The Unified Analytics Platform

Accelerating Innovation by Unifying Data Science, Engineering, and Business

Across industries today, companies are collecting more and more data of different kinds with the goal of mining insights that deliver tremendous business value. To drive this data-driven innovation across the company, you need various stakeholders to collaborate with each other through fast experimentation and validation. But, there are significant organizational and technological challenges that they need to overcome before they can create value.

Big Data and Analytics Challenges

**MASSIVE VOLUMES OF DIVERSE DATA**
Harnessing your data is impossible as it continues to grow across the business, while getting more distributed in different data stores across the cloud as well as your own data center.

**TOO MANY TECHNOLOGIES TO INTEGRATE**
Building infrastructure means identifying and integrating a myriad of technologies to make it all work. This forces data teams to spend more time on DevOps stitching together tools than on the data itself.

**INFRASTRUCTURE COMPLEXITY**
Big data infrastructure requires deep technical expertise and experience to setup, configure and fine-tune infrastructure to ensure high reliability and performance.

**INEFFICIENT, MANUAL WORKFLOWS**
The lack of automation between workflows to productionize models and distribute insights across the enterprise can cause error prone handoffs and significant slippage in innovation speed.

**DISJOINTED TEAMS THAT DON’T WORK TOGETHER**
The various teams that work with the data (data engineering, data science, business) have very different skills and reside in organizational silos which hurts cross-team collaboration and productivity.

**STRINGENT SECURITY AND COMPLIANCE**
The fragmented technology set supporting the analytics lifecycle and the increasing number of endpoints that needs to be secured makes it extremely hard for security stakeholders to protect one of the most valuable assets of the enterprise — its data.
Databricks Unified Analytics Platform
A Unified Approach to Data Analytics at Scale

Founded by the team who created Apache Spark, Databricks provides a platform that accelerates innovation by unifying data science, engineering, and business.

UNIFY ANALYTICS WITH APACHE SPARK
Eliminate the need for disparate tools.

INCREASE PRODUCTIVITY OF DATA SCIENCE TEAMS
With Databricks, they’ll be 4-5x more productive.

STREAMLINE ANALYTIC WORKFLOWS
Reduce deployment time to minutes.

REDUCE RISK
Enable innovation with out-of-the-box enterprise security and compliance.

Databricks’ unified platform has helped foster collaboration across our data science and engineering teams which has impacted innovation and productivity.

– John Landry, Distinguished Technologist at HP, Inc.
Accelerate Performance by 5x with Databricks Runtime

**OPTIMIZED SPARK**
The Databricks Runtime, built on top of Spark, is the highly tuned core of the Databricks Unified Analytics Platform that delivers data processing performance gains by up to 5x over vanilla Spark.

**DATABRICKS IO**
Leverages a vertically integrated stack to optimize the I/O layer and processing layer to significantly improve the performance of Spark in the cloud.

**DATABRICKS SERVERLESS**
A serverless architecture that democratizes infrastructure through the auto-configuration and fault-isolation of shared compute resources at scale — enabling best-in-class performance and reliability at dramatically lower costs.

**FULLY MANAGED IN THE CLOUD**
A cloud-native platform that abstracts the complexities of big data infrastructure and DevOps, resulting in a highly elastic, reliable and performant platform to build innovative products.

Increase Productivity by 4-5x through Interactive Data Science

**COLLABORATIVE WORKSPACE**
Speed up iterative model building and tuning with interactive notebooks purpose-built to instill collaboration across teams.

**SUPPORT FOR MULTIPLE PROGRAMMING LANGUAGES**
Interactively query large-scale data sets in R, Python, Scala, or SQL.

**BUILT-IN VISUALIZATIONS**
Visualize results with a wide variety of general-purpose data visualizations and publish insights for stakeholders via rich interactive dashboards.

**DASHBOARDS**
Share insights with your colleagues and customers, or let them run interactive queries with Spark-powered dashboards.

Streamline Processes from ETL to Production

**PRODUCTION WORKFLOWS**
A unified platform that streamlines end-to-end workflows from data ingest and ETL, to data exploration and model building, to productionizing models and data products.

**UNIFYING ALL ANALYTICS**
Move seamlessly across various types of analytics including batch, ad hoc, machine learning, deep learning, stream processing, and graph.

**ROBUST INTEGRATIONS**
Plug into a wide variety of AWS tools and data stores with built-in connectors and integrate with other data engineering services to facilitate CI/CD with comprehensive APIs.

Ensure Enterprise Security and Compliance

**STRONG DATA ENCRYPTION**
Benefit from best-in-class data protection at rest and in motion.

**INTEGRATED IDENTITY MANAGEMENT**
Seamless integration with enterprise identity providers via SAML 2.0 and Active Directory.

**ROLE-BASED ACCESS CONTROLS**
Fine-grained management access to every component of the enterprise data infrastructure, including files, clusters, code, application deployments, and dashboards.

**MONITOR AND AUDITING**
Tap into comprehensive audit logs to monitor and troubleshoot issues.

**COMPLIANCE STANDARD**
Databricks has successfully completed SOC 2 Type 2 certification and can offer a HIPAA-compliant solution.

**SECURE DEPLOYMENT**
Deploy in your own AWS VPC for full control over your data.
How Data Engineers and Data Scientists Get Value from Databricks

### Feature Importance for Data Engineers

<table>
<thead>
<tr>
<th>Feature Importance</th>
<th>Data Engineers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration and Automation</td>
<td>✔️</td>
</tr>
<tr>
<td>Interactivity and Collaboration</td>
<td>✔️</td>
</tr>
<tr>
<td>Performance</td>
<td>✔️</td>
</tr>
<tr>
<td>Simplicity / Ease of Use (Serverless Infrastructure)</td>
<td>✔️</td>
</tr>
<tr>
<td>Security</td>
<td>✔️</td>
</tr>
<tr>
<td>Expertise and Support</td>
<td>✔️</td>
</tr>
</tbody>
</table>

### Feature Importance for Data Scientists

<table>
<thead>
<tr>
<th>Feature Importance</th>
<th>Data Scientists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration and Automation</td>
<td>✔️</td>
</tr>
<tr>
<td>Interactivity and Collaboration</td>
<td>✔️</td>
</tr>
<tr>
<td>Performance</td>
<td>✔️</td>
</tr>
<tr>
<td>Simplicity / Ease of Use (Serverless Infrastructure)</td>
<td>✔️</td>
</tr>
<tr>
<td>Security</td>
<td>✔️</td>
</tr>
<tr>
<td>Expertise and Support</td>
<td>✔️</td>
</tr>
</tbody>
</table>


Data science and AI is transforming the way businesses across all industries deliver meaningful business value. The challenge is taking control of the data at scale to create competitive differentiation and grow revenue. The Databricks Unified Analytics Platform is designed to make AI approachable for all enterprises by eliminating the complexities of infrastructure, streamlining inefficient workflows and processes, and creating an environment that fosters collaboration across data teams in order to accelerate data-driven innovation.

Try Databricks for free: databricks.com/try-databricks
Contact us for a personalized demo: databricks.com/contact-databricks

© Databricks 2017. All rights reserved. Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation. Privacy Policy | Terms of Use