Yellowbrick

CASE STUDY

Yellowbrick Dramatically Improves Performance of a Critical Fraud-Detection Web Application

> Following extensive testing of data warehouses and Hadoop-based solutions, we found that Yellowbrick Data provided superior performance.
>
> **Matthias Baumhof**
> CTO, ThreatMetrix

One of the key business applications of ThreatMetrix (a LexisNexis Risk Solution), a global digital fraud and identity-authentication service, is an online portal accessed by thousands of users around the world. The portal serves over 5,000 global brands, helping them verify more than 20 billion financial transactions each year.

A NEED FOR DRAMATICALLY BETTER PERFORMANCE

Via the portal, ThreatMetrix customers query a 300TB multi-tenant database over 25,000 times per day, with up to 1TB of new data ingested daily in real-time. Often, several hundred users might be simultaneously writing queries. And many queries are also complex, accessing over six months of stored data spread across millions of records in a multi-billion-row table with over 1,000 columns.

To handle this heavy mixed workload, a legacy Apache Impala system required a complex table-partitioning method to manage contention between the complex queries and continual data inserts. Admins juggled over 1,000 partitions each day, trying to keep the database's performance at an acceptable level as required by SLAs. Even with this intricate partition management, the query engine couldn't keep up, and writes would queue up.

Performance: After replacing Impala with Yellowbrick, portal users noticed performance improvements immediately: Complex queries that used to take 30 seconds now execute in under 2 seconds, and short queries that executed in several seconds are now measured in milliseconds—all while Yellowbrick constantly ingests terabytes of new data.

Management: Yellowbrick also dramatically simplified data management. ThreatMetrix uses Yellowbrick pool management to easily reallocate resources and respond to spikes or unusual usage patterns, allowing it to optimally allocate resources to queries and database operations. Because Yellowbrick also does away with the need for complex partition management, a significant amount of admin time is saved.

CHALLENGES

- Slow queries, especially complex ones
- Frequent functional downtime
- High data management overhead
- Storage scalability issues

SOLUTION

Replacement of legacy analytic database with Yellowbrick on-premises data warehouse

BENEFIT

- 4x faster query execution, even during ingest
- Zero outages, all SLAs met
- Easy resource allocation, performance management, and system monitoring
- One-third less hardware, reduced energy and resource usage

High Availability and Resiliency: The previous ThreatMetrix portal experienced significant outages, often forcing the company to restart the database and taking its customers offline. ThreatMetrix deployed Yellowbrick in each of its global regions, allowing it to shift workloads seamlessly between clusters when needed.

Footprint: In addition to increased performance, simplified management, and better stability, the Yellowbrick footprint is significantly smaller: A single 15-node, 6u Yellowbrick on-premises instance replaced 31 40-core machines, reducing the amount of needed hardware by a third, requiring 20x less memory and a quarter of the compute core.

WINNING ON SPEED AND STABILITY

For ThreatMetrix customers, speed and stability improved dramatically, allowing them to run their businesses faster and more effectively. Best of all, Yellowbrick helped sharply reduce ThreatMetrix support costs and delivered its customers significant capital and operational savings.

ABOUT YELLOWBRICK DATA

Yellowbrick Data provides the world's fastest data warehouse for hybrid and multi-cloud environments. Enterprises rely on Yellowbrick Data Warehouse to power critical business outcomes and get answers to the hardest business questions for improved profitability, better customer loyalty, and faster innovation in near real time, and at a fraction of the cost of alternatives. Yellowbrick offers superior price/performance for thousands of concurrent users on petabytes of data, along with the unique ability to run analytic workloads on premises, in a private cloud, and/or in any public cloud and manage them in a simple, consistent way—all with predictable pricing via fixed-cost annual subscription.

Learn more at **yellowbrick.com**.

250 Cambridge Avenue, Suite 300, Palo Alto, California 94131 | USA | 1.650.687.0896