

DATABASES / KUBERNETES

Enterprise Data Platforms on Kubernetes Challenge Status Quo

Using Kubernetes operators, a SQL-based data platform becomes a K8s application that offers extensibility, resiliency, automation and portability.

Oct 2nd, 2024 7:06am by [Mark Cusack](#)



Image from natrot on Shutterstock

ANNUAL READER SURVEY

We are planning our coverage for 2025 and we need your input! Take 5 minutes to chime in and let us know what you're interested in.



BEGIN THE SURVEY

You could win a \$1,000 gift card for your participation!

CNCF sponsored this post.

Software as a Service (SaaS) data platforms have enjoyed significant adoption over the last few years. Growth has been driven by the perceived benefits of low management overhead, compelling user experience and faster time to business value.

However, as with all technologies, there are trade-offs that should be considered when deciding if a SaaS deployment model is right for your company. Let's explore the pros and cons of SaaS-based data platforms and dive deep into how **Kubernetes** is emerging as an alternative deployment option for a new breed of data platforms.

Will SaaS Eat the World?

Over the past few years, the data and analytics world has been obsessed with delivering and consuming SaaS offerings, and the market has been dominated by **Snowflake** and Databricks. Both offer public cloud-only SaaS platforms for building solutions for use cases such as data warehousing, data integration and preparation, and business intelligence and AI/machine learning.

The ease of consumption, on-demand elasticity and the broad set of features that these services offer have fueled its popularity. Much of the growth has been at the expense of legacy technologies from Teradata, Cloudera, Vertica, **Oracle** and others that have been unable to replicate the public cloud experience these vendors offer.

However, SaaS platforms such as these are not without challenges. They offer a one-size-fits-all model to deliver economies of scale. SaaS providers play a delicately balanced margin-stacking game. They buy compute capacity from a public cloud service provider (CSP), and then build software, operations, resiliency and support services on top.



The Cloud Native Computing Foundation (CNCF) hosts critical components of cloud native software stacks, including Kubernetes and Prometheus. CNCF serves as the neutral home for collaboration and brings together the industry's top developers, end users and vendors.

[Learn More →](#)

THE LATEST FROM CNCF

[Understanding initial states in cloud migration](#)

2 October 2024

[Reducing MTTD and increasing observability with Linkerd at loveholidays](#)

2 October 2024

[CNCF and sustainability: how cloud native projects are shaping the green tech movement](#)

1 October 2024

Excess compute capacity is also procured and maintained to deliver rapid expansion capabilities. Turning a profit requires that many customers consume data in the same way, in the same regions, with little customization. Hence, all users of the SaaS data platform have the same experience, the same security controls and defaults, and the same limitations regarding how and where their **data is stored and managed**.

When an organization signs up to use a SaaS data platform, it is handing over its data to the SaaS vendor to manage on its behalf. Basically, an organization is ceding control of the company's data to a third party and, in turn, paying the vendor to access it. This lack of control over how and where data is managed is troubling for many CIOs and CISOs, particularly those in the federal government sector and companies subject to data sovereignty regulations.

TRENDING STORIES

3. Automating the Security of Kubernetes Clusters in the Cloud
4. Build Kubernetes Security on Intent, Not Rules
5. How To Jump-Start Your Stalled Kubernetes Migration

Of concern is the lack of on-premises and hybrid cloud offerings to complement the SaaS vendors' public cloud platforms. To address **data security and sovereignty** issues, new data platforms are emerging that leverage the elasticity, resilience and portability of Kubernetes and provide a SaaS-like user experience within an organization's own data center or CSP account.

DIY Data and Analytics on Kubernetes

Kubernetes has been successful in supporting scale-out stateless web services. Its elasticity, resilience and extensibility make it an excellent choice as the orchestration platform for containerized web application fleets.

Equally important is the emergence of managed Kubernetes offerings from **AWS**, **Microsoft** Azure and **Google** Cloud Platform in the form of Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), and Google Kubernetes Engine (GKE). They allow developers to focus on application development and leave the task of managing Kubernetes to cloud providers.

In addition, Kubernetes can be used to support scale-out stateful database platforms, particularly those in **data warehousing and analytics**. Using Kubernetes operators, a SQL-based data platform becomes a Kubernetes application that offers the extensibility, resiliency, automation and portability of Kubernetes to deliver a data warehousing solution that can scale efficiently and run anywhere.

Kubernetes allows organizations to run data warehousing in any public cloud using managed offerings from cloud providers. It also allows them to run workloads on premises through enterprise offerings such as **Red Hat** OpenShift running on **Dell** hardware.

THE NEW STACK

~~ownership of its data~~ and still have access to all the characteristics and features that made SaaS data platforms popular in the first place.

But here is the critical point: There are fewer data sovereignty and **security issues as the data and processes that manage** the data are fully owned and run by the organization, rather than the SaaS vendor.

By running the data platform stack on Kubernetes in a company-owned cloud account or data center, organizations are free to tailor the authentication, **authorization and networking configuration** to suit their own information security policies. There are no one-size-fits-all requirements, as with SaaS data platforms. Also, because SaaS providers aim to maximize profits and minimize costs, geographic regions are limited to those in which they operate, meaning that certain companies cannot use them. With Kubernetes, if the public cloud provider does not exist within a given region, a company can still deploy data warehousing workloads in its own data center and maintain a cloud-like experience in terms of elasticity. An organization can also create hybrid and multicloud deployments of a Kubernetes-based data platform depending on data locality, residency or governance needs.

Large enterprises can also take advantage of discount plans already negotiated with CSPs by running a data platform on Kubernetes managed by cloud providers. In contrast, if an organization signs up with a SaaS provider such as Snowflake or BigQuery, they cannot do this and are forced to swallow the higher consumption prices that result from the stack of margins the SaaS vendors and the CSPs below them are forced to maintain to realize a profit.

Data on Kubernetes on Steroids

By natively integrating the data platform as a first-class Kubernetes application, customers can host and run a SaaS-like data platform for its own users. However, some may have a concern about the knowledge gap between Kubernetes and the data platform itself. Why should a **business analyst** or data scientist care about the internal workings of Kubernetes when they just want to analyze data?

One solution is to ensure that the details of Kubernetes are abstracted away from end users by deploying a data platform that has a SQL interface over Kubernetes. This simplifies provisioning and managing compute resources without having to

THE NEW STACK

Kubernetes expertise enables users to install and run a data platform in their own cloud account in just a matter of minutes. With this approach, it is highly possible that an organization can become a SaaS data platform provider in just under half an hour.

*To learn more about Kubernetes and the cloud native ecosystem, join us at **KubeCon + CloudNativeCon North America**, in Salt Lake City, Utah, on Nov. 12-15, 2024.*

TNS



Mark Cusack is CTO of Yellowbrick Data, an enterprise SQL data platform company that leverages the power of Kubernetes for data warehousing, business intelligence, streaming analytics and AI/ML workloads. A former research fellow at Newcastle University, Mark has held executive...

[Read more from Mark Cusack →](#)

AWS, Google, Microsoft, Red Hat, Snowflake and Dell Technologies are sponsors of The New Stack. TNS owner Insight Partners is an investor in: Databricks.

TRENDING STORIES

1. [Bring Storage and Databases Under Kubernetes Control](#)
2. [Interconnect Security Risks to Protect Your Kubernetes Environment](#)
3. [Automating the Security of Kubernetes Clusters in the Cloud](#)
4. [Build Kubernetes Security on Intent, Not Rules](#)
5. [How To Jump-Start Your Stalled Kubernetes Migration](#)

TNS DAILY NEWSLETTER

