# Cray XC System Diagnosability Roadmap

Jeffrey J. Schutkoske

Platform Services Group (PSG)

jjs@cray.com

# Safe Harbor Statement

This presentation may contain forward-looking statements that are based on our current expectations. Forward looking statements may include statements about our financial guidance and expected operating results, our opportunities and future potential, our product development and new product introduction plans, our ability to expand and penetrate our addressable markets and other statements that are not historical facts.  These statements are only predictions and actual results may materially vary from those projected. Please refer to Cray's documents filed with the SEC from time to time concerning factors that could affect the Company and these forward-looking statements.

4/27/15

# Overview

- **Introduction to System Diagnosability**
- **Diagnosability Enhancements**
  - Node Diagnostics
  - Nvidia GPU Diagnostics CUDA 6.5 Support
  - I/O Devices Diagnostics
  - HSS Diagnostics and Monitoring
  - System Notification
  - CLE Enhancements
- **Future Considerations**
- **Q&A**

# System Diagnosability Overview

- **System Diagnosability is a suite of software tools**
- **Features built into SMW and CLE commands**
- **Diagnosability is not just about Diagnostics**
- **HSS System Management Platform**
  - System discovery and inventory management
  - Node & High Speed Network (HSN) management
  - System infrastructure management
  - Relatively light footprint in terms of hardware and software
  - Performs monitoring and management Out-Of-Band (OOB)

# System Diagnosability

**CRAY**®

**Monitor** → **Diagnose** → **Notify**

4/27/15

5
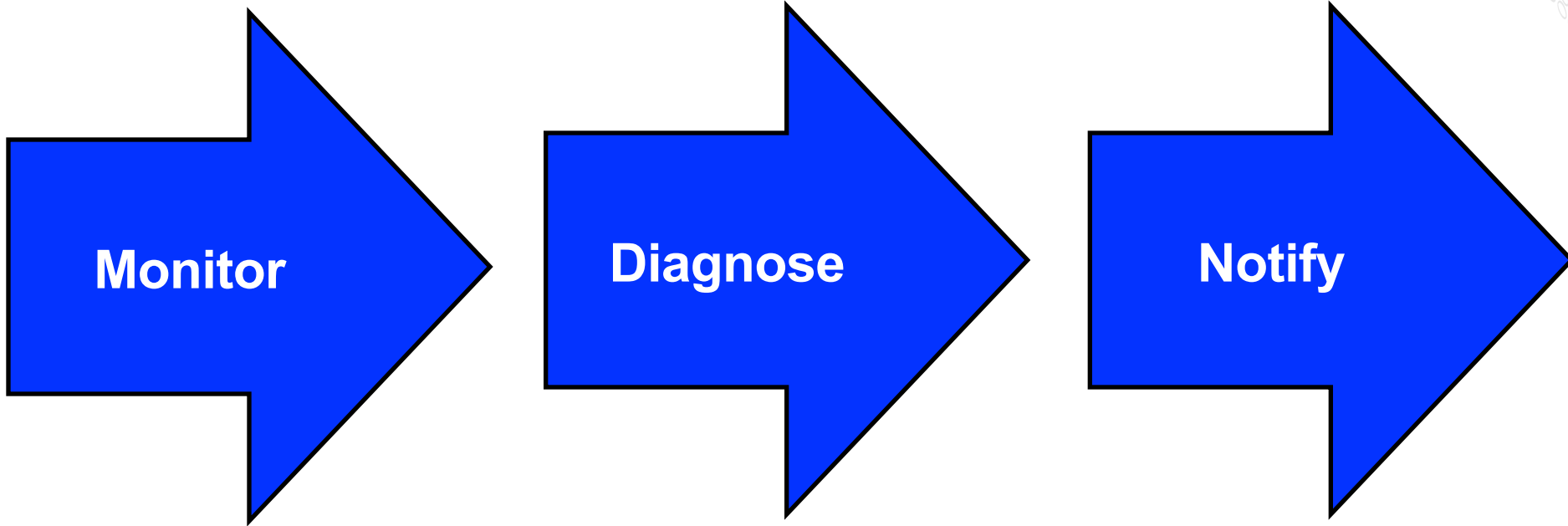
# Node Diagnostic Enhancements

- **Node Level Stress (NLS) diagnostic, *xtnls***

  - Collection of diagnostic programs and libraries

  - Diagnostic thread group executing test algorithms

- **Performance diagnostic, *xtcpudgemm***

  - Enhanced for greater control over each Intel processor core

  - Enhanced to utilize the Intel AVX2 instruction set

COMPUTE | STORE | ANALYZE

# Haswell Diagnostic Enhancements

- **Cray HSS, BIOS, and diagnostics enhanced**

- **MCA decode is a superset of previous Xeon**

- **Intel Haswell PECI support was enhanced**
  - Machine Check error log
  - Temperatures
  - Power management

# Nvidia GPU Diagnostic Enhancements

- ## Cray GPU Diagnostics enhanced to support CUDA 6.5

  - xkdgemm:  GPU performance

  - xkmemtest: GPU memory

  - xkbandwidth:  GPU PCIe performance

  - xkstress:  GPU Stress tests Streams, GEMM, and PCIe

  - xkcheck: GPU hardware and software verification utility

# I/O Devices Diagnostic Enhancements

- **HSS diagnostic utility, *xtcheckhss***

  - Validation of the PCIe I/O cards

  - Verifies PCIe speed, width, and missing

  - Must run xtbounce first

- **eXtreme DD (XDD), *xtxdd***

  - Perform data transfer operations between memory and disks

  - Creates multiple threads

  - Validates data and performance

# Workload Test Suite (WTS)

- **WTS control script, *xtsystest***

- **Tests that simulate a generic application workload**
  - Intel MPI Benchmarks (IMB)
  - High Performance Computing Challenge (HPCC)
  - High Performance Linpack (HPL)
  - Plus online diagnostics
- **Pre-compiled, pre-configured applications**
- **Baseline performance expectations**

# HSS Diagnostic Utilities

- **HSS diagnostic utility, *xtstresshss***

  - Executes processor and memory diagnostics on controllers
  - Checks for any errors encountered by controllers

*c0-0c1s2 Tolapai error register GLOBAL_FERR: 0x8000000*

*c0-0c1s2 Tolapai error register DRAM_FERR: 0x40*

*c0-0c3s7 Tolapai watchdog timeout*

# HSS Power and Cooling Diagnosability

- **Enhanced Cabinet power and cooling checks**

- **Full cabinet power cycle required after the repair**

- **HSS diagnostic utility, xtcheckhss**

    *xtcheckhss –health*

- **Includes Initialization, Power On, Run-time errors**

# HSS Telemetry Data

- **System Environment Data Collections (SEDC)**

- **Monitors and logs data to files on the SMW**

- **Enhanced to store the sensor data in the PMDB**

- **Tables for Cabinet and Blade Controller data**

# HSS Controller Monitoring

- **Controller Vitality Check (CVC) daemon, *cvcd***

- **Monitors the HSS blade and cabinet controllers**

- **Additional monitoring plugins**

  - Controller memory errors

  - Controller PCIe Advanced Error Reporting (AER) errors

  - Controller Kernel oops errors

- **Generates an HSS health event**

# System Notification

- **Simple Event Correlator (SEC) – Cray Doc: S-2542**

- **Alerts and alarms trigger appropriate rules**

  - Detect excessive cabinet power draw

  - Cabinet EPO and environmental alerts

  - Node memory errors

  - Aries PCIe link change

  - RDMA timeout

  - Gets ALPS Process ID (APID) on job failures

# CLE Enhancements

- **Cray Data Virtualization Service (DVS)**

- **Distributed network service provides access to file systems was enhanced as follows:**

  - Identify hung Cray DVS request processes

  - Improved log messages

  - Periodic sync file system data

  - Improved error recovery and failover

  - Added statistics to track the periodic sync when a file is closed

# DVS Enhancements (cont.)

- **Number of enhancements for ESTALE errno handling**

  - Periodic Sync

  - Close / Re-Open

  - Different Server

# Future Considerations

- **Provide additional workload tests**

- **Provide enhanced system level diagnostics**

- **Provide additional monitoring and logging**

- **Provide diagnostic data analysis tools**

- **HSS system dashboard within OpenStack**

# Diagnosability Summary

**Monitor** → **Diagnose** → **Notify**

4/27/15

# Summary

- ✓ **System Diagnosability Overview**
- ✓ **Diagnosability Enhancements**
  - ✓ **Node Diagnostics**
  - ✓ **Nvidia GPU Diagnostics CUDA 6.5 Support**
  - ✓ **I/O Devices Diagnostics**
  - ✓ **HSS Diagnostics and Monitoring**
  - ✓ **System Notification**
  - ✓ **CLE Enhancements**
- ✓ **Future Considerations**

# Q&A

# Legal Disclaimer

*Information in this document is provided in connection with Cray Inc. products. No license, express or implied, to any intellectual property rights is granted by this document.*

*Cray Inc. may make changes to specifications and product descriptions at any time, without notice.*

*All products, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.*

*Cray hardware and software products may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.*

*Cray uses codenames internally to identify products that are in development and not yet publically announced for release. Customers and other third parties are not authorized by Cray Inc. to use codenames in advertising, promotion or marketing and any use of Cray Inc. internal codenames is at the sole risk of the user.*

*Performance tests and ratings are measured using specific systems and/or components and reflect the approximate performance of Cray Inc. products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.*

*The following are trademarks of Cray Inc. and are registered in the United States and other countries: CRAY and design, SONEXION, URIKA, and YARCDATA. The following are trademarks of Cray Inc.:  ACE, APPRENTICE2, CHAPEL, CLUSTER CONNECT, CRAYPAT, CRAYPORT, ECOPHLEX, LIBSCI, NODEKARE, THREADSTORM.  The following system family marks, and associated model number marks, are trademarks of Cray Inc.:  CS, CX, XC, XE, XK, XMT, and XT.  The registered trademark LINUX is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.  Other trademarks used in this document are the property of their respective owners.*

COMPUTE    |    STORE    |    ANALYZE