



# **Experian Pandora Hadoop Configuration Help**

Versions 5.9.0, 5.8.0

THE WORD "EXPERIAN" AND THE GRAPHICAL DEVICE ARE TRADEMARKS OF EXPERIAN AND REGISTERED IN THE EU, USA AND OTHER COUNTRIES.

THIS DOCUMENT CONTAINS INFORMATION, PROPRIETARY TO EXPERIAN, WHICH IS PROTECTED BY INTERNATIONAL COPYRIGHT LAW. THE INFORMATION CONTAINED HEREIN MAY NOT BE DISCLOSED TO THIRD PARTIES, COPIED OR DUPLICATED, IN WHOLE OR IN PART, WITHOUT THE PRIOR WRITTEN CONSENT OF COPYRIGHT OWNER. PLEASE CONTACT EXPERIAN TO FOR ANY CONSENT ENQUIRIES.

(C) 2007 – 2018 EXPERIAN

## Overview

Pandora supports files stored in a remote HDFS file system. These are made accessible by configuring a Hadoop HDFS connection and are treated as Read-Only files.

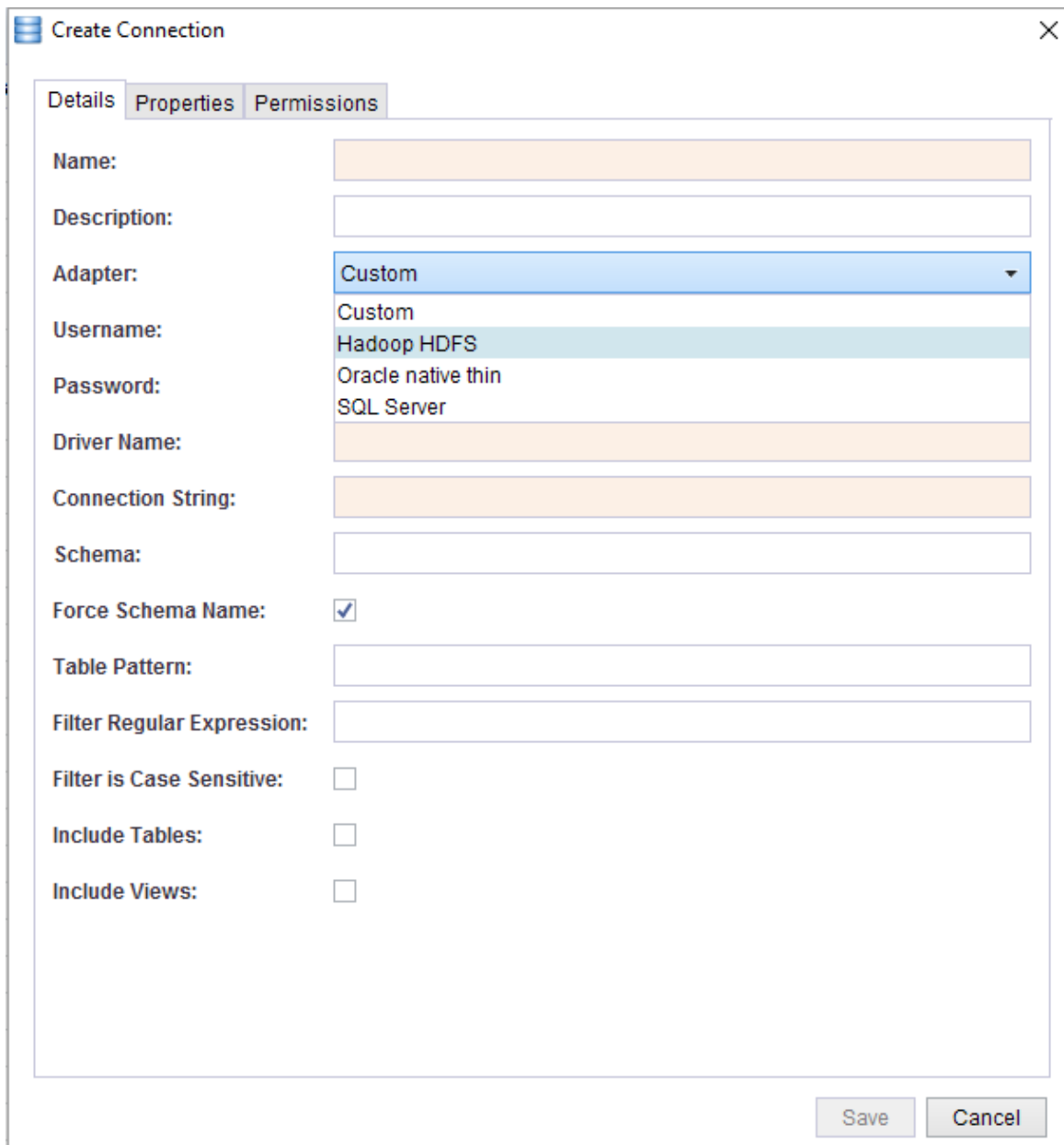
## Supported File Types

- Comma Separated Values files with an extension of .csv
- Fixed Length files with an extension of .dat
- Delimited Files with an extension of .txt

## Configuring a Hadoop Connection

Navigate to **Home** → **Connections** and right-click on **Connections** to add a new connection.

The Create Connection Dialog will appear and you should select **Hadoop HDFS** Adapter from the drop down list of available adapters.



**Create Connection**

Details Properties Permissions

Name:

Description:

Adapter: Custom

Username: Custom

Password: Hadoop HDFS

Driver Name: Oracle native thin

Connection String: SQL Server

Schema:

Force Schema Name:

Table Pattern:

Filter Regular Expression:

Filter is Case Sensitive:

Include Tables:

Include Views:

Save Cancel

Not all fields are mandatory; however you must supply at least a **Hostname** and **Port**. The Hostname is the IP address or DNS name of the Primary NameNode of your Hadoop cluster. The port is a site specific Hadoop property. Typically, you will find this in the Hadoop configuration file core-site.xml

If you are not familiar with Hadoop configurations ask your Hadoop administrator to provide this information.

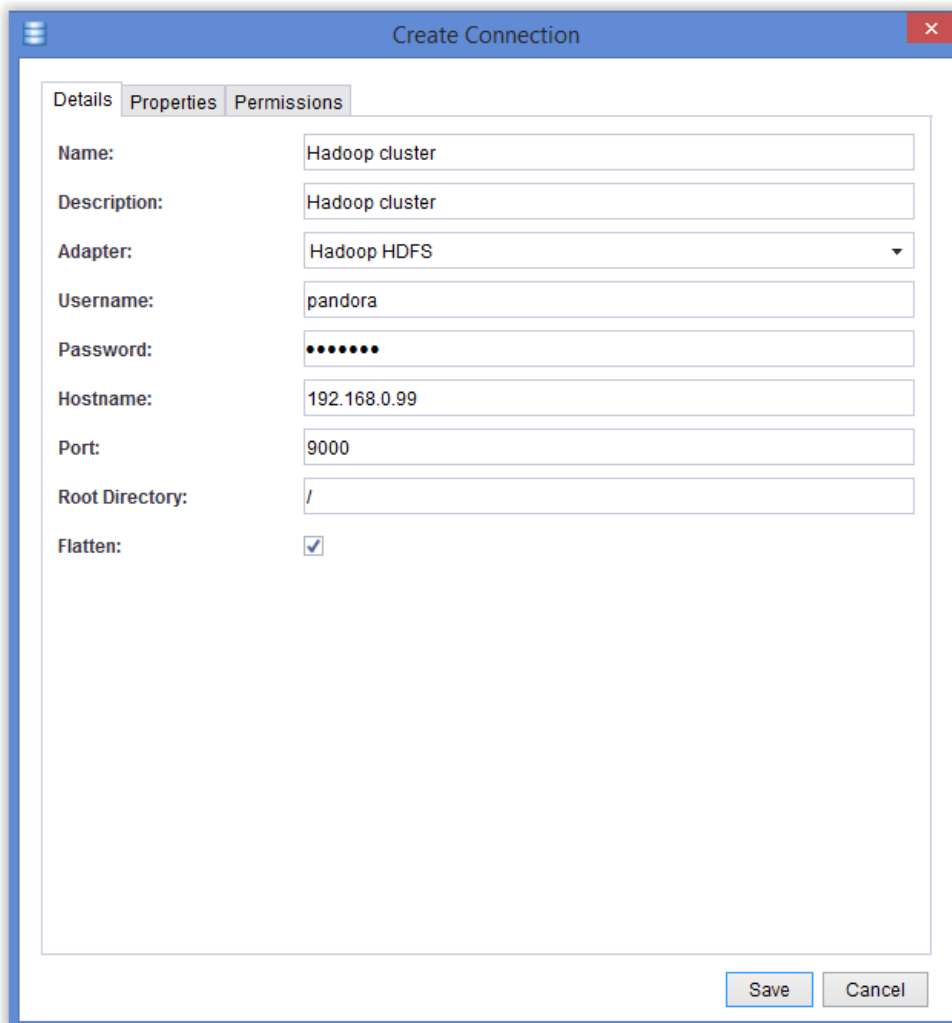
Below is a minimal example of the core-site.xml. You will find the HDFS port number under the property **fs.default.name**

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://0.0.0.0:9000</value>
  </property>
</configuration>
```

You will almost certainly require a **Username** and **Password** to access the files stored on the Hadoop cluster. This username and password are the Linux username and password on the Hadoop host machine. File access permissions for your username will determine which files are visible to Pandora. Only those to which you have read access and are supported file types will be visible.

The **Root Directory** is the starting directory from which to start the file discovery. If it is not specified it will default to the root of the HDFS file system ("/"). You should choose the starting directory carefully since HDFS file systems may contain a huge number of files. You also have the choice of recursing through all the directories below your starting directory. If you wish to include all other directories below your starting directory, then tick the box **Flatten**.

You should have everything filled in similar to this;



Field	Value
Name	Hadoop cluster
Description	Hadoop cluster
Adapter	Hadoop HDFS
Username	pandora
Password	••••••••
Hostname	192.168.0.99
Port	9000
Root Directory	/
Flatten	<input checked="" type="checkbox"/>

When Pandora connects to Hadoop it will list the available files in the Explorer pane. You may configure these files individually prior to loading them.

