Secure Migration of Sema4's Genetic Analysis Pipeline to AWS

aws partner network

Executive Summary

Sema4 is a patient-centered health intelligence company dedicated to advancing healthcare through data-driven insights. A key part of Sema4's business is providing state-of-the-art genomic testing to hundreds of thousands of patients a year. Sequencing data requires rigorous computational analysis to identify mutations of interest to a clinician. Sema4 ran an analysis pipeline for its flagship genetic analysis application via a high-performance compute grid. Sema4 found that the on-premise grid did not provide sufficient operational resilience or scale to support its growing business. Sema4 turned to Risk Focus to facilitate the secure migration of its application, analytics workloads, and associated data to AWS.

The Challenge

1. Sema4's entire genomic analysis pipeline had to be migrated to AWS without disruption of service. Patient samples are shipped into the laboratory where they are sequenced, and corresponding raw data is delivered into the analytics pipeline to identify mutations and run quality control checks. To ensure timely reporting of laboratory results, no downtime is possible.
2. The analytics pipeline was designed to run on a legacy HPC grid using IBM's Spectrum LSF. The pipeline would have to be re-engineered to run using AWS Batch.
3. A large fleet of servers would be required to run the analytics pipeline for the thousands of samples that Sema4 processes each week. Compute costs would have to be carefully managed.
4. The data produced by genomic sequencers is large. A data lake of sequenced genomic data would have to be created to allow for the long-term storage of both intermediate and final results.

Why AWS

Sema4 chose AWS because of its rigorous security standards, high availability, and capacity to scale quickly. Furthermore, the AWS Cloud offered significant cost savings compared to the company running its own data center.

Why Sema4 Chose Risk Focus

With its AWS DevOps competency, Risk Focus had the expertise of an AWS Advanced Consulting Partner that perfectly aligned to how Sema4 wanted to re-engineer and migrate its genomic analysis pipeline. Dr. Anatol Blass, Sema4's Vice President of Scientific Computing and Information Technology, shared: "We needed to move quickly to keep up with the growing demands of our business and engaged Risk Focus because of its successful track record with migrating mission-critical systems to AWS."

Partner Solution

Risk Focus re-platformed Sema4's analytics pipeline onto a secure AWS environment that offered several functional improvements. The infrastructure footprint was expanded to support a high volume of laboratory-run genomic tests. Spot and on-demand resources were used to optimize cost. The solution further improved resilience by providing failover and load balancing. At peak workloads, the application uses several thousand cores. Disaster recovery and archiving requirements were met with S3 and cross-region replication. As part of the engagement, Risk Focus conducted a Well-Architected Review of the application and implemented patch and resource-management tools that have become part of Sema4's standard pattern for Windows infrastructure deployed in AWS.

Results and Benefits

Risk Focus successfully helped migrate Sema4's genomic analysis pipeline to AWS without disrupting the company's daily workload of sequencing. Moving to AWS allowed Sema4 to spin up as many instances as required and generate analytical results in a standard timeframe of 12 hours after the sequenced data was available. This reduced the incidence of delays in providing clinical laboratory results to healthcare providers and patients. Dr. Blass stated: "As a result of running our genetic analysis pipeline on a secure AWS environment, we can now scale to meet the needs of our business and ensure high availability of this mission-critical application with significant operational cost savings."

Scalability and cost efficiency were improved by enabling the use of spot and on-demand resources that would allow the provisioning of more servers at peak use and scaling down afterword to save costs. Resilience was improved by providing transparent failover for each critical component, and business continuity was assured by deploying to multiple Availability Zones and by performing periodic cross-region copies of all backups. Operational excellence was improved through CloudWatch dashboards and automated log analysis to provide operational visibility into the execution of multi-hour batches. These AWS-native tools make it possible to identify errors and retrigger analysis. Finally, the infrastructure deployment was automated via CloudFormation, ensuring that both test and production environments are built consistently. Deploying Infrastructure as Code reduces the risks of failed deployments in production and increases the development velocity.

About Risk Focus

Founded in 2004, Risk Focus is a consultancy solving capital markets business problems with technology and insight. The firm combines business domain knowledge, technology expertise, and disciplined process to ensure the success of the most challenging projects in the industry. Some of the largest exchanges and investment banks operate on systems built by Risk Focus. The firm's practices include Custom Application Development, Regulatory Reporting & Compliance, DevOps & Cloud, Streaming Architectures, and IT Strategy. It is a Premier Confluent Systems Integrator and an Advanced AWS Consulting Partner with DevOps and Financial Services Competencies. Clients count on Risk Focus to provide outcomes that advance their objectives on time and on budget. For details, visit www.RiskFocus.com.