Saturn Cloud
End-to-end data science in Python

Why Saturn Cloud?

- Automate the DevOps and machine learning (ML) infrastructure engineering required to scale Python
- Run models on a cluster with Dask and Kubernetes to auto-scale compute resources
- Schedule tasks that launch asynchronously and can run in parallel with Prefect
- Build custom environments with Docker and share projects and templates across your team
- Deploy in your virtual private cloud (VPC) and integrate with your existing AWS environment

Develop and deploy data science models in Python at scale with automated DevOps and ML infrastructure engineering

Traditional methods for scaling and deploying Python involve multiple tools, multiple teams, and multiple programming languages. Saturn solves this by integrating the resources and frameworks for scalable Python as a managed service, so teams can focus on data science, and offload everything else.

Key features

Custom Jupyter environments
Spin up a virtual machine (VM) with the infrastructure services, compute resources, and packages ready for any Python workload:
• Build a custom Docker image that is maintained for you and available each time you startup in Saturn
• Share custom environments across your organization for easier collaboration and better model reproducibility

Cluster-computing with Dask
Scale your projects from laptop to server to cluster with Dask for Python-native parallel processing:
• Scale the most popular data science and ML libraries, incl. Pandas, Numpy, and Scikit-Learn, using 100% Python code
• Execute Dask on central processing unit (CPU) or graphics processing unit (GPU) hardware for processing speeds up to 3,000 tasks for parallel computations

Workflow orchestration with Prefect
Schedule and monitor workflows that execute in parallel across your cluster with Prefect:
• Run Prefect with Dask to launch tasks asynchronously and run in parallel up to the number of workers in your cluster
• Utilize Prefect’s task library to schedule the most common tasks without writing any code

Dask Performance

3,000 tasks executed per second

10x increase in throughput
**How it works**

Saturn Cloud runs as an application inside Kubernetes leveraging AWS services such as Amazon EC2, AWS Identity and Access Management (IAM) and Amazon Virtual Private Cloud (VPC) to provide secure and scalable infrastructure for running Data Science and Machine Learning workloads within your AWS environment. Saturn connects to AWS storage services, AWS real time data sources, and AWS management tools.

- Provision a VM in your environment with Docker, Kubernetes, and other infrastructure services
- Enable flexibility and scale to train and deploy data science and ML models using the full Python ecosystem
- Compatible with existing AWS and third-party data, security, and IT management tools

**Specifications**

<table>
<thead>
<tr>
<th>Product Specifications</th>
</tr>
</thead>
</table>
| **Hardware** | Processing: CPU instances range from 2 to 64 cores; GPU instances incl. 1 Nvidia T4 GPU  
RAM: CPU instance range from 4GB to 512GB RAM; GPU instances incl. 16GB GPU RAM  
Disk: Range from 10GB, 40GB, 256GB, and 1024GB of disk space |
| **Base Image** | Data Science: GeoPandas, Matplotlib, NLTK, Numba, NumPy, Pandas, SciPy  
Machine Learning: PyTorch, Scikit-learn, TensorFlow  
Parallel Processing: Dask  
Jobs Scheduling: Prefect  
NVIDIA RAPIDS: cuDF, cuML, cuGraph  
DataViz: Bokeh, Dash, Seaborn |
| **Image Config** | Python: environment.yaml, requirements.txt  
Linux: apt-get |
| **Cluster** | Dask: 1-click Dask cluster creation and auto-scaling |
| **Teams** | Collaboration: Share Jupyter environments; manage version control with a native Git integration  
Publishing: Display data, code, and visualizations on a single URL |
| **Production** | REST APIs: 1-click project deployment using REST APIs |