What is MLOps for IoT Edge?

7 minutes

MLOps (DevOps for machine learning) enables data science and IT teams to collaborate and increase the pace of model development and deployment via monitoring, validation, and governance of machine learning models. Your Enterprise can avail of a number of benefits by deploying MLOps:

- **Increasing the value of your model**: Unlike code, models will deteriorate over time because of data drift unless maintained. By deploying MLOps, you can realize the value of your models and retain that value over time by keeping the model up to date by retraining.

- **Integrating the developer and data science workflow**: MLOps aims to integrate the developer workflow (driven by DevOps) into the data scientist's processes. If the developer and data scientist workflows aren't integrated, it can be expensive and unproductive for the Enterprise. The workflow of a data scientist begins with experimentation. Data scientists first spend much time analyzing features, shaping data and determining the best algorithm. Only after this experimentation stage, do they build models. After the data scientist creates a viable model, they start to engage with the developers for integration of that model into an app. In contrast, the developer workflow begins by writing code and checking in the code into the DevOps pipeline. For developers, DevOps is the preferred way to manage application life cycles. DevOps provides a pipeline of version control, test, build, and deployment tools. Ideally, DevOps also provides a fully automated CI/CD pipeline typically through a GitHub repository. By creating a single workflow for data scientists and developers, you increase collaboration and improve productivity.

- **Deployment considerations**: By implementing MLOps, you can create an end-to-end pipeline that can support the deployment of new models to different kinds of platforms. You can deploy using formats like PNNL or ONNX using APIs. You can deploy to a range of target platform such as IoT Edge or to specialized hardware such as FPGAs.

- **Simplified configuration:** You can incorporate secrets management and service endpoint management to simplify configuration of models.

- **Increasing workflow efficiency:** By adopting MLOps, practices like CI/CD you can increase the efficiency of your model deployment workflows. You can increase the rate of

experimentation and development of models and at the same time maintain the quality.

- **Register and track your ML models:** You can register models, capture governance, log data for models, audit changes to models, store the run history of a model, notify, and alert on events in the ML lifecycle (ex: data drift detection) so on.

- **Automate the end-to-end ML lifecycle with Azure Machine Learning and Azure Pipelines**: You can turn your training process into a reproducible pipeline. You can use ML pipelines to stitch together the various steps involved in the model-training process such as data preparation, feature extraction, hyperparameter tuning, model evaluation so on. These steps can be reproducible as a unit (as a single pipeline).

- **Package and debug models:** Models are packaged into a Docker image for deployment. You can debug problems with the image locally before deployment.

- **Validate and profile models:** You can profile your models to determine ideal CPU and memory settings for deployment.

- **Continuous retraining:** Using pipelines, you can establish continuous retraining of your models based on certain criteria such as a specified retraining schedule availability of new data or data drift.

Next unit: How MLOps works for IoT Edge

Continue >