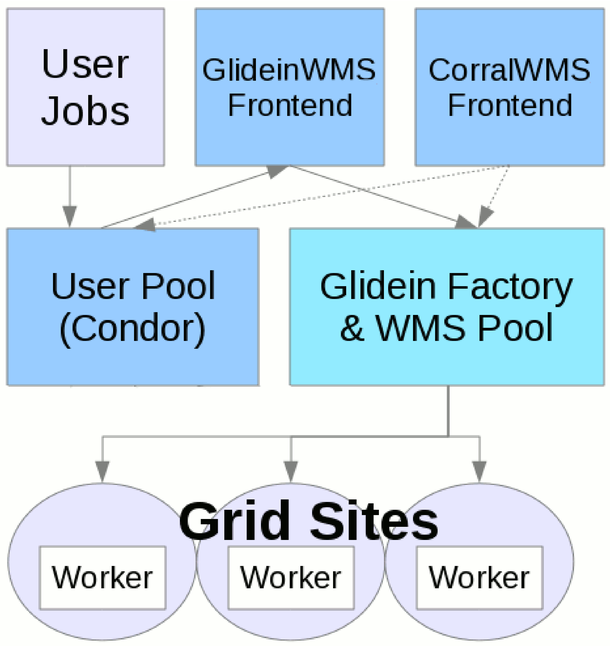
Harnessing the Power of AI and CI/CD to Fuel Scientific Discovery

Alex Umeasalugo, Grambling University – Mentors: Marco Mambelli, Namratha Urs, Bruno Coimbra, FNAL

GlideinWMS

Glidein Workload Management System is a pilot-based resource provisioning tool for distributed High Throughput Computing. It Provides reliable and uniform HTCondor virtual clusters and submits Glideins to heterogeneous resources.



The Distributed architecture of the GlideinWMS includes:

- Glidein: These set up resources and run tasks queued in by the users
- Factory: The Factory has a catalogue to determine if a computing resource has a GPU or not, so it does not send Glidein that won't be used.
- Frontend: The Frontend looks for how many Glidein needed to run a job and then it sends them and put the same number in each queue of the system.

CMS produces an enormous amount of data. Thousands of physicists submit **millions of jobs** for data reconstruction, simulation, and analysis.
Resources are spread across **the Worldwide LHC Computing Grid (WLCG)**, cloud providers, and HPC centers.

What GlideinWMS does:

- It dynamically provisions compute resources from these distributed systems.
- It uses pilot jobs to create a global HTCondor pool, where all jobs are managed centrally.
- It ensures efficient scheduling and better utilization of heterogeneous resources.

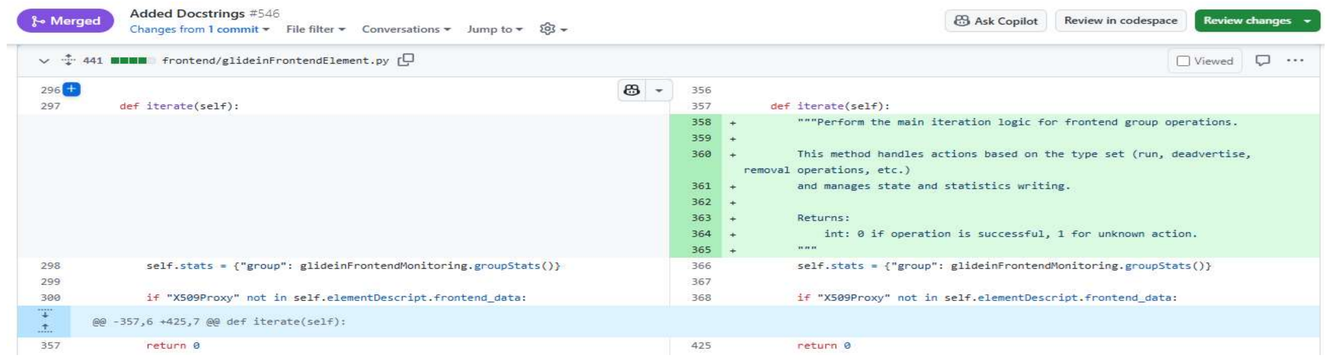
PROJECT ACHIEVEMENTS

- Created a Virtual Machine with Fermicloud to run container applications
- Used containers to run GlideinWMS
- Learned to use VSCode and GitHub to collaborate on software development
- Improved my Python knowledge and wrote docstrings to two GlideinWMS Python files using ChatGPT
- Boosted code coverage with AI: Used AI-powered tools and my coding skills to generate Python unit tests and improve software quality.
- Learned Bash scripting and wrote scripts to automate GlideinWMS testing

PROJECT RESULTS

```
[root@fermicloud752 ~]# podman images
REPOSITORY                                TAG         IMAGE ID
localhost/glideinwms/factory-workspace    latest      71fb95c21261
localhost/glideinwms/frontend-workspace   latest      b0a9300695cd
localhost/glideinwms/ce-workspace         latest      2dde08877e9a
localhost/glideinwms/gwms-workspace       latest      841281435b15
docker.io/library/almalinux               9           623706a2d956
```

```
[root@fermicloud752 ~]# podman container ls
CONTAINER ID  IMAGE                                       COMMAND
      NAMES
cdbe6d2d8e5f  localhost/glideinwms/ce-workspace:latest    /bin/bash
      ce-workspace.glideinwms.org
56aee320bd77  localhost/glideinwms/factory-workspace:latest  /bin/bash
/tcp  factory-workspace.glideinwms.org
bd11f9fd7973  localhost/glideinwms/frontend-workspace:latest  /bin/bash
/tcp  frontend-workspace.glideinwms.org
```



Acknowledgement

This work was supported by the DOE grant RENEW-HEP: U.S. CMS SPRINT - "A Scholars Program for Research Internship" at Tougaloo College, MS (DE-SC0023681), Brown University (DE-SC0023651), RI,– Madison, WI (DE-SC0023643) and U.S. CMS Operations at the Large University of Puerto Rico-Mayaguez, PR (DE-SC0023680), and University of Wisconsin Hadron Collider (NSF-2121686)