



Multi Pronged Approach to Pilot Log Anonymization in GlideinMonitor

Mirica Yancey

Supervisor: Dr. Marco Mambelli

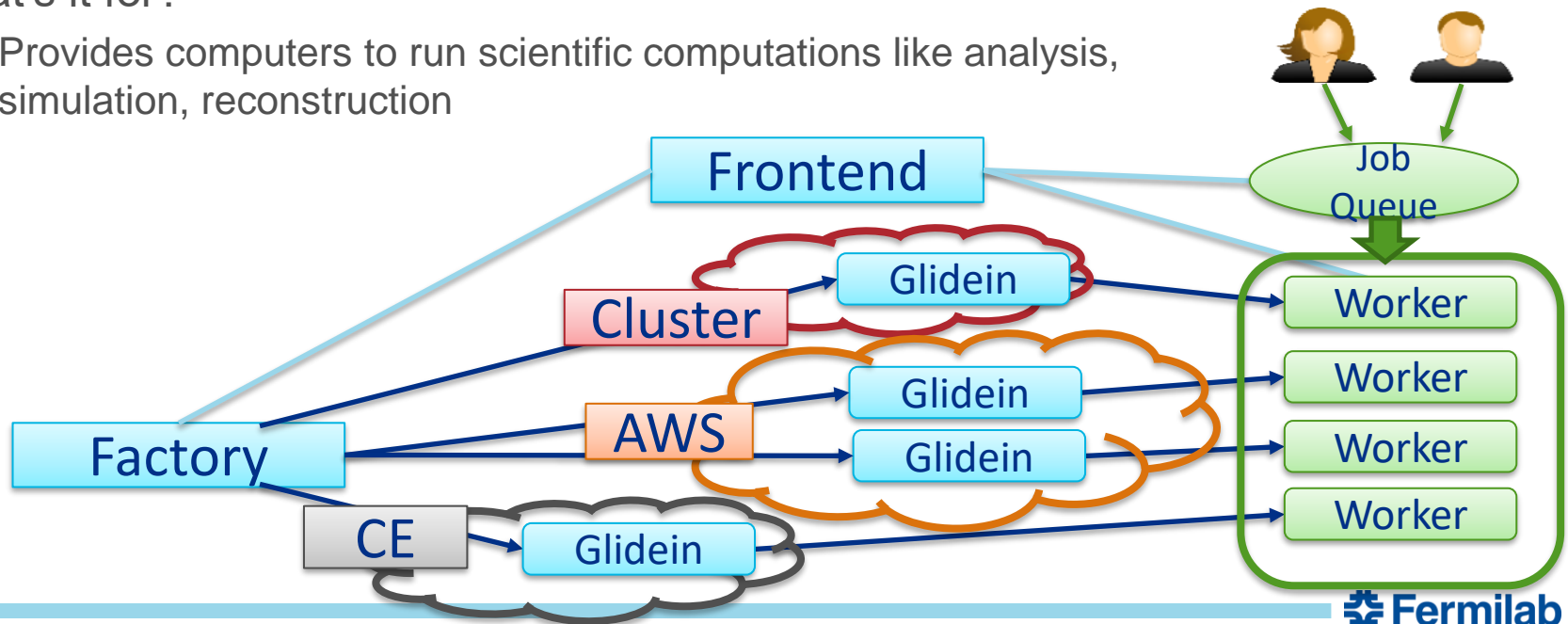
SIST Introduction Presentation

03 June 2020

GlideinWMS

GlideinWMS is a pilot-based resource provisioning tool for distributing HTCondor

- Provides reliable and uniform virtual clusters
- Components
 - Glidein: Execution node submitted as a job. Defines how to split resources
 - Frontend: Matches resources to user jobs and sends request to Factory
 - Factory: Receives requests from Frontend and submits glideins to grid sites
- What's it for?
 - Provides computers to run scientific computations like analysis, simulation, reconstruction



GlideinMonitor

- A Web application to view GlideinWMS's Glidein log files
 - A user interface tool
 - Tools for quick searches and decoding log content
 - An efficient, managed archive of log files
 - Framework to add log processing

- Two Components

- Indexer (glideinmonitor-indexer)
 - Collects, prepares and archives new log files
- Web server (glideinmonitor-webserver)
 - Serves and decodes archived log files

- Problem

- These log files have sensitive information (ex.IP addresses, information about current jobs) that need to be removed to make them easier to share and more secure.

GWMS Factory Job Logs Home Goto GUID

Factory Monitoring Job View

Filters below alter the data in the table
Click search once you have narrowed the query

Timestamp From: [] Timestamp To: 08/13/2019 1:47 AM Entry Name: entry_HCC_US_Omaha_crr

Copy Excel Print

Show 10 entries

JobID	FileSize	Timestamp	FrontendUsername	InstanceName	EntryName	Master
job.100.0	18061	2015-11-05T09:28:13-06:00	user_frontend	glidein_gfactory_instance	entry_HCC_US_Omaha_crane_gpu	False
job.101.0	18031	2015-11-05T09:43:31-06:00	user_frontend	glidein_gfactory_instance	entry_HCC_US_Omaha_crane_gpu	False
job.102.0	18027	2015-11-05T09:59:38-06:00	user_frontend	glidein_gfactory_instance	entry_HCC_US_Omaha_crane_gpu	False
job.103.0	18011	2015-11-05T10:09:37-06:00	user_frontend	glidein_gfactory_instance	entry_HCC_US_Omaha_crane_gpu	False
job.104.0	0	2015-11-05T10:09:37-06:00	user_frontend	glidein_gfactory_instance	entry_HCC_US_Omaha_crane_gpu	False

GWMS Factory Job Logs Home Goto GUID

Job 900

Time: 2019-07-20T23:40:15-05:00

General Information

Timestamp	1563684015
FileSize (err + out)	196219
Entry Name	entry_ITB_FC_CE2
Instance Name	glidein_gfactory_instance
Frontend Username	user_frontend
GUID	user_frontend@glidein_gfactory_instance@entry_ITB_FC_CE2@job.4622.37.0

Data Files

Open

Full Logs

- 900.out
- 900.err
- 900.tar.gz
- 900.json

Condor Logs

- Master Log
- Startd Log
- Starter Log
- StartdHist Log
- XML Description

Log Search

Output & Error Log Combined

error Search

ERROR_GEN_PATH /scratch/12131510.1.long/home_crn09_017103526/CREAM017103526/glide_n1060/main/error_gen.sh
error_gen-373908.sh: OK

Solution: Log Anonymization

- Problem
 - The log files contain sensitive information such as IP Addresses and information about the current running jobs
 - The goal is to leave only information for statistics and troubleshooting
- Research
 - Research and evaluate different log anonymization techniques
- Implementation
 - Implement reversible and irreversible anonymization
 - Develop code in Python to integrate log anonymization in GlideinMonitor
- Testing
 - Test my solution in distributed computing environments (Grids, Clouds, and Supercomputers)
 - Present my results.

Second Project: Computing Support in Python

- Support other Fermilab summer interns in Python!
 - Check out the new web pages, presentations, and lessons at:
 - <https://sites.google.com/view/targetpython/>
 - Join me to get support in the Fermi Coding Club slack and summer-interns channel:
 - fermicodingclub.slack.com

```
TARGETPython Home Start Lessons Projects Survey
7 def my_function(c): # c is a local variable bc it was defined in the function
8     d = 3 #This is a local variable
9     print(c)
10    print(d)
11    a = 5 #This is a global variable
12    print(a)
13
14 my_function(7) #This prints 7, 3, 5
15 print(a) #This prints 5 because it's accessing the GLOBAL a
16 print(b) #This will print GLOBAL b
```



What It's For

TARGETPython is an introductory course for Python that will take you through Python Basics, Computer Science Concepts, the use of Matplotlib and more! You will learn about Git, a tool to organize and share code, Machine Learning and find some guidelines about preparing a project required to complete the TARGET Python course.

```
16 self.race = race
17
18 def celebrate_birthday(self):
19     print("Happy Birthday!")
20     self.age += 1
21     self._message_attribute can be used in the function but NOT outside
22
23 def _checkAge(self): #in python, to say something is private you do underscores on either side
24     if self.age > 20:
25         print("Congrats! You are able to vote and drink!")
26     elif self.age > 17 and self.age < 21:
27         print("Congrats! You are able to vote!")
28     else:
29         print("You are a minor!")
30
31 #this is called a getter, getters get information from your object
32 def get_name(self):
33     return self.first_name + " " + self.last_name
34
35 #this is called a setter, this sets/overwrites information in your object
36 def change_full_name(self, new_first, new_last):
37     if new_first != None and new_last != "":
38         self.first_name = new_first
39         self.last_name = new_last
40
41
```

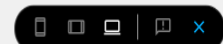
What You'll Find

Here you'll find presentations with that explain coding basics and self guided lessons with corresponding challenges to learn the basics of Python using Jupyter Notebooks.

For each lesson, you should first look over the presentation corresponding to the lesson, then do the Python Lesson, then the Challenges.

Finally, you will end this course by completing a final project.

In the menu you will find direct links to quickly go to the previously mentioned materials. This page also contains links to the lessons introduced in order.



Reminder:

You can always ask questions at (fermicodingclub.slack.com)!

I will be online frequently, and office hours are between 2pm and 3pm!