



---

## Destaques

- Os cientistas de dados e suas equipes de ciência de dados são encarregados de transformar dados brutos em ideias significativas usando análise de última geração. Fazer isso requer a melhor ferramenta, incluindo inovação de código aberto, juntamente com recursos sociais de compartilhamento e colaboração.
  - O IBM Data Science Experience proporciona um ponto único para equipes de ciência de dados aprenderem sobre novas ferramentas e tendências, criarem valor utilizando o melhor do código aberto e da IBM, e colaborarem em projetos uns com os outros e com a comunidade maior de ciência dos dados.
- 

# Domine a arte da ciência de dados com IBM Data Science Experience

Os dados estão por toda parte, mas cabe aos cientistas de dados e às equipes de ciência de dados transformar esses dados em algo significativo e valioso. Para atingir esse objetivo, é muito importante que eles tenham as ferramentas certas à sua disposição.

Ao buscar as ferramentas de que precisam, os cientistas de dados atuais frequentemente se voltam para tecnologias de código aberto. Embora elas possam ser uma importante fonte de inovação e valor, pode ser difícil para um cientista de dados reunir diferentes recursos de código aberto em um único ambiente funcional. Em vez disso, eles geralmente acabam com um conjunto de ferramentas desarticuladas, repletas de silos e gargalos. Essa experiência desarticulada pode prejudicar a colaboração e a produtividade.

O que os cientistas de dados de hoje e as equipes de ciências de dados precisam é de uma plataforma de ciência de dados. De acordo com um relatório recente da Gartner, uma oferta deve ser coesa, com módulos bem-integrados que sejam consistentes e interoperáveis, para ser considerada uma verdadeira plataforma de ciência de dados.<sup>1</sup> Fica evidente que uma coleção de ferramentas de código aberto reunidas de maneira apressada simplesmente não corresponderia a tal definição.

O IBM<sup>®</sup> Data Science Experience (DSX) é a plataforma colaborativa que os cientistas de dados e equipes de ciência de dados precisam para ser bem-sucedidos. Com o DSX, os cientistas de dados obtêm todas as ferramentas que precisam para fazer seus trabalhos, utilizando tecnologias de código aberto e da IBM para criar valor para seus negócios. Além disso, as equipes de ciência de dados obtêm acesso a uma comunidade de colegas da ciência de dados, permitindo que colaborem e utilizem recursos compartilhados, como conjuntos de dados, notebooks e artigos.



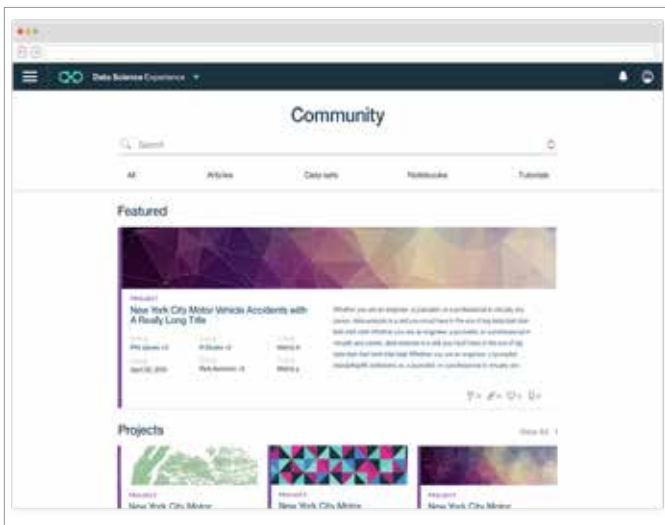


Figura 1: O recurso comunitário do IBM Data Science Experience conecta os usuários a conhecimentos e torna mais fácil começar, obter ajuda, aprender uma nova abordagem ou obter dados novos.

## Um conjunto crescente de ferramentas de código aberto para um ambiente de ciência de dados pré-configurado

O DSX atualmente inclui Apache Spark, Jupyter Notebooks e RStudio. Esse conjunto inicial de ferramentas continuará a crescer, mas já oferece uma base sólida para a ciência de dados.

O Apache Spark, um mecanismo rápido e geral de processamento de dados em larga escala, com mais de 80 operadores de alto nível, facilitando a criação de aplicativos paralelos; os cientistas de dados podem utilizá-lo interativamente a partir dos shells do Scala, Python e R. Todas as bibliotecas de aprendizado automatizado do Spark estão incluídas, bem como o SparkR – um front end leve para usar o Spark a partir do R, fornecendo uma implementação de estrutura de dados distribuída que suporta operações como seleção, filtragem, agregação e muito mais, mas em grandes conjuntos de dados.

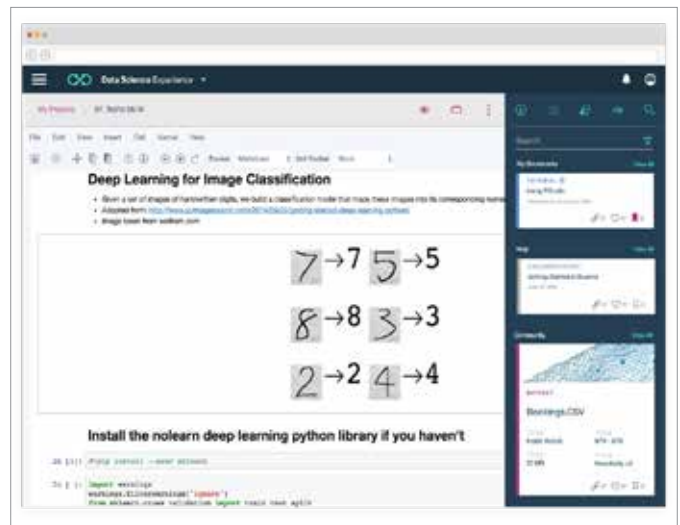


Figura 2: Os notebooks são uma ferramenta de análise e visualização interativa compatível com várias linguagens (Scala, Python, R) apoiados pela orientação integrada no IBM Data Science Experience.

Os cientistas de dados podem usar Notebooks ou RStudio para sua análise e aproveitar ambas as abordagens dentro de um projeto. Os Jupyter Notebooks permitem que cientistas de dados criem e colaborem em notebooks Python, R e Scala que contêm código e visualizações. O DSX também fornece acesso ao RStudio – um ambiente de desenvolvimento integrado (IDE, integrated development environment) popular de código aberto que permite o desenvolvimento rápido de scripts R e, portanto, maior produtividade.

Em última análise, a capacidade de usar cada uma dessas linguagens e uma variedade de técnicas é o que torna o DSX tão único.

Como explicado na próxima seção, ter essas ferramentas populares pré-configuradas no ambiente economiza tempo e esforço de cientistas de dados, os quais eles podem aplicar em trabalhos mais proveitosos, como a análise dos dados em si.

## Trabalhe com uma grande variedade de ferramentas e fontes de dados

Possivelmente, a maior característica de valor agregado do DSX é ser uma plataforma que unifica vários componentes de código aberto para ajudar os cientistas de dados e as equipes de ciência dos dados a se tornarem melhores no que fazem. Alguns dos desafios no uso de ferramentas de ciência de dados incluem instalar, configurar e fazer a manutenção delas. Como o DSX fornece uma plataforma pré-configurada de ciência de dados que vem pronta, os usuários não precisam se preocupar com a execução dessas tarefas.

O DSX também suporta uma grande variedade de fontes de dados que podem ser separadas ou conectadas, incluindo:

- 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

## Trabalhe em equipe com mais inteligência e rapidez

O DSX ajuda os cientistas de dados a colaborar com colegas em projetos para juntos encontrarem soluções melhores. Eles podem compartilhar seus conhecimentos e códigos, ajudar a acelerar o avanço da ciência de dados para os outros, ou receber opiniões de colegas sobre seu próprio trabalho.

Os cientistas de dados podem dividir seus notebooks e compartilhá-los com toda a comunidade para demonstrar abordagens bem-sucedidas para outras pessoas ou obter feedback sobre seu trabalho. Além disso, o DSX inclui conjuntos de dados compartilhados e uma variedade de tutoriais e artigos de instruções para garantir que os novos cientistas de dados e equipes de ciência de dados tenham o que precisam para começar. Cientistas de dados experientes também podem usar esses recursos para experimentar novas abordagens.

## Opções de implementação que funcionam para você

A IBM oferece várias opções para ajudá-lo a começar a utilizar o DSX.

### Nuvem pública

Quer você seja um usuário individual procurando apenas aprender sobre a ciência de dados, quer seja uma empresa desejando implementar um conjunto robusto, confiável e dimensionável de ferramentas de ciência de dados, o DSX em nuvem pública pode ser ideal para você. A implementação na nuvem pública oferece uma maneira simples de começar a aproveitar todas as ferramentas e fontes de dados mencionadas anteriormente, sem a necessidade de implementar ou gerenciar sua própria infraestrutura.

### Nuvem privada

O DSX em nuvem privada oferece os mesmos recursos que a opção de nuvem pública, incluindo a facilitação da colaboração e o fornecimento de acesso simplificado a ferramentas comumente usadas. No entanto, também dá às empresas a capacidade de implementação atrás de seu próprio firewall, o que pode ser fundamental para organizações com necessidades específicas de segurança. As empresas que implementam o DSX em nuvem privada criam e gerenciam suas próprias infraestruturas, dando-lhes o poder de garantir que todas as suas necessidades de segurança estejam sendo atendidas.

### Desktop

Usuários individuais têm a opção de baixar uma versão menor do DSX, que inclui os principais recursos da plataforma completa.

## Para obter mais informações

Para começar a usar, acesse [datascience.ibm.com](https://datascience.ibm.com).

Para saber mais sobre a IBM e a ciência de dados, visite [ibm.com/datascience](https://ibm.com/datascience).



---

© Copyright IBM Corporation 2017

IBM Corporation  
Route 100  
Somers, NY 10589

Produzido nos Estados Unidos da América  
Abril de 2017

IBM, o logotipo IBM, ibm.com, dashDB, DB2, Informix e Watson são marcas comerciais da International Business Machines Corp., registradas em muitas jurisdições em todo o mundo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual das marcas comerciais da IBM está disponível na web em “Copyright and trademark information” (Informações de copyright e marcas comerciais), no link [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Netezza é uma marca registrada da IBM International Group B.V., uma empresa da IBM.

Microsoft é uma marca registrada da Microsoft Corporation nos Estados Unidos, em outros países ou em ambos.

As declarações referentes a futuros projetos ou planos da IBM estão sujeitas a mudanças ou cancelamento sem aviso prévio, e representam apenas metas e objetivos.

É de responsabilidade do usuário avaliar e verificar a operação de todos os outros produtos ou programas com produtos e programas IBM.

AS INFORMAÇÕES DESTE DOCUMENTO SÃO FORNECIDAS "NA FORMA EM QUE SE ENCONTRAM", SEM NENHUMA GARANTIA, EXPRESSA OU IMPLÍCITA, INCLUINDO A NÃO GARANTIA DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM PROPÓSITO PARTICULAR E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO VIOLAÇÃO. Os produtos da IBM são garantidos de acordo com os termos e condições dos acordos sob os quais eles são fornecidos.

1 “Magic Quadrant for Data Science Platforms” (Quadrante Mágico para Plataformas de Ciência de Dados), Gartner, Inc. Fevereiro de 2017. <https://www.gartner.com/doc/3606026/magic-quadrant-data-science-platforms>.



Por favor, recicle

---