



Puntos destacados

- Los científicos de datos y sus equipos tienen la tarea de transformar los datos sin procesar en *insights* valiosos, utilizando análisis de última generación. Para ello, necesitan contar con las mejores herramientas, incluidas las innovaciones de código abierto, así como con recursos de redes sociales para poder compartir y colaborar.
 - IBM Data Science Experience proporciona una plataforma unificada que permite a los equipos de ciencia de datos obtener información sobre nuevas herramientas y tendencias, generar valor utilizando las mejores tecnologías de IBM y de código abierto, y colaborar en proyectos con otros equipos o con la comunidad más amplia de ciencia de datos.
-

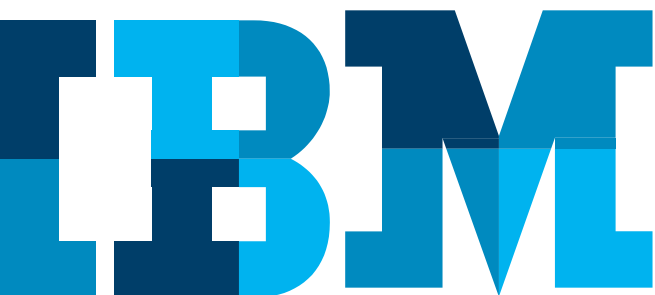
Domine el Arte de la Ciencia de Datos con IBM Data Science Experience

Los datos están en todas partes, pero les corresponde a los científicos de datos y a sus equipos transformarlos en algo comprensible y valioso. Y para que puedan lograr este objetivo, es muy importante que dispongan de las herramientas adecuadas.

Al tratar de obtener las herramientas que necesitan, los científicos de datos de hoy recurren con frecuencia a las tecnologías de código abierto. Si bien estas tecnologías pueden ser una fuente importante de innovaciones y valor, puede resultar difícil para los científicos de datos reunir diferentes capacidades de código abierto en un entorno único y eficaz. En lugar de eso, suelen terminar con un conjunto inconexo de herramientas, lleno de silos y cuellos de botella. Esta experiencia inconexa puede afectar la colaboración y la productividad.

Lo que los equipos de científicos de datos de hoy necesitan es una plataforma de ciencia de datos. Según un informe reciente de Gartner, un producto debe ser coherente, con componentes bien integrados, unidos e interoperables, para que se pueda considerar una verdadera plataforma de ciencia de datos.¹ Es fácil darse cuenta de que un conjunto de herramientas de código abierto ensamblado rápidamente no sería capaz de cumplir con esta definición.

IBM® Data Science Experience (DSX) es la plataforma colaborativa que los científicos de datos y sus equipos necesitan para tener éxito. Con DSX, los científicos de datos pueden obtener todas las herramientas que necesitan para realizar su trabajo, aprovechando tanto las tecnologías de código abierto como las tecnologías de IBM para generar valor para sus empresas. Además, los equipos de ciencia de datos pueden tener acceso a una comunidad de colegas científicos, lo que les permitirá colaborar y sacar provecho de recursos compartidos, como conjuntos de datos, blocs de notas y artículos.



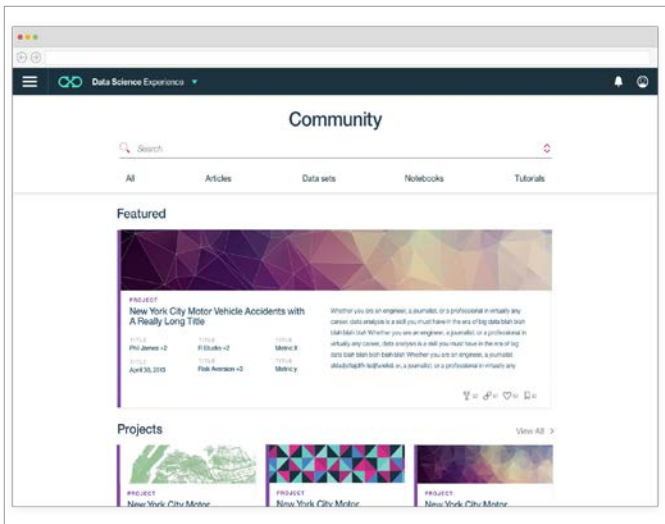


Figura 1: La función de comunidad de IBM Data Science Experience permite que los usuarios tengan acceso a conocimiento avanzado y hace que les resulte fácil comenzar, obtener ayuda, aprender nuevos enfoques u obtener datos actualizados.

Un conjunto creciente de herramientas de código abierto para un entorno preconfigurado de ciencia de datos

Actualmente, DSX incluye Apache Spark, Jupyter Notebooks y RStudio. Este conjunto inicial de herramientas continuará ampliándose, pero ya proporciona una base sólida para la ciencia de datos.

Apache Spark, un motor rápido y general para el procesamiento de datos a gran escala, proporciona más de 80 operadores de alto nivel, haciendo que sea fácil desarrollar aplicaciones paralelas; los científicos de datos pueden utilizarlo de manera interactiva a partir de shells Scala, Python y R. Se incluyen todas las bibliotecas de aprendizaje automático de Spark, así como también SparkR, un front-end ligero para el uso de Spark a partir de R que proporciona una implementación de dataframe distribuida, la cual admite operaciones como selección, filtrado, agrupación y mucho más, pero en grandes conjuntos de datos.

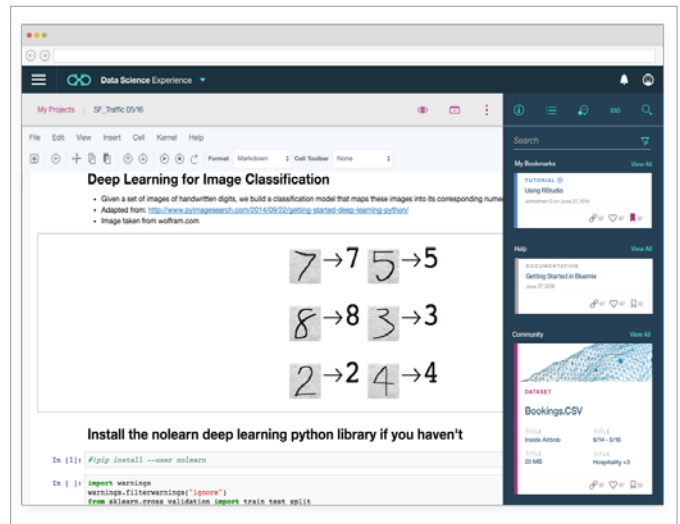


Figura 2: Los Notebooks son una herramienta interactiva de análisis y visualización que admite múltiples lenguajes (Scala, Python, R) y cuenta con el respaldo de una guía integrada en IBM Data Science Experience.

Los científicos de datos pueden utilizar Notebooks o RStudio para sus análisis, o aprovechar ambos enfoques en el ámbito de un proyecto. Los Jupyter Notebooks permiten que los científicos de datos creen y colaboren en Notebooks de Python, R y Scala que contienen código y visualizaciones. DSX también proporciona acceso a RStudio, un popular entorno de desarrollo integrado (IDE) de código abierto que posibilita un rápido desarrollo de scripts R y, por lo tanto, una mayor productividad.

En última instancia, la capacidad de utilizar cada uno de estos lenguajes y una variedad de técnicas es lo que hace que DSX sea tan especial. Como se explicará en la próxima sección, contar con estas herramientas populares preconfiguradas en el entorno hace que los científicos de datos ahorren tiempo y esfuerzo, los cuales pueden dedicarse a iniciativas más valiosas, como el análisis efectivo de los datos.

Trabaje con una amplia variedad de herramientas y orígenes de datos

Probablemente, la mejor característica de valor agregado de DSX es que se trata de una plataforma que unifica varios componentes de código abierto para ayudar a los científicos de datos y a sus equipos a mejorar su rendimiento. Algunos de los desafíos asociados a la utilización de herramientas de ciencia de datos incluyen su instalación, configuración y mantenimiento. Como DSX proporciona una plataforma de ciencia de datos preconfigurada y lista para usar, los usuarios no necesitan preocuparse por realizar estas tareas.

DSX también admite una amplia variedad de orígenes de datos, a partir de los cuales es posible obtener información o a los cuales es posible conectarse, incluyendo:

- 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

Trabaje en equipo de manera más inteligente y ágil

DSX ayuda a que los científicos de datos colaboren con colegas en sus proyectos para encontrar las mejores soluciones de manera conjunta. Los científicos pueden compartir conocimientos y código, ayudando a acelerar el avance de la ciencia de datos para los demás, o pueden recibir sugerencias de otros colegas con relación a su propio trabajo.

Los científicos de datos también pueden separar sus bloques de notas y compartirlos con toda la comunidad, para presentar enfoques exitosos a los demás u obtener comentarios sobre su trabajo. Además, DSX incluye conjuntos de datos compartidos y una variedad de tutoriales y artículos prácticos para garantizar que los nuevos científicos de datos cuenten con lo que necesitan para comenzar. Los científicos de datos experimentados también pueden utilizar estos recursos para probar nuevos enfoques.

Opciones de implementación adecuadas a sus necesidades

IBM ofrece varias opciones para ayudarle a empezar a utilizar DSX.

Nube pública

Independientemente de si usted es un usuario individual que desea aprender más sobre la ciencia de datos o si trabaja para una empresa que pretende implementar un conjunto sólido, confiable y escalable de herramientas de ciencia de datos, DSX en la nube pública puede responder a sus necesidades. La implementación en la nube pública ofrece una manera sencilla de empezar a sacar provecho de todas las herramientas y todos los orígenes de datos mencionados anteriormente, sin necesidad de implementar o gestionar una infraestructura propia.

Nube privada

DSX en la nube privada ofrece las mismas capacidades que la opción en la nube pública, incluyendo una colaboración fácil y un acceso simplificado a las herramientas utilizadas más frecuentemente. Sin embargo, también brinda a las empresas la capacidad de realizar implementaciones más allá de su firewall propio, lo cual puede ser fundamental para organizaciones con necesidades de seguridad específicas. Las empresas que implementan DSX en una nube privada pueden construir y gestionar su propia infraestructura, lo que les proporciona la capacidad de asegurarse de satisfacer todas las necesidades de seguridad.

Escritorio

Los usuarios individuales tienen la opción de descargar una versión más pequeña de DSX, que incluye las capacidades esenciales de la plataforma completa.

Más información

Para comenzar, visite datascience.ibm.com.

Para obtener más información sobre IBM y la ciencia de datos, visite ibm.com/datascience (contenido en inglés).



© Derechos Reservados IBM Corporation 2017

IBM Corporation
Route 100
Somers, NY 10589

Producido en los Estados Unidos de América
en abril de 2017

IBM, el logotipo de IBM, ibm.com, dashDB, DB2, Informix y Watson son marcas registradas de International Business Machines Corp. en varias jurisdicciones de todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas. Una lista actual de las marcas registradas de IBM está disponible en la Web, en “Copyright and trademark information” (Información sobre copyright y marcas registradas) en www.ibm.com/legal/copytrade.shtml.

Netezza es una marca registrada de IBM International Group B.V., una empresa de IBM.

Microsoft es una marca registrada de Microsoft Corporation en los Estados Unidos, en otros países, o en ambos.

Declaraciones con relación a intenciones y a la dirección futura de IBM están sujetas a cambios o anulación sin previo aviso, y representan solamente metas y objetivos.

Es responsabilidad del usuario evaluar y verificar la operación de cualquier otro producto o programa con los productos y programas de IBM.

LA INFORMACIÓN EN ESTE DOCUMENTO SE SUMINISTRA “COMO ESTÁ”, SIN NINGUNA GARANTÍA, EXPRESA O IMPLÍCITA, LO QUE INCLUYE GARANTÍAS DE COMERCIALIZACIÓN O ADECUACIÓN A UN PROPÓSITO ESPECÍFICO, Y CUALQUIER GARANTÍA O CONDICIÓN DE NO INFRACCIÓN. Los productos de IBM están garantizados conforme a los términos y condiciones de los contratos en virtud de los cuales se suministran.

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



Se ruega reciclar