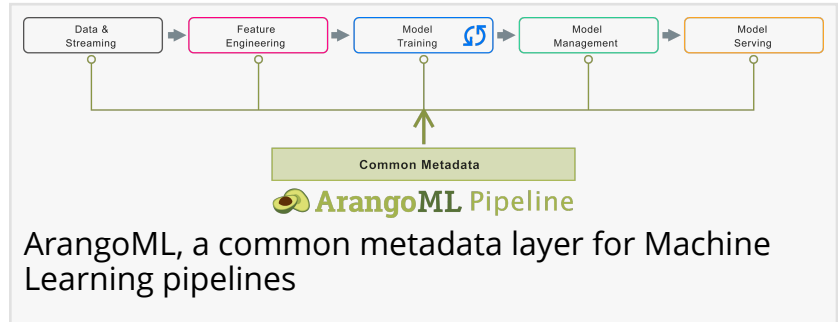# ArangoDB Extends Open Source Solution with ArangoML Pipeline

First Multi-Model Common Metadata Layer for Machine Learning Platforms

SAN FRANCISCO AND COLOGNE, GERMANY, October 2, 2019 /EINPresswire.com/ -- ArangoDB, the leading open source native multi-model database, today announced the release of ArangoML Pipeline, the first multi-model metadata layer for Machine Learning (ML) pipelines, an

ArangoML, a common metadata layer for Machine Learning pipelines

open source project that provides a common metadata layer for production-grade data science and ML platforms. ArangoML Pipeline is the first offering in ArangoDB's new extension, ArangoML.

> ArangoML Pipeline is the first solution on the market that captures, analyzes and monitors any kind of metadata, answers arbitrary complex questions and works for any kind of pipeline setup."
>
> *Jörg Schad, Head of Engineering and Machine Learning, ArangoDB*

The need to monitor, manage and audit ML pipelines is a common challenge. Machine learning pipelines are all different depending on the project and the company. They contain a number of different components -- from distributed training and Jupyter Notebooks, to hyperparameter optimization and feature stores. Most of these components produce metadata. For DataOps teams, it is critical to have a common view across their ML production platforms to answer questions such as, which models have been derived from which dataset, how can I reproducibly rebuild a particular model and more.

ArangoML Pipeline centralizes the metadata produced across the entire pipeline, allowing data scientists to have a history of how the ML models they are writing are trained and perform over time. As a native multi-model database, ArangoDB can easily accommodate and unite unstructured, highly-interlinked data, such as inference and model descriptions. This not only helps data scientists more easily access data that allows them to better optimize their ML models, but also it helps companies in highly-regulated industries meet auditing requirements, such as risk management or insurance incident handling. For example, consumers in many countries have the legal right to understand why their loan application or insurance claim has been declined. For ML- and AI-based case processing and risk analysis, enterprises need to provide the detailed audit trail which the common metadata of an ML pipeline can provide.

"In most machine learning production scenarios, Data Scientists and DataOps not only want to build a single accurate model, but also have a pipeline where they can build, rebuild and serve multiple machine learning models," said Jörg Schad, Head of Engineering and Machine Learning at ArangoDB. "The metadata produced by these pipelines is often overlooked but highly valuable in terms of, for example, finding lineage and audit information, as well as optimizing the model serving policy. ArangoML Pipeline is the first comprehensive solution on the market that

captures, analyzes and monitors any kind of metadata, answers arbitrary complex questions and works for any kind of pipeline setup."

As a native multi-model database, ArangoDB is a natural fit for ML use cases which involve unstructured data, but also the need to track the relationships between those different entities. ArangoDB unites graph, document, and key/value data models, along with a full-text search engine, natively in a single C++ core with the same query language. By uniting multiple data models in a single database, ArangoDB simplifies the process of accessing different data models, finding connections between them, and extracting value out of them -- which is what ML is all about.

Learn more
- To get started with ArangoML Pipeline, visit the GitHub repository: https://github.com/arangoml/arangopipe
- For more details on ArangoML, read the blog: https://www.arangodb.com/2019/09/arangoml-pipeline-common-metadata-layer-machine-learning-pipelines/
- To join a webinar for a more in-depth overview of ArangoML with Jörg Schad, ArangoDB Head of Engineering and Machine Learning, register here: https://www.arangodb.com/arangodb-events/from-data-to-metadata-for-machine-learning-platforms/

About ArangoDB
One database, one query language, and three data models. With more than 7 million downloads and over 8,000 stargazers on GitHub, ArangoDB is the leading open source native multi-model database. It combines the power of graphs with JSON documents, a key-value store, and a full-text search engine, enabling developers to access and combine all of these data models with a single, elegant, declarative query language.

Simplifying complexity and increasing productivity is the mission of ArangoDB Inc., the company behind the project. Founded in 2014, ArangoDB Inc. is a privately-held company backed by Bow Capital and Target Partners. It is headquartered in San Francisco and Cologne with offices and employees around the world. Learn more at www.arangodb.com.

Robyn Fernsworth
Reidy Communications for ArangoDB
+1 415-412-0300
email us here
Visit us on social media:
Facebook
Twitter
LinkedIn