Use of the Duke Compute Cluster in 2015: data, visualization, and lessons

Susan Starnes¹, Francisco Mejia²,³, Tom Milledge¹, Mark DeLong¹

¹ – Duke University Office of Information Technology, ² – North Carolina Central University

"It all comes out in the wash"

The cluster operates on the "condo" model, allowing researchers to use other researchers' machines at low priority (yielding to the buyer's use) as a "common partition" resource. The arrangement provides researchers intermittent bursts of compute power and dedicated cycles from purchased equipment (their "high-priority" use). This generally increases the overall average use of the computational power of the combined computers.

A. "Stacked" usage by group shows variability of usage over 2015.
B. Individual's use over 2015. Patterns often show intermittent use, which allows for an averaging of overall use.

Data & automation

Data for groups and users was captured using SLURM sreport. After we finalized visualization using Tableau (http://www.tableau.com/), we automated data capture with Pentaho (http://www.pentaho.com/) which updates data on a monthly basis. Tableau automatically renders new report graphs.

Finding balance

The framework of the cluster evens out the intermittent usage of many users, but requires a long-term balance of contributed and consumed cycles. That balance can be improved using data from the project by identifying groups with unbalanced usage.

C. Comparison of use v. contributions shows the balance of contribution and consumption. Red indicates higher consumption; green, higher contribution.

Future directions

• Deploy dashboard for groups to see compute cycle use
• Improve balance of use to contributions
• Deploy Open Science Grid to harvest unused cycles
• Add metrics for "cloud bursting" (AWS & Azure)
• Create similar system for storage

Every month, Duke researchers use a combined 290 CPU-years of cluster computation.