

HP XC How To



Installing Standard LSF on a Subset of XC Nodes

Version 1.0 July, 2005

© 2005 Hewlett-Packard Development Company, L.P.

The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Linux is a U.S. registered trademark of Linus Torvalds.

LSF, Platform Computing, and the LSF and Platform Computing logos are trademarks or registered trademarks of Platform.

Contents

Introduction

Requirements	5
Assumptions	5
Sample Case	5

Procedure

Prepare the setup files.....	6
Obtain standard LSF and install it into the existing LSF "tree"	6
Correct the setup files	6
Create a custom startup script for standard LSF	7
Adjust JOB_STARTER script for LSF-HPC for SLURM	8
Re-run <code>cluster_config</code> to update node roles and re-image	9
Restart the cluster with <code>startsys</code>	9
Verification	9

Revision history

Revision tables

Table 1 Revisions

Date	Edition	Revision
July 2001	V1.0	First Edition

Introduction

This document provides instructions for installing standard LSF on a subset of nodes in the XC cluster (in our example a set of large SMP nodes or "fat" nodes) while maintaining LSF-HPC integrated with SLURM on the rest of the nodes in the XC cluster (in our example the "thin" nodes).

This approach prevents jobs from running across both thin and fat nodes, but does offer full standard LSF support for these large SMP systems, particularly job scheduling based on the size and load of memory and/or cpu.

The existing XC `cluster_config` program allows you to decide which nodes have a "compute" role and which nodes have a "resource_management" role. The "compute" nodes become SLURM compute nodes. The "resource_management" nodes are where the SLURM master and backup daemons reside, and one of them is selected to run the LSF-HPC daemons. Thus the existing technology in XC allows you to "configure out" a subset of XC nodes that will **not** run LSF-HPC with SLURM.

Before running the `cluster_config` command to adjust the role assignments, you need to install standard LSF and perform some additional configuration to support standard LSF within XC.

Installing standard LSF on XC is straightforward, it just involves a few extra adjustments in order to work with the file system management on XC. The following procedures cover all the necessary adjustments:

Requirements

1. This procedure has the following requirements:
 - Standard LSF version must be the latest 6.0 version or later (with `schmod_slurm` module).
 - You must be familiar with LSF-HPC for SLURM installation and configuration on XC.
 - You must be familiar with standard LSF installation and administration procedures
 - **The XC head node cannot be configured to run standard LSF.** This is due to an unresolved issue in the LSF failover and setup mechanism on XC that will be corrected in the next XC release.

Assumptions

The following assumptions apply to this procedure:

- LSF-HPC for SLURM was installed by the `cluster_config` process using default values
- You have obtained a proper Platform LSF license.
- There is no desire to communicate with an external LSF cluster (this can be done, but involves additional procedures to prepare the external network connections).

Sample Case

The example in this HowTo considers an XC cluster of 128 nodes consisting of:

- A head node with a hostname of `xc128`
- 6 large SMP nodes or "fat" nodes with the hostnames `xc[1-6]`
- 1. 122 thin nodes. 114 of the "thin" nodes are compute nodes and have hostnames of `xc7-120`.

Procedures

Prepare the setup files

1. Log into the head node of the XC cluster as `root`. Do not log in through the cluster alias.
- Change directory to `/opt/hptc/lsf/top/conf` and rename the existing setup files:

```
# mv profile.lsf profile.lsf.xc
# mv cshrc.lsf cshrc.lsf.xc
```

Obtain standard LSF and install it into the existing LSF "tree"

These instructions assume that the user is familiar with the procedures to install standard LSF, which basically consist of configuring the `install.config` file and running `./lsfinstall -f install.config`.)

1. The required minimum settings for the installation config file are as follows (adjust as necessary for your system):

```
LSF_TOP="/opt/hptc/lsf/top"
LSF_ADMINS="lsfadmin"
LSF_CLUSTER_NAME="hptclsf"
LSF_ADD_SERVERS="xc1 xc2 xc3 xc4 xc5 xc6"
```

2. Set the hostnames of your "fat" nodes in the `LSF_ADD_SERVERS` entry.

Correct the setup files

2. Correct the setup files as follows:
 3. The new installation of standard LSF will create new setup files. Rename these files as follows:

```
# mv profile.lsf profile.lsf.notxc
# mv cshrc.lsf cshrc.lsf.notxc
```
 4. Create softlinks of the original files to the XC versions as follows:

```
# ln -s profile.lsf.xc profile.lsf
# ln -s cshrc.lsf.xc cshrc.lsf
```
 5. The setup files for standard LSF (the newly installed ones) need to be edited to workaroud a minor configuration bug. To determine whether or not LSF is running on a SLURM_based system,

these files look for an XC-specific file: `/etc/hptc-release`. If found, these files assume that the LSF daemons should be interfacing with SLURM.

For the fat nodes running standard LSF, this is not true. To work around this, we need to create an "identity" file that only exists on the thin nodes, and have the setup files for standard LSF look for this "identity" file.

The "identity" file will be `/var/lsf/lsfslurm`. Use the `pdsh` command to create this file on all nodes **except** the fat nodes. For our example the `pdsh` command is as follows:

```
# pdsh -a -x xc[1-6] touch /var/lsf/lsfslurm
```

6. Change the file name in the setup files by using the following `sed` commands:

```
# sed -e "s?/etc/hptc-release?/var/slurm/lsfslurm?g" < profile.lsf.notxc > profile.tmp
# sed -e "s?/etc/hptc-release?/var/slurm/lsfslurm?g" < cshrc.lsf.notxc > cshrc.tmp
```

7. Confirm that the only change was the filename:

```
# diff profile.tmp profile.lsf.notxc
120c120
< #           Currently we only check for HP-hptc: /var/slurm/lsfslurm
---
> #           Currently we only check for HP-hptc: /etc/hptc-release
127c127
<   _slurm_signature_file="/var/slurm/lsfslurm"
---
>   _slurm_signature_file="/etc/hptc-release"
# diff cshrc.tmp cshrc.lsf.notxc
266c266
<           if ( -f /var/slurm/lsfslurm ) then
---
>           if ( -f /etc/hptc-release ) then
```

8. Replace the old file with the new file:

```
# mv -f profile.tmp profile.lsf.notxc
# mv -f cshrc.tmp cshrc.lsf.notxc
```

9. Add the new LSF execution host(s) to `LSF_SERVER_HOSTS` in the `lsf.conf` file. In our example the new `LSF_SERVER_HOSTS` entry is as follows

```
:
LSF_SERVER_HOSTS="lsfhost.localdomain xc1 xc2 xc3 xc4 xc5 xc6"
```

Create a custom startup script for standard LSF

This startup script will be run on every node, so it needs to be selective and only operate on the nodes on which standard LSF is desired.

1. Create the following script in `/opt/hptc/lsf/etc/slsf` and edit it to contain your "fat" nodes:

```
#!/bin/bash
#
# chkconfig: 345 99 01
# description: standard LSF daemon management
# processname: slsf
```

```

# source LSF
. /opt/hptc/lsf/top/conf/profile.lsf.notxc

# valid hosts for standard LSF on this cluster
hosts="xc1 xc2 xc3 xc4 xc5 xc6"

hostname=`hostname`

valid=0
for i in $hosts; do
    if [ "$hostname" = "$i" ]; then
        valid=1
    fi
done

if [ "$valid" = "0" ]; then
    exit 0
fi

lsf_daemons "$1"

```

2. Save and exit the file. Then set permissions, create the appropriate softlink, and enable it:

```

# chmod 555 /opt/hptc/lsf/etc/slsf
# ln -s /opt/hptc/lsf/etc/slsf /etc/init.d/slsf
# chkconfig --add slsf
# chkconfig --list slsf
slsf          0:off  1:off  2:off  3:on   4:on   5:on   6:off

```

3. Edit `/opt/hptc/systemimager/etc/chkconfig.map` and add the following line to enable this new "service" on all nodes in the cluster:

```

slsf          0:off  1:off  2:off  3:on   4:on   5:on   6:off

```

Adjust JOB_STARTER script for LSF-HPC for SLURM

3. If the XC cluster version is earlier than v2.1 and LSF-HPC is configured with the recommended JOB_STARTER script, make the following small change to the JOB_STARTER script. At the top of the file change:

```

which srun > /dev/null 2> /dev/null
if [ "$?" != "0" ]; then

```

4. to the following:

```

if [ -z "$SLURM_JOBID" ]; then

```

5. This prevents the JOB_STARTER script from trying to invoke srun on the fat nodes.

Re-run `cluster_config` to update node roles and re-image

1. shutdown the rest of the cluster with `stopsys`.
2. Change directory to `/opt/hptc/config/sbin` and execute `./cluster_config` as follows:
 - a. select "Modify Nodes" and change the roles on the fat nodes to remove the "compute" and "resource_management" roles. Ensure there's at least one "resource_management" role remaining in the cluster (2 "resource_management" nodes are recommended).
 - b. do not re-install LSF.
3. When `./cluster_config` finishes, you may need to manually adjust the `/hptc_cluster/slurm/etc/slurm.conf` file to remove the fat nodes from the `NodeName` and `PartitionName` entries.
4. Run `scontrol reconfig` to update SLURM with the changed information.

Restart the cluster with `startsys`

6. When `startsys` is complete and the nodes have re-imaged, everything should be up and running:

```
# sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
lsf        up    infinite    6    idle xc[7-120]
# lshosts
HOST_NAME      type      model    cpuf  ncpus  maxmem  maxswp  server  RESOURCES
lsfhost.loc    SLINUX64 Itanium2 60.0  228   1973M   -       Yes (slurm)
xc1            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
xc2            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
xc3            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
xc4            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
xc5            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
xc6            LINUX64  Itanium2 60.0   8    3456M  6143M   Yes ()
```

7. Only those nodes on which role changes were made will be re-imaged. This means that standard LSF binaries and the `sfsf` script and `softlink` will not be present on the "thin" nodes. See the HP XC Administration Guide on the use of the `updateclient` command to update the "thin" nodes with these latest file changes.

Note that the "thin" nodes do not need to be updated with these files in order to complete this procedure. It is just a matter of consistency among all the nodes in the cluster. The "thin" nodes can be brought up-to-date with these changes at a later time. Refer to the HP XC documentation for more information on these commands.

Verification

Change to a non-root user and test the changes by running some jobs:

```
$ bsub -I -n1 -R type=LINUX64 hostname
Job <176> is submitted to default queue <normal>.
```

```
<<Waiting for dispatch ...>>
<<Starting on xc1>>
xc1
$ bsub -I -n1 -R type=SLINUX64 hostname
Job <177> is submitted to default queue <normal>.
<<Waiting for dispatch ...>>
<<Starting on lsfhost.localdomain>>
xc120
$ bsub -I -n2 -R type=SLINUX64 srun hostname
Job <178> is submitted to default queue <normal>.
<<Waiting for dispatch ...>>
<<Starting on lsfhost.localdomain>>
xc7
xc7
```