

Project Definition Document

End-To-End Solution using Globus Online Integrated with glideinWMS

1. Table of Contents

1.	Table of Contents	2
2.	Approvals.....	3
3.	Document Change Log	4
4.	Project Proposal Lead.....	4
5.	CD Strategy Document and Tactical Plans.....	4
6.	Problem Statement	4
7.	Requirements	5
8.	Project Description and Goals	5
9.	Project Scope	7
9.1	What is in Scope	7
9.2	What is out of Scope	7
10.	Project Deliverables and Milestones	7
11.	Project Organizational Structure	7
11.1	Sponsor(s).....	8
11.2	Stakeholders.....	8
11.3	Project Team.....	8
12.	Preliminary Project Plan / Statement of Work.....	8
12.1	WBS	8
12.2	Computer Security Considerations.....	9
12.3	Operations Responsibilities at Close of Project.....	9
13.	Estimated Resource Requirements	9
13.1	Personnel Cost.....	9
13.2	Hardware Cost	9
14.	Supporting Documentation.....	9
15.	Project Risks, Issues, and Assumptions	10

2. Approvals

glideinWMS Project Representative: Signature: _____ Date: _____
Print Name: Burt Holzman
Title: _____

Intensity Frontier Representative: Signature: _____ Date: _____
Print Name: Lee Lueking
Title: _____

REX Operations Representative: Signature: _____ Date: _____
Print Name: Margaret Votava
Title: _____

FermiGrid/FermiCloud Representative: Signature: _____ Date: _____
Print Name: Steve Timm
Title: _____

Computing Division Representative: Signature: _____ Date: _____
Print Name: Gabriele Garzoglio
Title: _____

Project Leader & CEDPS Representative: Signature: _____ Date: _____
Print Name: Parag Mhashilkar
Title: Application Developer and System Analyst

3. Document Change Log

Version	Date	Change Description	Prepared By
V 1.0	08/10/2011	First Version of the Document	Parag Mhashilkar
V1.0.1	08/24/2011	Incorporate feedback from the stakeholders	Parag Mhashilkar

4. Project Proposal Lead

Project Leader : Parag Mhashilkar
Department : Computing Division
Group : SCF/GRID/DOCS

5. CD Strategy Document and Tactical Plans

Projects referred in this document are covered in following documents in the Fermilab's docdb -

- CEDPS is covered in tactical plan document : CD-doc-3826
- FermiCloud is covered in tactical plan document : CD-doc-4401
- glideinWMS is covered in tactical plan document : CD-doc-4394

6. Problem Statement

Different experiments that are part of the Intensity Frontier (IF) at Fermilab have been increasing their use of the Fermilab grid computing resources. IF community is using glideinWMS to run most of their analysis workflows on the Grid and local clusters and has developed custom solutions to meet their data transfer needs. This increase in use has exposed the drawback that the ownership of the output files processed through current mechanism is not preserved correctly.

The computing/analysis model that is currently being used by experiments relies on the submission of large groups of simulation and analysis jobs by an individual user, but which run on the grid resources as a more general experimental account under the appropriate grid organization. This authentication chain produces output files that are not directly owned by the original user/group who submitted the jobs.

As a part of this project, we intend to provide an end-to-end solution to the IF community that would integrate well with glideinWMS and their custom data transfer solutions leveraging on the modern techniques and solutions like Globus Online as provided by the Grid community.

It is necessary to investigate and develop a proto-type system that can retain the ownership of the output files produced by user submitting the grid job. This system will provide an End-To-End solution by integrating with the current grid processing infrastructure and the tools that are currently in use for grid submission.

7. Requirements

This section lists a set of functional and performance requirements that a proposed solution should satisfy.

1. The system must be able to obtain the user's association with specific grid organizations based on their authentication and credentials.
2. The system must be able to determine the submitter's local user and group IDs based on the Fermilab maintained NIS services for the Intensity Frontier Experiments.
3. The system must preserve the file ownership properties of a transferred file. This includes username and the user group of the owner transferring the file. In case the user transferring the file belongs to multiple groups, membership should be determined based on the user credential that initiated the transfer.
4. The system must be compatible with existing grid processing infrastructure that is deployed and used by the IF experiments at Fermilab.
5. The system should support existing strong authentications that are deployed at Fermilab including Kerberos based authentication and X509 credentials.
6. The system must support transfer to the Fermilab central disk services (Bluearc) without adversely affecting the performance of the central disk services. This requirement will imply support for appropriate access locks and bandwidth throttling to the central disk.
7. The system must either work for the different IF experiments or it should be convenient enough to replicate for different IF experiments.
8. The system must be modular and support addition and subtraction of different Intensity Frontier experiments without impacting other experiments that are using the provided grid infrastructure.
9. Fermigrd members must have administrative rights on the machine.
10. The system must support a minimum transfer bandwidth of 50Mb/s per client, and an aggregated bandwidth sufficient to saturate the network connection to the storage destination.
11. The system must be able to scale to support a minimum of 500 simultaneous client jobs slots per experiment.
12. The system should include diagnostic tools and support for system monitoring which are compatible with existing tools that are deployed for grid processing, and should be integrated with these tools where possible.

8. Project Description and Goals

The work done for this project is Fermilab's contribution as part of multi-institution collaborative project, Center for Enabling Distributed Petascale Science (CEDPS). One of the aims of the CEDPS project is to identify and involve user communities that can benefit from Globus Online.

We have identified the Intensity Frontier Experiment as one of the user communities that can benefit from some or all the work done as a part of this user engagement process in the CEDPS projects.

The project to provide end-to-end solution is an investigative effort that focuses on providing the IF user community a working prototype of an end-to-end data transfers solution using Globus Online with glideinWMS. Globus Online is a service that optimizes data transfer parameters and provides convenient web-based monitoring interfaces. Internally Globus Online uses gridftp server to do the actual data transfers.

Setup a test bed for testing gridftp server

One of the goals for this project is to setup a test bed for testing gridftp server by the IF and REX community that will preserve the ownership information of the user transferring the files. To achieve this, investigate FermiCloud and use the FermiCloud services to create a Virtual Machine for deploying the test bed.

Develop tools to generate authentication/authorization list for accessing gridftp server

IF VO membership information is available in the Fermilab VOMS service. List of users affiliated with individual IF experiments, their UID and GID are available through the NIS server maintained by FES group. One of the goals for this project is to utilize and integrate this information to develop solution and tools creating the authentication/authorization list for accessing the gridftp server by different IF experiments.

The project understands that the number of individual IF experiments can change. Also, to simplify the operations, different VMs that would host the gridftp server should be identical.

Register the gridftp servers with Globus Online Service

In order to use Globus Online for transferring the files, the gridftp server deployed on VMs should be correctly registered with the Globus Online as valid Globus Online endpoint. The project will appropriately register the gridftp servers requested today and provide documentation for the registration of future possible instances of gridftp.

Develop condor-globusonline-transfer-plugin

Newer version of Condor has plugin architecture that lets the users write scripts to support custom file transfer protocols. Since condor is used as WMS inside glideinWMS, one of the goals is to write a plugin that lets condor handle the Globus Online transfers natively.

Integrate condor-globusonline-transfer-plugin with glideinWMS

In order for the condor in glideinWMS to work with custom transfer plugins, glideinWMS should be enhanced to support condor configuration with custom file transfer plugins.

Deploy glideinWMS for end-to-end testing

To enable end-to-end testing and the commissioning of a functional system for the IF community, one of the goals for this project is to deploy glideinWMS with support for custom condor file transfer plugins.

9. Project Scope

This section describes what is in scope and what is out of scope for the project. If the proto-type developed as a part of this project is accepted for production, this project should be closed and a new project should be opened to deploy and migrate the system to production.

9.1 What is in Scope

Scope of the project includes –

- Utilize FermiCloud resources to deploy services
- Develop administrative tools to provide authentication/authorization to the deployed services
- Develop tools to interface the solution with glideinWMS and with Globus Online
- If by the end of this project, stakeholders decide to accept the proto-type and migrate it to production -
 - Provide consulting level support to the stakeholders until the system is deployed in production.
- Archive the tools developed and the FermiCloud Virtual Machines and images as deemed necessary.

9.2 What is out of Scope

Tasks/items not listed in the section 9.1 are out of scope of this project.

10. Project Deliverables and Milestones

Some of the high level milestones and deliverables for the project are –

Milestones/Deliverables	Requester/ Stakeholder	Planned For
Develop tools to generate authentication and authorization list for accessing gridftp server	IF, REX, FermiCloud	06/27/2011
Setup a test bed for testing gridftp server	IF, REX, CEDPS	06/29/2011
Develop condor-globusonline-transfer-plugin	CEDPS, glideinWMS	08/03/2011
Integrate condor-globusonline-transfer-plugin with glideinWMS	CEDPS, glideinWMS	08/19/2011
Deploy glideinWMS for end-to-end testing and commissioning	IF, REX, CEDPS	08/26/2011
Deploy gridftp servers the required IF experiment(s)	IF, FermiCloud, CEDPS	08/28/2011
Register the gridftp servers with Globus Online Service	CEDPS	08/29/2011
Close the Project	CEDPS	09/30/2011

11. Project Organizational Structure

The program of work and efforts for this project are partially sponsored by CEDPS activity in the Fermilab.

11.1 Sponsor(s)

Effort for this project is partially funded by the CEDPS activity in the Fermilab.

11.2 Stakeholders

glideinWMS Project : Burt Holzman
 IF VO : Lee Lueking
 REX Operations : Margaret Votava
 FermiGrid/FermiCloud : Steve Timm
 Computing Division : Gabriele Garzoglio

11.3 Project Team

Name	Project Role	Effort (FTE Equivalent)
Burt Holzman	Project Manager	10% (07/01/2011 – 09/30/2011)
Parag Mhashilkar	Project Leader	30% (07/01/2011 – 09/30/2011)
Marko Slyz	Team Member	5% (07/01/2011 – 09/30/2011)
Xi Duan	Summer Student	50% (07/15/2011 – 09/15/2011)

12. Preliminary Project Plan / Statement of Work

12.1 WBS

- Project Documentation
 - Communicate with the stakeholders
 - Gather requirements
 - Write project definition document
- Develop tools to generate authentication and authorization list for accessing gridftp server
 - Script to generate gridmapfile for the gridftp server
 - Script to generate local users & groups by contacting the IF NIS
- Setup a test bed for testing gridftp server
 - Create a test VM on FermiCloud if-gridftp.fnal.gov
 - Deploy gridftp server on if-gridftp.fnal.gov
 - Deliverable: if-gridftp.fnal.gov VM with gridftp server that correctly preserves the group membership of transferred files
- Deploy one gridftp end point per IF experiment for testing as required
 - Create a VM with the gridftp server deployed for one experiment
 - Clone the VM for other experiments as required
 - Deploy multiple clones, one per IF experiments
- Register the gridftp servers with Globus Online
 - Register the gridftp servers used by IF experiments as individual GO end points
- Develop condor-globusonline-transfer-plugin
 - Develop condor-globusonline-transfer-plugin
 - Test condor-globusonline-transfer-plugin
 - Release condor-globusonline-transfer-plugin
- Integrate condor-globusonline-transfer-plugin with glideinWMS

- Integrate condor-globusonline-transfer-plugin with glideinWMS
 - Test the integration work
 - Release glideinWMS with support for globusonline-transfer-plugin
- Deploy glideinWMS for end-to-end testing
 - Deploy glideinWMS
 - Test the deployment
 - Commission the deployment
- End-To-End Test for running jobs on IF glideinWMS deployment
 - Do end-to-end testing using glideinWMS and Globus Online
- Close the Project
 - Write the closing document
 - Deliverable: Closing document

12.2 Computer Security Considerations

Hardware and software/services required for the project will be hosted at Fermilab and will follow the Computer security policies of Fermilab. The Globus Online Service is jointly hosted and maintained by the Globus team at Argonne National Laboratory and University of Chicago. Gridftp server that acts as the globusonline.org endpoint is deployed in the virtual machine hosted by the FermiCloud.

12.3 Operations Responsibilities at Close of Project

Hardware, virtual machine and the operating system that hosts gridftp endpoint will be hosted at Fermilab and will follow the Computer security policies imposed by Fermilab. Hardware and operating system for the virtual machine is supported by the FermiGrid operations group and will continue to remain the same. VDT installation required for the deployment of the gridftp and gridftp operations will be maintained and supported by the REX group.

13. Estimated Resource Requirements

This section describes the personnel and equipment cost for the project.

13.1 Personnel Cost

Personal cost is listed in the Project Organization structure section above.

13.2 Hardware Cost

Estimating the hardware cost is outside the scope of this document. The hardware and the resources utilized for this project are hosted by the FermiCloud project. Costing for the resources will be done by the FermiCloud project and presented to the IF community separately.

14. Supporting Documentation

The project web home page is <https://cdcvs.fnal.gov/redmine/projects/cedps-glideinwms>

This link includes documentation for two individual subprojects, IF Gridftp setup and Condor-globusonline-transfer-plugin.

15. Project Risks, Issues, and Assumptions

Risk	Impact Level	Risk Plan Actions
Support for Globus Online being dropped	Medium	Globus Online Services are maintained by the Globus Group at Argonne National Laboratory & University of Chicago. In case the Globus Online Services project is discontinued, the proposed solution will still work since it depends on the availability of the gridftp server. A new condor file transfer plugin needs to be written or the existing working model needs to be adjusted to use gridftp end point natively.
Delay in migrating the ownership of condor-globusonline-transfer-plugin the to Condor community	Medium	This will need working with the glideinWMS team so that they continue supporting the plugin as part of the glideinWMS project.
Delay in closing the project by 09/30/2011	High	This project is funded by the CEDPS activity in the Fermilab. If the work on this project is delayed, an alternate source of funds to support this activity past its original completion date of 09/30/2011 will be required.