

**Technical Note:
Distribute TIBCO® Enterprise Runtime for R
using Apache Hadoop™**

*Software Release 4.0
September 2015*

Important Information

SOME TIBCO SOFTWARE EMBEDS OR BUNDLES OTHER TIBCO SOFTWARE. USE OF SUCH EMBEDDED OR BUNDLED TIBCO SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED TIBCO SOFTWARE. THE EMBEDDED OR BUNDLED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER TIBCO SOFTWARE OR FOR ANY OTHER PURPOSE.

USE OF TIBCO SOFTWARE AND THIS DOCUMENT IS SUBJECT TO THE TERMS AND CONDITIONS OF A LICENSE AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED SOFTWARE LICENSE AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER LICENSE AGREEMENT WHICH IS DISPLAYED DURING DOWNLOAD OR INSTALLATION OF THE SOFTWARE (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH SOFTWARE LICENSE AGREEMENT OR CLICKWRAP END USER LICENSE AGREEMENT, THE LICENSE(S) LOCATED IN THE "LICENSE" FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.

This document contains confidential information that is subject to U.S. and international copyright laws and treaties. No part of this document may be reproduced in any form without the written authorization of TIBCO Software Inc.

TIBCO, Two-Second Advantage, TIBCO Spotfire, TIBCO Enterprise Runtime for R, TIBCO Spotfire Server, TIBCO Spotfire Web Player, TIBCO Spotfire Statistics Services, S-PLUS, and TIBCO Spotfire S+ are either registered trademarks or trademarks of TIBCO Software Inc. in the United States and/or other countries.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only.

THIS SOFTWARE MAY BE AVAILABLE ON MULTIPLE OPERATING SYSTEMS. HOWEVER, NOT ALL OPERATING SYSTEM PLATFORMS FOR A SPECIFIC SOFTWARE VERSION ARE RELEASED AT THE SAME TIME. SEE THE README FILE FOR THE AVAILABILITY OF THIS SOFTWARE VERSION ON A SPECIFIC OPERATING SYSTEM PLATFORM.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. TIBCO SOFTWARE INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "READ ME" FILES.

Copyright © 2010-2015 TIBCO Software Inc. ALL RIGHTS RESERVED.

TIBCO Software Inc. Confidential Information

Contents

Documentation and Support Services	4
Distribute TIBCO Enterprise Runtime for R using Apache Hadoop	6
Script Usage	6
Running the Script to Distribute TERR with Apache Hadoop	6
Example script: deployterr.sh	7

Documentation and Support Services

Documentation for this and other TIBCO products is available on the TIBCO Documentation site:

<https://docs.tibco.com>

Documentation on the TIBCO Documentation site is updated more frequently than any documentation that might be included with the product. To ensure that you are accessing the latest available help topics, please visit <https://docs.tibco.com>.

TIBCO Enterprise Runtime for R Documentation

The following documents for this product can be found in the TIBCO Documentation Library:

- *Release Notes*
- *Technical Guide*
- *Java with terrJava*
- *Package Management*
- *Language Reference*
- *Available Functions by Category and Subcategory*
- *Unavailable Function Lists*
- *Open-Source R Package Compatibility*
- *Differences Between TIBCO® Enterprise Runtime for R and Open-Source R*
- *Technical Note: Distribute TIBCO® Enterprise Runtime for R using Apache Hadoop™*
- *Technical Note: Configure KNIME to use TIBCO® Enterprise Runtime for R*
- *Technical Note: Configure RStudio to use TIBCO® Enterprise Runtime for R*
- *Technical Note: Configure SparkR to use TIBCO® Enterprise Runtime for R*
- *License Agreement*

Product System Requirements

For a list of system requirements for this product and other TIBCO Spotfire® products, visit this site:

<http://support.spotfire.com/sr.asp>

How to Contact TIBCO Support

For comments or problems with this manual or the software it addresses, contact TIBCO Support:

- For an overview of TIBCO Support, and information about getting started with TIBCO Support, visit this site:

<http://www.tibco.com/services/support>

- If you already have a valid maintenance or support contract, visit this site:

<https://support.tibco.com>

Entry to this site requires a user name and password. If you do not have a user name, you can request one.

How to Join TIBCOmmunity

TIBCOmmunity is an online destination for TIBCO customers, partners, and resident experts. It is a place to share and access the collective experience of the TIBCO community. TIBCOmmunity offers forums, blogs, and access to a variety of resources. To register, go to the following web address:

<https://www.tibcommunity.com>

Distribute TIBCO Enterprise Runtime for R using Apache Hadoop

You can deploy TIBCO® Enterprise Runtime for R to all tracker nodes in a Cloudera® or Hortonworks® cluster using Apache Hadoop™. This allows you to benefit from the performance and robustness of TIBCO Enterprise Runtime for R when running R language analytics on your Hadoop cluster.

By running the script provided in this document, you retrieve the list of active trackers from Hadoop, and then install the TIBCO Enterprise Runtime for R engine for each node to `/opt/tibco/TERRengine`.



For remote users, the user must have sudo privileges to create a directory under `/opt` and copy files to it. Additionally, because the script pulls the hostname and IP from tracker nodes directly, it must be executed as a user with privileges to run Hadoop commands.



The license for the Developer Edition of TIBCO Enterprise Runtime for R prohibits using TIBCO Enterprise Runtime for R in a Production setting. Development and prototyping with Hadoop is allowed, but production usage requires a commercial license.

Script Usage

```
deployterr.sh -k sshkey -u user TERRengine.tar.gz
```

This table contains the description of the options comprising the command to run the shell script to distribute TIBCO Enterprise Runtime for R engines using Hadoop to a Clouder or Hortonworks cluster.

Option	Description
-k	ssh key used for connecting to tracker nodes (Optional).
-u	user to log in as on tracker nodes (Required).
TERRengine.tar.gz	The TIBCO Enterprise Runtime for R engine archive (tar.gz) included with TIBCO Spotfire® Statistics Services (Linux build required).

Running the Script to Distribute TERR with Apache Hadoop

Follow these steps to run the TIBCO Enterprise Runtime for R deployment script using Hadoop to a Cloudera or Hortonworks cluster.

Prerequisites

- You must be able to connect to all nodes with passwordless key authentication.
- The remote user that the script requests must be a user with sudo privileges to create a directory under `/opt` and copy files to it.
- The script must be executed as a user that can run Hadoop commands.

Procedure

1. Copy the script in [Example script: deployterr.sh](#) to a file on the name node. Name the script file `deployterr.sh`.
2. Verify that the path to the Hadoop executable is correct in `deployterr.sh`.



if the Hadoop executable is not in `/usr/bin/hadoop` (for example, if it is either in the home directory or other location such as `/opt`), in the script, modify `HADOOPBIN` to use the correct path.

3. Execute the script `deployterr.sh` as a user with access to Hadoop.

Example script: `deployterr.sh`

Customize this example shell script to deploy multiple TIBCO Enterprise Runtime for R engines to a distributed Hadoop system.

To install TERR to a Linux system at `/opt/TIBCO/TERRengine`, customize and run the following script using the command described in [Script Usage](#), following the instructions in [Running the Script to Distribute TERR with Hadoop](#).

```
#!/bin/bash
HADOOPBIN=/usr/bin/hadoop

while getopts "hk:u:" opt; do
  case $opt in
    k)
      sshkey=$OPTARG
      ;;
    u)
      user=$OPTARG
      ;;
    h)
      echo "test"
      ;;
    :)
      echo "Option -$OPTARG requires an argument." >&2
      exit 1
      ;;
  esac
done

#Verify our arguments are valid

if test -n "$sshkey"; then
  REMOTESSH="ssh -i $sshkey -t -q -o StrictHostKeyChecking=no"
  REMOTESCP="scp -i $sshkey -q -o StrictHostKeyChecking=no"
else
  REMOTESSH="ssh -t -q -o StrictHostKeyChecking=no"
  REMOTESCP="scp -q -o StrictHostKeyChecking=no"
fi

if [[ -z "$user" ]]; then
  echo "You must specify the user for connecting to the tracker nodes"
  exit 1
fi

shift $(( $OPTIND - 1 ))
terrfile=$1

#Make sure we're working with a tar.gz
if [[ "$terrfile" != *.tar.gz || ! -f "$terrfile" ]]; then
  echo "Supplied file is not valid"
  exit 1
fi

#Get the list of active tracker nodes and pull the hostname/ip from their name,
these execute map reduce jobs

if ! trackers=$(("$HADOOPBIN" job -list-active-trackers|sed -e 's/^\.*tracker_\([^:]*\):.*$/\1/'); then
```

```

    echo "An error occurred while obtaining the tracker nodes"
    exit 1
fi

numtrackers=$(printf '%s\n' "$trackers"|wc -l)

#One final prompt before we proceed to copy and unpack TERR to each node

read -p "Ready to deploy TERRengine to $numtrackers tracker nodes. Proceed? [Y/N]"
yn
case $yn in
  [Yy]* ) echo "Proceeding"
          ;;
  [Nn]* ) exit
          ;;
  * ) echo "Please answer yes or no."
       ;;
esac

function deploy_terr {
  for node in $trackers; do
    $REMOTESCP $terrfile $node:/tmp/TERRengine.tar.gz
    $REMOTESSH $user@$node "sudo bash -c 'if [ -d /opt/TIBCO/TERRengine ]; then
rm -rf /opt/TIBCO/TERRengine; fi'"
    $REMOTESSH $user@$node "sudo mkdir -p /opt/TIBCO/TERRengine"
    $REMOTESSH $user@$node "sudo tar xf /tmp/TERRengine.tar.gz -C /opt/TIBCO/
TERRengine"
    $REMOTESSH $user@$node "sudo rm /tmp/TERRengine.tar.gz"
  done
}
if deploy_terr; then
  echo "TERR has been deployed to all nodes"
  exit
else
  echo "An error was encountered"
  exit 1
fi

```