

Azure HDInsight Kafka and Ambari

Sridhar Kothalanka (CSA-Data&AI)

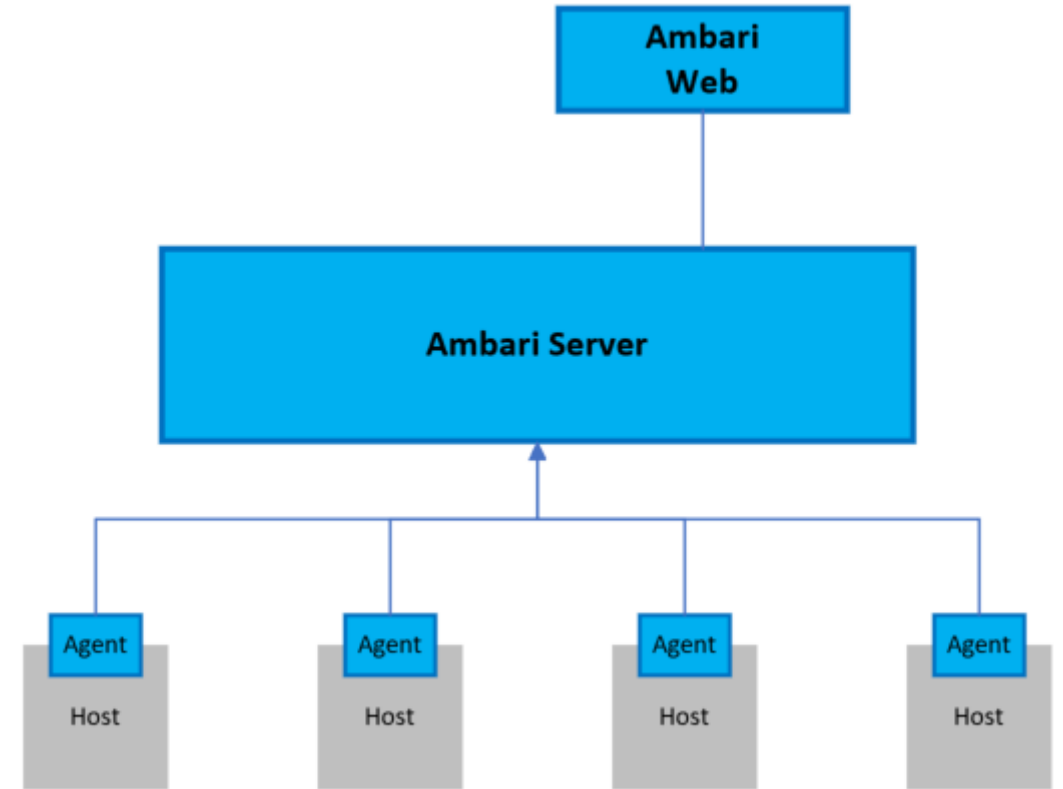
Feb 2020



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://CLUSTERNAME.azurehdinsight.net>, where CLUSTERNAME is the name of your cluster

Launch Ambari Web UI from Azure HDInsight Kafka

The screenshot displays the Microsoft Azure portal interface for an HDInsight cluster named 'sridharkafka'. The left sidebar contains navigation options such as Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Quick start, Tools, Settings, Cluster size, Quota limits, SSH + Cluster login, Data Lake Storage Gen1, Storage accounts, Applications, Script actions, HDInsight partner, Properties, Locks, Export template, Monitoring, Alerts, Metrics, Diagnostic settings, Azure Monitor, Support + troubleshooting, Resource health, and New support request.

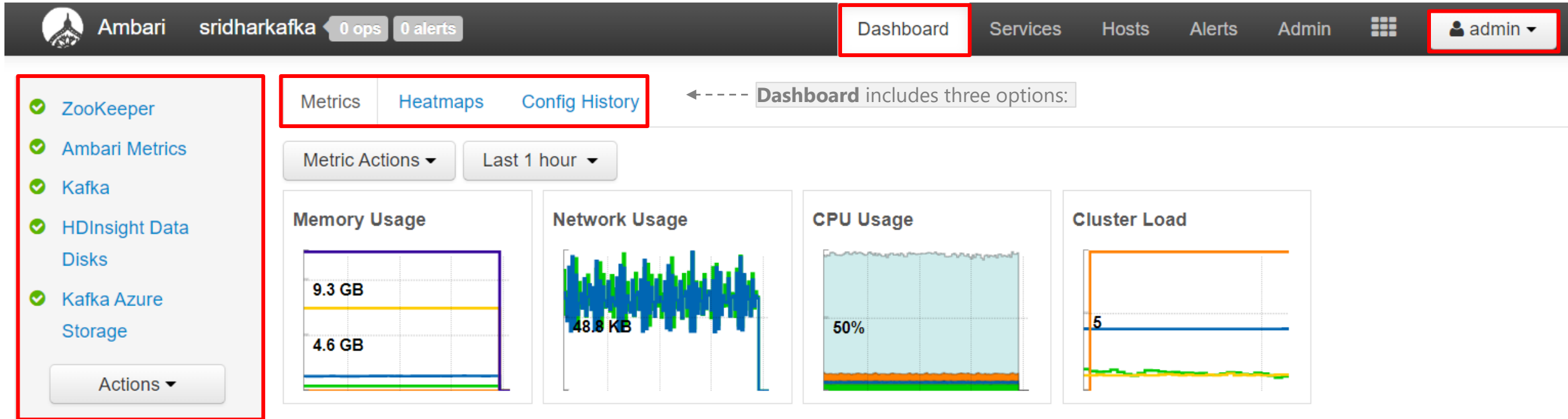
The main content area shows the cluster's overview, including the resource group 'Demo', status 'Running', location 'South Central US', subscription 'sridhar's internal subscription', and subscription ID '33124dac-d702-4d7d-851b-0e1159f01295'. A red box highlights the 'Cluster dashboards' section, which includes a link to 'Ambari home'. A yellow callout points to this link with the text 'Click to launch Ambari'.

Below the overview, the 'Cluster size' section shows 9 nodes. A table lists the node types and their specifications:

Type	Size	Cores
Head	D3 v2	8
Worker	D3 v2	16
Zookeeper	A4 v2	12

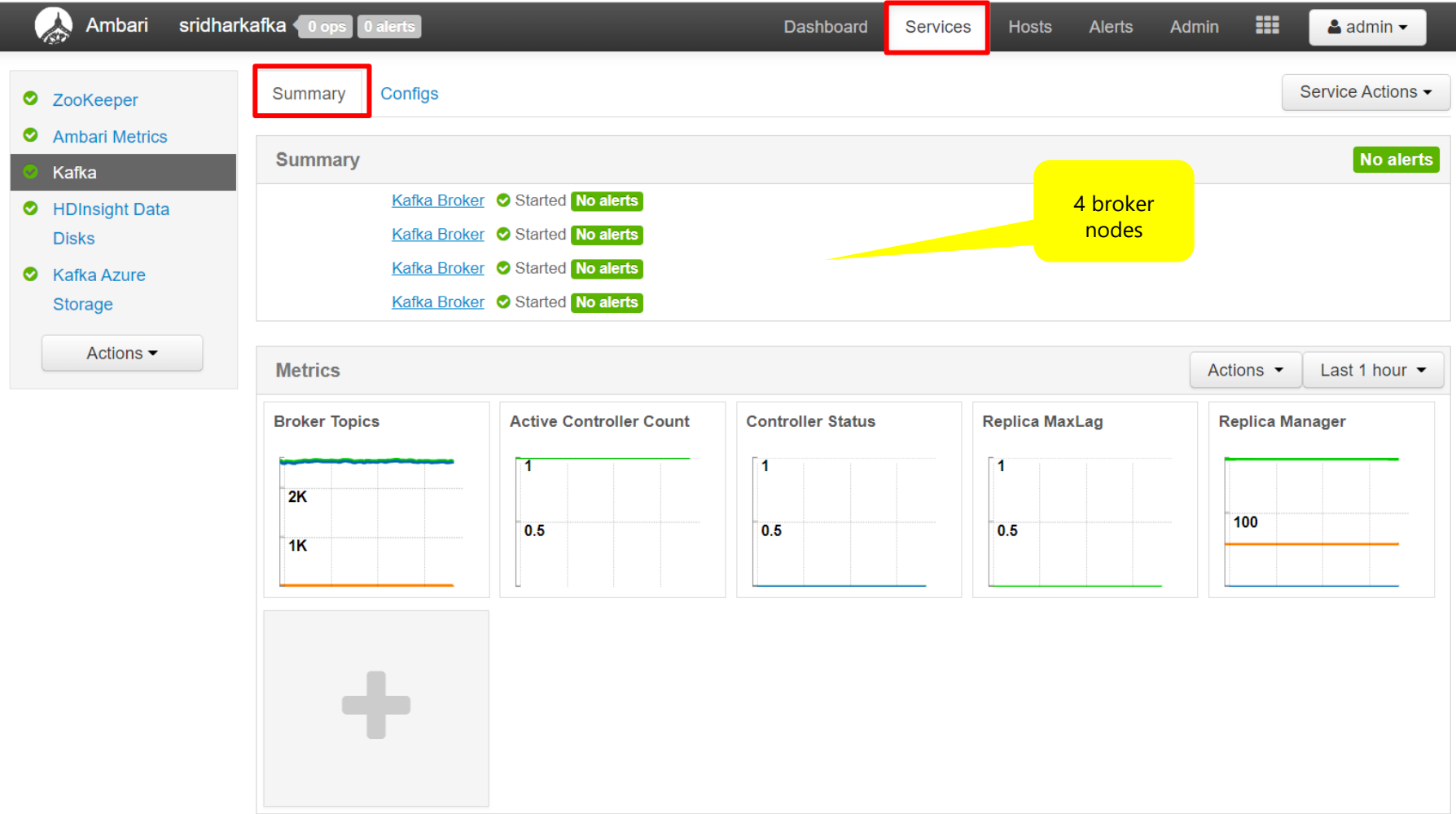
On the right side of the overview, a red box highlights the 'URL' field, which displays 'https://sridharkafka.azurehdinsight.net'. A yellow callout points to this URL with the text 'Cluster URL'.

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



Start

- Stop
- Restart All
- Run Service Check
- Turn On Maintenance Mode
- Add ZooKeeper Server
- Download Client Configs

Kafka Configs

Ambari

sridharkafka

0 ops0 alerts

Dashboard

Services

Hosts

Alerts

Admin

admin

ZooKeeper

Ambari Metrics

Kafka

HDInsight Data

Kafka Azure

Disks

Storage

Actions

Summary

Configs

Service Actions

Group

Default (9)

Manage Config Groups

Filter...

V2

internal

9 hours ago

HDP-2.6

V1

hdinsightwatchd...

9 hours ago

HDP-2.6

V2

internal authored on Sun, Feb 23, 2020 12:56

Discard

Save

Kafka Broker

Advanced kafka-broker

auto.create.topics.enable

false

- 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

▼ Kafka Broker

Kafka Broker hosts

wn0-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net and 3 others

zookeeper.connect

zk4-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net:2181,zk0-sridha.mddr2:

log.dirs

/data_disk_0/kafka-logs,/data_disk_1/kafka-logs

log.roll.hours

168

log.retention.hours

168

listeners

PLAINTEXT://localhost:9092

▶ 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-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

▼ Advanced kafka-env

is_supported_kafka_ranger

true

+

C

kafka_log_dir

/var/log/kafka

+

C

Kafka PID dir

/var/run/kafka

kafka_user_nofile_limit

128000

+

C

kafka_user_nproc_limit

65536

+

C

kafka-env template

```
#!/bin/bash

# Set KAFKA specific environment variables here.

# The java implementation to use.
export JAVA_HOME={{java64_home}}
export PATH=$PATH:$JAVA_HOME/bin
export PID_DIR={{kafka_pid_dir}}
export LOG_DIR={{kafka_log_dir}}
export KAFKA_KERBEROS_PARAMS={{kafka_kerberos_params}}
export JMX_PORT=${JMX_PORT:-9999}
export MAX_WAIT_TIME=600

# Add kafka sink to classpath and related dependencies
if [ -e "/usr/lib/ambari-metrics-kafka-sink/ambari-metrics-kafka-sink.jar" ]; then
    export CLASSPATH=$CLASSPATH:/usr/lib/ambari-metrics-kafka-sink/ambari-metrics-kafka-sink.jar
    export CLASSPATH=$CLASSPATH:/usr/lib/ambari-metrics-kafka-sink/lib/*
fi
if [ -f /etc/kafka/conf/kafka-ranger-env.sh ]; then
```

+

C

- ▶ 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 Drill Down

zk0-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net

Back

Summary

Configs

Alerts 0

Versions

Host Actions

Components

+ Add

ZooKeeper Server / ZooKeeper

Started

Metrics Monitor / Ambari Metrics

Started

Clients / ZooKeeper Client

Installed

Summary

Hostname: zk0-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net

IP Address: 10.0.0.14

Rack: /default-rack

OS: ubuntu16 (x86_64)

Cores (CPU): 4 (4)

Disk: 26.91GB/163.27GB (16.48% used)

Memory: 7.77GB

Load Avg: 2.52

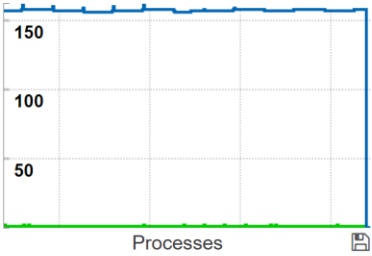
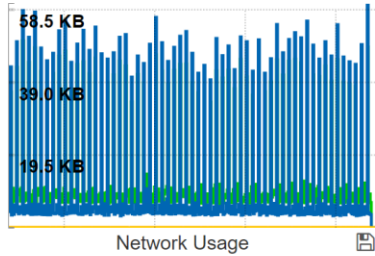
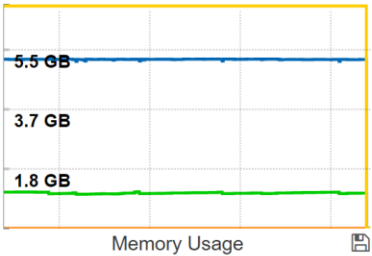
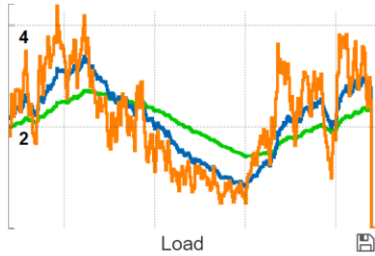
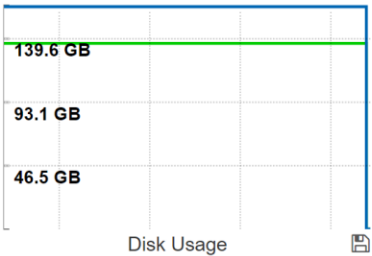
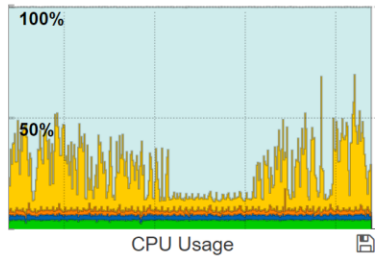
Heartbeat: less than a minute ago

Current Version: 2.6.5.3016-3

Unlimited JCE i... true

Host Metrics

Last 1 hour



- Start All Components
- Stop All Components
- Restart All Components
- Set Rack
- Turn On Maintenance Mode
- Delete Host
- Download Client Configs
- Recover Host

Hosts

Ambari

sridharkafka

0 ops

0 alerts

Dashboard

Services

Hosts

Alerts

Admin

admin

Actions

Filter by host and component attributes or search by keyword ...

<input type="checkbox"/>	Name	IP Address	Rack	Cores	RAM	Disk Usage	Load Avg	Versions	Components
<input type="checkbox"/>	hn0-sridha.mddr2xwli4wenc...	10.0.0.21	/default-rack	4 (4)	13.67GB		2.39	HDP-2.6.5.3016-3	5 Components
<input type="checkbox"/>	hn1-sridha.mddr2xwli4wenc...	10.0.0.20	/default-rack	4 (4)	13.67GB		0.88	HDP-2.6.5.3016-3	3 Components
<input type="checkbox"/>	wn0-sridha.mddr2xwli4wenc...	10.0.0.13	/default-rack	4 (4)	13.67GB		0.72	HDP-2.6.5.3016-3	4 Components
<input type="checkbox"/>	wn1-sridha.mddr2xwli4wenc...	10.0.0.11	/default-rack	4 (4)	13.67GB		0.84	HDP-2.6.5.3016-3	4 Components
<input type="checkbox"/>	wn2-sridha.mddr2xwli4wenc...	10.0.0.7	/default-rack	4 (4)	13.67GB		0.80	HDP-2.6.5.3016-3	4 Components
<input type="checkbox"/>	wn3-sridha.mddr2xwli4wenc...	10.0.0.4	/default-rack	4 (4)	13.67GB		0.88	HDP-2.6.5.3016-3	4 Components
<input type="checkbox"/>	zk0-sridha.mddr2xwli4wencz...	10.0.0.14	/default-rack	4 (4)	7.77GB		1.41	HDP-2.6.5.3016-3	3 Components
<input type="checkbox"/>	zk1-sridha.mddr2xwli4wencz...	10.0.0.9	/default-rack	4 (4)	7.77GB		0.87	HDP-2.6.5.3016-3	3 Components
<input type="checkbox"/>	zk4-sridha.mddr2xwli4wencz...	10.0.0.5	/default-rack	4 (4)	7.77GB		1.02	HDP-2.6.5.3016-3	3 Components

Show: 101 - 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

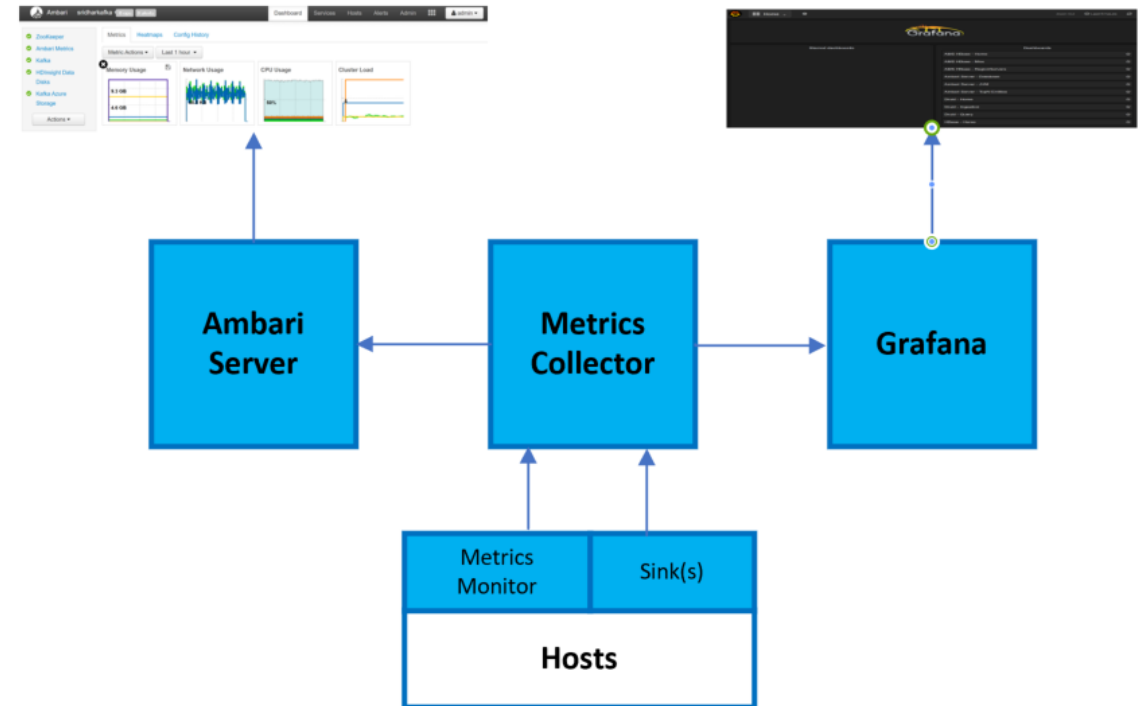
Azure HDInsight Data Disks Validator
Azure HDInsight Data Disks Watcher
Kafka Broker
Metrics Monitor

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://CLUSTERNAME.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	Status	Service	Last Status Changed	State
 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

Ambari

sridharkafka

0 ops

0 alerts

Dashboard

Services

Hosts

Alerts

Admin

admin

Kafka Broker Process

Back

OK (4)

Configuration

Edit

Description

This host-level alert is triggered if the Kafka Broker cannot be determined to be up.

Check Interval

5

Minute

Thresholds

OK

TCP OK - {0:.3f}s response on port {1}

WARNING

1.5

Seconds

TCP OK - {0:.3f}s response on port {1}

CRITICAL

5

Seconds

Connection failed: {0} to {1}:{2}

State:

Enabled

Service:

Kafka

Component:

Kafka Broker

Type:

PORT

Groups:

KAFKA Default

Last Changed:

Sun, Feb 23, 2020 13:08

Check Count:

3 (default)

Instances

Service	Host	Status	24-Hour	Response
Kafka	wn0-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net	OK for 20 hours	1	TCP OK - 0.000s response on port 9092
Kafka	wn1-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net	OK for 20 hours	1	TCP OK - 0.000s response on port 9092
Kafka	wn2-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net	OK for 20 hours	1	TCP OK - 0.000s response on port 9092
Kafka	wn3-sridha.mddr2xwli4wencz54kijmkpukh.jx.internal.cloudapp.net	OK for 20 hours	1	TCP OK - 0.000s response on port 9092

Show: 10 1 - 4 of 4

Kafka Broker Process

Back

Configuration

Description

This host-level alert is triggered if the Kafka Broker cannot be determined to be up.

Check Interval

5

Minute

Thresholds

OK

TCP OK - {0:.3f}s response on port {1}

WARNING

1.5

Seconds

TCP OK - {0:.3f}s response on port {1}

CRITICAL

5

Seconds

Connection failed: {0} to {1}:{2}

Cancel

Save

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

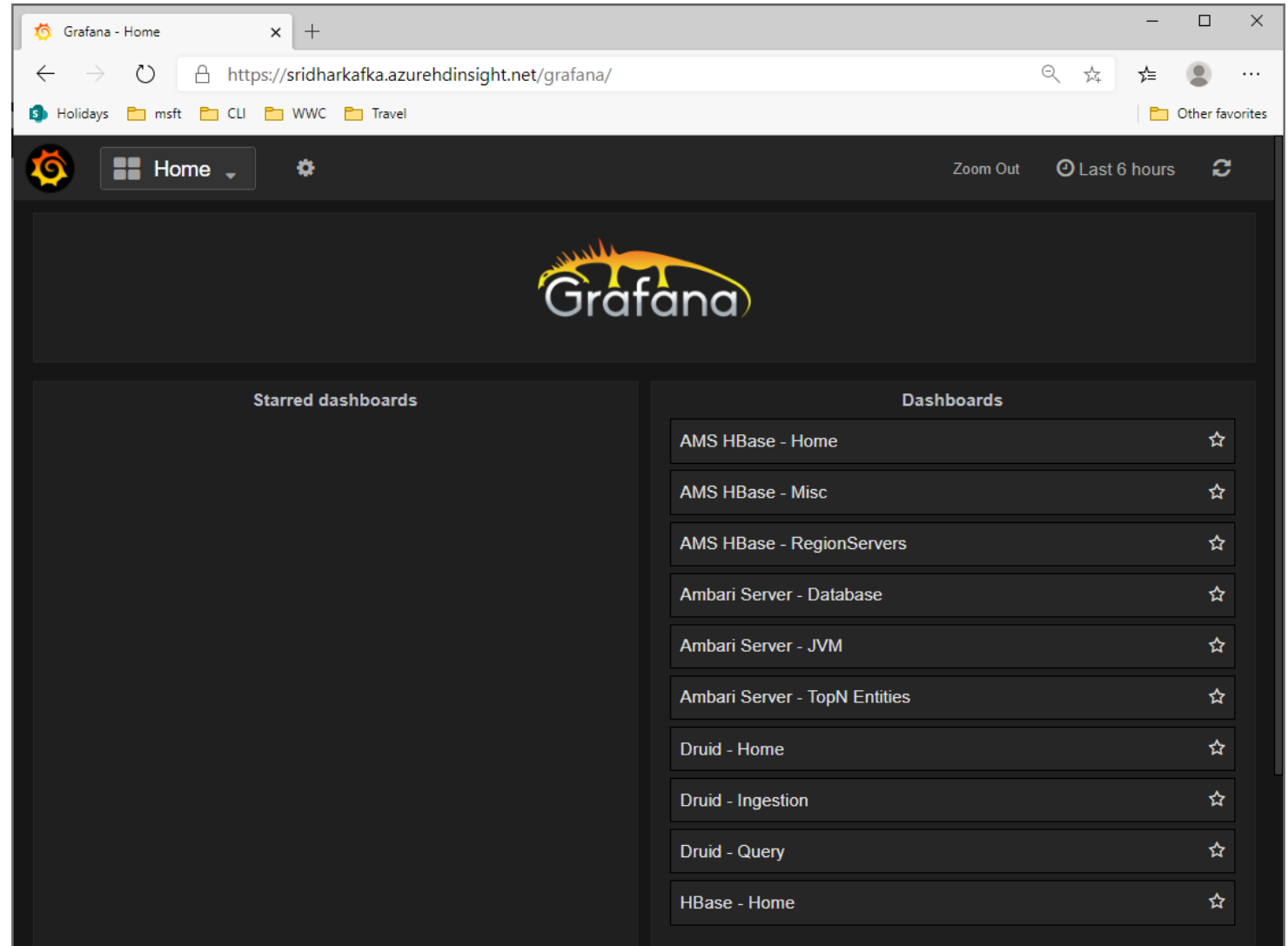
Save

Service		Component	Status
HDInsight Data Disks	●	Kafka Broker	<div>Enabled</div>
Kafka Azure Storage	●		
Kafka	●		Enable All Disable All
Ambari Metrics	●		
ZooKeeper	●		

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.

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



Q&A

Create a HDInsight Kafka Cluster

Microsoft Azure (Preview) [Report a bug](#) Search resources, services, and docs (G+)

Home > HDInsight_2020-02-23T18.49.15.318Z - Overview > sridharkafka - Cluster size

sridharkafka - Cluster size

HDInsight cluster

Search (Ctrl+/) Save Discard

This estimate does not include subscription discounts or costs related to storage, networking, or data transfer.

Node type	Node size	Number of nodes	Estimated cost/hour
Head node	D3 v2 (4 Cores, 14 GB RAM), 0.29 USD/hour	2	0.59 USD
Zookeeper node	A4 v2 (4 Cores, 8 GB RAM), 0.26 USD/hour	3	0.78 USD
Premium disks per worker node	P30 (1 TB per disk), 0.06 USD/hour	8	0.46 USD
Worker node	D3 v2 (4 Cores, 14 GB RAM), 0.29 USD/hour	<input type="text" value="4"/>	<input type="text" value="1.18 USD"/>

TOTAL COST 3.00
USD/HOUR (ESTIMATED)

Cluster size history

Number of Active Workers (Max) sridharkafka

Feb 17 Feb 19 Feb 21 Feb 23 UTC-06:00

--

Note the type and number of nodes in the kafka cluster