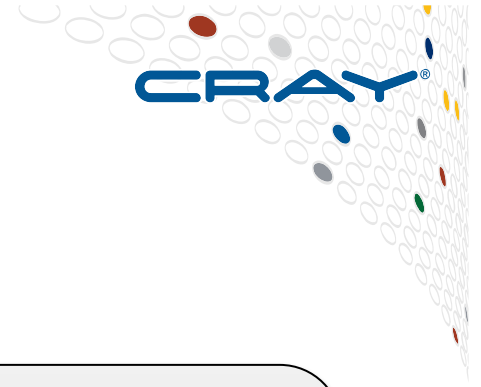


Cray XC System Level Diagnosability: Commands, Utilities and Diagnostic Tools for the Next Generation of HPC Systems

Jeffrey J. Schutkoske
Cray Administrative Environment (CAE Group)
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.



Overview

- **Introduction to System Diagnosability Toolkit**
- **Cray System Diagnosability Toolkit**
 - Aries High Speed Network
 - Intel Processor and Co-Processor
 - NVIDIA Processor
 - Cabinet Power and Cooling
- **Future Directions**



What is the System Diagnosability Toolkit?

- **System Diagnosability is a suite of software tools designed to provide Cray field support and end customers a tool chain to quickly and reliably identify hardware and software problems in the Cray XC system.**
- **Diagnostics are just one aspect of the tool chain that includes BIOS, user commands, power and thermal data and event logs.**
- **System Diagnosability features are built into the SMW and CLE commands so that hardware and software problems can be identified at the point of failure.**
- **System Diagnostics validate hardware and software are executed once a problem is suspected but can be periodically scheduled.**



System Diagnosability Toolkit

System Diagnosability Toolkit include:

- Command and Control tools
- Monitoring and Logging tools
- Recovery tools
- Notification tools
- System Diagnostics

System Diagnostics include:

- Boot level validation and tests
- Confidence level diagnostic tests
- Stress level tests
- Performance level tests
- Workload level tests

System Diagnosability Toolkit Usage



The right tool for the right job.

*“Once you eliminate the impossible, whatever remains,
no matter how improbable, must be the truth.”*

Sherlock Holmes

*System diagnosability toolkit allows system
administrators to eliminate the impossible....*

COMPUTE | STORE | ANALYZE



Aries HSN Diagnosability Toolkit

| Category | Toolkit |
|--------------------------------|---|
| Boot (HSS) | Power on MBIST BIOS PCIe training |
| Confidence (CLE) | BTE uGNI Tests: xtbte_ata and xtbte_ato FMA uGNI Tests: xtfma_ata and xtfma_ato FMA AMO uGNI Tests: xtfma_amo |
| Stress (CLE) | FMA & BTE Concurrent uGNI Test: xtfbc |
| Performance (CLE) | BTE All-To-All MPI/DMAPP Test: xta2a |
| Error and Data Reporting (SMW) | Linux Advanced Error Reporting (AER) PCIe Error Monitor: xtpcimon Aries Power and Thermal data |



Aries HSN Diagnosability Toolkit Example

xta2a results (no problem found)

| Bytes | Min | Mean | Max | Dev | Scaled |
|-------|------|------|------|-----|--------|
| 4096 | 5457 | 5603 | 5668 | 56 | 1.0% |

xta2a results (problem found)

| Bytes | Min | Mean | Max | Dev | Scaled |
|-------|------|------|------|-----|--------|
| 4096 | 3626 | 4897 | 4891 | 74 | 1.5% |
| 4096 | 3944 | 4916 | 4918 | 74 | 1.5% |
| 4096 | 4068 | 4916 | 4918 | 74 | 1.5% |
| 4096 | 3617 | 4915 | 4919 | 84 | 1.7% |

Bandwidth low for set 994 nodes 4056 4059: 3617GB/s

Bandwidth low for set 1051 nodes 4288 4291: 4012GB/s

xtpc output

Bad TLP/DLLP c1-0c1s10n2 0:3:0

COMPUTE | STORE | ANALYZE



Intel Processors Diagnosability Toolkit

| Category | Toolkit |
|--------------------------------|---|
| Boot (HSS) | BIOS processor, QPI, PCIe and memory |
| Confidence (CLE) | Processor Test: xtcpuburn Memory Tests: xtmemtester NUMA Tests: xtnumatest |
| Stress (CLE) | Processor Tests: xtcpudgemm |
| Performance (CLE) | Processor Tests: xtcpudgemm Memory Tests: xtnumatest |
| Error and Data Reporting (SMW) | Cray Embedded ITP scripts: xtitp MCA Errors: xthwerrlogd/xthwerrlog Intel MCA Decode: xtmcadecode Intel processor Power and Thermal data |



Intel Processor Diagnosability Toolkit Example

BIOS

QPI 0 Registers:
QPI Link Status: 0x0606FC00.
QPI MISC Status: 3.
QPI Error Counter[0]: 0x00000000.

xtnumatest -S 0x100 -s -l 3 -v

Running test08: ***** QPI Bandwidth Test *****
Expected socket 0 to socket 1 is: (22990 MB/s - 25410 MB/s)
Expected socket 1 to socket 0 is: (22990 MB/s - 25410 MB/s)

xtitp -t c0-0c0s7 qpi-status 1

Socket 0 QPI0:
Link Speed: 8.0 GT/s
Configured Tx Width: Full
Configured Rx Width: Full



Intel Co-Processors Diagnosability Toolkit

| Category | Toolkit |
|--------------------------------|--|
| Boot (HSS) | BIOS PCIe training |
| Confidence (CLE) | Co-processor Test: xtphidgemm PCIe Tests: xtphibandwidth |
| Stress (CLE) | Co-processor Test: xtphidgemm |
| Performance (CLE) | Co-processor Test: xtphidgemm |
| Workload (CLE) | GEMM, SHOC, and STREAMS |
| Error and Data Reporting (SMW) | MCA Errors: xthwerrlogd/xthwerrlog Intel MCA Decode: xtmcadecode Intel co-processor Power and Thermal data |



Co-Processor Diagnosability Toolkit Example

BIOS

PciBus: KNC detected @ B3|D0|F0

PciBus: KNC [B3|D0|F0] initialized, Width: x16, CurSpd: 5.0 Gbps

xtbounce

***** node_up *****

09:12:12 - Beginning to wait for response(s)

09:15:14 - Received 50 of 54 responses

ERROR: c0-0c0s8n0 - 370 – KNC PCIe link speed mismatch

xtphibandwidth -v -i 10

14:49:41 c0-0c0s8n0 nid00032 0, 128, 6223.58, 134184, 6481.61

14:49:41 c0-0c0s8n0 nid00032 1, 128, 6231.33, 107184, 6489.37

15:46:26 Failed:

Bandwidth actual: 107184,

Bandwidth expected greater than: 110000

COMPUTE | STORE | ANALYZE



Nvidia GPU SXM Diagnosability Toolkit

| Category | Toolkit |
|--------------------------------|--|
| Boot (HSS) | BIOS PCIe training |
| Confidence (CLE) | GPU Tests: xkdgemm GPU Memory Tests: xkmemtest PCIe Tests: xkbandwidth |
| Stress (CLE) | GPU, Memory, PCIe Tests: xkstress |
| Performance (CLE) | GPU Tests: xkdgemm |
| Workload (CLE) | HPL (Single NODE GPU) |
| Error and Data Reporting (SMW) | PCIe Error Monitor: xtpcimon GPU Memory Errors: xtaccecc Nvidia GPU Power and Thermal data |

COMPUTE | STORE | ANALYZE



Nvidia GPU Diagnosability Toolkit Example

```
xkdgemm -m 12100 -n 4096 -k 14100 -i 1 -l 10 -g -v
```

```
15:08:56 c0-0c2s13n2 nid00182 MPI mode enabled, root node  
Iteration, GFlops, Temp, Power, P-state, SMClockrate  
0, 1229.52, 44, 194.464, 0, 732
```

On Failure:

```
12:02:32 c0-0c1s5n1 nid00085 Failed: c0-0c1s5n1 nid00085  
8, 1295.29, 74, 194.934, 0, 745  
GFlops actual: 1295.29,  
GFlops expected greater than: 1300
```



Cray HSS Diagnosability Toolkit

| Category | Toolkit |
|--------------------------------|---|
| Boot (HSS) | HSS Microcontroller Power on validations HSS Microcontroller Pre-boot validations HSS Microcontroller Post-boot validations |
| Confidence (SMW) | System temp, power and status: xtcheckhss HSS tests: xtcheckhss |
| Error and Data Reporting (SMW) | HSS Faults, Errors & Warnings: xtconsumer System Environment Data Collections (SEDC) Processor Power & Thermal data |



HSS Diagnosability Toolkit Example

DIMM failure:

```
xthwerrlog -f ./hwerrlog.p0-20140503t150204 -M
```

| Node | Count | Chan | Type | DIMM BIT(s) |
|------------|-------|------|-------------|-------------|
| c0-0c0s7n2 | 26 | 0 | CORRECTABLE | J3000 |

Low Voltage Failure:

```
xtcheckhss --volts --blade=c0-0c0s7
```

| Component | Module | Sensor |
|------------|---------------------|------------------|
| HMIN | SMIN DATA | UNIT SMAX HMAX |
| c0-0c0s7n2 | qpdco_n0_s0_mem_vrm | vdd_vdr01_s0_c_i |
| 1200 | 1350 1339 | v*1000 1650 1800 |

COMPUTE | STORE | ANALYZE



Future Considerations

- **Additional workload tests**
- **Additional node stress tests**
- **Add end-to-end storage tests including the Infiniband**
- **Enhanced diagnostic data analysis tools**
- **HSS system dashboard**



Summary

- **Current focus has been on core HPC System**
 - Aries High Speed Network
 - Intel Processors and Co-Processors
 - Nvidia GPU
 - HSS Cabinet Power and Cooling
- **Future focus on the HPC System**
 - End-to-end storage tests
 - Additional Stress and workload tests
 - Data Analysis tools
 - System dashboard

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.