White Paper

Achieving Cloud-Based Digital Transformation: Customer Successes with Amazon DynamoDB

Sponsored by: Amazon Web Services

Carl W. Olofson
December 2016

EXECUTIVE SUMMARY

With the radical changes in information management technology that have taken place over the past 10 years, there has come a great variety and volume of new kinds of data and new ways to deploy applications that can ensure greater customer intimacy, operational efficiency, and the ability to offer new products and services and reach new markets. All this has placed great pressure on enterprises to move from limited, fixed IT assets in their own datacenters to far more flexible and scalable technologies that demand an elastic environment. Ultimately, this elastic environment is best realized in the cloud.

The process of digital transformation involves moving to fully embrace the 3rd Platform of computing, one that embraces a cloud platform for managing data and processes in the mobile platform, social media, and the internet of Things (IoT). While the established relational database management systems (RDBMSs) may continue to serve the enterprise well for the purpose of processing back-office transactions and supporting enterprise data warehouse workloads, these new platforms and data sources demand other, more nimble, more scalable, and faster loading data management systems.

IDC, at the behest of Amazon Web Services (AWS), has undertaken to examine Amazon DynamoDB, which is one such data management system, looking at its feature set and capabilities and researching the benefits realized by its users. This research has revealed a system that is well designed to address the demands of 3rd Platform workloads, with capabilities that have pleased the customers interviewed, each indicating that it could not accomplish what it has done with the RDBMS products it had been using previously. Furthermore, it seems clear that the cloud-based nature of the product delivers a level of flexible scalability that meets the extreme demands of customers' application scenarios.

SITUATION OVERVIEW

Digital Transformation

The IT world is in the midst of a radical transformation. Advances in data acquisition and management technologies, processors, memory systems, storage systems, and communications have opened up possibilities never before considered. We have transitioned from the 1st Platform of computing, which was based on automating back-office processes using closed systems (mainframes and minicomputers), to the 2nd Platform, which opened up those processes to more employees, partners, and even customers on PCs with graphical user interfaces and with browser-based functionality on the World Wide Web, to the 3rd Platform (see Figure 1).

The 3rd Platform integrates all those processes and their data with new classes of devices (including smart mobile devices and various kinds of internet-enabled interactive monitoring and control components) collectively called the IoT and new modes of operation, including the informality of social media. Managing the data from the 3rd Platform demands performance, scalability, and a level of flexibility that was never required for internally managed data. Enterprises are confronting a mandate to transition their operations to this new platform in order to compete and thrive. This transitional activity is called the digital transformation.

FIGURE 1

IDC's 3rd Platform
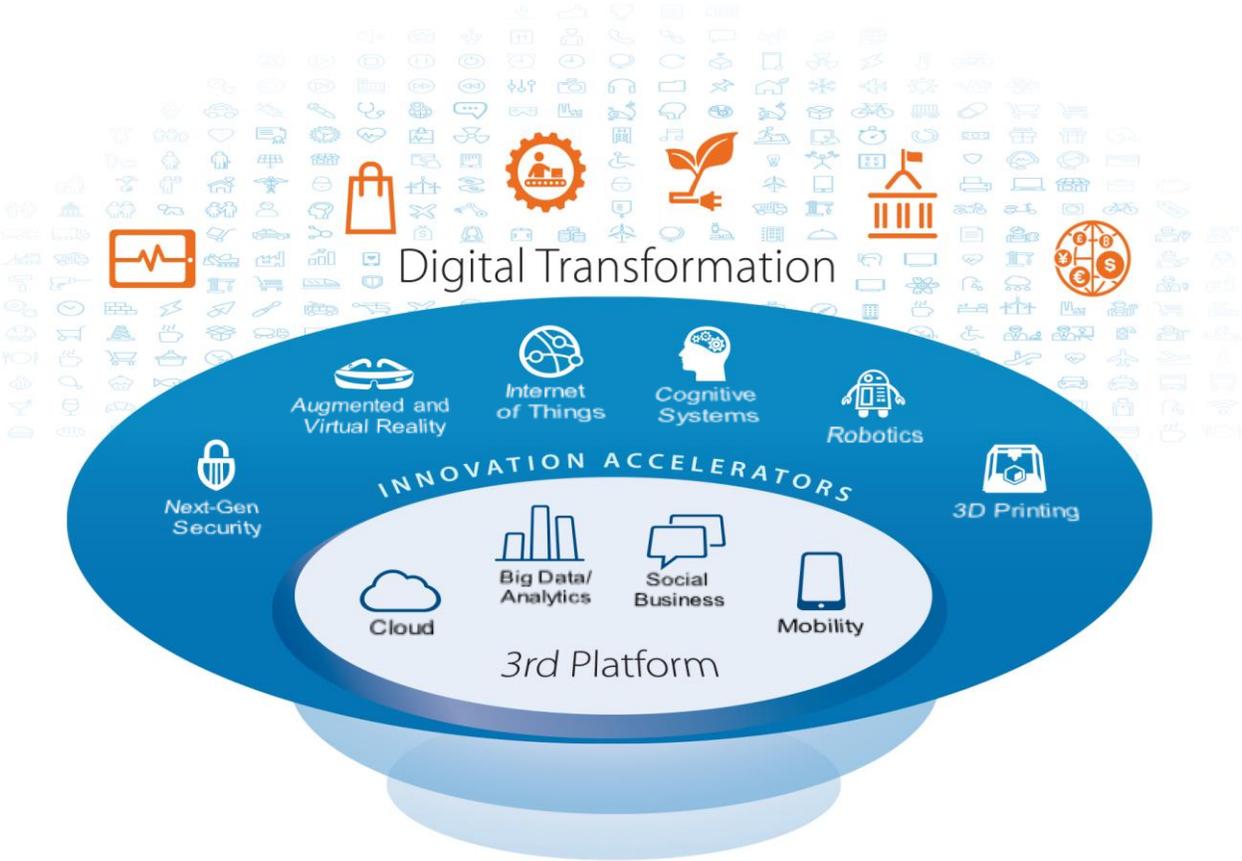
Source: IDC, 2016

New Data Management Requirements

The 3rd Platform makes new kinds of applications possible: applications that leverage a wide variety of data and are capable of establishing a previously unknown level of customer intimacy through various means of interaction like smartphones, the web, and other interfaces. These applications may absorb streaming data from various sources and integrate it with select internal data from the enterprise. The applications themselves are often in a constant state of flux, their data requirements changing day by day and week by week. The RDBMS that powers the fixed back-office processes and data warehouses of the

enterprise is ill-equipped to deliver the level of scalability and speed of data ingestion required. Also, because RDBMSs are structured and tuned based on fixed schemas, changes in the data structures can cause long delays as the databases are adjusted to support data changes.

What's required for such applications is a data management system that can operate without a schema, allowing the developers to adjust their processing as needed without a long database conversion process, and that can scale to unpredictable levels of data size. Such databases may also need to load data very rapidly, which may be done most efficiently when the data need not be parsed into a fixed table structure with the update of secondary indexes slowing the rate of ingest. Such data management systems are sometimes called "NoSQL" because they don't require the use of SQL and a schema to operate.

Benefits of the Cloud

A database that builds and grows to unpredictable sizes and that demands processing power on a highly variable basis is best deployed in a cloud environment. A cloud service is designed to deliver highly variable compute and storage resources on demand. It can do this because the virtual nature of its resources has made the cloud service highly fungible, meaning it can be shifted around at will to whichever workload requires it. The hallmarks of such a service include the ability to allocate resources on demand, to operate in a multitenant mode that ensures that no tenant either is aware of any other or affects the operations of any other, and to scale both compute and storage resources in an elastic manner, meaning resources can be dynamically scaled up or down as required.

Amazon DynamoDB

Amazon Web Services offers a NoSQL data management system called DynamoDB. This system enables developers to store data in blocks of variable size, identified by unique key values. This data can be loaded at rates of billions of records per day and can exceed just about any database in size. DynamoDB also includes native support for common document formats such as JSON.

Because DynamoDB is offered as a service on AWS, it can very efficiently acquire and release resources as required, and the user need not have any knowledge of the underlying hardware, such as server and storage systems. Also, because DynamoDB is maintained by a professional operations staff as part of the service, the user does not need to worry about database tuning, backup and recovery, or software upgrades.

CUSTOMER EXPERIENCES

IDC spoke with three Amazon DynamoDB customers to get their impressions of the technology and the service.

Telltale Games

Telltale Games is an online game service founded by people from the video game publisher LucasArts. It offers story-based games in which the user makes choices on behalf of various characters and observes what happens as a consequence of those choices. Popular games include *Batman, Game of Thrones,* and *The Walking Dead.* The company uses DynamoDB to hold the game content (elements relevant to the various places one can reach in the game story) and events (records of the choices users make and the consequences of those choices). Developers are constantly working on that story, to keep ahead of the users, and the data from DynamoDB is critical to the developers' understanding of what users are doing, and where the story should go, based on the patterns of the users' choices.

According to John Gamble of Telltale Games, who describes himself as a "back-end architect," the company tried other systems first, but they just couldn't keep up with the demand for size and speed. Gamble said, "You can put unlimited amounts of data in it." He indicated that scalability and reliability are the greatest benefits of DynamoDB.

Duolingo

Duolingo is a language learning online service. According to Duolingo, it is the leading offering in its category and the most downloaded education app on iOS and Android. The product guides the user through the process of learning a language using an intuitive and highly interactive method. Details regarding each student's progress, including scores in language exercises, are kept in DynamoDB. These details are used to predict the student's future success and make decisions as to whether to slow down or accelerate the process of learning.

Previously, Duolingo had attempted to use an RDBMS to manage this data but found that the combination of complexity in the data collection structures and the volume of data was causing it to fail. According to Severin Hacker, cofounder and CTO at Duolingo, the predictability of performance and the ability to scale have been impressive benefits of DynamoDB.

Even more important has been the nature of the cloud service on which DynamoDB runs. The worry-free nature of the service and the absence of any need for an operations staff are so significant, according to Hacker, that "the cloud is the natural habitat for any start-up."

DataXu

DataXu is a marketing cloud platform focused on delivering value to the world's biggest brand advertisers. It runs on a programmatic platform that optimizes advertising for 2 million items simultaneously. These items record *attribution,* which is an industry term referring to promotional actions that reach potential customers. Each contact is called an "impression." The impressions require acquiring space, in real time, on a website. Space on the website is auctioned through an exchange, so the ability to win business is key to competitive bidding. When the customer decides to make a purchase, this is called a "conversion event." All this data is kept in DynamoDB. These outcomes are retained for analysis, but the users are treated anonymously, and no personal information is stored. Over 2 billion anonymized user profiles are stored on DynamoDB.

According to Yekesa Kosuru, VP Engineering at DataXu, the company has pushed DynamoDB to extreme limits. He said, "We were never able to push any other systems to the levels that we pushed DynamoDB. The older system required lots of physical support. Not DynamoDB. It has delivered very consistent performance with very little downtime. We are very happy with the performance, and [DynamoDB] takes a lot of operations off our plate; works like a charm for our use case."

CHALLENGES/OPPORTUNITIES

These are early days for the cloud services business. New users and new competitors are coming online all the time. This will challenge AWS in terms of both competing with these new challengers and addressing the new workloads that are bound to arise as this space matures. AWS has the advantage of having entered this market early and has gained a compelling position as a cloud NoSQL player. Future opportunities are likely to emerge as more enterprises of all kinds come looking for scalable, flexible data management resources.

CONCLUSION

The lexicon and models of operation for enterprise IT up until recently have been based on a lexicon and model derived from the 1st Platform and 2nd Platform of computing. To address the 3rd Platform, with its demand for flexible data management and elastic scalability, enterprise IT must expand its horizons to embrace the new workload possibilities and the new technologies necessary to support those workloads.

NoSQL data management has become, and will continue to be, a key technology for addressing many such workloads, and in many, if not most, cases, such technology is best deployed in the cloud. The best of both worlds in such a case is a NoSQL data management service offered as part of a managed cloud service. DynamoDB is an outstanding example of such a data management service.

As a consequence, IT users should consider the following:

- Explore new workloads made possible by the 3rd Platform, and prepare for the digital transformation.
- In considering these workloads, be open to different data management technologies, including NoSQL technologies.
- Review existing workloads to determine whether at least some aspects of these workloads might be better served by NoSQL.
- Explore NoSQL data management options, particularly those offered as cloud-managed services.
- Consider Amazon DynamoDB as one such data management 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