

CMS@home: Enabling Volunteer Computing Usage for CMS

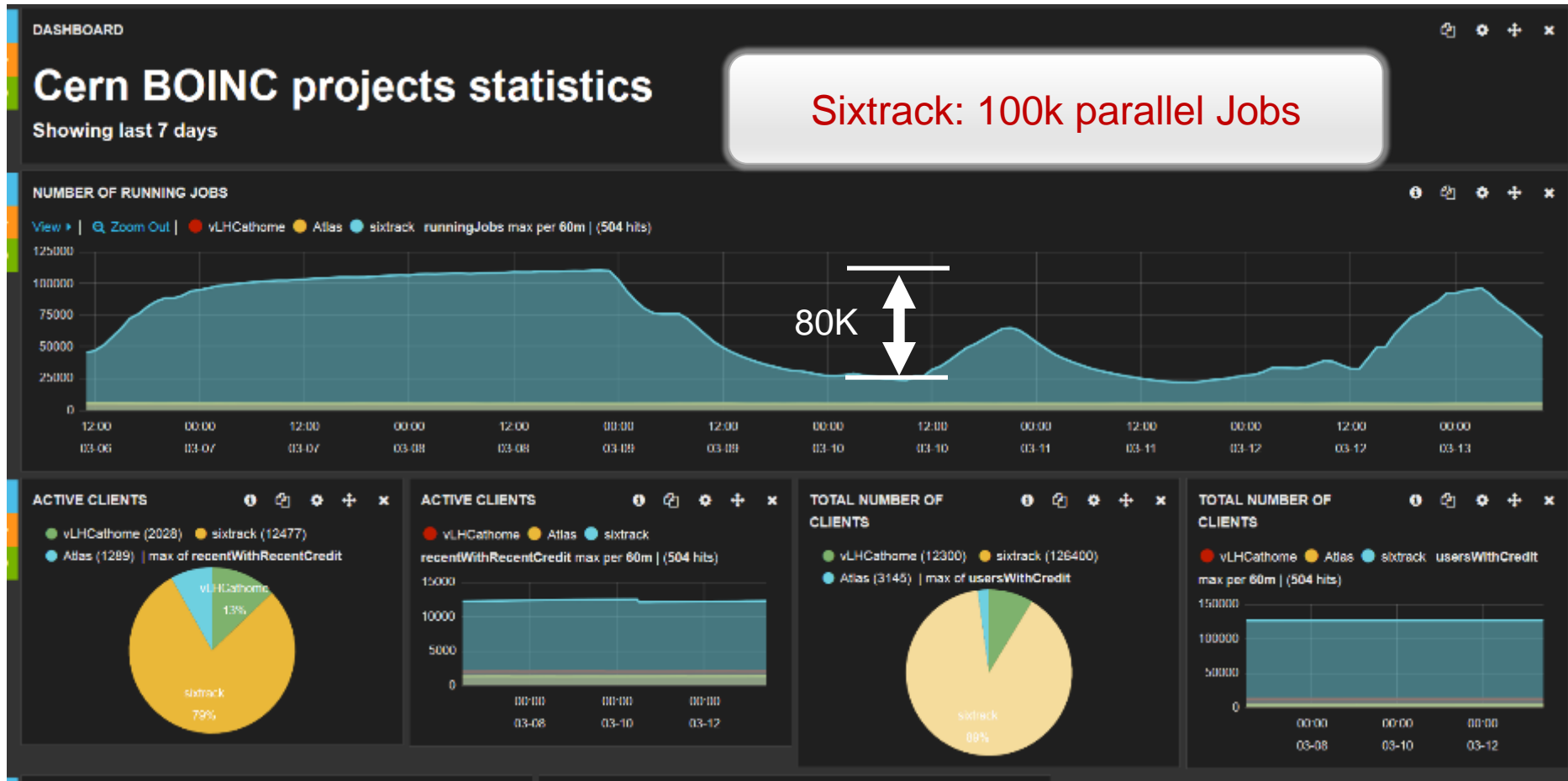
Laurence Field (CERN), Hendrik Borras (Heidelberg University), Daniele Spiga (CERN), Hassen Riahi (CERN)

Motivation

- Free* resources
 - 100K hosts achievable for large projects
 - Actual core count is higher
- 2014 Summer Student project
 - Investigate the feasibility
 - Develop a prototype
- Profited from a clustering of expertise in CERNs IT-SDC group
 - CRAB3
 - Async stage-out
 - Messaging
 - FTS
- And prior experience
 - Test4Theory etc.

* There are cost associated with their use

Opportunistic Resource

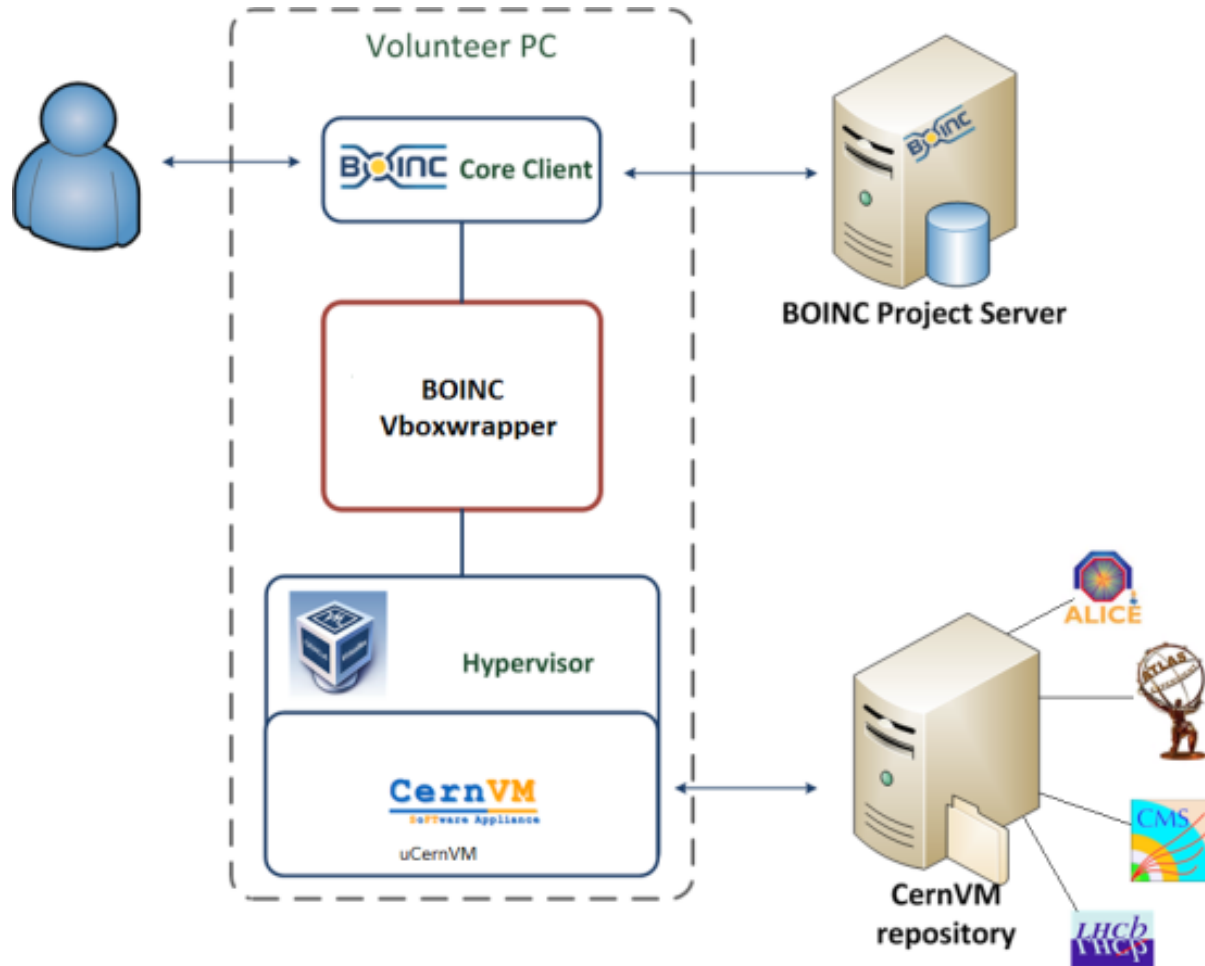


IT-PES <http://cern.ch/go/9nRz>

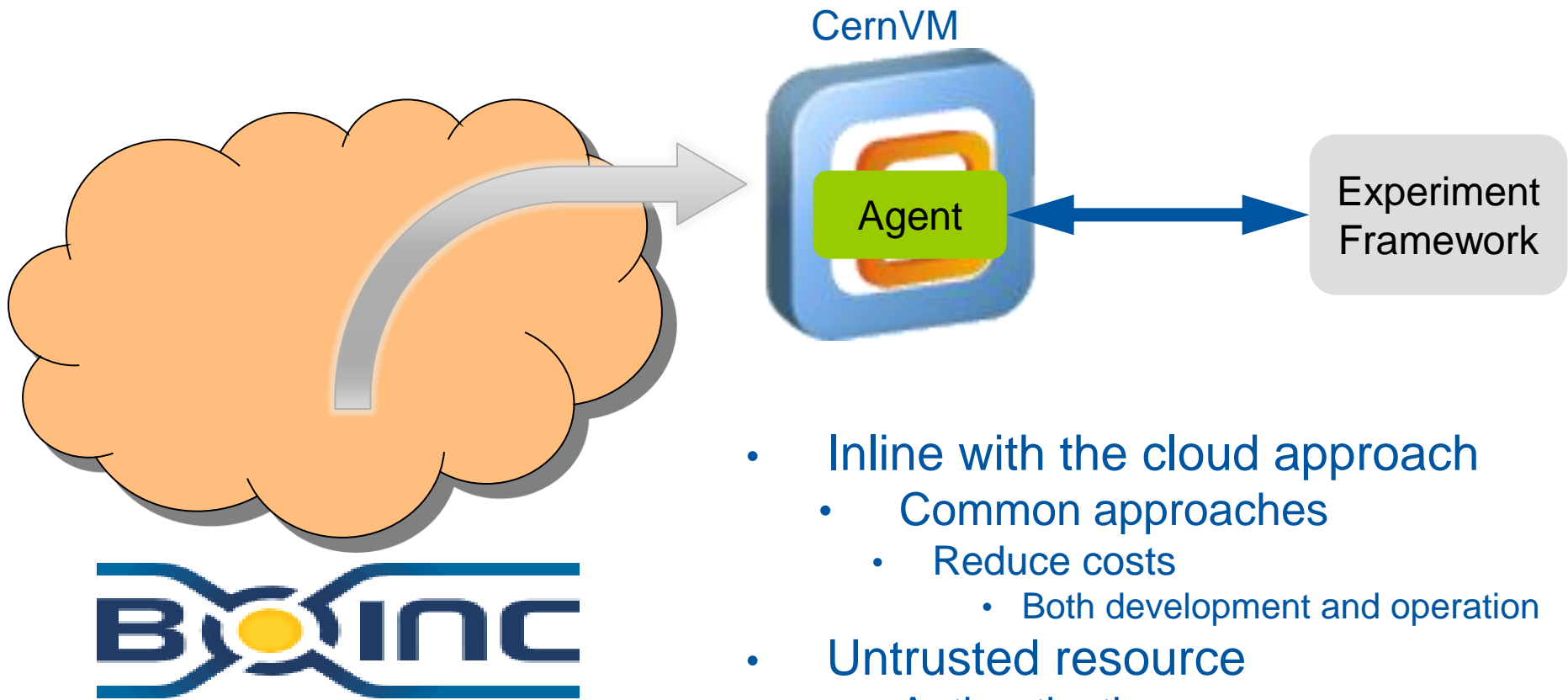
Challenges

- The cost of using the free resources
 - Initial integration requires investment
- Operations and maintenance
 - Public facing support
 - Lowered by community support
- Attracting and retention of volunteers
 - Advertisement
 - Engagement
- Low Level of Assurance
 - Anyone can register as a volunteer
 - Not the same level of trust as with Grid authentication
- Running HEP software on Windows
 - 85% of the resources

BOINC With Virtualization



The Vacuum Model



- Inline with the cloud approach
 - Common approaches
 - Reduce costs
 - Both development and operation
- Untrusted resource
 - Authentication
 - Validation

The Challenge

VO

CRAB3



Volunteer

Volunteer

VM

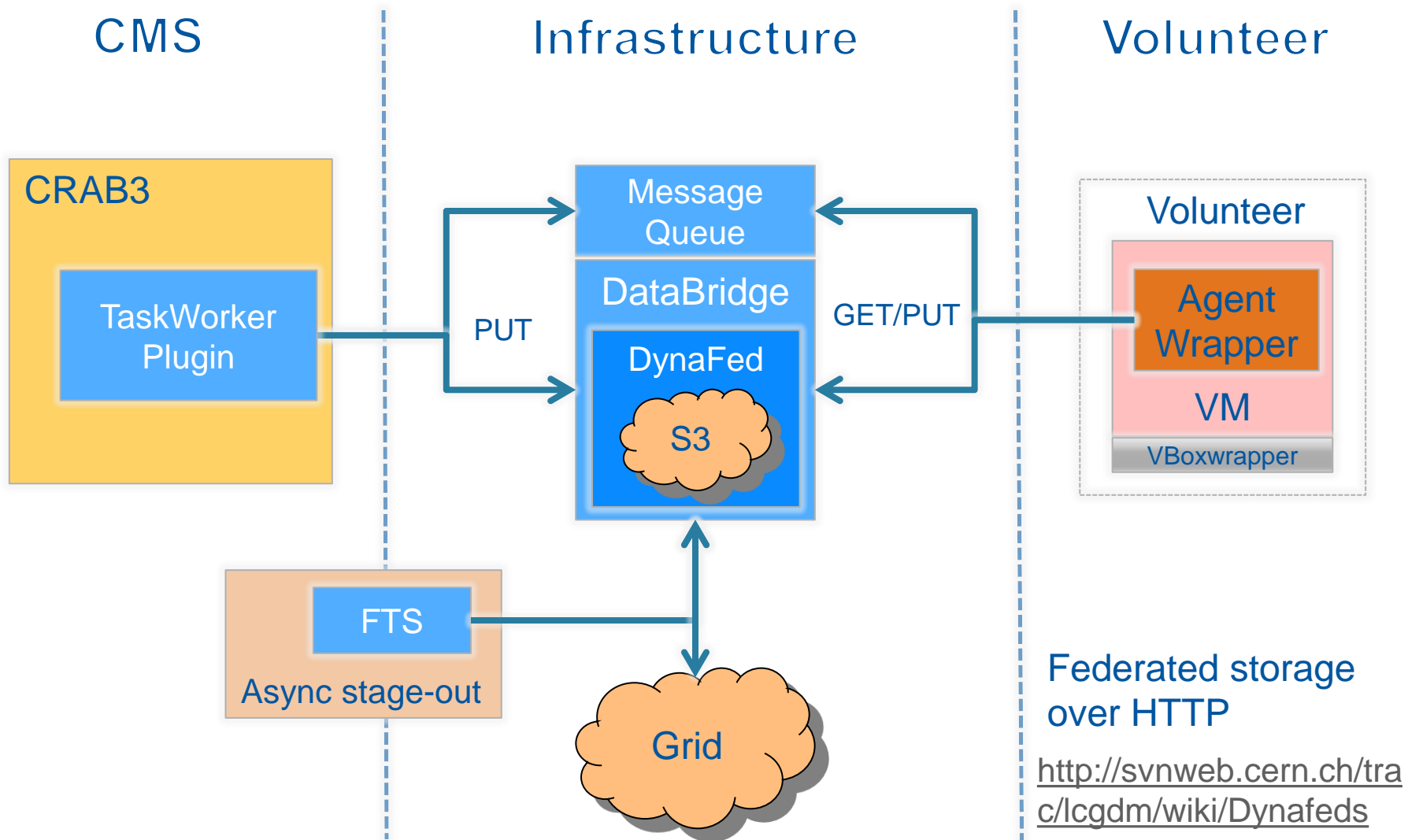
Job Wrapper



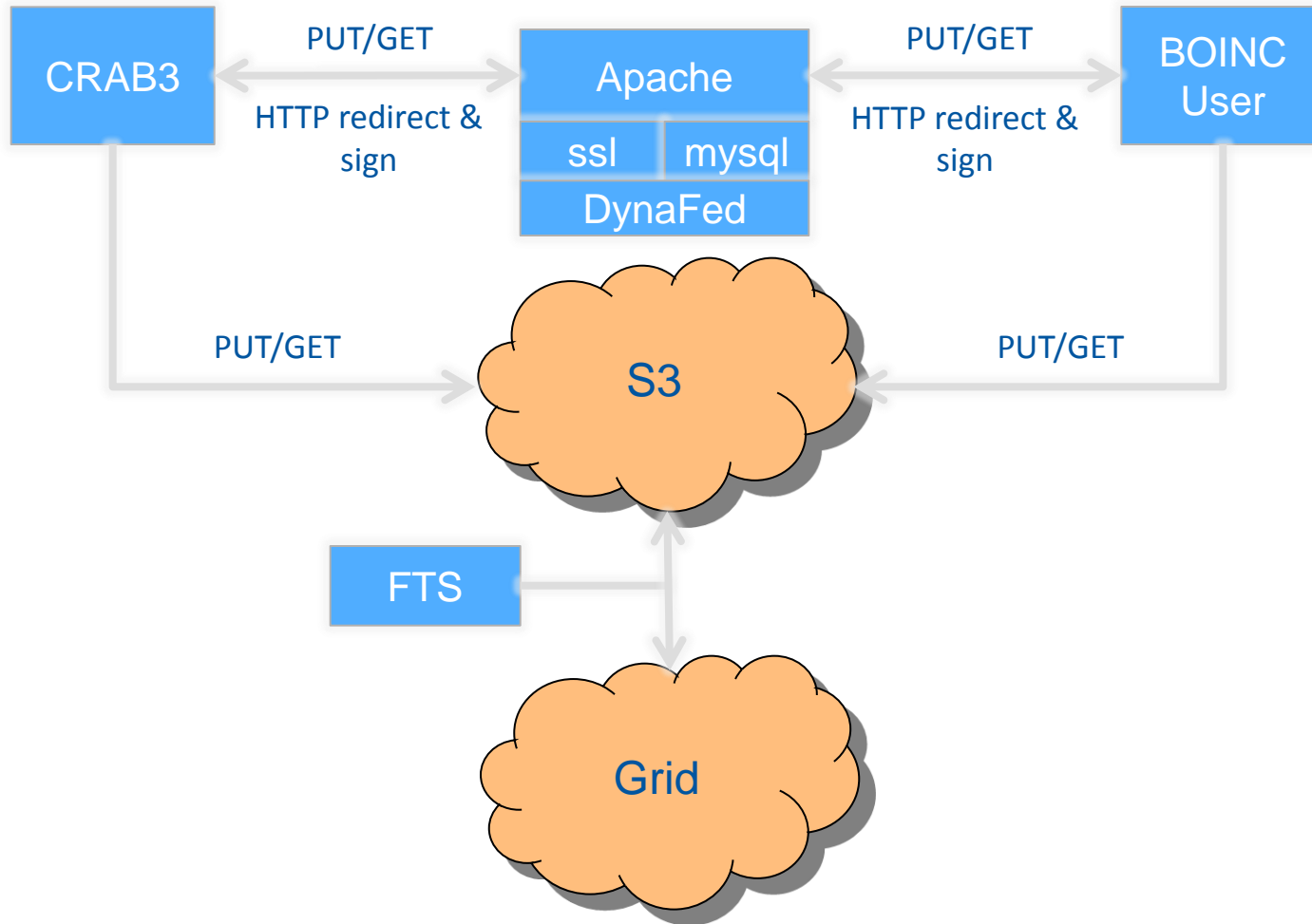
Authentication

- How to authenticate BOINC users?
 - In the VM, credential provided via /dev/fd0
 - BOINC_ID
 - BOINC_AUTHENTICATOR
- Identity Provider (IDP)
 - BOINC Project DB
 - MySQL
 - User Table
- mod_auth_mysql
 - Maps username/password to DB table
 - AuthMysqlUserTable user
 - AuthMySQLNameField id
 - AuthMySQLPasswordField authenticator
- Enables reuse of apache-based HTTP technology

Infrastructure



The DataBridge



Message Queue

- Messaging service does not support BOINC authentication
 - Not clear if it is possible or worthwhile to provide functionality
- Standard apache Web server approach
 - `mod_auth_mysql` to validate BOINC user's credential
 - `mod_auth_ssl` to validate CRAB3 server's x509 credential
- Two simple cgi scripts
 - `put-job.cgi`
 - `get-job.cgi`
- Simple file-based queue
 - `python-dirq`
- Job descriptions from CRAB3
 - Supports arbitrary file types
 - Garbage in, Garbage out
 - Extensible ☺



CMS Application

- CernVM3
 - Minimal Contextualization
 - Similar approach to CMS OpenData
- CVMFS configuration
 - Mount cms and grid
- Add BOINC user
 - Credentials read from /dev/fd0
 - id, authenticator
- CMSJobAgent.py workflow
 - Gets new job description
 - Downloads required input files
 - Runs job
 - Uploads output files
- CMSJobAgent.sh
 - Checks if CMSJobAgent.py is running
 - If not starts it
- cms-agent cron
 - Runs CMSJobAgent.sh every minute

BOINC Server/Project Status

Project status

Server status

Program	Host	Status
data-driven web pages	boincai05	Running
upload/download server	boincai05	Running
scheduler	boincai05	Running
feeder	boincai05	Running
transitioner	boincai05	Running
file_deleter	boincai05	Running
sample_bitwise_validator	boincai05	Running
sample_assimilator	boincai05	Running

Running: Program is operating normally

Not Running: Program failed or the project is down

Disabled: Program is disabled

Computing status

Work	#	Users	#
Tasks ready to send	5,111	with recent credit	21
Tasks in progress	127	with credit	23
Workunits waiting for validation	0	registered in past 24 hours	3
Workunits waiting for assimilation	0	Computers	#
Workunits waiting for file deletion	0	with recent credit	132
Tasks waiting for file deletion	0	with credit	135
Transitioner backlog (hours)	0	registered in past 24 hours	8
		current GigaFLOPs	73

Tasks by application

application	unsent	in progress	avg runtime of last 100 results in h (min-max)	users in last 24h
CMS Simulation	5,111	127	17.67 (1.00 - 24.10)	8

Top Users (Volunteers)

Top participants

Rank	Name	Recent average credit	Total credit	Country	Participant since
1	zombie67 [MM]	13,785	260,032	United States	26 Feb 2015, 6:57:48 UTC
2	jeidler	367	7,356	Austria	14 Feb 2015, 16:40:59 UTC
3	ivan	296	6,530	United Kingdom	20 Jan 2015, 14:40:41 UTC
4	lionelc	232	4,269	France	11 Feb 2015, 19:45:43 UTC
5	koll	210	3,477	Czech Republic	13 Feb 2015, 18:11:08 UTC
6	Crystal Pellet	91	1,741	Netherlands	13 Feb 2015, 21:57:49 UTC
7	Steve Hawker*	59	925	United States	6 Mar 2015, 10:08:57 UTC
8	DeleteNull	40	489	Germany	19 Feb 2015, 23:17:24 UTC
9	Bok	40	554	United States	5 Mar 2015, 14:56:55 UTC
10	philip-in-hongkong	32	651	Hong Kong	6 Feb 2015, 16:40:06 UTC
11	Hendrik	29	884	International	1 Aug 2014, 11:02:53 UTC
12	Jeff17	20	267	United States	15 Feb 2015, 0:45:00 UTC
13	Maxwell [MM]	17	240	United States	23 Feb 2015, 18:32:09 UTC
14	amazing	10	396	Korea, South	11 Feb 2015, 13:54:31 UTC
15	Nick Name	10	123	None	9 Mar 2015, 22:01:38 UTC
16	Laurence	7	97	Switzerland	12 Sep 2014, 9:23:16 UTC

Next Steps

- Try out more diverse workloads
 - Currently running MinBias MC generation
 - Requires further integrations effort
 - To simplify the job injection mechanism
- Be more attractive to volunteers
 - Create CMS web pages for BOINC
 - Improve Internal Web app
- Increase scale
 - Move towards Beta testing in vLHC@home

Summary

- Volunteer Computing can and is providing
 - Significant additional computing resources
 - Potentially $O(100K)$ machines
- Advanced prototype for CMS@home exists
 - Following the Vacuum model
 - Successfully runs CMS Jobs
 - Stage-out results
- Developed the concept of the DataBridge
 - Reused HTTP federation component for S3
 - Added BOINC authentication
- Extended CRAB3 to support his approach
 - Stage-in job description and input data
- Try out more diverse workloads
 - Increase Scale
- Working towards a common platform



www.cern.ch