

Urika®-GX System Overview (1.0.UP00) S-3017

Contents

1 About the Urika®-GX System Overview Publication	3
2 Features of Urika-GX High Performance Analytic Platform	5
3 Urika-GX Hardware Components	8
4 Urika-GX Network Components	10
5 Urika-GX Software Components	12
6 Urika-GX File Systems	13
7 Urika-GX Workload Management Tools	14
8 Urika-GX System Management and Monitoring Tools	15

1 About the Urika®-GX System Overview Publication

This publication describes the features, hardware and software of the Cray® Urika®-GX system. It also provides an overview of the analytic applications and third party tools installed on Urika-GX. This publication addresses product version 1.0.UP00 of the Urika-GX system.

Typographic Conventions

Monospace	Indicates program code, reserved words, library functions, command-line prompts, screen output, file/path names, key strokes (e.g., $Enter$ and $Alt-Ctrl-F$), and other software constructs.
Monospaced Bold	Indicates commands that must be entered on a command line or in response to an interactive prompt.
Oblique Or Italics	Indicates user-supplied values in commands or syntax definitions.
Proportional Bold	Indicates a graphical user interface window or element.
\ (backslash)	At the end of a command line, indicates the Linux® shell line continuation character (lines joined by a backslash are parsed as a single line). Do not type anything after the backslash or the continuation feature will not work correctly.

Scope and Audience

The audience of this publication is users and system administrators of the Urika®-GX system. This publication is not intended to provide detailed information about open source products used in the system. References to online documentation are included where applicable.

Record of Revision

Date	Description
March 2016	This version addresses release 0.5.UP00 of Urika-GX systems
August 2016	This version addresses release 1.0.UP00 of Urika-GX systems

Trademarks

The following are trademarks of Cray Inc. and are registered in the United States and other countries: CRAY and design, SONEXION, Urika-GX, Urika-XA, Urika-GD, and YARCDATA. The following are trademarks of Cray Inc.: APPRENTICE2, CHAPEL, CLUSTER CONNECT, CRAYDOC, CRAYPAT, CRAYPORT, DATAWARP, ECOPHLEX, LIBSCI, NODEKARE. 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.

Feedback

Visit the Cray publications portal at http://pubs.cray.com and make comments online using the Contact Us button in the upper-right corner or Email pubs@cray.com. Your comments are important to us and we will respond within 24 hours.

2 Features of Urika-GX High Performance Analytic Platform

The Cray® Urika®-GX system is a big data platform optimized for analytic work-flows. It combines a highly advanced hardware platform with a comprehensive analytic software stack to help derive optimal business value from data. The Urika-GX platform provides the tools required for capturing, analyzing and organizing a wide variety of data types from different sources.

Major features of the Cray® Urika®-GX system include:

Hardware and Networking

- Cray® GreenBlade™ based technology Urika-GX rack and sub-rack enclosures use GreenBlade technology, which provides reliable and high-quality blade-based solutions. GreenBlade's unique design decreases power consumption per node, thereby significantly reducing data center energy costs. For more information, see "Urika®-GX Hardware Guide".
- Aries High Speed Network (HSN) The Aries HSN provides high speed application and data network connectivity between system nodes. For more information, see S-3016, "Urika®-GX System Administration Guide".

• System Management

- Workload management tools There are a number of resource/workload management tools installed on the Urika-GX system, including Mesos, Marathon and YARN. These tools enable management of analytic workloads, dynamically allocate system resources to applications as needed, and provide the flexibility of running multiple jobs across the cluster concurrently. For more information, see S-3015, "Urika®-GX Analytic Applications Guide".
- Performance analysis tools Urika-GX features a set of performance analysis tools to analyze performance of Urika-GX's analytic applications. For more information, see S-3015, "Urika®-GX Analytic Applications Guide".
- Urika-GX CLI Urika-GX features a number of CLI commands that enable monitoring and managing analytic services. These commands are also integrated in to the system management UI for ease of use.
- Cray System Management Software (CSMS) CSMS is a system management tool that is based upon OpenStack, which is a scalable cloud computing platform that provides an Infrastructure-as-a-service (laaS) solution through a set of interrelated services, and features a dashboard and a CLI. To learn more about OpenStack, visit http://www.openstack.org.

The Cray System Management Software interfaces with the Cray Hardware Supervisory System (HSS) to provide provisioning as well as image and instance management capabilities on the Urika-GX system. For more information, see S-3016, "*Urika*®-*GX System Administration Guide*".

• Analytic Application Development Tools

 Urika-GX analytic programming environment - Urika-GX features an analytic programming environment that comprises of a set of tools that facilitate building, debugging, and compiling various types of analytic applications using various programming languages, including Python, R, JAVA, Spark etc. For more information, see S-3015, "Urika®-GX Analytic Applications Guide".

- Jupyter Notebook Urika-GX ships pre-installed with the Jupyter notebook, which enables creating
 executable documents. In additional, there are some cookbooks that Urika-GX ships with that provide
 information about using the system. For more information, see S-3015, "Urika®-GX Analytic Applications
 Guide".
- Cray Graph Engine (CGE) Urika-GX also features the Cray Graph Engine (CGE) for performing complex graph analytic tasks on RDF databases at interactive speeds.
- Cray Application Management Urika-GX features a user friendly UI for monitoring and managing jobs submitted to the system. In addition, the system features the Urika®-GX Applications Interface, which provides access to various system and analytic applications installed on the system. For more information, see S-3015, "Urika®-GX Analytic Applications Guide" and S-3016, "Urika®GX System Administration Guide".
- Additional analytic tools
 - Urika-GX system features a number of data analytic tools that help perform analytic tasks, such as
 executing Hadoop and SPARK jobs. For more information, see S-3015, "Urika®-GX Analytic
 Applications Guide".
 - mrun is a Cray-built application launcher program that enables running parallel jobs on Urika-GX using resources managed by Mesos/ Marathon. In addition to launching jobs, mrun also has the ability to display currently active Mesos Frameworks and Marathon applications, provide extensive details on running Marathon applications, as well as cancelling/ stopping currently active Marathon applications. The Cray Graph Engine (CGE) uses mrun to launch jobs under the Marathon framework on the Urika-GX system.
- **Support for external storage** Urika-GX systems feature support for external POSIX-compliant storage systems. Currently, Lustre® is supported as an external file system.
 - If the Cray Sonexion system is being used as an external storage system, refer to "Sonexion Administrator Guide" at https://pubs.cray.com
- **Fault tolerance** A number of Urika-GX's analytic applications are optimized for system resiliency and fault tolerance. For more information, see S-3015, "*Urika®-GX Analytic Applications Guide*".
- Flexible configurations The Urika-GX rack can support up to 3 sub-rack chassis, each of which contains 16 nodes. The Urika-GX system contains three types of nodes:
 - Compute nodes Compute nodes run application programs and services.
 - I/O nodes I/O nodes facilitate connecting to external storage and file systems. Urika-GX currently supports Lustre as an external file system.
 - Login nodes Login nodes enable users to log on to the Urika-GX system and facilitate launching jobs from the command line.

The Urika-GX system contains 2 login and 2 I/O nodes, whereas the remaining nodes in the system are all compute nodes.

- Node options
 - Processor: Intel® Broadwell 18C E5-2697 v4 or 8C E5-2620 v4 (2 per node)
 - Memory: 128, 256 or 512 GB per node (32 GB or 16 GB DDR-2400 DIMMS/DDR-2133 DIMMs) (8 or 16 per node)
 - HDDs: 2 TB or 1 TB SATA 7.2K 2.5" 6 Gb/s (2 per node) processors
 - SDDs: 0.8 TB, 1.6 TB or 4 TB Intel® PCIe (1 per compute node)
- Connectivity options

- I/O nodes Based on the external storage, the following PCIe HBA options are available for I/O nodes (only one can be selected out of the following):
 - LSI Logic LSI00343 9300-8E SGL 8-Ports Ext 12GB/S SATA SAS PCIE3.0
 - Mellanox MCX354A-QCBT ConnectX-3 VPI Adapter Card Dual-Port QSFP QDR IB (40Gb/s) and 10GbE PCIe3.0 x8
 - Mellanox MCX354A-FCBT ConnectX-3 VPI Adapter Card Dual-Port QSFP FDR IB (56Gb/s) and 40/56GbE PCIe3.0 x8
 - ConnectX-4 VPI cards
 - Mellanox MCX455A-ECAT ConnectX-4 VPI adapter card, EDR IB (100Gb/s) and 100GbE, single-port QSFP, PCIe3.0 x16
 - Mellanox MCX456A-ECAT ConnectX-4 VPI adapter card, EDR IB (100Gb/s) and 100GbE, dual-port OSFP, PCIe3.0 x16
- Login nodes Login nodes support on-board GigE connectivity by default. A PCIe card supporting
 10GigE connectivity can optionally be added to a login node for connectivity to an external network.
- Processor options (compute nodes) Intel[®] Broadwell 18C E5-2697 v4 or 8C E5-2620 v4 (2 per node)
- Hard disk drive options (compute nodes)
 - 2TB SATA storage per node provided by 2 x 1TB SATA 7.2K 2.5" 6GB/S HDDs
 - 4TB SATA storage per node provided by 2 x 2TB SATA 7.2K 2.5" 6GB/S HDDs. This is the default configuration.
- Memory options (compute nodes)
 - 128GB per node provided by 8 x 16GB DDR4-2133 DIMMs
 - 256GB per node provided by 8 x 32GB DDR4-2133 DIMMs. This is the default configuration.
 - 512GB per node provided by 16 x 32GB DDR4-2133 DIMMs
 - DDR4 2400MHz DIMM availability with broadwell processors

All compute nodes include 800GB of PCIe 3.0 SSD storage and are configured identically in terms of processor, memory, HDD, and SSD storage. I/O nodes do not support SSDs. Login nodes do not include SSD storage by default and may optionally include a PCIe SSD. If an SSD is included, it needs to be the same SSD that is installed on compute nodes. If an SSD is not needed on the login node, network bandwidth can be improved by installing one of the following optional PCIe network adaptors:

- Mellanox MCX312A-XCBT ConnectX-3 Network Interface Card Dual-Port 10GbE PCIe3.0 x8
- Mellanox MCX354A-QCBT ConnectX-3 VPI Adapter Card Dual-Port QSFP QDR IB (40Gb/s) and 10GbE PCIe3.0 x8
- Mellanox MCX354A-FCBT ConnectX-3 VPI Adapter Card Dual-Port QSFP FDR IB (56Gb/s) and 40/56GbE PCle3.0 x8
- Mellanox MCX455A-ECAT ConnectX-4 VPI adapter card, EDR IB (100Gb/s) and 100GbE, single-port QSFP, PCIe3.0 x16
- Mellanox MCX456A-ECAT ConnectX-4 VPI adapter card, EDR IB (100Gb/s) and 100GbE, dual-port QSFP, PCIe3.0 x16
- Mellanox MCX415A-CCAT ConnectX-4 Network Interface Card, 100GBE single-port QSFP, PCIe3.0 X16
- Mellanox MCX416A-CCAT ConnectX-4 Network Interface Card, 100GBE dual-port QSFP, PCIe3.0 X16

3 Urika-GX Hardware Components

Major hardware components of the Urika®-GX system rack include:

Sub-rack Enclosure

There are up to three GreenBlade 10.5 Rack Unit (RU) SR10216 sub-rack enclosures contained within the Urika-GX system rack.

Each Urika-GX sub-rack contains the following components:

- Nodes The Urika-GX system rack supports up to three sub-racks, each of which contains 16 GreenBlade™
 GB522XA blades/nodes. All Urika-GX nodes run the CentOS 7.2 operating system and portions of the Cray
 Linux Environment (CLE).
- Aries Network Card Controller (ANCC) Each sub-rack chassis of the Urika-GX rack contains two Dual Aries Network Card (dANC) boards. Each dANC board in turn contains 2 Aries Network Card Controllers (ANCCs), an Advanced RISC Machines (ARM) processor, and a number of environmental sensors. The ANCCs support High Speed Network (HSN) connectivity to all the nodes.
- Intelligent Server Control Board (iSCB) Each sub-rack contains one iSCB module, which reports the
 system's health status. The iSCB module interfaces with the Hardware Supervisory System (HSS), which is
 used for system monitoring and management.

In addition to the aforementioned components, each Urika-GX sub-rack contains 6 Power Supply Units (PSUs) and 6 cooling/fan modules.

System Management Workstation (SMW)

The SMW is a Dell R630 server that acts as a single-point interface to a system administrator's environment. The SMW provides a terminal window for performing administrative tasks, such as adding user accounts, changing passwords, and monitoring nodes and applications.

The SMW contains an Intel Xeon E5-2623 v3 3.0 GHz processor, 64GB of RAM and 3TB of physical storage. The base operating system of the SMW is CentOS 7.2

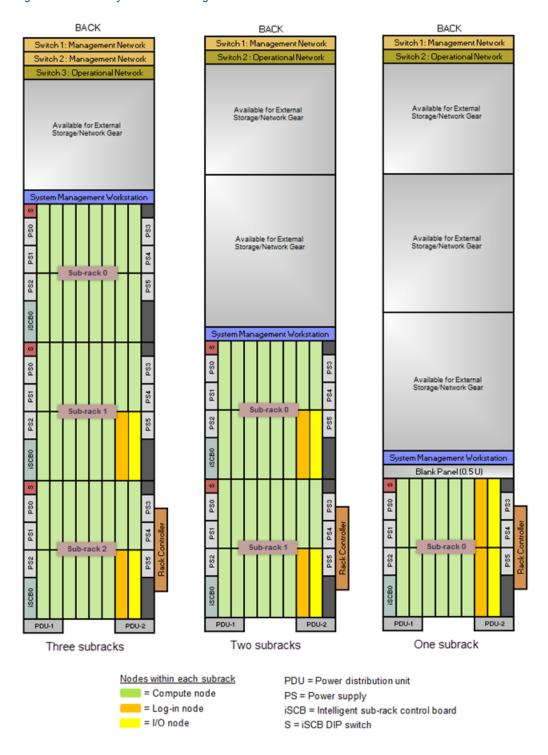
Switches

The Urika-GX rack contains three single RU 48-port GigE Brocade ICX 6450 Ethernet switches for management and external connectivity.

Rack Controller (RC)

The RC acts as a component of the HSS and manages communication between the SMW and the rack, sub-rack, dANCCs, and, in some cases, nodes. It monitors the sub-rack's temperature and other external environmental factors and is installed on the front interior of the rack side panel.

Figure 1. Urika-GX System Rack Configurations



4 Urika-GX Network Components

There are 3 networks deployed on the Urika®-GX platform:

- Aries High Speed Network (HSN) The Aries HSN provides high speed application and data network
 connectivity between nodes. This network provides node interconnect via high bandwidth, low latency DMA
 access. The hardware to support this network consists of an Aries Interface Board (AIB) connected to an
 available PCIe slot on each Urika-GX node and integrated into the node chassis assembly. The AIB is
 connected to the dANC integrated in the Urika-GX sub-rack. Copper cables provide an all-to-all connection of
 all dANCs in the system.
- Operational Ethernet network- The operational Ethernet network is used for ingesting user data. This network is comprised of a single unit 48-port GigE switch that provides dual 1GigE and/or dual 10GigE interfaces to the site network. Urika-GX's login nodes do not route through this switch and need to be directly connected to the site network. The operational network allows node connectivity externally from Urika-GX to the site network. The Urika-GX compute and I/O nodes are connected to a single managed Brocade ICX 6450-48, 48 port switch with a single power supply. Connectivity of this network to the site network is made possible by two available Gigabit Ethernet ports and/or two 10 Gigabit Ethernet ports on the ICX 6450-48 switch.
- Management Ethernet network The management Ethernet network is primarily used for system
 management, and not for user data. The management Ethernet network is comprised of two stacked 1U 48port switches, which are located at the top of the Urika-GX rack, and can optionally contain redundant switch
 power supplies. These switches provide GigE management Ethernet connectivity to every node, System
 Management Workstation (SMW), Rack Controller (RC), Intelligent Subrack Control Board (iSCB), Power
 Distribution Units (PDUs), Dual Aries Network Cards (dANCs) and to the operational network that connects to
 the nodes.

The Urika-GX system also contains the following subnets:

- SMW subnet, which provides connectivity to the SMW and the RC.
- Rack subnet, which provides connectivity to the dANCs and iSCB module.

This network is supported by two managed Brocade ICX 6450-48, 48 port switches stacked together with two 10gigE optical interconnects. Each switch contains a single power supply, and can optionally contain redundant switch power supplies. The following VLANs are defined for this network to support management network traffic:

- VLAN 102 Uses ports 1-5 on each ICX 6450-48 switch. This is a dual-mode (tagged dual-mode for VLAN 102 and tagged for VLAN 103) VLAN. Untagged traffic on these ports belongs to VLAN 102. Traffic can be tagged for VLAN 103. The SMW HSS interface, the RC for a given rack, and the PDUs for a given rack are connected to these ports.
- VLAN 103 Ports 6-12 on each ICX 6450-48 switch. Untagged traffic on these ports belongs to VLAN 103.
 The iSCBs and dANC cards are connected to these ports.
- VLAN 104 Ports 13-48 on each ICX 6450-48 switch.

NOTE: Traffic on this VLAN may be reduced if VLAN 105 is needed for storage as long as each compute node is connected to VLAN 104

- Untagged traffic on these ports belongs to VLAN 104. The compute nodes and the SMW node-side network are connected to these ports.
- VLAN 105 Some number of Ports 13-48 on each ICX 6450-48 switch, as needed for storage management. Untagged traffic on these ports belongs to VLAN 105. The Storage Management Ports are connected to these ports.
- o VLAN 1 (default) is unused.

Traffic from the SMW to the subcomponents in the rack subnet, and vice versa, is routed through the corresponding RC.

For additional information, see the *Urika®-GX Hardware Guide*.

5 Urika-GX Software Components

The Urika®-GX system runs a combination of Cray-developed, third-party, and open-source software components. The system software is optimized for applications that have fine-grain synchronization requirements, large processor counts, and significant communication requirements.

Major software components of Urika-GX are listed below:

Operating System

Urika-GX compute nodes and the System Management Workstation (SMW) run the CentOS 7.2 operating system. Along with CentOS, Urika-GX nodes also run portions of the Cray Linux Environment (CLE) for providing a number of Cray-specific features and functionalities.

Analytics Applications

- Cray Graph Engine (CGE) CGE is a highly optimized software application designed for high-speed processing of interconnected data. It is a SPARQL database engine that features an advanced platform for searching very large, graph-oriented databases and querying for complex relationships between data items with a database. It provides the tools required for capturing, organizing and analyzing large sets of interconnected data. CGE enables performing real-time analytics on the largest and most complex graph problems, and features highly optimized support for inference, deep graph analysis, and pattern-based queries. For more information, see S-3014, "Cray® Graph Engine (CGE) User Guide".
- Spark and Hadoop core and ecosystem components Urika-GX uses the Hortonworks Data Platform (HDP)
 for Hadoop support. In addition to the core Hadoop and Spark components, the Urika-GX system ships with a
 number of ecosystem components for increased productivity.
- Urika-GX analytics programming environment Urika-GX features an analytic programming environment that
 comprises of a set of tools that facilitate building, debugging, and compiling various types of analytics
 applications using various programming languages, including Python, R, JAVA, Spark etc. For more
 information, see S-3015, "Urika®-GX Analytic Applications Guide".
- Jupyter Notebook Urika-GX ships pre-installed with the Jupyter Notebook, which enables creating
 executable documents. There are a number of Jupyter cookbooks that are shipped with the Urika-GX system.
 For more information, see S-3015, "Urika®-GX Analytic Applications Guide". For more information, see
 S-3015, "Urika®-GX Analytic Applications Guide".

Additional Third Party Software

Urika-GX also features a number of system and workload management tools.

6 Urika-GX File Systems

Supported file system types on Urika-GX include:

- Internal file systems
 - Hadoop Distributed File System (HDFS) Hadoop uses HDFS for storing data. HDFS is highly faulttolerant, provides high throughput access to application data, and is suitable for applications that have large data sets. Urika-GX also features tiered HDFS storage. HDFS data is transferred over the Aries network.
 - Network File System (NFS) The Urika-GX SMW hosts NFS, which is made available to every node via the management network.
- External file system Urika-GX currently supports Lustre as an external file system. On the Urika-GX system,
 Cray Lustre clients and a Lustre server is included to support:
 - Direct Attach Lustre (DAL) In a DAL configuration, the Urika-GX I/O nodes are directly attached to an external block storage devices and configured as Lustre servers, whereas all the other node are configured as Lustre clients.
 - Cray Sonexion storage system When a Urika-GX system uses the Cray Sonexion as its external storage system, the Sonexion system is attached to Urika-GX I/O nodes as Lustre clients. In this type of configuration, the I/O nodes act as LNET routers to make serve Lustre to all other nodes in the system.

The Lustre file system is served over the Aries network when the DAL or Sonexion configuration type is used.

File Locations

- Home directories are mounted on (internal) NFS, with limited space
- Distributed filesystem (Lustre), if provisioned, is mounted at /mnt/lustre and is suitable for larger files.



CAUTION: Avoid using NFS for high data transfer and/or large writes as this will cause the network to operate much slower or timeout.

7 Urika-GX Workload Management Tools

The Urika-GX system features a number of workload management tools that enable management of disparate workload distributions. These tools include:

Apache™ Yet Another Resource Negotiator (YARN)

YARN is a cluster management tool, responsible for managing compute resources in a cluster and using those resources for scheduling of applications. YARN is designed to allow multiple, disparate applications to run on a multi-tenant platform. It is responsible for managing and monitoring workloads, maintaining a multi-tenant environment, implementing security controls, and managing high availability features of Hadoop.

YARN acts as the resource manager for Hadoop jobs on the Urika-GX system.

Apache™ Mesos™

Apache Mesos simplifies the process of deploying and managing applications in large-scaled clustered settings. It can run multiple applications on a dynamically shared pool of nodes. Mesos consists of a master daemon that manages agent daemons running on each cluster node, and mesos applications (also called frameworks) that run tasks on these agents. The master enables fine-grained sharing of resources across applications by making them resource offers and decides how many resources to offer each framework according to a given organizational policy.

Mesos acts as the primary resource manager on the Urika-GX system. As part of the Mesos ecosystem, Urika-GX features the Cray-developed mrun application launcher program. mrun enables running parallel jobs on Urika-GX, using resources managed by Mesos/Marathon.

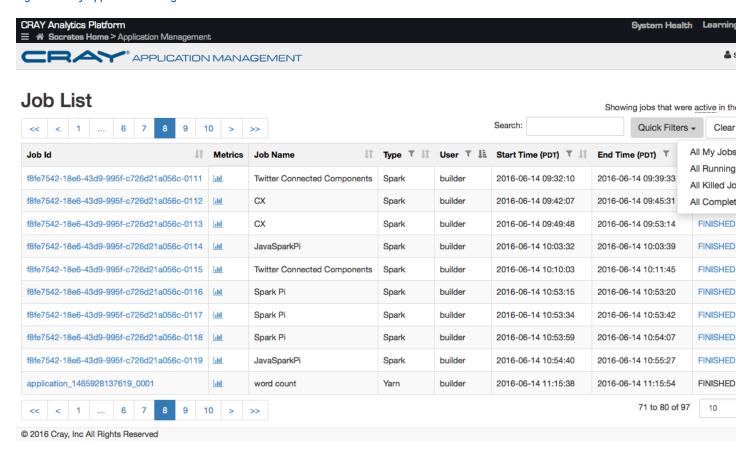
For more information, see S-3015, "Urika®-GX Analytic Applications Guide".

8 Urika-GX System Management and Monitoring Tools

There are a number of tools that can be used for managing and monitoring the Urika®-GX system. These tools include:

- Hardware Supervisory System (HSS) HSS is an integrated system of hardware and software components
 that monitors components, manages hardware and software failures, controls system start up and shutdown,
 manages the system interconnection network and maintains system states. The HSS communicates with
 nodes and management processors over an internal (private) Ethernet network that operates independently
 of the system interconnection network.
- Cray System Management Software (CSMS) CSMS is a system management tool that facilitates image
 provisioning and node management on the Urika-GX system. In addition, the Cray System Management
 Software features a user-friendly GUI as well as a command line for performing system management and
 monitoring tasks.
- Cray Application Management Urika-GX features the Cray Application Management UI that enables viewing, searching, filtering and performing actions on jobs submitted to the Urika-GX system.

Figure 2. Cray Application Management UI



• Analytic applications monitoring tools - There are a number of tools installed on the Urika-GX system that facilitate monitoring the analytic applications. These tools include Grafana, Hadoop Job History Server, and Spark History Server. For more information, see S-3015, "Urika®-GX Analytic Applications Guide".

Urika-GX features the **Urika**®-**GX Applications Interface**, which enables accessing many of the pre-installed analytic applications, the system management UI and Urika-GX documentation.

Figure 3. Urika®-GX Applications Interface UI

