



Highlights

- Data scientists and their data science teams are tasked with turning raw data into meaningful insights using state of the art analytics. Doing so requires the best tooling, including open source innovation, coupled with social features for sharing and collaborating.
 - IBM Data Science Experience provides a one-stop shop for data science teams to learn about new tools and trends, create value using the best of open source and IBM, and collaborate on projects with each other and the broader data science community.
-

Master the art of data science with IBM Data Science Experience

Data is all around us, but it's up to data scientists and data science teams to turn that data into something meaningful and valuable. In order to accomplish this goal, it's very important that they have the right tools at their disposal.

When sourcing the tools they need, today's data scientists frequently turn to open source technologies. While these can be an important source of innovation and value, it can be difficult for a data scientist to assemble different open source capabilities into a single functioning environment. Instead, they typically end up with a disjointed collection of tools, filled with silos and bottlenecks. This disjointed experience may impair collaboration and productivity.

What today's data scientists and data science teams need is a data science platform. According to a recent Gartner report, an offering must be cohesive, with well-integrated building blocks that are consistent and interoperable, in order to be considered a true data science platform.¹ It's plain to see that a hastily assembled collection of open source tools simply wouldn't meet this definition.

IBM® Data Science Experience (DSX) is the collaborative platform that data scientists and data science teams need to be successful. With DSX, data scientists get all the tools they need to do their jobs, drawing from across open source and IBM technologies to create value for their business. In addition, data science teams get access to a community of data science peers, allowing them to collaborate and draw from shared resources such as data sets, notebooks and articles.



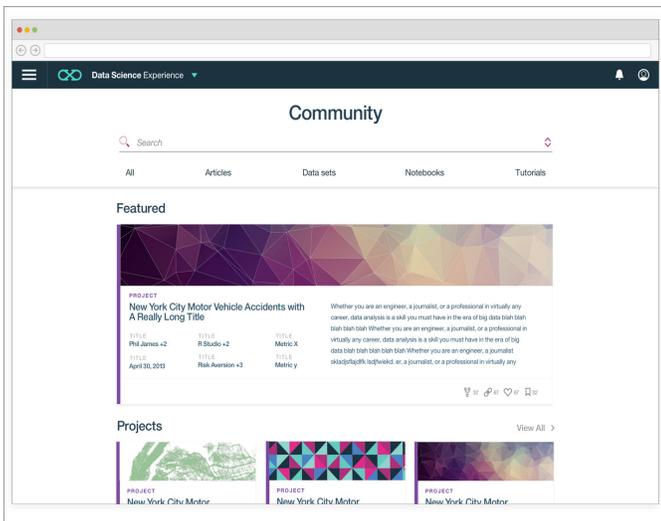


Figure 1: The community feature of IBM Data Science Experience connects users to expertise and makes it easier to get started, get help, learn a new approach, or get fresh data.

A growing set of open source tools for a preconfigured data science environment

DSX currently includes Apache Spark, Jupyter Notebooks and RStudio. This initial set of tools will continue to grow, but already provides a powerful foundation for data science.

Apache Spark, a fast and general engine for large-scale data processing, provides over 80 high-level operators, making it easy to build parallel apps; data scientists can use it interactively from the Scala, Python and R shells. All the machine learning libraries from Spark are included, as well as SparkR — a lightweight front end to use Spark from R, providing a distributed data frame implementation that supports operations like selection, filtering, aggregation, and more, but on large datasets.

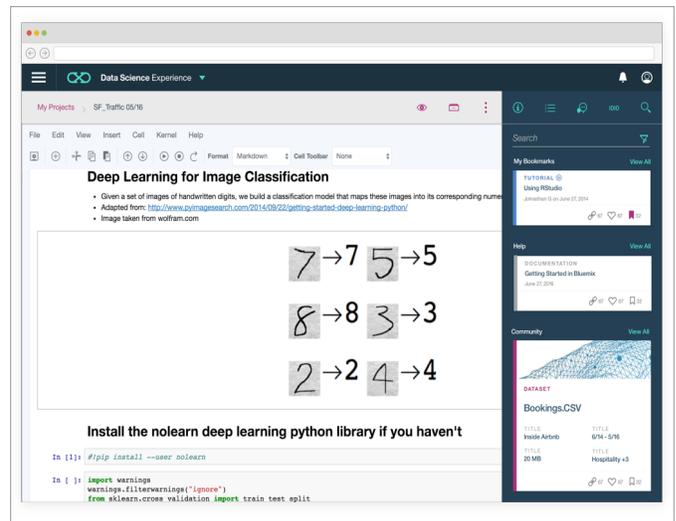


Figure 2: Notebooks are an interactive analysis and visualization tool that support multiple languages (Scala, Python, R) and are supported by integrated guidance within IBM Data Science Experience.

Data scientists can use Notebooks or RStudio for their analysis, and leverage both approaches within a project. Jupyter Notebooks enable data scientists to create and collaborate on Python, R and Scala notebooks that contain code and visualizations. DSX also provides access to RStudio — a popular open source integrated development environment (IDE) that allows for fast development of R scripts and, therefore, greater productivity.

Ultimately, the ability to use each of these languages and a variety of techniques is what makes DSX so unique. As explained in the next section, having these popular tools preconfigured in the environment saves data scientists time and effort that they can put toward more valuable endeavors, like the actual analysis of data.

Work with a wide variety of tools and data sources

Possibly the biggest value-added feature of DSX is that it is a platform that unifies numerous open source components to help data scientists and data science teams become better at what they do. Some of the challenges in using data science tools include installing, setting up, and maintaining them. Since DSX provides a preconfigured data science platform out of the box, users don't have to worry about performing these tasks.

DSX also supports a wide variety of data sources that can either be pulled from or connected to, including:

- Amazon Redshift
- Apache Hive
- Cloudera Impala
- IBM DB2®
- IBM Informix®
- IBM Netezza®
- IBM dashDB™
- IBM Watson™ Analytics
- Microsoft Azure
- Microsoft SQL Server
- MySQL
- Oracle
- Pivotal Greenplum
- PostgreSQL
- Salesforce.com
- Sybase
- Sybase IQ

Work smarter and faster as a team

DSX helps data scientists collaborate with peers on projects to find better solutions together. They can share their knowledge and code, help accelerate the advancement of data science for others, or get input from peers on their own work.

Data scientists can fork their notebooks and share them with the entire community to demonstrate successful approaches to others or get feedback on their work. In addition, DSX includes shared data sets and a wealth of tutorials and how-to articles to ensure that new data scientists and data science teams have what they need to get started. Experienced data scientists can also use these resources to try out new approaches.

Deployment options that work for you

IBM offers a number of options to help you get started with DSX.

Public cloud

Whether you're an individual user just looking to learn about data science, or an enterprise looking to deploy a robust, reliable and scalable set of data science tools, DSX on public cloud may be right for you. Deploying on the public cloud offers a simple way to start taking advantage of all the tools and data sources mentioned earlier, without the need to implement or manage your own infrastructure.

Private cloud

DSX on private cloud offers the same capabilities as the public cloud option, including facilitating collaboration and providing simplified access to commonly used tools. However, it also gives enterprises the ability to deploy behind their own firewall, which may be key for organizations with specific security needs. Enterprises deploying DSX on private cloud would build and manage their own infrastructure, giving them the power to ensure that all their security needs are being met.

Desktop

Individual users have the option to download a smaller version of DSX, which includes core capabilities from the complete platform.

For more information

To get started, visit datascience.ibm.com.

To learn more about IBM and data science, visit ibm.com/datascience.



© Copyright IBM Corporation 2017

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
April 2017

IBM, the IBM logo, ibm.com, dashDB, DB2, Informix, and Watson 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.

Netezza is a registered trademark of IBM International Group B.V., an IBM Company.

Microsoft is a trademark of Microsoft Corporation in the United States, other countries, or both.

Statements regarding IBM’s future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

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 NONINFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

1 “Magic Quadrant for Data Science Platforms”, Gartner, Inc. February 2017. <https://www.gartner.com/doc/3606026/magic-quadrant-data-science-platforms>.



Please Recycle