TIBCO Spotfire® offers native support for a number of Hadoop flavors. Hadoop integration points can be grouped into two categories: TIBCO Spotfire native data connectors and TERR-Hadoop integration. Both provide an extensive set of analytic features and security options.

Spotfire Hadoop connections can be quickly configured into analytic workflows, dashboards, or reports, which can then be shared, reused, and consumed across organizations. KPIs based on Hadoop data can be pushed to virtually any user device with TIBCO Spotfire® Metrics. Extensive geo analytic support within Spotfire makes it easy to generate insights from geographical data.

Data connectors allow accessing Hadoop data via in-database, in-memory, on-demand, or a combination of methods. In-database enables analysis of huge data sets by pushing aggregations into Hadoop. These in-database “master” visualizations are often combined with “details” visualizations where row level data slices are extracted from Hadoop on-demand and loaded into memory.

Spotfire connects, models, and visualizes Hadoop data without the need for scripting or manual query editing. However, to allow an advanced data modeler the freedom to create optimized queries for certain use cases, Spotfire offers custom query support in combination with in-database, in-memory, and on-demand.
VISUALIZE DATA IN HADOOP
Hadoop Data Connectors

<table>
<thead>
<tr>
<th>HADOOP</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache Hive</td>
<td>Supported via the Hortonworks connector</td>
</tr>
<tr>
<td>Cloudera Impala</td>
<td>Certified native connectivity with support for Kerberos SSO and LDAP with SSL</td>
</tr>
<tr>
<td>Cloudera Hive</td>
<td>Certified native connectivity with support for Kerberos SSO and LDAP with SSL</td>
</tr>
<tr>
<td>Hortonworks</td>
<td>Connectivity through Hive</td>
</tr>
<tr>
<td>Pivotal HAWQ</td>
<td>Native connectivity through Pivotal’s SQL query engine</td>
</tr>
<tr>
<td>Additional Hadoop Sources</td>
<td>Please contact your local TIBCO representative</td>
</tr>
</tbody>
</table>

Extract data from Hadoop for further analysis (either in Spotfire or TERR)
- Data on-demand (drill down from visualizations)
- Custom queries

TIBCO ENTERPRISE RUNTIME FOR R AND HADOOP: ADVANCED ANALYTICS
The powerful combination of Spotfire, TERR, and Hadoop enables deeper, more valuable analysis of Hadoop data:

SPOTFIRE FOR DATA SELECTION, TERR FOR ADVANCED ANALYSIS
Spotfire can exploit the power of Hadoop for in-datasource aggregations, presenting overview visualizations of the aggregated data, and then allowing users to extract selected slices of the data for deeper analysis. The detailed data can also be sent to TERR for advanced analysis, and the entire workflow can be encapsulated as a best practice in Spotfire, enabling less technical users to visualize and make predictions based on huge Hadoop data sets.

PRODUCTION ANALYSIS OF DATA FROM HADOOP
Once a data connection to Hadoop (or other data source) has been defined, TERR can reuse these same data connections to extract and analyze data directly, through simple or parameterized custom queries independent of Spotfire. TERR includes a set of R functions for this purpose, which can be called in simple scripts in a TERR session. This enables organizations to routinely analyze large amounts from Hadoop as part of scheduled, production applications.
TERR AS MAP/REDUCE ANALYTIC ENGINE
TERR can be used to execute native Map/Reduce calls directly on a Hadoop cluster via the Hadoop streaming interface. Faster, more scalable, and more robust than open source R, TERR enables you to process Hadoop data more quickly and more reliably. As performance gains are multiplied across the nodes, this approach produces analytic answers much faster and with fewer resources. The flexibility of the R language, combined with the reliability and performance of TERR, enables powerful applications such as ensemble modeling, where individual predictive models are created in parallel across the Hadoop cluster, and then combined to provide the best predictive model for your application.

PUTTING IT ALL TOGETHER
Combining these powerful features means that very sophisticated and powerful analytic use cases can be encapsulated in a Spotfire workflow, enabling a data scientist to create, analyze, and share the results of an analytic application without worrying about the details of the Hadoop architecture.

• Specify Map and Reduce operations through simple R scripts. TERR communicates directly to HDFS and executes MapReduce operations on the cluster using high-performance TERR engines.
• Parameterize and generate MapReduce code or HIVEQL queries from simple Spotfire interfaces.
• Combine and visualize the results within Spotfire.

Spotfire template for running a MapReduce job using TERR. The Spotfire user specifies keys and values from metadata along with a statistical model to run in parallel on the Hadoop data nodes. The R code for the mapper and reducer jobs is generated behind the scenes and run on the data nodes. Results of the statistical models are returned to HDFS for further interactive visual analysis in Spotfire. The user can test the models locally and dump results locally through Spotfire menu selections. The user can also edit the reduce code to take additional ETL or modeling steps.
TIBCO Spotfire
281 Summer Street,
3rd Floor
Boston, MA 02210
+1 866-240-0491 US
+44-800-520-0443 EUR
www.tibco.com
spotfire.tibco.com

TIBCO Software Inc. (NASDAQ: TIBX) is a global leader in infrastructure and business intelligence software. Whether it’s optimizing inventory, cross-selling products, or averting crisis before it happens, TIBCO uniquely delivers the Two-Second Advantage™—the ability to capture the right information at the right time and act on it preemptively for a competitive advantage. With a broad mix of innovative products and services, customers around the world trust TIBCO as their strategic technology partner. Learn more about TIBCO at www.tibco.com.

©2014 TIBCO Software Inc. All rights reserved. TIBCO, the TIBCO logo, TIBCO Software, and Spotfire are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

10/13/14

Under the hood of TERR MapReduce. The Spotfire point-click environment creates TERR scripts for Mapper and Reducer. These are run through the Hadoop streaming interface on TERR data nodes. Results are returned to HDFS and picked up in Spotfire for detailed drill-through analysis. This workflow enables complex statistical models to be run in parallel on the nodes through point-click operations in Spotfire, thus supplying the power of TERR and the scale of Hadoop to novice as well as expert users.

TERR PERFORMANCE AND R COVERAGE
Designed for high performance and scalability, TERR is compatible with thousands of R packages, as well as the popular R IDE, RStudio. On small to moderate size data sets, TERR is 2 to 10x faster than open source R for many common operations. On large data sets, where TERR performance and memory management really shine, it can be 10 to 100x faster than open source R, as shown in this example of fitting and scoring a Generalized Linear model.

<table>
<thead>
<tr>
<th>OS</th>
<th>R</th>
<th>TERR</th>
<th>SPEEDUP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>107.1 sec</td>
<td>17.5 sec</td>
</tr>
<tr>
<td>Model Fitting on 5 M rows</td>
<td>0.5 sec</td>
<td>1 sec</td>
<td>0.5 x</td>
</tr>
<tr>
<td></td>
<td>84.2 sec</td>
<td>1 sec</td>
<td>84.2 x</td>
</tr>
</tbody>
</table>

SPOTFIRE DATA CONNECTIONS
In addition to many Hadoop data sources, Spotfire and TERR can connect to many other enterprise data sources, enabling you to combine all your data and tackle analytic challenges. Refer to the full list of supported data sources: http://spotfire.tibco.com/resources/spotfire-data-sources