

Known Issues for Oracle Big Data Cloud

Learn about issues you may encounter when using Oracle Big Data Cloud and how to work around them.

Topics:

- [Supported Browsers](#)
- [Browsing OCI Storage Using the UI Results in Incorrect URLs for Jobs](#)
- [Changing IDCS Password While Running Jobs Could Lock Account](#)
- [Large .inprogress Files Fill Up HDFS Storage](#)
- [Can't Drop Tables \(Beeline, Spark, Zeppelin\)](#)
- [Spark 1.6 Job Fails When Writing Parquet Data](#)
- [Ambari Quick Links Do Not Work](#)
- [High Performance Storage Not Supported on Oracle Cloud Infrastructure](#)
- [Big Data File System \(BDFS\) Not Supported on Oracle Cloud Infrastructure](#)
- [ORAAH Does Not Pick Up Default Configuration](#)
- [Accessing Swift from spark-shell Using Spark 1.6 Requires Additional Classpath](#)
- [Basic Cluster Profile Shows Spark Thrift Server URL](#)
- [Properties Overwritten When Cluster Upgraded to 17.4.1](#)
- [HDFS Tab in Data Stores Page Shows "Forbidden, no data to display"](#)
- [Spark UI Opens with an Error](#)

Supported Browsers

Oracle Big Data Cloud supports the following minimum requirements for web browsers:

Web Browser	Version
Microsoft Internet Explorer	11 and later
Google Chrome	29 and later
Mozilla Firefox	24 and later
Apple Safari	7 and later

Browsing OCI Storage Using the UI Results in Incorrect URLs for Jobs

Browsing for files on Oracle Cloud Infrastructure (OCI) storage using the Big Data Cloud Console results in the wrong URL being used in job specifications. Jobs then fail because the specified files can't be found. This occurs if either the bucket name or the namespace is in mixed case or uppercase. Due to a problem with how OCI URLs are supplied to the Big Data Cloud Console, the OCI URL incorrectly forces the bucket name and the namespace to lowercase. This issue doesn't affect the browsing of OCI block storage or the accessing of storage, but rather just the display of the URL in the Big Data Cloud Console and the URLs that are constructed from the OCI file system browser.

Workaround

Manually edit the bucket name and the namespace to their proper form. For instance, an OCI bucket defined as `oci://MyBucketName@MyNameSpace` will be returned to the user interface as `oci://mybucketname@mynamespace`. The workaround is to manually edit the file URL returned by the OCI file system browser from `oci://mybucketname@mynamespace/MyFile` to `oci://MyBucketName@MyNameSpace/MyFile` in the input field prior to submitting a job. If you don't make this edit, the specified file will not be found and the job will fail after being submitted.

Changing IDCS Password While Running Jobs Could Lock Account

Changing the Oracle Identity Cloud Service (IDCS) password while running jobs could lock the IDCS account.

Big Data applications are executed on multiple mappers/reducers (multiple threads, cores, and nodes) for parallelism. If the IDCS password is changed while these mappers/reducers are running, the IDCS account could be locked.

The following sequence of events could result in a locked IDCS account:

1. Run a job or long-running job.
2. Before the job completes, attempt to change the IDCS password in IDCS.
3. The running job continues to access the object store with the old password, resulting in a 401 error.

4. Because there are multiple tasks for the same job, similar access occurs many times, leading to the account being locked.

Workaround

To avoid this scenario, do the following:

1. Ensure that no jobs are running on the Big Data Cloud cluster and stop any running jobs. To do so:

- a. Log in to the Ambari user interface at `https://Ambari_server_IP_address:8080` using the user name and password specified for the cluster when the cluster was created.

Ambari_server_IP_address is the IP address for the Ambari server host. This address is listed on the Instance Overview page for a cluster in the service console for Oracle Big Data Cloud.

- b. Click **Spocs Fabric Service** on the left.
- c. From the **Service Actions** drop-down menu at the top, select **Stop**.

2. Change the IDCS password in IDCS.

3. Update the IDCS password in Big Data Cloud. To do so, SSH to the cluster by using the private key and update the IDCS password:

```
ssh -i private_key_file -l opc Ambari_server_IP_address
sudo -u spoccs-fabric-server -s
hadoop credential delete fs.swift.service.default.password -provider jceks://
hdfs/system/oracle/bdcsce/associations/jceks
hadoop credential create fs.swift.service.default.password -provider jceks://
hdfs/system/oracle/bdcsce/associations/jceks -value new_IDCS_password
```

```
hadoop credential delete fs.swift2d.service.default.password -provider jceks://
hdfs/system/oracle/bdcsce/associations/jceks
hadoop credential create fs.swift2d.service.default.password -provider jceks://
hdfs/system/oracle/bdcsce/associations/jceks -value new_IDCS_password
```

where:

- *private_key_file* is the path to the SSH private key file that matches the public key associated with the cluster.
- *Ambari_server_IP_address* is the IP address for the Ambari server host.
- *new_IDCS_password* is the new IDCS password.

4. Log in to the Ambari user interface as described in step 1, only this time select **Start** from the **Service Actions** drop-down menu.

5. Resubmit any stopped jobs.

Large .inprogress Files Fill Up HDFS Storage

When executing long-running streaming applications, HDFS storage may get filled up with large (multiple GB) *.inprogress files in the `hdfs://spark-history/` directory.

This is due to a known limitation in Apache Spark, where the event log is written to the `.inprogress` file until the application is terminated without rotating the file.

Workaround

Disable Spark event logging.

To disable Spark event logging globally, set the `spark.eventLog.enabled` configuration property to `false`. This can be done either by using the Ambari management console (see [Accessing Big Data Cloud Using Ambari](#) in *Using Oracle Big Data Cloud*) or by editing the Spark configuration files. You should also set the following Spark properties as shown below to limit the accumulation of event logs for completed jobs:

- `spark.history.fs.cleaner.maxAge 24h`
- `spark.history.fs.cleaner.interval 1h`

If you don't want to disable Spark event logging globally, you can disable it just for desired (long-running streaming) applications. For example, you could do the following:

- **Spark shell:** Set `spark-shell --conf spark.eventLog.enabled=true`
- **Spark submit:** Create a `SparkConfig` with the `spark.eventLog.enabled` property set to `false` and use `SparkContext` instantiated with it.
- **Zeppelin:** Add/set the `spark.eventLog.enabled` property with the value of `false` under the `spark2` section in Notebook settings. For information about how to change notebook settings, see [Managing Notebook Settings](#) in *Using Oracle Big Data Cloud*.

Can't Drop Tables (Beeline, Spark, Zeppelin)

If you create a table from Beeline, that table cannot be dropped from Spark and vice versa (tables created from Beeline cannot be dropped in Spark and tables created in Spark cannot be deleted from Beeline). The same behavior is exhibited when Zeppelin is used for creating a cluster. That is, if you create a table in Zeppelin, it cannot be dropped from a Spark job or Beeline and vice versa. This behavior is exhibited on Basic Auth and IDCS-enabled clusters in Oracle Big Data Cloud.

Workaround

This issue is caused by the difference in identities in the cluster. Drop the table as the same user that created the table.

Spark 1.6 Job Fails When Writing Parquet Data

A Spark 1.6 job fails with the following messages when writing data in Parquet format from Spark to the object store using `hadoop-openstack-driver`:

```
ParquetOutputCommitter: could not write summary file for swift://
java.io.FileNotFoundException: Not Found swift://
```

Workaround

If this happens, set the following configuration in Spark jobs:

```
sc.hadoopConfiguration.setBoolean("parquet.enable.summary-metadata", false)
```

Ambari Quick Links Do Not Work

Ambari Quick Links do not work. There is no specific workaround for this issue. However, to access the specific resource references from Ambari, the appropriate access ports need to be enabled as well as the resource referenced directly. The URL can be derived from the configuration and topology information for the specific service. However, it is generally not advised to publicly expose service endpoints directly because proper security will not be enforced.

High Performance Storage Not Supported on Oracle Cloud Infrastructure

High performance storage is not supported for clusters on Oracle Cloud Infrastructure, even though the **Use High Performance Storage** option is available in the console when you're creating a cluster. Do not select this option when creating a cluster on Oracle Cloud Infrastructure. High performance storage is supported only on Oracle Cloud Infrastructure Classic.

Big Data File System (BDFS) Not Supported on Oracle Cloud Infrastructure

The BDFS feature (currently based on Alluxio) is not supported on Oracle Cloud Infrastructure.

ORAAH Does Not Pick Up Default Configuration

When using any of the ORAAH libraries, the default configuration is not selected.

Workaround

Explicitly set the `java.library.path` using the `spark.executor.extraJavaOptions` parameter when connecting to Spark. For example:

```
spark.connect(master="yarn-client", memory="1g", dfs.namenode=dfs.namenode,
spark.app.name="ORAAH Test",
spark.executor.extraJavaOptions="-Djava.library.path=/usr/lib64/R/lib",
spark.eventLog.dir="hdfs:///spark-history",
spark.eventLog.enabled="true")
```

Accessing Swift from spark-shell Using Spark 1.6 Requires Additional Classpath

If you're using Spark 1.6, `spark-shell` can't access Swift and results in the following when attempting to do so:

```
java.lang.ClassNotFoundException: Class
org.apache.hadoop.fs.swift.snative.SwiftNativeFileSystem
```

Workaround

Launch `spark-shell` with the Swift JAR on the classpath as follows:

```
spark-shell with "--jars /opt/oracle/bdcsce/current/lib/hadoop-openstack-spoc.jar"
```

Basic Cluster Profile Shows Spark Thrift Server URL

The Spark Thrift Server is not available for clusters created with the Basic deployment profile. A Spark Thrift Server URL is shown in the Oracle Big Data Cloud console but there is no such endpoint. Do not attempt to make use of the Spark Thrift Server for clusters with the Basic deployment profile. Spark Thrift Server is only made available in the Full deployment profile.

Properties Overwritten When Cluster Upgraded to 17.4.1

The default settings for the following cluster configuration properties have been modified in the 17.4.1 release. If you've customized any of these properties, your changes will be overwritten during patching when the cluster is upgraded to 17.4.1. You'll need to reapply your changes after the cluster is upgraded.

config-type	property-key	Default property-value
alluxio-env	alluxio.worker.memory	1024MB
alluxio-site	alluxio.underfs.address	swift://bdcsce/
alluxio-site	alluxio.user.file.writetype.default	CACHE_THROUGH
core-site	bdcsce.idcs.enabled	false
hbase-env	phoenix_sql_enabled	true
hbase-site	hbase.region.server.rpc.scheduler.factory.class	org.apache.hadoop.hbase.ipc.PhoenixRpcSchedulerFactory
hbase-site	hbase.regionserver.wal.codec	org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec
hbase-site	hbase.rpc.controllerfactory.class	org.apache.hadoop.hbase.ipc.controller.ServerRpcControllerFactory
hbase-site	phoenix.functions.allowUserDefinedFunctions	true
hdfs-site	alluxio.zookeeper.enabled	true
hive-site	alluxio.zookeeper.enabled	true

config-type	property-key	Default property-value
spark-defaults	spark.driver.extraClassPath	/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar:/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar:/u01/bdcsce/opt/alluxio/conf/
spark-defaults	spark.executor.extraClassPath	/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar:/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar:/u01/bdcsce/opt/alluxio/conf/
spark-env	spark_thrift_cmd_opts	--master yarn --driver-class-path /opt/oracle/bdcsce/current/lib/hadoop-openstack.jar:/opt/oracle/bdcsce/current/lib/joss-bdcsce.jar:/opt/oracle/bdcsce/current/lib/ojdbc7.jar:/opt/oracle/bdcsce/current/lib/spoccs-hdfs-zeppelin-0.5.1.jar:/opt/oracle/bdcsce/current/lib/stocator-bdcsce.jar --jars /opt/oracle/bdcsce/current/lib/hadoop-openstack-spoc.jar,/opt/oracle/bdcsce/current/lib/joss-bdcsce.jar,/opt/oracle/bdcsce/current/lib/ojdbc7.jar,/opt/oracle/bdcsce/current/lib/spoccs-hdfs-zeppelin-0.5.1.jar,/opt/oracle/bdcsce/current/lib/stocator-bdcsce.jar,/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar,/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar --driver-java-options -Dspark.local.dir=/data/var/tmp
spark-hive-site-override	alluxio.zookeeper.enabled	true
spark2-defaults	spark.driver.extraClassPath	/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar:/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar:/u01/bdcsce/opt/alluxio/conf/
spark2-defaults	spark.executor.extraClassPath	/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar:/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar:/u01/bdcsce/opt/alluxio/conf/
spark2-env	spark_thrift_cmd_opts	--master yarn --driver-class-path /opt/oracle/bdcsce/current/lib/hadoop-openstack.jar:/opt/oracle/bdcsce/current/lib/joss-bdcsce.jar:/opt/oracle/bdcsce/current/lib/ojdbc7.jar:/opt/oracle/bdcsce/current/lib/spoccs-hdfs-zeppelin-0.5.1.jar:/opt/oracle/bdcsce/current/lib/stocator-bdcsce.jar --jars /opt/oracle/bdcsce/current/lib/hadoop-openstack-spoc.jar,/opt/oracle/bdcsce/current/lib/joss-bdcsce.jar,/opt/oracle/bdcsce/current/lib/ojdbc7.jar,/opt/oracle/bdcsce/current/lib/spoccs-hdfs-zeppelin-0.5.1.jar,/opt/oracle/bdcsce/current/lib/stocator-bdcsce.jar,/u01/bdcsce/opt/alluxio/core/client/target/alluxio-core-client-1.3.0-jar-with-dependencies.jar,/u01/bdcsce/usr/hdp/current/zeppelin-server/lib/guava-15.0.jar --driver-java-options -Dspark.local.dir=/data/var/tmp
spark2-hive-site-override	alluxio.zookeeper.enabled	true

config-type	property-key	Default property-value
zeppelin-config	zeppelin.helium.enabled	false
zeppelin-env	additional_cp_paths	/u01/bdcsce/opt/alluxio/conf
hdfs-log4j	content	(Content properties contain a large amount of script text. The exact text is not documented here.)
hive-env	content	(Content properties contain a large amount of script text. The exact text is not documented here.)

HDFS Tab in Data Stores Page Shows "Forbidden, no data to display"

Sometimes the HDFS tab in the Data Stores page in the Big Data Cloud Console does not display files and directories from the Hadoop Distributed File System and Oracle Cloud Storage container, but instead shows "Forbidden, no data to display."

This happens when the HDFS NameNode fails over to the standby HDFS NameNode. In this case, the NGINX reverse proxy would not have updated itself to point to the newly active HDFS NameNode and the Big Data Cloud Console would be attempting to communicate to the NameNode that is no longer the active NameNode. Therefore, fetching no data.

Workaround

To resolve this issue, you should switch back to the old active NameNode. That is, you should set the currently active NameNode to standby and make the standby NameNode active.

Perform the following steps:

1. SSH to any HDFS data node. See [Connecting to a Node by Using PuTTY on Windows or Connecting to a Node by Using SSH on UNIX in *Using Oracle Big Data Cloud*](#).

2. Log in as the HDFS user.

```
$ sudo su hdfs
```

3. Determine the active and the standby NameNode.

```
$ hdfs haadmin -getServiceState nn1
```

This will return the status of the node as active or standby.

4. If the node is active, execute:

```
$ hdfs haadmin -failover nn1 nn2
```

Or, if the node is in standby, execute:

```
$ hdfs haadmin -failover nn2 nn1
```


Spark UI Opens with an Error

Accessing the Spark UI fails with the following message:

```
"An error occurred.
```

```
Sorry, the page you are looking for is currently unavailable.  
Please try again later.
```

```
If you are the system administrator of this resource then you should check  
the error log for details.
```

```
Faithfully yours,  
nginx."
```

Workaround

If this happens, restart NGINX by performing the following steps:

1. Open the Oracle Big Data Cloud console. See Accessing the Oracle Big Data Cloud Console in *Using Oracle Big Data Cloud*.

The console opens, showing a list of clusters.

2. Click the name of the cluster for which you want to restart NGINX.

An overview page with cluster details is displayed. Nodes are listed under Resources.

3. Locate Master nodes and make note of their public IP addresses.

4. SSH to each master node. See Connecting to a Node by Using PuTTY on Windows or Connecting to a Node by Using SSH on UNIX in *Using Oracle Big Data Cloud*.

5. Execute the following command for each Master node:

```
$ sudo service nginx restart
```

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

Oracle® Cloud Known Issues for Oracle Big Data Cloud, Release 18.4

E83737-17

Copyright © 2017, 2018, Oracle and/or its affiliates. All rights reserved.

Describes information about known software issues and their workarounds for this release of Oracle Big Data Cloud.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.