

Oracle® Big Data Discovery

Upgrade Guide

Version 1.1.1 • Revision B • December 2015

Copyright and disclaimer

Copyright © 2003, 2015, 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

Copyright and disclaimer	2
Preface	4
About this guide	4
Audience	4
Conventions	4
Contacting Oracle Customer Support	5
Chapter 1: What's Changed in this Release	6
New and updated features	6
Unsupported features	9
Chapter 2: Before You Upgrade	10
Supported upgrade paths	10
Hadoop cluster requirements	10
Downloading the upgrade packages	10
Obtaining the Hadoop client libraries	11
Running the upgrade hotfixes	12
Running the BDD 1.0 hotfix	12
Running the BDD 1.1.0 hotfix	13
Backing up your current cluster	13
Chapter 3: Upgrading Big Data Discovery	15
Overview	15
Silent upgrade	16
Merging the configuration file	17
Editing the configuration file	17
Upgrading the cluster	19
Troubleshooting a failed upgrade	19
Data Processing fails to start	20
Rolling back a failed upgrade	20
Restoring your 1.0 cluster	21
Chapter 4: After You Upgrade	22
Verifying the upgrade	22
Post-upgrade configuration	22
Clearing browser cache	23
Rewriting custom transformations	23

Preface

Oracle Big Data Discovery is a set of end-to-end visual analytic capabilities that leverage the power of Hadoop to transform raw data into business insight in minutes, without the need to learn complex products or rely only on highly skilled resources.

About this guide

This guide helps you upgrade your Oracle Big Data Discovery cluster and describes the major changes included in the new release.

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
User Interface Elements	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.

Symbol conventions

The following table describes symbol conventions used in this document.

Symbol	Description	Example	Meaning
>	The right angle bracket, or greater-than sign, indicates menu item selections in a graphic user interface.	File > New > Project	From the File menu, choose New, then from the New submenu, choose Project.

Path variable conventions

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

Path variable	Meaning
\$MW_HOME	Indicates the absolute path to your Oracle Middleware home directory, which is the root directory for your WebLogic installation.
\$DOMAIN_HOME	Indicates the absolute path to your WebLogic domain home directory. For example, if <code>bdd_domain</code> is the domain name, then the <code>\$DOMAIN_HOME</code> value is the <code>\$MW_HOME/user_projects/domains/bdd_domain</code> directory.
\$BDD_HOME	Indicates the absolute path to your Oracle Big Data Discovery home directory. For example, if <code>BDD1.1</code> is the name you specified for the Oracle Big Data Discovery installation, then the <code>\$BDD_HOME</code> value is the <code>\$MW_HOME/BDD1.1</code> directory.
\$DGRAPH_HOME	Indicates the absolute path to your Dgraph home directory. For example, the <code>\$DGRAPH_HOME</code> value might be the <code>\$BDD_HOME/dgraph</code> directory.

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

What's Changed in this Release

This section describes the changes made for this release of BDD, including new, deprecated, and unsupported features.

[New and updated features](#)

[Unsupported features](#)

New and updated features

The following features have been added, improved, or updated for the Oracle Big Data Discovery 1.1 release.

- [Hadoop-related features on page 6](#)
- [Data set loading and freshness, visualizations, transformations, and security on page 7](#)
- [Integration and operations on page 8](#)
- [Documentation on page 8](#)

Hadoop-related features

This release includes the following Hadoop-related features.

Feature	Description
Hortonworks Data Platform (HDP) support	BDD now supports the two most widely-used Hadoop distributions: Hortonworks Data Platform and Cloudera Distribution for Hadoop. For more information, see the <i>Installation and Deployment Guide</i> .
Support for multiple versions of Cloudera Distribution for Hadoop (CDH)	CDH support for BDD has been expanded to include multiple versions of CDH. BDD can be now installed on a wider range of systems and updated to incorporate improvements to the evolving Hadoop ecosystem. For more information, see the <i>Installation and Deployment Guide</i> .
Kerberos support	BDD now runs on Hadoop clusters secured by Kerberos authentication, allowing secure analysis of big data across whole organizations. For more information, see the <i>Installation and Deployment Guide</i> .
Spark on YARN support	BDD now uses the YARN resource scheduler, enabling it to run on production clusters that have many other Hadoop applications. For more information, see the <i>Data Processing Guide</i> .

Data set loading and freshness, visualizations, transformations, and security

This release includes the following changes to data set loading and freshness, visualizations, transformations, and security.

Feature	Description
Enhanced options for initial data loading	A new self-service data load experience makes it even easier for business analysts to build innovative data combinations that bring together data from Hadoop and their own desktops. For more information, see the <i>Data Exploration and Analysis Guide</i> .
Data set and project level security in Studio	Access to individual data sets and projects can now be restricted for specific users and user groups, allowing BDD to be rolled out to large groups of analysts securely. For more information, see the <i>Data Exploration and Analysis Guide</i> .
Ability to update and reload data	Data sets can now be updated from within Studio both manually and automatically, making it easy to keep projects and dashboards in synch with rapidly changing data. For more information, see the <i>Getting Started Guide</i> .
BDD applications	New BDD applications are projects that periodically receive new or refreshed data, and can be shared with a group of users to enable ongoing interactive analysis by a wider team. For more information, see the <i>Getting Started Guide</i> .
Ability to publish and share transformation scripts	Transformation scripts can now be published and shared, accelerating the process of data cleansing. For more information, see the <i>Data Exploration and Analysis Guide</i> .
Transform enhancements	<p>Studio's Transform page includes several enhancements:</p> <ul style="list-style-type: none"> • New support for Geotagger, Whitelist, and Entity Extraction transformations make it easier to derive structured data from messy Hadoop sources. Users can extract entities, terms, and locations from their data without writing code. • A redesigned Transform Editor provides even more help to business analysts who need to go beyond one-click transformations. • Improved performance tightens the cycle of improving data quality in Transform and analyzing the results in Explore and Discover. <p>For more information, see the <i>Data Exploration and Analysis Guide</i>.</p>
Improved data visualizations	Data visualizations have been standardized and enhanced to make them easier to interpret and interact with. Additionally, a refreshed configuration experience makes them faster to create and offers users more control. For more information, see the <i>Data Exploration and Analysis Guide</i> .
Custom visualization support	Studio's new Custom Visualization Component makes it easy for business users to build and deploy their own interactive data visualizations using JavaScript and EQL. For more information, see the <i>Extensions Guide</i> .

Integration and operations

This release includes the following changes related to integration and operations.

Feature	Description
Upgrade support	BDD now supports in-place upgrades to new releases, enabling administrators to quickly update their BDD clusters and user data to the latest version. Afterwards, users can go back to their projects with little to no reconfiguration. For information on upgrading, see Overview on page 15 .
JDBC connectivity support	The new JDBC connector in BDD allows IT to securely set up connections to gold-standard enterprise databases. Business users can then import data sets from JDBC and use them alongside data originating in Hadoop. For more information, see the <i>Data Exploration and Analysis Guide</i> .
Updated administration capabilities	The <code>bdd-admin</code> script has a number of new capabilities that give administrators more options for managing BDD. For more information, see the <i>Administrator's Guide</i> .
Enhanced language support	Language processing capabilities in BDD have been expanded to support over 80 languages. For more information, see the <i>Data Processing Guide</i> .
Updated logs	Many logs in BDD contain new properties that give administrators more control over their behavior and the information they provide. For more information on logging, see the <i>Administrator's Guide</i> .
Enhanced Enterprise Manager integration	The Enterprise Manager plug-in now includes new metrics and operations that make it easier for administrators to monitor and manage their BDD clusters. For more information on using the Enterprise Manager plug-in, see the <i>Administrator's Guide</i> .
Improved performance, stability, and reliability	BDD 1.1 includes fixes that improve its performance, reliability, and scalability.

Documentation

This release includes the following documentation changes.

Feature	Description
New guides	The BDD documentation set includes two new guides: <ul style="list-style-type: none"> The <i>Getting Started Guide</i> introduces BDD and provides an overview of how to use it. The <i>Upgrade Guide</i> (this guide) describes how to upgrade a BDD cluster to the latest version of BDD.

Feature	Description
Consolidation of transform documentation	The documentation related to transforms has been consolidated in the <i>Data Exploration and Analysis Guide</i> . Previously, this information was in the <i>Data Exploration and Analysis Guide</i> and the <i>Extensions Guide</i> .

Unsupported features

The following features are no longer supported in this release.

Feature	Description
Spark (Standalone)	BDD no longer supports Spark (Standalone). All data processing is now handled by Spark on YARN. For more information, see the <i>Data Processing Guide</i> .
Oozie	Because support for Spark (Standalone) has been removed, BDD no longer requires Oozie.
Index merging	BDD no longer supports index merging.



Chapter 2

Before You Upgrade

Before you can upgrade BDD, you must download the required software packages, run the hotfix script, and back up your current cluster.

[Supported upgrade paths](#)

[Hadoop cluster requirements](#)

[Downloading the upgrade packages](#)

[Obtaining the Hadoop client libraries](#)

[Running the upgrade hotfixes](#)

[Backing up your current cluster](#)

Supported upgrade paths

You can use the current upgrade script to upgrade your cluster from 1.0 to 1.1.x.

Hadoop cluster requirements

Before you upgrade your cluster, you should verify that both Spark (Standalone) and Spark on YARN are both installed on your Data Processing nodes. Additionally, your Hadoop cluster must be running throughout the upgrade process.

You shouldn't make any other changes to your Hadoop cluster. Although BDD 1.1 supports integration with Kerberos and multiple versions of CDH, you can't implement these until the upgrade is complete. Instructions for enabling these are available in the *Administrator's Guide*.



Note: You can't switch to HDP after you upgrade. This requires a fresh installation. For more information, see the *Installation and Deployment Guide*.

Downloading the upgrade packages

First, you must download the BDD upgrade packages and put them in a single directory (the upgrade source directory) on the Admin Server. 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.
 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 1.1.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**.

This downloads the following packages 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 (12.1.3.0.0) WebLogic Server and Coherence**

You should 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 the `/packages` directory within the upgrade source directory.
13. Rename the first BDD binary package `bdd1.zip` and the second `bdd2.zip`.

This ensures the upgrade script will recognize them.

14. Unzip the WebLogic Server package.

This creates a file called `fmw_12.1.3.0.0_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 should obtain the Hadoop client libraries.

Obtaining the Hadoop client libraries

Next, you must 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, which is distributed to all BDD nodes.

Before upgrading BDD, you must obtain these libraries and move them to the Admin Server. The location you put them in is arbitrary, as you will define the path to each in `bdd.conf`.

Download the following files from <http://archive-primary.cloudera.com/cdh5/cdh/5/> to the Admin Server and unpack 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`

Next, you can apply the upgrade hotfix.

Running the upgrade hotfixes

Next, you must run one of the BDD hotfixes to ensure your cluster can be backed up and restored properly. BDD includes two hotfixes; the one you run depends on your current version of BDD.

[Running the BDD 1.0 hotfix](#)

[Running the BDD 1.1.0 hotfix](#)

Running the BDD 1.0 hotfix

If you're upgrading from BDD 1.0 to 1.1.x, you must run `hotfix_EADMIN-850.sh`. This adds two scripts to `$BDD_HOME/BDD_manager/bin`:

- `bdd-backup.sh`, which backs up your cluster data and metadata to a single tar file.
- `bdd-restore.sh`, which restores your cluster from backup.

Both scripts are recommended for the upgrade procedure: you should back up your cluster before upgrading so that you can restore it if the upgrade fails.

To run the BDD 1.0 hotfix:

1. On the Admin Server, go to `<upgrade_source_dir>/installer/hotfix/hotfix_EADMIN-850`.
2. Run the hotfix script:

```
hotfix_EADMIN-850.sh <path/to/bdd.conf>
```

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

When the script finishes, the `$BDD_HOME/BDD_manager/bin` directory will contain the `bdd-backup.sh` and `bdd-restore.sh` scripts.

Running the BDD 1.1.0 hotfix

If you're upgrading from BDD 1.1.0 to 1.1.x, you must run `hotfix_EADMIN-1279.sh`. This fixes an issue with your current backup and restore scripts that could either cause the upgrade to fail or create issues in your cluster after you upgrade.

When the hotfix runs, it updates your backup and restore scripts with the following:

- `bdd-backup1_1_0.sh`
- `bdd-restore1_1_0.sh`



Note: The hotfix only updates the contents of your current backup and restore scripts; their names remain the same.

To run the BDD 1.1.0 hotfix:

1. On the Admin Server, go to `<upgrade_source_dir>/BDD1.1.x/installer/hotfix/hotfix_EADMIN-1279`.
2. Run the hotfix script:

```
./hotfix_EADMIN-1279.sh <path/to/bdd.conf>
```

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

Backing up your current cluster

Next, you must back up your current cluster by running the `bdd-backup.sh` script. This backs up the following to a single tar file:

- Configuration files
- Studio database
- Schema and data for Hive tables created in Studio
- Dgraph indexes
- Sample files in HDFS

If your upgrade fails, you can use this file to restore your current cluster.

Before you run the backup script, verify the following:

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

To back up your cluster:

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

2. Stop your cluster.

- For BDD 1.0, run:

```
./bdd-admin.sh stop --all
```

- For BDD 1.1.0, run:

```
./bdd-admin.sh stop
```

3. Run the backup script.

- For BDD 1.0, run:

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

- For BDD 1.1.0, run:

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

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.

When the script completes, your current cluster will be backed up to the specified file.



Chapter 3

Upgrading Big Data Discovery

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

[Overview](#)

[Merging the configuration file](#)

[Editing the configuration file](#)

[Upgrading the cluster](#)

[Troubleshooting a failed upgrade](#)

[Rolling back a failed upgrade](#)

[Restoring your 1.0 cluster](#)

Overview

You upgrade your cluster to BDD 1.1 by running two separate scripts.

The first script merges your current `bdd.conf` file with the version from the new release. This ensures that your upgraded cluster will retain most of your current configuration settings. Note that you will still need to manually edit the new properties that were added in this release, as well as any properties the script was unable to merge.

The second script upgrades your cluster. When this script runs, it:

1. Deletes all transient data from your cluster.
2. Deploys the new version of BDD:
 - Deploys the Data Processing libraries and DP CLI.
 - Deploys the OLT.
 - Creates a new WebLogic Server domain for Studio and the Dgraph Gateway.
 - Creates a new JDBC pool within the domain.
 - Deploys the Dgraph HDFS Agent.
3. Merges the current configuration files for all BDD components with their counterparts from the new release.
4. Upgrades the format of your Dgraph index.
5. Upgrades your BDD cluster's metadata, including the DataSet Inventory used by Data Processing and Studio, as well as Studio's views and projects.
6. Rebuilds the spelling index.

7. Removes the **All** search interface.
8. Upgrades the BDD sample files.

\$BDD_HOME

The upgrade script installs the new version of BDD in `BDD_<version number>`. This is the new `$BDD_HOME` directory. The script doesn't remove the previous version's `$BDD_HOME` directory. You can delete this if you want, but it is not required.



Important: If you decide to uninstall BDD, the uninstallation script will delete the contents of all versions of `$BDD_HOME`, not just the most recent. If there is anything in an older version's `$BDD_HOME` directory that you want to keep, you should back it up to a different location before uninstalling. See the *Installation and Deployment Guide* for more information.

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, it will obtain that information from environment variables you set beforehand.

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

- The username and password for your Hadoop cluster manager UI (Cloudera Manager or Ambari).
- The username and password for the WebLogic Server admin.
- The username and password for the Studio database.

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

Environment variable	Description
<code>BDD_HADOOP_UI_USERNAME</code>	The username for Cloudera Manager/Ambari.
<code>BDD_HADOOP_UI_PASSWORD</code>	The password for Cloudera Manager/Ambari.
<code>BDD_WLS_USERNAME</code>	The username for the WebLogic Server admin.
<code>BDD_WLS_PASSWORD</code>	The password for the WebLogic Server admin.
<code>BDD_STUDIO_JDBC_USERNAME</code>	The username for the Studio database.
<code>BDD_STUDIO_JDBC_PASSWORD</code>	The password for the Studio database.

Merging the configuration file

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

You merge `bdd.conf` by running the `merge-bddconf.sh` script from the upgrade source directory on the Admin Server. When the script runs, it populates the new version of `bdd.conf` with the settings from your current version.

Although the merged file will contain most of your current settings, you'll need to manually edit it to set any properties the script couldn't merge, as well as those that have been added for the new release. For more information, see [Editing the configuration file on page 17](#).

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 $BDD_HOME/BDD_manager/conf/bdd.conf bdd.conf
```

The first argument in the command above is the absolute path to the current version of `bdd.conf`, and the second is the path to the new one.

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

Editing the configuration file

After you merge `bdd.conf`, you need to 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 updating the file:

- The accepted values for some properties are case-sensitive and must be entered exactly as they appear in the table.
- 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 must have 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.

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

Property	Description
UPGRADE	Determines whether you're performing an upgrade. This should be set to <code>TRUE</code> .
BDD_OLD_CONFIG	The absolute path to your current version of <code>bdd.conf</code> . This will already be populated.

Property	Description
INSTALLER_PATH	The absolute path to the BDD software packages in the upgrade source directory.
HADOOP_CLIENT_LIB_PATHS	<p>A comma-separated list of the absolute paths to the Hadoop client libraries on the install machine. The orchestration script will use these libraries to generate the Hadoop fat jar.</p> <p>The comments above this property contain templates for the value you should provide. The one you use depends on your Hadoop distribution.</p> <p>To set this property:</p> <ul style="list-style-type: none"> • If you're installing on a CDH cluster, copy and paste the value from the first template to the actual property and replace each instance of UNZIPPED_<COMPONENT>_BASE with the absolute path to the component's client library. • If you're installing on an HDP cluster, copy and paste the value from the second template to the actual property and update the paths to point to the libraries you copied to the install machine. Don't change the order of the paths in this list, as they <i>must</i> be specified in the order they appear in. <p>Be sure to include all of the required libraries. If this list is incomplete, the orchestration script will fail.</p> <p>For more information on the client libraries and how to obtain them, see Obtaining the Hadoop client libraries on page 11.</p>
STUDIO_JDBC_URL	<p>The JDBC URL for your database.</p> <p>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.</p>
YARN_QUEUE	The YARN queue that Data Processing jobs are submitted to.
SPARK_ON_YARN_JAR	The absolute path to the Spark on YARN jar on Hadoop nodes. This will be added to the CLI classpath.
FORCE	Determines whether the upgrade script will remove files and directories left over from previous installations when it runs. This can be set to TRUE or FALSE.
WLS_NO_SWAP	<p>Controls whether the upgrade script will check for the required amount of free swap space (512MB) on the Admin Server and all Managed Servers before upgrading WebLogic Server.</p> <p>If set to TRUE, the script won't check the amount of swap space. Use this value if you're running WebLogic Server on nodes that don't meet the swap space requirement.</p>

Property	Description
DP_ADDITIONAL_JARS	<p>Optional. A comma-separated list of the absolute paths to custom SerDe jars you want to use during data processing. These will be added to the CLI classpath.</p> <p>Note that you must manually copy each SerDe jar to the same location on all cluster nodes before upgrading.</p>

Upgrading the cluster

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

Before you run the upgrade script, you should verify that:

- You downloaded the required BDD upgrade packages and Hadoop client libraries to the Admin Server
- You ran the upgrade hotfix script
- You backed up your cluster
- You merged and edited `bdd.conf`
- Perl is installed on the Admin Server
- Python is installed on all Dgraph nodes
- Your Hadoop cluster is running
- Spark on YARN is running

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 username and password for your Hadoop cluster manager (Cloudera Manager or Ambari).
 - The username and password for the WebLogic Server admin.
 - The username and password for the Studio 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`.

[Data Processing fails to start](#)

Data Processing fails to start

Data Processing may be unable to start on production clusters that have been upgraded more than once.

If this occurs, you'll receive an error similar to the following:

```
[Upgrader] Switch soft link Success
[Upgrader] Starting BDD cluster...
[2015/10/08 01:07:45 -0400] [Admin Server] Starting BDD components...
[2015/10/08 01:13:17 -0400] [web009.us.example.com] Starting Dgraph.....Success!
[2015/10/08 01:13:22 -0400] [web009.us.example.com] Starting Dgraph HDFS Agent.....Success!
[2015/10/08 01:15:42 -0400] [web009.us.example.com] Starting Dgraph Gateway.....Success!
[2015/10/08 01:16:30 -0400] [web009.us.example.com] Starting Data Processing.....Fail!
```

To fix this issue, edit the following properties in `edp.properties` on all Managed Servers and Dgraph nodes:

```
oltHome=/opt/bdd/edp-<version>/olt
bddHadoopFatJar=/localdisk/Oracle/Middleware/BDD-<version>/common/hadoop/lib/bddHadoopFatJar.jar
edpJarDir=/opt/bdd/edp-<version>/lib
```

where `<version>` is the version of BDD you upgraded to, as defined by `BDD_VERSION` in `bdd.conf`.

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 the upgrade script made to your cluster. It also restores your old cluster from backup, starts it, and performs a healthcheck.



Note: The rollback script removes the logs created by the upgrade script, so be sure to check them beforehand to determine why the upgrade failed.

When the script finishes running, your cluster will be back to the state it was in before you ran the upgrade and will be running. You can then either rerun the upgrade process or restore your old cluster.

If you decide to rerun the upgrade process, you must rerun the merge script, as the merged version of `bdd.conf` wasn't included in the backup.

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.

Restoring your 1.0 cluster

If you want to restore your 1.0 cluster after a failed upgrade, you can do so by running the restore script that was added by the hotfix.

Before you restore your cluster, you must rollback the upgrade. For more information, see [Rolling back a failed upgrade on page 20](#).

You must provide the restore script with the absolute path to the backup of your 1.0 cluster. When the script runs, it restores your old index, database, sample files, and configuration from backup.

To restore your 1.0 cluster:

1. On the Admin Server, go to `$BDD_HOME/BDD_manager/bin`.
2. Run the restore script:

```
./bdd-restore.sh <path>
```

Where `<path>` is the absolute path to the tar file you backed up your 1.0 cluster to.



Chapter 4

After You Upgrade

This section describes the changes the upgrade script made to your cluster, as well as BDD's new and deprecated features.

[Verifying the upgrade](#)

[Post-upgrade configuration](#)

[Clearing browser cache](#)

[Rewriting custom transformations](#)

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:

```
[2015/06/19 04:18:55 -0700] [Admin Server] Checking health of BDD cluster...
[2015/06/19 04:20:39 -0700] [web009.us.example.com] Check BDD functionality.....Pass!
[2015/06
/19 04:20:39 -0700] [web009.us.example.com] Check Hive Data Detector health.....Hive Data
Detector has previously run
[2015/06/19 04:20:39 -0700] [Admin Server] Successfully checked statuses.
```

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.

Please refer to the *Installation and Deployment Guide* for more information on the changes you should make.

Clearing 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.

Rewriting custom transformations

Many of BDD's transformation functions have been refactored in this release; however, the upgrade script doesn't update your custom transformation scripts accordingly. Because of this, you need to update your scripts manually before you can use them in your projects.

You can edit your transformation scripts in Studio's **Transform** component. For more information, see the *Data Exploration and Analysis Guide*.

The following table lists the changes made to the functions.

Function in release 1.0.0	Changes in release 1.1.x
geotagIPAddressGetCity geotagIPAddressGetCountry geotagIPAddressGetPostCode geotagIPAddressGetRegion geotagIPAddressGetRegionID geotagIPAddressGetSubRegion geotagIPAddressGetSubRegionID	Replaced by the <code>geotagIPAddress</code> function.
geotagAddressGetCity geotagAddressGetCountry geotagAddressGetPostcode geotagAddressGetRegion geotagAddressGetSubRegion geotagAddressGetRegionID geotagAddressGetSubRegionID	Replaced by the <code>geotagUnstructuredAddress</code> function.
getLocationEntities getOrganizationEntities getPersonEntities	Replaced by the <code>getEntities</code> function.

Index

B

- backup script 13
- browser cache, clearing 23

C

- configuration file
 - editing 17
 - merging 17
- custom transformations, rewriting 23

F

- failed upgrade
 - restoring cluster 21
 - rollback 20
 - troubleshooting 19

H

- Hadoop
 - client libraries, obtaining 11
 - cluster requirements 10
- hotfixes
 - 1.0 script 12

- 1.1.0 script 13

P

- post-upgrade configuration 22

R

- release overview
 - new and updated features 6
 - unsupported features 9

T

- troubleshooting when Data Processing fails to start 20

U

- upgrade
 - overview 15
 - silent mode 16
 - supported paths 10
 - upgrade packages, downloading 10
 - upgrade script 19
 - verifying 22