White Paper

Breaking Free from On-Premises Constraints: Cloud Database Services from AWS

Sponsored by: Amazon Web Services

Carl W. Olofson
September 2020

IN THIS WHITE PAPER

As enterprises contemplate the move to the cloud, they are confronted with a major challenge: what to do with legacy systems. IDC research has found that "lift and shift" approaches, involving redeploying databases and applications in the cloud through virtual machines, are often cost prohibitive, but more importantly, they inhibit users from taking full advantage of the flexibility and elastic scalability of the cloud. As a consequence, some sort of migration to a more cloud-native architecture is the better option. At the same time, one may want to consider moving from a general-purpose DBMS to a more purpose-built DBMS.

This white paper examines the implications of this option, especially from a database perspective, and investigates the opportunity it represents to do something new and ultimately more in keeping with the long-range goals of the enterprise by embracing new database architectures in the cloud. It also considers the cloud database services of Amazon Web Services (AWS), showing how they may offer a way forward in this regard.

SITUATION OVERVIEW

A big wave of technical change is heading toward most enterprises as they engage in digital transformation with an eye toward moving to the cloud. These enterprises have been using proprietary relational DBMSs (RDBMSs) in their datacenters for 30 years or more and have felt constrained by the lack of flexibility in deployment and the ongoing cost of license and maintenance fees. As enterprises make plans to move database workloads to the cloud, they must necessarily ask these questions:

- Will our current DBMS meet our data management needs moving forward, or should we consider something else?
- What new or different data management challenges will we face that our current DBMS cannot address?
- Would a cloud-native DBMS, one that can leverage all the features of the desired cloud platform, be preferable to a DBMS that is less well integrated?
- Does our current relationship with our DBMS vendor serve our best interests, or is it time for a change?

The established workloads of relational databases, particularly in the online transaction processing (OLTP) realm, will remain relevant, and the RDBMSs, or other RDBMSs like them, will continue to fulfill that role, albeit in the cloud rather than on premises. On the other hand, there are application areas where a change is necessary. In this regard, many enterprises are also looking at DBMSs with well-established APIs that

are in common use, such as those of open source DBMSs, both for relational data access (SQL) and other types of access, such as key value. Adopting databases with such APIs should make the task of finding developers to build and maintain database applications easier because knowledge of such APIs is pervasive. This yields greater flexibility from a staffing perspective.

New Data Optimization Opportunities

For the most part, the incumbent DBMSs that are most commonly used by enterprises on premises have been designed to serve two classes of use cases: online transaction processing and business intelligence analytics. New classes of applications are emerging, however, that require database support. These include "edge data"-focused customer experience applications, deep and recursive data analysis that is not well served by SQL, support for artificial intelligence (AI) and machine learning (ML), complex relationship analysis using data graphs, and streaming data capture and real-time event handling. Such new workloads require purpose-built databases that feature better scalability and performance while supporting a more agile development process.

In some cases, these workloads are best served by specialized data management systems such as document, memory-optimized, key value, graph, and other database technologies that provide data management beyond the relational paradigm. But acquiring such technologies from different vendors and knitting them together can be problematic, creating complex integration projects in the short term and complicated maintenance challenges in the long term. Many enterprises are looking to rewrite their applications to embrace an API-driven microservices architecture, which also puts extra pressure on the database to perform and operate in a highly distributed manner that the incumbent system may not be able to manage.

Limitations of Traditional Data Management Licensing and Deployment

The large incumbents not only are centered on established database technology but also tend to offer two choices: either run in the traditional manner in your own datacenter or run in their cloud. The on-premises version tends to be offered under a perpetual use license and maintenance agreement that requires a significant up-front expense and continual increases in fees as your use of the metered resource (usually CPU) increases. In some cases, this agreement is enforced using periodic invasive audits meant to ensure that you are complying with the terms of the agreement.

The cloud version comes under a subscription but runs best only on that vendor's cloud platform. This means the user is put in the position of needing to put all database applications on the vendor's cloud or find a way to connect the applications to the database, which is technically problematic. A DBMS vendor's cloud platform tends to lack the breadth of features offered by the three main cloud platform vendors, so the user must often locate some applications on that platform and others on a major cloud platform and coordinate them somehow.

Breaking Down the Boundaries in Data Management

A better approach may be to start by committing to one of the major cloud platforms and then building on it. One could do this by selecting a variety of products from different vendors or from open source, each offering the specialized functionality required and integrating them, but this approach leads to operational complexity. Each vendor tends to take an approach that works well within the domain of the database in question but does not necessarily support the integration of data and applications for other database systems. This lack of interoperability leads to frustrating boundaries in data management.

Another route would be to examine the database technology offered by the cloud platform provider. Such technology should be easier to integrate and produces a simpler vendor relationship.

The Freedom and Challenges of Open Source

Some enterprises have turned entirely to open source as their route to the cloud. Because the source code of the DBMS is available, even if the DBMS is not fully compatible with the intended deployment environment, any operational difficulties can be addressed in the code. Also, the leading open source projects are well understood among developers, so finding developers who can work with the leading open source technologies, including DBMSs, is generally not a problem.

But going the pure open source route poses its own challenge. Open source solutions, in and of themselves, also tend to lack capabilities required by most enterprises in areas such as service-level agreement (SLA) guarantees of availability, scalability, and performance, as well as data security and privacy support. Also, in the absence of a professional support organization, enterprises must support themselves entirely. When problems arise, their engineers must pore over the open source code and find an answer. If the problem is in the code, they must code a solution and submit it to the open source community.

Most enterprises do not have staff with the technical depth to do this, so they contract with an open source support service. But these support services are, for the most part, also software vendors, and as one determines the level of functionality needed to use the open source DBMS in production, in a full enterprise context, with additional scalability, performance, security, and reliability, one finds that the subscription agreement has expanded to include other code that is not open source. And the price has gone up.

There is nothing wrong with this in and of itself, but if one subscribes to a different vendor's service for each DBMS, one soon has the old problem again: how to operationally integrate and optimize an application environment when each technology element is provided by a different vendor. A cleaner approach is to find a vendor that can provide support and enterprise-class functionality and features for all the databases involved across the board.

The Business Benefits of Data Modernization

Data modernization requires a review of the current DBMS technology. Is the incumbent RDBMS the best option in moving toward a future that is cloud based and involves a range of emerging database models and the workload types they support? These models and workload types can lead to a revolutionizing of the business and its processes, with better intelligence, better automation, faster time to market, and smarter decisions both within the organization and at the edge, but sticking with an incumbent that is not able to deliver the functionality required on the chosen cloud platform, especially where a range of database types is involved, can lead to operational difficulties.

To take advantage of all the cloud has to offer, one must also have dynamic scalability as well as agility in the provisioning of compute or storage resources. Such provisioning must also deliver reliable performance and the means, through platform integration, of faster innovation. All this must be delivered automatically without operational effort on the user's part and at an affordable, predictable price.
Also, many enterprises require functionality on a global scale, so the flexibility of deploying and moving workloads among availability zones around the world is critical for them. This again speaks to the need for tight integration with the underlying cloud platform and its exploitation of emerging application models such as microservices architectures.

Data Management in the Cloud

Some data operations are occasional and project based, whereas others are ongoing and require continuous availability. In the datacenter, one provisions both the same way, allocating systems in the former case that are used only from time to time and others that are scaled to the "high-water mark" of usage when data volumes are high but run at 20–40% capacity the rest of the time. In the cloud, by contrast, resources can be dynamically allocated on an as-needed basis and can scale up and down as required, ensuring that no money is "left on the table." Also, with managed services in the cloud, the staff is relieved of such tasks as patch application, systems maintenance, and database backups.

Beyond all this, the cloud offers the opportunity to introduce new data management technologies and address new workloads and integrate them with existing systems in an additive way, without the complexity that normally accompanies such projects when done on premises.

Using the Cloud Platform to Bring Differing Data Models Together

Moving to the cloud should be a way not only of running existing workloads more efficiently and affordably but also of deploying new database workloads and taking advantage of capabilities offered by the cloud platform itself to switch data management paradigms. Such paradigm switches might include moving some relational tables to documents. The cloud platform should also support adding new data management paradigms (such as are needed for streaming data processing or data graph analysis) and blend the various database operations in an environment that is managed altogether by a single cloud platform provider.

AWS Cloud-Native DBMS Options

Amazon Web Services offers a broad range of database management technologies that, together, address all the requirements described previously. The aim is to provide a transformative move that AWS calls "refactoring."

Refactoring the Databases

To really get the benefit of database cloud deployment, one must use DBMS technology built into the cloud environment with tools that enable faster innovation. Whether your need is for relational, document, graph, or other database functionality, the choice should involve technologies that enjoy cloud platform support and the commonality of open source and that offer a net reduction in license and maintenance costs. This means more than deploying an on-premises DBMS in the cloud (AWS calls this rehosting) or moving to a managed version of the same (for AWS, this is replatforming); it means choosing a DBMS technology that is built for the cloud platform and is operationally integrated with cloud services to deliver maximum efficiency and flexibility. One could simply replatform by moving data to a cloud managed form of the DBMS currently in use or similar to it, such as MySQL or MongoDB. But if new workloads or functionality are envisioned, perhaps refactoring is in order. Refactored data is not merely moved but transformed into a format and mode of operation that serves the new workloads made possible in the cloud.

AWS DBMSs are offered in a single managed environment and are optimized to take full advantage of the operational characteristics and facilities of the AWS cloud platform. These technologies are offered with APIs that support the APIs of commonly used DBMSs in the relational and NoSQL realms, which means that developers should have no problem understanding how to use them to maximum advantage.

AWS DBMSs That Refactor Your Database Workloads

Refactoring data means putting it in a format that is usable for all the workloads to be supported, both old and new. This may mean moving beyond relational databases designed for OLTP or OLAP to something that anticipates new challenges. In such cases, data will be enhanced or transformed to serve additional purposes. Sometimes, the best option is to implement a more agile approach, such as a key value store. In some cases, all the data from the on-premises system may stay together in a homogeneous format; in other cases, the data may be allocated to a variety of heterogeneous formats. The principal technologies in question include the following:

- **Amazon Aurora.** This managed relational DBMS is both MySQL and PostgreSQL compatible, provides the full capabilities of an enterprise RDBMS, and is optimized for performance, scalability, and high availability beyond what is possible with community open source relational database technology. Amazon Aurora is managed by Amazon Relational Database Service (Amazon RDS), which automates administrative tasks such as database setup, patching, and backups. It supports six open source and commercial database engines.
- **Amazon Redshift.** This DBMS is designed for analytics workloads, with a columnar architecture that stores data efficiently for fast query performance. It also has features to easily query data in operational databases and data lakes.
- **Amazon DynamoDB.** This fully managed NoSQL database service provides fast and predictable performance with seamless scalability. It is a multiregion, multimaster, and durable database for internet-scale applications with built-in security, backup, and restore. Key use cases are serverless web apps, massively multiplayer games, mobile back ends, media metadata, event-driven transactions, and microservices.

In some cases, the data may be applied to more specialized purposes. AWS offers Amazon EMR for data lake management, Amazon DocumentDB (with MongoDB compatibility) for document database management, Amazon Neptune for graph analysis, Amazon KeySpaces (for Apache Cassandra), and Amazon ElastiCache, which is compatible with Redis or Memcached, for in-memory shared operational data.

AWS Data Migration Services

AWS offers a full range of data migration services to help in breaking free from legacy environments to AWS data management in the cloud. These services, driven by self-service tools, guide clients through the process of the following:

- Assessing the data and application portfolio to determine what to move and what to transform
- Achieving overall organizational commitment to the migration, including full support from the database administration team
- Working out a plan and timeline for stepwise migrations that deliver continuous benefits while minimizing risk

In addition, there is a plethora of partners in the Amazon Partners Network (APN) and programs such as Database Freedom and Amazon Database Migration Accelerator to help enterprises on the way to an efficient, integrated, and cloud-based future.

Benefits of Cloud Transformation

Moving workloads to AWS, which AWS calls "modernization," and embracing new workloads on AWS offer significant business benefits:

- IDC research has shown that moving workloads to the cloud produces substantial cost savings in terms of hardware and software.
- In a recent research project conducted for AWS, IDC interviewed seven AWS customers that collectively reported a 264% three-year ROI and 39% lower three-year cost of operations as a result of moving database deployments to RDS.
- Moving workloads to the cloud frees valuable staff from routine operational tasks, increasing developer and DBA productivity and allowing them to perform more high-value jobs for the enterprise.
- Development costs are greatly reduced because development teams can spin up and shut down test instances at will, paying for only what they use.
- New functionality in areas such AI/ML, streaming data, edge data management, and data science, which would have been cost prohibitive on premises, yields improved efficiency and greater business opportunity for the enterprise overall.

FUTURE OUTLOOK

In the past, enterprises were bogged down with complex data integration solutions that limited their options for new uses of data and new types of data. This has to change. Workloads based on AI/ML, streaming data (including data from the Internet of things [IoT]), time series analysis, and other kinds of data and analysis must be integrated and managed together. These technologies are made necessary to address new business opportunities in such areas as supply chain optimization, a better overall customer experience powered by artificial intelligence, dynamic pricing, and logistics optimization. AWS has a jump on this, offering support in all these areas. Clearly, a unified cloud platform environment is best for fully embracing this new integrated future.

CHALLENGES/OPPORTUNITIES

In moving established IT users to this new environment, AWS faces several significant challenges. For one thing, AWS may offer customers too many choices without enough guidance. This is largely a messaging challenge, but it must be addressed. For another, many enterprises use packaged applications, and those applications may not be fully or, in some cases, even partially compatible with the AWS database offerings. If those applications are also not cloud native, moving them would not be a good option in any event. There is an opportunity here for AWS, however. Although AWS already has an active partnership program, work needs to continue in forming partnerships with cloud conversion services and with application vendors, helping them refactor their applications to operate as cloud-native and, indeed, AWS-native applications that can support AWS database systems. AWS already offers services directly and through partners that help application developers build for the AWS platform, so this should not be difficult.

CONCLUSION

Enterprises either are in the process of cloud migration or are in the midst of planning for a move to the cloud. The reasons have to do with increased efficiency, simplified operations, and overall cost savings in the areas of hardware, software, and staff time. The greater challenge has to do with choosing a migration target and calculating the costs and risks associated with migration. Many enterprises, unhappy with the financial relationship they have with their current DBMS supplier, are also looking to make a change on that front.

AWS offers compelling options in this regard. In addition to providing a range of DBMS options to serve as targets for refactoring, AWS offers data integration and database management services along with support and related technology that are optimized to work at peak efficiency on the AWS cloud platform. The company's platforms also deliver interfaces that are compatible with the leading open source DBMS technologies available, eliminating the need for application code transformations in cases where those technologies are in use and enabling users to adopt those technologies without locking themselves into a narrow, vendor-specific interface. AWS also provides a range of feature, tools, and services that ease the transformation process.

In planning for a move to the cloud, enterprises should consider the following:

- Assess the applications in the datacenter and their databases to determine what will be involved in moving them to the cloud and making them more cloud efficient.
- Determine whether to continue with existing DBMS vendor suppliers or plan to go in another direction and assess the costs and risks associated with such a move.
- If you decide to move to another DBMS, consider whether to adopt a third-party database cloud service or use one that is part of the cloud platform portfolio.
- In considering a DBMS that is part of the cloud platform portfolio, think about the simplicity and efficiency to be gained by using technologies that are designed to work together and to be supported by a single organization.
- Investigate the various DBMS options available on AWS and consider what is involved in moving to an AWS-native DBMS, bearing in mind the tools and services offered by AWS to make such a move simpler and less risky than possible alternatives.

AWS provides a comprehensive cloud platform in addition to a comprehensive set of database technology options. As such, it represents a compelling option.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com