Cray XC System Node Diagnosability

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.

Overview

- Introduction to System Diagnosability
- Node Level Diagnosability
  - Initialization & Hardware Errors
  - Performance Errors
  - Out-Of-Band Diagnosis & Debug
- Q&A

What is the System Diagnosability?

- System Diagnosability is a suite of software tools
- Diagnostics are just one aspect of the toolset
- Features built into SMW and CLE commands
- System Diagnostics validate hardware and software
- System Diagnostics can be periodically scheduled

System Diagnostics

- **Reporting: Report and log errors, warnings, and faults**
- **Workload: Simulate customer workload**
- **Performance: Measure component performance**
- **Stress: Maximize hardware stress**
- **Confidence: Validate individual functionality**
- **Boot: Performed prior to booting CLE**

On-line Diagnostic Execution

- **On-line diagnostics are installed with CLE**
  - **Node/Aries**: /opt/cray/diag/default
  - **GPU**: /opt/cray/cray-nvidia/default
  - **KNC**: /opt/cray/cray-intel/default

- **Submit jobs through the batch or interactive mode**

```
aprun -n 2 -N 1 -L 28,29 ./xtfma_ata –R 2
```

Node Initialization

- **BIOS initializes the Intel processor & memory**
- **BIOS also discovers, initializes, and trains**
  - QPI bus
  - Aries PCIe bus
  - Nvidia GPU or Intel Co-Processor PCIe bus
  - I/O card PCIe bus
- **Reports link width, speed, and status**
- **BIOS logs are copied to the SMW on failure**

BIOS Aries Initialization Example

- **Aries Detected**

  Aries NIC detected @ B1|D0|F0, RevId=0x10

- **Aries Initialized**

  Aries NIC [B1|D0|F0] initialized, Width: x16, CurSpd: 8.0 Gbps

- **Aries Trained Successfully**

  Aries (B0:D2:F0) completed all PCIe Gen3 Phases successfully, LNKSTS2=0x1f

BIOS Aries Initialization Failure Examples

- **Aries Not at PCIe Gen 3 Speed**

  Aries (B3:D0:F0) not running at Gen3, PCIe Phase 1 did not complete, LNKSTS=0xx, LNKSTS2=0xx

- **Aries PCIe Link training failed**

  Aries (B3:D0:F0) link training failed, LNKSTS=0xx

- **Aries Not at PCIe Gen 3 Speed**

  Aries (B3:D0:F0) did not train to Gen3, LNKSTS=0xx

BIOS PCIe Initialization Failure Example

- **xtbounce indicates a PCIe Link Speed mismatch**

```
***** node_up *****
ERROR: c0-0c0s15n0 - 370 - SXM (GPU) PCIe link speed mismatch
```

- **Ensures device functionality at system boot time**

Node Hardware Errors

- The CLE kernel captures node hardware errors
- CLE Kernel console log
- CLE Kernel sends the errors to HSS via RCA
- Hardware Error Log Channel connected to the BC
- Logs and Errors saved on the SMW

Xthwerrlog DIMM Output Example

Node	Count	Bank	Type	DIMM
c1-0c0s1n0	1	8	CORRECTABLE	J10
c1-0c0s1n1	16	9	CORRECTABLE	J7
c1-0c0s7n2	50	9	CORRECTABLE	J11
c1-0c0s7n2	1	10	CORRECTABLE	J12
c1-0c1s1n1	24	9	CORRECTABLE	J11

Advanced Error Reporting (AER)

- **Enabled in the CLE kernel by default**
  - Aries
  - Nvidia GPU
  - Intel KNC

```
c0-0c0s7a0n1   CorrectableMemErr   0:0:0
        AER Correctable: Non-fatal error (mask bit: 1)

c0-0c1s6a0n2   CorrectableMemErr
        Link CRC error (cnt: 3)
```

Node Performance

- **Node validated using DGEMM**
- **Validates performance and data miscompares**

```
Cname:              c0-0c0s7n0
NID:                nid00028
Iteration:          0
GFlops:             24.1141
Power (W):          89
Processor actual:   515.110009109461
Processor expected: 514.110009109461
```

- **Supported on Node, GPU, and KNC**

Node Performance Example

Node	GFlops Min	GFlops Max	GFlops Avg	Bin	Eff Bin
c0-0c2s8n0	503.813	506.223	505.066	-332.98	-77.2488
c0-0c2s9n0	501.591	504.382	503.217	-341.659	-87.0555
c0-0c2s9n1	495.502	498.846	497.501	-365.447	-113.934
c0-0c2s8n1	489.865	493.652	492.368	-387.466	-138.814
c0-0c2s8n2	477.858	481.309	480.039	-434.366	-191.809
c0-0c2s8n3	477.336	479.782	478.613	-436.406	-194.114
c0-0c2s9n3	473.895	477.102	475.889	-449.848	-209.302
c0-0c2s9n2	**472.54**	**474.801**	**473.787**	**-455.14**	**-215.283**

Aries HSN Performance

- **Aries All-To-All performance test, *xta2a***
- **Measures performance on all-to-all communication**

Bytes	Min (GB/s)	Mean (GB/s)	Max (GB/s)	Dev	Scale
4096	*5457*	*5603*	*5668*	*56*	*1.0%*
4096	***3626***	***4887***	***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%***

Aries HSN Performance Results Analysis

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

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

Out-Of-Band Diagnosis

- Validate the HSS hardware and software
- HSS diagnostic utility, *xtcheckhss*
  - Cabinet
    - Blade
      - Aries Network Card (ANC)
      - Processor Daughter Card (PDC)
        - Node
        - GPU
        - KNC

xtcheckhss Example

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

Component: c0-0c0s7n2
Module: qpdc0_n0_s0_mem_vrm
Sensor: vdd_vdr01_s0_c_i

HMIN	SMIN	Data	Unit	SMAX	HMAX
1200	1350	**1339**	V*1000	1560	1800

Intel In-Target Probe (ITP) Debug

- **Intel In-Target Probe (ITP) is a JTAG bus**
- **Scripts reside on the SMW ➔ *xtitp***
- **Provide useful hardware and software debug information**
  - PCIe configuration and status
  - QPI configuration and status
  - Processor information, MCA errors, and MSR data
  - Package Power Limit (turbo) registers
- **Executing this command on the SMW temporarily pauses the processor**

ITP Debug Example

xtitp -t c0-0c0s7 qpi-status 1

```
Socket 0
        QPI0:
                Link Speed: 8.0 GT/s
                Configured Tx Width: Full
                Configured Rx Width: Full
                Tx Lane Status: 0xfffff
                Rx Lane Status: 0xfffff
                Error Counter 0: 0
                Error Counter 1: 0
```

Summary

- ✓ System Diagnosability Overview
- ✓ Node Level Diagnosability
  - ✓ Node Initialization & Hardware Errors
  - ✓ Performance Errors – Processor & Aries
  - ✓ HSS At Scale Out-Of-Band Diagnosis & Debug

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.