

Configure for EC2 Role-Based Authentication


Contents:

- *IAM roles*
- *AWS System Mode*
- *Additional AWS Configuration*
- *Use of S3 Sources*

When you are running the Trifacta platform on an EC2 instance, you can leverage your enterprise IAM roles to manage permissions on the instance for the Trifacta platform. When this type of authentication is enabled, Trifacta administrators can apply a role to the EC2 instance where the platform is running. That role's permissions apply to all users of the platform.

IAM roles

Before you begin, your IAM roles should be defined and attached to the associated EC2 instance.

 **NOTE:** The IAM instance role used for S3 access should have access to resources at the bucket level.

For more information, see <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/iam-roles-for-amazon-ec2.html>.

AWS System Mode

To enable role-based instance authentication, the following parameter must be enabled.

```
"aws.mode": "system",
```

Additional AWS Configuration

The following additional parameters must be specified:

Parameter	Description
<code>aws.credentialProvider</code>	Set this value to <code>instance</code> . IAM instance role is used for providing access.
<code>aws.hadoopFsUseSharedInstanceProvider</code>	Set this value to <code>true</code> for CDH. The class information is provided below.

Shared instance provider class information

Hortonworks:

```
"com.amazonaws.auth.InstanceProfileCredentialsProvider",
```

Cloudera:


```
"org.apache.hadoop.fs.s3a.SharedInstanceProfileCredentialsProvider"
```

In the future:

CDH is moving back to using the `Instance` class in a future release. For details, see <https://issues.apache.org/jira/browse/HADOOP-14301>.

Use of S3 Sources

To access S3 for storage, additional configuration for S3 may be required.

 **NOTE:** Do not configure the properties that apply to `user` mode.

Output sizing recommendations:

- Single-file output: If you are generating a single file, you should try to keep its size under 1 GB.
- Multi-part output: For multiple-file outputs, each part file should be under 1 GB in size.
- For more information, see https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html