Linux Differences

Table of Contents

Preface

Scope: This Linux Differences guide summarizes the Linux-based computing resources available to LC users and explains the differences (among Linux and AIX and Tru64 UNIX) in tool and library availability, tool location (path), within-tool options and features, and operating-system features that can affect code portability and the general usefulness of local Linux computers. One section also discusses local support for POSIX threads on LC Linux machines, while another focuses on Linux-relevant environment variables. For a systematic introduction to the design goals and local enhancements built into the version of Linux that LC is itself developing (called CHAOS), see the local document called <u>CHAOS: Linux at Livermore</u> (URL: http://www.llnl.gov/LCdocs/chaos). For information on the job scheduler coming into use on LC Linux clusters, see <u>Moab at LC</u> (URL: http://www.llnl.gov/LCdocs/moab).

Availability: Comparing the availability of tools and features between Linux and other LC operating systems is the prime goal of this whole manual.

Consultant: For help contact the LC customer service and support hotline at 925-422-4531 (open e-mail: lc-hotline@llnl.gov, SCF e-mail: lc-hotline@pop.llnl.gov).

Printing: The print file for this document can be found at

```
OCF: http://www.llnl.gov/LCdocs/linux/linux.pdf
SCF: http://www.llnlgov/LCdocs/linux/linux_scf.pdf
```

Introduction

This section introduces Linux, its role in providing computing resources for LC users, and the role of this document in interpreting Linux locally.

Linux Overview

GENERAL BACKGROUND.
The UNIX operating system comes in many historically related varieties. It originated at Bell Laboratories in the late 1960s, and as early as 1975 it diverged into Berkeley Software Distribution (BSD) and System V versions (with similar basic kernels for hardware interaction but somewhat different service programs). In the late 1970s, Digital Equipment Corporation developed a UNIX version called ULTRIX, enhanced and tuned to suit Digital hardware. Digital later replaced this with one of the industry's first 64-bit UNIX versions, called Tru64 (still in use on Compaq Alpha machines, such as LC's secure Furnace cluster and open LX cluster, today). After Sun and HP also offered UNIX systems, IBM decided to do the same, producing a UNIX version for their RS/6000 machines (such as the LC ASCI machines) called AIX. AIX blends into System V the BSD software features as well as POSIX 1003 standards conformance.

Meanwhile, in 1991, Helsinki University student Linus Torvalds wrote his own UNIX kernel, intended to be freely distributed for hobbiest and hacker experimentation. This was soon called Linux and its source code became available on the Internet. Richard Stallman's Free Software Foundation had already been giving away source code for a growing body of utility and service programs during the 1980s under the "GNU General Public License" (which guaranteed their free and continuing public availability). All GNU software was compatible with UNIX, but no one had actually written a free UNIX kernel to support it. Thus the happy confluence of Torvald's 1991 experiment and the on-going GNU-tools project yielded a complete open-source operating system package (or "distribution"): the Linux-based GNU system.

There are now at least half a dozen commercial providers of some variation on this open-source Linux-GNU distribution. LC's version happens to come from Red Hat, Inc.

LINUX AT LIVERMORE.
Late in 2001, LC began seriously increasing its production computing capacity for institutional (nonASCI) users with machines that run Linux. There are three reasons for this shift in emphasis toward the Linux operating system:

- COST.
  These Linux systems have much better price/performance ratios than do comparable systems running vendor-branded versions of UNIX (such as AIX).

- VENDOR INDEPENDENCE.
  Using Linux reduces LC dependence on any specific hardware/software vendor, which can allow more flexible installation and service arrangements (such as for network gear, global file systems, or job-control support).

- SOURCE ACCESS.
  Linux can be more easily modified (here, or in collaboration with others) as needed to support such special LC needs as gang scheduling of massively parallel jobs. Although vendors charge for packaging Linux, the source remains open. In fact, during 2002 LC began deploying its own locally modified version of Linux tuned for high-performance computing on many-node, Intel-chip clusters. The local version is called CHAOS, or Cluster High Availability Operating System, and it now runs on all LC Intel-chip clusters, open and secure. CHAOS version 2.0, with further LC-relevant enhancements, deployed during the summer of 2004. CHAOS version 3.0, with the Linux exec-shield (URL: http://www.llnl.gov/LCdocs/chaos/index.jsp?show=s7) security feature enabled, deployed late in 2005.

This Linux Differences guide introduces the Linux resources available at LC (hardware and software). It alerts you to available tools (such as Linux compilers) and also to familiar tools not available under Linux (such as MATLAB). In some cases tool locations (paths) are also different under Linux. The text points out the dozens of UNIX utility programs (such as CP and LS) that either behave differently under Linux (same option, different result), or that have extra options only in the Linux environment. Internal operating-system differences among varieties of UNIX are noted here as well. Finally, a special section discusses support for POSIX threads and for multi-threaded programs under Linux at LC.

OTHER RELEVANT DOCUMENTATION.
For a systematic introduction to the design goals and local enhancements built into the version of Linux that LC is itself developing (named CHAOS), see the additional local document called CHAOS: Linux at Livermore. (URL: http://www.llnl.gov/LCdocs/chaos)

For a locally maintained summary of current technical tips called "Linux Basics," which emphasizes the Fortran compilers, libraries, debuggers, and other Fortran-related tools recommended for use on LC's Linux computers, see this pair of web sites:

```
OCF: http://www.llnl.gov/linux/linux_basics.html
SCF: https://lc.llnl.gov/linux/linux_basics.html
```

An official online source for large amounts of Linux-specific documentation is the web site of the Linux Documentation Project at

```
http://www.ibiblio.org/mdw
```

This is not, however, a user-friendly resource for easily answering questions, but rather a place to download big sets of MAN pages, PDF reference files, or other material that a system administrator might park on a running Linux system for reuse.

Those concerned about improving the usability of Linux (in particular and UNIX in general) may want to consult the analysis at

```
http://developer.gnome.org/projects/gup/ut1_report/report_main.html
```

where Sun Microsystems staff members invoke standard usability principles and the reactions of working technical professionals to evaluate a proposed user-friendly interface for Linux, targeted at nonprogrammers.

In addition, IBM now also acknowledges Linux as an alternative to its own AIX and offers some documentation to help those (such as LC users) who have access to both varieties of UNIX and who want to move more easily between them. IBM reference material is available online (in HTML and PDF formats) through IBM's Redbooks web site.

For a comparative survey of all IBM online manuals relevant to LC (a tiny subset of what they publish), consult LC's <u>IBM Documentation Directory.</u> (URL: http://www.llnl.gov/LCdocs/ibmdir) The manuals most specifically related to Linux at Livermore are:

```
AIX and Linux Interoperability
http://www.redbooks.ibm.com/redbooks/SG246622
```

which focuses on problems of cross-platform data sharing and "user management" (authentication, networking, file-system sharing, etc.), and

```
Linux Handbook
http://www.redbooks.ibm.com/redbooks/SG247000
```

which is more focused on buying and installing Linux system software on various kinds of IBM hardware (not an LC strategy).

LC's Linux Resources

Although once an exotic frontier in high-performance computing, Linux/CHAOS systems now form a central part of LC's computing resources. Over 90% of LC's open (OCF) compute nodes now run Linux/CHAOS, and all of LC's secure (SCF) compute nodes except for Purple/UV/UM/Tempest run Linux/CHAOS as well. Linux clusters at LC vary in their specific features, however, so attention to configuration details may be necessary for you to pick the system most appropriate for your computing needs:

- SWITCH:
  Most LC Linux/CHAOS clusters have a high-speed switch (interconnect) to support communications for massively parallel programs, but some (e.g., YANA) lack such a switch and are best suited for serial computing.

- CHIPS:
  Some LC Linux/CHAOS clusters (ALC, Thunder) rely on Intel chips (usually Xeon or Itanium), while others (Atlas, etc.) use AMD Opteron chips. A few clusters built on IBM (xSeries) chips at LC also run Linux rather than AIX (such as Lilac). This may affect your choice of compilers or libraries as well.

- WORD LENGTH:
  Some LC Linux/CHAOS machines support 32-bit words by default (e.g., ALC), while others support 64-bit words (e.g., Thunder). Software accomodations for other word lengths are often available, however.

- SPONSOR:
  Multiprogrammatic and Institutional Computing (M&IC) is by far the most frequent sponsor of LC Linux/CHAOS clusters, but a few (e.g., Lilac) are ACS supported. This may affect machine scheduling and availablity (especially on weekends).

- SLURM:
  All LC Linux/CHAOS clusters, regardless of the above differences, use the Simple Linux Utility for Resource Management or SLURM (URL: http://www.llnl.gov/LCdocs/slurm) to take the place of IBM's LoadLeveler for underlying job and node management tasks. (Note that in 2006 LC also began *replacing* LoadLeveler with SLURM even on its production AIX machines. So those AIX systems end up with job-control features much more like Linux/CHAOS than like a traditional IBM batch environment.) Moab (URL: http://www.llnl.gov/LCdocs/moab) began slowly replacing LCRM for across-machine job scheduling on LC Linux clusters in early 2007.

For additional details on and comparisons among the currently available Linux/CHAOS machines managed by Livermore Computing, consult the feature-comparison tables (with supporting links to specific technical information for each different cluster) at these LC web sites:

```
OCF: http://www.llnl.gov/computing/hpc/resources
SCF: http://www.llnl.gov/computing/hpc/resources/hardware_conf.html
```

Software Available

Unix Utilities

The usual hundreds of UNIX utilities are also available under Linux, but a few have different names (VI becomes VIM, for example) and quite a few have different (or differently behaving) internal options or features.

See the Within-Tool Differences (page 18) section below for an alphabetical, explanatory list of known differences in tool behavior and available options between IBM's AIX version of UNIX and Linux.

See the next subsection for a list of LC-only or localized utilities and their status on LC Linux systems. Note that Red Hat Linux 7.3, from which LLNL's CHAOS 1.2 derives, stopped supporting web browser Netscape. Consequently, all LC Linux machines that run CHAOS 1.2 or later (deployed gradually in the fall of 2003) offer Mozilla rather than Netscape as the client for web access.

See the Software Locations (page 16) section below for noteworthy differences in location (and hence in pathname) for software tools and libraries under Linux and AIX.

For running batch jobs under Linux, see the comments and references in the Batch and Resource Management (page 41) section below.

LC Utilities

Some well-known and important LC-developed (or LC-adapted) software tools are not available under Linux, or are emerging incrementally. Here is a (preliminary) list of LC-only (or LC-modified) utility programs and their current Linux/CHAOS status:

```
LC Tool       Under Linux?
-------       ------------
 BRLIM         YES

 CHMODSTG      YES
 CHGRPSTG      YES

 DCECP         NO (see LDAPSEARCH)
 DDT           YES
 DIRCOLORS     YES (Linux only, not AIX)

 GIVE          YES (also across machines)
 GUIDE         NO

 HTAR          YES

 JJ            NO
 JS            NO
 JU            NO

 KINIT         YES

 LDAPSEARCH    YES (replaces DCECP)
 LDEDIT        NO
 LFT           YES
 LIB76         YES
 LPOPTIONS     YES (OCF Linux only)
 LSTORAGE      YES

 MATLAB        YES (v7 OCF only,
                   AIX frozen at v6.5)
 MAXJOBS       NO
 MOLE          YES
 MOZILLA       YES

 NETSCAPE      NO (use MOZILLA)
 NFT           YES

 PCSUSAGE      YES
 PDSH          YES
 PHIST         YES

 QUOTA         YES (omits /nfs/tmp* on SCF)

 SCANCEL       YES
 SFTP          YES
 SINFO         YES
 SMAP          YES (BlueGene/L only)
 SPJSTAT       NO  (use SQUEUE instead)
 SQUEUE        YES
 SRUN          YES
 SUPER         YES
```

```
TAKE        YES (also across machines)
TSB         YES

UINFO       YES

WHATSUP     YES (Linux only, not AIX)

XSSH        YES
```

Compilers

Besides the usual CCC (Linux for CC) and C++ compilers, the OCF Linux machines also offer the GNU-project compilers GCC, G++, and G77, all in /usr/local/bin.

On the SCF Linux clusters (such as ACE and QUEEN), LC also offers the corresponding Intel compilers ICC and IFC as well. Because of subtle system, library, and compiler differences, you should recompile all codes before running them under Linux, and you should use the Intel compilers whenever you can.

Compiler Availability

On the various LC Linux clusters, the available compilers on the login nodes can include:

Compiler Name	Directly Invoked By	Invoked By MPI Script(*)
Intel Fortran	ifc	mpiifc
Intel C++	icc	mpiicc
PGI C	pgcc	mpipgcc
PGI C++	pgCC	mpipgCC
PGI HPF	pghpf	mpipghpf
PGI F90	pgf90	mpipgf90
PGI F77	pgf77	mpipgf77
KCC++	/usr/local/KAI/KCC_BASE/bin/KCC	
GCC	see below	mpicc
G77	see below	(none)
PATHF90	/usr/local/tools/pathscale/pathscale/bin/pathf90	
PATHFCC	/usr/local/tools/pathscale/pathscale/bin/pathfCC	

(*)MPI scripts reside in /usr/lib/mpi/bin.

Many alternative GCC-related compilers are available, and you can use these execute lines to invoke the specific GNU-compiler version and variety that best suits your needs (note that under CHAOS 2.0, starting in the summer of 2004, the former GCC-2.96 compiler version is no longer available):

```
/usr/local/bin/gcc
/usr/local/bin/gcc-3.2.3
/usr/local/bin/gcc-3.3.3
/usr/local/bin/g77
/usr/local/bin/g77-3.2.3
/usr/local/bin/g77-3.3.3
/usr/local/bin/g++
/usr/local/bin/g++-3.2.3
/usr/local/bin/g++-3.3.3
```

Compiler Differences

(1) LD_LIBRARY_PATH, --AUTO_RPATH

Environment variable LD_LIBRARY_PATH (on Linux/CHAOS machines only) specifies a colon-delimited, ordered list of directories (pathnames) to search first for *dynamically* linked libraries, before the standard set of Linux library directories. This is especially useful for testing a new, nonstandard, or debug-version of a library. (Dynamically linked libraries are those not incorporated into your binary file at link time, but instead referenced by their location. All have .so suffixes, such as libnew.so.) On AIX systems, environment variable LIBPATH fills the same role as LD_LIBRARY_PATH does under Linux.

Using LD_LIBRARY_PATH correctly on LC machines calls for careful attention to several interdependencies. Invoking the setup script for the Intel compilers (ICC, IFC) on the SCF Linux clusters, namely

```
source /opt/intel/compiler50/ia32/bin/iccvars.sh
```

sets the environment variable LD_LIBRARY_PATH to specific Intel libraries. If you alternate between using ICC and the GNU compiler GCC, you should unset this variable by hand each time you use GCC. Note that sometimes Linux system updates leave LD_LIBRARY_PATH undefined, so your scripts should always test for this possibility before trying to append (or prepend) other path values to whatever this environment variable contains.

Under CHAOS 3.0 (deployed late in 2005), the compiler scripts link with -rpath for the MPI libraries by default. Consequently:

(A) If you previously set LD_LIBRARY_PATH to /usr/lib/mpi/lib or /usr/lib/mpi/mpi_gnu/lib by hand, this step is no longer needed.

(B) If you need "side installs" to handle MPI-vendor bugs, you should link with -rpath=no and then set LD_LIBRARY_PATH to your preferred nondefault directory.

(C) Remember that like all environment variables, LD_LIBRARY_PATH is inherited by all programs called by the one for which you set it. This may sometimes produce unintended errors or problems for child processes.

(D) Remember also that LCRM *unsets* LD_LIBRARY_PATH on the execution machine at the start of every batch job that it manages. If, for any of the above reasons, your batch job needs a specific value in LD_LIBRARY_PATH, your own job script must provide that value. For details, see the "How LCRM Uses Environment Variables" section (URL: http://www.llnl.gov/LCdocs/ev/index.jsp?show=s3.4.2) of LC's Environment Variables user guide.

(E) Starting in December, 2006, LC enabled the LLNL-developed GNU linker (LD) option pair

```
-Wl,--auto_rpath
```

on all (and only) Xeon (e.g., ALC, ACE, QUEEN, LILAC) and Opteron (e.g., Yana, Zeus) Linux/CHAOS production machines (OCF and SCF). The LD options -L*dir* or --library-path=*dir* insert *dir* into the list of search paths that LD searches for libraries *at link time*. The -Wl,--auto_rpath option pair inserts all such -L directories also into the list of search paths that LD searches for dynamically linked libraries *at run time*

(sometimes called the executable's RPATH list). Specifically, invoking --auto_rpath inserts the -L directories into the RPATH list in the following order:

```
SEARCH ORDER for run-time libraries:

1. Directories specified by user or compiler
   with --rpath options.

2. Directories specified by -L and added by
   using -Wl,--auto_rpath.

3. Directories specified in environment variable
   LD_LIBRARY_PATH or in /etc/ld.so.conf.
```

You can check an executable's current ordered RPATH list (or confirm the operation of -Wl,--auto_rpath to expand that list) by using the OBJDUMP utility:

objdump -p *executable* | grep RPATH

(2) VECTORIZATION (INTEL)

Intel Pentium 4 CPUs have 8 128-bit registers that are used as vector registers by the SSE2 (Streaming SIMD Extensions 2) instructions. When code is vectorized, these registers are not just used for floating-point calculations but packed with multiple values prior to an add or multiply step. This packing leaves no room for extra precision in vectorized code. Hence, roundoff differences can and do occur between floating-point results produced from vectorized and from nonvectorized executables built from the same source code. Only you can decide if the precision of Intel vectorized code is sufficient for your needs.

(3) VECTORIZATION (PGI)

The Linux PGI compilers do *not* use the SSE2 vectorization features described above. If your code vectorizes well, therefore, using the Intel compilers can provide significant performance gains over using the PGI compilers. For nonvectorized code, the performance results are usually very much the same.

(4) NONSTANDARD OUTPUT

Compiling with the ICC or IFC compilers without invoking the -mp ("maintain precision") option can produce assembler code that does not conform to IEEE 754 or ANSI C standards (and this nonconformance may alter your floating-point results).

(5) KCC/GCC SLOW DOWN

On LC Linux systems, KCC uses only GCC as its back-end compiler, not PGCC or ICC. Tests at LC suggest that GCC-compiled scientific codes generally run at one-third to one-half the speed of the same code compiled with ICC.

(6) STACK SIZE

Starting in April, 2005, the default stack size on all OCF production Linux clusters (such as MCR) is "unlimited" (the default stack size on LC's AIX machines is 256 Mbyte).

(7) "SECOND UNDERSCORE" PROBLEM

On LC Linux machines the Intel and PGI compilers use one convention for appending underscores to symbol names, while the GNU and Pathscale compilers use a different convention (that often yields two underscores). This incompatibility most often affects Fortran codes linked to libraries compiled with the "opposite" underscore convention. To solve this problem, LC replaced (August, 2006) the g77 and pathf90 compiler executables with wrapper scripts that automatically invoke the -fno-second-underscore option, and replaced the FFTW libraries with new versions also built using -fno-second-underscore (users must recompile their codes to benefit from these changes). If you are sure that you want to override this standing change in a particular case, then invoke the special option -funderscoring when you compile to deactivate -fno-second-underscore.

(8) LIBRARY PRECEDENCE

Starting in November, 2006, links to the libraries PNG and JPEG were moved to /usr/local/tools to eliminate possible conflicting-precedence issues when compiling under Linux. See the Software Locations section (page 16) below for details.

Affinity Program

Compaq (at www.compaq.com/linux) advertises a package of software (an "affinity program") to facilitate switching between Tru64 and Linux versions of UNIX on Compaq machines. LC has NOT purchased and does NOT provide this conversion package on its Linux computers.

NETMON Relevance

LC's locally developed <u>NETMON</u> (URL: http://www.llnl.gov/LCdocs/netmon) web sites (open and secure) list and plot network traffic for several popular "services" (such as FTP and SSH) between pairs of LC network nodes.

Starting in 2006, NETMON began including at least one Linux/CHAOS production cluster (namely, thunder.llnl.gov on the open network) in the set of "perspective nodes" *from* which network traffic is monitored. Relevant NETMON-monitored "host" (target) nodes *to* which traffic is monitored include various network routers, storage.llnl.gov, and again the Linux/CHAOS machine thunder.llnl.gov.

Software Locations

Software available under Linux is sometimes located differently (along a different pathname) than under other versions of UNIX on LC machines. As with availability, Linux software location is an evolving story. Currently important Linux directories include:

/dfs   is *not* mounted on any LC Linux node (but is on all AIX 5.2 (not 5.3) and Tru64 nodes).

/usr/bin   holds the standard Red Hat distribution tools (such as MAKE and NROFF) as well as their GNU-project counterparts (such as GMAKE and GNROFF), along with all of the Linux compilers. WARNING: Starting in 2002, IBM keeps its AIX compilers in /usr/local/bin rather than in /usr/bin (where they reside under Linux), a difference to remember if you invoke compilers with their full pathname (hence, this practice is not recommended).

/usr/local/bin   does not duplicate /usr/local elsewhere because of license restrictions on some software tools. This is the Linux home for TOTALVIEW, MOLE, and YORICK, for example.

/usr/gapps   replaces /usr/apps as the place for user-developed shared software on LC's Linux computers.

/usr/give   is a file system mounted across OCF (only) "production machines" to hold the files that one user GIVEs to another. In general, this enables across-machine GIVEs on OCF (between Blue and GPS, for example). However, although /usr/give is mounted on ILX, the local GIVE and TAKE clients look only in /usr/local/give for files to exchange. So you *cannot* (yet) perform across-machine GIVEs (or TAKEs) to or from open Linux machines.

/usr/tmp   is a small work directory on each node. You should use /nfs/tmp1 or /nfs/tmp3 on SCF nodes, or use /nfs/tmp3 alone on OCF nodes as a better global file system(s) to hold your large working files or output.

/usr/kerberos/bin

contains a special version of FTP on all CHAOS 3.0 or later systems. But for accessing LC's archival storage (HPSS), you should always use /usr/bin/ftp instead. Under some conditions, /usr/kerberos/bin/ftp (but not /usr/bin/ftp) refuses to log you into HPSS or needlessly asks you to "please log in with USER and PASS."

/usr/local/hdf5   contains the available HDF5 I/O library and helper programs. Under Linux, the latest library version is hdf5-1.4.2 and there is no Fortran counterpart to the C and C++ libraries. HDF5 tools (such as H5DUMP) reside in another subdirectory several layers from the top. See the separate <u>I/O Guide for LC</u> (URL: http://www.llnl.gov/LCdocs/ioguide) for a comparison of HDF5 features under AIX and Linux.

/usr/local/lib   contains much less than under AIX or Tru64, but is still the place for the vendor-independent precompiled mathematics and other library files that are available (such as PMATH and SPRNG). See also /usr/local/intel/mkl.

/usr/local/intel/mkl

(on Intel Linux machines only, such as ILX, Adelie, or Emperor) contains Intel's own Math Kernel Library (MKL), a good source for BLAS and LAPACK routines in the Linux environment. The relevant subdirectories are:
/lib/32 (contains the library files),
/include (contains the include files),
/doc (contains vendor documentation files).

/usr/local/tools   contains specialized job-support software for many purposes. Included here are compilers (GNU, Pathscale), math tools (Matlab), general script tools (Perl), and even documentation design aids (Docbook). Access to subdirectories here may be restricted, or may vary by cluster.
Starting in November, 2006, library PNG links from
/usr/local/tools/libpng
to /usr/global/tools/libpng/latest/$SYS_TYPE
and library JPEG links from
/usr/local/tools/libjpeg
to /usr/global/tools/libjpeg/latest/$SYS_TYPE
to eliminate any conflicting precedence choices between the operating system and some compilers when selecting library versions.

Within-Tool Differences

Some software tools or utility programs are available under both Linux and other versions of UNIX (such as IBM's AIX), but the syntax or function of one or more internal options differs by operating-system version. Sometimes the Linux version of a tool offers more options than AIX supports. This section lists the utility programs that have such within-tool (option) differences, and it briefly explains the difference in syntax or function for each option where they diverge. The first subsection below covers the most widely used UNIX tools. The second covers the less common or more specialized tools, tools whose intricate differences defy brief summary here, or tools for root users or system administrators only. (Each set is in alphabetical order.)

Often-Used Tools

These popular software tools differ in their Linux implementation from their implementation under other versions of UNIX (especially IBM's AIX) in the ways summarized below. See the next subsection for more obscure or specialized tools.

at	runs commands at a later time. Same Option, Different Role: -c AIX: runs the job with csh. Linux: cats the job to standard output. Options in Linux ONLY: -d.....is an alias for ATRM. -v.....shows the time the job will run. -V.....prints the version number.
banner	outputs large character strings. Same Option, Different Role: None. Options in Linux ONLY: -w nnn.....uses output width of nnn characters (instead of 80).
cal	displays a calendar. Same Option, Different Role: None. Options in Linux ONLY: -m.....makes Monday the first day of the week. -j.....displays Julian dates (numbered from Jan. 1). -y.....displays a calendar for the current year.
chgrp	changes a file's group. Same Option, Different Role: None. Options in Linux ONLY: -c,-changes.....like verbose, but reports only changes. --reference=rfile.....assigns the group to match rfile's. -v,--verbose.....reports on every file processed.

chmod   changes a file's permissions.
Same Option, Different Role: None.
Options in Linux ONLY:
-c,-changes.....like verbose, but reports only changes.
--reference=rfile.....assigns the permissions to match rfile's.
-v,--verbose.....reports on every file processed.

chown   changes a file's ownership.
Same Option, Different Role: None.
Options in Linux ONLY:
-c,-changes.....like verbose, but reports only changes.
--dereference.....acts on a symbolic link's referent, not on the link.
--reference=rfile.....assigns the ownership to match rfile's.
-v,--verbose.....reports on every file processed.

compress   compresses a file.
Same Option, Different Role: None.
Options in Linux ONLY:
-r.....recursively compresses every file in a specified directory.

cp   copies one file or directory to another.
Same Option, Different Role: None.
Options in Linux ONLY:
-a,--archive.....same as -dpR.
-b,--backup.....makes a backup before removal.
-d,--no-dereference.....preserves links but removes existing destinations.
--help.....lists options.
-l,--link.....links files instead of copying (see -s).
-P,--parents.....appends source path to a directory.
--sparse=when.....controls creation of sparse files.
-s,--symbolic-link.....makes symbolic links instead of copying.
-S,--suffix=sss.....overrides the usual backup suffix.
-u,--update.....copies only when source is newer than sink.
-v,--verbose.....explains what is happening.
-V,--version-control=vvv.....overrides unual version control.
-x,--one-file-system.....stays on current file system.
--version.....outputs version info and ends.

cut   writes selected parts from each line of a file.
Same Option, Different Role: None.
Options in Linux ONLY:
--help.....lists options.
--output-delimiter=sss.....uses string sss to delimit output.
--version.....outputs version info and ends.

date   displays current date and time.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                -d,--date=string.....displays date for string, not now.
                -f,-file=dfile.....same as -d for every line in dfile.
                -I,--iso8601.....outputs an ISO-8601 compliant date and time.
                -r,--reference=rfile.....displays when rfile was last modified.
                -R,--rfc822.....outputs an RFC-822 compliant date and time.
                --help.....displays brief help.
                --version.....displays version info and ends.

df   ("disk free") reports free space on mounted file systems.
                Same Option, Different Role:
                -k
                AIX: reports in 1-kbyte blocks (default is 512-byte blocks).
                Linux: unneeded, 1-kbyte blocks are the default.

du   summarizes your disk usage.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                -b,--bytes.....prints amount used in bytes.
                -c,--total.....prints a grand total.
                -D,--dereference.....reports on symbolic links instead of their referents.
                --exclude=pat.....excludes files that match pat.
                -h,--human-readable.....prints sizes with units (e.g., 2G).
                -H,--si.....same as -h but in powers of 1000, not 1024.
                -L,--dereference.....dereferences all symbolic links.
                -m,--megabytes.....uses a block size of 1048576.
                --max-depth=n.....reports only files n or fewer levels below DU.
                -S,--separate-dirs.....excludes the size of subdirectories.
                -X fff,--exclude-from=fff.....excludes files that match any pattern in fff.
                --help.....displays brief help.
                --version.....displays version info and ends.

echo   writes character strings to standard output.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                -n.....omits the trailing newline character.

env   displays or sets environment variables.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                -u,--unset=name.....removes the named variable from the environment.

expand   expands tabs to spaces in output.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                -i,--initial.....converts tabs only if preceded by white space.
                -t,--tabs=nnn.....converts tabs to nnn spaces, not 8 spaces.

ftp   transfers files to or from a remote host.
                (Under CHAOS 3.0 or later, /usr/kerberos/bin/ftp is available, but always use
                /usr/bin/ftp for the most troublefree transfers to or from HPSS archival storage.)
                Same Option, Different Role:
                AIX: site umask|chmod.
                Linux: chmod.....changes file permissions.
                Linux: umask.....sets remote default umask.
                parallel
                AIX: transfers with AIX *target* are automatically parallel.
                Linux: transfers with Linux *target* need PARALLEL toggle.
                Options in Linux ONLY:
                -e.....disables command editing and history.
                -p.....uses passive mode (if server supports PASV).

head   displays the first few lines of a file.
                Same Option, Different Role: None.
                Options in Linux ONLY:
                --help.....displays an option list.
                -q,--quiet.....never prints headers giving file names.
                -v,--verbose.....always prints headers giving file names.
                --version.....outputs version info and ends.

join   joins the data fields of two files.
                Same Option, Different Role: None
                Options in Linux ONLY:
                --help.....displays brief help message.
                --version.....outputs version info and ends.

kill   signals a running process to stop.
                Same Option, Different Role: None
                Options in Linux ONLY:
                -p.....prints process ID but does not kill.

ldapsearch   reports user and group properties, such as storage groups.
                  Same Option, Different Role: None
                  AIX: uses the complex syntax shown in the storage groups <u>section</u> (URL:
                  http://www.llnl.gov/LCdocs/ezstorage/index.jsp?show=s5.1) of the EZSTORAGE
                  guide.
                  SCF Linux: requires an extra execute-line option, namely
                  -H ldaps://dir-server.llnl.gov
                  to yield the same results as on AIX or OCF Linux.

ln   links files or directories.
                  Same Option, Different Role: None
                  Options in Linux ONLY:
                  -b,--backup.....makes backup of each destination file.
                  -n,--no-dereference.....treats symlink directory targets as normal files.
                  -i,--interactive....prompts whether to remove destinations.
                  -S,--suffix.....overrides the usual backup suffix.
                  -v,--verbose.....prints name of each file before linking.
                  -V,--version-control.....overrides usual version control.
                  --help.....displays the LN man pages.
                  --version.....outputs version info and ends.

lp   sends files to network printers.
                  Same Option, Different Role:
                  -o
                  AIX: offers a handful of output-control suboptions.
                  Linux: offers many complex output-control suboptions from the Common UNIX Print
                  System, the most useful 20 of which are explained in a <u>section</u> (URL:
                  http://www.llnl.gov/LCdocs/ezoutput/index.jsp?show=2.4.4) of the EZOUTPUT
                  guide.

lpr   sends files to network printers.
                  Same Option, Different Role:
                  -o
                  AIX: offers a handful of output-control suboptions.
                  Linux: offers many complex output-control suboptions from the Common UNIX Print
                  System, the most useful 20 of which are explained in a <u>section</u> (URL:
                  http://www.llnl.gov/LCdocs/ezoutput/index.jsp?show=2.4.4) of the EZOUTPUT
                  guide.

ls   lists file in a directory.
              Same Option, Different Role: None.
              Options in Linux ONLY:
              --block-size=nnn.....uses nnn-byte blocks.
              -B,--ignore-backups.....skips implied entries ending with tilde.
              --color=never|always|auto.....specifies when color distinguishes file types.
              -D,--dired.....outputs in Emacs dired mode.
              --format=across|commas|long|vertical.....same as -x, -m, -l, -C.
              --full-time.....lists full date and full time.
              -G,--no-group.....supresses group information.
              -h,--human-readable.....prints sizes with units (e.g., 2G).
              -H,--si.....same as -h but uses powers of 1000, not 1024.
              --indicator=none|classify|file-type.....appends specified indicator.
              -I,--ignore=pattern.....skips entries matching pattern.
              -k,--kilobytes.....same as --block-size=1024.
              -Q,--quote-name.....encloses entry names in double quotes.
              --quoting-style=literal|shell-always|c|escape....specifies quoting style.
              -S.....sorts files by size.
              --sort=extension|none|size|time|version|status|use....specifies sorting.
              --time=access|status|use.....changes the time displayed.
              -T,--tabsize=cols.....sets tab stops at cols instead of 8.
              -U.....lists entries unsorted, in directory order.
              -v.....sorts files by version.
              -w,--width=cols.....sets screen width to cols.
              -X.....sorts alphabetically by file extension.
              --help.....shows MAN pages for ls.
              --version.....shows version info and ends.

mail   sends and receives e-mail.
              Same Option, Different Role: None.
              Options in Linux ONLY:
              -b.....sends blind carbon copies to a comma-separated list.
              -I.....runs interactively even when input is not a terminal.

man   displays "manual" pages online.
               Same Option, Different Role:
               -t
               AIX: formats the output with TROFF.
               Linux: formats the output with /usr/bin/groff -Tps -mandoc.
               Options in Linux ONLY:
               -a.....displays ALL man pages that match request.
               -C cfile.....specifies a configuration file other than /etc/man.config.
               -d.....shows debug info instead of man pages.
               -D.....shows debug info AND man pages.
               -F,--preformat.....formats output but does not display.
               -h.....prints a one-line help message.
               -K.....searchs ALL man pages for string.
               -m mmm.....searches an alternate set of man pages specified by mmm.
               -p ppp.....specifies which preprocessors to run if their code appears in ppp: eqn(e),
               grap(g), pic(p), tbl(t), vgrind(v), refer(r).
               -P pager.....specifies a pager instead of /usr/bin/less-is.
               -w,--path.....displays the location of man pages instead of their content.
               -W.....same as -w but output one per line.

more   shows text one screen at a time.
               Same Option, Different Role:
               -l
               AIX: pauses display after each page break.
               Linux: pauses display after each form feed.
               -p
               AIX: executes a specified subcommand.
               Linux: avoids scroll, clears screen and shows text.
               Options in Linux ONLY:
               -f.....counts logical, not screen, lines (no folding).
               +num.....starts display as line num.
               +/.....searches for a specified string.

mv   moves one file or directory to another.
               Same Option, Different Role: None.
               Options in Linux ONLY:
               -b,--backup.....makes a backup before removal.
               -S,--suffix=sss.....overrides the usual backup suffix.
               -u,--update.....moves only when source is newer than sink.
               -v,--verbose.....explains what is happening.
               -V,--version-control=vvv.....overrides unual version control.

pr   writes a file to standard output.

Same Option, Different Role: None.

Options in Linux ONLY:

-c,--show-control-chars.....uses hat (up-arrow) notation.

-J,--join-lines.....merges full lines, disables -W.

-l,--length=ppp.....sets page length to ppp lines.

-N,--first-line-number=nnn.....counts from nnn as first line, first page.

-S[sss],--sep-string=sss.....separates columns using string sss.

-v,--show-nonprinting.....uses octal backslash notation.

ps   shows current status of processes.

Same Option, Different Role:

-c

AIX: reports only on specified processes classes.

Linux: changes scheduler info for -l.

-s

AIX: displays size of kernel stack.

Linux: displays signal format.

-v

AIX: displays 8 special fields.

Linux: displays virtual memory format.

Options in Linux ONLY:

-A.....shows all processes.

-C.....selects by command name.

--cols,--columns.....sets screen width.

--cumulative.....includes some dead child processes.

-g,--group,--Group.....selects by group.

-H,--forest.....shows process hierarchy.

-h,--no-header.....omits header lines.

--headers.....repeats header lines.

--help.....prints help message.

--info.....prints debugging information.

-j.....gives extra job-control details.

--lines,--rows.....sets screen height.

--nul,--null.....outputs unjustified.

-o,-O.....allows user-defined output format.

-r.....reports running processes only.

--sort.....specifies sorting order.

--User.....selects by real user name.

-V,--version.....prints version number.

--width.....same as --cols.

-y.....does not show flags.

--zero.....same as --null.

rm removes files or directories.
 Same Option, Different Role: None.
 Options in Linux ONLY:
 -r.....recursive removal, always works for Linux, fails when path exceeds PATH_MAX
 under AIX.

sort sorts files in specified ways.
 Same Option, Different Role: None.
 Options in Linux ONLY:
 -g.....compares by numerical value (implies -b).
 -M.....compares by date (implies -b).
 -s.....stabilizes a sort by disabling last resort comparison.

split divides a file into pieces.
 Same Option, Different Role: None.
 Options in Linux ONLY:
 -C,--line-bytes=sss.....puts at most sss bytes of lines per output file.

sum displays checksum and block count of a file.
 Same Option, Different Role: None.
 Options in Linux ONLY:
 -s,--sysv.....uses 512-byte blocks (System V algorithm).

tail displays the last few lines of a file.
 Same Option, Different Role: None.
 Options in Linux ONLY:
 --help.....displays an option list.
 -q,--quiet.....never prints headers giving file names.
 --retry.....keeps trying to open an inaccessible file.
 -v,--verbose.....always prints headers giving file names.
 --version.....outputs version info and ends.

tar   makes and manages file archives.

Same Option, Different Role:

-F

AIX: checks file type before archiving.

Linux: runs script at end of each tape.

-L

AIX: writes specified directories NOT recursively.

Linux: changes tape after writing N*1024 bytes.

-l

AIX: reports errors if links not all archived.

Linux: stays within the local file system.

-N

AIX: allows large block clusters (default: 20).

Linux: only stores files newer than DATE (--newer).

-s

AIX: tries to create a symbolic link.

Linux: sorts list of names to extract (--same-order).

-S

AIX: specifies using 512-kb blocks/volume.

Linux: handles sparse file efficiently (--sparse).

Options in Linux ONLY:

-A,--concatenate.....appends files to archive.

--atime-preserve.....keeps access times on dumped files.

--block-compress.....blocks compressed output for tape.

--checkpoint.....prints dir names while reading archive.

-d,--diff,--compare.....finds archive/file-system differences.

--delete.....deletes from an archive.

--exclude fff.....omits named file fff.

--force-local.....always keeps archive on local disk.

-G,--incremental.....invokes old GNU-format backup.

-g,listed-incremental.....invokes new GNU-format backup.

--ignore-failed-read.....ignores unreadable files.

-k,--keep-old-files.....does NOT overwrite existing files from acrhive.

-K,--starting-file fff.....starts at fff in archive.

-M,--multi-volume.....signals a multi-volume archive.

-O,--to-stdout.....extracts files to standard output.

-P,--absolute-paths.....leaves leading /s on file names.

-R,--record-number.....shows each within-archive record number.

--remove-files.....removes files after adding to archive.

--same-owner.....keeps ownership on extracted files.

-T,--files-from fff.....gets file names from file fff.

--use-compress-program ppp.....filters archive through ppp.

-V,--label nnn.....calls archive volume nnn.

--version.....prints version information.

-W,--verify.....trys to verify the archive after writing it.

-X,--exlude-from fff.....excludes all files listed in fff.

-Z,--compress,--uncompress.....filters archive through COMPRESS.
-z,--gzip,--unzip.....filters archive through GZIP.

telnet   connects interactively to a remote host.
Same Option, Different Role:
-e
AIX: specifies a terminal type (vt100, 3270).
Linux: specifies an escape character.
Options in Linux ONLY:
-8.....specifies 8-bit data path on input (see -L).
-a.....trys automatic login with current user name.
-b alias.....binds local socket to aliased address.
-c.....disables reading the user's .telnetrc file.
-E.....eliminates any escape character.
-K.....specifies no automatic login to remote host.
-L.....specifies 8-bit data path on output (see -8).
-r.....specifies user interface like RLOGIN.
-S.....sets the IP type-of-service option.
-X atype.....disables atype authentication.
-x.....enables encryption of data stream if possible.

time   prints a command's execution time.
Same Option, Different Role: None.
Options in Linux ONLY:
TIME is a system call, not a command, under Linux.

tr   translates characters.
Same Option, Different Role: None.
Options in Linux ONLY:
-t.....first truncates string1 to length of string2.

type   describes a specified command.
Same Option, Different Role: None.
Options in Linux ONLY:
--all ccc.....prints all places that contain ccc.
--path ccc.....prints the name of disk file that ccc executes.
--type ccc.....prints one of: alias, keyword, function, builtin, file.

uname   reports system name and details.
Same Option, Different Role:
-a
AIX: reports same as -mnrsv.
Linux: reports all information.
Options in Linux ONLY:
--help.....displays a help message.
--version.....displays version info and ends.

uncompress	uncompresses a file. Same Option, Different Role: None. Options in Linux ONLY: -r.....recursively uncompresses every file in a specified directory.
uniq	deletes duplicated lines in a file. Same Option, Different Role: None. Options in Linux ONLY: -D,--all-repeated.....prints all duplicate lines. -i,--ignore-case.....ignores case differences when comparing. -w,--check-chars=nnn.....compares no more than nnn characters/line.
userinfo	opens a separate window and reports the full name, office telephone, and default shell of the user who runs it. NOTE: LC's local tool UINFO avoids the separate window and reports the same information plus numerical UID, home directory, group memberships, and bank memberships (for any specified user).
vi|vim	edits files with full-screen display. Same Option, Different Role: AIX: executed by vi [options][file] Linux: executed by vim [options][file] Options in Linux ONLY: [many and complex]
wc	counts words, lines, bytes in a file. Same Option, Different Role: None. Options in Linux ONLY: -L,--max-line-length.....reports length of the longest line.
which	locates files under C shell. Same Option, Different Role: None. Options in Linux ONLY: -a,--all.....prints all matching files, not just first. -i,--read-alias.....reads aliases from stdin, reports matches on stdout. --show-dot.....prints ./match, not full path. --show-tilde.....outputs tilde when directory matches HOME. --skip-alias.....ignores --read-alias. --skip-dot.....ignores directories that start with dot. --skip-tilde.....ingores directories that start with tilde. --tty-only.....stops processing if not on tty. --version.....prints version info and ends.

who   reports currently logged-in users.
         Same Option, Different Role:
         -l
         AIX: lists only LOGIN processes.
         Linux: lists originating machine for each user.
         Options in Linux ONLY:
         --help.....displays a help message.
         --version.....displays version info and ends.

Specialized Tools

The software tools listed here have known internal (option) differences in implementation between Linux and other versions of UNIX, but those differences are very specialized, too complex to usefully summarize here, or relevant only to root users or system administators. If you know these tools elsewhere and plan to use them under Linux, consult one of the online files below to see how IBM explains their complex differences under AIX and Linux:

    http://ftp.software.ibm.com/aix/freeSoftware/aixtools/docs/cmdsdiff_a_n.html
    http://ftp.software.ibm.com/aix/freeSoftware/aixtools/docs/cmdsdiff_o_z.html

See the previous subsection for summarized, option-level differences in the most common, often-used UNIX software tools.

```
ac
ar
atq
atrm
awk

bc
bsh

chroot
cmp
cpoi
crontab
csplit
ctags
cu

dd
diff
diff3

ed

file
find
fortune
fuser

gprof
grep

halt

id
indent
init
install
ipcs

jobs

killall
```

```
ksh

last
lastcomm
lex
logger
look


make
man
mknod
mount
mt

passwd
patch

rdist
reboot
red
restore
rmt
rsh

sa
sdiff
sed
sh
shutdown
strings
stty

telinit
test

ulimit
units

vmstat

w

xargs

yacc
```

Operating System Differences

Operating system kernels, including UNIX, have many features to manage a computer's hardware on behalf of user (application) software. Most kernel features are shared among most varieties of modern UNIX, but a few are not. Relevant here are the specific features that Linux lacks but that other UNIX varieties used at LC, such as AIX or Tru64, usually possess.

This table, based on a comparative analysis of Linux kernel features published by Cesar de Silva in June, 2001, for the most recent Linux version at that time, shows which widely shared features Linux lacks.

Kernel Feature	Linux	AIX	Tru64
Dynamic Processor Resilience: isolates faulty CPUs, continues without them.	NO	YES	YES
Dynamic Memory Resilience: isolates faulty memory areas with single-bit errors.	NO	YES	YES
Dynamic Page Sizing: uses variable page sizes for I/O when file sizes vary.	NO	NO	YES
Access Control List (ACL): assigns each file and directory an access list to control who has what type of access.	NO	YES	YES
CacheFS: improves efficiency of NFS-mounted disk access by caching.	NO	YES	NO
IPsec: provides standard security for network or packet layer during communications.	NO	YES	YES

Note also that CHAOS 2.0 used the XFree86 X-window system, while CHAOS 3.0 (deployed late in 2005) uses Xorg instead. This is because XFree86 is incompatible with the CHAOS 3.0 exec-shield security feature (see the CHAOS Manual (URL: http://www.llnl.gov/LCdocs/chaos/index.jsp?show=s7)).

POSIX Threads Under Linux

Besides allowing execution of multiple processes at the same time, Linux (as with most other versions of UNIX) now supports multiple threads (independent flows of control) within a single process (all sharing the same process resources, such as address space and files). Communication among threads is much easier that among separate processes, but thread synchronization to avoid trouble demands careful planning.

The Red Hat Linux (CHAOS) systems at LC support standard POSIX 1003.lc threads (usually called "pthreads"). While originally used with a native batch system (RMS) that was *not* thread aware, CHAOS now supports a locally designed, thread-aware resource manager called SLURM (details below).

GENERAL BACKGROUND.
You should consult the POSIX Threads Overview (For LC) (URL: http://www.llnl.gov/LCdocs/pthreads), a locally maintained introductory guide to pthreads use here, if you need primarily:

- A conceptual review of process/thread relationships,

- A summary of general pthreads benefits and pitfalls, with special attention to five different thread-synchronization alternatives (and cross references to code examples elsewhere), or

- A comparison of different vendor implementations of pthreads on LC machines (including comparative system defaults for AIX and Linux).

LINUX THREADS TUTORIAL.
You should consult the POSIX Threads Tutorial (URL: http://www.math.arizona.edu/swig/pthreads/threads.html) by Mark Hays (at

```
http://www.math.arizona.edu/swig/pthreads/threads.html)
```

if you need primarily a tour of the most common pthreads issues, routines, and design techniquues with C-language examples that have been tested exclusively on Linux systems (using the glibc compiler and the LinuxThreads library). The emphasis here is on general pthreads topics that happen to be shown by Linux-tested cases, *not* on the specific Linux resources that support pthreads (or on how pthreads under Linux differ from pthreads elsewhere).

EXCLUSIVE LINUX/THREADS ISSUES.
You should consult The LinuxThreads Library (URL: http://pauillac.inria.fr/~xleroy/linuxthreads) by Xavier LeRoy and colleagues (at http://pauillac.inria.fr/~xleroy/linuxthreads) if you need primarily:

- Help with the glibc (GNU) compiler and links to background on compiling pthreads specifically under Linux.

- Help with the gdb thread-aware debugger and links to background on debugging pthreads codes under Linux.

- Information on (getting or) using the LinuxThreads library that is already on all LC Red Hat Linux systems, including a frequently asked questions file and links to several emulators that run on top of LinuxThreads. LinuxThreads implements standard POSIX threads for Linux.

SLURM.

The Simple Linux Utility for Resource Management (SLURM) has replaced RMS as the low-level job-control scheme on all LC Linux clusters. The SLURM tool that allocates resources and launches parallel jobs is called SRUN. SRUN now offers several options intended to help you properly assign compute resources to pthreads applications. SRUN's -c (lowercase) option lets you specify multiple CPUs per task, while -T (uppercase) lets you specify the number of threads per job. See the SRUN sections of the SLURM Reference Manual (URL: http://www.llnl.gov/LCdocs/slurm) for details on the complex ways in which SRUN options (these two and others) can interact. (Note also that as LC gradually *replaces* LoadLeveler with SLURM even on its production AIX systems, pthreads support there will become much more like it is under Linux/CHAOS.)

NATIVE THREADS.

CHAOS version 2.0, deployed during the summer of 2004, also supports a "native threading model." This thread-support alternative can increase possible parallelism (by allowing threads to migrate to differernt CPUs in a multiprocessor), but using it sometimes reduces code portability.

Environment Variables

You should set these environment variables (with SETENV for the C shell or EXPORT for the SH or KSH shells) when you run parallel batch jobs on LC Linux systems. (On BlueGene/L only, you must pass environment variable values to MPIRUN-executed application programs by using MPIRUN's own -env option, whose argument is a quoted, blank-delimited string of *variable=value* equations.) See the POE User Guide (URL: http://www.llnl.gov/LCdocs/poe) for corresponding IBM environment variables relevant to MPI jobs. (All those that mention LIBELAN support your program's use of Quadrics Elan, a library of low-level functions that provides a message-passing interface for the local network of Linux nodes.) See the "Environment Variables for Batch Jobs" section of the LCRM (DPCS) Reference Manual (URL: http://www.llnl.gov/LCdocs/dpcs) for a helpful set of environment variables (all called PSUB_*name*) that LCRM supports in *every* LCRM-managed production environment to provide a uniform context for batch runs among AIX and Linux systems. NOTE: all Elan-related environment variables have no role on the open-network ILX cluster, because it has no Quadrics fast interconnect.

The CHAOS Simple Linux Utility for Resource Management (SLURM) uses its own additional environment variables (all with names that begin with "SLURM_") to store resource-allocation values for the jobs that SLURM manages. See the "Environment Variables" section of the SLURM Reference Manual (URL: http://www.llnl.gov/LCdocs/slurm) for an explanatory list. A general, comparative analysis of the roles that environment variables play on LC machines appears in the Environment Variables (URL: http://www.llnl.gov/LCdocs/ev) user guide.

BGL_APP_L1_SWOA

> (BG/L only) controls overwriting the L1 cache memory. On BG/L, each CPU has a fast, 32-kbyte local cache (L1), and both CPUs on a chip share a larger but slower prefetch cache (L2) as well as another 4-Mbyte cache (L3). BGL_APP_L1_SWOA toggles the "store-without-allocate" feature of the L1 cache as follows:

> 0   (default, disables SWOA) instructs each CPU to load the cache line into L1 and perform a read-modify-update. This sacrifices the former contents of L1.

> 1   (enables store-without-allocate, SWOA) loads the cache line into the L3 cache instead of L1. L3 performs the read-modify-update and so this preserves the contents of L1. Codes that depend heavily on the use of L1 sometimes perform much better with SWOA enabled.

BGL_APP_L1_WRITE_THROUGH

> (BG/L only) controls writethrough (immediate backend copying) of the L1 cache. Jobs on BG/L that encounter a parity error in the 32-kbyte L1 cache (that is local to each CPU) promptly die. At LLNL, full-system runs have only about a 6-hour mean time to failure from such L1 errors. BGL_APP_L1_WRITE_THROUGH toggles your response to this problem as follows:

0   (default) disables L1 writethrough (copying waits for a threshold so restoring L1 from a copy usually fails). In this case, every L1 parity error kills your job.

1   enables L1 writethrough for the current job. This can be inefficient (it often incurs a 10% to 40% performance penalty), but it virtually eliminates machine stops and job death when L1 parity errors occur.

BGLMPI_COLLECTIVE_DISABLE

(BG/L only) controls which implementation is used for eight MPI collective operations (ALLGATHER, ALLGATHERV, ALLREDUCE, ALLTOALL, ALLTOALLV, BARRIER, BCAST, and REDUCE) managed *as a set*. The choices are:

0   (default) uses optimized collective routines for all eight operations (usually improves application performance).

1   disables optimization and uses MPICH routines instead for all eight operations. This can let you work around unexpected errors that seem related to collective operations, but with a performance penalty. To change implementation details for any *single* specific collective routine on BG/L, use one of the eight specific environment variables (with a shared syntax) described at BGLMPI_*COLLECTIVE* below.

BGLMPI_*COLLECTIVE*

(BG/L only) represents eight environment variables (listed below) with a common syntax whose role is to each control which implementation is used for one of eight MPI collective operations on BG/L (to toggle all eight collectives *at once*, use BGLMPI_COLLECTIVE_DISABLE above). Each BGLMPI_*COLLECTIVE* environment variable can be used in three ways:

(1) Disable Optimization.
Set the variable's value to MPICH to disable the corresponding optimized collective routine and instead invoke the "safe" MPICH algorithm for that collective operation (same effect individually that BGLMPI_COLLECTIVE_DISABLE=1 has globally).

(2) Specify a Network.
Set the variable's value to the BG/L network that you prefer for this collective operation (for example, BGLMPI_BCAST=TREE). Available network choices are GI (global interrupt), TREE, and TORUS (but not all values work for all eight environment variables). The default network for each collective operation comes first in the list assigned to it below. Some collectives accept an optional "packet inject parameter" (whose default is 3) for the TORUS value.

(3) Provide a Failover Sequence.
Leave the variable's value blank (unset) and BG/L will try each alternative *in order* in the comma-delimited list below associated with each of the eight MPI collective

operations (for example, for ALLREDUCE the failover sequence will be first TREE, then TORUS, then MPICH).

The eight BG/L collective environment variables and the associated possible values for each one are:

BGLMPI_ALLGATHER={TORUS:3,MPICH}

BGLMPI_ALLGATHERV={TORUS:3,MPICH}

BGLMPI_ALLREDUCE={TREE,TORUS,MPICH}

BGLMPI_ALLTOALL={TORUS:3,MPICH}

BGLMPI_ALLTOALLV={TORUS:3,MPICH}

BGLMPI_BARRIER={GI,TREE,TORUS,MPICH}

BGLMPI_BCAST={TREE,TORUS}

BGLMPI_REDUCE={TREE,TORUS,MPICH}

BGLMPI_EAGER

(BG/L only; default is 1000) specifies (with an integer value *nnnn*) the size in bytes of the interprocess message at which BG/L switches from using the *eager* to the *rendezvous* protocol. The eager protocol sends data asynchronously and immediately to the destination (fast but sometimes problem prone), while the rendezvous protocol sends data only when requested by the destination. By default, messages sized less than or equal to 1000 bytes use the eager protocol, and messages sized 1001 bytes or greater use the rendezvous protocol.

BGLMPI_MAPPING

(BG/L only) controls the mapping of MPI processes to physical CPUs to make the most of BG/L's unusual features for point-to-point communication. The two mapping alternatives are:

XYZT   (default) maps the MPI tasks to the first CPU on each node using (x,y,z) order (for co-processor mode, where only one CPU/node computes), and then uses the same order for assigning tasks to the second CPU on each node (for virtual-node mode, where both CPUs compute).

TXYZ	maps the MPI tasks in consecutive pairs to the two CPUs on the first node, then by pairs to each other node in (x,y,z) order (so that task0 and task1 go to the two CPUs on node0, task2 and task3 go to the two CPUs on node1, etc.). This method guarantees two tasks/node in virtual-node mode, sometimes a performance enhancer.

BGLMPI_PACING

(BG/L only) toggles the use of flow control (packet pacing) for interprocess messages larger than 1000 bytes (BG/L "rendezvous messages," see also BGLMPI_EAGER). MPI applications that heavily use collectives (such as LINPACK) seldom benefit from packet pacing, but applications that send large messages point to point may see big performance gains. The choices are:

YES	(default) enables packet pacing with a "pacing window" size of 16.
NO	disables packet pacing.

BGLMPI_RVZ (BG/L only) is the same as BGLMPI_EAGER, explained <u>above</u> (page 38).

BGLMPI_RZV (BG/L only) is the same as BGLMPI_EAGER, explained <u>above</u> (page 38).

LIBELAN_GALLOC_EBASE

(default value: 0xb0000000) resizes the Elan global memory heap for MPI collective operations. EBASE is a pointer to a base virtual address in Elan memory to be used for the global heap. (Set this variable and the next two if you use MPI collectives, such as REDUCE, GATHER, SCATTER, or their ALL versions, with more than about 100 processes.)

LIBELAN_GALLOC_MBASE

(default value: 0xb0000000) resizes the Elan global memory heap for MPI collective operations. MBASE is a pointer to the main memory base in Elan memory to be used for the global heap.

LIBELAN_GALLOC_SIZE

(default value: 16777216) resizes the Elan global memory heap for MPI collective operations. SIZE is the size in bytes of the Elan global heap.

LIBELAN_WAITTYPE

(suggested value: POLL) specifies how a blocking MPI process will share computing resources (comparable to MP_WAIT_MODE under POE on IBM machines). Possible values are:

POLL	(default) has the receiving thread actively poll for incoming messages. Use this choice for all MPI jobs on clusters that have a Quadrics interconnect.

SLEEP has the receiving thread sleep and thus remove itself from the active dispatching queue.

YIELD has the receiving thread stay in the queue but yield the processor if it has no work to do.

MALLOC_TRIM_THRESHOLD

(suggested value: -1) see the next item for joint use.

MALLOC_MMAP_MAX

(suggested value: 0) when combined, MALLOC_TRIM_THRESHOLD and MALLOC_MMAP_MAX force MALLOC to use SBRK() rather than MMAP() to allocate memory. This improves performance, but it may reduce the total amount of memory available to your user processes (to no more than 1 Gbyte/process).

MPI_USE_LIBELAN

toggles the Elan library optimizations. Possible values and their roles are:

1 (default) enables the Elan library optimizations.

0 disables the Elan library optimizations. Use this only for debugging, if you suspect problems with the Elan libraries themselves.

OMP_NUM_THREADS

controls the number of threads spawned by Intel's Math Kernel Library (MKL) routines when you invoke this threaded library on any of LC's Intel Linux machines (such as MCR, ALC, or ACE). By default, MKL sets the number of threads equal to the number of processors.

Batch and Resource Management Systems

BATCH SYSTEMS.

LC's Livermore Computing Resource Management (LCRM, formerly DPCS) system is periodically modified to provide enhanced support for parallel jobs on local Linux clusters. Among the relevant enhancements are:

TBS

(Trivial Batch System) has replaced NQS as the underlying batch system for managing queued and competing batch jobs on Linux systems *without a switch*. The former NQS (Network Queueing System) in no longer used as the native batch system on any LC Linux machines.

SLURM

(Simple Linux Utility for Resource Management) is a collaborative project between LLNL and vendor Linux NetworX to manage "compute resources" (available nodes, time to run on them, interconnect access) in the face of conflicting requests for those resources in Linux environments. SLURM replaces RMS on Linux systems *with a switch*, including those with exotic architectures such as BlueGene/L.

In 2006 LC began *replacing* LoadLeveler with SLURM even on its AIX production systems, thus making the batch environment with POE under AIX much more like that under Linux/CHAOS.

Once your SLURM-managed batch job starts to run on some LC machine's compute nodes, you have been allowed to log in to those nodes to execute additional processes (usually ones for monitoring or interactively guiding the batch job). When the batch job ended, these extra login sessions and nonbatch processes were allowed to continue. If these tasks were CPU intensive, however, they sometimes caused problems for subsequent batch runs by other users of the same compute nodes. Starting in August, 2007, on LC Linux (but *not* AIX) clusters where SLURM is the resource manager underlying LCRM or Moab, *all* of a user's processes on any compute node will terminate whenever a SLURM-managed batch job completes on that node. This guarantees that the next user's job will see no interference from stray processes that accompanied your job. If your login session or interactive process on a Linux compute node unexpectedly ends, your batch job to which that node was allocated has probably just completed.

See also the SLURM section of the local manual called <u>CHAOS: Linux at Livermore</u> (URL: http://www.llnl.gov/LCdocs/chaos) and the separate <u>SLURM Reference Manual</u> (URL: http://www.llnl.gov/LCdocs/slurm). Job-control features customized to take advantage of TBS and SLURM (such as special PSUB options) are discussed in the <u>EZJOBCONTROL</u> (URL: http://www.llnl.gov/LCdocs/ezjob) basic guide. Note that the PSUB -g ("geometry") option is only implemented on AIX machines (regardless of whether their underlying batch system is LoadLeveler or SLURM), but does not work on Linux/CHAOS machines.

In the spring of 2007, Moab began gradually replacing LCRM as the across-cluster job scheduler on LC Linux machines (starting with Atlas). See Moab at LC (URL: http://www.llnl.gov/LCdocs/moab) for deployment details and contrasts with how LCRM behaves on AIX machines.

SCF JOB DOMAINS.
In September, 2005, LCRM divided all SCF (secure-network) production machines into two disjoint job-scheduling domains. Interestingly, these domains are based *not* on operating system (AIX or Linux/CHAOS) but instead on underlying batch system (IBM's LoadLeveler or LC's SLURM, described above). All machines in the LoadLeveler LCRM domain also run AIX, as one would expect (and when used in that domain, PSUB and PSTAT apply only to jobs running under AIX). But while most machines in the SLURM LCRM domain also run Linux/CHAOS, PU is an exception: it is an AIX machine that nevertheless manages its jobs using SLURM. So you cannot assume that SLRUM job management entails a Linux/CHAOS environment, and you cannot submit or track PU jobs from the LCRM LoadLeveler domain on SCF. Starting in 2006, SLURM began spreading to all SCF AIX 5.3 machines.

Parallel File Systems

LC deploys parallel file systems on both its IBM/AIX and its Linux/CHAOS clusters to support efficient parallel I/O:

- Access to relevant files from every compute (and login) node in the cluster, even when there are thousands of nodes (but *not* from other clusters),

- Good use of the available network switch and other techniques to handle very large files quickly and reliably, and

- The ability of different nodes to read from or write to different parts of a single file at the same time without corrupting the file.

LC uses IBM's proprietary GPFS to provide parallel file service on AIX machines, but it uses an open-source, collaboratively developed alternative called Lustre to provide parallel file service on its Linux/CHAOS machines.

Lustre's design and implementation at LC are rather different from that used by IBM for GPFS. So while some user aspects of GPFS and Lustre appear the same (such as directory names, portability aliases, and purge policies), other user aspects diverge significantly. Only Lustre, for example, uses logical "object storage targets" (OSTs) to hide the actual location of files from parallel application programs...a move sometimes beneficial and sometimes confusing. Lustre and GPFS also interact differently with the MPI (message passing interface) I/O library, sometimes giving wildly divergent performance results.

To see a comparative analysis of the features, behavior, and pitfalls of the Lustre parallel file system, consult the I/O Guide for LC (URL: http://www.llnl.gov/LCdocs/ioguide/index.jsp?show=s7). Also included there is a cluster-by-cluster comparison of currently deployed Lustre file-system properties on LC's Linux/CHAOS machines.

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government thereof, and shall not be used for advertising or product endorsement purposes.

(C) Copyright 2007 The Regents of the University of California. All rights reserved.

Keyword Index

To see an alphabetical list of keywords for this document, consult the

```
Keyword                      Description
-------                      -----------
entire                       This entire document.
title                        The name of this document.
scope                        Topics covered in this document.
availability                 Applies to Linux at LC.
who                          Who to contact for assistance.

introduction                 Linux at LC introduced.
  linux-overview             Linux background, role at LC.
  linux-resources            Comparison of LC Linux machines.

software                     Software present, absent under Linux.
  unix-utilities             Standard UNIX tools under Linux.
  lc-utilities               LC-only or localized tools.
  compilers                  Compilers on LC Linux machines.
    compiler-availability    Compiler names and paths.
    compiler-differences     Compiler complexities, unusual features.
  affinity                   Compaq Linux affinity program.
  netmon                     NETMON's relevance to Linux nodes.

locations                    Software paths under Linux.

tool-options                 Within-tool option, feature differences.
  often-used-tools           Within-tool differences (common).
  specialized-tools          Within-tool differences (exotic).

os-differences               Basic Linux kernel differences.

pthreads                     POSIX threads under Linux.

environment-variables        Linux-relevant env. variables.

batch-system                 Linux-relevant batch, resource systems.

parallel-file-systems        Lustre parallel file system for Linux.

index                        The structural index of keywords.
a                            The alphabetical index of keywords.
date                         The latest changes to this document.
revisions                    The complete revision history.
```

Alphabetical List of Keywords

```
Keyword                    Description
-------                    -----------
a                          The alphabetical index of keywords.
affinity                   Compaq Linux affinity program.
availability               Applies to Linux at LC.
batch-system               Linux-relevant batch, resource systems.
compilers                  Compilers on LC Linux machines.
compiler-availability      Compiler names and paths.
compiler-differences       Compiler complexities, unusual features.
date                       The latest changes to this document.
entire                     This entire document.
environment-variables      Linux-relevant env. variables.
index                      The structural index of keywords.
introduction               Linux at LC introduced.
lc-utilities               LC-only or localized tools.
linux-overview             Linux background, role at LC.
linux-resources            Comparison of LC Linux machines.
locations                  Software paths under Linux.
netmon                     NETMON's relevance to Linux nodes.
often-used-tools           Within-tool differences (common).
os-differences             Basic Linux kernel differences.
parallel-file-systems      Lustre parallel file system for Linux.
pthreads                   POSIX threads under Linux.
revisions                  The complete revision history.
scope                      Topics covered in this document.
software                   Software present, absent under Linux.
specialized-tools          Within-tool differences (exotic).
title                      The name of this document.
tool-options               Within-tool option, feature differences.
unix-utilities             Standard UNIX tools under Linux.
who                        Who to contact for assistance.
```

Date and Revisions

```
Revision    Keyword        Description of
Date        Affected       Change
--------    --------       ------
05Sep07     batch-system   New compute node use policy.
                           Moab spreads on Linux machines.
            linux-resources
                           Many details and comparisons updated.

04Jan07     netmon         Linux/CHAOS nodes now tracked by NETMON.
            compilers      Section subdivided for better access.
            compiler-differences
                           LD's --auto_rpath option explained.
            index          New keywords for new sections.

08Nov06     locations      PNG, JPEG libraries now moved to
                           /usr/local/tools.
            often-used-tools
                           -H needed for SCF Linux LDAPSEARCH.

06Sep06     locations      /usr/local/tools added.
            compilers      "Second underscore" problem addressed.

25Jul06     lc-utilities   LDAPSEARCH replaces DCECP.
            locations      Two paths to FTP compared.
            often-used-tools
                           Two versions of FTP compared.
            batch-system   SLURM spreads under AIX 5.3.

14Jun06     linux-resources
                           SLURM spreads to AIX also.
            pthreads       SLURM thread support spreads to AIX.
            environment-variables
                           Many BG/L-only variables added.
            batch-system   PSUB -g works only on AIX.

13Feb06     linux-resources
                           New summary, cross refs added.
            lc-utilities   DIRCOLORS added.
            compilers      LD_LIBRARY_PATH details updated.
            environment-variables
                           Cross ref to new manual added.

28Nov05     linux-overview CHAOS 3.0 deployed.
            compilers      How CHAOS 3.0 affects LD_LIBRARY_PATH.
            os-differences Xorg replaces XFree86 for CHAOS 3.0.

21Sep05     batch-system   SCF SLURM/LoadLeveler domain split.

25Apr05     lc-utilities   SMAP, LDEDIT added.
            compilers      Unlimited stack size on OCF Linux.
            batch-system   TBS, NQS status updated.

09Feb05     lc-utilities   WHATSUP for Linux only.
            parallel-file-systems
                           New section for Lustre.
```

```
           index         New keyword for new section.

05Oct04    lc-utilities  MATLAB diverges on CHAOS, AIX.

05Aug04    lc-utilities  LPOPTIONS only under OCF Linux.
           compilers     Warning added on LD_LIBRARY_PATH.
           often-used-tools
                         LP, LPR -o differences added.

28Jun04    linux-overview CHAOS 2.0 deployed.
           lc-utilities  TSB available under Linux.
           compilers     Many GNU changes for CHAOS 2.0.
           pthreads      Native threads under CHAOS 2.0.

13Apr04    linux-resources
                         /nfs/tmp1, 3 replace 0, 2.
           lc-utilities  QUOTA SCF limitations noted.
           compilers     GCC version 3.3.2 added.

13Jan04    linux-overview More support sites added.
           lc-utilities  GIVE, TAKE now cross machines.

04Nov03    lc-utilities  More comparisons added.
           pthreads      SLURM replaces RMS, thread aware.
           environment-variables
                         SLURM env. vars. cross referenced.
           batch-system  Deployment comments updated.

17Sep03    unix-utilities Mozilla replaces Netscape.
           lc-utilities  Four more status changes.

25Aug03    batch-system  Cross ref to SLURM manual added.

02Jul03    linux-overview Linux-relevant IBM documents noted.

14May03    netmon        New section on NETMON relevance.
           index         New keyword for new section.
           lc-utilities  No across-machine GIVEs, TAKEs.
           locations     /usr/give present but not used.

22Apr03    lc-utilities  HTAR now on Linux.
           often-used-tools
                         DF, FTP more differences noted.
           locations     DFS not on any Linux nodes.

11Mar03    linux-overview Cross ref to CHAOS manual added.
           lc-utilities  PDSH, SUPER added, others revised.
           environment-variables
                         Details updated and clarified.
           batch-system  Cross ref to CHAOS manual added.

27Jan03    linux-overview CHAOS deployment noted.
           locations     MKL Intel library pathnames added.
           environment-variables
                         OMP_NUM_THREADS role added.
           lc-utilities  SFTP added.
           batch-system  New section on TBS, SLURM.
           index         New keyword for new section.

04Dec02    linux-resources
```

```
                        ILX replaces LX cluster on OCF.
           lc-utilities  GIVE, TAKE now under Linux.
           locations     PMATH, SPRNG now in /usr/local/lib.

22Aug02    locations     More on compilers, HDF5 library.

08Jul02    lc-utilities  Two more LC tools added.
           often-used-tools
                         USERINFO, UINFO compared.

14May02    software      Subdivided, many more details.
           compilers     New subsection on compiling.
           lc-utilities  Availability updated, list expanded.
           pthreads      Linux RMS not thread aware.
           linux-resources
                         SCF Adelie and Emperor added.
           environment-variables
                         New section added.
           index         New keywords for new sections.

19Mar02    software      Five more LC tools listed.

12Feb02    software      Three LC-only storage routines listed.

09Jan02    linux-overview Linux usability study added.
           linux-resources
                         More OCF LX nodes; o.s. upgrade.
           software      LFT added to list.

06Dec01    software      Four more LC-only tools listed.

19Nov01    entire        First edition of Linux manual.


TRG (05Sep07)
```

UCRL-WEB-201326