



# Yellowbrick

## Yellowbrick Data Warehouse

### Technical Overview





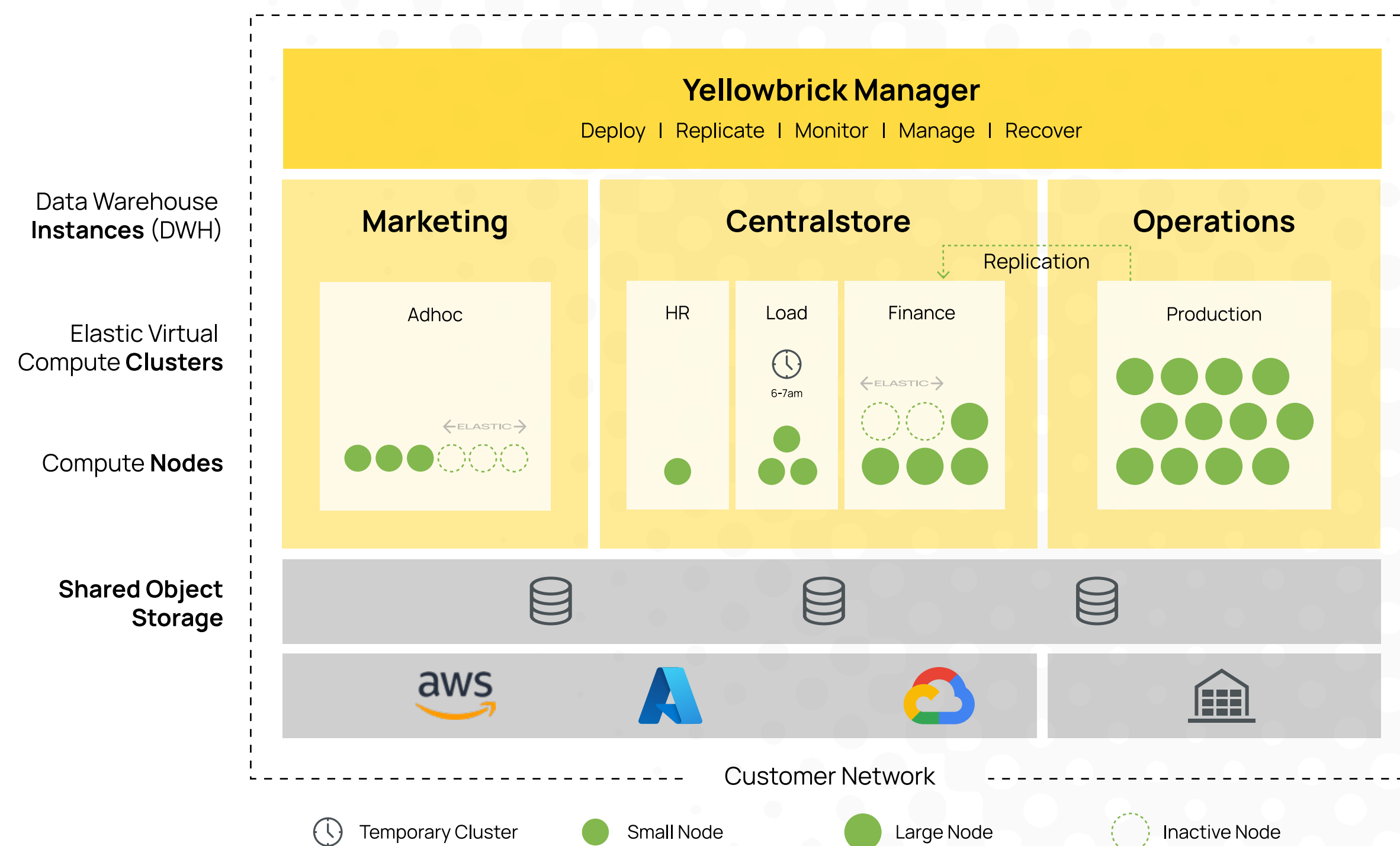
# Yellowbrick Data Warehouse

## The only modern enterprise cloud data warehouse.

The Yellowbrick Data Warehouse is an elastic, massively parallel processing (MPP) SQL database that runs on-premises, in the cloud, and at the network edge. It was designed for the most demanding batch, real-time, ad hoc, and mixed workloads and can run complex queries at up to petabyte-scale with guaranteed sub-second response times. Yellowbrick is proven, providing business-critical services at many large global enterprises with thousands of concurrent users. It is available on AWS, Azure, and Google Cloud as well as on-premises.

### Elasticity in Your Cloud Account

The Yellowbrick Data Warehouse runs in your cloud account without data ever leaving your network to an external SaaS provider. This eliminates compliance and security risks. Running costs are lowered by paying for the cloud infrastructure (both storage and compute) using your own enterprise cloud agreements.



Yellowbrick Data Warehouse: Logical Enterprise Architecture





In an industry-first, full SQL-driven elasticity with separate storage and compute is available within your own cloud account as well as on-premises. Compute resources – elastic, virtual compute clusters (VCCs) – are created, resized, and dropped on-demand through SQL, and cache data persisted on shared cloud object storage. For example, ad-hoc users can be routed to one cluster, business-critical users to a second cluster, and more clusters created and dropped on demand for ETL processing.

Each data warehouse instance runs completely independently of one another. There is no single point of failure or metadata shared across instances. Global outages – when deployed with replication across multiple public clouds and/or on-premises – are impossible.

Yellowbrick is secure by default with no external network access to your database instance. Encryption of data at rest is standard with keys you manage. Columnar encryption, granular role-based access control, column masking, OAuth2, Active Directory, and Kerberos authentication are built in. Integrations with best-in-class enterprise data protection solutions secure PII data. Enterprise-class high availability, backups for data retention, and asynchronous replication for disaster recovery are standard.

### Simple and Predictable Pricing

We support both on-demand and subscription-based pricing. All pricing is based on consumption of vCPUs for compute; we do not charge for storage since data is persisted on object storage in your own cloud account. On-demand pricing caters to short-term burst needs and is billed monthly in arrears without credits. Subscription pricing is predictable, works across cloud and on-premises, and allows efficient acquisition of capacity that you know you'll need. Models can be mixed and matched to meet business objectives.

### Designed for Performance

Yellowbrick was conceived with the goal of optimizing price/performance. The storage engine is a hybrid column and row store: Most data is persisted in the column store while the row store supports real-time streaming ingest of hundreds of thousands of records per second from CDC tools and Kafka. Yellowbrick's patented Direct Data Path Architecture is an OS bypass technology enabling in-memory analytics performance at petabyte-scale without requiring a typical database buffer cache – leading to more predictable response times and massive cost reductions.



Rules-based workload management makes sure business objectives are met in an environment with dynamically changing workloads. Long-running ad-hoc queries do not obstruct business-critical queries. Queries operate with their own set of assigned resources and hard resource controls are present to make sure queries can't starve each other. Yellowbrick supports thousands of concurrent users and, through provisioning of multiple elastic compute clusters, can scale to thousands of queries per second.

### Open Standards Support

Yellowbrick's database engine is fully ACID compliant. A deliberate design choice was to make use of PostgreSQL's SQL grammar, wire protocols, and metadata schema to avoid vendor lock-in and provide compatibility with a database familiar to modern developers. We've enhanced the PostgreSQL grammar with compatibility functions for other databases as well as improved manageability. All core SQL data types are present (numeric, UTF-8 varchar, dates and times, etc.). Further semi-structured data support is a key roadmap item. Views and PL/pgSQL stored procedures are fully supported as are cursors for data retrieval.

Access to Yellowbrick is through PostgreSQL ODBC, JDBC, and ADO.NET drivers. A substantial number of commercial and open-source tools, including Python, R, Kafka, and Spark interoperate with Yellowbrick.

### Availability

The Yellowbrick Data Warehouse is designed for business-critical data warehouse workloads and has no single points of failure. It is resilient to storage, server, and network outages. Data is persisted on shared object storage for the highest possible availability in the cloud and on erasure-coded local storage for on-premises deployments.

Full, cumulative, and incremental backups allow businesses to meet off-site data retention requirements. Transactionally consistent, asynchronous replication is built in and supports failover and failback; replication of DDL, data, and metadata allow provisioning of read-only hot standby databases for disaster recovery which may be in the same cloud, a different cloud, or on-premises.



### Modern with Minimal Management

The Yellowbrick Data Warehouse largely runs itself on autopilot. Minimal administrative activities are required: there's no need for creating and maintaining indexes, vacuuming, keeping statistics up to date, and defragmenting; provisioning and managing storage is completely unnecessary.

A friendly web UI called Yellowbrick Manager surfaces all information needed to keep the instances running, configure integrations and control, and optimize workloads. For developers, Yellowbrick Manager provides a simple way to execute queries, develop and maintain schemas, and profile query plans. All management and monitoring functionality can be accomplished through SQL and system tables as well as the web UI.

### Migration

Migration from legacy data warehouses and SQL-on-Hadoop is largely automated. We partner with Next Pathway to offer their SHIFT™ Migration Suite. Shift features a workload profiler to automatically isolate workloads and identify their dependencies in complex data warehouses and Hadoop clusters, allowing estimation of migration effort and cost ahead of time. SHIFT enables >95%

automated migration of the vast majority of database objects as well as ETL, BI, and even BTEQ scripts. Testing and validation services are offered alongside. We also partner with KPMG, Capgemini, Accenture, ZS, Systech, and Cognizant for other ongoing development and migration work.

Thanks to Yellowbrick's unique distributed data cloud architecture, cloud migrations can be staged to reduce risk. Staging allows you to incrementally migrate from legacy on-premises warehouses to Yellowbrick, replicating data to the cloud and moving workloads as needed. Organizations with complex on-premises ecosystems prefer this approach: In particular, Yellowbrick is supported by both Informatica Powercenter and Informatica Cloud to enable easier ETL migrations. Our Customer Success Managers have first-hand migration experience and will be by your side throughout the process.



## Summary

Yellowbrick is the modern data warehouse designed to solve today's analytics challenges. It provides full elasticity in your own cloud account as well as on-premises, with separate storage and compute. Pricing is simple and predictable, and our architecture, optimized for performance, means that nobody runs data warehouses faster or more cost effectively than Yellowbrick.

Yellowbrick is built on open standards to avoid lock-in, meets the availability needs of business critical, ad-hoc workloads, and is easy to use. Migration from legacy data warehouse platforms or Hadoop is largely automated.



# Yellowbrick

Intrigued? Find your biggest, most complex data set and most important business challenge and invite us in for a Yellowbrick test drive. We'll show you how high-speed data analytics can redefine your business opportunities.

[Book a Test Drive](#)