

Databricks for Data Science

Accelerate Innovation through Unification

Businesses are generating data at a faster pace than ever: 90% of the world’s data was generated within the last two years. The increased data volume is rapidly outpacing our ability to consume it. Data science allows businesses to efficiently predict future outcomes, and even preemptively take action, based on insights from terabytes of business data.

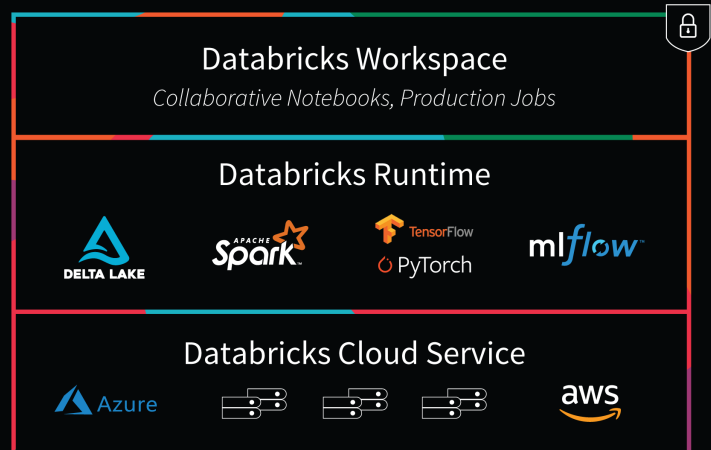
However, as the data continues to grow in volume, new challenges arise that can impede time-to-insight and innovation:

- Spending too much time maintaining infrastructure rather than the data.
- Complexity and cost to train machine learning models at scale.
- Poor collaboration among team members and across the organization.

By combining big data with data science techniques such as machine learning and deep learning, businesses can build and train scalable models that drive new and extraordinary business use cases.

Better Data Engineering with Databricks

Founded by the team who created Apache Spark™, Databricks provides a Unified Analytics Platform that accelerates innovation by unifying data science, engineering, and business. With Databricks, data engineers can securely and reliably deploy production data pipelines with ease.



Automated Infrastructure

Databricks’ serverless and highly elastic cloud service is designed to remove operational complexity while ensuring reliability and cost efficiency at scale, so you can focus on your data instead of DevOps. Through the first serverless API for Apache Spark, organizations can remove the barriers of infrastructure for both end-users and DevOps.

Key benefits of Databricks’ serverless infrastructure are:

AUTO-CONFIGURATION

The Spark version deployed in serverless pools is automatically optimized for interactive SQL and Python workloads.

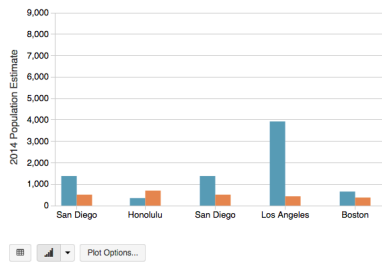
ON-DEMAND ELASTICITY

Databricks automatically scales the compute and local storage resources in the serverless pools in response to Apache Spark’s changing resource requirements for user jobs.

RELIABLE FINE-GRAINED SHARING

Serverless pools embed preemption and fault isolation into Spark, enabling a pool’s resources to be shared among many users in a fine-grained manner without compromising on reliability.

Top Cities by 2017 Median Sales Price



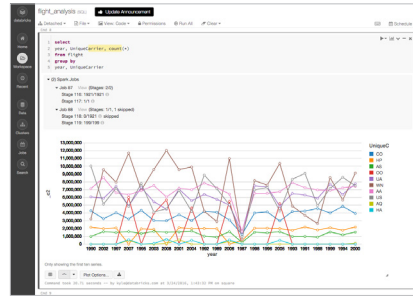
Add comment

dave_wang @ 5/14 3:48 PM
@parviz - why not try a bar chart?

parviz@da @ 5/14 3:48 PM
we need to add Boston as well

parviz@da @ 5/14 3:49 PM
ok I can try that @dave

parviz@da @ 5/14 3:49 PM
how about now?



Collaborative Workspace

Databricks provides an interactive workspace that eliminates the need to integrate third party tools and libraries. Support for multiple programming languages (R, Python, Scala, and SQL) ensures you use the right tool for the job. Improve team productivity by enabling team members to collaborate on the data and models in real time, while tracking usage through viewer logs and revision history.

Databricks' interactive workspace allows data science teams to leverage the following capabilities:

COLLABORATION

Built-in collaboration features to increase productivity across the entire data science team.

POINT-AND-CLICK VISUALIZATIONS

Data visualizations allow you to easily create and embed a wide range of point-and-click visualizations into your notebook or use powerful scriptable options like matplotlib, ggplot, and D3.

ROLE-BASED ACCESS CONTROL

Granular levels of permissions for notebooks which enables secured and managed shared collaboration with team members.

VERSION CONTROL

Revision history and Github integration for version control which is extremely useful when building notebooks for production and ad-hoc querying.

Interactive Dashboards

Turn your analysis from a notebook into a dynamic dashboard with one click. Databricks Dashboards allow you to easily share insights with your colleagues and customers, or let them run interactive queries with Spark-powered dashboards.

Key capabilities built into the Databricks Dashboards include:

ONE CLICK PUBLISHING

Create shareable dashboards from notebooks with a single click. One notebook can be tailored into multiple dashboard views.

CONTINUOUS UPDATES

Publish dashboards and schedule the content to be updated continuously.

PARAMETERIZED DASHBOARDS

Enable non-technical users to perform scenario analysis directly from published dashboards.

DASHBOARD WIDGETS

Input widgets allow you to parameterize your dashboards.

*We have this pipeline now where **instead of the three months** that it initially took us to deploy any machine learning model, **we can deploy new models within a day.***

— Craig Kelly, Group Product Manager 

Machine Learning and AI for the Masses

Databricks' Unified Analytics Platform takes the complexity out of data science at scale, allowing data scientists of all backgrounds and levels of experience to tap into the power of advanced analytic techniques such as machine learning and deep learning.



Get started with Databricks for data science today with a **free trial**.

