

AI/ML

AI&MLOps Platform

Kubernetes-based Machine Learning Platform

AI&MLOps Platform automates repetitive work in the overall pipeline for development, learning, and deployment processes of the machine learning model. AI/MLOps¹⁾ environments based on the machine learning platform are provided, offering integrated management of learning data and models as well as operational data.

Cloud Native MLOps Environments

AI&MLOps Platform offers ML model development environments optimized for cloud, enabling Kubernetes-based linking with various open source software.

Usage Convenience for Big Data

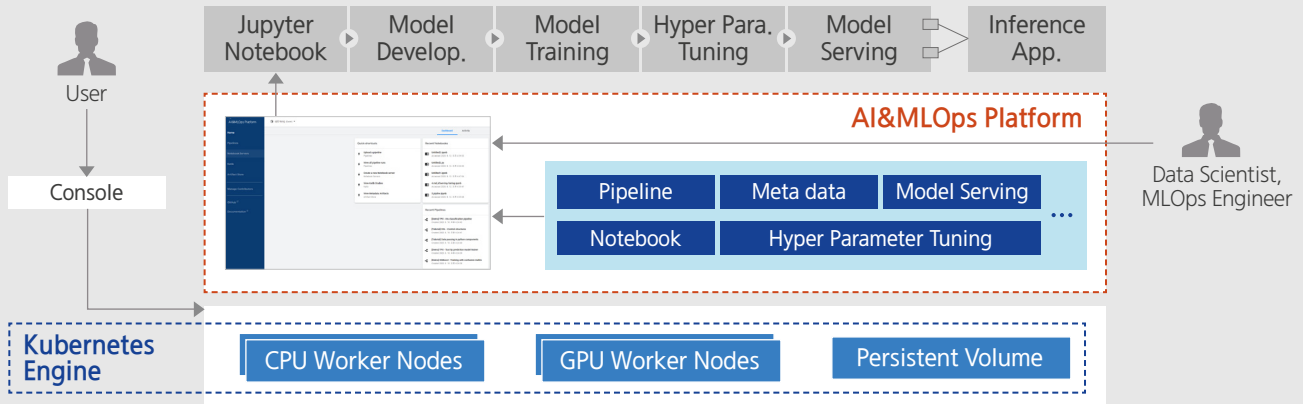
The standardized environments support a range of machine learning frameworks from TensorFlow, PyTorch, scikit-learn, and Keras. The pipeline for the entire development, learning and deployment processes of machine learning models are automated to ensure simple configuration/creation as well as reuse of the models.

Add-on Features

AI&MLOps Platform provides various features for configuring MLOps environments, including distributed learning job execution and monitoring, inference service management and analysis, and job queue management. Users can also enjoy job schedulers(FIFO, Bin-packing, and Gang-based), GPU fraction, GPU resource monitoring and more add-on features for efficient GPU resource utilization. In particular, a BM-based multi-node GPU and GPUDirect RDMA(Remote Direct Memory Access) help achieve faster processing for large language model(LLM) and natural language processing(NLP).

¹⁾ MLOps : A ML engineering discipline that aims to unify machine learning development(Dev) and machine learning system operation(Ops)

Service Architecture



Key Features

- **Basic function**
 - Create AI platform(auto-deployment/configuration), view(platform version, resource status), and delete
 - Provide Jupyter Notebook : Model development, learning, inference
 - Automate machine learning pipeline workflow
- **Additional feature(Available on AI&MLOps Platform)**
 - Advanced AI/ML platform dashboard
 - AI/ML notebook server : Base image, user-defined image
 - AI/ML job : Job creation, template, archive, scheduling, execution, monitoring
 - ※ Support GPU resource monitoring, GPU fraction
 - ※ Providing job operator for Large Language Model training(DeepSpeed)
 - Build and manage user image
 - AI JumpStarter and ETM(Experiment Tracking Management)
 - Serving : Dashboard, register/manage model, inference, predictions visualization
 - Managing platform resource : Manage resource usage by project, monitor resource usage
 - Manage project user/permissions, admin feature, adjust platform configuration

Pricing

- **Offering**
 - SW packaging for configuring AI&MLOps environments
- **Billing**
 - Charged by the hour for the scale and usage of deployed AI&MLOps Platform
 - ※ Samsung Cloud Platform for user environment configuration charged additionally

※ This service utilizes Kubeflow, an open source machine learning tool.

FOR MORE INFORMATION

SAMSUNG SDS

www.samsungsds.com / cloud.samsungsds.com
contact.sds@samsung.com / scp_sales@samsung.com
youtube.com/samsungsds

