SSA, ClusterStor Call-home Service Actions, and an Introduction to Cray Central Telemetry and Triage Services (C2TS)

Jeremy Duckworth, Tim Morneau
Cray Inc.
Seattle, Washington, USA
{jeremyd,tmorneau}@cray.com

***Abstract*— The ClusterStor platform is designed to minimize a customer's operational burden (time and money) by offering an enterprise ready reliability, availability, and serviceability (RAS) solution. ClusterStor RAS, via SSA, can securely submit comprehensive diagnostic information to Cray in near real time, over the Internet. Using this data stream, Cray plans to generate proactive service opportunities and, in select cases, automate repair part shipment and service dispatch. As a complementary serviceability feature, Cray also plans to make it easier for customers to capture and securely transfer support data to Cray – by utilizing SSA on ClusterStor as a triage data collection framework. Finally, as a foundation for the future of call-home systems at Cray, this paper introduces the Cray Central Telemetry and Triage Services (C2TS) – including key motivations for Cray's work in this area and how C2TS relates to ClusterStor, SSA, and future products.**

Keywords: Customer Support, Telemetry, Remote Product Monitoring

I. INTRODUCTION

A. Overview

The Cray® System Snapshot Analyzer (SSA) provides a common set of product support data capture and call-home services. SSA and connected systems provide a support process automation platform that affords Cray and its customers more opportunities for proactive support, improved time to support issue resolution, and an opportunity for data-driven product quality improvements.

ClusterStor™ [1] is a highly-scalable, HPC storage system product line, designed with a robust reliability, serviceability and availability (RAS) feature set – including a key design focus on call-home telemetry. Cray is enhancing ClusterStor RAS call-home features through integration with SSA and connected services.

Cray Central Telemetry and Triage Services (C2TS) is a next generation, cloud-based, call-home system. C2TS is the evolution of Cray's call-home data storage and processing system. Cray plans to refactor SSA and key connected systems to interface with C2TS. Conceptually, C2TS affords Cray an opportunity to better align products and teams across its call-home service portfolio.

B. Motivation

The motivation for this paper is to open a dialogue with current and prospective customers on the product serviceability features and concepts discussed herein.

C. Paper Organization

For those not yet familiar with SSA or ClusterStor, Section II of this paper provides background information and references for further reading. Section III defines ClusterStor RAS telemetry types and examples. Section IV introduces Cray's plans to automate select repair part ordering, case management, and service actions using ClusterStor and SSA. Section V covers improvements Cray is planning to triage (support data) collection features on ClusterStor, using SSA. Section VI introduces Cray's next generation call-home architecture (C2TS). Section VII discusses future work, and the conclusion follows in section VIII.

II. BACKGROUND

This section provides background on Cray® SSA and ClusterStor, including a brief integration history of the two technologies.

A. Cray SSA Overview

SSA[2] is a Cray Global Technical Service (GTS) System designed to accelerate proactive service delivery across our product offerings. SSA is focused on the secure submission of product telemetry and support data from a customer system, to Cray.

As of May 10, 2018, SSA call-home is active at 21 customer accounts, with a total of 78 systems actively transmitting data to Cray on at least a daily basis. From June 2015 (initial release) to date of publish, the call-home system for SSA has ingested roughly 80,000 snapshots, processed 48,000 product-specific events and 414 service actions (cases created or triage case associations). There are approximately 7,500 snapshots in storage by SSA at any

[1] https://www.cray.com/products/storage/clusterstor

[2] https://www.cray.com/support/cray-system-snapshot-analyzer

given time (this number is relative to the reporting system population).

B. Cray SSA Value to Customer Support Process

SSA provides value to the support process by automating (1) the collection, submission and analysis of product diagnostic information (2) the collection, submission and analysis of product health information and (3) key aspects of the customer support process. This automation is enabled by a simple architecture that is based on making it easy to understand what information is being collected, how the information is transmitted to Cray and how Cray will then use the information. SSA can be used after a problem with a product is identified or, in some cases, SSA can proactively identify a problem. Through the collection of product configuration data on a routine basis, SSA can also track historical changes in a systems's configuration – offering another source of diagnostic information to aide in problem resolution.

C. Getting Started with SSA

SSA is currently available free of charge to customers with a Cray service contract. Supported platforms include the Cray® XE6™, XK6™, XK7™, XC30™, XC40™ and XC50™; the Cray® Sonexion™ 900™, 1600™, 2000™ and 3000™; and the Cray® ClusterStor™ 1500™, 6000™, 9000™, 300™ and L300N™. See Cray Field Notice (FN) 6122 for full compatibility information.

SSA is designed to make client installation and update easy. An SSA control panel in CrayPort[3] also makes SSA call-home activation easy. CrayPort Knowledge Article ID 4546, "*Getting Started with the Cray System Snapshot Analyzer (SSA)*" [1] contains procedures for downloading the SSA client, accepting the end user license agreement (EULA) [2] and managing activation of the upload service account. The SSA EULA also covers acceptable use of the information uploaded to Cray, via SSA. Release notes for each release and other release-specific documentation are available for download in CrayPort. The release notes also contain a reference to the SSA User Guide associated with the release.

D. Cray ClusterStor

Cray ClusterStor is a highly-scalable, HPC storage system product line. ClusterStor is designed to make high-performance storage simpler. The ClusterStor platform is designed to minimize a customer's operational burden (time and money), by offering an enterprise-ready reliability, availability and serviceability (RAS) solution. The ClusterStor RAS design goals include complete detection and isolation of customer and field replacement hardware components (i.e., CRUs and FRUs), with 95% isolation to a single hardware component. It is designed as part of the system, with the goal of keeping the system working when failures occur (reliability and availability). A foundational element for ClusterStor RAS features is the hierarchical, model-based inventory. This model facilitates automated, causal analysis based on structured relationships inherent in the design.

The ClusterStor platform has several interfaces that support integration into a larger data-center monitoring environment. Options include Nagios® and Ganglia plugins for fault and performance monitoring; simple network management protocol (SNMP) support and hyper-text transport protocol (HTTP), representative state transfer (REST)-ful application programming interfaces (APIs) for customized data-center integration. ClusterStor also supports local, e-mail based 'service alerts' and optional call-home support via Cray SSA. When hardware fails on the system, a guided repair facility is available for local use by properly trained personnel. Collectively, the RAS features on a ClusterStor system present a cohesive view of system status.

E. Brief History of ClusterStor and SSA Integration

Cray acquired the ClusterStor product in 2017. Prior to 2017, Cray was a reseller of the ClusterStor product under the Cray brand Sonexion. In September of 2016, SSA released the first client for Sonexion. In June 2017, SSA became the official RAS call-home transport for Sonexion. In March 2018, the first unified SSA client was released to support both Sonexion and ClusterStor systems. Cray now recommends that all ClusterStor and Sonexion customers configure their systems for call-home, via SSA.

Prior to the Cray acquisition of the ClusterStor product, SSA and ClusterStor were developed independently, leading to overlapping functionality in some cases and considerable divergence in others. Now that ClusterStor and SSA are being developed at Cray, Cray is working towards leveraging the best attributes of each solution going forward.

III. CLUSTERSTOR RAS TELEMETRY

This section introduces ClusterStor™ RAS telemetry types with supporting examples and focuses on standalone (local, on-product) RAS telemetry. Section IV builds upon this information to describe call-home behavior.

A. Oveview: Types of RAS Telemetry

The ClusterStor RAS subsystem generates three types of RAS-related telemetry messages; interesting event messages (IEMs), service event messages (SEMs), and machine reportable product data (MRPD). By design, IEMs and MRPD packages are not intended for customer use. Service events (a representational view of SEMs) are available local to a customer data center, via e-mail notifications, Nagios®, SNMP, and the on-product guided repair user interface.

B. ClusterStor RAS Rules Engine

The ClusterStor RAS subsystem contains an on-product rules engine. Abstractly, this rules engine is responsible for:

[3] https://portal.cray.com

a) analyzing incoming RAS telemetry events

b) updating rule engine state and context

c) orchestrating response actions based on defined rules, which may include the generation of additional RAS telemetry events

The telemetry type definitions follow. The examples provided are based on actual call-home events from a production system. These events captured a disk failure and replacement scenario in May 2018. Some of the information has been redacted or modified as to not identify the reporting system or event.

C. IEMs Defined

IEMs are perhaps best understood as structured, highly contextual diagnostic events that cover a broad range of topics (e.g., Lustre, disk and enclosure monitoring services, systems management). In **Figure 1**, an IEM expressing initial detection of a disk failure event is shown -- in JavaScript Object Notation[4] (JSON). Specifically, this IEM triggered disk drive failure evaluation processing. (i.e., verify a disk has failed over an evaluation window).

D. SEMs Defined

SEMs are generated when the rules engine asserts that a service action is needed and resolved when service is no longer needed (e.g., part replaced). The JSON in **Figure 2** illustrates a disk drive failure SEM. This SEM was ultimately triggered from the event sequence starting with the IEM shown in **Figure 1**. Notice in the SEM example that model-based location information (rack, enclosure, enclosure index) and disk drive model and configuration information is included. Each SEM is registered and tracked by the rules engine as a globally unique event. While not shown here, subsequent IEMs are generated and processed by the rules engine that resolved and closed this SEM after the disk drive was replaced.

E. Machine Reportable Product Data (MRPD)

Whereas IEMs and SEMs represent small, discrete events, MRPD represents a structured system snapshot, similar to a daily SSA snapshot. So, where IEMs and SEMs are typically smaller in size and focused, MRPD are larger, span multiple blobs and are generally a more comprehensive view of the system state at a certain point in time. MRPD packages are triggered on at most a daily basis, and contain a wide array of inventory, status and diagnostic information.


```
{
"local_time": "Thu, 03 May 2018 14:25:01 CST",
"message_type": "iem",
"messages": [
{
"data": {
 "confirmed_time": 0.0,
 "creation_time": 1525375281.394192,
 "event_code": "001001001",
 "event_data": {
  "dcs_timestamp": 1525375281,
  "enclosure_serial_number": "REDACTED-ENCL-SERIAL",
  "fru_serial_number": "REDACTED-DRIVE-SERIAL",
  "index": "13",
  "status": "failed",
  "type": "disk"
 },
 "id": "REDACTED-ENCL-SERIAL:disk:13",
 "resolved_time": 0.0,
 "state": "EVALUATION",
 "version": 3
},
"event_code": "001002001",
"event_description": "Rules Engine Event changed state. Event
created and set to EVALUATION for event.",
"host": "redactedn000",
"program": "plex",
"severity": 6,
"timestamp": 1525375281
}
],
"sequence_number": 1028,
"system_identifier": "[not-set]",
"system_serial_number": "REDACTED-SYS-SERIAL",
"utc_timestamp": 1525375501,
"version": 4
}
```


Figure 1. JSON IEM Example: Possible disk drive failure, confirmed via later IEM

[4] https://www.json.org/

```
{
"local_time": "Thu, 03 May 2018 14:39:45 CST",
"message_type": "sem",
"messages": [
{
"completion_time": 0.0,
"confirmed_time": 1525376181.398169,
"creation_time": 1525376181.3996589,
"dcs_timestamp": 1525375281,
"event_code": "002005001",
"event_description": "Disk drive needs replacement",
"fru": {
 "disk_installed": "1",
 "dm_report_t10": "11110111100",
 "firmware": "E0G5",
 "manufacturer": "SEAGATE",
 "part_number": "ST6000NM0034",
 "sect": "512",
 "serial_number": "REDACTED-DRIVE-SERIAL",
 "status": "Failed",
 "t10_enabled": true
},
"location": {
 "enclosure_location": "16U",
 "enclosure_model": "5U84",
 "index": "13",
 "rack": "R1C4"
},
"re_event_code": "001001001",
"state": "SVC_CREATED",
"type": "disk",
"uuid": "41eaa29e-4f09-11e8-8bc0-000000000000"
}
],
"sequence_number": 1031,
"system_identifier": "[not-set]",
"system_serial_number": "REDACTED-SYS-SERIAL",
"utc_timestamp": 1525376385,
"version": 6
}
```


Figure 2. JSON SEM Example: Disk drive needs replacement

IV. CALL-HOME SERVICE ACTIONS

This section provides a high-level description of how ClusterStor, SSA and connected systems are used to orchestrate service actions. It also introduces Cray's plans to automate aspects of the field replaceable unit (FRU) repair part ordering, and provide proactive service alerts based on call-home data, for ClusterStor.

A. Overview: Components and Orchestration

Figure 3 illustrates the key relationships amongst ClusterStor RAS components on a standalone ClusterStor system. ClusterStor components produce IEMs which are then analyzed by the RAS rule engine, using predefined rules. The rule engine updates internal state and can optionally generate one or more IEM or SEM events. The rule engine can also use the events it generates as input for future engine operation.

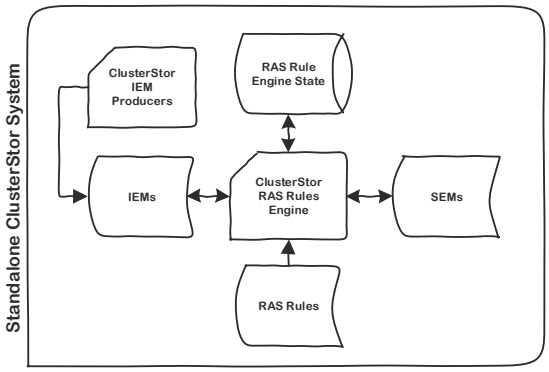

Figure 3. RAS Rule Engine Flow – standalone ClusterStor System

Figure 4 extends **Figure 3** through introduction of call-home processing for IEMs and SEMs. **Figure 4** illustrates SSA as both a secure network transport and as a component that orchestrates Cray service automation tasks. Note that the ClusterStor call-home rules and, to some degree, the rule engine can diverge from the on-product ClusterStor RAS instance. The flexibility afforded in this configuration allows Cray to quickly add rules for emergent issues and ultimately improve upon the on-product instrumentation with insights gained since the product was installed.

B. Automating Field Replaceable Unit (FRU) Repair Part Orders

In 2017, approximately 1,932 repair part orders were placed for ClusterStor Systems and processed by Cray[5]. Approximately 82% of these orders were for disk drives. Each repair part order was largely a manual process, requiring Cray and often customer time and effort to complete. To reduce the time required and human errors involved in this process, Cray plans to begin automating FRU replacement orders on ClusterStor systems with SSA enabled[6]. Early feature availability is scheduled to start in 2018.

Considering the longest possible repair part lead time, the canonical FRU failure, replacement, and return merchandize authorization (RMA) process follows:

1) A FRU fails on a system, possibly leading to degraded system state

2) An operator notices the FRU failure, ideally via a monitoring interface (e.g., service events as e-mailed from the ClusterStor System)

[5] This number does not include orders shipped by Seagate, prior to Cray's acquisition of the ClusterStor product

[6] Other requirements will be announced coincident with feature availability

3) An operator assesses the impact of the FRU failure, typically including the location of the failure, the severity of the failure and any identifying information on the FRU

4) An operator opens a Cray service case with details on the FRU failure and requests a replacement part

5) Once the replacement part arrives, an operator replaces the FRU[7]

6) An operator monitors for the system to exit a degraded state, if applicable

7) An operator completes RMA processing and ships the faulty FRU back to Cray Inc.

The first three steps of this process are largely automated today, for ClusterStor FRUs. A ClusterStor system can be configured to notify operations staff of service events. These service events contain identifying information about the FRU and its location. As previously stated, SEMs are the basis for service events.

Cray is developing a feature set to automate the fourth step in the process. This feature will leverage IEMs, SEMs, context from SSA, and a centralized ClusterStor RAS rules engine (as shown in **Figure 4**). The rules engine will perform secondary validation of FRU failures, and, assuming the failure is validated, submit the failure to SSA to orchestrate case and part order tasks. FRU failure validation is included to ensure the most appropriate action is taken based on the best information available at Cray, reducing noise in the service process. The orchestration will also ensure appropriate team members are notified as the process progresses (e.g., as the part ships, is tracked, etc.). The goal of this work is to eliminate the time required by an operator to create and populate a service case and create a part order.

C. Proactive Service Recommendations

By combining Cray product service and engineering expertise with call-home intelligence from multiple customer systems and a centralized rule engine, Cray plans to introduce proactive service recommendations for ClusterStor. As an example, the system may recommend a software update to a specific ClusterStor system based on the availability of a newer release or a critical issue in a previous release. The recommendations may also include suggestions to change a system configuration setting, look closer at a marginal hardware component or investigate other, more subtle issues.

[7] An immediate replacement can occur if local spares are available

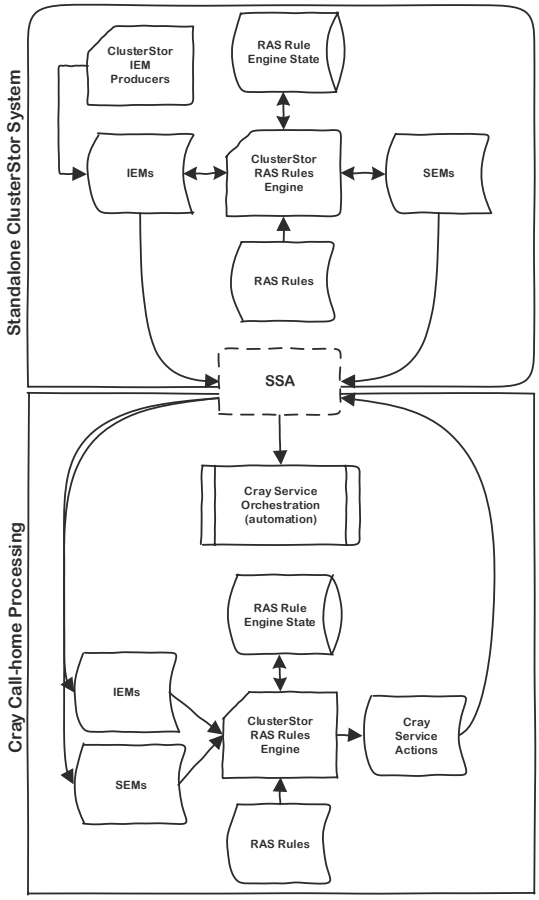

Figure 4. Call-home model for ClusterStor Service Actions

If Cray determines that customer contact is warranted for a proactive service event, Cray service will initiate contact directly with the customer after first evaluating the recommendation. Over time, this process is expected to become more automated and less gated for trusted recommendations. Intelligence leveraged in this process may also be selectively included in future on-product ClusterStor rule engine updates – with the potential to benefit all customers.

V. CLUSTERSTOR SUPPORT DATA CAPTURE AND SSA

The ClusterStor™ platform currently has a support data collection sub-system that generates 'support bundles'. Support bundles contain data used by Cray customer service and product engineering to identify root cause for product support issues. Support bundles can be triggered automatically by the system (e.g., in response to a node failover event) or they can be triggered by systems administrators. Support bundles are currently available directly on the platform and integrated with the common user interfaces within ClusterStor. At the request of Cray support, support bundles can be uploaded to Cray, manually, via SFTP or FTP.

In 2018, the SSA and ClusterStor teams will begin the process of replacing the current support bundle mechanism

with SSA triage collections and we plan to release incremental improvements to both the collection interfaces (mechanics) and manifest (what is included in a target collection). SSA can capture support data from ClusterStor systems today, but Cray plans to improve and consolidate this facility over time.

VI. Cray: Next Generation Call-home Architecture

SSA was first introduced at CUG 2015, after nearly two years of internal development at Cray. As stated in section II, SSA is now actively processing data for 45 customers and 78 Cray systems globally – with support for numerous Cray platforms. As Cray provides 24x7x365 support to customers globally, SSA and ancillary systems must be supported in the same way. Over the past five years, Cray has made incremental improvements to SSA through careful consideration of feedback from Cray service and product teams, customers and the SSA team. However, the initial architecture of the SSA back-end, and hence SSA, was found to be an ill-suited foundation for necessary, longer term enhancements – including closer integration with current and future Cray platforms. In response, architectural design for the Cray Central Telemetry and Triage Services (C2TS) platform started in December 2017. Cray motivations for C2TS, its relationship to SSA and Cray platforms, and an architectural overview for C2TS are presented in this section.

A. Motivations for C2TS

Many of the motivations below necessarily overlap or are interdependent. An effort is made to minimize redundancy in the treatment of each category.

Availability

SSA and connected systems have a target service level agreement (SLA) north of 99%. With the existing architecture, attainment of this SLA has not been realistic with current maintenance requirements and scaling properties. To address this issue, Cray plans to deploy C2TS and critical ancillary systems in a cloud environment; employ highly-resilient, horizontally-scaled services; and utilize load-balanced, scaling tiers that enable rolling service updates.

Scalability

The existing architecture is, for the most part, vertically scaled. As the number of systems, amount of data and size of data has increased, the scaling limits of the current architecture have become apparent. C2TS will address performance scalability through the use of horizontally-scaled services and utilize load-balanced, scaling tiers. C2TS will address snapshot-size specific scale using a content-addressable storage schema that provides flexible storage options for system 'snapshots'. The content-addressable snapshot schema also has the potential to provide better network transfer resiliency and performance for customer snapshot uploads.

Usability

Cray service teams use SSA snapshots routinely to proactively research the configuration of a reporting system, while working customer issues (e.g., what is the software release level and hardware inventory). They also use SSA to request triage snapshots (support data) to support root cause analysis and ultimately to help a customer resolve a support issue. The SSA team continuously works with service and product teams to add or alter 'what goes in a support data collection'. In the current architecture, the snapshot data model is not an ideal medium through which to express complex system topologies. The data model is also constrained to a few primitive data types that do not allow desired expression of various data sources on Cray products. Since it is difficult to fully express product data in the current architecture, Cray service teams have requested several usability enhancements related to snapshot search-ability, format and exception reporting. In response, C2TS introduces a new snapshot schema. This snapshot schema includes (content-addressable) blob storage for contents and JSON metadata to fully describe attributes of the reporting system and data collection services, collection qualifiers (e.g., reason, exception level, and start and end times to gather data within), and service case associations. As a single entity, the JSON metadata will also fully describe snapshot contents (manifest) using base data types and extended attributes. Perhaps most importantly, tooling will be provided that translates a machine-readable snapshot to a view – using the model-view-controller (MVC) pattern – making it possible to present multiple views of the same snapshot. Based on this model, C2TS is designed to provide search service indexing and interfaces for flexible snapshot metadata search. The enhancements introduced in this area are also expected to make SSA snapshot data considerably easier to use when SSA is operating in a disconnected mode (e.g., not calling-home).

The second usability-related motivation relates to data processing models. In the current architecture, all data that comes back to Cray via SSA is reported in a batch snapshot format. For several use cases (product health, alerts, and inventory extraction), snapshot data is transformed via an extract, transform and load (ETL) process. This transformation most often results in a contextualized event format that is used as input into a complex event processing (CEP) service and ultimately used in more complex orchestration tasks. C2TS embraces streaming data processing for this reason, ultimately with the vision that data requirements for several current and future use cases will be provided in this way.

Finally, and perhaps most importantly, Cray desires to make call-home information more accessible to customers. Along with changes to schemas and processing models, this requires enhancements in the pipeline between SSA, C2TS and CrayPort.

Programmability

C2TS, as a platform, is planned to be considerably more programmable for Cray developers. Notable improvement here will be better documentation of data schemas and developer documentation, the use of well-known interfaces and the use of better architectural standards.

Serviceability

The current architecture is based on a closed source, proprietary data warehousing system at its core and Cray services developed to interface with said system. Plans for C2TS are based on open source and Cray developed services that should ostensibly be easier to support.

B. SSA's Relationship to C2TS

Cray does not plan to replace SSA with C2TS. C2TS will enhance SSA and provide a solid foundation for future services. In the existing architecture, SSA encompasses both the call-home interfaces on Cray Products (e.g., the SSA client) and services in the Cray back-end. In the latter area, SSA is also connected to various Cray internal and business systems. C2TS replaces the core data processing and storage services for SSA. SSA will use data lake services in C2TS for upload of product information and to orchestrate back-end customer support automation workflows.

C. Early C2TS Architecture

The first architecture for C2TS is code-named Metis, and is under active development. Initial deliverables for C2TS, internal to Cray, are focused on snapshot processing. While the architecture is designed to generalize across and benefit all Cray platforms, initial work on C2TS will support Cray's next generation supercomputing platform. Specifically, the SSA client is currently being refactored to leverage a new snapshot schema (as previously described) and schema-compliant collection APIs. A functional architecture diagram for C2TS (Metis) illustrating an instrumented Cray platform (SSA Client) and the C2TS data lake is provided in **Figure 5**. Applications that use data lake services are not included in the diagram for brevity.

VII. FUTURE WORK

Future work includes the call-home service actions for ClusterStor and support data capture improvements introduced in this paper. The development of and migration to C2TS are also focal areas for future work. Cray also plans to better integrate call-home information into CrayPort. Finally, Cray plans to continue making incremental improvements in SSA for currently supported Cray platforms.

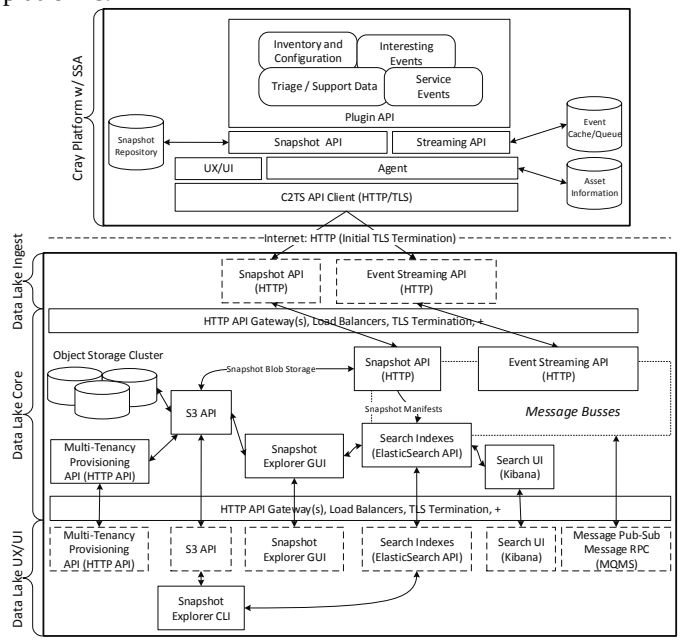

Figure 5. Early C2TS Functional Architecture

VIII. CONCLUSION

The Cray ClusterStor RAS sub-system is designed to enable proactive product support via call-home telemetry. SSA provides a secure call-home transport and a service orchestration platform to complement the ClusterStor RAS sub-system, yielding an enhanced proactive support capability. SSA also provides a consistent support data capture platform that Cray plans to better incorporate into ClusterStor and its next generation supercomputing platform. The Central Telemetry and Triage Services (C2TS) is Cray's next generation call-home platform. Design for the C2TS Architecture is heavily informed from five years of lessons learned in SSA. C2TS requirements are being driven by SSA, Cray's next generation supercomputing platform and ClusterStor.

ACKNOWLEDGEMENT

The authors would like to thank the Cray® User Group (CUG) for the opportunity to present this paper at the 2018 conference and Cray customers for using and providing valuable feedback on Cray's call-home services. The authors would also like to thank their colleagues at Cray – who provided expertise for, and critical review of, this paper.

AUTHORS

Jeremy Duckworth works on the System Snapshot Analyzer (SSA) team at Cray® and is responsible for SSA and Central Telemetry and Triage (C2TS) Architectures.

Tim Morneau works on the ClusterStor Reliability, Availability and Serviceability (RAS) team at Cray. Tim has been closely involved in the development and deployment of call-home systems for ClusterStor.

REFERENCES

[1] Cray Customer Portal (CrayPort) KB 4546, "Getting Started with the Cray System Snapshot Analyzer (SSA), (2018, March 28). [Online].
Available:
https://portal.cray.com/Support/apex/ka_how_to?Id=kA3330000008ToACAU

[2] Cray System Snapshot Analyzer (SSA) End User License Aggrement (EULA) (2018, March 26). [Online].
Available:
http://www.cray.com/sites/default/files/resources/Cray-SSA-EULA.pdf