

# Tools to manage batch queue for science processing

Dear BESSIG folks:

I'm looking for tools to organize the data processing for an atmospheric research instrument: perhaps a "batch queue manager", "resource manager" or "scientific workflow system."

One recommendation has been:

- the Maui scheduler, which runs on the
- Torque resource manager, a fork of the
- Portable Batch Manager (PBM)

There was also a recommendation for "gridengine".

Do you have thoughts or personal experience with this or any competing systems?

We have four Linux boxes--dual quad-core Xeon systems running Scientific Linux. The team is small: a few developers and scientists.

We use cron and Perl scripts to launch the jobs today. This works for us, but the CPU utilization is not as high as I would like and the system is hard to understand and manage. The system is big enough that it has grown complex, but this is not what I would call high-performance computing.

There ought to be a better way. This is a well-studied problem and I expect there are some standard solutions.

Desirable:

- cheap/free
- relatively simple to install and experiment with; we cannot afford to make this a large project
- scheduling based on priority and resource availability:
- CPUs
- memory
- ability to monitor and manage the queue
- distribute jobs to multiple compute servers
- command-line interface for ease of integration
- modular/light-weight enough to be adapted to our existing structure

I'd be grateful for any comments or suggestions.

Thanks,  
Vince

-----Vince Dean

SIPS Operations Manager - HIRDLS Project

Data Manager - MOPITT Project

National Center for Atmospheric Research

3090 Center Green Drive., Boulder, CO 80307

Email: [vdean@ucar.edu](mailto:vdean@ucar.edu)

Phone: +1 303-497-8077

URL: <http://acd.ucar.edu/~vdean/>