Oracle Big Data Manager User's Guide

For Oracle Big Data Appliance

E96163-02
June 2018

Oracle Big Data Manager User's Guide, For Oracle Big Data Appliance

E96163-02

Primary Author: Ben Gelernter, Frederick Kush

Contents

A Keyboard Shortcuts for Oracle Big Data Manager

List of Tables

1

Overview of Oracle Big Data Manager

Oracle Big Data Manager makes it easy to copy data between data sources.

What is Oracle Big Data Manager?

Oracle Big Data Manager is a browser-based tool that gives you broad capabilities to manage data across your enterprise. You can use it to connect to and interconnect a range of supported Oracle and non-Oracle data storage providers, including Oracle Database, Oracle Object Store, MySQL, as well as Hadoop, S3, and GitHub. After you register storage providers with Big Data Manager, you can preview data and (depending upon the accessibility of each storage provider) compare, copy, and move data between them. With a Hadoop storage provider, you can also move data internally within HDFS, do data import/export and analytics with Apache Zeppelin, and import data into Hive tables. You can also upload data from your local computer to a selected storage provider.

Oracle Big Data Manager provides several methods for data transfer. You can use the console, which includes drag and drop data selection. Python and Java SDKs are available for building data management scripts and applications. There is also CLI for creating and administering data management jobs and tools for monitoring job status.

The Oracle Big Data Manager administrator can create other user accounts and assign roles to those accounts.

Feature Summary

The full list of Oracle Big Data Manager features is as follows:

- The Oracle Big Data Manager console, accessible through a browser-based GUI.
- Graphical tools for:
  - Comparing, copying, and moving data between storage providers.
  - Uploading files, extracting data from ZIP archives, and browsing data in Oracle Database and MySQL database.
  - Scheduling, managing, and monitoring copy, move, and compare jobs.
  - Importing data into Apache Hive.
  - Importing and exporting Apache Zeppelin notes; and creating and running notes.
  - Managing storage providers, users, and roles.
  - Monitoring the health of the cluster and the services running on it.
  - Processing and analyzing data via Apache Zeppelin notes.
- The `bdm-cli` utility, for copying data and managing copy jobs from the command line.
- Python and Java SDKs, for integrating Oracle Big Data Manager operations into applications

Supported Storage Providers

Oracle Big Data Manager supports the following storage providers, although not all tasks are supported in every provider:

- Hadoop Distributed File System (HDFS)
- Oracle Cloud Infrastructure Object Storage Classic
- Amazon Simple Storage Service (S3)
- Github
- Oracle Database
- Apache Hive
- MySQL database

How is Big Data Manager Installed and Configured?

Oracle Big Data Manager is installed automatically by the Mammoth installation of the Oracle Big Data Appliance software release. By default, it is installed on the same node where Cloudera Configuration Manager runs (usually node 3). No manual configuration is needed except to register storage providers.

The default port is 8890. The default password for the administrative account is the same as the Configuration Manager password. These are specified by the `BDP_PWD` and `BDM_PORT` parameters in the `<cluster name>-config.json` file. This file is one of the outputs generated when you use the Oracle Big Data Appliance Configuration Generation Utility to define your cluster and rack configuration.

> **See Also:**
>
> The chapter Using the Oracle Big Data Appliance Configuration Utility in the Oracle Big Data Appliance Owner's Guide describes `<cluster name>-config.json` , which contains the Mammoth installation parameters.

Limitations on use in Kerberos-Secured Clusters

In this release of Oracle Big Data Appliance, Oracle Big Data Manager is not available for clusters secured by Active Directory Kerberos.

MIT Kerberos is supported, except for clusters that use an external KDC.

2
Getting Started with Big Data Manager

Oracle Big Data Manager is installed and configured during the Mammoth installation of the Oracle Big Data Appliance software. No further configuration is required.

> **Note:**
>
> Some tasks described in this section require administrator privileges. When you are getting started, use the default `bigdatamgr` administrator account. Later on as `bigdatamgr` , you can add other users and selectively grant administrator privileges.

Topics:

2.1 Opening the Oracle Big Data Manager Console

The Oracle Big Data Manager console can be accessed from your web browser.

The Oracle Big Data Manager console is on the Cluster Manager host. The default port on Oracle Big Data Appliance is 8890.

`https://<cm_host>:8890`

Log on with the `bigdatamgr` administrator account. The password is the same as the Cloudera Manager password.

As `bigdatamgr` you can create login accounts for other users.

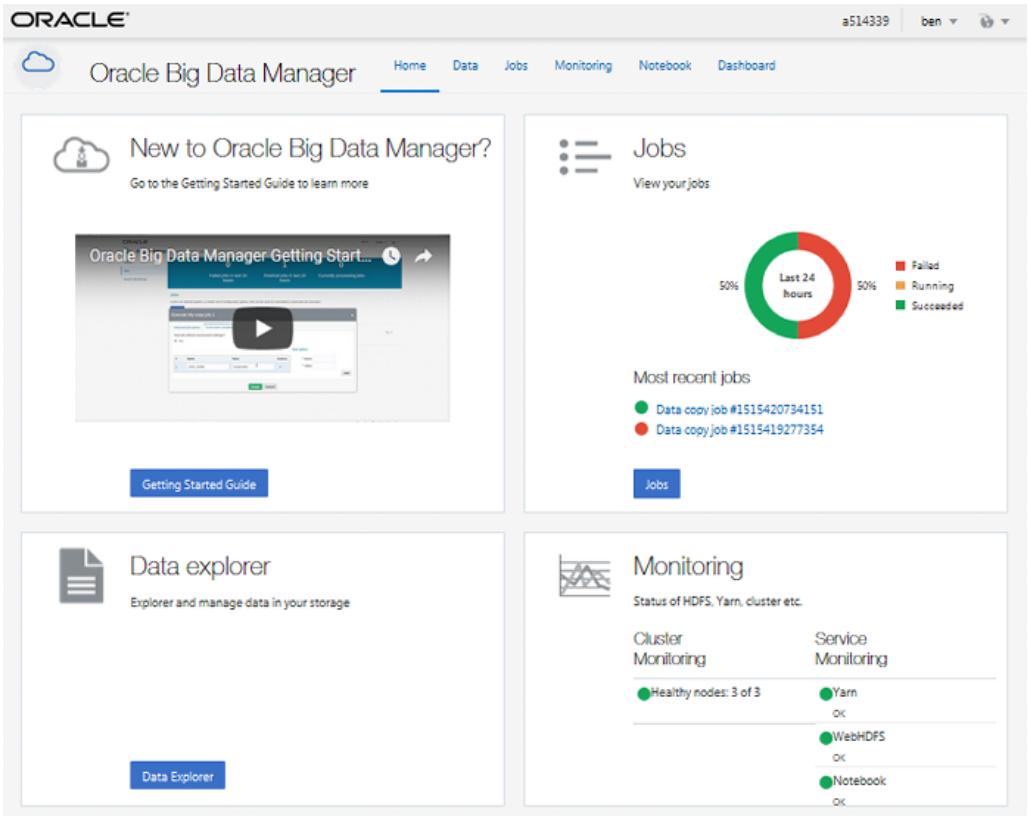
2.2 Navigating the Oracle Big Data Manager Console

The Oracle Big Data Manager console is displayed in a web browser and contains graphical tools for transferring and analyzing data and managing data providers, and for managing users, and roles.

The console has five main sections, which you can access by clicking the links in the tab bar at the top of the page.

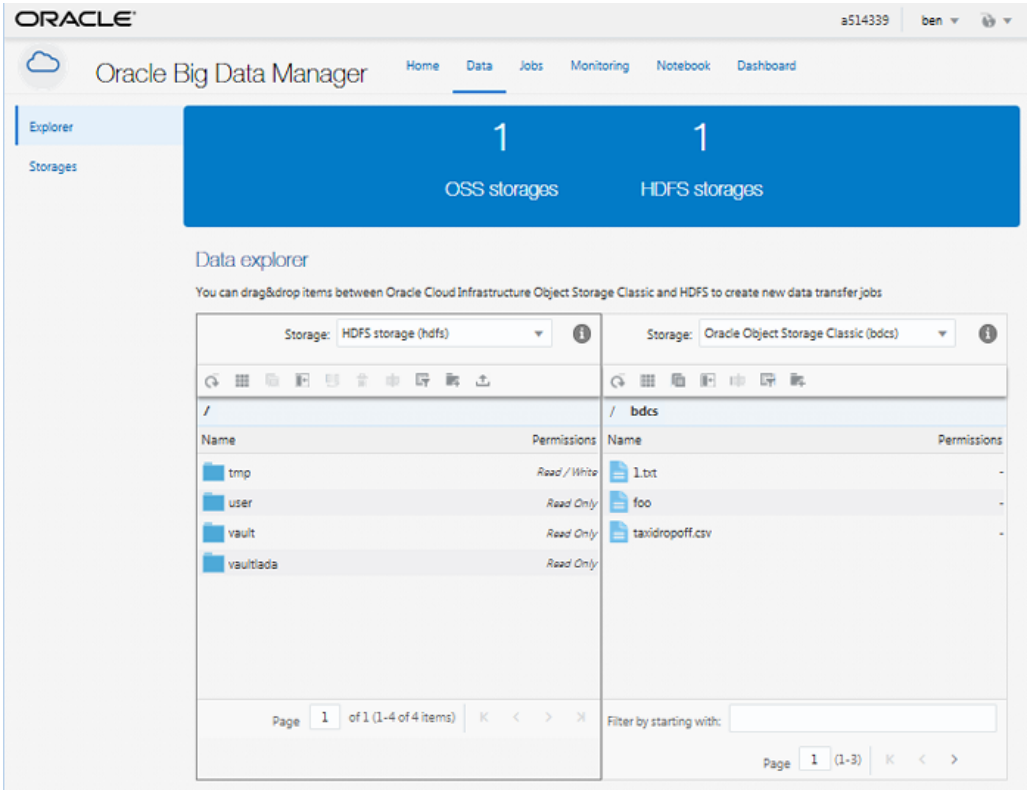
Home

The **Home** page includes basic instructions on how to use Oracle Big Data Manager, and some overview information about jobs and monitoring, along with links to the other main sections of the console.



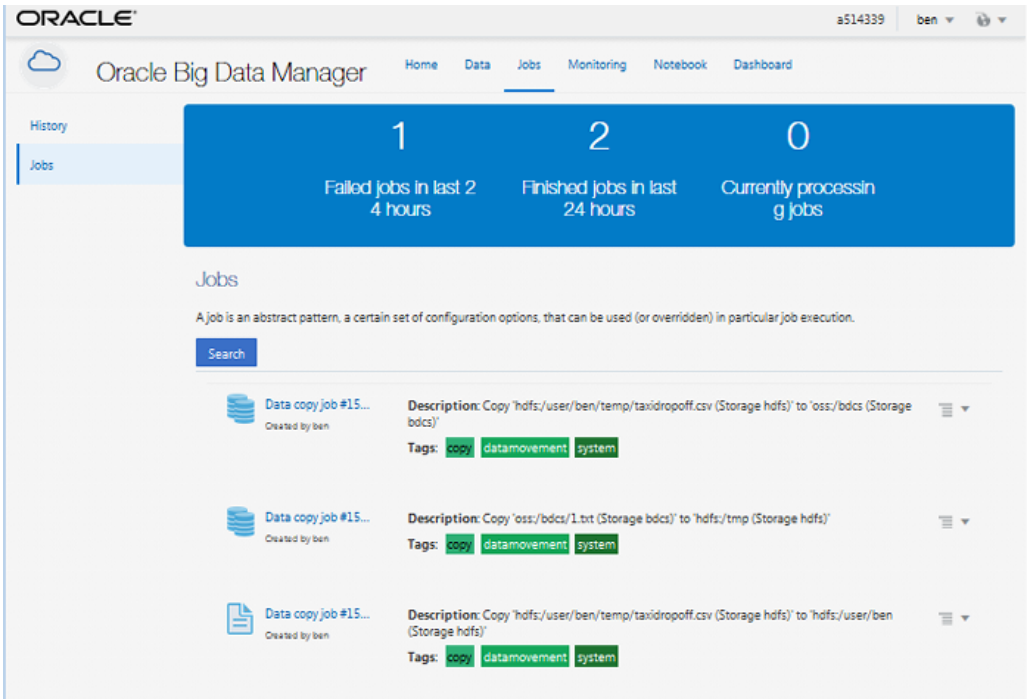
Data Explorer

Use the **Data** pages to compare, copy, and move files and containers between data sources, including HDFS. You can also upload files, extract the contents from a ZIP archive, import data to Apache Hive, and import and export Apache Zeppelin notes, among other tasks.



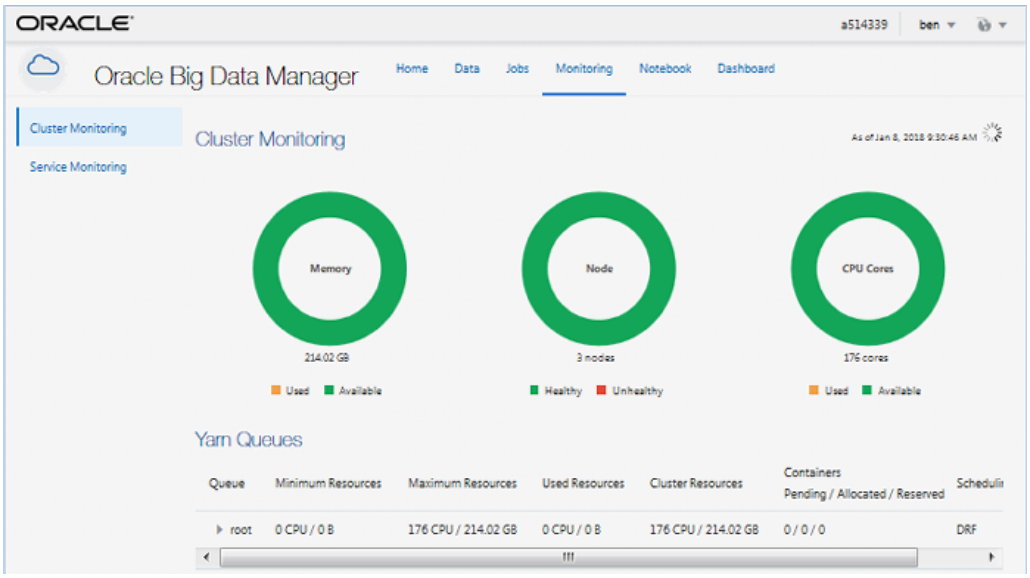
Jobs

Use the Jobs pages to review and manage copy jobs.



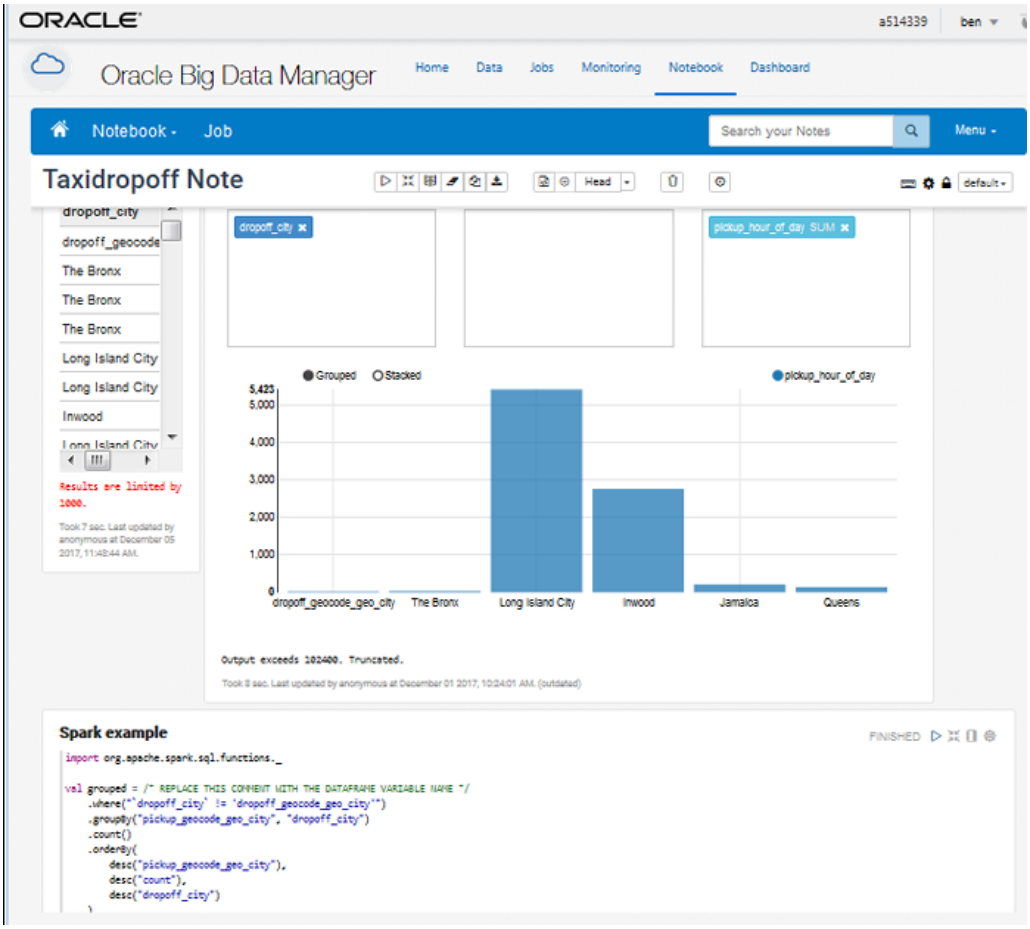
Monitoring

Use the **Monitoring** pages to monitor the performance of your cluster and the services running on it.



Notebook

Use the **Notebook** pages to process and analyze data by using Apache Zeppelin notes.



When a note is open in the console, you can click **Home** in the page banner to return to the **Notebook** home page.

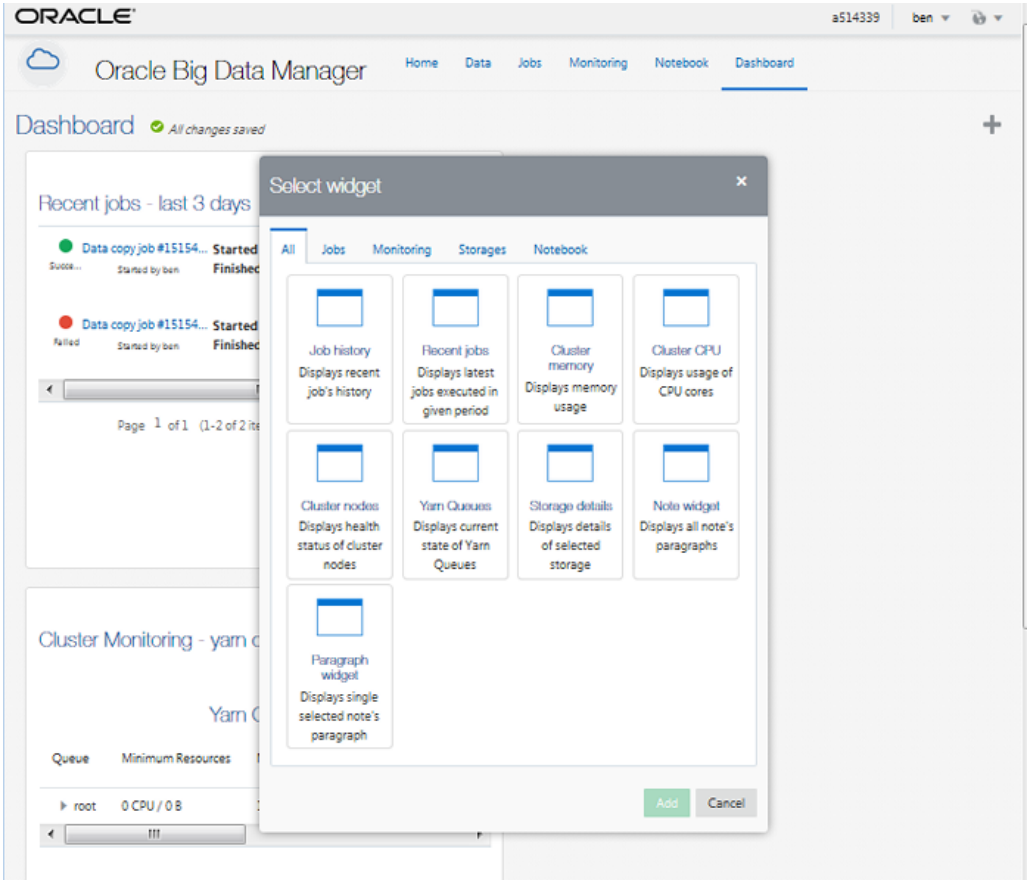
Administration

Use the **Administration** pages to manage users, roles, and storage providers.

> **Note:**
>
> Only users with administrative privileges have access to the Administration pages. The default `bigdatamgr` user that was created when the cluster was provisioned has administrative privileges.

Dashboard

Add widgets to your **Dashboard** to display details about status, history, and current activity.



2.3 Managing Oracle Big Data Manager Users, Roles, and Access

An administrator must create Oracle Big Data Manager users at the command line. Once they've been created, you can edit user details and manage access in the Oracle Big Data Manager console.

> **Note:**
>
> By default, the `bigdatamgr` user is created and granted the administrator role in Oracle Big Data Manager. This user should be used to grant roles and register providers. The `bigdatamgr` user has the same password as the Cloudera Manager administrator that was defined in Create Instance wizard when creating the cluster.

2.3.1 Adding Oracle Big Data Manager Users

An administrator must create Oracle Big Data Manager user accounts on the Linux command line. After creating a user account, the administrator can use the Oracle Big Data Manager console to edit user details and manage access.

To add a user:

1. Open a command shell and use SSH to connect to a cluster node as the `bigdatamgr` user (or another user with administration privileges).
2. Open a `root` shell:

   ```
   su -
   ```

3. Export the new user's password to a password file:

   ```
   user_password > user_password_file
   chmod 600 user_password_file
   ```

   where:

   - *`user_password`* is the password for the new user.
   - *`user_password_file`* is the password file for the new user. This file must have permissions `600`.

   > **Note:**
   >
   > It is a safer practice to define the user password as an environment variable and then pass that value to the command. When the value is passed as the value of the environment variable, the actual value won't be visible in the bash history. In this case, use the following, instead of the commands listed above.
   >
   > ```
   > echo ${USER_PASSWORD}>${USER_PASSWORD_FILE}
   > chmod 600 ${USER_PASSWORD_FILE}
   > ```
   >
   > where:
   >
   > - *`USER_PASSWORD`* is the environment variable containing the value of the password for the new user. The name of the environment variable can be any valid environment variable name.
   > - *`USER_PASSWORD_FILE`* is the environment variable containing the value of the password file for the new user. The name of the environment variable can be any valid environment variable name. This file has to have permissions `600`.

4. Add the user and create a home directory for the user in the cluster's HDFS file system:

   ```
   /usr/bin/bdm-add-user--create-hdfs-home new_user user_password_file
   ```

   where *`new_user`* is the new user name.

5. On the node where Oracle Big Data Manager runs, enter the following command to restart Oracle Big Data Manager. This reloads the user configuration from the database.

```
service bigdatamanager restart
```

> **Note:**
>
> On Oracle Big Data Appliance, Oracle Big Data Manager is by default hosted on the same node as Cloudera Manager and is accessed on port `8890`..

2.3.2 Editing User Details and Managing Roles

A user with administrator privileges can edit user details and manage roles in the Oracle Big Data Manager console.

To access and modify user details and manage user roles:

1. Sign in to the Oracle Big Data Manager console as the `bigdatamgr` user, or as another user with administrator privileges..
2. Click **Administration** at the top of the page to open the Administration page.
3. Click **Users** on the left of the page to show the list of users that have been added.
4. Edit details as needed.

2.3.3 Controlling Access to Specific Providers

A user with administrator privileges can control access to storage containers.

> **Note:**
>
> When a new cluster is created, the `bigdatamgr` user is created and granted the Oracle Big Data Manager Administrator role.

To control access to storage containers:

1. Sign into the Oracle Big Data Manager console as the `bigdatamgr` user, or another user with administrator privileges.
2. Click **Administration** at the top of the page to open the Administration page.
3. Click **Storages** on the left of the page to show a list of registered storage providers.
4. Click the menu icon to the right of the provider you are providing access to, and select **Manage Users**.
5. Use the arrows to move users from the left panel to the right panel to create an access list of users who will be able to see that provider in the web application. This doesn't give Write access to the storage. Users must have appropriate permissions to work with data in the provider.

2.4 Registering Storage Providers with Oracle Big Data Manager

You must register storage providers with Oracle Big Data Manager to be able to see and use them in the console.

To register a new provider:

1. Sign in to the Oracle Big Data Manager console as the `bigdatamgr` user, or as another user with administrator privileges.
2. Click **Administration** at the top of the page to open the **Administration** page.
3. Click **Storages** on the left of the page to show a list of registered storage providers.
4. Click the **Register new storage** button.
5. On the **General** page of the **Register New Storage** wizard, enter a name and description for the provider, select the storage type, and then click **Next.**
6. On the **Storage Details** page, provide details for accessing the provider.
7. On the **Access** page, specify which users can access this storage from within Oracle Big Data Manager. To add a user or user, select the name(s) in the left panel and click one of the arrows in the center, or drag the selected names(s) to the right panel.
8. Review the details on the **Confirmation** page and click the **Register** button.

2.4.1 Registering an Oracle Database Storage Provider

For an Oracle Database storage provider, use the **Storage Details** wizard page to supply the information needed to build a JDBC Thin Driver connection string.

Be sure that on the previous **General** page, you selected Oracle Database as the storage type.

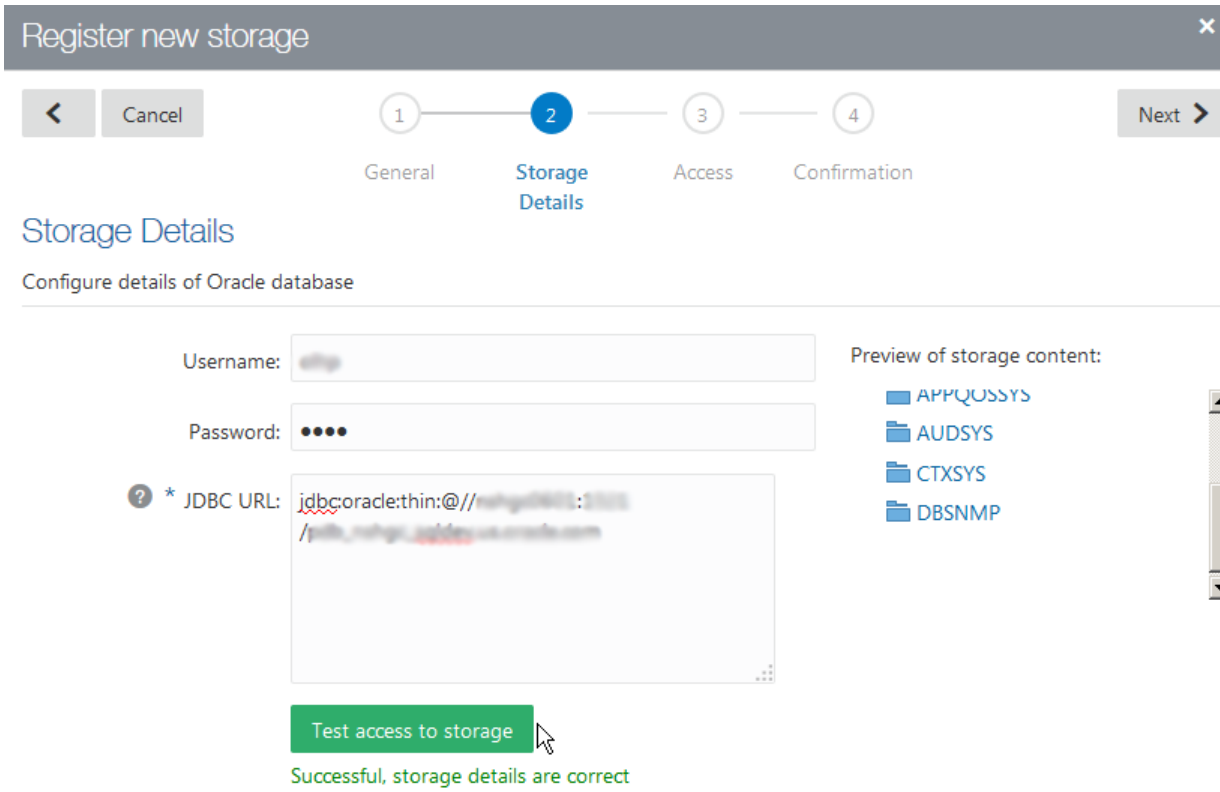
Provide the Storage Details

1. Enter your Oracle database username and password in the **Username** and **Password** fields.
2. In the **JDBC URL** field, edit the Oracle Database connection string template:

   ```
   jdbc:oracle:thin:@//host:port/service_name
   ```

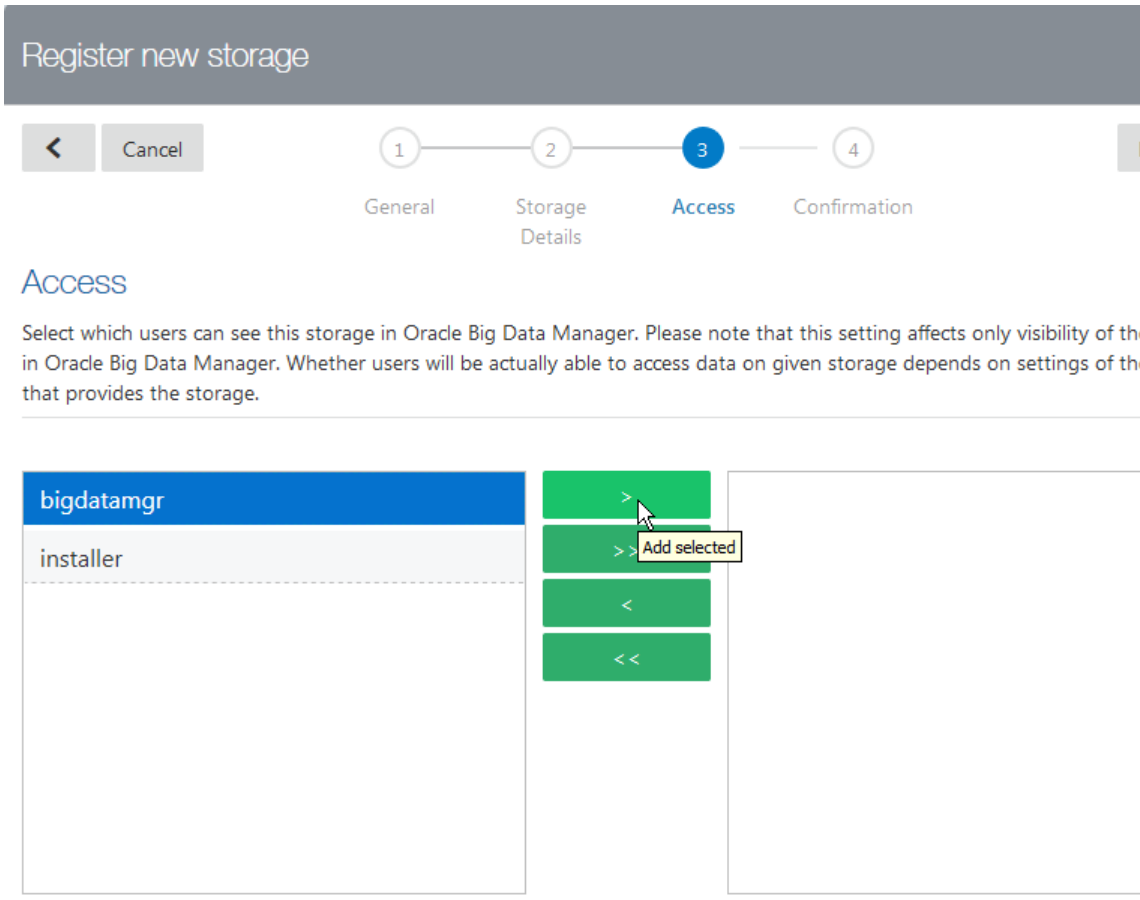
   Replace `host`, `port`, and `service_name` with the appropriate values.
3. Click **Test access to storage** to make sure that you can access the Oracle Database storage. If the storage details that you provided are correct, the **Successful, storage details are correct** message is displayed.

   If the connection is successful, the **Preview of storage content** section displays the schemas accessible to the user.

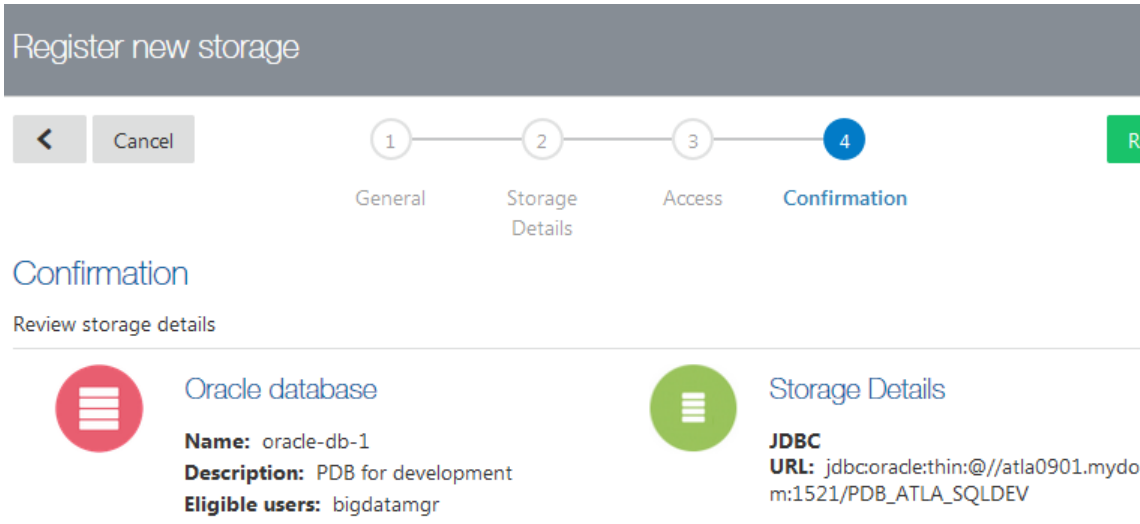


4. Click **Next** to go the **Access** page of the wizard.

   Build the list of Big Data Manager users that should have access. Each of these users will have access to the Oracle Database storage provider through the same JDBC connection.



5. Click **Next** to go to the last page of the wizard.
6. Check that the information you entered is correct and then click **Register**. The registered storage provider will be immediately available for selection by authorized users.



3
Viewing Data in Oracle Big Data Manager

You can view data sources, data, and data properties in the Oracle Big Data Manager console.

Topics:

3.1 Displaying and Navigating Storage Providers

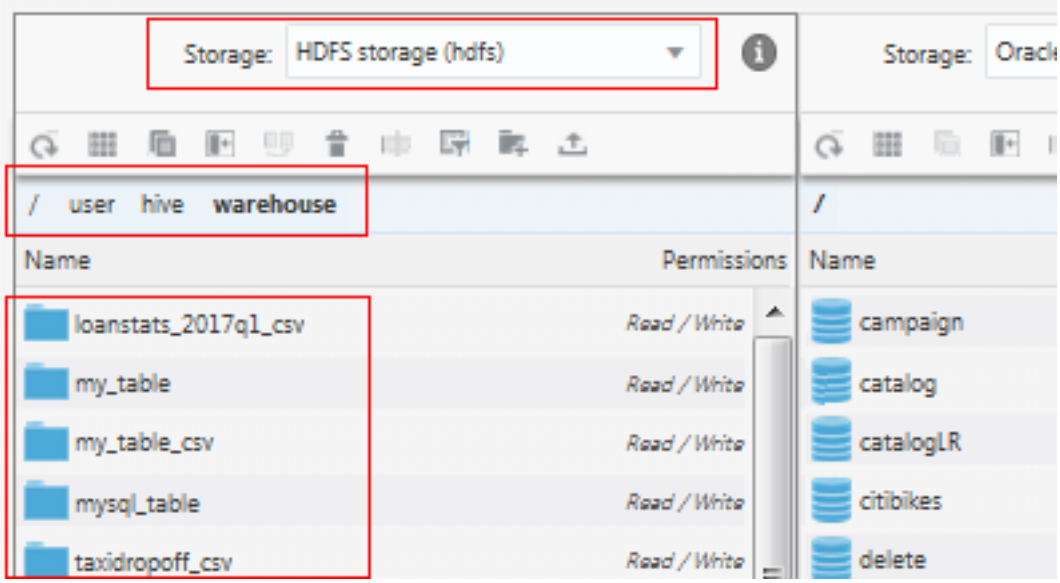
You can display and navigate through storage providers in the Data section of the Oracle Big Data Manager. console.

1. Click the **Data** tab on the top of the page.
2. If it isn't already selected, click the **Explorer** tab on the left side of the page.

   The **Explorer** page contains two panels, each of which shows a data source. (One way to copy data is to drag items from one panel to the other.)

   To display and navigate through a data provider:

   - Display a storage provider by clicking the **Storage** list at the top of the panel and selecting the storage provider.
   - Drill down by double-clicking items (folders, etc.) under **Name** in the panel.
   - Navigate back up the hierarchy by clicking on an item in the "breadcrumbs" below the toolbar, for example: **/ user hive warehouse**



3.2 Previewing Content from Github

In the **Data Explorer**, you can preview the contents of data in Github, including table data presented in a table viewer.

To preview content:

1. Click **Data** on the menu bar to go to the data Explorer.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. Right-click the item in Github whose content you want to preview and select the command to show the data. The menu command varies depending on the type of data; for example, **Show file content** or **Show table data**.

3.3 Viewing Data Properties

In the Oracle Big Data Manager console, you can view properties of data objects and containers.

To view the properties:

1. Click **Data** on the console menu bar to go to the Data explorer.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. Navigate to the object or container, right-click it, and select **Properties** from the menu.

Depending on what kind of item you examined, properties such as the following are dsplayed:

- Location
- Size
- Modified date
- Owner

- Roles
- Read/write permission

4
Transferring and Comparing Data

In the Oracle Big Data Manager console, you can create jobs to copy, move, and compare data. You can run the jobs once or repeatedly, on a set schedule. You can also upload files from your local machine and upload data into Hive.

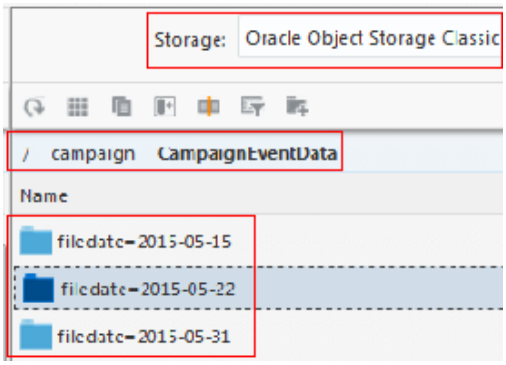
Topics:

4.1 Copying Data (Including Drag and Drop)

In the Oracle Big Data Manager console, you can copy data between storage providers by creating copy jobs.

To copy data from one storage to another,

1. Click **Data** on the console menu bar to go to the Data explorer.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. In one panel, select a destination data provider from the **Storage** list, and navigate to a folder or container by selecting a location in the breadcrumbs or by drilling down in the list below it, for example:



4. In the other panel, select a source data provider from the **Storage** drop-down list, navigate to the folder or container containing the file, folder, or container you want to copy.
5. Do any of the following:
   a. Drag the source file, folder, or container from the source and drop it on the target. If you drop a file from the source on a single file in the target, that file will be replaced by the one being copied. If your drop an item on a folder or container, it will be copied into the folder or container.
   b. Right-click the item you want to copy and select **Copy** from the menu. If a folder or container is selected in the target, the item will be copied into the folder or container. If a single item is selected in the targey, it will be replaced. If nothing is selected in the target, the item will be copied into the current folder or container.
   c. Click **Copy**  . If a folder or container is selected in the target, the item will be copied into the folder or container. If a single item is selected in the targer, it will be replaced. If nothing is selected in the target, the item will be copied into the current folder or container
6. In the **New copy data job** dialog box, choose or enter values as described below.

   General tab

   - **Job name:** A name is provided for the job, but you can append to it or replace it with a different name.
   - **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer — import from HTTP.**
   - **Run immediately:** Select this option to run the job immediately and only once.
   - **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job.

   Advanced tab

   - **Number of executors:** Select the number of executors from the drop-down list. The default number is 3. If you have more then three nodes you can increase execution speed by specifying a higher number of executors. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of executors to increase performance.
   - **Number of CPU cores per executor:** Select the number of cores from the drop-down list. The default number is 5. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of cores to increase performance.
   - **Memory allocated for each execution:** Select the amount of memory from the drop-down list. The default value is 40 GB. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the memory to increase performance.
   - **Memory allocated for driver:** Select the memory limit from the drop-down list.
   - **Custom logging level:** Select this option to log the job's activity and to select the logging level.
7. Click **Create**.

The **Data compare job *job_number* created** dialog box shows minimal status information about the job. Click the **View more details** link to show more details about the job in the **Jobs** section of console.

8. Review the job results. In particular, in the **Jobs** section of the console, click the **Comparison results** tab on the left side of the page to display what's the same and what's different about the compared items.

4.2 Copying Data (Including from Multiple Sources)

In the Oracle Big Data Manager console, you can create, schedule, and run job that includes multiple sources. You can also copy via HTTP(S).

1. Click **Data** on the menu bar to go to the **Data Explorer**.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. In either panel of the **Data Explorer**, select a target location as the destination for the copy job.
4. On the toolbar for that panel, click **Copy here from HTTP(S)**  .
5. In the **New copy data job** dialog box, enter information in the **Sources** row, as follows::
   a. From the first drop-down list, select **Direct link** to copy a single file or select **Link to list of files** to copy multiple files that are listed in a manifest file containing the list in comma-separated values (CSV) format.
   b. From the second drop-down list, select the data source from which you are copying. This list shows the data providers registered with Oracle Big Data Manager.
   c. The last control in the Sources row depends on the type of data source selected in the second drop-down list. For HTTP(S), enter the URL of the source in the **Enter a valid HTTP(S)** text box. For other types of data sources, click the **Select file** button to navigate to and select a file.
6. If you want to copy from multiple sources in the same copy job, click the **Add source** button and repeat the tasks in the previous step.
7. If you want to change the destination for the copy job, click in the **Destination** field and edit the current location.
8. In the tabs of the **New copy data job** dialog box, enter the following values.

   General tab

   - **Job name:** A name is provided for the job, but you can append to it or replace it with a different name.
   - **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer — import from HTTP.**
   - **Run immediately:** Select this option to run the job immediately and only once.
   - **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job.

   Advanced tab

   - **Number of executors:** Select the number of executors from the drop-down list. The default number is 3. If you have more then three nodes you can

increase execution speed by specifying a higher number of executors. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of executors to increase performance.

- **Number of CPU cores per executor:** Select the number of cores from the drop-down list. The default number is 5. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of cores to increase performance.
- **Memory allocated for each execution:** Select the amount of memory from the drop-down list. The default value is 40 GB. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the memory to increase performance.
- **Memory allocated for driver:** Select the memory limit from the drop-down list.
- **Custom logging level:** Select this option to log the job's activity and to select the logging level.
- **HTTP proxy:** If this data transfer type is HTTP(S) and if you have HTTP(S) header information stored in a file, you can use that header information in the HTTP(S) request header. From the **HTTP headers file** drop-down list, select the storage that contains the file. If it's via HTTP(S), enter the URI for the file in the **Enter a valid HTTP(S) URI** field. If it's a different kind of provider, click the **Select File** button and navigate to and choose the file.

9. Click **Create**.

   The **Data compare job *job_number* created** dialog box shows minimal status information about the job. Click the **View more details** link to show more details about the job in the **Jobs** section of console.

10. Review the job results. In particular, in the **Jobs** section of the console, click the **Comparison results** tab on the left side of the page to display what's the same and what's different about the compared items.

4.3 Uploading Files from a Local Computer

In the Oracle Big Data Manager console, you can upload files from a local computer to a registered data provider.

To upload files from a local computer:

1. Click **Data** on the menu bar to go to the Data explorer.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. In either of the Data explorer panels, select a destination for the files.

   Do this by first selecting the data provider from the **Select** list at the top of the panel and then by navigating (drilling down) to the location where you want to upload the files. The folder or container that you select will be used as the destination.

4. On the toolbar of the panel you chose above, click **Upload Files**  .
5. In the **Files Upload** dialog box, click **Choose files to upload** to select the files from your computer's file system. Alternatively, you can drag files from your computer's file system to the **Or drop files here** box.

You can upload multiple files at one time by using either or both of the above methods.

6. Click **Upload** to upload the selected files, and then click **Close** to close the dialog box.

4.4 Moving Data in HDFS

In the Oracle Big Data Manager console, you can move data from one HDFS location to another.

1. Click **Data** on the menu bar to go to the **Data Explorer**.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. In either panel of the **Data Explorer**, select a target location as the destination for the copy job.
4. From the **Storage** drop-down list in one of the panels, select **HDFS Storage** and navigate to the target location.
5. From the **Storage** drop-down list in the other panel, select **HDFS Storage**, navigate to the item you want to move, and select it.
6. On the toolbar for the panel containing the item to be moved, click **Move**  .
7. When prompted, click **Move**.

4.5 Copying Data Via HTTP

In the Oracle Big Data Manager console, you can create, schedule, and run jobs that copy data from a source on a web server by using the HTTP protocol.

1. Click **Data** on the menu bar to go to the **Data Explorer**.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).
3. On the toolbar, click **Copy here from HTTP(S)**  .
4. In the **New copy data job** dialog box, enter information in the **Sources** row, as follows:
   a. From the first drop-down list, select **Direct link** to copy a single file or select **Link to list of files** to copy multiple files that are listed in a manifest file containing the list in comma-separated values (CSV) format.
   b. From the second drop-down list, select **HTTP(S)** if it isn't already selected.
   c. In the last control on the **Sources** row, enter the URL of the source in the **Enter a valid HTTP(S) URL** box.
5. If you want to copy from multiple sources in the same copy job, click the **Add source** button and repeat the tasks in the previous step.
6. If you want to change the destination for the copy job, click in the **Destination** field and edit the current location.
7. In the tabs of the **New copy data job** dialog box, enter the following values.

   General tab

- **Job name:** A name is provided for the job, but you can append to it or replace it with a different name.
- **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer — import from HTTP.**
- **Run immediately:** Select this option to run the job immediately and only once.
- **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job.

Advanced tab

- **Number of executors:** Select the number of executors from the drop-down list. The default number is 3. If you have more then three nodes you can increase execution speed by specifying a higher number of executors. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of executors to increase performance.
- **Number of CPU cores per executor:** Select the number of cores from the drop-down list. The default number is 5. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of cores to increase performance.
- **Memory allocated for each execution:** Select the amount of memory from the drop-down list. The default value is 40 GB. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the memory to increase performance.
- **Memory allocated for driver:** Select the memory limit from the drop-down list.
- **Custom logging level:** Select this option to log the job's activity and to select the logging level.
- **HTTP proxy:** If you have HTTP(S) header information stored in a file, you can use that header information in the HTTP(S) request header. From the **HTTP headers file** drop-down list, select the storage that contains the file. If it's via HTTP(S), enter the URI for the file in the **Enter a valid HTTP(S) URI** field. If it's a different kind of provider, click the **Select File** button and navigate to and choose the file.

8. Click **Create**.

   The **Data copy job *job_number* created** dialog box shows minimal status information about the job. Click the **View more details** link to show more details about the job in the **Jobs** section of console.

9. Review the job results. In particular, in the **Jobs** section of the console, click the **Comparison results** tab on the left side of the page to display what's the same and what's different about the compared items.

4.6 Importing Data into Hive

In the Oracle Big Data Manager console, you can import `.csv` files, Apache Avro files, and Apache Parquet files from HDFS into HiveServer2.

To import one of the supported files:

1. Click **Data** on the console menu bar to go to the Data explorer.
2. If it isn't already selected, click the **Explorer** tab (on the left side of the page).

3. From the storage drop-down list in one of the panels, select **HDFS Storage**.

   Apache Hive import might not work, depending on the access rights of the file and its parent directories. If so, you can copy or move the file to the **/tmp** directory and import from there.

4. Navigate to the file you want to import, right-click it, select **Import into Hive**, and select how to import it: **Import as CSV**, **Import as Apache Avro**, or **Import as Apache Parquet**.

   When you import a `.csv` file, a table containing the data is shown as a preview.

4.7 Copying Data Between Oracle Database and Apache Hive

In the Oracle Big Data Manager console, you can copy data from Oracle Database to Apache Hive and also in the reverse direction – from Apache Hive to the database.

Oracle Big Data Manager data transfer between Oracle Database and Apache Hive requires the following:

- An Oracle Database registered as a storage provider in Oracle Big Data Manager. A database connector is not included with the installation on Oracle Big Data Appliance. You must create your own through the **Register New Storage** wizard described in this document.
- An Oracle Database account with access to the schema that you want to work with.
- An Oracle Big Data Manager account with rights to the same Oracle Database account.
- For data transfers from Apache Hive to Oracle Database, columns in the Hive table and columns in the target Oracle Database table must match. This is not a requirement for database-to-Hive data transfers.
- Licenses:
  - Copying data from Apache Hive to Oracle Database requires an Oracle Loader for Hadoop license.
  - Copying data from Oracle Database to Apache Hive requires a license to run Copy to Hadoop.

> **Note:**
>
> These licensing requirements are specific to use of the tool on Oracle Big Data Appliance and may not apply to Oracle Big Data Manager in Oracle cloud-based product offerings.

> **See Also:**
>
> Registering Storage Providers with Oracle Big Data Manager
>
> Adding Oracle Big Data Manager Users

4.8 Comparing Data Sets

In the Oracle Big Data Manager console, you can create, schedule, and run jobs that compare large data sets in different storage providers.

A compare job uses the `odiff` utility on Oracle Big Data Appliance, and the computation runs as distributed Spark application.

1. Click **Data** on the menu bar to open the **Data Explorer** .
2. Click the **Explorer** tab (on the left side of the page).
3. Select an item in the left panel and an item in the right panel to compare. You can only compare like items, for example file to file or directory to directory.
4. On the toolbar, click **Compare**  .
5. In the **New compare data job** dialog box, enter the following values.

   General tab

   - **Job name:** A name is provided for the job, but you can append to it or replace it with a different name.
   - **Job type:** This read-only field describes the type of job. In this case, it's **Oracle Distributed Diff — compare.**
   - **Run immediately:** Select this option to run the job immediately and only once.
   - **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job.

   Advanced tab

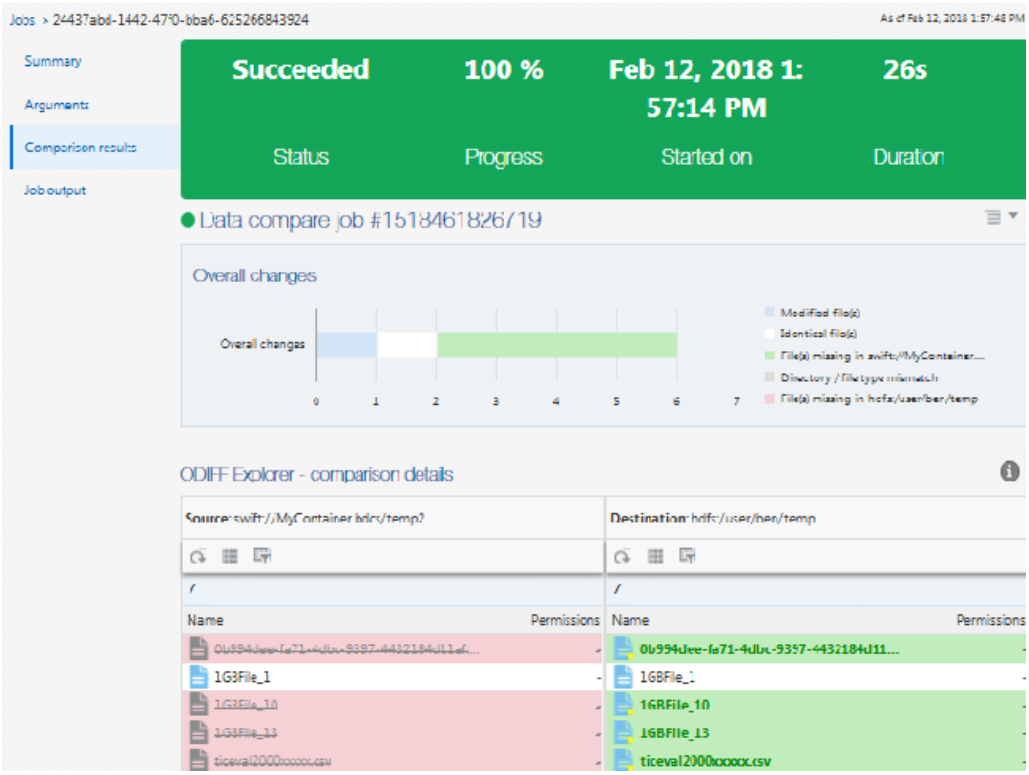
   - **Number of executors:** Select the number of executors from the drop-down list. The default number is `3`. If you have more then three nodes you can increase execution speed by specifying a higher number of executors. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of executors to increase performance.
   - **Number of CPU cores per executor:** Select the number of cores from the drop-down list. The default number is 5. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the number of cores to increase performance.
   - **Memory allocated for each execution:** Select the amount of memory from the drop-down list. The default value is 40 GB. If you want to execute this job in parallel with other Spark or MapReduce jobs, decrease the memory to increase performance.
   - **Memory allocated for driver:** Select the memory limit from the drop-down list.

- **Custom logging level:** Select this option to log the job's activity and to select the logging level.

6. Click **Create**.

   The **Data compare job *job_number* created** dialog box shows minimal status information about the job. Click the **View more details** link to show more details about the job in the **Jobs** section of console.

7. Review the job results. In particular, in the **Jobs** section of the console, click the **Comparison results** tab on the left side of the page to display what's the same and what's different about the compared items.



5

Managing Jobs in Big Data Manager

Copying and comparing data in Oracle Big Data Manager is handled by creating jobs.

Topics

5.1 Viewing Execution History of All Jobs

You can view the execution history of all jobs on the **Jobs** area of the Oracle Big Data Manager console.

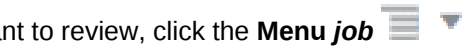
To view the execution history of all job:

1. Click **Jobs** on the menu bar of the console.
2. Click **History** on the left side of the page.

5.2 Viewing Summary Information About a Job

You can view summary information about running and completed jobs in the Jobs section of the Oracle Big Data Manager console.

To view summary information about a job:

1. Click **Jobs** in the Oracle Big Data Manager console menu bar.
2. If it's not already selected, click **History** on the left side of the page.
3. In the row for the job you want to review, click the **Menu *job***  icon, and then select **View Details**.

The information shown can include the following:

- Name
- Source and destination
- Description
- Schedule
- Status
- Progress
- Start and stop times
- Duration

5.3 Viewing the Arguments for a Job

You can view the parameters that were passed to a job in the Jobs section of the Oracle Big Data Manager console.

To view the arguments for a job:

1. Click **Jobs** in the Oracle Big Data Manager console menu bar.
2. If it's not already selected, click **History** on the left side of the page.
3. In the row for the job you want to review, click the **Menu *job***  icon, and then select **View Details**.

The arguments defined for the job are shown, for example number of executors, block size, etc.

6
Analyzing Data Interactively With Notes

Use *notes* to explore and visualize data iteratively.

Oracle Big Data Manager uses Apache Zeppelin as its notebook interface and coding environment. The following topics tell how to do some of the most common tasks with notes in Oracle Big Data Manager. For complete documentation, see Apache Zeppelin. (Not all Apache Zeppelin features are supported in Oracle Big Data Manager.)

Topics:

6.1 Working with Notes

Import, create, and run notes in the Notebook section of the Oracle Big Data Manager console.
The Notebook Home page lists your existing notes, along with controls for importing and creating new notes. When you open a note, it's displayed in its own Note page as a collection of *paragraphs* that contain snippets of code for accessing services, running jobs, and displaying results. You can define and run the code quickly and iteratively, which provides flexibility for analyzing and visualizing your data.

Commands for performing actions on the entire note are on the toolbar at the top of each Note page.

Commands for performing actions on individual paragraphs are on the toolbar on the right side of each paragraph on the Note page. Paragraphs contains a *code* section, where you enter your source code, and an *output* section, which displays the output from executing that code.

6.1.1 Using the Commands on the Note Toolbar

Use the toolbar at the top of the Note page to perform actions on the entire note.

Item	Action
Run all paragraphs	Executes all the paragraphs in the note sequentially, in the order they're displayed.
Show/hide the code	Shows or hides the code sections of all paragraphs in the note.
Show/hide the output	Shows or hides the output sections of all paragraphs in the note.
Clear output	Clears the output sections of all paragraphs in the note.
Clone note	Makes a copy of the note.
Export this note	Exports the code and output sections of all the paragraphs in the note to a JSON file in your web browser's default download directory. If the output sections are very long, consider clearing the output before exporting the note, to save space.
Version control	Commits the content of the note to the current repository. When you click this button, you're prompted for a commit message. The message you enter here is displayed in the **Head** list, described below.
Head (revision) drop-down list	Displays a list of previously committed revisions of the note, if any. By default, the head revision is selected. If you want to view a previous revision, select it from the list. Click **Set revision** to set the head to the current revision.
Move note to trash	Deletes the note.
Run scheduler	Schedule the execution of all paragraphs in the note with a `cron` scheduler. When you select this option, a pop-up window displays the following options:: • **Preset** —A list of preset intervals. If one of the presets is adequate for your needs, click the link for the interval. It's added as an expression to the **cron expression** field. Options are **None 1m**, **5m**, **1h**, **3h**, **6h**, **12h**, **1d**. Select **None** to remove any expressions that were added. • **Cron expression**—Enter a custom `cron` expression, if you need something other than the above presets. • **Cron executing user**— Enter the name of the user for running the cron job, if other than `root`. • **Auto-restart interpreter on cron execution** — Terminates the interpreter after the `cron` job completes running. The interpreter restarts automatically the next time a paragraph is run.

6.1.2 Using the Commands on the Paragraph Toolbar

Use the toolbar on the right side of a paragraph panel to perform actions on that paragraph only.

Item	Action
Status	Shows the status of the paragraph. It can be one of the following: • READY • FINISHED • ABORT • ERROR • PENDING • RUNNING
Run this paragaph (Shift +Enter)	Executes the code in the code section of the paragraph.
Show/hide editor (Control +Option+E)	Shows or hides the code section of the paragraph.
Show/hide output (Control +Option+O)	Shows or hides the output section of the paragraph.
Menu	Opens a menu with the following options that apply to the current paragraph: • ***nnnnnnnnn-nnnnnn_nnnnnnnnnn***—The paragraph ID. Click the ID to copy it to the clipboard. • **Width**—Select a number from the drop-down list to set a width for the paragraph on a grid of 12 units. This allows you to organize the paragraphs in the grid. • **Move down**—Move the paragraph one level down. • **Insert new**—Insert a new paragraph below the current one. • **Clone paragraph**—Create and show a copy of the current paragraph. • **Show/Hide title**—Show or hide the title of the paragraph. You can edit the title when it's shown. • **Show/Hide line numbers**—Show or hide line numbers in the code section of the paragraph. • **Disable run**—Disable the **Run** button for this paragraph. • **Link this paragraph**—Export the paragraph as an **iframe** and open the **iframe** in a new window. • **Clear output**—Clear the output section for this paragraph. • **Remove**—Delete the paragraph.

6.2 Importing a Note

You can import a note in the Notebook section of the Oracle Big Data Manager console.

To import a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. On the Notebook Home page, click the **Import note**  link.
3. In the Import new note dialog box, do the following:
   a. Leave the **Import As** field blank to keep the original name of the note, or enter a new name to replace the original name.
   b. Click **Choose a JSON here** to upload a file from your local computer, or click **Add from URL** to upload from a location on the internet.

6.3 Exporting a Note

You can export a note from the Notebook section in the Oracle Big Data Manager console.

To export a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner at the top of the page.
2. On the Notebook home page, click the name of the note you want to export. The note is opened.
3. On the toolbar next to the note's title, click the **Export this note**  icon. The note is exported to a JSON file in your web browser's default download directory. The exported note has the same name as the original note.

6.4 Creating a Note

You can create a note in the Notebook section in the Oracle Big Data Manager console.

To create a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. Click the **Create new note** link at the head of the list of notes, or click the **Notebook** drop-down list and select **Create new note**.
3. In the Create new note dialog box, enter a name in the **Note Name** field. If you want to save the note to a different location, you can specify a path to a folder. If the folder doesn't exist, Oracle Big Data Manager will create it.

   For example, to create a note named `my_note` in a new or existing directory named `my_notes_dir`, enter the following in the **Note Name** field:

   ```
   my_notes_dir/my_note
   ```
4. Select an interpreter from the **Default Interpreter** drop-down list. The available choices are **spark**, **md**, **sh**, **python**, **jdbc**, and **mysql**.

5. Click **Create Note**. The note is displayed with an empty paragraph. Each note is composed of one or more paragraphs.

6.5 Renaming a Note

You can rename a note in the Notebook section of the Oracle Big Data Manager console.

Renaming a Note Without Displaying the Note

To rename a note without displaying the note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon at the top of the page to display the Notebook home page.
2. In the list of notes on the home page, hover the mouse pointer over the note you want to rename, and then click the **Rename note**  icon.
3. In the **Rename note** dialog box, enter the new name for the note, and then click **Rename**. If you want to save the note to a different location, you can specify a path to a folder. If the folder doesn't exist, Oracle Big Data Manager will create it.

   For example, to rename a note named `my_note` to `my_note_001` and move it to a directory named `project_notes`, enter the following in the **Note Name** field:

   ```
   project_notes/my_note_001
   ```

Renaming a Note That's Currently Displayed

To rename a note that's currently displayed:

1. Click the name of the note under the banner at the top of the page, and edit as needed. If you want to save the note to a different location, you can specify a path to a folder, as described above.
2. Click anywhere in the note or press the **Enter** key to accept the changes

6.6 Clearing the Output from Paragraphs in a Note

You can clear the output from a note that's been run in the Notebook section of the Oracle Big Data Manager console

When you run the paragraphs in a note, the results are displayed beneath the code in each paragraph. To clear that output from all the paragraphs in a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. In the Notebook Home page, hover over the note for which you want to clear the output, click **Clear output,**t  and then click **OK.**
3. Re-open the note and confirm that the output is cleared from the result section of all the paragraphs.

6.7 Deleting a Note

You can delete a note ih the Notebook section of the Oracle Big Data Manager console.

To delete a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. On the Notebook Home page, hover over the note that you want to delete, click the **Move note to Trash** icon, and then click **OK.**

6.8 Viewing and Editing a Note

You can view and edit a note in the Notebook section of the Oracle Big Data Manager console.

To view and edit a note:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. The Notebook Home page lists all existing notes, Select the note from that list. Alternatively, from anywhere in the Notebook section, click the **Notebook** drop-down list from the banner at the top of the page, and then select the name of the note.
3. Edit the note as desired. You can modify, add, remove, and run paragraphs. You can also perform other actions on the note and its paragraphs by using the Note and Paragraph toolbars on the page. When you make changes to a note or a paragraph, the changes are automatically saved. See Working with Notes.

6.9 Running a Note

You can run a note in the Notebook section of the Oracle Big Data Manager console. You can run an entire note or individual paragraphs in the note.

To run the note:

1. If you're not already in the Notebook section of the console, click the Notebook tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon to display the Notebook Home page.
2. On rhe Notebook home page, click the name of the note you want to run. The note is opened,
3. Click the **Run all paragraphs** ▷ icon in the toolbar at the top of the page to execute all the paragraphs in the note sequentially, in the order they're displayed. If a paragraph contains code in the code section, the output of the code section is displayed beneath it.

To run an individual paragraph, click the ▷ icon in the toolbar for the paragraph.

6.10 Organizing Notes

You can organize notes into directories in the Notebook section in the Oracle Big Data Manager console.

You give a name to a note when you create it, and you can change the name of an existing note. See Creating a Note and Renaming a Note.

To specify that the note should be contained in a directory, add a qualifying path to the name. For example, to put a note named `note1` into the `Demo` directory, specify its name as `Demo/note1`. To move that note to the `Test` directory, rename it as `Test/note1`. If the directory doesn't exist, Oracle Big Data Manager creates it.

6.11 Managing Notebook Interpreters Settings

You can configure interpreters for running notes in the Notebook section of the Oracle Big Data Manager console.

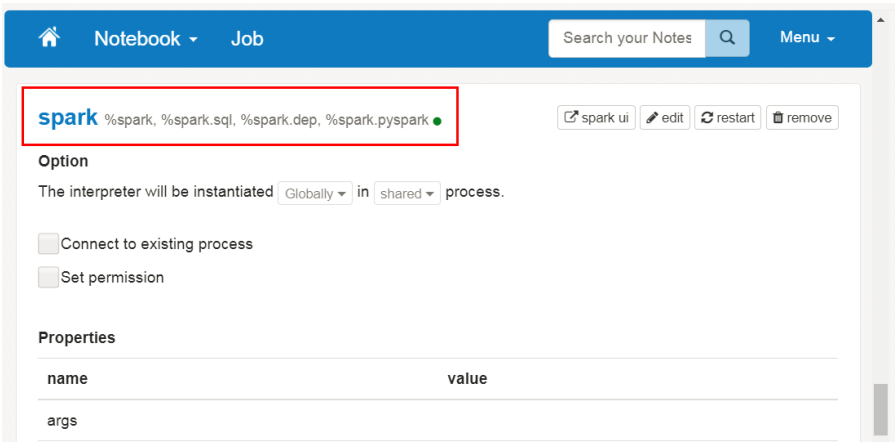
Interpreters are bindings for how code should be interpreted and where it should be submitted for execution. The Zeppelin interpreter allows any language and data processing back end to be plugged into Zeppelin. Oracle Big Data Manager supports the following interpreters:

- JDBC
- Markdown language (md)
- MySQL
- Python
- Unix shell (sh)
- Spark

To configure interpreters:

1. If you're not already in the Notebook section of the console, click the **Notebook** tab at the top of the page. If you're already in the Notebook section, click the **Home**  icon in the banner near the top of the page.
2. On the Notebook Home page, click the **Menu** drop-down list, and then select **Interpreters.**
3. Use the Interpreters page to manage the available interpreters' settings. You can create, edit, and remove settings. You can also restart interpreters.

   Every Interpreter belongs to a single interpreter group; however, an interpreter group can contain several interpreters. For example, the Spark interpreter group includes the highlighted interpreters in the following image:


Notebook
Job
Search your Notes
Menu
spark %spark, %spark.sql, %spark.dep, %spark.pyspark
spark ui
edit
restart
remove
Option
The interpreter will be instantiated Globally in shared process.
Connect to existing process
Set permission
Properties
name
value
args

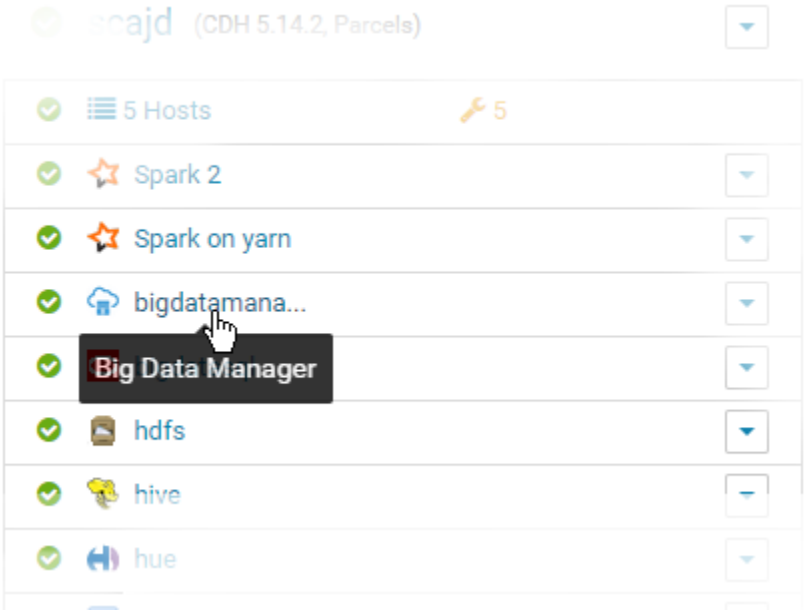
7

Using Cloudera Manager to Work With Oracle Big Data Manager

Oracle Big Data Manager is automatically added as a service Cloudera Manager.

As with other services in Cloudera Manager, you can use the interface to monitor, stop, start, and change the configuration of Oracle Big Data Manager.

1. Log on to Cloudera Manager.
2. On the Home page, find **bigdatamanager** in the list of services.



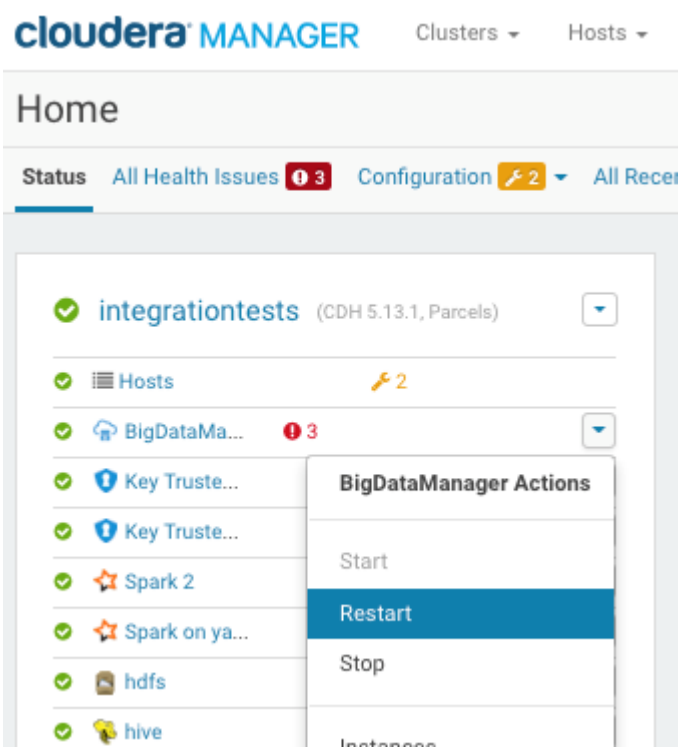
Oracle Big Data Manager Roles

Cloudera Manager supports four Oracle Big Data Manager roles:

- BigDataManager
- Big Data Manager Notebook
- Big Data Manager Proxy
- Hosts

Stopping and Starting Oracle Big Data Manager

On the Cloudera Manager Home page, you can you can stop, start (or restart) the Oracle BigDataManager service from the Actions pulldown menu.

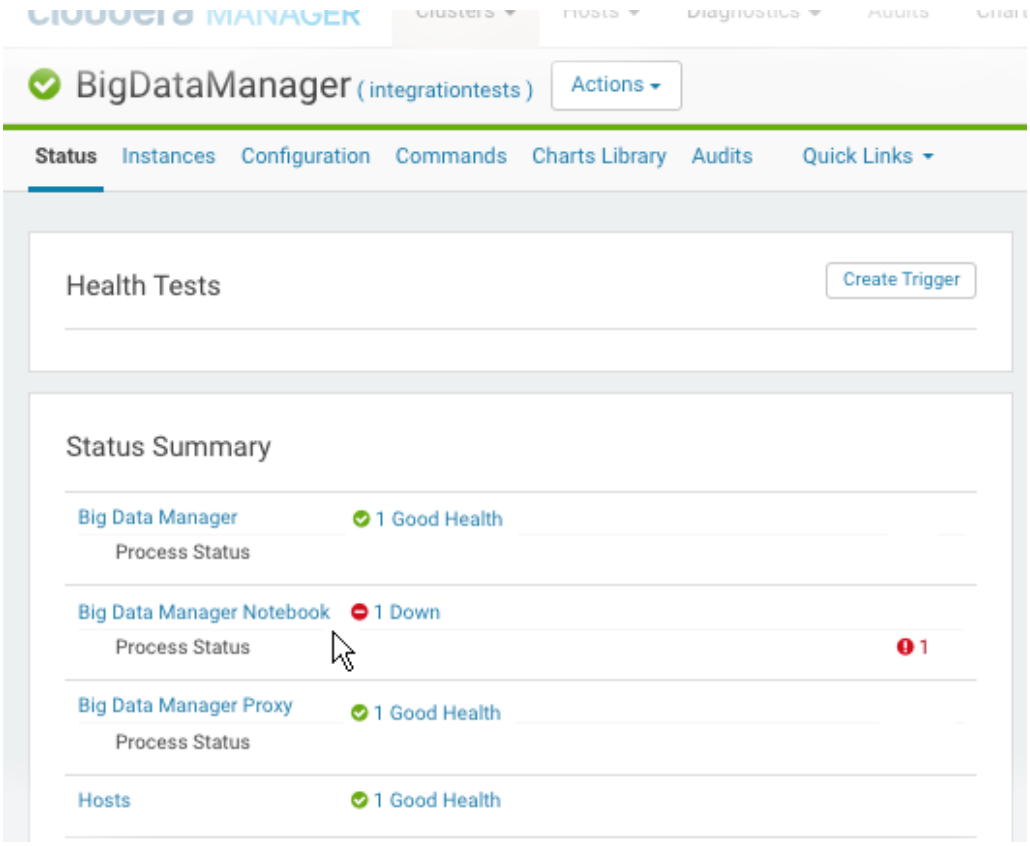


Tip:

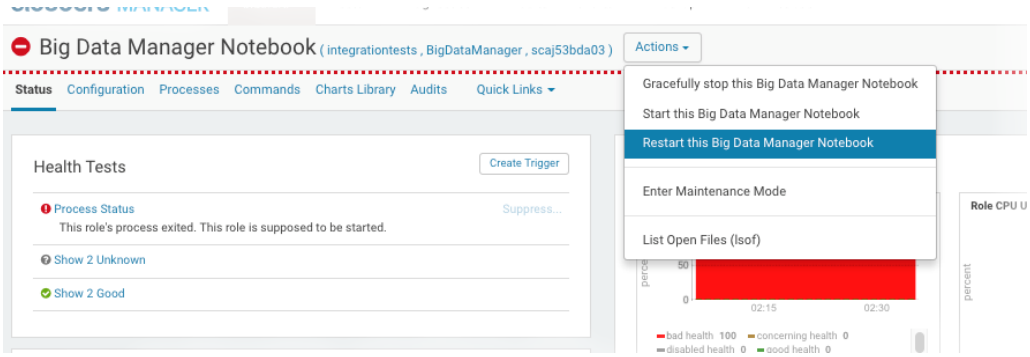
If you stop or start BigDataManager, the same action is applied to all Big Data Manager roles. If there are problems that require a restart, you may only need to restart one role.

Running Health Tests on BigDataManager Roles

You can test the health of individual roles within the BigDataManager service. In the example below, the Big Data Manager Notebook service is down.



The **Actions** pulldown menu provides the actions for BigDataManager roles that are shown in the screen below.



Tip:

In the case of the Big Data Manager Notebook role, the restart through Configuration Manager is equivalent to the following shell commands, which would need to be executed on the node where Confiiguration Manger is hosted.

```
sudo su -
/etc/init.d/bdm-notebook restart
```

8

Managing Data and Copy Jobs With the Oracle Big Data Manager SDKs

You can use the Oracle Big Data Manager SDKs to manage data and copy jobs from applications.

The Oracle Big Data Manager SDKs are available from GitHub:

- Python SDK
- Java SDK

9 Using the Oracle Big Data Manager bdm-cli Utility

Use the `bdm-cli` (Oracle Big Data Manager Command Line Interface) utility to copy data and manage copy jobs at the command line.

`bdm-cli` has several commands that duplicate `odcp` commands , but `bdm-cli` also includes additional commands for scheduling and managing copy jobs and other administrative tasks.

You have to download and install `bdm-cli` yourself, either on a node of the cluster or on a remote operating system. If you install it on your cluster, you must use SSH to connect to the cluster. If you install it on a remote system, you can run the commands without SSH. See Installing the bdm-cli Utility.

There are no special requirements for using `bdm-cli` when it's installed outside the cluster.

9.1 Installing the bdm-cli Utility

The `bdm-cli` (Big Data Command Line Interface) is a command line utility for copying data and managing copy jobs. You can download and install `bdm-cli` from GitHub. You can install it on a remote operating system, so you don't have to use SSH to connect to the cluster.

To install `bdm-cli`:

1. If you use a proxy server, first call:

   ```
   export http_proxy="your_proxy_server"
   export https_proxy="your_proxy_server"
   ```

2. Then call:

   ```
   curl -L https://github.com/jazeman/bdm-python-cli/blob/1.0/install-rpm?
   raw=true | bash
   ```

9.2 Usage

You can use `bdm-cli` at the command line to create and manage copy jobs.

Syntax

`bdm-cli [global_options] subcommand [options][arguments]...`

Supported Storage Protocols and Paths

The protocols and paths to the file systems and storage services supported by `bdm-cli` are:

- HDFS:

```
hdfs:///
```

- Oracle Cloud Infrastructure Object Storage Classic (formerly known as Oracle Storage Cloud Service):

  ```
  swift://container.provider/
  ```

- Oracle Cloud Infrastructure Object Storage (formerly known as Oracle Bare Metal Cloud Object Storage Service):

  ```
  oss:///container
  ```

  For operations with Oracle Cloud Infrastructure Object Storage, you must specify the provider by using the options `src-provider` and `dst-provider`. For example, those options are used with bdm-cli create_job when used with Oracle Cloud Infrastructure Object Storage.

Finding a Job's UUID

A number of bdm-cli subcommands require that you identify a job by its Universally Unique Identifier (UUID). To find UUIDs, execute bdm-cli list_all_jobs.

Specifying Source and Destination Paths

When specifying sources and destinations, fully qualify the paths:

- *source* ...

  File name qualified by protocol and full path, for example: `hdfs:///user/oracle/test.raw`

- *destination*

  Directory name qualified by protocol and full path, for example: `swift://container.storagename/test-dir`

Setting Environment Variables

You can set some `bdm-cli` options as environment variables. For example, you can set Oracle Big Data Manager URL and user password file, as follows:

```
export BDM_URL=https://hostname:8888/bdcs/api && export BDM_PASSWORD=/tmp/password_file
```

All the `bdm-cli` options that can be set as environment variables are documented in the sections below.

Getting Help

To get help for `bdm-cli` use:

```
bdm-cli --help
```

To get help for a specific command use:

```
bdm-cli command --help
```

For example:

```
bdm-cli edit_job_template --help
```

9.3 Options

Options that can be used by all bdm-cli commands are explained below.

Option	Description
`--bdm-passwd` *`path_to_password_file`*	Path to the Oracle Big Data Manager user password file. Environment variable: `BDM_PASSWORD`
`--bdm-url` *`bdm_url`*	Oracle Big Data Manager server URL. Environment variable: `BDM_URL`
`--bdm-username` *`username`*	Oracle Big Data Manager server user name. Default value: `oracle` Environment variable: `BDM_USERNAME`
`-f [table|csv|json]`	Specify the output format: • `table` (default) Each field is displayed in a separate column. • csv Each record is displayed as a comma-separated list on a single line. • `json`: The output is displayed in JavaScript Object Notation (JSON) format.
`--fields` *`fields`*	Specifies comma-separated fields depending on the type of object.
`-h` `--help`	Show this message and exit.
`--no-check-certificate`	Don't validate the server's certificate.
`--proxy` *`proxy`*	Proxy server.
`--tenant-name` *`tenant_name`*	Name of the tenant. Default value: `admin`
`-v`	Print the REST request body.
`--version`	Show the Oracle Big Data Manager version and exit.

9.4 Subcommands

The following table summarizes the bdm-cli subcommands. For more details on each, click the name of the command.

Command	Description
bdm-cli abort_job	Abort a running job.
bdm-cli copy	Execute a job to copy sources to destination.
bdm-cli create_job	Execute a new job from an existing template.
bdm-cli create_job_template	Create a new job template.
bdm-cli get_data_source	Find a data source by name.

Command	Description
bdm-cli get_job	Get a job by UUID.
bdm-cli get_job_log	Get a job log.
bdm-cli list_all_jobs	List all jobs from the execution history.
bdm-cli list_template_executions	List all jobs from the execution history for the given template.
bdm-cli ls	List files from a specific location.

9.5 bdm-cli abort_job

Abort a running job.

Syntax

```
bdm-cli abort_job [options] job_uuid
```

Options

Option	Description
`--force`	Force abort job.
`-h` `--help`	Show this message and exit.

Example

Abort a job.

```
/usr/bin/bdm-cli -f json --no-check-certificate --bdm-url ${DATA_HOST}:8888/bdcs/api
--bdm-username ${DATA_USER} --bdm-passwd ${USER_PASSWORD_FILE} abort_job
24ef30e8-913b-4402-baf8-74b99c211f50
```

9.6 bdm-cli copy

Execute a job to copy sources to destination.

Syntax

```
bdm-cli copy [options] source... destination
```

Options

Option	Description
`block-size` *block_size*	Specify the block size in bytes.
`--description` *description*	Data source description.
`--driver-memory-size` *driver_memory_size*	Specify the maximum amount of memory for the Oracle Storage Cloud Service driver.

Option	Description
`--dst-provider` *oss_destination_provider*	Specify the provider of the destination, when using Oracle Cloud Infrastructure Object Storage Classic destination.
`-h` `--help`	Show this message and exit.
`--memory-size-per-node` *memory_size_per_node*	Specify the Spark executors memory limit in GB per node, for example, `40GB`.
`--number-of-executor-nodes`*number_of_executors_per_node*	Specify the maximum number of Spark executors per node, for example, `10GB`.
`--number-of-threads–per-node`*number_of_threads_per_node*	Specify the maximum number of threads per node.
`part-size` *part_size*	Specify the part size in bytes.
`--recursive` `--no-recursive`	Recursively copy (enabled by default).
`--retry` `--no-retry`	Retry data transfer in case of failure.
`--src-provider` *oss_source_provider*	Specify the provider of the source, when using for Oracle Cloud Infrastructure Object Storage Classic.
`--sync` `--no-sync`	Synchronize the source with the destination.

Example

Copy a file from HDFS to Oracle Storage Cloud Service:

```
/usr/bin/bdm-cli  -f json  --no-check-certificate  --bdm-url ${DATA_HOST}:8888/
bdcs/api --bdm-username ${DATA_USER} --bdm-passwd ${USER_PASSWORD_FILE}   copy
hdfs:///user/${DATA_USER}/1MFile.raw oss:///${DATA_USER} --dst-provider $
{OSS_PROVIDER}
```

9.7 bdm-cli create_job

Execute a new job from an existing template.

Syntax

`bdm-cli create_job [`*options*`] `*job_template_name*

Options

Option	Description
`--run-now`	Execute job immediately if job scheduling is set. Ignored otherwise.
`--source` *source*	Source file, for example: `hdfs:///user/oracle/test.raw`

Option	Description
`--destination destination`	The destination directory, for example: `swift://container.storagename/test-dir`.
`--driver-memory-size driver_memory_size`	Specify the maximum amount of memory for an Oracle Storage Cloud Service driver.
`--memory-size-per-node memory_size_per_node`	Specify the Spark executors memory limit in GB per node, for example: `40G`.
`--number-of-executor-nodes number_of_executors_per_node`	Specify the maximum number of Spark executors per node, for example: `10G`.
`--number-of-threads-per-node number_of_threads_per_node`	Specify the maximum number of threads per node.
`--block-size block_size`	Specify the block size in bytes.
`--part-size part_size`	Specify the part size in bytes.
`--retry` `--no-retry`	Retry data transfer in case of failure.
`--sync` `--no-sync`	Synchronize the source with the destination.
`--recursive` `--no-recursive`	Recursively copy (enabled by default).
`--job-executable-class job_executable_class`	Main Java class used for the Spark job execution.
`--src-provider oss_source_provider`	Specify the provider of the source, when using an Oracle Cloud Infrastructure Object Storage Classic source.
`--dst-provider oss_destination_provider`	Specify the provider of the destination, when using an Oracle Cloud Infrastructure Object Storage Classic destination.
`-h` `--help`	Show this message and exit.

9.8 bdm-cli create_job_template

Create a new job template.

Syntax

```
bdm-cli create_job_template [options] job_template_name source ... destination
```

Options

Option	Description
`--abort-running-job` `--no-abort-running-job`	Abort an already running execution if the next scheduled execution is started.
`--block-size block_size`	Specify block size in bytes.
`--data-source-name data_source_name`	Job's data source name.

Option	Description
`--description` *description*	Job template description.
`--dst-provider` *destination_provider*	Specify for `oss:///` destination.
`--environment` *environment*	Environment in JSON format: `{"envName1": "envValue2", "envName2": "envValue2"}`
`-h` `--help`	Show this message and exit.
`--history-size` *history_size*	Count of executions history log.
`--job-executable-class` *job_executable_class*	Main Java class used for the Spark job execution.
`--job-schedule` *job_schedule*	Specify cron-like job schedule, for example: `"0 56 8 * * ?"` means run every day at 08h 56m UTC time.
`--job-template-type` *job_template_type*	Specify job template type. Allowed values are: • `DATA_MOVEMENT_COPY` • `GENERAL`
`--libraries` *libraries*	Hadoop libraries, for example: `OdcpLibraries`. This option can have multiple values, for example: `--libraries OdcpLibraries --libraries OdcpLibraries`
`--memory-size-per-node` *memory_size_per_node*	Specify the Spark executors memory limit in GB per node, for example: `40G`.
`--number-of-executor-nodes` *number_of_executor_per_node*	Specify the maximum number of Spark executors per node, for example: `10G`.
`--number-of-threads-per-node` *number_of_threads_per_node*	Specify the maximum of threads per node.
`--part-size` *part_size*	Specify part size in bytes.
`--recursive` `--no-recursive`	Recursively copy (enabled by default).
`--retry` `--no-retry`	Retry data transfer in case of failure.
`--src-provider` *oss_source_provider*	Specify the provider of the source, when using for Oracle Bare Metal Cloud Object Storage Service.
`--sync` `--no-sync`	Synchronize source with destination.
`--tags` *tags*	User defined tag. This option can have multiple values, for example: `--tags system --tags datamovement --tags copy`

9.9 bdm-cli get_data_source

Find a data source by name.

Syntax

```
bdm-cli get_data_source [options] data_source_name
```

Options

Option	Description
`-h` `--help`	Show this message and exit.

9.10 bdm-cli get_job

Get a job by UUID.

Syntax

```
bdm-cli get_job [options] job_uuid
```

Options

Option	Description
`-h` `--help`	Show this message and exit.

Example

Get information on a job.

```
/usr/bin/bdm-cli  -f json  --no-check-certificate  --bdm-url ${DATA_HOST}:8888/
bdcs/api --bdm-username ${DATA_USER} --bdm-passwd ${USER_PASSWORD_FILE}
get_job ${JOB_UUID}
```

9.11 bdm-cli get_job_log

Get a job log.

Syntax

```
bdm-cli get_job_log [options] job_uuid
```

Options

Option	Description
`-h` `--help`	Show this message and exit.

9.12 bdm-cli list_all_jobs

List all jobs from the execution history.

Syntax

```
bdm-cli list_all_jobs [options]
```

Options

Option	Description
`-h` `--help`	Show this message and exit.
`--limit limit`	Specify the size of the page.
`--offset offset`	Specify the paging offset.

Example

List all jobs.

```
/usr/bin/bdm-cli  -f json  --no-check-certificate  --bdm-url ${DATA_HOST}:8888/bdcs/
api --bdm-username ${DATA_USER} --bdm-passwd ${USER_PASSWORD_FILE}   list_all_jobs
```

Use the `--offset` and `--limit` options to restrict the results. For example to get the eighth page when there are 20 rows per page, do the following:

```
bdm-cli list_all_jobs --offset 8 --limit 20
```

9.13 bdm-cli list_template_executions

List all jobs from the execution history for the given template.

Syntax

```
bdm-cli list_template_executions  [options] job_uuid
```

Options

Option	Description
`-h` `--help`	Show this message and exit.

9.14 bdm-cli ls

List files from a specific location.

Syntax

```
bdm-cli ls [options] path_1 ... path_n
```

Options

Option	Description
`-h` `--human-readable`	Human readable file sizes.
`-d` `--dirs-only`	List directories only.
`--provider` *`oss_provider`*	Specify for Oracle Bare Metal Cloud Object Storage Service paths.
`-h` `--help`	Show this message and exit.

Examples

List HDFS content under selected user.

```
/usr/bin/bdm-cli  -f json  --no-check-certificate  --bdm-url ${DATA_HOST}:8888/bdcs/
api --bdm-username ${DATA_USER} --bdm-passwd ${USER_PASSWORD_FILE}   ls hdfs:///
user/${DATA_USER}/integration_in --provider hdfs
```

List Oracle Cloud Infrastructure Object Storage Classic content under selected user.

```
/usr/bin/bdm-cli  -f json  --no-check-certificate  --bdm-url ${DATA_HOST}:8888/bdcs/
api --bdm-username test20170324113533 --bdm-passwd ${USER_PASSWORD_FILE}    ls
oss:///${OSS_CONTAINER}/ --provider ${OSS_PROVIDER}
```

A

Keyboard Shortcuts for Oracle Big Data Manager

You can use the keystroke shortcuts to perform actions in the Oracle Big Data Manager console, as described below.

Table A-1 Keyboard Shortcuts in the Big Data Manager Console

Task	Keyboard Shortcut
Change the currently selected item	Up/Down/Left/Right Arrow
Open the selected directory/container	Enter
Navigate back to parent directory/container	Backspace
Select the first item in list	Home or PageUp
Select the last item in list	End or PageDown
Switch between left and right panel in the Data Explorer	Tab
Deselect the currently selected item	Esc
Open the **Rename** dialog (supported only on HDFS)	F2
Reload the content of the current panel (same as the Refresh button)	F5 or Ctrl+R
Invokes copy/move/paste actions	Ctrl+C/X/V

If you're using a Mac, use the Command key instead of the Control (Ctrl) key.