cnvrg.io | Playtika

·SUCCESS STORY·

Playtika

Delivering real-time machine learning in production with Apache Kafka

ABOUT OUR CUSTOMER

Playtika

Playtika Holding is a leading gaming company with over 30 million monthly active users playing its titles.

INDUSTRY

Gaming/Entertainment

USE-CASE

Real-time Streaming Endpoints

RESOURCES

- AWS EC2
- CPU Cluster
- DGX-1
- Bare Metal Servers

TOOLS

kafka

kubernetes

TensorFlow

Spark

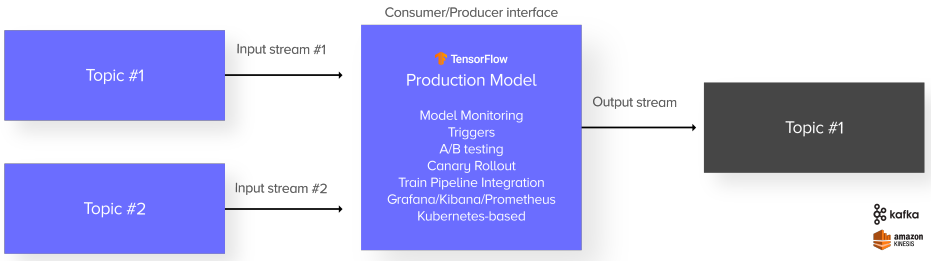
OVERVIEW

Playtika is a leading Game-Entertainment company that provides audiences around the world with a wide variety of games based on quality, and personalized content. Playtika uses massive amounts of data to reshape the gaming landscape by tailoring UX based on in-game actions. With over 10 million daily active users (DAU), 10 billion daily events and over 9TB of daily processed data, Playtika is able to provide its scientists the data they need to create a continuously changing and adaptive game environment for its users based on their ingame behavior, earning them top-grossing games for the past 5 years.

CHALLENGES

SCALING PRODUCTION OF REAL-TIME MACHINE LEARNING

Playtika's large-scale ML processes gather millions of users and thousands of features every minute. The engineering team deployed through batch, and web services, but neither solution could scale to Playtikas needs, or produce predictions in real-time. REST APIs in their ML Pipelines caused service exhaustion, client starvation, handling failures and retries, and performance tuning of bulk size for batch partitioning. Playtika's event-driven ecosystem required a solution that could support real-time streaming of their production models and scale without downtime to ensure the best player experience. Playtika's models also required various integrations including Airflow and Spark processes which would gather features for millions of players and then send requests and events for the model to predict. As demand to answer business needs with ML models increased, Playtika's AI Research Engineering teams found themselves in need of a robust infrastructure that scales horizontally and can handle bursts, peaks, and fast creation of new ML Pipelines. They needed a solution that integrates quickly and can easily handle all types of deployments including batch, REST API and Kafka streams. That is when they reached out to cnvrg.io as an all purpose deployment solution, that could scale and monitor all models in production.



With cnvrg.io we were able to increase our model throughput by up to 50% and on average by 30% when comparing to RESTful APIs. cnvrg.io also allows us to monitor our models in production, set alerts and retrain with high-level automation ML pipelines

Avi Gabay
Director of Architecture
at Playtika

FEATURES

- MLOps for model-training
- Experiment tracking
- Flows
- Notebooks
- Collaboration
- Model Serving

FIND OUT MORE

Website: **www.cnvrg.io**
Twitter: **@cnvrgio**
Blog: **https://cnvrg.io/blog/**
Email: **hi@cnvrg.io**

SOLUTION

ONE-CLICK STREAMING ENDPOINTS WITH APACHE KAFKA

cnvrg.io AI OS was the perfect solution to handle Playtika's experiments, scaling and deployments. With cnvrg.io, Playtika was able to execute real time predictions with advanced model monitoring features (logs, a/b testing, canary deployment, integration with flows, continual learning and more). cnvrg.io organizes every stage of Playtika's data-science projects, including research, information collection, model development, and model optimization at scale and is used to unify their ML workflow. Using cnvrg.io's MLOps solutions, Playtika was able to bridge the work between their Data Scientists and ML Data Engineers and continuously write, train and deploy machine learning models to various stages (i.e. staging, production) in one click. cnvrg.io delivered a scalable solution for streaming endpoints with Apache Kafka, leading to a massive increase in successful throughput, and little to no latency. cnvrg.io streaming endpoints provided Playtika with:

- Event-at-a-time processing (not microbatch) with millisecond latency
- Exactly once processing
- Distributed processing and fault-tolerance with fast failover
- Reprocessing capabilities so you can recalculate output when your code changes
- Enabled Kubernetes backed autoscaling for Kafka streams
- Bring Kubernetes Pods to their most optimal deployment

RESULTS

cnvrg.io delivered a simple, and quick solution to Playtika's unique ML production challenges. cnvrg.io reduced technical debt in Playtika's workflow, and connected data scientists from engineering teams. cnvrg.io MLOps solutions, enabled Playtika's engineering team to easily deploy, update and monitor ML in production to ensure peak performance, and reduced complex triggering and scheduling as data came in. Their data scientists are able to visualize results and business impact in real-time in a unified platform. With multiple deployment options, cnvrg.io offered a one click deployment solution which resulted in frictionless deployment solutions for every model in production. cnvrg.io enabled Playtika to instantly deploy with Kafka and easily integrated into their existing system. Playtika's ML services perform better than ever, with native integration with Kubernetes and Apache Kafka allowing them to successfully handle any spike in incoming demand and predict and handle workloads and scale consistently and linearly by adding more pods. cnvrg.io streaming endpoints solution increases successful throughput, and reduces latency/errors to zero compared to RESTful API's.

- Increased performance by 40%
- Gained up to 50% increase in successful throughput
- Reduced latency and error rates to zero
- Handled real-time predictions with zero downtime and zero batching
- Maximized model computing performance
- Reduced divergence from 5TPS to 0.5TPS compared to Web Services
- Gain better model computing performance with single thread/process run on single CPU