Use of Common Technologies between XT and Black Widow Bill Sparks (Jim Harrell)

III Sparks (Jim Harre

Cray Inc

ABSTRACT: This talk will present how Cray uses common technologies between XT3 and Black Widow. How past proven technologies and new emergent technologies help in administration and configuration of large scale systems..

KEYWORDS: BlackWidow, Linux, XT

1. Introduction

The Software Development teams at Cray working on Black Widow and the XT products are working towards using common technologies in their respective products. The drive towards commonality has several sources. We will introduce the topics here and then go into more detail.

First, there are changes in the system roadmap for Cray in the past year. The architectural changes for the hardware systems are providing a more common base for systems going forward. These changes made a desirable direction more of a requirement.

There are rationales for using common technologies and common techniques for both Software Development and the end customers. Chief among the advantages is interesting benefit of having software that has been used extensively by other architectures.

The Software Stack is the best place to show what technologies are being used in common. There are different views that can be used to show what technologies are available and how they are being combined. The use of common technologies will evolve over time as technologies evolve.

The status of the work to have more common base is already underway. A short summary of the work is described at the close of this paper.

2. System Architecture Directions

There are changes between 2005 and 2007 in the Hardware System Roadmap. The systems that are in the process of developed move to Black Widow being and Hood from X1E, XD1, and XT3. The infrastructure for these systems moves to a Hood base and to HSS/CRMS which are a combination of the Hardware Supervisory System for both the Black Widow and future systems, and the CRMS implementation of "HSS" for XT3 and Hood. There is work to reuse as much of CRMS as practicable in HSS for both Black Widow and future systems.

One major change in the systems that are being produced in the future is the use of the XT3, recently renamed Rainier, infrastructure as a base for Vector, Scalar, and Threaded systems. base, and The administrative T/O capabilities are all provided to these systems through Rainier. Rainier will be the focus for all I/O efforts for products. This should the help substantially in giving customers better more complete offerings.

As the products move forward to using Rainier there is an opportunity to start using heterogeneous systems since the overall architecture will allow Vector, Scalar, and Threaded processor The use of heterogeneous types. leads directly capabilities into Cascade System Architecture. Cascade's organization provides low level access to scalar, vector, and threaded processor capabilities. Thus, we can step forward into Cascade and fine grained Heterogeneity.

3. Software Development Rationale

Software development will benefit from merging technologies. This will allow us to focus on one technology to support several different products. At the same time Software development is working to create a single set of sources, a single set of tools for developers to use for development of code, and one set of development processes. This will allow us to provide a consistent look, feel, and quality to all the products. It will also allow Cray to move developers from one product line to another with much greater ease because only the product characteristics are different.

The heterogeneous systems will require more common technologies if software is to make the systems easy to use and manage. Obviously different interfaces for the different processor or products will inhibit the adoption of even multiple different products. Preparation for Cascade is another important rationale for merging the technologies.

4. Customer Rationale

Customers will be able to use fewer interfaces on mixed systems. This will make the user work and administration of the systems easier. Common interfaces will reduce costs at customer sites as well.

The larger value that is projected is that the software, having been used on other systems, will be more stable. The value of used software is greater than new software when the application of the code enhances the stability of the system.

5. Software Stack

The software stack over the next several years mirrors the changes in the hardware systems. Software will support the same hardware systems, but the components of the software stack that are the same are dramatically greater by 2007. By 2007 Black Widow and Hood are using much the same software except for the compilers, which are different because of the processor architecture. The use of common technologies will enhance both products quality.

6. Simple Example of Sharing Technologies

There are currently four different compute node operating systems under development. In the case of Eldorado the architectural differences in the processor are sufficiently different that a unique operating environment is required. Catamount, likewise, has been built around a scalar base and would require substantial work to port to a different architecture. However, both the Black Widow Linux compute node operating system and the XT Linux compute node operating system can be merged into a single source base and a single product. Clearly there are differences in the machine dependent sections of Linux and there are different drivers, but the majority of the code is the same and can be shared.

7. Comparing BW and XT Software Stacks Now and in the Future

Currently the software stacks for XT and BW are entirely separate. This is shown in the following diagram:



CUG 2006 Proceedings 2 of 4

The XT has, as a set of examples of components in the stack, such software as the Event Router Demon (ERD), Cray Remote Management System (CRMS), Portals, Quintessential Kernel (QK), Process Control Thread (PCT, Linux 2.6, ColdStart, MPI, and SHMEM. Currently these are only used by XT. The Black Widow software stack currently contains a set of software components such as the Hardware Supervisory System (HSS), Application Launch Placement System (ALPS), Portals, Mazama, Linux 2.6 (but a different revision than the XT), MPI, and SHMEM. These examples were chosen because they represent examples of components that provide similar functionality across the two systems.

The goal of the work to make common the software is to choose one base and migrate the base to support both products and migrate to that base.

Currently the XT stack is being used on a number of sites and has been run for over a year. This time in use has a benefit because it is tested software. Tested by being used. On the other hand, there are new concepts in Black Widow and some of the software components that will be the base moving forward. So, that the choices of components are not obvious without knowing the direction of development in the future.

The result of having the software made more common will look something like -



The diagram shows that most of the components are used on both systems in the future. There are certain components that are not used in common. One example is ColdStart. ColdStart is the processor initialization that is particular to the XT3. Another example is QK/PCT, the light weight kernel for compute nodes, which is also specific to XT and is not being ported to BW. Many of the components have machine dependent code but are essentially common.

8. Status

Cray is currently working to build common software repositories that will be used for development in the future. The transition to the new repositories will take place soon. There is also work to use common tools and processes which will have obvious benefits.

About the Authors

Bill Sparks is a member of the Operating System Team. He is currently leading the Mazama Project. He can be reached at jsparks@cray.com. Jim Harrell is Director of the Software Architecture Group. He can be reached at ejh@cray.com.