

# Domino vs. Databricks

## An Enterprise MLOps Comparison



### Introduction

You manage a team of professional data scientists. Your team's mission is clear:

- Take on complex and demanding use cases.
- Deliver actionable results. And do it faster.
- Provide measurable value to the business.

Wouldn't it be great if you didn't have to spend time dealing with platforms and infrastructure? You could focus all of your attention on things that matter:

- Recruit, hire, and retain the best people
- Spend time with clients
- Investigate cutting-edge tools and techniques

Wishful thinking, right? You'd like to spend more time on the big picture, but reality keeps getting in the way.

You may have heard about Domino, the leading Enterprise MLOps Platform. But your IT leadership wants you to consider Databricks for your data science team. Databricks is a leading platform for data engineering in the cloud. If Databricks can meet the needs of your data science professionals, IT can solve two needs with one deed.

### Your Complex Reality



**Diverse tools, infrastructure, and data.** Your data scientists use a wide range of tools and software on different computing platforms. Data is everywhere, and it's a mess. Your CDO says the new data platform will fix everything. Uh huh. You've heard that before.



**Documentation and compliance.** Data science that isn't reproducible isn't trustworthy. Reproducibility means exact reproducibility. Your team spends a lot of time and effort documenting and curating results, so you can survive an audit.



**Short model shelf life.** Model accuracy can drift in production for many different reasons. Time spent monitoring model accuracy can take up a lot of expert time. Some teams have dedicated resources doing nothing but tracking model performance.



**Client insights.** Every model has stakeholders. Most stakeholders aren't trained in data science and machine learning. Showing them logloss statistics isn't enough; they need to see charts and graphs that illustrate the behavior of a model in business terms.

## Will Databricks meet your needs?

Like Domino, Databricks is available on AWS, Azure, and GCP. With Databricks, your users can spin up Apache Spark™ clusters on demand and connect to most data sources. They can do that with Domino, too. So far, it's an even match.

But what if members of your team use MATLAB® or SAS®? Your MATLAB and SAS users can work on Domino together with data scientists who use tools such as Python, R, TensorFlow, and PyTorch.

Databricks has an answer for SAS users who want to work on their platform. It's "learn to code." Your data scientists use many different notebooks and IDEs: Jupyter, Visual Code, RStudio, and many others.

With Domino, they can keep using their preferred development environment. Databricks will force them to switch to the Databricks Notebook. Do you really want your team to spend time learning another notebook? Domino users can launch Ray or Dask clusters on demand. Ray and Dask are designed for machine learning. Spark, on the other hand, is best for data processing tasks like ETL. It's possible to set up a Ray cluster in Databricks, but you're going to need an administrator and a lot of time.

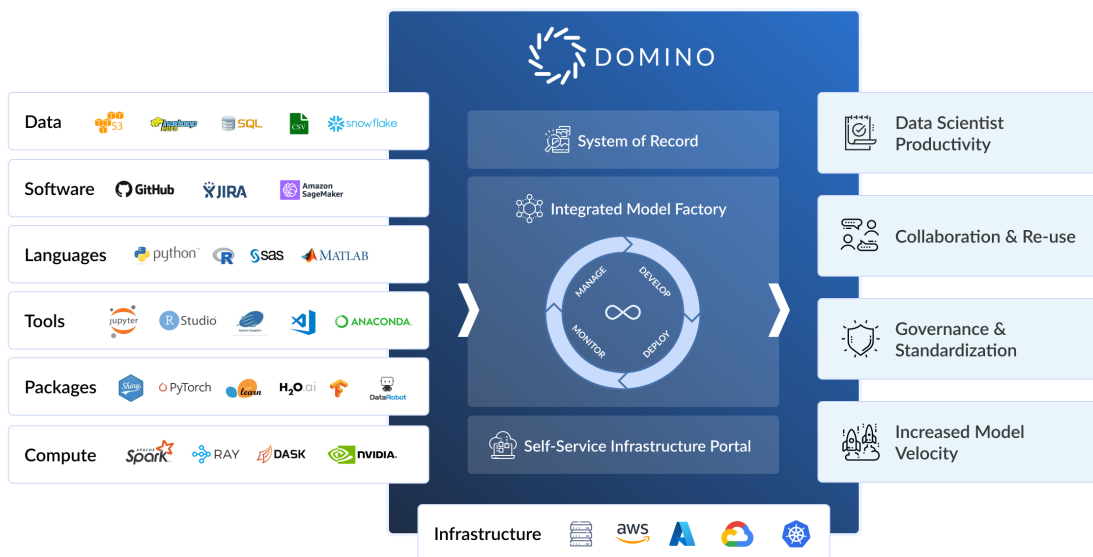
Domino reproducibility is exact reproducibility. To deliver this, Domino automatically tracks eight key project assets: data, code, environments, compute,

experiments, models, batch jobs, and web apps. The key word is automatic. Your users don't have to do anything or remember anything. They can't "forget" to check in the latest version. Since Domino tracking is comprehensive and automatic, you can trust that everything your team does is tracked, captured, and reproducible.

Databricks, in contrast, tracks some of those assets – if the user invokes tracking. That's a big "if." You may have to give your team sticky notes that say "Don't Forget to Track!"

Monitoring models in production is one of your biggest headaches. Domino offers model monitoring out of the box. It automatically captures instrumented training, prediction, and ground truth data, and delivers a dashboard of key metrics. Domino monitors your models wherever you deploy them, so you can monitor all of your models from a single dashboard.

Databricks says you can build a custom solution to monitor models on their platform, but do you really want your team to spend time building tools? Your team uses tools like Shiny, Dash, and Flask to create insights for stakeholders. With Domino, your web apps are easily integrated with your modeling workflow. And Domino tracks them with all of the other project assets.



We mentioned earlier that Databricks runs on the three leading public cloud platforms, and so does Domino. But what if your team runs on customer-managed infrastructure? Domino is 100% native Kubernetes; it runs on the leading cloud K8S distributions, plus Red Hat OpenShift. With Domino, you can leverage NVIDIA AI Enterprise® to orchestrate virtualized GPU-accelerated environments deployed on mainstream servers or in the public cloud, or NVIDIA DGX® hardware on NVIDIA Certified Systems® available through NVIDIA's partner ecosystem. And you can deploy models to the edge with NVIDIA Fleet Command®.

**“The paid platform we used briefly—a unified data analytics platform—was too reliant on Apache Spark™ and couldn't provide the support, security, or flexibility our data engineers, data scientists, and ML engineers needed.”**

**BIZ PHILLIPS**

Senior Health Data Scientist at Evidation

## Summary of Key Differences

We see quite a few customers who already use Databricks as a data engineering platform. It's natural for these organizations to consider whether they can use Databricks as an MLOps platform as well. Customers choose Domino over Databricks when they want to:

- ☑ Support expert users on diverse tools, including SAS and MATLAB
- ☑ Ensure exact reproducibility
- ☑ Monitor model quality for all production models
- ☑ Deliver insights with Shiny, Dash, or Flask web apps
- ☑ Leverage customer-managed computing platforms
- ☑ Tap the power of NVIDIA DGX hardware or NVIDIA AI Enterprise
- ☑ Deploy models to the edge with NVIDIA Fleet Command

But even with the differences in capabilities, we sometimes see companies try to make Databricks work for MLOps. For example, at [Evidation](#), they were trying to replace an existing homegrown data science platform and found that Databricks simply couldn't provide the support, security, or flexibility their data science team needed.



## The Choice is Clear

Databricks makes a great platform for data engineering, and some of our customers use it. But when they consider their requirements for data science and MLOps, Databricks comes up short. Instead, Domino is the choice for Enterprise MLOps for over 20% of the Fortune 100.

[LEARN MORE](#)

