Yellowbrick

Modernize your data warehouse for near-real time insights across distributed clouds

The problem with data warehouses

Designed to enable insights from enterprise data across multiple lines of business in conjunction with business intelligence, data science, and data visualization tools, data warehouses can create value from data at a scale that would otherwise be impossible.

But all too often, legacy data warehouse platforms like Teradata, IBM Netezza, Oracle, and Microsoft SQL Server, which are based on decades-old architectures, are challenged to keep up with modern workloads. (See Figure 1.) They're too inflexible, too expensive to expand and scale, and require too many technical resources to manage and optimize. Facing huge data volumes, growing numbers of users, increasingly complex queries, and more real-time data, their owners are often hesitant to extend their investments in what is likely a losing cause.

What these organizations need is a modern platform that not only supports today's requirements but provides a path to the future, with flexible deployment options and expand-as-you-grow architecture. And although cloud-only options are often perceived to be the obvious choice for reaching those goals, the truth is usually more complicated.

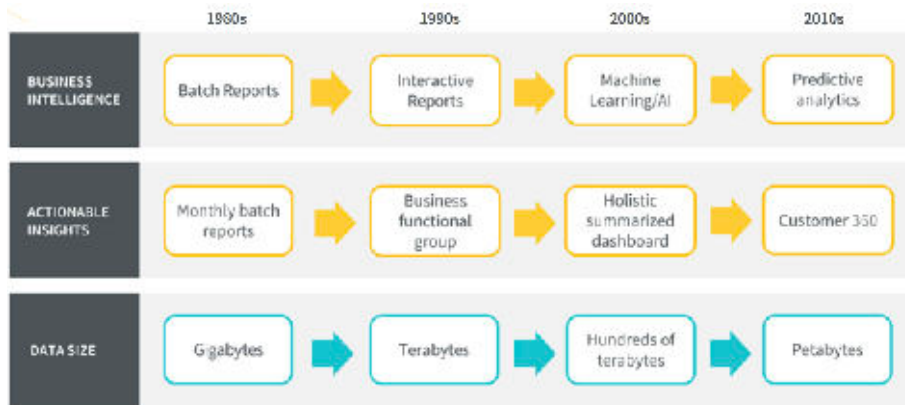**A modern data strategy supports today's needs, while providing a path to the future**

Figure 1. Data warehouse requirements have evolved beyond the capabilities of legacy architectures that were designed decades ago.

What is modernization?

Data warehouse modernization requirements vary across organizations and industries, but common themes include:

- **Excellent price/performance economics**
  Legacy data warehouse vendors have struggled to refresh their platforms to produce good price/performance as data volumes grow and concurrent users increase in number. Cloud-only options have revolutionized the user experience, but the combination of plain-vanilla performance and consumption pricing always leads to exorbitant and unpredictable bills.

- **Linear scalability**
  The one thing that's constant with data warehouse deployments is that the volume of data will continue grow, as will the number of users and types of queries. Therefore, when evaluating a more modern platform, it's critical to understand how easy it is to add more data or support more users without adding more cost.

- **Native support for real-time data**
  The ability to ingest and query real-time data alongside at-rest data is now a critical requirement for use cases like fraud detection, Customer 360, and IoT analytics.

- **All workloads**
  Strong support for ad hoc, batch, real-time, interactive, and mixed workloads—not just recurring cookie-cutter ones—is important for meeting emerging business needs.

- **De-risked cloud and IoT journeys**
  A modern platform should support a flexible range of deployment options, so that organizations can de-risk migrations to the cloud (e.g., to respect security and data gravity concerns). Furthermore, your data warehouse architecture needs to support future use cases like IoT analytics under the same policies and control plane.

- **Predictable pricing**
  While most enterprises now avoid CAPEX, their need for accurate forecasting is incompatible with the hidden and complex costs typical of cloud-only options. A predictable pricing model that solves for both needs is important.

The ability to ingest and query real-time data alongside at-rest data is now a critical requirement

Re-imagining data warehousing

Yellowbrick Data Warehouse is the first MPP analytics database with adaptive architecture designed to take advantage of whatever physical (e.g., optimized instances) or virtualized (e.g., Kubernetes stacks) infrastructure it runs on. We've added on top of that a modern, standards-based database interface that's familiar to users (PostgreSQL) for ecosystem compatibility, as well as support for open standards like Apache Spark, Apache Kafka, Python, and R.

The architecture of Yellowbrick was conceived with a goal of optimizing price/performance: New SQL analytic use cases are evolving all the time and more and more concurrent users are making more use of ad-hoc analytics. More users and larger data sets require more performance per dollar spent, and Yellowbrick architecture leapfrogs the industry in this respect. It's not uncommon for users to see their workloads run tens or hundreds of times faster at a fraction of the cost of cloud-only and legacy data warehouses.

The result is a modern, easy-to-deploy, and easy-to-use solution that blows the doors off rivals in price/performance economics, and that can be deployed anywhere across distributed clouds (private, public, and edge networks) along with simple, unified management. That makes Yellowbrick your best choice for any data warehouse modernization, BI acceleration, or data lake augmentation project.

Read our Technical White Paper for complete details.

The result is a modern, quickly provisioned, and easy-to-use solution that blows the doors off rivals in price/ performance economics

As a result, Yellowbrick enables:

- Best-in-class price/performance economics, with 100x faster time-to-insight versus alternatives at a fraction of their cost

- Sub-second ANSI SQL queries across billons of rows of real-time and at-rest data—increasing the richness (for example, spanning multiple months of historical data) and rate of insights

- Parallel queries by hundreds or thousands of users

- Deployment in any environment (private clouds, public clouds, at the edge, on physical or virtualized infrastructure), along with simple, unified management

- Rapid ingestion of data in bulk (up to 10TB/hour), as a real-time stream from Kafka (millions of rows/sec) for data lake integration, or incrementally for CDC (continuous data capture) from OLTP systems, with data immediately query-able and actionable

- Operational simplicity that makes manual indexes, tuning, partitioning, and reclaiming storage space unnecessary

- Compatibility with enterprise BI and data motion ecosystems, as well as open source tools like Kafka, Spark, R, and Python

- Fast and easy migrations from any platform

Modernization and distributed clouds

The Gartner *Top Strategic Technical Trends for 2021* report suggests that distributed clouds will emerge to address the explosion of data growth, particularly at the network edge. Distributed clouds are characterized by the deployment of cloud software and hardware stacks outside of the public cloud provider's data center to provide a mesh of interconnected cloud resources to form a best-of-breed logical cloud. These stacks enable the ability to run applications developed for the public cloud in a company's own data center and in other locations, such as in multi-access edge computing centers connected to 5G cell tower groups, or on the factory floor in support of IoT applications in manufacturing.

The distributed cloud model offers reduced latency, increased data sovereignty, higher security, addresses data gravity requirements, and provides uniformity in terms of infrastructure and services. Distributed clouds bring analytic applications to remote locations (such as factories or even shipping) that may be only intermittently connected to the internet, or not connected at all.

Distributed clouds also enable analytic applications in remote locations that may be sporadically connected to the internet, if at all

To take full advantage of distributed clouds, we must rethink our approach to how data is managed in such a homogeneous, geographically separated and logically interconnected environment. Distributed clouds will provide the common foundational hardware and software infrastructure on which new applications that are federated across clouds will be deployed. In order to support these applications, the underlying data management services higher up in the stack must also be federated.

Yellowbrick has embraced Kubernetes as core, cloud-native architecture to help customers deploy, manage, and orchestrate data warehouse workloads across this architecture. This approach complements our Andromeda optimized instances for use cases that demand the ultimate in price/performance inside private data centers.

Customer experiences

Here are some examples of how customers are using Yellowbrick to modernize their data warehouses with proven success.

BMW Financial Services previously used a patchwork of SQL Server instances to support loan approval and Customer 360, but the system couldn't keep up with the workload. With Yellowbrick, BMW can consolidate its infrastructure for much simpler operations and expects to see an average 182x improvement in query performance. The company now also has use cases that were impossible before, like sophisticated risk modeling.

Catalina Marketing is the market leader in shopper intelligence and targeted in-store and digital media. The company delivers $6.1 billion in consumer value annually, combining the richest buyer history database in the world with its own deep analytics and insights to help retailers, CPG brands, and agencies optimize every stage of media planning, execution, and measurement. The company's legacy IBM Netezza system lacked the capacity to support growing workloads, with analysts having to wait 20 minutes or more for their queries to run—leaving only a small window

Yellowbrick has embraced Kubernetes as core, cloud-native architecture to bring data warehouse workloads to distributed clouds

of time (25% of the day) for complex analysis. Now, Yellowbrick delivers up to 182x better performance than Netezza, and queries that used to take up to 30 minutes—if they weren't killed first—now complete in seconds.

TEOCO is a leading provider of analytics, assurance, and optimization solutions to the telecom industry. In late 2017, TEOCO evaluated Yellowbrick Data Warehouse and found it could handle the 30 -40 billion records TEOCO loads each day, along with performance improvements of up to 100x for some queries. With Yellowbrick, TEOCO can find insights that were previously impossible and expects to save $5 million in data center costs alone over the next several years.

Symphony RetailAI helps retailers and CPG manufacturers drive profitable revenue growth through AI-enabled decision-making. Its customers include 15 of the world's 25 largest grocery retailers, thousands of retail brands. Yellowbrick helped the company consolidate disparate Amazon Redshift, 1010 Data, and SQL Server data warehouses into a single solution that offers unparalleled price/performance at massive scale— delivering reports to customers twice as fast on 10x more data.

Conclusion

If your data warehouse isn't keeping up with your business needs, it's time to modernize with Yellowbrick. Its unique combination of industry-leading price/performance economics and flexible, future-proof architecture will make you re-imagine what a data warehouse can do.

Try our free 7-day test drive: yellowbrick.com/test-drive

660 W. Dana Street Mountain View, CA 94041 I USA I 1.650.687.0896