A merit based priority scheme to optimize the use of shared computing infrastructure

Dr. Gowtham

Director of Research Computing
Michigan Technological University
(906) 487-3593 · g@mtu.edu · http://hpc.mtu.edu

2015/02/03

Michigan Tech
Create the Future

Goals and objectives

#1

Design, implement and manage an easy to use, responsive and stable shared computing infrastructure that the research computing community can feel at home with, take pride in using it well with minimum, if not none, system administrative tasks while establishing accountability, responsibility and transparency at every possible level for all involved parties.

#2

Design and implement a semi-automated workflow to measure the meaningful tangibles in an easily understandable way to reflect the return on investment, reward consistent and productive researchers, and attract the attention of potential faculty candidates, funding agencies and donors.

Michigan Tech
Create the Future

Journey through the ages

* Pre- June 2013

    * 8 mini to medium sized clusters ($\sim$1,000 cores)

    * Neither well used ($\sim$20% busy) nor shared with researchers in need

* June 2013 and beyond

    * Superior (research; 1400 cores)

    * Portage (HPC proving grounds and education; 100 cores)

    * Immersive Visualization Studio (research and education)

    * 90+% busy, comprehensive documentation and end user training

Michigan Tech
Create the Future

Greatest good for the greatest number

– Warren Perger and Gifford Pinchot

Much is said of the questions of this kind, about greatest good for the greatest number. But the greatest number too often is found to be one. It is never the greatest number in the common meaning of the term that makes the greatest noise and stir on questions mixed with money …

– John Muir

Michigan Tech
Create the Future

Cannot manage what cannot be measured

Not everything that is (easily) measurable is (really) meaningful

Not everything that is (really) meaningful is (easily) measurable

Implementation of driving philosophies

* Every PI interested in using Superior will submit a short proposal

    * Résumé

    * Title, abstract and preliminary results

    * Nature of data sets and required resources

    * User population, and source of funding

* Chair of HPC Committee reviews and assigns a tier

    * A: new faculty or established researchers with funding

    * B: established researchers with no (immediate) funding

http://superior.research.mtu.edu/account/
Unequivocal support from the executive team has helped not make any exceptions to anyone under any circumstance.

Michigan Tech
Create the Future

Implementation of driving philosophies

* Software as requestable and consumable resources

    * Licensed as well as free and open source suites

    * One mandatory license per job

* User accounts with uniquely identifiable

    * username (same as ISO; must exist in Michigan Tech banner system)

    * primary group (e.g., `jane-users`)

    * department (e.g., `ME-EM` or `Chemistry`)

    * college (e.g., `COE` or `CSA`)

Rocks Cluster Distribution with Grid Engine queuing system is used to build HPC clusters.
User information (ISO username, primary group, department and college affiliation) are stored in a MySQL database.
Grid Engine log contains username, primary group and mandatory resources for every job along with other information.

Michigan Tech
Create the Future

Implementation of driving philosophies

* Easily measurable quantities

    * User information (new/established faculty, post-docs, students, etc.)

    * # of CPUs, total CPU time and software suite used

* Really meaningful entities

    * Publications and their citations

    * Graduated students (and the degree earned)

    * Successful proposals, preferably from external sources

http://superior.research.mtu.edu/projects/
http://superior.research.mtu.edu/publications/
Researchers are expected to periodically report the really meaningful entities that result from the use of Superior to the chair of HPC committee, and are stored in a MySQL database.

Michigan Tech
Create the Future

Implementation of driving philosophies

* Human engineering

  * New user training sessions

  * Tips, conferences/workshops, webinars and tutorials

  * Scientific Computing courses (UN5390 and UN5395)

* Keeping track of violations

  * Running programs in login nodes

  * Exceeding allocated quota for disk usage

  * Other behavior deemed not in compliance with the expected etiquette

http://superior.research.mtu.edu/tips/ | http://superior.research.mtu.edu/courses/
http://superior.research.mtu.edu/webinars/ | http://superior.research.mtu.edu/conferences/
A set of scripts perform the self-policing tasks, and informs the respective PI automatically.

Michigan Tech
Create the Future

Implementation of driving philosophies

✳ Transparency via value added usage report

    ✳ PI gets it every week

    ✳ Executive team gets it every quarter, end of year and on demand

---

$0.10 per CPU core per hour

Researchers are not currently charged any fee to use the shared resource. The amount in the report (along with # of jobs and CPU time for each user in every research group) is to be interpreted as computing cost if Superior wasn't available, and may be used in budgeting externally funded proposals.

Implementation of driving philosophies

$$\text{Job priority} = g\ (\text{Raw CPU time, Production}) - u\ (\text{Violations})$$

✳ Raw CPU time (35%)

    ✳ Extracted from Grid Engine log and retained in time units

✳ Production (65%)

    ✳ Based on funded proposals, publications and their citations

    ✳ Extracted from MySQL database and converted to time units

    ✳ Conversion factor depends on the type of publication

    ✳ Every citation counts as 0.10 publication, and every \$ as 10 CPU hours

Raw CPU time and production are at the research group level.
Violations, extracted from a MySQL database (total count is a number), are at the individual user level.

Michigan Tech
Create the Future

Implementation of driving philosophies

* Job priority

  * Built-in feature of Grid Engine

  * An integer between -1023 and 1024

  * Higher the number, higher is the priority

  * Requires admin privileges for 0 through 1024

  * Users can control from -1023 to -1

  * Once assigned, users can only reduce it

Michigan Tech
Create the Future

Implementation of driving philosophies
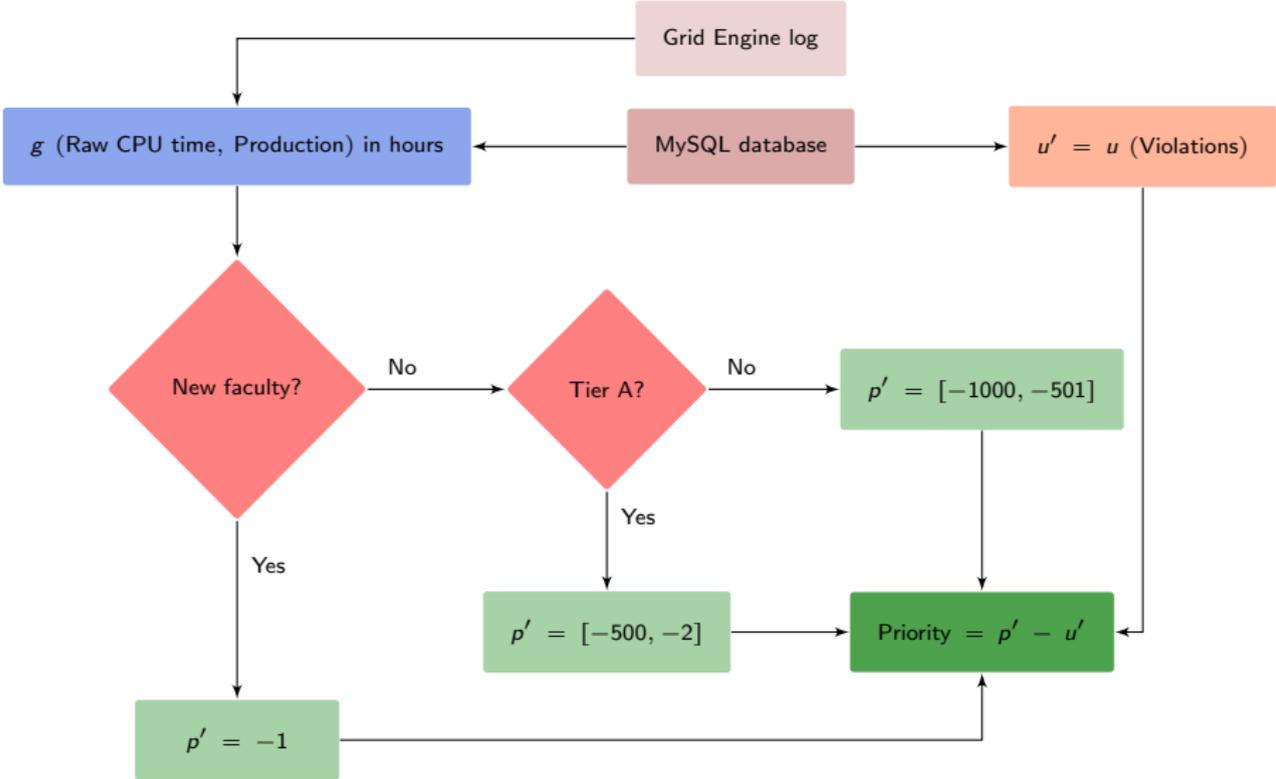
Implementation of driving philosophies

* Observable results

  * 40 projects ($\sim$50% each of tier A and B)

  * 30 publications ($\sim$20 additional manuscripts under review)

  * 90+% busy on most days

  * $1.2M worth of usage ($750k initial and $875k total investment)

  * Increased sense of ownership, accountability and responsibility

  * One mostly happy research computing community

Michigan Tech
Create the Future

Near future work

Methods discussed and associated code are under review as two potential publications: *Metrics4HPC: A tool set for analysis and visual representation of HPC cluster usage information* and *Metrics4Merit: A merit based priority scheme to optimize the use of shared computing infrastructure*.

* Not all publications are created equal
    * Impact factor can be integrated into computing job priority
    * XML file with annual impact factor of all journals

* Automated citation collection
    * Google Scholar is somewhat helpful
    * API that generates an XML file with all citations for a given DOI

Michigan Tech
Create the Future

Thanks be to

❊ Philip Papadopoulos, Luca Clementi and Rick Wagner (SDSC)

❊ Thomas Reuti Reuter (Phillips Universität Marburg)

❊ Rocks and Grid Engine mailing lists

❊ Rich Brueckner (insideHPC) and Amber Harmon (iSGTW)

❊ Friends and collaborators in academia, industry and media