



AWS re:Invent 2021: Data analytics and machine learning

By Jon Southby
Senior Cloud Consultant

In the last week of November, AWS re:Invent took place in Las Vegas. For HeleCloud engineer Jon Southby, it was the third time he attended the event in person and he was happy to be back in person and connect to many other builders. The event and campus were smaller this year, which reflected the reduced number of attendees, estimated at around 25,000. The event kicked off as usual on Sunday evening with Midnight Madness, which included a marching band, performers and an air guitar contest. In this post, Jon details the main announcements around Data Analytics and Machine Learning (ML).

Data analytics and ML are becoming more and more important for businesses. The importance of data, data analytics and ML can be seen from the fact that it had its own keynote with Swami Sivasubramanian. Swami explained "...if you look at what's powering the ML revolution, it is all about how customers are reinventing their business with data. Data is the underlying force that fuels the insights and the predictions that helps you make better decisions and spur completely new innovations".



Three phases

Swami spoke of the 3 phases to moving to a modern end-to-end data strategy
Modernize, Unify, Innovate



Modernize phase

The modernize phase is about moving to scalable managed databases, and using the right database for the right job, new announcements included AWS DMS Fleet Advisor for automated discovery and analysis of database and analytics workloads (in preview). Fleet Advisor collects data from all your on-premises database and analytic servers from one or more central locations without the need to install agents on every computer. This is aimed at users with a large fleet of traditional database and analytic servers looking to move them to the cloud.

In addition, Amazon RDS Custom for SQL Server - a managed database service for legacy, custom, and packaged applications that require access to the underlying OS and DB environment - will help where previously RDS SQL server could not be used due to the lack of access to the underlying operating system. Amazon DynamoDB - a NoSQL serverless database - now supports Standard-Infrequent Access table class, which helps you reduce your DynamoDB costs by up to 60 percent. And, from a performance perspective, Amazon DevOps Guru for RDS was announced - an ML-powered capability that automatically detects and diagnoses performance and operational issues within Amazon Aurora.

Unify Phase

The unify phase is bringing all your siloed data into a single place by implement a data lake that can then be your single source of truth.



The big announcement here was for AWS Lake Formation, which now supports Governed Tables, storage optimization and row-level security. Governed Tables, a new type of table on Amazon S3, simplifies building resilient data pipelines with multi-table transaction support. As data is added or changed, Lake Formation automatically manages conflicts and errors to ensure that all users see a consistent view of the data. Governed Tables monitor and automatically optimize how data is stored so query times are consistent and fast. Lastly, Lake Formation now supports row and cell-level permissions. Amazon Athena – which lets you query data in S3 using SQL – now supports new Lake Formation fine-grained security and reliable table features. And in the serverless and on-demand analytics a number of new serverless offerings were announced (in Preview) Amazon Redshift Serverless, Amazon EMR Serverless, Amazon MSK Serverless, Amazon Kinesis Data Streams On-Demand – these are great features where capacity planning is difficult and for those looking to cut costs or move to serverless.

Innovate Phase

The third phase, Innovate, presents new opportunities for savings, to innovate or personalise. Bratin Saha spoke in his leadership session about the rapid rise in investment in ML, the key drivers of ML innovation and detailed a number of new announcements around Sagemaker.



Amazon Sagemaker

The two biggest announcements for me were, Amazon SageMaker Studio Lab (currently in preview), a free, no-configuration ML service and Amazon SageMaker Canvas - a visual, no-code interface to build accurate machine learning models. Studio Lab is a great *completely free* way for anyone to start learning and experimenting with ML (with no concerns that you might get a nasty bill at the end of the month). Whilst Canvas puts ML into the hands of business users and offers a no-code solution for machine learning. Also of note is that the AWS AI & ML Scholarship Program was launched in collaboration with Intel and Udacity to help bring diversity to the future of the AI and ML workforce.

The focus of this year's AWS re:Invent was on incremental updates rather than new services. Over the previous month, we've seen over 350 announcements from AWS.