Installation and Administration of Software on a Cray SV1 SuperCluster from the SWS

Scott E. Grabow, Silicon Graphics, Inc., grabowc@sgi.com

ABSTRACT: *With the introduction of the CRAY SV1 SuperCluster the installation process for the initial and upgrade installations of UNICOS has been updated to provide a level of flexibility and consistency for sites that have a CRAY SV1 SuperCluster system. This paper will outline the two processes available to perform each installation and address some administration issues when working with SV1-8 and larger SuperClusters.*

Introduction

The introduction of the CRAY SV1 SuperCluster has necessitated the creation of an installation process which allows a site to install software package on multiple systems concurrently rather than having the site install software on each system individually. This document is divided into several sections dealing with the changes and impact of the new installation process for the CRAY SV1 SuperCluster.

SuperCluster configurations

SuperCluster Building Block

The basic CRAY SV1 SuperCluster is called a SuperCluster building block, and is commonly referred to as a CRAY SV1-4 SuperCluster. The CRAY SV1-4 is the smallest grouping of mainframes that make up a SuperCluster. The figure below shows a SuperCluster building block.
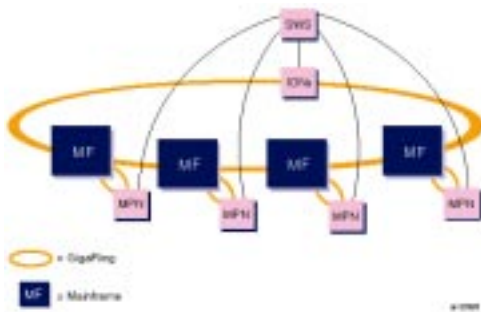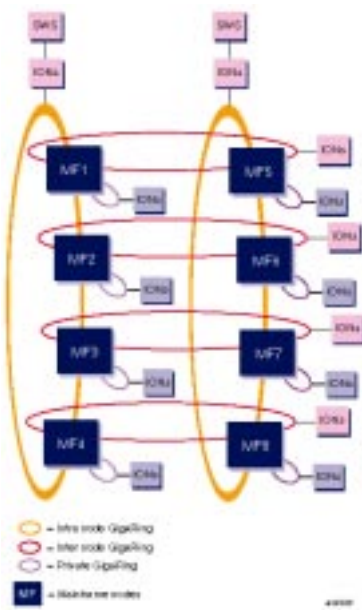
A SuperCluster building block has four CRAY SV1 mainframes that exist on a shared GigaRing with I/O Node(s) (IONs). The shared GigaRing is commonly referred to as the Intra GigaRing and serves two duties. The first being to provide a connection between I/O devices on the shared GigaRing while the second duty is to provide a TCP/IP network for communication between mainframes.

Each mainframe in the SuperCluster building block also has at least one private GigaRing that provides I/O gear to boot the mainframe and connect to external networks. Additional private GigaRings may be present depending upon the configuration.

The mainframes in the SuperCluster building block are connected in a direct connection configuration. The direct connection configuration means that each mainframe is connected to all other mainframes in the SuperCluster, and that a message from one mainframe to another mainframe does not have to be forwarded by a third mainframe.

Merging SuperCluster Building Blocks

The making of larger SuperCluster configurations is achieved by merging multiple SuperCluster building blocks. For example to create a CRAY SV1-8 SuperCluster is created by merging two SuperCluster building blocks and adding four new shared GigaRing connections between some of the mainframes in each building block with additional shared I/O Node(s). The figure below shows a CRAY SV1-8 SuperCluster.

A CRAY SV1-8 SuperCluster has eight CRAY SV1 mainframes that exist on six shared GigaRings with IONs. There are two Intra GigaRings which are parts of the SuperCluster Building Block. The other four shared GigaRings with IONs are Inter GigaRings and are used to connect specific mainframes from one building block to a similar mainframe in the other building block. All shared GigaRings serve the same purpose as the Intra GigaRings. The first being to provide a connection between I/O devices on the shared GigaRing while the second duty is to provide a TCP/IP network for communication between mainframes.

All SuperClusters larger than the CRAY SV1-4 system have a matrix connection configuration. To access mainframes that are not on the same shared GigaRings, one mainframe from the shared GigaRing IP forwards the communication to another GigaRing where the target mainframe resides. There is at most one IP forwarding of communication between any two mainframes in the SuperCluster.

Another advantage to this design is that there are multiple SWSs which can be provide redundancy for administration and control the SuperCluster.

Installation processes

Prior to the introduction of the CRAY SV1 SuperCluster, there was only one method to install UNICOS and other software products on multiple mainframes; it was one mainframe at a time.

The there are two supported installation processes for the SuperCluster: the standard single system installation and a new SuperCluster building block installation process.

Standard Single System Installation

The standard single system installation is the installation process that has been in place since 1997 using the Common Installation Tool (CIT).

This process is still available for SuperCluster sites that want to evaluate a software release on one system of their Super-Cluster before placing the software on out on the entire Super-Cluster. There may be other reasons a site may want software specific to one mainframe, and the single system installation process will allows sites to tailor their installation process to suit their needs.

Multiple System Installs

The new installation process being introduced for a Super-Cluster is a multiple system install. This new process will allow a site to install software upon the mainframes of a SuperCluster building block concurrently, rather than doing multiple single installs of software.

One can think of the multiple system installation for a Super-Cluster building block as a wrapper that performs each step of the standard single system installation four times. The advantage to doing the installs in parallel is that the amount of time necessary to do the installation is roughly the same as the amount of time to perform a single system installation. This process also allows a site to keep the software in sync across a SuperCluster building block.

Note if the SuperCluster is larger than a CRAY SV1-4 Super-Cluster, then multiple installs of the SuperCluster installation process must take place. This approach of limiting the number of systems done in parallel was done for several reasons. An important reason was to allow for sites with SuperCluster to stage software beyond one system to multiple systems, and also ensures that the number of processes and windows taking place concurrently on an SWS during a SuperCluster installation are limited.

The multiple installation process only works with four systems that make up a SuperCluster building block, if a site has made a cluster of two or three systems the approach will not work. In these instances, the site will need to perform the standard single system installation multiple times.

The process for controlling a multiple system installation has a new installation information file called SuperClusterMap. This file is similar to the sysinfo file that is used in UNICOS installs, but does not contain the detailed hardware information that is present in the sysinfo file. Instead, the SuperClusterMap file contains the following information for each mainframe in the SuperCluster building block:

• SerialNumber

• System name

• Mainframe IP address and hostname

• A private GigaRing's Ring number, and IONs

In addition.to the above information, there is information about the SuperCluster building block in the file such as:

- SWS information

- SuperCluster Name and SerialNumber information

- Information about the Intra GigaRing connections and IONs

- Information about the Inter GigaRing connections and IONs

The above information can accurately describe the minimal configuration of a SuperCluster building block.

One can think of the SuperClusterMap file as a complement, or starting point for filling out the sysinfo file for each system.

Potential SuperCluster Issues with Multiple UNICOS Releases

In implementing either installation approach in a Super-Cluster environment there are potential issues that the site must be aware of. The way in which a site addresses these issues is rather site specific and mainly depending upon how the Super-Cluster will be used in day to day operations.

What is an appropriate resolution for one site may not be satisfactory to another and sites are urged to create a consistent SuperCluster operational procedure for software upgrades. The following are potential issues that a site may encounter during the operation of a SuperCluster:

- New features the site wants to implement are limited to a select number of mainframe(s) in the SuperCluster while the UNICOS release is being evaluated for production on the entire SuperCluster.

- Software interaction between mainframe(s) running different software levels may produce unexpected results. A prime example of this would be the recent addition of the ability to turn off the priority queues for I/O on select disk types or disk devices at a specific mainframe.

- The potential for different user environments in a SuperCluster as new async software is installed for evaluation is possible.

Overall, SGI will work to expose such issues in the UNICOS release materials such as the UNICOS Release Letter, UNICOS Release Overview, and UNICOS Errata. While working to providing such information to our customers is necessary, there still may exist the potential for these and other software issues to arise due to software combinations that were not available during the UNICOS release cycle.

SuperCluster Initial Install

To support an initial install on a SuperCluster building block as a concurrent process with multiple installs taking place, the current single system initial install process had some additional steps wrapped around the process. These additional steps are in place to deal with the repetitive nature of the installation process in a SuperCluster building block.

Before starting, one must copy the SuperClusterMap file from the CD-ROM, and save the file as `/opt/CYRIos/sn`*`SuperClusterSerial-`*`Number/SuperClusterMap`. Next, the user should edit the SuperClusterMap file, and provide the necessary information to describe the SuperCluster building block.

Once a user has provided the necessary information, the SuperCluster.setup script is executed which creates an options, options.ram, topology, and boot.ions_rings files and then does the following for each mainframe in the SuperCluster building block:

- take information from the SuperClusterMap file and fill in the basic parts of the sysinfo file

- create a console and boot.ram commands

- create a param and param.ram file

Once the SuperCluster.setup script has successfully completed, each mainframe in the building block will then be booted with the just created RAM file system parameter file.

The last step then left is the installation of the software on the SuperCluster building block which will then take place concurrently. The process includes the staring of the networks on each mainframe, wiping clean the disk slices that will have UNICOS installed upon, load the software, verify the installation, setup the just installed software for going to multi-user mode, and then transfer back the new UNICOS kernel to the SWS.

At this point, the site configuration for the SuperCluster should be defined. The definition of the SuperCluster includes which resources on shared GigaRings are to be shared with all or some of the systems, and which resources are to be kept private.

If the SuperCluster is a CRAY SV1-8 system or larger there are several tasks that must be completed prior to booting the entire SuperCluster. These tasks include:

- Merging SuperCluster building block `topology` and `options` files from the multiple SWSs into a singe `topology` and `options` files and mirroring the files across the multiple SWSs

- Each Cray mainframe will need to have it's /etc/gr#.arp files updated to contain the network node name and arp address for each mainframe on the Inter GigaRing connections between two or more SuperCluster building blocks.

Once finished with the integration of multiple SWS building blocks, it will become necessary to update each Cray mainframe's param file and other site specific configuration issues. These tasks include:

- Addition of shared GigaRing I/O Nodes resources and private resources on shared GigaRings to the param file

- Addition of Inter GigaRing connection information to the param file in the hardware section and GigaRing Host-to-Host communication information to the network section.

- Updating the `/etc/hosts` file as necessary to contain Cray mainframe's network node names

- Addition of /etc/gr#.arp file(s) to contain the network node name and arp address for each mainframe on an Intra GigaRing connection.

After finishing all of these tasks, the site should continue to perform any necessary site modifications to the basic UNICOS configuration that was installed.

SuperCluster Upgrade Install

An approach similar to that of the SuperCluster Initial Install was taken for the SuperCluster Upgrade Install Process. Here again the desired end result was to take the existing single system installation process and place a wrapper around it to perform repetitive and parallel tasks where possible. Here again, the wrapper for the upgrade installation process will only work with the mainframes in the SuperCluster building block. If there are multiple building blocks, the process must be repeated for each building block.

The preparation tasks are the same as for a single UNICOS system installation, and must be repeated for each mainframe in the SuperCluster being upgraded. These tasks include allocating disk space, making sure the system is running a level of security that allows PRIV_SU and setting up the network communications between the SWS of each SuperCluster building block and the Cray mainframes.

Once the general preparation tasks have been completed, it will become necessary to verify the information inside the SuperClusterMap file on each of the SWSs that make up part of the SuperCluster building block.

With the SuperClusterMap file updated, the upgrade installation process is ready to be started on the building block. The first task to be performed is the automatic updating of each main-frame's sysinfo file to the newest version included with the UNICOS release. When the automatic updating of the sysinfo takes place, the default sysinfo file will be updated to include the same system information as the original file.

Once each mainframe's sysinfo file has been updated to use the newest sysinfo file, each file will need to be verified to make sure the critical disk information and where the configuration information should be coming from during the upgrade is correct.

Once the information in each file has been reviewed, it will be possible to start the setup process for the UNICOS upgrade. During the setup stage, the file systems will be dumped from the mount point identified as providing the configuration information and then restored upon the disks that will have the new UNICOS release installed upon. The last task of the setup process will be to perform a remove of old executables and old nmake files that ship with UNICOS.

After completing the setup stage, it will be possible to start the Common Installation Tool(CIT) for each mainframe in the SuperCluster building block concurrently, and perform the installation in parallel.

Final Remarks

It is hoped that the addition of the frame work put in place to perform multiple installs concurrently from an SWS of a Super-Cluster building block will provides sites a way to minimize the amount of additional time necessary to install UNICOS.