Lights Out Operations at Cray Research

Don Grooms, Worldwide Customer Service, Chippewa Falls, Wisconsin

ABSTRACT: *"Lights Out" operation may imply a variety of system capabilities. It means different things to different customers. Automated power on/off, system health monitors, automatic problem notification, and automated system recovery/reboot are capabilities that commonly come into play in a computer center's Lights Out operational strategy. This talk will begin with an examination of the meaning of Lights Out operation in terms of these various customer requirements. The standard features of current Cray products that support Lights Out operation will be summarized. The new SIO/GigaRing and System Workstation products will also be highlighted relative to the changes and possibilities they introduce into a Lights Out operational strategy.*

Introduction

The term "Lights Out" describes a computer center environment in which little or no direct contact is required during the startup, normal operation, or shutdown of the system involved. Within the industry the term blankets several operating environments and system features. For example, a customer that requires a secure environment may value attributes that allow the system to operate in a vault; remote monitoring and system redundancy may very well be the features needed most. In other environments, a customer may value unattended system startup and shutdown due to their concern over total power consumption. In these and other cases the term Lights Out can be applied as a general term, but the exact definitions vary with each customer's individual application and requirements.

Maintenance versus Customer Features

Features that support Lights Out operations tend to require design changes to existing equipment and specific efforts when the features are developed for new equipment. In some situations, though, the features that are designed for maintenance and general troubleshooting have the side benefit of doubling as Lights Out features. This report refers to system features in a very generic sense and makes little to no distinction between what was designed specifically for Lights Out operations, what has been developed for other purposes that lends itself conveniently to Lights Out, and the issue of what was designed for customer use versus system maintenance.

Scope

The family of Cray Research products available today includes many of the features that are often considered oriented to Lights Out operations. Due to the evolutionary nature of system development, each system, and in some cases each individual configuration, has its own set of Lights Out features. This paper will address these features system by system; however, the scope is limited to the Cray products available today. Products that are no longer being produced, products from non-Cray divisions of Silicon Graphics Inc., and the products that were sold through Cray's Business Systems Division, are not covered in this text. Designed-in system resiliency is a part of Lights Out operations. Most Cray products today include features that provide built-in resiliency. These features are included in this discussion.

System Power

Another important note is the definition of what power on/off really is. From a remote operations point of view, a chassis that is in a powered-off state still has some power applied to it. In this state, there is power applied within the chassis Warning And Control System (WACS) and other minimal electrical components. The system workstation, wherever it is located, is in a powered-up state as well. The modules themselves and all or most of the supporting power supplies are off.

Air and Liquid Cooling

These terms can be misleading depending on the product that is being discussed. For this discussion, air cooled refers to those systems that use the computer room environment as a source for

cool air and as an exhaust point for warm air. Liquid cooled refers to the use of a computer room chilled water system as the eventual heat exchange point for system cooling needs.

Definitions

The following are some of the terms and system attributes often considered part of Lights Out operations.

- Unattended Operation: In this environment the intention is to limit or eliminate the need for physical contact with the system. A perfect example of this is vault operations in which the system is located in a secure area. The word "unattended" does not have any specific limits in terms of what features are included; however, it is assumed that a system that falls into this category can be powered on, booted, and monitored to some extent from a remote location.
- Remote Operation: Similar in most respects to Unattended Operation, Remote Operation also implies a significant distance between the center of operations and the computer center where the chassis is physically located.
- Remote Monitoring: This system feature allows for system warning and alarm conditions to be set, viewed, and addressed from a remote location.
- Automatic System Start Up and Boot: An extension of Remote Monitoring, this feature performs the power up, system clears, and OS start functions. Ideally the sequence is initialized by a single keyboard command.
- Remote Support: This is a maintenance platform that allows a support engineer to dial in to a system from a remote location. In situations where a system is inoperative or degraded, it provides a means for support activity to begin almost immediately. Systems that are customer maintained can receive dial-in support when necessary. Systems that are in need of escalated support can be accessed by tech support from virtually any location. Cray provides two Remote Support platforms to choose from. These choices are the same for the OWS/MWS and SWS supported systems. As a default, a basic modem is available and there is the option of a NetBlazer which provides the same interface with enhanced network security. As a rule these options are available on all of the products discussed in this document. The details of what can be monitored, reset, and tested remotely depends completely on the product itself.
- Auto-Notification: Generally considered part of Remote Support, Auto-Notification refers to features that alert the operations staff or a Remote Support team that an event has taken place. The nature of the event depends on the system, but can include extreme conditions such as power and temperature out of tolerance or resiliency events such as a channel or disk that has been deconfigured by the system. The System Monitoring And Remote Troubleshooting Environment (SMARTE) was developed by Cray several years ago as a Remote Support tool. SMARTE is in limited use today as an Auto-Notification tool for disk maintenance.
- Designed-in System Resiliency: These are features that are added to a product in anticipation of a failure event. The resiliency feature accommodates the failure in a way that allows the failure event to take place without an interrupt to the system. Examples of this include single bit error correction, N+1 power supply design, redundant control systems, etc.
- NWACS: Warning And Control System (WACS) generally refers to the hardware that resides in and around a chassis that monitors points such as voltages and temperatures. NWACS is the software component of the system. The N prefix refers to the fact that the software supports nearly all Cray products, or N devices. It monitors from either the MWS/OWS or the SWS depending on what support platform is provided with the chassis. For systems that are equipped with remote power on/off capability, the NWACS software and WACS hardware together control power to the system.
- MWS/OWS and the SWS: One of the transitions that is taking place at Cray is the move toward a single support device. The MWS/OWS approach has been around for many years. The Maintenance Work Station (MWS) has been the platform for service related functions like maintenance channels, WACS information, etc., while the OWS has been the platform for operational functions. The System Work Station, or SWS, is the combination of both products into one. Unlike the MWS/OWS, the SWS supports multi-system and multiple I/O node environments. The details of this transition and its effect on Lights Out operations is provided in the product-by-product discussion and in the future section at the end of this document.
- Field Replaceable Unit (FRU): A Field Replaceable Unit is a sub-component of a system. Generally these components are processor and memory circuit boards, power supplies, disk drives, etc.

Lights Out Features

The following is a system-by-system discussion which addresses the remote power, remote monitoring, remote support, and system resiliency features within the Cray products that are available today.

J90

The J90 family of products has not been designed with features that support remote power on/off. The rationale for this decision comes from two points: cost of design and manufacture (the need to control the implementation of features that may not be of value to all customers) and cost of operation. The CMOS-based J90s are much less expensive to operate in terms of power consumption compared to their larger and more expensive ECL-based siblings. A side-by-side comparison between a fully configured J90se and T90 put the J90se at about 1/40 of the operating costs in terms of power consumption. The two points, cost of operation and cost of manufacture, interlock to some

extent. Cray has proceeded from the theory that a system with low power consumption and a low price tag is not a likely candidate for Lights Out operations.

Remote system monitoring is limited on the J90 with noticeable changes on the J90se. On the J90, the system console doubles as the maintenance platform, providing feedback on system memory errors. Chassis status like AC/ DC power and temperature information is only available at the chassis. Another factor is the reset of I/O hang conditions: on the J90 the reset is a manual button on the CCU box itself.

The J90se measurably improves upon the J90's limitations. First, the system console will be replaced by the SWS. Secondly, the J90se voltage and temperature status signals will be available to the SWS.

The J90se introduces compatibility to the SIO/GigaRing system. When sold with the J90se, the SIO/GigaRing can be purchased with one of two cabinets. The PC-10a cabinet provides limited remote monitoring capability, as the chassis has to be examined directly for DC and Fan status indicators. The PC-10b cabinet provides status from the power supplies, fan failure, and N+1 supply status to the SWS console. The J90se also supports remote reset of individual CPUs and I/O MPNs.

Many of these features, like the relationship between system monitoring and reset from the console, as compared to chassis based buttons and indicator lights, play an important role in how the machines can be supported. The remote support platform for the J90 and J90se is keyed to the local console or SWS, so what can be viewed and reset locally can also be performed by a technician in a remote location.

J90/J90se Lights Out Summary

The J90 and J90se are extensions of the ELS line of products. The ELS, or Entry Level Systems, approach is, and must be, different from the classic Cray. These differences can be seen in the system and maintenance prices and in the cost of operation. The differences can also be seen in the careful selection of maintenance features that are designed into the systems. The features are limited; however, they are matched to the environments the systems operate in, the general simplicity of the systems in terms of FRUs, and a step toward decreased cost of ownership.

C90

For this discussion it is important to note that the C90 will continue to be offered in its standard configuration of IOS-E and MWS/OWS. There are no plans to offer the C90 with either the new SWS, which will supersede the MWS/OWS on some products, or the SIO/GigaRing, which is the follow-on to the IOS-E.

The features that support remote system power on/off differ slightly based on the type of cooling system installed. All air cooled systems are delivered with remote on/off capability. For liquid cooled C90 systems the feature is not shipped by default; however, it is available as a field-installable option. The Field Change Order (FCO) for this option includes a new PC board in the system chassis, a minor wiring addition, and an addition to the MWS software. Today, several liquid cooled C90s in the field have this option installed.

Power on the IOS-E for the C90 is keyed to the state of the power on the C90 chassis. Regardless of the motor generator (MG) configuration, the same remote power on/off commands sent to the C90 will be mimicked by the related IOS-E. As a result, the remote power on/off features available for the C90 also control the IOS-E even if the IOS-E is in a stand alone cabinet and/or is operating on its own MG.

Remote monitoring on the C90 is quite comprehensive. Virtually all of the WACS information from the display on the chassis is also available on the MWS console. This information includes the incoming AC voltage readings, the DC voltages of each internal power supply, module temperatures and voltages, fluorinert temperatures and pressures, etc. The NWACS software on the MWS supports the display and controls the power on/off features.

The C90 HM4M and M4M memory modules, containing spare memory chips, offer an opportunity for added system resiliency for the Lights Out environment. If the memory error correction reports a consistent solid single bit error, the system can be shut down, the failing chip configured out of the system remotely, and the system restarted without the module being removed. This is an option that was not available on earlier Cray systems where the choices were limited to either leaving the module in and perhaps running the chance of double bit error and possible system halt, or taking the system down for a longer period of time and removing the module physically from the chassis.

C90 Lights Out Summary

The C90 can best be viewed as a stepping stone toward Lights Out capable systems. The C90 systems can be powered on and off remotely, but modifications are required if the system is liquid cooled. In short, the remote on/off features are available, but not offered with all C90 systems by default. Remote monitoring of the chassis is supported through the NWACS program on the MWS. This provides a comprehensive view of the chassis status and can be viewed from wherever the MWS is located. From a resiliency perspective, the C90 HM4M and M4M modules offer spare chips. From a remote location, for example vault operations, a solid single bit can be virtually removed from the system and the system returned to the customer without physical contact with the system itself.

T90

The T90 family of products is offered today with the IOS-E I/O product and is operated and maintained with the OWS/MWS platforms. In upcoming months the T90 products will eventually switch to the SIO/GigaRing platform and SWS. These changes will not impact the T90 chassis in a way that affects Lights Out capability.

The T90 products are powered on and off from the MWS. In fact, unlike the C90 in which console driven power on/off is an option, on the T90 products it is the only means of controlling

the system power other than an emergency off button. The NWACS software on the MWS also displays chassis information such as the input AC power readings, DC bus voltage readings, fluorinert pressures and temperatures, and standard error log information like single bit and register parity errors. NWACS also reports the status of the N+1 power supply system.

Another feature of the T90 products is the single start command. From the OWS console, a single command powers the system on, clears the system, and boots the operating system. The same is available for system shutdown: a single command performs an operating system shutdown and then drops the system power. While not directly Lights Out features, they identify a step toward a less procedure-intensive approach to system administration.

The T90 products boast several resiliency features. Power is distributed through an N+1 power supply system. This design accommodates a single power supply failure without interrupting the operation of the system. The failed supply is reported to the console through the NWACS system and can be replaced when it is most convenient to the customer. The chassis control system is also redundant. A command leaving the primary control system simply informs the backup system of an internal system failure. The backup system in turn takes control of the chassis. As in the C90, the T90 comes with a spare memory chip format. Like the HM4M and M4M modules in the C90, the feature allows solid single bit failures to be mapped out of the system with the module still in the chassis. The internal clock system in the T90 products is also built around a redundant system. While both of these features require a system restart after the transition, the entire process can be performed remotely from the workstation.

T90 Lights Out Summary

The T90 family of products represents a significant step toward Light Out capable systems. The T90 is equipped for remote power control and system monitoring and has a simplified start up procedure. In addition, it boasts an array of redundancy features that allow for ongoing operations in some failure modes, like a single power supply or control system failure, and limited downtime in other failure modes.

T3D

The T3D is a fairly mature product. It has been available for several years and is now being replaced by the T3E. Sales of the T3D have included the OWS/MWS operation and support platforms and IOS-E I/O product.

Like many Cray products, the T3D has been available in air cooled and liquid cooled chassis. For the air cooled systems, remote power on/off is performed from the MWS or OWS utilizing a link from the OWS to the NWACS software on the MWS. The link that supports this feature from the OWS to the MWS has not been made available for the liquid cooled systems. The reason for this is the pairing of the smaller air cooled T3Ds with small Y-MP and C90 configurations.

For both air and liquid cooled systems the NWACS program provides a comprehensive display of chassis monitoring features like module voltages and temperatures and general cooling system status.

T3D Lights Out Summary

The pairing of C92-A and C94-A systems with air cooled T3D systems and the overall offering of power control from the MWS/OWS system marks a significant event in the availability of Lights Out capable systems from Cray.

T3E

Note: While the introduction of the SWS and SIO/GigaRing adds a complication to the Lights Out features offered on some products, it does not on the T3E. The T3E is only available with the SWS and SIO/GigaRing devices.

The T3E air cooled and liquid cooled chassis are both capable of supporting remote power on/off and remote monitoring. The features are standard and are operated within the NWACS system on the SWS. In multi-cabinet configurations, power to each cabinet is controlled individually by the NWACS system. Module temperatures, voltages, and other chassis specific data is available to support remote status monitoring of the system. With air cooled systems the internal cabinet air temperature is monitored, and within the liquid cooled systems the Heat Exchange Unit (HEU) is monitored for temperature and flow conditions.

The potential for a single command to power on and boot an entire T3E system is being investigated. Multi-cabinet configurations pose a major challenge for this as the design today requires each cabinet to be powered up individually and in a specific sequence. Once the power up is complete, a boot script is used to boot the entire system. The boot script can also perform boundary scan and a quick diagnostic procedure as part of the boot process.

From a resiliency standpoint, the T3E brings forth an interesting set of features and the promise of more to come. Today, the T3E comes equipped with extra nodes. The number of spares is proportional to the number of nodes in the system. For example, a T3E sold as a 128 processing element (PE) system contains 136 actual PEs. These are considered to be a combination of operating system software PEs and spares. By the end of the calendar year it is planned that a failing PE will be accessible with diagnostics while the operating system is running concurrently.

The mapping of nodes in and out of the system is an important step toward the next goal, concurrent repair of the chassis. Also known as hot swap, this is being considered for implementation in late 1997. To perform a concurrent repair, a technician would first map a failing PE out of the system and activate a spare PE to fill its position. Diagnostics would then confirm the failure and capture the related information, and finally the failing PE is removed from the chassis and replaced by a site or depot held spare. The system could then be reconfigured back to its normal operating format.

T3E Lights Out Summary

One must keep in mind that the T3E is very new. At first glance many of the features mentioned above appear to lend themselves directly to Lights Out operations, especially concurrent system diagnosis and reconfiguration. From the lights out perspective this system looks very well suited. But, in reality, some of the features are untested at this time.

SIO/GigaRing

The SIO and GigaRing together form the latest I/O product from Cray. It is, however, an I/O device and as such will always be associated with another chassis. For Lights Out purposes, the SIO/GigaRing must then be evaluated along with the chassis to which it is connected.

The SIO/GigaRing, when connected to a T3E or T90 product, matches well with many of the Lights Out features provided by these systems. The PC-10b cabinets, the chassis building block of the SIO/GigaRing system, can be powered on and off individually and monitored from the SWS utilizing the NWACS system. The PC-10b information is limited to the N+1 power supply status and fan failure status on the Single Purpose Nodes (SPNs) and disk subracks.

For remote monitoring, an exception exists on the J90se product in the type of SIO/GigaRing cabinet that is provided by default with a J90se configuration. See the J90 section of this document for details on this issue.

The Future of Lights Out Operations

From the product information contained in this document, it can be seen that as Cray products have evolved, so have the features that support Lights Out operations.

Near term software development plans for the SWS, specifically in the implementation of the *resource* concept, will have a direct impact on the way Lights Out operation is made available to customers. As the SWS platform has been developed, the need to control multiple system and multiple I/O node environments from a single SWS has been addressed. To accommodate this complex problem, the concept of a *resource* has been introduced. A resource binds specific hardware and software together and allows configuration of the combined resource. For example, a specific mainframe, operating system, and operating system kernel can be defined and treated together as a resource. Commands for the resource function include reserving and relinquishing specific system resources, starting and stopping resources, and, where it is supported by the hardware, resource power on/off functions as well. Control of power on/off functions from the SWS introduces a new level of Lights Out opportunity like crontab(1) driven power on/off functions. Implementation of these features is planned for 1997.

Enhancing this picture is the wave of resiliency features that have been added. Features like redundant nodes, spare memory chips, and N+1 power supply designs reduce the impact of interrupts. Beyond that, these features allow for uninterrupted operation in failure modes that would have otherwise resulted in system downtime and on-site repair.

The Cray remote support program plays an important role in Lights Out as well. As the monitoring, diagnostic, and reset features have made their way out of the system chassis and onto the system workstation, these features have also been available to the remote support engineer.

There is potential for more. With strong customer support the next steps could be toward compatibility with host systems for host-driven Lights Out operation. Further enhancements in the area of one-keystroke power on/boot procedures are possible as well.