



HOW TO BUILD AN EFFICIENT ENTERPRISE ESG DATA STRATEGY IN BANKING WITH CLOUD AND MACHINE LEARNING

TABLE OF CONTENTS

- 3** Introduction
- 5** Current ESG Situation in Banking
- 6** Data & Technology Challenges in Banking
- 9** What's Needed in an Enterprise ESG Strategy
- 10** Banking Use Cases
- 12** Building ESG Data Workflows with Snowflake, Microsoft Azure, & Intel
- 14** Conclusion
- 15** About Snowflake, Microsoft, and Intel



INTRODUCTION

Since the United Nations created the Standing Committee on Finance 12 years ago within its broad sustainability program, financial markets have embraced environmental, social, and governance (ESG). Capital market institutions and banks raise debt, offer loans, and provide advice to direct private funds toward the companies and projects that can help hit sustainability targets. Those targets include the U.N.'s Sustainable Development Goals (SDGs) and the associated Paris Agreement on climate mitigation.

The movement has won support from investors too—particularly new, younger investors with many years of investing activity still ahead of them—prompting a clamor among financial institutions to differentiate themselves with ESG products and services. Pressure from consumers means that ESG has also become a matter of reputational importance.

Acceptance of sustainability markets has accelerated over the past 24 months. A milestone came at the COP26 in Glasgow in 2021, when a coalition of asset managers, banks, and other financial markets participants, with a **combined \$130 trillion** in assets under management, committed to help fund the transition to a decarbonized future.

However, the ESG wave has so far been characterized by a lack of standards and transparency. This is partly due to the fragmented nature of the many organizations promoting its adoption. Unlike other more established investment methodologies, ESG lacks clear definitions, processes, and data types. Nevertheless, organizations face increasing pressure from regulators to be more transparent about their impacts on the world to prevent greenwashing.

The key to changing this situation lies in data. However, while there is no shortage of ESG data—the volume of information being generated is increasing at a fantastic pace—by its very nature, that information is difficult to ingest into analytical systems.

Despite the tsunami of data going into banks' systems, many of the data sets are incomplete, not in time sequence, or simply erroneous. Without clean, high-quality and well-governed data, financial institutions are unable to provide the transparency that investors demand to run due-diligence processes on their portfolios, devise risk management strategies, or comply with a growing body of regulatory disclosure requirements.

Consequently, efficient and effective data management is essential. More than simply making data manageable and usable, chief data officers need to ensure that information is accessible across the entire enterprise and to clients, counterparties, and others. By migrating ESG data and analytics into modern cloud platforms while also leveraging machine learning platforms, banks and other financial institutions will be better positioned to understand and extract insight from the data already at their fingertips.

This paper discusses the challenges facing financial institutions with respect to ESG data, and offers guidance on how to implement enterprise-wide ESG investment and risk-management strategies along with the technologies to build out ESG data pipelines. It will also explain how banks and other financial institutions can harness the capabilities of Snowflake and Microsoft Azure Machine Learning in adopting a cloud-first data mesh strategy with an ESG data domain, underpinned by self-service capabilities to support a persuasive set of use cases across the enterprise.



CURRENT ESG SITUATION IN BANKING

ESG investing is acknowledged to be far more important than its initial perception as a box-ticking exercise suggested. Early appeal to the institutional marketplace centered on the “good PR” from being seen to support ESG initiatives and the social/diversity and environmental benefits they were designed to promote.

But as ESG has gained traction from the investing public, institutional investors are recognizing the opportunity for financial gain from maintaining a robust ESG program. With regulators beginning to address inconsistencies in reporting frameworks as they seek to snuff out greenwashing, firms are planning for new responsibilities with respect to their own activities as well as those of their clients and the firms whose securities they trade or hold.

Increasing regulations present a complex set of organizational, data sourcing, and data management challenges. For banks, that has meant rethinking how they address the data management requirements of the business units that have been most affected by the sustainability transformation, particularly commercial banks and capital markets.

To give some perspective on why banks are getting involved, a recent report by the Climate Bonds Initiative found that sustainable debt exceeded \$1 trillion in 2021, up 57% from the previous year, taking the total volume of such loans issued since the market was created in 2007.

The changing regulatory landscape is also putting pressure on banks to maintain strong and agile data management setups; in turn, that is having an impact on customer relationships. The E.U.'s Markets in Financial Instruments Directive (MiFID II), for instance, has made it compulsory for advisers to seek the sustainability appetites of their customers and outline the products into which they could invest. This has highlighted the social pressures from investors who want only to invest in companies that have made public commitments to lowering carbon emissions and have set ESG goals. As a result, banks have been compelled to bring ESG into their sales functions, where data is now consumed not only as a marketing tool but also as a regulatory obligation.

ESG is not the only factor that has prompted banks to reshape their data and technology processes to improve customer relationships, but it is an illustration of how sustainability is no longer the focus solely of risk-management teams; it is now an overlay that must be applied to decisions made in every part of the business.

DATA & TECHNOLOGY CHALLENGES IN BANKING

DIVERSE AND FRAGMENTED LEGACY SYSTEMS

The growing computational demands of new data-driven sustainability analytics and risk management platforms is putting strain on the fragmented legacy infrastructures that many banks still operate. Not only are they slow and lacking in power and capacity, many also lack integration with other systems within the enterprise. A by-product of that has been the pooling of data in disconnected silos.

Without the connective tissue that comes with cloud architectures, this data will remain largely unused with poor governance and restricted permissioning, making large swathes of information inaccessible to the front office and data science teams that need it most. This puts the organization at a disadvantage because it impedes the enterprise-wide utilization of data that will inevitably lead to unlocking its full value. One example of this is the inability to leverage machine learning in the process of automating data pipelines and management processes, as well as downstream data visualization or business intelligence workflows.

To address these issues, banks are investing heavily in new technologies to help with ESG disclosure and reporting, underwrite ESG-related products like green bonds or sustainability-linked bonds, run climate risk models and analytics, and process unaudited emissions data from their corporate clients.

POOR INTEGRATION

For sustainability data to be of value, it must be mapped and integrated with other data sets, including market and financial data. But the breadth of data formats in which ESG is reported, which can include satellite imagery, web-scraped sentiment data, and carbon-emissions calculations, as well as the variable quality in which it is packaged, makes this difficult. It's only by integration with the organization's broader data that organizations can mine valuable insights from their ESG data. To enable that, banks must subject its ESG data to pre-ingestion cleaning processes that require analytical and computational power that many simply don't have within older, fragmented infrastructures. And some banks are even investing in machine learning capabilities that automate their entity resolution workflows.

Data integration is the biggest challenge to accurate climate risk analysis. According to technology advisory firm **Avanade and the European Financial Management Association (EFMA)**, almost one-third of banks are struggling to embed climate risk data into their risk management frameworks.

POOR DATA DEFINITION

There are no hard and fast definitions of what sustainable investing is, but a plethora of frameworks exist for reporting the data. Consequently, ESG can mean different things to different organizations. ESG investing has been subject to a fragmented regulatory environment and diverging standards, which present obstacles for firms seeking to comply with ESG initiatives directed from senior management of their organizations and external bodies seeking to establish best practices.

The market expectation is that recent regulatory moves are finally bringing a more viable structure to ESG compliance. Since the introduction of the European Union's Sustainable Finance Disclosure Regulation (SFDR) in 2021, the Securities and Exchange Commission (SEC) has issued its own set of climate-related disclosure proposals. And in terms of reporting standards, the creation of the International Sustainability Standards Board (ISSB) under the IFRS at COP26 has seen the coalescence of several standards-setting bodies. Nevertheless, few predict a full convergence of reporting standards soon, if ever.

COMPLEX DATA SOURCING

The lack of standardization has also contributed to the necessity for organizations to engage with multiple vendors. The fact that no single vendor can provide a full range of ESG data sets is also driving this trend. Both create data sourcing issues.

Third-party risk has been an issue for some time, and so-called Know Your Vendor practices have proliferated throughout financial services firms as they seek to mitigate risk posed by defective technology products and the like. This kind of scrutiny now applies from an ESG perspective.

DATA MANAGEMENT CHALLENGES

Vendors provide data in a variety of ways and compile scores and ratings according to proprietary methodologies that are rarely made public. Making sense of these in the round, therefore, is impossible without a clear program of data normalization and mastering. Physically, it's not always convenient to engage some vendors because their distribution practices may not be compatible with the bank's own systems.

INCONSISTENT DATA QUALITY

That leads us to another key challenge: data quality. Financial professionals need to make day-to-day decisions, but this is at odds with the frequency of public nonfinancial disclosures, which can be reported in a wide array of formats, often incompletely and occasionally corrupted. This is partly the consequence of the novelty of ESG reporting; most companies had never needed to report this kind of information before and, for many, adjusting to this new landscape has been troublesome. The European Parliament's recent ratification of the **Corporate Sustainability Reporting Directive** is expected to have a galvanizing effect on companies.

Gaps in the data also exist because different jurisdictions have adopted ESG disclosures at their own pace, producing a patchwork quilt of regulatory rigor. That has meant companies in some places, like the U.S., report less information than their counterparts in Europe, where oversight is more stringent.

RESPONSIBLE AI

Machine learning models developed with these inconsistent financial training data to predict fraud or normalize data could generate biased predictions and lead to negative consequences for the business. It's critical to keep track and visualize factors such as which data was used to train models, how features are impacting model predictions, and whether there was sufficient data representation from different ESG data providers.

DATA DISTRIBUTION ISSUES

Ensuring that ESG considerations are applied to the firm's entire activities is a challenge. Trading and investing teams need accurate and accessible ESG data for their own research and to benchmark their portfolios. More than that, they'll want access to the data that underlies research models, scores, and ratings.

They aren't the only parts of a bank with such demands. Compliance teams will require the same data to fulfill the bank's reporting obligations. That will include the third-party information that complements an organization's own datasets. In Europe, client advisory teams are obliged to provide sustainable investing strategies to individuals.

Each of those use cases depend on easy access to the fullest ESG data sets available. It calls for processing, secure storage, and management of that data, and the maintenance of the golden source so that each part of the bank is reading from the same script.

OTHER CONSIDERATIONS

Another challenge is organizational. In many cases, the person responsible for delivering the obligation under a policy or regulation doesn't have the proper authority or resources to deliver on that obligation. In such instances, it falls upon various intermediaries within the firm to carry out the required tasks, which can be difficult to manage and often leads to quality of delivery issues.

The need to assemble the right people, processes, and data to meet ESG obligations is challenging in several ways. For a start, the expertise and IP required to understand the requirement and translate it into actions, workflows, and responsibilities is expensive.

Firms are also struggling to deal with the many competing and changing standards and regulations, which are difficult to manage and process. While the emergence of the E.U.'s SFDR represents a stick in the ground to some extent, many other initiatives may be pursuing the same end solution but are traveling by different routes. This can be distracting and confusing for those charged with translating the requirements into actions and tasks.

For smaller banks, some of this may be moot—they may not have the resources to buy in data from third-party providers and will need to rely on information directly reported or obtained through engagement. Additionally, they may lack the technology stack or automated workflows that can help in the integration of ESG data with financial data. A consequence is likely an inability to create proprietary ESG metrics that can provide a dependable and consistent internal benchmark against which to gauge performance.

In short, banks are not ready to fully embrace ESG workflows. Data problems mean that only one-quarter of banks have a climate risk model ready today, according to the survey of 51 banks by Avande and EFMA. And while a further one-third plan to have one in place in six months, more than 40% will not be able to test climate scenarios for at least another year, and 12% will have to wait two years.





WHAT'S NEEDED IN AN ENTERPRISE ESG STRATEGY

It's clear that the ESG project is placing a huge data and organizational burden on many financial services companies. It's also clear that ESG success is going to be dependent on robust and agile data management structures, including the implementation of machine learning and MLOps, to streamline and scale ESG workloads.

The good news is that all of these issues can be overcome. Cloud-based management architectures and machine learning capabilities provide the scale and flexibility to manage an organization's sustainability program efficiently, at low latency, and cost-effectively.

A well-managed and constructed architecture within the cloud provides banks with the foundations to harness the huge possibilities of their ESG data. Systems that are integrated within the cloud and leverage machine learning platforms make it easier to automate processes across an enterprise, ensuring that ESG data is properly linked to the necessary workflows, generating insights that better inform decision makers across the enterprise to drive innovation, and meeting business objectives.

The cloud offers the compute power and storage capacity to absorb the volume of data necessary to run ESG operations effectively. With that sort of firepower, banks can aggregate data to create their own sustainability scores for assets and host the analytics needed to arrive at those metrics. Further, it enables individual teams to build and work from their own ESG workflows without taking bandwidth from others working separately or creating troublesome governance bottlenecks.

Probably the single-most important result of a migration to the cloud is the breaking down of data silos. Enabling information to flow freely across teams, departments, and functions is best achieved by creating a single golden source of data that can be widely assessed and utilized. When cloud-hosted, data governance and management protocols can be quickly and easily applied and updated to suit, maintaining a nimbleness that will enable enterprises to respond rapidly to market, operational, and regulatory changes.

That flexibility and compute power is vital for the effective pre-ingestion processing of ESG data to put into a form that is usable for the entire enterprise. Machine learning capabilities and integrated data visualization dashboards will not only deliver scale and efficiencies across data science teams but also facilitate faster data and model consumption across the business or in customer-specific applications.

Finally, to address the challenges of ESG, substantial dedicated programs and changes to workflows and practices are needed. Transparency is fundamental to the success of financial services practitioners' ESG programs, both internally and externally to supervisory bodies, their clients, and the public. Without evidence of their efforts, market participants may be accused of greenwashing and will have to face the reputational repercussions that entails.

BANKING USE CASES

Banks are using ESG data in many ways. Examples include:

COMMERCIAL FINANCE AND DEBT ISSUANCE

Most importantly for the aims of the U.N.'s SDGs, banks help to direct capital to companies and projects that are best positioned to bring about change. They issue loans and other credit facilities to fund corporate programs that reduce carbon emissions, ease social ills, and strengthen corporate governance.

Similarly, they coordinate the packaging and sale of green bonds and sustainability-linked credit that corporates can sell to raise funds for their sustainability projects. This part of the market has surged in the past couple of years, **tripling to more than \$1.5 trillion** between 2019 and 2021, according to Bloomberg NEF data.

None of these activities can be achieved without the use of ESG data and efficient data sharing capabilities. When properly managed and processed, that information drives analytical models to identify which assets are having the most beneficial impact (and those that aren't), what activities should be funded to achieve sustainability targets, and what the return on those investments might be.

SUSTAINABILITY PERFORMANCE ANALYSIS

Sell-side firms have identified the need for their clients to validate the sustainability claims of the entities they invest in. Emerging regulations like the E.U.'s SFDR will require asset management firms to file disclosures for funds classified as sustainable. In turn, this will require them to conduct due diligence on components of these funds.

To date, the response to this due diligence requirement among the buy side has been lackluster; many firms opt for box-checking solutions in the form of environmental ratings and rankings offered by mainstream suppliers. But innovators among sell-side securities services and fund administration groups are beginning to establish themselves as experts in the space by developing data and analysis aimed at helping their buy-side clients with this due diligence.

RESEARCH

Apart from incorporating ESG data into their fundamental research services, banks are also building out quantitative research capabilities that bring together larger volumes of structured and unstructured ESG data with quantitative analyses and machine learning capabilities. This is enabling them to build multi-factor models, backtest and perform historical analyses, and ultimately, meet the needs of their clients. For these teams, the ability to access and map ESG data to a vast portfolio of securities, as well as to collaborate with that data, is critical.

To power ESG research workflows, banks are:

- Streamlining their data management processes—including leveraging machine learning to support entity resolution requirements—necessary to minimize time and resource-intensive data wrangling work and maximize more value-add activities like model building and backtesting
- Investing in cloud platforms to benefit from near-unlimited scale and compute, enabling them to run backtests more efficiently and in a timely way
- Building out machine learning capabilities to help create and test complex investment ideas that translate to alpha generation



BUILDING ESG DATA WORKFLOWS WITH SNOWFLAKE, MICROSOFT AZURE, & INTEL

Migrating ESG data, processes, and analytics to the cloud and leveraging machine learning tools provide the optimal environment to achieve those aims. Snowflake and Microsoft, in collaboration with Intel, are able to bring high quality end-user value for scalable cloud services. Together, they offer the performance, flexibility, and analytical power that senior data managers need to successfully meet the mammoth task of managing their ESG systems.

To capitalize on the opportunity for closer customer relations and an improved understanding of client requirements, banks are investing in cloud transformation initiatives to modernize the data experience of both their internal users and customers. The partnership between Snowflake and Microsoft Azure, and their collaboration with Intel, enables banks to leverage capabilities like data access, data collaboration, and machine learning to improve business-critical banking workflows, including debt underwriting, sustainability performance analysis, and research—all supported by a modern customer experience.

The synergies from this close association apply equally to the management of ESG data. With near-infinite storage and huge compute scalability, sustainability-focused research is enriched by the team-up's AI and machine learning analytical power and seamless regulatory reporting capabilities. With those insights gained, portfolios can be monitored and constructed, and risk-mitigation strategies built and updated. All the while, the nuts-and-bolts business of underwriting debt and loan issuance can be further turbocharged with enterprise-wide data sharing and collaboration.



KEY STEPS TO BUILDING AN EFFICIENT ENTERPRISE ESG DATA STRATEGY

1 Build a single, fully managed data platform that enables data collaboration

Say goodbye to data silos created by legacy on-premises applications with Snowflake's Financial Services Data Cloud; you can integrate and analyze data sets that were previously difficult to obtain. Store and access your structured, semi-structured, and unstructured data in one location and gain seamless access to external data with similar scale and speed.

Your golden source version of the truth can be managed safely for use across teams, ensuring that the version one individual uses is identical to another's, while also allowing for certain data to be restricted, where necessary.

By working across all cloud platforms, Snowflake brings the best of all worlds to clients—from high-performance analytics and models to governance and storage solutions.

2 Leverage machine learning applications

Machine learning applications like Microsoft Azure Machine Learning empower data scientists and developers to build, deploy, and manage high-quality models faster and confidently. It accelerates time to value with industry-leading MLOps, open-source interoperability, and integrated tools. This trusted platform is designed for responsible AI applications in machine learning, thereby enabling financial professionals to model and calculate investment portfolio analyses, value at risk from transitions, portfolio decarbonization, and financed emissions so they know where they stand and can make smarter investment decisions.

Additionally, Microsoft Azure Cognitive Services like Form Recognizer can help easily translate physical documents into digital data sets, platforms like Azure Machine Learning can be used by data professionals to connect disparate data sets, and machine learning models can be built, deployed, and refined at scale to understand ESG data and trends.

3 Enable access to ESG data

Snowflake enables banks to quickly explore and access ESG data via the Snowflake Marketplace. With a growing list of industry leading ESG data providers, organizations can not only find new data sets that suit their needs but also monetize their own data by marketing it to other institutions with similar needs. Snowflake Marketplace makes data collaboration possible, an attribute of great value in an era when open-source research and development is helping to advance ESG capabilities.

CONCLUSION

ESG is a driver of revenue opportunity and risk for financial institutions, as well a key requirement of a growing number of jurisdictions around the world. It's a movement underpinned by a burning desire among consumers of financial institutions' managed products to do more for the world than simply accrue value.

For that reason, ESG data is becoming as important to financial institutions as trade and price data, as well as other forms of information, on which they have relied for decades. But getting the ESG piece correct will all come down to accurate, easily accessible, and well-managed data.

The problems that banks and other financial institutions have in getting it right, however, center on technological capacity and capability. Many are running infrastructure or operating on data-management architectures that are unable to accommodate the huge volume of ESG information that their new obligations require.

For that reason, it has never been more important to migrate those processes to modern cloud structures. With near-infinite storage and compute power, managed cloud services can provide the technology at scale needed to tame and manage banks' growing ESG data operations. That includes hosting artificial intelligence and machine-learning analytical tools that can discern key insights from the wealth of ESG data ingested into banks' systems.

They also provide the distributive functionality to ensure all parts of the bank have access to those insights—and the underlying data—to ensure they are working from the same high-quality, properly managed golden source.





ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 543 of the 2022 Forbes Global 2000 (G2K) as of October 31, 2022, use Snowflake Data Cloud to power their businesses. Learn more at [snowflake.com](https://www.snowflake.com)

ABOUT MICROSOFT

Microsoft (Nasdaq "MSFT" @microsoft) enables digital transformation for the era of an intelligent cloud and an intelligent edge. Its mission is to empower every person and every organization on the planet to achieve more. Microsoft Azure is a cloud computing platform operated by Microsoft that can be used for various purposes from database and application migration to data analytics and machine learning workflows. Azure Machine Learning offers an end-to-end enterprise-grade machine learning platform for data scientists and machine learning engineers to build, train, and operationalize machine learning models at scale, responsibly and securely. Learn more at [Azure Machine Learning](https://azure.microsoft.com/en-us/machine-learning/).

ABOUT INTEL

Founded in 1968, Intel's technology has been at the heart of computing breakthroughs. We are an industry leader, creating world-changing technology that enables global progress and enriches lives. We stand at the brink of several technology inflections—artificial intelligence (AI), 5G network transformation, and the rise of the intelligent edge—that together will shape the future of technology. Silicon and software drive these inflections, and Intel is at the heart of it all. Learn more at [intel.com](https://www.intel.com).



© 2023 Snowflake Inc. All rights reserved. Snowflake, the Snowflake logo, and all other Snowflake product, feature and service names mentioned herein are registered trademarks or trademarks of Snowflake Inc. in the United States and other countries. All other brand names or logos mentioned or used herein are for identification purposes only and may be the trademarks of their respective holder(s). Snowflake may not be associated with, or be sponsored or endorsed by, any such holder(s).