EBOOK

Taking ESG from Buzzword to Reality with Data Analytics and AI

Unified data analytics enables ESG to make a real-world impact.

databricks

The State of ESG

Consider these numbers. In 2006, when the United Nations launched an initiative aimed at promoting responsible investing, 64 companies committed to factor in environmental, social, and governance (ESG) issues when they made investment decisions. Representing some $6.5 billion under management, it was a good start for the nascent ESG movement. But few would have predicted that a little more than a dozen years later, the Principles for Responsible Investment would count more than **3,000 organizations and $80 trillion committed to ESG investing**.

ESG moved more into the mainstream when Larry Fink, the CEO of BlackRock, one of the largest asset management companies in the world, challenged his fellow CEOs in January 2020 to make sustainability a key measure of corporate performance.

And as many early adopters of ESG have already discovered, the two aims are not in competition. On the contrary, ESG is a solid path to achieving resilience and sustaining profitability in a rapidly evolving economic, political, and environmental landscape.

> "The world has woken up —
> [ESG] is good business. This is about resilience."
>
> **KATHLEEN MCLAUGHLIN**
> EVP and Chief Sustainability Officer, Walmart

What's the difference between Corporate Social Responsibility and ESG

- CSR is a framework for the principles organizations adopt in order to be more ethical, sustainable, and transparent.
- ESG is the empirical, data-driven measurement of an organization's fulfillment of CSR goals.
- ESG is the CSR report card.

According to McKinsey, ESG links to cash flow in five important ways:

1. Facilitating top-line growth
2. Reducing costs
3. Minimizing regulatory and legal interventions
4. Increasing employee productivity
5. Optimizing investment and capital expenditures

THE BOTTOM LINE:

93% of the world's largest companies focus on ESG

Examples of ESG

ENVIRONMENTAL	SOCIAL	GOVERNANCE
Climate change	Human rights	Bribery & corruption
Waste	Modern slavery	Executive pay
Pollution	Child Labor	Board diversity & structure
Resource depletion	Working conditions	Political lobbying & donations
Deforestation	Employee relations	Tax strategy

Source: UNPRI

The Threat of ESG 'Greenwashing'

While there is growing awareness that ESG is the right thing for companies and investors to focus on, the movement lacks standards. Unlike financial disclosures, there are no generally accepted principles on how ESG metrics are disclosed by companies. As a result, it has become nearly impossible today for investors, partners and the public to independently measure and verify how companies are progressing towards their sustainability goals at scale.

Standards will eventually come but until that time there's a danger that Sustainability efforts increasingly becomes a marketing term companies use to appease investors and consumers, a kind of greenwashing that undermines the goals of the movement and devalues the hard work of organizations who take ESG seriously.

The solution to this challenge starts with data. It's only through data analytics and AI, and an ability to process vast amounts of ESG data (often unstructured data), that organizations and investors can prove that ESG actually means something - actually can make a positive difference in the world. The ultimate goal should not just be a yearly report that addresses an organization's ESG priorities. But rather an ability for companies to measure and report ESG progress in a timely manner and for investors and other stakeholders to be able to verify it with data. This is possible now, but many organizations lack the tools to perform data analytics and AI at this scale.

> "Without solving the foundational data problem, companies will not have an accurate understanding of their own ESG metrics. This is why ESG could be best viewed in the prism of technology, not policy."
>
> **JUNTA NAKAI**
> Global Industry Leader for Financial Services, Databricks

Real-time ESG Use Cases

CEOS: With a wary eye on their own jobs and bonus provisions, CEOs will want to know how well they are tracking versus stated ESG goals before board members, investors, regulators, and the media do.

SUPPLY CHAIN MANAGERS: Companies will no longer have to rely on self-reporting to know if vendors are meeting ESG standards so they can navigate emerging risks and enhance supply-chain resilience.

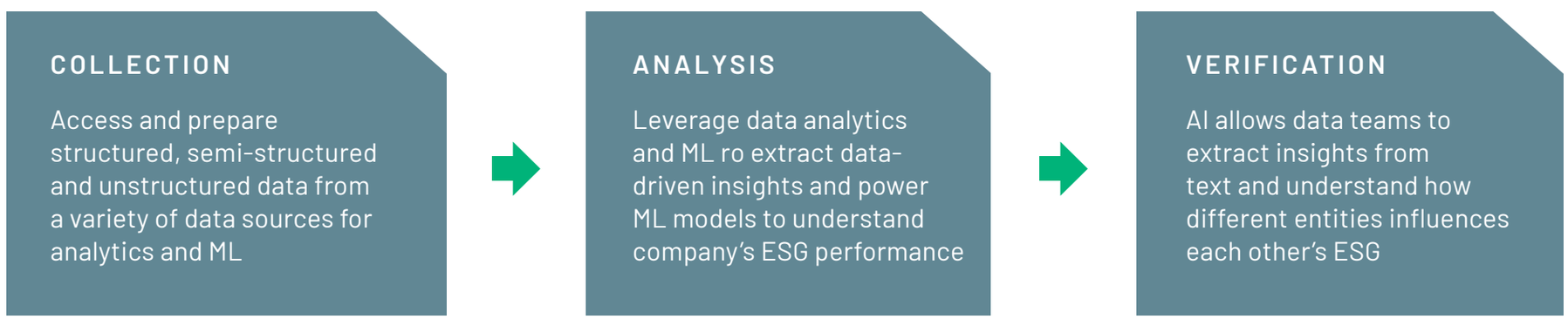
INVESTORS: Analyzing big data and alternative data at scale to make responsible investing more transparent and impactful.

CONSUMERS: With more confidence that brands are living up to the ESG ideals they advertise, consumers can make more informed decisions about their purchases.

RISK MANAGEMENT: What is the company's exposure to the E, the S, and the G? Understanding sustainability performance of a company you lend money to get a more complete view of risks.

INVESTOR RELATIONS/CFOS/CMOS: Obtaining real time indications of how the market/media is perceiving a company's ESG performance and having the data points to be able to push back on unfavorable rating/coverage.

Data Science in Service of ESG: Collection, Analysis, and Verification



The process starts with collection, which in itself is a massive undertaking. Even a relatively small company could be drawing on wildly diverse data sets in multiple formats, covering everything from water use to carbon footprint of a key supplier or exit interviews from employees who left the company. For most organizations, these datasets are spread across 100s of operational systems, data warehouses and vendors. Finding, organizing, and making certain all of this data is of good quality is in practical terms impossible without data cleansing and preparation.

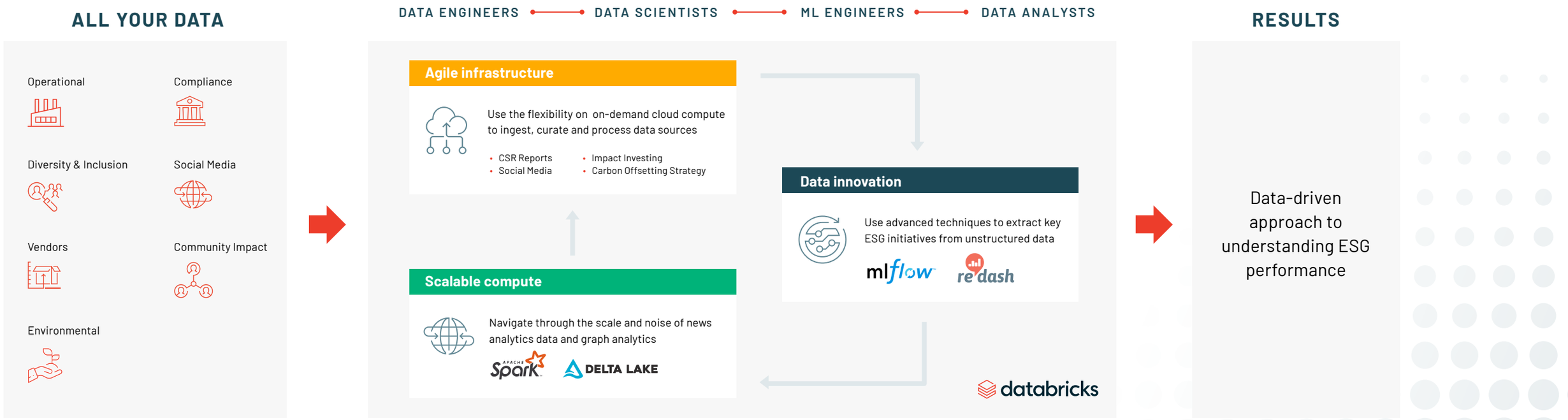
Once the data is assembled, it has to be analyzed. Advanced data analytics and dashboarding allow organizations to know if they are meeting ESG goals and make necessary corrections if they aren't. But data analytics on this scale calls for the ESG data to be centralized in a modern, cloud-based storage infrastructure and data pipelines that ensure reliability and data integrity. A modern cloud data architecture that can deal with structured, semi-structured and unstructured data coming in both batch and real-time is a foundational component. Analysis of this sort also requires that data scientists collaborate effectively and efficiently across business lines.

ESG is meaningless without verification. This is true across all use-cases from investing to supply-chain analytics. At the moment, ESG verification is essentially an honor code by which organizations pledge to follow the rules. At best, also taking data vendors ratings at face value. History shows companies don't always live up to their marketing and that ESG vendors don't always catch bad actors. AI can add a new dimension to verification by using techniques from natural language processing (programmatically extracting information from text) to graph analytics (learning how different entities influence each other's ESG). While AI provides an opportunity to programmatically improve ESG, organizations now face the challenge of adopting and managing the broad set of machine learning tools available today to make this promise a reality.

All of this can be accomplished with a Unified Data Analytics Platform that allows organizations to easily and quickly gather data from disparate sources and store it for accurate analysis in a highly reliable and high performing data lake that accelerates the rate at which teams can turn data into insights using machine learning and AI at scale. Big data and AI can augment the ESG ratings data that many companies buy today or drive completely novel insights.

How Databricks enables a data-driven approach to ESG for sustainability

Databricks offers the ideal solution to address ESG challenges by providing organizations and investors a holistic and data-driven view of ESG performance with a Unified Data Analytics Platform.



databricks

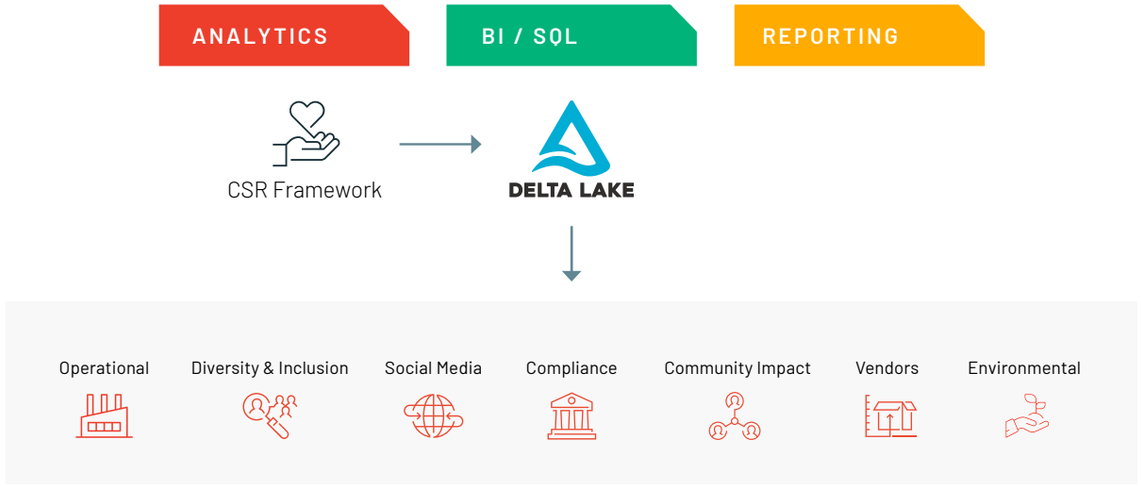
Preparing ESG data for insights at scale

Managing your entire data journey with a unified data service

ESG analytics starts by aggregating and processing large collections of diverse data from vendors, IOT, news, geospatial and emissions data sources. For many organizations this requires a large infrastructure build out. The Databricks Unified Data Service provides a reliable and scalable platform for your data pipelines, data lakes, and data platforms. Manage your full data journey, so you can ingest, process, store, and expose data throughout your organization through machine learning or data analytics.

DATA INGEST	DATA PIPELINES	DATA LAKES	DATA CONSUMERS
Pull data across all your different data sources, data storages and data types, including batch and streaming. Leverage a library of connectors, integrations and APIs for all your needs.	Run scalable and reliable data pipelines. Use Scala, Phython, R or SQL to run processing jobs quickly on distributed Spark runtimes, without having to worry about the undelying compute.	Build reliable data lakes at scale. Improve data quality, optimize storage performance and manage stored data, all while maintaining data lake compliance and security.	Use your data lake as a shared source of truth across Data Science, Maching Learning and Business Analytics teams — BI dashboards, production models and everything in-between.

Build Reliable Data Pipelines at Scale



One of the most common problems organizations face when dealing with massive volumes of real-world data across disparate sources is that it can become unreliable, low quality, and challenging to manage. Many organizations turn to data lakes to aggregate their big data cost-effectively, but this poses its own challenges.

Delta Lake, an open-source storage layer for data lakes, is natively integrated in Databricks to provide ACID transactions for big data use cases, including batch and streaming ingests, fast interactive queries, and machine learning. Designed to help data professionals build robust production data pipelines at scale, Delta Lake brings data reliability and performance to your data lakes so companies can make use of the real-world data so critical to ESG. A single source of truth also allows organizations to time travel and track progress over time.

Extracting business insights from Delta Lake

With data centralized for easy access, data analysts can easily connect directly to their most complete and recent data in the data lake with Delta Lake and SparkSQL, and use your preferred BI visualization and reporting tools for more timely business insights.

For a completely seamless experience, you can leverage Redash to easily visualize and share your data via intuitive dashboards and queries.

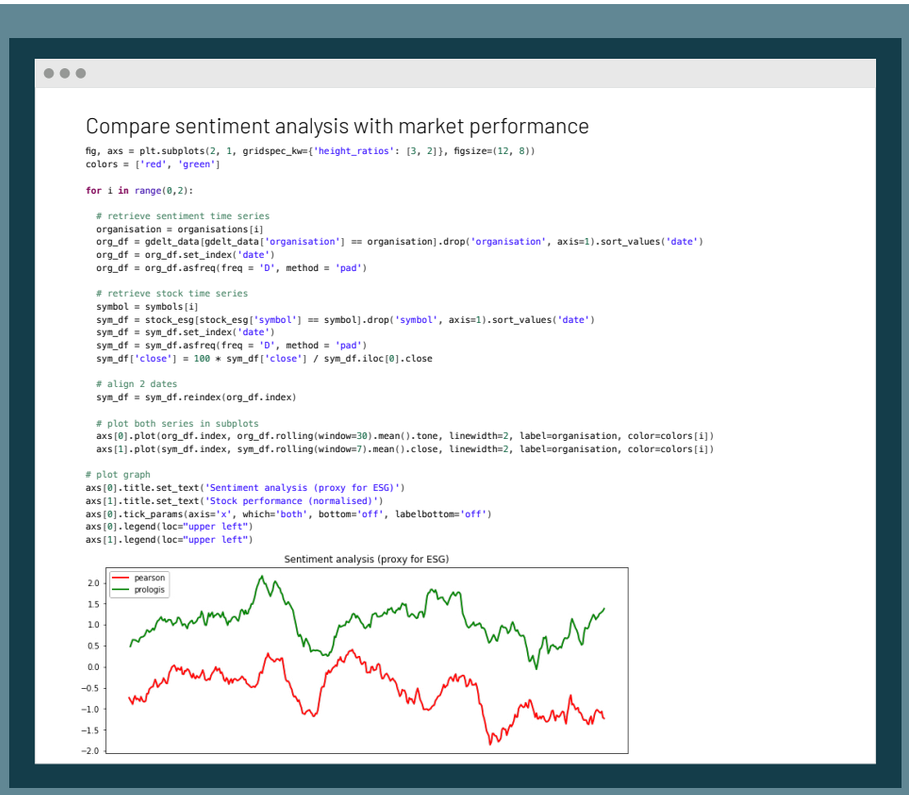
databricks

Applying advanced analytics to ESG data

Bring teams together to accelerate innovation

Key to ensuring a rapid pace of data-driven innovation is to foster a collaborative environment that empowers data teams to work better together across the enterprise. When data teams work together effectively, they are able to more easily focus their ideas, skills and energy toward accomplishing amazing things.

Through an interactive workspace, data scientists can easily collaborate on data, share models and code, and manage the entire machine learning lifecycle in one place. Databricks notebooks natively support Python, R, SQL, and Scala so practitioners can work together with the languages and libraries of their choice to discover, visualize and share insights with stakeholders.



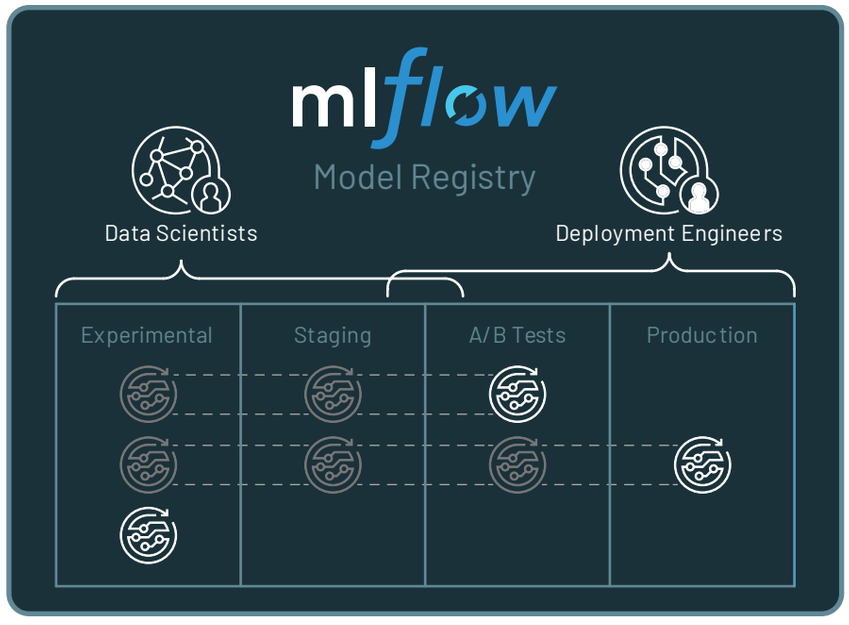
Collaboratively build and manage models from experimentation to production.

databricks

Applying advanced analytics to ESG data

Managing the complete machine learning lifecycle

Successfully building and deploying a machine learning model can be difficult but essential to ESG analytics and reporting. Think of all the text that needs analyzing in corporate sustainability disclosure or the insights that need to be mined from social noise. MLflow is an open-source framework that streamlines the machine learning lifecycle — allowing data scientists to reproduce a pipeline, compare the results of different versions, track what's running where, and redeploy and rollback updated models. MLflow is natively integrated in Databricks Unified Data Analytics Platform allowing your teams to seamlessly connect their data pipelines to their models in development and track across the entire ML lifecycle.



MLflow streamlines the entire ML lifecycle with enterprise reliability, security, and scale.

Conclusion

When organizations obtain a real-time picture of the diverse data sets that constitute an accurate understanding of ESG, they can analyze it, individualize for executives and specific teams, predict issues in advance, make corrections as quickly as possible, and lay the foundation for a virtuous cycle of resilience. In short, unified data science is the key to realizing the world-changing promise of ESG.

> Read **this blog** to dive deeper into how data and AI can help with ESG.

Deliver a data-driven approach to ESG with Databricks.

REQUEST A DEMO TODAY

databricks