cloudera

DATA SHEET

CLOUDERA
DATA SCIENCE WORKBENCH

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 on-demand compute and secure CDH data access

_ **Work together**
Share reproducible research with your whole team

_ **Deploy with confidence**
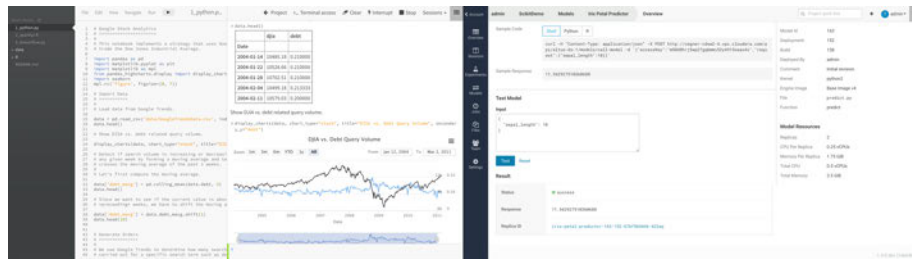Get to production repeatably and without recoding

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 Cloudera 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 model building with self-service access to 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 2 and connectivity to secured Hadoop clusters and any data store

_ Complete model lifecycle support with experiment tracking and easy model deployment

_ Lightweight job and pipeline system for data science workloads that supports real-time monitoring, results tracking, and email alerting

Deploy trusted machine learning models faster

Data scientists need to bring tools to their data, whether it's stored in CDH, HDP or the cloud. Using a choice of 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 Cloudera 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.

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.

Visit our website and learn more at

cloudera.com/products/data-science-and-engineering/data-science-workbench