



# Yellowbrick

## Stacking up versus the competition







## Yellowbrick Data Warehouse Overview

The Yellowbrick Data Warehouse delivers a modern data warehouse experience in your own cloud account as well as on-premises. We support all the benefits of truly cloud-native data warehousing – separate storage and compute and built-in elasticity, managed through SQL – without your data ever needing to leave your VPC or data center.

Yellowbrick customers are the world's largest global enterprises, running ad-hoc analytics, and operational data warehouses supporting business-critical workloads with high concurrency. Wherever you run Yellowbrick, you'll find rich workload management, real-time data ingest, the ability to load and query data together on the same instance, high availability, and replication across clouds and on-premises for disaster recovery. You can say "goodbye" to the instability of Hadoop and "hello" to a stable, reliable, and trustworthy data warehouse.

Pricing with Yellowbrick is open, simple, and predictable. We've saved many customers millions of dollars per year in legacy and cloud spend. You use and pay for your storage and compute, making use of your cloud credits without paying

smaller companies to mark up your infrastructure costs. You can be comfortable meeting your security, regulatory, and compliance requirements.

If you're considering a new migration to the cloud, Yellowbrick avoids lock-in to any particular cloud vendor or database. We use open-source PostgreSQL as our SQL syntax of choice and are backward compatible with on-premises ETL tools like Informatica PowerCenter and CDC tools, Oracle GoldenGate or Qlik (Attunity) Replicate. We don't do lock-in. We don't want to be the next Oracle (nor did our founders come from there) and we use open standards to ensure the portability of your data and workload.

We're proud to have the best performance in the industry, at the lowest possible cost. No one runs data warehousing workloads faster than Yellowbrick.

Keep reading to see how we stack up versus our competitors in the market.







# Snowflake

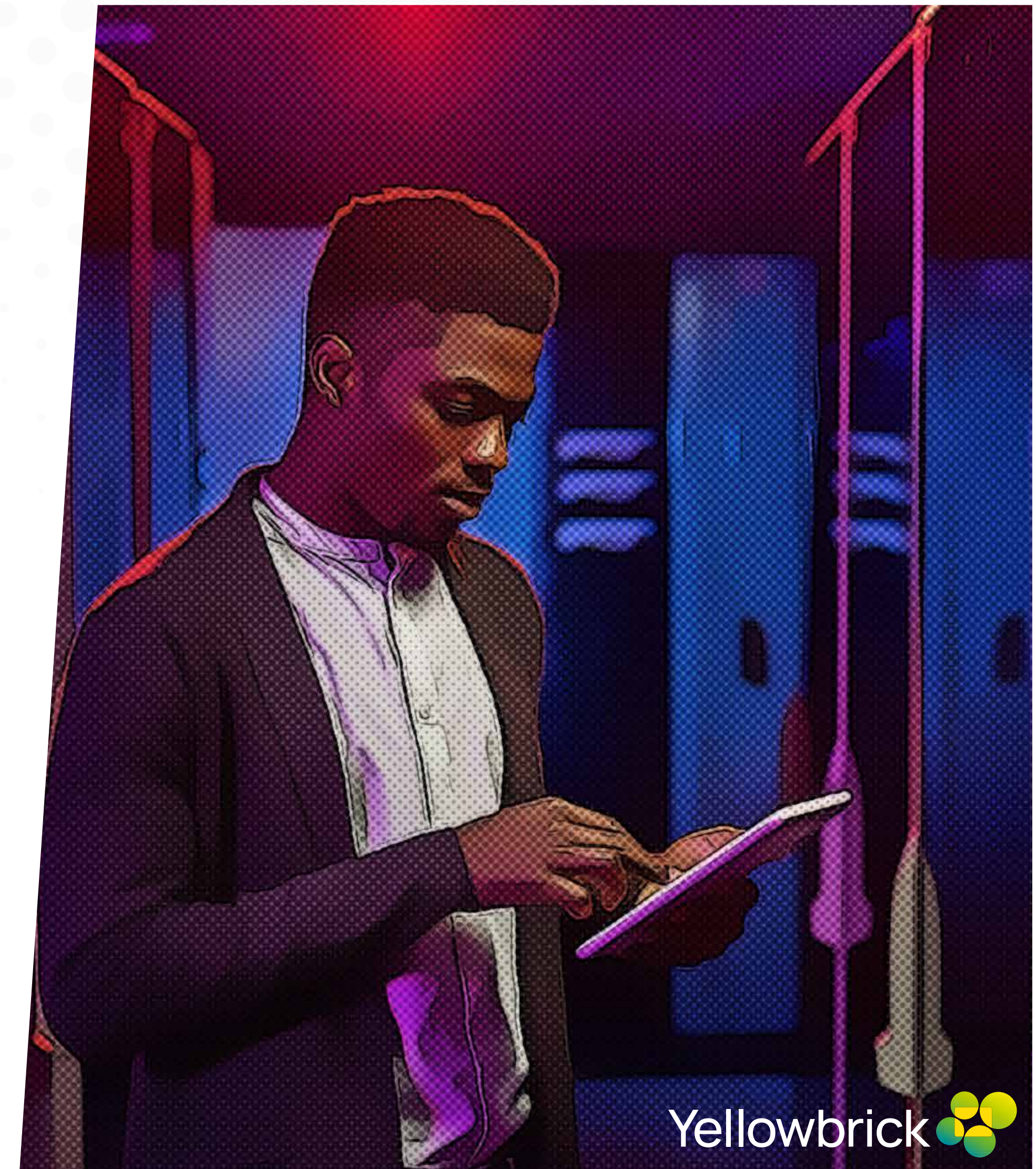
Yellowbrick runs both on-premises as well as in your own cloud account. Snowflake is a cloud-only offering. For many enterprises, the sensitivity of their data and workloads, as well as factors such as data gravity, means that not all data warehousing operations can be moved fully to the cloud.

Our pricing model is predictable – you won't need to worry about buying proprietary credits, or if they will get canceled on you if you don't renew. We don't have Snowflake's low query concurrency limits that force you to run even more clusters to burn even more credits. With Yellowbrick, you simply decide what size data warehouse you want (vCPUs) and pay a flat subscription over a 1- or 3-year term. You can consume as much of Yellowbrick as you like, 24x7 and 365 days of the year, without worrying about cost overruns.

Snowflake scales in large expensive increments. With Yellowbrick you add one node at a time and are charged by the second for the extra vCPUs – no huge step changes in cost. With our advanced workload management, you'll end up needing to scale much less frequently and needing a fraction of the infrastructure you'd need to get the equivalent performance for mixed workloads in Snowflake, greatly reducing spend.

With Snowflake you lose ownership of your data and the infrastructure it runs on. Yellowbrick runs in your own cloud account and network. You benefit from your enterprise agreements with cloud providers. And still, we've made the process of operating Yellowbrick easy – it's data warehousing as a service.

Nobody runs data warehousing workloads faster than Yellowbrick, and we've saved some Snowflake customers millions of dollars per year.







# Snowflake

A digital marketing customer conducted query performance tests using 12 representative SQL workloads against its own data. The data consisted of 3.5 trillion records distributed across 82 tables, with a volume of approximately 500 TB. The query tests compared a 32-node 2XL Snowflake cluster with a 30-node Yellowbrick cluster. Yellowbrick was 6x faster on average across all 12 queries and 147x faster on one of the longest running queries.

	Yellowbrick	Snowflake
Predictable, low cost	✓	✗
Consumption pricing with dollars, not credits	✓	✗
Per-second billing	✓	✓
Runs and secured in your own VPC/cloud account	✓	✗
On-premises deployments	✓	✗
Disaster recovery & protection from global outages	✓	✗
Elastic with separate storage/compute	✓	✓
Kubernetes-based for private cloud	✓	✗
High concurrency, predictable latency for operational workloads	✓	✗
Workload management	✓	✗
Real-time streaming data ingest at no additional cost	✓	✗
Compatibility with on-premises ETL (Informatica PowerCenter, etc.)	✓	✗
Full control over network path and firewalls	✓	✗
Granular scale by single node	✓	✗
Integrated machine learning	Roadmap	✓
Data sharing	Roadmap	✓
Support for semi-structured & geospatial data	Roadmap	✓





# Teradata

We’ve done away with the things that annoy you the most about Teradata: Fork-lift upgrades, massive on-premises appliances, complex database administration, and – most importantly – the sky-high prices... while supporting the things you’ve grown to love the most about Teradata: Support for complex, mixed workloads running at high concurrency, predictable response times, a broad portfolio of ecosystem integrations, sophisticated rules-based workload management, advanced SQL support, and the SLAs needed for operational, business-critical workloads.

From there, we’ve plugged key feature gaps like built-in, asynchronous replication for disaster recovery and support for real-time data streams, and we offer superior query performance at a much lower cost. We wrap this in a modern, truly elastic, cloud-native data warehouse that runs

same the in the cloud and on-premises.

Migration is far more straightforward than it used to be. We have tools that automatically migrate the vast majority of stored procedures, BTEQ scripts and ETL, and services that can help with the remainder. We can even survey your existing system up-front to tell you how long migration will take.

A top 10 financial services company compared a 15-node, 6U Yellowbrick system against a 20-rack Teradata 6800. Today, every new transaction across the company’s lines of business is loaded into Yellowbrick, instantly queryable for thousands of daily users.

	Yellowbrick	Teradata
Low cost	✓	✗
Autonomous without indexing/tuning	✓	✗
Scale without downtime	✓	✗
Real-time streaming data ingest	✓	✗
High concurrency, predictable latency for operational workloads	✓	✓
Workload management, ad-hoc query	✓	✓
Cloud-native Kubernetes architecture – run in your private cloud as well as on-premises	✓	✗
Built-in asynchronous replication for disaster recovery	✓	✗
Support for semi-structured & geospatial data	Roadmap	✓
Integrated machine learning	Roadmap	✓
Hardware instance size per 1PB warehouse	24 U	16 racks



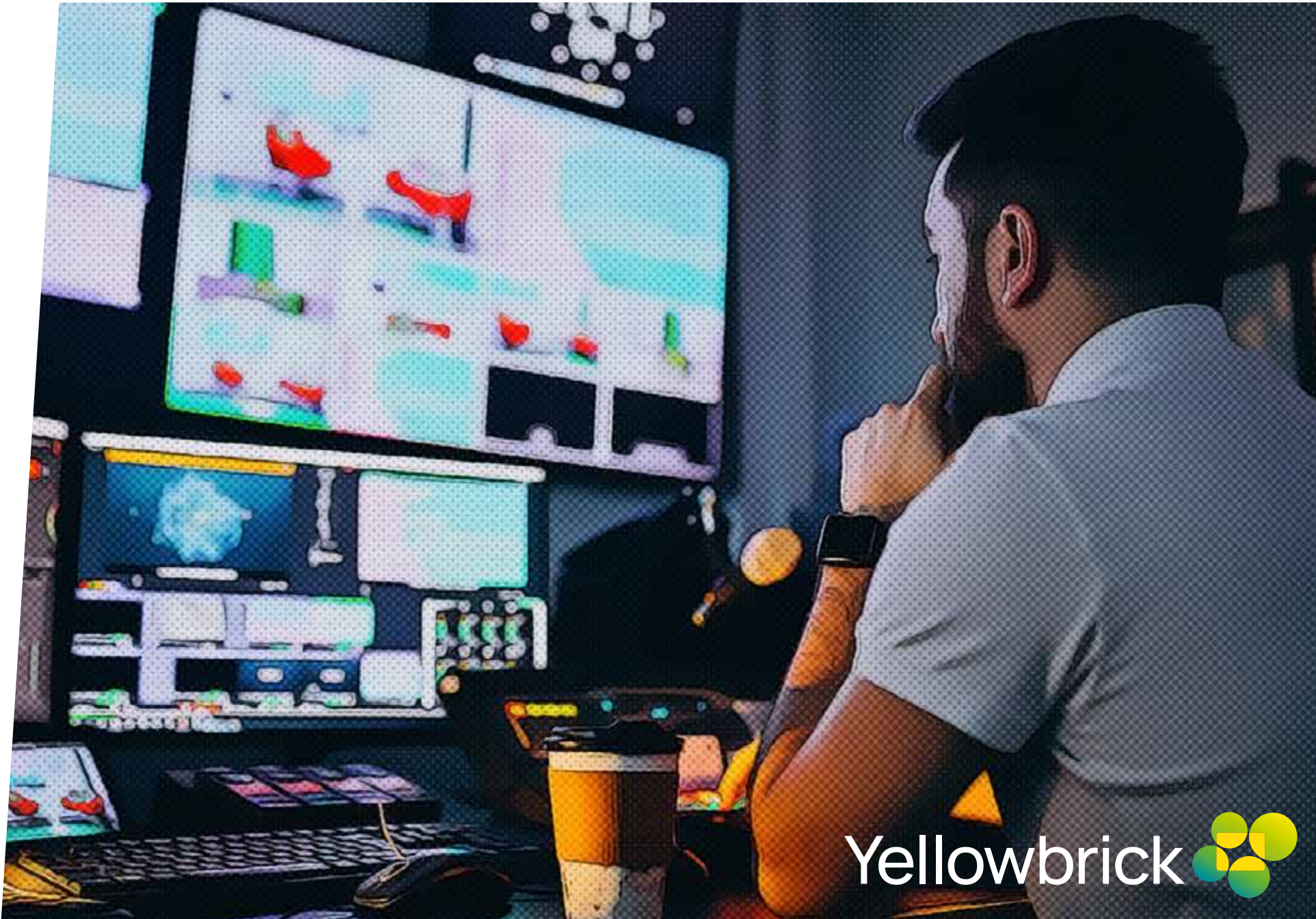


# Amazon Redshift

If you've been looking at or using Amazon Redshift, you must be intrigued by the benefits of true elasticity and cross-cloud portability of competing solutions, yet worried about more vendor lock-in, unpredictable performance, and run-away cost. The Yellowbrick Data Warehouse takes the best of both worlds. Our data warehouse is autonomous, requiring little to no DBA involvement, inherently multi-cloud (and on-premises), has simple transparent pricing, and has built-in asynchronous replication for disaster recovery.

	Yellowbrick	Redshift
PostgreSQL-compatible syntax	✓	✓
Low cost	✓	✗
Runs in GCP, AWS, and Azure	✓	✗
Runs on-premises	✓	✗
Autonomous without indexing/tuning	✓	✗
Built-in asynchronous replication for disaster recovery	✓	✗
Real-time streaming data ingest	✓	✗
Scan acceleration support	✓	✓
Support for semi-structured & geospatial data	Roadmap	✓
Compatibility with on-premises ETL (Informatica PowerCenter, etc.)	✓	✗
Automated migration from legacy technology	✓	✓

We do all of this without the quirks and complexities that come with a 15-year-old code base like Redshift: We support higher concurrency and mixed workloads without having to worry about things like “vacuuming” or if your cluster can scale without downtime. You can take CDC or Kafka data and stream it into Yellowbrick in real-time, with great performance, and our SLAs are designed for operational, business-critical workloads.





# Google BigQuery

The Yellowbrick Data Warehouse combines the best features of Google BigQuery with a true, enterprise-grade database.

At Yellowbrick, we've built a sophisticated cost-based query optimizer to enable complex queries across highly variable schemas. We've coupled that with the world's most efficient database execution engine that thrives on running workloads consisting of highly concurrent, complex queries that need to be completed with predictable latency. Google forces you to pick between

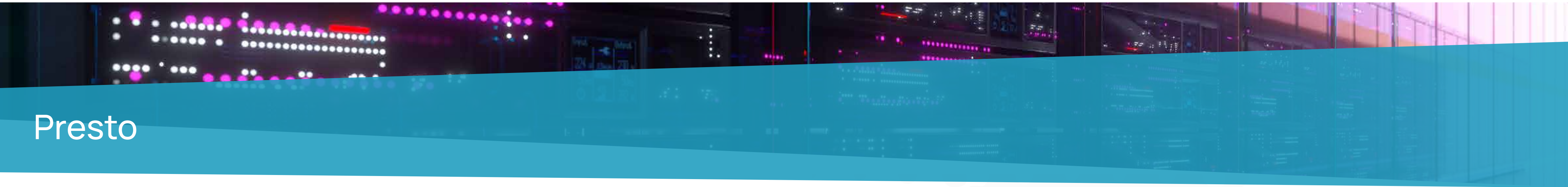
	Yellowbrick	BigQuery
Predictable transparent pricing, low cost	✓	✓
Advanced cost-based query optimizer for complex schemas	✓	✗
Cloud data warehouse in your VPC on Azure, AWS, and GCP	✓	✗
Runs on-premises and multi-cloud	✓	✗
Kubernetes-based for private cloud	✓	✗
High concurrency, predictable latency for operational workloads	✓	✗
Workload management	✓	✗
Built-in asynchronous replication for disaster recovery	✓	✗
Compatibility with on-premises ETL (Informatica PowerCenter, etc.)	✓	✗
PostgreSQL-compatible syntax	✓	✗
Elastic with separate storage/compute	✓	✓
Full control over network path and firewalls	✓	✗
Integrated machine learning	Roadmap	✓
Support for semi-structured & geospatial data	Roadmap	✓

“batch” and “interactive” mode – with different tradeoffs and pathetic concurrency limits. At Yellowbrick, we support advanced workload management for thousands of concurrent users to make sure that business-critical SLAs can be met. We even have asynchronous replication – between clouds or between on-premises installations and the cloud – built into the database.

We're proud to have the best performance in the industry, at the lowest possible cost with predictable pricing that doesn't take a Ph.D. to understand or leave you with a hole in the wallet due to a runaway query. No one runs data warehousing workloads faster than Yellowbrick.







# Presto

The Yellowbrick Data Warehouse fits easily into your Data Lake architecture, without needing to replace it. Most importantly, Yellowbrick is an enterprise-quality database. We have enterprise-class stability; you don't need to manage storage or file layouts or partitions; we have built-in multi-site replication, transactional consistency, and the ability to plan complex SQL queries, run stored procedures, role-based access control; and all the other things you'd expect from a "database."

Yellowbrick supports high concurrency and low-latency interactive queries. Badly written run-away queries can't disrupt other users' queries, and you can have thousands of users querying your data

at the same time. Yellowbrick guarantees correct results, isn't written in Java – a platform choice that leads to fundamentally unpredictable performance – and doesn't periodically crash. Furthermore, it interoperates with all of your existing ETL. At one of the world's largest logistics companies, a 15-node Yellowbrick cluster ran ad-hoc, interactive queries 30x faster than a 120-node Presto cluster.

If you have a need for operational data warehousing, high concurrency, ad-hoc queries, predictable response times, or simply software that doesn't crash all the time, Presto – like other SQL-on-Hadoop products – will do nothing for you. Follow in the footsteps of the world's largest insurance, logistics, credit card, and risk companies and use Yellowbrick to enable business users to get value out of your Data Lake.

At one of the world's largest logistics companies, a 15-node Yellowbrick cluster ran ad-hoc, interactive queries 30x faster than a 120-node Presto cluster.

	Yellowbrick	Presto
Is it a database?	✓	✗
Enterprise-class stability	✓	✗
Built-in multi-site replication	✓	✗
Fast inserts, updates, deletes	✓	✗
Autonomous without indexing/tuning	✓	✗
Hardware-specific, native code performance optimizations	✓	✗
Cloud data warehouse in your VPC	✓	✓
SaaS cloud data warehouse (data elsewhere)	✓	✓
Elastic with separate storage/compute	✓	✓
Kubernetes cloud-native architecture	✓	✓
High concurrency, predictable latency for operational workloads	✓	✗
Advanced Workload management	✓	✗
Real-time streaming data ingest	✓	✗





If you’ve used the IBM Netezza platform, IBM will try to sell you “Cloud Pak for Data with Netezza Performance Server.” IBM has succeeded in positioning this as a new, modern data warehouse offering; but behind the charade is the ancient, buggy Netezza database engine with the same old limitations: Poor concurrency; poor write performance; largely missing workload management; weak business continuity; a support team that can’t debug or fix core issues; performance that’s weak compared to modern cloud offerings; no working replication or disaster recovery strategy; and most importantly, no real roadmap at all.

IBM will position this “new” offering as working in the cloud, but it’s just a “lift and shift, bundle in

	Yellowbrick	Netezza
High software reliability and excellent support	✓	✗
Actively maintained database engine with a roadmap	✓	✗
Cloud data warehouse in your VPC	✓	✗
Elastic with separate storage/compute	✓	✗
Data stored on S3, no backups required	✓	✗
High concurrency, predictable latency for operational workloads	✓	✗
Workload management	✓	✗
Real-time streaming data ingest	✓	✗
High density, small footprint solution	✓	✗
Runs in AWS, your VPC	✓	✓
Runs on-premises	✓	✓
Autonomous without indexing/tuning	✓	✓
Scale without downtime	✓	✗
Built-in asynchronous replication for disaster recovery	✓	✗

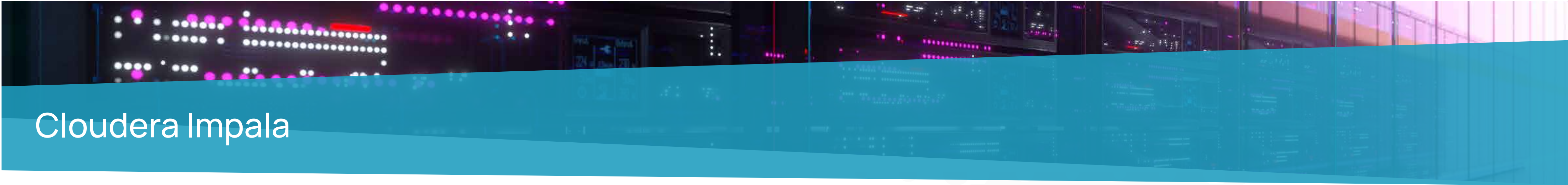
containers” of legacy technology built on Postgres 7 from the year 1999. Truly cloud-native databases have separate storage and compute, storing persistent data on object stores like S3. IBM doesn’t. Truly cloud-native databases have SQL-managed elasticity and modern, web-based query tools. IBM doesn’t. Truly cloud-native databases authenticate with OAuth2. IBM doesn’t. The development teams of truly cloud-native databases believe that now is the most exciting era in database architecture in decades – they don’t put a relic in a container and re-spin it as a reincarnation of a dead brand.

If you’re going to invest in a new data warehouse, you want Netezza compatibility, but with a truly cloud-native architecture, elasticity, a modern management experience, high concurrency, business continuity, the ecosystem, and ETL integrations you know and love – that runs in any cloud and on-premises. There’s no need to look at IBM anymore: Choose Yellowbrick.

“In our testing of Yellowbrick, we compared the performance of a six-rack TwinFin to the 6U baseline Yellowbrick system. And performance was anywhere from 3 to 50 to 100 times faster.”

RICK MAHUSON Vice President, R&D Business Analytics, TEOCO





# Cloudera Impala

The Yellowbrick Data Warehouse fits easily into your Data Lake architecture, without needing to replace it. Most importantly, Yellowbrick is an enterprise-quality database. We have enterprise-class stability; you don't need to manage storage or file layouts or partitions; we have built-in multi-site replication, transactional consistency, and the ability to plan complex SQL queries, run stored procedures, role-based access control; and all the other things you'd expect of a "database."

Yellowbrick supports high concurrency and low-latency interactive queries. Badly written run-away queries can't disrupt other users' queries, and you can have thousands of users querying your data

at the same time. Yellowbrick guarantees correct results, yet most importantly we have enterprise support and a development team that can fix problems: Most Impala users are used to regular crashes, and Cloudera can't fix them. Yellowbrick, being a database, interoperates with all of your existing ETL.

If you have any need for operational data warehousing, high concurrency, ad-hoc queries, predictable response times, or simply software that doesn't crash all the time, Impala – just like other failed SQL-on-Hadoop products – will do nothing for you. Follow in the footsteps of the world's largest insurance, logistics, credit card, and risk companies, and use Yellowbrick to enable business users to get value out of your Data Lake. We have many customers who have happily said goodbye to Impala, delighted their user communities, and saved the time, money, and effort investments in Impala that ultimately proved to be futile.

	Yellowbrick	Impala
Buggy and poor support	✗	✓
Is it a database?	✓	✗
Enterprise-class stability	✓	✗
Fast inserts, updates, deletes	✓	✗
Real-time streaming data ingest	✓	✗
Autonomous without indexing/tuning	✓	✗
On-premises deployments	✓	✓
Hardware-specific performance optimizations	✓	✗
Cloud data warehouse in your VPC	✓	✓
Elastic with separate storage/compute	✓	✓
Kubernetes-based, for private cloud	✓	✓
High concurrency, predictable latency for operational workloads	✓	✗
Advanced workload management	✓	✗

