TIBCO Spotfire Data Science

Enterprise platform for advanced analytics on big data

BENEFITS

- Allows data scientists, data engineers, and business users to collaborate on big data analytic projects
- Provides an intuitive web-based interface to create data science and data pipeline workflows
- Runs the analytics within big data platforms without data movement using in-cluster technology
- Optimizes machine learning algorithms for parallel processing on Hadoop and traditional databases
- Supports the entire lifecycle, including data discovery, exploration, modeling, and deployment

TIBCO Spotfire® Data Science is an enterprise big data analytics platform that can help your organization become a digital leader. The platform allows data scientists, data engineers, and business users to collaborate on advanced analytics projects. These cross-functional teams can build machine learning workflows in an intuitive web interface with a minimum of code, while still leveraging the power of big data platforms.

Spotfire Data Science provides a complete array of tools (from visual workflows to Python notebooks) for the data scientist to work with data of any magnitude, and it connects natively to most sources of data, including Apache™ Hadoop®, Spark®, Hive®, and relational databases. The collaboration interface then allows the analytics team to share insights and data with the rest of the organization, white providing security and governance, driving action for the business.

DATA DISCOVERY & EXPLORATION

Spotfire Data Science connects to any data source remotely, allowing users to provision sandboxes and build models without needing to move their data. The platform dynamically indexes metadata about each project, creating a living data dictionary that is accessible by each stakeholder in the process.

DATA BLENDING AT HADOOP SCALE

The visual drag-and-drop interface allows business users and data scientists to interrogate their data without writing SQL queries or writing MapReduce, Scala, and R code. The interface contains specific operators for functions like value replacement and filtering, which allow the user to construct complex workflows to clean, blend, transform, and prepare their data process.

ADVANCED ANALYTICS AT SCALE

IN-CLUSTER PROCESSING

Spotfire Data Science algorithms are optimized to push computations into any analytical source. When users execute analytic workflows, the system's distributed execution engine stores the workflow logic and sends instructions to multiple database systems automatically. This capability allows analysts and scientists to run algorithms at scale without moving the data or optimizing their algorithms based on their database logic.

PARALLEL COMPUTING

The system was designed to fully utilize parallel compute clusters to run analytics efficiently. Its Parallel Workflow Optimizer optimizes an analytic workflow based on whether the environment is Hadoop or in-database, taking advantage of Spark, MapReduce, or SQL where it is appropriate. The distributed execution engine then pushes the code into the data platform in-parallel across the entire cluster.

ACCESSIBLE PREDICTIVE ANALYTICS

The platform contains a comprehensive collection of predictive data mining and modeling algorithms that empower businesses to manipulate, model, and leverage big data in a business cycle. The code-free interface guides users from data exploration and transformation, to predictive modeling and evaluation.

EXTENSIBILITY AND LEVERAGING OPEN SOURCE

Spotfire Data Science provides a flexible and extensible environment for advanced analytics. The publicly available SDK enables customers to build their own custom operators for non-standard transformations or proprietary algorithms. Additionally, the platform has deep integrations with R and Python for data science users who are seeking more flexibility in their analytics process.

EMBED INSIGHTS INTO THE BUSINESS

MANAGE ANALYTICS PROJECTS

Users can create development sandboxes, or workspaces, that can be shared with any key stakeholder. Each resource involved in the project can be attached to the workspace including workflows, data definition documents, and project plans. Workflow history provides a chronological summary of each version of a workflow within any given workspace, allowing users to revert to an earlier model version.

FOSTER CROSS-TEAM COLLABORATION

Spotfire Data Science provides a single platform for each stakeholder to interact with their data and provide comments on the status of each project. For the first time, business users can see the real-time status along the entire journey of an analytics project, without needing to use multiple tools or rely on various teams to deliver status reports.

DEVELOP OPERATIONAL MODELS

Users are able to easily develop models and push them into production via PMML from inside the platform. Workflows and scripts that are developed for R and SQL are stored for future use. The platform also offers an API extension for embedding Spotfire Data Science logic into applications and processes.

SYSTEM REQUIREMENTS & SELECTED PLATFORMS

WEB REQUIREMENTS

- Firefox
- Chrome

SERVER REQUIREMENTS

- Dedicated server
- Quad core CPU (Multiple recommended)
- 48GB of RAM or higher recommended
- 500GB storage (RAID 1 mirroring)

OPERATING SYSTEM

- RHEL/CENTOS

INTEGRATIONS

- R
- Python (Jupyter Notebooks)
- Tableau
- PMML
- PFA
- MADlib

SUPPORTED HADOOP DISTRIBUTIONS

- Amazon EMR
- Cloudera CDH
- Hortonworks
- MapR
- Pivotal HD
- IBM Big Insights

SUPPORTED DATA PLATFORMS AS DATA SOURCES

- Amazon Redshift
- Greenplum Database
- MySQL
- PostgreSQL
- Oracle Database (11g, Exadata)
- SAP HANA
- Teradata
- SQL Server
- Vertica

SUPPORTED DATA PLATFORMS AS ANALYTICAL SOURCES

- Amazon EMR
- Cloudera CDH
- Apache Hive
- Hortonworks
- MapR
- Pivotal HAWQ
- PIvotal HD
- Greenplum Database
- PostgreSQL
- Oracle Database (11g, Exadata)

TIBCO Spotfire® Data Science

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO fuels digital business by enabling better decisions and faster, smarter actions through the TIBCO Connected Intelligence Cloud. From APIs and systems to devices and people, we interconnect everything, capture data in real time wherever it is, and augment the intelligence of your business through analytical insights. Thousands of customers around the globe rely on us to build compelling experiences, energize operations, and propel innovation. Learn how TIBCO makes digital smarter at www.tibco.com.

04/26/18