Domino doubles down on flexibility, openness and model monitoring for enterprise data science

September 16 2021

Krishna Roy
The company has been busy cementing a longtime strategy to provide large organizations with a platform that enables them to preserve investments already made in programming languages, tools, compute environments and infrastructure employed for data science. Domino is also reinforcing its model-monitoring capabilities in support of enterprise MLOps.
Introduction

Domino Data has been busy cementing a longtime strategy to provide large organizations with a platform that enables them to preserve investments they’ve already made in programming languages, tools, compute environments and infrastructure employed for data science. The company is also reinforcing its model-monitoring capabilities in support of enterprise MLOps. Domino is delivering flexible and open-code-first data science to teams of data scientists in large organizations so they can use its platform to create, train and deploy models into production, as well as keep them up to date – a focus that is reflected in the company’s latest two releases.

THE 451 TAKE

Domino’s longtime commitment to tech-agnosticism is impressive because it involves keeping abreast of an ever-expanding array of third-party products used for every aspect of data science – not just for model creation but also for hardware and other infrastructure used to operationalize models once they have been created. Domino has also been widely emulated; nonetheless, the company remains a ‘go to’ name for collaborative code-first data science for large teams of expert data scientists, and with recent model-monitoring enhancements, it is further addressing the critical area of enterprise MLOps, too. Domino continues to be one of a few vendors to offer a code-based data science platform capable of providing the same experience in the cloud as on-premises. That’s vital because not every enterprise is ‘all in’ with the cloud, even though public clouds are popular for machine learning. Companies that already have machine learning projects in play chose the public cloud as the most popular primary venue for storage (51%), training (39%) and predictions (37%) in our Voice of the Enterprise: AI & Machine Learning, Infrastructure 2021 survey.

Details

As an early pioneer of agnosticism and heterogeneity, Domino was instrumental in shaping the criteria for enterprise data science platforms. Indeed, when the company was founded in 2013, it embraced a mixture of open source and proprietary data science tools, languages and frameworks from the get-go so data scientists didn’t have to learn new ones to use its platform for their enterprise data science initiatives. The company continued with this strategy in Domino 4.5, unleashed in June, and Domino 4.6 in September. Furthermore, in 2018 Domino introduced a model management framework as an add-on module, recognizing that large organizations require support for the full data science lifecycle, and that requires model operationalization, which is where Domino Model Monitor fits in.

Domino’s early positioning was as a ‘GitHub’ – or system of record – for data science. Indeed, the vendor continues to use the comparison to communicate its value to data scientists as a version-control platform for their workflows to make them reproducible by using Git as a repository. Furthermore, Domino integrated with Git in 2017 to understand how data scientists work with Git so that it could provide the tools and workflows to make using Git codebases as easy as possible.

GitLab integration in Domino 4.6 is all about building on the love data scientists already have for Git through integration with GitLab, which is used as a centralized environment for features associated with models. Domino 4.6 essentially enables GitLab users to create a new GitLab repository during the Git-based project-creation process. Domino has also embraced Git in other ways. Domino 4.5 enables data scientists to browse the content of linked Git repositories natively from within its platform, as well as streamline the setup of Git-based projects. The latter is delivered through hooks into CodeSynch, a third-party tool used to stream code changes in real time to the cloud. When data scientists create a project in Domino, they have the option to link to an existing Git repository or create a new GitHub or GitHub Enterprise repository for Git-based projects using CodeSynch.
For about a year, Domino has provided the ability to provision and orchestrate Spark clusters in a point-and-click manner so data scientists don't have to rely on IT to do it on their behalf. Data scientists require a Spark cluster to support compute-intensive workloads such as deep learning. Domino 4.6 adds easily deployable access to Ray.io, an open source offering for scaling Python from single machines to large clusters, as well as Dask, an open source library for parallel computing written in Python, in order to embrace other types of scaling environments for data science workloads using a point-and-click approach. Furthermore, Domino had previously undertaken an engineering effort to support containers by enveloping Kubernetes to engender cloud and infrastructure portability. Domino 4.6 is certified for Amazon EKS, which is designed to make it easy to run Kubernetes on the AWS cloud and on-premises.

Finally, Domino continues to bolster its model monitoring add-on. Domino Model Monitor is now designed to automatically compute data drift and model quality across billions of daily predictions courtesy of the new Domino Elastic Monitoring Engine that underpins it. Domino Model Monitor is also now available on cloud infrastructure from AWS, Google and Microsoft, and has a new configuration setting that enables single sign-on between it and the rest of the vendor’s platform.