Derek Morris Jr.
Science Gateways Community Four-Week Coding Institute
7/15/17

The Practice & Experience in Advanced Research Computing (PEARC17)

Going into The Practice & Experience in Advanced Research Computing (PEARC17) conference I did not know what to expect. It turned out to be extraordinary. The PEARC meetings provide conversations regarding difficulties, openings, and arrangements among HPC focus executives and directors, computational researchers, end clients, students, facilitators, instructors, HPC framework chairmen and client support staff, and additionally industry and government office agents from over the United States and around the globe.

During my stay at PEARC conference we had to do volunteer hours. The requirement was a 6 hours minimal. You were able to do more hours if you wanted. I think it was a good idea because this helps the student volunteers come out there bubble. Mingling with other people is a big part of society now of days. Mingling gives you a better chance of getting an internship or a job. Most agency and schools were looking how is your social skill; are you capable interacting with people/difference race.

A presentation I attend to was the High-Throughput Computation on the Open Science and AWS. The presenters were Robert Gardner and Mats Rynge.The Open Science Grid gives a texture of administrations to empower appropriated, high-throughput processing abilities over the US. When all is said in done, OSG is sorted out into VOs (Virtual Organizations) around expansive analyses, yet singular PIs and gatherings too little for the VO model can utilize the facilitated administrations given by OSG Connect. The configuration of this instructional exercise is a blend of address and hands-on works out, so we needed to bring a tablet/computer and we had to ensure that we had SSH customer introduced. The objective of this session is to figure out how to run and scale up workloads, deal with your information and furthermore overflow into Amazon cloud assets. In this area, we had learned the basics HTCondor in submitting and checking workloads, or "jobs". The occupations are submitted through the submit host. The submitted occupations are executed on the telecommuter node(s) and the logs and, if configured, yields are transferred back to the login hub. In the HTCondor job submit document, we need to depict how to execute the program and exchange information.

First thing was the job execution script. We signed with the username and password provided. Then we ssh into the email that was also provided to us. Now we started to get the tutorial file. Once that happened it created a directory “tutorial-quickstart”. Once cd into the tutorial-quickstart directory we open the short.sh and tutorial.submit to take a look at the shell script. You can use any editor you are comfortable with; in this case we open the files in nano. After looking thru the files we used a command called condor_submit and of the file you wanted to submit. When submitted it will tell you how any jobs submitted and how many clusters there are. Clusters are numerous of data sent to the one database. Then we were looking for the status; to do that we used the command condor_q. The condor_q command tells the status of currently running jobs. Please note, that the condor_q command line interface has changed in recent

HTCondor versions, and in this tutorial we are using the new version. Condor_q shows the number of query, username, and all users. Inside each one it also shows how many files complete, remove, idle, running, held, and suspended.

If you wanted to see individual job you had to condor_q –nobatch and name. That will output everything from that specific job. You can also get the status on a specific job condor_ -nobatch and ID. Once your job has finished, you can get information about its execution from the condor_history command. You can see much more information about your job's final status using the -long option. Once your job has finished, you can look at the files that HTCondor has returned to the working directory. If everything runs okay; it will output fie for each job's output, an error file for each job's errors, and a log file for each job's log. The cat is short for concatenate; command is one of the most frequently used command in Linux/Unix like operating systems. Cat command allows multiple files or us to create single, view contain of file, concatenate files and redirect output in terminal or files. Once we knew how to create files, we want to know how to delete them. And once we can schedule workloads across thousands of computers simultaneously, we needed to know how to remove them. The command for that is condor_rm.

In addition, to that I had present two poster that I did in the previous years which are "A continuing study of the water quality in the Pasquotank Watershed in Northeastern North Carolina" and "Update of the CERSER TeraScan Cataloguing System and the TeraScan Image Processing". Update of the CERSER TeraScan Cataloguing System and the TeraScan Image Processing focus was, " The Center of Excellence in Remote Sensing Education and Research (CERSER) on the campus of Elizabeth City State University is currently tasked with the responsibility of receiving remotely sensed data from National Oceanic and Atmospheric Administration (NOAA) Polar Operational Environmental Satellites (POES) and Geostationary Operational Environmental Satellites (GOES). This data is collected by SeaSpace TeraScan systems installed in the CERSER labs in Dixon-Patterson Hall. In 2005, the processing system underwent a major update due to a migration to a new operating system. A minor update was needed at this time to deal with a second operating system migration and display of the processed images on the CERSER web site. Since then, a second transfer to a new server was made in 2013. The cataloguing system went down at this time and was not repaired due to technical issues with the TeraScan system. The 2014 team corrected issues within the current server directory system and updated the data script to process images from the GOES-13 satellite received by the TeraScan system. Software and languages utilized for this task included ImageMagick, PHP, HTML, Dreamweaver, phpMyAdmin, and MySQL. Along with this operating system update, a major script development was needed on the TeraScan processing equipment due to an upgrade in hardware. The ground station upgrades included a 3.7m X/L band, a 3.6m C band, and a 5.0m L band dishes, along with accompanying computing hardware. This project implemented new script on the TeraScan and CERSER servers. This new script processes both infrared and visible light images received from the GOES-13 satellite into the Tagged Image File (TIFF) Format." The A continuing study of the water quality in the Pasquotank Watershed in Northeastern North Carolina poster focuses on "The Pasquotank River Watershed is found in Northeast North Carolina beginning in the Great Dismal Swamp at the Virginia/North Carolina border and flows into the Albemarle Sound. The watershed provides a transition between the Great Dismal Swamp and the waters of the Albemarle

Sound. The watershed is surrounded by a variety of landforms including swamps, farmland, and suburban developments. These produce a variety of runoff into the watershed affecting both the aquatic vegetation and marine life in the waters. This project built on the previous analysis of the four tributaries and the Pasquotank River completed in 2011, 2013, 2014, and 2015. The 2016 Research Experience for Undergraduates Pasquotank River Watershed Team completed one set of tests of the watershed. These test points originated from the 2011 and 2013 research projects with the addition of four points created in 2014 to sample further downstream in the Pasquotank River. The results were compared with previous readings utilizing a Water Quality Index (WQI), a unitless number ranging from 1 to 100 with higher numbers denoting better water quality. The waterways tested were the Pasquotank River, Newbegun Creek, Knobbs Creek, Areneuse Creek, Mill Dam Creek, and Sawyers Creek. These creeks, along with the Pasquotank River, cover a large portion of the watershed and provided a wide area of study for the watershed. Tests performed in the laboratory on this year's samples include pH, salinity, total dissolved solids, and conductivity. Air/water temperature, dissolved oxygen, wind speed/direction, and turbidity/clarity measurements were taken in the field. The results collected were placed online and displayed in correlation to their position utilizing Google Maps. The data was then compared to the previous projects results. The overall WQI for all of the tested waterways combined remained in a constant pattern. All of the waterways, except for Sawyers Creek, were at their lowest WQI when being compared to past research ranging from 2011 to 2015. The Lower Pasquotank still has the best water quality index, but the waterway had a significant decrease in its water quality score. When compared to the results of the 2014 team and the 2015 team, the Lower Pasquotank water quality went from a good Water Quality Index score to a medium Water Quality Index score, dropping 21 points. Newbegun Creek, Areneuse Creek, Mill Dam Creek, Sawyers Creek, Knobbs Creek, and the Pasquotank River all had bad Water Quality Index scores this year."

Furthermore, we had a modeling day event; where had to manipulate a code that was provided. We were in a couple groups; my group members were Derek Morris Jr, Ke'darius Whitley, Reggie Kelley, George Kurian, and Ebrahim Alareqi. We have chosen the fall of the skydiver Felix Baumgartner and his famous "stratos" jump in 2012. They would have to write a program that would plot Felix's position and velocity in freefall and after he opens his parachute. We used a program similar to the code of throwing a ball into the air to calculate the velocity and position of Felix. Once we finished the program, it would not run because it was saying the float numbers were too large. The problem was we continued calculating the drag force once the free faller had landed, and that was not necessary. When we stopped calculating the drag force once Felix had landed, the program ran with no errors. We put our heads together and fixed the problem. Afterwares we created a powerpoint for we can present our studies for that day. It turned out being exquisite. I would recommend doing it again.