Database Solutions on AWS

Leveraging ISV AWS Marketplace Solutions

November 2016
# Table of Contents

- Introduction ......................................................................................................................................3
- Operational Data Stores and Real Time Data Synchronization ...........................................................5
- Data Warehousing ............................................................................................................................7
- Data Lakes and Analytics Environments ............................................................................................8
- Application and Reporting Data Stores ..............................................................................................9
- Conclusion ......................................................................................................................................10
Introduction

Amazon Web Services has a number of database solutions for developers. An important choice that developers make is whether or not they are looking for a managed database or if they would prefer to operate their own database. In terms of managed databases, you can run managed relational databases like Amazon RDS which offers a choice of MySQL, Oracle, SQL Server, PostgreSQL, Amazon Aurora, or MariaDB database engines, scale compute and storage, Multi-AZ availability, and Read Replicas. You can also run managed NoSQL databases like Amazon DynamoDB or Amazon ElastiCache, a managed in-memory cache cloud service that offers a choice of two in-memory caching engines, Memcached and Redis. If you’re looking for a managed data warehouse, Amazon Redshift is a service that makes it simple and cost-effective to efficiently analyze all your data using your existing business intelligence tools. If your existing database is running on Amazon EC2 or on-premises and you want to take advantage of the managed database services from AWS, one of the challenges is migration of the schema and the data stored inside the database. AWS Database Migration Service (DMS) is a managed offering, that makes such migrations easier. Using DMS, customers can migrate data to and from all widely used database platforms, such as Oracle, SQL Server, MySQL, PostgreSQL, Amazon Aurora, and MariaDB.

If you want to operate your own database or database management tools, you can do so in AWS using Amazon EC2 and Amazon EBS. This will provide scalable compute and storage, control over instances, and more. Whether an organization needs a way to implement a traditional data warehousing environment or a next-generation analytics platform, AWS Marketplace has a wide variety of database solutions from popular software vendors that can be deployed quickly and easily, offering you the flexibility of the pay-as-you-go model and in some cases also allows you to bring your own licenses (BYOL). Many of the products available in AWS Marketplace can be combined with Amazon services you’re already using such as Amazon S3, Amazon EBS, Amazon CloudWatch, and AWS Lambda to help you integrate with your existing solutions.

<table>
<thead>
<tr>
<th>IF YOU NEED</th>
<th>CONSIDER USING</th>
</tr>
</thead>
<tbody>
<tr>
<td>A relational database service with minimal administration</td>
<td>Amazon RDS, a managed relational database service that offers a choice of MySQL, Oracle, SQL Server, PostgreSQL, or Amazon Aurora Database engines, scale compute &amp; storage. Multi-AZ availability, Read, Replace, and more.</td>
</tr>
<tr>
<td>A fast, highly scalable nonrelational database service</td>
<td>Amazon DynamoDB, a managed NoSQL database that offers extremely fast performance, seamless scalability and reliability, low cost and more.</td>
</tr>
<tr>
<td>An easy-to-operate in-memory cache</td>
<td>Amazon Elasticache, a managed in-memory cache cloud service that offers a choice of two in-memory caching engines, Memcached and Redis</td>
</tr>
<tr>
<td>A fast, petabye-scale datawarehouse</td>
<td>Amazon Redshift, a managed data warehousing service that makes it simple and cost-effective to efficiently analyze all of your data using your existing business intelligence tools</td>
</tr>
<tr>
<td>A database you can manage on your own</td>
<td>Your choice of relational AMIs on Amazon EC2 and EBS that provide scale compute &amp; storage, complete control over instances, and more</td>
</tr>
</tbody>
</table>

Figure 1: mapping of database solutions based on your needs.
In the past, organizations were stuck deploying on-premises hardware solutions, which meant they had to accurately predict usage, purchase hardware resources, and acquire software licenses before implementation could even begin. If they guessed too low, the performance of their applications and storage solutions would suffer. If they guessed too high, they found themselves paying for hardware that was sitting idle. With solutions available in AWS Marketplace you can quickly deploy new services on AWS on the fly, scale them as your business grows, and pay only for what you use. No longer must you wait to have hardware purchased, shipped, configured, and installed in your datacenter. With AWS Marketplace, your database solution can be up and running at the speed of a few mouse clicks.

Deploying database solutions from AWS Marketplace provides a number of benefits versus deploying on-premises hardware solutions:

- **Faster deployment**: Eliminate the lengthy process of procuring database hardware and software licenses so you can deploy in minutes. Easily spin up a proof of concept, and quickly move into production.

- **Reduced costs and risk**: Replace large, upfront hardware investments and expensive license agreements with one consolidated hourly, monthly, or yearly rate. Only pay for what you use.

- **Integration capability**: Seamlessly integrate with existing on-premises hardware solutions and incrementally migrate existing infrastructure to AWS on your timeline.

- **Business intelligence connectivity**: Leverage the business intelligence tools you already use or deploy new ones from AWS Marketplace.

- **Improved flexibility/scalability**: Additional virtual machine instances and storage can be provisioned on-demand, which eliminates the guesswork of hardware capacity planning. You can also move into new regions much more rapidly for your disaster recovery (DR) strategy and thereby increase high availability and fault tolerance.

- **Improved security**: A 2015 IDC study found that organizations can be more secure on AWS than they could in an on-premises datacenter. AWS infrastructure is designed to be optimized for security and compliance regulations across financial and government to healthcare and more, including SOC 1, SOC 2, SOC 3, and ISO27001. If your applications are built for vertical industries like finance, healthcare, government or if you have other compliance requirements, this may help in your path to gain such certification.

---

1 IDC: Assessing the Risk: Yes, the Cloud Can Be More Secure Than Your On-Premises Environment.
Building Your Platform: Choosing AWS Marketplace Seller Solutions to Augment and Enhance Your Existing Data Infrastructure

The following section goes through each component shown in Figure 3 and how they can be implemented using AWS Marketplace. It also features some of the popular AWS Marketplace solutions. There are plenty others that may be more suited for your business needs but are not covered in this document. AWS Marketplace covers 23 Server Categories today with up to 2,500+ listings from popular software vendors.

Operational Data Stores and Real Time Data Synchronization

Operational data stores traditionally store transactional data sources from one or more business systems. They offer strong transactional integrity and are capable of applying sophisticated business rules. They excel at heavy Online Transaction Processing (OLTP) workloads. Recently, a new class of these data stores has become available that are designed for real time ingestion and synchronization. They offer users the choice of utilizing more flexible data models and SQL or NoSQL querying capabilities. In addition to the capabilities of traditional methods which ingest data from transaction systems such as Point of Sale, these newer databases specialize in ingestion from mobile devices or streaming Internet of Things (IoT) applications. Real-time data ingestion also benefits applications where state is needed quickly. For example, consider a telematics device placed onto a vehicle streaming location data. This data could be needed as soon as it is available by an application, such as Uber. It could also be used for analytics, such as a shipping company looking to analyze vehicle location data to improve route efficiency. Other examples may include real-time fraud detection, inventory alerts that keep a retailer from going out of stock for a particular item, or possibly detecting device failure. Often, the main driver for such real-time analysis of the data is to provide a corrective action in response. For example, if your software can detect fraud as it is happening, you have an opportunity to prevent it. If hardware issues are identified in logs that indicate that a failure is imminent, it may be able to stop the machine before real damage is done and notify an operator immediately.

AWS Marketplace Seller Solutions for Operational Data Stores and Real Time Data Synchronization

1. **Use Case:** You need to ingest large volumes of data in real time from Mobile or IoT streaming data sources: *Couchbase Server and Sync Gateway*

Specifically engineered to work with AWS, Couchbase is an open source, scalable, high performance NoSQL database. Couchbase can be deployed as a distributed cache engine, key/value store, or document database. It is simple to deploy and scale, and has a sophisticated

---

configuration and monitoring user interface. Known for its low-latency and high throughput, Couchbase is an easy decision when considering a solution for real time data ingestion. Like most database solutions, Couchbase can also be spread across Availability Zones in order to mitigate any single point of failure, ensuring that your real time data stream is landing data without worry.

2. **Use Case: You need a transactional database system to capture OLTP transactions and implement complex business rules:** *Oracle Database 11g or Oracle Database 12c*

A time tested and mature database, Oracle DB is still a popular option for many organizations. With Oracle DB you can archive your data directly to S3 using RMAN with Oracle Secure Backup to save on backup costs. You can also significantly improve Recovery Time Objective and Recovery Point Objective by utilizing Amazon’s global infrastructure with Data Guard or Oracle Golden Gate to create ready-to-go standby databases. With Oracle DB you can quickly deploy new database workloads or migrate existing on-premises databases to AWS using existing licenses. With the press of a button you can easily scale up your database to handle whatever kind of traffic you’re seeing, and scale down to save on cost.

Both 11g and 12c options are available in Standard and Enterprise editions from AWS Marketplace.

Oracle Database12c simplifies database management for scalable cloud applications by introducing Pluggable Databases. Pluggable Databases enable multi-tenant architectures and cut down on administrative overhead with large production deployments.

3. **Use Case: You need a transactional database system to capture OLTP transactions and implement complex business rules:** *Microsoft SQL Server 2016 Enterprise*

Amazon EC2 enables you to run Microsoft SQL Server Enterprise on Windows Server 2012 R2; this is designed to provide a high-performance, reliable, cost-effective cloud computing platform. Common Windows use cases include Enterprise Windows-based application hosting, website and web-service hosting, data processing, media transcoding, distributed testing, ASP.NET application hosting, and any other application requiring Windows software. With SQL Server 2016 you can store transactional and analytical data. It also provides in-memory performance for OLTP, columnstore indexes including updateable nonclustered columnstore indexes, many newer features for operational analytics, and additional security features. In addition to Microsoft SQL Server 2016 enterprise, using solutions available from popular software vendors in AWS Marketplace, you can also run other versions of SQL Server 2016, such as Express, Standard, and Web.
Data Warehousing

A Data Warehouse contains various types of data from several different data sources which help guide an organization’s management decisions. This type of data isn’t necessarily exposed to an end-user, but is more often used by the business for analysis and reporting. On-premises data warehouses are typically appliance based, which can make expanding them expensive and time consuming. AWS Marketplace has a variety of different data warehouse solutions from popular software vendors that can be deployed in a matter of minutes.

AWS Marketplace Seller Solutions for Data Warehousing

1. **Use Case:** You need a large scale system for storing and processing large amounts of data: *Teradata Database Base Edition or Teradata Database Base+ Edition*

   One of the most popular options offered by a software vendor in AWS Marketplace is Teradata Database, which is an extremely powerful data warehouse and analytic solution that many companies employ around the world. Applications are designed to port seamlessly between cloud and on-premises solutions, so no retraining is required. The Teradata Database provides quick query response and support for complex analytics for large data sets at high levels of concurrency. One particular capability available with Teradata Database is Teradata QueryGrid, which is designed to enable you to work seamlessly across all of your data and analytics engines for no-hassle analytics. Your subscription to Teradata Database also includes rights to use the following products, which are listed in AWS Marketplace: Teradata Data Stream Utility; Teradata REST Services; Teradata Server Management; and Teradata Viewpoint (Single Teradata System). Currently Teradata Database is available in all Regions except for South America, and by launching your instances in multiple Availability Zones you can protect your application from any single point of failure.

2. **Use Case:** You need an enterprise scale data warehousing system that is compatible with most Oracle applications: *EnterpriseDB Postgres Plus Cloud Database Advanced or Basic*

   Another popular option is EnterpriseDB’s Postgres Plus Cloud Database. It is an open source-based database with high scalability with compatibility for Oracle. With EnterpriseDB you have access to an enterprise-scale Postgres database running in Amazon EC2. Since it is compatible with most Oracle applications, it is designed to allow you to preserve existing code bases and skill sets if you have on your existing applications that run on Oracle DB. Additionally, Postgres Plus includes ACID-compliant NoSQL document and key-value stores to handle diverse workloads with a single database. To reduce database management burden, it automates connections pooling, load balancing, memory scaling, and failover to other AWS Availability Zones, backup scheduling, and cluster health monitoring. Postgres Plus Cloud Database is available in both hourly and annual pricing models.
Data Lakes and Analytics Environments

Having the ability to extract insights from all of your data, whether it be structured or unstructured, can have a huge effect on the path your organization will take in the future. Vast amounts of structured and unstructured data demand new tools and new processing approaches. The good news is that employing analytics and machine learning algorithms on data in AWS has never been easier with solutions available in AWS Marketplace. Many packages offered by software vendors in AWS Marketplace feature in-hadoop filesystems which can be used to store large amounts of unstructured data. Storage for structured and unstructured data coupled with high performance querying and analytics engines result in low-latency predictions and near real time access to the kinds of insights that your organization needs to succeed.

AWS Marketplace Seller Solutions for Data Lakes and Analytics Environments

1. **Use Case**: You need a data lake with a flexible storage layer and complex analytic capabilities: *MapR Enterprise Edition*

MapR Converged Enterprise Edition Plus Spark includes 24/7 support for the MapR Converged Enterprise Edition plus support for either Apache Drill or Impala. Use MapR Standard Cluster with VPC Support delivery method to launch your cluster. This edition provides full access to the MapR Data Platform, including MapR File System, a distributed Hadoop and POSIX-compliant file system, MapR Database, a distributed NoSQL database compatible with HBase and MapR Streams, a distributed publish-subscribe messaging system compatible with Kafka. MapR-FS, MapR-DB and MapR Streams are standards-based and enterprise-class, complete with high availability and disaster recovery features. Also included are a broad range of technologies like SQL-on-Hadoop with Apache Drill or Impala, data processing with MapReduce, and YARN for resource management. With the browser-based management console, MapR Control System, you can monitor and manage your Hadoop cluster easily and efficiently.

2. **Use Case**: You need a scalable and redundant NoSQL solution for storing large amounts of data and running frequent aggregations: *DataStax Cassandra Enterprise*

Well known for its lightning fast read/write capabilities, in addition to DataStax version of Cassandra, Apache Cassandra is also available as an option in AWS Marketplace. Cassandra is an open source, highly available, and scalable database solution. Spreading the cluster throughout Availability Zones on AWS will ensure that performance will never suffer from any single point of failure. With the right configuration for distribution, replication, partitioners and the right snitch, clusters with extremely large write throughput can be deployed on AWS. It is also relatively simple to achieve multi AZ and multi Region with DataStax Cassandra. Cassandra is linearly scalable so as you add new instances you’ll observe a noticeable increase in performance. Further, the use of
Dynamic columns and distributed counters will aggregate statistics as they’re written so you won’t have to wait for a typical map reduce aggregation scenario.

Application and Reporting Data Stores

Stable backend data storage is mission critical when developing any kind of application. Unlike a traditional database, a data store is a repository that persists and manages an assortment of different types of data, including documents, images, emails, and more. Simpler objects are also held in data stores, such as metadata that enrich applications or workflows. Because application traffic can spike, the need for scalable data stores is paramount, and AWS Marketplace has a number of different options offered by popular software vendors.

1. **Use Case:** You need a simple relational data store to power a web or mobile application: *MySQL or MariaDB*

   One of the most popular data store options, MySQL and its open source counterpart MariaDB, are available for quick deployment on AWS through AWS Marketplace. MySQL has over 11 million installations and gained popularity by being a stable, fast, robust, easy to use, multithreaded SQL database server. Most commonly, MySQL has been leveraged as a backend storage database for web pages and blogs.

   MariaDB is a recognized leader in open source database solutions for SaaS, cloud, and on-premises applications that require high availability, scalability and security. In addition to MariaDB Enterprise with MaxScale, which is available as a subscription, MariaDB also provides a MariaDB Connector J database connector that supports Amazon Aurora.

2. **Use Case:** You need a scalable NoSQL document store to power a purpose built application: *MongoDB*

   MongoDB rounds out the list as a very popular NoSQL database solution. Unlike traditional relational databases, MongoDB is designed to store data as JSON-like documents with dynamic schemas. Storing certain types of data in this fashion tend to make integration with certain applications much easier and faster. When deploying MongoDB in AWS through AWS Marketplace you can take full advantage of MongoDB’s Cloud Manager. Cloud Manager allows you to focus on apps rather than ops. It automates operational tasks like deployment, configuration, and maintenance. Cloud Manager make disaster recovery very simple by providing point-and-click backup and restore ability. MongoDB also delivers major advances in the critical areas of data governance, high availability, and disaster recovery.
Conclusion

Migration to a Hybrid Cloud Architecture will help enable you to expand at your business’ rate of expansion, deliver information and solutions at the speed your employees and customers need them, and provide the database required for a variety of different types of applications.