Evaluating Data Management Platforms for Cloud Applications

Understanding what’s required and how to be successful
CONTENTS

Introduction ................................................................. 3
The Multi-Faceted Cloud Application ........................................ 4
The Promises Vs. The Reality of a Cloud Application’s Data Management Options ........................................ 4
Data Management Requirements for a Cloud Application ................................................................. 5
  Contextual ........................................................................ 5
  Always On .......................................................................... 5
  Distributed ......................................................................... 6
  Scalable .............................................................................. 7
Additional Capabilities ................................................................. 8
  Functionally Cohesive ......................................................... 8
  Low Latency ......................................................................... 8
  Operationally Mature .......................................................... 9
Customer Examples ................................................................. 10
  eBay ................................................................................. 10
  ProtectWise ....................................................................... 10
  Capital One ....................................................................... 11
  Sony ................................................................................. 11
Conclusion ............................................................................ 12
About DataStax ........................................................................ 13
INTRODUCTION

For any business that wants to successfully compete in today’s digital economy, it is not a question of if but rather how much of their business will be done with cloud applications.

A cloud application is one with many endpoints including browsers, mobile devices, and/or machines that are geographically distributed. The application is intensely transactional (high velocity reads and/or writes), always available, and instantaneously responsive no matter the number of users or machines using the application. In short, businesses need cloud applications to deliver real-time value at epic scale.

The move towards cloud application vs traditional application delivery models is unmistakable with Gartner noting a current trend in the need for systems to support the “development of new digital - in most cases end-consumer facing – applications that need to be able to scale and transform in a similar agile and dynamic fashion as consumer-oriented social media applications.”

Not surprisingly, as modern businesses move more towards cloud computing and its application model, there is a noticeable impact on the data management solutions being used to run the business. This fact is highlighted in a Goldman Sachs technology report, which highlights the effect cloud computing has made on legacy database vendors like Oracle and Microsoft: “Again [with] this survey, spend on databases is expected to have the biggest negative impact over the next two years, as more people expect to spend less on Oracle (13% more, 52% less) and Microsoft (9% more, 38% less).”

While statistics such as these show how businesses are looking beyond traditional relational database management systems (RDBMS) with data management offerings that better handle cloud applications. But, what isn’t as obvious is the fact that these same businesses struggle with determining the data management requirements needed to meet the critical characteristics of their cloud applications.

This paper discusses the general characteristics of a cloud application and then examines the key data management requirements that make cloud applications successful.

---

THE MULTI-FACETED CLOUD APPLICATION

Today’s cloud applications are far from being one-dimensional in nature; instead they are intensely multi-faceted and include a variety of components, each with its own set of data requirements.

For example, a modern retail cloud application now includes various modules such as product catalogs, user profile management, fraud detection, recommendation engine, shopping cart, clickstream/log analysis, and others. Behind each one of these application modules are data needs which if not met can negatively impact the user’s overall experience.

Given how legacy and Web 1.0 application’s data-handling functioned in the past, the data demands of today’s cloud applications appear intimidating. For instance, a recommendation engine must track every transaction a user makes on a web or mobile application, analyze those transactions in real-time, search the user’s historical behavior, identify potential recommendations, and provide those recommendations to the users. If that weren’t difficult enough, the entire sequence described above must happen within stringent performance and data security service level agreements.

All of this data-related activity involves numerous pieces of data management technology that must be blended together to act as one – transactional, analytical, search, in-memory, and other similar engines – each of which must also run in a highly performant way without impacting the other’s speed. In other words, the system must be functionally cohesive within a single architecture.

Lastly, the cloud application's data model is increasingly becoming multi-faceted, which means a one-size-fits-all approach will no longer suffice. Certain components of a cloud application may require a tabular data model while another component might need a graph model to function properly. Add to that the desire that each model be stored in the same persistent datastore in order to avoid the headaches associated with sharding a system and trying to keep dependent data sets in sync across multiple models.³

THE PROMISES VS. THE REALITY of a Cloud Application’s Data Management Options

The data requirements of a cloud application can certainly be challenging. A lot of solutions or vendors claim to be able to meet these requirements, but when examined closely, most don’t fully deliver.

Anyone who has read about cloud computing and its applications knows that a lot of promises and claims are made, especially where data management is concerned. These promises are often times prefaced with the word ‘transparent’. Transparent elasticity, transparent scalability, transparent redundancy, transparent manageability, and so on.

The truth is there is no magic behind data management of a cloud application. While a software or cloud vendor may attempt to mask the architectural limitations of a legacy RDBMS with their own custom framework that tries its best to meet the requirements of a cloud application, it will fall short almost every time in terms of performance, scalability, and availability.

If a cloud application’s data management is going to achieve transparency <fill in the buzzword> it has to start with the right architectural foundation, one that was built from the ground up to handle a cloud application.

³For examples of this type of processing, see Gartner's report on hybrid transaction/analytical processing (HTAP):
DATA MANAGEMENT REQUIREMENTS FOR A CLOUD APPLICATION

While each cloud application is unique, there is a foundational set of technical database requirements that must be met for the business to successfully compete in its market. These include the database being:

- Contextual
- Always-on
- Real time
- Distributed
- Scalable

Legacy RDBMSs will struggle with all but the last requirement in general, and even then, they will not be able to exhibit operational maturity across the other criteria. With relational technology not being a good fit, architects and IT professionals are looking to distributed data management platforms like DataStax Enterprise, and DataStax Managed Cloud to satisfy their cloud application’s requirements.

The sections that follow examine each data management requirement of a cloud application in more detail and discuss how DataStax Enterprise (DSE) and DataStax Managed Cloud deliver the functionality, tooling, security and support required by today’s enterprises to power their cloud applications. DSE, built on the best distribution of Apache Cassandra™, is an always-on data management platform with search, analytics, tooling, administration, and security all integrated into a unified platform. DataStax Managed Cloud is DSE delivered as a fully managed, secure, white-glove service, so you can simply focus on innovation that matters most to you and your customers and offload all your operations to DataStax.

Contextual

The data flowing through cloud applications is very complex, ever changing, large in volume, and highly connected (i.e. a large number of relationships exists between data elements).

With integrated search, analytics, and graph capabilities, DSE allows businesses to make sense of large volume of data from disparate sources and deliver a highly relevant and personalized customer experience.

Always On

A cloud application’s ability to remain continuously available even in the face of the most catastrophic infrastructure outages can be tied directly back to both high customer satisfaction and revenue for the business. Depending on the business, downtime for a cloud application either equates to tens of thousands of dollars lost per minute or, in the case of some financial institutions, tens of millions of dollars. And more importantly, an unavailable cloud application could result in loss of customer trust, the impact of which is even bigger and long lasting.

While legacy RDBMS may achieve high availability via failover procedures, true continuous availability for a cloud application requires an architecture that was designed with no single point of failure.

Unlike traditional relational and NoSQL databases, DataStax Enterprise and DataStax Managed Cloud provide true continuous availability to keep your essential line-of-business applications going around the clock. Thanks to their peer-to-peer masterless architecture, in which all nodes in a cluster are equal, both DSE and DataStax Managed Cloud have
no single point of failure, allowing you to automatically add new commodity nodes to an existing cluster without having to take it down, automatically write data across multiple nodes in a cluster, and maintain as many copies of the data as the user desires. This continuous availability extends to any combination of multi-cloud region, on-premise, or multi-datacenter resources.

Real Time

A cloud application needs to handle multiple types of data, varied data streams and different workloads in a simultaneous fashion. The data complexity involved is typically required to enable the application to make immediate decisions and provide insightful responses back to a web or mobile user at transactional speed.

While an enterprise architect can traverse down the system sharding route of dividing workloads and DBMS providers onto separate systems, this approach normally leads to greater expense, overhead, complication, and slower time to market for the application. An integrated, single data management platform is preferred if it is capable of delivering real time decisions in a performant way.

DataStax Enterprise and DataStax Managed Cloud do just this by providing a full range of data manipulation and mixed workload capabilities in a tightly integrated platform. Both the platforms provide full data ingestion and interrogation across transactional, analytical, and search components so that no matter where data is added, modified, or deleted, it is synchronized everywhere and automatically maintained by the data platform.

Distributed

A cloud application presents a follow-you-everywhere data management challenge. A user wants access to their requested data 24x7 and expects it to be delivered fast regardless of their location. This is sometimes referred to as active everywhere or location independence: the ability to deliver active data everywhere a customer needs it. Having full geographically distributed capabilities built into a cloud application allows a business to quickly expand into new markets and also simplifies legal compliance with geographic data distribution (or data sovereignty) laws.

While it is widely acknowledged that the RDBMS architecture isn’t well suited for active-everywhere use cases, DataStax Enterprise and DataStax Managed Cloud were built from the ground up to handle widely distributed data that is accessed and synchronized across multiple data centers and cloud providers. Multiple copies of data can be replicated across as many locations as needed and should a specific locale go offline, client requests are automatically re-routed to the other, nearest live location.

DataStax Enterprise also fully supports hybrid cloud environments where an enterprise splits its design across on-premise datacenters and one or more cloud providers. A single DataStax Enterprise cluster can span each location, with the data platform appearing to the application as a single datastore.
Scalable

It’s not uncommon to hear developers of successful cloud applications say how they were caught off guard by how quickly their data and number of application users grew. Part of a cloud application’s success depends on its ability to scale in a linear fashion, so the application continues to deliver predictable performance regardless of the data volume or number of users.

A related point to note about scalability is that it should not come at the expense of availability. The ability to increase or decrease compute capacity should be accomplished in an online fashion so that no application downtime is experienced.

Linear scalability is something that has been demonstrated by numerous DataStax customers. DataStax Enterprise and DataStax Managed Cloud have scale baked into the core architecture, allowing additional capacity to be added in an online fashion with no disruption being experienced by application users. As additional nodes come online, DSE and DataStax Managed Cloud evenly and automatically redistribute data so incremental growth is accomplished with little to no effort.

Where planning for the future is concerned, DataStax Enterprise supplies its Capacity Service, which lets customers forecast what additional data or compute resources will be needed down the road. For example, the Capacity Service can predict when a cluster will run out of disk space or when current RAM allocations will be exhausted. Such functionality aids both IT staff and their management as the information provided by the Capacity Service allows them to easily plan and budget for future needs. DataStax Managed Cloud being a fully managed service, allows you to easily provision and scale on-demand based on your capacity needs.

In addition to the data management requirements covered above, there are several additional requirements which a data management platform should meet in order to successfully run cloud applications.
ADDITIONAL CAPABILITIES

Functionally Cohesive

No architect or operator wants to tackle a database platform that consists of a dizzying array of technologies that are clumsily bolted together. Instead, they desire an elegant, functionally cohesive database that smartly weaves together the right technologies in a way where they work together as one.

Both DSE and DataStax Managed Cloud have search, analytics, graph, monitoring, and security integrated within, which simplifies the cloud application’s data dependencies and flattens the database architecture into a single layer.

The end result is a platform that is less expensive, removes complexity and overhead, and accelerates the application’s time to market.

DataStax Enterprise and DataStax Managed Cloud also provide the ability to blend together transactional, analytical, and search commands into single requests so that what previously took three different commands on three different systems can now be more quickly accomplished in one. For example, a DSE Analytics query can utilize DSE Search syntax to create one statement sent against the database that performs both an analytics and search task at the same time. Having such capabilities in the data management platform delivers exactly the type of hybrid transactional-analytical-search functionality and decision-making abilities that cloud applications require.

In addition, as mentioned previously, cloud applications are increasingly using multiple data models for the various components that make up the application as a whole. The use of a flexible data model framework might practically equate to a tabular model being used for a couple of the application components while a graph model may be used for other components of the same application.

However, while the use of innovative functionality like flexible data model delivers very real benefits, the cloud application’s developers will want it to be operationally easy to manage and come with standard best practice functionality that includes strong security, comprehensive workload coverage (i.e. run all types of workloads – transactions, analytics, search – regardless of the underlying data model), and excellent commercial software support. All of these capabilities are found in DSE and DataStax Managed Cloud.

Low Latency

A cloud application user expects the response times (i.e. the speed at which they interact with the system and data) to be predictable no matter where they are located. They also want the wait times (or latency) to be as short as possible for both read and write operations.

For a cloud application, a scale-up design is prohibitive not only because it doesn’t support a geographically diverse user base, but also because it lacks the divide-and-conquer computing capabilities of a scale-out architecture needed to keep performance at expected levels.

DataStax Enterprise and DataStax Managed Cloud deliver short wait times and predictable performance by utilizing a number of built-in mechanisms. For example, one of Cassandra’s hallmarks is its extreme write and read performance made possible by its log-based write engine and its node-aware routing that is able to automatically direct client requests to the least busy and most closely located nodes.
Evaluating Data Management Platforms for Cloud Applications

DataStax Enterprise and DataStax Managed Cloud have search and analytics integrated in the platform, which avoids the performance slowdowns associated with sharding a system across multiple RDBMS and NoSQL providers. Moreover, since mixed workload management is native to the architecture transactional, analytical, and search workloads can be executed without competing for data or compute resources, thereby resulting in higher performance. Both DataStax Enterprise and DataStax Managed Cloud also provide an in-memory option so the read data enjoys fastest possible response times.

DataStax Enterprise also provides automated management services that ensure the system operates efficiently without involving operations team. Through its Performance and Best Practice Services, key diagnostic metrics are automatically and constantly collected, and potential problems are proactively identified along with recommendations for rectifying them. All of DataStax Enterprise’s automatic management services may be handled via the command line or visually through DataStax OpsCenter, a web-based management and monitoring solution that can be utilized from any laptop or mobile device. DataStax Managed Cloud, on the other hand, allows businesses to offload all the day-to-day administrative tasks to the experts at DataStax, allowing them to simply focus their efforts on business innovation and get new cloud applications to market quickly.

Operationally Mature

The companies behind cloud applications face the dilemma of needing to utilize innovative new technologies that give them a competitive edge in the market, while at the same time needing those technologies to exhibit the same maturity and dependability of legacy software that have been around for decades.

Solutions like DataStax Enterprise and DataStax Managed Cloud demystify the usage of these powerful, new data management technologies while also providing enterprise-ready data management for cloud applications. This ability showcases an important distinction between the open source components that comprise DSE and DataStax Managed Cloud and the platform itself.

For example, Apache Cassandra is widely known to be an innovative, fast-moving NoSQL database project driven by the open source community. On the other hand, DataStax Enterprise and DataStax Managed Cloud are data management platforms built on the best distribution of Cassandra that deliver the functionality, tooling, certification, training, and support required by today’s enterprises to power their cloud applications.
CUSTOMER EXAMPLES

Below are four examples of customers using DataStax Enterprise for their cloud applications. For other customer case studies, visit https://www.datastax.com/customers

eBay

eBay is the world's largest online marketplace, enabling the buying and selling of practically anything. eBay currently serves over 162 million active users and has 800M+ live listings.

One of the keys to eBay's extraordinary success is its ability to turn the enormous volumes of data it generates into useful insights that its customers can glean directly from the pages they frequent. To accommodate eBay’s explosive data growth (its data centers perform billions of reads and writes each day) and the increasing demand to process data at blistering speeds, eBay needed a solution that did not have the typical bottlenecks, scalability issues and transactional constraints associated with common relational database approaches. The company also needed to perform rapid analysis on a broad assortment of the structured and unstructured data it captured.

In order to address these challenges DataStax Enterprise with integrated search and real-time analytics was deployed across multiple data centers. The end result allows eBay to more cost effectively process massive amounts of data at very high velocities, and achieve far more than they were able to with the higher-cost proprietary relational system they had been using. eBay is managing a sizable portion of its data center needs—250TBs+ of storage—in DataStax Enterprise clusters. The end result allows eBay to more cost effectively process massive amounts of data at very high velocities, and achieve far more than they were able to with the higher-cost proprietary relational system they had been using.

ProtectWise

ProtectWise, a provider of the industry's first Cloud Network DVR for complete visibility and detection of enterprise threats and accelerated incident response, has revolutionized the way enterprise network security is handled, giving security professionals a real-time picture of what is happening on their network even if they might have missed the initial signs of intrusion. They do this by deploying clusters of lightweight sensors at points on a network that record network data and then ship it securely to Amazon’s cloud where the ProtectWise platform performs a variety of analyses to identify out of the ordinary patterns and anomalies.

When ProtectWise embarked on this journey to take enterprise network security to the cloud, they realized that they needed to build their application on a database platform that could not only handle the enormous volumes of high velocity, streaming, time-series data, but they needed to do so without taking a hit on performance and availability. ProtectWise realized that to meet these demands, they needed to shift their database approach to support the solution they were looking to build. ProtectWise had some previous experience with Apache Cassandra and felt that DataStax Enterprise was the distributed data management platform that would allow them to store and manage time-series data in a way that legacy systems just couldn't do. With DataStax Enterprise, ProtectWise is handling millions of transactions and writes per second and without downtime or data loss.
Capital One

Capital One is a diversified bank that offers a broad array of financial products and services to consumers, small businesses, and commercial clients. A Fortune 500 company, Capital One is one of the most recognized brands in United States.

Capital One’s applications generate huge amounts of data, the mining of which in real time became of paramount importance for making critical business decisions. Capital One was storing this data in multiple warehouses and databases and using traditional batch analysis for decision making. To meet its real-time needs, Capital One decided to build a new reporting and analytics platform to quickly process and analyze its data. The new platform also needed continuous availability.

A stringent SLA requirement on uptime and need to process large amounts of data in real time made DataStax Enterprise (DSE) a natural choice for Capital One. In addition, DSE’s reliable scalability and multi-data center replication capability made it easy for Capital One to grow data clusters, support new features, and attain a distributed environment. Furthermore, DSE Search allowed it to run flexible and customized querying.

Powered by DSE, the new reporting and analytics platform was able to kick-start Capital One’s big data intelligence by quickly and easily incorporating a year’s worth of data – 450 million rows – from a legacy Oracle database, consolidating this into far less rows (11.5 million), all handled on a six-node cluster at the lightning-fast rate of 21,000 transactions per second, and meeting Capital One’s always on requirement.

Sony

Sony Corporation, commonly referred to as Sony, is one of the world’s largest media conglomerates and a leading manufacturer of electronics, video, communications, video game consoles, and information technology products for the consumer and professional markets.

With over 70 million monthly active users, Sony PlayStation is the leader in the gaming consoles market. Sony uses DataStax Enterprise (DSE) to power: 1) PlayStation Now, a cloud-based gaming streaming service that offers video game content, 2) LiveArea, where gamers can get more information about the game, their friends’ recent activities and a view of in-game play streamed live from friends, and 3) PlayStation Plus, which is a requirement for online multi-player gaming on PS4, and allows gamers to interact with each other for insights on new games.

Sony chose DSE as the backbone data management platform to deliver these new features in PS4, which provides gamers a revolutionary social gameplay experience and reliable performance with very low latency. The read/write speeds on DSE reaches over 200K per second on average with huge spikes during peak hours. DSE’s masterless architecture and simplified operational support empower Sony to provide an always-on application experience to the PlayStation users. Thanks to DSE, PS4 gamers can now play in a socially connected environment with their network, share their status and activities, and experience multi-player games with others without worrying about performance or downtime issues.
CONCLUSION

With evidence consistently indicating the rise in digital commerce and consumerism and the demise of companies that fail to embrace the Internet and cloud economy, the importance that a business puts on its cloud applications cannot be overstated. As discussed earlier, a cloud application needs to meet certain data requirements to be successful, so it is essential to have a data management platform in place that meets those requirements.

While legacy database technology is unable to meet the key criteria needed for today’s cloud applications, DataStax Enterprise and DataStax Managed Cloud with the always-on architecture delivers a comprehensive data management layer that accelerates the ability of enterprises, government agencies, and systems integrators to power their exploding number of cloud applications. DSE and DataStax Managed Cloud well serve cloud applications that require data distribution across datacenters and clouds, through the use of a secure, operationally simple platform that is built on the best distribution of Apache Cassandra.

For more information about DataStax Enterprise, DataStax Managed Cloud, and downloads of DataStax software, visit https://www.datastax.com/products/datastax-enterprise
ABOUT DATASTAX

DataStax powers the Right-Now Enterprise with the always-on, distributed cloud database built on Apache Cassandra™ and designed for hybrid cloud. The foundation for real-time applications at massive scale, our flagship product, DataStax Enterprise, makes it possible for companies to exceed expectations through consumer and enterprise applications that provide responsive and meaningful engagement to each customer wherever they go. Our product also gives businesses full data autonomy, allowing them to retain control and strategic ownership of their most valuable asset in a hybrid cloud world. DataStax helps more than 400 of the world’s leading brands like Capital One, Cisco, Comcast, eBay, McDonald’s, Microsoft, Safeway, Sony, UBS, and Walmart transform their businesses through right-now applications focused on enterprise optimization and customer experience. For more information, visit DataStax.com and follow us on @DataStax.

DataStax is a registered trademark of DataStax, Inc. and its subsidiaries in the United States and/or other countries. Apache Cassandra is a trademark of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries.
Evaluating Data Management Platforms for Cloud Applications

Understanding what's required and how to be successful