



VS

Amazon SageMaker



When data scientists are choosing a data science platform, they look for several qualities and capabilities that will ensure their work is productive and valuable. They seek a platform that is easy to use, scales with their team, and allows them to use their preferred tools and languages.

Saturn Cloud and SageMaker are two popular platforms among data scientists. They both allow data scientists to do their work in the cloud using hosted notebooks, but they differ significantly in their features and ease of use.

Saturn Cloud

Saturn Cloud is a platform that allows data scientists and teams to work with scalable resources in the cloud. With just a few clicks, Saturn Cloud provides access to computing resources with customizable amounts of memory and power, including GPUs and Dask distributed computing clusters, in a completely hosted environment. In Saturn Cloud, data scientists can use their preferred languages, IDEs, and machine learning libraries. It offers full Git integration, shared custom images, and secure credential storage as well, making it easy to scale and build any data science team in the cloud. And with features like jobs and deployments, it supports the entire machine learning lifecycle from experimentation to production. With its intuitive user interface and enhanced configurability, data science in the cloud becomes effortless and work becomes more productive.

SageMaker

SageMaker, from Amazon Web Services, allows data scientists to build, train, and deploy machine learning models in the cloud. It gives users a Jupyter notebook interface to work in, and it supports Python as well as other programming languages (e.g., R, Scala, Julia). In addition to model development and deployment, SageMaker features data labeling tools, custom algorithms designed or implemented by AWS, model interpretability tooling, and feature storage products. It has tight integration to other AWS tools, which can be advantageous for a team exclusively using AWS products. However, it lacks several key features that data science teams look for, such as ease of use and interoperability.

Choosing Your Data Science Platform

As you prepare to choose a data science platform for your team, consider how the following features align with your business priorities and needs:

Ease of Use

The computing power of any data science platform is only useful if it is easily accessible.

Saturn Cloud's user interface is not only intuitive and simple to use, but it also provides the capabilities and configurability that data scientists need. The start-up process for new users is only two steps: create an account, which can use Github or Google credentials, and, with one click each, create and start a template project.

From there, data scientists have access to a Jupyter environment as well as straightforward instructions for adding their own code, customizing their workspace, scaling up compute resources to Dask clusters, and more. The Saturn Cloud environment is as familiar to the users as their own local environment.

SageMaker's start-up process is much more complex for new users. First, they must have an AWS account as well as the appropriate permissions. To get started, they need to understand concepts such as VPC creation, subnets, and AWS credentialing just to start up an instance. Once they do manage to get started in SageMaker, more challenges await them. In particular, they are confronted with a UI experience in the Jupyter workspace that can be busy, confusing, and stressful, especially if users already have a perfect setup in a local environment.

What's more: AWS advocates separating machine learning training and inference from its hosted notebook, which can provide flexibility in deployment but nonetheless requires additional user training and experience. Frequently, extra assistance from DevOps professionals is required to securely and successfully complete the set-up processes and get them working.

Open and Interoperable Computing

Users should be able to choose their own tools and workflows in a data science platform.

IDEs

Data science teams looking at new software often question how well it will mesh with their existing, preferred tools. Few teams are interested in completely scrapping their current workflow, but they do have a gap or many gaps in functionality that need to be filled. For example, users may like developing in a local environment but require compute resources in the cloud to power that work.

Saturn Cloud makes SSH access both versatile and easy. Users can use SSH to connect a local IDE to a Saturn Cloud Jupyter instance and work from there, or they can use a local Jupyter installation and just SSH to a Dask cluster hosted in Saturn Cloud. This means that users have access to powerful resources without needing to leave the development environment where they do their best work.

SageMaker does not natively support this type of connection. Notably, SSH connectivity from local to cloud development settings require workarounds or are impossible.

Images

Consistent, robust development images make data science practice easier, reproducible, and testable. If images are difficult to make or edit, users will likely struggle to reproduce analyses or bugs, slowing down development and potentially introducing avoidable errors.

In Saturn Cloud, the default images are designed to be as slim as possible, and they are regularly updated to keep users' tools current. Custom images may be created from a simple list of packages, YAML, or conda environment specification, among other options. If desired, the images can even be based on Docker containers that a data science team already has available. Images may be shared with one or many users, and they can be viewed without having to start up a project.

In SageMaker, several default images are provided to users as kernels, but these images often have outdated versions of libraries. For example, at the time of this writing, the Python 3 Data Science kernel in SageMaker contains outdated versions of pandas, numpy, scikit-learn, ipykernel, jupyterlab, and other standard libraries. Outdated images require more time investment for updates, reducing productivity. Using a custom image in SageMaker is a series of lengthy multi-step processes: first to create the image, and then to attach the image to a project. Custom images must all be stored as a container image in Amazon ECR, and they may not be created in the UI.

Libraries

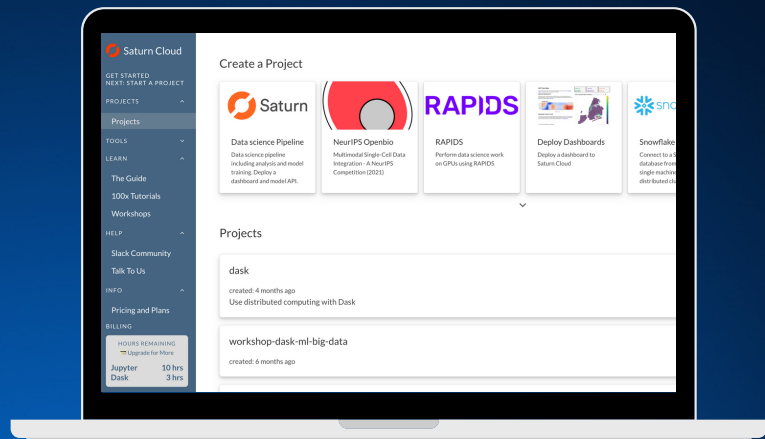
Data scientists and analysts work best using the tools they want to use.

In Saturn Cloud, users can use their preferred machine learning libraries. Saturn Cloud explicitly and actively supports different machine learning frameworks, including scikit-learn, as well as deep learning tools such as PyTorch and TensorFlow. Moreover, it's designed so that data scientists can adapt their local or single node code to the product. Here, code is transferable: Code written outside Saturn Cloud can be brought in, and code written inside Saturn Cloud can run in other workspaces.

While SageMaker does passively support a variety of machine learning approaches, users are strongly encouraged to utilize AWS internal libraries, which abstract away large and important pieces of the modeling process. These abstractions might be acceptable for users who do not have existing workflows in their local environment, but they might not be ideal for users who already do. Moreover, if users or teams ever depart from the SageMaker ecosystem, the code they wrote using the AWS libraries might become obsolete, resulting in re-work and wasted time.

Ready to give it a try? [Join Saturn Cloud for free](#) and get 30 hours of free compute each month.

Your open data science cloud environment



ALL-IN-ONE WORKSPACE

- Scalable data science with Python, R, & more
- Use up to 4TB RAM, large GPUs, and Dask clusters
- Cloud-hosted JupyterLab and RStudio
- Easily share work with colleagues
- Straightforward deployment environments
- Run jobs plus deploy dashboards and APIs
- Connect from existing cloud resources
- Admin reporting tools

Trusted by



And more

Join thousands of data scientists today and get 30 hours of free compute each month:

[JOIN FOR FREE](#)

www.saturncloud.io