



Highlights

- Helps improve performance and scalability significantly for even the most complex analytical problems
 - Provides an integrated, accessible predictive analytics platform designed to handle big data and improve outcomes
 - Users to implement machine learning algorithms on massive data
 - Distributes analytics processing into Hadoop environments with support for IBM® BigInsights® for Apache Hadoop, Hortonworks HDP, Cloudera CDH and MapR
 - Enables users to access structured and unstructured data from sources throughout your organization, such as traditional RDBMSs, Hadoop distributions, social media and more
 - Provides access to an ever-growing IBM SPSS® Predictive Analytics Community, which delivers additional information, assistance and extensions to help expand the value of SPSS Predictive Analytics
-

IBM SPSS Modeler and SPSS Analytic Server: Big data simplified

Overview

In today's business climate, achieving business success includes gaining deeper insight from data. Such insight enables analytics maturity, or the ability to move from a descriptive to a predictive and then to a prescriptive approach for solving the business problems that are associated with customer, operational, and threat and fraud analytics.

Big data presents a world of new opportunities but also new analytics challenges: analyzing streaming data, exploiting social media, making fast decisions from massive data volumes and more. Currently, unlocking the value of big data can be complex and requires expert skills. IBM aims to remove this complexity, make predictive analytics for big data accessible to business users and provide the solutions needed to capitalize on the promise of big data.

When IBM SPSS Modeler is combined with IBM SPSS Analytic Server, analysts can develop and deploy predictive analytics over big data without extensive technical skills. SPSS Modeler is a powerful data mining and text analytics workbench that helps you build accurate predictive models quickly and intuitively without the need for programming. Data mining is the process of using predictive techniques to uncover hidden patterns and trends to reveal insights that can improve decision-making and business processes. SPSS Modeler enables the creation of predictive models that organizations can use proactively and repeatedly to reduce costs and increase productivity.



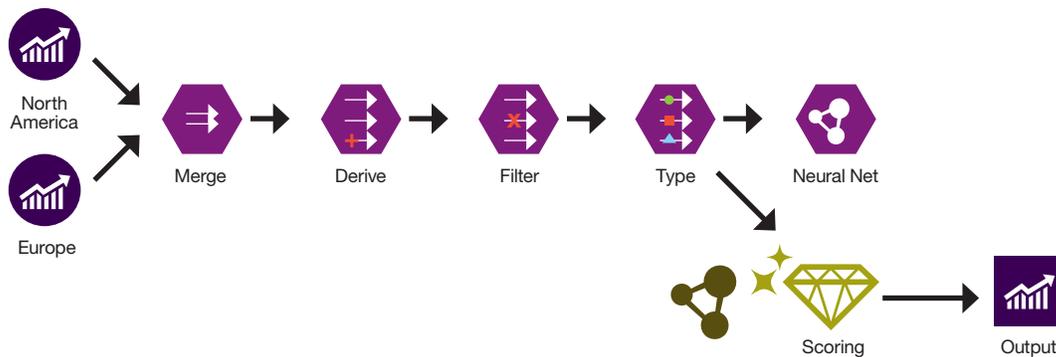


Figure 1: A sample SPSS Modeler and SPSS Analytic Server stream

The advantages of a combined solution

Together, SPSS Modeler and SPSS Analytic Server provide an integrated, accessible predictive analytics platform that enables you to use big data as a source for predictive modeling within SPSS Modeler (Figure 1). Users can discover insights in data stored in big data frameworks and traditional relational database management systems (RDBMS) without the need to write complex code or scripts.

In addition, this combined solution can help improve performance and scalability for even the most complex analytical problems while:

- Providing an integrated, accessible predictive analytics platform that is designed for big data and can help improve decision outcomes
- Distributing analytics processing into IBM BigInsights for Apache Hadoop and other common Hadoop environments¹
- Enabling users to access structured and unstructured data from sources throughout their organization, such as traditional RDBMSs, Hadoop distributions, social media and more
- Integrating with Apache Spark for efficient processing and modeling of big data using both native machine learning algorithms and those available through the Spark MLlib machine learning library

Providing an integrated, accessible predictive analytics platform

Predictive analytics includes running numerous iterations of the most relevant data to attain optimal results. When data sets grow to hundreds of millions of records, the time needed to perform such tasks becomes prohibitive. And the volume of data can be so large that it's impractical if not impossible to use traditional analytics platforms. As a result, accessing big data can be complex and requires specific skills and scripts.

The SPSS Modeler and SPSS Analytic Server solution is designed to support unstructured or semi-structured predictive analytics for big data. SPSS Modeler features a graphical interface that puts the power of data mining in the hands of business users, and with SPSS Analytic Server, enables you to combine structured and unstructured data to improve model accuracy. As a result, you can glean insights from your big data quickly and efficiently without complex programming packages or scarce expertise.

Distributing analytics processing into Hadoop environments

Hadoop is a Java-based programming framework that supports the processing of large data sets in a distributed computing environment. Hadoop makes it possible to run applications on systems with thousands of nodes involving thousands of terabytes. Its distributed file system facilitates rapid data transfer rates between nodes and enables the system to continue operating uninterrupted in case of a node failure. This approach lowers the risk of catastrophic system failure, even if multiple nodes become inoperative.

A main advantage of implementing a Hadoop distribution is the ability to scale. As a result, large amounts of data are being stored in the Hadoop environment. With the SPSS Modeler and SPSS Analytic Server combination, processing is distributed into the Hadoop environment, which eliminates the need to move data so you get optimal performance on large volumes of varied data. As a result, your organization can analyze enormous amounts of data by pushing the analytics to the data rather than taking the data to the analytics. And, most importantly, the solution supports multiple Hadoop platforms, including IBM BigInsights for Apache Hadoop, Hortonworks HDP, Cloudera CDH and MapR.

Enabling users to access data stored in sources throughout your organization

Not all data used for modeling is found in Hadoop. In many cases, it can also be found in traditional RDBMSs or flat files. The combined SPSS Modeler and SPSS Analytic Server solution enables you to access data from Hadoop and combine it with external data from other sources. As a result, you can not only discover insights in data stored in Hadoop distributions, but also through federation with a traditional RDBMS.

In addition, you can use the SPSS Modeler interface to add data sources to a Hadoop distribution. This and the accessibility to data wherever it is stored can help your business users get a complete picture and analyze all the data available to them that is relevant to the problem they are trying to solve.

Powering faster machine learning and real-time processing

IBM SPSS Modeler integrates with Apache Spark, an open source engine built specifically for data science, helping to simplify algorithm development and accelerate analytics results. Using Spark, you can better extract value from big data, conduct deeper analyses and deliver results faster, all while reducing the time and effort required for coding. Spark accelerates analytics on Hadoop, delivering the necessary speed and agility to data scientists and developers working with big data at scale.

This capability is particularly important for real-time stream processing and machine learning, which requires iterative computation, a task that is normally prohibitively time consuming with massive data. Spark provides in-memory computing, which speeds the analytical process and enables you to take advantage of big data machine learning algorithms available both natively within SPSS Modeler, and through the Spark machine learning library, MLlib.

Using the power of SPSS Analytic Server

SPSS Analytic Server enables the IBM predictive analytics platform to use data from Hadoop distributions and Spark applications to improve decisions and outcomes. SPSS Analytic Server provides an open, integrated data-centric architecture that uses big data systems and is scalable to problems of almost any size. It supports popular Hadoop distributions and features a defined interface that incorporates new statistical algorithms designed to go to the data. In addition, the familiar IBM SPSS user interface hides the details of big data environments so that analysts can focus on analyzing the data.

Conclusion

The combination of SPSS Modeler and SPSS Analytic Server provides an integrated, accessible predictive analytics platform that helps improve decision outcomes. Users of all levels can discover insights in data that is stored in Hadoop distributions, apply machine learning to big data with Apache Spark, and use all accessible information through federation with traditional RDBMS.

SPSS Modeler includes a visual interface that is supported by statistical algorithms, which can help you build accurate predictive models quickly and intuitively without the need for programming. SPSS Analytic Server provides support for unstructured or semi-structured predictive analytics in a Hadoop- or Spark-based environment, eliminating the need to move data for optimal performance on large volumes of data. Together, they can help you develop and deploy predictive analytics for big data without extensive technical skills or coding.

Learn more

Explore more predictive analytics software and resources at ibm.biz/predictive.

For more information on the advantages of combining the power of SPSS Modeler and SPSS Analytic Server, read the Ziff-Davis white paper, “[Big data, little data and everything in between: IBM SPSS solutions help you bring analytics to everyone.](#)”

Take advantage of our active and growing open-source community, where you can find resources to help you expand the use of IBM predictive analytics software. Resources include blogs, videos, tutorials, and an extensive library of more than 6,000 predictive extensions to help you take advantage of popular programming languages such as R, Python and Java. Join the community at: developer.ibm.com/predictiveanalytics

About IBM Analytics

IBM Analytics software delivers data-driven insights that help organizations work smarter and outperform their peers. This comprehensive portfolio includes solutions for business intelligence, predictive analytics and decision management, performance management, and risk management.

IBM Analytics solutions enable companies to identify and visualize trends and patterns in areas, such as customer analytics, that can have a profound effect on business performance. They can compare scenarios, anticipate potential threats and opportunities, better plan, budget and forecast resources, balance risks against expected returns and work to meet regulatory requirements. By making analytics widely available, organizations can align tactical and strategic decision-making to achieve business goals. For further information please visit ibm.com/analytics

Request a call

To request a call or to ask a question, go to ibm.com/analytics/contactus. An IBM representative will respond to your inquiry within two business days.



© Copyright IBM Corporation 2016

IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
February 2016

IBM, the IBM logo, ibm.com, BigInsights and SPSS are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

1 Refer to the Software Product Compatibility Reports available from the IBM website for information on supported operating environments. To create a report, visit: ibm.co/1Jqu9gT



Please Recycle