

The Distributed Data Cloud is an architectural pattern for data management and analytics that abstracts away the details of the cloud, on-premises and network edge infrastructure from the end-user. It is designed to address the challenges that enterprises face in terms of cloud concentration risk, efficiency and modernization as they progress on their journey to the cloud. To align with the Distributed Data Cloud blueprint, a data management and analytics technology must display five traits:

1. A platform-agnostic runtime allowing the provisioning of data and analytics anywhere
2. A common user experience anywhere
3. Common security and governance features on any deployment target
4. Cost and technology efficiency anywhere – minimize resources and allow for strong cost management (FinOps) and spend guardrails
5. A single control plane, tying all deployments together, public cloud, on-premises and at the network edge

Yellowbrick has created a data warehouse that embodies the five traits of the Distributed Data Cloud:

### **1. Platform agnostic runtime**

Yellowbrick has embraced the latest serverless and microservices paradigms to deliver an elastic enterprise data warehouse that runs everywhere. This means you can choose to deploy data warehousing in any public cloud or in your own data center, minimizing cloud concentration risk. With Yellowbrick, you don't have to lock into a single cloud or SaaS vendor for your data warehousing needs, and you retain ownership of your data at all times.

We've taken best-of-breed approaches developed over decades by the on-premises data warehousing industry, where watchword was efficiency and reliability, and combined it with the elasticity and self-service user experience of the new cloud data warehouses. Furthermore, we've delivered a platform that reaches new levels of concurrency and high availability. The Yellowbrick enterprise data warehouse delivers the best of all data warehousing worlds.

Gartner predicts that by 2023, 50% of all data will be generated outside of the public cloud or data center.

Deploying data warehousing closer to where the data is generated will be increasingly important. The idea of backhauling edge data to a data center or public cloud will be a non-starter, due to bandwidth limitations. Yellowbrick's ability to deploy at the network edge, consume and process streaming data, and replicate to the public cloud, means that it is ready to address emerging IoT/edge use cases.

## **2. Common user experience**

It's not enough just to provide a compelling data warehousing user experience on a single cloud. The types of customers that are impacted by concentration risk need to ensure the same experience is available anywhere – in their on-premises data center, in public clouds, and increasingly at the network edge.

The end consumer shouldn't have to care where their data warehouse is physically running, and they shouldn't have to change their working processes or tools to deal with the idiosyncrasies of a given cloud or on-prem environment. Instead, what is required is a user experience that is the same wherever the data warehouse is deployed.

Yellowbrick was designed with common, open standard compatibility in mind. From the outside, Yellowbrick looks and feels like a PostgreSQLs database. It's easy to hire the skillsets need to use it, and easy to migrate onto and off Yellowbrick. It can happily co-exist with other data warehouses and databases and there are a range of options in place to support data and workload migration between other data management and analytics platforms.

## **3. Common security and governance**

The differences in authentication and authorization approaches represents one of the largest impediments to the deployment of data warehousing on different cloud platforms. On-prem enterprises enter the journey to cloud with a whole set of legacy baggage, whether it's locally managed user accounts, LDAP, PKI or Kerberos. Today's enterprise users expect a login experience that mirrors the one they get from their consumer devices. They expect authentication based on an SSO experience delivered through an identity service provider such as Google or Microsoft.

Yellowbrick provides a common framework for authentication and authorization that can be set up for any IDP, in any cloud and in any on-premises legacy environment. Our OAuth2/OIDC compatibility means you can use the same mechanism to support single sign-on wherever the data warehouse is deployed.

At the database level, every object in Yellowbrick is subject to access control with the security posture that everything is locked down by default. When combined with our column-level encryption capability, this means you can open up data to users on a fine-grained, as required basis.

#### **4. Cost and technology efficiency**

Yellowbrick's pricing model consists of an easy-to-understand blend of capacity billing and on-demand billing. If your workload is well understood, then a capacity-based subscription offers the best value. In this model, you reserve a set of virtual CPUs, over a 1- or 3-year term, and are free to use that set of vCPUs as you like – even for “always on” applications. This means your spend is completely predictable. If the workload is variable in nature, perhaps due to seasonal events or the need for additional compute capacity at the end of each month, then you are free to burst beyond the fixed capacity and consume vCPUs elastically.

Although on-demand billing naturally introduces a level of uncertainty into the monthly spend, the efficiencies inherent in the Yellowbrick software mean that the magnitude of the overall spend will be lower compared to the alternatives. We achieve these efficiencies by taking a holistic approach to the database management problem. We optimize not only the database software, but also the operating system kernel. We've introduced our own memory management, threading model, network and storage device drivers, workload management, as well as data path optimizations that route data directly from NVMe SSDs into the L3 caches on the CPUs themselves, to deliver incredible levels of performance at high concurrency. We detail these optimizations elsewhere.

Another feature of Yellowbrick that helps keep overall cloud spend down is our ability to run in your VPC and cloud account. This means you can apply any discounts for cloud infrastructure spend your enterprise receives from the CSP directly to your spend with Yellowbrick. Contrast this with SaaS data warehouse vendors, where your data resides in their AWS, Azure or Google Cloud account, and they get the benefits from CSP discounting. What's more, when you run Yellowbrick in your own VPC, you now own your own data rather than having to hand it over to a SaaS vendor to manage. This is attractive to businesses in highly regulated industries because it removes one component out of the chain of potential single point failures. Your data is more available and more secure in the cloud with Yellowbrick.

Unfortunately, the economies of scale that the cloud vendors achieve through their immense infrastructure buying power do not necessarily translate into cost savings for the consumers of cloud data warehousing.

Legacy data warehouse vendors that have simply lifted-and-shifted their on-premises database software into the cloud find that they must use expensive compute instances and persistent block storage to work. In some cases, it can be cheaper to carry on running the data warehouse in your own data center rather than move it to the cloud. It's advisable to target cloud data warehouse solutions built around the central design goal of efficiency. Whether that's through using fewer compute instances and more efficient software stacks, or using cheap and deep object storage to persist data, rather than expensive block storage. This is the approach we've taken at Yellowbrick. We don't think it's cost-efficient or environmentally appropriate to have to spin up a massive cloud compute cluster to support a data warehouse, and we use low cost S3, ADLS gen2, and Google Cloud Storage to persist data.

## 5. Single control plane

We provide a single pane of glass for provisioning, managing, monitoring and billing with Yellowbrick. It's designed to support the deployment of data warehouses simultaneously in any public cloud, on-premises environment, and at the network edge. This unifying control plane is a critical feature of the Distributed Data Cloud, tying together data warehouses deployed over a geographically disparate set of locations. With Yellowbrick you can deploy data warehouses at the point of need, based on data gravity, latency, sovereignty, and security requirements for your business, all managed across a single control plane, with no single points of failure.

## Summary

The Yellowbrick Enterprise Data Warehouse was built to address the concentration risk, efficiency and modernization challenges businesses face as they journey to the public cloud. Yellowbrick helps to enable the Distributed Data Cloud, and delivers industry-leading levels of concurrency, efficiency, availability, flexibility, elasticity and usability to the largest enterprises in the process.