

DiscoverX Partners with ClearScale to Build a Cloud Native SaaS Delivery Platform



Executive Summary

Founded in 2000, [DiscoverX Corporation](#) is a privately held, venture-backed company headquartered in Fremont, California, with additional offices in San Diego, South San Francisco and Birmingham, England. DiscoverX is dedicated to the development and commercialization of high-value solutions for drug discovery research.

DiscoverX is an innovative company that develops, manufactures, and commercializes an integrated suite of products and services for the global pharmaceutical and biotechnology industry, allowing customers to streamline results and improve the safety and efficacy of drugs.

The Challenge

DiscoverX Bioinformatics had been delivering their applications and services to internal customers based on client-server architecture for a number of years. Employees would install a thin client application that connected to application servers and databases hosted in the DiscoverX datacenter. With advancements in technology and cloud services, DiscoverX saw an opportunity optimize their operations. They envisioned a [cloud-based SaaS delivery model](#) that would eliminate the Technical Support overhead associated with supporting a desktop-based thin client. DiscoverX IT management costs could also be reduced by leveraging AWS services. The stability and scale of AWS was a bonus for infrastructure resiliency.

High level requirements:

- Implement cloud-based SaaS delivery model
- Reduce IT management overhead
- Reduce Technical Support overhead
- Reduce outages/increase availability
- Scalability and elasticity of services
- Recoverability/Disaster Recovery

“When we first started to explore rebuilding our analytical software, it was clear that moving to a web application would enable further collaboration with customers and partners, as well as provide a modern platform on which to continue to build in the coming years. ClearScale was selected based on their extensive, proven experience in building out scalable platforms for web applications. They provided the overall platform architecture design and build, as well as quickly became an integrated member of our DevOps team. It’s a model that has proven successful in executing our software development roadmap.”

Christopher W. Clarke, Vice President, DiscoverX Corporation

The ClearScale Solution

DiscoverX partnered with ClearScale to design and implement their new SaaS platform on AWS. New customers were directly on-boarded to the SaaS platform and existing customers needed time to migrate. ClearScale helped DiscoverX design a strategy to synchronize the legacy Oracle database with the Postgres RDS database servicing the SaaS solution. This pipeline allowed current customers to access both systems while transitioning to the new platform. This was a great opportunity for DiscoverX to build a **cloud-native solution** from the ground up. Infrastructure-as-code, DevOps, and AWS best practices were applied as foundational principals of the architecture design.

AWS services implemented:

Elastic Load Balancers (ELBs), CloudWatch, CloudTrail, AWS Config, Identity and Access Management (IAM), Virtual Private Networks (VPCs), Availability Zones (AZs), Simple Storage Service (S3), EC2 Container Service (ECS), Amazon EC2 Container Registry (ECR), Auto-scaling, Elastic Beanstalk, PostgreSQL Multi-AZ RDS, NAT Gateway, VPN Gateway

Infrastructure Design

ClearScale recommended a proven, yet simple, infrastructure design for DiscoverX. The Production VPC was striped across two AZs for both redundancy and availability of the SaaS application. Each AZ was configured with only two subnets, Public subnet and Private subnet. The Public subnet hosts the Elastic Load Balancers (ELBs), NAT Gateways, and a Bastion host when needed. The Private subnets host the EC2 Container Services (ECS) clusters running Java and AngularJS containers. A multi-AZ deployment of Postgres RDS configured with a Master in the first AZ, Standby and Read Replica in the second AZ. This RDS configuration yields redundancy and availability for the database tier. Auto-scaling provided elasticity of application containers and infrastructure (ECS) based on utilization. All application services are load balanced for High Availability across AZs. Staging and Production environments were deployed separately to provide isolation, security, and control. Building on AWS services for all mission critical components leverages all the benefits of the world class infrastructure/services and offloads IT overhead required for infrastructure management.

Automation

The automation designed to deliver the DiscoverX solution with [best practices and DevOps principals](#) includes three distinct layers. Base infrastructure is provisioned and managed using CloudFormation templates which deliver consistent, version controlled changes to underpinning services. The Elastic Beanstalk service managed the deployment of the application infrastructure components. This included creation of ELBs, ECS clusters, Docker containers, and Auto-scaling groups. Elastic Beanstalk provided a simple, effective means for centralized management of application dependent services. Once these layers were tuned and tested, Continuous Integration (CI) and Continuous Delivery (CD) workflows were integrated using Jenkins. The tool chain that developed (Code Repository -> Jenkins -> Docker -> ECS) is a proven formula for success. This solution provides flexibility to perform fully automated builds and deployment for CI, to staged deployment for CD.

To allow users to access both the legacy and new SaaS systems in parallel, ClearScale developed an Oracle -> Postgres data pipeline to keep the legacy Oracle database and Postgres RDS in sync. This was a critical component to ensure a smooth migration of customers from legacy to the SaaS environment.

Base infrastructure automation layer managed by CloudFormation:

CloudFormation is leveraged by DiscoverX for creating and managing the virtual network infrastructure and creation of Postgres RDS instances. This lays down the foundational components for the application tier and can be re-used to create subsequent VPC or deployments in other AWS Regions.

VPC and Base AWS CloudFormation Template

RDS CloudFormation Template

Application infrastructure automation layer:

Application related infrastructure is managed with Elastic Beanstalk which has great integration with ECS, Container deployment, ELBs, and Auto-scaling. It is an easy to use tool that has proven reliability and works very effectively for blue/green deployment methodologies.

- Create/Manage ECS cluster
- Create/Deploy Docker Containers
- Create/Manage ELBs
- Configure Auto-scaling

CI/CD automation layer:

Jenkins was used to orchestrate four workflows to drive CI/CD across Staging and Production environments. These workflows enable manual and automated deployments to Staging and direct or blue/green deploys to Staging and Production.

Workflow 1 – Build Image

This workflow builds a new Docker image based on the selected code branch, verifies the image is correct, then commits the image to the Docker repository. The image is ready for staging or production deployments.

Workflow 2 – Update Staging

This workflow has options for deploying the new Docker images directly to a running environment, or creating a new environment for blue/green deployment. For blue/green deployments, the workflow swaps the environments after all health checks are passed on the new deployment.

Workflow 3 – Update Production

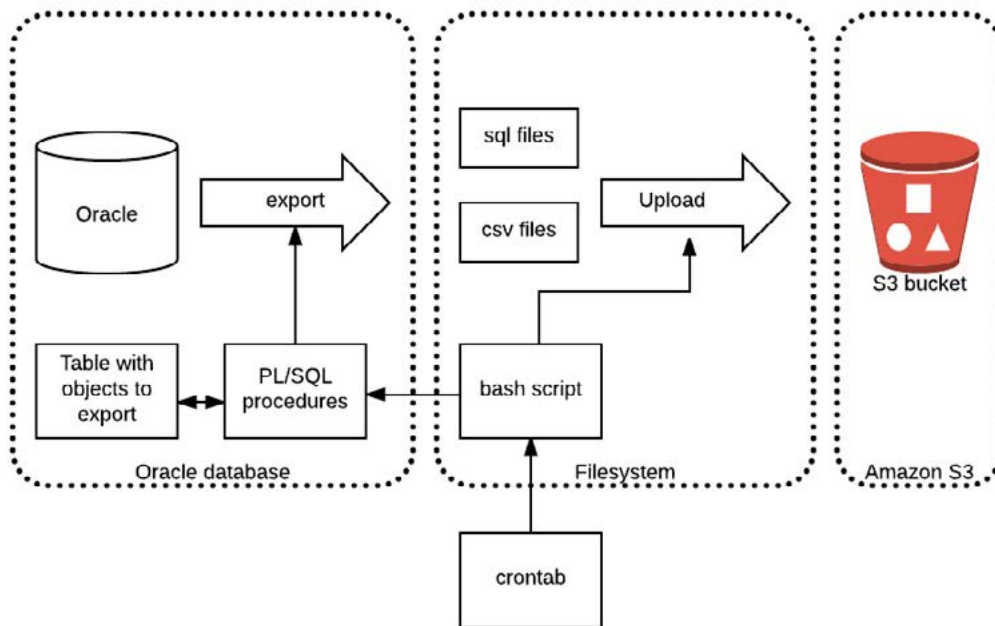
This workflow has the same basic components as “Update Staging,” but executes against the Production environment.

Oracle -> Postgres Data Pipeline

The goal of this data pipeline was to continuously sync data and schema from the legacy client/server application to the SaaS platform in AWS.

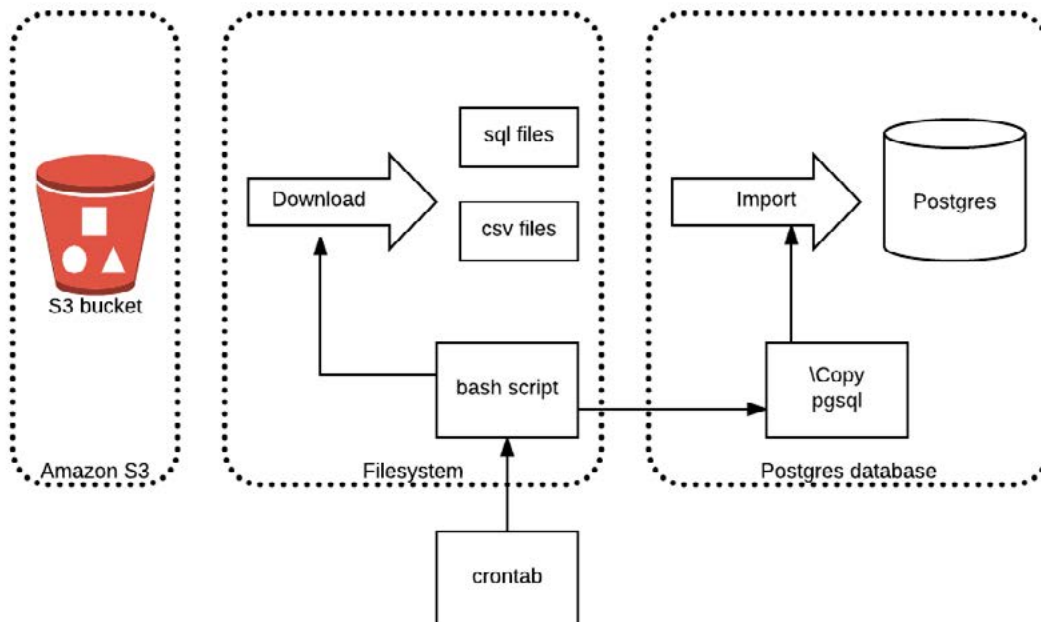
The export from Oracle was divided in two parts. The first part of the export process involved extracting data from the Oracle tables. The second part of the export process involves extracting the data from the archive logs. A custom solution was developed by ClearScale to manage the details of the two processes. The solution synchronizes the Oracle data/schema with the target Postgres RDS database.

Export overview:



The data load process is simplified since the custom data export solution stores metadata, required for Postgres import, directly in the exported files. The S3 bucket is used as a transitory data store.

Data Loading Overview:



Implementation Summary

The architecture design and implementation of this solution was the result of an iterative collaboration effort by DiscoverX and ClearScale. The high level of discussion and feedback throughout the project helped refine the initial requirements into specific implementation details. This process yielded a solution that is in full alignment with the business objectives.

The Benefits

DiscoverX now has a solid [cloud platform for SaaS delivery](#). The service reaps the benefits of High Availability and Redundancy of the AWS platform. Elastic capabilities of auto-scaling delivers cost efficiencies and enables infinite scale. Continuous integration and infrastructure automation drives quality and repeatability of deployments. By selecting AWS services, DiscoverX has also reduced IT overhead and maintenance costs.

Most importantly, DiscoverX is able to significantly reduce the time it takes to deliver new analytical products and features into the hands of their customers. This accelerated development/release cycle is elevating the level of innovation and further differentiating the DiscoverX name.