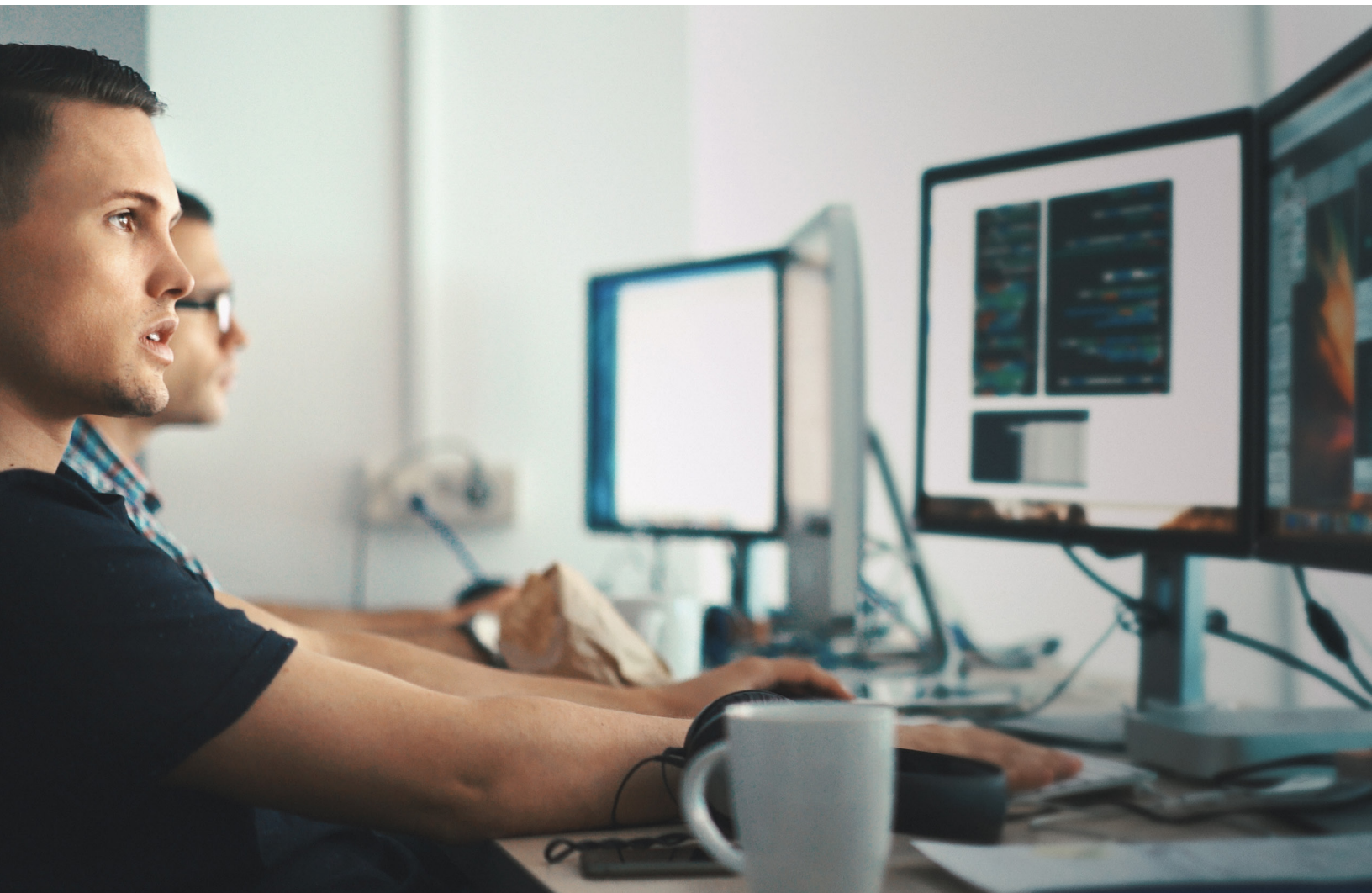




# Cloudera Data Science Workbench with IBM

Accelerate machine learning  
from research to production



## Accelerate data science and machine learning from research to production.

### For data scientists

- Experiment faster  
Use R, Python, or Scala with security-rich, self-service access to compute and data stored in CDH, Hortonworks Data Platform (HDP) or virtually anywhere.
- Work together using your favorite tools  
Collaborate and share reproducible work done in your preferred editing tools, including Jupyter Notebook, RStudio® and more.
- Deploy with confidence  
Get to production repeatably and without recoding.

### For IT professionals

- Bring data science to the data  
Give your data science team more freedom while reducing the risk and cost of silos.
- Security rich by default  
Leverage common security and governance across workloads.
- Run essentially anywhere  
On premises or in the cloud.

The Cloudera Data Science Workbench with IBM (CDSW) is an enterprise data science platform that accelerates data science and machine learning (ML) projects by providing a robust yet familiar environment for data science and data engineering with self-service access to governed business data. It provides a complete solution for data exploration, analysis, visualization, modeling and model deployment. CDSW makes security-rich, collaborative data science at scale a reality for the enterprise, accelerating the delivery of new data products and unifying IT, data science, developers, analysts and business stakeholders across the data science workflow.

#### Key features:

- Collaborative hub for enterprise data science with isolated projects, security-rich collaboration and simple dependency management
- Integrated development environment for Python, R and Scala with support for Spark and connectivity to security-rich CDH and HDP clusters, and data stored virtually anywhere
- Complete model lifecycle support with experiment tracking and model deployment designed for ease of use
- Lightweight analytics job and pipeline management system for data science and data engineering workloads that supports real-time monitoring, results tracking, and email alerting

# Cloudera Data Science Workbench with IBM

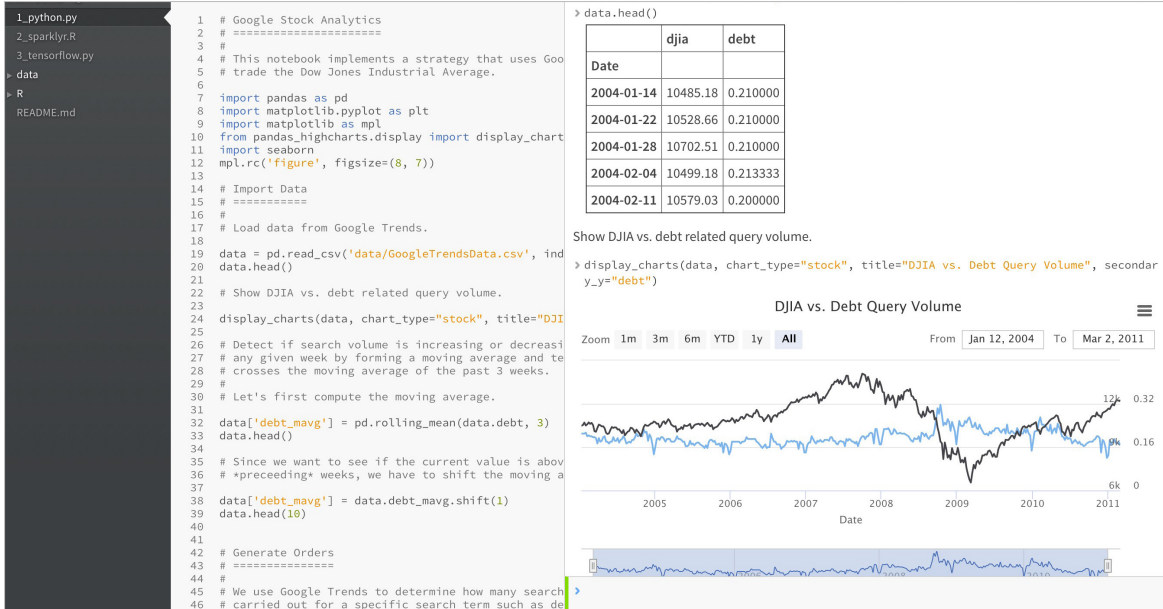


Figure 1: Work together using your favorite data science tools.

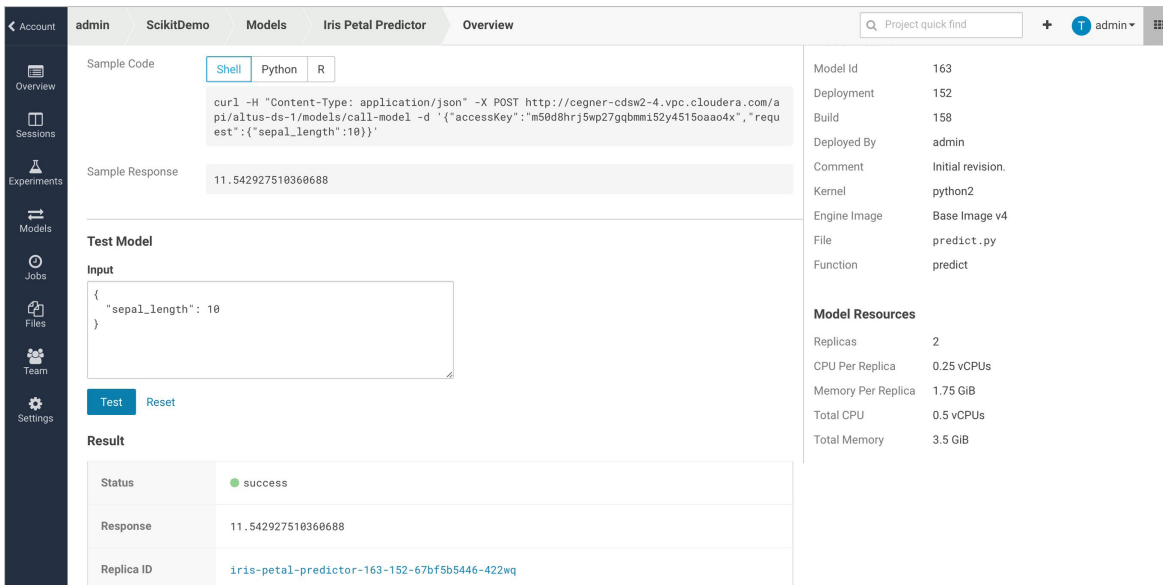


Figure 2: Collaborate and share reproducible work done in your preferred editing tools including Jupyter Notebooks, RStudio and more.

## Deploy trusted ML models faster using your preferred development tools

Data scientists need to bring tools to their data, whether it's stored in CDH, HDP or the cloud. Using a choice of editors, including Jupyter Notebook, RStudio and more, and languages, including Python, R or Scala directly from the web browser, CDSW delivers a rich, self-service experience for data scientists.

Download the latest libraries and frameworks in customizable project environments. Beyond the Python and R ecosystems, as open data science expands to deep learning frameworks like TensorFlow, PyTorch, Caffe 2, MX-Net, DL4J, BigDL, scikit-learn and more, CDSW is designed to deliver a safe, secure environment to combine the latest open source innovations with the unified platform Cloudera customers trust.

Collaborative, shareable project environments help ensure diverse data science teams can work together toward standard, reproducible research and production-ready models, and easily deploy and manage them across stakeholders and end users.

## Get better use of your core data management investments

IT groups often struggle to onboard data scientists to big data systems because of their diverse needs, especially where open source tools are involved. The result is duplication and analytic silos with limited security and governance. Meanwhile, data scientists look to scale their work to larger data sets and more powerful compute platforms. CDSW helps you break down analytic silos and drive value from your enterprise data platform, whether on premises or in the public cloud.

For modern IT organizations, CDSW can:

- Up-level your existing analytics stack with a collaborative environment that supports the latest open source languages and libraries.
- Get better use of your core data management investments while opening up new supported tools for data science
- Expand data science outside of siloed environments and enable self-service data science access.

## A complete data science solution, security rich and compliant by default

IT departments are responsible for compliance with corporate directives like security and governance. This process can be difficult even when every user is accessing your environment through a common interface, such as Structured Query Language (SQL). It becomes much harder when every team, user and project uses a different set of open source tools. Managing so many environment variations against a secured cluster is almost impossible. IT departments, forced to balance enterprise data security against the benefits of data science, are often compelled to lock the data up and the data scientists out.

Cloudera is helping IT groups and data scientists work together, bringing more users to shared environments in a way that delivers flexibility and compliance. CDSW allows data science work to happen within governed and secure CDH and HDP platforms. Users can achieve full data visibility and platform-wide governance, auditing and lineage.

## For more information

To learn more about Cloudera Data Science Workbench with IBM, visit the [IBM and Cloudera webpage](#) or [contact an IBM data management expert](#).

© Copyright IBM Corporation 2019

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
December 2019

IBM, the IBM logo, and [ibm.com](http://ibm.com) 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. "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

RStudio is a registered trademark of RStudio, Inc.

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.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.

