
HPCBIOS Documentation

Release 0.9.9rc5

Fotis Georgatos

21 June 2017

Contents

1	HPC Baseline Configuration	3
2	HPCBIOS_05-01: Multiple-Version Software Policy	7
3	HPCBIOS_05-05: Common Queue Names	9
4	HPCBIOS_05-06: Baseline Set of Login Shells	11
5	HPCBIOS_06-01: Common Set of Open Source Math Libraries	13
6	HPCBIOS_06-04: Baseline Editors and Scripting Tools	17
7	HPCBIOS_06-05: Baseline Set of Debuggers	19
8	HPCBIOS_06-15: Sample Code Repository	21
9	HPCBIOS_06-17: Use of Modules for Accessing Multiple Versions of Software	23
10	HPCBIOS_06-19: Common Set of Open Source Utilities	25
11	HPCBIOS_07-02: Common Open Source Performance and Profiling Tools	27
12	HPCBIOS_07-03: Common Set of Open Source Compilers	29
13	HPCBIOS_10-01: New/Returning User Welcome Letter	31
14	HPCBIOS_10-02: Common Open Source High Productivity Languages	33
15	HPCBIOS_2012-80: Common Set of DFT codes	35
16	HPCBIOS_2012-90: Software Tools and Development Environment	39
17	HPCBIOS_2012-91: Modules Namespace for HPC sites	43
18	HPCBIOS_2012-92: EasyBuild HPC Software Development Environment	47
19	HPCBIOS_2012-93: Life Sciences Productivity Environment	57
20	HPCBIOS_2012-94: Bioinformatics & Comp. Biology Productivity Environment	59
21	HPCBIOS_2012-95: Molecular Dynamics Productivity Environment	63
22	HPCBIOS_2012-96: Common Set of Commercial Compilers	65

23 HPCBIOS_2012-97: Climate Science Productivity Environment	67
24 HPCBIOS_2012-98: Common Set of Environment Variables	69
25 HPCBIOS_2012-99: GPU Productivity Environment	71
26 HPCBIOS_2013-01: Common Dependencies for Life Science Applications	73
27 HPCBIOS_2015-01: Biocompressors, for NGS data and bioinformatics	75
28 Indices and tables	77
29 Reference Pages	79

Contents:

HPC Baseline Configuration

Introduction

HPCBIOS is an effort to setup a common, well-documented and reproducible environment, spanning across multiple HPC systems & sites, in a user-friendly way.

HPC Baseline Configuration (BC) is an effort to define a common set of capabilities and functions, so that users can work more productively and collaboratively when using the HPC resources at multiple computing sites within the HPCBIOS compatibility context. This is accomplished by defining and developing a common set of capabilities and functions so that users can focus energy on their research effort, as opposed to wrestling with site-specific policies and procedures.

It is modeled after a similar effort taking place among 6 HPC sites in the US and it is, in effect, an extension upon it (notably, its Compliance Matrix):

HPC-BC.

This effort keeps marching on helping users work more productively, by updating existing policies to accommodate emerging needs, integrating new capabilities that must be consistent across the HPC enterprise or, just providing a sounding board for users experiencing difficulties with an HPC environment.







- The keywords *MUST*, *MUST NOT*, *REQUIRED*, *SHALL*, *SHALL NOT*, *SHOULD*, *SHOULD NOT*, *RECOMMENDED*, *MAY*, and *OPTIONAL* are to be interpreted as described in [RFC 2119](#).

About the Baseline Configuration

There is a growing community of HPCBIOS users who need to work on data and computation-intensive problems, spread across multiple computing sites and services. Recent discussions show that over half of users eventually need accounts at more than one site.

The main goal of the HPC Baseline Configuration activity is to define and establish a common set of capabilities and functions so that users can compute productively and collaboratively, when using HPC resources at multiple computing sites.

Compliance Matrix

Use notation       respectively standing for: *yes, no, agree, disagree, information notice, warning notice.*

Policy tag #	Policy Topics	C	G	K
[HPCBIOS_05-01]	Multiple-Version Software			
[HPCBIOS_05-05]	Queue Names			
[HPCBIOS_05-06]	Login Shells			
[HPCBIOS_06-01]	Open Source Math Libraries			
[HPCBIOS_06-04]	Editors and Scripting Tools			
[HPCBIOS_06-05]	Baseline Set of Debuggers			
[HPCBIOS_06-15]	Sample Code Repository			
[HPCBIOS_06-17]	Multiple-Version Software Access via Modules			
[HPCBIOS_06-19]	Open Source Utilities			
[HPCBIOS_07-02]	Open Source Performance and Profiling Tools			
[HPCBIOS_07-03]	Common Set of Open Source Compilers			
[HPCBIOS_10-01]	New/Returning User Welcome Letter			
[HPCBIOS_10-02]	Common Open Source High Productivity Languages			
[HPCBIOS_2012-80]	Common Set of DFT codes			
[HPCBIOS_2012-90]	Software Tools and Development Environment			
[HPCBIOS_2012-91]	Modules Namespace for HPC sites			
[HPCBIOS_2012-92]	EasyBuild HPC Software Development Environment			
[HPCBIOS_2012-93]	Life Sciences Productivity Environment			
[HPCBIOS_2012-94]	Bioinformatics and Computational Biology Productivity Environment			
[HPCBIOS_2012-95]	Molecular Dynamics Productivity Environment			
[HPCBIOS_2012-96]	Common Set of Commercial Compilers			
[HPCBIOS_2012-97]	Climate Science Productivity Environment			
[HPCBIOS_2012-98]	Common Set of Environment Variables			
[HPCBIOS_2012-99]	GPU Productivity Environment			
[HPCBIOS_2013-01]	Common Dependencies for Life Science Applications			
[HPCBIOS_2015-01]	Biocompressors, for NGS data and bioinformatics			

Note: Non-Compliance Disclaimer

Site compliance assures that the minimum guidelines of the policy have been met by a site. Sites may, at their discretion, exceed the minimum guidelines and still remain in compliance. Periodic checks may need to be performed to ensure compliance is maintained.

Note that, there can be valid reasons for not being compliant with a policy. For example, non-compliance may be acceptable especially if it implies removal of an existing capability or function, from a given system; also, for generic operational reasons or, if the HPC system affected is scheduled for decommissioning in the near future.

Non-compliance is not an issue, as long as the documentation above is consistent and the users are informed about status, in a prompt way.

This document was last rebuilt on: 21 June 2017.

HPCBIOS_05-01: Multiple-Version Software Policy

Note: A minimum set of versions of open source and third party software are recommended at all HPCBIOS compatible sites.

- BC Policy: HPCBIOS_05-01
 - Date of Policy: 2012-12-15
-

These versions are:

- A primary production version; (MUST)
- A previous production version, for validation and verification purposes; (SHOULD)
- Beta or Release-Candidate version to be made available within first month of that release. (MAY)

These three versions MAY vary by package and system, yet users SHOULD have the ability to test newer versions as they become available and still have access to production versions. No attempt/effort will be made to force vendors/upstream providers to provide three versions.

See also:

The primary concern of this policy is that, when a new production version of a package is being deployed at an HPC site it SHALL NOT disturb the existing userbase. It remains at site's discretion to decide which version is the current default yet, it SHOULD still allow at least one previous production version to serve gracefully users who have incompatibilities with newer software and, also allow for testing in advance.

HPCBIOS_05-05: Common Queue Names

Note: Common queue names will allow users to understand HPCBIOS compatible resources faster, select appropriate queues and, move workflows across sites without related issues.

- BC Policy: HPCBIOS_05-05
 - Date of Policy: 2012-12-15
-

The goal of this policy is to institute a core namespace of batch queues with the same names at each of the systems participating in a Baseline Configuration. The queue names were chosen to align with collective experience from multiple HPC sites. In addition, each system or site is explicitly allowed to offer customized batch support and service levels for specialized workloads or projects of each own. Systems MAY have more queues on as needed basis and SHOULD extend this policy in a backward-compatible way.

The core set of common queue names will allow users to more easily move their jobs and scripts among systems. This improves user productivity by requiring less time and effort to support many different scripts and to learn a different queue environment at each site.

The Compute queue names defined by HPCBIOS and their respective priority classes are:

Queue Name	Priority Class	Cost factor	Policy SLA	Description	Specifications to meet (ie. minimum maxima)
urgent	Urgent	1	MAY	Approved by Site	as needed, for fulfilling site or project objectives
interac- tive	Interac- tive	1	MAY	For interactive jobs	maxsize @ 4 nodes; max. walltime @ 2 hours
debug	Debug	0	SHOULD	Time/Resource limited for debug purposes	max. walltime @ 30 minutes
gpu	Standard	1	MAY	For approved projects needing GPUs	Normal queue for GPU resources
batch	Standard	1	MUST	For approved projects - GENERIC	maxsize @ 50% of resources ; max. walltime @ 24hours
transfer	Standard	1	MAY	Data stage-in/stage-out jobs	used for jobs heavy in I/O, meant for data transfers
gpu-low	Low Priority	0	MAY	For approved projects needing GPUs	No allocation subtraction, possibly best-effort
batch-low	Low Priority	0	MAY	For approved projects	No allocation subtraction, possibly best-effort

The above list of queues, scaling among *urgent* and *batch-low*, denotes different level of priorities, ranging from the highest to lowest.

A site still maintains the right to choose which users have access to which queues and at which times, so essentially compliance with this policy is quite flexible.

Potential sources of information for further future upgrades of this policy include:

- <http://centers.hpc.mil/consolidated/bc/policies.php?choice=queue>
- <http://centers.hpc.mil/consolidated/bc/policies.php?choice=common>
- http://centers.hpc.mil/consolidated/bc/commonCommands.php?choice=show_queues
- <http://www.nersc.gov/users/computational-systems/hopper/running-jobs/queues-and-policies/>
- <http://trac.mcs.anl.gov/projects/cobalt/wiki/DebugQueueConfig>

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_05-06: Baseline Set of Login Shells

Note: A user MAY specify any login shell among the Login Shells baseline set as the default shell, via a new user application.

- BC Policy: HPCBIOS_05-06
 - Date of Policy: 2012-12-15
-

Sites MAY provide their users with login shell choices beyond this list, but all sites MUST provide every login shell in the current baseline set.

Users MAY request a change to the current login shell for their active accounts at any site via email, telephone, or walk-in request.

This policy is not a guarantee that all shells will be working under all circumstances (fi. in a modulefiles context) or, that a particular version will be available.

Sites MAY force changes in available shell versions, due to security or other operational needs.

Sites MAY opt to use modules for providing certain shell/version combinations.

Sites MAY choose to restrict users from changing default login shells, esp. if these are not originally specified during allocation application. (ie. specific dependencies on shells MUST be declared and tested in advance, before a large allocation is made. Users are advised to test their codes under a preparatory scheme)

The Login Shells baseline set and user selection policy MAY vary; in that case it MUST be documented at each site's website or system (fi. MOTD).

Default Login Shell Baseline Set

- Bourne (sh)
- Korn (ksh)
- C shell (csh)
- Extended C-shell (tcsh)
- GNU Bourne-again shell (bash)
- Z shell (zsh)

HPCBIOS_06-01: Common Set of Open Source Math Libraries

Note: This policy defines a set of open source math libraries and their consistent maintenance (compatible version and configuration) across resources of HPCBIOS compatible sites. Having same configuration is dependent on system compatibility, and also, some libraries may have slightly different configurations due to architectural differences. These differences, however, **SHOULD NOT** affect the use of the libraries.

A user moving from one resource to another is not expected to notice significant difference between the installations and it is expected to find an optimized version of that library.

- BC Project: HPCBIOS_06-01
 - Date of Policy: 2012-12-15
-

The suite of maintained packages consists of:

Library API	Purpose	Compliance level	References
FFTW	Library for computing Discrete Fourier Transforms Both MPI and non-MPI versions should be provided.	MUST	http://www.fftw.org/
PETSc	Suite of scientific computing routines	MUST	http://www.mcs.anl.gov/petsc/
LAPACK	Software library of numerical linear algebra routines	MUST	http://www.netlib.org/lapack/
ScaLAPACK	Subset of parallelized LAPACK routines	MUST	http://www.netlib.org/scalapack/
BLAS optimized stack	Basic linear algebra routines; this could be any of: <ul style="list-style-type: none">• ATLAS• MKL• OpenBLAS or Goto-BLAS	MUST	http://www.netlib.org/blas/ http://math-atlas.sourceforge.net/ http://software.intel.com/en-us/intel-mkl http://www.tacc.utexas.edu/tacc-projects/#blas https://github.com/xianyi/OpenBLAS
GSL	C/C+ library with a wide variety of mathematical routines	MUST	http://www.gnu.org/s/gsl/
SPRNG	Pseudo-Random Number Generators for Stochastic computation via Monte Carlo methods et al	MUST	http://sprng.cs.fsu.edu/
ARPACK	Solution of Eigenvalues/Eigenvectors	MAY	http://www.caam.rice.edu/software/ARPACK/
SuperLU	Direct solution of sparse linear systems of equations	MAY	http://crd.lbl.gov/~xiaoye/SuperLU

Change notification

Guidelines:

- This policy MUST provide for the default/preferred implementation for LINPACK/BLAS functions
- Notification MUST be sent via the system's helpdesk, announcing the new version and describing how to access it and use it
- The helpdesk MUST be given a minimum of 2 weeks notice when the new version becomes the default version
- It is the responsibility of the individual sites to notify their users of any changes

See also:

Native Libraries

As regards packages ScaLAPACK, LAPACK and/or ATLAS, they MAY not be installed whenever a system includes native libraries such as PESSL (Parallel ESSL) ACML or, MKL (Math kernel Libraries) that fully provide the same functionality and interface. However, if an explicit request is made to have open source ScaLAPACK, LAPACK and/or ATLAS available on a system, then these packages MUST also be installed.

Support

A helpdesk team (User Support) maintains the suite of software listed in this policy, and it also provides the needed support channels.

HPCBIOS_06-04: Baseline Editors and Scripting Tools

Note: Each site **MUST** install and maintain the following baseline set of editors and scripting tools. New versions of these tools **SHOULD** be installed within 30 days of their availability on all compatible HPC systems and non-storage user-accessible support servers, and they **SHOULD** be installed in the users default path OR provided via modulefiles.

References for obtaining information and downloading these packages are included below. Sites **MAY** choose to provide certain versions via modules framework, at their discretion.

- BC Policy: HPCBIOS_06-04
 - Date of Policy: 2012-12-15
-

N.B. Baseline shells are covered by policy [HPCBIOS_05-06]

Baseline Editors

Emacs	http://www.gnu.org/software/emacs/emacs.html	MUST
Vi and Vim	http://ex-vi.sourceforge.net http://www.vim.org	MUST
Eclipse	http://www.eclipse.org	SHOULD
nano	http://www.nano-editor.org	SHOULD
Nedit	http://www.nedit.org	SHOULD
Pico	http://www.washington.edu/pine/	SHOULD
Joe	http://freshmeat.net/projects/joe	SHOULD
Jove	http://freshmeat.net/projects/jove	SHOULD

Baseline Scripting Tools

Perl	http://www.perl.org	MUST
Expect	http://expect.nist.gov	SHOULD
Java	http://www.java.com	SHOULD
Ruby	http://www.ruby-lang.org	SHOULD
PHP	http://www.php.org	SHOULD
Tcl	http://www.tcl.tk	MUST
Tk	http://www.tcl.tk	MUST

See also:

Python and its scientific add-on components are fully covered in Baseline Configuration policy [*HPCBIOS_10-02*] entitled “Common Open Source High Productivity Languages.”

HPCBIOS_06-05: Baseline Set of Debuggers

Note: This policy defines a baseline set of debuggers which can assist HPC users in handling parallel codes.

- BC Policy: HPCBIOS_06-05
 - Date of Policy: 2012-12-15
-

The lack of a baseline set of debuggers can compound what is arguably one of the most difficult tasks in scientific code development - debugging new algorithms. The process of tracking down bugs in large scientific codes is frustrating and time consuming. Having to learn a new tool to do so limits efficiency and productivity of valuable HPC programmers.

Somebody can argue that code should be written without bugs yet, until that day arrives, there is a dire need for using debuggers and building expertise upon them.

For these reasons a baseline set of debuggers, preferred to be installed at all systems, has been identified by examining what is in use at different sites and what can be used across a variety of platforms. At this time, these packages are included in the RECOMMENDED baseline set:

De-bug-ger Name	Description	Compli-ance Level	References
GDB	A freeware debugger provided by the GNU project, covered by the GNU public license Supported architectures: any	MUST	http://www.gnu.org/software/gdb/gdb.html
Valgrind	Memory management analysis and profiling	MUST	http://valgrind.org/
Allinea DDT	A popular suite for debugging in the HPC realm, provided by Allinea Supported architectures: x86_64, GPUs, ARM	SHOULD (commercial)	http://www.allinea.com/products/ddt/
Allinea MAP	An MPI profiler built on simplicity Supported architectures: x86_64	SHOULD (commercial)	http://www.allinea.com/products/map/
Totalview	A long standing de-facto standard for debugging in the HPC realm provided by TotalView Technologies Supported architectures: x86_64, GPUs	SHOULD (commercial)	http://www.roguewave.com/products/totalview.aspx
IDB	Intel Debugger provides support for debugging programs written in C, C++ & Fortran It provides a choice of command-line and graphical user interface (GUI) on Linux. Supported architectures: x86_64,	SHOULD (commercial)	https://software.intel.com/en-us/articles/idb-linux
MemoryScape	Dynamic memory debugging for monitoring heap allocations, memory leaks memory overruns, and memory usage; by Rogue Wave software Supported architectures: x86_64,	MAY (commercial)	http://www.roguewave.com/products/memoriescape.aspx
Eclipse PTP	Parallel Tools Platform supports MPI, OpenMP & UPC; as well as OpenSHMEM & OpenACC Supported architectures: any	SHOULD	http://www.eclipse.org/ptp/

In order to be compliant, a site is expected have at least the “MUST” class of these debuggers installed (ie. GDB & Valgrind) and in the standard login path across all nodes or, available via modules mechanism; the latter is very suitable for the latest version.

In addition, the GUI interface for TotalView/DDT, when available, MUST be operational and accessible from the compute nodes of the allocated system.

Potential sources of information for further future upgrades of this list include:

- <http://wiki.scinethpc.ca/wiki/images/c/ce/Best-practice-debug.pdf>
- <http://www.nersc.gov/users/software/debugging-and-profiling/>
- http://sebastien.dinot.free.fr/liens_dev_c_cpp.html
- http://user.cscs.ch/fileadmin/user_upload/customers/CSCS_Application_Data/Files/Presentations/Courses_Ws_2011/Multi-Threaded_Course_Feb11/debug.pdf
- <http://www.scalalife.eu/book/export/html/115>
- http://www.acrc.a-star.edu.sg/10/code_development.html
- A list of memory debuggers and more specialized tools such as: ElectricFence, dmalloc, Purify, **Intel Inspector**, IBM Rational Purify

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_06-15: Sample Code Repository

Note: Each compatible site shall maintain a user-accessible Sample Code Repository of frequently-used procedures, routines, scripts, and codes on each project affiliated system. Users of all skill levels can use the samples in the Repository to find best practices for various situations. The samples will be especially helpful to users who are new to HPCBIOS sites or to a particular system.

- BC Policy: HPCBIOS_06-15
 - Date of Policy: 2012-12-15
-

An environment variable named `$SAMPLES_HOME` will point to the Repository base directory. Each sample will be findable via its own subdirectory under the base directory. The exact architecture of the sample directories and the formatting style of the contents will be left to the discretion of the sites. However, each sample, including all scripts, codes, Makefiles, etc., shall be well-documented and shall be as self-contained as possible.

The top-level `$SAMPLES_HOME` directory shall contain an index file named `INDEX.txt`. This file includes the name and a brief explanation of each sample. The entries in the index file shall be organized by category. A description of the Code Repository and the `INDEX.txt` listing from each system shall be placed on the site's website or other suitable documentation area.

Samples in the Repository will be created primarily by each site's staff. However, a site may choose to allow users to submit prospective samples for consideration as new entries in the Repository, subject to approval by the site staff. The submitting user shall provide all necessary files and documentation, as well as future updates. Site staff shall be responsible for placing new samples and updates into the Repository.

In addition to the above-mentioned samples, each site may also have a collection of simple, stand-alone tools and scripts for the convenience of users of the systems. The tools and scripts, if present, shall be available in the users' default search path or, retrievable via modules.

The `$SAMPLES_HOME` directory shall contain a text file named `AVAILABLE_TOOLS.txt`. This file will contain a list of the tools that are available on the machine on which the file is stored. In addition to listing the tools, a short description of each will be provided. If there is no man page available for a particular tool, then a description on how to use that tool will be included in the listing file.

If a site has no such tools for a particular resource, the `AVAILABLE_TOOLS.txt` file must still exist, and it must specify that no such tools are available for that resource.

HPCBIOS_06-17: Use of Modules for Accessing Multiple Versions of Software

Note: Use of environment modules is the standard technology on HPC sites for handling multiple software versions, in a manner which offers orthogonality.

Any compliant system is required to install and use some ENVIRONMENT MODULES technology to support multiple versions of compilers, associated libraries, and heavily used application software (fi. OpenMPI, HDF5, NetCDF, WRF et al) on the available resources of HPC sites. Compliance is determined by sites ensuring that the *module* command is in the users' path and that modulefiles available by support staff are accessible in some way.

- BC Policy: HPCBIOS_06-17
 - Date of Policy: 2012-12-15
-

The ENVIRONMENT MODULES package provides for the dynamic modification of a user's environment via modulefiles. Each modulefile contains the information needed to configure the shell environment for an application. Once the Modules package is initialized, the environment can be modified on a per-module basis, using the module command which interprets modulefiles.

Typically, modulefiles instruct the module command to alter or set shell environment variables such as PATH, MANPATH, define aliases etc. Modulefiles may be shared by many users on a system and users may have their own collection to supplement or replace the shared modulefiles.

Scope

ENVIRONMENT MODULES can be loaded and unloaded dynamically and atomically in a clean fashion. Common shells such as bash, sh, tcsh/csh **MUST** be supported - while ksh, zsh (Ref: [[HPCBIOS_05-06](#)]), as well as some scripting languages such as Python or Perl (Ref: [[HPCBIOS_06-04](#)]) **MAY** be supported, as per each site's needs/capabilities.

Implementation

The function of the modulefiles is to modify a user's environment such that a user may reference the selected version of an application using the default commands for that application without hardcoding paths or version

information. This allows to maximize code reuse and pave the way for uniform documentation across software versions.

Implementations which are known to be compliant with the above needs are:

- *environment-modules-c* # version $\geq 3.2.10$
- *environment-modules-tcl* # version ≥ 1.562
- *Lmod* # version $\geq 5.9.0$

The individual site MAY combine compliant environment modules implementations and modulefiles with more existing site-specific methods (e.g. additional modulefiles or profile initialization or wrapper scripts) to select versions of applications to help define the user's environment (fi. storage locations, project variables etc).

Dependencies

In the cases where the modulefiles for applications depend on other modulefiles (an application may require compiler and library modulefiles to be loaded), these dependencies SHOULD be controlled, upon load time. A reasonably default modulefile MAY be loaded, if the dependency is not met already.

A site MAY automatically unload dependencies of a given modulefile, upon unload. Note that different user communities COULD well have different valid reasons for or against the default attitude upon unloading, therefor this is deliberately left open.

Escape path

An additional modulefile MAY exist, called OPT-OUT, which will remove the system's common-link directories from the associated *PATH* environment variables. Users will then need to load individual modulefiles to add the application specific directories to the *PATH* environment variables.

See also:

The currently suggested namespace format is the one provided via EasyBuild v2 (see [HPCBIOS_2012-92]), and a site MAY use aliases for default modules, as defined via UNITE, version May 2010. UNITE suite provides definitions with examples under section 1.2 of its documentation: <http://apps.fz-juelich.de/unite/files/unite-installguide.pdf> The above namespaces are essentially *flat*, meanwhile sites MAY prefer *hierarchical* namespaces.

HPCBIOS_06-19: Common Set of Open Source Utilities

Note: Sites will provide a core set of open source utilities on all allocated HPC systems and non-storage user-accessible support servers. These should be installed in the users' default path or, provided via modulefiles.

- BC Project HPCBIOS_06-19
 - Date of Policy: 2012-12-15
-

The current core set of open source utilities is:

Autoconf (GNU)	Produces shell scripts to automatically configure software source code packages.
Automake (GNU)	Generates makefile.in files from makefile.am input files. These ini files are compliant with the GNU coding standards and build process.
Bison (GNU)	Converts a grammar description into a C program to parse that grammar.
cvs (GNU)	Provides a version control system for managing source files, and documents.
Subversion	Provides a version control system for managing files and directories
Tar (GNU)	Creates tar archives, as well as providing various other kinds of manipulation.
gzip (GNU)	Is a data compression utility.
bzip2	Block-sorting data compression/uncompression utilities. Command line options are deliberately similar to those of gzip.

Other software utilities can be incorporated into this core set with Baseline Configuration team consensus.

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_07-02: Common Open Source Performance and Profiling Tools

Note: Performance and profiling tools are used on HPC sites for keeping codes execution times tamed.

This policy defines a set of performance and profiling tools and promotes their consistent maintenance (ie maintain compatible versions and configuration) across all systems of compatible sites. Same configuration is dependent on system compatibility, and also, some libraries may have slightly different configurations due to architectural differences. These differences, however, SHALL NOT affect the use of the libraries. A user moving from one resource to another is not expected to notice significant difference between the installations.

- BC Policy: HPCBIOS_07-02
 - Date of Policy: 2012-12-15
-

The suite of maintained packages consists of:

Tool	Description	Compliance level	Reference
gprof	The GNU Profiler	MUST	http://www.gnu.org/software/binutils/
PAPI	Consistent interface to hardware counters and events	MUST	http://icl.cs.utk.edu/papi
Valgrind	Memory management analysis and profiling	MUST	http://valgrind.org/
Intel Vtune Analyzer	performance optimization tool NOTE: not open source but still relevant	SHOULD (commercial)	http://software.intel.com/en-us/articles/intel-vtune-amplifier-xe/
UNITE	UNiform Integrated Tool Environment	SHOULD	http://apps.fz-juelich.de/unite

UNITE

The goal of UNITE is to provide a robust, portable, and integrated environment for the debugging and performance analysis of parallel MPI, OpenMP and hybrid MPI/OpenMP programs on high-performance compute clusters. It consists of a set of well-accepted portable, mostly open-source tools. UNITE itself includes:

mpiP	Lightweight, Scalable MPI Profiling	http://mpip.sourceforge.net/
SCALASCA	Scalable trace analysis package	http://www.scalasca.org/
Vampir	Performance Optimization	http://www.vampir.eu/
TAU	Performance analysis of parallel applications <ul style="list-style-type: none">• Support for PAPI and SCALASCA• PDT: Source-level auto-instrumentation	http://www.cs.uoregon.edu/Research/tau
PerfSuite	software performance analysis	http://perfsuite.ncsa.illinois.edu/
other	hpctoolkit periscope marmot itac Extrae Paraver kcachegrind	

The packages will be installed and maintained as part of a consistent computational environment.

Potential sources of information for further future upgrades of this list include:

- <http://apps.fz-juelich.de/unite>
- <http://www.vi-hps.org/tools/>

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_07-03: Common Set of Open Source Compilers

Note: As open source software becomes increasingly prevalent, the number of packages that are configured to compile under the GNU compiler suite grows each year. Although codes compiled under GNU suite often do not match the performance of codes generated by commercial compilers, the general availability and familiarity with the GNU suite is consistently drawing a larger HPC user base. Many packages that compile automatically under GNU may require large amounts of support time to compile under other compiler suites. Providing a minimum set of open source compilers is imperative.

- BC Policy: HPCBIOS_07-03
 - Date of Policy: 2012-12-15
-

Thus, to increase user productivity and expand commonality across HPCBIOS compatible sites, **a site MUST install the GNU compiler suite** and make it easily available, either through a user's default path or the availability of a modulefile. In addition, in order to fully take advantage of the HPC environment, *MPI support for the GNU compiler suite MUST be implemented* for compliance with this policy.

GCC version 4.6.x of the suite is strongly encouraged for maximum compatibility, hence it has become a requirement for being compliant with this policy:

Name	Preferred version(s)	Compliance level
GCC	v4.1.2	SHOULD
GCC	v4.3.5	SHOULD
GCC	v4.4.x	SHOULD
GCC	v4.6.2 or later	MUST
GCC	v4.7.2 or later	MUST
GCC	v4.8.2 or later	SHOULD
GCC	v4.9.0 or later	SHOULD
GCC	v5.0.0 or later	SHOULD
Clang	v3.2 or later	SHOULD
TCC	v0.9.26 or later	MAY

It is at the discretion of each site which (if any) libraries/software will be compiled to support GNU based codes. Ultimately, it will be the responsibility of each user of the GNU compilers to create the libraries/software that they require for their GNU compiled software.

Clang/LLVM is another open source compiler of general interest, not so necessarily for HPC purposes but rather for cross-checking compiled code and the remarkable ability to be extendible at will. Furthermore, Clang is very

extensive in testing aspects, revealing early on potential issues.

Finally, TCC may be used in scripting efforts, via the convenience of C code or, simply as experimentation environment.

HPCBIOS_10-01: New/Returning User Welcome Letter

Note: Each compatible HPC site SHOULD provide new and returning users with a standard minimum amount of information about their account and the services that are being made available to them. The minimum required information includes, username, (Kerberos principle if available,) subproject ID, default group, login host names for the systems where access is granted, the email address where information will be sent, links to local documentation, and contact information for help.

- BC Policy: HPCBIOS_10-01
 - Date of Policy: 2012-12-15
-

A sample minimum welcome letter is as follows

Sample Welcome Letter

Welcome to an HPCBIOS environment.

Below we outline the procedure required for obtaining access on HPCBIOS HPC resources.

Login Node Names for HPC Resources If you have been successful in your request, you (and your collaborators) are being granted access to the HPCBIOS HPC resource(s) as indicated in your notification email. The login nodes of the available systems are the following:

- System1 (Location1) : node1.example.org
- System2 (Location2) : node2.example.org
- System3 (Location3) : node3.example.org

Obtaining an Account

To gain access to one of the above HPC resources please do the following:

a. Print, complete, sign and fax the Acceptable Usage Policy (AUP) for the resource you have been granted access to (fax no: +XYZ-AB-CDEFGHIJ). The AUP is available in .pdf & .doc formats from the HPC website (URL). Insert "Project Reference code" as appropriate.

b. Each registered project member should request a user account on the allocated HPC resource(s) by completing the following form: (URL)

Please note:

- You will not be granted a user account on a HPC resource if you have not completed this form.
- If accepted project is allocated on a specific HPC site, you will be allowed access to your project allocated site(s)
- As an external, you may not be granted access unless you are registered as a collaborator in an accepted project. Additional collaborators can be registered by the project leader by contacting XYZ.
- If you are permitted access to more than one HPC resource you will have to complete the forms for each permitted site – each time choosing the appropriate (and different) HPC resource you are entitled to use.

Once the above has been done, you will be contacted by the HPC site to request your ssh public key. If you need instructions on how to generate a public key, please visit the following link: (URL)

User Resources

(URL) The User Support website has a section on User Resources, please familiarize yourself with its content. There you will find useful and extensive information - such as, for example, how to address the helpdesk and report issues, and information on available training materials.. Here are some initial pointers from this information that should prove useful:

User Support & helpdesk procedures (URL) [user-support/basics.html](#)

Quickstart - Getting Started on HPC Resources course: (URL) Quickstart

Training Material <http://www.citutor.org/> (please make an account)

For any questions, please do not hesitate to contact us via the helpdesk system: XYZ

HPCBIOS_10-02: Common Open Source High Productivity Languages

Note: High Productivity Languages have recently become an important part of scientific computing and visualization environment. Leading high productivity languages include R, the commercially available MATLAB, open source Python (with its scientific computing add-ons) and open source Octave.

Each of these high productivity languages forms an integral part of a standard library that includes text processing, file I/O, data compression, and a host of capabilities ranging from basic numerical linear algebra to complex data visualization. As such they are an indispensable component of any modern HPC platform.

- BC Policy: HPCBIOS_10-02
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common open source high productivity languages environment across all HPCBIOS compliant resources:

Pack- age/Tool	Description	Compliance level	Reference
Python	General Purpose Scripting Language	MUST	http://python.org/
NumPy	Numerical Arrays and Linear Algebra in Python	MUST	http://numpy.scipy.org/
SciPy	Scientific Python	MUST	http://www.scipy.org/
PyMPI	Python Message Passing Interface	SHOULD	http://pympi.sourceforge.net/
Octave	MATLAB Clone	MUST	http://www.gnu.org/software/octave/
matplotlib	Scientific 2-D and 3-D Plotting	MUST	http://matplotlib.sourceforge.net/
R	R language	MUST	http://www.r-project.org/
MATLAB	MATLAB is a high-level language for computation	SHOULD (commercial)	http://www.mathworks.nl/products/matlab/
NAG	Numerical library /Numerical Algorithms Group	SHOULD (commercial)	http://www.nag.co.uk/

This high productivity languages environment will be supplemented with more open source productivity languages, as they become available on allocated systems.

There are three key sources which MAY be used to provide support to the scripting language Python and its scientific add-ons NumPy, PyMPI and SciPy, as well as more tools from the above list. These are:

- EasyBuild. See policy [[HPCBIOS_2012-92](#)] ; in fact, this is the preferred way, since integration with the other scientific software is better guaranteed.
- Computational Science Environment (CSE). The CSE is an integrated production level environment developed at *ARL DSRC*. CSE consists of open source tools, libraries as well as Python support (Python, NumPy, SciPy and PyMPI). All CSE components are assembled in an extensible framework that handles software dependencies at compile and runtime.
- Parallel Tools Runtime Environment (PToolsRTE). The PToolsRTE is a collection of Python, NumPy, PyMPI, SciPy and related packages, including Matplotlib, developed by ParaTools, Inc. PToolsRTE provides a hybrid binary and source distribution where a majority of the components are pre-packaged in a binary form and a minimal number of platform and compiler specific components are built on the target system. This simplifies the complexity of software installation and the runtime environment can be easily installed on external systems to provide a consistent Python based environment for developing and deploying scientific packages that rely on a common runtime environment.

At least one of EasyBuild, CSE or PToolsRTE, SHOULD be made available on HPC systems.

EasyBuild is maintained by *Ghent University* and is available as open source python codebase, while the CSE package is being maintained by the *ARL CSE Team* and the PToolsRTE is being maintained through *PETTT* project. Sites are encouraged to used reproducible techniques for herding software components!

HPCBIOS_2012-80: Common Set of DFT codes

Note: Providing (a subset of) popular DFT codes is nearly imperative for most general purpose HPC sites. This is a work-in-progress effort to provide a candidate list, as a single set. There is no particular level of compliance as of yet (all list is at “SHOULD” status).

- BC Policy: HPCBIOS_2012-80
 - Date of Policy: 2012-12-15
-

Introduction

The following is a list of tables that document various density-functional theory based electronic structure codes, both periodic and molecular, and atomic pseudopotential codes. This list is no way comprehensive.

DFT production codes

Periodic codes - Plane wave and related methods

VASP	http://cms.mpi.univie.ac.at/vasp
CASTEP and CETEP	http://www.tcm.phy.cam.ac.uk/castep
CPMD	http://www.cpmd.org
ABINIT	http://www.abinit.org
BigDFT	http://inac.cea.fr/L_Sim/BigDFT
Quantum-Espresso	http://www.quantum-espresso.org
PEtot	http://www.nersc.gov/~linwang/PEtot/PEtot.html
DACAPO	https://wiki.fysik.dtu.dk/dacapo
Socorro	http://dft.sandia.gov/Socorro/mainpage.html
DFT++	http://dft.physics.cornell.edu
Octopus	http://www.tddft.org/programs/octopus
Paratec	http://www.nersc.gov/projects/paratec
DoD Planewave	http://cst-www.nrl.navy.mil/people/singh/planewave/v3.0
PARSEC	http://www.ices.utexas.edu/parsec
CP2K	http://cp2k.berlios.de
GPaw	https://wiki.fysik.dtu.dk/gpaw
SPHINX	http://www.sphinxlib.de
QBOX	http://eslab.ucdavis.edu/software/qbox

Periodic codes - Local orbital basis codes

Quest	http://dft.sandia.gov/Quest
SeqQuest	http://www.cs.sandia.gov/~paschul/Quest
SIESTA	http://www.icmab.es/siesta
CRYSTAL/CSE	http://www.cse.dl.ac.uk/Activity/CRYSTAL
AIMPRO	http://aimpro.ncl.ac.uk
FPLO	http://www.fplo.de
OpenMX	http://www.openmx-square.org

Periodic codes - All-electron (augmented methods) codes

ELK	http://elk.sourceforge.net
EXCITING	http://exciting-code.org
FLEUR	http://www.flapw.de
RSPT	http://www.rspt.net
WIEN2k	http://www.wien2k.at

Molecular codes

Gaussian	http://www.gaussian.com
NWChem	http://www.emsl.pnl.gov/docs/nwchem/nwchem.html
DMol3	http://people.web.psi.ch/delley/dmol3.html
Jaguar	http://www.schrodinger.com/Products/jaguar.html
GAMESS-US	http://www.msg.chem.iastate.edu/gamess
QCHEM	http://www.q-chem.com
NRLMOL	http://quantum.utep.edu/nrlmol/nrlmol.html
MondoSCF	http://www.t12.lanl.gov/home/mchalla
ADF	http://www.scm.com
deMon	http://www.demon-software.com
CADPAC	http://www-theor.ch.cam.ac.uk/software/caddoc.html
PyQuante	http://pyquante.sourceforge.net
TURBOMOLE	http://www.cosmologic.de/QuantumChemistry/main_qChemistry.html

DFT atomic pseudopotential codes

fhi98PP	http://www.fhi-berlin.mpg.de/th/fhi98md/fhi98PP
OPIUM	http://opium.sourceforge.net
APE	http://www.tddft.org/programs/APE
USPP	http://www.physics.rutgers.edu/~dhv/uspp
ADPACK	http://www.openmx-square.org/adpack/adpack.html
ATOMPAW	http://www.wfu.edu/~natalie/papers/pwpaw/man.html
Virtual Vault for PP	http://www.nnin.org/nnin_comp_psp_vault.html

DFT functionals

AM05	http://dft.sandia.gov/functionals/AM05.html
PBE	http://dft.uci.edu/pubs/PBE.asc
PBEsol	http://dft.uci.edu/pubs/PBEsol.html
Truhlar functionals	http://comp.chem.umn.edu/info/dft.htm
libxc	http://www.tddft.org/programs/octopus/wiki/index.php/Libxc

Useful links

- http://dft.sandia.gov/Quest/DFT_codes.html
- <http://www.nersc.gov/users/software/applications/materials-science/>
- http://www.c3se.chalmers.se/index.php/Software_Beda#Atomistic_Simulation
- <https://doc.zih.tu-dresden.de/hpc-wiki/bin/view/Compendium/NanoscaleSimulations>
- <http://www.psi-k.org/codes.shtml>
- http://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid-state_physics_software

Note: The current form of this policy does not specify 32 bit vs 64 bit or any other architecture, though it is expected that the native architecture of a given system (typically a 64 bit variant) is supported as the default target.

CHAPTER 16

HPCBIOS_2012-90: Software Tools and Development Environment

Note: Software Development environment includes a set of tools which are needed for building and using scientific computing software. The following is an attempt to define which ones are relevant for the HPCBIOS communities and set an action plan.

- BC Policy: HPCBIOS_2012-90
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common productivity environment across any HPCBIOS resources:

Name	Preferred version(s)	Compliance level	Subcategory
bash sh tcsh ksh csh zsh	latest	MUST	1. unix shells
vim emacs nano ed sed	latest	SHOULD	2. editors / stream editors
coreutils bc par grep gawk findutils parallel	latest	SHOULD	3. GNU variations of common Unix tools
more less screen script tmux	latest	SHOULD	4. user interaction
man help2man texinfo par	latest	SHOULD	5. online help
zlib openssl openssh	latest	SHOULD	6. zlib/openssl are typically required allover

Continued on next page

Table 16.1 – continued from previous page

Name	Preferred version(s)	Compliance level	Subcategory
wget curl lynx links rsync unison ftp lftp bbcp	latest	SHOULD	7. downloaders / file fetchers
subversion rcs cvs git	latest	SHOULD	10. version control
ssh ldap nfs nscd telnet netcat ntp ntpdate	latest	SHOULD	11. system intercon- nection essentials
strace ltrace lsof finger pwgen procs fuser who	latest	SHOULD	12. process & user in- spection
dstat sysstat iotop htop host ifconfig arp iptraf	latest	SHOULD	13. system inspection tools
nmap tcpdump dig bind- utils whois icmpinfo	latest	SHOULD	14. network inspection tools
ping traceroute ping6 traceroute6 tcptrace mtr ss	latest	SHOULD	15. network diagnos- tics - basics
iperf nuttcp bwctl ndp ndt thrulay vnstat	latest	SHOULD	16. network diagnos- tics - extras
patch diff diffstat m4	latest	SHOULD	17. patching & macros
bzip2 zip unzip pax arc lzip rzip cpio tar gzip	latest	SHOULD	18. archivers
dmidecode lspci blkid fdisk hdparm ethtool	latest	SHOULD	19. hardware & devices inspection
make cmake autoconf automake autotools	latest	SHOULD	20. build tools
yasm nasm dev86 elfutils	latest	SHOULD	21. assembly tools
bison byacc flex	latest	SHOULD	22. parsing tools
Continued on next page			

Table 16.1 – continued from previous page

Name	Preferred version(s)	Compliance level	Subcategory
GMP MPFR MPC PPL antlr expect	latest	MUST	23. http://gcc.gnu.org/install/prerequisites.html
gcc g++	v4.4.x OR v4.6.x	MUST	24. gcc compiler: http://gcc.gnu.org
g77 gfortran libf2c libc6- prof libgfortran3	latest	SHOULD	25. compiler related
libtool libltdl-dev ia32- libs	latest	SHOULD	26. lib or arch related
gdb valgrind gprof oprofile time objdump	latest	SHOULD	27. debugging, profiling, timing
makeself.sh mc	latest	SHOULD	28. misc platform
libstdc++	latest	MAY	30. base c++ libraries
ncurses ncurses-dev	latest	SHOULD	31. tui development
binutils-dev libssl-dev libtiff4-dev libsm-dev	latest	SHOULD	32. devel libs
libmysqlclient-dev libpng12-dev libpq-dev	latest	SHOULD	33. devel libs, part II
xutils-dev libx11-dev libxt-dev	latest	SHOULD	34. devel, X related
x11proto-xext-dev libxext-dev libqt4-dev	latest	SHOULD	35. devel, X related, part II
tk gtk2 QT	latest	SHOULD	36. X related
asciidoc doxygen doxygen-doc graphviz	latest	SHOULD	40. documentation related
sqlite3 sqlite3-dev	latest	SHOULD	50. db
Continued on next page			

Table 16.1 – continued from previous page

Name	Preferred version(s)	Compliance level	Subcategory
python ruby ri perl ipython ocaml tcl	latest	SHOULD	60. scripting & proto- typing
pychecker pyflakes pylint	latest	SHOULD	61. python code checkers
python-git python-svn	latest	SHOULD	62. scripting integra- tion with SCM
python-setuptools	latest	SHOULD	63. python-setuptools
openjdk / java	v1.6.0 or later	MAY	70. java SDK/Sun JDK
openmpi mpich2 mpich mvapich intelmpi	latest	MAY	80. mpi stacks
environment-modules	v3.2.10 or later	MAY	90. env. modules (Tcl/C)
lmod	v5.9.0 or later	MAY	91. env. modules (lua/lmod)
lsb lsb-release	latest	MAY	99. Linux Standard Base

This productivity environment will be supplemented with other related tools.

Potential sources of information for future upgrades of this list include:

- http://wiki.gumstix.org/index.php?title=Build_Environment_openSUSE_11.0
- http://www.grid5000.fr/mediawiki/index.php/FC11_Tlse
- http://sebastien.dinot.free.fr/liens_dev_c_cpp.html

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2012-91: Modules Namespace for HPC sites

Note: The main objective of this policy is to provide a consistent module namespace across any HPCBIOS compatible resources.

- BC Policy: HPCBIOS_2012-91
 - Date of Policy: 2012-12-15
-

Description

This document specifies a syntax for defining the names of packages as they become available via the modules mechanism of HPC sites. Such definition is important when sites perform under a common project or, share their user base or, simply, systems administrators need to swap configuration.

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in [RFC 2119](#).

Introduction

When the users enter an HPC system, they need to tune their software environment variables (to find available programs, library & include paths etc). The most common solution for that issue has been [modules](#) ; In practice though, HPC sites act in isolation, with lots of divergence in their module namespace. This does not allow good training examples or working code reuse by users so, work has to be done to bring convergence of the namespace across HPC sites.

This document is an attempt to define basic guidelines for modules namespace in order to address homegeneity arguments and reduce maintenance effort.

Scope

The standard should capture the maximum possible compatibility across sites, reducing unnecessary divergence, yet being practical and allowing some flexibility for local HPC sites needs & policies. There is a deliberate effort to be compatible with existing PRACE , DEISA , UNITE & PARMA-ITEA2 guidelines.

Standard to follow

The following approach is being suggested:

Package Lists and software guidelines:

- Any Package List SHOULD supply Package names in a non-ambiguous case-sensitive format
- The case-sensitive Package name SHOULD be derived from software author's naming preference

Conforming sites SHOULD adhere to these principles:

- MUST provide either the original Package name or, the lower case version
- MUST place the Version string after the "/" separator (to permit operation *module swap*)
- MUST document their module configuration & namespace policy, explicitly
- SHOULD provide a default Version when multiple ones exist (or, let one be implied)
- SHOULD prefer numbered Version strings with dots notation (as opposed to _ or - separators)
- MAY further expand the Version string or provide aliases, as per local needs; - examples: `-goolf-*`, `-openmpi-intel-32bit`
- MAY provide multiple buildsets of modules, as per local needs; - examples: `HPCBIOS.20130902`, `HPCBIOS.20131224`
- MAY provide multiple subcategories of modules, as per local needs; - examples: `bio cae chem compiler debugger devel ide lib math mpi numlib performance phys system tools vis`

For sites planning to be conforming to this standard, it is very important to adhere at the stated requirement levels and rigidly document implementation and deviations, so that users receive little to no surprise when moving from site to site.

Definitions

If definitions are required for what some terms mean, they can be considered as defined via UNITE standard (p.2/1.2) The relevant text is copied hereby, with only minor adjustments:

Package A software product, tool, or component which has its own name, a specific source and, is available / can be used / can be installed as separate entity.

Version Basic version naming SHOULD be in the form `<MajorVersion>.<MinorVersion>` or `<MajorVersion>.<MinorVersion>.<PatchLevel>` optionally followed by either `b<Number>` in case of a beta version or `rc<Number>` for release candidates. Examples: `2.1`, `4.1.3`, `3.2b2`, `4.5rc1`

Specialization Optional constraints which limit the applicability of a version of a package. Example: `WRF/3.3.1-goolf-1.4.10-dmpar` (ie. the suffix after the version) This MAY be necessary on computing systems which provide multiple MPI libraries or compiler suites (e.g., very commonly on Linux clusters). The alias of the default version MAY take the form `-<MPILibrary>-<Compiler>-<Precision>`. If so, the three components MUST always be specified in this order, however unnecessary constraints (not provided by a system) can be left out. Examples: `-openmpi-64bit`, `-mvapich-gnu-32bit` etc

Fully-Qualified Version (FQV) A fully-qualified version is the complete specification of a version of a package installed under specific constraints, i.e., it is the combination of a version followed by a specialization separated by a dash (“-”). Example: 1.4.3-mpich2-intel.

Platform The target computer system on which Package is installed. A platform is characterized by its vendor, architecture, and set of compiler suites and MPI libraries available for parallel programming.

References

- <http://modules.sourceforge.net/>
- <http://apps.fz-juelich.de/unite/files/unite-installguide.pdf>
- <http://www.deisa.eu/usersupport/user-documentation/deisa-common-production-environment/modules-usage>
- <http://www.deisa.eu/usersupport/user-documentation/deisa-common-production-environment/deisa-common-production-environment.pdf>
- <http://www.prace-ri.eu/prace-common-production/>

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2012-92: EasyBuild HPC Software Development Environment

Note: EasyBuild HPC Software Development environment allows to build and use Scientific Computing software in a consistent and reproducible way: it can manage Software and Modules in a coherent manner.

Though building HPC software can be done in alternative ways, either manual or structured, EasyBuild can incorporate the experience across multiple HPC sites and provides an advanced service, which simplistic solutions -such as Makefiles and regular shell scripts- can never match, since good code factorization is necessary in order to tackle the complexity of building software for multiple HPC platforms.

- BC Policy: HPCBIOS_2012-92
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide EasyBuild, because it allows to:

- build reproducible and rigorously defined software codes and,
- have consistent module namespaces across all HPCBIOS resources.

Objective

EasyBuild is a tool developed over the course of recent years and has been released for production use during SuperComputing'12, in Nov. 2012, by the HPC group of Ghent University.

Although other alternatives to EasyBuild could eventually appear in the open, so far it has been the most self-consistent publicly available tool in yielding reproducible Scientific Computing environments, transferable across multiple HPC platforms. Therefore it is considered necessary across HPC sites for both users and systems engineering work. EasyBuild runs fine in user-space and so, even if it is not provided in a system by default, users can and will eventually use it.

To be compliant with this policy, an HPC site **MUST** provide at least one production version of EasyBuild. This can be done either in the default \$PATH or, available via modules (as per `module load EasyBuild`). The default provided version **MAY** be customized as per local needs and **SHOULD** have out of the box functionality, as its developers describe in the documentation of the tool.

Example usage

```
sw@gaia-1:~$ module avail EasyBuild

----- /opt/apps/HPCBIOS/modules/all -----
EasyBuild/1.10.0
EasyBuild/1.11.0

sw@gaia-1:~$ module load EasyBuild/1.10.0
sw@gaia-1:~$ eb --version
== This is EasyBuild 1.10.0
sw@gaia-1:~$ which eb
/opt/apps/HPCBIOS.20131224/software/EasyBuild/1.10.0/bin/eb
```

Example build tree of WRF/WPS, along with GCC/OpenMPI & dependency libraries

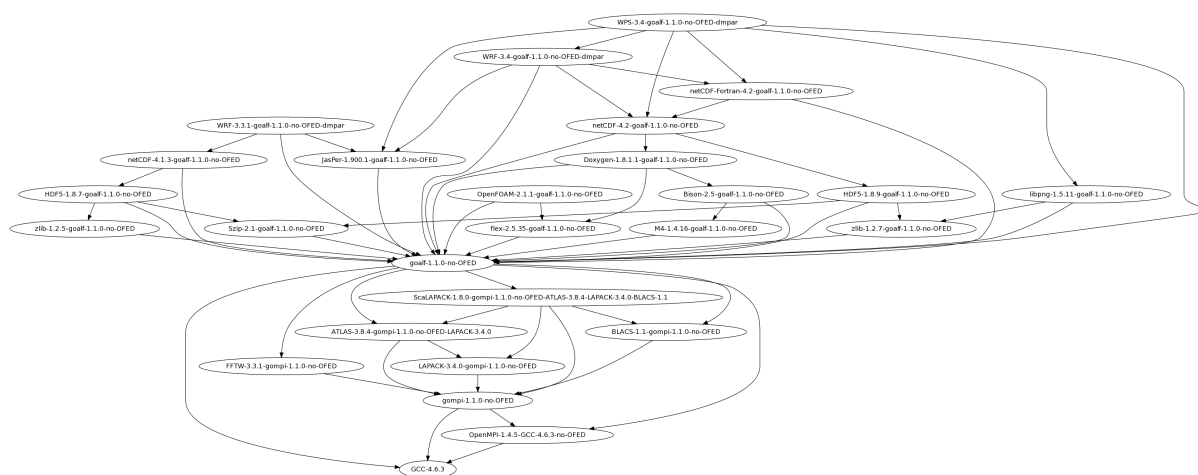


Fig. 18.1: Build tree of WRF/WPS & OpenFOAM

The Weather Research and Forecasting WRF model is a reference code for Weather and Climate communities. You can see how it is built along with its multiple dependent modules, with EasyBuild. More details about WRF are visible at [WRF's wikipedia page](#), including further pointers.

WRF & OpenFOAM are prime examples why a tool like EasyBuild is truly essential for an HPC site; not only can it manage the set of dependencies which are necessary in a structured way, it will also do so in a manner which is very consistent, eg. in relation to GNU/Intel compilers etc. For more detailed information why such documentation of parameters is important, see this issue: [EBnWRF](#).

Example module namespace

A site MAY provide the following list of modules. There is no need to provide the complete list of the following packages to be in compliance with this policy, since this is only provided for demonstration of the features and can be customized fully, as per site needs. Notably, future versions of EasyBuild MAY permit module categories, alternative versions, module hierarchies, improved integration with environment modules, filtered lists - defined as per user preferences - etc.

```
sw@gaia-25:~$ module avail

----- /opt/apps/HPCBIOS/modules/all -----
↪
```

ABINIT/7.0.3-x86_64_linux_gnu4.5	Szip/2.1-ictce-5.3.0
ABINIT/7.0.5-x86_64_linux_gnu4.5	Tar/1.26-goolf-1.4.10
ABINIT/7.2.1-x86_64_linux_gnu4.5	Tar/1.26-ictce-5.3.0
ABYSS/1.3.4-goolf-1.4.10-Python-2.7.3	Tcl/8.5.12-goolf-1.4.10
ABYSS/1.3.4-ictce-5.3.0-Python-2.7.3	Tcl/8.5.12-ictce-5.3.0
ALLPATHS-LG/46968-goolf-1.4.10	Tcl/8.5.14-goolf-1.4.10
AMOS/3.1.0-goolf-1.4.10	Tcl/8.5.14-ictce-5.3.0
AMOS/3.1.0-ictce-5.3.0	Theano/0.5.0-goolf-1.4.10-
↪Python-2.7.3	
ASE/3.6.0.2515-goolf-1.4.10-Python-2.7.3	Theano/0.5.0-ictce-5.3.0-
↪Python-2.7.3	
ASE/3.6.0.2515-ictce-5.3.0-Python-2.7.3	TiCCutils/0.3-goolf-1.4.10
ATLAS/3.8.4-gompi-1.1.0-no-OFED-LAPACK-3.4.0	TiCCutils/0.3-ictce-5.3.0
AnalyzeFMRI/1.1-15-ictce-5.3.0-R-2.15.2	TiMBL/6.4.3-goolf-1.4.10
Armadillo/2.4.4-goolf-1.4.10-Python-2.7.3	TiMBL/6.4.3-ictce-5.3.0
Armadillo/2.4.4-ictce-5.3.0-Python-2.7.3	TinySVM/0.09-goolf-1.4.10
Autoconf/2.69-goolf-1.4.10	TinySVM/0.09-ictce-5.3.0
Autoconf/2.69-ictce-5.3.0	Tk/8.5.12-goolf-1.4.10
Automake/1.13.4-goolf-1.4.10	Tk/8.5.12-ictce-5.3.0
Automake/1.13.4-ictce-5.3.0	TopHat/2.0.4-goolf-1.4.10
BFAST/0.7.0a-goolf-1.4.10	TopHat/2.0.8-goolf-1.4.10-
↪biodeps-1.6-extended	
BFAST/0.7.0a-ictce-5.3.0	TopHat/2.0.8-ictce-5.3.0
BLACS/1.1-gompi-1.1.0-no-OFED	TopHat/2.0.8-ictce-5.3.0-
↪biodeps-1.6-extended	
BLAST/2.2.27-goolf-1.4.10	TotalView/8.11.0-0-linux-x86-64
BLAST/2.2.28-goolf-1.4.10	TotalView/8.11.0-2-linux-x86-64
BLAST/2.2.28-ictce-5.3.0	Trilinos/10.12.2-goolf-1.4.10-
↪Python-2.7.3	
BLAT/3.5-goolf-1.4.10	Trilinos/10.12.2-ictce-5.3.0-
↪Python-2.7.3	
BWA/0.6.2-goolf-1.4.10	Trinity/2012-10-05-goolf-1.4.10
BWA/0.6.2-ictce-5.3.0	Trinity/2012-10-05-ictce-5.3.0
BWA/0.7.4-goolf-1.4.10	Trinity/2013-02-25-goolf-1.4.10
BWA/0.7.4-ictce-5.3.0	UDUNITS/2.1.24-goolf-1.4.10
BamTools/2.2.3-goolf-1.4.10	UDUNITS/2.1.24-ictce-5.3.0
Bash/4.2-goolf-1.4.10	UFC/2.0.5-goolf-1.4.10-Python-
↪2.7.3	
Bash/4.2-ictce-5.3.0	UFC/2.0.5-ictce-5.3.0-Python-2.
↪7.3	
BioPerl/1.6.1-goolf-1.4.10-Perl-5.16.3	UFL/1.0.0-goolf-1.4.10-Python-
↪2.7.3	
Biopython/1.61-goolf-1.4.10-Python-2.7.3	UFL/1.0.0-ictce-5.3.0-Python-2.
↪7.3	
Biopython/1.61-ictce-5.3.0-Python-2.7.3	VSC-tools/0.1.2-goolf-1.4.10-
↪Python-2.7.3	
Bison/2.5-goolf-1.1.0-no-OFED	VSC-tools/0.1.2-ictce-5.3.0-
↪Python-2.7.3	
Bison/2.5-goolf-1.4.10	VTK/5.10.1-goolf-1.4.10
Bison/2.5-ictce-3.2.2.u3	VTK/5.10.1-ictce-5.3.0
Bison/2.5-ictce-5.3.0	Valgrind/3.8.1-cgoolf-1.1.7
Bison/2.6.5-ictce-5.3.0	Valgrind/3.8.1-goolf-1.4.10
Bison/2.7-GCC-4.7.3	Velvet/1.2.07-goolf-1.4.10
Bison/2.7-goolf-1.4.10	Velvet/1.2.07-ictce-5.3.0
Bonnie++/1.03e-goolf-1.4.10	Velvet/1.2.09-goolf-1.4.10
Bonnie++/1.03e-ictce-5.3.0	Velvet/1.2.09-ictce-5.3.0
Boost/1.49.0-goolf-1.1.0-no-OFED-Python-2.7.3	ViennaRNA/2.0.7-goolf-1.4.10
Boost/1.49.0-goolf-1.4.10-Python-2.7.3	ViennaRNA/2.0.7-ictce-5.3.0
Boost/1.49.0-ictce-4.0.6-Python-2.7.3	Viper/1.0.0-goolf-1.4.10-
↪Python-2.7.3	
Boost/1.49.0-ictce-5.3.0-Python-2.7.3	Viper/1.0.0-ictce-5.3.0-Python-
↪2.7.3	

Boost/1.51.0-goolf-1.4.10	WPS/3.3.1-goolf-1.4.10-dmpar
Boost/1.51.0-goolf-1.4.10-Python-2.7.3	WPS/3.4-goolf-1.4.10-dmpar
Boost/1.51.0-ictce-5.3.0	WPS/3.4-ictce-5.3.0-dmpar
Boost/1.51.0-ictce-5.3.0-Python-2.7.3	WRF/3.3.1-goalf-1.1.0-no-OFED-
↪dmpar	
Boost/1.53.0-goolf-1.4.10	WRF/3.3.1-goolf-1.4.10-dmpar
Bowtie2/2.0.2-goolf-1.4.10	WRF/3.3.1-ictce-3.2.2.u3-dmpar
Bowtie2/2.0.2-ictce-5.3.0	WRF/3.3.1-ictce-5.3.0-dmpar
CCfits/2.4-goolf-1.4.10	WRF/3.4-goalf-1.1.0-no-OFED-
↪dmpar	
CD-HIT/4.5.4-ictce-5.3.0-2011-03-07	WRF/3.4-goolf-1.4.10-dmpar
CFITSIO/3.34-goolf-1.4.10	WRF/3.4-ictce-3.2.2.u3-dmpar
CGAL/4.0-goolf-1.4.10-Python-2.7.3	WRF/3.4-ictce-5.3.0-dmpar
CGAL/4.0-ictce-5.3.0-Python-2.7.3	XCrySDen/1.5.53-goolf-1.4.10
CLHEP/2.1.1.0-goolf-1.4.10	XCrySDen/1.5.53-ictce-5.3.0
CLHEP/2.1.1.0-ictce-5.3.0	XML/3.95-0.1-goolf-1.4.10-R-2.
↪15.2	
CMake/2.8.4-GCC-4.7.3	XML/3.95-0.1-ictce-5.3.0-R-2.
↪15.2	
CMake/2.8.4-goalf-1.1.0-no-OFED	YamCha/0.33-goolf-1.4.10
CMake/2.8.4-goolf-1.4.10	YamCha/0.33-ictce-5.3.0
CMake/2.8.4-goolfc-1.3.12	Yasm/1.2.0-goolf-1.4.10
CMake/2.8.4-ictce-4.0.6	Yasm/1.2.0-ictce-5.3.0
CMake/2.8.4-ictce-5.3.0	ZeroMQ/2.2.0-goolf-1.4.10
CPLEX/12.4	ZeroMQ/2.2.0-ictce-5.3.0
CRF++/0.57-goolf-1.4.10	ZeroMQ/3.2.2-goolf-1.4.10
CUDA/5.0.35-1	ZeroMQ/3.2.2-ictce-5.3.0
CVXOPT/1.1.5-goolf-1.4.10-Python-2.7.3	a2ps/4.14-goolf-1.4.10
CVXOPT/1.1.5-ictce-5.3.0-Python-2.7.3	a2ps/4.14-ictce-5.3.0
Chapel/1.6.0-goolf-1.4.10	ant/1.8.4-Java-1.7.0_10
Chapel/1.7.0-goolf-1.4.10	ant/1.9.0-Java-1.7.0_15
Clang/3.2-GCC-4.7.3	aria2/1.15.1-goolf-1.4.10
ClangGCC/1.1.3	aria2/1.15.1-ictce-5.3.0
ClustalW2/2.1-goolf-1.4.10	bam2fastq/1.1.0-goolf-1.4.10
ClustalW2/2.1-ictce-5.3.0	bam2fastq/1.1.0-ictce-5.3.0
Corkscrew/2.0-goolf-1.4.10	bbFTP/3.2.0-goalf-1.1.0-no-OFED
Corkscrew/2.0-ictce-5.3.0	bbFTP/3.2.0-goolf-1.4.10
Cufflinks/2.0.2-goolf-1.4.10	bbcp/12.01.30.00.0-amd64_
↪linux26	
Cython/0.16-goolf-1.4.10-Python-2.7.3	bbftpPRO/9.3.1-goalf-1.1.0-no-
↪OFED	
Cython/0.16-ictce-5.3.0-Python-2.7.3	bbftpPRO/9.3.1-goolf-1.4.10
DL_POLY_Classic/1.9-ictce-5.3.0-no-gui	bbftpPRO/9.3.1-ictce-5.3.0
Diffutils/3.2-goolf-1.4.10	beagle-lib/20120124-goolf-1.4.
↪10	
Diffutils/3.2-ictce-5.3.0	beagle-lib/20120124-ictce-5.3.0
Docutils/0.9.1-goolf-1.4.10-Python-2.7.3	binutils/2.22-goolf-1.4.10
Docutils/0.9.1-ictce-5.3.0-Python-2.7.3	biodeps/1.6-goolf-1.4.10
Doxygen/1.8.1.1-goalf-1.1.0-no-OFED	biodeps/1.6-goolf-1.4.10-
↪extended	
Doxygen/1.8.1.1-goolf-1.4.10	biodeps/1.6-ictce-5.3.0
Doxygen/1.8.1.1-ictce-3.2.2.u3	biodeps/1.6-ictce-5.3.0-
↪extended	
Doxygen/1.8.1.1-ictce-5.3.0	byacc/20120526-goolf-1.4.10
Doxygen/1.8.2-ictce-5.3.0	byacc/20120526-ictce-5.3.0
Doxygen/1.8.3.1-goolf-1.4.10	bzip2/1.0.6-cgoolf-1.1.7
ELinks/0.12pre5-goolf-1.4.10	bzip2/1.0.6-goalf-1.1.0-no-OFED
ELinks/0.12pre5-ictce-5.3.0	bzip2/1.0.6-goolf-1.4.10
EMBOSS/6.5.7-goolf-1.4.10	bzip2/1.0.6-ictce-4.0.6
EMBOSS/6.5.7-ictce-5.3.0	bzip2/1.0.6-ictce-5.3.0
ESPResSo/3.1.1-goolf-1.4.10-parallel	bzip2/1.0.6-ictce-5.5.0
ESPResSo/3.1.1-goolf-1.4.10-serial	cURL/7.27.0-goolf-1.4.10
ESPResSo/3.1.1-ictce-5.3.0-parallel	cURL/7.27.0-ictce-5.3.0

ESPResSo/3.1.1-ictce-5.3.0-serial	cURL/7.29.0-cgoolf-1.1.7
EasyBuild/1.10.0	cURL/7.29.0-goolf-1.4.10
Eigen/3.1.1-goolf-1.4.10	cairo/1.12.14-goolf-1.4.10
Eigen/3.1.1-ictce-5.3.0	ccache/3.1.9-goolf-1.4.10
FASTA/36.3.5e-goolf-1.4.10	ccache/3.1.9-ictce-5.3.0
FASTA/36.3.5e-ictce-5.3.0	cflow/1.4-goolf-1.4.10
FASTX-Toolkit/0.0.13.2-goolf-1.4.10	cflow/1.4-ictce-5.3.0
FASTX-Toolkit/0.0.13.2-ictce-5.3.0	cgdb/0.6.5-goolf-1.4.10
FCM/2.3.1	cgdb/0.6.5-ictce-5.3.0
FFC/1.0.0-goolf-1.4.10-Python-2.7.3	cgmpich/1.1.6
FFC/1.0.0-ictce-5.3.0-Python-2.7.3	cgmpolf/1.1.6
FFTW/2.1.5-ictce-5.3.0	cgmvpich2/1.1.12rc1
FFTW/3.3.1-gompi-1.1.0-no-OFED	cgmvolff/1.1.12rc1
FFTW/3.3.1-ictce-5.3.0	cgompi/1.1.7
FFTW/3.3.3-cgmpich-1.1.6	cgoolf/1.1.7
FFTW/3.3.3-cgmvpich2-1.1.12rc1	expat/2.1.0-cgoolf-1.1.7
FFTW/3.3.3-cgompi-1.1.7	expat/2.1.0-goolf-1.4.10
FFTW/3.3.3-gmvapich2-1.7.12	expat/2.1.0-ictce-5.3.0
FFTW/3.3.3-gmvapich2-1.7.12rc1	findutils/4.2.33-goolf-1.4.10
FFTW/3.3.3-gompi-1.3.12	findutils/4.2.33-ictce-5.3.0
FFTW/3.3.3-gompi-1.4.10	flex/2.5.35-goalf-1.1.0-no-OFED
FFTW/3.3.3-ictce-5.3.0	flex/2.5.35-goolf-1.4.10
FFTW/3.3.3-ictce-5.3.0-single	flex/2.5.35-ictce-3.2.2.u3
FIAT/1.0.0-goalf-1.1.0-no-OFED-Python-2.7.3	flex/2.5.35-ictce-5.3.0
FIAT/1.0.0-goolf-1.4.10-Python-2.7.3	flex/2.5.37-goolf-1.4.10
FIAT/1.0.0-ictce-4.0.6-Python-2.7.3	flex/2.5.37-ictce-5.3.0
FIAT/1.0.0-ictce-5.3.0-Python-2.7.3	fmri/1.4-8-ictce-5.3.0-R-2.15.2
FSL/4.1.9-goolf-1.4.10	fontconfig/2.10.91-goolf-1.4.10
FSL/4.1.9-ictce-5.3.0	freetype/2.4.10-goolf-1.4.10
GCC/4.6.3	freetype/2.4.10-ictce-5.3.0
GCC/4.6.4	freetype/2.4.11-goolf-1.4.10
GCC/4.7.2	g2clib/1.2.3-goolf-1.4.10
GCC/4.7.3	g2clib/1.2.3-ictce-5.3.0
GDB/7.5.1-cgoolf-1.1.7	g2lib/1.2.4-goolf-1.4.10
GDB/7.5.1-goolf-1.4.10	g2lib/1.2.4-ictce-5.3.0
GEOS/3.3.5-goolf-1.4.10	gawk/4.0.2-goolf-1.4.10
GEOS/3.3.5-ictce-5.3.0	gawk/4.0.2-ictce-5.3.0
GHC/6.12.3	gettext/0.18.2-cgoolf-1.1.7
GHC/7.4.2-goolf-1.4.10	gettext/0.18.2-goolf-1.4.10
GHC/7.6.2-goolf-1.4.10	gettext/0.18.2-ictce-5.3.0
GIMPS/p95v279.linux64	git/1.7.12-goolf-1.4.10
GLIMMER/3.02b-goolf-1.4.10	git/1.7.12-ictce-5.3.0
GLIMMER/3.02b-ictce-5.3.0	git/1.8.2-cgoolf-1.1.7
GMP/5.0.5-goolf-1.4.10	git/1.8.2-goolf-1.4.10
GMP/5.0.5-ictce-5.3.0	glproto/1.4.16-goolf-1.4.10
GPAP/0.9.0.8965-goolf-1.4.10-Python-2.7.3	glproto/1.4.16-ictce-5.3.0
GPAP/0.9.0.8965-ictce-5.3.0-Python-2.7.3	gmvpich2/1.7.12
GROMACS/4.6.1-goolf-1.4.10	gmvpich2/1.7.12rc1
GROMACS/4.6.1-goolfc-1.3.12	gmvolff/1.7.12
GSL/1.15-goolf-1.4.10	gmvolff/1.7.12rc1
GSL/1.15-ictce-5.3.0	gnuplot/4.6.0-goolf-1.4.10
Greenlet/0.4.0-goolf-1.4.10-Python-2.7.3	gnuplot/4.6.0-ictce-5.3.0
Greenlet/0.4.0-ictce-5.3.0-Python-2.7.3	goalf/1.1.0-no-OFED
HDF/4.2.7-patch1-goolf-1.4.10	gompi/1.1.0-no-OFED
HDF/4.2.7-patch1-ictce-5.3.0	gompi/1.3.12
HDF5/1.8.10-ictce-5.3.0	gompi/1.4.10
HDF5/1.8.10-ictce-5.3.0-gpfs	google-sparsehash/2.0.2-goolf-
↪ 1.4.10	
HDF5/1.8.10-patch1-goolf-1.4.10	google-sparsehash/2.0.2-ictce-
↪ 5.3.0	
HDF5/1.8.7-goalf-1.1.0-no-OFED	goolf/1.4.10
HDF5/1.8.7-goolf-1.4.10	goolfc/1.3.12
HDF5/1.8.7-ictce-3.2.2.u3	gperf/3.0.4-goolf-1.4.10

HDF5/1.8.7-ictce-5.3.0	gperf/3.0.4-ictce-5.3.0
HDF5/1.8.9-goalf-1.1.0-no-OFED	grib_api/1.10.0-goalf-1.4.10
HDF5/1.8.9-goalf-1.4.10	grib_api/1.9.18-goalf-1.4.10
HDF5/1.8.9-ictce-3.2.2.u3	grib_api/1.9.18-ictce-5.3.0
HDF5/1.8.9-ictce-5.3.0	guile/1.8.8-goalf-1.4.10
HH-suite/2.0.16-goalf-1.4.10	gzip/1.5-cgoalf-1.1.7
HMMER/3.0-goalf-1.4.10	gzip/1.5-goalf-1.4.10
HMMER/3.0-ictce-5.3.0	gzip/1.5-ictce-5.3.0
HMMER/3.1b1-goalf-1.4.10	h5py/2.0.1-goalf-1.4.10-Python-
↪2.7.3	
HMMER/3.1b1-ictce-5.3.0	h5py/2.0.1-ictce-5.3.0-Python-
↪2.7.3	
HPCBIOS_Bioinfo/20130717-goalf-1.4.10	h5utils/1.12.1-goalf-1.4.10
HPCBIOS_Bioinfo/20130717-ictce-5.3.0	hwloc/1.5.1-GCC-4.6.3
HPL/2.0-cgoalf-1.1.7	hwloc/1.6-iccifort-2011.13.367
HPL/2.0-goalf-1.4.10	hwloc/1.6.2-ClangGCC-1.1.3
HPL/2.0-ictce-5.3.0	hwloc/1.6.2-GCC-4.6.4
Harminv/1.3.1-goalf-1.4.10	hwloc/1.6.2-GCC-4.7.2
Harminv/1.3.1-ictce-5.3.0	icc/11.1.073
Hypre/2.8.0b-goalf-1.1.0-no-OFED	icc/11.1.075
Hypre/2.8.0b-goalf-1.4.10	icc/2011.10.319
Hypre/2.8.0b-ictce-4.0.6	icc/2011.13.367
Hypre/2.8.0b-ictce-5.3.0	icc/2011.6.233
Infernal/1.1rc1-goalf-1.4.10	icc/2013.1.117
Infernal/1.1rc1-ictce-5.3.0	icc/2013.2.146
Instant/1.0.0-goalf-1.4.10-Python-2.7.3	icc/2013.3.163
Instant/1.0.0-ictce-5.3.0-Python-2.7.3	icc/2013.4.183
Iperf/2.0.5-goalf-1.4.10	icc/2013.5.192
Iperf/2.0.5-ictce-5.3.0	iccifort/2011.13.367
JUnit/4.10-Java-1.7.0_10	ictce/3.2.2.u3
JUnit/4.11-Java-1.7.0_15	ictce/4.0.10
JasPer/1.900.1-goalf-1.1.0-no-OFED	ictce/4.0.6
JasPer/1.900.1-goalf-1.4.10	ictce/4.1.13
JasPer/1.900.1-ictce-3.2.2.u3	ictce/5.1.1
JasPer/1.900.1-ictce-5.3.0	ictce/5.2.0
Java/1.7.0_10	ictce/5.3.0
Java/1.7.0_15	ictce/5.4.0
Java/1.7.0_21	ictce/5.5.0
Jinja2/2.6-goalf-1.4.10-Python-2.7.3	ifort/11.1.073
Jinja2/2.6-ictce-5.3.0-Python-2.7.3	ifort/11.1.075
LAPACK/3.4.0-gompi-1.1.0-no-OFED	ifort/2011.10.319
LZO/2.06-goalf-1.4.10	ifort/2011.13.367
LZO/2.06-ictce-5.3.0	ifort/2011.6.233
LibTIFF/4.0.3-goalf-1.4.10	ifort/2013.1.117
LibTIFF/4.0.3-ictce-5.3.0	ifort/2013.2.146
Libint/1.1.4-goalf-1.4.10	ifort/2013.3.163
Libint/1.1.4-ictce-5.3.0	ifort/2013.4.183
M4/1.4.16-GCC-4.7.3	ifort/2013.5.192
M4/1.4.16-goalf-1.1.0-no-OFED	imkl/10.2.6.038
M4/1.4.16-goalf-1.4.10	imkl/10.3.10.319
M4/1.4.16-ictce-3.2.2.u3	imkl/10.3.12.361
M4/1.4.16-ictce-5.3.0	imkl/10.3.6.233
MATLAB/2013a	imkl/11.0.1.117
MCL/12.135-goalf-1.4.10	imkl/11.0.2.146
MCL/12.135-ictce-4.1.13	imkl/11.0.3.163
MCL/12.135-ictce-5.3.0	imkl/11.0.4.183
MDP/3.3-goalf-1.4.10-Python-2.7.3	imkl/11.0.5.192
MDP/3.3-ictce-5.3.0-Python-2.7.3	impi/4.0.0.028
MEME/4.8.0-goalf-1.4.10	impi/4.0.0.028-32bit
MEME/4.8.0-ictce-5.3.0	impi/4.0.2.003
METIS/4.0.1-goalf-1.4.10	impi/4.1.0.027
METIS/4.0.1-ictce-5.3.0	impi/4.1.0.030
METIS/5.0.2-goalf-1.1.0-no-OFED	impi/4.1.1.036

METIS/5.0.2-goolf-1.4.10	libctl/3.2.1-goolf-1.4.10
METIS/5.0.2-ictce-4.0.6	libdrm/2.4.27-goolf-1.4.10
METIS/5.0.2-ictce-5.3.0	libdrm/2.4.27-ictce-5.3.0
MPFR/3.1.0-goolf-1.4.10	libffi/3.0.11-goolf-1.4.10
MPFR/3.1.0-ictce-5.3.0	libgtextutils/0.6.1-goolf-1.4.
↪10	
MPICH/3.0.3-ClangGCC-1.1.3	libgtextutils/0.6.1-ictce-5.3.0
MTL4/4.0.8878	libharu/2.2.0-goolf-1.4.10
MUMmer/3.23-goolf-1.4.10	libharu/2.2.0-ictce-5.3.0
MUMmer/3.23-goolf-1.4.10-Perl-5.16.3	libidn/1.27-goolf-1.4.10
MUMmer/3.23-ictce-5.3.0	libmatheval/1.1.8-goolf-1.4.10
MVAPICH2/1.7-GCC-4.6.3	libpciaccess/0.13.1-goolf-1.4.
↪10	
MVAPICH2/1.7-GCC-4.6.3-hwloc-chkpt	libpciaccess/0.13.1-ictce-5.3.0
MVAPICH2/1.8.1-GCC-4.7.2	libpng/1.5.10-goolf-1.4.10
MVAPICH2/1.9-GCC-4.7.3	libpng/1.5.10-ictce-5.3.0
MVAPICH2/1.9a2-GCC-4.7.2	libpng/1.5.11-goolf-1.4.10
MVAPICH2/1.9b-GCC-4.7.2	libpng/1.5.11-ictce-5.3.0
MVAPICH2/1.9rc1-ClangGCC-1.1.3	libpng/1.5.13-goolf-1.4.10
MVAPICH2/1.9rc1-GCC-4.7.3	libpng/1.5.13-ictce-5.3.0
Meep/1.2-goolf-1.4.10	libpng/1.5.14-goolf-1.4.10
Mercurial/2.3.2-goolf-1.4.10-Python-2.7.3	libpng/1.6.2-goolf-1.4.10
Mercurial/2.3.2-ictce-5.3.0-Python-2.7.3	libpng/1.6.2-ictce-5.3.0
Mercurial/2.5.2-cgoolf-1.1.7-Python-2.7.3	libpthread-stubs/0.3-goolf-1.4.
↪10	
Mercurial/2.5.2-goolf-1.4.10-Python-2.7.3	libpthread-stubs/0.3-ictce-5.3.
↪0	
Mercurial/2.5.2-ictce-5.3.0-Python-2.7.3	libreadline/6.2-cgoolf-1.1.7
Mesa/7.11.2-goolf-1.4.10-Python-2.7.3	libreadline/6.2-goalf-1.1.0-no-
↪OFED	
Mesa/7.11.2-ictce-5.3.0-Python-2.7.3	libreadline/6.2-goolf-1.4.10
MetaVelvet/1.2.01-goolf-1.4.10	libreadline/6.2-ictce-4.0.6
MetaVelvet/1.2.01-ictce-5.3.0	libreadline/6.2-ictce-5.3.0
Mothur/1.30.2-goolf-1.4.10	libtool/2.4.2-goolf-1.4.10
Mothur/1.30.2-goolf-1.4.10-biodeps-1.6	libtool/2.4.2-ictce-5.3.0
Mothur/1.30.2-ictce-5.3.0	libungif/4.1.4-goolf-1.4.10
Mothur/1.30.2-ictce-5.3.0-biodeps-1.6	libungif/4.1.4-ictce-5.3.0
MrBayes/3.1.2-goolf-1.4.10	libunistring/0.9.3-goolf-1.4.10
MrBayes/3.1.2-ictce-5.3.0	libunistring/0.9.3-ictce-5.3.0
MrBayes/3.2.0-goolf-1.4.10	libxc/2.0.1-goolf-1.4.10
MrBayes/3.2.0-ictce-5.3.0	libxc/2.0.1-ictce-5.3.0
NASM/2.07-goolf-1.4.10	libxcb/1.8-goolf-1.4.10-Python-
↪2.7.3	
NASM/2.07-ictce-5.3.0	libxcb/1.8-ictce-5.3.0-Python-
↪2.7.3	
NCBI-Toolkit/9.0.0-goolf-1.4.10	libxml2/2.8.0-goolf-1.4.10
NEURON/7.2-goolf-1.4.10	libxml2/2.8.0-goolf-1.4.10-
↪Python-2.7.3	
NEURON/7.2-ictce-5.3.0	libxml2/2.8.0-ictce-5.3.0
NWChem/6.1.1-goolf-1.4.10-2012-06-27-Python-2.7.3	libxml2/2.8.0-ictce-5.3.0-
↪Python-2.7.3	
Oases/0.2.08-goolf-1.4.10	libxml2/2.9.0-goolf-1.4.10
Oger/1.1.3-goolf-1.4.10-Python-2.7.3	libxml2/2.9.0-ictce-5.3.0
Oger/1.1.3-ictce-5.3.0-Python-2.7.3	libxml2/2.9.1-goolf-1.4.10
OpenBLAS/0.2.6-cgmpich-1.1.6-LAPACK-3.4.2	libxml2/2.9.1-ictce-5.3.0
OpenBLAS/0.2.6-cgmvapich2-1.1.12rc1-LAPACK-3.4.2	libxslt/1.1.28-goolf-1.4.10
OpenBLAS/0.2.6-cgomp1-1.1.7-LAPACK-3.4.2	libxslt/1.1.28-ictce-5.3.0
OpenBLAS/0.2.6-gmvapich2-1.7.12-LAPACK-3.4.2	libyaml/0.1.4-goolf-1.4.10
OpenBLAS/0.2.6-gmvapich2-1.7.12rc1-LAPACK-3.4.2	libyaml/0.1.4-ictce-5.3.0
OpenBLAS/0.2.6-gomp1-1.3.12-LAPACK-3.4.2	likwid/3.0.0-goolf-1.4.10
OpenBLAS/0.2.6-gomp1-1.4.10-LAPACK-3.4.2	likwid/3.0.0-goolf-1.4.10-
↪pinomp	
OpenFOAM/2.1.1-goolf-1.4.10	lxml/3.1.2-goolf-1.4.10-Python-
↪2.7.3	

OpenFOAM/2.1.1-ictce-5.3.0 ↪2.7.3	lxml/3.1.2-ictce-5.3.0-Python-
OpenFOAM/2.2.0-goolf-1.4.10	make/3.82-goolf-1.4.10
OpenMPI/1.4.5-GCC-4.6.3	make/3.82-ictce-5.3.0
OpenMPI/1.4.5-GCC-4.6.3-no-OFED	makedepend/1.0.4-goolf-1.4.10
OpenMPI/1.6.3-iccifort-2011.13.367	makedepend/1.0.4-ictce-5.3.0
OpenMPI/1.6.4-ClangGCC-1.1.3 ↪Python-2.7.3	matplotlib/1.1.1-goolf-1.4.10-
OpenMPI/1.6.4-GCC-4.6.4 ↪Python-2.7.3	matplotlib/1.1.1-ictce-5.3.0-
OpenMPI/1.6.4-GCC-4.7.2 ↪Python-2.7.3	matplotlib/1.2.0-ictce-5.3.0-
OpenPGM/5.2.122-goolf-1.4.10 ↪Python-2.7.3	matplotlib/1.2.1-goolf-1.4.10-
OpenPGM/5.2.122-ictce-5.3.0	mc/4.6.1-goolf-1.4.10
OpenSSL/1.0.0-goolf-1.4.10	mc/4.6.1-ictce-5.3.0
PAML/4.7-goolf-1.4.10 ↪2.7.3	mpi4py/1.3-goolf-1.4.10-Python-
PAML/4.7-ictce-5.3.0 ↪2.7.3	mpi4py/1.3-ictce-5.3.0-Python-
PAPI/5.0.1-goolf-1.4.10	mpiBLAST/1.6.0-goolf-1.4.10
PAPI/5.0.1-ictce-5.3.0	mpiBLAST/1.6.0-ictce-5.3.0
PCRE/8.12-goolf-1.4.10	nano/2.2.6-goolf-1.4.10
PCRE/8.12-ictce-5.3.0	nano/2.2.6-ictce-5.3.0
PETSc/3.3-p2-goolf-1.1.0-no-OFED-Python-2.7.3 ↪1.7	ncurses/5.9-20130406-cgoolf-1.
PETSc/3.3-p2-goolf-1.4.10-Python-2.7.3	ncurses/5.9-GCC-4.7.3
PETSc/3.3-p2-ictce-4.0.6-Python-2.7.3	ncurses/5.9-goalf-1.1.0-no-OFED
PETSc/3.3-p2-ictce-5.3.0-Python-2.7.3	ncurses/5.9-goolf-1.4.10
PLINK/1.07-ictce-5.3.0	ncurses/5.9-goolfc-1.3.12
ParMETIS/3.1.1-goolf-1.4.10	ncurses/5.9-ictce-4.0.6
ParMETIS/3.1.1-ictce-5.3.0	ncurses/5.9-ictce-5.3.0
ParMETIS/4.0.2-goalf-1.1.0-no-OFED ↪OFED	netCDF/4.1.3-goalf-1.1.0-no-
ParMETIS/4.0.2-goolf-1.4.10	netCDF/4.1.3-goolf-1.4.10
ParMETIS/4.0.2-ictce-4.0.6	netCDF/4.1.3-ictce-3.2.2.u3
ParMETIS/4.0.2-ictce-5.3.0	netCDF/4.1.3-ictce-5.3.0
Pasha/1.0.3-goolf-1.4.10	netCDF/4.2-goalf-1.1.0-no-OFED
Pasha/1.0.3-ictce-5.3.0	netCDF/4.2-goolf-1.4.10
Pasha/1.0.5-ictce-5.3.0	netCDF/4.2-ictce-3.2.2.u3
Perl/5.16.3-goolf-1.4.10	netCDF/4.2-ictce-5.3.0
Perl/5.16.3-goolf-1.4.10-bare	netCDF/4.2.1.1-goolf-1.4.10
Perl/5.16.3-ictce-4.1.13	netCDF/4.2.1.1-ictce-5.3.0
Perl/5.16.3-ictce-5.3.0	netCDF-C++/4.2-goolf-1.4.10
Perl/5.16.3-ictce-5.3.0-bare ↪no-OFED	netCDF-Fortran/4.2-goalf-1.1.0-
Primer3/2.3.0-goolf-1.4.10	netCDF-Fortran/4.2-goolf-1.4.10
Primer3/2.3.0-ictce-5.3.0 ↪u3	netCDF-Fortran/4.2-ictce-3.2.2.
PyYAML/3.10-goolf-1.4.10-Python-2.7.3	netCDF-Fortran/4.2-ictce-5.3.0
PyYAML/3.10-ictce-5.3.0-Python-2.7.3	ns/2.35-goolf-1.4.10
PyZMQ/2.2.0.1-goolf-1.4.10-Python-2.7.3-zmq2	ns/2.35-ictce-5.3.0
PyZMQ/2.2.0.1-goolf-1.4.10-Python-2.7.3-zmq3	numactl/2.0.8-goolf-1.4.10
PyZMQ/2.2.0.1-ictce-5.3.0-Python-2.7.3-zmq2	numactl/2.0.8-ictce-5.3.0
PyZMQ/2.2.0.1-ictce-5.3.0-Python-2.7.3-zmq3 ↪Python-2.7.3	numexpr/2.0.1-ictce-5.3.0-
Python/2.7.3-cgoolf-1.1.7 ↪Python-2.7.3	numpy/1.6.2-goolf-1.4.10-
Python/2.7.3-goalf-1.1.0-no-OFED ↪2.7.3	numpy/1.6.2-ictce-5.3.0-Python-
Python/2.7.3-goolf-1.4.10 ↪Perl-5.16.3	orthomcl/2.0.8-goolf-1.4.10-
Python/2.7.3-ictce-4.0.6 ↪Perl-5.16.3	orthomcl/2.0.8-ictce-4.1.13-

Python/2.7.3-ictce-5.3.0	orthomcl/2.0.8-ictce-5.3.0-
↪Perl-5.16.3	
Python/2.7.5-ictce-5.3.0	otcl/1.14-goolf-1.4.10
Python/3.2.3-goolf-1.4.10	otcl/1.14-ictce-5.3.0
Python/3.2.3-ictce-5.3.0	pandas/0.11.0-goolf-1.4.10-
↪Python-2.7.3	
QuantumESPRESSO/4.2-goolf-1.4.10	pandas/0.11.0-ictce-5.3.0-
↪Python-2.7.3	
QuantumESPRESSO/4.2-goolf-1.4.10-hybrid	parallel/20130122-goolf-1.4.10
QuantumESPRESSO/5.0.2-goolf-1.4.10	parallel/20130122-ictce-5.3.0
QuantumESPRESSO/5.0.2-goolf-1.4.10-hybrid	paycheck/1.0.2-goolf-1.4.10-
↪Python-2.7.3	
R/2.15.2-goolf-1.4.10	paycheck/1.0.2-ictce-5.3.0-
↪Python-2.7.3	
R/2.15.2-goolf-1.4.10-bare	petsc4py/3.3-goalf-1.1.0-no-
↪OFED-Python-2.7.3	
R/2.15.2-ictce-5.3.0	petsc4py/3.3-goolf-1.4.10-
↪Python-2.7.3	
R/2.15.2-ictce-5.3.0-bare	petsc4py/3.3-ictce-4.0.6-
↪Python-2.7.3	
R/3.0.1-goolf-1.4.10-bare	petsc4py/3.3-ictce-5.3.0-
↪Python-2.7.3	
R/3.0.1-ictce-5.3.0-bare	phonopy/1.6.4-goolf-1.4.10-
↪Python-2.7.3	
RNAz/2.1-goolf-1.4.10	pixman/0.28.2-goolf-1.4.10
RNAz/2.1-ictce-5.3.0	pkg-config/0.27.1-goolf-1.4.10
ROOT/v5.34.01-goolf-1.4.10	pkg-config/0.27.1-ictce-5.3.0
SAMtools/0.1.18-goolf-1.4.10	problog/1.1-goolf-1.4.10
SAMtools/0.1.18-ictce-5.3.0	problog/1.1-ictce-5.3.0
SCOOP/0.5.3-goolf-1.4.10-Python-2.7.3	python-meep/1.4.2-goolf-1.4.10-
↪Python-2.7.3	
SCOOP/0.5.3-ictce-5.3.0-Python-2.7.3	scikit-learn/0.13-ictce-5.3.0-
↪Python-2.7.3	
SCOTCH/5.1.12b_esmumps-goalf-1.1.0-no-OFED	scipy/0.11.0-goolf-1.4.10-
↪Python-2.7.3	
SCOTCH/5.1.12b_esmumps-goolf-1.4.10	scipy/0.11.0-ictce-5.3.0-
↪Python-2.7.3	
SCOTCH/5.1.12b_esmumps-ictce-4.0.6	setuptools/0.6c11-goolf-1.4.10-
↪Python-2.7.3	
SCOTCH/5.1.12b_esmumps-ictce-5.3.0	setuptools/0.6c11-ictce-5.3.0-
↪Python-2.7.3	
SCOTCH/6.0.0_esmumps-goolf-1.4.10	tbb/4.0.5.339
SHRiMP/2.2.3-goolf-1.4.10	tcclcl/1.20-goolf-1.4.10
SHRiMP/2.2.3-ictce-5.3.0	tcclcl/1.20-ictce-5.3.0
SLEPc/3.3-p1-goolf-1.4.10-Python-2.7.3	tcsh/6.18.01-goolf-1.4.10
SLEPc/3.3-p1-ictce-5.3.0-Python-2.7.3	tcsh/6.18.01-ictce-5.3.0
SOAPdenovo/1.05-goolf-1.4.10	util-linux/2.22.2-goolf-1.4.10
SOAPdenovo/1.05-ictce-5.3.0	util-linux/2.22.2-ictce-5.3.0
SQLite/3.7.17-goolf-1.4.10	wiki2beamer/0.9.5-goolf-1.4.10-
↪Python-2.7.3	
SQLite/3.7.17-ictce-5.3.0	xcb-proto/1.7-goolf-1.4.10-
↪Python-2.7.3	
SWIG/2.0.4-goolf-1.4.10-Python-2.7.3	xcb-proto/1.7-ictce-5.3.0-
↪Python-2.7.3	
SWIG/2.0.4-ictce-5.3.0-Python-2.7.3	xorg-macros/1.17-goolf-1.4.10
ScaLAPACK/1.8.0-gompi-1.1.0-no-OFED-ATLAS-3.8.4...	xorg-macros/1.17-ictce-5.3.0
ScaLAPACK/2.0.2-cgmpich-1.1.6-OpenBLAS-0.2.6-LA...	xproto/7.0.23-goolf-1.4.10
ScaLAPACK/2.0.2-cgmvapich2-1.1.12rc1-OpenBLAS-0...	xproto/7.0.23-ictce-5.3.0
ScaLAPACK/2.0.2-cgomp1-1.1.7-OpenBLAS-0.2.6-LAP...	zlib/1.2.5-goalf-1.1.0-no-OFED
ScaLAPACK/2.0.2-gmvapich2-1.7.12-OpenBLAS-0.2.6...	zlib/1.2.5-goolf-1.4.10
ScaLAPACK/2.0.2-gmvapich2-1.7.12rc1-OpenBLAS-0...	zlib/1.2.5-ictce-3.2.2.u3
ScaLAPACK/2.0.2-gompi-1.3.12-OpenBLAS-0.2.6-LAP...	zlib/1.2.5-ictce-4.0.6
ScaLAPACK/2.0.2-gompi-1.4.10-OpenBLAS-0.2.6-LAP...	zlib/1.2.5-ictce-4.1.13

ScientificPython/2.8-goalf-1.1.0-no-OFED-Python...	zlib/1.2.5-ictce-5.3.0
ScientificPython/2.8-goolf-1.4.10-Python-2.7.3	zlib/1.2.7-cgmpolf-1.1.6
ScientificPython/2.8-ictce-4.0.6-Python-2.7.3	zlib/1.2.7-cgmvolf-1.1.12rc1
ScientificPython/2.8-ictce-5.3.0-Python-2.7.3	zlib/1.2.7-cgoolf-1.1.7
Shapely/1.2.15-goolf-1.4.10-Python-2.7.3	zlib/1.2.7-goalf-1.1.0-no-OFED
Shapely/1.2.15-ictce-5.3.0-Python-2.7.3	zlib/1.2.7-goolf-1.4.10
Sphinx/1.1.3-goolf-1.4.10-Python-2.7.3	zlib/1.2.7-ictce-3.2.2.u3
Sphinx/1.1.3-ictce-5.3.0-Python-2.7.3	zlib/1.2.7-ictce-4.0.10
Stow/1.3.3-goolf-1.4.10	zlib/1.2.7-ictce-4.0.6
Stow/1.3.3-ictce-5.3.0	zlib/1.2.7-ictce-4.1.13
SuiteSparse/3.7.0-goalf-1.1.0-no-OFED-withparmetis	zlib/1.2.7-ictce-5.3.0
SuiteSparse/3.7.0-goolf-1.4.10-withparmetis	zlib/1.2.8-goolf-1.4.10
SuiteSparse/3.7.0-ictce-4.0.6-withparmetis	zlib/1.2.8-ictce-4.1.13
SuiteSparse/3.7.0-ictce-5.3.0-withparmetis	zlib/1.2.8-ictce-5.3.0
Szip/2.1-goalf-1.1.0-no-OFED	zlib/1.2.8-ictce-5.5.0
Szip/2.1-goolf-1.4.10	zsync/0.6.2-goolf-1.4.10
Szip/2.1-ictce-3.2.2.u3	zsync/0.6.2-ictce-5.3.0

References

EasyBuild: <http://hpcugent.github.com/easybuild/>

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2012-93: Life Sciences Productivity Environment

Note: Life Sciences Productivity Environment includes a set of HPC tools which are needed for scientific computing and visualization in the respective domain. The following is an attempt to define which ones are relevant for the HPCBIOS communities and set an action plan.

- BC Policy: HPCBIOS_2012-93
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common productivity environment across any HPCBIOS resources:

Name	Category	version(s)	Compliance level	Reference URL
CPMD	Quantum Chemistry	x	SHOULD	http://www.cpmc.org/
Dalton	Quantum Chemistry	x	SHOULD	http://dirac.chem.sdu.dk/daltonprogram.org/
GAMESS-US	Quantum Chemistry	x	SHOULD	http://www.msg.ameslab.gov/gamess/
Gaussian	Quantum Chemistry	x	MAY	http://www.gaussian.com/
OpenEye	Chemical Informatics	x	SHOULD	http://www.eyesopen.com/
BLAST/mpiBLAST	Sequence Analysis	x	MUST	http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download
BWA	Sequence Analysis	x	MUST	http://bio-bwa.sourceforge.net/
CLUSTALW	Sequence Analysis	x	MUST	http://www.ebi.ac.uk/Tools/msa/clustalw2/
HMMER	Sequence Analysis	x	MUST	http://hmmer.janelia.org/
MrBayes	Sequence Analysis	x	MUST	http://mrbayes.sourceforge.net/
GROMACS	Molecular Modeling	x	MUST	http://www.gromacs.org/
NAMD	Molecular Modeling	x	SHOULD	http://www.ks.uiuc.edu/Research/namd/
AMBER	Molecular Modeling	x	SHOULD	http://ambermd.org/
AutoDock /Vina	Molecular Modeling	x	SHOULD	http://autodock.scripps.edu/ / http://vina.scripps.edu/
Glide	Molecular Modeling	x	SHOULD	http://www.schrodinger.com/
Dock	Molecular Modeling	x	SHOULD	http://dock.compbio.ucsf.edu/
Flexx	Molecular Modeling	x	SHOULD	http://www.biosolveit.de/FlexX/
FTDock	Molecular Modeling	x	SHOULD	http://www.sbg.bio.ic.ac.uk/docking/ftdock.html
LigandFit	Molecular Modeling	x	SHOULD	http://www.accelrys.com
Rosetta	Molecular Modeling	3.5	SHOULD	http://www.rosettacommons.org/

Reference documents:

- PRACE – The Scientific Case for HPC in Europe 2012-2020 <http://www.prace-ri.eu/PRACE-The-Scientific-Case-for-HPC>
- European Exascale Software Initiative Working Group report on Life Science and Health activities http://www.eesi-project.eu/media/download_gallery/EESI_D3.6_WG3.4-Report_R2.0.pdf

This productivity environment may be supplemented with other related productivity tools as they become available on allocated systems.

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

CHAPTER 20

HPCBIOS_2012-94: Bioinformatics & Comp. Biology Productivity Environment

Note: Bioinformatics & Computational Biology productivity environment includes a set of HPC tools which are needed for scientific computing and visualization in the respective domain. The following is an attempt to define which ones are relevant for the HPCBIOS communities and set an action plan.

- BC Policy: HPCBIOS_2012-94
- Date of Policy: 2013-12-22

The main objective of this policy is to provide the following common productivity environment across any HPCBIOS resources:

Name	Preferred version(s)	Compliance level	Reference URL
Python	v2.7.3	MUST	http://www.python.org/
Biopython	1.61	MUST	http://biopython.org
scipy	0.10.1	MUST	http://www.scipy.org
numpy	1.6.1	MUST	http://numpy.scipy.org
Cython	0.16	MUST	http://pypi.python.org/pypi/Cython/
R language	v2.14.2 or v3.2.x or later	SHOULD	http://www.r-project.org
Bioconductor	v2.10	SHOULD	http://www.bioconductor.org
Perl	v5.16.3	MUST	http://www.perl.org/
BioPerl	1.6.1	MUST	http://www.bioperl.org/
NCBI-Toolkit	9.0.0	MUST	http://www.ncbi.nlm.nih.gov/toolkit
BLAST+	v2.2.26 or later	MUST	http://blast.ncbi.nlm.nih.gov/
mpiBLAST	v1.6.0	MUST	http://www.mpiblast.org/
HMMER	v3	MUST	http://hmmer.janelia.org/
MrBayes	v3.1.2	MUST	http://mrbayes.sourceforge.net
AMOS	3.1.0	MUST	http://sourceforge.net/apps/mediawiki/amos
Boost	1.49 or later	MUST	http://www.boost.org
Bowtie2	2.0.2	MUST	http://bowtie-bio.sourceforge.net/bowtie2
BWA	0.6.2	MUST	http://bio-bwa.sourceforge.net
ClustalW2	2.1	MUST	http://www.ebi.ac.uk/Tools/msa/clustalw2
Cufflinks	2.0.2	MUST	http://cufflinks.cbc.umd.edu
Eigen	3.1.1	MUST	http://eigen.tuxfamily.org/index.php?title=Main_Page

Continued on next p

Table 20.1 – continued from previous page

Name	Preferred version(s)	Compliance level	Reference URL
FASTX-Toolkit	0.0.13.2	MUST	http://hannonlab.cshl.edu/fastx_toolkit
libgtextutils	0.6.1	MUST	http://hannonlab.cshl.edu/fastx_toolkit
Infernal	1.1	MUST	http://infernal.janelia.org
MUMmer	3.23	MUST	http://mummer.sourceforge.net
MetaVelvet	1.2.01	MUST	http://metavelvet.dna.bio.keio.ac.jp
RNAz	2.1	MUST	http://www.tbi.univie.ac.at/~wash/RNAz
SOAPdenovo	1.05	MUST	http://soap.genomics.org.cn/index.html
SAMtools	0.1.18	MUST	http://samtools.sourceforge.net
TopHat	2.0.4	MUST	http://tophat.cbcb.umd.edu
ViennaRNA	2.0.7	MUST	http://www.tbi.univie.ac.at/~ronny/RNA/vrna2.html
Velvet	1.2.07	MUST	http://www.ebi.ac.uk/~zerbino/velvet
BFAST	0.7.0a	MUST	http://bfast.sourceforge.net/
MCL	12.135	MUST	http://micans.org/mcl/
MEME	4.8.0	MUST	http://meme.nbcr.net/
HH-suite	2.0.16	MUST	ftp://toolkit.lmb.uni-muenchen.de/HH-suite/
BLAT	3.5	MUST	http://genome.ucsc.edu/FAQ/FAQblat.html
MUSCLE	3.8.31	MUST	http://www.drive5.com/muscle/
EMBOSS	6.5.7	MUST	http://emboss.sourceforge.net/
GATK	2.6-5 or later	MUST	http://www.broadinstitute.org/gatk/
PLINK	1.07	MUST	http://pngu.mgh.harvard.edu/purcell/plink/
CD-HIT	4.5.5	MUST	http://weizhong-lab.ucsd.edu/cd-hit/
RAxML	7.7.5 or later	MUST	https://github.com/stamatak/standard-RAxML
Oases	0.2.08	MUST	http://www.ebi.ac.uk/~zerbino/oases
Glimmer	3.0.2	MUST	http://ccb.jhu.edu/software/glimmer
PANDAsseq	2.5	MUST	https://github.com/neufeld/pandaseq
ABYSS	1.3.4	MUST	http://www.bcgsc.ca/platform/bioinfo/software/abyss
ALLPATHS-LG	46968	MUST	http://www.broadinstitute.org/software/allpaths-lg/blog/
ARB	5.5	MUST	http://www.arb-home.de/
bam2fastq	1.1.0	MUST	http://www.hudsonalpha.org/gsl/information/software/
BamTools	2.2.3	MUST	https://github.com/pezmaster31/bamtools
BEDTools	2.17.0	MUST	http://code.google.com/p/bedtools/
BiSearch	20051222	MUST	http://biseach.enzim.hu/
Bowtie	1.1.0	MUST	http://bowtie-bio.sourceforge.net/index.shtml
CAP3	20071221	MUST	http://seq.cs.iastate.edu/
cutadapt	1.3	MUST	http://code.google.com/p/cutadapt/
FASTA	36.3.5e	MUST	http://fasta.bioch.virginia.edu
fastahack	20110215	MUST	https://github.com/ekg/fastahack
FoldX	2.5.2	MUST	http://http://foldx.crg.es/
FRC_align	20130521	MUST	https://github.com/vezzi/FRC_align
FreeSurfer	2.5	MUST	http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki
FSL	4.1.9	MUST	http://www.fmrib.ox.ac.uk/fsl/
GLIMMER	3.02b	MUST	http://www.cbcb.umd.edu/software/glimmer/
Mothur	1.30.2	MUST	http://www.mothur.org/
orthomcl	2.0.8	MUST	http://orthomcl.org/
PAML	4.7	MUST	http://abacus.gene.ucl.ac.uk/software/paml.html
PASHA	1.0.3	MUST	http://pasha.sourceforge.net/
Picard	1.100	MUST	http://picard.sourceforge.net/
Primer3	2.3.0	MUST	http://primer3.sourceforge.net
RNAz	2.1	MUST	http://www.tbi.univie.ac.at/~wash/RNAz
SHRiMP	2.2.3	MUST	http://compbio.cs.toronto.edu/shrimp/
Stacks	1.03	MUST	http://creskolab.uoregon.edu/stacks/
Stride	1.0	MUST	http://structure.usc.edu/stride/
SURF	1.0	MUST	http://www.ks.uiuc.edu/Research/vmd/vmd-1.7/ug/node65

Continued on next page

Table 20.1 – continued from previous page

Name	Preferred version(s)	Compliance level	Reference URL
Trinity	2013-02-25	MUST	http://trinityrnaseq.sourceforge.net/
IDBA-UD	1.1.1	SHOULD	http://i.cs.hku.hk/~alse/hkubrg/projects/idba_ud
Ruby	v1.9.3-p194 or later	SHOULD	http://www.python.org/
HOMER	4.2	SHOULD	http://homer.salk.edu/homer/
PhyML	latest	SHOULD	http://www.atgc-montpellier.fr/phyml/
Clustal	latest	SHOULD	http://www.clustal.org/
PyMOL	latest	SHOULD	http://pymol.sf.net/
FastQC	0.10.0	SHOULD	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Future potential candidates for this policy:

T-Coffee	latest	SHOULD	http://www.tcoffee.org/
GPU-BLAST	latest	SHOULD	http://eudoxus.cheme.cmu.edu/gpublast/gpublast.html
CUDA-BLASTP	latest	SHOULD	https://sites.google.com/site/liuweiguohome/software
mpiCUDA-BLASTP	latest	SHOULD	https://sites.google.com/site/liuweiguohome/mpicuda-blastp
mr-mpi-blast	latest	SHOULD	https://github.com/andreito/mr-mpi-blast
BioRuby	latest	SHOULD	http://bioruby.org
PolyPhen-2	2.2.2r405	SHOULD	http://genetics.bwh.harvard.edu/pph2/dokuwiki/start
SSAHA	31c	SHOULD	http://www.sanger.ac.uk/resources/software/ssaha/
RepeatMasker	latest	SHOULD	http://www.repeatmasker.org/
ABC-SysBio	latest	SHOULD	http://sourceforge.net/projects/abc-sysbio/
Annovar	latest	SHOULD	http://www.openbioinformatics.org/annovar/
PEMer	latest	SHOULD	http://sv.gersteinlab.org/pemer/
deFuse	latest	SHOULD	http://defuse.sf.net/
TopHat-Fusion	latest	SHOULD	http://tophat.cbcb.umd.edu/fusion_index.html
BreakDancer	latest	SHOULD	http://breakdancer.sourceforge.net/
PANTHER	latest	SHOULD	http://www.pantherdb.org/tools/
HHPred	latest	SHOULD	http://toolkit.lmb.uni-muenchen.de/hhpred
pysam	latest	SHOULD	http://code.google.com/p/pysam/
ReadSeq	latest	SHOULD	http://iubio.bio.indiana.edu/soft/molbio/readseq/java/
Phrap	latest	SHOULD	http://www.phrap.org/phredphrapconsed.html
Phred	latest	SHOULD	http://www.phrap.org/phredphrapconsed.html
Consed	latest	SHOULD	http://www.phrap.org/consed/consed.html
nesoni	latest	SHOULD	http://bioinformatics.net.au/software/nesoni.shtml
galaxy	>ca0c4ad2bb39	SHOULD	http://galaxy.psu.edu/
swarm	latest	SHOULD	http://biowulf.nih.gov/apps/swarm.html /download/swarm.tar
WU-BLAST	latest	SHOULD	http://blast.wustl.edu/
PAUP	latest	SHOULD	http://paup.csit.fsu.edu/
PHYLIP	latest	SHOULD	http://evolution.genetics.washington.edu/phylip.html
QuEST	2.4	SHOULD	http://mendel.stanford.edu/sidowlab/downloads/quest/

This productivity environment will be supplemented with other related productivity tools as they become available on allocated systems.

Potential sources of information for further future upgrades of this list include:

- <http://www.vital-it.ch/software/tools.php>
- <http://wiki.hpc.ufl.edu/doc/Category:Bioinformatics>
- http://www.csc.fi/english/research/sciences/bioscience/programs/index_html
- <http://bacpathgenomics.wordpress.com/software/>
- <http://umbc.rnet.missouri.edu/general/software/alphabetical.html>

- <http://www.broadinstitute.org/software/bsi-sig/>
- <http://ncgas.org/software>, genome assembly, assembly, genomics
- <http://confluence.rcs.griffith.edu.au:8080/display/GHPC/qiime#qiime-Alignment%2Ctreebuilding%2Ctaxonomyassignment%2COTUpicking%2Candotherdatagenerationsteps%28requiredforalternativepipelines%29>
- <http://www.bioplexity.org/lectures/ebi-s10.pdf>
- <http://helix.nih.gov/Applications/>
- <http://openwetware.org/wiki/Image:Tools.png>
- http://en.wikipedia.org/wiki/List_of_sequence_alignment_software
- <http://wiki.galaxyproject.org/Admin/Tools/Tool%20Dependencies>
- <http://www.eaglegenomics.com/2012/04/the-elements-of-bioinformatics/>
- <http://nebc.nerc.ac.uk/tools/bio-linux-5/package-list>

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2012-95: Molecular Dynamics Productivity Environment

Note: Molecular Dynamics productivity environment includes a set of tools which are needed for scientific computing and visualization in the respective domain. The following is an attempt to define which ones are relevant for the HPCBIOS communities and set an action plan.

- BC Policy: HPCBIOS_2012-95
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common productivity environment across any HPCBIOS resources:

Name	Preferred version(s)	Compliance level	Reference
GROMACS	v4.5.5 or later	MUST	http://www.gromacs.org
CP2K	20111205 or latest	MUST	http://www.cp2k.org
CHARMM	c37b2	SHOULD	http://www.charmm.org
NAMD	v2.8	MUST	http://www.ks.uiuc.edu/Research/namd
VMD	v1.9	SHOULD	http://www.ks.uiuc.edu/Research/vmd
DL_POLY	v4.02	SHOULD	http://www.ccp5.ac.uk/DL_POLY
LAMMPS	v1Feb2014 or later	SHOULD	http://lammps.sandia.gov
CPMD	v3.15.1	SHOULD	http://www.cpmc.org
NWCHEM	v6.1	SHOULD	http://www.nwchem-sw.org
XMakemol	v5.16	SHOULD	http://www.nongnu.org/xmakemol
TINKER	v6.0.08	SHOULD	http://dasher.wustl.edu/tinker/
HOOMD-blue	v0.10.0	SHOULD	http://codeblue.umich.edu/hoomd-blue/
AMBER	v11	SHOULD	http://ambermd.org

This productivity environment will be supplemented with other related productivity tools as they become available on allocated systems.

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2012-96: Common Set of Commercial Compilers

Note: Providing commercial compilers is imperative for many user groups.

- BC Policy: HPCBIOS_2012-96
 - Date of Policy: 2012-12-15
-

Thus, to increase user productivity and expand commonality across HPCBIOS sites, sites shall install the following commercial compilers, either through a user's default path or the availability of a module file.

In addition, in order to truly take advantage of the modern HPC environment, *MPI support for the commercial compilers suite MUST be implemented* for compliance with this policy, with any MPI stack.

Name	Preferred version(s)	Compliance level	Comments
Intel C/C++ Compiler	v11, v12.1, v13.x or later	MUST	icc/11.1.073 (v11), icc/2011.6.233 (v12.1), icc/2013.3.163 (v13.1.1), icc/2013.5.192 (v13.1.3)
Intel Fortran Compiler	v11, v12.1, v13.x or later	MUST	ifort/11.1.073 (v11), ifort/2011.6.233 (v12.1), ifort/2013.3.163 (v13.1.1), ifort/2013.5.192 (v13.1.3)
Lahey/Fujitsu Fortran	v6.2 & v8.1	SHOULD	This may be desired by certain climate commu- nity members

It is at the discretion of each site which (if any) libraries will be compiled with the commercial compilers. Ultimately, it will be the responsibility of each user of the commercial compilers to create the libraries that they require for their compiled software (at user or group context).

Note: The current form of this policy does not specify 32 bit vs 64 bit or any other architecture, though it is expected that the native architecture of a given system (typically a 64 bit variant) is supported as the default target.

HPCBIOS_2012-97: Climate Science Productivity Environment

Note: Climate Science productivity environment includes a set of tools which are needed for scientific computing and visualization in the respective domain. The following is an attempt to define which ones may be relevant for HPCBIOS sites.

- BC Policy: HPCBIOS_2012-97
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common productivity environment across any HPCBIOS resources that need it:

Name	Preferred version	Compliance level	Reference
Ferret	v6.10 / v6.72	MUST	http://ferret.wrc.noaa.gov/Ferret/
NCL	v6.0.0-nodap	MUST	http://www.ncl.ucar.edu/
WRF	v3.3 or later	MUST	http://www.wrf-model.org/
UDUNITS	2.1.24	MUST	http://www.unidata.ucar.edu/software/udunits/
NCO	v4.0.0 or later	MUST	http://www.nco.sourceforge.net/
ncview	v1.93g or later	MUST	http://www.nersc.gov/users/software/vis-analytics/ncview/
GrADS	v1.9b4 / v2.0.1	SHOULD	http://www.iges.org/grads/
OpenGrADS	v0.1.9.4	SHOULD	http://www.opengrads.org/
nc-tools/ANTS	v1.1 ?	MAY	http://science.arm.gov/~cflynn/ANTS/
OpenDX	v4.3.2	SHOULD	http://www.opendx.org/
ParaView	v3.8.0 or later	MUST	http://www.paraview.org/
CDO	v1.6.0	MUST	https://code.zmaw.de/projects/cdo
gnuplot	v4.4.4	MUST	http://www.gnuplot.info/
VTK	v5.4.2 or later	MUST	http://www.vtk.org/
R-language	v2.14.1 or later	MUST	http://www.r-project.org/
Python	v2.x AND v3.x	MUST	http://www.python.org/
h4toh5	v2.2.0	MUST	http://www.hdfgroup.org/h4toh5/
Open-FOAM	v2.0.1 or v2.1.1	MUST	http://www.openfoam.com/
IDL	8.0 or later	SHOULD	http://www.exelisvis.com/ProductsServices/IDL.aspx
MATLAB	v2012a or later	SHOULD	http://www.mathworks.com/products/matlab/

This productivity environment will be supplemented with other related productivity tools as they become available on allocated systems.

There is also an effort to support for the Climate community the following data formats:

Name	Prefered version(s)	Compliance level	References
NetCDF	v3.6.3, v4.0, v4.1	MUST	http://www.unidata.ucar.edu/netcdf
HDF5	v1.8.2 & v1.8.8	MUST	http://www.hdfgroup.org/HDF5/
HDF4	v4.2.6	MUST	http://www.hdfgroup.org/products/hdf4/
Grib	v1 & v2	MUST	http://www.grib.us/

Note: Please notify is this policy is incomplete both your local technical representative & user-support instead of the respective HPC site.

HPCBIOS_2012-98: Common Set of Environment Variables

Note: This policy is written to define a core set of environment variables that represent a resource definition at each of the HPCBIOS compatible sites participating in a Baseline Configuration (BC) Initiative. These variables shall be pre-defined in all login scripts, making them automatically available to users at each site, whenever the relevant functionality is in place.

- BC Policy: HPCBIOS_2012-98
 - Date of Policy: 2012-12-15
-

The current list of variables has as follows:

BC_CORES_PER_NODE (MUST) This variable contains the number of cores per node for the default compute node type to which a job is being submitted.

BC_MEM_PER_NODE (MUST) This variable contains the approximate maximum memory per node available to an end user program (in integer MBs) for the default compute node type to which a job is being submitted. Users are advised to not consume more than 95% of the stated limit and consider well any implications when node sharing is in place (this aspect MAY be queue-specific).

BC_MPI_TASKS_ALLOC (SHOULD) This variable, intended to be referenced from inside a job script, shall contain the number of MPI tasks that are allocated for a particular job.

BC_NODE_ALLOC (SHOULD) This variable, intended to be referenced from inside a job script, shall contain the number of nodes allocated for a particular job.

HOME (MUST) This is defined automatically in a Unix environment and MUST correspond to a location where a user can define his initialization scripts. This area is backed up at a defined frequency and policy per each site. Users are strongly advised to increase redundancy of critical items by ensuring an own copy; Using Source Version Control Systems (git, hg, svn, cvs) is highly RECOMMENDED for custom software codes, which may represent many manhours of work.

JAVA_HOME (SHOULD) This variable contains the path to the base directory of the default installation of JAVA on a particular compute platform. If the platform does not have JAVA installed, this variable should not be defined. It can also be provided dynamically via the modules mechanism (ie. loading a modulefile).

SAMPLES_HOME (SHOULD) This variable contains the path to the base directory for the sample codes and scripts installed by a site's staff under BC policy Sample Code Repository.

SOFT_HOME (MUST) This variable contains the path to the base directory for the application codes installed on a system.

SCRATCH (SHOULD) This is a shared parallel filesystem optimized for large file and high bandwidth access. This file system is not backed up and files on this system **MAY** be purged as soon as a job finishes. Users **SHOULD** use this filesystem for jobs that seek to have high aggregate bandwidth. A given system **MAY** not provide it or, restrict its usage to certain users/groups.

TMPDIR (MUST) This is a directory on a local temporary file system (i.e., local high speed disk) that **SHOULD** be available on all HPCBIOS compatible systems. TMPDIR is intended to be used by executing programs to perform file I/O that is local to that system in order to avoid slower file I/O across a network mounted file system, such as a user's HOME or ARCHIVE directories. It is not intended to be used as a file storage area by users, rather as working area within tasks. Accordingly, this file system **SHALL NOT** be backed up or exported to any other system. In the event of file or directory structure deletion or a catastrophic disk failure, such files and directory structures are lost. Thus, it is the user's responsibility to transfer files that need to be saved to a location that allows for longer term file storage, such as the user's \$HOME, \$WORK or \$ARCHIVE directory locations - if these are made permanent via a backup policy - or, another external system.

WORK (MUST) This is an individual user's directory for stage-in, stage-out of running jobs. Also, this is the correct location to store checkpointing state. This file system **MAY** not be backed up or exported to any other system. In the event of file or directory structure deletion or a catastrophic storage failure, such files and directory structures **MAY** be lost. Thus, it is the user's responsibility to increase redundancy and transfer files that need to be saved to a location that allows for longer term file storage, such as \$HOME or \$ARCHIVE directory locations - if they are permanent via a backup policy - or, another external system. Fitness of filesystem type is users' responsibility.

WORKDIR (SHOULD) Same as WORKDIR, see above

This list of environment variables is subject to change, and should be reviewed on a periodic basis for potential additions, modifications or deletions.

HPCBIOS_2012-99: GPU Productivity Environment

Note: GPU productivity environments have recently become an important part of scientific computing and visualization. The following is an attempt to define which ones are relevant for HPCBIOS compliant sites and set an action plan.

- BC Policy: HPCBIOS_2012-99
 - Date of Policy: 2012-12-15
-

The main objective of this policy is to provide the following common GPU productivity environment across all HPCBIOS resources:

Package/Tool	Requested version(s)	Compliance level	Reference
CUDA Toolkit	v4.2 or later	MUST	https://developer.nvidia.com/cuda-downloads
rCUDA	v4.0.1	SHOULD	http://www.rCUDA.net
PyCUDA	2012.1	SHOULD	http://mathematician.de/software/pycuda
PyOpenCL	2012.1	SHOULD	http://mathematician.de/software/pyopencl
MAGMA	v1.3	SHOULD	http://icl.cs.utk.edu/magma
ViennaCL	v1.4.2	SHOULD	http://viennacl.sourceforge.net
AMD math library for OpenCL (APPML)	v1.10	SHOULD	http://developer.amd.com/libraries/appmathlibs/Pages/default.aspx
Rgputools for R language	v0.28	SHOULD	http://cran.r-project.org/web/packages/gputools
Intel OpenCL SDK	v1.5	SHOULD	http://software.intel.com/en-us/articles/vcsource-tools-opencl-sdk
CULA & CULASPARSE	R16a & S4	SHOULD	http://www.culatools.com
VCL	v1.19	SHOULD	http://www.mosix.org/txt_vcl.html

This GPU productivity environment will be supplemented with other such productivity tools, as they become available on allocated systems.

Potential sources of information for further future upgrades of this list include:

- <http://brainarray.mbni.med.umich.edu/brainarray/rgpgpu/>
 - <http://www.nvidia.com/object/gpu-test-drive.html>
-

HPCBIOS_2013-01: Common Dependencies for Life Science Applications

Note: As the number of Life Science applications increases, it becomes more and more pronounced the need to stabilize on a substrate of common dependencies; this permits to mix together software which depends on multiple (dynamic) libraries, with **predictable results and no conflicts**. This need is particularly important when an HPC platform delivers Life Science and Bioinformatics packages via *modules*, which need to be combined in a pipeline or simply loaded as one unit, for reproducibility purposes.

- BC Policy: HPCBIOS_2013-01
 - Date of Policy: 2013-06-15
-

A list has been compiled with the most common dependencies for Life Science applications; some room for (justified) deviations is provided by softening the compliance level to **SHOULD**:

Name	Preferred version(s)	Compliance level
bzip2	v1.0.6	MUST
zlib	v1.2.7	MUST
libreadline	v6.2	MUST
ncurses	v5.9	MUST
Python	v2.7.3	SHOULD
Boost	v1.51.0-Python-2.7.3	SHOULD
SAMtools	v0.1.18	SHOULD
Perl	v5.16.3	SHOULD
Java	v1.7.0_10	SHOULD
libpng	v1.5.13	SHOULD

N.B. The *GNU-compatible suite* **MUST** be implemented for compliance with this policy on a given site. Furthermore, *libreadline* **SHOULD** be build with *ncurses* as dependency.

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

HPCBIOS_2015-01: Biocompressors, for NGS data and bioinformatics

Note: As the volume of NGS data increases, it becomes more and more pronounced the need to optimally store and handle data coming from Next Generation Sequencing pipelines, in efficient manner. This is particularly true when such data reside in a common shared storage back-end which sustains big I/O pressure from hundreds or thousands of concurrent accesses.

Maintaining the volume of data in compressed format, may achieve the following objectives:

- Reduces considerably the required storage space (possibly as much as 5 or 6 times)
- Provides an incentive for users, to perform correct stage-in/stage-out operations
- May increase the speed with which I/O operations take place, if slow disk operations are minimized
- Saturation of I/O subsystems may become less pronounced, or be postponed towards heavier loads

The aim of this policy is to ensure that a sufficiently rich palette of entropy-conscious tools are available on machines which may be flooded by NGS data.

- BC Policy: HPCBIOS_2015-01
 - Date of Policy: 2015-02-02
-

A list has been compiled with the most common compression/decompression tools for Bioinformatics data;

Biocompressors:

DSRC	2.0rc	SHOULD	http://sun.aei.polsl.pl/dsrc
fastqz	1.5	SHOULD	http://mattmahoney.net/dc/fastqz/
fqzcomp	4.6	SHOULD	http://sourceforge.net/projects/fqzcomp/
Quip	1.1.8	SHOULD	http://homes.cs.washington.edu/~dcjones/quip
SCALCE	2.7	SHOULD	http://scalce.sourceforge.net/Home
ZPAQ	7.00	SHOULD	http://mattmahoney.net/dc/zpaq.html

Kindly notify -if this policy is inadequate for your work- both your local site technical representative & HPCBIOS maintainers.

Reference documents:

- Compression of FASTQ and SAM Format Sequencing Data, by James K. Bonfield & Matthew V. Mahoney
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0059190>

CHAPTER 28

Indices and tables

- `genindex`
- `modindex`

CHAPTER 29

Reference Pages

- Search
-

This document was last rebuilt on: 21 June 2017.

R

RFC

RFC 2119, [3](#), [43](#)