

Success Story

Automating infrastructure provisioning for a maritime research firm leading to a significantly reduced cloud cost

Customer

Drewry

Country

United Kingdom






Industry

Supply Chain
Logistics

About The Client

Drewry is a leading provider of research and consulting services to the maritime & shipping industry. With over 3,000 clients across 100+ countries, it's trusted for impartial market insights, industry analysis, and advice. Drewry serves its clients through its diverse business units which include maritime research, maritime advisory, supply chain advisory, and maritime financial research.

Technology Stack

 Apache Airflow	 AmazonEKS	 AWS CloudFormation
 Amazon S3	 Amazon EC2	

Business Situation

Enterprises operating within the maritime sector often deal with vast quantities of data concerning freight rates. This data, while potentially valuable, often remains underutilized due to the sheer volume and complexity involved in its analysis, which is essential for anticipating market trends and making strategic decisions swiftly.

Drewry's vision for creating the Container Freight Intelligence Portal (CFIP) stemmed from this pressing need to empower maritime sector stakeholders with advanced data-driven decision-making tools. Recognizing the volatility of freight rates and the competitive nature of the shipping industry, Drewry sought to provide clients with a platform that could deliver real-time, accurate market insights and analytics. The CFIP was designed to harness large volumes of data to ensure that shippers, freight forwarders, and other industry players could maintain a competitive edge through access to timely freight rate intelligence.

To manage and interpret this data, Drewry employed ETL pipelines for data cleansing, data manipulation, and particularly resource-intensive benchmarking within the CFIP. The core application was executed on an EKS-based Kubernetes cluster, which necessitated manual provisioning of large EKS node groups prior to running the ETL pipelines. This approach often led to underutilized or idle cloud resources, contributing to cloud waste.

Recognizing the need to optimize infrastructure for monthly processes and reduce static load, Drewry sought a solution that would facilitate the creation of an on-demand infrastructure for ETL pipelines. This would enable a more efficient use of resources, leading to a reliable, efficient, and cost-effective infrastructure management solution, ultimately enhancing the performance and scalability of the Container Freight Intelligence Portal (CFIP).

While the CFIP promised to be a game-changer, its implementation surfaced a host of challenges that Drewry had to navigate.

The primary challenge was the sheer volume of data that needed to be processed. The analysis of such extensive datasets was not only time-consuming, taking several days to complete, but it also required significant computing power and memory resources. This led to heightened operational costs that were unsustainable in the long run. Moreover, the infrastructure in place was not optimized; servers were running round the clock, regardless of the fact that the intensive data analysis was only required once a month. This resulted in a continuous financial drain and underutilized computing resources, manifesting as cloud waste.

Faced with these obstacles, Drewry sought to partner with a technology expert that could provide the sophisticated solutions required to overcome their data processing and infrastructure challenges.

By zeroing-in on Daffodil Software as their technology partner, Drewry aimed to tap into their wealth of knowledge and innovative strategies to refine the CFIP's functionality.

The Solution

Key expectations from Daffodil Software were:

- ✓ Utilize advanced AWS services to expedite the transformation of raw data into valuable insights within CFIP.
- ✓ Optimize CFIP's data processing workflows, particularly the ETL pipelines for data cleansing, data manipulation, and benchmarking, to automate infrastructure provisioning and refine operations for enhanced performance and reliability.
- ✓ Reduce operational expenses by implementing scalable cloud solutions and managing resources to avoid cloud waste.
- ✓ Develop a strong, flexible infrastructure for CFIP, capable of handling current and future data demands efficiently, with a focus on establishing an on-demand environment that minimizes idle resources and aligns with the monthly processing schedule.

Daffodil Software's team of seasoned DevOps engineers undertook an in-depth evaluation of Drewry's existing resource provisioning framework. With a clear understanding of Drewry's need for agility and efficiency, Team Daffodil set out to create a responsive, on-demand infrastructure that could seamlessly adapt to Drewry's unique operational needs, ensuring optimal performance and cost-effectiveness for their data processing activities.

Team Daffodil implemented a comprehensive suite of solutions that involved:

✓ Automated infrastructure provisioning with AWS CloudFormation

The DevOps team at Daffodil ensured that infrastructure is created on demand whenever there is a need to run processes with high resource utilization. To automate infrastructure provisioning, AWS CloudFormation was utilized. AWS CloudFormation is an Infrastructure as Code (IaC) service that allows automating and managing resources across all AWS accounts & regions through a single operation.

✓ Comprehensive CloudFormation templates

Daffodil crafted CloudFormation templates that were capable of launching EKS node groups in various sizes to meet specific requirements. These templates were meticulously stored in Amazon S3, providing version control and the ability to download at the time of stack creation, thus ensuring consistency and reliability in infrastructure deployment.

✓ Integration with Airflow and Boto3 for streamlined operations

The AWS SDK for Python (Boto3) was utilized to integrate these CloudFormation templates with Apache Airflow DAGs. Team Daffodil designed two types of DAGs: one for creating and another for deleting the CloudFormation stacks, which included the necessary EKS Node Groups for running ETL pipelines. This allowed for precise control over the infrastructure during periods of high demand.

✓ Container orchestration with Amazon EKS

Amazon Elastic Container Service (EKS) was utilized to deploy Apache Airflow [an open-source platform used for orchestrating, scheduling, and monitoring complex workflows or data pipelines] and its components, ensuring the portability and consistent management of containerization tasks. This strategic deployment on Amazon EKS enhanced the deployment process, contributing to the high availability and scalability of the infrastructure.

✔ CI/CD pipelines for efficient and automated software deployment

Daffodil streamlined the software development and deployment processes through AWS CodeCommit and Jenkins, an open-source CI/CD tool hosted on AWS EC2. CodeCommit securely and efficiently stores application's source code, while Jenkins efficiently handles code building, testing, and deployment. This approach has significantly expedited software delivery, making it both fast and reliable.

✔ Cost optimization and processing efficiency

In a quest to reduce processing time, the Daffodil team adopted a strategy of breaking down workflow pipelines into multiple parallel executions. While this approach greatly enhances efficiency, it necessitates the use of high CPU and memory-intensive servers (EKS Worker Nodes). However, this could lead to increased infrastructure costs.

To address this challenge and optimize costs CloudFormation was utilized to help in automating the infrastructure provisioning. It provided the necessary control to the product admin to create infrastructure as per their need when they wished to run a highly intensive process once or twice in the month & delete it when the process completes. Thus aiding in cost optimization as it would reduce the need to run servers round the clock. The result was improvement in operational efficiency and resource management.

The implementation of a robust infrastructure and efficient automated software deployment catalyzed high improvements in the maritime and shipping industry's data processing capabilities. The industry achieved a remarkable 97.02% improvement in data processing time for their customers. Where earlier, processing 150,000 records once took a week, it was now completed in just 4-5 hours. The upgraded infrastructure proved to be highly scalable, capable of handling 7-8 times more datasets.

The automation of infrastructure provisioning using CloudFormation enabled customers and administrators to create and delete resources as needed, streamlining operations for memory and CPU-intensive workloads, and ensuring error-free scalability along with cost reductions. This comprehensive approach led to a significant increase in industry revenue, underscoring the success of the project.

The Impact

97.02%	Improvement in data processing time
7x	Data scalability
20x	Faster freight costing calculations

Services Used

DevOps

AWS Consulting

AWS CloudFormation

Have a software product vision in mind?

Setup a personalized consultation with our technology expert.

Let's Talk