# Enable Integration with Compressed Clusters

**Contents:**



---

The Trifacta® platform can be configured to integrate with fully compressed Hadoop clusters. The following cluster compression methods are supported:

- Gzip
- Bzip2
- Snappy

Supported running environments:

- Trifacta Photon
- Spark

For more information, see *Running Environment Options*.

Hadoop clusters can be configured to enable compression of intermediate and/or final output data by default. The settings that are usually used to do so can be found in `mapred-site.xml` and `core-site.xml`.

## Pre-requisites

> **NOTE:** If you have not done so already, you must retrieve cluster configuration files and store them on the Trifacta node. For more information, see *Configure for Hadoop*.

## Enable integration with compression

**Steps:**

1. Edit the local version of `mapred-site.xml`. This file is typically located in `/etc/conf/hadoop`.
2. Add the following properties:

```
<configuration>
  ...
  <property>
    <name>mapreduce.map.output.compress</name>
    <value>true</value>
  </property>

  <property>
    <name>mapreduce.map.output.compress.codec</name>
    <value>org.apache.hadoop.io.compress.SnappyCodec</value>
  </property>

  <property>
    <name>mapreduce.output.fileoutputformat.compress</name>
    <value>true</value>
  </property>

  <property>
    <name>mapreduce.output.fileoutputformat.compress.type</name>
    <value>BLOCK</value>
  </property>

  <property>
    <name>mapreduce.output.fileoutputformat.compress.codec</name>
    <value>org.apache.hadoop.io.compress.SnappyCodec</value>
  </property>
  ...
</configuration>
```

3. Save the file and complete the following steps.

## Specify codecs

One or more compression/decompression methods (codecs) must be specified in `core-site.xml`.

**Steps:**

1. Edit the local version of `mapred-site.xml`. This file is typically located in `/etc/conf/hadoop`.
2. Specify the codecs to use in the `io.compression.codecs` property. Supported values:

| Code | Value |
|---|---|
| Gzip | `org.apache.hadoop.io.compress.GzipCodec` |
| Bzip2 | `org.apache.hadoop.io.compress.BZip2Codec` |
| Snappy | `org.apache.hadoop.io.compress.SnappyCodec` |

3. In the following example, all three codecs have been specified:

```
<configuration>
  ...
  <property>
    <name>io.compression.codecs</name>
   <value>org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.compress.BZip2Codec,org.apache.
hadoop.io.compress.SnappyCodec</value>
  </property>
  ...
</configuration>
```

4. Save the file.

## Configure platform

Apply the following changes from within the application to enable the Trifacta platform to communicate with the compressed cluster.

**Steps:**

1. Login to the application.
2. In the Admin Settings page, set the following settings:

| Setting | Description |
|---|---|
| `hadoopDefaultClusterCompression.enabled` | To enable integration with a compressed cluster, set this value to `true`. |
| `hadoopDefaultClusterCompression.compression` | Set this value to the type of compression applied on the cluster:<br>`none` - (default) no cluster compression<br>`gzip`<br>`bzip2`<br>`snappy` |

3. Save your changes and restart the platform.