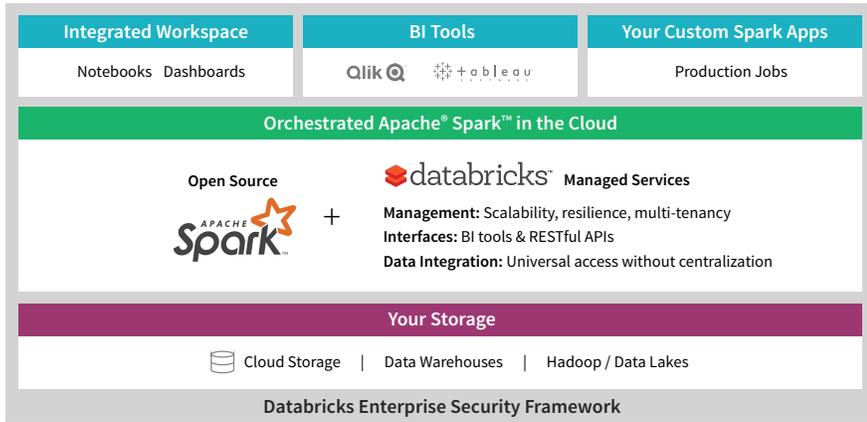


Making Big Data Simple with

THE IDEAL DATA PLATFORM

Databricks offers a cloud platform powered by Apache® Spark™ that makes it easy to turn data into value, from ingest to production, without the hassle of managing complex infrastructure, systems, and tools. A complete solution for data scientists and engineers.



DATABRICKS AVAILABLE FOR GOV TODAY

Databricks, a cloud-hosted, secure big data platform based on Apache Spark for Data Science and Engineering at scale, is available for the public sector today. Databricks runs natively on AWS, supports GovCloud, and is currently working towards assessment and authorization for use with classified workloads.



HOW DATABRICKS CAN HELP YOUR ORGANIZATION

<p>1 Just-in-Time Data Warehouse</p> <ul style="list-style-type: none"> • Connect to your data in AWS. • Integrate structured and unstructured data. • Scale compute independently from storage. <p>ETL acceleration</p>	<p>2 Zero Management Apache Spark</p> <ul style="list-style-type: none"> • Tuned and managed by Databricks Spark experts. • Self service launch and scale. • Deploy ML and streaming jobs <p>Production Spark Applications</p>	<p>3 Integrated Workspace</p> <ul style="list-style-type: none"> • Interactive notebooks, real-time dashboards. • Data exploration and visualization, model tuning. <p>Agile Data Science</p>
---	---	---

END-TO-END SECURITY

The Databricks Enterprise Security (DBES) Framework provides a holistic approach to solve the enterprise security challenge.

Encryption At-Rest and In-Flight	Integrated Identity Management	Role-Based Access Control	Data Governance	FedRAMP Compliance
				
				(coming soon)

Want to learn more about the Databricks platform or get started on a POC?

[CONTACT US](#)

databricks.com/contactgov

About Databricks

Databricks' vision is to empower anyone to easily build and deploy advanced analytics solutions. The company was founded by the team who created Apache® Spark™, a powerful open source data processing engine built for sophisticated analytics, ease of use, and speed. Databricks is the largest contributor to the open source Apache Spark project providing 10x more code than any other company. The company has also trained over 20,000 users on Apache Spark, and has the largest number of customers deploying Spark to date. Databricks provides a just-in-time data platform, to simplify data integration, real-time experimentation, and robust deployment of production applications. Databricks is venture-backed by Andreessen Horowitz and NEA. For more information, contact info@databricks.com.

+ Add Cluster

Name	Memory	Type	State	Notebooks	Dashboards	Options
Dedicated ETL cluster	1000 GB	On-demand	Running	2 Notebooks	Attached	<ul style="list-style-type: none"> Configure Restart Terminate
Exploration cluster - eng	100 GB	On-demand	Running	5 Notebooks	Make Dashboard Cluster	<ul style="list-style-type: none"> Configure Restart Terminate
ML cluster - Training	300 GB	On-demand	Running	12 Notebooks	Make Dashboard Cluster	<ul style="list-style-type: none"> Configure Restart Terminate

EFFORTLESSLY MANAGE LARGE-SCALE SPARK CLUSTERS

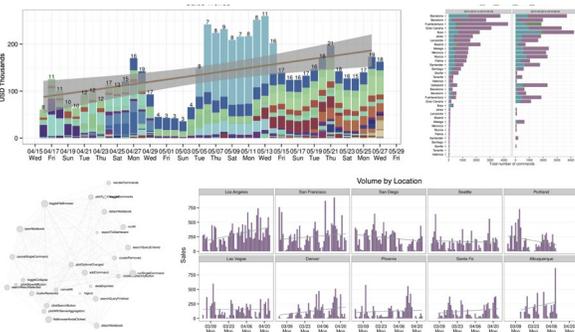
Spin up and scale out clusters to hundreds of nodes and beyond with just a few clicks, without IT or DevOps. Easily harness the power of Spark for streaming, machine learning, graph processing, and more.

prod rec - training set ingest

Task: Set Notebook / Set JAR
Cluster: 300 GB Spot, On-demand
Schedule: Every day at 1:00am (L)
 Advanced **Schedule Job**
 Every day at 02:00:00 US/Pacific
 Alerts: Edit
 On start: dave@databricks
 On success: dave@databricks.com
 On error: dave@databricks.com
 Timeout: 15 minutes Edit / Remove
 Retries: Limit 3x total, 15 min delay Edit / Remove

RUN YOUR PRODUCTION JOBS AT SCALE

Put new applications in production with one click by scheduling either notebooks or JARs. Monitor the progress of production jobs and set up automated alerts to notify you of changes.



PUBLISH YOUR ANALYSIS WITH CUSTOMIZED DASHBOARDS

Build and articulate your findings in dashboards in a few clicks. Set up dashboards to update automatically through jobs.

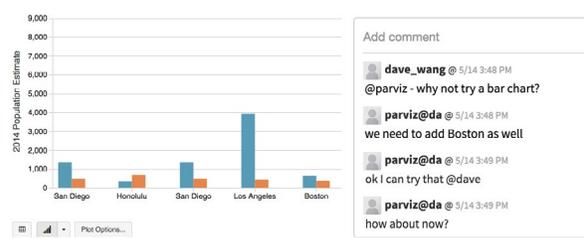
Mobile Devices by Geography



ACCELERATE YOUR WORK WITH AN INTERACTIVE WORKSPACE

Work interactively while automatically documenting your progress in notebooks — in R, Python, Scala, or SQL. Visualize data in just a few clicks, and use familiar tools like matplotlib, ggplot or d3.

Top Cities by 2015 Median Sales Price



COLLABORATE INTERACTIVELY

Seamlessly share notebooks, collaborate in the same code base, comment on each other's work, and track activities.



CONNECT YOUR FAVORITE APPS

Run your favorite BI tools or sophisticated third-party applications on Databricks.