

GCS Fine-Grained Access Control (Preview)

Date published: 2023-06-29

Date modified: 2025-02-25

Legal Notice

© Cloudera Inc. 2023. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 ("ASLv2"), the Affero General Public License version 3 (AGPLv3), or other license terms.

Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners. Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER'S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Contents

Legal Notice	2
Contents	3
Introduction to RAZ in GCP environments	4
Supported use cases for RAZ in GCP environments	5
Limitations for using RAZ in GCP environments	5
GCP requirements for RAZ-enabled GCP environments	6
Registering a RAZ-enabled GCP environment	7
Using Cloudera UI to register RAZ-enabled GCP environment	7
Using Beta CDP CLI to register RAZ-enabled GCP environment	8
Ranger policy options for RAZ-enabled GCP environment	9
Troubleshooting	11
Creation of Kudu table is failing	11
RAZ fails to authorize request	11

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Introduction to RAZ in GCP environments

Cloudera on cloud defaults to using cloud storage which might be challenging while managing data access across teams and individual users. The Ranger Authorization Service (RAZ) resolves this challenge by enabling Google Cloud Storage (GCS) users to use fine-grained access policies and audit capabilities available in Apache Ranger similar to those used with HDFS files in an on-premises or IaaS deployment.

Many of the use cases that RAZ for GCS enables are cases where access control on files or directories is needed. Some examples include:

- Per-user home directories.
- Data engineering (Spark) efforts that require access to cloud storage objects and directories.
- Data warehouse queries (Hive/Impala) that use external tables.
- Access to Ranger's rich access control policies such as date-based access revocation, user/group/role-based controls, along with corresponding audit.
- Tag-based access control using the classification propagation feature that originates from directories.

Prior to the introduction of RAZ, controlling access to GCS could be enforced at coarse-grained group level using [IDBroker mappings](#). This required rearchitecting the implementation of important file-centric activities as well as admin-level access to both the Google Cloud account and the Cloudera account.

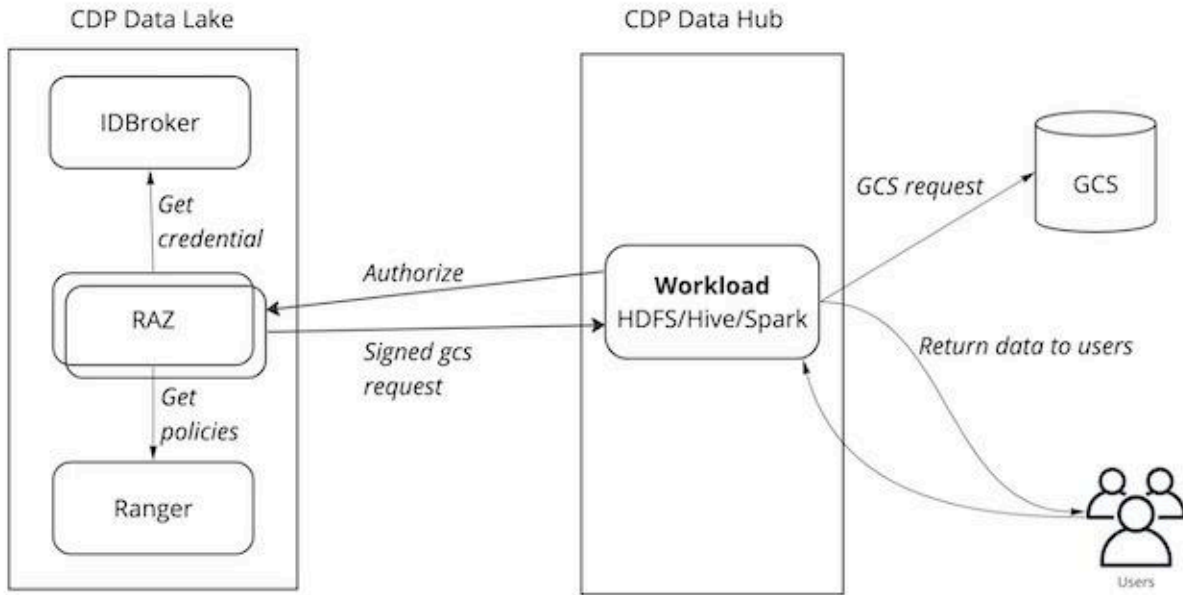
In HDP and CDH deployments, files and directories are protected with a combination of HDFS Access Control Lists (ACLs) (in CDH, HDP) and Ranger HDFS policies (in HDP). Similarly, in a Cloudera environment on GCP with RAZ for GCS enabled, Ranger's rich access control policies can be applied to Cloudera's access to GCS buckets, directories, and files and can be controlled with admin-level access to Cloudera alone.

Important

It is recommended that you do not set up IDBroker mappings for workload users in a RAZ-enabled GCP environment.

The following diagram illustrates RAZ authorization process for controlling user access to GCS:

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.



Supported use cases for RAZ in GCP environments

The core RAZ for GCP for Data Lakes and several Cloudera Data Hub templates are available. The following Cloudera Data Hub cluster types are supported with Cloudera Runtime 7.2.17 or newer:

- Data Engineering
- Data Engineering HA
- Data Engineering Spark3
- Data Mart
- Flow Management Light Duty
- Operational Database with SQL
- Streams Messaging HA

Limitations for using RAZ in GCP environments

The following limitations and known issues have been identified:

- RAZ on GCP uses [HMAC keys](#) for authentication. When multiple Cloudera environments are created on GCP with RAZ enabled, the account limit of 10 HMAC keys per service account may be reached. To avoid reaching the limit, you should create a new [Ranger Cloud Access Authorizer](#) service account for each Cloudera environment on GCP.
- Due to the fact that Google HMAC key generation cannot be limited in scope, the Ranger Cloud Access Authorizer service account has elevated credentials to create HMAC keys for any service account in the same GCP project. Therefore, for security reasons, it is

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

recommended that you create all Cloudera service accounts and storage buckets in a separate dedicated project.

- There is currently no automated way to enable RAZ in an existing CDP environment which does not have RAZ enabled. If you have a non-RAZ environment, you should create a new environment with RAZ, and then migrate your data using the steps described in [Data Lake backup and restore documentation](#).
- Hue file browser is not supported.
- Flink is not supported.
- S3A access and GCS access may not work from the same cluster without configuration changes, because there are overlapping configuration parameter names for GCS and S3.
- Enabling RAZ for GCP is supported only with Cloudera Runtime 7.2.17 or newer.

GCP requirements for RAZ-enabled GCP environments

The minimal GCS cloud storage setup recommended for production is described in the [Minimum setup for cloud storage for GCP](#). In addition to this, you should create the following service account for RAZ:

Service account name	Description	Required IAM roles or permissions	Scope
Ranger Cloud Access Authorizer	This service account will be used by Cloudera services to access workload data via RAZ. It provides full access to the data storage location.	<p>Choose one of the following two approaches:</p> <ul style="list-style-type: none"> - Option A: Use both built-in and custom roles - Option B: Use only a custom role <p>Option A: Use both built-in and custom roles:</p> <ul style="list-style-type: none"> • Assign the storage.objectAdmin built-in GCP role • Assign the storage.hmacKeyAdmin built-in GCP role • Create and assign a custom role that grants the storage.buckets.get permission. <p>Option B: Use only a custom role:</p>	<p><i>Data storage bucket</i></p> <p>For Data Lake backup and restore: <i>Backups</i> bucket, if different from the main data storage bucket</p>

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

		<p>Create a custom role with the below permissions and assign it:</p> <ul style="list-style-type: none"> • storage.buckets.get • storage.hmacKeys.create • storage.hmacKeys.delete • storage.hmacKeys.get • storage.hmacKeys.list • storage.hmacKeys.update • storage.multipartUploads.abort • storage.multipartUploads.create • storage.multipartUploads.list • storage.multipartUploads.listParts • storage.objects.create • storage.objects.delete • storage.objects.get • storage.objects.getIamPolicy • storage.objects.list • storage.objects.setIamPolicy • storage.objects.update 	
--	--	--	--

Use the usual instructions to create GCS buckets, create the service accounts, create custom roles, and add service accounts as members and users to other service accounts. Refer to the instructions in the [Minimum setup for cloud storage for GCP](#).

Note that this additional **Ranger Cloud Access Authorizer** service account mentioned above also needs to be added as a member to the *Data storage* bucket by using the steps described in [Add service accounts as members to buckets](#).

Registering a RAZ-enabled GCP environment

You can use the Cloudera web interface or Beta CDP CLI to register a RAZ-enabled GCP environment.

You can enable RAZ on the latest available version of Cloudera Runtime. The minimum version supporting RAZ for GCP environments is Cloudera Runtime 7.2.17.

Using Cloudera UI to register RAZ-enabled GCP environment

Follow the usual instructions to [register a GCP environment](#) in Cloudera. Make sure to do the following:

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

1. Select GCP as the cloud provider and then provide the necessary details.
2. Make sure to select the Cloudera Runtime version 7.2.17 or newer.
3. To enable Ranger RAZ authorization, click on **Enable Authorization for Google Cloud Storage** and then select the **Ranger Cloud Access Authorizer** service account (that you created earlier) from the dropdown:

Environments / Environments

✓ Register Environment

- Name your environment
- Select a cloud provider
- Provide a credential for your cloud provider account

2 Data Access and Data Lake Scaling

- Provide Data Lake name
- Add Data Access and Audit
- Choose Data Lake scale

3 Region, Networking and Security

- Select a region
- Define security access
- Add your SSH settings

4 Storage

- Add Logs Storage

Data Lake Settings

Data Lake Name *

Data Lake version *

Runtime 7.2.17

Fine-grained access control on GCS

Choose how to manage cloud storage access control for this environment

☒ Enable Ranger authorization for Google Cloud Storage

Select GCP Service Account for Ranger authorizer *

[Click here](#) to refresh GCP Service Account from the cloud provider.

4. Continue registering the environment as usual.

Using Beta CDP CLI to register RAZ-enabled GCP environment

RAZ for GCP is supported via Beta CLI. Below are the steps to configure RAZ via Beta CDP CLI:

1. Install Beta CLI. See [Install Beta CDP CLI](#).
2. Create a GCP environment with CDP CLI by using the `cdp environments create-gcp-environment` command. For more information, see [Register a GCP environment from CDP CLI](#).
3. Use the following command to set empty IDBroker mappings for the environment. Since RAZ is to be configured, `ranger-cloud-access-authorizer-role` also needs to be set. Below is a CLI command example:

Unset

```
cdp environments set-id-broker-mappings \
--environment-name myraz-gcp-env \
```

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.


```
--data-access-role
datalakeadmin@gcp-dev-cloudbreak.iam.gserviceaccount.com \
--set-empty-mappings \
--ranger-audit-role
rangeraudit@gcp-dev-cloudbreak.iam.gserviceaccount.com \
--ranger-cloud-access-authorizer-role
datalakeadmin@gcp-dev-cloudbreak.iam.gserviceaccount.com
```

4. Create RAZ enabled Data Lake using Beta CLI. Note that the `--enable-ranger-raz` option is provided to enable RAZ. Runtime version should be greater than or equal to 7.2.17. Below is a CLI command example:

```
Unset
./cdp.sh datalake create-gcp-datalake \
--environment-name myraz-gcp-env \
--datalake-name myraz-gcp-dl \
--runtime 7.2.17 \
--cloud-provider-configuration
"serviceAccountEmail=idbroker@gcp-dev-cloudbreak.iam.gserviceaccount.com,storageLocation=gs://perf-team-west2-bucket" \
--enable-ranger-raz --scale LIGHT_DUTY
```

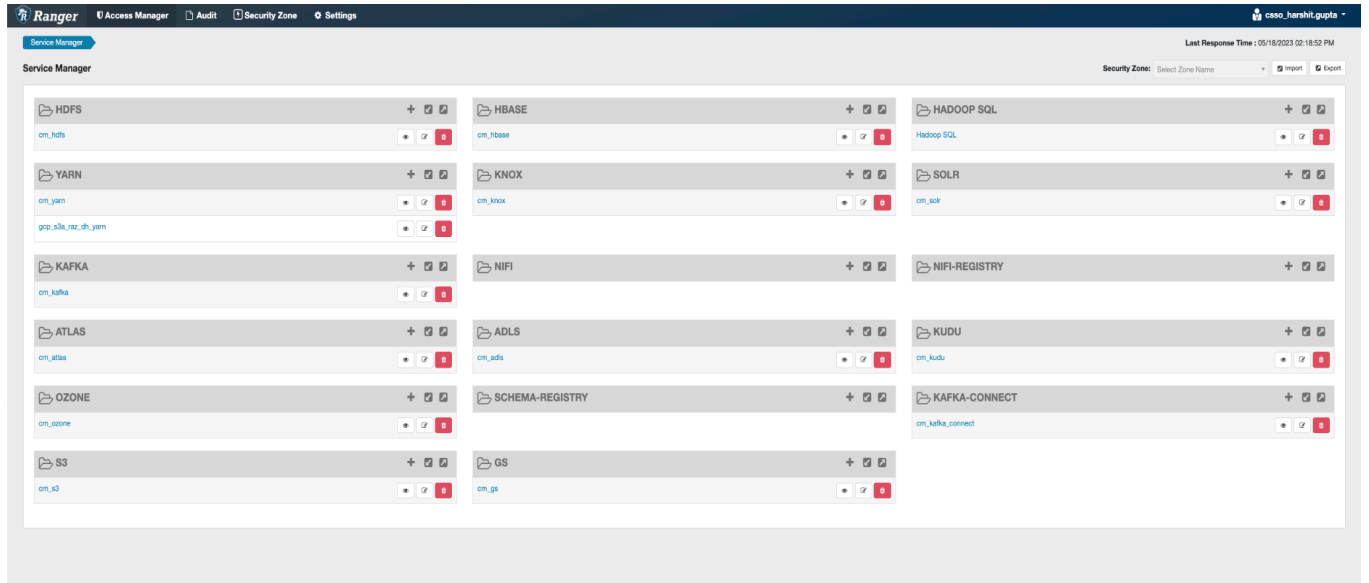
5. Once Data Lake is created successfully, log in to Cloudera Manager web UI and verify that ranger-RANGER_RAZ server has been configured.

Ranger policy options for RAZ-enabled GCP environment

The default Ranger policies for GCS are located under the **GS Tab (ServiceDef)** in the **cm_gs repo**:

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

CLOUDERA TECHNICAL PREVIEW DOCUMENTATION



This is the default view of the **cm_gs** repo with the default Ranger policies specified:

The screenshot shows the 'List of Policies' page for the **cm_gs** repository. It features a search bar and a table with columns: Policy ID, Policy Name, Policy Labels, Status, Audit Logging, Roles, Groups, Users, and Action. The table lists 13 policies, all with a status of 'Enabled' and 'Audit Logging' set to 'Enabled'. The policies include default audit logs, temporary files, Hadoop App Logs, Zappeln notebooks, HBase, Schema Registry, User Home, Data Catalog, Data Catalog Audit Logs, S3, S3 - user, and Automatic backup.

Policy ID	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
88	all - bucket_path	--	Enabled	Enabled	--	cm_ranger_admins,1st473ee	reguser	+ -
89	Default: Hive warehouse locations	--	Enabled	Enabled	--	--	hive,impala,spark	+ -
90	Default: Ranger Audit Logs	--	Enabled	Enabled	--	--	hive,atlas,atlasquery,atlas,atlas...	+ -
91	Default: S3 Audit Logs	--	Enabled	Enabled	--	--	s3	+ -
92	Default: Temporary files	--	Enabled	Enabled	--	hadoop	--	+ -
93	Default: Hadoop App Logs	--	Enabled	Enabled	--	--	impala,yarn	+ -
94	Default: Hadoop App Logs - user	--	Enabled	Enabled	--	--	USER	+ -
95	Default: Zappeln notebooks	--	Enabled	Enabled	--	--	spark	+ -
96	Default: HBase	--	Enabled	Enabled	--	--	hbase	+ -
97	Default: Schema Registry	--	Enabled	Enabled	--	--	schemaregistry	+ -
98	Default: User Home	--	Enabled	Enabled	--	--	USER	+ -
99	Default: Data Catalog	--	Enabled	Enabled	--	--	spark	+ -
100	Default: Data Catalog Audit Logs	--	Enabled	Enabled	--	--	spark	+ -
101	Default: S3 - user	--	Enabled	Enabled	--	--	USER	+ -
102	Default: S3 - user	--	Enabled	Enabled	--	--	USER	+ -
103	Default: Automatic backup	--	Enabled	Enabled	--	--	hive,s3,s3,atlas,atlas...	+ -

The Ranger policy creation works similar to that of RAZ for S3. You just specify the GCS bucket name, the path, and the policy parameters:

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Policy Details:

Policy Type: Access

Policy Name: Enabled Normal Add Validity Period

Policy Label:

GS Bucket:

Path: Recursive

Description:

Audit Logging: Yes

Allow Conditions:

Select Role	Select Group	Select User	Permissions	Delegate Admin	
<input type="text" value="Select Roles"/>	<input type="text" value="Select Groups"/>	<input type="text" value="cisco_harshit.gupta"/>	Permissions	<input type="checkbox"/>	✖
+					
⚠ Exclude from Allow Conditions:					
<input type="text" value="Select Roles"/>	<input type="text" value="Select Groups"/>	<input type="text"/>	Add Permissions	<input type="checkbox"/>	✖
+					

Troubleshooting

Refer to this section to learn about errors that you may encounter and steps to resolve them.

Creation of Kudu table is failing

Problem: The Kudu service user is not authorized to access Hive warehouse locations on cloud object stores, which under certain conditions can prevent Kudu tables from being created. This results in the following error:

```
ImpalaRuntimeException: Error creating Kudu table 'default.truckspeedevents'
CAUSED BY: NonRecoverableException: failed to create HMS catalog entry for table
[id=b764bceb167746b7bb3dc1e8722e66e6]: failed to create Hive MetaStore table:
TException - service has thrown: MetaException(message=Got exception:
java.nio.file.AccessDeniedException
```

Workaround: Add "kudu" service user to the allow list for "Default: Hive warehouse locations" in the `cm_gs` Ranger repository.

RAZ fails to authorize request

Problem: RAZ fails to authorize a request with the following error.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

StackTrace on Client side:

```
hdfs dfs -ls gs://perf-team-west2-bucket/
```

```
...
```

CLOUD_PROVIDER_ERROR:

org.apache.ranger.raz.s3.lib.exceptions.RCSUECloudProviderException: Could not determine region for bucket: [perf-team-west2-bucket]

```
org.apache.ranger.raz.s3.lib.signer.util.ExceptionUtils.createUnableToFindRegionException(ExceptionUtils.java:29)
```

```
org.apache.ranger.raz.processor.s3.signer.S3RemoteSigner.getRegion(S3RemoteSigner.java:217)
```

```
org.apache.ranger.raz.processor.s3.signer.S3RemoteSigner.getAuthorizationRequest(S3RemoteSigner.java:379)
```

```
org.apache.ranger.raz.processor.s3.S3RazProcessor.getRangerAccessRequests(S3RazProcessor.java:183)
```

```
org.apache.ranger.raz.processor.s3.S3RazProcessor.preProcess(S3RazProcessor.java:120)
```

```
org.apache.ranger.raz.RangerRemoteAuthorizer.authorize(RangerRemoteAuthorizer.java:151)
```

```
org.apache.ranger.raz.rest.AuthzREST.authorizeAccess(AuthzREST.java:98)
```

```
ls: gs://perf-team-west2-bucket/:
```

```
org.apache.ranger.raz.hook.s3.RazS3ClientCredentialsException:
```

```
....
```

```
23/05/16 08:10:50 INFO impl.MetricsSystemImpl: Stopping s3a-file-system metrics system...
```

```
23/05/16 08:10:50 INFO impl.MetricsSystemImpl: s3a-file-system metrics system stopped.
```

23/05/16 08:10:50 INFO impl.MetricsSystemImpl: s3a-file-system metrics system shutdown complete.

StackTrace on RAZ server side:

```
com.amazonaws.SdkClientException: Unable to load AWS credentials from any
provider in the chain:
[org.apache.ranger.raz.processor.gs.RazGSCredentialProvider@4eeb9a91:
RazGSCredentialsProvider.java: Exception while creating HMAC Keys in GCP. If service
account HMAC key limit reached, manually remove the key from GCP console and
restart the service.
```

Root cause: { com.google.cloud.storage.StorageException: Exception from GCP: Service account HMAC key limit reached }

```
com.amazonaws.auth.AWSCredentialsProviderChain.getCredentials(AWSCredentials
ProviderChain.java:136)
```

```
org.apache.ranger.raz.s3.lib.utils.S3RegionUtils.lookupRegion(S3RegionUtils.java:118
)
```

```
org.apache.ranger.raz.processor.s3.signer.S3RemoteSigner.getRegion(S3RemoteSign
er.java:212)
```

```
org.apache.ranger.raz.processor.s3.signer.S3RemoteSigner.getAuthorizationRequest(
S3RemoteSigner.java:379)
```

```
org.apache.ranger.raz.processor.s3.S3RazProcessor.getRangerAccessRequests(S3Ra
zProcessor.java:183)
```

```
org.apache.ranger.raz.processor.s3.S3RazProcessor.preProcess(S3RazProcessor.jav
a:120)
```

```
org.apache.ranger.raz.RangerRemoteAuthorizer.authorize(RangerRemoteAuthorizer.ja
va:151)
```

```
org.apache.ranger.raz.rest.AuthzREST.authorizeAccess(AuthzREST.java:98)
```

Workaround: As mentioned in the stack trace of server logs above, manually remove the key from the GCP console and restart the RAZ service.