Yellowbrick

WHITE PAPER

# Creating an Enterprise-Wide Data Fabric to Underpin Digital Transformation in Capital Markets

Digital transformation requires firms to set in place a common data fabric that embeds their single view of the truth, and to underpin analytics, reporting and regulatory processes. But many organizations are saddled with legacy data systems that require modernization to meet the demands of their digital ambitions. Data modernization is not just about upgrading legacy platforms one-for-one. Organizations can build toward a coherent data fabric that unlocks opportunities and differentiation. This white paper outlines key considerations and sets out a strategy for Capital Markets organizations to maximize these benefits

## Top 5 takeaways for data modernization in Capital Markets

1. Firms need to put in place a modern data infrastructure if they are to unlock the potential of digital transformation, in terms of analytics, streamlined operations, and improved customer outcomes. Build a data fabric, not just a one-for-one replacement.
2. Cloud is not a panacea. Getting cloud right takes time and needs re-skilling and re-thinking of IT processes.
3. Plan for hybrid: Hybrid is proving to be the new normal with some data needing to reside on-premises for the foreseeable future.
4. Prepare for an explosion in data platform utilization as the business catches on to the potential it is unlocking.
5. Plan for real-time: Real-time presents real opportunities for firms to identify and respond rapidly to new business situations and client requests.

Managing different data platforms on-premises and in the cloud can add complexity that firms simply cannot afford. The Distributed Data Cloud is an emerging pattern that seeks to address these challenges.

## 1. Introduction

Capital Markets firms are recognizing the promise of potential benefits from digital transformation, in terms of rapid and streamlined client and product onboarding, fast and consistent trading and investment analytics, and robust processes for transaction, risk, and regulatory reporting.

The potential benefits of adopting a data-driven approach are numerous. Firms can benefit from advanced trading analytics across all asset classes, helping them achieve best execution for themselves and for clients, take advantage of the electronification of the fixed income and foreign exchange markets, and build and test unique trading models that give them competitive edge. They can draw upon deep data sets on-demand, enabling financial advisors in the field to respond in real time to client queries and inquiries.

They can optimize the quality of regulatory data, confident that transaction reports for the US Consolidated Audit Trail (CAT) and CFTC derivatives, the EU's EMIR and MiFIR, and countless other global requirements meet their obligations, while preparing for the onset of perpetual KYC and other FinCrime requirements.

Standing still is not an option; there are risks involved in failing to modernize. Firms will suffer from an ability to respond to requests from ever more sophisticated clients. They will suffer

operational disruption as their legacy systems struggle to keep up with exploding data volumes. And they risk reputational damage from failing to meet regulatory reporting obligations.

Firms are shifting appropriate core activities from on-premises legacy platforms to cloud and SaaS-delivered services models, giving them the capacity, flexibility, and speed to respond to emerging opportunities. In this way, firms are seeking not only to deal with the five Vs of Big Data – volume, velocity, variety, veracity, and value – but also to re-cast themselves as data-driven enterprises.

With many legacy data platforms, repositories and applications having been designed – and often deployed – more than a decade ago, firms are finding existing data IT infrastructure is not up to the job of supporting the kind of transformation they have in mind. To realize their aspirations, firms are recognizing that they need to put in place a modern, enterprise-wide data fabric that can accommodate new cloud-hosted systems consistently.

They also understand that not everything can shift off-premises: Some data sets are too sensitive or are subject to regulatory restrictions that require them to remain on-premises. As a result, capital markets firms will likely continue to run key systems, data platforms, and repositories on-premises or at their own data centers.

This translates into a requirement for a data fabric that's able to provide seamless access to both on-premises and in-cloud applications and data sets.

## 2. Current Situation: Pain Points

Recent years have witnessed an explosion in the types and volumes of data created, with research house Statista estimating that the total amount of data created and consumed reached 64.2 zettabytes (64.2 trillion gigabytes) in 2020 and is projected to more than triple to 180 zettabytes by 2025. In financial markets, this mushrooming of available data has coincided with unprecedented levels of volatility in response to the global Covid pandemic. The expansion of data volumes, velocities, and volatility is expected to continue as businesses embrace Big Data and investors' expectations follow trends in consumer markets.

By harnessing these vast quantities of data, financial institutions have realized they can draw unique insights into client preferences, market trends, and risk parameters, all of which can allow them to respond rapidly to emerging market opportunities, boost customer satisfaction and mitigate again market, liquidity, and operational risk.

Buy- and sell-side firms see the value in equipping their teams with up-to-the-minute analysis to support investment storytelling and drive sales of new funds and products. This may involve sharing the estimated impact of a proprietary trading model with customers, to illustrate how it will generate alpha or otherwise contribute to the performance of a portfolio or fund. Or it may give financial advisors and salespeople the data needed to immediately respond to prospects' questions about a financial security or other offering.

More widely, many see data infrastructure modernization as a key step toward creating digital experiences for customers, leveraging insights, and transforming the business. From an operational standpoint, modernization is seen as essential to deal with legacy platform obsolescence, performance issues, and the risk of security breaches, through cyber-attacks and data leaks. But there's also an opportunity cost issue, with firms fearing that without investment

in modern systems they will be unable to deliver on the strategies customers are demanding, and risk falling behind peers in terms of competitiveness.

This opportunity to modernize is made all the more compelling by the fact that financial institutions across the board are facing growing pressure to perform. Notwithstanding the windfalls they derived from trading in the high volatility markets of the pandemic, sell-side firms like investment banks and brokerages have struggled in recent years to generate positive equity value from their activities. Many have sought to cut costs by outsourcing, eliminating duplication, or automation of repetitive processes. But these efforts often restrict a firm's options and deprive them of the flexibility they need to exploit emerging opportunities.

Buy-side firms including hedge funds and asset managers, meanwhile, are facing pressures on margins as investors question the true value of their activities and justification of substantial management fees. With limited internal resources, many buy-side firms have struggled to tap into the data and analytics they need to boost investment performance.

As they seek to adopt modern capabilities that match consumers' experiences in their personal lives and help them address the financial pressures they are facing, capital markets firms are finding that legacy applications and platforms aren't fit for purpose in terms of performance and scalability to meet growth in data volumes and complexity. As they explore new opportunities for turning expansive data inventories into opportunities, firms are struggling to overcome the hurdles presented by legacy infrastructure.

Many are experiencing system disruptions, often the result of data quality or integration issues, that create bottlenecks, outages, and other performance and data integrity issues. Even firms that have adopted some form of first-generation data lake or data warehouse are finding the fragmented nature of business operations is hindering progress on the journey to become data-driven organizations. Different data solutions deployed in different geographies, across different operating units, or split between on-premises and cloud environments, often make agile innovation difficult. And firms are finding that investing on tactical fixes to operational and infrastructure inefficiencies don't move things forward in terms of digital transformation. What's more, emerging regulations aimed at addressing operational risk, such as the EU's Digital Operational Resilience Act (DORA) will introduce penalties for firms failing to maintain robust systems.

Lack of a strategic approach to data infrastructure can inhibit financial institutions' ability to move swiftly to react to emerging opportunities in terms of clients, products, and geographies. And while the seemingly universal shift to the cloud offers some hope of modernization, it's not necessarily the panacea many believe it to be. Existing cloud-based data warehouse solutions often run on single cloud operator stacks, creating the prospect of cloud concentration risk and threatening to lock firms into a single operator.

Deployment of cloud-based data warehouse platforms can also mask future costs, with little predictability possible, risking the ire of finance departments desperate to nail down expenses as markets become more competitive. Performance is a constant issue, with existing cloud-based data management platforms presenting only marginal gains in terms of latency and throughput at any given cost point as compared with legacy systems. Skillset shortages – particularly cloud technology expertise – are adding to the challenge.

Finally, even firms with the greatest pretensions of becoming data-driven need to retain some element of their legacy infrastructure on-premises, due to management comfort levels or regulatory imperatives. Firms are questioning whether they can afford to host their regulatory reporting or capital risk calculations off-site, where a cloud outage could disrupt critical regulatory processes.

This acceptance that some data and applications are simply too important to host outside of the firm's infrastructure, means that financial institutions should plan to continue to maintain some critical systems on-premises for some time to come. But firms need to recognize that these on-premises systems also need modernization; legacy platforms will continue to struggle to deal with the five Vs of Big Data, and will ultimately fail to meet the needs of the business going forward.

## 3. Realizing the Promise of a Data-Driven Organization

Despite early reticence among capital markets firms, mostly on security and privacy grounds, many financial institutions have initiated a shift to cloud hosting and/or SaaS delivery of services. This was partly due to the home working forced on market participants by governments' response to the Covid-19 pandemic. The pandemic showed firms that it was possible to operate digitally and demonstrated that all but the most sensitive and performance-intensive activities could be carried out by staff working remotely.

But firms also started to realize that cloud technologies would allow them to take advantage of the explosion in data sources and volumes that has unfolded as populations embraced Big Data in their everyday lives. At the same time, many are finding that cloud hosting opens possibilities for modernizing IT infrastructures and adopting high-performance applications, while flexibly handling bursts and spikes in data volumes.

Deployed carefully, the cloud can also help reduce costs – although cloud itself is not a universal remedy to firms' operational cost issues. While acknowledging the dangers of deploying cloud without appropriate checks on processing rates and costs, practitioners say

they expect to see operational efficiencies in terms of lower overall operating costs from cloud deployment in the medium term – perhaps a three- to four-year timeframe.

For many firms, the modernization of IT infrastructures afforded by migration to cloud platforms presents the opportunity to adopt data warehousing and management systems that can provide consistency of proprietary and external data sets when and wherever required. This consistency of data delivery and availability opens further opportunities for firms to drive analytics for better business outcomes and improve data quality for regulatory reporting, among other benefits.

This approach can answer financial institutions' long-standing calls for a single view of the truth that can be used to drive business applications, market and client analytics, and regulatory obligations. By drawing from the same consistent data set to meet these requirements, firms can start to derive the benefits from operating as a data-driven organization, improving client outcomes and ultimately generating greater returns for their stakeholders.

By establishing a common data architecture across all its operations, a Capital Markets firm can create a single view of the truth that can be used across the spectrum of business activities. This approach ensures consistency of data for driving trading and investment analytics; whether a trading model or strategy that triggers a market execution, or an analytical framework to support the narrative around a client's portfolio or fund product.

It also ensures firms are consistent in regulatory reporting, starting with transaction reporting obligations under CAT, EMIR, CFTC or the Monetary Authority of Singapore (MAS) requirements, through to KYC and Anti-Money Laundering, Countering the Financing of Terrorism (CFT), Russian and Chinese sanctions, and the emerging ESG (environmental, social, governance) standards.

Security and privacy are key elements to protecting this kind of unified data architecture from cyber-attacks and other abuse and ensuring data sets are protected where required. This is particularly true of client information, where breaches can incur substantial regulatory penalties and significantly damage reputations. Firms need to retain control over their data to meet regulatory requirements – particularly those relating to privacy, such as the EU's GDPR and California's privacy laws – and be able to demonstrate control. This idea of control also must extend to permissioning of data sets only to those staff internally or to clients that are authorized to receive them.

Finally, any modern data infrastructure must be capable of delivering data in the right format, with the right granularity and under appropriate timeframes, to be of use to the final consumer, whether a human or a computer application. The requirement varies depending on the task at hand; for some trading-related situations, real or near real-time delivery is needed to keep up with market movements; in other areas, such as liquidity risk calculations, latency won't be so significant.

The key imperatives are consistency, flexibility, and availability, with all three attributes supported by an integrated, multi-cloud environment that seamlessly includes on-premises data and systems

## 4. Obstacles to Establishing a Modern Data Fabric

On paper, this kind of capability has been available for some time. In practice, it's a tall order, as financial services practitioners are finding firms need to overcome a series of obstacles that can interrupt the journey to establishing a data-driven organization.

One consideration, particularly for early cloud adopters, is the danger of concentration risk through reliance on a single cloud operator. Firms that moved to embrace the cloud early on – whether through AWS, Google Cloud, or Microsoft Azure – now find they are exposed to potential downtime should a single platform operator experience technical issues. This was demonstrated at the end of 2021 when AWS experienced a series of service interruptions that dovetailed into the non-availability of services for affected customers.

To mitigate against this situation, firms need to embrace the concept of adopting a multi-cloud approach, while continuing to support those on-premises systems and platforms that won't be making the journey to the cloud. But this approach typically involves building a common services layer across multiple data technology platforms, which can have repercussions in terms of inconsistent capabilities and features, as well as the challenge of finding expertise and skills to build and operate the solution.

As they migrate to the cloud, practitioners face an interoperability challenge as they attempt to co-manage two separate technology stacks and the associated skill sets. At the same time, they struggle to establish a single consistent user experience across multiple cloud operators and on-premises systems. The result is a danger of fragmentation of analytics across multiple incompatible platforms requiring a complex web of integration that is both expensive and nonperformant.

Meanwhile, those pioneering firms that are already attempting to modernize their data warehousing capabilities in a cloud environment are running into performance and cost issues. Practitioners are finding that first-generation cloud data lake technologies are the new legacy platforms, as the explosion of data sets and volumes stretches their capabilities. As a result, these systems will likely fail to meet the increasingly onerous needs of complex operational reporting. Meanwhile, those practitioners who have succeeded in putting in place a target data platform for supporting agile analytics are being hit by unexpected costs as the promise of unlimited compute results in unpredictable and often unsustainable costs.

Finally, the human element can present even the most pioneering data engineering teams with insurmountable obstacles. At the base level is the general lack of skills. Many firms dived into cloud without expert technical staff: ETL developers, analysts, and data engineers experienced in running complex data platforms in cloud environments. The global skillset shortage is making it difficult for firms to find and retain staff with cloud engineering and analytics skills, which can disrupt the best digital transformation plans.

On top of that, many firms are finding the shift to a cloud environment requires a change of mindset throughout the organization. This starts at the top, where practitioners suggest having C-level representation with cloud expertise is a major help when seeking buy-in and funding for data modernization and digital transformation projects.

But it goes further than this. The associated change can be daunting and leaders must consider risk appetite and current skillsets when deciding how much disruption they can afford. As one data practitioner at a Tier 1 sell-side firm puts it: “Cloud is very different. It needs a mindset change. People need to unlearn what they know about data infrastructure if they are to succeed in their journey.”

## 5. Best-Practice Approaches to Data Infrastructure Modernization

Practitioners agree that an important first step in your firm’s data infrastructure modernization journey is the setting of a clear and achievable target state. This should reflect the reasons behind the initiative, which necessarily include improved commercial performance, better client outcomes, and a positive impact on equity value. In real terms, these involve providing the data foundation for exploiting new opportunities, reducing operational costs, and protecting the enterprise against regulatory and operational risk.

Adopting a single infrastructure for data warehousing requires a robust approach to data security. A breach of this kind of setup has the potential to do far more damage than a breach of a single data repository among many fragmented data siloes. Getting security right is a key step in putting executives’ minds at ease.

With security in place, it's important to create a narrative describing the journey that ends invaluable analytics for driving new business, improving customer experiences, and mitigating risk. Getting management buy-in, as well as the support of individual lines of business and the IT organization, is essential if the project is to proceed beyond the starting post. It's worth researching and understanding others' journeys, learning from mistakes, and seeing what worked.

Part of the narrative needs to discuss the benefits of modernization, in terms of improved insights across the board and the creation of a single view of the truth. A modern data fabric can offer infinite compute on-demand, allowing for the kind of data wrangling that simply isn't possible on legacy on-premises solutions. It also provides the kind of agility that supports rapid response to market opportunities, giving lines of business speed to market when first-mover advantage matters. But the narrative also needs to talk about the problems of duplicated processes and inconsistent data, and the governance and security issues that come from a fragmented approach to data in the enterprise.

In terms of implementation, practitioners recommend segregation of data storage from compute power and taking steps to ensure all data touching the new infrastructure is as clean as possible, whether originating from inside or outside of the firm. They suggest ensuring on-premises data warehousing capabilities are fit for purpose before expanding into a multi-cloud environment. They stress the need for simplicity, pointing out that many practitioners tend to over-complicate the design of their solution, leading to stasis.

Finally, it's important to set milestones along the way, and use them to take stock of progress. At each stage, progress should be reviewed and plans revised to reflect reality. These architectural hops can be used to cement buy-in and exploit new learnings for an improved outcome.
By taking these steps, practitioners can hope to put in place a modern data fabric that optimizes the hosting capacity available to it to provide a single, consistent version of the truth that can be used by the institution for its analytical and regulatory reporting needs, while enjoying the performance, cost, security, and reliability

## 6. Introducing Yellowbrick

Yellowbrick is a modern and open elastic Data Warehouse that runs consistently across cloud and on-premises with predictable and controllable costs. For firms across the spectrum of financial services – from banks and insurers, buy and sell-side firms, through to custodians, service providers, and other intermediaries – Yellowbrick is the catalyst for modernizing data warehouse infrastructures, allowing firms to reduce concentration risk in terms of exposure to cloud operators, increase efficiency through predictable spend and faster time to analysis and insight, and enjoy the benefits of digital transformation.

The Yellowbrick Data Warehouse is built to meet the needs of the modern Distributed Data Cloud (DDC). Yellowbrick Data Warehouse is a fabric for enterprise data across cloud platforms that recognizes that different data may have different ownership in the same cloud region, across different regions, or spanning on-premises and cloud. In line with the principles of DDC, Yellowbrick offers a single operating model and base technology across the entire data fabric for simplified data architecture, lowering the cost of operation, and enabling the flexibility to seamlessly shift data and functions to different cloud or on-premises platforms.

Yellowbrick's "Your Data Anywhere" approach supports the financial institution's journey from on-premises deployment to hybrid private and public cloud, enabling the firm to streamline and modernize the applications and IT/data infrastructures supporting Capital Markets activities. Yellowbrick de-risks firms' efforts to migrate to the cloud for appropriate capital markets activities and provides the performance and cost-efficiency needed to handle mission-critical time-sensitive tasks with a high cost of failure.

Whatever their scale of operations, banks, investment banks and brokerages, hedge funds and asset managers, fund administrators, and data vendors – all need to bring legacy data IT estates up to date, to support world-class trading and investment technology. Yellowbrick's innovative architecture means that the solution can achieve comparable performance to legacy solutions on a significantly smaller footprint, resulting in cost savings, energy savings, and sustainability benefits

## 7. Five Ways Yellowbrick Can Help Your Firm Modernize Its Data Infrastructure

1. Through its Your Data Anywhere approach, Yellowbrick helps financial institutions put in place a modern data infrastructure that allows them to unlock the potential of digital transformation, in terms of analytics, streamlined operations and improved customer outcomes. Yellowbrick can form the foundation for a true data fabric that de-risks business growth going forward through improved performance and predictable costs.

2. Recognizing that cloud computing is not a panacea, Yellowbrick's deep experience of cloud technologies and data engineering ensures clients can adopt cloud at a pace that is right for them, allowing them to set in place the necessary re-skilling and re-thinking of IT processes.

3. Hybrid is here to stay. Yellowbrick seamlessly integrates on-premises processes and platforms into the overall cloud or multi-cloud data infrastructure, ensuring continuity of mission-critical or sensitive systems while mitigating cloud concentration risk.

4. Infinitely scalable and commercially transparent, Yellowbrick allows customers to plan for rapid uptake of new services supported by the modernized data platform – with no nasty cost surprises for the CFO.

5. The Yellowbrick Data Warehouse offers unparalleled performance and supports high granularity and low latency data delivery, making your data available whenever it's needed and allowing the business to respond to emerging opportunities.

Book a Demo with us:
Phone 877.492.3282
**info@yellowbrick.com**
**sales@yellowbrick.com**