

REPORT ON OpenDreamKit DELIVERABLE D5.3

Sun Grid Engine support for Project JUPYTER Hub

STUART MUMFORD, NEIL LAWRENCE & MIKE CROUCHER



Due on	2016-08-31 (month 12)
Delivered on	2016-08-22
Lead	University of Sheffield (USFD)
Progress on and finalization of this deliverable has been tracked publicly on: https://github.com/OpenDreamKit/OpenDreamKit/issues/116	

DELIVERABLE DESCRIPTION, AS TAKEN FROM GITHUB ISSUE'S #116 ON 2016-09-02

- **WP5:** High Performance Mathematical Computing
- **Lead Institution:** University of Sheffield
- **Due:** 2016-08-31 (month 12)
- **Nature:** Other
- **Task:** T5.8 #106
- **Proposal:** p.51
- **Final report**

It is common for academic High Performance Computing (HPC) clusters to make use of schedulers based on Sun Grid Engine with Son of Grid Engine as one of the most popular. It is used, for example, on the institutional HPC systems in the Universities of Sheffield and Manchester in the UK. It is also used on the regional N8 HPC facility, a system shared by the 8 most research intensive universities in the North of England.

In this deliverable, we have developed and demonstrated a Sun Grid Engine notebook spawner for Project Jupyter, allowing users to easily access Jupyter notebooks on HPC clusters directly from the web-browser. This development allows users with no background in High Performance Computing to easily migrate workflows from laptop to HPC cluster, allowing them to access greater resources with no additional training required.

CONTENTS

Deliverable description, as taken from Github issue's #116 on 2016-09-02	1
1. Demonstration install	2
2. Technology Overview	2
3. Future Work	2

1. DEMONSTRATION INSTALL

The software developed as a result of this task has been installed on The University of Sheffield's High Performance Computing system and is available to all staff and students there.

User-facing documentation is available at <http://docs.iceberg.shef.ac.uk/en/latest/using-iceberg/accessing/jupyterhub.html>

2. TECHNOLOGY OVERVIEW

We make use of JupyterHub, a multi-user proxy server for the Jupyter notebook developed by the Jupyter team. The JupyterHub is run as a front end to the HPC cluster, allowing users to login and run Jupyter Notebooks and terminals through their web browsers. For a description of the architecture of the JupyterHub see: <https://jupyterhub.readthedocs.io/en/latest/howitworks.html>

We have made the following contributions in this area:

- The Sun Grid Engine (SGE) interface to JupyterHub. Released on Github at <https://github.com/cadair/sgespawner>, this extends Jupyter Hub such that it can spawn notebook servers on HPC worker nodes. It does this by interacting with the SGE queuing system.
- A Jupyter plugin which enables the automatic detection of conda or virtualenv environments as kernels. https://github.com/Cadair/jupyter_environment_kernels. This allows users of the HPC cluster to install specialised environments in their home directory and subsequently make use of them in the Jupyter notebook on the system.
- An install of the above software on University of Sheffield's HPC cluster that is now in production use by staff and students there for teaching and research. User-facing documentation is at <http://docs.iceberg.shef.ac.uk/en/latest/using-iceberg/accessing/jupyterhub.html>

3. FUTURE WORK

- Assist other UK sites that make use of Son of Grid Engine to install and document this method of access. Initial sites identified include University of Manchester and the regional N8 HPC cluster in the UK.
- Make use of the Spawner options form implemented in JupyterHub 0.4 to allow users to specify resources on the system such as memory and number of CPU cores.
- The current implementation is based on version 0.3 of JupyterHub. An update is necessary to bring it into line with the latest version.
- Extend the work so that it can be used with other schedulers.