Azure HDInsight Kafka and Ambari

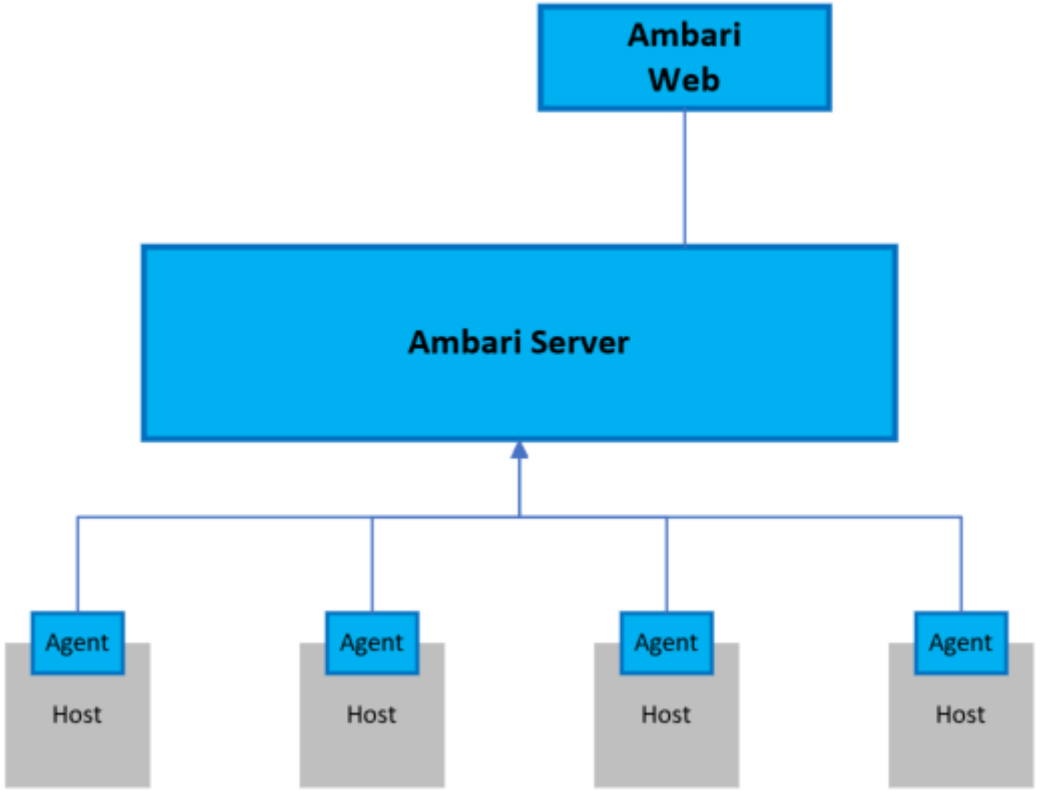
Sridhar Kothalanka (CSA-Data&AI)

Feb 2020

What is Apache Ambari

Apache Ambari simplifies the management and monitoring of an Apache Hadoop cluster by providing an easy to use web UI and REST API.

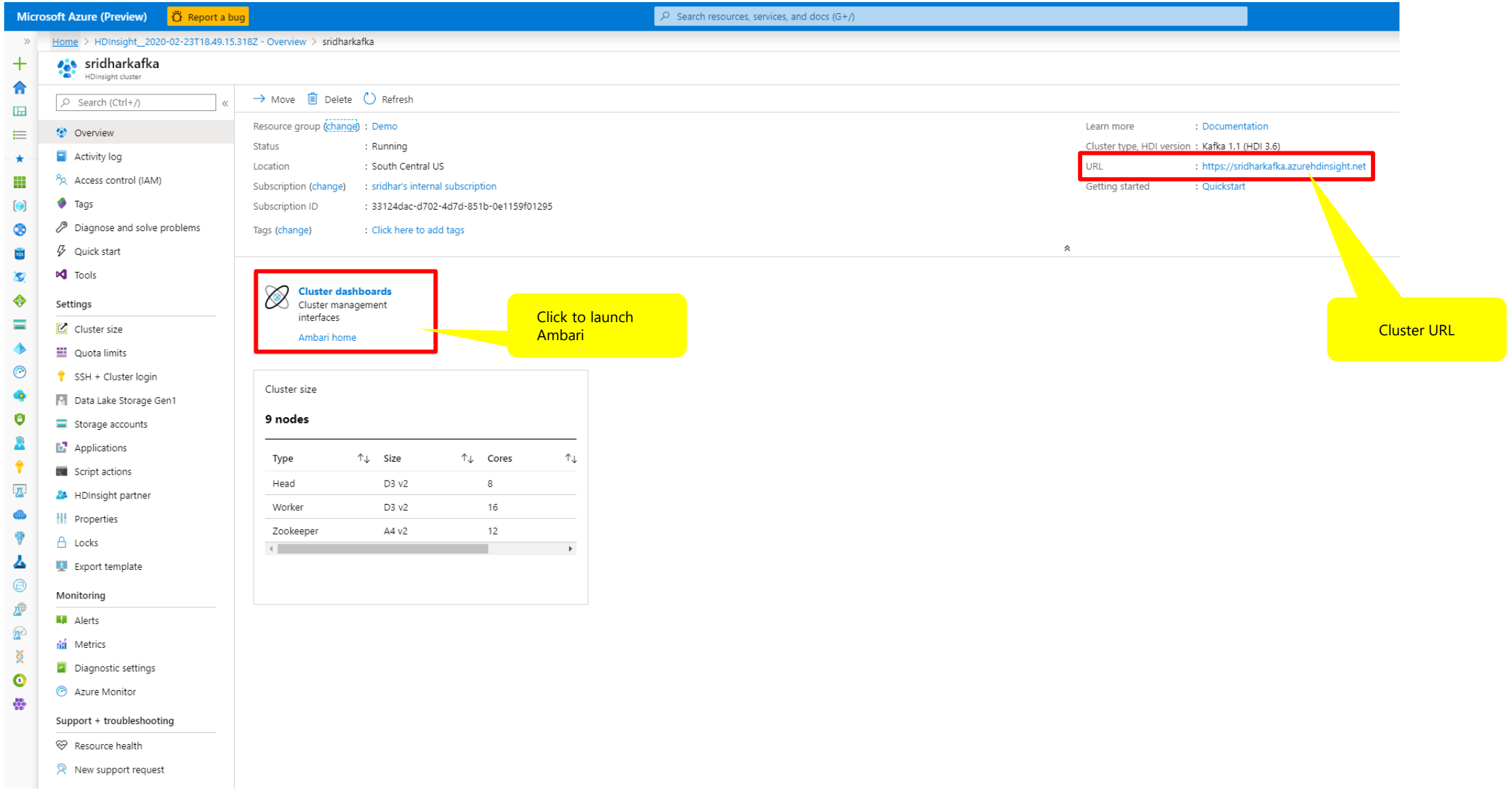
Ambari is included on HDInsight clusters, and is used to monitor the cluster and make configuration changes.



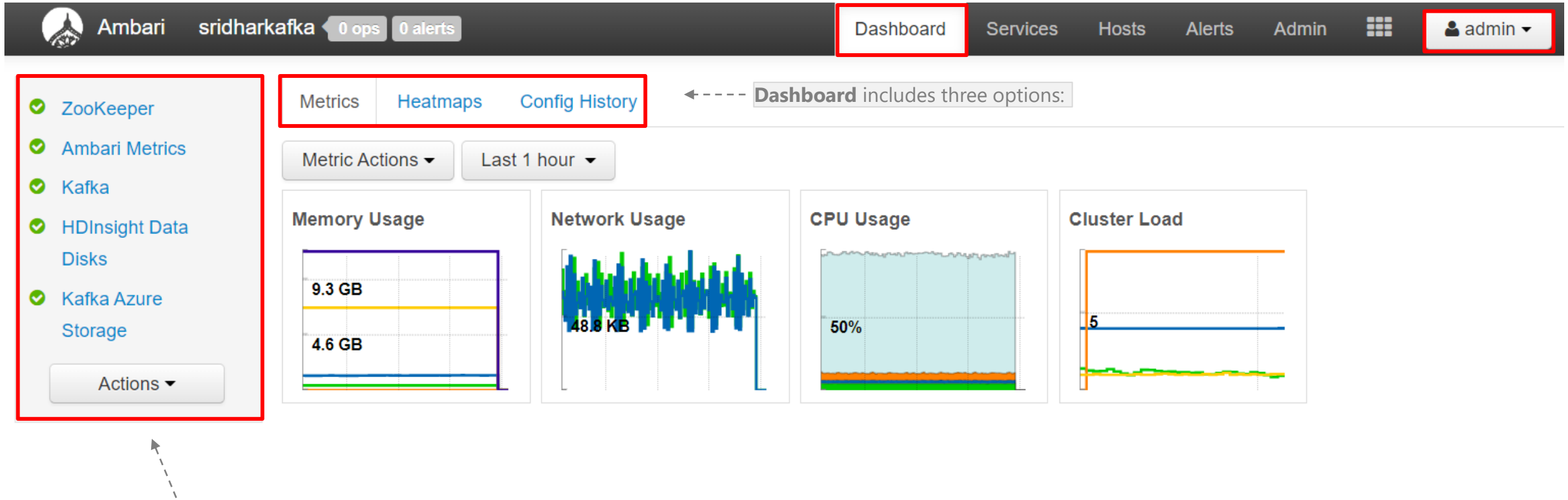
Reference:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-manage-ambari

The Ambari Web UI is available on your HDInsight cluster at http://CLISTERNAME.azurehdinsight.net, where CLUSTERNAME is the name of your cluster

Launch Ambari Web UI from Azure HDInsight Kafka

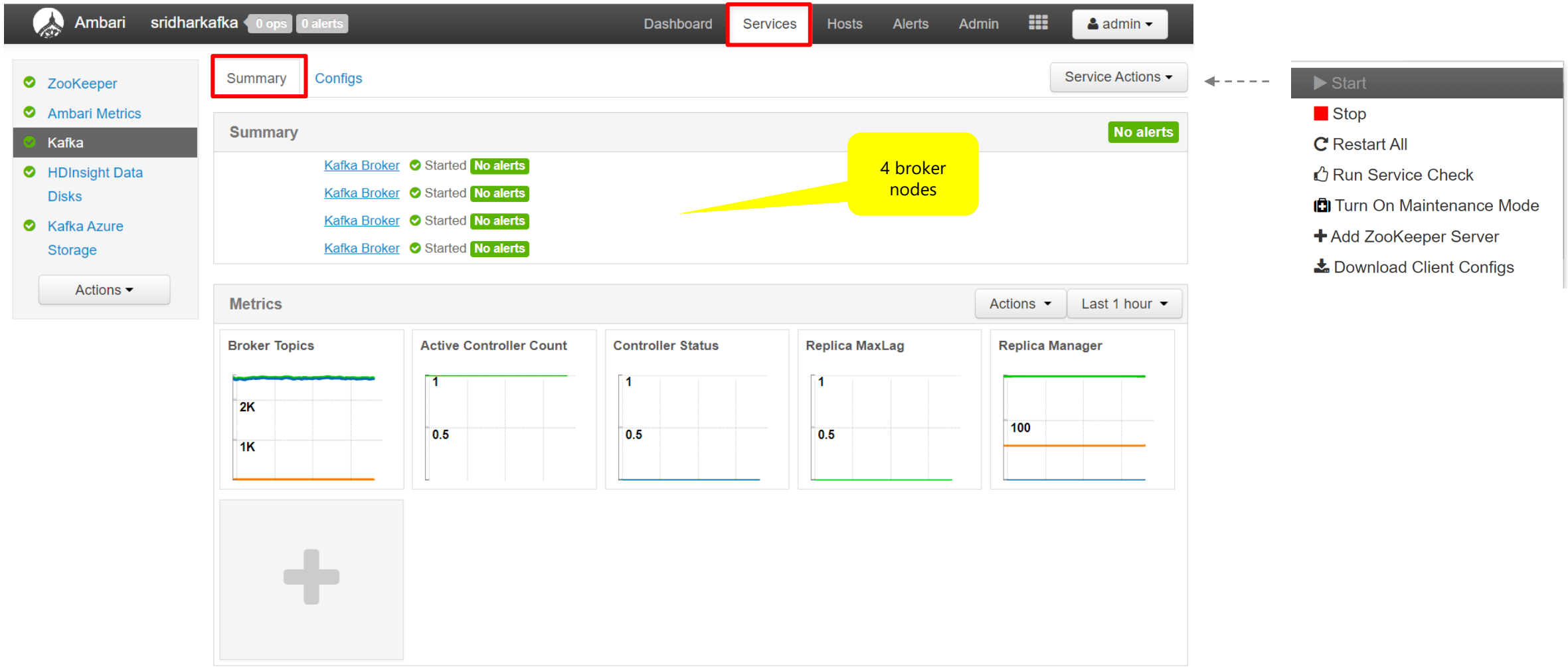


Ambari Dashboard

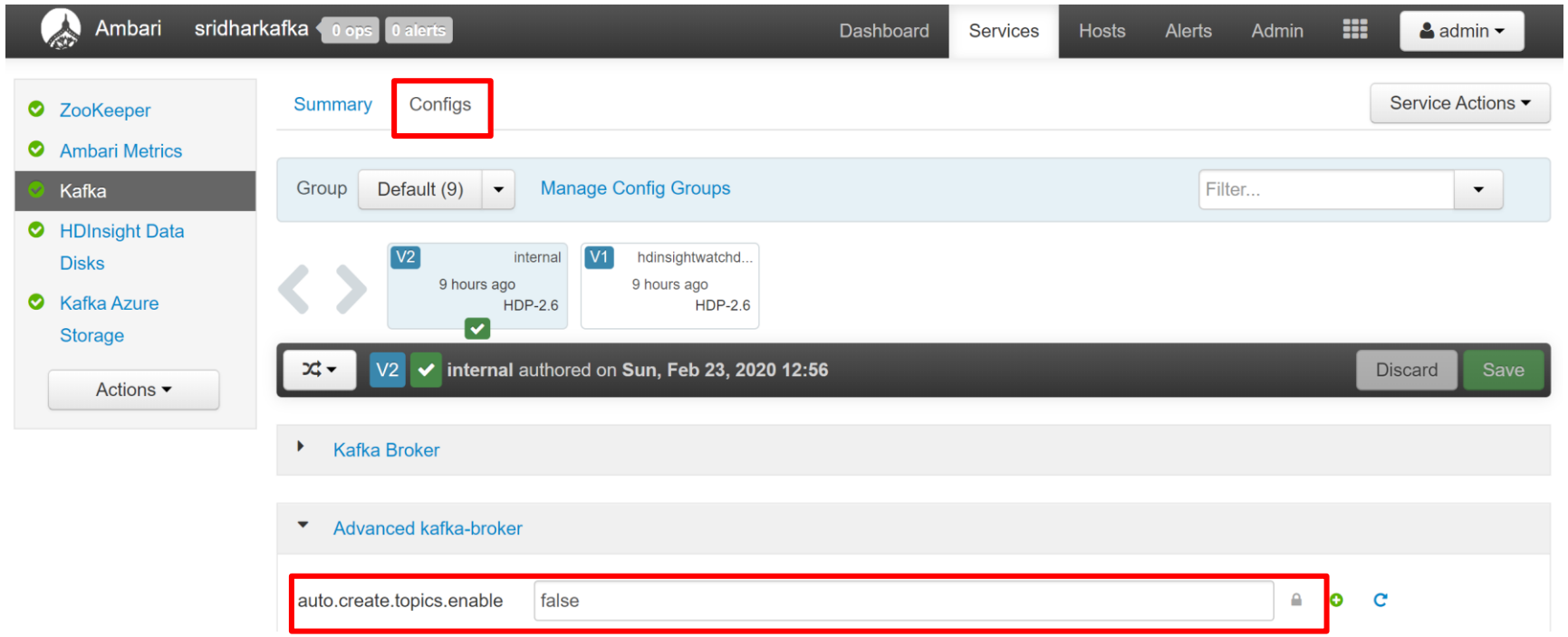


The **services** sidebar on the dashboard provides quick insight into the status of the services running on the cluster. Various icons are used to indicate status or actions that should be taken. For example, a yellow recycle symbol is displayed if a service needs to be recycled.

Kafka Service Summary

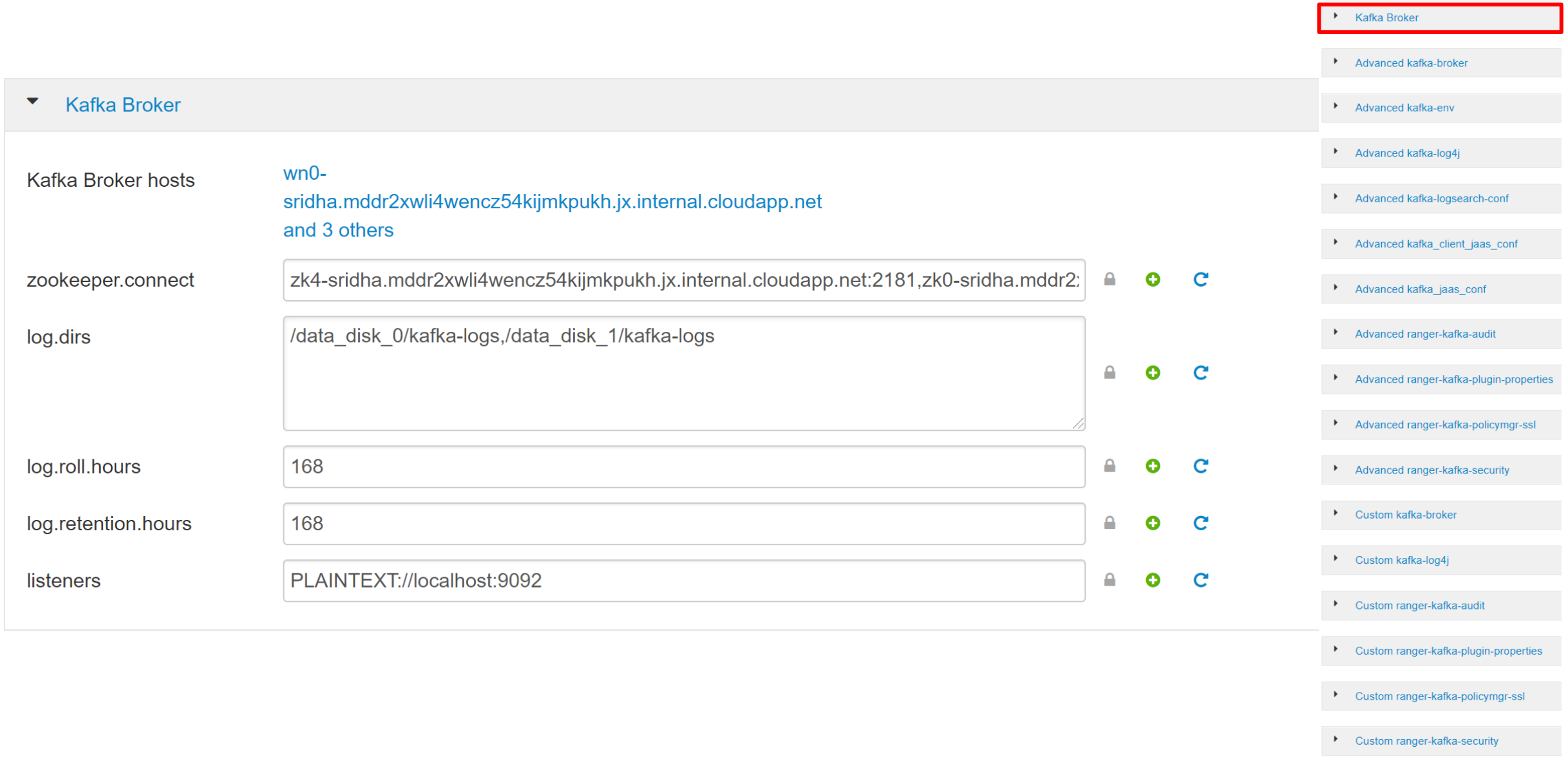


Kafka Configs

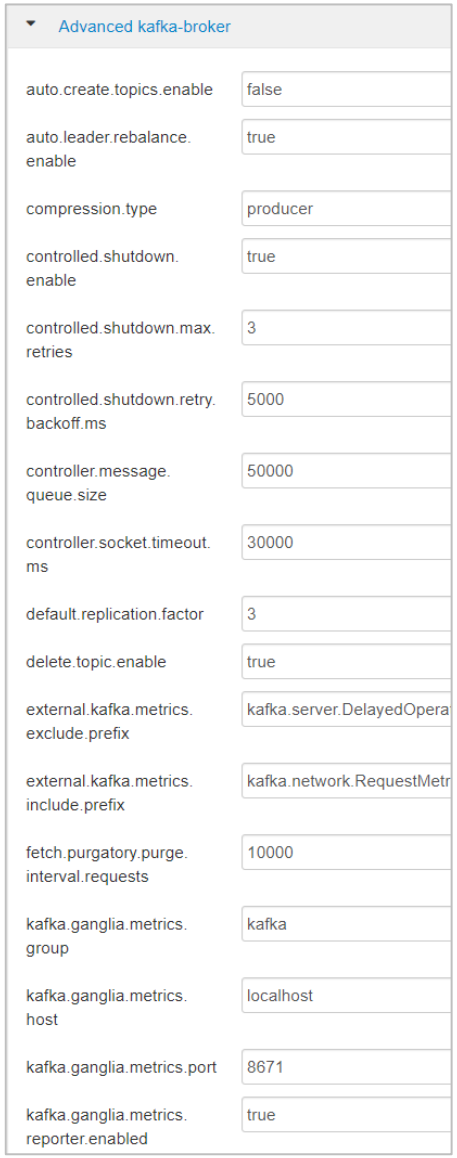


- Kafka Broker
- Advanced kafka-broker
- Advanced kafka-env
- Advanced kafka-log4j
- Advanced kafka-logsearch-conf
- Advanced kafka_client_jaas_conf
- Advanced kafka_jaas_conf
- Advanced ranger-kafka-audit
- Advanced ranger-kafka-plugin-properties
- Advanced ranger-kafka-policymgr-ssl
- Advanced ranger-kafka-security
- Custom kafka-broker
- Custom kafka-log4j
- Custom ranger-kafka-audit
- Custom ranger-kafka-plugin-properties
- Custom ranger-kafka-policymgr-ssl
- Custom ranger-kafka-security

Kafka Broker config

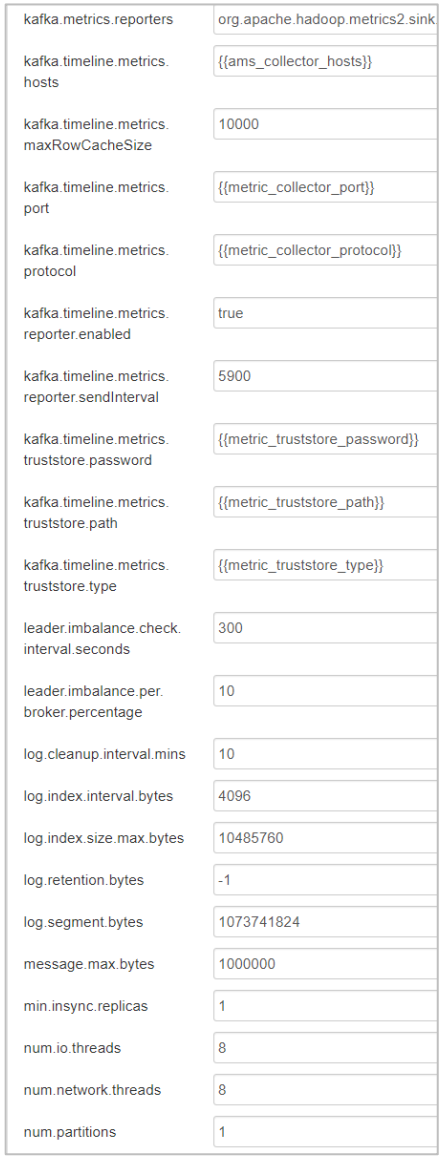


Advanced Kafka-broker config

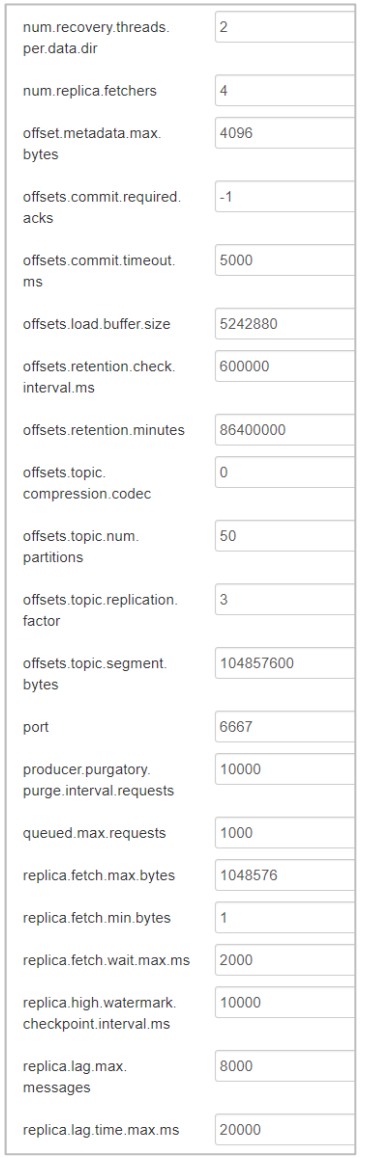


Advanced kafka-broker

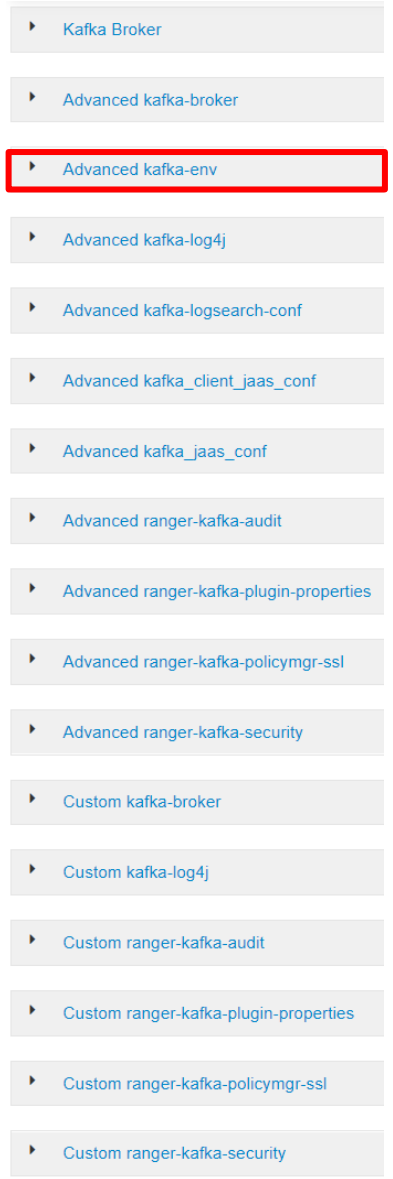
auto.create.topics.enable	false
auto.leader.rebalance.enable	true
compression.type	producer
controlled.shutdown.enable	true
controlled.shutdown.max.retries	3
controlled.shutdown.retry.backoff.ms	5000
controller.message.queue.size	50000
controller.socket.timeout.ms	30000
default.replication.factor	3
delete.topic.enable	true
external.kafka.metrics.exclude.prefix	kafka.server.DelayedOpera
external.kafka.metrics.include.prefix	kafka.network.RequestMetr
fetch.purgatory.purge.interval.requests	10000
kafka.ganglia.metrics.group	kafka
kafka.ganglia.metrics.host	localhost
kafka.ganglia.metrics.port	8671
kafka.ganglia.metrics.reporter.enabled	true



kafka.metrics.reporters	org.apache.hadoop.metrics2.sink
kafka.timeline.metrics.hosts	{{ams_collector_hosts}}
kafka.timeline.metrics.maxRowCacheSize	10000
kafka.timeline.metrics.port	{{metric_collector_port}}
kafka.timeline.metrics.protocol	{{metric_collector_protocol}}
kafka.timeline.metrics.reporter.enabled	true
kafka.timeline.metrics.reporter.sendInterval	5900
kafka.timeline.metrics.truststore.password	{{metric_truststore_password}}
kafka.timeline.metrics.truststore.path	{{metric_truststore_path}}
kafka.timeline.metrics.truststore.type	{{metric_truststore_type}}
leader.imbalance.check.interval.seconds	300
leader.imbalance.per.broker.percentage	10
log.cleanup.interval.mins	10
log.index.interval.bytes	4096
log.index.size.max.bytes	10485760
log.retention.bytes	-1
log.segment.bytes	1073741824
message.max.bytes	1000000
min.insync.replicas	1
num.io.threads	8
num.network.threads	8
num.partitions	1

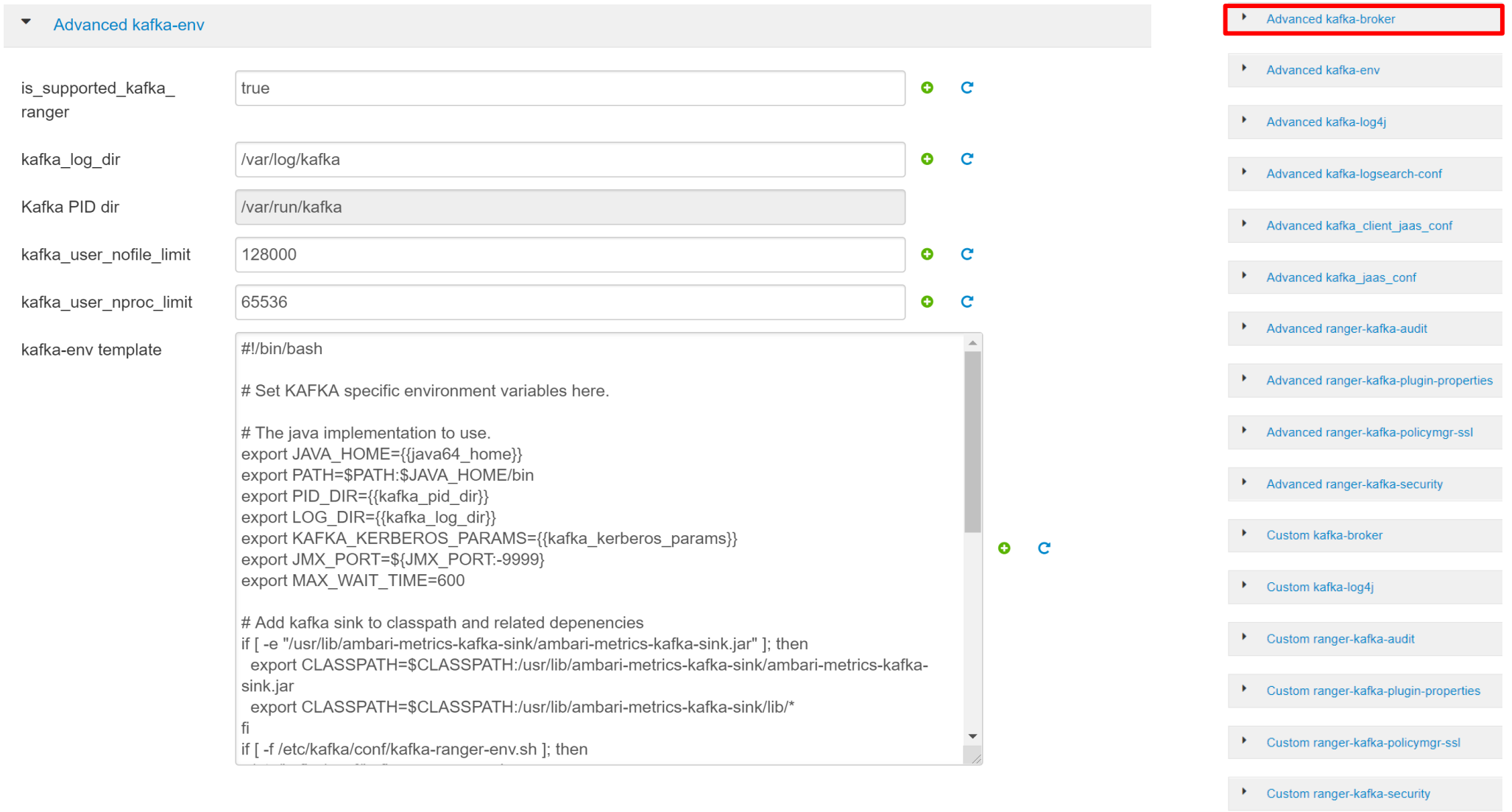


num.recovery.threads.per.data.dir	2
num.replica.fetchers	4
offset.metadata.max.bytes	4096
offsets.commit.required.acks	-1
offsets.commit.timeout.ms	5000
offsets.load.buffer.size	5242880
offsets.retention.check.interval.ms	600000
offsets.retention.minutes	86400000
offsets.topic.compression.codec	0
offsets.topic.num.partitions	50
offsets.topic.replication.factor	3
offsets.topic.segment.bytes	104857600
port	6667
producer.purgatory.purge.interval.requests	10000
queued.max.requests	1000
replica.fetch.max.bytes	1048576
replica.fetch.min.bytes	1
replica.fetch.wait.max.ms	2000
replica.high.watermark.checkpoint.interval.ms	10000
replica.lag.max.messages	8000
replica.lag.time.max.ms	20000

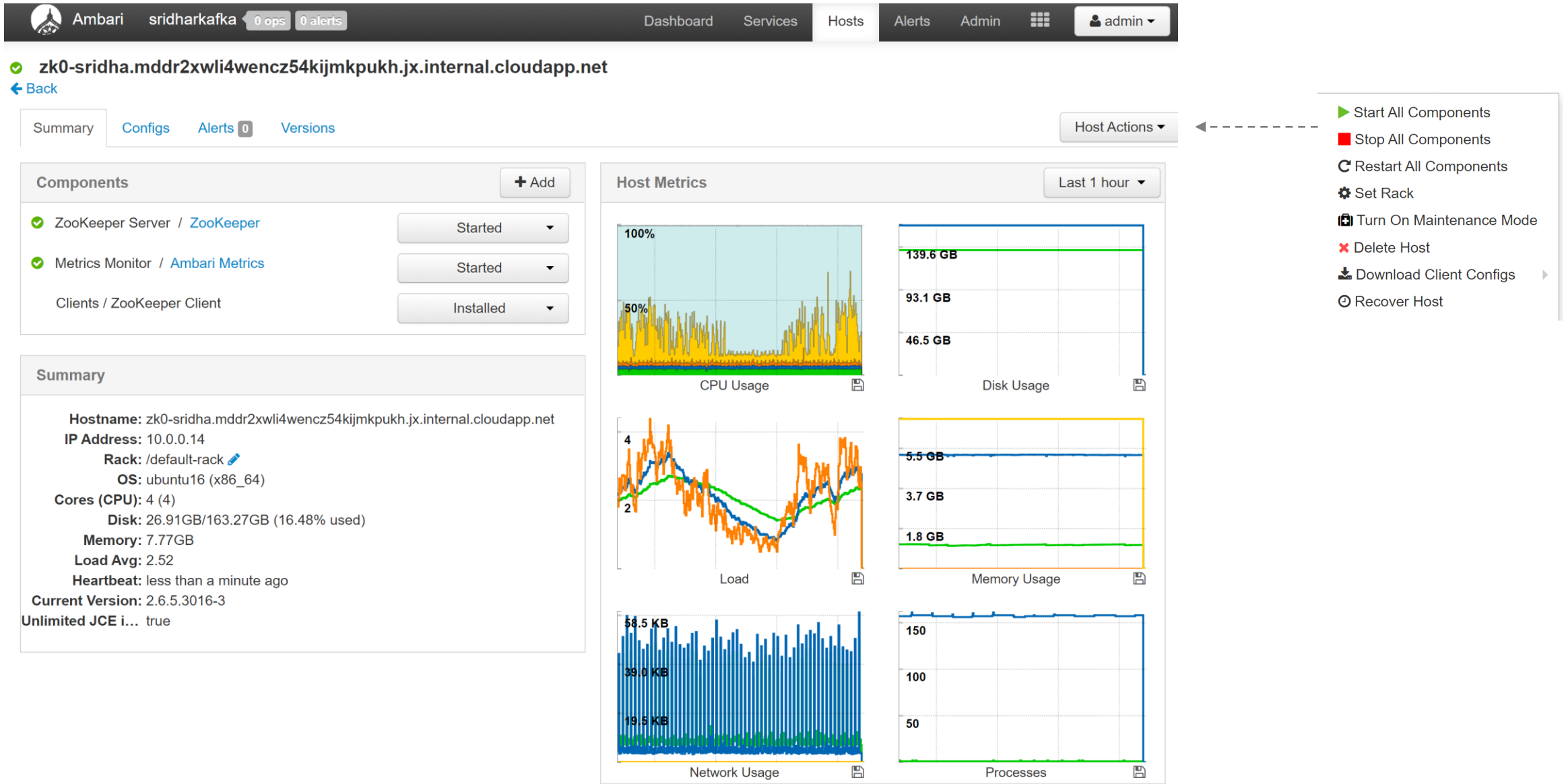


- Kafka Broker
- Advanced kafka-broker
- Advanced kafka-env
- Advanced kafka-log4j
- Advanced kafka-logsearch-conf
- Advanced kafka_client_jaas_conf
- Advanced kafka_jaas_conf
- Advanced ranger-kafka-audit
- Advanced ranger-kafka-plugin-properties
- Advanced ranger-kafka-policymgr-ssl
- Advanced ranger-kafka-security
- Custom kafka-broker
- Custom kafka-log4j
- Custom ranger-kafka-audit
- Custom ranger-kafka-plugin-properties
- Custom ranger-kafka-policymgr-ssl
- Custom ranger-kafka-security

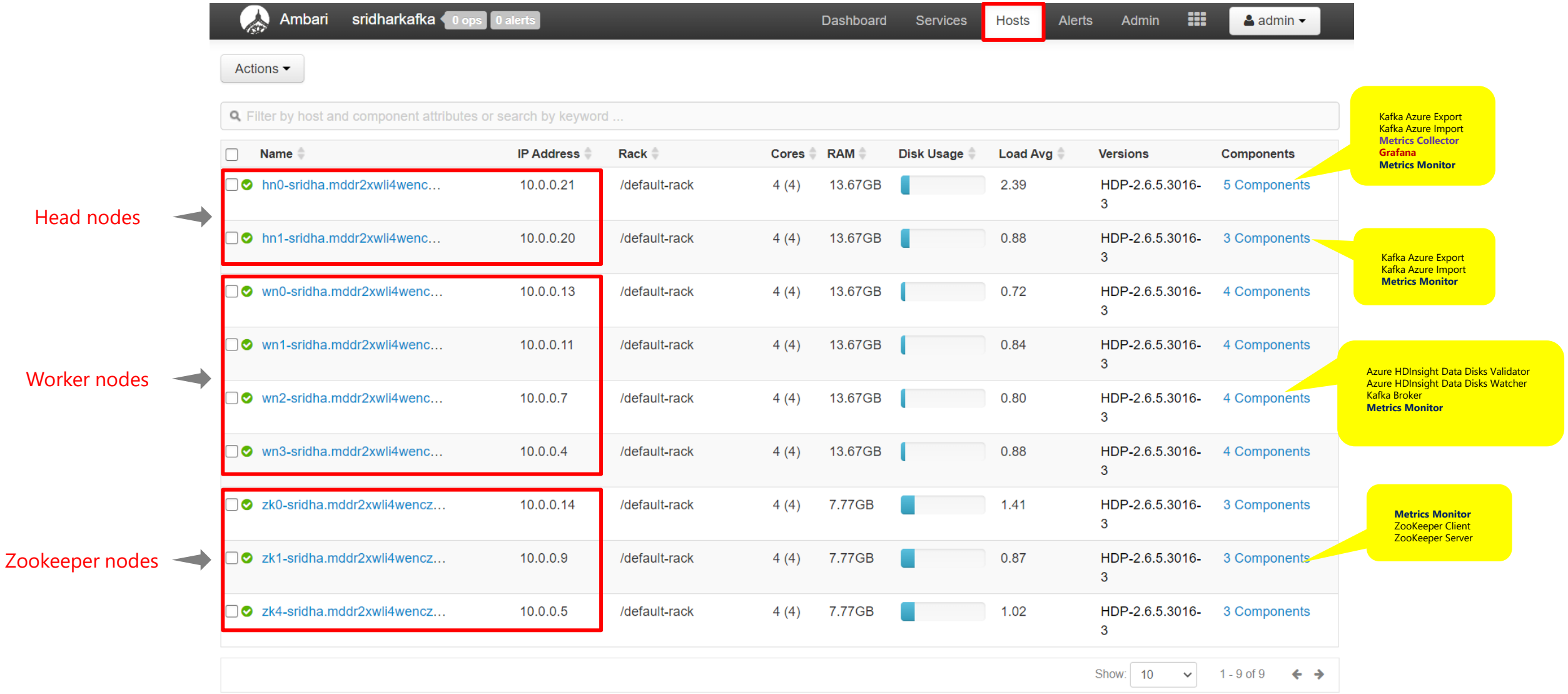
Advanced Kafka-env config



Kafka Broker Drill Down



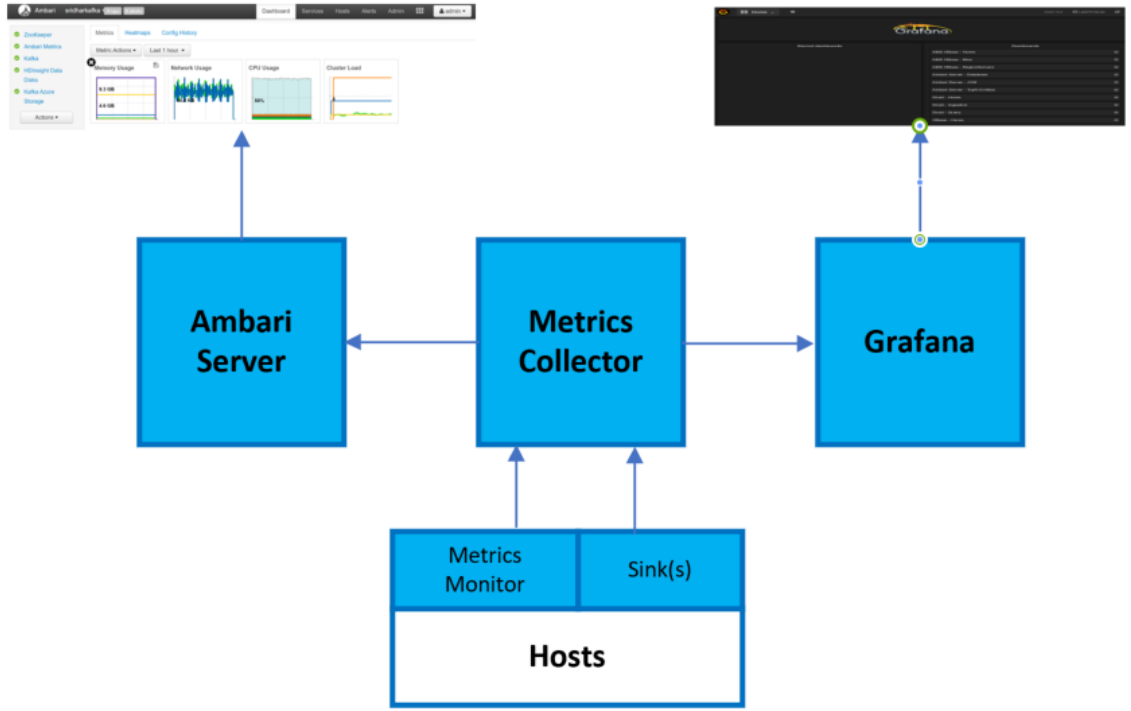
Hosts


Ambari
sridharkafka
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
admin
Actions
Filter by host and component attributes or search by keyword ...
Name
IP Address
Rack
Cores
RAM
Disk Usage
Load Avg
Versions
Components
hn0-sridha.mddr2xwli4wenc...
10.0.0.21
/default-rack
4 (4)
13.67GB
2.39
HDP-2.6.5.3016-3
5 Components
hn1-sridha.mddr2xwli4wenc...
10.0.0.20
/default-rack
4 (4)
13.67GB
0.88
HDP-2.6.5.3016-3
3 Components
wn0-sridha.mddr2xwli4wenc...
10.0.0.13
/default-rack
4 (4)
13.67GB
0.72
HDP-2.6.5.3016-3
4 Components
wn1-sridha.mddr2xwli4wenc...
10.0.0.11
/default-rack
4 (4)
13.67GB
0.84
HDP-2.6.5.3016-3
4 Components
wn2-sridha.mddr2xwli4wenc...
10.0.0.7
/default-rack
4 (4)
13.67GB
0.80
HDP-2.6.5.3016-3
4 Components
wn3-sridha.mddr2xwli4wenc...
10.0.0.4
/default-rack
4 (4)
13.67GB
0.88
HDP-2.6.5.3016-3
4 Components
zk0-sridha.mddr2xwli4wencz...
10.0.0.14
/default-rack
4 (4)
7.77GB
1.41
HDP-2.6.5.3016-3
3 Components
zk1-sridha.mddr2xwli4wencz...
10.0.0.9
/default-rack
4 (4)
7.77GB
0.87
HDP-2.6.5.3016-3
3 Components
zk4-sridha.mddr2xwli4wencz...
10.0.0.5
/default-rack
4 (4)
7.77GB
1.02
HDP-2.6.5.3016-3
3 Components
Show: 10
1 - 9 of 9
Head nodes
Worker nodes
Zookeeper nodes
Kafka Azure Export
Kafka Azure Import
Metrics Collector
Grafana
Metrics Monitor
Kafka Azure Export
Kafka Azure Import
Metrics Monitor
Azure HDInsight Data Disks Validator
Azure HDInsight Data Disks Watcher
Kafka Broker
Metrics Monitor
Metrics Monitor
ZooKeeper Client
ZooKeeper Server

Metrics Collector

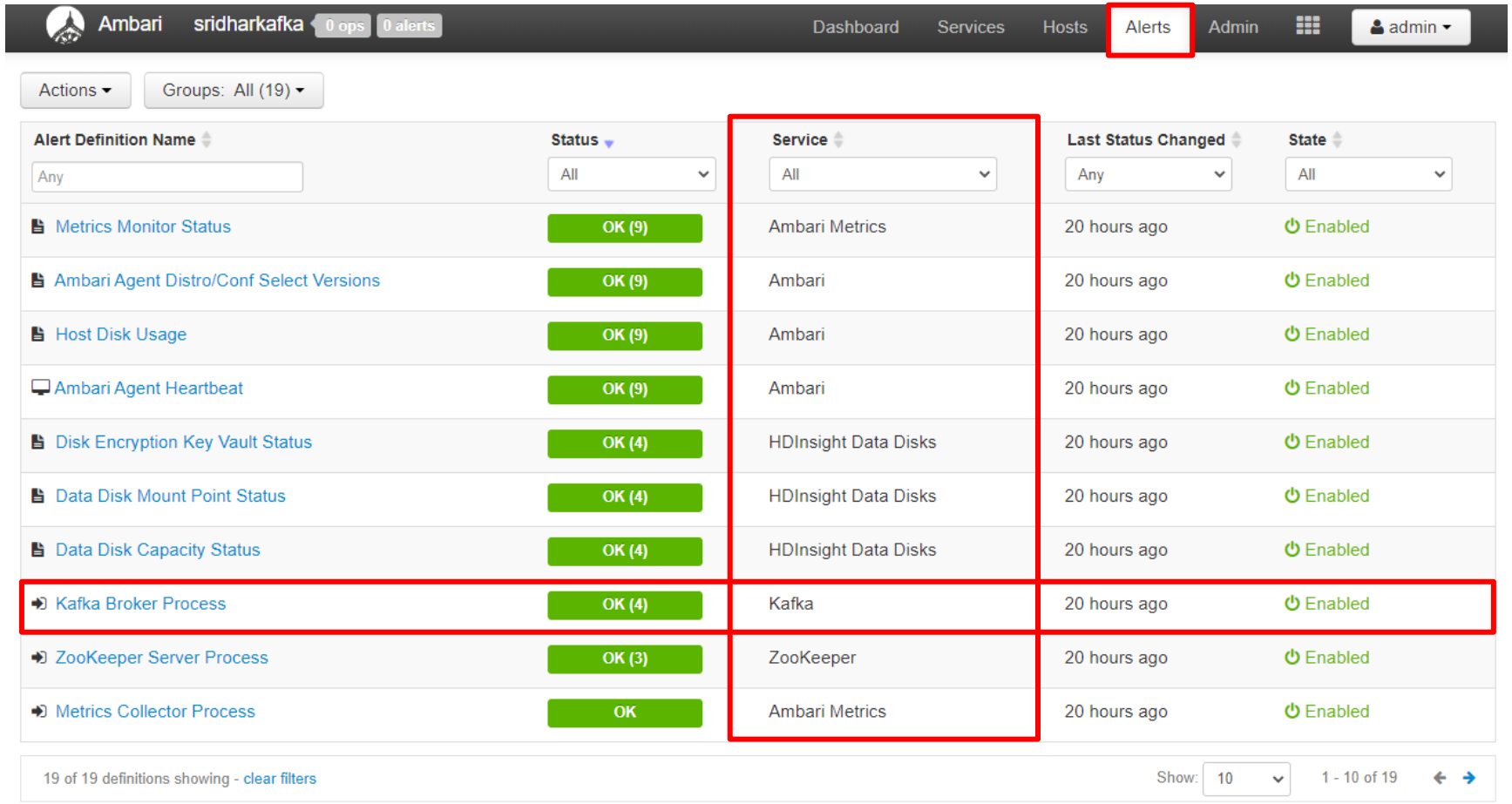
Ambari Metrics System has four components:

- **Metrics Monitors** on each host in the cluster collect system-level metrics and publish to the Metrics Collector.
- **Hadoop Sinks** plug in to Hadoop components to publish Hadoop metrics to the Metrics Collector.
- The **Metrics Collector** is a daemon that runs on a specific host in the cluster and receives data from the registered publishers, the Monitors, and the Sinks.
- **Grafana** is a daemon that runs on a specific host in the cluster and serves prebuilt dashboards for visualizing metrics collected in the Metrics Collector.

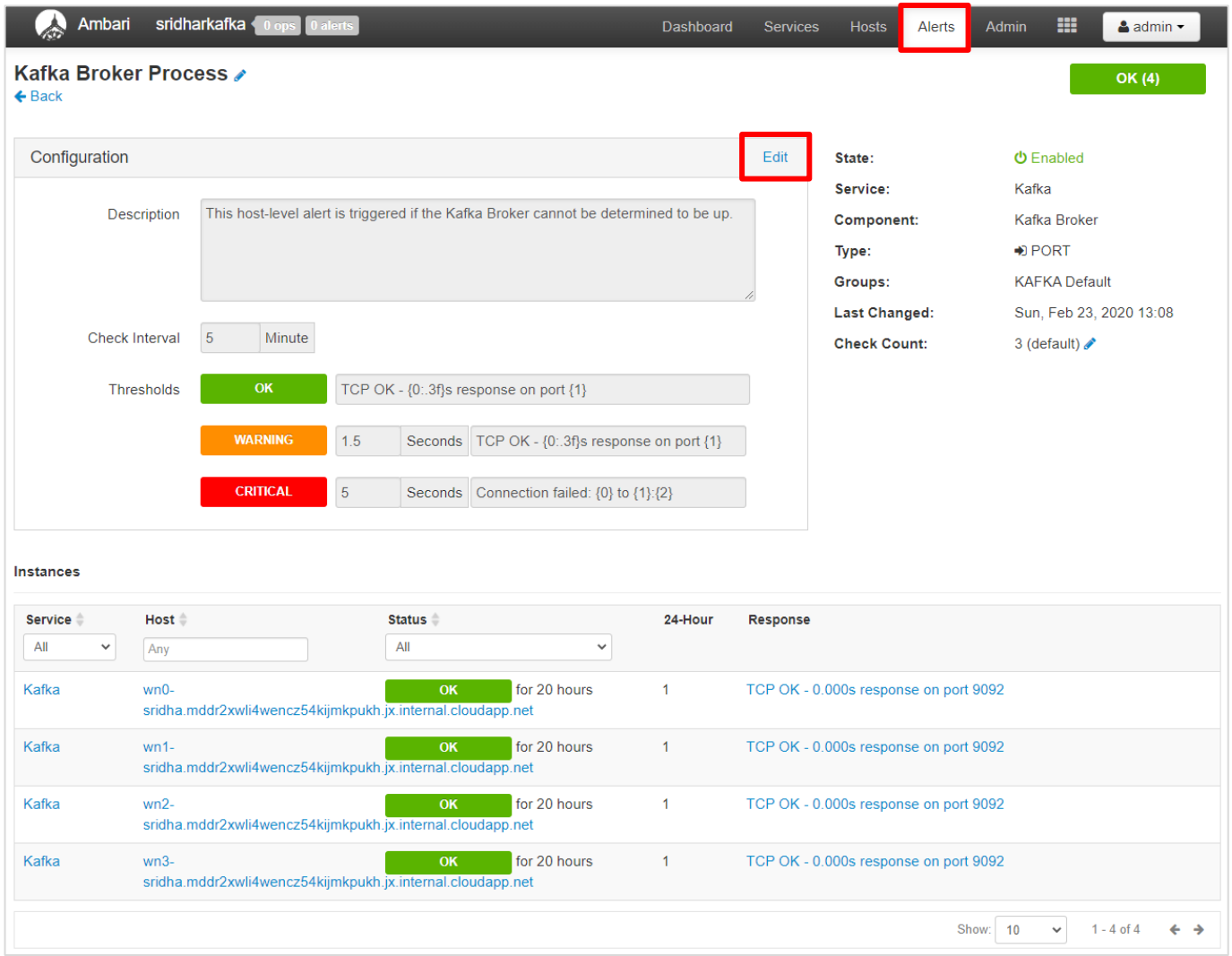


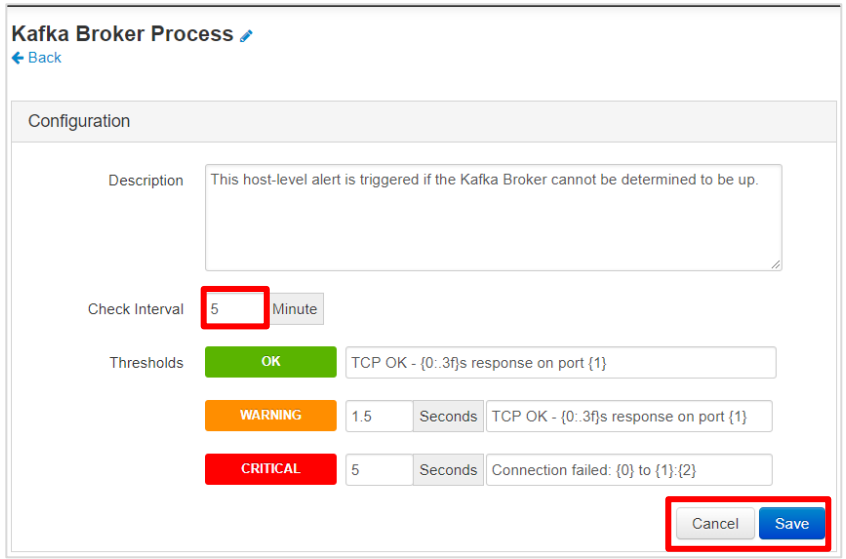
AMBARI: http://CLISTERNAME.azurehdinsight.net
GRAFANA: https://CLUSTERNAME.azurehdinsight.net/grafana/

Alerts

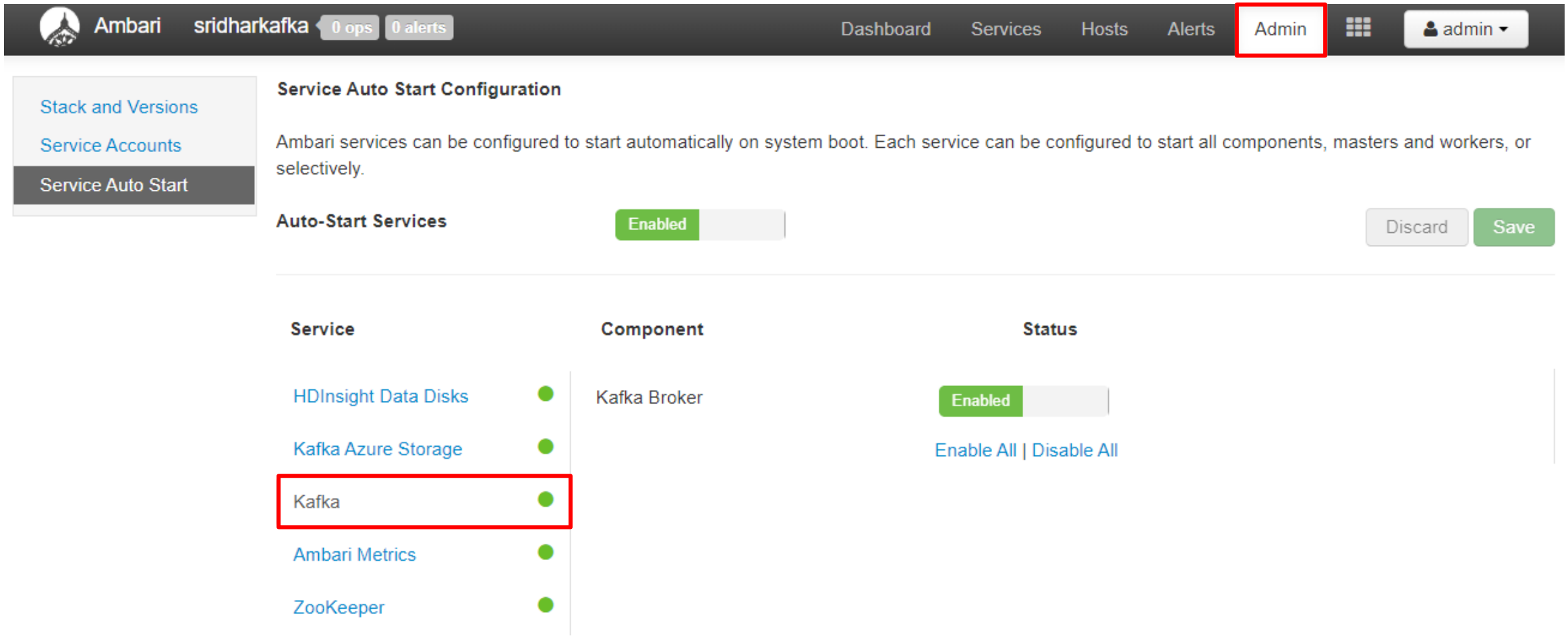

Ambari
sridharkafka
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
admin
Actions
Groups: All (19)
Alert Definition Name
Any
Status
All
Service
All
Last Status Changed
Any
State
All
Metrics Monitor Status
OK (9)
Ambari Metrics
20 hours ago
Enabled
Ambari Agent Distro/Conf Select Versions
OK (9)
Ambari
20 hours ago
Enabled
Host Disk Usage
OK (9)
Ambari
20 hours ago
Enabled
Ambari Agent Heartbeat
OK (9)
Ambari
20 hours ago
Enabled
Disk Encryption Key Vault Status
OK (4)
HDInsight Data Disks
20 hours ago
Enabled
Data Disk Mount Point Status
OK (4)
HDInsight Data Disks
20 hours ago
Enabled
Data Disk Capacity Status
OK (4)
HDInsight Data Disks
20 hours ago
Enabled
Kafka Broker Process
OK (4)
Kafka
20 hours ago
Enabled
ZooKeeper Server Process
OK (3)
ZooKeeper
20 hours ago
Enabled
Metrics Collector Process
OK
Ambari Metrics
20 hours ago
Enabled
19 of 19 definitions showing - clear filters
Show: 10
1 - 10 of 19

Kafka Service Alert





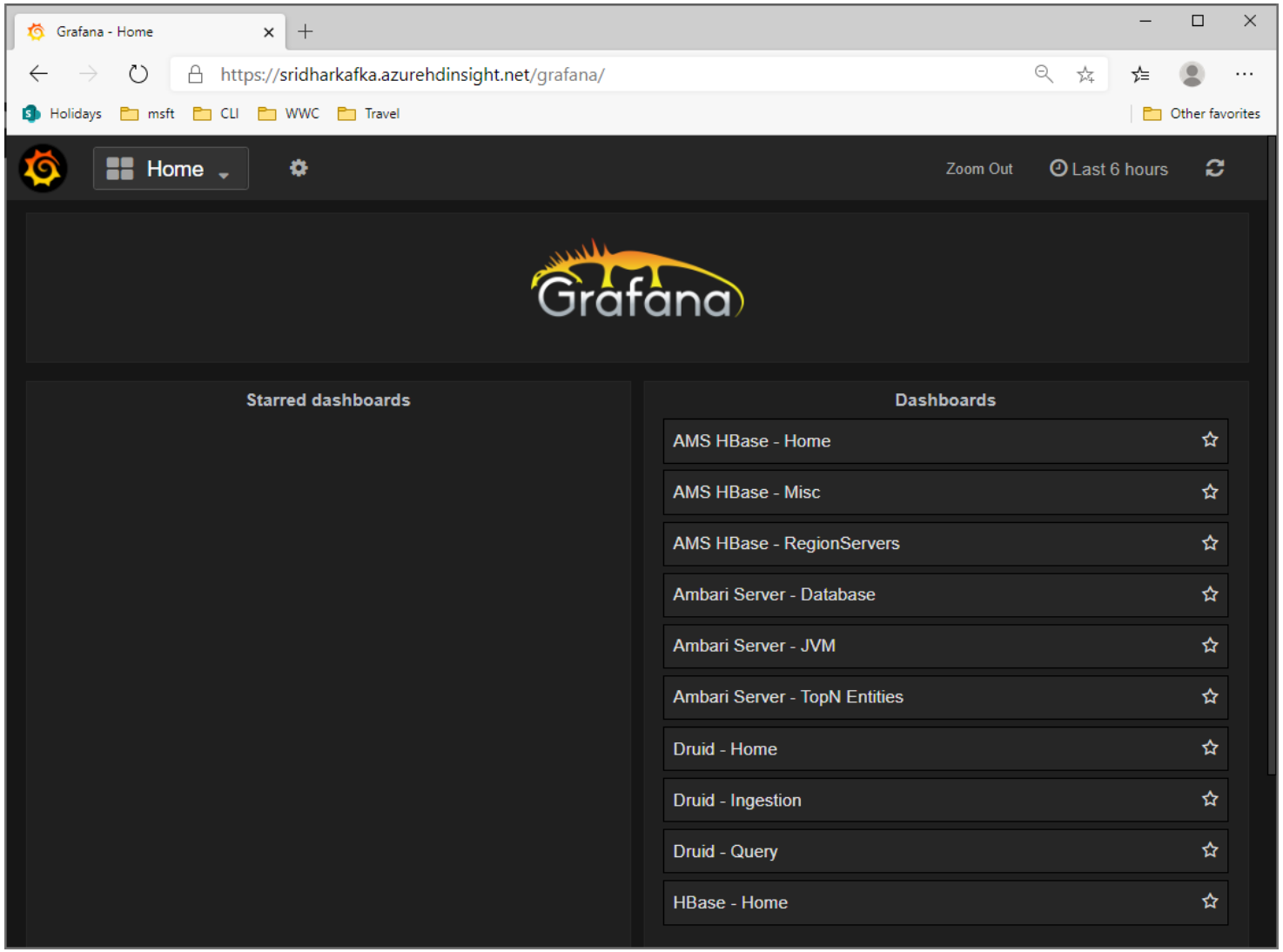
Admin


Ambari
sridharkafka
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
admin
Stack and Versions
Service Accounts
Service Auto Start
Service Auto Start Configuration
Ambari services can be configured to start automatically on system boot. Each service can be configured to start all components, masters and workers, or selectively.
Auto-Start Services
Enabled
Discard
Save
Service
Component
Status
HDInsight Data Disks
Kafka Azure Storage
Kafka
Ambari Metrics
ZooKeeper
Kafka Broker
Enabled
Enable All | Disable All

Grafana

GRAFANA: https://CLUSTERNAME.azurehdinsight.net/grafana/

Grafana is a popular, open-source graph and dashboard builder. Grafana is feature rich; not only does it let users create customizable and shareable dashboards, it also offers templated/scripted dashboards, LDAP integration, multiple data sources, and more.



https://docs.microsoft.com/en-us/azure/hdinsight/interactive-query/hdinsight-grafana

Microsoft

Q&A

Create a HDInsight Kafka Cluster