

# Using Oracle Big Data Manager



with Oracle Big Data Appliance

F39028-01

February 2021

The Oracle logo, consisting of a solid red square with the word "ORACLE" in white, uppercase, sans-serif font centered within it.

ORACLE®

Using Oracle Big Data Manager, with Oracle Big Data Appliance

F39028-01

Copyright © 2018, 2021, Oracle and/or its affiliates.

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 embedded, installed or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software" or "commercial computer software documentation" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. 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 Inside 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, Epyc, and the AMD 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.

# Contents

## Preface

---

Audience	vi
Documentation Accessibility	vi
Related Resources	vi
Conventions	vi

## 1 Overview of Oracle Big Data Manager

---

1.1 Differences Between Big Data Manager 2.0 and 1.0	1-1
1.2 How to Tell if a Cluster is Running Big Data Manager 2.0 or 1.0	1-2
1.3 Storage Providers Supported by Oracle Big Data Manager	1-2

## 2 Get Started with Oracle Big Data Manager

---

2.1 Open the Oracle Big Data Manager Console	2-1
2.2 Navigate the Oracle Big Data Manager Console	2-2
2.3 Manage Oracle Big Data Manager Users, Roles, and Access	2-6
2.3.1 Add Oracle Big Data Manager Users	2-6
2.3.2 Edit User Details and Manage Roles	2-8
2.3.3 Control Access to Specific Providers	2-8
2.4 Register Storage Providers with Oracle Big Data Manager	2-9
2.5 View Oracle Big Data Manager Settings	2-11

## 3 View Data in Oracle Big Data Manager

---

3.1 Display and Navigate Storage Providers	3-1
3.2 Preview Content from GitHub	3-2
3.3 View Data Properties	3-2

## 4 Transfer and Compare Data

---

4.1 Copy Data (Including Drag and Drop)	4-1
4.2 Copy Data (Including from Multiple Sources)	4-3

4.3	Copy Data Via HTTP	4-5
4.4	Upload Files from a Local Computer to HDFS	4-7
4.5	Move Data in HDFS	4-8
4.6	Import Data into Apache Hive	4-8
4.7	Compare Data Sets	4-9

## 5 Manage Jobs in Oracle Big Data Manager

---

5.1	View Execution History of All Jobs	5-1
5.2	View Summary Information About a Job	5-1
5.3	View the Arguments for a Job	5-2
5.4	View Comparison Results for a Job	5-2
5.5	View Job Output	5-2

## 6 Monitor

---

6.1	Monitor the Status of Services	6-1
6.2	Monitor the Status of Clusters	6-1
6.3	Monitor YARN Queues in Clusters	6-1

## 7 Analyze Data Interactively With Notes

---

7.1	Import a Note	7-1
7.2	Export a Note	7-2
7.3	Create a Note	7-2
7.4	Rename a Note	7-2
	Rename a Note Without Displaying the Note	7-3
	Renaming a Note That's Currently Displayed	7-3
7.5	Clear the Output from Paragraphs in a Note	7-3
7.6	Delete a Note	7-3
7.7	View and Edit a Note	7-4
7.8	Run a Note	7-4
7.9	Organize Notes	7-5
7.10	Manage Notebook Interpreters Settings	7-5

## 8 Using Cloudera Manager to Work With Oracle Big Data Manager

---

## 9 Use the Oracle Big Data Manager bdm-cli Utility

---

9.1	Usage	9-1
9.2	Options	9-2

9.3	Subcommands	9-3
9.4	bdm-cli abort_job	9-4
9.5	bdm-cli copy	9-4
9.6	bdm-cli create_job	9-5
9.7	bdm-cli create_job_template	9-6
9.8	bdm-cli get_data_source	9-8
9.9	bdm-cli get_job	9-8
9.10	bdm-cli get_job_log	9-8
9.11	bdm-cli list_all_jobs	9-9
9.12	bdm-cli list_template_executions	9-9
9.13	bdm-cli ls	9-10

## A Keyboard Shortcuts for Oracle Big Data Manager

---

# Preface

This document describes how to use Oracle Big Data Manager to move data between Hadoop Distributed File System (HDFS) and other storage providers.

**Topics:**

- [Audience](#)
- [Documentation Accessibility](#)
- [Related Resources](#)
- [Conventions](#)

## Audience

This guide is intended for administrators of Oracle Big Data Manager and for users who want to use the tool to transfer data between storage providers.

## 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.

## Related Resources

See the complete Oracle Big Data Appliance documentation set as well as training resources for Oracle Big Data Appliance at <https://docs.oracle.com/en/bigdata/big-data-appliance/index.html>. Use the pulldown menu to find your release of Oracle Big Data Appliance.

## Conventions

The following text conventions are used in this document:

<b>Convention</b>	<b>Meaning</b>
<b>boldface</b>	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

# 1

## Overview of Oracle Big Data Manager

Oracle Big Data Manager is a browser-based tool that gives you broad capabilities to manage data across your enterprise. Oracle Big Data Manager is installed and configured by default when you create a cluster.

You can use Oracle Big Data Manager to connect to and interconnect a range of supported Oracle and non-Oracle data storage providers. After you register storage providers, 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. You can also upload data from your local computer to HDFS.

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's also a CLI for performing copy operations and for creating and administering data management jobs.

The Oracle Big Data Manager administrator can create other user accounts and selectively grant access privileges to storage providers.

By default, Oracle Big Data Manager runs on port 8890 on the same node as Cloudera Manager (usually the third node in the cluster).

### 1.1 Differences Between Big Data Manager 2.0 and 1.0

Both Big Data Manager 2.0 and 1.0 are included in Big Data Appliance 5.2

#### Important:

Big Data Manager 2.0 is installed on new clusters only. Clusters upgraded from an earlier Big Data Appliance release to release 5.2 retain Big Data Manager 1.0. When you move between new and upgraded clusters you will see some differences in the way Big Data Manager 2.0 and 1.0 look and the features that are available.

Big Data Manager 2.0 on new clusters does not include the following features. These features are available in Big Data Manager 1.0 on upgraded clusters.

- Big Data Manager Notebook  
Runs Apache Zeppelin Notebook
- Big Data Manager Proxy  
Provides a reverse proxy to route requests from the web client to the Big Data Manager and Big Data Manager Notebook
- Pipelines  
A feature that lets you chain Big Data Manager jobs together into pipelines that define a specific workflow.

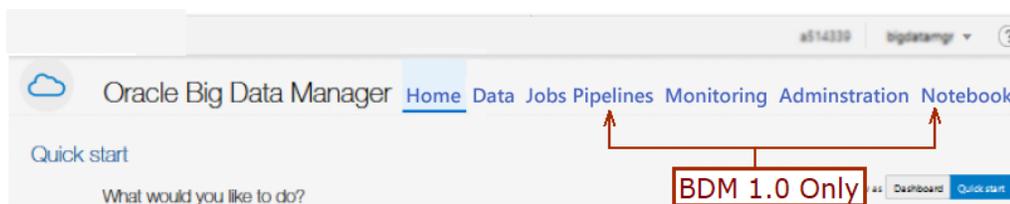
## How to Spot Features That Apply to Big Data Manager 1.0 Only

Throughout this guide, content that applies to Big Data Manager Only is marked with the following badge.

**Applies to:**  Big Data Manager 2.0  Big Data Manager 1.0

## 1.2 How to Tell if a Cluster is Running Big Data Manager 2.0 or 1.0

Big Data Manager 2.0 and 1.0 have different features sets. When you log on to Big Data Manager on a cluster, an easy way to determine which version of Big Data Manager is running is to look at the tabs on the **Home** page. The **Pipeline** and **Notebook** tabs appear in Big Data Manager 1.0 only.



## 1.3 Storage Providers Supported by Oracle Big Data Manager

You can use Oracle Big Data Manager to connect to and interconnect a range of Oracle and non-Oracle data storage providers.

The following storage providers are supported by Oracle Big Data Manager. Note that the Hive and HDFS storage providers are registered for you by default when a cluster is created.

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

Storage providers must be registered with a cluster before they can be accessed through Oracle Big Data Manager.

You can register all providers through the Oracle Big Data Manager console. See [Register Storage Providers with Oracle Big Data Manager](#).

# 2

## Get Started with Oracle Big Data Manager

Oracle Big Data Manager is installed and configured by default when you create a cluster. Before using it, you must configure a few more details, as described in the following topics.

### Note:

You must have administrative privileges to perform some of these tasks. User `bigdatamgr` is created by default when a cluster is created, and that user has the appropriate administrative privileges.

### Topics:

- [Open the Oracle Big Data Manager Console](#)
- [Navigate the Oracle Big Data Manager Console](#)
- [Manage Oracle Big Data Manager Users, Roles, and Access](#)
- [Register Storage Providers with Oracle Big Data Manager](#)
- [View Oracle Big Data Manager Settings](#)

## 2.1 Open the Oracle Big Data Manager Console

Oracle Big Data Manager is integrated with Oracle Big Data Appliance and can be accessed either from Cloudera Manager or from a browser.

- To open the Big Data Manager console from Cloudera Manager, navigate to `CM home > bigdatamanager service > Big Data Manager role > Click Big Data Manager Web UI`. From there, you are directed to the Big Data Manager console login page:

- To open the Big Data Manager console from a browser:

```
https://host_name.domain:port
```

where:

- `host_name` is the host name of the node that's running Big Data Manager.
- `domain` is the domain name.
- `port` is the port used for Big Data Manager (port 8890 by default).

For example, if the host name of the first node is `node01`, and the port is left at the default, the URL used to open the Big Data Manager console would be this:

```
https://node01.domain:8890
```

Sign on using your Oracle Big Data Manager credentials.



**See Also:**

[Manage Oracle Big Data Manager Users, Roles, and Access.](#)

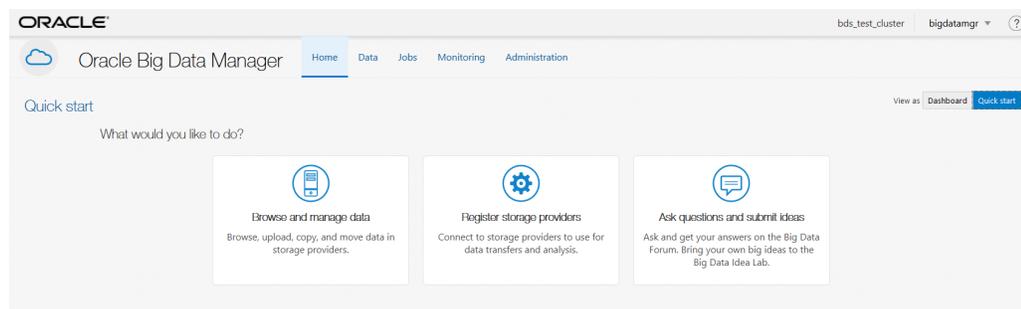
## 2.2 Navigate the Oracle Big Data Manager Console

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

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

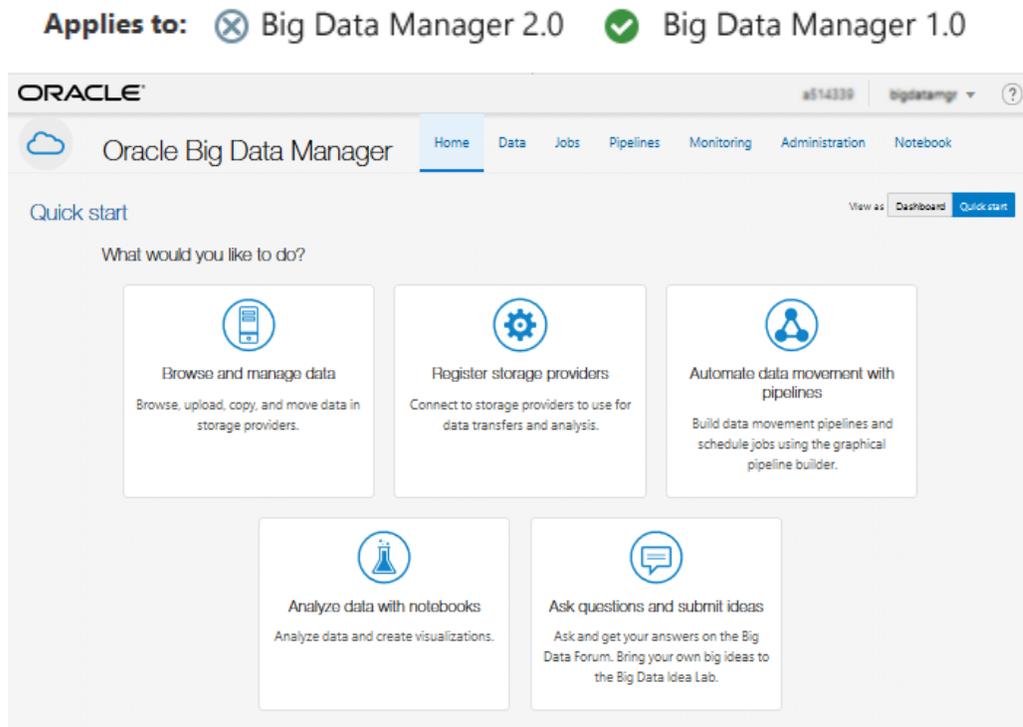
### Home

**Home** has two pages that provide different ways to quickly access tools and controls for viewing and working with data. The **Quick Start** page directs you to the pages for managing data and registering storage providers. It also provides a quick link for asking questions.

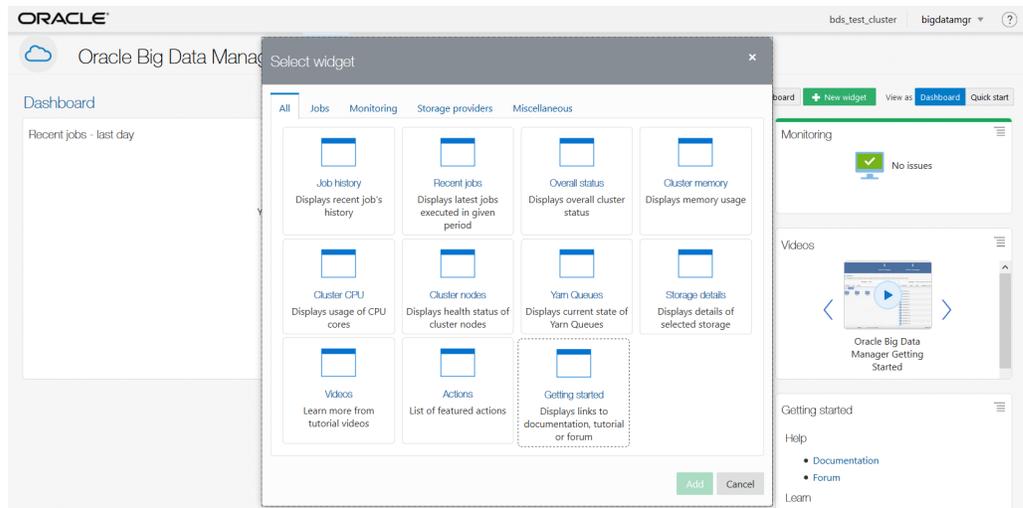


**Note:**

Clusters created in Big Data Appliance 5.2 run Big Data Manager 2.0, the Quickstart view for this version of Big Data Manager is show above. Clusters upgraded from an earlier release of Big Data Appliance run Big Data Manager 1.0. There are some differences in the console between these two versions of Big Data Manager. The Big Data Manager 1.0 console includes the Notebook and Pipelines features, which are visible in the QuickStart view as shown below. Big Data Manager 2.0 does not include these features.

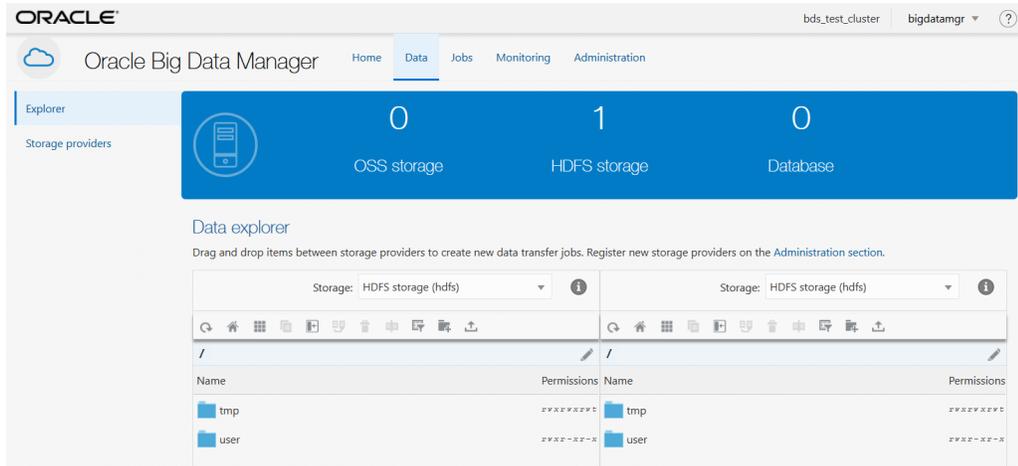


Use the **Dashboard** to build and access custom views of data, activity, status, and history.



## Data

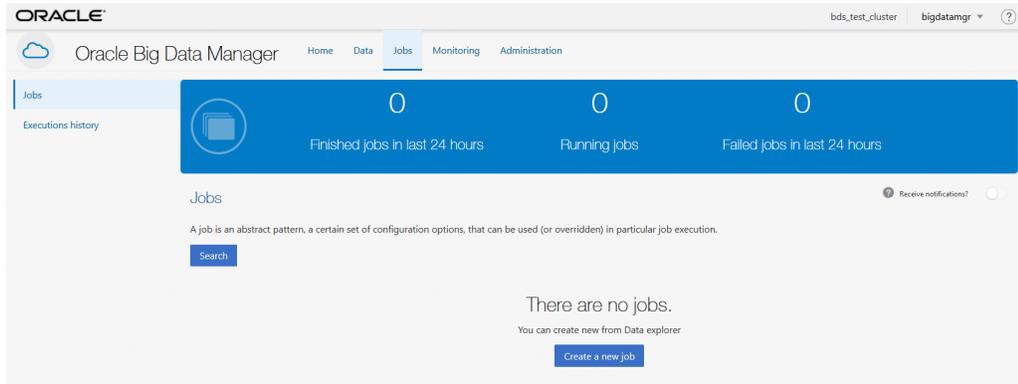
Use the **Data** pages to compare, copy, and move files and containers between data sources, including HDFS. You can also upload files, extract the contents from a ZIP archive, and import data to Apache Hive, among other tasks. The counters in the banner show the number of registered storage providers of a given type in Oracle Big Data Manager.



## Jobs

Use the **Jobs** pages to review, manage, and create data transfer jobs.

Users with the `user` role will see just the jobs they've created.



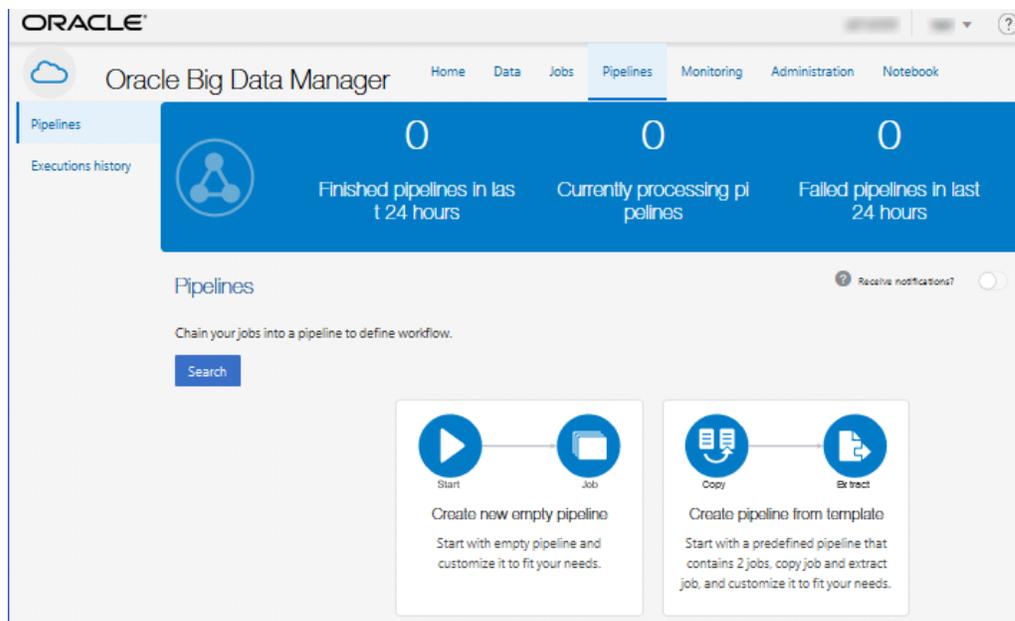
## Pipelines

**Applies to:**  Big Data Manager 2.0  Big Data Manager 1.0

Use the **Pipelines** pages to chain jobs into pipelines to define workflows.

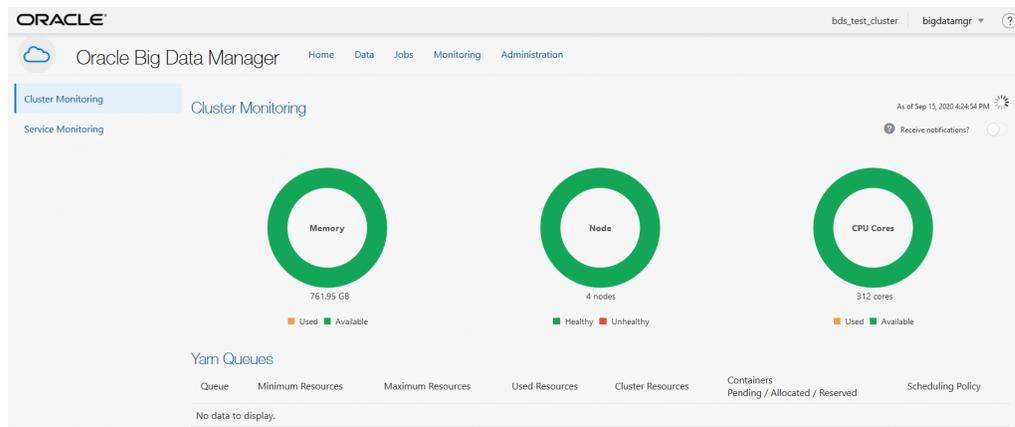
Users with the `user` role will see just the pipelines they've created.

 [Tutorial](#)



## Monitoring

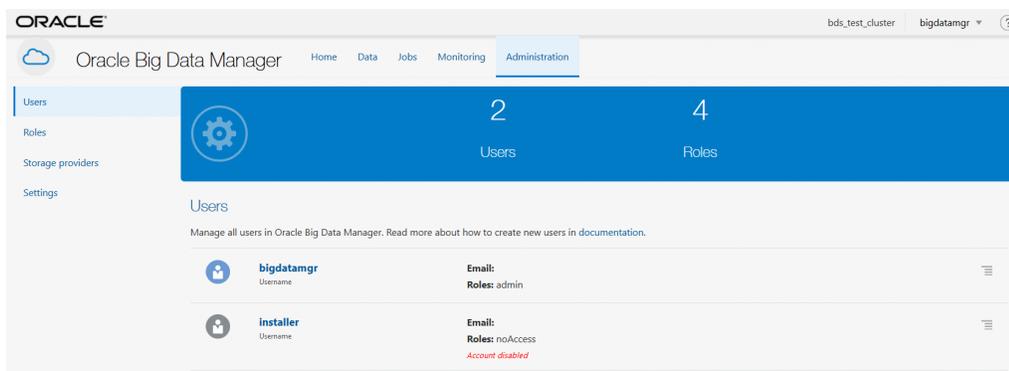
Use the **Monitoring** pages to monitor the performance of your cluster and the services running on it.



## Administration

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

Users with the `admin` role, such as the default `bigdatamgr` user created when a cluster is provisioned, have full administrative privileges and will see all of the tabs on the left as shown below. Users with the `user` role have access to just the storage provider pages. These users can register new storage providers, and manage the storage providers they've added.



## 2.3 Manage Oracle Big Data Manager Users, Roles, and Access

An administrator creates Oracle Big Data Manager users on 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 storage providers. The `bigdatamgr` user has the same password that was assigned to the Cloudera Manager administrator in the Create Instance wizard when the cluster was created.

### 2.3.1 Add Oracle Big Data Manager Users

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

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:

```
echo 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. On the node where Oracle Big Data Manager runs, 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.

 **Note:**

By default, Oracle Big Data Manager is started on port 8890 on the same node that hosts Cloudera Manager.

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.

```
service bigdatamanager restart
```

## 2.3.2 Edit User Details and Manage Roles

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

 **Note:**

User accounts can't be permanently deleted but they can be disabled. Users with disabled accounts won't be able to sign in to Oracle Big Data Manager.

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 the **Administration** tab at the top of the page, and then click **Users** on the left side of the page to show the list of users that have been added.
3. Click the  menu in the row for the user, and then select **Edit**. To disable the user account, select **Disable account**.
4. Edit user details as desired. Click inside the **Roles** field to see and select from available roles.

The **Roles** page on the **Administration** tab lists the roles available for users and the permissions associated with each. The permissions for each role are configured by default in Oracle Big Data Manager and cannot be changed.

## 2.3.3 Control Access to Specific Providers

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

 **Note:**

When you create a new cluster, the `bigdatamgr` user is automatically created within the cluster and granted the Oracle Big Data Manager Administrator role.

To control access to storage providers:

1. Sign in to the Oracle Big Data Manager console as the `bigdatamgr` user, or another user with administrator privileges.
2. Click the **Administration** tab at the top of the page, and then click the **Storage providers** tab on the left side of the page to show a list of registered storage providers.
3. Click the menu icon to the right of the storage provider you are providing access to, and select **Manage users**.
4. 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 Register 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 the **Administration** tab at the top of the page, and then click the **Storage providers** tab on the left side of the page to show a list of registered storage providers.
3. Click the **Register a new storage provider** button.
4. On the **General** page of the **Register storage providers** wizard, enter a name and description for the storage provider, select the storage type, and then click **Next**.
5. On the **Storage Details** page, provide details for accessing the storage provider. You can test connection details by clicking the **Test access to storage** button.
  - **Amazon S3**
    - **Access Key:** Enter the access key ID.
    - **Secret Key:** Enter the secret key.See [About Access Keys](#) in the Amazon S3 documentation.
  - **GitHub**
    - **Access token:** Enter the access token for the GitHub repository. The token should have a scope of `repo` or `public_repo`. See [Creating a personal access token for the command line](#) in the GitHub documentation.
  - **HDFS**

No settings are required because HDFS is part of the cluster where Oracle Big Data Manager is running and is registered automatically. Registration of remote HDFS (HDFS running on a different cluster than where Oracle Big Data Manager is running) is not supported.
  - **Oracle Cloud Infrastructure Object Storage**
    - **Authentication type:** Select the authentication type to be used for data access. Choices are **Credentials** or **Oracle Cloud Instance Principal**. See [Calling Services from an Instance](#) in the Oracle Cloud Infrastructure documentation. If you use an instance principal, you'll need to specify just the **Storage URL** and **Compartment** settings listed below.
    - **Storage URL:** Enter the Oracle Cloud Infrastructure Object Storage URL or select one from the list.
    - **Tenant:** Enter the Tenancy OCID.
    - **User:** Enter the User OCID.

See [Where to Get the Tenancy's OCID and User's OCID](#) in the Oracle Cloud Infrastructure documentation.

- **Access to:** The **Compartment** option is selected by default and is the only option that's currently available.
  - **Compartment:** Enter the OCID of the compartment you want to register. The compartment OCID can be found in the Oracle Cloud Infrastructure console. Once this storage provider is registered, the compartment and its contents (buckets) can be accessed from the Oracle Big Data Manager console.
  - **Public key fingerprint:** Enter the fingerprint of the public key. See [Required Keys and OCIDs](#) in the Oracle Cloud Infrastructure documentation.
  - **Private key:** From the drop-down list, choose how you want to provide the private key. Depending on your choice, either paste the private key value in the text box, or upload the private key file. This is the private key in PEM format that corresponds to the public key whose fingerprint you entered above. See [Required Keys and OCIDs](#) in the Oracle Cloud Infrastructure documentation.
  - **Key passphrase:** (Optional) If the private key is encrypted with a password, enter that password here so the private key file can be decrypted and used for authentication.
  - **Automatic proxy detection:** Select this check box if the cluster is running behind a proxy server and you want to use automatic proxy detection.
- **Oracle Database**
    - **Username:** Enter a user name that corresponds to a user account for the database. This is the user name that should be used to connect to the database.
    - **Password:** Enter the password associated with the user name.
    - **JDBC URL:** Enter the Oracle database connect string in the following format: `jdbc:oracle:thin:@//host:port/service_name`, where *host* and *port* are the database host and port and *service\_name* is the database service name. For example: `jdbc:oracle:thin:@//myhost:1521/myservername`
  - **Oracle MySQL**
    - **Username:** Enter a user name that corresponds to a user account for the database. This is the user name that should be used to connect to the database.
    - **Password:** Enter the password associated with the user name.
    - **JDBC URL:** Enter the MySQL database connect string in the following format: `jdbc:mysql@host:port/[database_name]`, where *host* and *port* are the database host and port and *database\_name* is optional. For example: `jdbc:mysql://myhost:3306` or `jdbc:mysql://myhost:3306/mydatabase`
6. On the **Access** page, use the arrows to specify which users can see this storage from within Oracle Big Data Manager. This setting only affects whether users can see the storage. Whether they can access data on a given storage depends on the settings of the service that provides the storage.

7. Review the details on the **Confirmation** page, and then click the **Register** button.

## 2.5 View Oracle Big Data Manager Settings

Oracle Big Data Manager is installed and configured by default when you create a cluster. Users with administrator privileges can manage some configuration settings if necessary (such as for troubleshooting and so on), but in general the settings should be left at the defaults.

To view and manage configuration settings:

1. Sign in to the Oracle Big Data Manager console as the `bigdatamgr` user, or as another user with administrator privileges.
2. Click the **Administration** tab at the top of the page, and then click the **Settings** tab on the left side of the page to show configuration settings.
3. Manage settings as desired. You can:
  - Use the text box to filter settings
  - Double-click values to edit settings
  - Click the **Restore**  icon to restore the default value of a single setting
  - Click the **Restore all** button to restore the default values of all settings

# 3

## View Data in Oracle Big Data Manager

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

### Topics:

- [Display and Navigate Storage Providers](#)
- [Preview Content from GitHub](#)
- [View Data Properties](#)

### 3.1 Display and Navigate Storage Providers

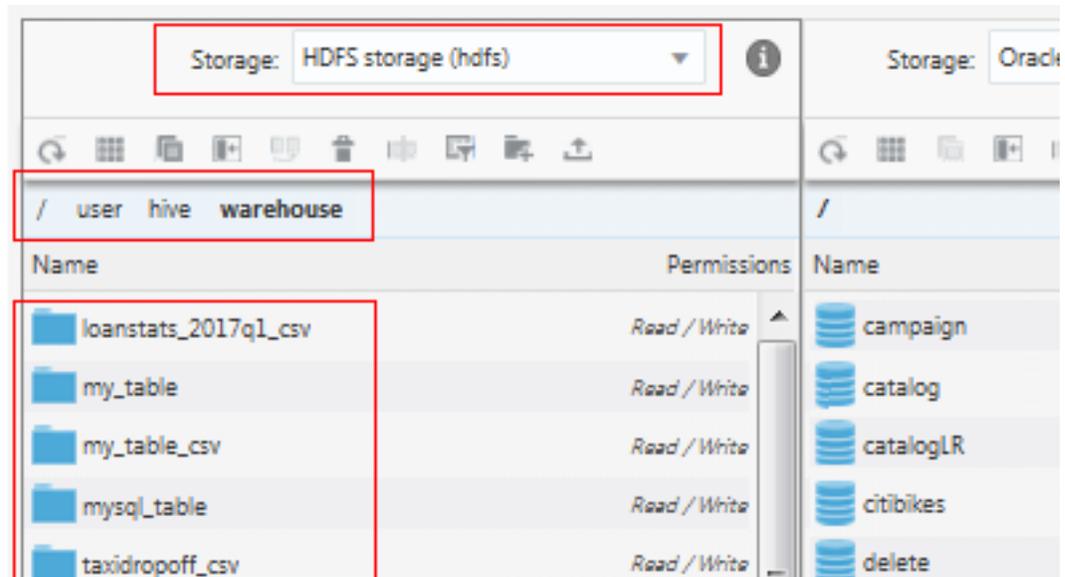
You can display and navigate through storage providers in the Data section of the Oracle Big Data Manager console.

- Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.

The **Data 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, and so on) 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 Preview Content from GitHub

In the Data section of the Oracle Big Data Manager console, you can preview the contents of data in GitHub, including table data presented in a table viewer.

To preview content:

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. 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 View Data Properties

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

To view the properties:

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. 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 displayed:

- Location
- Size
- Modified date
- Owner
- Roles

- Read/write permission

# 4

## Transfer and Compare 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 computer to HDFS and upload data into Apache Hive.

### Topics:

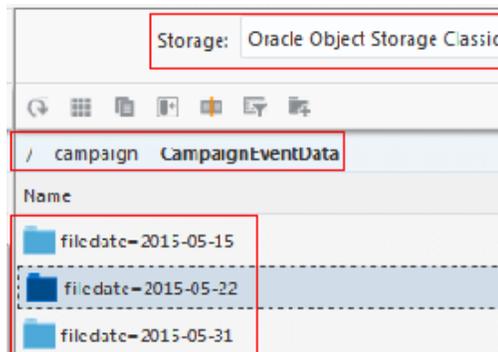
- [Copy Data \(Including Drag and Drop\)](#)
- [Copy Data \(Including from Multiple Sources\)](#)
- [Copy Data Via HTTP](#)
- [Upload Files from a Local Computer to HDFS](#)
- [Move Data in HDFS](#)
- [Import Data into Apache Hive](#)
- [Compare Data Sets](#)

### 4.1 Copy Data (Including Drag and Drop)

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

To copy data from one storage provider to another:

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. In one panel, select a destination storage provider for the copy job from the **Storage** drop-down list, and then navigate to a folder or container by selecting a location in the breadcrumbs or by drilling down in the list below it. For example:



3. In the other panel, select a source storage provider from the **Storage** drop-down list, and then navigate to the folder or container containing the file, folder, or container you want to copy.

4. Do any of the following:
  - a. Drag the source file, folder, or container from the source and drop it on the target.
  - b. Right-click the item you want to copy and select **Copy** from the menu.
  - c. Select the item you want to copy and click **Copy**  in the toolbar.
5. In the **New copy data job** dialog box, provide values as described below.

#### General tab

- **Job name:** A name is provided for the job, but you can change it if you want.
- **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer - copy**.
- **CPU utilization:** Use the slider to specify CPU utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. If you set this to a higher value, you'll have more CPUs for the job, which can mean better performance when you're copying a large number of files. But assigning more CPUs to a job also means there will be fewer CPUs in the cluster available for other tasks.
- **Memory utilization:** Use the slider to specify memory utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. Assigning more memory to a job can increase its performance, but also leaves less free memory available for other tasks. If the job is given too little memory, it will crash. If the job is given more memory than what's currently available, it remains in a PENDING state until the requested amount of memory becomes available.
- **Synchronize destination with sources:** Select this option to synchronize the destination with sources so files or parts of files that have already been copied aren't copied again. This check box is deselected by default. Generally, if you're copying data that's not in the destination location at all, it's better to leave this option deselected so faster data transfer takes place. If you're copying data that's already in the destination location and you just want to update some of it, it's better to select this check box so only new and updated data is detected and transferred.
- **Overwrite existing files:** Select this option to overwrite existing files of the same name in the target destination. This is selected by default.
- **Run immediately:** Select this option to run the job immediately and only once. This is selected by default.
- **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job. You can specify a simplified entry, or click **Advanced entry** to enter a cron expression.

#### Advanced tab

- **Block size:** Select the file chunk size in HDFS from the drop-down list. This setting defaults to the default block size in Hadoop.
- **Number of executors per node:** Specify the number of CPU cores. The default 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 driver:** Select the memory limit from the drop-down list. Memory allocated for the driver is memory allocated for the Application Driver responsible for task scheduling. The default is 1 GB.
  - **Custom logging level:** Select this option to log the job's activity and to select the logging level. The default logging level is INFO.
6. Click **Create**.
- The **Data copy job *job\_number* created** dialog box shows minimal status information about the job. When the job completes, click **View more details** to show more details about the job in the Jobs section of the console. You can also click this link while the job is running.
7. Review the job results. The tabs on the left provide different types of information. You can also stop or remove running jobs and rerun or remove completed jobs from the  menu for the job on each tab.
- The **Summary** tab shows summary information for the job.
  - The **Arguments** tab shows the parameters passed to the job.
  - The **Job output** tab shows job output, which you can also download.
- Also see [Manage Jobs in Oracle Big Data Manager](#).

## 4.2 Copy Data (Including from Multiple Sources)

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

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. In either panel, select a destination storage provider for the copy job from the **Storage** drop-down list.
3. On the toolbar for that panel, click **Copy here from HTTP(S)** .
4. In the **New copy data job** dialog box, select the source to be copied:
  - a. Click the **Select file or directory** button next to **Sources**.
  - b. From the **Location** drop-down list, select the storage provider from which you are copying. The list shows the storage providers registered with Oracle Big Data Manager.
  - c. If desired, select **Use file as link to list of files** to link to a list of files instead of linking directly to a single file.

If you select this option, the file must be a `.csv` file, and each line of the file must satisfy the following schema: `link_to_file[,http_headers_encoded_in_Base64]`, where `,http_headers_encoded_in_Base64` is optional. For example:

```
http://172.16.253.111/public/big.file
https://172.16.253.111/public/small.file
http://172.16.253.111/private/
secret.file,QXV0aG9yaXphdGlvbjogQmFzaWMgYjNkaFkyeGxPa2cwY0hCNVJqQ
```

```
jQK  
https://oracle:passwd@172.16.253.111/private/small.file
```

- d. Navigate to and select the item you're copying. You can also specify a path manually in the **Path** field. For HTTP(S), enter the URI of the source in the **URI** field.
  - e. Click **Select**.
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 inside the **Destination** field and edit the current location.
  7. In the **New copy data job** dialog box, provide values as described below.

#### General tab

- **Job name:** A name is provided for the job, but you can change it if you want.
- **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer - import from HTTP**.
- **CPU utilization:** Use the slider to specify CPU utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. If you set this to a higher value, you'll have more CPUs for the job, which can mean better performance when you're copying a large number of files. But assigning more CPUs to a job also means there will be fewer CPUs in the cluster available for other tasks.
- **Memory utilization:** Use the slider to specify memory utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. Assigning more memory to a job can increase its performance, but also leaves less free memory available for other tasks. If the job is given too little memory, it will crash. If the job is given more memory than what's currently available, it remains in a PENDING state until the requested amount of memory becomes available.
- **Overwrite existing files:** Select this option to overwrite existing files of the same name in the target destination. This is selected by default.
- **Run immediately:** Select this option to run the job immediately and only once. This is selected by default.
- **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job. You can specify a simplified entry, or click **Advanced entry** to enter a cron expression.

#### Advanced tab

- **Block size:** Select file chunk size in HDFS from the drop-down list.
- **Number of executors per node:** Specify the number of cores. The default 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 driver:** Select the memory limit from the drop-down list. Memory allocated for the driver is memory allocated for the Application Driver responsible for task scheduling. The default is 1 GB.
- **Custom logging level:** Select this option to log the job's activity and to select the logging level. The default logging level is INFO. If this check box is not

selected, the logging level for the job defaults to whatever is configured in the cluster.

- **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. Click **Select file**, select the storage provider from the **Location** drop-down list, and navigate to and select the file that contains the HTTP(S) header information. For HTTP(S), enter the URI for the file in the **URI** field. Click **Select file**.

The structure of the files with HTTP(S) headers is *regex\_pattern,http\_headers*. For example, the following file applies specific HTTP(S) headers for files that contain `image` in their path or name. Note that HTTP(S) headers must be Base64 encoded:

```
. *image.* ,QXV0aG9yaXphdG1vbjogQmFzaWMgYjNkaFkyeGxPa2cwY0hCNVJqQjQK
```

#### 8. Click **Create**.

The **Data copy job *job\_number* created** dialog box shows minimal status information about the job. When the job completes, click **View more details** to show more details about the job in the Jobs section of the console. You can also click this link while the job is running.

#### 9. Review the job results. The tabs on the left provide different types of information. You can also stop or remove running jobs and rerun or remove completed jobs from the menu for the job on each tab.

- The **Summary** tab shows summary information for the job.
- The **Arguments** tab shows the parameters passed to the job.
- The **Job output** tab shows job output, which you can also download.

Also see [Manage Jobs in Oracle Big Data Manager](#).

## 4.3 Copy Data Via HTTP

In the Data section of 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 the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. In either panel, select a destination storage provider for the copy job from the **Storage** drop-down list.
3. On the toolbar for that panel, click **Copy here from HTTP(S)** .
4. In the **New copy data job** dialog box, select the source to be copied:
  - a. Click the **Select file or directory** button next to **Sources**.
  - b. From the **Location** drop-down list, select the storage provider from which you are copying. The list shows the storage providers registered with Oracle Big Data Manager.
  - c. If desired, select **Use file as link to list of files** to link to a list of files instead of linking directly to a single file.

If you select this option, the file must be a `.csv` file, and each line of the file must satisfy the following schema: `link_to_file[,http_headers_encoded_in_Base64]`, where `,http_headers_encoded_in_Base64` is optional. For example:

```
http://172.16.253.111/public/big.file
https://172.16.253.111/public/small.file
http://172.16.253.111/private/
secret.file,QXV0aG9yaXphdGlvbjogQmFzaWMgYjNKaFkyeGxPa2cwY0hCNVJqQ
jQK
https://oracle:passwd@172.16.253.111/private/small.file
```

- d. Navigate to and select the item you're copying. You can also specify a path manually in the **Path** field. For HTTP(S), enter the URI of the source in the **URI** field.
- e. Click **Select**.
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 inside the **Destination** field and edit the current location.
7. In the tabs of the **New copy data job** dialog box, provide values as described below.

#### General tab

- **Job name:** A name is provided for the job, but you can change it if you want.
- **Job type:** This read-only field describes the type of job. In this case, it's **Data transfer - import from HTTP**.
- **CPU utilization:** Use the slider to specify CPU utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. If you set this to a higher value, you'll have more CPUs for the job, which can mean better performance when you're copying a large number of files. But assigning more CPUs to a job also means there will be fewer CPUs in the cluster available for other tasks.
- **Memory utilization:** Use the slider to specify memory utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. Assigning more memory to a job can increase its performance, but also leaves less free memory available for other tasks. If the job is given too little memory, it will crash. If the job is given more memory than what's currently available, it remains in a PENDING state until the requested amount of memory becomes available.
- **Overwrite existing files:** Select this option to overwrite existing files of the same name in the target destination. This is selected by default.
- **Run immediately:** Select this option to run the job immediately and only once. This is selected by default.
- **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job. You can specify a simplified entry, or click **Advanced entry** to enter a cron expression.

#### Advanced tab

- **Block size:** Select file chunk size in HDFS from the drop-down list.

- **Number of executors per node:** Specify the number of cores. The default 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 driver:** Select the memory limit from the drop-down list. Memory allocated for the driver is memory allocated for the Application Driver responsible for task scheduling. The default is 1 GB.
- **Custom logging level:** Select this option to log the job's activity and to select the logging level. The default logging level is INFO. If this check box is not selected, the logging level for the job defaults to whatever is configured in the cluster.
- **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. Click **Select file**, select the storage provider from the **Location** drop-down list, and navigate to and select the file that contains the HTTP(S) header information. For HTTP(S), enter the URI for the file in the **URI** field. Click **Select file**.

The structure of the files with HTTP(S) headers is *regex\_pattern,http\_headers*. For example, the following file applies specific HTTP(S) headers for files that contain `image` in their path or name. Note that HTTP(S) headers must be Base64 encoded:

```
. *image.* ,QXV0aG9yaXphdG1vbjogQmFzaWMgYjNkaFkyeGxPa2cwY0hCNVJqQjQK
```

#### 8. Click **Create**.

The **Data copy job *job\_number* created** dialog box shows minimal status information about the job. When the job completes, click **View more details** to show more details about the job in the Jobs section of the console. You can also click this link while the job is running.

9. Review the job results. The tabs on the left provide different types of information. You can also stop or remove running jobs and rerun or remove completed jobs from the  menu for the job on each tab.
  - The **Summary** tab shows summary information for the job.
  - The **Arguments** tab shows the parameters passed to the job.
  - The **Job output** tab shows job output, which you can also download.

Also see [Manage Jobs in Oracle Big Data Manager](#).

## 4.4 Upload Files from a Local Computer to HDFS

In the Data section of the Oracle Big Data Manager console, you can upload files and directories from a local computer to HDFS.

To upload files from a local computer to HDFS:

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. From the **Storage** drop-down list in either panel, select **HDFS storage (hdfs)** and navigate to the destination for the uploaded files.
3. On the toolbar of the panel you chose above, click **File upload** .

4. In the **Files upload** dialog box, click **Select files** and then **Files upload** or **Directory upload** to select the files or directories from your computer's file system. Alternatively, you can drag files or directories from your computer's file system to the **Or drop files or directories here** box.

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

5. Click **Upload** to upload the selected files or directories. You can expand **Details** to see details of the running or completed job, or click **Close** to close the dialog box. If you click **Close** while the job is running, you also have the option to cancel the job.

## 4.5 Move Data in HDFS

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

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. From the **Storage** drop-down list in one of the panels, select **HDFS storage (hdfs)** and navigate to the target location.
3. From the **Storage** drop-down list in the other panel, select **HDFS storage (hdfs)**, navigate to the item you want to move, and select it.
4. On the toolbar for the panel containing the item to be moved, click **Move** .
5. When prompted, click **Move**, and then click **Close** to close the dialog box.

## 4.6 Import Data into Apache Hive

In the Data section of the Oracle Big Data Manager console, you can import `.csv` files, Apache Avro files, and Apache Parquet files from HDFS to the HiveServer2 on the cluster where Oracle Big Data Manager is running.

To import one of the supported files into Hive:

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. From the **Storage** drop-down list in one of the panels, select **HDFS storage (hdfs)**.

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.

3. Navigate to the file you want to import, right-click it, select **Import to Apache Hive**, and select how to import it: **Import as CSV**, **Import as Apache Avro**, or **Import as Apache Parquet**.
4. Provide import details.
  - For **Import as CSV**, provide values on each tab of the **Create a new job** wizard and then click **Create**.

- For **Import as Apache Avro** and **Import as Apache Parquet**, specify the Hive table in the **Table name** field, and select the Hive database from the **Database name** drop-down list. The table name defaults to the name of the file you selected to import. The database defaults to the `default` Hive database. If other databases have been created in your Hive metastore, those databases will also be listed in the drop-down list. Click **Import to Hive**.

## 4.7 Compare 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.

1. Click the **Data** tab at the top of the page, and then click the **Explorer** tab on the left side of the page.
2. Navigate to and 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.
3. On the toolbar, click **Compare** .
4. In the **New compare data job** dialog box, provide values as described below.

### General tab

- **Job name:** A name is provided for the job, but you can change it if you want.
- **Job type:** This read-only field describes the type of job. In this case, it's **Oracle Distributed Diff - compare**.
- **CPU utilization:** Use the slider to specify CPU utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. If you set this to a higher value, you'll have more CPUs for the job, which can mean better performance when you're copying a large number of files. But assigning more CPUs to a job also means there will be fewer CPUs in the cluster available for other tasks.
- **Memory utilization:** Use the slider to specify memory utilization for the job. The proper job configuration will be calculated based on the cluster's shape. This is set to 30 percent by default. Assigning more memory to a job can increase its performance, but also leaves less free memory available for other tasks. If the job is given too little memory, it will crash. If the job is given more memory than what's currently available, it remains in a PENDING state until the requested amount of memory becomes available.
- **Run immediately:** Select this option to run the job immediately and only once. This is selected by default.
- **Repeated execution:** Select this option to schedule the time and frequency of repeated executions of the job. You can specify a simplified entry, or click **Advanced entry** to enter a cron expression.

### Advanced tab

- **Diff file block size:** Select a value from the drop-down list to specify the level of comparison. The lower the value the more detailed the comparison will be and the more time it will take. The default value is 512 MB.

- **Number of executors per node:** Specify the number of CPU cores. The default 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 driver:** Select the memory limit from the drop-down list. Memory allocated for the driver is memory allocated for the Application Driver responsible for task scheduling. The default is 1 GB.
  - **Custom logging level:** Select this option to log the job's activity and to select the logging level. The default logging level is INFO. If this check box is not selected, the logging level for the job defaults to whatever is configured in the cluster.
5. Click **Create**.
- The **Data compare job *job\_number* created** dialog box shows minimal status information about the job. When the job completes, click **View more details** to show more details about the job in the Jobs section of console. You can also click this link while the job is running.
6. Review the job results. In particular, 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.

Jobs > 24437abd-1442-47f0-bba6-625266843924 As of Feb 12, 2018 1:57:48 PM

Summary

Arguments

Comparison results

Job output

Succeeded    100 %    Feb 12, 2018 1:57:14 PM    26s

Status	Progress	Started on	Duration
--------	----------	------------	----------

● Data compare job #1518461826/19 ⌵

Overall changes

Modified file(s)	1
Identical file(s)	1
File(s) missing in swift://MyContainer...	4
Directory / file type mismatch	0
File(s) missing in hdfs://user/ber/temp	0

ODIFF Explorer - comparison details i

Source: swift://MyContainer/ber/temp/	Destination: hdfs://user/ber/temp																								
<table style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 60%;">Name</th> <th style="width: 40%;">Permissions</th> </tr> <tr style="background-color: #ffe6e6;"> <td>0b994dee-fa71-4d1c-9357-4432184d11af...</td> <td>-</td> </tr> <tr> <td>1G8File_1</td> <td>-</td> </tr> <tr style="background-color: #ffe6e6;"> <td>1G8File_10</td> <td>-</td> </tr> <tr style="background-color: #ffe6e6;"> <td>1G8File_13</td> <td>-</td> </tr> <tr style="background-color: #ffe6e6;"> <td>ticeval2000xxxx.csv</td> <td>-</td> </tr> </table>	Name	Permissions	0b994dee-fa71-4d1c-9357-4432184d11af...	-	1G8File_1	-	1G8File_10	-	1G8File_13	-	ticeval2000xxxx.csv	-	<table style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 60%;">Name</th> <th style="width: 40%;">Permissions</th> </tr> <tr style="background-color: #e6ffe6;"> <td>0b994dee-fa71-4d1c-9357-4432184d11af...</td> <td>-</td> </tr> <tr> <td>1G8File_1</td> <td>-</td> </tr> <tr style="background-color: #e6ffe6;"> <td>1G8File_10</td> <td>-</td> </tr> <tr style="background-color: #e6ffe6;"> <td>1G8File_13</td> <td>-</td> </tr> <tr style="background-color: #e6ffe6;"> <td>ticeval2000xxxx.csv</td> <td>-</td> </tr> </table>	Name	Permissions	0b994dee-fa71-4d1c-9357-4432184d11af...	-	1G8File_1	-	1G8File_10	-	1G8File_13	-	ticeval2000xxxx.csv	-
Name	Permissions																								
0b994dee-fa71-4d1c-9357-4432184d11af...	-																								
1G8File_1	-																								
1G8File_10	-																								
1G8File_13	-																								
ticeval2000xxxx.csv	-																								
Name	Permissions																								
0b994dee-fa71-4d1c-9357-4432184d11af...	-																								
1G8File_1	-																								
1G8File_10	-																								
1G8File_13	-																								
ticeval2000xxxx.csv	-																								

# 5

## Manage Jobs in Oracle Big Data Manager

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

### Topics

- [View Execution History of All Jobs](#)
- [View Summary Information About a Job](#)
- [View the Arguments for a Job](#)
- [View Comparison Results for a Job](#)
- [View Job Output](#)

### 5.1 View Execution History of All Jobs

You can view the execution history of all jobs in the Jobs section of the Oracle Big Data Manager console.

To view the execution history of all jobs:

1. Click the **Jobs** tab at the top of the page.
2. Click **Executions history** on the left side of the page.

### 5.2 View 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 the **Jobs** tab at the top of the page, and then click **Executions history** on the left side of the page.
2. In the row for the job you want to review, click the  menu, and then select **View Details**. You can also rerun and remove completed jobs from this menu, or stop and remove running jobs.
3. Click the **Summary** tab on the left.

The information shown can include the following:

- Name
- Source and destination
- Description
- Schedule
- Status
- Progress

- Start and stop times
- Duration

## 5.3 View 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 parameters for a job:

1. Click the **Jobs** tab at the top of the page, and then click **Executions history** on the left side of the page.
2. In the row for the job you want to review, click the  menu, and then select **View Details**. You can also rerun and remove completed jobs from this menu, or stop and remove running jobs.
3. Click the **Arguments** tab on the left.

The parameters defined for the job are shown, for example number of executors, block size, and so on.

## 5.4 View Comparison Results for a Job

You can view comparison results for a compare job in the Jobs section of the Oracle Big Data Manager console.

To view comparison results for a compare job:

1. Click the **Jobs** tab at the top of the page, and then click **Executions history** on the left side of the page.
2. In the row for the compare job you want to review, click the  menu, and then select **View Details**. You can also rerun and remove completed jobs from this menu, or stop and remove running jobs.
3. Click the **Comparison results** tab on the left.

Job details are shown.

## 5.5 View Job Output

You can view output for a job in the Jobs section of the Oracle Big Data Manager console.

To view job output:

1. Click the **Jobs** tab at the top of the page, and then click **Executions history** on the left side of the page.
2. In the row for the job you want to review, click the  menu, and then select **View Details**. You can also rerun and remove completed jobs from this menu, or stop and remove running jobs.
3. Click the **Job output** tab on the left.

The output of the job is shown. You can also download the output.

# 6

## Monitor

In the Oracle Big Data Manager console, you can monitor the status of services, clusters, and YARN queues in clusters.

### Topics:

- [Monitor the Status of Services](#)
- [Monitor the Status of Clusters](#)
- [Monitor YARN Queues in Clusters](#)

### 6.1 Monitor the Status of Services

To monitor the status of services:

1. Click **Monitoring** in the Oracle Big Data Manager console menu bar.
2. Click **Service Monitoring** on the left side of the page.

In the Service Monitoring page:

- All services are listed with the current status of each - either **OK** or **Down**.
  - The Service URL(s) for each service are listed along with the last update day and time.
3. Optionally, enable the **Receive notifications** option to receive background notifications for finished jobs and cluster monitoring.

### 6.2 Monitor the Status of Clusters

To monitor the status of clusters:

1. Click **Monitoring** in the Oracle Big Data Manager console menu bar.
2. If it's not already selected, click **Cluster Monitoring** on the left side of the page.

A graphic and numeric display of cluster status is displayed, including:

- Memory usage (Used and Available)
  - Nodes (the number of nodes and current status: Healthy or Unhealthy)
  - CPU Cores (Used and Available)
3. Optionally, enable the **Receive notifications** option to receive background notifications for finished jobs and cluster monitoring.

### 6.3 Monitor YARN Queues in Clusters

To monitor YARN queues:

1. Click **Monitoring** in the Oracle Big Data Manager console menu bar.

At the bottom of the Monitoring page, the YARN queue for **root** is displayed.

2. In the **Yarn Queues** section, drill down on the **root** queue.
3. Then, drill down on **root.users** to display YARN queues for all `root` users.

The following YARN queue information is provided for each user:

- Minimum and maximum resources for CPU and memory
  - Used resources for CPU and memory
  - Cluster resources for CPU and memory
  - The number of containers that are pending, allocated, and reserved
4. Optionally, enable the **Receive notifications** option to receive background notifications for finished jobs and cluster monitoring.

# 7

## Analyze Data Interactively With Notes

Use notes to explore and visualize data iteratively.

**Applies to:**  Big Data Manager 2.0  Big Data Manager 1.0

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

### Topics:

- [Import a Note](#)
- [Export a Note](#)
- [Create a Note](#)
- [Rename a Note](#)
- [Clear the Output from Paragraphs in a Note](#)
- [Delete a Note](#)
- [View and Edit a Note](#)
- [Run a Note](#)
- [Organize Notes](#)
- [Manage Notebook Interpreters Settings](#)

### 7.1 Import 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 **Select JSON File** to upload a file from your local computer, or click **Add from URL** to upload from a location on the internet.

## 7.2 Export a Note

You can export a note from the Notebook section of 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 near the top of the page.
2. On the Notebook home page, click the name of the note you want to export. The note opens.
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.

## 7.3 Create a Note

You can create a note in the Notebook section of 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, or click the **Notebook** drop-down list from the banner at the top of the page 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**, and **jdbc**.
5. Click **Create**. The note displays with an empty paragraph. Each note is composed of one or more paragraphs.

## 7.4 Rename a Note

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

## Rename 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 in the banner near the top of the 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:

```
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.

## 7.5 Clear the Output from Paragraphs in a Note

You can clear the output from a note 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 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**,  and then click **OK**.
3. Reopen the note and confirm that the output is cleared from the result section of all paragraphs.

## 7.6 Delete a Note

You can delete a note in 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**.

## 7.7 View and Edit 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 toolbars on the page. When you make changes to a note or a paragraph, the changes are automatically saved.

## 7.8 Run 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 in the banner near the top of the page.
2. On the Notebook home page, click the name of the note you want to run. The note opens.
3. Click the **Run all paragraphs**  icon in the toolbar at the top of the page to execute all 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 **Run this paragraph**  icon in the toolbar for the paragraph.

## 7.9 Organize Notes

You can organize notes into directories in the Notebook section of 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 [Create a Note](#) and [Rename 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.

## 7.10 Manage 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)
- 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 **Interpreter**.
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:

The screenshot shows the Oracle Data Science IDE interface. At the top, there is a blue navigation bar with a home icon, 'Notebook', 'Job', a search bar labeled 'Search your Notes', and a 'Menu' dropdown. Below the navigation bar, a red box highlights the 'spark' interpreter name and its dependencies: '%spark, %spark.sql, %spark.dep, %spark.pyspark'. To the right of this box are four action buttons: 'spark ui', 'edit', 'restart', and 'remove'. Below the interpreter name, the 'Option' section shows 'The interpreter will be instantiated' with a 'Globally' dropdown, 'in' a 'shared' dropdown, and 'process.'. There are two checkboxes: 'Connect to existing process' and 'Set permission', both of which are currently unchecked. The 'Properties' section contains a table with two columns: 'name' and 'value'. The table has one row with the value 'args'.

name	value
args	

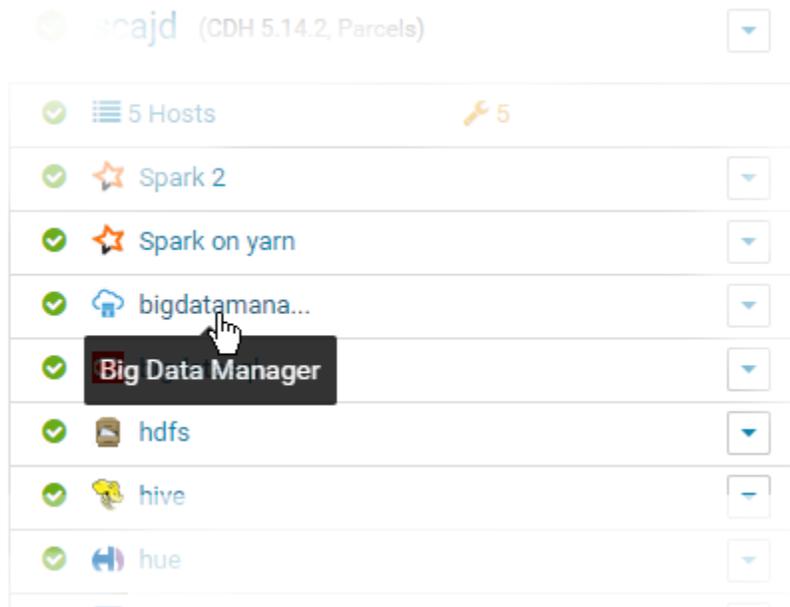
# 8

## Using Cloudera Manager to Work With Oracle Big Data Manager

Oracle Big Data Manager is automatically included as a service in 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

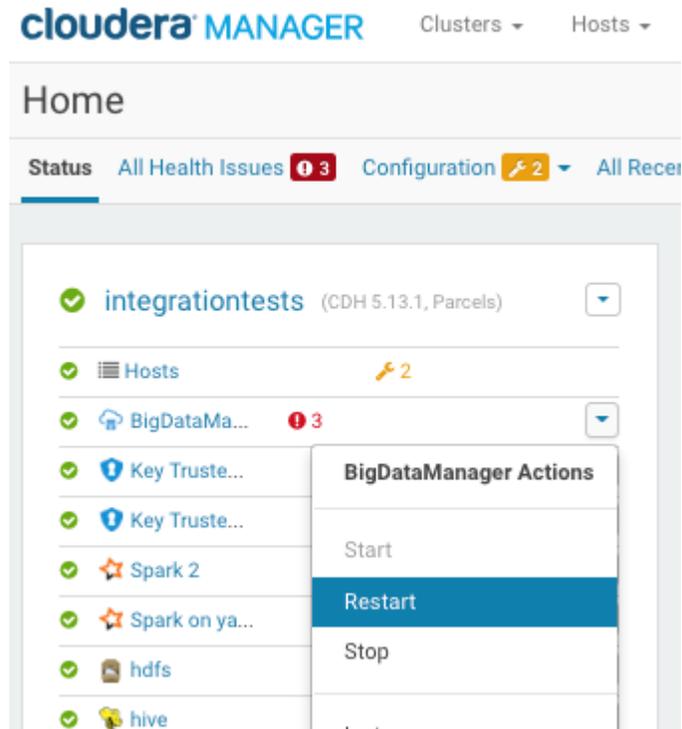
Cloudera Manager supports four Oracle Big Data Manager roles. Cloudera Manager provides the usual administrative, monitoring and reporting facilities for each of these roles.

Oracle Big Data Manager Role	Description
Hosts	Provides representations of physical or virtual nodes where Oracle Big Data Manager is running. (By default, this product runs on the third node of the cluster. )
Big Data Manager	Runs the Oracle Big Data Manager web interface and the back-end server.
Big Data Manager Notebook	Runs Apache Zeppelin Notebook. <b>Present in Big Data Manager 1.0 only.</b>

Oracle Big Data Manager Role	Description
Big Data Manager Proxy	Provides a reverse proxy to route requests from the web client to the Big Data Manager and Big Data Manager Notebook roles. <b>Present in Big Data Manager 1.0 only.</b>

### Stopping and Starting Oracle Big Data Manager

On the Cloudera Manager Home page, 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. Note that the Big Data Manager Notebook and Proxy roles are present only on clusters running Big Data Manager 1.0 (clusters upgraded from an earlier version of Big Data Appliance ).

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 Configuration Manager is hosted.

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

# 9

## Use 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. The `bdm-cli` utility is installed by default on all nodes of the cluster. You don't need to install it.

`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.

### 9.1 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 (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:8890/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
```

## 9.2 Options

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

Option	Description
<code>--bdm-passwd</code> <i>path_to_password_file</i>	Path to the Oracle Big Data Manager user password file. Environment variable: <code>BDM_PASSWORD</code>
<code>--bdm-url</code> <i>bdm_url</i>	Oracle Big Data Manager server URL. Environment variable: <code>BDM_URL</code>
<code>--bdm-username</code> <i>username</i>	Oracle Big Data Manager server user name. Default value: <code>oracle</code> Environment variable: <code>BDM_USERNAME</code>

Option	Description
<code>-f [table csv json]</code>	Specify the output format: <ul style="list-style-type: none"> <li><code>table</code> (default) Each field is displayed in a separate column.</li> <li><code>csv</code> Each record is displayed as a comma-separated list on a single line.</li> <li><code>json</code>: The output is displayed in JavaScript Object Notation (JSON) format.</li> </ul>
<code>--fields fields</code>	Specifies comma-separated fields depending on the type of object.
<code>-h</code> <code>--help</code>	Show this message and exit.
<code>--no-check-certificate</code>	Don't validate the server's certificate.
<code>--proxy proxy</code>	Proxy server.
<code>--tenant-name tenant_name</code>	Name of the tenant. Default value: <code>admin</code>
<code>-v</code>	Print the REST request body.
<code>--version</code>	Show the Oracle Big Data Manager version and exit.

## 9.3 Subcommands

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

Command	Description
<a href="#">bdm-cli abort_job</a>	Abort a running job.
<a href="#">bdm-cli copy</a>	Execute a job to copy sources to destination.
<a href="#">bdm-cli create_job</a>	Execute a new job from an existing template.
<a href="#">bdm-cli create_job_template</a>	Create a new job template.
<a href="#">bdm-cli get_data_source</a>	Find a data source by name.
<a href="#">bdm-cli get_job</a>	Get a job by UUID.
<a href="#">bdm-cli get_job_log</a>	Get a job log.
<a href="#">bdm-cli list_all_jobs</a>	List all jobs from the execution history.
<a href="#">bdm-cli list_template_executions</a>	List all jobs from the execution history for the given template.
<a href="#">bdm-cli ls</a>	List files from a specific location.

## 9.4 bdm-cli abort\_job

Abort a running job.

### Syntax

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

### Options

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

### Example

Abort a job.

```
/usr/bin/bdm-cli -f json --no-check-certificate --bdm-url $
{DATA_HOST}:8890/bdcs/api --bdm-username ${DATA_USER} --bdm-passwd $
{USER_PASSWORD_FILE} abort_job 24ef30e8-913b-4402-baf8-74b99c211f50
```

## 9.5 bdm-cli copy

Execute a job to copy sources to destination.

### Syntax

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

### Options

Option	Description
block-size <i>block_size</i>	Specify the block size in bytes.
--description <i>description</i>	Data source description.
--driver-memory-size <i>driver_memory_size</i>	Specify the maximum amount of memory for the Oracle Storage Cloud Service driver.
--dst-provider <i>oss_destination_provider</i>	Specify the provider of the destination, when using Oracle Cloud Infrastructure Object Storage Classic destination.
-h	Show this message and exit.
--help	
--memory-size-per-node <i>memory_size_per_node</i>	Specify the Spark executors memory limit in GB per node, for example, 40GB.

Option	Description
<code>--number-of-executor-nodesnumber_of_executors_per_node</code>	Specify the maximum number of Spark executors per node, for example, 10GB.
<code>--number-of-threads-per-nodenumbers_of_threads_per_node</code>	Specify the maximum number of threads per node.
<code>part-size part_size</code>	Specify the part size in bytes.
<code>--recursive</code> <code>--no-recursive</code>	Recursively copy (enabled by default).
<code>--retry</code> <code>--no-retry</code>	Retry data transfer in case of failure.
<code>--src-provider oss_source_provider</code>	Specify the provider of the source, when using for Oracle Cloud Infrastructure Object Storage Classic.
<code>--sync</code> <code>--no-sync</code>	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}:8890/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}
```

## 9.6 bdm-cli create\_job

Execute a new job from an existing template.

### Syntax

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

### Options

Option	Description
<code>--run-now</code>	Execute job immediately if job scheduling is set. Ignored otherwise.
<code>--source source</code>	Source file, for example: hdfs:///user/oracle/test.raw
<code>--destination destination</code>	The destination directory, for example: swift:// container.storagename/test-dir.
<code>--driver-memory-size driver_memory_size</code>	Specify the maximum amount of memory for an Oracle Storage Cloud Service driver.

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

## 9.7 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
<code>--abort-running-job</code> <code>--no-abort-running-job</code>	Abort an already running execution if the next scheduled execution is started.
<code>--block-size</code> <i>block_size</i>	Specify block size in bytes.
<code>--data-source-name</code> <i>data_source_name</i>	Job's data source name.
<code>--description</code> <i>description</i>	Job template description.

Option	Description
<code>--dst-provider</code> <i>destination_provider</i>	Specify for <code>oss:///</code> destination.
<code>--environment</code> <i>environment</i>	Environment in JSON format: {"envName1": "envValue2", "envName2": "envValue2"}
<code>-h</code> <code>--help</code>	Show this message and exit.
<code>--history-size</code> <i>history_size</i>	Count of executions history log.
<code>--job-executable-class</code> <i>job_executable_class</i>	Main Java class used for the Spark job execution.
<code>--job-schedule</code> <i>job_schedule</i>	Specify cron-like job schedule, for example: "0 56 8 * * ?" means run every day at 08h 56m UTC time.
<code>--job-template-type</code> <i>job_template_type</i>	Specify job template type. Allowed values are: <ul style="list-style-type: none"> <li>DATA_MOVEMENT_COPY</li> <li>GENERAL</li> </ul>
<code>--libraries</code> <i>libraries</i>	Hadoop libraries, for example: <code>OdcpLibraries</code> . This option can have multiple values, for example: <code>--libraries OdcpLibraries --libraries OdcpLibraries</code>
<code>--memory-size-per-node</code> <i>memory_size_per_node</i>	Specify the Spark executors memory limit in GB per node, for example: 40G.
<code>--number-of-executor-nodes</code> <i>number_of_executor_per_node</i>	Specify the maximum number of Spark executors per node, for example: 10G.
<code>--number-of-threads-per-node</code> <i>number_of_threads_per_node</i>	Specify the maximum of threads per node.
<code>--part-size</code> <i>part_size</i>	Specify part size in bytes.
<code>--recursive</code> <code>--no-recursive</code>	Recursively copy (enabled by default).
<code>--retry</code> <code>--no-retry</code>	Retry data transfer in case of failure.
<code>--src-provider</code> <i>oss_source_provider</i>	Specify the provider of the source, when using for Oracle Cloud Infrastructure Object Storage Service.
<code>--sync</code> <code>--no-sync</code>	Synchronize source with destination.
<code>--tags</code> <i>tags</i>	User defined tag. This option can have multiple values, for example: <code>--tags system --tags datamovement --tags copy</code>

## 9.8 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	Show this message and exit.
--help	

## 9.9 bdm-cli get\_job

Get a job by UUID.

### Syntax

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

### Options

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

### Example

Get information on a job.

```
/usr/bin/bdm-cli -f json --no-check-certificate --bdm-url $
{DATA_HOST}:8890/bdcs/api --bdm-username ${DATA_USER} --bdm-passwd $
{USER_PASSWORD_FILE} get_job ${JOB_UUID}
```

## 9.10 bdm-cli get\_job\_log

Get a job log.

### Syntax

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

## Options

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

## 9.11 bdm-cli list\_all\_jobs

List all jobs from the execution history.

### Syntax

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

### Options

Option	Description
-h	Show this message and exit.
--help	
--limit <i>limit</i>	Specify the size of the page.
--offset <i>offset</i>	Specify the paging offset.

### Example

List all jobs.

```
/usr/bin/bdm-cli -f json --no-check-certificate --bdm-url $
{DATA_HOST}:8890/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
```

## 9.12 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	Show this message and exit.
--help	

## 9.13 bdm-cli ls

List files from a specific location.

### Syntax

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

### Options

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

### Example

List HDFS content under selected user.

```
/usr/bin/bdm-cli -f json --no-check-certificate --bdm-url $
{DATA_HOST}:8890/bdcs/api --bdm-username ${DATA_USER} --bdm-passwd $
{USER_PASSWORD_FILE} ls hdfs:///user/${DATA_USER}/integration_in --
provider hdfs
```

# 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 Oracle 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 <b>Rename</b> 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.