

Adopting a Modern Data Infrastructure for Trading Analytics

Commissioned by:

Yellowbrick 

Table of Contents

Outline	
Adopting a Modern Data Infrastructure for Trading Analytics	03
<hr/>	
Section 1	
Introduction	04
<hr/>	
Section 2	
The Data Challenge for Trading Firms	06
<hr/>	
Section 3	
The Drive to Modernize	09
<hr/>	
Section 4	
What's Needed	12
<hr/>	
Section 5	
Obstacles to Progress	16
<hr/>	
Section 6	
Deploying Yellowbrick for High-Performance Trading and Investment Analytics	19
<hr/>	
Section 7	
Seven Ways Yellowbrick Can Help Build a Winning Analytics Capability	22
<hr/>	

As they seek an edge in their trading and investment activities, hedge funds and other sophisticated trading firms are looking to harness new sources of insight that can allow them to identify and exploit emerging market and industry opportunities. But many organizations are saddled with a disconnected sprawl of legacy systems that are unable to handle the demands of today's fast and often volatile markets.

The agility of cloud technologies – which can offer rapid temporary access to massive compute resources – offers firms a way of driving the analytics they need to deal with the massive volumes and extraordinary volatility the markets have witnessed during and after the Covid pandemic.

To exploit this opportunity, hedge funds and other trading firms can build toward a more productive data fabric that unlocks emerging opportunities and helps differentiate the firm's offering. This white paper outlines key considerations and sets out a strategy for taking a modern, agile approach to trading and investment analytics.

Top six takeaways for establishing a modern trading and investment analytics framework:

- 1** Trading firms need to put in place a modern data infrastructure to handle the deluge of new and emerging Big Data sets that can be used to generate new insights into securities, markets, and opportunities.
- 2** Having multiple disconnected legacy systems stifles innovation by limiting firms' flexibility to add new products and bring together data sets, resulting in lost opportunities.
- 3** Modern data infrastructures that embrace cloud technologies can offer consistent, flexible, and on-demand access to massive data sets in support of sophisticated analytics.
- 4** Cloud is not a panacea. Getting cloud right takes time and needs re-skilling and re-thinking of IT processes.
- 5** Plan for hybrid: Hybrid is proving to be the new normal with some data and processes needing to reside on-premises or in colocation facilities for the foreseeable future.
- 6** Build a data fabric that can support consistent access throughout the trading and investment analytics lifecycle.

1. Introduction

Introduction

Driven by the ongoing need to identify and exploit new business opportunities in the face of growing market uncertainty, proactive trading firms – including hedge funds, active asset managers, non-bank liquidity providers (NLPs), market-makers, and prime brokers – are drawing on more datasets than ever before. The digitalization of many industries – and the broad acceptance of Big Data into mainstream business activities – are generating vast quantities of data that can be used as inputs into the analytics used to drive firms' trading and investment strategies.

Many market participants today employ teams of data scientists charged with leveraging these new datasets to identify new and unique market opportunities – for their own funds and investments, or on behalf of clients. These teams are responsible for identifying market opportunities, and in many cases for developing trading strategies and algorithmic models to exploit them.

This requires access to consistent and timely data and data histories, as well as the capacity to handle the growing volume and volatility of data experienced post-pandemic. In line with the broader financial services segment's embrace of cloud technologies, many trading firms are finding cloud offers them substantial benefits as they grapple with these challenges, in terms of flexible access to compute power and the freedom to experiment with new trading strategies and techniques using "sandbox" facilities. Many new data sets originate in the cloud, making cloud-based processing more efficient.

But managing different data platforms on-premises, in the cloud, and at existing colocation or proximity hosting facilities, can add complexity that many firms simply cannot afford. As they shift key activities, systems, and data repositories to the cloud, organizations need to adopt a modern data warehousing infrastructure that ensures consistent access to data by trading models and investment analysts as they strive to gain a competitive edge.

2. The Data Challenge for Trading Firms

The Data Challenge for Trading Firms

The datasets required to power financial institutions' execution systems and trading analytics originate from a wide and growing variety of sources, many of which are increasing in terms of both volume and frequency. Real-time market information is mushrooming as liquidity fragments across different trading volumes. The explosion in data in US options markets – and specifically the OPRA data feed, which peaked in December 2021 at an average of 128.5 billion transactions a day – is just one of several data sources forcing firms to upgrade their data systems in order to cope.

Hedge funds and other proactive trading firms seeking to execute on a strategy or algorithm often need historical data to back-test their trading models against various market scenarios. This may include historical tick data from price changes in the marketplace as well as corresponding chronologies of other data that may impact the market in question. This is supplemented by the reference data required to identify securities, venues, issuers, and counterparties.

In designing and running their investment models, trading firms may use data from across a wide range of asset classes to support analysis aimed at understanding market pricing, identifying trends and correlations, and predicting how a company might react to some kind of market event. This approach is also used to compare securities across different asset classes, including equity and credit derivatives for example.

“

In designing and running their investment models, trading firms may use data from across a wide range of asset classes to support analysis aimed at understanding market pricing, identifying trends and correlations, and predicting how a company might react to some kind of market event.”

Hedge funds and active investment managers in particular often require access to a broader range of data to help them identify opportunities to identify potential returns that justify the fees they charge clients. These firms often use large data sets to execute on a trading strategy that may be based on hours, days, or even weeks or months of analysis. The primary objective is to boost the performance of a given trading strategy in pursuit of alpha generation and/or best execution.

As well as requiring access to pricing, news, and analysis relating to the marketplace itself, proactive trading firms increasingly draw on so-called alternative data, which can give them insights into the workings of the individual companies, sectors, and industries they invest in. With the digitalization of industries of all kinds, there has been an explosion in data creation, resulting in an upsurge in data sets that may impact the performance of any given investment.

This can make these data sets essential for hedge funds and other trading firms, proactively mining the next market opportunity. Often described as the “exhaust” of corporates’ mainstream business activities, alternative data is often unstructured, making it difficult to ingest, manage, and apply to trading strategies.

Meanwhile, to support buy-side trading and execution strategies, prime brokers and other sell-side firms often use trading algorithms that are driven by a range of signals and trends, including liquidity levels, changes to market structure, and order book or trading venue activity. Best execution has emerged as a key regulatory driver for sell-side firms in this aspect of their business ever since its introduction under MiFID II in the EU and Dodd-Frank in the US. In either instance, best execution obligates firms to report their execution performance against promises made in the publicly available best execution policies. Delivering the analytics resources to prove compliance is challenging.

Finally, significant market volumes particularly in the US are being driven by NLPs or market-makers – many of them former high-frequency trading (HFTs) – which are tapping into massive retail volumes from electronic brokers like Robinhood, Charles Schwab, and TD Securities. This approach requires processing of huge volumes of order data, and rapid execution turnaround to generate tiny incremental gains.

Whatever their role in the trading lifecycle, firms are struggling to deal with the volume and complexity of the data they need, whether to identify trading opportunities or to execute on those opportunities. And many firms continue to be constrained by legacy data and analytics systems and infrastructure that are unable to keep up with market demands.

3. The Drive to Modernize

The Drive to Modernize

Lack of investment in systems infrastructure has led to the continued use of legacy platforms that limit firms' ability to respond to emerging market opportunities and other requirements, leading to lost business, poorer client outcomes, and heightened risk. While firms often see a clearer need for updating core trading systems, the case for upgrading their data and analytics capabilities are not always so clear, or are often further down the priority list, and this dampens the agility firms need to respond to challenges and opportunities.

This lack of agility, for example, can make it difficult to onboard new clients, a process that can involve many manual processes, such as Know Your Client (KYC) due diligence, that straddle multiple applications and systems. Firms are also beginning to grapple with the concept of Perpetual KYC, a state of constant due diligence of clients' and counterparties' business activities, supply chains, and funding sources, aimed at addressing FinCrime and other regulatory considerations, all of which require access to large sets of consistent and high-quality data. In particular, the inflexibility of legacy data infrastructures also hinders the addition of new financial products, which often involves extensive documentation and associated metadata before they can be added to systems, traded, or offered to clients. Delay can result in lost opportunities if firms are unable to meet a client request for a new fund or other financial instruments.

The presence of multiple, disconnected legacy data platforms also makes it difficult for firms to add new data sets that would allow them to exploit emerging market opportunities. Proactive trading firms, like hedge funds and prime brokers, are constantly seeking new opportunities and need to leverage emerging data sets to identify and exploit them. These data sets can be used to develop trading models aimed at achieving best execution for the strategy or investment plan in question. By developing unique trading strategies, these firms' data scientists can help empower the trading desk, helping them make better decisions, find new sectors, segments, and markets to invest in or trade, and where to direct orders for best execution.



These data teams are constantly seeking to innovate by fine-tuning and optimizing trading models. As they do this, dealing with changes to market structure and new data types is a major challenge, complicated by ongoing digital transformation initiatives, which are transferring big processing functions to the cloud.

To ease the transition, operators like Microsoft Azure, Google Cloud, and Amazon Web Services (AWS) are offering tools that can help onboard unstructured data and apply it to quantitative research teams' development and testing work. However, firms should be wary of tying themselves into cloud-only platforms or closed ecosystems that limit innovation. Without the right data infrastructure in place, firms may not be able to optimize their cloud usage in line with the operators' commercial models or avoid unexpected costs associated with these endeavors.

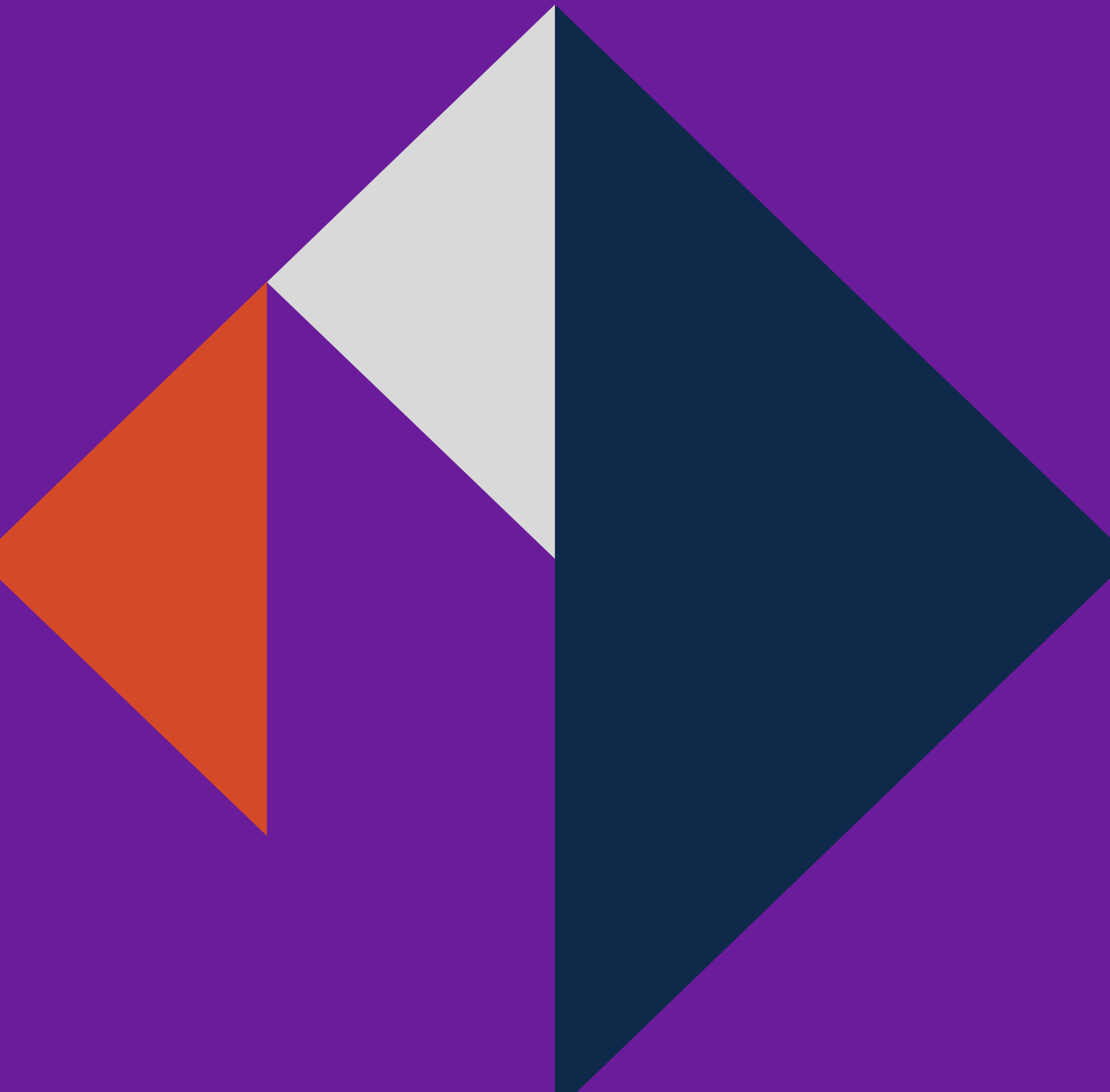
Investment Analytics

The need for flexibility in data infrastructures extends beyond the trading environment. More widely across the financial enterprise, investment management firms and fund managers are seeking to improve their understanding of client behaviors and preferences, and need effective analytics to get the insights they require.

According to a recent A-Team Group poll of 60 buy-side practitioners, firms understand the value of client behavioral analysis to optimize business outcomes, with two-thirds of respondents identifying potential business improvements of between 10% and 20% as a result of better understanding customers and serving their needs more effectively. But many firms have yet to fully exploit their customer data to improve cross- and up-selling potential, with only 50% of respondents claiming they are deriving measurable benefits from their client data analysis efforts. And of these, one-third had systems in place but had yet to benefit from processes for methodically collecting and analyzing client data.

As a result of these factors, trading and investment firms are facing growing pressures to modernize their legacy data infrastructures. Many are now putting in place the underlying data foundation upon which to build a flexible and highly performant analytics framework that allows them to take advantage of new business opportunities.

4. What's Needed



The Data Challenge for Trading Firms

Proactive trading firms often use a combination of on-premises, colocation, and increasingly cloud-based applications and data sets to identify market opportunities and execute on them. Much of the pre-trade analysis historically has been conducted on-premises with buy/sell signals sent to colocated trading applications for final execution. Cloud is being adopted by market participants to leverage on-demand access to massive data sets for designing, testing, and running trading strategies, sometimes in designated “sandbox” facilities created for the express purpose of experimentation.

Hedge funds, NLPs, and other proactive trading firms need the ability to deal with siloed data sources and incompatible data formats and orchestrate between the various staging points of the pre-trade/trade lifecycle. They need access to both real-time and historical data to support real-time data processes, market access, and back-testing of the trading strategies. This requires straightforward and cost-effective onboarding of new datasets to support rapid testing of new trading ideas and models. When necessary, these firms need rapid scaling of systems to meet new opportunities, typically by accessing cloud-based data and integrating it with on-premises or colocated systems in a flexible and agile way.

Buy-side use of analytics, particularly among those with higher turnover in portfolios and shorter investment windows, is characterized by high levels of customization and experimentation, making data quality and consistency a significant consideration when designing the underlying data fabric and related data infrastructure.

“

Cloud is being adopted by market participants to leverage on-demand access to massive data sets for designing, testing, and running trading strategies.”

Meanwhile, fund managers seeking to deliver the kinds of customer experiences that consumers are growing used to – and also be able to cross-sell in an increasingly complex regulatory environment – need to integrate siloes of data in a way that provides not just a genuinely holistic view of a customer’s engagement with the firm but delivers actionable insights into customer preferences.

There are other considerations to take into account. Hedge funds, investment managers, and other buy-side firms need systems that support compliance with appropriate regulatory jurisdictions. While the buy-side enjoys less regulation than the sell-side and advisory side of the business, increased regulatory scrutiny of marketing messaging on the fund side, as well as the Investment Firms Prudential Regime (IFPR) capital adequacy regulations in the UK – and its EU equivalents, the Investment Firms Directive/Regulation (IFD/IFR) – are posing data challenges to buy-side firms. For US hedge funds, meanwhile, the SEC's Form PF has proved challenging. Sanctions against Chinese military companies and people and entities associated with Russian President Vladimir Putin also add to the complexity.

Data privacy is a growing issue, particularly in light of the EU's GDPR and California's privacy laws, both of which raise issues around data sovereignty and hosting locations. These are particularly pertinent for firms that turned to the cloud to support staff and clients during the Covid-19 pandemic, which became a catalyst for more rapid adoption of cloud technologies in the previously reticent capital markets segment by allaying security, performance, and resilience fears.

Embracing the cloud

For firms seeking to improve their analytics capabilities – whether for trading or portfolio analysis – cloud hosting opens up possibilities for modernizing IT infrastructures and adopting high-performance applications. At the same time, it is capable of flexibly handling bursts and spikes in data volumes. This allows them to deal with increased market volatility and try new ideas, test new strategies, build new products, and forecast more accurately. But it also allows firms to devote IT resources to specific data-intensive analytical tasks without having to commit that resource to ongoing overhead. This is further supported by the advent of low-touch managed analytics services in the cloud, which reduce the need to manage infrastructure.

With margins under pressure, though, institutional firms need predictability of cost. This is often lacking in first-generation or single-operator cloud implementations. While firms expect to see lower overall operating costs from their cloud migrations in the medium term – with practitioners citing a three- to four-year timeframe – cloud is not the direct cost panacea many have suggested it is.

Cost management is an important differentiator, and those firms getting a grip on the cost of their analytics programs may benefit from the ability to pass on those cost savings in terms of lower fees, or to devote resources to develop products and capabilities that would otherwise not have been feasible.

The evergreen nature of cloud provider infrastructure makes it easier to procure and experiment with the latest technological advances such as processors, memory, and high-speed storage without up-front capital expenditure and months of procurement and planning. Attractive cloud provider SLAs for managed services provide some assurance that service failures will be rare and data will not be lost. Firms should be aware that these SLAs are for individual components and not for the end-to-end service, resulting in the need to carefully consider business continuity strategies.

In short, trading firms of all kinds need consistent availability of proprietary and external data sets wherever required and permissioned for. They need to retain control over their data to meet regulatory requirements and be able to demonstrate control. They need the appropriate granularity and latency of delivery for enterprise, real time and near real-time capabilities, to optimize market execution and to support up-to-date market analytics. And they need a secure infrastructure capable of withstanding the barrage of cyber-attacks and other potential breaches that can undermine reputations and trigger disruptive mitigation actions.

While the cloud provides exciting opportunities, analytics architectures need to actively incorporate datasets that are necessarily hosted outside of the public cloud, in private colocation facilities, and in firms' own data centers.

5. Obstacles to Progress

Obstacles to Progress

While the direction of the journey toward a high-performance data warehousing and analytics framework is clear for many firms, the path toward getting there is laden with potential tripwires.

Many firms – particularly small to medium-sized boutique trading firms – are saddled with the ongoing maintenance cost of legacy systems, which can monopolize IT budgets and leave little scope for innovation. Many suffer from the barrier of high operating costs as a result of integration work, data cleansing, and data orchestration workarounds to their existing systems. And it's often the case that senior management is still reticent about investing in technology and infrastructure unless it can be demonstrated to be required by regulatory considerations.

These firms lack the expertise to put in place an advanced analytics database framework that's capable of handling huge volumes and a variety of often-unstructured data. Firms can be daunted by the prospect of adopting emerging cloud technologies. The ongoing skills shortage in cloud and distributed technologies means that even larger firms lack the expertise to deploy a sophisticated hybrid on-premises and multi-cloud implementation. For some, this kind of project is simply too difficult.

“

Many firms – particularly small to medium-sized boutique trading firms – are saddled with the ongoing maintenance cost of legacy systems, which can monopolize IT budgets and leave little scope for innovation.”

For those who have already embarked on migrating analytical applications to a cloud environment, there are other potential pitfalls. Early adopters of single operator cloud solutions face the danger of concentration risk through reliance on a single point of failure. The December 2021 outage at Amazon Web Services (AWS) is a case in point that saw service disruption for many financial services firms that based key functions there.



More generally, firms have suffered performance issues from adopting first-generation cloud data lake technologies, which likely will fail to meet the needs of complex operational reporting, not to mention the ability to support truly integrated data for consistency of analytics.

Finally, for firms looking to deploy multi-cloud data infrastructures – or those involving on-premises systems – there is the danger of fragmentation of analytics across multiple incompatible platforms. This can require a complex web of integration often involving a common services layer across multiple data technology platforms with inconsistent capabilities and features, not to mention the effort, skills and resources need to deploy and maintain them. Add to that the challenge of establishing a single consistent user experience across multiple cloud operators and on-premises systems, and the task looks daunting indeed.

These challenges can be avoided by using a high-performance data warehouse solution that's built for multi-cloud environments. Firms should avoid the temptation to lift and shift legacy systems to cloud infrastructure, an approach that delivers limited benefits, and instead invest in managed offerings built natively for the cloud.

6. Deploying Yellowbrick for High-Performance Trading and Investment Analytics



Deploying Yellowbrick for High-Performance Trading and Investment Analytics

Yellowbrick Data Warehouse is a modern, elastic, massively parallel (MPP) SQL relational database delivering high performance on massive volumes of live data across cloud and on-premises for better informed business decisions with predictable and controlled costs. For hedge funds and other trading firms, Yellowbrick underpins high-performance analytics in support of the trading and investment process, 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.

Yellowbrick delivers on the vision of a Distributed Data Cloud (DDC), which 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. DDC promotes a single operating model and base technology across the entire data fabric for a 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 their trading and investment analytics. By alleviating the need to build out common service layers that traverse multiple clouds, Yellowbrick de-risks firms' efforts to migrate to the cloud for appropriate capital markets activities and provides the performance and cost efficiency firms need to handle mission-critical time-sensitive tasks with a high cost of failure.

The Distributed Data Cloud powered by YellowBrick Data Warehouse



Whatever their scale of operations, hedge funds and other trading firms all need to bring their 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. Seven Ways Yellowbrick Can Help Build a Winning Analytics Capability

- 1** Yellowbrick Data Warehouse provides a large-scale, high performance, massively parallel (MPP), SQL database with industry-leading performance and enterprise-grade security enabling firms to consolidate multiple legacy data platforms, simplifying data architecture, and reducing run costs.
- 2** Using open standards and the familiar SQL interface means analytics teams do not need new skills and ensures Yellowbrick plays nicely with the broader data eco-system, assuring future open innovations can be incorporated.
- 3** As a modern, truly hybrid, multi-cloud solution, built using cloud-native technologies, firms can take advantage of Yellowbrick's same, low-touch, managed platform, wherever it is deployed: in the public cloud, in your data center, or in colocation facilities.
- 4** With Yellowbrick, firms can adopt cloud analytics at their own pace, without complexity, reducing the need to acquire new cloud infrastructure skills, and easily incorporating data that remains on-premises. The ability to run anywhere alleviates concerns over single-cloud solutions and cloud concentration risk.
- 5** Re-imagined from the ground up and engineered for efficiency, Yellowbrick extracts more performance from the underlying infrastructure than rivals, delivering faster results in a smaller footprint, and resulting in lower run costs, lower energy costs, and unbeatable price-performance economics.
- 6** Yellowbrick is built to cope with the full range of firms' analytics needs, including demanding complex mixed workloads consisting of batch jobs, real-time data, 24x7 operational applications, and complex, high-volume ad hoc queries. Analysts and quants can be let loose across any data without fear of breaking other processes or long run times.
- 7** Yellowbrick's pricing structure and advanced workload management deliver lower and predictable costs through fixed subscription models while retaining the ability to experiment or burst demand with on-demand consumption pricing. Yellowbrick customers avoid runaway costs that trouble many other cloud platforms.

Yellowbrick

Yellowbrick is a modern Data Warehouse, built on an open, cloud-native architecture that enables large-scale enterprises to eliminate complexity, reduce risk, and predict and control costs by running all their data anywhere, across all cloud and on-premises instances within a single management plane. Yellowbrick powers the most resource-efficient, high-performing, large-scale data warehouses. Enabling complex queries on live data at petabyte scale, while supporting high concurrency with fast, interactive query response to customers' most challenging business questions.



 | www.yellowbrick.com  | [@YellowbrickData](https://twitter.com/YellowbrickData)  | [Yellowbrick Data](https://www.linkedin.com/company/yellowbrick-data)

A-TEAMGROUP

For Marketing: www.a-teamgroup.com

For Content: www.a-teaminsight.com