



Volunteer Computing

Laurence Field
IT/SDC

21 November 2014



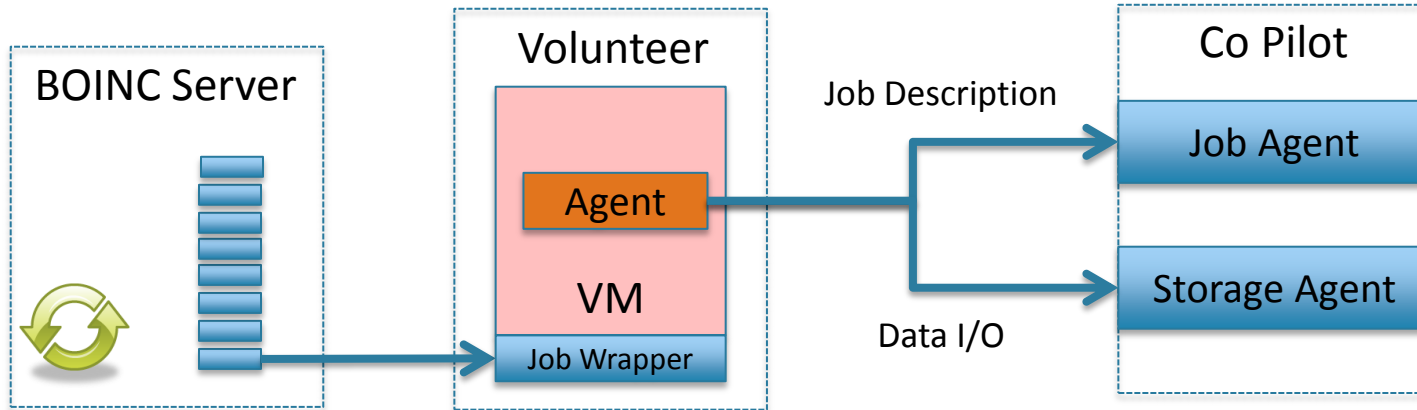
Overview Of Existing Activities

- LHC@home
 - The umbrella for our volunteer computing actives
- SIXTRACK
 - Original classic BOINC project
 - Accelerator beam simulations
 - 10 years old
 - 11K users, 18K hosts with recent credit
- Test4Theory
 - Pioneered the VM approach
 - Runs HEP simulations
 - In production since 2011
 - 2K users, 3K hosts with recent credit

Overview Of Existing Activities

- ATLAS@home
 - Prototype running in Beijing Jan 2014
 - Moved activity to CERN in July 2014
 - Open but not advertised
 - 1K users, 1.2K hosts with recent credit (2K running jobs)
- Beauty@home
 - LHCb's prototype
 - Since June 2013
 - CVMFS and Security Issues
 - Internal testing/use
- CMS@home
 - Started August 2014
 - Advanced prototype

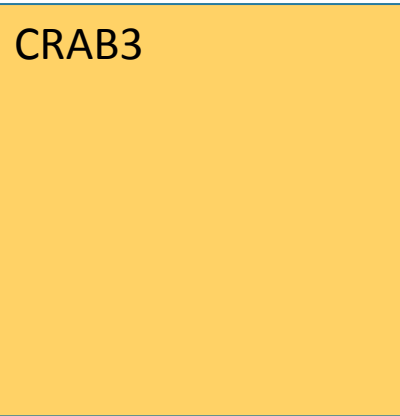
Test4Theory Model



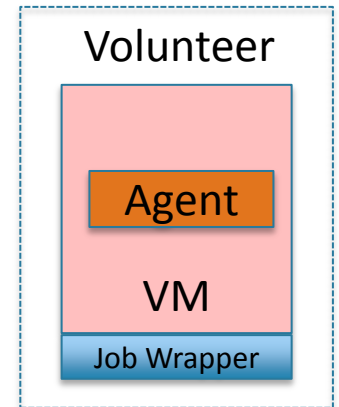
- Avoid restarting the VM for every job
 - Reduces CVMFS related network traffic
- Separate VM management and job management
 - Inline with the cloud model
 - Can reuse cloud related tooling
- CoPilot support challenges
 - Aging codebase
 - Dependencies not available in the standard repositories
 - New standardized components available for some functions

The Challenge

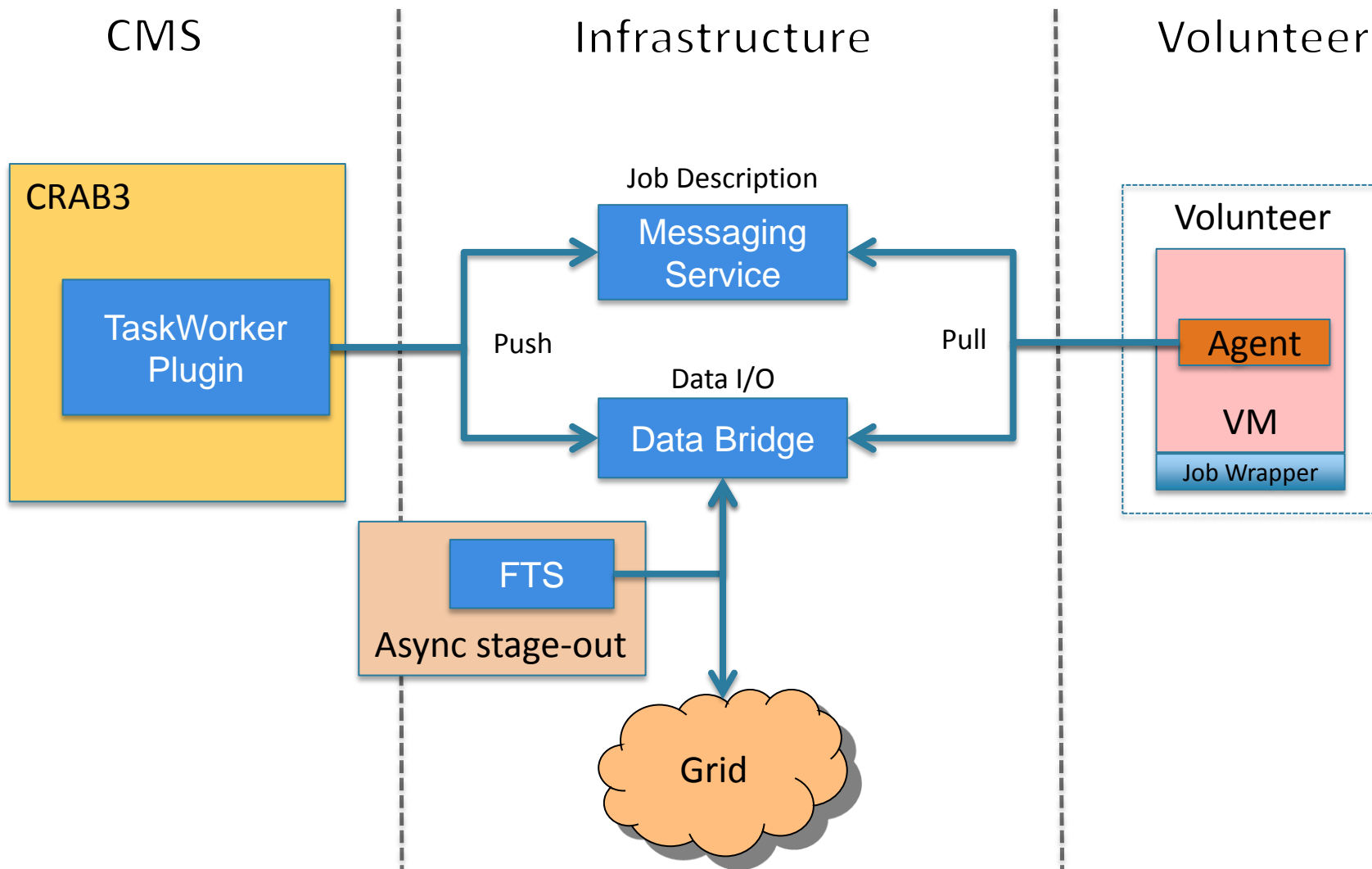
CMS



Volunteer



The Initial Architecture

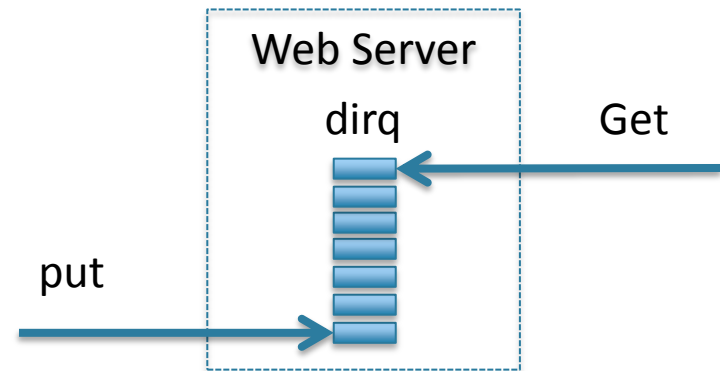


Authentication

- 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

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 ☺



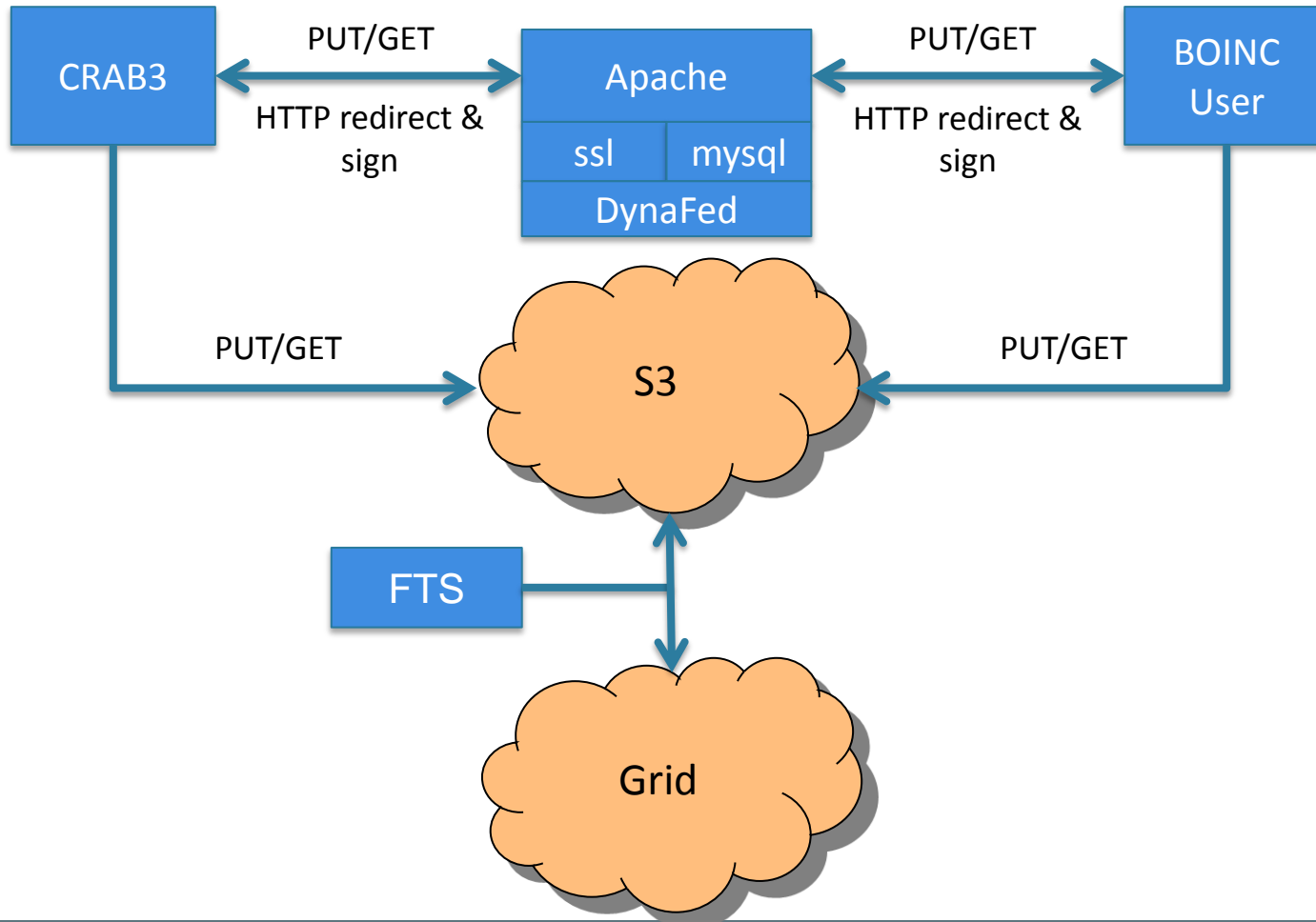
The Data Bridge

- Spans authentication domains
 - BOINC user's credential
 - Grid x509 credentials
- Scalable data I/O
 - With sandboxing capabilities
 - Data Isolation
- Simple apache-based prototype
 - Supports HTTP PUT/GET
 - mod_auth_mysql to validate BOINC user's credential
 - mod_auth_ssl to validate CRAB3 server's x509 credential
- HTTP Federation
 - Possibility to reuse standard DM tools

Dynamic HTTP Federations

- The *Dynafed* implements federated storage over HTTP
 - In testing in LHCb and Canada (ATLAS)
 - Federates WebDAV or S3 enabled storage systems
 - Apache front end
- Can be used as a data bridge
 - S3 storage backend(s)
 - Acts as a security gateway between X509 or *BOINC Auth*
 - Clients then redirected directly to the storage
 - Great scalability potential
 - Global system, smart replica selection (availability, proximity)
- <http://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds>

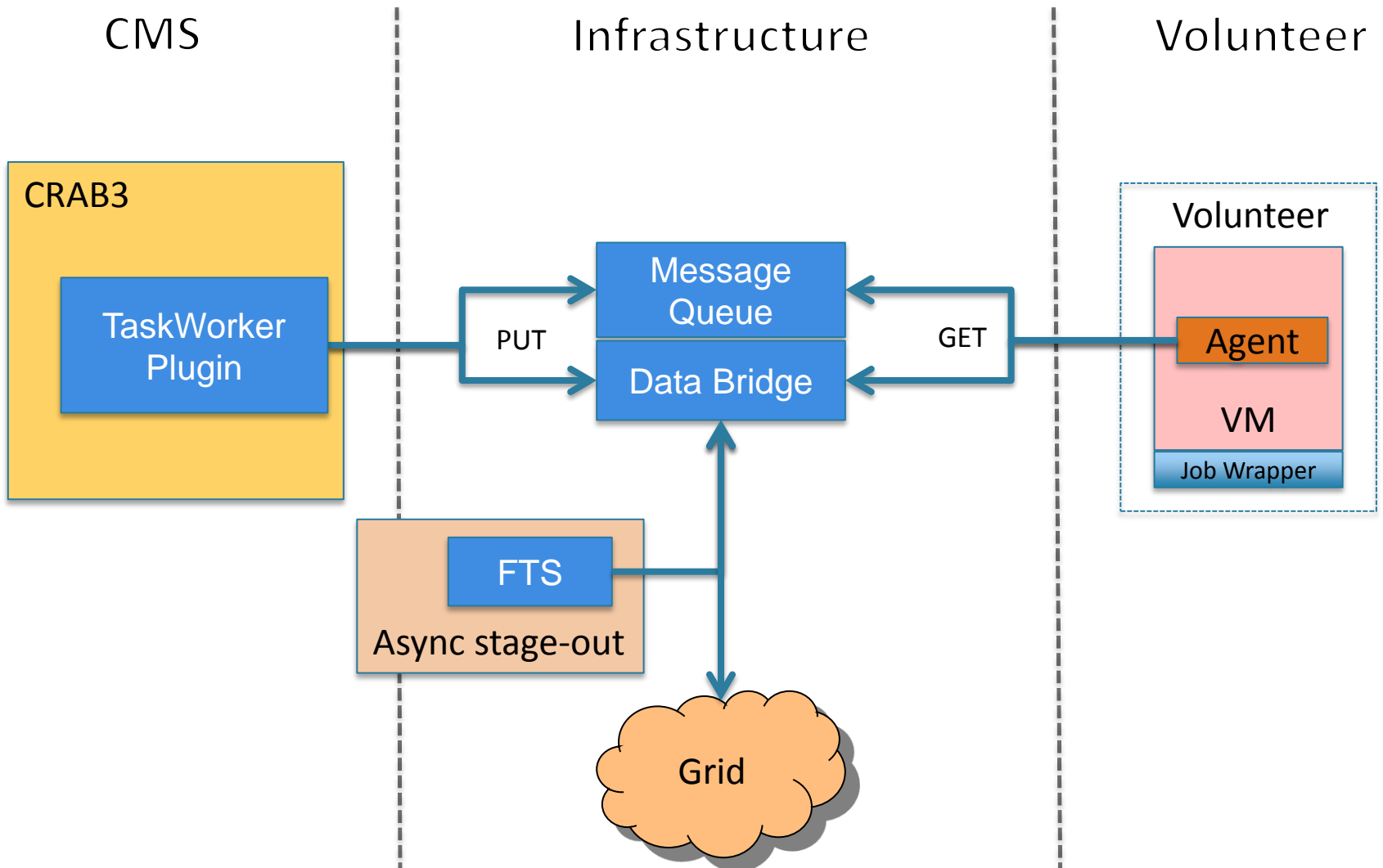
The Data Bridge



VM and Job Agent

- CernVM3
 - Contextualized using CernVM Online
- 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

Final Architecture



ALICE@home?

- Build upon vLHC@home and the data bridge
- Require:
 - Stage input data to the data bridge
 - if needed for simulation jobs
 - Job Agent
 - Method to get job description
 - GET input from the data bridge
 - Run job
 - PUT output on the data bridge
 - Read data from the output *bucket*
 - Similar to HLT
 - Fitter bad data etc.

Summary

- Many ongoing volunteer computing activities
 - Most following the proven Test4Theory model
- Data Bridge concept recently developed
 - Reuses HTTP federation component for S3
 - Added BOINC authentication
- Added a simple message delivery function
 - For the job description
- Working towards a common platform
 - Only require a custom job agent
 - Job description
 - Data stage-in, stage-out

Links

- <http://lhcatlhome.web.cern.ch>
- <http://lhcbatlhome.cern.ch/Beauty>
- <http://boincai05.cern.ch/CMS-dev>