



databricks

G-Cloud 15

Service Definition

Databricks Data Intelligence Platform

Lot 2b – Software as a Service (SaaS)

1. Introduction	4
Company Overview	4
Value Proposition	4
Customer References	7
What the Service Provides	8
Social Value	10
Climate Change	10
Overview of the G-Cloud Service	12
Associated Services	24
2. Data Protection	27
Information Assurance	27
Physical and Environmental Controls	29
Privacy by Design	33
3. Using the service	35
Ordering and Invoicing	35
Pricing Overview	36
Trial Service	38
On-Boarding, Off-Boarding, Service Migration, Scope etc.	38
Training & Enablement	40
Implementation Plan	43
Service Management	46
Service Constraints	48
Service Levels	49
Outage and Maintenance Management	52
Financial Recompense Model for not Meeting Service Levels	52
4. Provision of the service	52
Customer Responsibilities	52
Technical Requirements and Client-Side Requirements	52
Outcomes/Deliverables	52
Development life cycle of the solution	53
After-sales Account Management	54
Termination Process	55
5. Our experience	57
Clients	57
Case Studies	57
Contact Details	60



1. Introduction

Company Overview

With origins in academia and the open source community, Databricks was founded in 2013 by the original creators of Apache Spark™, Delta Lake and MLflow. All seven co-founders, at the time of conceiving Databricks, were UC Berkeley researchers, who were able to successfully bank on the proposition: when combined with A.I., data holds the promise of curing diseases, saving lives, reversing climate change, and even changing the way we live.

As the world's first and only lakehouse platform in the cloud, Databricks combines the best of data warehouses and data lakes to offer an open and unified platform for data and AI.

Databricks is the founder of three of the most popular open source projects Apache Spark, Delta Lake and MLFlow and has pioneered the Lakehouse paradigm which provides a simple, open platform capable of handling all data workloads, from ETL through to advanced Data Science and Machine Learning. Available on all 3 main public cloud platforms (AWS, Microsoft Azure and Google Cloud Platform) and based on open source technologies, processes and data within Databricks are inherently portable.

More than 9,000 organisations worldwide — including Ordnance Survey, Department for Education, DEFRA, US Air Force, European Maritime Safety Authority and NASA — rely on Databricks to enable massive-scale data engineering, collaborative data science, full-lifecycle machine learning and business analytics.

Headquartered in San Francisco, with offices around the world and hundreds of global partners, including Microsoft, Amazon, Tableau, Informatica, Capgemini, Accenture, Deloitte and EY, Databricks is on a mission to simplify and democratise data and AI, helping data teams solve the world's toughest problems.

Value Proposition

Data and data-related capabilities play a crucial role in informing policy decisions within the public sector, offering insights that can enhance governance, improve service delivery, and address societal challenges more effectively.

Evidence-based Decision Making: Data provides empirical evidence that policymakers can use to understand societal trends, assess the impact of existing policies, and identify areas requiring intervention. By analysing data on demographics, economics, health, education, crime, and more, policymakers can make informed decisions grounded in evidence rather than intuition. Databricks platform enables users to process country level data volumes that are traditionally too big to handle with ease. The platform provides scalable compute that does not require code changes to scale up with increased data volumes.

Predictive Analytics: Advanced data analytics techniques, such as predictive modelling and machine learning, enable policymakers to forecast future trends and anticipate potential outcomes of policy interventions. By analysing historical data and identifying patterns, governments can predict the potential effects of various policy options before implementation, thereby minimising risks and maximising positive impacts. Databricks provides end to end capabilities for managing ML/AI/GenAI workloads leveraging open source technologies such as MLFlow for easy exporting/importing of models in/out of the platform fully embracing FAIR principles.

Performance Monitoring and Evaluation: Data allows policymakers to monitor the performance of government programs and initiatives in real-time. By tracking key performance indicators (KPIs) and conducting comprehensive evaluations, policymakers can assess the effectiveness and efficiency of policies, identify areas for improvement, and make necessary adjustments to achieve desired outcomes. Databricks platform provides seamless integration of streaming data into data pipelines, the code used to process batch and streaming data has minimal differences and only where explicitly required. This in fact allows for build once pipelines and use in both batch and streaming use cases. Data coming from streaming sources can easily be integrated with historical data to build both retrospective, predictive and real-time reporting.

Targeted Interventions: Data-driven insights enable policymakers to target interventions more precisely, ensuring that resources are allocated efficiently and effectively. By identifying vulnerable populations, geographic areas with the highest need, or specific risk factors, policymakers can tailor policies and programs to address underlying issues and achieve greater impact. Databricks platform provides state of the art ETL tooling and allows for easy slice and dice over various data assets that can come in any number of open formats. The data can be processed and served using python, R, SQL or scala as programming languages.

Transparency and Accountability: Making data publicly available promotes transparency and accountability in government decision-making processes. By publishing data on government spending, performance metrics, and policy outcomes, policymakers can engage citizens, civil society organisations, and other stakeholders in the policymaking process, fostering greater trust and accountability. Databricks provides out of the box sharing service – delta sharing, through delta sharing organisations can easily provide any data product both internally and externally without the need to manually build Rest API services. This in effect is increasing transparency and shortening the sharing cycle.

Cross-Sector Collaboration: Data can facilitate collaboration across different sectors and levels of government. By sharing data and insights across agencies, departments, and jurisdictions, policymakers can develop holistic solutions to complex challenges that require coordinated efforts across multiple domains. Databricks platform both through Delta Sharing and through Databricks Marketplace can expose data assets and data derivatives such as notebooks, code, dashboards, models, etc. to organisations both in Public and Private sectors. These assets can be shared free of charge or could be monetized if there is a need for monetization.

Risk Management and Disaster Response: Data enables governments to assess and mitigate risks, particularly in areas such as public health, natural disasters, and national security. By analysing data on disease outbreaks, weather patterns, or social vulnerabilities, policymakers can develop proactive strategies to mitigate risks, enhance resilience, and improve emergency response.

Generative AI and Agentic Systems: Databricks is a unified platform that enables organizations to build and deploy quality Agent Systems. Mosaic AI is built on the same platform as the lakehouse, secured by Unity Catalog which gives end-to-end governance for all data and AI assets, combined with a robust way to build, deploy, and evaluate agents. This allows organizations to deliver high quality GenAI applications using Agent Systems.

On the infrastructure side, a common challenge across public sector organisations is the time and effort required to produce the data and insights required for decision and policy makers. This is driven in part by a complicated IT and data infrastructure that has evolved over many years across many different generations of technologies, and a wider market challenge of attracting the required skills. The Databricks Lakehouse Platform simplifies your data processing architecture, reduces the data infrastructure management overhead and provides a centralised security and permissions and model for controlling and monitoring access to data assets, such as PII or machine learning models. This creates a significant administrative and maintenance burden, running technologies that are outdated, hard to find talent for, and no longer represent the best solution to the problem they address. Furthermore, these disparate systems create distinct silos of data within the organisation that make it difficult and time consuming to consolidate information for the analysis and insights necessary to inform decision and policy makers. In these highly federated environments, Databricks can act as a single point of entry for data users by peering into different cloud tenancies or on-prem servers. This can reduce the administrative burden on granting data access requests, by consolidating all requests into a single place. Combined with security and governance policies and monitoring this can democratise data across the organisation whilst also ensuring the controls who can access what data are maintained and monitored. This allows an organisation to reap the benefits of increased data access, reduced time to insight and a reduced administrative burden without requiring an initial long term migration.

Another common challenge faced across the public sector is the attraction and retention of engineering and analytical talent with modern skillsets such as python, combined with an existing talent pool of SQL and Excel users. Databricks allows the former to build using the latest data tooling, whilst for the latter, product features such as Databricks SQL provide streamlined experience native within Databricks which abstracts away the underlying technologies such as Spark and data lakes, without sacrificing performance and cost benefits.

To simplify the learning curve and ease the journey to the cloud, product features such as Databricks SQL offers a route to cloud for users with traditional Business Intelligence

or Data Warehousing skills such as SQL. Because the platform is accessible to users of all skill levels, this leads to reduced friction when migrating users off their traditional tools and patterns, such as working in Excel on their local laptop in an ungoverned and unaudited fashion to the cloud where data access and use is easily controlled and monitored.

Customer References

NHS Digital utilises the Databricks Lakehouse to consolidate 8PB of clinical data, facilitating the provision of publication data for planning purposes, and ensuring secure researcher access. In Population Health, it offers national and official statistics on health system usage, along with a Vaccination API for COVID-19. In Research and Innovation, it enables direct data sharing through Trusted Research Environments (TREs) to support cross-sector research and provides data services for clinical trials. Data access is ensured through Data Science and Business Intelligence (BI) integration, allowing secure analysis within the NHS while supporting various use cases. Its scalable and elastic architecture provides rapid processing power, reducing operational maintenance costs through consistency and automation. Notably, during the COVID-19 pandemic, the Shielded Patient List facilitated the identification of individuals at higher risk of severe illness, with Databricks significantly enhancing its performance compared to legacy systems.

The **Department for Education (DfE)** implemented a new system which automatically collects daily attendance data from schools. This method of data collection is revolutionary for the department and, because it is automated, it creates no additional burden for schools. Using Databricks schools, trusts, local authorities, and the PM's office are provided with daily reports with powerful visualisations. These allow any issues with individual pupils or groups, such as vulnerable children and those eligible for free school meals, to be easily identified and action taken. The Databricks Lakehouse makes it possible to report accurate and reliable data across 18k+ schools in the UK, bringing a 45x improvement in processing efficiency over previous solutions and fine-grained governance for mission-critical data.

Ordnance Survey has been working closely with Databricks on the architectural strategy and data engineering capabilities that underpin the Core Data Services Platform. This platform enables OS to migrate geospatial data processing that has traditionally been carried out on on-prem machines in single-threaded processes to a more distributed cloud compute platform, that is available and scalable on-demand – thus, achieving the processing and analysis of geospatial data at scale. OS is using Azure Databricks to add Apache Spark™ capability to the cloud platform, and this brings the opportunity to re-think how to optimise both data and approach to perform geospatial joins at scale using parallelized processing. Ordnance Survey data outputs are countrywide and handle complex objects such as geometries describing building, greenspaces, roads, or aerial imagery that can be as fine grained as 12.5 cm size pixels. This data can be very difficult to handle using traditional geospatial tools. In databricks we have access to an end to end data platform that brings together capabilities for all



stages of data processing including geospatial, ml, AI, ETL, etc. Combining these tools we were able to improve the speed of joining UPRNs to OS Master Map Building shapes over 70 times.

What the Service Provides

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

This unified approach simplifies your modern data stack by eliminating the data silos that traditionally separate and complicate data engineering, analytics, BI, data science and machine learning. It's built on open source and open standards to maximise flexibility. And, its common approach to data management, security and governance helps you operate more efficiently and innovate faster.

Databricks covers the following aspects of your data platform:

- **Data Sharing:** Users can share live data sets — as well as models, dashboards and notebooks — across platforms, clouds and regions without dependencies on specific data-sharing services, including Databricks. This flexibility is achieved with strong security and governance by integrating Delta Sharing with Databricks Unity Catalog to govern, track and audit access to shared data sets. This unlocks secure and efficient collaboration within and across organisational boundaries.
- **Data Engineering:** Databricks Data Intelligence Platform allows easy ingestion and transformation of batch and streaming data. Databricks automatically manages your infrastructure at scale and allows users to orchestrate reliable production workflows. Built-in data quality testing and support for software development best practices Increase the productivity of your teams. This simplification allows teams to focus more on high value activities,.
- **Data Governance:** Databricks Unity Catalog offers a unified governance layer for data and AI within the Databricks Data Intelligence Platform. With Unity Catalog, organisations can seamlessly govern their structured and unstructured data, machine learning models, notebooks, dashboards and files on any cloud or platform. Data scientists, analysts and engineers can use Unity Catalog to securely discover, access and collaborate on trusted data and AI assets, leveraging AI to boost productivity and unlock the full potential of the lakehouse architecture. This unified approach to governance accelerates data and AI initiatives while simplifying regulatory compliance.
- **Data Warehousing:** Databricks SQL is a serverless data warehouse built on lakehouse architecture that lets you run all your BI and ETL workloads at scale with up to 12x better price/performance, a unified governance model, open formats and APIs, and your tools of choice — no lock-in.
- **AI/ML:** Databricks Mosaic AI provides unified tooling to build, deploy and monitor AI and ML solutions — from building predictive models to the latest GenAI and large language models (LLMs). Built on the Databricks Data Intelligence Platform,

Mosaic AI enables organisations to securely and cost-effectively integrate their enterprise data into the AI lifecycle.

- **Data Science:** Streamline the end-to-end data science workflow — from data prep to modelling to sharing insights — with a collaborative and unified data science environment built on an open lakehouse foundation. Get quick access to clean and reliable data, preconfigured compute resources, IDE integration, multi-language support, and built-in advanced visualisation tools for maximum flexibility for data analytics teams.
- **Real-time Streaming:** The Databricks Data Intelligence Platform dramatically simplifies data streaming to deliver real-time analytics, machine learning and applications on one platform. Enable your data teams to build streaming data workloads with the languages and tools they already know. Simplify development and operations by automating the production aspects associated with building and maintaining real-time data workloads. Eliminate data silos with a single platform for streaming and batch data.
- **Data Marketplace:** Databricks Marketplace is an open marketplace for all your data, analytics and AI, powered by open source Delta Sharing standard. The Databricks Marketplace expands your opportunity to deliver innovation, and advance all your analytics and AI initiatives.

Overview of the G-Cloud Service

The Databricks Data Intelligence Platform allows your entire organisation to use data and AI. It's built on a lakehouse to provide an open, unified foundation for all data and governance, and is powered by a Data Intelligence Engine that understands the uniqueness of your data.

Databricks combines generative AI with the unification benefits of a lakehouse to power a Data Intelligence Engine that understands the unique semantics of your data. This allows the Databricks Platform to automatically optimise performance and manage infrastructure in ways unique to your business.

The Data Intelligence Engine understands your organisation's language, so search and discovery of new data is as easy as asking a question like you would to a coworker. Additionally, developing new data and applications is accelerated through natural language assistance to write code, remediate errors and find answers.

Data and AI applications require strong governance and security, especially with the advent of generative AI. Databricks provides an end-to-end MLOps and AI development solution that's built upon our unified approach to governance and security. You're able to pursue all your AI initiatives — from consuming frontier AI models provided by OpenAI, Anthropic and others directly on the platform, to custom-built models — without compromising data privacy and IP control.

Data Intelligence Platform is an all-purpose data platform, currently deployed in a number of Public Sector organisations including Ordnance Survey, DEFRA, The Home Office, and Department for Education. The use cases Databricks is deployed in range from high volume and complexity geospatial processing, peatland identification, real time reporting for immigration and border security, and creating a single analytical front door for over 1000 analysts with a diverse array of skills across a highly siloed data estate.

Data sharing: Delta Sharing, developed by Databricks and the Linux Foundation, offers an open-source solution for sharing data across data, analytics, and AI platforms.

This enables users to share live data sets, models, dashboards, and notebooks across platforms, clouds, and regions, without reliance on specific data-sharing services like Databricks. Integration with Databricks Unity Catalog ensures robust security and governance, allowing for the management, tracking, and auditing of access to shared data sets.

With Delta Sharing's open approach, users can efficiently leverage their data assets for quicker insights discovery.

Furthermore, Delta Sharing facilitates the discovery and monetization of a wide range of data assets, including data sets, notebooks, ML models, applications, and dashboards, through an open marketplace. This broadens opportunities for innovation and revenue generation by delivering transformative data products.



Additionally, Delta Sharing enables users to securely source high-quality data both internally and externally, fostering collaboration at scale with customers, suppliers, and partners. Centralised governance, auditing, and monitoring of data usage are facilitated through Unity Catalog. Collaboration with multiple partners is supported through scalable Databricks Clean Rooms, allowing for confidential computation in any language on existing lakehouse data. In the image above we describe the Delta Sharing Protocol, and how Delta Sharing can provide an open, governed protocol for sharing data assets, through popular tools and languages such as PowerBI, Java, Pandas, Spark and across multiple clouds (Azure, AWS & GCP).



Data Engineering: Within the Databricks Data Intelligence Platform, batch and streaming data ingestion and transformation are efficiently handled. The platform automates workflow orchestration and supports team productivity by incorporating integrated data quality testing and adherence to software development best practices.

Data is seamlessly ingested into the platform to support analytics, AI, and streaming applications. Auto Loader simplifies file processing in cloud storage, managing new files without the need for state management. Additionally, it automates schema inference and adaptation over time. Analysts can perform batch file ingestion into Delta Lake using the SQL COPY INTO command. We also offer out of the box connectors to tools such as SQL server, Salesforce and Service Now with Lakeflow connect.

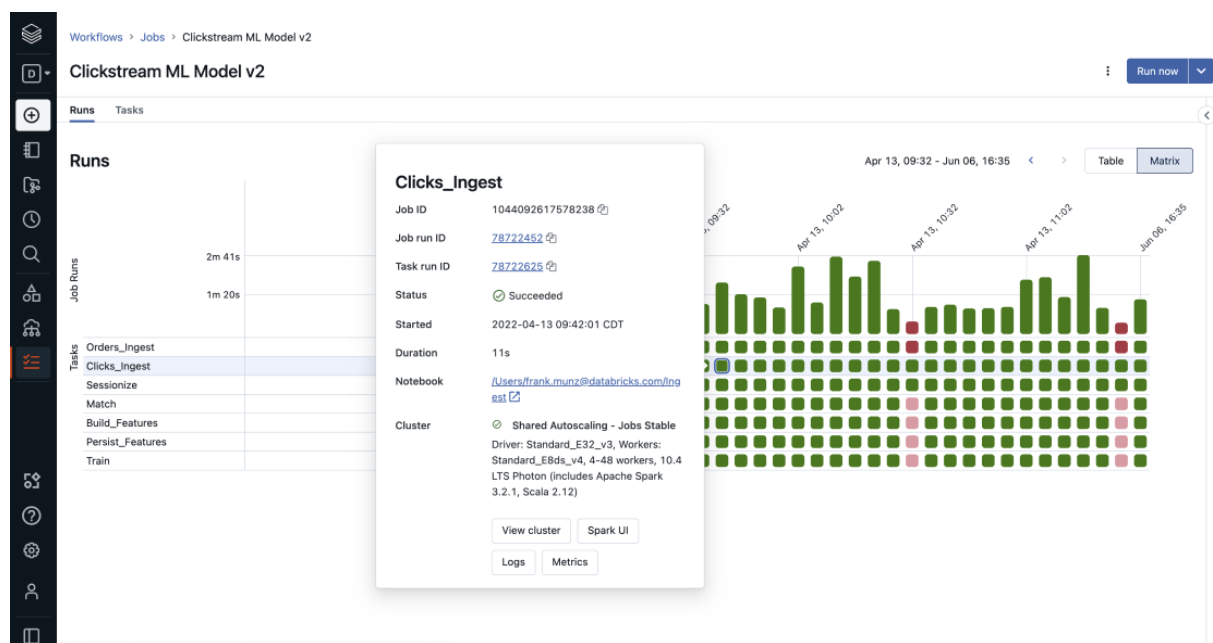


Fig 1 Databricks Jobs Interface

Following ingestion, raw data is transformed to prepare for analytics and AI usage. Databricks provides robust ETL capabilities through Lakeflow Declarative pipelines, enabling the creation of pipelines for both batch and streaming data. Lakeflow handles operational complexities such as infrastructure management, task orchestration, error handling, and performance optimization. Engineers can treat data as code, applying software engineering best practices to deploy reliable pipelines at scale.

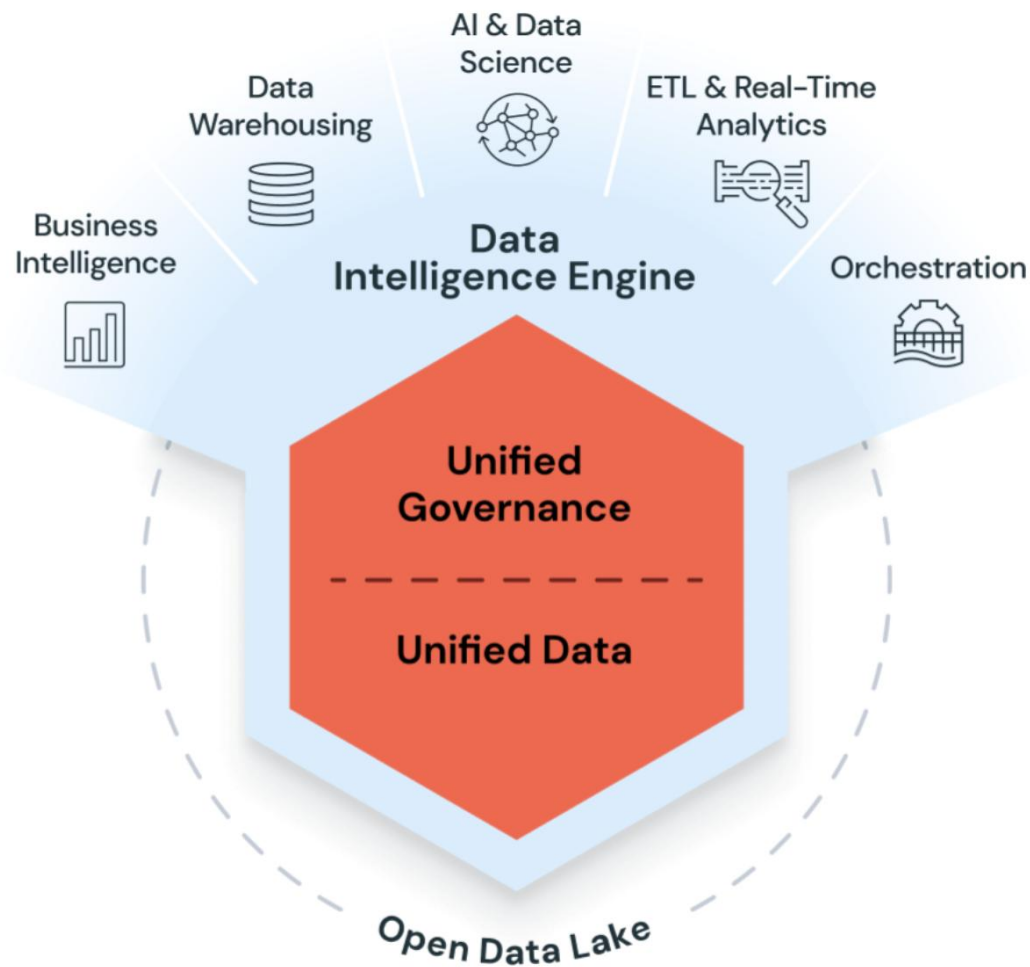


Fig 2 Databricks Tooling Overview

Data engineering on Databricks relies on foundational components like Unity Catalog and Delta Lake. Delta Lake optimises raw data storage with ACID transactions and high performance, while Unity Catalog provides detailed governance for all data and AI assets. This streamlined approach simplifies governance, ensuring consistent data discovery, access, and sharing across clouds. Moreover, Unity Catalog offers native support for Delta Sharing, facilitating secure data sharing with other organisations through an open protocol.

Governance: The Databricks Unity Catalog serves as a centralised governance hub for managing data and AI assets within the Databricks Data Intelligence Platform. It enables organisations to effectively oversee a wide range of data types, machine learning models, notebooks, dashboards, and files across various cloud environments and platforms. Through the Unity Catalog, AI Engineers, data scientists, analysts, and engineers can securely explore, access, and collaborate on trusted data and AI

resources, leveraging AI capabilities to enhance productivity and maximise the benefits of the lakehouse architecture. This unified governance approach streamlines the execution of data and AI initiatives while ensuring adherence to regulatory compliance requirements.

With Unity Catalog, users can efficiently discover and categorise diverse data assets, including structured and unstructured data, Generative AI models, ML models, notebooks, dashboards, and files, across different cloud environments. The platform enables seamless consolidation and querying of data from multiple sources, such as MySQL, PostgreSQL, Amazon Redshift, Snowflake, Azure SQL, Azure Synapse, and Google's BigQuery, providing users with a unified interface for data exploration and analysis. This centralised access point accelerates data and AI projects while facilitating secure search, comprehension, and insight extraction through natural language capabilities.

Access management is simplified through Unity Catalog's unified interface, which allows users to define and enforce access policies for data and AI assets consistently across various cloud platforms and data repositories. The platform supports secure data access from different computing environments through open interfaces, while offering granular control over data access at the row and column levels through attribute-based access policies. Security measures are further enhanced by automated monitoring and error diagnosis capabilities, which enable proactive identification of personally identifiable information (PII) data and model drift detection for timely issue resolution.

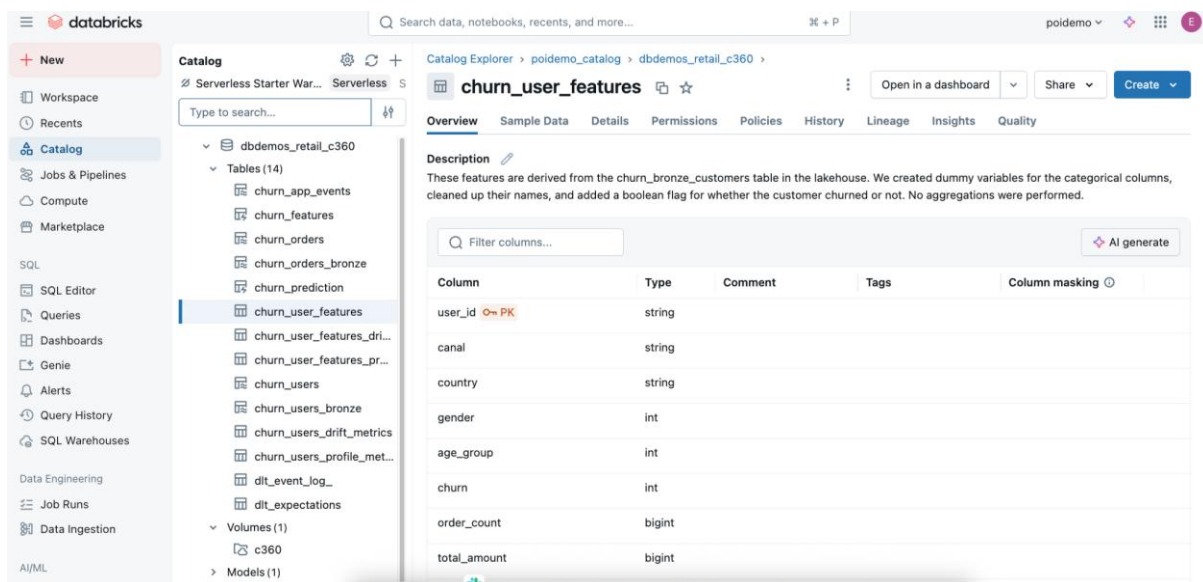


Fig 3 Unity Catalogue Data Exploration

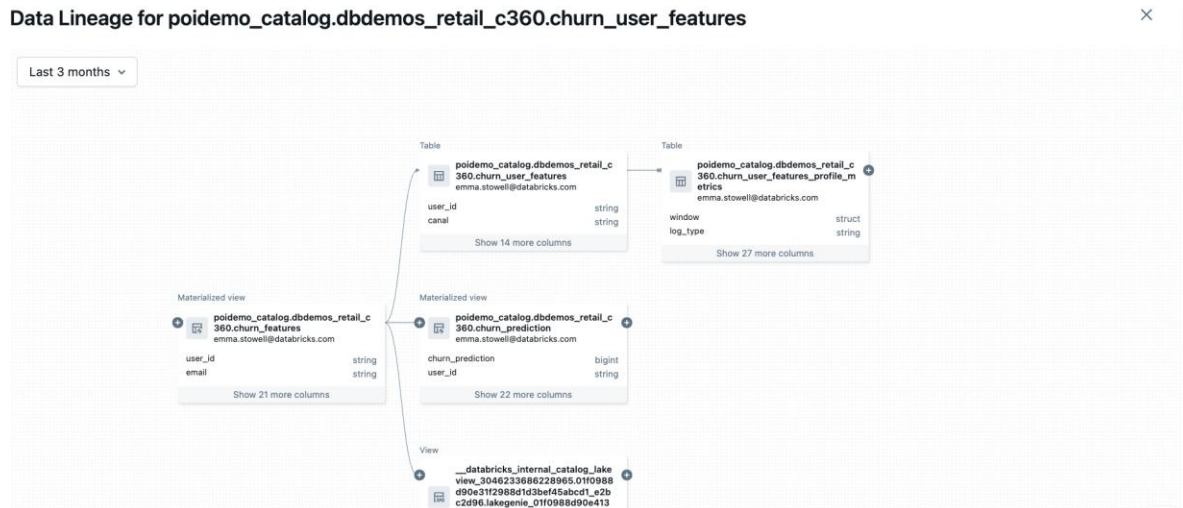


Fig 4 Automated Table Lineage

Additionally, Unity Catalog provides automated data lineage tracking at the column level, facilitating streamlined debugging, root cause analysis, and impact assessment processes. Users can gain comprehensive insights into data and AI operations through operational intelligence features, leveraging built-in system tables for billing, auditing, lineage tracking, and other functionalities.

Integration with existing data catalogues, storage systems, and governance solutions is seamless, allowing organisations to leverage their current investments and establish a robust governance framework without incurring significant migration costs.

SQL Warehouse: A versatile module designed to streamline data management across various sources. Fivetran integration simplifies the incorporation of data from cloud storage or enterprise applications such as Salesforce, Google Analytics, or Marketo, ensuring a seamless process. Dependency management and data transformation are effortlessly handled within the lakehouse environment through built-in ETL capabilities. With low code offerings such as Lakeflow connect. Alternatively, users can leverage familiar tools like dbt on Databricks SQL for enhanced performance.

With built in capabilities such as AI/BI dashboards and Genie rooms, users can create dashboards with no additional licencing costs. If an organisation has existing tools for BI, collaboration with popular BI tools like Power BI and Tableau is seamless, providing analysts with the flexibility to explore insights using their preferred interfaces. Databricks SQL enhances collaboration by facilitating joint querying, discovery, and sharing of insights through its integrated SQL editor, visualisations, and dashboards.

With Databricks SQL Serverless, infrastructure management becomes a breeze, allowing data teams to focus on core responsibilities without the hassle of configuring or scaling cloud resources. SQL warehouses within Databricks adapt automatically to support varying levels of concurrency, ensuring seamless operation for high-traffic scenarios.

Powered by the Photon vectorized query engine, Databricks SQL delivers exceptional performance across a wide array of tools, query types, and real-world applications.

Leveraging AI-driven predictive I/O, the platform intelligently prefetches data using neural networks, eliminating the need for manual performance tuning.

Centralise data in Delta Lake to maintain accessibility and avoid vendor lock-in, enabling in-place analytics and ETL/ELT operations within the lakehouse environment. Simplify data governance, security, and management with fine-grained controls and standard SQL capabilities across clouds through Databricks Unity Catalog.

When designing solutions around large language models, it's essential to consider four architectural patterns: prompt engineering, retrieval augmented generation (RAG), fine-tuning, and pre-training. Databricks distinguishes itself as the sole provider offering all four generative AI architectural patterns, ensuring unparalleled flexibility to meet evolving business needs.

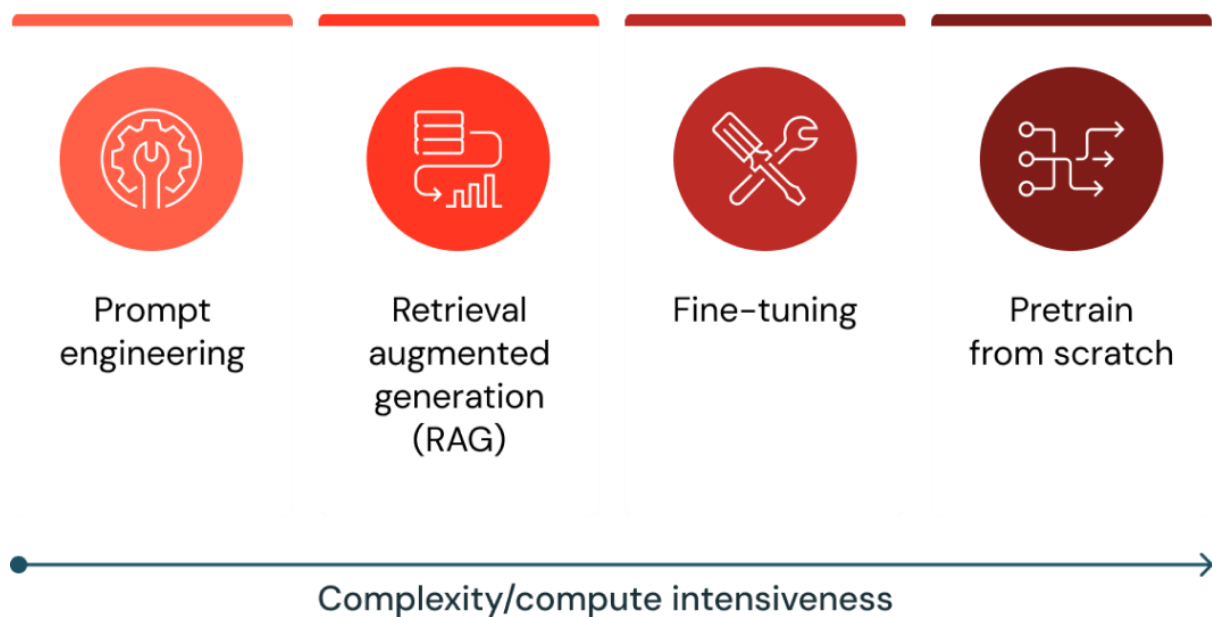


Fig 5 - AI on Databricks

Mosaic AI plays a pivotal role within the Databricks Data Intelligence Platform, seamlessly integrating data, model training, and production environments. This integration enables safe utilisation of enterprise data to augment, fine-tune, or develop custom machine learning and generative AI models. By ensuring semantic alignment with business objectives, organisations can maintain data privacy and protect intellectual property within organisational boundaries.

With Mosaic AI, organisations can develop custom large language models tailored to their specific domains, leveraging proprietary data and intellectual property to create unique models. Databricks Mosaic AI Training offers an optimised training solution capable of building new multibillion-parameter LLMs in a matter of days, with significantly reduced training costs compared to traditional methods.

Model Serving provides a comprehensive service for deploying, governing, and querying AI models, supporting various types including custom ML models like PyFunc and scikit-learn, as well as foundation models (FMs) hosted on Databricks or elsewhere. This

unified approach simplifies the experimentation and productionization of models, streamlining model management processes.

Lakehouse Monitoring offers a centralised monitoring solution within the Databricks Data Intelligence Platform, facilitating oversight of the statistical properties and quality of all tables with minimal effort. For generative AI applications, it includes capabilities to scan outputs for potentially harmful content and efficiently diagnose errors.

Databricks Unity Catalog enables the enforcement of proper permissions, rate limits, and lineage tracking to meet stringent security and governance standards. With this tool, organisations can seamlessly govern all ML assets, ensuring consistent oversight and control throughout the ML lifecycle, from development to deployment and maintenance.

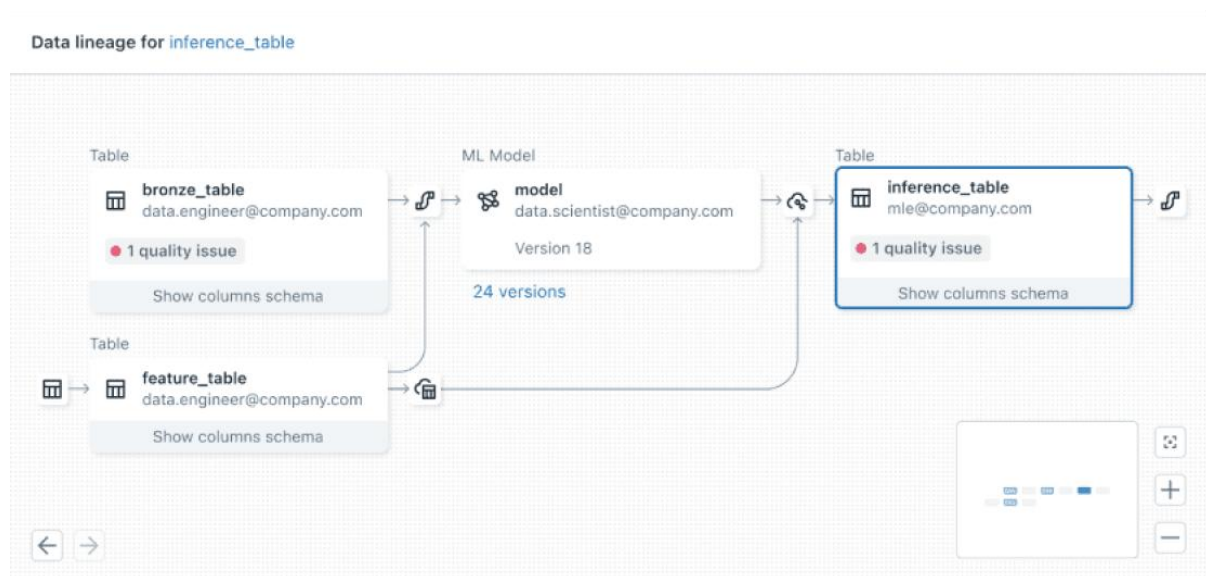


Fig 6 - Model Inference on Databricks

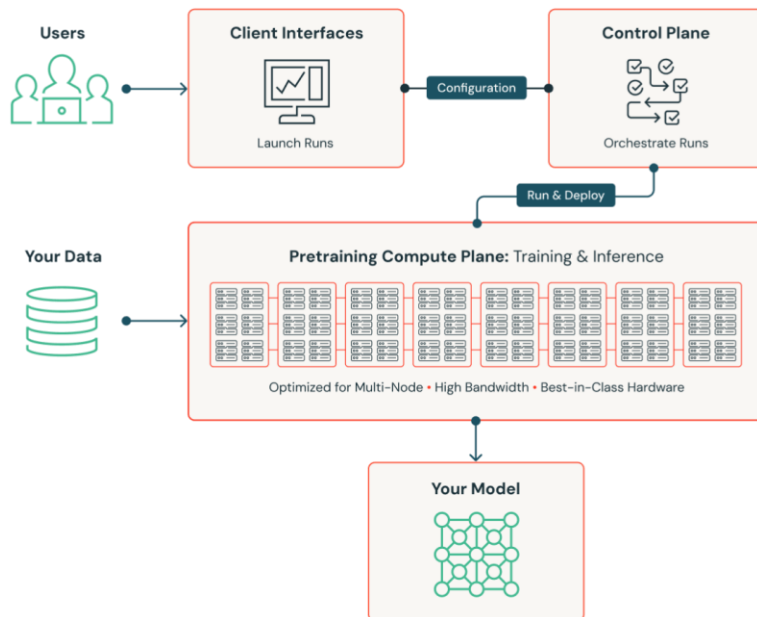


Fig 7 - Model Development in Databricks

With Databricks Notebooks, users can effortlessly write code in Python, R, Scala, and SQL to explore data and uncover insights. Collaborate securely with co-authoring, commenting, automatic versioning, Git integrations, and role-based access controls.

Users can liberate themselves from the constraints of local environments and computing resources by seamlessly migrating workflows to the cloud. Connect notebooks to personal compute resources or auto-managed clusters for enhanced scalability and performance.

Customise the development environment to suit individual preferences by connecting preferred IDEs to Databricks. Access limitless data storage and compute resources, or utilise built-in support for RStudio or JupyterLab within Databricks.

Centralise data management with Delta Lake, empowering users to clean and catalogue all data types—batch, streaming, structured, or unstructured—in one place. Automatic quality checks validate incoming data, and data versioning ensures compliance and evolution tracking.

Transform analyses into dynamic dashboards effortlessly, allowing users to share and export results seamlessly. Dashboards remain up-to-date and support interactive queries, while role-based access controls ensure secure sharing.

Users can unleash the potential of real-time analytics, machine learning, and application development across the organisation with unified batch and streaming APIs in SQL and Python. Leverage familiar languages and tools to drive innovation and derive valuable insights.

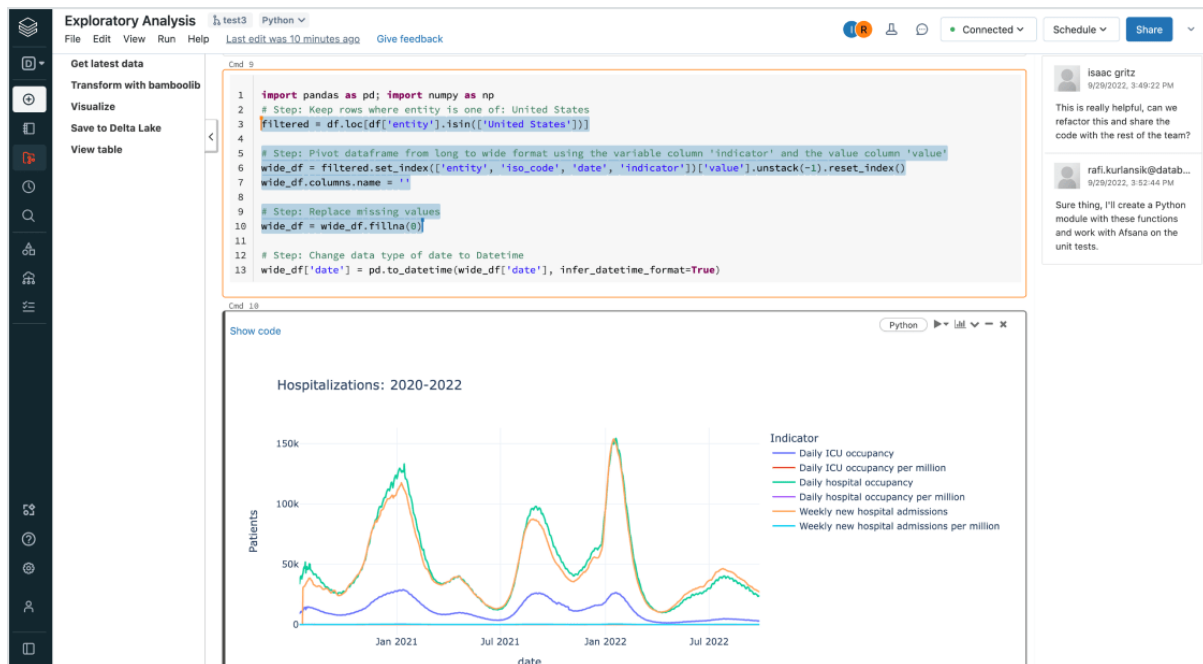


Fig 8 - Notebooks for Exploratory Data Analysis in Databricks

Deploying and managing real-time pipelines and applications in production is facilitated by Databricks through automated tooling. This automation streamlines tasks such as orchestration, fault tolerance and recovery, checkpointing, performance optimization, and autoscaling, enhancing operational efficiency for users.

Unity Catalog serves as a unified governance model for all streaming and batch data within Databricks. This centralization simplifies data discovery, access, and sharing processes across different data types, providing users with a cohesive experience.

Simplify data ingestion and ETL processes for streaming data pipelines with **Lakeflow**. This feature enables teams to utilise a simple declarative approach to data engineering using familiar languages like SQL and Python. Users can manage batch and streaming data pipelines in a single location, with options for controlled and automated refresh settings, reducing complexity and saving time.

Real-time insights from streaming data pipelines immediately enhance the accuracy and utility of analytics and AI applications. Businesses can benefit from faster decision-making and more precise predictions, gaining a competitive advantage in their respective industries.

Databricks automates various operational tasks required for building and deploying streaming data pipelines. These tasks include infrastructure scaling, pipeline dependency orchestration, error handling, recovery, and performance optimization. Enhanced Autoscaling ensures efficient cluster resource allocation, optimising compute resources for each workload.

Spark Structured Streaming, the core technology driving data streaming on the Databricks platform, offers a unified API for both batch and stream processing. With a proven uptime track record, Databricks provides a reliable environment for running

Apache Spark workloads. Photon, the next-generation engine compatible with Apache Spark APIs, delivers impressive performance-per-cost, automatically scaling to accommodate workload demands.

Data streaming on Databricks leverages foundational components such as Unity Catalog and Delta Lake. Delta Lake optimises raw data storage for both streaming and batch data, while Unity Catalog facilitates integrated governance for all data and AI assets. This includes seamless data discovery, access, and sharing across multiple clouds. Additionally, Unity Catalog supports Delta Sharing, an open protocol for secure data sharing with external organisations.

Access data sets, AI, and analytics assets—such as ML models, notebooks, applications, and dashboards—without being tied to proprietary platforms, intricate ETL processes, or costly data replication. This open approach allows for swift utilisation of data across various cloud platforms with the tools of your preference, facilitating efficient data-driven workflows.

Advance your organisation's AI, ML, and analytics initiatives by tapping into a diverse range of assets beyond conventional data sets. These include readily available ML models, notebooks, applications, and customised solutions tailored to meet specific requirements.

Utilise pre-built notebooks and sample data sets to effectively assess and gain confidence in the suitability of data products for your AI, ML, or analytics projects. This streamlined evaluation process helps expedite decision-making and reduces the time needed for implementation.

Significantly reduce the time required to derive insights and avoid being locked into a particular vendor by leveraging open and seamless sharing and collaboration capabilities across different clouds, regions, or platforms. Integrate directly with preferred tools and workflows, empowering users to work efficiently within their familiar environments.

The data intelligence platform is open and interoperable with many popular data providers available commercially, or those that are willing to provide in an open protocol parquet format.

Openness is a core design principle of the data intelligence platform, with data residing in commodity cloud storage owned by the customer. It stores data as delta tables, or iceberg tables – two popular open source, highly performant data formats. This approach to data removes the risk of vendor lock in and reduces operational inefficiencies by removing barriers to collaboration.

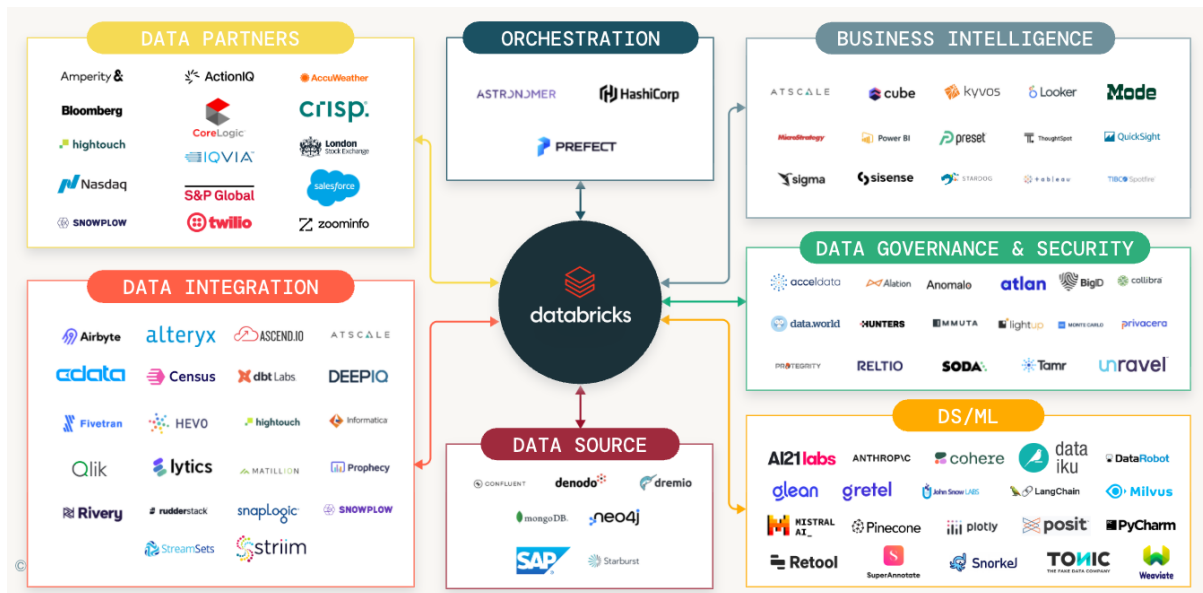


Fig 9 - Databricks Partner Ecosystem Fig 9

How does the Data Platform Services work?

The Platform Services architecture enables you to retain significant control of your data and cloud environment. The architecture consists of computing resources controlled and operated by you on your chosen cloud provider, and computing resources that are controlled and operated by Databricks. Your data is processed with your computing resources unless you choose the serverless architecture, in which case your data is processed by the computing resources within Databricks' control and operation.

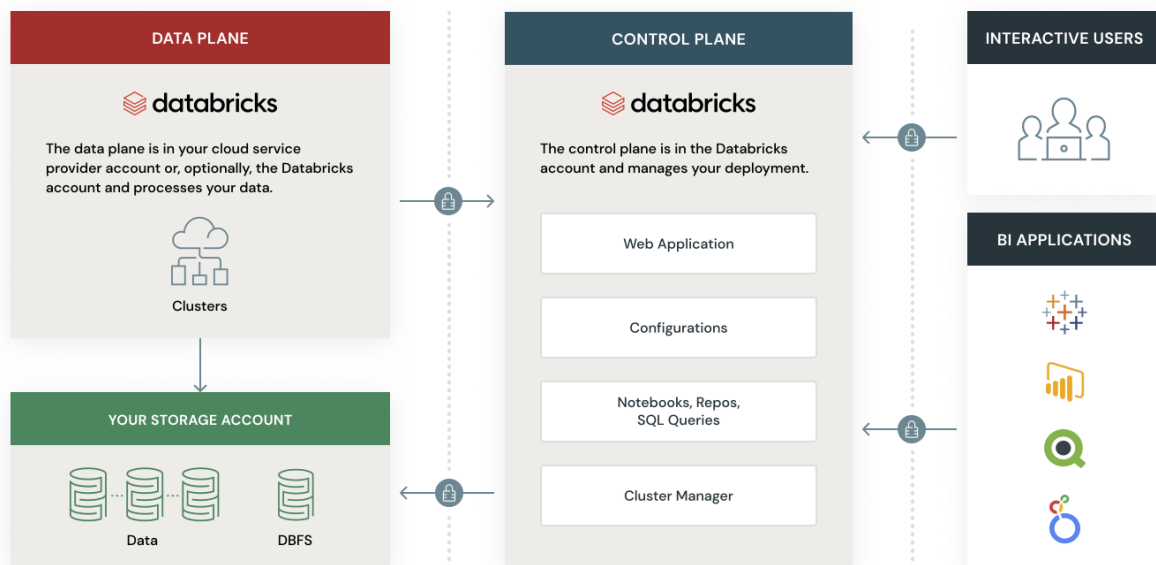


Fig 10 - Databricks Platform Services Architecture

What data is in my cloud environment? What about the Databricks' cloud environment?



All of your data sets (e.g., files and tables) are stored in your chosen cloud provider storage. For example, when you are running jobs, jobs data from the UI will be stored in Databricks File System (DBFS) in your cloud storage. You can also connect your streaming sources, such as Apache Kafka or Amazon Kinesis, as well as other data sources hosted on the internet.

The Databricks environment will contain notebooks, configuration information, user information, logs and Databricks account information. The Databricks environment by default will also include interactive query results, which you can configure to instead be stored in a storage location controlled by you:

- For AWS see this [Documentation](#)
- For Google Cloud see this [Documentation](#)

When you're using Databricks resources to process your data, the resources will send basic telemetry and usage data to Databricks. As a platform operator, we use this data to operate, improve, support and develop our services. This data is encrypted in transit and at rest, and data in transit between your resources and Databricks resources is also encrypted. There are additional controls available detailed in the [Security & Trust Overview Whitepaper](#).

Associated Services

Databricks Professional Services (refer to our rate card/pricing for detailed pricing)

Delivering rapid success on projects with world-class data engineering, data science and project management expertise. Databricks Professional Services can help you at any point in your data and AI journey:

Key benefits of Databricks Professional Services:

- Accelerate your data and AI journey: Our offerings and expert services help you in your data and AI journey to provide optimal and tailored acceleration from the initial workspace onboarding to building out DataOps and Center of Excellence practices at enterprise scale.
- De-risk your project: Whether you are migrating from legacy workloads to Databricks, building out new data products or data and AI pipeline and machine learning projects, let us be your partner and trusted advisor to minimise risk and maximise value at each step of your journey.
- Operationalise at scale: Building a data pipeline POC or single node model development is relatively manageable. The challenge lies in how you successfully adopt and scale data and AI practices across the company. Let us help you meet this challenging goal with our prescriptive offerings and expertise.

Databricks' prepackaged service offerings includes:

- Architecture and Best Practices Assessment
- Architectural Design Review
- Enterprise-Readiness Review



- DevOps best practices
- MLOps best practices
- Delta Lake – Use Case Acceleration
- Streaming Use Case acceleration
- Data Estate Optimisation Assessment
- Migration Assurance
- Unity Catalog Reference Implementation
- SQL MVP
- Databricks Jumpstart
- Hadoop Migration
- EMR Migration
- Multicloud migration
- Databricks Lakehouse Buildout
- Streaming Analytics
- GenAI MVP
- Machine Learning – Model MVP
- Machine Learning – MLOps optimised

The overall Databricks Professional Services offerings can be broadly categorised into four areas – Services, Coaching, Training and Support. The customer usually buys a pot of success credits which they can then spend across the various areas.

They can leverage the success credits to obtain assurance or delivery services, by leveraging the professional services team of architects or consultants, who work on helping with areas like hadoop migration, Lakehouse build out, streaming analytics or architectural reviews and best practices etc. The customer can also leverage the services of a Delivery Solution Architect to help with onboarding and to provide enablement and hand holding to the teams. Training, in person or virtual along with certification is also obtained by using the success credits. Tiered business support is also provided but is bought separately. The below diagram provides a holistic approach of the various offerings that Databricks provides its customers.

The Databricks team continues to support the customer on their journey in leveraging the technology to accelerate the value in delivery.

This usually comprises four parts.

1. Support smooth onboarding onto the platform via–
 - a. Establishing goals of partnership
 - b. Create Success plans for use cases
 - c. Provide continuous industry leading best practice guidance and suggestions for improvement
 - d. Provide new features and issue reviews
2. Provide long term Training and Enablement –
 - a. Facilitate Rapid start sessions for new users

- b. Provide customised learning path for different personas like Business users, Platform Admins, Data Analysts, Data Scientists and Data Engineers
 - c. Continuous updates on new features and relevant content
 - d. Product Roadmap reviews
- 3. Business Reviews –
 - a. Track progress against Mutual Success Plan
 - b. Track derived value from Databricks
 - c. Update on roadmaps
 - d. Understand and support needs of new use cases and users
- 4. Cross Functional –
 - a. Single point of contact to provide the right support
 - b. Provide industry expertise collaboration
 - c. Provide Go live / production tracking
 - d. Update and access to early feature previews

Specialised resource capability can be provided through Delivery Solution Architects and Resident Solution Architect, where service subscriptions are the most value for money way in which to purchase this highly specialised capability.

2. Data Protection

Information Assurance

Our trusted platform is built by embedding security throughout the software development and delivery lifecycle. We follow rigorous operational security practices such as penetration testing, vulnerability assessments and strong internal access controls. We believe transparency is the key to winning trust — we publicly share how we operate, and work closely with our customers and partners to address their security needs. We have offerings for Cyber Essentials Plus, PCI-DSS, HIPAA, DOD IL5 and FedRAMP compliance, and we are ISO 27001, ISO 27017, ISO 27018 and SOC 1, 2 & 3 Type II compliant.

Beyond the documentation and best practices that you will find in our Security and Trust Center, we also provide a contractual commitment to security written in plain language to all our customers. This commitment is captured in the Security Addendum of our customer agreement, which describes the security measures and practices that we follow to keep your data safe.

The Security Addendum is incorporated into and made a part of the written agreement between Databricks, Inc. (“Databricks”) and Customer that references this Security Addendum (“Agreement”).

Databricks maintains a comprehensive documented security program that is based on industry standard security frameworks including NIST 800-53 and ISO 27001 (the “Security Program”). Pursuant to the Security Program, Databricks implements and maintains administrative, physical, and technical security measures to protect the Platform Services and Support Services and the security and confidentiality of Customer Content (including any Personal Data that may be contained therein) (each as defined in the Agreement) under Databricks’ control that is processed by Databricks in its provisioning of the Platform Services or Support Services (the “Security Measures”). Databricks’ compliance with this Addendum shall be deemed to satisfy any more general measures included within any Agreement, including the Service Specific Terms.

Databricks regularly tests and evaluates its Security Program, and may review and update this Security Addendum at any time without notice, provided that such updates are equivalent (or enhance) security and do not materially diminish the level of protection afforded to Customer Content by these Security Measures.

Architecture. Databricks is a hybrid platform-as-a-service offering. The components responsible for managing and controlling the Platform Services are referred to as the ‘Databricks Control Plane’. The compute resources that perform data processing operations are referred to as the “Data Plane”. For certain Cloud Service Providers, the Data Plane may either be deployed in Customer’s Cloud Service Provider account (known as the ‘Customer Data Plane’) or, for Databricks Serverless Compute, in a Databricks-controlled Cloud Service Provider account (known as the ‘Databricks Data

Plane’). Data Plane shall refer to both Customer Data Plane and Databricks Data Plane unless otherwise specified.

Shared Responsibility. Databricks operates in a shared responsibility model, where both Databricks and Customer maintain security responsibilities.

Most Customer Content is stored within Customer’s own Cloud Service Provider account at rest (e.g., within Customer’s AWS S3 bucket) or within other Systems under Customer’s control. Please see the Documentation for more details.

Databricks Control. Small amounts of Customer Content may be stored within the Databricks Control Plane, including metadata about Customer Content. Databricks offers Customers options regarding the storage of certain Customer Content within the Platform Services such as query results. Please see the Documentation for more details.

Customers may specify the region(s) where their Platform Services Workspaces are deployed. Customers can choose to deploy the Data Plane into any supported Databricks region. The Databricks Control Plane is deployed into the same region. Databricks will not, without Customer permission, move a Customer Workspace into a different region.

Databricks’ Audits & Certifications. Databricks uses independent third-party auditors to assess the Databricks Security Program at least annually, as described in the following audits, regulatory standards, and certifications:

Cyber Essentials Plus

HITRUST

SOC 1 Type II

SOC 2 Type II

SOC 3 Type II

ISO 27001

ISO 27017

ISO 27018

ISO 27036

HIPAA

PCI DSS

FedRAMP High

DoD IL5

G-Cloud 15



Databricks' Chief Security Officer leads the Databricks' Information Security Program and develops, reviews, and approves (together with other relevant internal stakeholders) Databricks' Security Policies and Procedures.

Databricks maintains a documented change management policy, reviewed at least annually.

Databricks has implemented a formal Information Security Management System ("ISMS") in order to protect the confidentiality, integrity, authenticity, and availability of Databricks' data and information systems, and to ensure the effectiveness of security controls over data and information systems that support operations.

Databricks employs monitoring and logging technology to help detect and prevent unauthorised access attempts to its network and equipment.

Active users with privileged access to the Platform Services are reviewed at least quarterly and are promptly removed upon termination of employment. As part of the personnel offboarding process, all accesses are revoked and data assets are securely wiped.

Databricks maintains a comprehensive Third Party Risk Management program that assesses the security compliance of applicable third parties, including vendors and subprocessors, in order to appropriately measure and manage risk.

Personnel receive comprehensive training on the Security Policies upon hire and refresher training annually. Personnel are required to certify and agree to the Security Policies and personnel who violate the Security Policies are subject to disciplinary action, including warnings, suspension and up to (and including) termination.

All personnel undergo background checks prior to onboarding (as permitted by local law), which may include, but are not limited to, criminal record checks, employment history verification, education verification, and global sanctions and enforcement checks. Databricks uses a third-party provider to conduct screenings, which vary by jurisdiction and comply with applicable local law. Personnel are required to sign confidentiality agreements.

Physical and Environmental Controls

Databricks Corporate Offices. Databricks has implemented administrative, physical, and technical safeguards for its corporate offices. These include, but are not limited to, the below:

Visitors are required to sign in, acknowledge and accept an NDA, wear an identification badge, and be escorted by Databricks personnel while on premises

Databricks personnel badge into the offices

Badges are not shared or loaned to others without authorization

Physical entry points to office premises are recorded by CCTV and have an access card verification system at every door, allowing only authorised employees to enter the office premises

Equipment and other Databricks-issued assets are inventoried and tracked

Office Wi-Fi networks are protected with encryption, wireless rogue detection, and Network Access Control

Cloud Service Provider Data Centers. Databricks regularly reviews Cloud Service Provider audits conducted with compliance.

Security controls include, but are not limited to the list below:

Biometric facility access controls

Visitor facility access policies and procedures

24-hour armed physical security

CCTV at ingress and egress

Intrusion detection

Business continuity and disaster recovery plans

Smoke detection sensors and fire suppression equipment

Mechanisms to control temperature, humidity and water leaks

Power redundancy with backup power supply

Databricks leverages multiple layers of network security controls, including network-level isolation, for separation between the Databricks' development and production environments.

Firewalls are implemented as network access control lists or security groups within Databricks' production environment.

Databricks employs industry standards to harden images and operating systems under its control that are deployed within the Platform Services, including deploying baseline images with hardened security configuration such as disabled remote root login, isolation of user code, and images are regularly updated and refreshed.

For Systems under Databricks control supporting the production data processing environment, Databricks tracks security configurations against industry standard baselines such as CIS and STIG.

Encryption of data-in-transit. Customer Content is encrypted using cryptographically secure protocols (TLS v.1.2 or higher) in transit between (1) Customer and the Databricks Control Plane and (2) the Databricks Control Plane and the Data Plane. Additionally, depending on functionality provided by the Cloud Service Provider, Customers may

optionally encrypt communications between clusters within the Data Plane (e.g., by utilising appropriate AWS Nitro instances).

Customer Content is encrypted using cryptographically secure protocols (AES-256 bit, or the equivalent or better) while at rest within the Databricks Control Plane.

Cryptographic standards are periodically reviewed and selected technologies and ciphers are updated in accordance with assessed risk and market acceptance of new standards.

Customer(s) may choose to leverage additional encryption options for data in transit within the Customer Data Plane or Databricks Data Plane as described in the Documentation (e.g., Customer may utilise AWS Nitro EC2 instances within the Customer Data Plane to provide additional encryption in transit). Customer shall, based on the sensitivity of the Customer Content, configure the Platform Services and Customer Systems to encrypt Customer Content where appropriate (e.g., by enabling encryption at rest for data stored within AWS S3).

Databricks leverages security capabilities provided natively by Cloud Service Providers for security detection.

Databricks generates audit logs from Customer's use of the Platform Services. The logs are designed to store information about material events within the Platform Services.

Customers may, depending on the entitlement tier of the Platform Services, enable delivery of audit logs. It is Customer's responsibility to configure this option.

Databricks stores audit logs in a manner designed to protect the audit logs from tampering.

Databricks stores audit logs for at least one year.

Databricks conducts third-party penetration tests at least annually, in addition to maintaining in-house offensive security personnel and a public bug bounty program.

Databricks regularly runs authenticated scans against representative hosts in the SDLC pipeline to identify vulnerabilities and emerging security threats that may impact the Platform Services. Databricks will use commercially reasonable efforts to address critical vulnerabilities within 14 days, high severity within 30 days, and medium severity within 60 days measured from, with respect to publicly declared third party vulnerabilities, the date of availability of a compatible, vendor-supplied patch, or for internal vulnerabilities, from the date such vulnerability is confirmed. Databricks leverages the National Vulnerability Database's Common Vulnerability Scoring System (CVSS), or where applicable, the U.S.-Cert rating, combined with an internal analysis of contextual risk to determine criticality.

Databricks deploys new code to the Databricks Control Plane on an ongoing basis.

New Data Plane virtual machines use the latest applicable source code and system images upon launch. Customers are encouraged to restart always-on clusters on a periodic basis to take advantage of security patches.

Databricks personnel are authenticated through single sign-on (SSO), 802.1x (or similar) where applicable, and use a unique user ID and password combination and multi-factor authentication. Privileges are consistent with least privilege principles. Security Policies prohibits personnel from sharing or reusing credentials, passwords, IDs, or other authentication information. If your identity provider supports the SAML 2.0 protocol, you can use Databricks' SSO to integrate with your identity provider.

Databricks enforces RBACs (based on security groups and access control lists). Only authorised roles, which are defined based on the principle of least privilege and segregation of duties, are allowed to access production systems.

Databricks enforces certain security controls on its workstations used by personnel, including:

Full-disk encryption

Anti-malware software

Automatic screen lock after 15 minutes of inactivity

Secure VPN

Breach Detection & Response

Databricks' dedicated Detection engineering team deploys and develops intrusion detection monitoring across its computing resources, with alert notifications sent to the Security Incident Response Team (SIRT) for triage and response. The SIRT employs an incident response framework to manage and minimise the effects of unplanned security events.

"Security Breach" means a breach of security leading to any accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to Customer Content under Databricks control. A Security Breach shall not include an unsuccessful attempt or activity that does not compromise the security of Customer Content, including (without limitation) pings and other broadcast attacks of firewalls or edge servers, port scans, unsuccessful log-on attempts, denial of service attacks, packet sniffing (or other unauthorised access to traffic data that does not result in access beyond headers) or similar incidents. Databricks maintains a record of Security Breaches that includes description, dates and times of relevant activities, and disposition. Suspected and confirmed security incidents are investigated by security, operations, or support personnel; and appropriate resolution steps are identified and documented. For any confirmed Security Breaches, Databricks will take appropriate, reasonable steps to minimise product and Customer damage or unauthorised disclosure.

In accordance with applicable data protection laws, Databricks will notify Customer of a Security Breach for which that Customer is impacted without undue delay after

G-Cloud 15



becoming aware of the Security Breach, and take appropriate measures to address the Security Breach, including measures to mitigate any adverse effects resulting from the Security Breach.

Upon written request and at no additional cost to Customer, Databricks shall provide Customer, and/or its appropriately qualified third-party representative, access to reasonably requested documentation evidencing Databricks' compliance with its obligations under Security Addendum in the form of the relevant audits or certifications listed above. Such audits are performed (a) at least once annually; and (b) by independent third-party security professionals selected by Databricks. Such audits result in the generation of a confidential audit report collectively, "Audit Reports").

Only to the extent Customer cannot reasonably satisfy Databricks compliance with Security Addendum through the Audit Reports, Customer may send a written request to conduct an audit of Databricks applicable controls during the term of the Agreement on an annual basis. Following receipt by Databricks of such a request, Databricks and Customer shall mutually agree in advance on the details of the audit, including the reasonable start date, scope and duration of, and security and confidentiality controls applicable to, any such audit. The Audit Report, audit, and any information arising therefrom shall be considered Databricks Confidential Information and may only be shared with a third party (including a third party controller) with Databricks' prior written agreement.

Notwithstanding any other audit provisions in the Agreement, Customer requests for audits are limited to once per year.

Databricks Business Continuity (BC) and Disaster Recovery (DR) plans are reviewed and drills are conducted annually.

Data Resiliency. Databricks performs backups for the Databricks Control Plane, generally managed by the Cloud Service Provider capabilities, for data resiliency purposes in the case of a critical systems failure. While Databricks backs up certain service elements that persist in the Databricks Control Plane as part of its systems resiliency, those backups are maintained only for emergency recovery purposes and are not available for Customer.

No Data Restoration. Due to the hybrid nature of the Databricks Platform, Databricks does not back up Customer Content. To assist Customers in performing their own backups, Databricks provides certain self-service features within the Platform Services (like the ability to synchronise notebooks with a customer's Github or Bitbucket account).

The Platform Services provide Customers with functionality that permit Customer to delete Customer Content under Databricks' control.

Customer Content contained within a Customer Workspace is permanently deleted within thirty (30) days following cancellation of the Workspace.

MosaicML Security Standards. Notwithstanding anything to the contrary in the Agreement including its schedules or annexes, the security standards and measures set forth in the Agreement (“Security Measures”) do not apply to products and services related to MosaicML (“MosaicML Services”). Instead, the following section comprises the Security Measures for the purpose of MosaicML Services. For clarity, the Security Measures described below do not apply to any other Databricks Services.

Databricks implements and maintains commercially reasonable technical, physical, administrative, and organisational safeguards, and other applicable aspects of system and content management, designed to protect the security, confidentiality, integrity, and availability of all Customer Content. Provider shall use commercially reasonable efforts to promptly notify Customer in the event of any breach of security impacting Customer Content (“Security Breach”) within three (3) business days from the date the Security Breach is confirmed.

Privacy by Design

Databricks takes privacy seriously. We understand that the data you analyse using Databricks is important both to your organisation and your customers, and may be subject to a variety of privacy laws and regulations. We incorporate privacy by design into our product development methodology, with input from our legal privacy team throughout the development lifecycle.

Databricks itself is compliant with key privacy regulations such as the GDPR, CCPA, Swiss Data Protection Act, and UK Data Protection Laws, which are included in our default DPA language (<https://databricks.com/dpa>). Additionally, due to the way the Databricks platform is architected, customers retain control of most of their Customer Data within their own cloud service provider storage buckets and generally store little Customer Personal Data within the Databricks services themselves.

3. Using the service

Ordering and Invoicing

When a customer considers deploying Databricks would recommend getting in contact via gcloud@databricks.com to discuss requirements and how best to provide a solution. Databricks can then align account management, technical solutions support and a delivery partner as appropriate to requirements.

Customers can enable trial services as described in the section below (Trial Service). As Databricks is a first party service in Microsoft's Azure Platform, if a customer has access to Microsoft Azure's platform services they are able to deploy Azure Databricks on a pay as you go, self service way much like any other Microsoft Azure cloud services, leveraging their direct Microsoft Agreements for ordering/invoicing.

If a customer is using significant volumes of Azure Databricks and wants to get additional benefits to support their deployment when making 1 or 3 year commitments (through Customer Success credits to support Training, Enablement and Delivery Partners) then please contact via gcloud@databricks.com.

Customers who wish to deploy Databricks into their Amazon Web Services (AWS) / Google Cloud Platform (GCP) account can do so by contacting us directly (at gcloud@databricks.com).

During trial usage of Databricks Community Edition, AWS/Azure/GCP will bill you directly for the EC2/VM/GCE instances created in Databricks. At the end of the trial, customers will need to contact Databricks to enable services going forward, which will require a Call-Off Contract.

By default monthly billing based on per-second usage (expressed in Databricks Units or DBUs) is how our charges are applied. Customers can look at alternative billing options, such as billing upfront via annual plan (which has the potential for further savings much like reservations for cloud services).

A Databricks Unit ("DBU") is a unit of processing capability per hour, billed on per-second usage. Databricks supports many AWS EC2 / Azure VMs / GCP GCE instance types. The larger the instance is, the more DBUs you will be consuming on an hourly basis. For example, 1 DBU is the equivalent of Databricks running on a c4.2xlarge machine for an hour.

Databricks can provide assistance to the customer in completing the Call-Off contract once the buyer makes a request for our software services, using terms that are consistent with existing agreements across other government organisations.

Pricing Overview

Databricks operates under a consumption model, where the amount the service costs depends wholly on the amount the service is used. The service can be used under a Pay As You Go model, where each month's usage is billed in arrears, or in direct commitment model where an amount of usage time is purchased up front at a discounted rate (as specified in Pricing Document), with additional benefits provided through Success Credits (which can be used towards Professional Services, Training and Enablement and as specified in Pricing Document).

Databricks on AWS is available in Premium & Enterprise editions. See below a listed feature comparison.

		Premium	Enterprise
Included features			
Databricks Workspace		Workspace for production jobs, analytics, and ML	Workspace for production jobs, analytics, and ML
Managed Apache Spark™		✓	✓
Optimized Delta Lake		✓	✓
Cluster Autopilot		✓	✓
Connectors and Integration		✓	✓
Databricks Runtime for ML		✓	✓
Managed MLflow		✓	✓
Performance		Optimized Runtime Engine	
Governance and Manageability		Audit logs & automated policy controls	Audit logs & automated policy controls
Administration Console		✓	✓
Unity Catalog (Cross-Workspace Data Governance)		✓	✓
Unity Catalog (Automated Data Lineage)		✓	✓
Managed Delta Sharing		✓	✓
Audit Logs		✓	✓
Cluster Policies		✓	✓
Enterprise Security		Extend your cloud-native security for company-wide adoption	Advanced compliance and security for mission critical data
Single Sign-On (SSO)		✓	✓
Secure Cluster Connectivity		✓	✓
Role-based Access Control		✓	✓
Federated IAM		✓	✓
Customer Managed VPC		✓	✓
Token Management API		✓	✓
Customer Managed Keys		✓	✓
IP Access List			✓
PrivateLink			✓
Enhanced Security And Compliance			Available as a platform add-on. See below.
Enhanced Security and Compliance		Premium	Enterprise
Provides enhanced security and controls for your compliance needs			15% of Product Spend before discounts
Enhanced Security Monitoring			✓
HIPAA Compliance Controls			✓
PCI-DSS Compliance Controls			✓
FedRAMP-Moderate Compliance Controls			✓

Databricks on GCP is available in Premium edition only. See below a listed feature comparison.

Included features	Premium
Databricks Workspace	Workspace for production jobs, analytics, and ML
Managed Apache Spark™	✓
Optimised Delta Lake	✓
Cluster Autopilot	✓
Connectors and Integration	✓
Databricks Runtime for ML	✓
Managed MLflow	✓
Performance	Optimised Runtime Engine
Governance and Manageability	Audit logs & automated policy controls
Administration Console	✓
Audit Logs	✓
Cluster Policies	✓
Unity Catalog (Cross-Workspace Data Governance)	✓
Managed Delta Sharing	✓
Enterprise Security	Extend your cloud-native security for company-wide adoption
Single Sign-On (SSO)	✓
Secure Cluster Connectivity	✓
Role-Based Access Control	✓
Token Management API	✓
IP Access List	✓
HIPAA Compliance	Available as a platform add-on. See below.
	Premium
GCP HIPAA Compliance	
Provides enhanced security and controls for your HIPAA compliance needs	10% of Product Spend before discounts
HIPAA Compliance Controls	✓

Trial Service or Free Edition

2 week free trials are available, giving access to the full product suite for evaluation. We would recommend customer to sign up to the trial service using the following link:

<https://databricks.com/try-databricks> .

For non-commercial / educational purposes, Databricks also offers a free edition which provides a compute-restricted sandbox for users to familiarise themselves with the features of the platform. This can be found at this link:

<https://www.databricks.com/learn/free-edition>

On-Boarding, Off-Boarding, Service Migration, Scope etc.

Getting started with Databricks is straightforward, with slight differences depending on your cloud provider. The simplest way to get started is through the free trial, after which it can convert into a Pay as you Go account type (subject to Call-Off Contract Agreement).

For off-boarding, the end user is required to cancel databricks subscription (for Google and AWS) and to delete databricks account, in that order. After you cancel your subscription:

- You can no longer access workspaces, notebooks, or data in your Databricks account.
- In accordance with the Databricks terms of service, any customer content contained within workspaces tied to your subscription will be deleted within 30 days of cancellation.
- You can't sign up for a new subscription by using the same email address. You must provide a new email address in the sign-up form.

To cancel your subscription you will need to cancel your subscription in the marketplace, either in Google or AWS. Please email gcloud@databricks.com and a member of Databricks account team will help you execute the steps in the cloud platform of your choice.

Databricks is a general purpose computing platform, and as such there is no single path to onboarding, offboarding, migrations etc – the process used is dependent on the workload being addressed. Migrating from an on-premise SQL server is different to migrating from another cloud. Databricks can assist in planning and executing these, using experience and process from other engagements.

A typical process would be the following. The Databricks Solution Architect would work with the customer to profile and understand the business requirements and architect a solution that meets the need. A delivery partner would then work to implement the solution, with reviews, training, and enablement provided by Databricks Professional Services. This outline would be the same regardless of whether it would be a new workload or migrating an existing workload. The variables would be the type and experience of the resource engaged from the Databricks and delivery partner side – for

Machine Learning workloads, an appropriately experienced architect would be used, for example.

Databricks provides an Account Manager to support our customers and build an ongoing relationship and commercial requirements. In addition, our account teams include Solution Architects who can provide technical advice and guidance so that customers can leverage the Databricks platform to deliver their business outcomes. The Databricks management team closely engages with the customer during the entire lifecycle of delivery, from the initial scoping and planning stage to the final handover.

As part of the customer journey with us, we follow the below broad approach to help us drive successful outcomes.

1. A Strategy and planning session is held between customer stakeholders and the Core Databricks team of Account Manager, Solution Architect, Delivery Solution Architect and the Engagement Manager to define the criteria for success, resulting in a high level Mutual success plan with a clearly defined scope. The stakeholders governance tracking is also set up, including monthly and quarterly reviews, to ensure the engagement is tracked and so any course correction can be applied, if required.
2. This is followed by a Business Kickoff and Prioritisation meeting between the Databricks team and the Customer Business owners, with the aim to drive maximum value and return on investment, which is then added to the Mutual Success Plan
3. This is followed by a technical kick off meeting to define the approach where the Databricks' Solution Architect and the Resident Solution Architect works with the customer's Technical owners, which results in a technical design document. At this point the timelines for delivery along with the cadence for governance is also set and meetings scheduled.
4. Databricks' Resident Solution Architect then works on studying current architecture review and definition of the desired state followed by helping setup the environments. The deliverable for this milestone is a final architecture document, reviewed and approved by the customer's Technical and Security team.
5. The Databricks' team of Resident Solution Architect and the Resident Solution Engineer or Consultants then design, build and test the agreed solution using the Agile Methodology in 2-week cycles. The Scrum Master is from the customer side and the Databricks team attends the daily stand ups to ensure things are on track and raises any concerns or issues. In addition to this weekly or monthly reports, as agreed, are provided to the stakeholders to ensure health and status of the project. The entire process is shadowed by the customer team and later documented for future reference.
6. After the initial delivery of an Minimum Viable Product, the Databricks team focuses on security, optimization of cost and performance to ensure driving

maximum value and also provide documentation of recommendation for a roadmap to further drive value of the products in the future

7. The solution then goes live into production with the entire process being documented and shared with the customer by the Databricks delivery team
8. Post this the consumption of data is tracked by the Delivery Solution Architect , measured and reported, as per agreed schedules, to ensure adherence to initially agreed parameters
9. In parallel to the delivery planning the Databricks Engagement Manager, Customer Enablement Architect and the Delivery Solution Architect provide work on accessing the learning needs of the customer team and help plan for an Enablement Framework, leveraging our courses, workshops and user communities, to ensure the customer team is trained and ready to use our products
10. Finally our Core Databricks team are also available post production release, to help address and support any issues and future product roadmap updates, as applicable

Each step in the cycle and the corresponding milestone represents key elements of the experience of becoming a customer of Databricks, ending in successful realisation of the platform's potential for business value to an organisation.

During this entire process the health and status of the engagement is tracked and reported back to the stakeholders in the form of weekly and monthly meetings and reports. We also track the overall success of the engagement and value delivered to the customer through the quarter business reviews and implement any course corrections, as required.

Training & Enablement (please refer to our Pricing Document)

As part of our engagement with the customer we aim to train the customer and enable them to upskill and become self-sufficient in using our technology. We do this via the Databricks Academy which is an online service available to all customers offering training courses on the Databricks platform for all the primary user roles likely to interface with it: Data Engineers, Data Scientists / Machine Learning Engineers, SQL Analysts and Platform Admins.

It has multiple free and paid course content available to be used for this purpose. We have dedicated program managers to handhold customers through the journey and provide support at all stages of the Talent Management program.

Talent Management

Databricks can support you at every step of the way

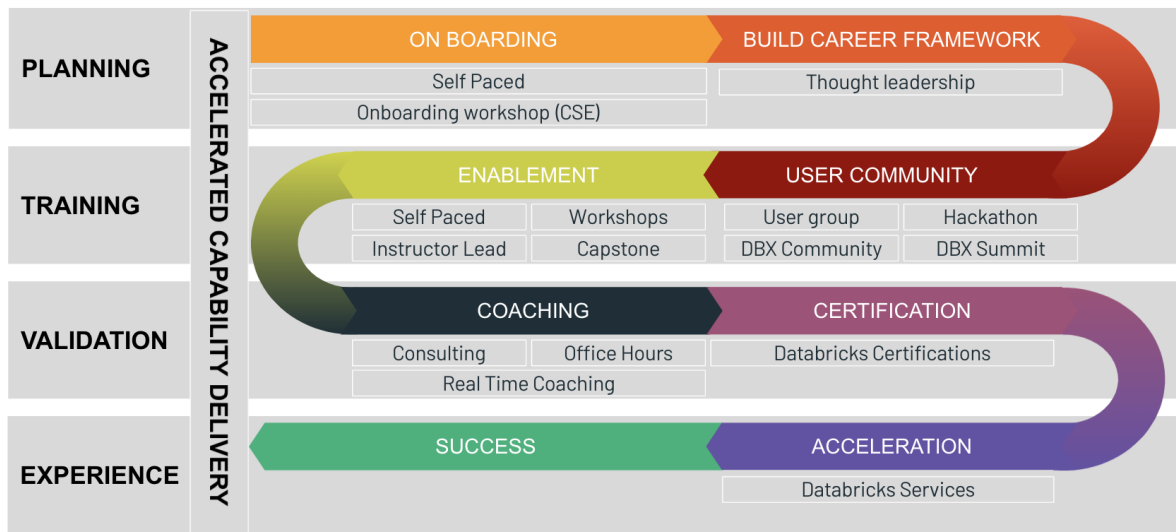
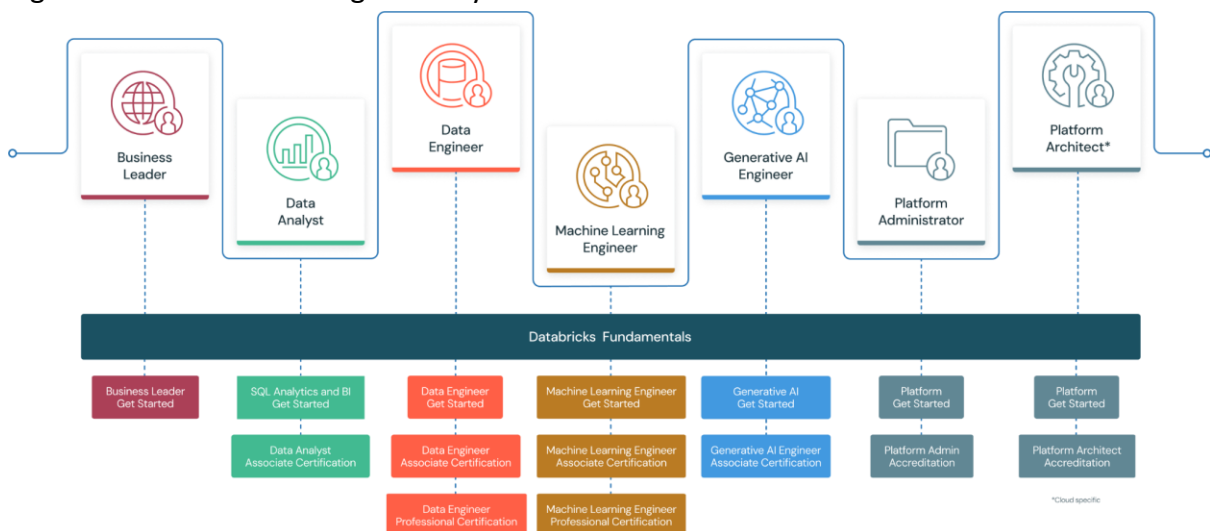


Fig 11 - Databricks Enablement Pathways

Databricks training is intended to be both hands-on and flexible. Enablement of users on the characteristics, functionality and constraints of the platform is conducted through self-paced or instructor-led learning sessions available through the Databricks Academy, or through smaller, focussed workshops aimed at solving a particular problem. The workshops are organised on an ad-hoc basis as the needs arise, conducted either in person or remotely. Typically, a workshop is led by a Databricks Solution Architect with 5-10 attendees working on a specific issue, such as deploying ML models to production.

We have developed learning paths for Business Leaders, Platform Administrators, SQL Analysts, Data Engineers, and Data Scientists, a sample of which can be found below.

Fig 12 - Databricks Learning Pathways



We have executed large and small training and enablement programs for our customers of all scales and needs. We work with customers to understand their training needs and build learning paths for their organisation. This includes well-defined steps and modules for your team based on:

- Defining personas, segments, and proficiency levels
- Conducting a skills assessment (in collaboration with your the customer's learning and development team)
- Building proficiency-based, multi-step skills development plans
- Accreditations and certifications

During the delivery process itself, we place emphasis on using “two in a box approach” where possible to ensure ongoing and smooth knowledge transfer all through the project.

We also encourage shadowing during the delivery and capture the process via recording or through documentation and allocate time for a further handover, to ensure all aspects for the work done and knowledge are handed over to the customer.

Finally, all Databricks' documentation is comprehensive, freely available and widely accessible.

Implementation Plan

Databricks provides an Account Manager to support our customers and build an ongoing relationship and commercial requirements. In addition, our account teams include Solution Architects who can provide technical advice and guidance so that customers can leverage the Databricks platform to deliver their business outcomes.

The core account team of our customers are supported through their entire journey via various programs, to ensure they are constantly engaged and updated with the latest developments, so they can decide on an optimised method to ensure their success. A sample of the various options available to customers are:

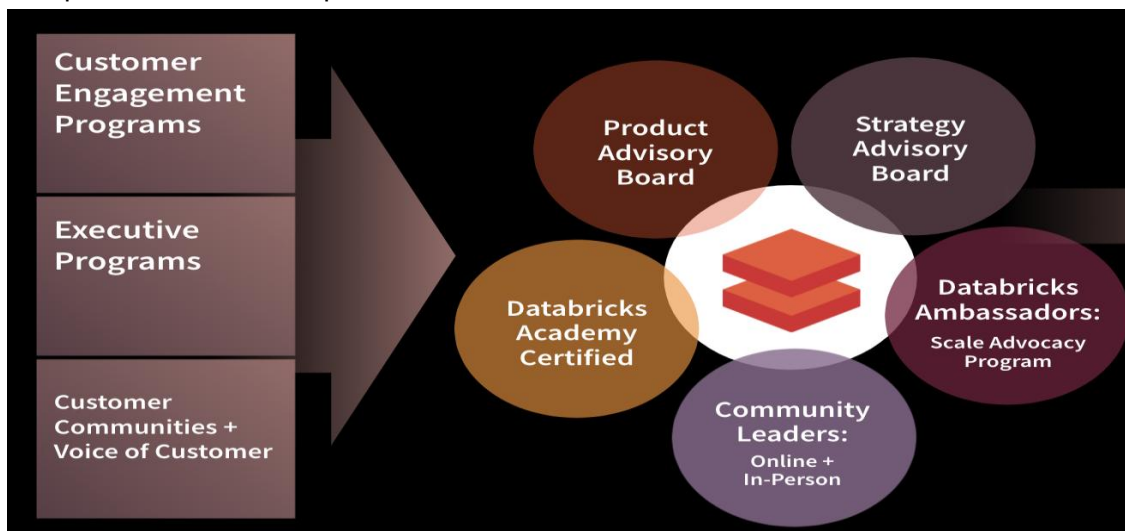


Fig 13 - Databricks Engagement Model

G-Cloud 15



The ongoing customer engagement is also maintained via a governance process, agreed during the initial kickoff meeting. In addition to the Account Manager, the Databricks team comprises of Solution Architects, Delivery Solution Architects, Engagement Managers, Project Managers, Resident Solution Architects, Solution Engineers along with a plethora of other functional teams who are available to work with the customer throughout the lifecycle to ensure success.



Fig 14 - Databricks Platform Personas

Project Management and Delivery

As part of the customer journey, we follow the below milestones based approach. Each milestone represents the key elements of the experience of becoming a new customer of Databricks, ending in successful realisation of the platform's potential for business value to an organisation. Each step in the customer journey is a necessary component in the natural maturation of a customer from onboarding through to production and achieving value.



Fig 15 - Project Management Lifecycle

The overall customer success process can be summarised as below:

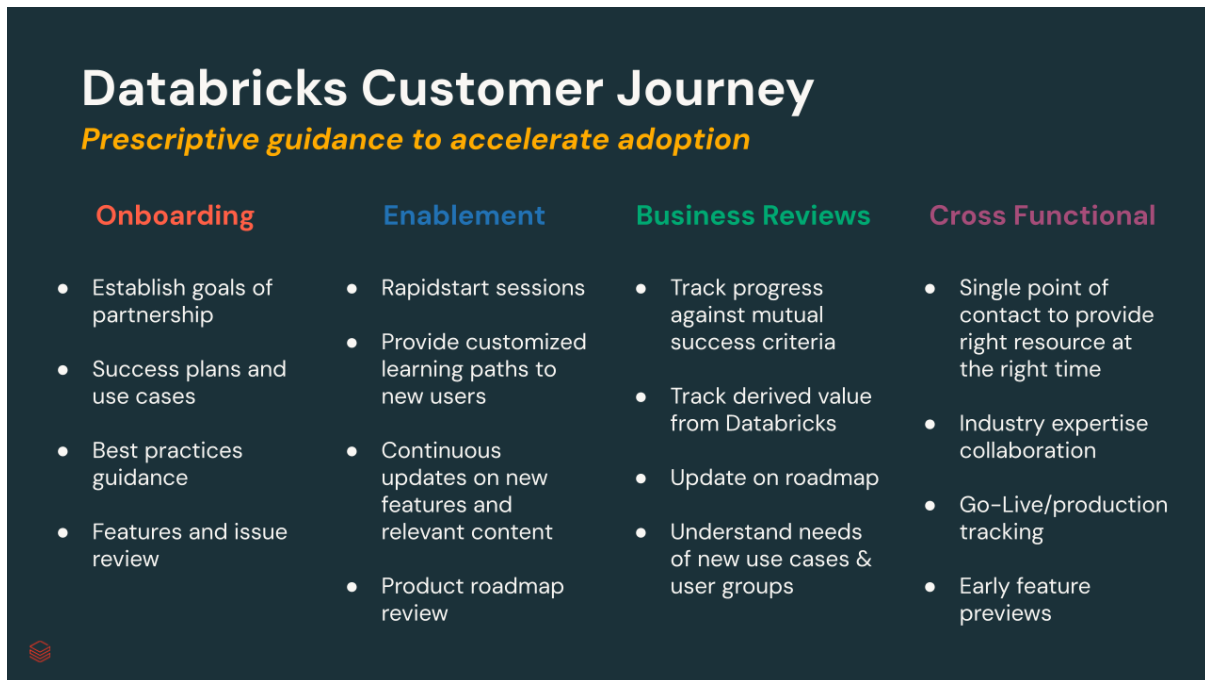


Fig 16 - Databricks Customer Journey

During this process, after the customer has been onboarded to Databricks and the appropriate solution has been scoped, the actual delivery is carried out by our team of experienced Consultants and Architects. Our team of Delivery Solution Architects, Engagement Managers and Project Managers provide service reports at agreed timelines to the appropriate stakeholders. Some suggested sample project milestones followed by Databricks are:

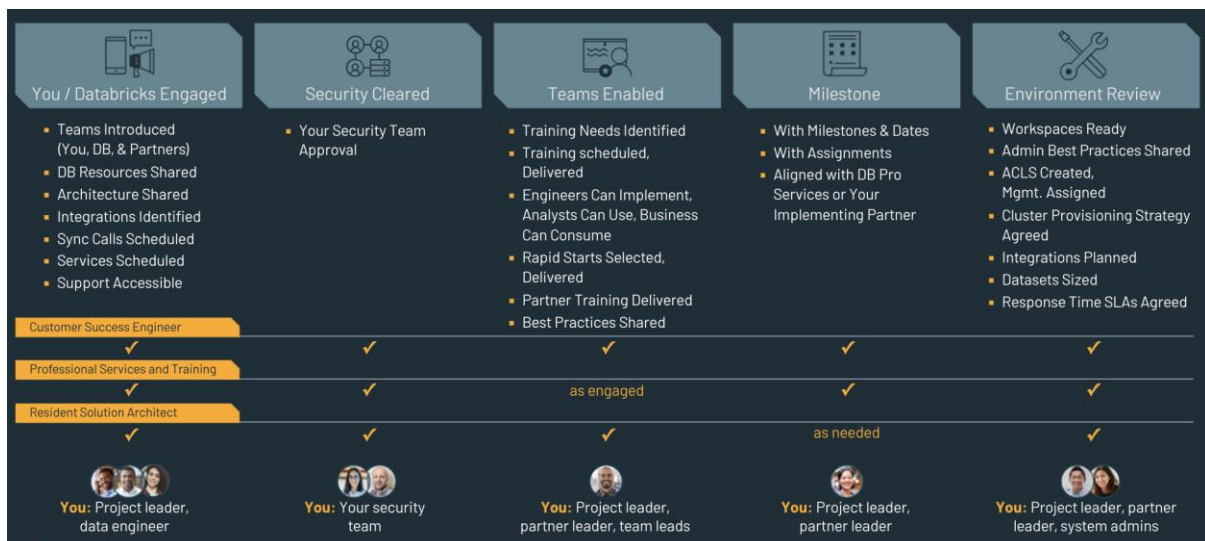


Fig 17

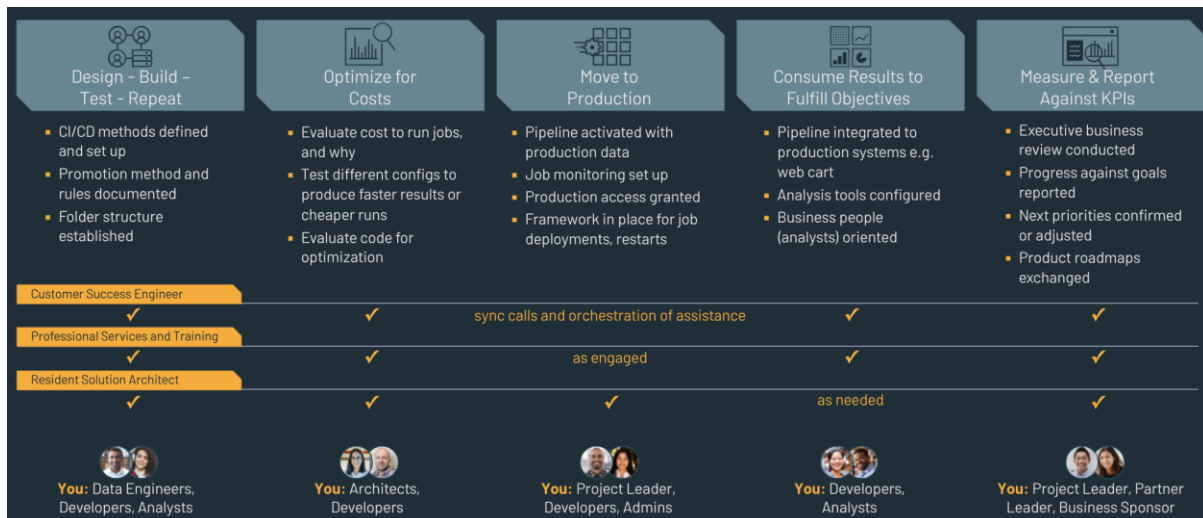


Fig 18

Methodologies

We follow an agile methodology as our implementation approach. This enables us to demonstrate value within short timescales. For instance, our recommended timeline for a POC is around 2 weeks.

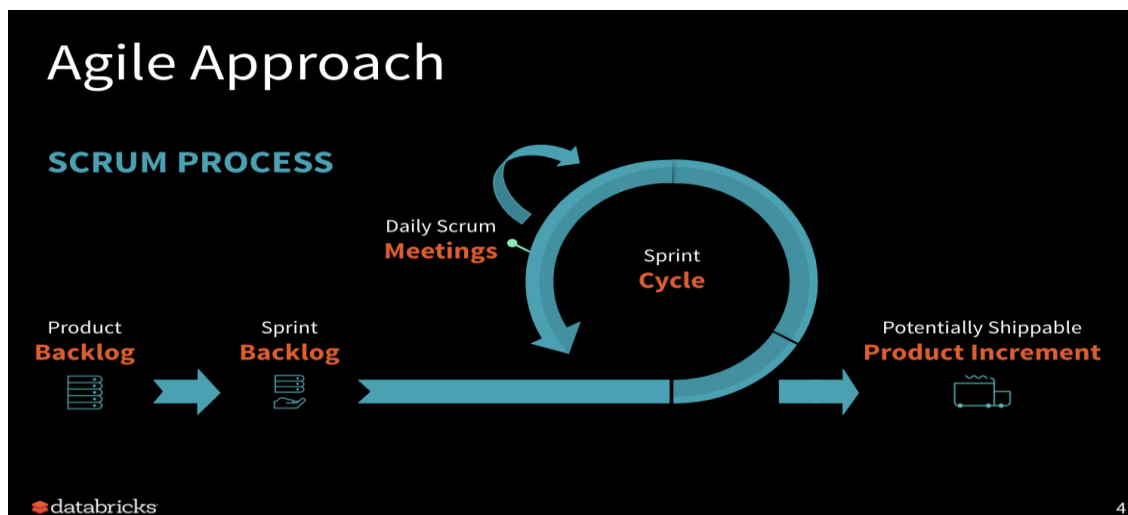


Fig 19 - Agile Development Process

Reporting

The health and status of the projects are tracked and reported back to the stakeholders in the form of weekly and monthly meetings and reports. We also track the overall success of the engagement and value delivered to the customer through the quarter business reviews and implement any course corrections, as required.

Service Management

There are many components to the Databricks product - the user front end, APIs, the Databricks Runtimes (DBRs), and the backend architectural / configuration options. Each of these receives regular updates. As an example, the DBR is the execution engine which

powers workloads running on Databricks. The DBR is regularly updated with bug fixes and performance updates. Whilst there is no strict schedule, with releases occurring when ready, historically minor versions are released monthly, with major versions released biannually or less. Databricks releases runtimes as Beta and GA versions. Databricks supports GA versions for six months, unless the runtime version is:

- A Long Term Support (LTS) version. LTS versions are released every six months and supported for two full years.
- A Databricks Light runtime. Light versions are supported until either 12 months after release or two months after the next Databricks Light release, whichever comes first.

Service Constraints

Customisation

The Databricks platform has a degree of customisation which allows which features are made available to groups of users. This is controlled by the administrators of the platform. For example, administrators may allow the Data Scientists to only access the Data Science interface, and the BI team to only access Databricks SQL. This is done in such a way that users are unaware of the areas they are unable to access; the unavailable options are simply not shown.

The Databricks Runtime (DBR)

Workloads on unsupported runtime versions may continue to run, but they receive no Databricks support or fixes.

Bugs can be reported to Databricks via support channels or by raising with the account team, which will work with the customer to create a short term solution. Internally bugs are tracked through Jira. The severity of the bug will influence how a fix is prioritised by engineering, whilst the customer's support agreement will determine the SLA for a response from Databricks. As a managed cloud service, upgrades and patches are automatically applied without any interruption of service.

To mitigate these risks, Databricks provides a predictable lifecycle and industry-leading long-term support for the Databricks Runtime, ensuring stability and security for your data and AI workloads.

Databricks Runtime Long Term Support (LTS) versions are released every six months and each LTS version receives full support—including critical security and stability fixes—for three years from its release date. Customers are encouraged to adopt LTS versions for the optimal lifespan and peace of mind in production environments, as these releases provide the most robust and reliable foundation. Non-LTS Databricks Runtime versions are supported for six months from their general availability. After end-of-support (EoS) dates, no further updates or fixes are provided, so timely upgrades are strongly recommended to maintain support coverage.

Support levels

Databricks provides a number of plans that provide dedicated support and timely service for the Databricks platform and Apache Spark. These plans are split into 3 levels of service; Business, Enhanced, and Production. At the basic level, the plans offer a response to a ticket within 1 working day, access to all updates & patches to the Databricks platform, support portal access and 2 hours per month access to Spark experts for troubleshooting problems using the products and services. At the highest level, the shortest SLA, for when Production systems are down or severely impacted such that routine operation is impossible, is a response within 1 hour of submitting the request, 24 hours a day, 7 days a week, all year. A dedicated customer support channel is available on Slack for informal communication

between the two teams, and 8 hours per month access to Spark experts. For full details <https://databricks.com/support>.

Informal support is also available through the Solutions Architect in the customer's account team. The Solutions Architect has technical expertise across the Databricks platform and will in many cases be the first point of contact for technical issues

Service Levels

Service uptime is agreed and documented in the Order Form between both parties.

Support levels are contingent on the support version purchased. There tiers of support available are here:

Databricks provides a number of support options to customers with pre-paid consumption plans that provide them with dedicated support and timely service for the Databricks platform.

FEATURE	BUSINESS	ENHANCED	PRODUCTION	MISSION CRITICAL
MULTI-CLOUD SUPPORT Support for Databricks on permitted Cloud Service Providers and Databricks-powered services;	-	✓	✓	✓
SINGLE-CLOUD SUPPORT Support for platform services on a single-chosen Cloud Service Provider	✓	✓	✓	✓
SUPPORT PORTAL ACCESS Online repository of documentation, guides, best practices and more	✓	✓	✓	✓
UPDATES and PATCHES Receive updates, bug fixes and patches without impact to your business	✓	✓	✓	✓
SERVICE LEVEL AGREEMENT (SLA) Receive support responses according to issue severity	Initial Response Time			

MISSION CRITICAL SEVERITY A mission critical production system is down or severely impacted such that routine operation is impossible	-	-	-	Within 15 minutes
STANDARD SUPPORT SEVERITY 1 Production system is down or severely impacted such that routine operation is impossible	Within 1 business day	Within 4 hours	Within 1 hour	Within 1 hour
STANDARD SUPPORT SEVERITY 2 Production issue where the system is functioning but in degraded or restricted capacity		Within 1 business Day	Within 4 hours	Within 2 hours
STANDARD SUPPORT SEVERITY 3 Issue where minor functionality is impacted or a development issue occurs			Within 1 business Day	Within 4 hours
STANDARD SUPPORT SEVERITY 4 Request for information or feature request with no impact on business operations				Within 1 business Day
SUPPORT SERVICE HOURS Live support during customer's choice of time zone	9 AM–6 PM, business days	9 AM–6 PM, business days	Severity 1 and 2: 24x7x365 Severity 3 and 4: 9 AM–6 PM, business days	Severity 1 and 2: 24x7x365 Severity 3 and 4: 9 AM–6 PM, business days
DATABRICKS STANDARD SUPPORT Limited to break-fix support for the Databricks platform	✓	✓	✓	✓
DATABRICKS CHAT SUPPORT CHANNEL Per customer dedicated real-time messaging (e.g., Slack, Microsoft Teams) channel available during	-	Optional Upgrade	✓	✓



business hours for informal communication between the two teams, such as basic questions and information exchange				
MAX NUMBER OF TECHNICAL CONTACTS The number of technical contacts with access to the Databricks Help Center or Chat Support channel (if available)	2	4	8	16
DATABRICKS SUPPORT FOR SPARK Prioritised access to the world's leading Spark technical experts for troubleshooting problems using the product and services	Up to 2 hours/month	Up to 4 hours/month	Up to 8 hours/month	Up to 8 hours/month
ADDITIONAL BENEFITS — MISSION CRITICAL				
PROACTIVE MONITORING Proactive monitoring of workspace-level health metrics by the Databricks Support Operations team	-	-	-	Yes
ESCALATION MANAGEMENT Direct access to Databricks Escalation Managers for cases with mission critical severity	-	-	-	Yes

Definitions and Terms

- Support business hours are from Monday through Friday, from 9 AM to 6 PM (North America time zones (EST/CST/PST/MST), Central Europe (CET), Singapore/China (SGT/CST) and Australia Eastern (AET) Timezone). Support business hours exclude local holidays in each timezone.
- Chat Support channel is not covered under the Support SLA response times.
- “Contact” means a unique named user at customer (whether by email address, chat ID); accounts may not be shared.
- For reported mission critical severity issues, Databricks will assign an Escalation Manager to monitor and oversee case resolution and provide updates to the customer on a 15-minute recurring basis.
- Additional assistance beyond what is included in a support plan can be purchased as Advisory Services which will be delivered by our Professional Services team.

- If no support is specified in an order, the customer's support is limited to public documentation and open resources on the Databricks Help Center.
- View the resources available through our Customer Support Handbook for more information on support definitions, processes and terms.

Outage and Maintenance Management

Databricks does not provide SLA details as they are contractual documents managed through the legal process. The current status of all Databricks availability regions can be checked and subscribed to from <https://status.databricks.com>.

Maintenance and updates to the Databricks platform are carried out without interruption to the service due to the hybrid nature of the Databricks Platform as a Service - new features, updates and maintenance releases are integrated without impact.

Downtime SLAs are negotiated on a per contract basis.

Financial Recompense Model for not Meeting Service Levels

If Azure Databricks does not achieve and maintain the guaranteed Service Levels, you may be eligible for a credit towards a portion of your monthly service fees (based on your Microsoft agreements). For GCP and AWS deployments, this will be negotiated on a per contract basis.

4. Provision of the service

Customer Responsibilities

It is the responsibility of the customer to install Databricks. Databricks support is available to assist with this. The simplest way of doing this is through the free trial, which will automatically set up Databricks in your cloud account.

Technical Requirements and Client-Side Requirements

Getting started with Databricks is simple. The free trial will set up Databricks up in the customer's cloud account, or if no cloud account is available provide access to Databricks Community Edition.

Databricks is usable by people with a wide range of skills, from advanced Data Scientists and Data Engineers through to entry level analysts and developers just beginning their careers.

Outcomes/Deliverables

Databricks' initial assessment focuses on co-developing a Proof of Concept (PoC) document with the customer. The PoC's aim is to establish the validity and value of Databricks as a platform upon which the solution to the buyer's requirements can be built. The PoC document serves as reference defining the current state of affairs (for example what is the current data landscape, what is the current architecture, defining the problem the buyer has

today) ; the short term aim of the PoC (typically conducted over two weeks, it is important to set realistic, achievable goals that sufficiently demonstrate how with further effort a complete solution could be built on the platform); how the PoC will run in order to deliver this aim (typically resourced by the buyer, with 2-3x a week touchpoints with Databricks experts to guide and assist); how the PoC relates to the larger organisation's requirements. The document concludes with a jointly agreed definition of the outcome of the PoC, typically a minimally viable product which can be put into production with a minimum of effort.

In addition to the stated aim of the PoC, where the PoC is typically being resourced by the customer the customer's technical teams get hands-on experience with the set up, administration and use of the Databricks platform as place building their solution.

Development life cycle of the solution

There are many components to the Databricks product - the user front end, APIs, the Databricks runtimes, and the backend architectural / configuration options. Databricks uses an agile development methodology.

A new feature is developed based on customer or market requirements. Customer requirements are understood via regular interactions between the Databricks team and the customer team to identify product gaps. These are relayed to the product team, and tracked using the tool Aha!. To decide if a requested feature is added to the development roadmap, factors such as the business value the customer will realise from the feature, the number of customers requesting the feature, and engineering effort are used. Once development starts, a select few customers are offered early access to the feature as Private Preview. These customers are chosen based on the benefit they will see from the feature, their willingness to be early adopters, and their relationship with Databricks. During the Private Preview the feature is still under active development, and users may experience bugs and shortcomings, which can be reported back to Databricks via regular discussions with the product team. As the feature matures, the access criteria for the preview are loosened and more customers are admitted access. For example, the business value the customer may realise from the feature may be lower, or the customer may be interested in testing the feature to decide on whether to incorporate it into their plans. Eventually the feature is released into Public Preview - this is available to all customers, but highlighted as being a Preview feature. This allows customers to adhere to policies around use of Preview features in production environments, for example. The final step after Public Preview is General Availability (GA), at which point the feature is considered mature enough for production workloads.

The Databricks runtime is regularly updated with bug fixes and performance updates. Whilst there is no strict schedule, with releases occurring when ready, historically minor versions are released monthly, with major versions released biannually or less. Databricks releases runtimes as Beta and GA versions. Databricks supports GA versions for six months, unless the runtime version is:

- A Long Term Support (LTS) version. LTS versions are released every six months and supported for two full years.
- A Databricks Light runtime. Light versions are supported until either 12 months after release or two months after the next Databricks Light release, whichever comes first.
- Workloads on unsupported runtime versions may continue to run, but they receive no Databricks support or fixes.

Bugs can be reported to Databricks via support channels or by raising with the account team, which will work with the customer to create a short term solution. Internally bugs are tracked through Jira. The severity of the bug will influence how a fix is prioritised by engineering, whilst the customer's support agreement will determine the SLA for a response from Databricks. As a managed cloud service, upgrades and patches are automatically applied such that they become available for use by the customer without any action required. Existing customer workloads are not affected.

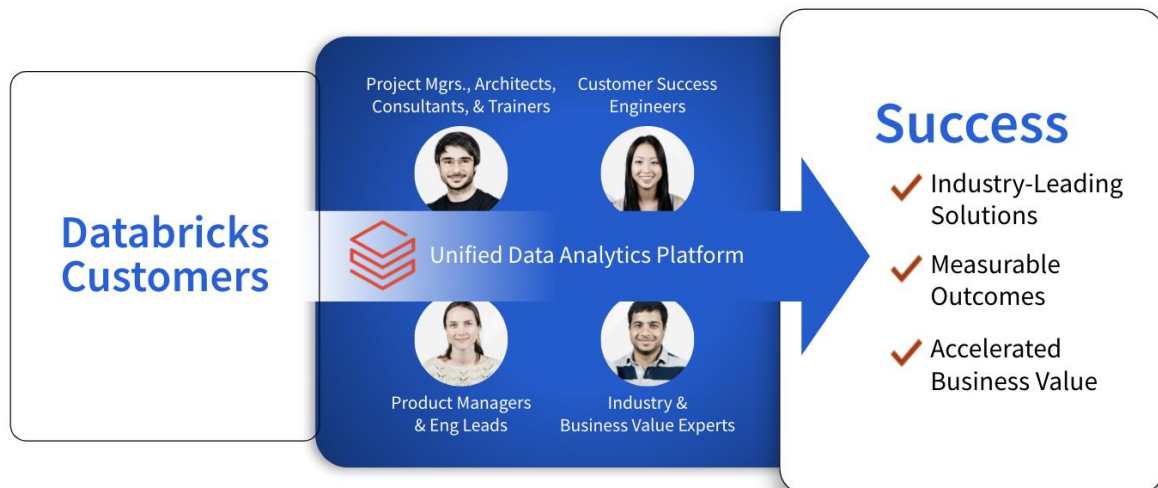
After-sales Account Management

Databricks is a value-led, software and services business. and values customer obsession as a key business objective. It has an ongoing engagement with customers, with resources are based globally, but crucially for UKG, in the UK and Europe, including security cleared personnel.

The lead role for after-sales engagement is the Databricks Delivery Solutions Architect, who engages in order to provide:

- Technical account manager
- User onboarding & Value Realisation
- Project management & triaging customer issues
- Technical Escalation and Resolution
- Joint Success Planning and Realisation
- Business and stakeholder Liaison

Their engagement is part of a wider supporting ecosystem with Databricks:



Your Team + The Databricks Team = Success

Fig 20

For high value customers, Databricks delivers after-sales engagement through the Strategic Customer Program.

Strategic Customer Program

An invitation only, strategic partnership between you and Databricks to maximize your value from Databricks consumption with ongoing collaboration to uncover, architect and execute on critical business use cases.

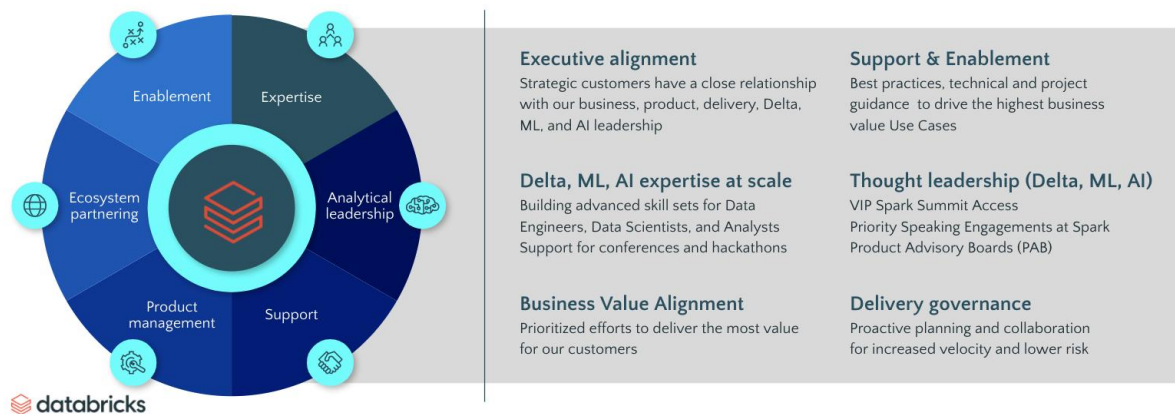


Fig 21

Termination Process

As mentioned above, Databricks is used on either a Pay As You Go basis, or a direct commitment basis. If a customer's contract for direct commitment expires, they automatically switch to a Pay As You Go model – there is no loss of functionality. Should a customer decide to fully move away from Databricks, the migration process would be

controlled by them. Upon account cancellation, Databricks deletes their account and all data within 30 days.

5. Our experience

Clients

Below is an example of UK Public Sector customers who benefit from the power of Data and AI through the Databricks platform.



Fig 22 - Public Sector Customers

Global adoption

OVER 12,000 CUSTOMERS, ACROSS EVERY INDUSTRY



Fig 23 – Private Sector Customers

Contact Details

gcloud@databricks.com