Yellowbrick Data
Product Overview

Yellowbrick Data Overview

Yellowbrick was founded in 2014 by experts in database and flash memory technologies with a mission to solve the challenges of multi-user analytics for modern enterprise structured data sets. Specifically, the Yellowbrick goal was to retain the feature sets, reliability and predictability of mature enterprise data warehouses while offering economics and price points suitable for today's large data volumes.

The company achieved its purpose by building a next-generation data warehouse, the Yellowbrick Data Warehouse, delivered it in a small compact form factor that is quick to deploy and easy to expand. This fit with the goal to provide end-to-end analytics across the hybrid cloud from the data center to the cloud to the edge.

Problems with traditional data warehouses

The data warehouses you use today are probably overloaded. In particular, they are expensive, scarce resources, originally built for smaller data sets, with increasing numbers of concurrent users as business demand for more analytics skyrockets. They probably were not designed for 'big data,' and the continued expansion of ad-hoc use cases jeopardize business critical reporting.

Traditional data warehouses are likely too slow to efficiently run ad-hoc queries against raw fact data, meaning that more and more cycles are spent building cubes, slowing and complicating data loading, inhibiting ad-hoc data science. Low capacity points mean older data is deleted or moved into the data lake, making it hard to gain business value from seasonality, monetize previous events, or back-test new data science models.

The majority of today's data warehouses cannot ingest real-time data efficiently. They require data to be bulk loaded or micro-batched, further complicating ETL and prohibiting analytics of real-time event or device data. They may also be based on technology that prohibits effectively running mixed workloads – concurrent queries on continually changing data or complicated mixes of reads, updates, loads, short and long queries.

Finally, the support costs are likely going through the ceiling as the systems age, or you run an increasing number of cloud instances where bills from your cloud provider are not yielding the economic benefits once envisioned.

For enterprises facing these data warehousing challenges, the Yellowbrick Data Warehouse merits in depth consideration.

Yellowbrick Data: Solving the hard problems first

There has been an explosion of new databases, SQL engines, and open source technologies in the last several years. All these products fail to solve the really difficult problems that separate the amateurs from the experts in data warehousing. Yellowbrick began by focusing on the most intense challenges first. This includes

- **Continuous availability**: Built-in redundancy ensures potential hardware failures or software issues do not impact system uptime. Yellowbrick maintains no physical single point of failure and three levels of redundancy for your data. Support for expanding capacity on-the-fly means you can grow compute and storage in-place without fork lift upgrades. This ensures you can trust Yellowbrick for business-critical workloads.

- **Ad-hoc SQL**: Yellowbrick supports all of the rich SQL constructs you expect including complex multi-thousand-line queries, as well as running against large fact tables where the questions asked are not always known in advance. Data scientists and interactive business explorers get answers to new questions quickly.

- **Correct answers on any schema**: The real world is not full of highly numeric star or snowflake schemas. Yellowbrick understands this, and our sophisticated query planner gives correct results quickly on any schema. Statistics are automatically kept up-to-date using incremental big data algorithms, so you do not have to worry about them becoming stale. No vacuuming or grooming is required to maintain the system.

- **Scale from terabytes to petabytes**: Yellowbrick systems support data warehouses from single digit terabytes through to multiple petabytes. A 500TB data warehouse still fits in 12 inches, or 30cm, of rack space and installs and provisions in under an hour. With Yellowbrick, you can deploy in locations beyond a full data center, providing more environmental flexibility.

- **Mixed workload support**: Your data warehouse might be running perpetually heavy ELT workloads to wrangle data. You might have batch queries and interactive queries running together. Of course, you will have new data loads running. With Yellowbrick, you can handle all of this at the same time, using Workload Management functionality to guarantee business-critical tasks are not impacted and run in a timely fashion.

- **Concurrent user support**: You can run all of the workloads mentioned previously while supporting thousands of concurrent users.

These are the challenging problems in data warehousing that go beyond just building a SQL engine or running queries on a big data platform. Yellowbrick has solved all of these problems to give you confidence that our data warehouse will support all your business needs and stay running to do so.

Delivering industry-leading price/performance

Having built the underpinnings of a reliable and predictable enterprise data warehouse, Yellowbrick focused on building the world's fastest data storage and SQL engines designed natively for high bandwidth flash memory, high speed RDMA (remote direct memory access) networking, and the latest large multi-core CPUs. In doing so, we designed for both low system price **and** high query performance, to produce a system that wins on price, offers best-in-industry performance for large data sets, and is unmatched on price/performance.

Based on customer workloads and standard TPC-DS benchmarks against 100TB data sets, the Yellowbrick database delivers winning results. Compared to systems with comparable capacity, CPU counts and (for cloud-based databases) three-year reserved instances:

Yellowbrick results compared to	Batch query price/performance	Interactive query price/performance
Traditional Data Warehouses	5x	40x
Cloud-only Data Warehouses	4x	10x

A data warehouse offering superior price/performance means you are able to deploy a new data warehouse with lower acquisition costs, lower maintenance costs, larger capacity, and superior performance compared to traditional solutions. Larger capacity points mean you can retain more historical data. Support for mixed workloads with higher performance means more users and use cases consolidated onto one system. All of this makes business users and data scientists happier with highly responsive, interactive applications.

In some cases, Yellowbrick customers have been able to procure their system for less than the cost of support from legacy enterprise vendors, or less than one year's usage fees in the cloud. Installation typically takes less than two hours, and the systems are small enough to not require a dedicated data center installation.

Moving ahead with Yellowbrick

Yellowbrick solves the hard problems in data warehousing – high availability, complex mixed-workload support, support for ad-hoc SQL, correct answers on any schema, scalability from terabytes to petabytes, and on-the-fly capacity expansion while supporting large numbers of concurrent users. From there, Yellowbrick offers the most compelling economics in the industry to deliver the lowest acquisition costs and best performance for your business users.

ABOUT YELLOWBRICK DATA

The Yellowbrick Data Warehouse delivers unprecedented analytic speed, simplicity, and savings. Unlike traditional solutions that force organizations to compromise on performance, scale, or ease of management, Yellowbrick Data's unique architecture enables users to quickly analyze all relevant data including real-time and deep historical data. Led by industry experts and backed by top-tier funding, Yellowbrick Data is redefining how data can help you drive change in your organization.

Want to learn more about how Yellowbrick Data can meet your most demanding data warehousing needs? Contact us to schedule a meeting or demo: **info@yellowbrickdata.com**