

Databricks for the Public Sector

Big data is transforming the government sector. Insights buried deep within troves of data hold the key to improving public services and creating a more secure homeland. Yet, extracting these insights from diverse datasets quickly and efficiently is a challenge made even more complex by the strict regulatory environment. Databricks helps government agencies overcome these challenges with a Unified Analytics Platform built to handle the most sensitive data while simplifying data science with a scalable and collaborative platform.

Building a Better Government with Big Data and AI

Built by the creators of Apache Spark, the Databricks Unified Analytics Platform is a secure, cloud-based platform hosted on AWS and Azure that makes it easy to build, scale, and deploy artificial intelligence and machine learning applications in minutes.



DEFENSE AND PUBLIC SAFETY

Bolster national defense and law enforcement with advanced machine learning algorithms to uncover, predict and prevent criminal activities, cyber attacks and potential national threats.



SOCIAL SERVICES

Analyze public records data such as tax, financial, welfare claims, and public assistance statistics to identify new ways to collect revenue, reduce operating costs and fight fraud.



GEOSPACIAL AND IOT ANALYTICS

Process petabytes of data from GPS satellites, cell towers, and IoT devices to better predict weather patterns, expedite disaster recovery, improve urban planning, evaluate crop production and more.



HEALTH SERVICES

Asses healthcare needs by analyzing large-scale datasets (e.g. claims, hospital data, disease center reports, geographic data, etc) resulting in improved outcomes and lower costs.

Built Secure for the Public Sector



- Hosted on AWS and Azure with a tightly-integrated security model
- AWS Public Sector Partner with support for AWS GovCloud
- Authority to Operate in AWS region for the US Intelligence Community
- HIPAA Compliant; Working towards FedRamp
- Robust Encryption, User Management and Data Governance
- Trusted by 500+ companies and government entities

How Databricks Can Help Your Organization

JUST-IN-TIME DATA WAREHOUSE

- Connect to your data in AWS or Azure.
- Integrate structured and unstructured data.
- Scale compute independently from storage.

ZERO MANAGEMENT APACHE SPARK

- Tuned and managed by Databricks Spark experts.
- Self service launch and scale.
- Deploy ML and streaming jobs

INTEGRATED WORKSPACE

- Interactive notebooks, real-time dashboards.
- Data exploration and visualization, model tuning.

1 ETL Acceleration

2 Production Spark Applications

3 Agile Data Science

The Databricks Unified Analytics Platform

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.

+ Add Cluster

Name	Memory	Type	State	Notebooks	Dashboards	Options
Dedicated ETL cluster	1000 GB	On-demand	Running	2 Notebooks	Attached	Configure Restart Terminate
Exploration cluster - eng	100 GB	On-demand	Running	5 Notebooks	Make Dashboard Cluster	Configure Restart Terminate
ML cluster - Training	300 GB	On-demand	Running	12 Notebooks	Make Dashboard Cluster	Configure Restart Terminate

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.

< All Jobs

prod rec - training set ingest

Task: Set Notebook / Set JAR

Cluster: 300 GB Spot, On-demand

Schedule: Every day at 1:00am (L)

Advanced

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

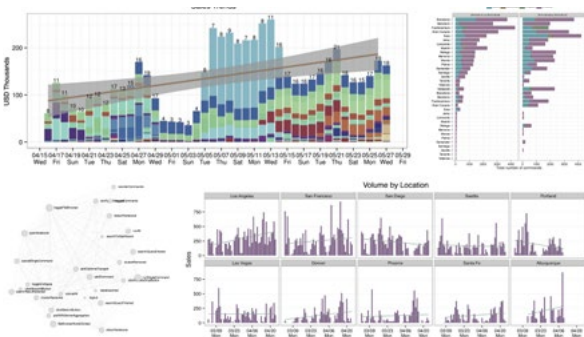
Schedule Job

Every day at 1:00am (L) US/Pacific

Cancel OK

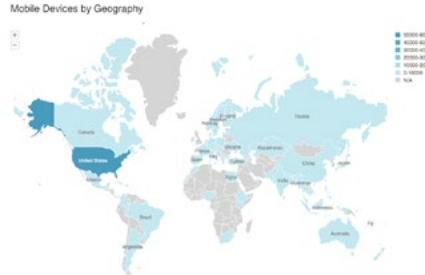
PUBLISH YOUR ANALYSIS WITH CUSTOMIZED DASHBOARDS

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



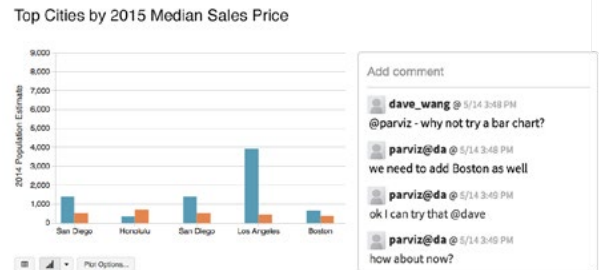
ACCELERATE YOUR WORK WITH AN INTERACTIVE WORKSPACE

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.



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.



Learn more about the Databricks Unified Analytics Platform or get started on a POC today at databricks.com/contactgov.

