

# Oracle® Big Data Discovery

Upgrade Guide

Version 1.6.0 • Revision A • January 2018

# Copyright and disclaimer

Copyright © 2015, 2018, Oracle and/or its affiliates. All rights reserved.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners. UNIX is a registered trademark of The Open Group.

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, the following notice is applicable:

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

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

This software or hardware and documentation may provide access to or information on 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. 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.

# Table of Contents

<b>Copyright and disclaimer .....</b>	<b>2</b>
<b>Preface.....</b>	<b>4</b>
About this guide .....	4
Audience .....	4
Conventions.....	4
Contacting Oracle Customer Support.....	5
<b>Chapter 1: Before You Upgrade.....</b>	<b>6</b>
Supported upgrade paths .....	6
Upgrade requirements.....	7
Downloading the upgrade packages .....	9
Downloading a WebLogic Server patch .....	10
Obtaining the Hadoop client libraries.....	10
Running the BDD hotfix.....	12
Backing up your current cluster .....	12
<b>Chapter 2: Upgrading Big Data Discovery .....</b>	<b>14</b>
Overview .....	14
Silent upgrade .....	15
Merging the configuration files.....	15
Editing the configuration file .....	16
Upgrading the cluster .....	19
Troubleshooting a failed upgrade.....	20
Studio fails to start .....	20
Rolling back a failed upgrade .....	20
<b>Chapter 3: After You Upgrade .....</b>	<b>21</b>
Verifying the upgrade .....	21
Changes made by the upgrade script.....	21
Post-upgrade configuration .....	22
Clearing the browser cache .....	22

## Preface

Oracle Big Data Discovery is a set of end-to-end visual analytic capabilities that leverage the power of Apache Spark to turn raw data into business insight in minutes, without the need to learn specialist big data tools or rely only on highly skilled resources. The visual user interface empowers business analysts to find, explore, transform, blend and analyze big data, and then easily share results.

## About this guide

This guide describes how to upgrade your Big Data Discovery cluster to version 1.6.0.

## Audience

This guide is intended for system administrators and developers who are upgrading Oracle Big Data Discovery.

## Conventions

The following conventions are used in this document.

### Typographic conventions

The following table describes the typographic conventions used in this document.

Typeface	Meaning
<b>User Interface Elements</b>	This formatting is used for graphical user interface elements such as pages, dialog boxes, buttons, and fields.
Code Sample	This formatting is used for sample code segments within a paragraph.
<i>Variable</i>	This formatting is used for variable values. For variables within a code sample, the formatting is <i>Variable</i> .
File Path	This formatting is used for file names and paths.

### Path variable conventions

This table describes the path variable conventions used in this document.

Path variable	Meaning
<code>\$ORACLE_HOME</code>	Indicates the absolute path to your Oracle Middleware home directory, where BDD and WebLogic Server are installed.

Path variable	Meaning
\$BDD_HOME	Indicates the absolute path to your Oracle Big Data Discovery home directory, \$ORACLE_HOME/BDD-<version>.
\$DOMAIN_HOME	Indicates the absolute path to your WebLogic domain home directory. For example, if your domain is named bdd-<version>_domain, then \$DOMAIN_HOME is \$ORACLE_HOME/user_projects/domains/bdd-<version>_domain.
\$DGRAPH_HOME	Indicates the absolute path to your Dgraph home directory, \$BDD_HOME/dgraph.

## Contacting Oracle Customer Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. This includes important information regarding Oracle software, implementation questions, product and solution help, as well as overall news and updates from Oracle.

You can contact Oracle Customer Support through Oracle's Support portal, My Oracle Support at <https://support.oracle.com>.



## Chapter 1

---

# Before You Upgrade

Before upgrading BDD, you must download the required software packages and prepare your cluster.

[\*Supported upgrade paths\*](#)

[\*Upgrade requirements\*](#)

[\*Downloading the upgrade packages\*](#)

[\*Downloading a WebLogic Server patch\*](#)

[\*Obtaining the Hadoop client libraries\*](#)

[\*Running the BDD hotfix\*](#)

[\*Backing up your current cluster\*](#)

## Supported upgrade paths

The following upgrade paths are supported.

- 1.1.x to 1.6.0
- 1.2.x to 1.6.0
- 1.3.x to 1.6.0
- 1.4.x to 1.6.0
- 1.5.x to 1.6.0

You can't upgrade to 1.6.0 from 1.0.x. If you have 1.0.x, you must first upgrade to 1.1.x or higher.

## Upgrade requirements

Before upgrading, make sure your system meets the following requirements.

Requirement	Description
Hadoop	<p>BDD supports the following:</p> <ul style="list-style-type: none"> <li>• Cloudera Distribution for Hadoop (CDH) 5.10.x for BDA 4.8</li> <li>• Cloudera Distribution for Hadoop (CDH) 5.11.x for BDA 4.9</li> <li>• Cloudera Distribution for Hadoop (CDH) 5.12.x for BDA 4.10</li> <li>• Cloudera Distribution for Hadoop (CDH) 5.13.x for BDA 4.11</li> <li>• Hortonworks Data Platform (HDP) 2.6</li> <li>• MapR 5.2.1+</li> </ul> <p>If you don't have one of the above installed, upgrade your Hadoop cluster before upgrading BDD. For instructions, refer to the documentation for your Hadoop distribution.</p> <p>Additionally, BDD requires Spark 1.6.x. Be sure to check your current version of Spark and upgrade it, if necessary.</p> <p>Before upgrading your Hadoop cluster, be aware of the following:</p> <ul style="list-style-type: none"> <li>• You can't switch to a different Hadoop distribution without reinstalling BDD. For example, if you currently have CDH, you can't switch to HDP.</li> <li>• You should stop your BDD cluster before you upgrade Hadoop. Once Hadoop has been upgraded, follow the procedure described in "Switching Hadoop versions" in the <i>Administrator's Guide</i> to enable BDD to work with the new version.</li> </ul>
Security	<p>BDD can run on Hadoop clusters secured with TLS/SSL and HDFS data at rest encryption. If you don't currently have these enabled, you can configure them in your Hadoop cluster before upgrading BDD and then enable them for BDD at upgrade time.</p> <p>See the section "Security options" in the <i>Installation Guide</i> for instructions. Also, be sure to set the <code>HADOOP_CERTIFICATES_PATH</code> property in <code>bdd.conf</code> before upgrading.</p>
Operating system	<p>BDD supports the following operating systems:</p> <ul style="list-style-type: none"> <li>• Oracle Enterprise Linux 6.4+, 7.1, 7.2</li> <li>• Red Hat Enterprise Linux 6.4+, 7.1, 7.2</li> </ul> <p>If you don't have one of the above installed, upgrade your OS before upgrading BDD.</p> <p>Additionally, you must clear the <code>\$http_proxy</code> environment variable. You can do this by running the following command:</p> <pre>export http_proxy=</pre>
OS user	<p>The entire upgrade process must be performed by the <code>bdd</code> user, the OS user that installed BDD and runs all BDD processes. This user must have the same UID on all nodes in the cluster, including Hadoop nodes.</p>

Requirement	Description
Dgraph databases (index)	<p>The Dgraph database (index) directory must contain enough free space to double your databases, as this may occur temporarily during the upgrade process.</p> <p>Although the current version of BDD enables you to store your Dgraph databases on HDFS, you can't move them there until after the upgrade. For instructions on doing this, see the <i>Administrator's Guide</i>.</p>
Transform Service	<p>If you're upgrading from 1.1.x, the upgrade script will install the Transform Service on the nodes you specify in <code>bdd.conf</code>. For best performance, these should be Managed Servers.</p> <p>The Transform Service requires at least 11GB of RAM and 39GB of virtual memory, and may require more depending on the size of its workloads. Be sure that the nodes you install it on meet these requirements.</p>
Studio database	<p>Studio supports the following database versions:</p> <ul style="list-style-type: none"> <li>• Oracle 11g, 12c (12.1.0.1.0+)</li> <li>• MySQL 5.5.3+</li> </ul>
Workflow Manager Service	<p>The upgrade script will install the Workflow Manager Service on the node you specify in <code>bdd.conf</code>. Note that you can't install it on more than one node.</p> <p>The Workflow Manager Service requires a database (including username, password, and schema) of one of the following types, which you must create before upgrading:</p> <ul style="list-style-type: none"> <li>• Oracle 11g, 12c (12.1.0.1.0+)</li> <li>• MySQL 5.5.3+</li> </ul> <p>For instructions, see "Workflow Manager Service database requirements" in the <i>Installation Guide</i>.</p>
Perl modules	<p>Verify that the following Perl modules are installed on the Admin Server:</p> <ul style="list-style-type: none"> <li>• <code>Mail::Address</code></li> <li>• <code>XML::Parser</code></li> </ul> <p>If any is missing, install it according to the instructions in "Installing the required Perl modules" in the <i>Installation Guide</i>.</p>
Linux utilities	<ul style="list-style-type: none"> <li>• curl 7.19.7+ (with support for the <code>--tlsv1.2</code> and <code>--negotiate</code> options) and Network Security Services (NSS) 3.16.1+ must be installed on all Studio nodes.</li> <li>• Apache Ant 1.7.1+ must be installed and added to the <code>PATH</code> on all nodes in the cluster, including Hadoop nodes.</li> </ul>



Requirement	Description
Browser	<p>BDD supports the following browsers:</p> <ul style="list-style-type: none"> <li>• Microsoft Edge</li> <li>• Internet Explorer 11</li> <li>• Firefox ESR</li> <li>• Chrome for Business</li> <li>• Safari Mobile 9.x</li> </ul>

## Downloading the upgrade packages

Once you have satisfied all prerequisites, you can download the upgrade packages.

These are available on the Oracle Software Delivery Cloud. After downloading them, you must put them in a single directory on the Admin Server, called the *upgrade source directory*. You will perform the entire upgrade process from this location.

To download the upgrade packages:

1. On the Admin Server, create a new directory or select an existing one to be the upgrade source directory.
2. Within the upgrade source directory, create a subdirectory named `packages`.
3. Go to the [Oracle Software Delivery Cloud](#) and sign in.
4. Accept the Export Restrictions.
5. Check **Programs** if it isn't already selected.
6. In the **Product** text box, enter `Oracle Big Data Discovery`.
7. Click **Select Platform** and check **Linux x86-64**.  
**Oracle Big Data Discovery** displays in the **Selected Products** table.
8. Click **Continue**.
9. Verify that **Available Release** and **Oracle Big Data Discovery 1.6.x.x.x for Linux x86-64** are both checked, then click **Continue**.
10. Accept the Oracle Standard Terms and Restrictions and click **Continue**.
11. In the **File Download** popup, click **Download All**.

The following packages are downloaded to your machine:

- **First of two parts of the Oracle Big Data Discovery binary**
- **Second of two parts of the Oracle Big Data Discovery binary**
- **Installer for Oracle Big Data Discovery**
- **SDK for Oracle Big Data Discovery**
- **Documentation for Oracle Big Data Discovery**
- **Oracle Fusion Middleware 12c (12.2.1.0.0) WebLogic Server and Coherence**

Also, make a note of each package's part number, as you will need this information to identify it.

12. On the Admin Server, move the packages you downloaded to `/<upgrade_source_dir>/packages`.
13. Rename the first BDD binary package `bdd1.zip` and the second `bdd2.zip`.  
This is required for the upgrade script to recognize them.
14. Extract the WebLogic Server package.  
This creates a file called `fmw_12.2.1.0.0_wls.jar`, which contains the WebLogic Server installer.
15. Move up a directory to the upgrade source directory and unzip the installer package:

```
unzip packages/<installer_package>.zip
```

This creates a new directory within the upgrade source directory called `/installer`, which contains the scripts and files required to perform the upgrade.

Next, you can optionally download a WebLogic Server patch for the upgrade script to apply. If you don't want to patch WebLogic Server, move on to [Obtaining the Hadoop client libraries on page 10](#).

## Downloading a WebLogic Server patch

You can optionally download a WebLogic Server patch for the upgrade script to apply when it runs. **(Note: If JDK 1.8.0\_161+ is used, then the PSU released January 2018 and later which includes the Bug 27117282 should be applied. More information could be found at [Doc ID 2357811.1](#))**

You can only apply one patch when upgrading. If the patch fails, the upgrade script will remove it and continue running.

For more information on patching WebLogic Server, see [Oracle Fusion Middleware Patching with OPatch](#).

To download a WebLogic Server patch:

1. Within the upgrade source directory, create a new directory called `WLSPatches`.  
Don't change the name of this directory or the upgrade script won't recognize it.
2. Go to [My Oracle Support](#) and log in.
3. On the **Patches & Updates** tab, find and download the patch you want to apply.
4. Move all ZIP files associated with the patch to `WLSPatches/`. Don't extract the files. The upgrade script will do this when it runs.

Next, you must obtain the Hadoop client libraries.

## Obtaining the Hadoop client libraries

Next, obtain the Hadoop client libraries and put them on the Admin Server.

BDD requires a number of client libraries to interact with Hadoop. In a normal Hadoop cluster, these libraries are spread out, making it difficult for BDD to find them all. To solve this issue, the upgrade script adds the required libraries to a single JAR, called the Hadoop fat JAR, and distributes it to all BDD nodes.

The specific libraries you need depend on your Hadoop distribution. The location you put them in is arbitrary, as you will define it in `bdd.conf`.

- **CDH:** Download the following files from <http://archive-primary.cloudera.com/cdh5/cdh/5/> to the Admin Server and extract them:

- spark-`<spark_version>`.cdh.`<cdh_version>`.tar.gz
- hive-`<hive_version>`.cdh.`<cdh_version>`.tar.gz
- hadoop-`<hadoop_version>`.cdh.`<cdh_version>`.tar.gz
- avro-`<avro_version>`.cdh.`<cdh_version>`.tar.gz

Be sure to download the files that correspond to the component versions you currently have installed.

- **HDP:** Copy the following libraries from your Hadoop nodes to the Admin Server. Note that these directories might not all be on the same node.
  - /usr/hdp/`<version>`/hive/lib/
  - /usr/hdp/`<version>`/spark/lib/
  - /usr/hdp/`<version>`/hadoop/
  - /usr/hdp/`<version>`/hadoop/lib/
  - /usr/hdp/`<version>`/hadoop-hdfs/
  - /usr/hdp/`<version>`/hadoop-hdfs/lib/
  - /usr/hdp/`<version>`/hadoop-yarn/
  - /usr/hdp/`<version>`/hadoop-yarn/lib/
  - /usr/hdp/`<version>`/hadoop-mapreduce/
  - /usr/hdp/`<version>`/hadoop-mapreduce/lib/
- **MapR:** Locate the following directories on your Hadoop nodes and copy them to the Admin Server. Note that they might not all be on the same node.
  - /opt/mapr/spark/spark-1.6.1/lib
  - /opt/mapr/hive/hive-1.2/lib
  - /opt/mapr/zookeeper/zookeeper-3.4.5
  - /opt/mapr/zookeeper/zookeeper-3.4.5/lib
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/common
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/common/lib
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/hdfs
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/hdfs/lib
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/mapreduce
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/tools/lib
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/yarn
  - /opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/yarn/lib

If you're upgrading from 1.1.0 or 1.1.1, you should apply the upgrade hotfix now. If you have a later version, move on to [Backing up your current cluster on page 12](#).

## Running the BDD hotfix

If you're upgrading from 1.1.0 or 1.1.1, you must apply one of the hotfixes to install the backup and restore scripts required during the upgrade. **Note that this step isn't required if you're upgrading from 1.1.3 or higher.**

There are separate hotfixes for 1.1.0 and 1.1.1. Run the script that corresponds to the version you currently have installed.

To run the BDD hotfix:

1. On the Admin Server, go to  
`<upgrade_source_dir>/installer/hotfix/<version>/hotfix_EADMIN-1503.`  
Where `<version>` is the version of BDD you currently have installed.
2. Run the hotfix script:

```
./hotfix_EADMIN-1503.sh <path/to/bdd.conf>
```

Where `<path/to/bdd.conf>` is the absolute path to your current `bdd.conf` file under `$BDD_HOME/BDD_manager/conf`.

The hotfix adds two new scripts to `$BDD_HOME/BDD_manager/bin`: `bdd-backup.sh` and `bdd-restore.sh`.

## Backing up your current cluster

Next, back up your current cluster.

If you're upgrading from 1.1.0 or 1.1.1, run the backup script added by the hotfix. If you're upgrading from 1.1.3 or higher, use `bdd-admin's backup` command. All versions of the script back up the following data to a single TAR file, which you can use to restore your current cluster if the upgrade fails.

- Configuration files
- Studio database
- Schema and data for Hive tables created in Studio
- Dgraph databases (index)
- Sample files in HDFS

Before you run the backup script, verify the following:

- The `BDD_STUDIO_JDBC_USERNAME` and `BDD_STUDIO_JDBC_PASSWORD` environment variables are set. If they aren't, the script will prompt you for the username and password of the Studio database at runtime.
- The database client is installed on the Admin Server. For MySQL databases, this should be MySQL client. For Oracle databases, this should be Oracle Database Client, which must be installed with a type of Administrator. Note that the Instant Client isn't supported.
- If you have an Oracle database, the `ORACLE_HOME` environment variable is set to the parent directory of the `/bin` directory the `sqlplus` executable is located in. For example, if the `sqlplus` executable is located in `/u01/app/oracle/product/11/2/0/dbhome/bin`, set `ORACLE_HOME` to `/u01/app/oracle/product/11/2/0/dbhome`.

Additionally, make sure the temporary directories used during the backup operation contain enough free space. These requirements only need to be met for the duration of the backup operation.

- If you have 1.1.x, both the destination of the backup TAR file and the location defined by `TEMP_FOLDER_PATH` in `bdd.conf` must contain enough space to store the Dgraph databases, `$HDFS_DP_USER_DIR`, and the `edpDataDir` (defined in `edp.properties`) at the same time.
- If you have 1.2.x or later:
  - The destination of the backup TAR file must contain enough space to store the Dgraph databases, `$HDFS_DP_USER_DIR`, and the `edpDataDir` (defined in `edp.properties`) at the same time.
  - The location on the Admin Server defined by `BACKUP_LOCAL_TEMP_FOLDER_PATH` in `bdd.conf` also requires enough space to store all three items simultaneously.
  - The location in HDFS defined by `BACKUP_HDFS_TEMP_FOLDER_PATH` in `bdd.conf` must contain enough free space to accommodate the largest of these items, as it will only store them one at a time.

To back up your current cluster:

1. On the Admin Server, open a command prompt and go to `$BDD_HOME/BDD_manager/bin`.
2. Stop your cluster :

```
./bdd-admin.sh stop [-t <minutes>]
```

3. Run the backup script.

- BDD 1.1.0 and 1.1.1:

```
./bdd-backup.sh -o -v <backup_tar_file>
```

- All other versions:

```
./bdd-admin.sh backup -o -v <backup_tar_file>
```

Where `<backup_tar_file>` is the absolute path to the backup TAR file. This file must not exist and its parent directory must be writable.

4. Enter the username and password for the Studio database, if prompted.

The script backs up your current cluster to the specified TAR file.



## Chapter 2

# Upgrading Big Data Discovery

---

Once you've obtained the required BDD and Hadoop packages and prepared your cluster, you can begin upgrading BDD.

[Overview](#)

[Merging the configuration files](#)

[Editing the configuration file](#)

[Upgrading the cluster](#)

[Troubleshooting a failed upgrade](#)

[Rolling back a failed upgrade](#)

## Overview

You upgrade your cluster by running two separate scripts.

The first script merges the current version of `bdd.conf` (under `$BDD_HOME/BDD_manager/conf`) with the version from the new release. This ensures that your cluster will retain most of its current configuration. However, you still need to manually edit the file to set new properties added in this release and any that couldn't be merged.

The second script upgrades your cluster and installs the new components and features. When it finishes running, your cluster will be completely upgraded and ready to use.

The upgrade script moves the old version of BDD installation with `/BDD_<old_version>` into a tar file as `BDD_OLD_ARCHIVE_TAR` specified at [Editing the configuration file](#), then installs the new version of BDD in `/BDD_<version>` into `$ORACLE_HOME`, which is the new `$BDD_HOME`. The script also installs the new Weblogic Server package into `$ORACLE_HOME`.



**Important:** If you decide to uninstall BDD, the uninstallation script will only delete `$ORACLE_HOME` which contains the current instance of `/BDD <version>`, it will not delete the tar file `BDD OLD ARCHIVE TAR`. You must delete this tar file manually after uninstallation. You must not delete this tar file before the current BDD is uninstalled, otherwise it will make rollback script failing to recover BDD to its pre-upgrade state.

[Silent upgrade](#)

## Silent upgrade

You can optionally run the upgrade script in silent mode. This means that instead of prompting you for information it requires at runtime, the script obtains it from environment variables you set beforehand.

Normally, when you run the script, it prompts you to enter:

- The user name and password for your Hadoop cluster manager (Cloudera Manager or Ambari).
- The user name and password for the WebLogic Server admin.
- The user name and password for the Studio database.
- The use rname and password for the Workflow Manager Service database.

You can avoid these steps by setting the following environment variables before running the script.

Environment variable	Description
BDD_HADOOP_UI_USERNAME	The user name for Cloudera Manager/Ambari.
BDD_HADOOP_UI_PASSWORD	The password for Cloudera Manager/Ambari.
BDD_WLS_USERNAME	The user name for the WebLogic Server admin.
BDD_WLS_PASSWORD	The password for the WebLogic Server admin.
BDD_STUDIO_JDBC_USERNAME	The user name for the Studio database.
BDD_STUDIO_JDBC_PASSWORD	The password for the Studio database.
BDD_WORKFLOW_MANAGER_JDBC_USERNAME	The user name for the Workflow Manager Service database.
BDD_WORKFLOW_MANAGER_JDBC_PASSWORD	The password for the Workflow Manager Service database.

## Merging the configuration files

The first step in the upgrade process is to merge the current version of `bdd.conf` with the one from the new release.

You do this by running `merge-bddconf.sh`, which populates the new version of `bdd.conf` with values from the current version under `$BDD_HOME/BDD_manager/conf`. Although the merged file will contain most of your current settings, you'll need to manually edit any properties the script couldn't merge, as well as those that were added in the new release.

To merge the configuration files:

1. On the Admin Server, open a command prompt and go to the upgrade source directory.
2. Run the merge script:

```
./merge-bddconf.sh <current_bdd.conf> <new_bdd.conf>
```

Where `<current_bdd.conf>` is the absolute path to the current version of `bdd.conf` under `$BDD_HOME/BDD_manager/conf`, and `<new_bdd.conf>` is the relative path to the new one. On most systems, this should be:

```
./merge-bddconf.sh $BDD_HOME/BDD_manager/conf/bdd.conf bdd.conf
```

Next, update the new and unmerged properties in `bdd.conf`.

## Editing the configuration file

After you merge `bdd.conf`, you need to manually edit the new properties and any that weren't merged properly.

The upgrade script validates `bdd.conf` at runtime and fails if the file contains any invalid values. To avoid this, keep the following in mind when editing the file:

- The accepted values for some properties are case-sensitive and must be entered exactly as they appear in the table below.
- All hostnames must be fully qualified domain names (FQDNs).
- Any symlinks included in paths must be identical on all nodes. If any are different or don't exist, the upgrade may fail.
- Each port setting requires a unique value. You can't use the same port number more than once.
- Some of the directories defined in `bdd.conf` have location requirements. These are specified in the table below.

The properties you need to edit are described below. You should also review the rest of the file to verify that all other settings are still accurate. Additional information on the properties in `bdd.conf` is available in the *Installation Guide*.

Property	Description
BDD_OLD_VERSION	The version of your current BDD. This will already be populated.
BDD_OLD_ARCHIVE_TAR	The absolute path to the tar file to archive the current BDD installation.
INSTALLER_PATH	The absolute path to the software packages of BDD and Weblogic Server in the upgrade source directory.
STUDIO_JDBC_URL	The JDBC URL for your database.  This isn't a new property, but the URL for your database may have changed since you installed. You should verify that this property is still accurate and edit it if it's not.



Property	Description
WORKFLOW_MANAGER_JDBC_URL	<p>The JDBC URL for the Workflow Manager Service database.</p> <p>There are two templates for this property. Copy the template that corresponds to your database type to <code>WORKFLOW_MANAGER_JDBC_URL</code> and update the URL to point to your database.</p> <ul style="list-style-type: none"> <li>For MySQL databases, use the first template and update the URL as follows: <pre data-bbox="662 590 1445 674">jdbc:mysql://&lt;database hostname&gt;:&lt;port number&gt; /&lt;database name&gt;?useUnicode=true&amp;characterEncoding=UTF-8&amp;useFastDateParsing=false</pre> </li> <li>For Oracle databases, use the first template and update the URL as follows: <pre data-bbox="662 768 1445 821">jdbc:oracle:thin: @&lt;database hostname&gt;:&lt;port number&gt;:&lt;database SID&gt;</pre> </li> </ul> <p>If you're installing on more than one machine, be sure to use the database host's FQDN and not <code>localhost</code>.</p>
HADOOP_CLIENT_LIB_PATHS	<p>A comma-separated list of the absolute paths to the Hadoop client libraries.</p> <p>To set this property, copy the template to <code>HADOOP_CLIENT_LIB_PATHS</code> and update the paths to point to the client libraries you copied to the install machine.</p> <p>Don't change the order of the paths in the list as they <i>must</i> be specified as they appear.</p> <p>For more information on the client libraries and how to obtain them, see <a href="#">Obtaining the Hadoop client libraries on page 10</a>.</p>
HADOOP_CERTIFICATES_PATH	<p>Only required for Hadoop clusters with TLS/SSL enabled. The absolute path to the directory on the install machine where you put the certificates for HDFS, YARN, Hive, and the KMS services. For instructions on exporting these, see the <i>Installation Guide</i>.</p>
SPARK_ON_YARN_JAR	<p>The absolute path to the Spark on YARN JAR on Hadoop nodes. Verify that the value matches the correct version of your Hadoop distribution.</p>
WORKFLOW_MANAGER_SERVERS	<p>The Workflow Manager Service node. Note that you can only define one.</p>
WORKFLOW_MANAGER_PORT	<p>The port the Workflow Manager Service listens on for data processing requests.</p>

Property	Description
KERBEROS_KEYTAB_PATH	<p>The absolute path to the BDD keytab file on the install machine, <b>this keytab file shouldn't be located under ORACLE_HOME, or it will be deleted.</b></p> <p>The upgrade script will rename this to bdd.keytab and copy it to \$BDD_HOME/common/kerberos/ on all BDD nodes.</p> <p>This property is only required if ENABLE_KERBEROS is set to TRUE.</p>

## BDD 1.2.x upgrades

If you're upgrading from BDD 1.2.x, you should also verify the following properties.

Property	Description
ENABLE_CLUSTERING_SERVICE	For use by Oracle Support only. Leave this property set to <code>FALSE</code> .
CLUSTERING_SERVICE_SERVERS	For use by Oracle Support only. Don't modify this property.
CLUSTERING_SERVICE_PORT	For use by Oracle Support only. Don't modify this property.
STUDIO_JDBC_CACHE	Enables/disables database caching in Studio.
DGRAPH_ENABLE_MPP	For use by Oracle Support only. Don't modify this property.
DGRAPH_MPP_PORT	For use by Oracle Support only. Don't modify this property.
RECORD_SEARCH_THRESHOLD	The minimum number of characters the average value of a String attribute must contain to be record searchable. For more information, see the <i>Data Processing Guide</i> .
VALUE_SEARCH_THRESHOLD	The minimum number of characters the average value of a String attribute must contain to be value searchable. For more information, see the <i>Data Processing Guide</i> .

## BDD 1.1.x upgrades

If you're upgrading from BDD 1.1.x, you should also verify the following properties.

Property	Description
ZOOKEEPER_INDEX	The index of the Dgraph cluster in the ZooKeeper ensemble, which ZooKeeper uses to identify it.  This used to be called <code>COORDINATOR_INDEX</code> .
TRANSFORM_SERVICE_SERVERS	A comma-separated list of the Transform Service nodes. For best performance, these should all be Managed Servers. In particular, they shouldn't be Dgraph nodes, as both the Dgraph and the Transform Service require a lot of memory.  Each of these nodes must have at least 11GB of RAM and 39GB of available virtual memory.
TRANSFORM_SERVICE_PORT	The port the Transform Service listens on for requests from Studio.
BACKUP_LOCAL_TEMP_FOLDER_PATH	The absolute path to the default temporary folder on the Admin Server used during backup and restore operations. This can be overridden on a case-by-case basis by the <code>bdd-admin</code> script.

Property	Description
BACKUP_HDFS_TEMP_FOLDER_PATH	The absolute path to the default temporary folder on HDFS used during backup and restore operations. This can be overridden on a case-by-case basis by the <code>bdd-admin</code> script.
DGRAPH_USE_MOUNT_HDFS	Specifies whether the Dgraph databases are stored on HDFS. When upgrading from 1.1.x and lower, this must be set to <code>FALSE</code> .
DGRAPH_HDFS_MOUNT_DIR	The absolute path to the local directory where the Dgraph mounts the HDFS root directory.  Use a nonexistent directory when installing. If this location changes after installing, the new location must be empty and have read, write, and execute permissions for the <code>bdd</code> user.  This setting is only required if <code>DGRAPH_USE_MOUNT_HDFS</code> is set to <code>TRUE</code> .
DGRAPH_ENABLE_CGROUP	Enables cgroups for the Dgraph. This must be set to <code>TRUE</code> if you use cgroups on your Dgraph nodes.  If set to <code>TRUE</code> , <code>DGRAPH_CGROUP_NAME</code> must also be set.  For more information on setting up cgroups for the Dgraph, see the <i>Installation Guide</i> and the <i>Administrator's Guide</i> .
DGRAPH_CGROUP_NAME	The name of the cgroup that controls the Dgraph. This is required if <code>DGRAPH_ENABLE_GROUP</code> is set to <code>TRUE</code> .

## Upgrading the cluster

Once you've updated `bdd.conf`, you can upgrade your cluster.

Before you run the upgrade script, verify that you satisfied all prerequisites described in [Before You Upgrade on page 6](#). Additionally, your Hadoop cluster must be running throughout the upgrade process.

To upgrade your cluster:

1. On the Admin Server, open a command prompt and go to the upgrade source directory.
2. Run the upgrade script:

```
./upgrade.sh bdd.conf
```
3. If you're not running the script in silent mode, enter the following when prompted:
  - The user name and password for your Hadoop cluster manager (Cloudera Manager or Ambari).
  - The user name and password for the WebLogic Server admin.
  - The user name and password for the Studio database.
  - The user name and password for the Workflow Manager Service database.

## Troubleshooting a failed upgrade

If the upgrade fails, you should first determine why it failed. To do this, check the script's log files in `<upgrade_source_dir>/packages/installer/upgrade.log.xxx`.

### *Studio fails to start*

## Studio fails to start

In some cases, Studio may fail to start after the upgrade.

If this occurs, restart Studio by going to `$BDD_HOME/BDD_manager/bin` on the Admin Server and running:

```
./bdd-admin.sh restart -c bddServer
```

## Rolling back a failed upgrade

Once you've determined why the upgrade failed, you can roll it back and restore your old cluster.

The rollback script removes all of the changes made by the upgrade script. It also restores your old cluster from backup, starts it, and performs a healthcheck.

Before running the script, verify that the temporary directories it uses contain enough free space. These requirements only need to be met for the duration of the rollback operation.

- If you have 1.1.x, the location defined by `TEMP_FOLDER_PATH` in `bdd.conf` must contain enough space to store the Dgraph databases, `$HDFS_DP_USER_DIR`, and the `edpDataDir` (defined in `edp.properties`) at the same time.
- If you have 1.2.x or later:
  - The location on the Admin Server defined by `BACKUP_LOCAL_TEMP_FOLDER_PATH` in `bdd.conf` must contain enough space to store the Dgraph databases, `$HDFS_DP_USER_DIR`, and the `edpDataDir` (defined in `edp.properties`) at the same time.
  - The location on HDFS defined by `BACKUP_HDFS_TEMP_FOLDER_PATH` in `bdd.conf` must contain free space equal to the largest of these items, as it will only store them one at a time.

When the script finishes running, your cluster will be back to its pre-upgrade state and running. You can then either rerun the upgrade script or restore your old cluster. If you decide to rerun the upgrade, note that you *don't* need to rerun the merge script.

To roll back a failed upgrade:

1. On the Admin Server, open a command prompt and go to the upgrade source directory.
2. Run the rollback script:

```
./rollback.sh bdd.conf [--silent]
```

You can optionally include the `--silent` flag to avoid the confirmation step.

3. Confirm that you want to roll back the upgrade.
4. When prompted, enter the absolute path to your old cluster's backup TAR file.



## Chapter 3

# After You Upgrade

---

This section describes some of the changes made to your cluster, as well as tasks you should perform immediately after upgrading.

*Verifying the upgrade*

*Changes made by the upgrade script*

*Post-upgrade configuration*

*Clearing the browser cache*

## Verifying the upgrade

After the upgrade script completes successfully, you should verify your upgrade by running the `health-check` script.

To verify the upgrade:

1. On the Admin Server, open a new terminal window and go to `$BDD_HOME/BDD_manager/bin`.
2. Run the `health-check` script:

```
./bdd-admin.sh status --health-check
```

If your BDD cluster is healthy, the script's output should be similar to the following:

```
[2016/03/24 04:18:55 -0700] [Admin Server] Checking health of BDD cluster...
[2016/03/24 04:20:39 -0700] [web009.us.example.com] Check BDD functionality.....Pass!
[2016/03
/24 04:20:39 -0700] [web009.us.example.com] Check Hive Data Detector health.....Hive Data
Detector has previously run
[2016/03/24 04:20:39 -0700] [Admin Server] Successfully checked statuses.
```

## Changes made by the upgrade script

The upgrade script made a number of changes to your cluster, but there are a few in particular that you should be aware of.

- The Workflow Manager Service is now installed on the node you specified in `bdd.conf`.
- All sample files have been converted from Avro format to Parquet.
- The following changes have been made to the `bdd-admin` script. For more information, see the *Administrator's Guide*.
  - It contains a number of new options for administering the Workflow Manager Service.
  - The `restore` command supports two types of restore operations: full and data-only.

- The `publish-config` command has a new configuration type called `database`, which can be used to modify component database settings.
- For the `start`, `stop`, and `restart` commands, the `dp` component only affects the DP CLI and doesn't have an effect on current DP jobs. To cancel DP jobs, the `wm` option should be used.
- Similarly, for the `get-log-levels`, `set-log-levels`, and `get-logs` commands, the `dp` component only affects the DP CLI logs, and not Data Processing logs. Information on administrating Data Processing logs is available in the *Data Processing Guide*.
- `$BDD_HOME` contains several new directories, including `/opss`, `/opss_standalone`, and `/workflowmanager`. For more information, see the *Installation Guide*.
- The configuration files related to Data Processing have changed:
  - The DP CLI is configured by `edp-cli.properties`. In this release, this file is renamed from `edp.properties` which was its name in the previous release. It is located under `$BDD_HOME/dataprocessing/edp_cli/config`.
  - The Workflow Manager Service is configured by `edp.properties`, under `$BDD_HOME/workflowmanager/dp/config`.

In this release, for the most part, the data set's workflow configuration is controlled by `edp.properties`. However, there are a few exceptions, where you should modify settings in `edp-cli.properties`, or in Studio. For a list of these exceptions, see "What's New and Changed in this Release" in the *Getting Started Guide*.

For a complete lists of configuration settings for `edp-cli.properties`, and `edp.properties`, see the *Data Processing Guide*.

If you upgraded from 1.2.x or lower the following changes were also made:

- Your Dgraph database (index) was upgraded to the new format and `/DGRAPH_INDEX_DIR` now includes a separate database directory for each project. For more information, see the *Administrator's Guide*.
- Your sample files were upgraded to the new format and moved from `$edpDataDir/.collectionData/<collectionName>` to `$edpDataDir/.collectionData/<databaseName>.<collectionName>`.
- The list of supported Dgraph flags has changed in this release—some flags were renamed and a few were removed. Any that you had added to the `DGRAPH_ADDITIONAL_ARG` property in `bdd.conf` were updated accordingly. For more information, see the *Administrator's Guide*.

## Post-upgrade configuration

If you run into any performance issues after you upgrade, you might need to adjust the configuration of your BDD or Hadoop cluster.

Refer to the *Installation Guide* for more information on the changes you should make.

Note: If you have made any custom changes to your WLS domain configuration, they will have to be re-applied after the upgrade, since they will not be preserved by the upgrade process.

## Clearing the browser cache

After an upgrade, all BDD users should clear their browser caches before logging in to Studio. This ensures they'll be able to open their projects successfully.