos¹ and F Fanzago³

¹University of California San Diego

²Università e INFN di Trieste

³Università e INFN di Padova

Agenda

- How we got there?
- What we did?
- How it worked out?

- CMS is one of the 4 experiments at the Large Hadron Collider
- CMS Analysis Operations (AnaOps) is a computing task focused on the operational aspects of **enabling physics data analysis**
 - i.e. provides the computing infrastructure used by the actual physicists



CRAB

CMS AnaOps submission tool

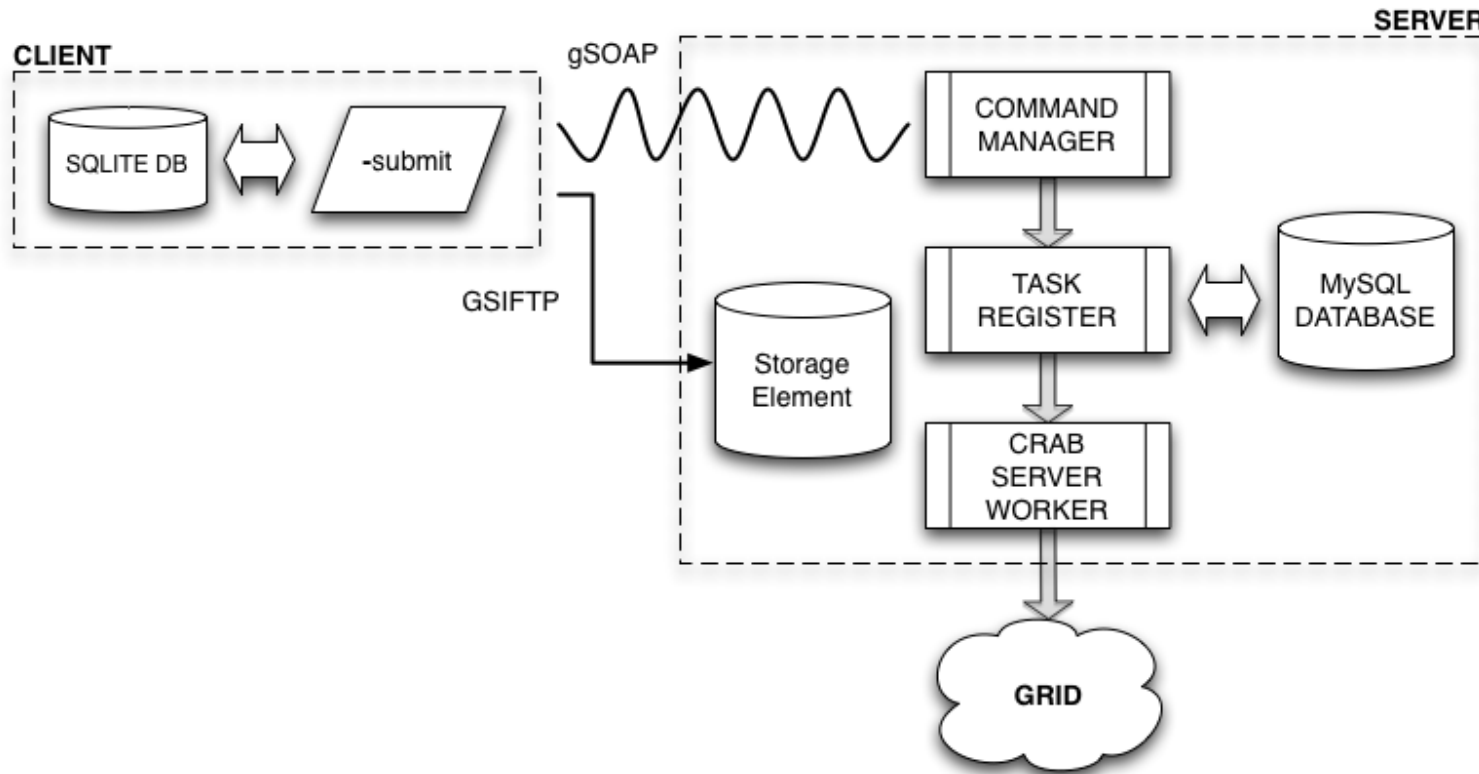


- CMS AnaOps has been using CRAB as **the user interface to the Grid** since 2004
- Initially was a **purely client interface**
 - CRAB basically just a fancy wrapper
 - So the Client machine had to have the **full Grid stack to function**



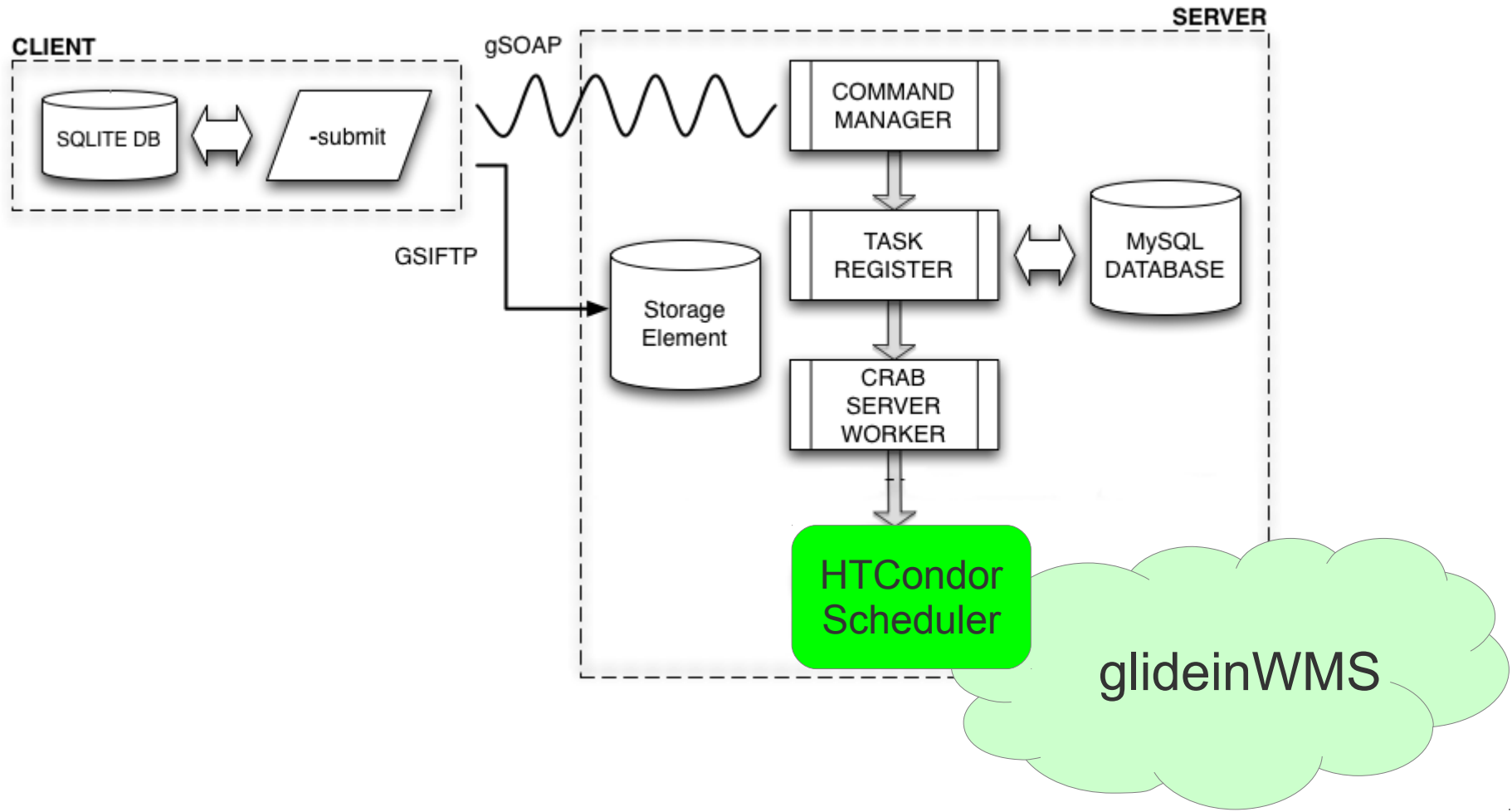
CRAB₂ Server

- CRAB₂ Server introduced in 2007
- Client does not need the Grid stack anymore



- In 2009 CMS added glideinWMS to the list of supported middleware
- glideinWMS, based on HTCondor, **does not support remote job submission**
 - CRAB₂ Server had to sit on the same node as (one of) the HTCondor scheduler(s)

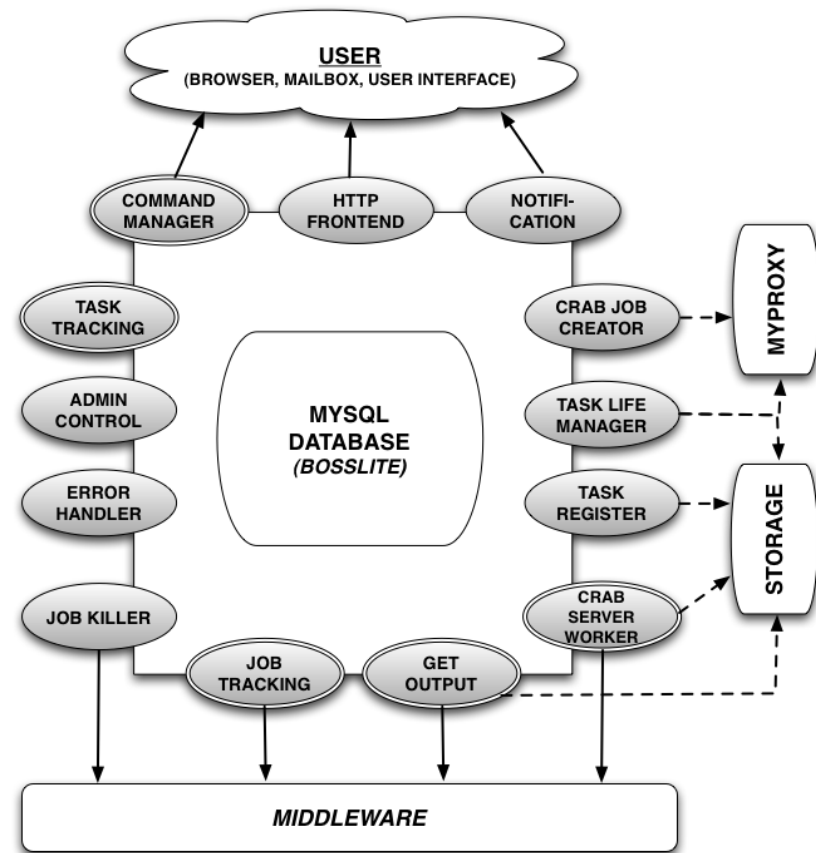
A schematic view



Operational experience

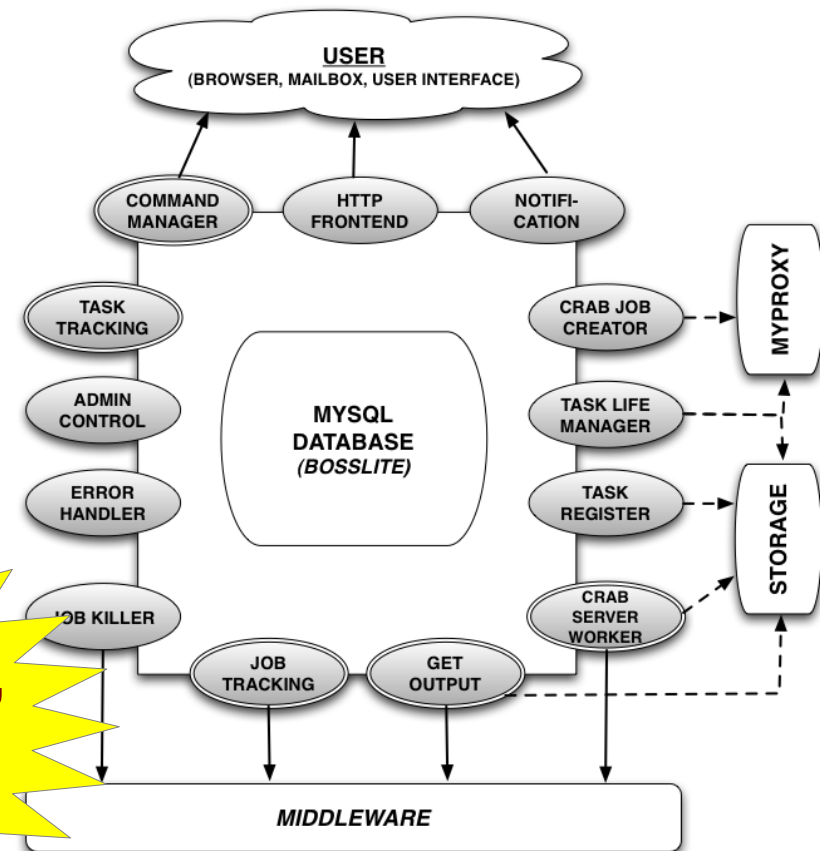
- CMS AnaOps experienced **a lot of operational issues** with CRAB₂ Server submitting through glideinWMS
 - One major problem being the server losing track of job status
- **Users were complaining a lot**

- Designed to be very modular and flexible
 - But that made it complex, too
- CMS decided to **give up on fixing it**
 - CRAB3 would simply replace it



- Designed to be very modular and flexible
 - But that made it complex, too
- CMS decided to **give up on fixing it**
 - CRAB3 would simply replace it

But by late 2012, CRAB3 still did not exist



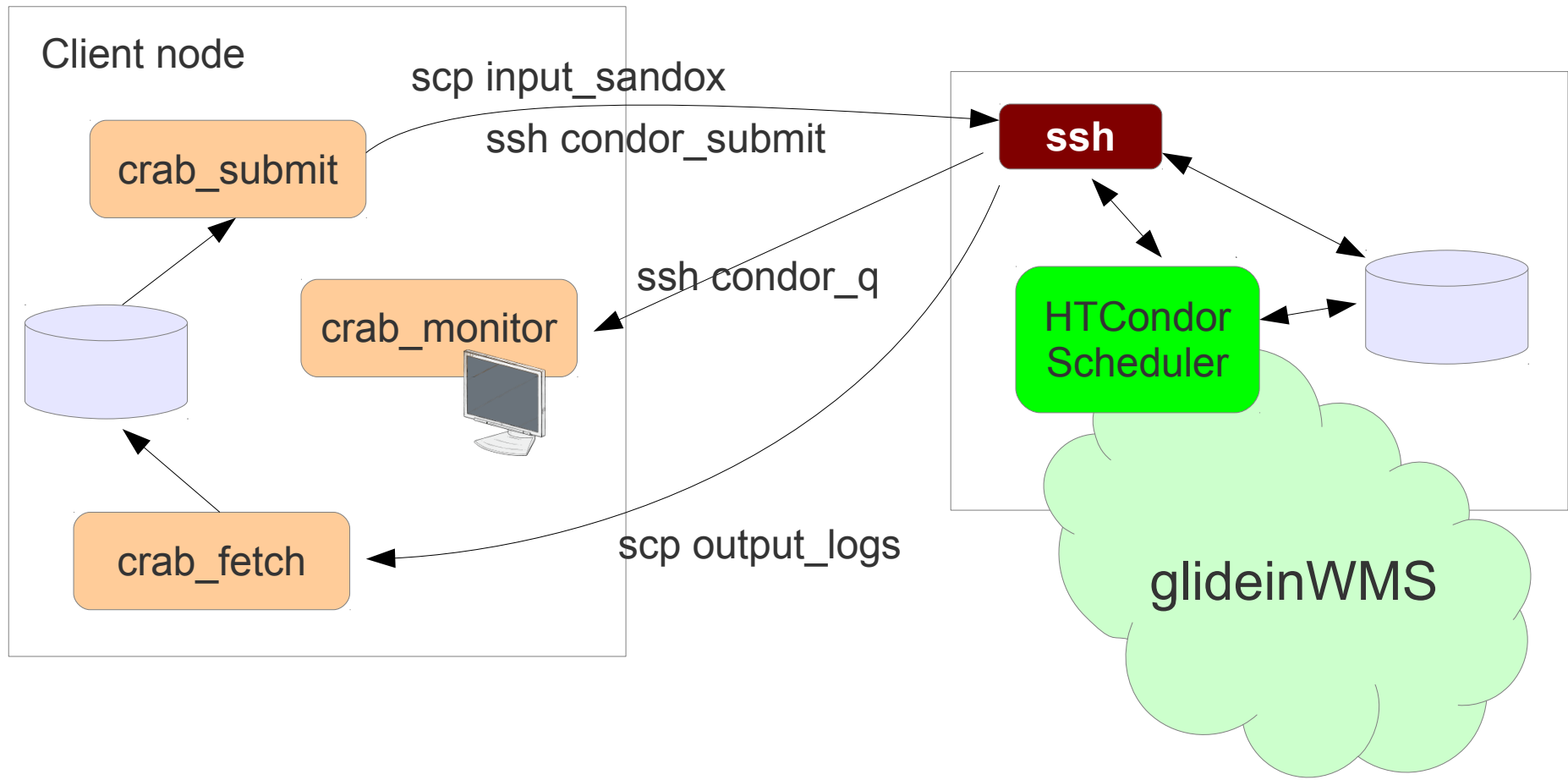
- **The CRAB₂ Client never went away**
 - A significant fraction of CMS users kept using it after CRAB₂ Server was put in production
 - And they were, by and large, liking it
- **It just could not be used to submit to glideinWMS**
 - Due to lack of (native) remote submission to HTCondor

Why not just use ssh?

- The major stated benefit of CRAB₂ Server was “remote submission using a standard interface”
(there were of course other stated benefits, but they had very little impact when paired with glideinWMS)
- But what users normally use for remote access?
 - **SSH**
- So, what if we used ssh in CRAB₂ client as well?
 - The gsissh dialect to support x509 proxies

- There are three stages in CRAB2 Client
 - Job submission
 - Includes uploading the input sandbox
 - Job monitoring
 - Output log fetching
- The Client uses **ssh and/or scp** to talk to a node running the HTCondor scheduler

A schematic view

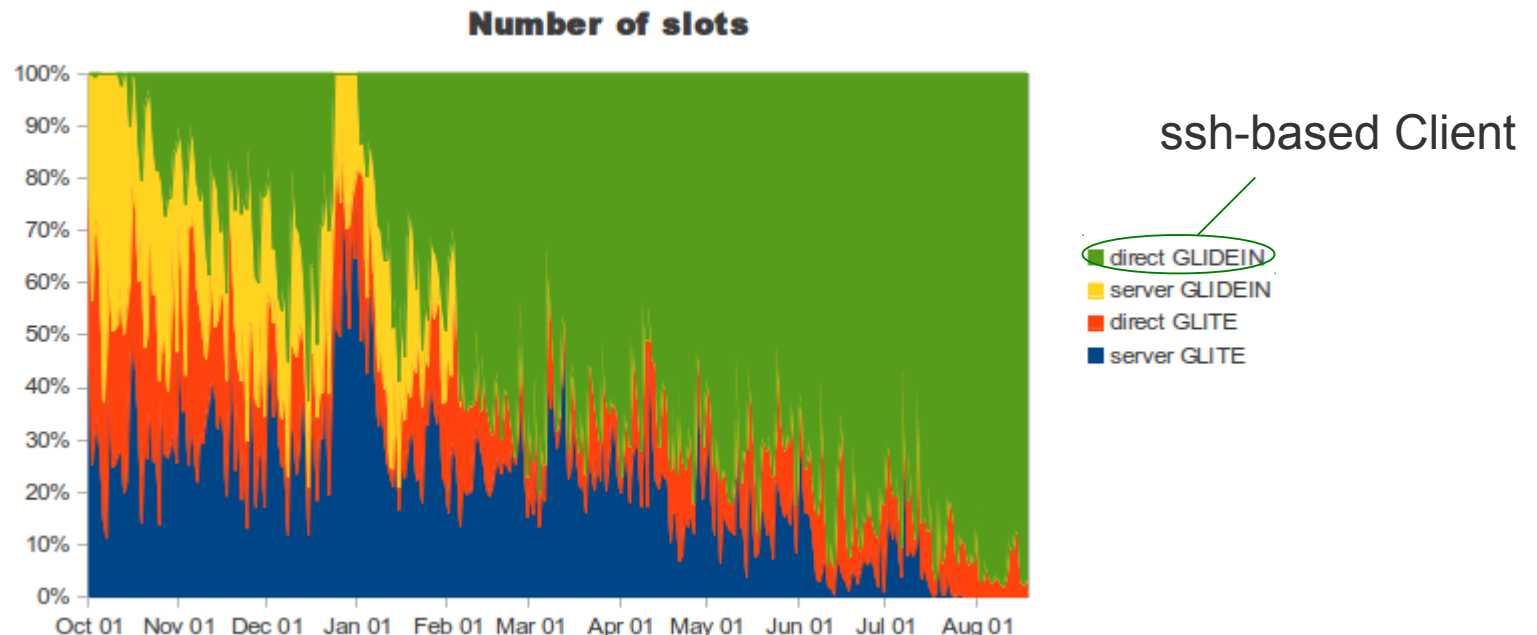


Server side setup

- gsisshd is a standard package
 - Just installed it from RPM
- Using lcms for authorization
 - With callbacks to GUMS in the USA, and Argus in Europe
- Each user gets a standard Linux account
 - No technical limits on what the user can run, but we would kill any offenders, if spotted

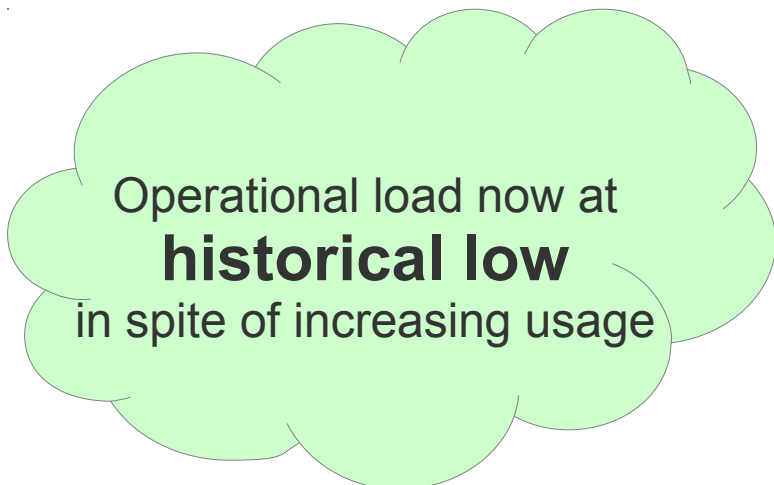
Users liked it!

- Took only 4 months for all users to completely switch from Server to ssh-Client for glideinWMS
- And another 4 months to get 90% of all work



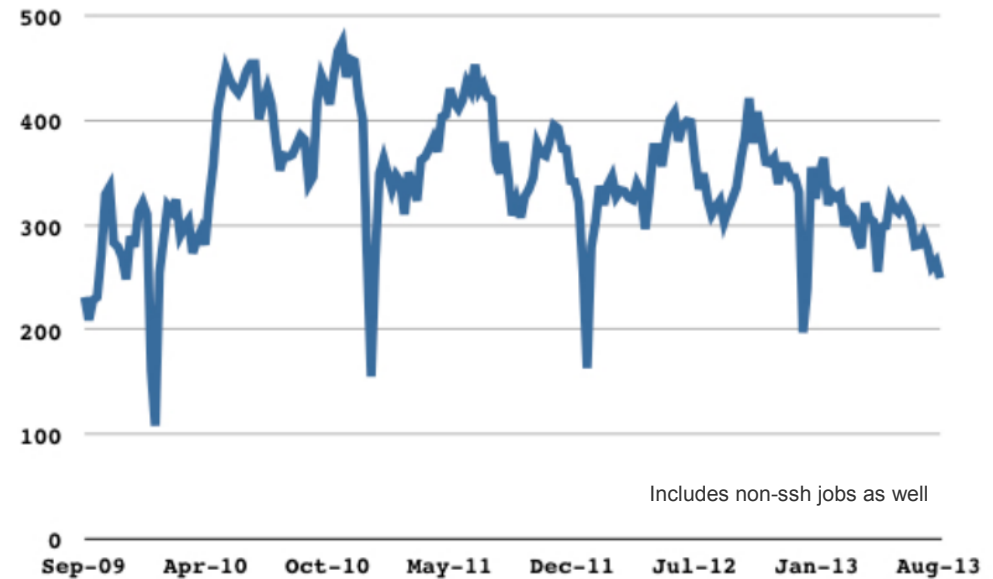
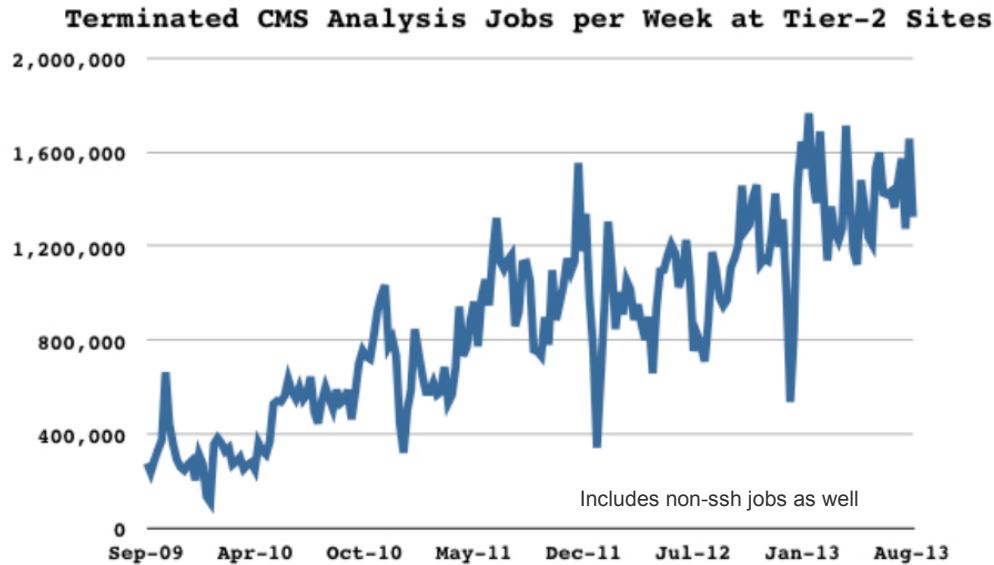
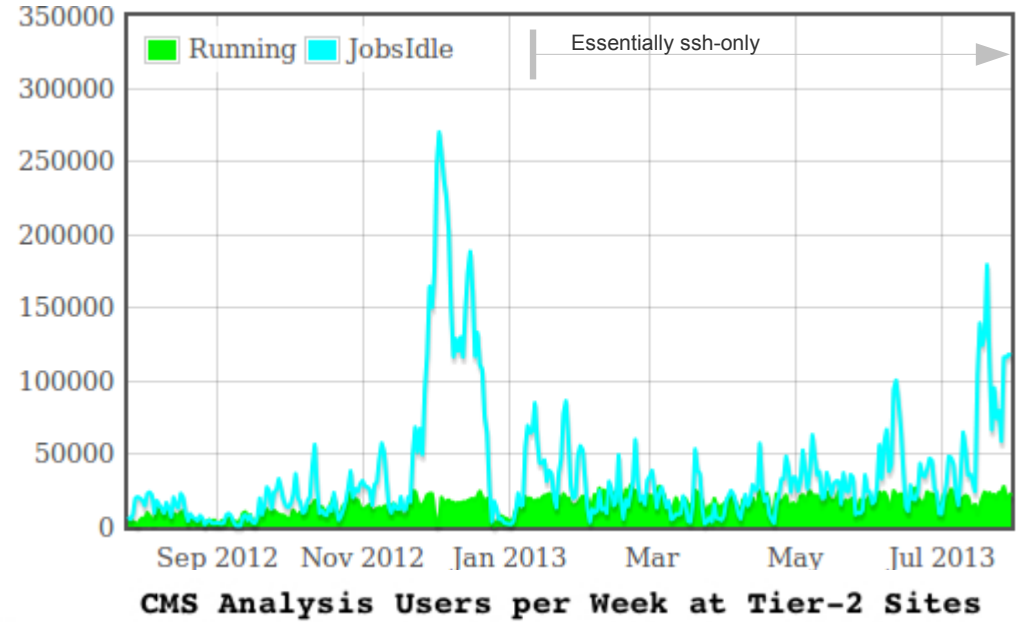
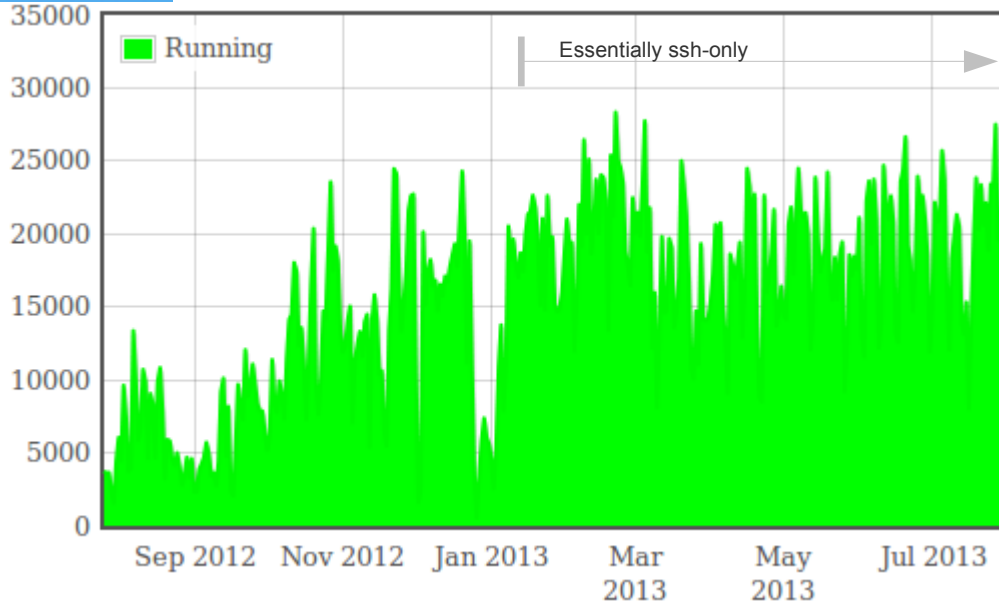
Smooth operations

- We have **not seen any major problems** with ssh-based CRAB₂ client itself
 - Now almost all problems are in the Grid layer
- Maybe not too surprising
 - CRAB₂ Client is really just a fancy wrapper
 - SSH is a very mature tool



Operational load now at
historical low
in spite of increasing usage

Works at scale



Hickups during the voyage

- We occasionally hit a ssh quirk
 - In order to speed up ssh connections, we re-use the ssh control session (-S)
 - If it gets stuck, user needed to manually remove it
- CRAB₂ Client now automatically detects and fixes the problem

Remaining issue

- The major remaining issue is the distribution of the CRAB₂ Client code
 - This still must be installed on the client node
 - Any change must be pushed to ALL the users
 - Making server-side changes challenging

Conclusions

- Using (gsi)ssh as a portal has proven to be very effective
- By removing complexity out of the code we have to maintain, life got much easier

Acknowledgments

- The idea of using ssh came from the RCondor package (see poster #322)
- This work was partially sponsored by the US National Science Foundation under Grants No. PHY-1148698 and PHY-1120138.