Bosco – A simple interface for managing jobs on both XSEDE and campus computing resources

Derek Weitzel
University of Nebraska – Lincoln
dweitzel@cse.unl.edu

Dan Fraser
Argonne National Laboratory
fraser@anl.gov

Miha Ahronovitz
University of Chicago
mihaa@uchicago.edu

Mats Rynge
USC / Information Sciences Institute
rynge@isi.edu

Submit Locally. Compute Globally.

Bosco makes it easy for the great majority of dedicated scientists - who don't have time to think about the computing side of it - to access High Throughput Computing and High Throughput Parallel Computing resources on your campus, XSEDE, or the cloud from their desktops.

Can your workload be classified as HTC or HTPC?

Bosco mainly supports two types of high throughput workloads:

- **HTC (High Throughput Computing)** workloads are made up of single threaded tasks.
- **HTPC (High Throughput Parallel Computing)** workloads are made with codes designed to utilize many cores on a *single* compute node. Internally the codes can use any single node, parallelization system (e.g. MPI, OpenMP, threads, ...)

Easier High Throughput Computing
Bosco is designed to manage high throughput computing jobs where hundreds to thousands of jobs run simultaneously on remote clusters. The researcher submits jobs from their desktop, and from there Bosco manages all the remote aspects of running the jobs and collecting the data.

Submit Locally
In Bosco, there is no need to remotely log into a cluster to manage job submissions and data. Bosco handles this automatically.

All you need is a user account and a password
Bosco does the rest. For example it automatically identifies the maximum number of jobs that can be submitted to each cluster, and throttles the jobs accordingly.

Bosco and XSEDE
We are here to be part of future solutions that will help XSEDE be used by non-power user scientists and students. We can contribute by lowering the ability threshold, increasing user desire, and creating triggers to actively engage new users. Ask us how! Send an email to:

bosco-discuss@opensciencegrid.org

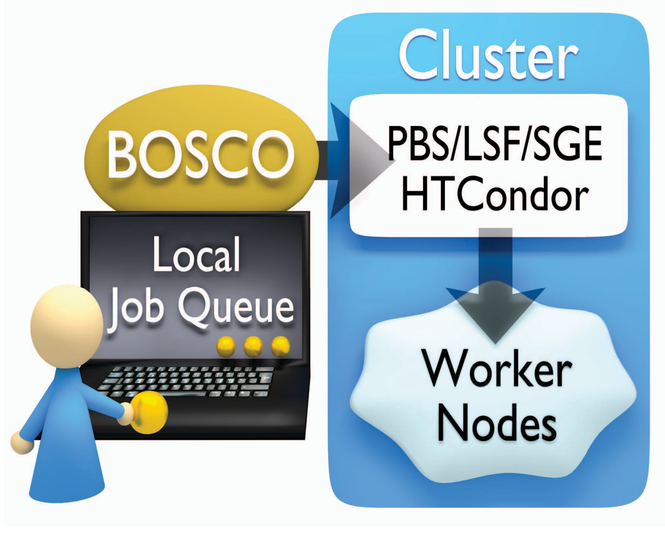
What is Bosco?
Bosco is a tool that allows researchers to quickly connect from their desktops to multiple clusters where they have SSH access. If offers a superior user experience (UX).

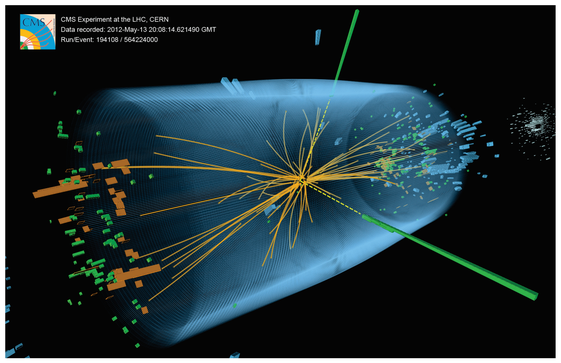
Bosco Empowers the User
All a user needs is an account on the cluster -- no additional configuration is needed from a system administrator.

Submission through SSH
No need for gatekeeper software, Bosco submits jobs over SSH to the remote clusters.

Multiple Remote Clusters
Submission to multiple remote clusters is automatic. Processing will be balanced between the clusters to maximize throughput. The remote clusters can run PBS, LSF, SGE, or HTCondor, but will all appear the same to the researcher.

Secure Execution
Jobs and data are transferred using strongly secured connections between the worker nodes and the submit node.





Using Bosco, the Compact Muon Solenoid was able to process 125TB of collision data. Bosco enabled the seamless integration of the XSEDE resource into the CMS framework.

Bosco and glideinWMS provided the crucial "glue" to make job submissions to Gordon possible to process 125TB of CMS data. This accelerated availability of this data for doing science from 6 months to 1 month.
–Frank Wuerthwein, CMS

BOSCO R

Project R is an open source statistical modeling package with over two million users. BoscoR is a software solution utilizing Bosco and GridR to enable remote processing of R programming language functions. Utilizing BoscoR, you can submit remote processing from within your R environment.

Acknowledgement
This research was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

http://bosco.opensciencegrid.org/