

Accelerate machine learning from research to production

CLUDERA DATA SCIENCE WORKBENCH



Accelerate data science and machine learning from research to production.

FOR DATA SCIENTISTS

- **Experiment faster**
Use R, Python, or Scala with secure, self-service access to compute and data stored in CDH, HDP or anywhere
- **Work together using your favorite tools**
Collaborate and share reproducible work done in your preferred editing tools including Jupyter Notebooks, RStudio and more
- **Deploy with confidence**
Get to production repeatably and without recoding

FOR IT PROFESSIONALS:

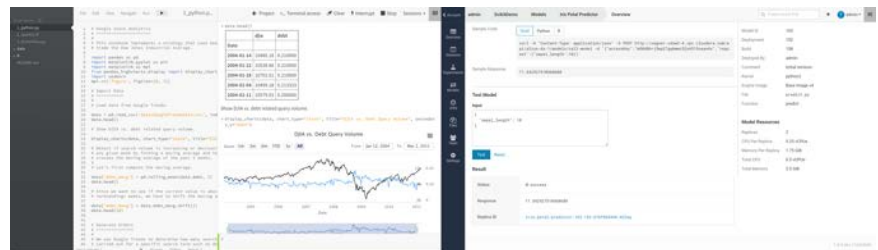
- **Bring data science to the data**
Give your data science team more freedom while reducing the risk and cost of silos
- **Secure by default**
Leverage common security and governance across workloads
- **Run anywhere**
On premises or in the cloud



The Cludera Data Science Workbench (CDSW) is an enterprise data science platform that accelerates data science and machine learning 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 secure, 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, secure collaboration, and simple dependency management
- Integrated development environment for Python, R, and Scala with support for Spark and connectivity to secured CDH and HDP clusters, and data stored anywhere
- Complete model lifecycle support with experiment tracking and easy model deployment
- Lightweight analytics job and pipeline management system for data science and data engineering workloads that supports real-time monitoring, results tracking, and email alerting



Deploy trusted machine learning 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 Notebooks, 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, MXNet, DL4J, BigDL, scikit-learn, and more, CDSW delivers a safe, secure environment to combine the latest open source innovations with the unified platform Cludera customers trust.

Collaborative, shareable project environments 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.

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

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, secure and compliant by default

IT is responsible for compliance with corporate directives like security and governance. This can be difficult even when every user is accessing your environment through a common interface, such as 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, forced to balance enterprise data security against the benefits of data science, is 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 both 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.