

Our Databricks Partnership

Unified data, analytics and AI that accelerates innovation

Businesses today collect a massive amount of data from multiple sources and from across domains and functions. They need to build and run big data jobs quickly and easily, with multiple processors running simultaneously. They also need to eliminate data redundancy, ensuring quality, reliability, and integrity of their data while providing analytics that help improve forecasting and outcomes.

Databricks can help businesses manage and analyze their big data using applications that require large-scale data processing, such as real-time analytics, machine learning (ML), and data warehousing.

Our Databricks Accelerator

As, AI Enablers, leverage our subject matter expertise and long-standing partnership with Databricks to utilize our comprehensive suite of services to expedite your digital transformation journeys. Our Databricks Accelerator includes:

▶ Data Architecture

Using Databricks, we design data architectures that provide access to high quality, relevant and accessible data, while reducing costs and improving efficiencies.

▶ Database Platform Migration

We help our clients migrate their data assets from legacy platforms to Databricks and stay with them throughout their migration journey—from assessment to production.

▶ Solutions Implementation

As AI Enablers recommend futuristic data architecture to enable clients implement the Apache Spark's AI advances effectively.

▶ Streamline Data Governance Framework with Unity Catalog

Unity Catalog allows configuring role-based access controls, row-level and column-level access controls, dynamic data masking, data encryption, enabling data lineage tracking, and ensuring secure data sharing within Databricks services helping to enable compliance needed. Our Data Governance experience with Unity Catalog expertise, efficiently establish Data Governance framework for an enterprise.

Databricks Advantage

The Databricks Lakehouse Platform combines the best elements of data lakes and data warehouses to deliver reliability, strong governance, and performance of data warehouses with the openness, flexibility, and ML support of data lakes.

The system platform can verify the data, ingest it into in-house systems, process data and run business logic. This platform is powered by Apache Spark, a unified analytics engine that enables distributed computing on multiple compute nodes with implicit data parallelism and fault tolerance.

The Databricks Lakehouse Platform integrates easily with popular BI tools like Tableau and Power BI. This enables users to quickly visualize data or explore the results generated by ML models. Moreover, Databricks' ability to return complex queries in a few seconds enables teams to build live responsive analytics dashboards for business teams.

Our Expertise

- Develop analytics solutions on Databricks platform using SparkSQL and PySpark to process and transform big data.
- Integrate existing cloud and on-premise data storages to build a unitform data processing architecture using Databricks.
- Enable advanced analytics using leading Enterprise BI tools like Power BI, Tableau, for faster data analysis and visualizations.
- Orchestrate big data pipelines powered by Databricks streamlining and accelerating the data processing for analytics.
- Automate big data analytics using powerful ML models enhancing data science team's efficiencies. Powered by the Spark engine the ML lifecycle is automated for several sub components like model training, validations, deployment.

Addressing Industry Problems with Databricks

We have helped many businesses to leverage Databricks. Here's a glimpse of the type of problems we have helped our clients address:

Pharmaceutical

Client Challenges

- Data inconsistencies and errors sourced from various sources (i.e., IoT equipment, laboratory machines, and pharma lifecycle data from external vendors). No one version of the truth for timely decision-making pertaining to market strategy and products.

Our Solution

- Leveraged Databricks to ingest nearly 30 Terabytes of market scan data for analysis from varied sources.
- Assessed and transformed data inconsistencies using PySpark on Databricks cluster to generate comprehensive report.
- Built live Tableau dashboards enabling real-time queries.

Client Challenges

- No mechanism for ingestion of medical data generated via IoT devices to the client's cloud system for the real time analytics for usage patterns, refills, maintenance etc.

Our Solution

- Used Azure Data Factory pipelines to bring the data generated by medical devices to the source system.
- Leveraged Azure Data Factory to automate the entire process of data ingestion, transformation and load to existing tables to obtain real time insights on Tableau dashboard.

Client Challenges

- Extensive human efforts needed in reviewing marketing material on a particular drug by running an ML model.

Our Solution

- Automated process using Azure Data Factory pipeline and Databricks for new files ingestion
- Built an ML model to review data and project the results in a Tableau dashboard.
- Implemented a write-back mechanism on the Tableau dashboard to capture feedback from users and push back to the database.

Asset Management

Client Challenges

- Disparate data for Customer, Distributor and Products data
- Significant processing time to refresh data, resulting in higher operational costs.
- Data latency in sharing the trusted data to Analytics Applications and CRM

Our Solution

- Azure and Databricks based Delta Lakehouse Medallion Architecture with Data Governance layer leveraging Databricks Unity Catalog
- Progressively refined data with functional Customer, Distributor Datamart using PySpark and SPARK SQL based data transformations and orchestrations
- Tailor-sized SQL warehouses for downstream applications as per the need to optimize the operating costs

Databricks Implementation Options

Blue Altair's certified Databricks experts help clients expedite their data processes and analytics so business teams can gain accurate insights and make timely decisions.

Options	Source Systems	Workflows	Reports	Lakehouse Deployment	Duration	Estimated Cost
Starter	Up to 2 sources	2-4 simple transformations	2 reports	Non-production environment	10 business days	\$25K
Comprehensive	Up to 4 sources	Workflows for each source and associated transformations	Up to 10 reports	Production deployment Included	10-12 weeks	\$75K
Tailored	No Limit	Workflows for each source and associated transformations including custom transformations	No Limit	Production deployment Included	Custom Duration - Contact Blue Altair	Custom Pricing - Contact Blue Altair

About Blue Altair

Blue Altair is a niche, industry-recognized business and technology consulting firm that assists our clients with digital transformations. We offer Assessment and Strategy, Technology Implementation, and Managed Services in API Management and Integration; Data Management; Digital Application Development; and Data Science and AI. Our Client Success capability ensures a higher-than-industry rate of successfully delivered projects, with a primary focus on program and project management, business analysis, and quality assurance. Blue Labs is our innovation hub, where we use cutting-edge technology to build offerings that deliver accelerators and solutions. Our culture is the heart of our existence, and our core values are the key drivers for our handpicked, top-tier performers.

Blue Altair plays a pivotal role alongside enterprise’s IT and business teams to ensure critical business pain points are addressed, and enterprise business goals are achieved.

Contact us to learn more about how Blue Altair can help implement Databricks in your organization.

info@bluealtair.com | www.bluealtair.com

1530 Wilson Blvd., Suite 650
Arlington, Virginia 22209 USA
+1 703 721 7275



United States | India | Australia | Argentina