Hopper Queue Prediction Tools for Efficient Use of NERSC Systems
Rachel Harred
Dept. Computer Science, University of North Carolina at Greensboro, Greensboro, NC 27412

**Project Description**
- **Examine** the showstart command, discover which time is most accurate, why it is often inaccurate and determine a better way to predict wait time
- **Create a web tool** for users to help plan jobs based on node size, wall time and overall CPU usage to give optimal queue wait time

**Background**
- NERSC is one of the largest centers for supercomputing in the world
- Supercomputers are used to run large calculations for scientific purposes
- Hopper is a Cray supercomputer at NERSC with 6,384 compute nodes, 24 cores in each node
- Hopper is used for running parallel scientific programs, or jobs, for research projects funded by the U.S. Department of Energy

- Hopper has 10 queue classes to handle the large volume of jobs submitted each day
- Node size is the number of nodes that a job will utilize when it runs
- Wall time is the amount of time a user sets for their job to run
- The Hopper queue is accessed by Linux command line environment
- Queue commands include the “qs” command, which displays attributes of each job in the queue, and the “showstart” command, which displays predicted wait times for a given job

**Methods**
- Wrote a script that created and submitted jobs of various node sizes and wall times to Hopper, collected qs and showstart data for each job while in the queue
- Stored this information in a database
- Graphs were created to display the predicted wait time from showstart, the queue position and the actual job start time
- Wrote a script that merged two databases to show accurate completed job statistics
- Web tool to allow users to select job parameters and reduce queue wait time

**Results**
- Showstart prediction is poor for jobs using low and high numbers of nodes and better for jobs using a midsize range of nodes
- Best showstart time is "rsv"
- Overall, showstart predictions are inaccurate and further testing is needed to find a function that will better predict job start times
- Web tool shows difference between former completed jobs database and new database with more accurate queue wait time

**Future work**
- Run more jobs to compile more statistics on the showstart predictions and find a function for better prediction of queue wait times
- Over time the new database will accumulate enough data to replace the former database for the web tool

**Skills Learned**
- Mac/Terminal
- Unix/Linux
- BASH
- MYSQL
- Python
- JavaScript, HTML, PHP
- Torque/MOAB

**References**

**Acknowledgements**
Thank you to my mentor, Jack Deslippe, for his help and encouragement during the project. Special thanks to TRUST Director Aimee Tabor, Elizabeth Bautista and the staff at NERSC. This work was supported by the TRUST Center (NSF award number CCF-0424422).