



## Top-10 Financial Services Company Unlocks Data Lake Insights for 1000s of Users

WITH YELLOWBRICK, **SOME QUERIES ARE NOW UP TO 100 TIMES FASTER** — resulting in thousands of daily users who can now run queries that simply weren't possible before.

### BUSINESS CHALLENGE

Financial services are heavily data driven. Every day, they must capture and analyze extremely large amounts of information—making the speed and scalability of their analytics environments of paramount importance.

This customer is a multinational financial services corporation and member of the S&P 100, with hundreds of billions of dollars in annual revenue. Every day, the company must capture millions of transactions and analyze petabytes of data across its partner and co-branding programs, merchant services, reward points programs, and more. Thousands of analysts across the company require immediate access to this data throughout the day—via both ad-hoc, interactive queries and prebuilt reports.

To support those analysis needs, the company maintains a data lake containing several petabytes of detailed transaction data, along with multiple data warehouses that map to its various lines of business. And like most financial services companies, its top requirements for these systems include speed, scalability, and price-performance. These are the same reasons why the company chose Yellowbrick for data lake augmentation and data warehouse modernization.

### DATA LAKE AUGMENTATION WITH YELLOWBRICK

The company's first look at Yellowbrick was when it began looking to augment its MapR-based data lake with a modern data warehouse—the primary goal being faster ad-hoc query performance. At the time, the company was using Hive for reporting, but, as data-driven businesses across all industries have come to realize, this architecture—a SQL query layer on top of a data lake—couldn't support thousands of analysts running interactive, ad-hoc queries on large data sets.

### OVERVIEW

Industry: Financial services

### CUSTOMER PROFILE

A multinational financial services company and member of the S&P 100.

### SITUATION

The company's Hive-on-MapR data lake couldn't support its thousands of users, who complained about slow response times for their ad-hoc, interactive queries. In addition, the company needed better performance from its aging Netezza data warehouse environment.

### SOLUTION

The company augmented its data lake by moving the data it contains into a modern data warehouse based on Yellowbrick—and has since deployed several additional Yellowbrick systems to modernize existing Netezza and DB2 systems.

### BENEFITS

- **Up to 100x faster query speeds**—resulting in much more productive users
- **Support for larger data sets**—enabling new insights into customer behavior
- **Faster data loads**—with some that took 7-10 hours now done in 20 minutes
- **Rapid migration**, with minimal code and stored procedure rewrites
- **Lower data center costs**—achieved through vastly reduced rack space, power, and cooling requirements

**Instead, this customer envisioned a better approach:** continue leveraging what its data lake did well—cost-effective storage for petabytes of raw data—and augment it by moving the data it contains into a modern data warehouse capable of delivering far greater speed, scalability, and price-performance.

In evaluating its options for data lake augmentation, the company compared a 15-node, 6-U Yellowbrick system against a 20-rack Teradata 6800. The company loaded two years of historical data from MapR into each system and then evaluated their performance across a set of 240 queries—with Yellowbrick beating the Teradata system on all but two.

Shortly thereafter, the company purchased its first Yellowbrick system. Every new transaction across the company's lines of business is loaded into Yellowbrick, which, thanks to its 5 PB capacity, is able to store five full years of transaction data. Even better, thanks to Yellowbrick's unique architecture, all of this data remains "hot" at all times, making it instantly queryable for the system's thousands of daily users.

## **DATA WAREHOUSE MODERNIZATION – REPLACING NETEZZA AND DB2 WITH YELLOWBRICK**

Following its successful initial deployment of Yellowbrick, the company began looking at where else it could leverage Yellowbrick's unparalleled speed, capacity, and overall price-performance. Additional Yellowbrick deployments have included the following use cases:

- **Accounts receivable and auditing** – where the company replaced a three-rack Netezza system with a single 10-U, 16-node Yellowbrick system that supports both reporting and ad-hoc, interactive queries. Again, data is loaded using Informatica and is accessed via MicroStrategy and other tools.
- **Merchant reporting** – where the company replaced a one-rack Netezza system with a single 6-U, 8-node Yellowbrick system. Reporting is done via MicroStrategy and SAS.
- **Partner program reporting** – where the company replaced three racks of Netezza with three 6-U, 15-node Yellowbrick systems for dev/test, production, and disaster recovery. Data for this environment comes from a myriad of sources and is loaded into Yellowbrick using Informatica, against which reports are run using MicroStrategy and Tableau. Data set size is about 60 terabytes.
- **Reward points** – where the company replaced a DB2 database used to analyze how customers spend their rewards with a single 6-U, 8-node Yellowbrick system that serves both ad-hoc queries and interactive reports.
- **Consumer loans** – where the company replaced a two-rack Netezza system with a single 6-U, 8-node Yellowbrick system.

Today, the company is actively working with Yellowbrick on several new initiatives, including one for fraud detection. In addition, although the company's current use of Yellowbrick has been entirely on-premises, it plans to investigate how it can take advantage of Yellowbrick's unique hybrid-cloud architecture to run its analytics workloads wherever it makes the most sense: on-premises, in a private cloud, in the public cloud, or any combination thereof—with the same predictable price-performance.

## BENEFITS / RESULTS

Through its use of Yellowbrick, this customer is benefiting in several ways.

- **Unmatched price-performance.** The main reason this customer chose Yellowbrick is because it's fast. In the data lake augmentation use case, some queries are now up to 100 times faster—resulting in thousands of daily users who no longer complain about query performance. What's more, they can now run queries that simply weren't possible before—enabling entirely new insights into customer behavior through deeper detail and increased predictive analysis.
- **Faster data loads.** Some data loads that used to take 7 to 10 hours are now completed in less than 20 minutes, reducing the time required before new insights can be derived from that information. Even better, because Yellowbrick looks and acts like a standard PostgreSQL database, the customer was able to achieve these gains using existing ETL processes and tools.
- **Rapid migration.** Most of the company's data warehouse modernizations—especially those where it moved from Netezza to Yellowbrick—were completed in about a month, with minimal code and stored procedure rewrites required. Not only has this accelerated time-to-value, but it has also helped minimize the costs and effort associated with having to maintain the old environment during the migration.
- **Reduced data center costs.** In several of its Yellowbrick deployments, the company replaced entire racks of Netezza with a single 6-U or 10-U Yellowbrick appliance. This has helped the company reduce data center costs through reduced rack space requirements, power and cooling needs, and so on—all while delivering increased query and report performance for the end users of those systems.

The company's continued adoption of Yellowbrick across new parts of its business is clear proof of the value that Yellowbrick is delivering—and how it can drive new business value across both data lake augmentation and data warehouse modernization scenarios.