The Journey to Hybrid Cloud with DataStax Enterprise

DATASTAX®

CONTENTS

INTRODUCTION

Hybrid and multi-cloud computing environments incorporate infrastructure from multiple platforms and data centers and are typically a combination of on-premises resources and resources offered by one or more cloud service providers.

Hybrid and multi-cloud offer many benefits, and DataStax Enterprise (DSE) can be an important component in achieving those benefits. This white paper discusses a number of success patterns for moving to a hybrid or multi-cloud environment, and the role that DSE can play in each pattern.

HYBRID AND MULTI-CLOUD USE CASES

There are many reasons to adopt hybrid or multi-cloud. For example, an organization may want to avoid being locked into a specific cloud vendor for financial or strategic business reasons. They may also want to ensure that they avoid entire cloud outages.

An organization may also want increased agility in their development cycle. For example, they could partition their application in the cloud based on the application lifecycle. In this case, the organization would utilize services in a public cloud such as Google Cloud Platform (GCP), Microsoft Azure, or Amazon Web Services (AWS) during their development and test cycles. For production, they would then deploy their application to local on-premises infrastructure and turn off those used in the public cloud, resulting in cost savings and reduced time to market.

Hybrid cloud also enables enterprises to choose where to place workloads and data based on compliance, policy, and/or security requirements. For example, an enterprise may want to keep certain data assets on premises. Security policies could require user authentication and authorization data to be stored and accessed via on-premises servers while the content is managed via a public cloud provider, as is the case for Sony, which runs databases for individual PlayStation games in the cloud but takes care of user authentication on premises.

Accordingly, hybrid cloud allows you to put the data closer to additional cloud services an enterprise may want to leverage. Perhaps the enterprise wants to use Microsoft Azure Cognitive Services, Google Cloud Machine Learning, or both. Machine learning requires a lot of data and processing; moving the data sets from one cloud to the next isn't typically a straightforward process.

However, DSE provides the flexibility needed to seamlessly share data across disparate resources. Because of this, enterprises use DSE not only as the database of record to collect data for training purposes when building their AI models, but also as a way to store features for continued use by data science and engineering teams in both development and production.

HOW TO ACHIEVE HYBRID CLOUD SUCCESS

Over 60% of DataStax customers leverage resources from cloud service providers. In helping our users achieve their hybrid cloud objectives, we see common patterns for achieving a successful outcome.

These include:

- Using a distributed database management system (DBMS) that spans multiple clouds
- Eliminating single points of failure
- Creating a data management strategy
- Creating a security strategy
- Monitoring the database via a single view

In the following sections, we'll explore these keys to success and how DataStax addresses them.

Use a distributed DBMS that spans multiple clouds

If you have multiple database management systems in multiple clouds, then you likely have many different applications and APIs in place to access and update them. This creates an overly complex environment to use and maintain.

With DSE, you can deploy a single distributed database (one schema) spanning multiple clouds and accessed through one or more microservices. The microservices do not need to know that there are multiple clouds in place, just that they utilize a single API. In fact, the microservice could be architected for a single cloud and then later an additional cloud could be added without a major rewrite of the microservice. In other words, it's easy to add an additional cloud service provider.

Eliminate single points of failure

An active everywhere architecture such as the masterless architecture of DSE is a distributed database that has no single point of failure. It utilizes a peer-to-peer design and can include multiple data centers in a hybrid cloud.

This matters because, rather than relying on a typical master/slave architecture common with relational databases and most NoSQL databases, an active everywhere architecture has no master point that services write traffic. In master/slave environments, if the master is not available, a new master needs to be elected and write requests will be temporarily suspended while this happens.

With a peer-to-peer architecture, requests can go to any one of the database nodes. If a server disappears, requests will simply not be sent to that node. If an entire data center is not available, perhaps due to a cloud outage, requests automatically go to another data center.

It's also important to remember that having a distributed database architected to avoid single points of failure does not guarantee protection. There are still entire cloud outages and/or hardware configurations configured with a single point of failure built in. For example, a single network connection between a microservice and an application that uses that microservice could cause an outage if the connection isn't available.

The architecture must ensure that there are always multiple paths for connections between microservices, applications, and data. In the case of cloud outages, if your microservices are deployed in only one cloud, they will not be available during the outage. In a similar fashion, if your business logic is dependent on a specific management service that is only available through your cloud provider, then that management service is not likely going to be available in another cloud. (Note: In some cases, companies decide to accept that risk in order to take advantage of a great service provided by the specific public cloud.)

Create a data management strategy

Data management will be the most challenging part of your hybrid cloud project, but it's also the key to avoiding the creation of data silos as you move your applications to hybrid cloud.

It's often tempting for enterprises to forgo establishing a data management strategy, but they shouldn't. Ignoring this critical element increases the risk of creating another data silo with redundant data from other applications that need to be kept synchronized by multiple players.

For example, suppose you have several databases that contain customer information and each database allows applications to create, read, update, and delete (CRUD) customer data? To support this, there is likely a feed or sync program that keeps the customer databases in sync. In this situation, if you decide to build microservices in the cloud, it's important that you resist the temptation to create another database—and therefore another data silo—that contains customer data on which applications perform CRUD operations.

Instead, you should design microservices that perform CRUD operations. Each micro service operates on, for example, customer data in one database. This micro service becomes the conduit for updates for every other application that updates the same data in another database. There are many examples ways for doing this, but that is the subject of another white paper. The transition for an application to use micro service to obtain data updates will be different for each application. This transition has a higher impact on systems that are costly to change, such as a COBOL application on a mainframe.

If you must create a new database in the cloud, first consider the corporate data view. Identify company systems that already have the data that you plan to store in the cloud and use these systems to source the data. If your company has a conceptual data model, use it in the design of the new database.

DSE's unique masterless architecture, which is natively architected for hybrid cloud, allows you to build a consistent data layer across on-premises, hybrid cloud, and multi-cloud infrastructures. It eliminates data silos and accelerates hybrid and multi-cloud application deployments.

As a hybrid cloud database, DSE provides all the operational, or transactional, data in a single logical layer for global data availability. The data is highly distributed, always available, and highly secure, delivering the data and ingesting massive amounts of writes in real time.

With this architecture, enterprises can deploy part of the data layer on premises, part of it in one public cloud, and part of it in another public cloud. Enterprises can also move the data in and out of any cloud at any time with no downtime or rewriting of their applications.

Create a security strategy

An IT organization always needs to comply with its enterprise's security policy, and this extends to hybrid cloud.

For example, the enterprise may be using AWS in conjunction with Azure or GCP and wish to keep data associated with its retail store business unit in Azure and GCP but not in AWS.

This is easily accomplished with DSE.

DSE includes advanced security, which fortifies the database against potential harm from deliberate attacks or user error. DSE has advanced mechanisms for authentication and authorization; encryption of data in-flight and at rest; and data auditing.

DSE is also compatible with various partner security solutions to meet industry-specific and other advanced requirements. It leverages enterprise standards to integrate cohesively with existing technology such as Active Directory (AD), Lightweight Directory Access Protocol (LDAP), Kerberos, Public Key Infrastructure (PKI), and Key Management Interoperability Protocol (KMIP).

Monitor your distributed database via a single view

A best practice for the building and deployment of any application—whether simple or complex or deployed in a private, public, or hybrid cloud—is to create, test, and maintain (by keeping it up to date after going live) an operational runbook.

This extremely important document contains critical operational procedures, including monitoring, security, access control, and configuration. The operational runbook is used not just for the database but also for microservices, backup and restore, event handling of all types (such as if a node goes down or a cloud region is not available), SLAs not being met, background jobs not running, how to handle failures (in detail), and upgrades.

However, monitoring database workloads, activity, auditing, and performance metrics throughout an enterprise is very tricky and becomes even more complicated in a hybrid cloud environment, particularly if it includes public clouds.

DSE configured in a hybrid cloud allows for monitoring and distributing the database across clouds in a single view. The DSE Metrics Collector is built on collectd, a popular, well-supported, open source metric collection agent. With over 90 plugins, you can tailor the solution to collect metrics most important to your organization. When DSE starts, it automatically begins sending metrics and other structured events to the DSE Metrics Collector, where the frequency and type of metrics collected are easily configured. These metrics can aid in the identification of read and write performance latency that is outside of a defined SLA. The aggregated metrics can also be sent to many metrics visualization and monitoring tools such as Prometheus, Graphite, Splunk, and Grafana.

Of course, you can use DataStax OpsCenter as your visual management and monitoring solution for DSE. OpsCenter provides architects, database administrators, and operations staff with the capabilities to intelligently and proactively ensure their database clusters are running optimally across multiple clouds with a single view. Furthermore, DataStax Lifecycle Manager allows enterprises to easily configure and upgrade DSE in hybrid cloud environments, with everything exposed via RESTful API so that it's fully compatible with all tooling.

CONCLUSION

Most of today's enterprises are planning to deploy hybrid and multi-cloud data architectures, if they're not doing so already. But this transition often comes with drawbacks, such as data silos and data governance issues.

DSE makes it easy to fully exploit hybrid and multi-cloud environments without having to rearchitect cloud applications. A masterless architecture can provide a consistent data fabric between on-premises and cloud-based resources, without compromising availability, scalability, and security required for modern applications.

The distributed NoSQL database allows you to implement a successful hybrid or multi-cloud strategy because it:

- Offers distributed database management across multiple clouds
- Eliminates single points of failure in your hybrid or multi-cloud deployment
- Helps you optimize your data management strategy
- Enforces your security strategy throughout your hybrid cloud
- Lets you monitor your distributed database with a single view of your hybrid cloud

With these five key abilities, any enterprise can take full advantage of everything hybrid cloud has to offer.

WANT TO LEARN MORE?

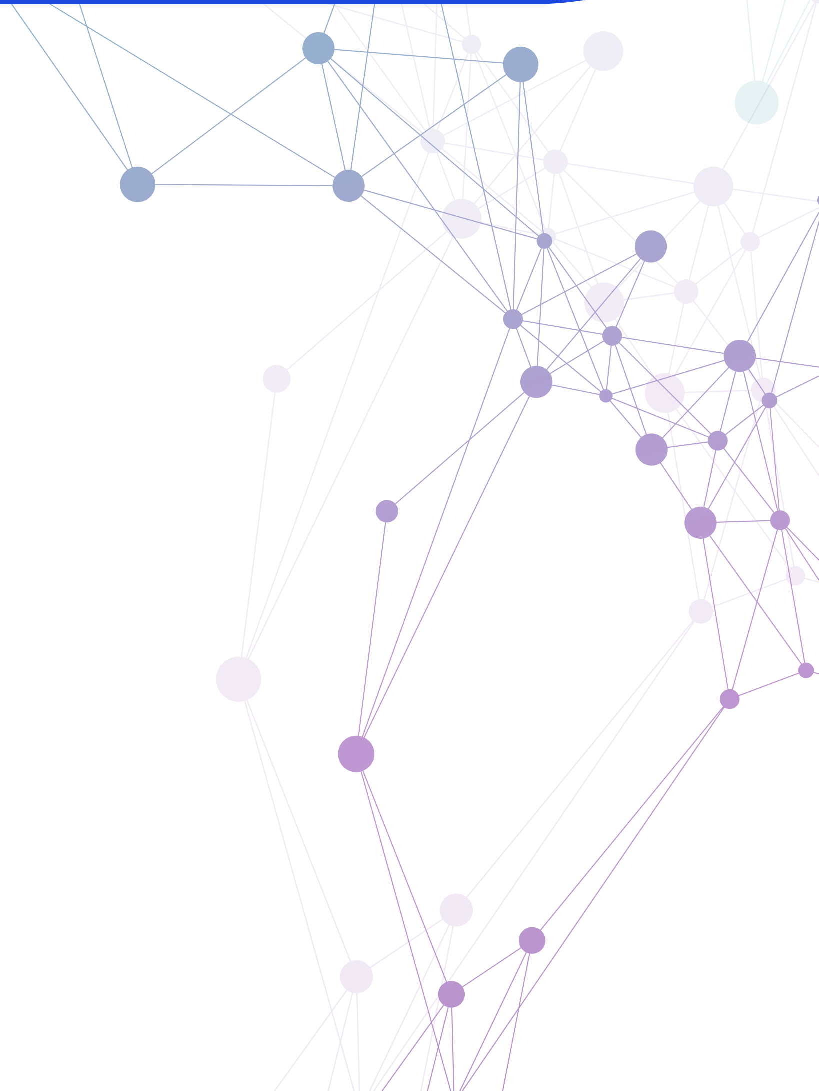
Here are some additional resources you might want to check out:

- Five Ways to Secure Your Distributed Data
- The Always-On, Active Everywhere, Distributed Hybrid Cloud Database Built on Apache Cassandra™

ABOUT DATASTAX

DataStax delivers the always-on, active-everywhere distributed hybrid cloud database built on Apache Cassandra™. The foundation for personalized, real-time applications at scale, DataStax Enterprise makes it easy for enterprises to exploit hybrid and multi-cloud environments via a seamless data layer that eliminates the issues that typically come with deploying applications across multiple on-premises data centers and/or multiple public clouds.

Our product also gives businesses full data visibility, portability, and control, allowing them to retain strategic ownership of their most valuable asset in a hybrid/multi cloud world. We help many of the world's leading brands across industries transform their businesses through an enterprise data layer that eliminates data silos and cloud vendor lock-in while powering modern, mission-critical applications. For more information, visit www.DataStax.com and follow us on Twitter @DataStax.



Last Rev: JULY2019