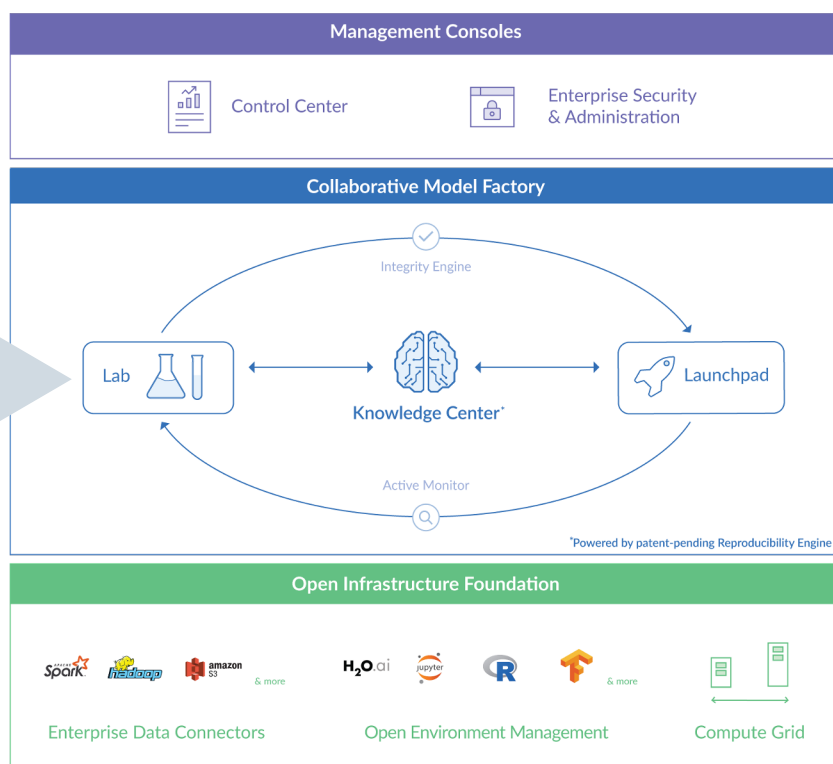
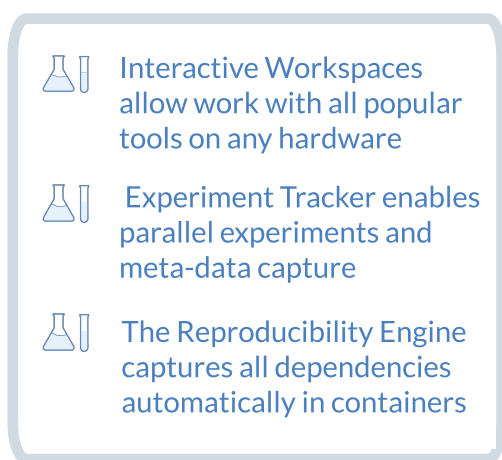


Accelerate Research and Maximize Data Science Productivity

Domino Lab is where data scientists develop models, working interactively with their preferred tools and then experimenting systematically to rapidly train and tune models. An integral component of Domino's data science platform, Lab leverages [Domino Compute Grid](#) to provide scalable compute and access to any tool via Environment

Management. Models built in Lab are delivered to production in [Domino Launchpad](#), while the [Reproducibility Engine](#) automatically preserves and organizes the full experimental record. Data science and IT leaders can manage activity in Lab via [Domino Control Center](#).

The Domino Data Science Platform



Lab addresses the three key challenges data science organizations face when developing models:

- **DevOps headaches** Data science teams spend time manually juggling hardware and software infrastructure, rather than focusing on developing models with their favorite tools.
- **Inefficient experimentation process** Developing models is nonlinear and can be hard to track and reproduce. Data science teams struggle to track key metrics and the complex chain of dependencies underlying each experiment.
- **Redundant work** Models are rarely the output of a single individual, but teams face constant challenges building models collaboratively, including orphaned work, missing context, and package mismatches.

TRUSTED BY



Domino Lab Features and Benefits

Domino Lab has three key features, integrated into one seamless experience:

1. **Workspaces** provide one-click access to both popular open-source tools like Jupyter, RStudio, and Zeppelin, as well as commercial tools such as SAS Studio and H2O Driverless AI.
2. **Experiment Tracker** can run, track, and compare many experiments simultaneously to rapidly develop models in all popular scripting languages (Python, R, SAS, etc.)
3. **Reproducibility Engine** automatically captures all experiments done in Lab, including their dependencies and their results, making it easy to document, share, and build on past work.

These key Lab features are supported by our Infrastructure Foundation, including Compute Grid, which offers vertical and horizontal scaling and access to specialized resources such as GPUs, along with Compute Environments, which provide flexible, revisioned Docker containers to support open-source and proprietary software tools.

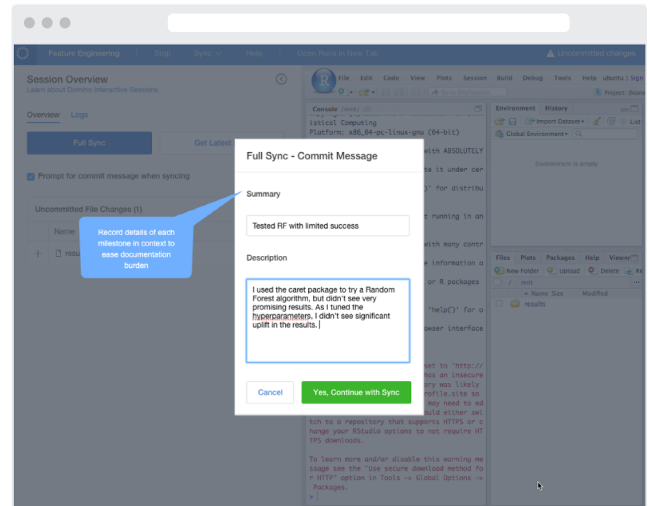
Workspaces

Data science teams are often constrained to a few centrally mandated, infrequently updated tools. This leads to frustrated data scientists who can't use the right tool for the job and, eventually, an increase in team turnover. Lab natively supports popular interactive open-source tools like Jupyter, RStudio, and Zeppelin. Lab also provides support for commercial IDEs like SAS and H2O Driverless AI. Given Lab's extensible Compute Environments with Docker containers, users can rapidly adopt the rich and evolving ecosystem of tools and libraries.

- **Many tools, one project** Workspaces enable data scientists and their collaborators to choose the appropriate tool for the problem at hand. One user may prefer RStudio for tuning models, while another user prefers Jupyter for EDA. If users want to automate pieces of the lifecycle, they can use automation tools such as DataRobot and H2O Driverless, all within the platform.

Workspaces							
Choose from the options below to launch a new workspace. Don't see what you need? Learn more about Workspaces.							
Launch a new Workspace							
<div><div></div><div></div><div></div><div></div><div></div><div></div></div>							
Workspaces							
Name	Type	Status	Start Time	Run Time	Creator	Run	
Hyperparameter Optimization	H2O	RUNNING	Today at 5:10 pm	—	mac	#68	Stop Open
Trying different model types	RStudio	RUNNING	Today at 4:12 pm	—	mac	#67	Stop Open
Feature Engineering Customer Lead Data	Jupyter	RUNNING	Today at 3:18 pm	—	mac	#66	Stop Open
Initial Data Exploration	SAS Studio	RUNNING	Today at 1:43 pm	—	mac	#65	Stop Open

- **Context preservation** Lab supports midsession syncing in an Interactive Workspace (e.g., Jupyter or RStudio) to mark the state of a project at key milestones. These milestones can have a commit message preserved and indexed for later search by the model builder or future collaborators.



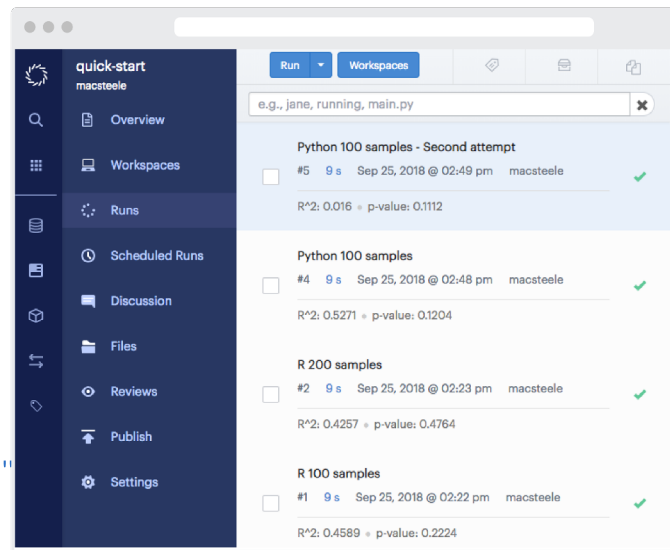
- **Hardware flexibility** In addition to using their tool of choice, users can configure Workspaces to work on the hardware most appropriate for the task at hand. Teams can easily scale up to larger machines or utilize specialized hardware like GPUs. Domino suggests which tiers have availability, offers configurable cost transparency, and automatically spins down machines to maximize cost efficiency.

Hardware tier	
Medium 4 cores · 16 GB RAM · \$0.0033/min	<1 MIN
P3 2XLarge 8 cores · 61 GB RAM · \$0.0510/min	<7 MIN
i3 Storage Optimized 64 cores · 488 GB RAM · \$0.0832/min	<7 MIN
GPU (p3.2xlarge, dedicated to 1 run) 8 cores · 61 GB RAM · \$0.0510/min	<7 MIN
Large (m4.4xlarge, dedicated to 1 run) 16 cores · 64 GB RAM · \$0.0133/min	<1 MIN
domino-c5.4xlarge 16 cores · 30 GB RAM · \$0.0113/min	<7 MIN
High Memory 16 cores · 122 GB RAM · \$0.0177/min	<7 MIN
GPU Large 32 cores · 60 GB RAM · \$0.0054/min	<7 MIN

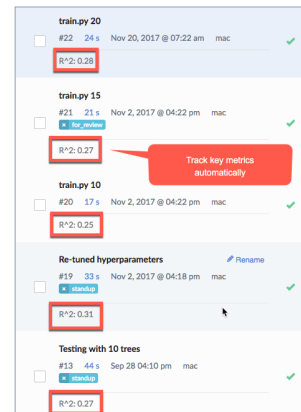
Experiment Tracker

The Lab lets you execute, track, and measure programmatic experiments like an automated lab notebook. There is support for Python, R, SAS, MATLAB, and many other scripting languages on a single, revisioned set of files. Teams composed of data scientists with different backgrounds can all track and compare their results, allowing seamless coordination between one language (e.g., SAS) and another (e.g., R).

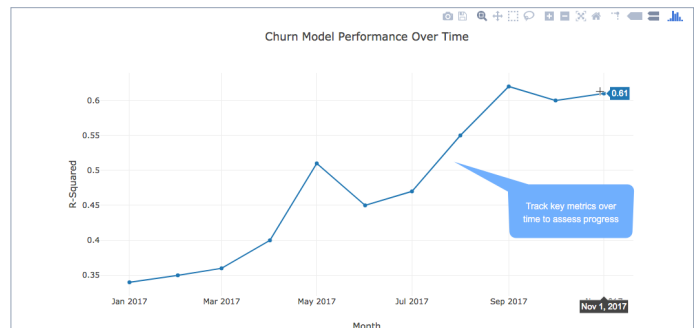
- **Run many experiments in parallel** Often data scientists want to try multiple—or even dozens—of experiments simultaneously to see what combination of features and parameters performs best. The Lab allows users to scale their work “horizontally” by kicking off many experiments at once and tracking all of the key results simultaneously. IT can configure limits on each tier depending on the team’s evolving needs, and Lab provides reporting to track usage costs. These experiments can also be initiated via Domino’s CLI or [platform APIs](#).



- **Domino stats** Data science projects are often open-ended and involve significant trial and error. Many users ignore the difficult after-the-fact documentation such projects create, even though this means they lose out on past progress. Experiment Tracker acts as a modern lab notebook that supports the automated tracking of key results generated during batch experiments. All that is required is that the results are written to a [dominostats.json file](#), which is detected and rendered in the UI. For example, a team may track the r-squared of model variations over time.



Leading data science teams extract these key metrics via our platform API and perform “meta-analysis.” For example, teams can render each experiment’s key metrics visually to see how model performance has evolved over time. One hedge fund uses this feature to showcase the incremental lift new models provide to the aggregate investment portfolio based on weekly backtests.



- **Experiment comparison** In the process of dozens of experiments, the ability to easily compare results allows users to understand meaningful changes and preserve insights. The Lab allows data scientists and any collaborators they invite to their projects to compare two results side-by-side and then leave comments.



Reproducibility Engine

The Lab leverages our patent-pending Reproducibility Engine to automatically keep track of the relevant components of every ongoing project. The full history of a model encompasses every version of the data, packages, code, and parameters used to develop that model, in addition to the results and any subsequent discussion. The system automatically persists the state of the files, the hardware used to execute the experiment, exact software versions and drivers, dependencies (e.g., imported repos), and results of the experiment.

For data science teams, this dramatically reduces the need for post hoc documentation. For collaborators, the series of these snapshots provides rich context rather than requiring them to forensically reconstruct what happened months earlier.

Naive Bayes Test **Description of Experiment** ✓ Succeeded

[> Console output](#) [Resource usage](#) [Details](#) [Discussion](#)

Code
Command
`evaluate_model.R nb iris Species`
[Schedule This to Repeat](#)

Data
Input
[View Input Files](#)
Output
[View Output Files](#)

Results
Duration
8 s
Queued
Oct 20, 2017 @ 03:21 pm
Started by
mac
Results
[View Results](#)

Tools and Libraries
Hardware tier
Default
Environment
Anaconda - Revision #18
Executor
vip-20170322-exec-default-14

Repositories **Dependencies**

Name	Mount point	Ref	Starting commit ID	Finished commit ID
scikit-learn	/repos/scikit-learn	(Default branch)	d9eba75	d9eba75

In addition to the Reproducibility Engine, Domino has tightly woven collaboration capabilities throughout Lab, sparking more idea sharing and reducing redundant work amongst data science teams. For more details on these features, see our [Reproducibility Engine](#) and [Collaboration](#) tech briefs.

- **Import/Export Projects** The Lab supports the import of key modular components from one project, such as datasets or helpful packaged code utilities (e.g., backtesting tools or standard sensitivity analyses), into other projects to save time and disseminate best practices. Downstream consumers can then use the latest version of components or specific releases determined by the owner.

Exports [Learn more](#)

See what projects depend on this.

Export	Description
Environment variables	If checked, all projects which import this project will be able to access this project's environment variables.
Files	If checked, all projects which import this project will be able to access this project's files.
Code Package	If selected, an R or Python package will be available to projects which import this one.

Select what components to export from Project #1

Import desired version of relevant components to Project #2

Local **Other Projects** **Git Repositories**

Artifacts From Other Projects
This project can view and use files, Environment Variables, and Code Packages from other projects. The files are linked from other projects and are Read Only. You are able to link to different releases using the project dropdown. [Learn more about linking projects.](#)

[Import](#)

Project	Location	Release	Imports	View	Remove
mac/clickstream_data	/mnt/mac/clickstream_data	✓ Latest	Files	View	Remove
mac/customer_data	/mnt/mac/customer_data	#1 test.py (Dec 12 2017 11:18AM)	Files	View	Remove

- **Git Integration** Lab is also open to how data science teams prefer to store and version their code—some are 100% Git, while others use manual file-naming conventions, and some have no established system at all. The Lab integrates with popular tools like GitHub while also allowing other users to version files directly in Domino. Different data scientists can work as they prefer, ensuring tool siloes don't limit work collaboration and throughput.

Conclusion

The Lab accelerates research for data science organizations by eliminating DevOps headaches, reducing knowledge management burdens, and fostering collaboration across all types of data scientists. By allowing users to access their preferred tools and reducing the collaboration and experiment tracking burdens, users naturally gravitate towards the platform. As more activity moves on platform, all users benefit from the additional work that becomes accessible and managers can more easily coach their teams. The Lab creates an end result of accelerating throughput and delivers greater business value with more efficiency.

About Domino

Domino Data Lab provides an open data science platform to help companies run their business on models. Model-driven companies like Allstate, Instacart, Dell, and Bayer use Domino to accelerate breakthrough research, increase collaboration, and rapidly deliver high-impact models. Founded in 2013 and based in San Francisco, Domino is backed by Sequoia Capital, Coatue, Bloomberg Beta, and Zetta Venture Partners. To learn more, visit dominodatalab.com.