

Accelerating Machine Learning from Research to Production with MLOps Automation

Case Study



Company

Seagate Technology

Location

Normandale, MN

Contact

seagate.com

Industry

Data Storage

LYVELABS™
ISRAEL

cnvrg.io

Overview

Seagate Technology has been a global leader offering data storage and management solutions for over 40 years. Seagate's technology has transformed business results across sectors, powering AI/ML initiatives, modernizing backup infrastructure, and delivering private cloud solutions. Teams of leading data science professionals and machine learning engineers build advanced deep learning scripts to solve business problems and drive results for Seagate.

Challenge

Low efficiency from siloed, manual workflow and utilizing hybrid cloud resources

Like many companies, Seagate was challenged by legacy workflows that were siloed and manually intensive. Because of this development was slow and required a lot of time spent on technical tasks reducing the efficiency of their data science resources. They needed to find an optimal solution for unifying and sharing their machine learning workflows and connecting them with their hybrid cloud infrastructure.

Seagate has a global footprint with human and compute resources in many different areas as well as the cloud. The teams required a solution to help them coordinate work across their global machine learning teams and resources to maximize efficiency and effectiveness. This would ensure all teams could have access to the GPUs in their DGX and Apollo servers as well as their cloud vendors. The solution also helps build an optimal pipeline for completing the work and managing the scheduling.

At the production level, Seagate required advanced, flexible deployments that could serve on multiple endpoints like TensorFlow and Kafka. Their goal was to maximize the productivity and efficiency of their workflow, and modernize their pipelines to maximize utilization of their hybrid cloud. Seagate required a unified MLOps solution that could automate and streamline their deep learning work from research to production, operate at maximum efficiency in a hybrid cloud environment, and deliver advanced endpoints with automated updates. Seagate understood that in order to achieve full and efficient deployment of machine learning and AI models onto Edge RX in each Seagate Factory, an Enterprise Machine Learning Operations (MLOps) capability was required.

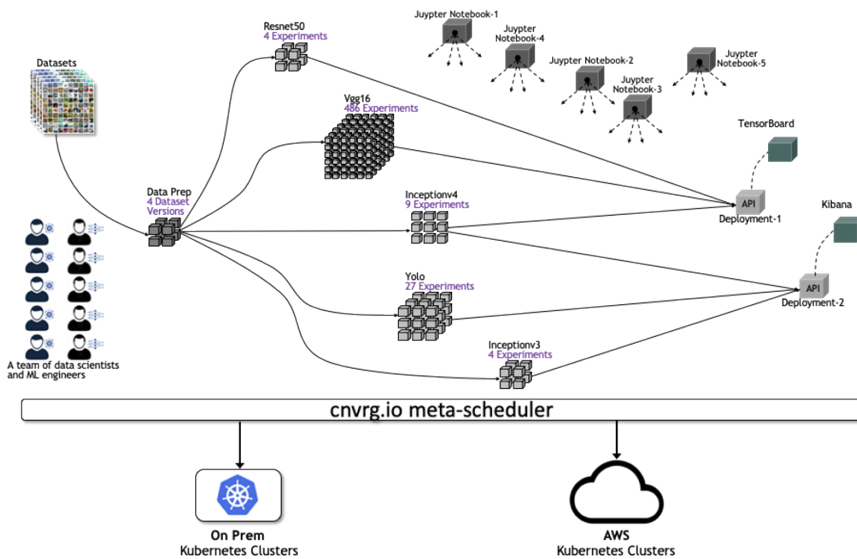
The Lyve Labs Israel team's mission is to connect external innovation in the Israeli ecosystem with these Seagate challenges. When they learned of this challenge, they set out to find a startups with potential solutions. In this way they found cnvrg.io, who offered an end-to-end machine learning platform to build and deploy AI models at scale.

Delivering automated MLOps pipelines from research to production

Cnvr.io's platform is both code-first and light code, full stack, container / Kubernetes and open. cnvr.io accelerates data science from research to production across any platform in any cloud or on-prem. A key cnvr.io concept is an ML flow that encompasses onboarding data, managing data versions, running model experiments, version control and deployments.

Seagate has on-prem DGX and Apollo 6500 machines and GPU/CPU clusters in the cloud, and needed to automate the flow components, such that the resources will be scheduled automatically, in real-time, with maximum efficiency. Cloud bursting has been incorporated – whenever the on-prem GPU machines were 100% utilized, cnvr.io scheduled more experiments in the cloud – therefore minimizing costs, and driving productivity.

cnvr.io is an MLOps platform transforming the way enterprises manage, scale and accelerate AI and data science development from research to production. The code-first platform is built by data scientists, for data scientists and offers unrivaled flexibility to run on-premise or cloud. From MLOps automation to continual learning, cnvr.io brings top of the line technology to data science teams so they can spend less time on DevOps and focus on the real magic - algorithms



Project in action: Defect detection at scale with ML

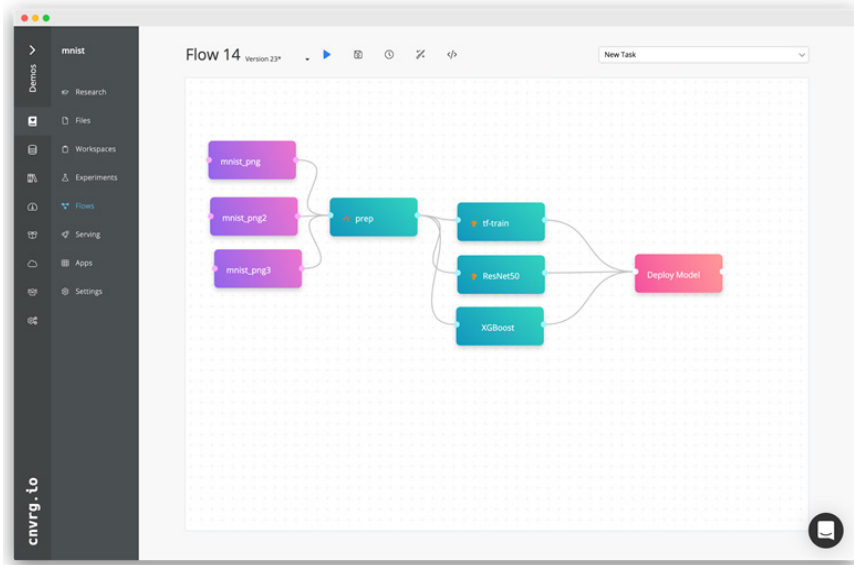
Designing an end to end flow, which will be automatically executed, is the centerpiece of MLOps pipelines. Seagate has demonstrated this with an automatic defect detection use case supporting its manufacturing process. Data can be fed directly from the manufacturing process and used for training and for real-time inference.

Seagate automated MLOps for both training and inference which can be deployed in a production system that includes distributed training on multi GPUs using Horovod and MPI backend, Kafka streaming with TensorFlow serving for real-time predictions and end-to-end model management and versioning. The pipeline runs recurrently whenever new data is ingested to keep the model up-to-date and fresh.

The Seagate team used the following functionalities of cnvr.io as part of their ML activities:

- Hybrid Cloud support - training hardware resource management for on-premises GPU servers and Cloud compute instances
- Support Cloud Kubernetes "scale to zero" for both CPU and GPU worker nodes
- Model training and evaluation with source code version control with the option to use cnvr's native git integration or Seagate's git repository.
- Model Management and reproducible experimentation - trained Model Files and artifact version control
- Model deployment - production deployments are scalable, version controlled and automated with canary rollouts.
- Application deployment – one-click web application deployment
- Global support for collaboration between Data Science Teams
- Model monitoring - inference performance
- Model retraining and updating capabilities
- Data Management - version control of datasets used to train and validate models with data movement and caching between sites over the network

An example of a data science flow is presented in the following chart.



Lyve Labs Israel, Seagate's Innovation Center was opened in February 2020 in Tel Aviv, Israel. Lyve Labs' mission is to form partnerships with external innovation in the Israeli ecosystem in order to enable innovations by providing simple, secure, and efficient ways to work with exabytes of data. The initiative empowers the seamless movement of data, optimizing its business value both in flight and at rest.

When they found cnvrg.io who offered an end-to-end machine learning platform to build and deploy AI models at scale, they thought their solution could be a good fit with Seagate's data scientists.

Results

Accelerated ML pipeline by 50% and achieved modern workflow transformation

Using cnvrg.io in this use case, Seagate has demonstrated that they are able to transform their current AI workflow into a scalable modern automated pipeline. cnvrg.io MLOps demonstrated the capability to increase a data scientist's efficiency by up to 50%. This could allow them to address 30% more business use cases by replacing days of repetitive work with one automated pipeline that maintains optimized performance. By enabling customized environments for each workload, Seagate is able to accelerate their training with an MPI to achieve the best results possible. Their modern automated pipeline delivers the ability to release and manage models in production using TensorFlow endpoints and Kafka endpoints seamlessly. cnvrg.io delivered unified data management with shared datasets, version control, data caching, querying and metadata management capabilities. This provides the potential for Seagate to run and manage hundreds of experiments in parallel on optimized compute, and support model serving with canary rollout to deliver peak performing models. With cnvrg.io MLOps platform, Seagate is able to achieve:

- Collaboration globally across advanced analytics, engineering and IT teams
- Maximized hybrid cloud node utilization with scale to zero
- Optimized on-premises hardware utilization
- Improved efficiency of data scientists by 50%
- Potential peak performance of models in production with zero downtime
- Potential decreased IT technical debt with MLOps capabilities

With these positive results, the innovation center model successfully connected a startup solution with a Seagate challenge. "The reason Seagate created Lyve Labs is because we understand that innovation cannot happen in silos," said Seagate's CEO Dr. Dave Mosley. "It's a work of collaboration. The innovators at Lyve Labs are indebted to others. In turn—drawing on over 40 years of Seagate's research and development—we want to help enable innovations that use data for the good of humanity."

"Working in a hybrid cloud environment has major advantages but can be increasingly complex to manage, especially for AI workloads. cnvrg.io has the potential to enable us to operate in a hybrid cloud environment seamlessly. The new ML infrastructure dashboard could fill a major need in connecting our infrastructure to ML projects. It provides visibility into our on-prem GPU clusters and cloud resources, paving the way for increasing the utilization and ROI of our GPUs."

Bruce King
Data Science Technologist
at Seagate Technology Advanced
Analytics Group

seagate.com

AMERICAS
ASIA/PACIFIC
EUROPE, MIDDLE EAST AND AFRICA

Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
Seagate Singapore International Headquarters Pte. Ltd. 90 Woodlands Avenue 7, Singapore 737911, 65-6412-3666
Seagate Technology SAS 16-18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00