The HPC Orchestrator v1.0 enables users to manage jobs for their accounts on HPC (high-powered computer) systems such as Raijan.

[Pre-requirements for using the HPC Orchestrator:](#_o2u4c74nrvjj)

[Getting started with the HPC Orchestrator v1.0:](#_97t4j12xvjat)

[Working with the HPC Orchestrator v1.0:](#_uqzb58iayt3f)

[Submitting a job with the HPC Orchestrator:](#_17l1na5e5o9t)

[Querying your job/s status with the HPC Orchestrator:](#_7mrnuxdva18e)

[Moving jobs from the HPC Orchestrator “Completed” queue or deleting jobs from the “Error” queue:](#_nyyowm1woev)

|  |
| --- |
| Pre-requirements for using the HPC Orchestrator: 0.1 User HPC project remote access account setup.  0.1.1 User has accounts set up on a Remote HPC (e.g. Raijin) and local host.  0.1.2 Set up a password-less access between the HPC and local host. You can ask the HPC provider if you’re unsure about this.  0.2 Assure your computer can use Python - some of the HPC Orchestrator code is Python & is required.  0.2.1 Check if you already have Python installed: Instructions can be found here <https://wiki.python.org/moin/BeginnersGuide/Download>  0.2.2 If Python is not already installed, instructions for installing it can be found at <https://www.python.org/downloads/>  0.3 Install the HPC Orchestrator from github:  0.3.1 Download <https://github.com/IntersectAustralia/HPC-Orchestrator.git>  0.3.2 Follow the instruction at hpc\_orchestratorinstall.md to install the code.  0.4 Check you have (or create) an ssh key.  0.4.1 To check if you have a public ssh key, open a terminal and enter:  $ ls ~/.ssh/\*.pub  (If you have no .ssh public key see <https://help.github.com/articles/generating-a-new-ssh-key-and-adding-it-to-the-ssh-agent/>)  0.4.2 Copy your ssh public key  $ pbcopy < ~/.ssh/id\_rsa.pub (# Copies the contents of the id\_rsa.pub file to your clipboard. Note: the name of your .pub file may vary, take a note of the name & location of this .pub file). Open your ~/.ssh/authorized\_keys file on the remote machine and paste the public key at the end.  0.5 Assure that the Application required to run your job exists on the HPC that you are using Getting started with the HPC Orchestrator v1.0: 1. Setup the global environment configuration variables.  1.1 Open Terminal and cd to the installed HPC Orchestrator (from at step 0.3 above)  1.2 vi into the Orchestrator’s config file and edit your details into the file  $ vi .hpc-orchestrator.config  Recommended config settings:  Note: KEY: xxxx = update with your specific details (explanation of field)  cat ~/.hpc-orchestrator.config  export user=xxxx #\_Input=User's name at HPC host like:=g66  export rhost=raijin.nci.org.au #\_Input= HPC host like:=raijin.nci.org.au. Relevant HPC to be used here  export luser=xxxxx #\_Input=Name of the User on the local machine like:=sam-admin  export lhost=xxxx #\_Input=ip address of the local machine like:122.29.212.244  export RHPC\_PROJECT=xxxx #\_Input=Name of the User's project at the HCP like:=sd999  export RHPC\_LOCATION\_HOME=/\*\*\*\*/raijin\_space #\_Input=Location on HPC like: =/short/g66/sd999/raijin\_space  export JOB\_LIMIT=5 #\_Input=User's hpc target for number of parallel Jobs - Integers only  export ORC\_HOME=/home/xxxx/hpc-orchestrator #\_Input=Location for Orchestrator-created Job directory on local machine  export ORC\_BIN=/home/xxxx/hpc-orchestrator/bin #\_Input=Location for Orchestrator-created application bin directory on local machine like:=/home/sam/hpc-orchestrator/bin  export ORC\_DATA\_HOME=/home/xxxx/hpc-orchestrator/my\_project\_name #\_Input=Location for Orchestrator-created application storage directory on local machine like: =/home/sam/hpc-orchestrator/myjob-workspace  export ORC\_HOME\_MIDDLEWARE=/home/xxxx/hpc-orchestrator/middleware #\_Input=Location for Orchestrator-created application middleware directory on local machine like:=/home/sam/hpc-orchestrator/middleware  export PBS\_MAIL=yes #\_Input=Should the hpc email me? <yes/no> like:=yes  Example config (Note: example below sets up for Raijin HPC. Relevant HPC to be used for each user)  export user=N42  export rhost=raijin.nci.org.au  export luser=sam-admin  export lhost=122.29.212.244  export RHPC\_PROJECT=sd999  export RHPC\_LOCATION\_HOME=/short/N42/sd999/raijin\_space  export JOB\_LIMIT=5  export ORC\_HOME=/home/sam-admin/hpc-orchestrator  export ORC\_BIN=/home/sam-admin/hpc-orchestrator/bin  export ORC\_DATA\_HOME=/home/sam-admin/hpc-orchestrator/sam-genome-workspace  export ORC\_HOME\_MIDDLEWARE=/home/sam-admin/hpc-orchestrator/middleware  export PBS\_MAIL=yes  Once edited to your required values, save & close the file  <Shift>ZZ  Note: This file can be edited and saved at any time if the User wishes to use a different HPC host or account.  2. Open the ~/.bashrc file and add the following source global definitions  $ vi ~/.bashrc # Source global definitions  Then edit the file with the following:  $ if [ -f /etc/bashrc ]; then  . /etc/bashrc  fi  if [ -f ~/.hpc-orchestrator.config ]; then  source ~/.hpc-orchestrator.config  fi  export PATH=$PATH:~/hpc-orchestrator/bin/  # Once edited to your required values, save & close the file  <Shift>ZZ  3. Set up the cron job/job scheduler (This sets the period for querying the job status where “\*/2\* \* \* \*” = run the cron job for status query every two minutes. @ minutes is our recommendation for this setting, but you may change it to another setting if you require.):  Cronjob configuration example:  $ crontab -e # This runs like a vi command  $ \*/2 \* \* \* \* sh /home/ec2-user/hpc-orchestrator/bin/move\_output.sh >>/home/ec2-user/hpc-orchestrator/move.log 2>&1  \*/2 \* \* \* \* /home/ec2-user/hpc-orchestrator/middleware/auto\_submit.sh >>/home/ec2-user/hpc-orchestrator/autosubmit.log 2>&1 Working with the HPC Orchestrator v1.0:Submitting a job with the HPC Orchestrator: 4. To run a job, input “orcsub --input” and input your source file/s location/s and the script. The command will:  4.1 Queue your job on the HPC Orchestrator until your allocation is available at the HPC.  4.2 Send your job to the HPC queue when there is space on your allocation.  4.3 Run your job at the HPC when there’s available space  4.4 Return your files to the HPC Orchestrqator when the Job is complete  *NOTES:*   * *Each HPC Orchestrator job is named using the following protocol: <local file directory as per configfile>/hpc-orchestrator/<workspace>/<date/time of “orcsub” command like YYYY-MM-HHTMM/SS.SSSSSS>\_<name of first file in orcsub command>. Because of this very specific time labelling, every job will have a unique name.* * *If the user uses the “orcsub” command using a file that is not at the specified location, the command will answer with a response like: “<filelocation/name> is not a file, quit”* * *In special cases, Intersect will supply an alternative “orcsub2” command for processing with 4 arguments instead of the current 7 arguments in “orcsub”*   *Example job submission:*  orcsub --input /home/sam-admin/RDSI/NSW-RDSI22-TCGA/NSW-RDSI22-TCGA/s01024-tcga/TCGA/cghub\_data/tcga\_pipeline\_output/TCGA-A6-2671-01A-01D-1405-02/bb0c5c11-285a-4f0d-a9fc-dcd82900871e/TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam /home/sam-admin/RDSI/NSW-RDSI22-TCGA/NSW-RDSI22-TCGA/s01024-tcga/TCGA/cghub\_data/tcga\_pipeline\_output/TCGA-A6-2671-10A-01D-1405-02/b8b04320-5de6-429d-828c-53dde9705b97/TCGA-A6-2671-10A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam /home/sam-admin/eu9-files/Region\_indel\_analysis\_3\_noM.BED /home/sam-admin/eu9-files/ucsc.hg19.fasta /home/sam-admin/eu9-files/ucsc.hg19.fasta.fai /home/sam-admin/TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam.bai /home/sam-admin/TCGA-A6-2671-10A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam.bai Querying your job/s status with the HPC Orchestrator: 5. Run the status query command:  $ orcstat  This will output a series of Statii:  The following explains each status:  **+Query remote job status:** Lists your job/s running or queued at HPC like:    *Note: Key to terms used:*  *“r-man2” may be replaced with a different name defining the job queue for your job at the HPC*  ***NDS****= # of Nodes required*  ***TSK****= # of Tasks*  ***S****= Status (Outputs are: Queued, Held or Exited/Completed = Q/H/E)*  ***+Query generated:*** *= Job is stored locally until HPC is available to upload job.*  ***+Query submitted jobs:*** *Job/s stored locally & waiting for User's available allocation at HPC.*  *&/or copying to HPC and awaiting allocation there.*  ***+Query completed jobs:*** *Lists job/s completed at HPC and available in local (specified) directories.*  *+****Query failed jobs:*** *List job/s completed at HPC with errors and available in local (specified) directory.*  *Sample* orcstat output:  +Query remote job status:  r-man2:  Req'd Req'd Elap  Job ID Username Queue Jobname SessID NDS TSK Memory Time S Time  --------------- -------- -------- ---------- ------ --- --- ------ ----- - -----  4602125.r-man2 wxf569 copyq eu9.job-3. -- 1 1 4096mb 01:00 H --  4602126.r-man2 wxf569 copyq eu9.job-4. -- 1 1 4096mb 01:00 H --  +Query generated jobs:  /home/sam-admin/hpc-orchestrator/myjob-workspace/error/2017-05-15T10:45:13.600038\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  +Query submitted jobs:  /home/sam-admin/hpc-orchestrator/myjob-workspace/error/2017-05-09T10:50:13.600038\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  +Query completed jobs:  /home/sam-admin/hpc-orchestrator/myjob-workspace/completed/2017-05-09T11:51:43.661073\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  /home/sam-admin/hpc-orchestrator/myjob-workspace/completed/2017-05-09T11:52:30.835593\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  /home/wei-admin/hpc-orchestrator/myjob-workspace/completed/2017-05-09T11:53:01.265570\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  +Query failed jobs:  /home/sam-admin/hpc-orchestrator/myjob-workspace/error/2017-05-09T10:25:13.600038\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  /home/sam-admin/hpc-orchestrator/myjob-workspace/error/2017-05-09T10:38:05.667807\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam  /home/sam-admin/hpc-orchestrator/myjob-workspace/error/2017-05-09T15:10:18.075178\_TCGA-A6-2671-01A-01D-1405-02\_IlluminaHiSeq-DNASeq\_whole\_bw2\_uniq.bam Moving jobs from the HPC Orchestrator “Completed” queue or deleting jobs from the “Error” queue: 6.To move jobs from the “+Query completed jobs” or “+Query failed jobs” status queues:  6.1 To move jobs from the “+Query completed jobs” queue:  In the Terminal, from the hpc-orchestrator directory  $ cd <yourworkspacename>  $ mkdir oldcompleted #This makes a directory to contain the unwanted items from the queue. *Note: This step only needs to done for the first time you move completed jobs.*  $ cd completed  $ ls #This will show the directories for all jobs in the “completed” directory  # copy the name of the job you wish to move  $ mv <jobname> mv /oldcompleted/  # repeat for each job you wish to no longer remain in “completed” queue  6.2 To delete jobs from the “+Query failed jobs” queue:  In the Terminal, from the hpc-orchestrator directory  $ cd <yourworkspacename>/error  $ ls #This will show the directories for all jobs in the “error” directory  # delete the name of the job you wish to remove  $ rm -rf <jobname>  # repeat for each job you wish to no longer remain in “+Query failed job” queue |

## Frequently asked questions:

Q: How do I see how much memory a job has used?

A: Go to ???? jobfill name/##

Q: How do I “stop” a job on the HPC?

A: Login to your HPC account and … rdel.<jobID> What are the risks with doing this, what happens to the Orchestrator in these cases. Is there some sort of clean-up required afterwards?

Q: I want to use a new tool for use at the HPC, how do I install it?

A: Contact [help@intersect.org.au](mailto:help@intersect.org.au) for details on how to customize for new tools.