Get started

Centralized and Collaborative Data Science Platform with Domino Data Lab on AWS

One platform to accelerate data science from end-to-end, from exploring data sets to managing production models

Domino Data Lab is an AWS advanced technology partner with the AWS Machine Learning competency.

Technology Challenges

Companies looking to capitalize on the revolution in artificial intelligence (AI) and machine learning (ML) face many roadblocks as they go from developing initial concepts to deploying and managing production models. In the initial exploration phase, researchers need access to specialized hardware and flexibility in their choice of tools and approaches. As they test ideas, they need access to scalable compute power and collaboration tools. When they are ready to put models into production, they need a low friction deployment path that doesn't cost a fortune in DevOps resources and take weeks to complete. Even after models are put into production, data scientists need the ability to monitor and update production models easily. Throughout the lifecycle, there are requirements for regulatory compliance and documentation. Failure in any part of this process can amplify risk to unacceptable levels. To be successful with ML, companies need an approach that combines the process control and scalability of software engineering with the flexibility and creativity of basic research environments.

Solution Overview

To provide companies with the infrastructure they need to excel in data science, Amazon Web Services (AWS) works with leading data science platform providers to simplify infrastructure management, accelerate projects, and help focus the researchers on what they do best. AWS delivers efficient, secure, and scalable compute and storage resources to data scientists, including innovative high-performance compute instances, and sophisticated platform services like Amazon Elastic Block Store (Amazon EBS) and Amazon Elastic File System (Amazon EFS).

Organizations struggling to manage and scale data science as a core business function can rely on AWS to provide the foundation they need, and a long-term capability they can build on. Leveraging a cloud environment like AWS makes it easy to connect distributed scientific teams, build a foundation for sharing secure data, and instilling a process of validating and deploying production models that is faster and more cost effective.

Domino Data Lab – Centralizing a Collaborative Data Science Platform on AWS

Domino provides a central system of record that tracks all data science activity across an organization, and acts as an orchestration layer on the AWS compute/storage foundation. The result is a platform that all stakeholders can rely on. Data scientists get flexibility and scalable compute. IT gets visibility into, and management over, resource consumptions. DevOps gets a scalable deployment platform. Management gets a reliable, repeatable process for implementing model driven business programs.

With this joint solution, Domino and AWS deliver self-service cloud resource provisioning, enabling rapid scalability while reducing operational risk. The Domino platform includes core services, as well as user and management layers to deliver a comprehensive, customizable solution. The core services layer includes collaboration, automatic version control, and compute environment management. At the user level, the Domino platform delivers cloud based interactive workspaces, an experimentation engine, and deployment services. Governance, activity tracking, and resource controls are provided in the management layer.

Customer Ready Solutions

Discover scalable solutions that help you achieve your business needs through a combination of AWS services and APN Partners that have attained AWS Competency designations. Based on architectures validated by AWS to accelerate your cloud transformation, you can deploy solutions quickly with AWS Quick Starts and optional Jumpstart consulting offers provided by APN Partners. Visit here for more information.

Open Source and Container Technologies Help Drive Productivity

Using container and open source technology delivers tremendous productivity advantages. Teams can share environments, freeing people from managing their own packages. Researchers are free to use the analytics packages that they know and love, and can easily collaborate with those using other tools. New researchers can access current and past work on their first day, and explore historical projects without worrying about what they were built with. Leveraging Domino's reproducibility engine, past research is always available, and experiments can easily be re-run months or even years later.

This container-driven approach also makes it easy to migrate workloads onto cloud-based compute resources and specialized hardware on an as-needed basis. Data scientists can easily leverage scalable AWS compute instances like Amazon Elastic Compute Cloud (Amazon EC2) and highly parallelized P3 graphics processing units (GPUs) to run more experiment cycles per day, streamlining and accelerating the scientific process. Additionally, with this flexible solution, the journey to deployment has been streamlined. Models can be deployed quicker to production as applications or REST APIs without a costly, time-consuming DevOps process, and can be supported by AWS high availability infrastructure.

Benefits of Domino Data Lab on AWS

Accelerate Research
Streamline collaboration, sharing and communication

Increase Productivity
Scale resources and run more experiments per day

Create reproducibility
Save experiment setups, quickly re-establish test conditions

Streamline deployment
Choose from flexible production options

Reduce cost
Visualize and intelligently manage your resources

Summary

With this centralized, collaborative solution, organizations get support for a low friction, high throughput development process. Data scientists get better insights from their data and more time for research, while other stakeholders get the functionality they need to support a model-driven enterprise.

Availability in AWS Marketplace

Domino's data science platform with its Domino Model Delivery product is built to run natively on AWS, and is available as a Software as a Service (SaaS) offering now in AWS Marketplace. Learn more today at Domino Data Lab.