Resource Utilization Reporting

Two Year Update

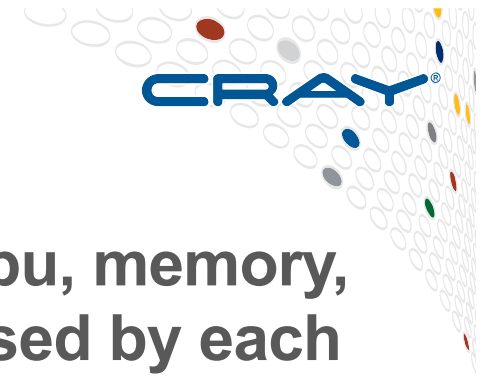Andrew Barry, Cray Inc.

Agenda

- **What Is RUR?**
- **How RUR works**
- **RUR in CLE**
  - 4.2up02 & 5.1up00
  - 5.1up01
  - 5.2up00
  - 5.2up01
  - 5.2up03
- **RUR in the datacenter**

COMPUTE | STORE | ANALYZE

What is RUR

- **RUR collects data about what resources (cpu, memory, energy, filesystem, accelerators, etc.) are used by each user application**

- **Runs before & after the application, thus low-noise**

- **Plugin architecture allows arbitrary data to be collected**

- **Output plugins allows storage to arbitrary destination**

- **Replaced a number of legacy accounting tools.**

How RUR works

- **RUR is a scalable framework for staging, collecting, post-processing, and reporting utilization data**
- **Data plugins for compute nodes**
- **Output plugins on login nodes**
- **Sites may create custom plugins**

RUR in CLE 4.2up02 & 5.1up00

- **Data plugins:**
  - Taskstats – CPU, memory, filesysetem I/O
  - Energy – Joules used, XC hardware only
  - Gpustats – Nvidia compute and memory resources
- **Output plugins:**
  - LLM – log file on SMW
  - File – Text file on some filesystem

RUR in CLE 5.1up01

- **Updated Taskstats with better exitcode reporting**
- **Timestamp plugin: application start / end times**
- **User output plugin: data to file in user's home directory**
  - ~/.rur/User_output_redirect: user want data elsewhere
  - ~/.rur/User_output_report_type: per-job, per application, one-file

RUR in CLE 5.2up00

- **Kncstats plugin – basic process accounting for KNC accelerators in autonomous mode**
- **Memory plugin – Lots of memory statistics, very high output volume**
- **Taskstats updates**
  - Xpacct – collect more detailed process accounting data
    - Pageswap count – minor/major, read/write calls, coremem/rss/vm, nice-value, block-IO-Wait-time, etc.
  - Per-process – collect details on every process on every node

RUR in CLE 5.2up01

- **Json-dict option for taskstats, energy**
  - {"coremem": 8421, "pgswapcnt": 0, "rcalls": 4…}
  - Json-list still available: ['wchar', 52, 'rchar', 78,…]
- **Expanded details in energy plugin with json-dict**
  - Accelerator usage, throttling information, etc.
- **User plugin gets opt-in option**
  - Only write to user's home directory if they want the data
  - Touch ~/.rur/user_output_report_type|user_output_redirect| user_output_optin

RUR in CLE 5.2up03

- Job-scope RUR
- Service node RUR
- RUR scalability enhancements
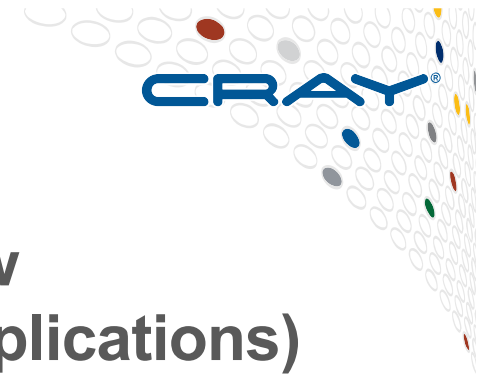
Job-scope RUR

- **Capture RUR data per-job, not per-application**
- **Capture utilization in reservation, idle nodes**
- **Idle nodes still consume energy**
- **Example: Application RUR reports 47172j of 50292j**

	Job 50292j	
Sum 47172j	Setup 4165j	Compute 43007j
Node1	Setup 4165j	Compute 10752j
Node2	Idle 1040j	Compute 10752j
Node3	Idle 1040j	Compute 10752j
Node4	Idle 1040j	Compute 10752j

Job-scope RUR

- **Many plugins (e.g. taskstats) don't give new information per-job. (sum of constituent applications)**

- **If using both job & application scope, use two config files, or use config-sets**

- **Tag RUR config file plugins with config-set names, will only be called when RUR Prologue/Epilogue called with matching config-set name**

Service Node RUR

- **Possible to target RUR to service nodes**
- **Should use alternate config file or config-set**
- **No current Cray plugins for service nodes, several are planned**

RUR scalability enhancements

- **RUR config, infrastructure, plugins, and python on /dsl**
- **Problem if too many computes per DVS server**
- **config file now sent in launch message**
- **Compute_local_python option "compiles" RUR with pyInstaller, stored in compute node ramfs**
- **Reduces DVS load and improves RUR scalability**

RUR in the Datacenter
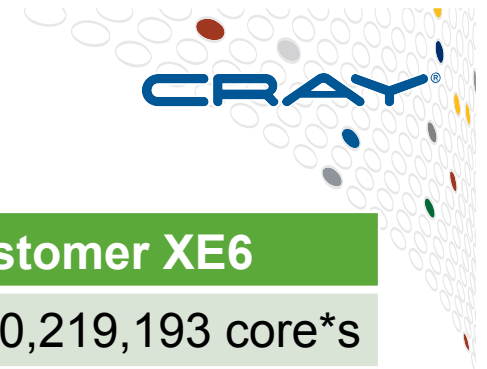
	Cray XC30	Customer XE6
Max CPU usage	8,862,578 core*s	1,730,219,193 core*s
Average CPU usage	6,706 core*s	12,356 core*s
Max FS writes	5277 GB	3460 TB
Average FS writes	1.5 GB	3.0 GB

- **Not Surprisingly: Big machines can and do run big jobs**
- **That said, even big machines run a lot of small jobs**

Summary

- **RUR has incrementally added functionality, and will continue to do so**
- **Now used by a number of customers, including some very large systems**
- **We would like to hear how you are using RUR, and what you think it is missing**

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