

webMethods Adapter for HDFS Installation and User's Guide

Version 9.8

February 2018

This document applies to webMethods Adapter for HDFS 9.8 and to all subsequent releases.

Specifications contained herein are subject to change and these changes will be reported in subsequent release notes or new editions.

Copyright © 2016-2021 Software AG, Darmstadt, Germany and/or Software AG USA, Inc., Reston, VA, USA, and/or its subsidiaries and/or its affiliates and/or their licensors.

The name Software AG and all Software AG product names are either trademarks or registered trademarks of Software AG and/or Software AG USA Inc. and/or its subsidiaries and/or its affiliates and/or their licensors. Other company and product names mentioned herein may be trademarks of their respective owners.

Detailed information on trademarks and patents owned by Software AG and/or its subsidiaries is located at <https://softwareag.com/licenses/>.

Use of this software is subject to adherence to Software AG's licensing conditions and terms. These terms are part of the product documentation, located at <https://softwareag.com/licenses/> and/or in the root installation directory of the licensed product(s).

This software may include portions of third-party products. For third-party copyright notices, license terms, additional rights or restrictions, please refer to "License Texts, Copyright Notices and Disclaimers of Third Party Products". For certain specific third-party license restrictions, please refer to section E of the Legal Notices available under "License Terms and Conditions for Use of Software AG Products / Copyright and Trademark Notices of Software AG Products". These documents are part of the product documentation, located at <https://softwareag.com/licenses/> and/or in the root installation directory of the licensed product(s).

Document ID: ADAPTER-WHD-IUG-98-20210326

Table of Contents

About this Guide.....	5
Document Conventions.....	6
Online Information and Support.....	7
Data Protection.....	7
 1 Overview of webMethods Adapter for HDFS.....	 9
About the adapter.....	10
Architecture Overview.....	10
 2 Installing and Uninstalling webMethods Adapter for HDFS.....	 11
Overview.....	12
Requirements.....	12
The Integration Server Home Directory.....	12
Installing Adapter for HDFS.....	12
Uninstalling Adapter for HDFS.....	13
 3 Package Management.....	 15
Overview.....	16
Adapter for HDFS Package Management.....	16
 4 Managing Adapter for HDFS Connections.....	 19
Overview.....	20
Before Configuring or Managing Adapter Connections.....	20
Configuring Adapter Connections.....	21
Testing Adapter Connections.....	26
Viewing Adapter Connection Parameters.....	26
Editing Adapter Connections.....	27
Deleting Adapter Connections.....	27
Enabling Adapter Connections.....	28
Disabling Adapter Connections.....	28
 5 Adapter Logging.....	 31
Overview.....	32
Adapter Logging Levels.....	32
Adapter for HDFS Message Logging.....	33
 A Built-In Services.....	 35
Overview.....	36
wm.hadoop.hdfs:open.....	36
wm.hadoop.hdfs:read.....	37
wm.hadoop.hdfs:write.....	39
wm.hadoop.hdfs:close.....	40

About this Guide

■ Document Conventions	6
■ Online Information and Support	7
■ Data Protection	7

This guide describes how to install, configure, and use the webMethods Adapter for HDFS. This guide also describes the built-in services provided by the Adapter for HDFS. It contains information for administrators and application developers who want to exchange data with HDFS clusters.

To use this guide effectively, you should be familiar with:

- The basic concepts and tasks for working with HDFS
- The terminology and the basic operations of your operating system
- The setup and operation of webMethods Integration Server
- How to perform basic tasks with Software AG Designer

Document Conventions

Convention	Description
Bold	Identifies elements on a screen.
Narrowfont	Identifies service names and locations in the format <i>folder.subfolder.service</i> , APIs, Java classes, methods, properties.
<i>Italic</i>	Identifies: Variables for which you must supply values specific to your own situation or environment. New terms the first time they occur in the text. References to other documentation sources.
Monospace font	Identifies: Text you must type in. Messages displayed by the system. Program code.
{ }	Indicates a set of choices from which you must choose one. Type only the information inside the curly braces. Do not type the { } symbols.
	Separates two mutually exclusive choices in a syntax line. Type one of these choices. Do not type the symbol.
[]	Indicates one or more options. Type only the information inside the square brackets. Do not type the [] symbols.
...	Indicates that you can type multiple options of the same type. Type only the information. Do not type the ellipsis (...).

Online Information and Support

Software AG Documentation Website

You can find documentation on the Software AG Documentation website at <http://documentation.softwareag.com>.

Software AG Empower Product Support Website

If you do not yet have an account for Empower, send an email to empower@softwareag.com with your name, company, and company email address and request an account.

Once you have an account, you can open Support Incidents online via the eService section of Empower at <https://empower.softwareag.com/>.

You can find product information on the Software AG Empower Product Support website at <https://empower.softwareag.com>.

To submit feature/enhancement requests, get information about product availability, and download products, go to [Products](#).

To get information about fixes and to read early warnings, technical papers, and knowledge base articles, go to the [Knowledge Center](#).

If you have any questions, you can find a local or toll-free number for your country in our Global Support Contact Directory at https://empower.softwareag.com/public_directory.aspx and give us a call.

Software AG TECHcommunity

You can find documentation and other technical information on the Software AG TECHcommunity website at <http://techcommunity.softwareag.com>. You can:

- Access product documentation, if you have TECHcommunity credentials. If you do not, you will need to register and specify "Documentation" as an area of interest.
- Access articles, code samples, demos, and tutorials.
- Use the online discussion forums, moderated by Software AG professionals, to ask questions, discuss best practices, and learn how other customers are using Software AG technology.
- Link to external websites that discuss open standards and web technology.

Data Protection

Software AG products provide functionality with respect to processing of personal data according to the EU General Data Protection Regulation (GDPR). Where applicable, appropriate steps are documented in the respective administration documentation.

1 Overview of webMethods Adapter for HDFS

■ About the adapter	10
■ Architecture Overview	10

About the adapter

webMethods Adapter for HDFS is an add-on to webMethods Integration Server that enables you to exchange data with an HDFS cluster. The adapter provides seamless and real-time communication with the HDFS cluster without requiring changes to your existing application infrastructure.

By using Adapter for HDFS, clients can create and run services that execute programs to read data from and write data to HDFS clusters. For example, you can use Adapter for HDFS to write a customer record to an HDFS data warehouse that stores and processes millions of customer records.

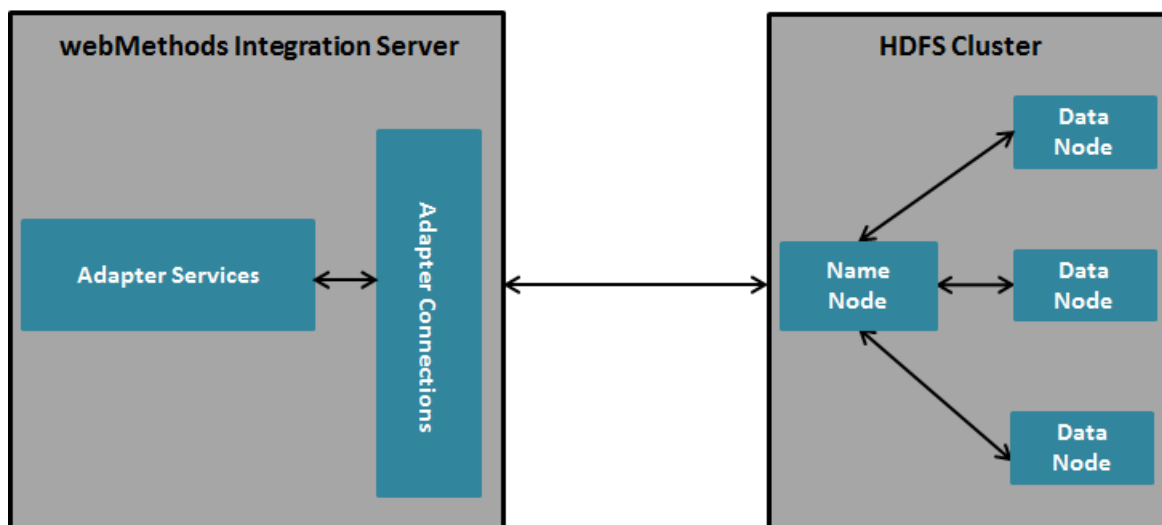
Architecture Overview

Adapter for HDFS provides a set of user interfaces and services that enable Integration Server to connect to, and perform operations on the HDFS cluster. The adapter is provided as a single package that must be installed on Integration Server.

Adapter for HDFS enables you to configure the following components:

- **Adapter Connections:** Enable Integration Server to connect to the HDFS cluster at run time. You must configure an adapter connection before you can use adapter services. For a detailed description of adapter connections, see [“Managing Adapter for HDFS Connections” on page 19](#).
- **Adapter Services:** Perform operations on the HDFS cluster using the built-in services provided with the Integration Server. You can use services to read data from and write data to HDFS clusters. For a detailed descriptions of adapter services, see [“Built-In Services” on page 35](#).

The following diagram illustrates how Adapter for HDFS uses these components to interact with the HDFS cluster.



2 Installing and Uninstalling webMethods Adapter for HDFS

■ Overview	12
■ Requirements	12
■ The Integration Server Home Directory	12
■ Installing Adapter for HDFS	12
■ Uninstalling Adapter for HDFS	13

Overview

You must use Software AG Installer for installing webMethods Adapter for HDFS. If you want to uninstall the adapter, use the Software AG Uninstaller.

For complete information about the wizards or other installation methods, or to install other webMethods products, see *Installing webMethods Products On Premises* for your release.

Requirements

For a list of operating systems, RDBMSes, and webMethods products supported by Adapter for HDFS, see *webMethods Adapters System Requirements*.

Adapter for HDFS has no hardware requirements beyond those of its host Integration Server.

The Integration Server Home Directory

You can create one or more Integration Server instances under a single installation directory. Each Integration Server instance has a home directory under *Integration Server_directory* \instances*instance_name* that contains the packages, configuration files, log files, and updates for the instance.

Note:

For more information about running multiple Integration Server instances, see the *webMethods Integration Server Administrator's Guide* for your release.

Installing Adapter for HDFS

If you are installing Adapter for HDFS on an existing Integration Server, you must shut down the Integration Server before you begin the installation process.

➤ **To install Adapter for HDFS**

1. Download Installer from the [Empower Product Support website](#).
2. Start the Installer wizard.
3. Choose the webMethods release that includes the Integration Server on which you want to install the adapter. For example, if you want to install the adapter on 9.8, choose the 9.8 release.
4. Specify the installation directory as follows:
 - If you are installing on an existing Integration Server, specify the webMethods installation directory that contains the host Integration Server.
 - If you are installing both the host Integration Server and the adapter, specify the installation directory to use.

5. In the product selection list, select **Adapters > webMethods Adapter 9.8 for HDFS**.

You can choose to install the package in the default instance. In this case, Software AG Installer installs the adapter in both locations, *Integration Server_directory* \packages and the default instance packages directory located in *Integration Server_directory* \instances\default\packages.

6. To download the documentation for the adapter, go to [Software AG Documentation website](#).
7. After the installation completes, close the Installer and start the host Integration Server.

Uninstalling Adapter for HDFS

1. Shut down the host Integration Server.

You do not need to shut down any other webMethods products or applications that are running on your computer.

2. Start Software AG Uninstaller, selecting the webMethods installation directory that contains the host Integration Server.
3. In the product selection list, select **Adapters > webMethods Adapter 9.8 for HDFS**.

You can also choose to uninstall documentation.

4. After Uninstaller completes, restart the host Integration Server.

Uninstaller removes the installed files related to Adapter for HDFS. However, Uninstaller does not delete the files created after you installed the adapter (for example, user-created or configuration files) nor the adapter directory structure. Therefore, you can delete the WmHDFSAdapter directory from the *Integration Server_directory* \packages directory and *Integration Server_directory* \instances\default\packages directory.

3 Package Management

■ Overview	16
■ Adapter for HDFS Package Management	16

Overview

Adapter for HDFS is available as a package called WmHDFSAdapter. You can manage the WmHDFSAdapter package as you would manage any package on Integration Server.

Adapter for HDFS Package Management

When you create adapter connections and services, define them in user-defined packages for easier package management.

As you create user-defined packages in which to store connections and adapter services, use the package management functionality provided in Software AG Designer and set the user-defined packages to have a dependency on the WmHDFSAdapter package. Therefore, when the WmHDFSAdapter package loads or reloads, the user-defined packages load automatically.

Package management tasks include the following:

- [Setting package dependencies \(see “Package Dependency Requirements and Guidelines” on page 16\)](#)
- [“Enabling Packages” on page 17](#)
- [“Disabling Packages” on page 17](#)
- [“Importing and Exporting Packages” on page 18](#)
- [“Group Access Control” on page 18](#)

Package Dependency Requirements and Guidelines

You must follow certain package dependency guidelines for effectively managing the adapter packages.

- A user-defined package must have a dependency on its associated adapter package, WmHDFSAdapter.
- Package dependencies ensure that Integration Server automatically loads or reloads all packages at startup in the proper order: the adapter package first, and the user-defined packages next.
- If the connections and services of an adapter are defined in different packages, then:
 - A package that contains the connections must have a dependency on the WmHDFSAdapter package.
 - Packages that contain adapter services must have a dependency on their associated connection package.
- Integration Server does not allow you to enable a package if it has a dependency on another package that is disabled. That is, before you can enable your package, you must enable all packages on which your package depends. For information about enabling packages, see [“Enabling Packages” on page 17](#).

- Integration Server does not allow you to disable a package if its dependent packages are enabled. Therefore, you must manually disable any user-defined packages that have a dependency on the adapter package before you disable the adapter package. For information about disabling packages, see [“Disabling Packages” on page 17](#).
- You can provide the same name to connections, adapter services, and notifications only if they are in different folders and packages.

Enabling Packages

All packages are automatically enabled by default. Use the following procedure when you want to enable a package that was previously disabled.

➤ To enable a package

1. Open the Integration Server Administrator if it is not already open.
2. In the **Packages** menu of the navigation area, click **Management**.
3. Click **No** in the **Enabled** column. The Integration Server Administrator displays a ✓ symbol and **Yes** in the **Enabled** column.

Note:

Enabling an adapter package does not cause its associated user-defined packages to be reloaded. For information about reloading packages, see the *webMethods Service Development Help* for your release.

Important:

Before you manually enable a user-defined package, you must first enable its associated adapter package (WmHDFSAdapter).

Disabling Packages

When you want to temporarily prohibit access to the elements in a package, disable the package. When you disable a package, the server unloads all of its elements from memory. Disabling a package prevents Integration Server from loading that package at startup.

Important:

If your adapter has multiple user-defined packages, and you want to disable some of them, disable the adapter package first (WmHDFSAdapter). Otherwise, errors will be issued when you try to access the remaining enabled user-defined packages.

➤ To disable a package

1. Open the Integration Server Administrator if it is not already open.

2. In the **Packages** menu of the navigation area, click **Management**.
3. Click **Yes** in the **Enabled** column for the package that you want to disable.

The Integration Server Administrator issues a prompt to verify that you want to disable the package.

4. Click **OK** to disable the package.

When the package is disabled, the Integration Server Administrator displays **No** in the **Enabled** column.

The disabled package remains disabled until you enable it again. In addition, the package is not listed in Designer.

Importing and Exporting Packages

You import and export packages using Designer. Exporting allows you to export the package to a .zip file and save it to your hard drive. The .zip file can then be imported for use by another package.

Important:

Do not rename packages you export; the rename function is comparable to moving a package, and when you import the renamed package, you lose any triggers, connections, and notifications associated with this package.

For details about importing and exporting packages, see the *webMethods Service Development Help* for your release.

Group Access Control

You must use access control lists (ACLs) to control which groups have access to which adapter services. For example, you can use ACLs to prevent one development group from inadvertently updating the work of another group, or to allow or deny access to services that are restricted to one group but not to others.

For information about assigning and managing ACLs, see the *webMethods Service Development Help* for your release.

4 Managing Adapter for HDFS Connections

■ Overview	20
■ Before Configuring or Managing Adapter Connections	20
■ Configuring Adapter Connections	21
■ Testing Adapter Connections	26
■ Viewing Adapter Connection Parameters	26
■ Editing Adapter Connections	27
■ Deleting Adapter Connections	27
■ Enabling Adapter Connections	28
■ Disabling Adapter Connections	28

Overview

Adapter connections enable Integration Server to connect to the HDFS cluster at run time. An Adapter for HDFS connection contains parameters that Integration Server uses to connect to the HDFS cluster. The number of connections you create depends on your integration needs and the number of users who need to connect to the HDFS cluster. You must create a separate adapter connection for each unique user accessing the HDFS cluster.

You must have administrator privileges on Integration Server to access the administrative screens for Adapter for HDFS and configure the adapter connections. For information about setting user privileges, see the *webMethods Integration Server Administrator's Guide* for your release.

Before Configuring or Managing Adapter Connections

Perform the following steps before configuring or managing adapter connections.

➤ To prepare to configure or manage adapter connections

1. Install webMethods Integration Server and Adapter for HDFS on the same computer.

For information about the installation process, see [“Installing and Uninstalling webMethods Adapter for HDFS” on page 11](#).

2. Ensure that you have administrator privileges on Integration Server so that you can access the administrative screens of Adapter for HDFS.
3. Start your Integration Server and Integration Server Administrator, if they are not already running.
4. Ensure that the WmHDFSAdapter package is enabled.

For information about enabling the package, see [“Enabling Packages” on page 17](#).

5. By using Designer, create a user-defined package to include the connection.

For information about managing packages for the adapter, see [“Package Management” on page 15](#).

6. Configure Integration Server to use Kerberos, if you intend to implement Kerberos authentication for users accessing the HDFS cluster over Integration Server.

For Kerberos authentication to work with the adapter, include the following lines of code in *Integration Server_directory\instances\instance_name\config\is_jaas.cnf*:

```
HdfsClient {  
    com.sun.security.auth.module.Krb5LoginModule required  
    client=TRUE;  
};
```

For more information about configuring Kerberos in Integration Server, see the *webMethods Integration Server Administrator's Guide* for your release.

Configuring Adapter Connections

When you configure Adapter for HDFS connections, you specify information that Integration Server uses to connect to an HDFS cluster. You must configure Adapter for HDFS connections manually by using the Integration Server Administrator.

Note:

Ensure that you have performed the prerequisite tasks before configuring Adapter for HDFS, as specified in [“Before Configuring or Managing Adapter Connections” on page 20](#).

➤ To configure an adapter connection

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click **Configure New Connection**.
3. In the **webMethods Adapter for HDFS** section, provide following information about the package and folder in which you are creating the connection:

Field	Description/Action
Package	<p>The package in which to create the connection. You must have created the package before you can select it from the drop-down list. For general information about creating packages, see the <i>webMethods Service Development Help</i> for your release.</p> <div> <p>Note:</p> <p>Configure the connection in a user-defined package. For other important considerations when creating packages for Adapter for HDFS, see “Package Dependency Requirements and Guidelines” on page 16.</p> </div>
Folder Name	The folder in which to create the connection.
Connection Name	The name you want to provide for the connection. Connection names cannot have spaces or use special characters reserved by Integration Server and Designer. For more information about the use of special characters in package, folder, and element names, see the <i>webMethods Service Development Help</i> for your release.

4. In the **Connection Properties** section, provide following information about the HDFS NameNode to which Integration Server connects:

Field	Description/Action
Description	The description you want to provide for the connection. This is an optional field.
NameNode URI	URI of the NameNode in the HDFS cluster to which Integration Server connects. Specify this value in the <i>hdfs://<server>:<port></i> format. If you have configured High-Availability (HA) for your HDFS cluster, you must specify the NameNode URI in the <i>hdfs://<nameservices></i> format.

5. In the **File System Properties** section, provide following information about the data blocks added to or updated in the HDFS cluster:

Field	Description/Action
Replication	The number of copies of the data blocks into which a file that is updated in or added to the HDFS cluster is split. The default value of this parameter is one.
Block Size (MB)	The size, in MB, of each data block of a file that is updated in or added to the HDFS cluster. The default value of this parameter is 1MB.
Buffer Size (KB)	The size, in KB, of the buffer that Integration Server uses for each read or write operation to the HDFS cluster. The default value of this parameter is 4KB.

6. In the **Authentication Properties** section, select the authentication scheme for users accessing the HDFS cluster.

You can select one of the following authentication schemes from the drop-down list:

- **Simple (No Authentication):** Specify user and group names with valid access to the HDFS cluster.
- **Kerberos (Using Keytab):** Integration Server checks the contents of a keytab file for information about the specified principal, and then authenticates the principal.
- **Kerberos (Using Password):** Integration Server checks the KDC for information about the specified principal, and then authenticates the principal.

You can select either of the Kerberos authentication schemes only if you have configured Integration Server to use Kerberos. For more information about configuring Kerberos in Integration Server, see the *webMethods Integration Server Administrator's Guide* for your release.

7. In the **Authentication Properties** section, provide information about the selected authentication scheme.

The information that you need to provide depends on the type of authentication scheme that you select.

For this scheme...	Specify...								
Simple (No Authentication)	Select this parameter value to specify user and group names with valid access to the HDFS cluster.								
	<table> <tr> <th>Option</th><th>Description</th></tr> <tr> <td>User</td><td>A user on the host operating system with valid access to the HDFS cluster.</td></tr> <tr> <td>Group</td><td>The group to which the specified User belongs.</td></tr> </table>	Option	Description	User	A user on the host operating system with valid access to the HDFS cluster.	Group	The group to which the specified User belongs.		
Option	Description								
User	A user on the host operating system with valid access to the HDFS cluster.								
Group	The group to which the specified User belongs.								
Kerberos (Using Keytab)	<p>Select this parameter to enable Kerberos authentication for a principal based on information provided in the keytab file.</p> <p>You can select this scheme only if you have configured Integration Server to use Kerberos. For more information, see the <i>webMethods Integration Server Administrator's Guide</i> for your release.</p>								
	<table> <tr> <th>Option</th><th>Description</th></tr> <tr> <td>Kerberos Principal</td><td> <p>The name of the principal to use for Kerberos authentication.</p> <p>Specify the fully qualified domain name of the principal in the <i>hdfs/fully.qualified.domain.name@REALM</i> format.</p> </td></tr> <tr> <td>Path to Keytab File</td><td> <p>The local path to the keytab file that contains information about the specified principal.</p> <p>For example, you can specify a keytab file at the following path: <i>C:/users/foo/hdfs.keytab</i></p> </td></tr> <tr> <td>Group</td><td>The group to which the user associated with the specified Kerberos principal belongs.</td></tr> </table>	Option	Description	Kerberos Principal	<p>The name of the principal to use for Kerberos authentication.</p> <p>Specify the fully qualified domain name of the principal in the <i>hdfs/fully.qualified.domain.name@REALM</i> format.</p>	Path to Keytab File	<p>The local path to the keytab file that contains information about the specified principal.</p> <p>For example, you can specify a keytab file at the following path: <i>C:/users/foo/hdfs.keytab</i></p>	Group	The group to which the user associated with the specified Kerberos principal belongs.
Option	Description								
Kerberos Principal	<p>The name of the principal to use for Kerberos authentication.</p> <p>Specify the fully qualified domain name of the principal in the <i>hdfs/fully.qualified.domain.name@REALM</i> format.</p>								
Path to Keytab File	<p>The local path to the keytab file that contains information about the specified principal.</p> <p>For example, you can specify a keytab file at the following path: <i>C:/users/foo/hdfs.keytab</i></p>								
Group	The group to which the user associated with the specified Kerberos principal belongs.								

For this scheme...	Specify...
Kerberos (Using Password)	<p>Select this parameter to enable Kerberos authentication for a principal based on the credentials available in the KDC.</p> <p>You can select this scheme only if you have configured Integration Server to use Kerberos. For more information, see the <i>webMethods Integration Server Administrator's Guide</i> for your release.</p>
	Option Description
	<p>Kerberos Principal The name of the principal to use for Kerberos authentication.</p> <p>Specify the principal name in the <i>hdfs/fully.qualified.domain.name@REALM</i> format.</p>
	<p>Principal Password The password used to authenticate the principal to the KDC.</p>
	<p>Group The group to which the user associated with the specified Kerberos Principal belongs.</p>

8. In the **Advanced Settings** section, specify the local paths to the HDFS configuration files `core-site.xml` and `hdfs-site.xml`.

Based on your requirements, you can use these files for advanced configurations on the HDFS cluster.

Important:

Software AG recommends that you maintain identical versions of both these advanced configuration files on the HDFS cluster and the computer on which webMethods Adapter for HDFS is running.

Field	Description/Action
Path to core-site.xml	<p>The local path to the <code>core-site.xml</code> file.</p> <p>For example, you can specify <code>core-site.xml</code> at the following path: <code>C:/users/foo/core-site.xml</code></p>
Path to hdfs-site.xml	<p>The local path to the <code>hdfs-site.xml</code> file.</p> <p>For example, you can specify <code>hdfs-site.xml</code> at the following path: <code>C:/users/foo/hdfs-site.xml</code></p>

Note:

- The values of the connection parameters that you specify on this screen override certain properties in the configuration files `core-site.xml` and `hdfs-site.xml` on the HDFS cluster. For details, see [“Adapter Connection Parameters that Override the Configuration File Properties” on page 25](#)
- For the clients authenticated to access the HDFS cluster using Kerberos, you can configure their communication with the HDFS cluster through Integration Server by setting values of certain properties in `core-site.xml` and `hdfs-site.xml`. For more information, see [“Encrypting Communication with the HDFS Cluster” on page 25](#).

9. Click **Save Connection**.

If the parameters are valid, the connection you created appears on the adapter's Connections screen.

10. Click **Return to webMethods Adapter for HDFS Connections** to return to the Connections screen.

Adapter Connection Parameters that Override the Configuration File Properties

The parameter values that you specify to configure the Adapter for HDFS connections from the Integration Server Administrator override certain properties in the configuration files `core-site.xml` and `hdfs-site.xml` on the HDFS cluster.

The following table lists the adapter connection parameters and the properties in the HDFS configuration files that they override:

The connection parameter...	Overrides the property...	In the configuration file...
NameNode URI	<code>fs.defaultFS</code>	<code>core-site.xml</code>
Replication	<code>dfs.replication</code>	<code>hdfs-site.xml</code>
Block Size (MB)	<code>dfs.blocksize</code>	<code>hdfs-site.xml</code>
Buffer Size (KB)	<code>file.buffer.size</code>	<code>core-site.xml</code>
Scheme	<code>hadoop.security.authentication</code>	<code>core-site.xml</code>

Note:

For information about configuring the adapter connection parameters, see [“Configuring Adapter Connections” on page 21](#).

Encrypting Communication with the HDFS Cluster

For the clients authenticated to access the HDFS cluster using Kerberos, you can encrypt their communication with the HDFS cluster by setting values of certain parameters in the `core-site.xml` and `hdfs-site.xml` files.

Important:

Software AG recommends that you maintain identical versions of both `core-site.xml` and `hdfs-site.xml` on the HDFS cluster and the computer on which webMethods Adapter for HDFS is running.

1. In `core-site.xml`, set the value of the `hadoop.rpc.protection` property to `privacy`.
2. In `hdfs-site.xml`, set the value of the `dfs.encrypt.data.transfer` property to `true`.

Testing Adapter Connections

You can check if Integration Server can connect to the HDFS cluster by testing the Adapter for HDFS connections from the Integration Server Administrator.

➤ **To test a connection**

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click the  icon for the connection you want to test.

Integration Server attempts to connect to the HDFS cluster based on the information specified in the adapter connection. If the connection is successful, the Integration Server Administrator displays a confirmation message.

Viewing Adapter Connection Parameters

You can view the connection parameters for Adapter for HDFS from the Integration Server Administrator. You cannot view these parameters from Designer.

Viewing Adapter Connection Parameters from the Integration Server Administrator

➤ **To view the parameters for a connection from the Integration Server Administrator**

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click the  icon for the connection you want to see.

The View Connection screen displays the parameters for the connection. For descriptions of the connection parameters, see [“Configuring Adapter Connections” on page 21](#).

3. To return to the Connections screen, click **Return to webMethods Adapter for HDFS Connections**.

Editing Adapter Connections

If the login information for the HDFS cluster changes, or if you want to redefine parameters that a connection uses when accessing the HDFS cluster, you can update the connection's parameters by using the Integration Server Administrator.


Note:

You must ensure that the connection is disabled before editing it. For instructions, see [“Disabling Adapter Connections” on page 28](#).

➤ **To edit a connection**

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click the  icon for the connection you want to edit.

The Edit Connection screen displays the current parameters for the connection. Update the connection's parameters by typing or selecting the values you want to specify.

Note:

From the Edit Connection screen for a particular connection, you can edit values of all the connection parameters except **Package**, **Folder Name**, and **Connection Name**.

3. To save your changes and return to the Connections screen, click **Save Changes**.

Deleting Adapter Connections

If you do not want to use a particular Adapter for HDFS connection, you can delete it by using the Integration Server Administrator.

If you delete an Adapter for HDFS connection, the adapter services that are defined to use the connection will not work. However, you can assign a different connection to an adapter service and re-use the service.

Note:

You must ensure that the connection is disabled before deleting it. For instructions, see [“Disabling Adapter Connections” on page 28](#).

➤ **To delete a connection**

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click the  icon for the connection you want to delete.

Integration Server deletes the adapter connection.

Enabling Adapter Connections

An Adapter for HDFS connection must be enabled before you can configure any adapter service by using the connection, or before an adapter service can use the connection at run time. You enable adapter connections by using the Integration Server Administrator.

Note:


When you reload a package that contains enabled connections, the connections are automatically enabled when the package reloads. If the package contains connections that are disabled, they remain disabled when the package reloads.

> To enable a connection

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click **No** in the **Enabled** column for the connection you want to enable.

The adapter connection is enabled. In addition, the Integration Server Administrator displays a  symbol and **Yes** in the **Enabled** column.

Disabling Adapter Connections

Adapter for HDFS connections must be disabled before you can edit or delete them. You disable adapter connections by using the Integration Server Administrator.

> To disable a connection

1. In the **Adapters** menu of the Integration Server Administrator's navigation area, click **webMethods Adapter 9.8 for HDFS**.

The Connections screen appears.

2. On the Connections screen, click **Yes** in the **Enabled** column for the connection you want to disable.

The adapter connection is disabled. In addition, the Integration Server Administrator displays a **No** in the **Enabled** column.

5 Adapter Logging

■ Overview	32
■ Adapter Logging Levels	32
■ Adapter for HDFS Message Logging	33

Overview

Adapter for HDFS uses the Integration Server server log for logging of error messages. You can configure different logging levels for the adapter error messages.

Adapter Logging Levels

Adapter for HDFS uses the Integration Server logging mechanism for error messages. You can configure and view the Integration Server logs to monitor and troubleshoot Adapter for HDFS. For detailed information about logging in Integration Server, including instructions for configuring and viewing the server log, see the *webMethods Integration Server Administrator's Guide* for your release.

Accessing Adapter Logging Levels

You can configure different logging levels for Adapter for HDFS by using Integration Server.

➤ To access the adapter's logging information

1. From the Integration Server Administrator screen, select **Settings > Logging**.

The Logging Settings screen appears.

2. In the **Logger List** section, click **Server Logger**.

The View Server Logger Configurations screen appears. The **Server Logger Configuration** section has **HDFS Adapter** included in the **Facility** section.

3. Expand the **HDFS Adapter** tree to see the logging levels for the adapter.

Changing Logging Levels

You must use Integration Server to change the error logging levels for Adapter for HDFS.

➤ To change logging settings for the adapter

1. Click **Edit Logging Settings**. Select the required **Level of Logging** for Adapter for HDFS.
2. After making your changes, click **Save Changes**.

For complete information about specifying the amount and type of information to include in the log, see the *webMethods Audit Logging Guide* for your release.

Adapter for HDFS Message Logging

Integration Server maintains several types of logs; however, Adapter for HDFS logs messages only to the server logs.

The log messages for the adapter appear in the following format:

- WHD.000n.*nnnn*

where WHD indicates that the message is from Adapter for HDFS, 000n indicates the facility for which the error is logged, and *nnnn* represents the error code.

To monitor the log messages for the adapter in the server logs, ensure that the logging settings are configured to monitor the following facilities:

- 0001 Administration
- 0002 HDFS Connection
- 0003 HDFS

A Built-In Services

■ Overview	36
■ wm.hadoop.hdfs:open	36
■ wm.hadoop.hdfs:read	37
■ wm.hadoop.hdfs:write	39
■ wm.hadoop.hdfs:close	40

Overview

The built-in services available with Adapter for HDFS enable you to read data from and write data to the HDFS cluster. These services are located in the WmHDFSAdapter package.

wm.hadoop.hdfs:open

The `wm.hadoop.hdfs:open` service enables you to open a stream to an HDFS file to perform read and write operations. In addition, the service allows you to specify rollover properties for writing to the file based on its size or the duration of the write operation.

Input Parameters

<i>ConnectionName</i>	String Name of the Adapter for HDFS connection for connecting to the HDFS cluster.
<i>hdfsFilePath</i>	String Path to the HDFS file to open. You can specify either the relative path to the file (for example, <code>files/abc.txt</code>) or its absolute path (for example, <code>/users/hduser/files/abc.txt</code>).
<i>rolloverType</i>	String Type of rollover to be performed. You can specify either the file size or duration for the write operation. Specify the following parameters only if you want the write operation to be performed until the file reaches a specified size or for a specified duration. After the specified file size or duration is reached, the write operation continues in a new file. Set to: <ul style="list-style-type: none">■ <code>FileSizeRollover</code> to create a new file when the file reaches a certain size . The unit of size for rollover is in bytes for the rollover.■ <code>TimeRollover</code> to create a new file for the write operation after a specified duration. The unit of time for rollover is in seconds for the rollover.
<i>openType</i>	String Specifying the type of operation to be performed for which the file is being opened. Set to: <ul style="list-style-type: none">■ <code>Write</code> to write data to a specified HDFS location.■ <code>Read</code> to read data and make the contents of the HDFS file available to subsequent services.
<i>rolloverUnit</i>	String Specify the unit of the rollover for write operations.
<i>writeMode</i>	String The mode in which to write to the specified HDFS file.

Specify this parameter only if *openType* is set to *write*.

Set to:

- *create* to create a new HDFS file.
- *append* to append to the contents of an existing HDFS file.
- *overwrite* to overwrite the contents of an existing HDFS file.

This is the default value.

<i>permissions</i>	String Name of a registered, IANA character set. If you specify an unsupported encoding, the system throws an exception
--------------------	--

Output Parameters

<i>status</i>	String The status of the open operation. The valid value of this parameter can either be <i>success</i> or <i>failure</i> .
<i>message</i>	String A detailed message specifying the reason for the success or failure of the open operation.
<i>handle</i>	String Handle of the HDFS file to which the stream is opened.

wm.hadoop.hdfs:read

The `wm.hadoop.hdfs:read` service connects to the HDFS cluster and reads a specified file. The service returns the contents of the read file in a specific format.

Input Parameters

<i>handle</i>	String Handle of the HDFS file to which the <code>wm.hadoop.hdfs:read</code> service reads data.
<i>loadAs</i>	String Format in which you want the <code>wm.hadoop.hdfs:read</code> service to make the contents of the HDFS file available to subsequent services. Set to: <ul style="list-style-type: none"> ■ <i>reader</i> to return the contents as a <code>java.io.Reader</code> object. ■ <i>file</i> to return the contents to a file on the local computer. ■ <i>stream</i> to return the contents as a <code>java.io.InputStream</code> object. ■ <i>string</i> to return the contents as a <code>java.io.String</code> object. ■ <i>bytes</i> to return the contents as a byte array.

<i>localFilePath</i>	String Optional. The complete path to the file to be created in the local file system. Specify this parameter only if <i>loadAs</i> is set to <i>file</i> .	
<i>readOptions</i>	Document Details to specify for the read operation.	
	Option	Description
	<i>encoding</i>	String Optional. Name of a registered, IANA character set. If you specify an unsupported encoding, the system throws an exception.
	<i>localFileCreationMode</i>	String Optional. The mode in which to save the contents of the HDFS file to the local file system. Specify this parameter only if <i>loadAs</i> is set to <i>file</i> . Set to: <ul style="list-style-type: none">■ <i>create</i> to create a new local file.■ <i>append</i> to append to the contents of an existing local file.■ <i>overwrite</i> to overwrite the contents of an existing local file. This is the default value.

Output Parameters

<i>status</i>	String The status of the read operation. The valid value of this parameter can either be <i>success</i> or <i>failure</i> .
<i>message</i>	String A detailed message specifying the reason for the success or failure of the read operation.
<i>data</i>	Document The data returned by the <code>wm.hadoop.hdfs:read</code> service. The format of the returned data depends on the value of the <i>loadAs</i> input parameter. <ul style="list-style-type: none">■ <i>bytes</i> Returns the data as a byte array.

- `stream` Returns the data as a `java.io.InputStream` object.
- `string` Returns the data as a `java.io.String` object.
- `reader` Returns the data as a `java.io.Reader` object.
- `filePath` Returns the data to a file path on the local computer.

wm.hadoop.hdfs:write

The `wm.hadoop.hdfs:write` service connects to the HDFS cluster and writes to a specified file. To ensure that the contents of the file are not updated by other users or services at the time of the write operation, the service changes the file extension to that of a temporary file (`.tmp`), writes to the temporary file, and reverts the extension to the original value after the write operation is complete. The rollover logic renames the `.tmp` file to have the following filename conventions if the rollover is based on the file size or time duration:

```
{Filename}+{_hostname}+{_currentdatetime}.{format}
```

The `wm.hadoop.hdfs:write` service returns the status of the write operation.

Input Parameters

<i>handle</i>	String Handle of the HDFS file to which the <code>wm.hadoop.hdfs:write</code> service writes data.	
<i>writeAs</i>	Format of the HDFS file to which the <code>wm.hadoop.hdfs:write</code> service writes data.	
<i>data</i>	String Input source to write to HDFS. Set to: <ul style="list-style-type: none"> ■ <code>bytes</code> to write a byte array. ■ <code>stream</code> to write a <code>java.io.InputStream</code> object. ■ <code>string</code> to write a <code>java.io.String</code> object. ■ <code>localFilePath</code> to write a <code>java.io.localFilePath</code> string. ■ <code>reader</code> to write a <code>java.io.Reader</code> object. 	
	Note: If you specify multiple input sources, the service considers the first of the specified sources for writing to the HDFS file.	
<i>writeOptions</i>	Document Details to specify for the write operation.	
	Option	Description
	<i>encoding</i>	String Optional. Name of a registered, IANA character set. If you specify an

		unsupported encoding, the system throws an exception.
	<i>permissions</i>	String Optional. File permissions in either octal or symbolic format.
<i>execution</i>	Document Optional. Details to specify for executing the <code>wm.hadoop.hdfs:write</code> service.	
	Option	Description
	<i>async</i>	Boolean Optional. Flag that indicates whether or not the service is to be executed asynchronously. Set to: <ul style="list-style-type: none">■ <code>true</code> to execute the service asynchronously.■ <code>false</code> to execute the service synchronously.
	<i>callbackServiceName</i>	String Optional. Fully qualified name of the service to call after the <code>wm.hadoop.hdfs:write</code> service is executed asynchronously. Specify this value only if the <i>async</i> option is set to <code>true</code> .

Output Parameters

<i>status</i>	String The status of the write operation. The valid value of this parameter can either be <code>success</code> to indicate success of the write operation, <code>failure</code> to indicate failure of the write operation, or <code>submitted</code> to indicate that the write request is submitted to HDFS but has not yet started.
<i>message</i>	String A detailed message specifying the reason for the success or failure of the write operation.

wm.hadoop.hdfs:close

The `wm.hadoop.hdfs:close` service closes the stream to the file with the file handle. The service returns the message specifying the reason for the success or failure of the close operation.

Input Parameters

<i>handle</i>	String Handle of the HDFS file to which the stream is opened.
---------------	--

Output Parameters

<i>status</i>	String The status of the close operation. The valid value of this parameter can either be <code>success</code> or <code>failure</code> .
---------------	---

<i>message</i>	String A detailed message specifying the reason for the success or failure of the close operation.
----------------	---
