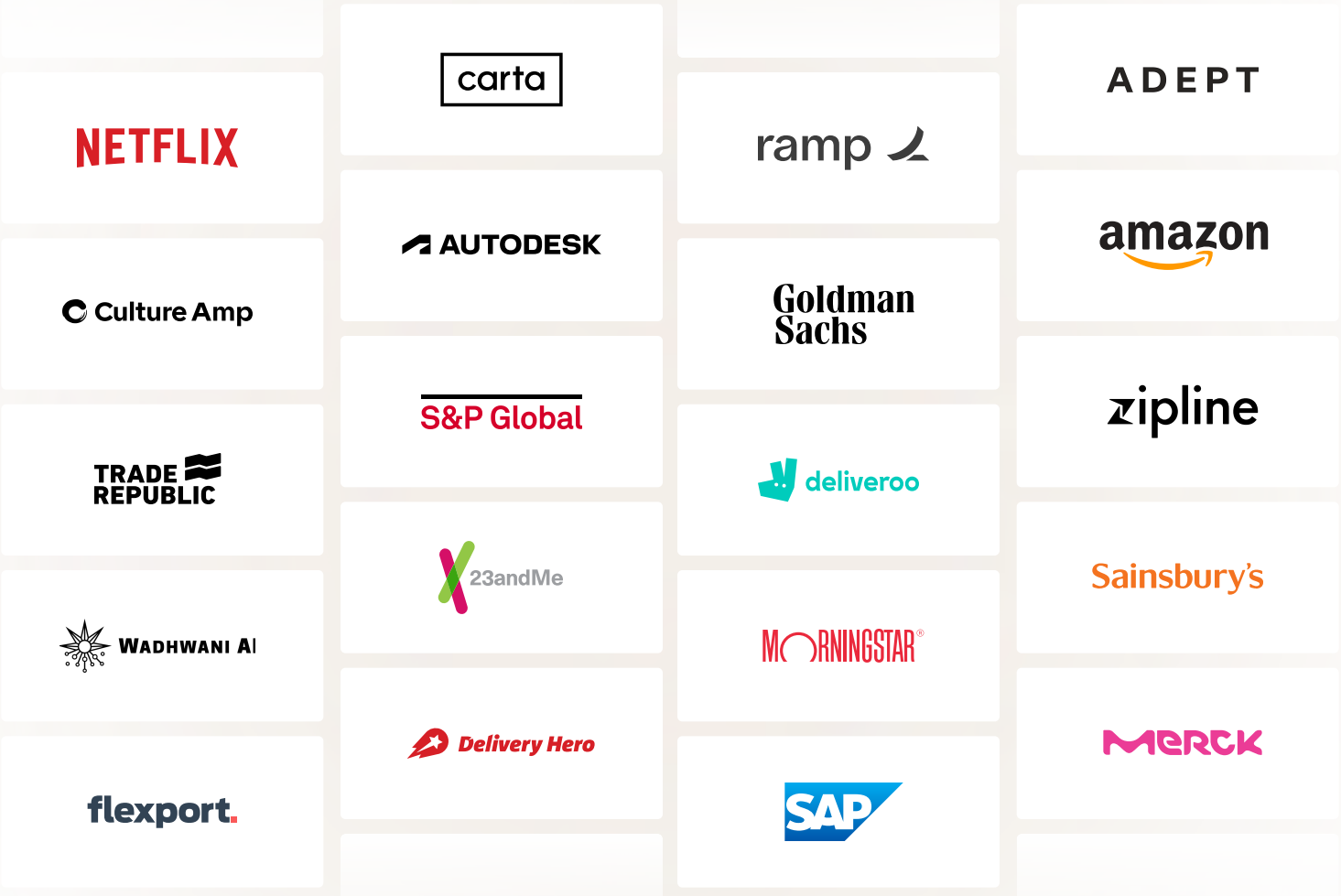
A developer-friendly platform for ML+AI systems

Background

Outerbounds was spun off from Netflix in 2021. At Netflix, Outerbounds' founders led ML and AI infrastructure, encoding the best practices of rapid ML/AI development into an open-source library Metaflow, with a particular focus on human-centric, productivity-boosting developer experience.

In addition to powering most ML/AI projects at Netflix today, Metaflow has become an industry-standard tool for production ML/AI systems, adopted by hundreds of leading companies. It powers a wide range of use cases from financial fraud detection and biotech to autonomous drones and custom large language models.

Outerbounds builds on the foundation laid by Metaflow by offering it as a part of a fully managed, secure, cost-effective ML and AI platform.



Scenario

Continuously updating ML with structured data

Let's take a look at a typical business-oriented ML system. The system ingests data from a data warehouse, trains models for classification or forecasting, and uses them to provide up-to-date inferences, integrating results to surrounding systems. Consider common questions that raise during development of a system like this.

How to access data quickly and securely?

How to interface data engineering and ML workflows?

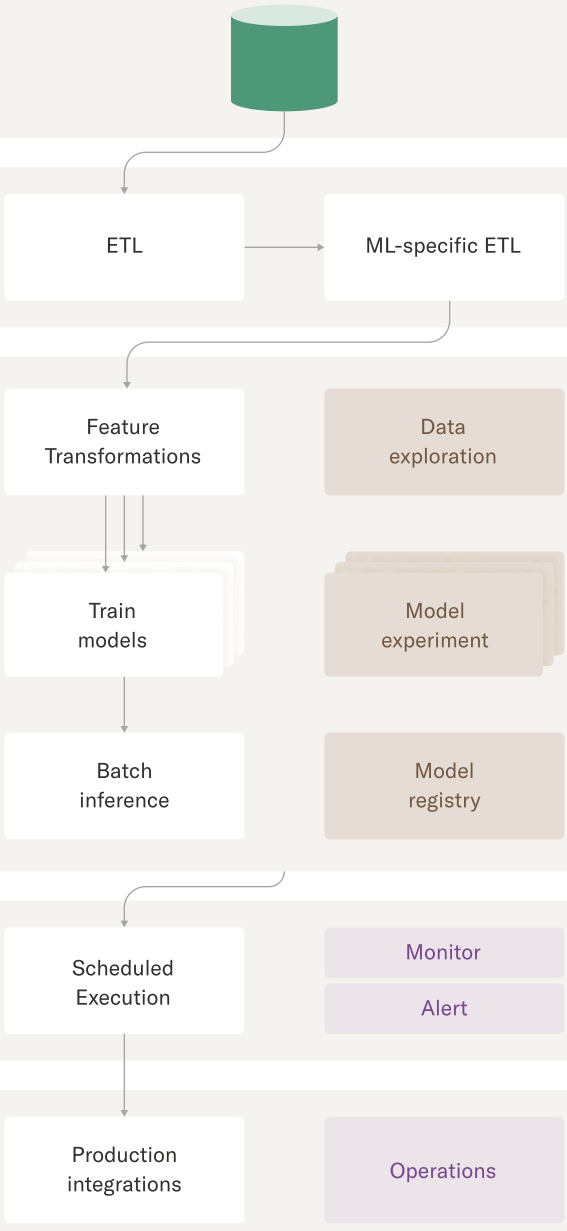
How to allow rapid development of models and features?

How to experiment and train models at scale, leveraging the cloud cost-efficiently?

How to keep track of all code, data, and models, enabling continuous improvement?

How to deploy all of the above in production in a highly-available manner?

How to integrate the results into surrounding systems reliably and operate them effortlessly?



The challenge is not to ship the above system once, but to **develop systems like this routinely, continuously improving results.**

Solution

Human-Centric Infrastructure for ML and AI

Based on our experience from working with hundreds of companies, real-world ML and AI systems end up including a these four foundational layers of infrastructure - sometimes organically, sometimes by design:

Data	Accessing data efficiently and securely.
Compute	Leveraging compute resources to process data, train models, and run inference.
Orchestration	Orchestrating the system, keeping it running in a highly-available manner.
Tracking and Versioning	Observing and keeping track of code, data, and models across experiments and production.

There are many valid technical solutions to each of these layers. While not all approaches are equal, ultimately human factors - **the ease of experimentation, development, and operations** - tend to dominate the effectiveness of the overall solution.

Developer UX	Enabling developers to experiment rapidly, develop effectively, ship to production confidently, and improve results continuously.

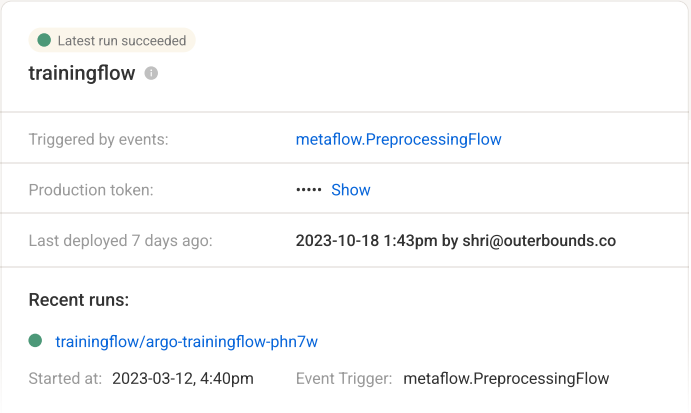
Outerbounds provides a full stack of ML/AI infrastructure, addressing the above layers holistically - take a look how.

Data

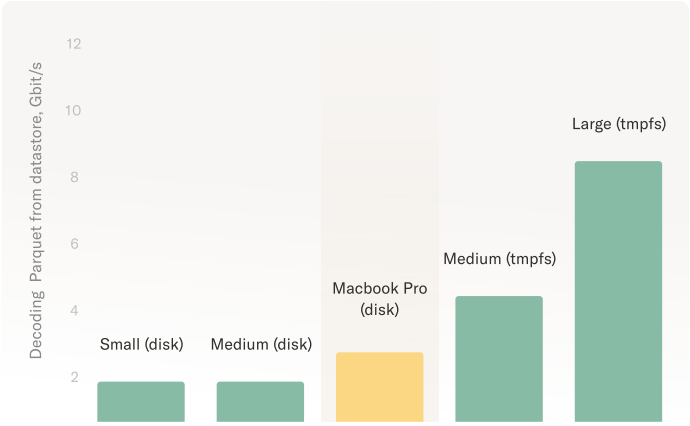
Outerbounds integrates to popular data lakes and warehouses which excel at storing and processing structured data. Access data quickly and securely, interface with ETL, process features, and use modern tools for data, while building ML/AI systems cost-efficiently.

```python
@trigger (event="data_update")
class DataFlow(FlowSpec):

    @secrets
    @snowflake
    @step
    def start(self):
      self.x = 124
      self.next(self.end)
```

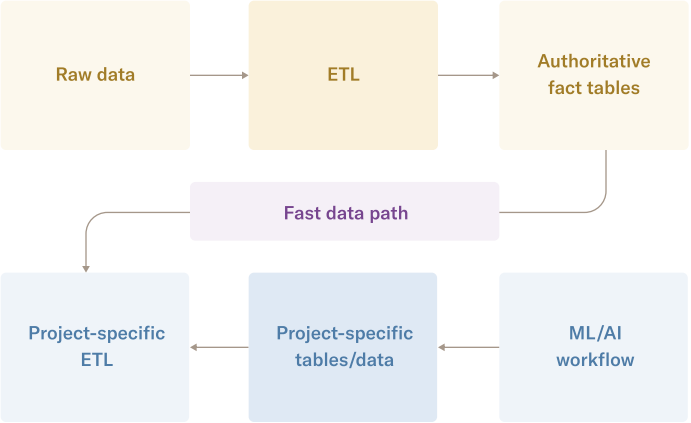
With a few lines of Python code, you can read data securely from various data sources, following preconfigured paved paths that conform with policies.



Workflows can be configured to run automatically whenever new data is available, enabling continuous training and inference.



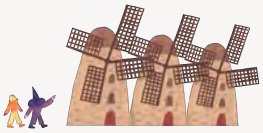
Load both structured and unstructured data at blazing speeds, up to tens of gigabits per second, using a built-in, high-performance data layer.

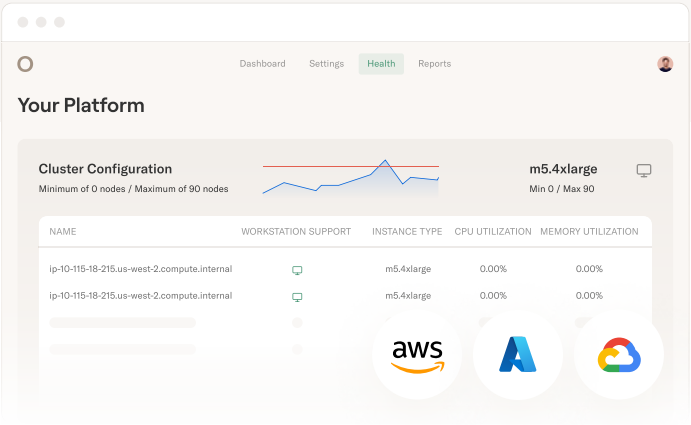


Divide responsibilities clearly between data engineering and ML/AI/data science teams, balancing stability and experimentation.

Compute

Outerbounds provides a uniquely capable and cost-efficient compute layer. Leverage unlimited compute resources across clouds, covering all instance types, including a wide range of GPUs and other hardware accelerators at the lowest cost for a flat, unmetered fee.

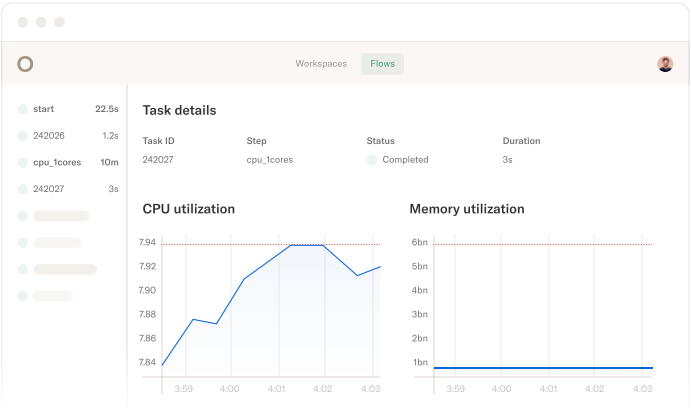


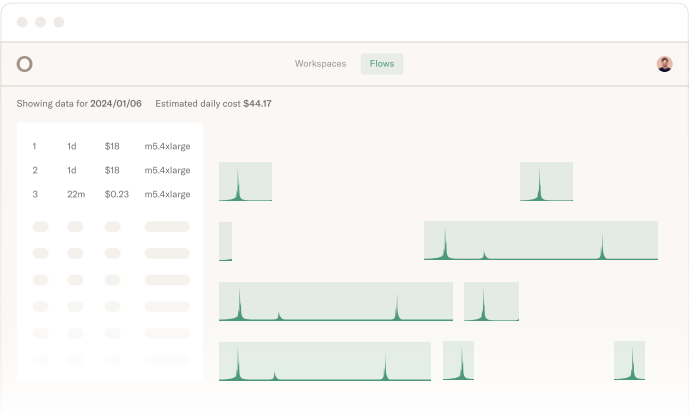



```
@card (type="blank",id="gen_ai_results")
@gpu_profile(type=1
@kubernetes(gpu=4)
@step
def generate(self):
    from utils import load_model
    model = load_model('Stable Diffusion')
    result = .
    current.card.append(result)
    self.next(self.end)
```

Use your existing cloud accounts for compute without extra margin, moving between clouds effortlessly. You can also bring on-prem resources in the mix.

Scale your systems up and out in plain Python using your existing code and favorite libraries. No need to learn new paradigms, complex APIs, or limited environments.



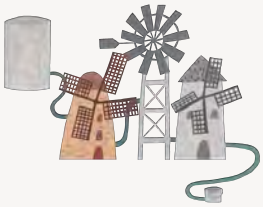


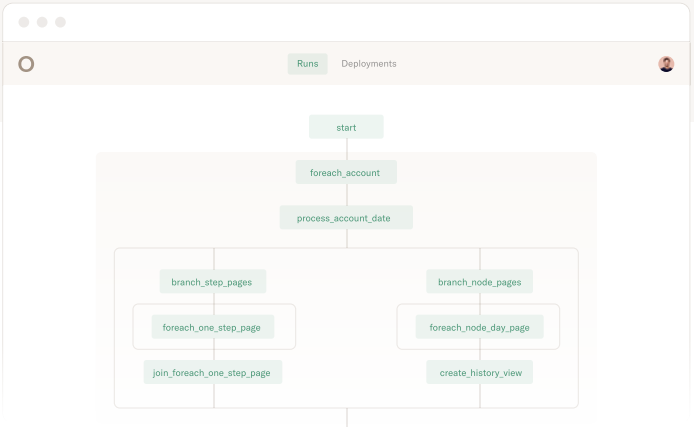
Train demanding models or fine-tune LLMs with fleets of GPUs with distributed training - Ray, Deepspeed, PyTorch, and MPI are supported out of the box.

Compute costs are readily observable in the UI. You can attribute costs down to individual workflows and flows, minimizing costs where it matters.

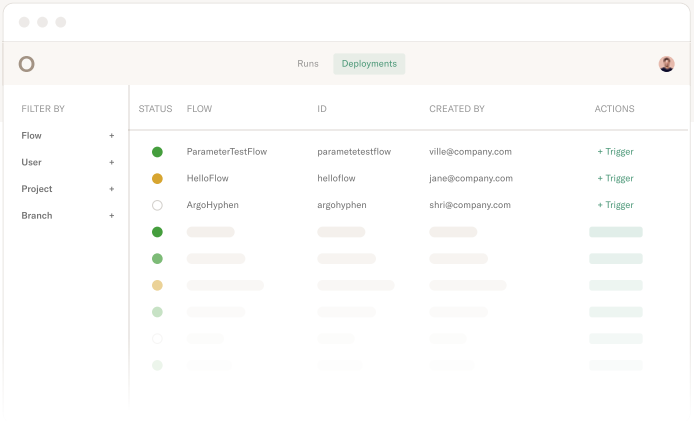
Orchestration

Outerbounds supports projects from experimentation to production. Compose complex, real-life systems from modular components and deploy them in a highly-available production environment with a single click - or more often, through a CI/CD system.

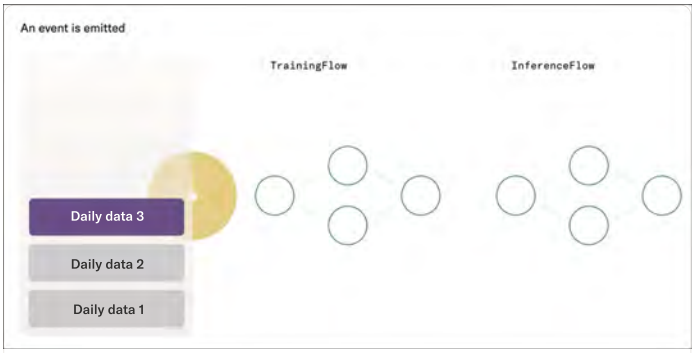




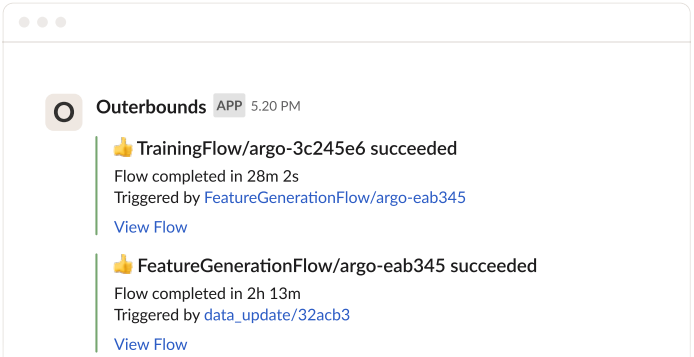
Develop production-ready workflows quickly with open-source Metaflow that has been battle-hardened for years at Netflix and other leading companies.



Deploy workflows with a single click and make them run automatically in stable, isolated execution environments, connected to other systems upstream and downstream.



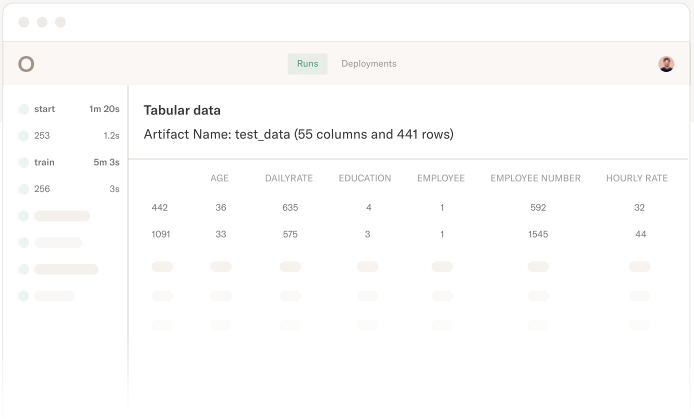
Build increasingly advanced systems incrementally by composing larger flows from individual components, dividing responsibilities across teams.



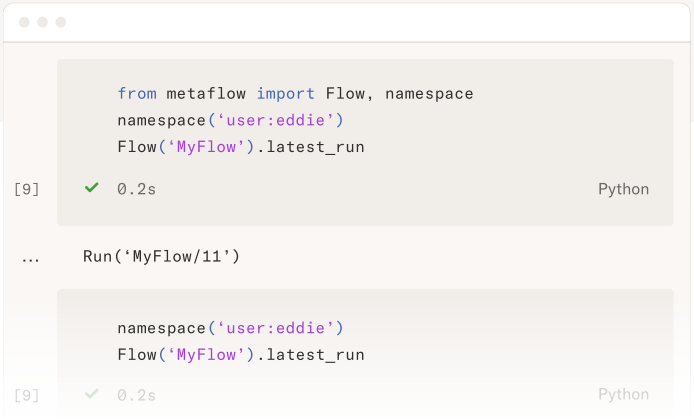
Focus on operating your data, models, and applications with full visibility - Outerbounds keeps the foundational infrastructure running.

Versioning & Observability

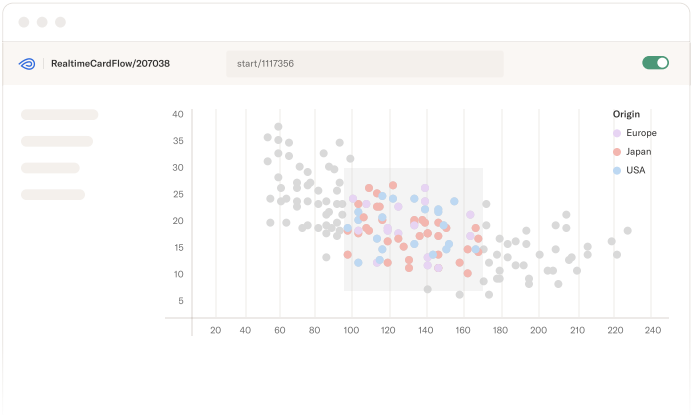
Outerbounds tracks and organizes projects - and models and data they contain - automatically. Build observable ML/AI systems, while drawing secure boundaries between projects, versions, and teams.



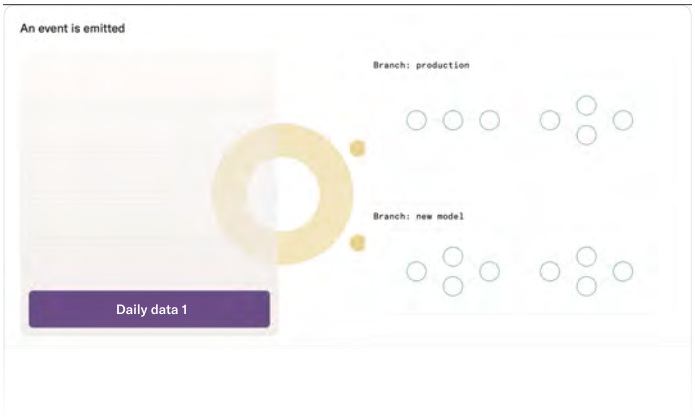
Metaflow tracks, records, and versions all data, code, and models automatically, providing a built-in model registry and experiment tracker.



Access past results with a simple Python API, reuse them in workflows and explore, analyze, and debug them in notebooks or programmatically.



Visualize custom metrics and KPIs with real-time dashboards and reports, which are natively integrated in the system through a simple Python API.



Deploy and operate any number of system variants, e.g. for A/B testing, knowing that deployments are safely isolated from each other.

Boost productivity and happiness with

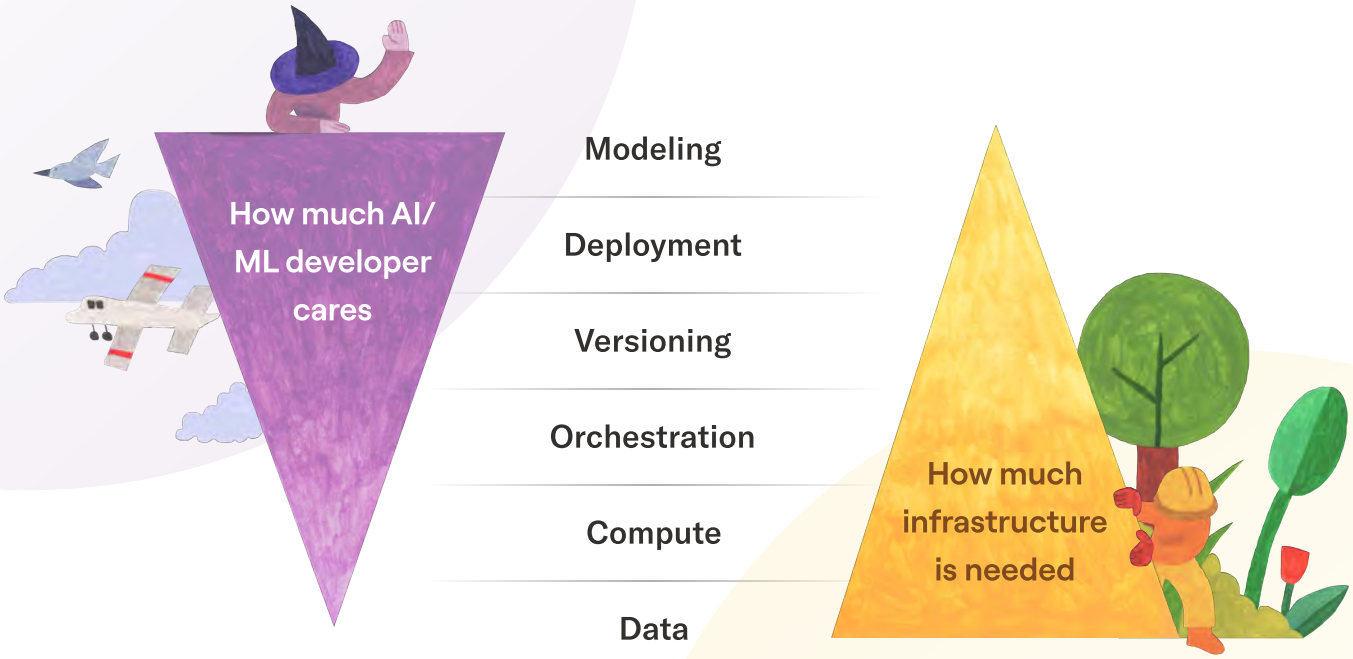
Delightful developer experience

TRADE REPUBLIC

Outerbounds has increased our appetite for model based solutions. We can use ML in places where we hadn't considered using it before, now that we know that we can have a reliable model in production quickly.

Thanasis Noulas | VP of Engineering

Instead of having to navigate disparate tools, **Outerbounds offers a single developer-friendly API and a coherent UI covering the full ML and AI stack,** so developers with diverse expertise can focus on building and deploying systems rapidly.



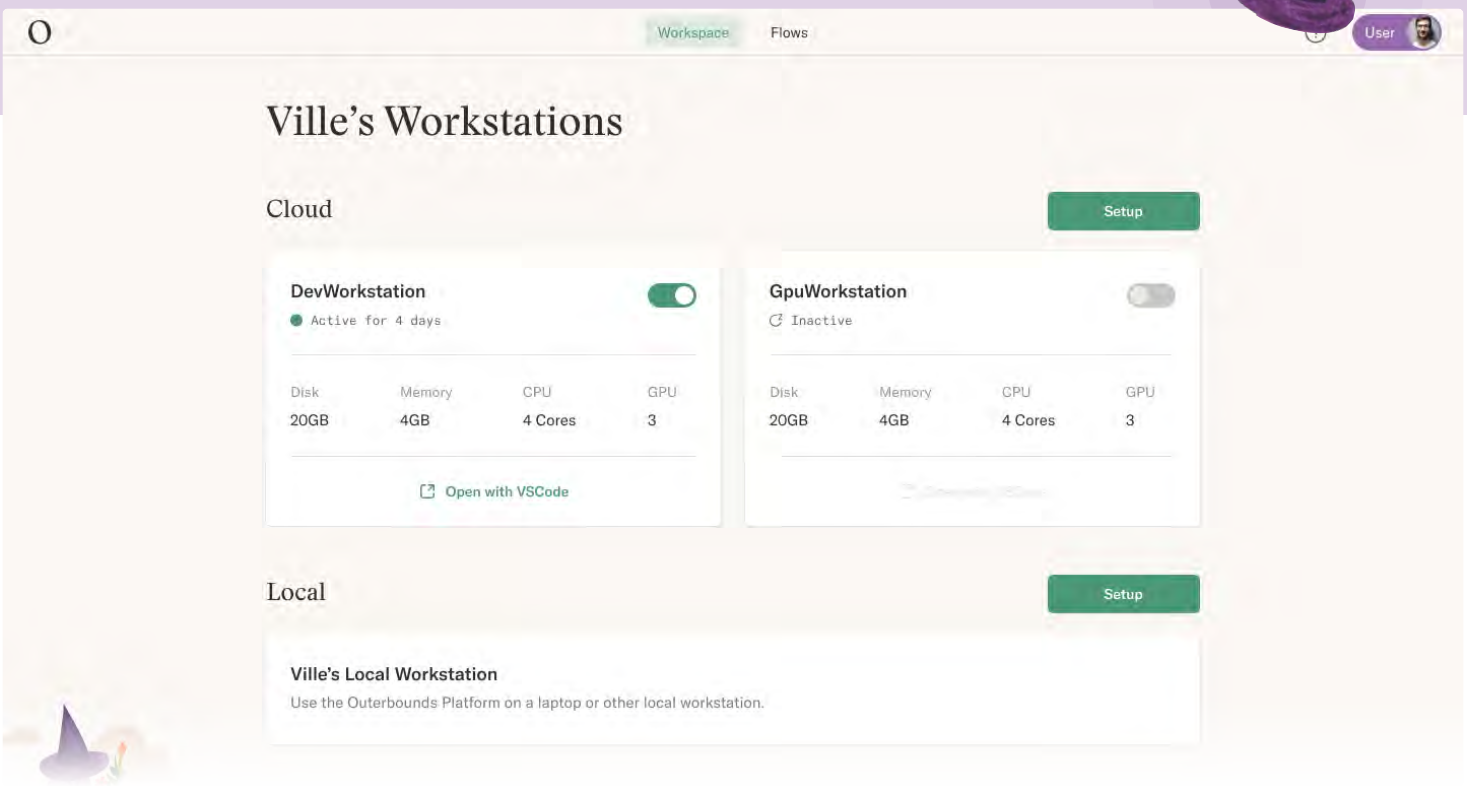
Gone are the days when AI/ML systems were treated as islands, separate from other production systems. Today's systems are **built on infrastructure and policies which make DevOps, SREs, and other engineers happy** too.

TALA

Our goal has been maximizing the utilization of AWS resources and therefore minimizing costs. For us, Outerbounds is a very big efficiency gain, just in terms of efficient utilization of AWS resources.

Will High | Head of ML

Making ML/AI developers productive



Say bye to mismatching dependencies, underpowered laptops, clunky configurations, and yesterday's IDEs. Outerbounds workstations are accessible from local VSCode, so you can **develop code quickly and conveniently with a modern toolchain**, while analyzing data with familiar, managed Jupyter notebooks on the side.

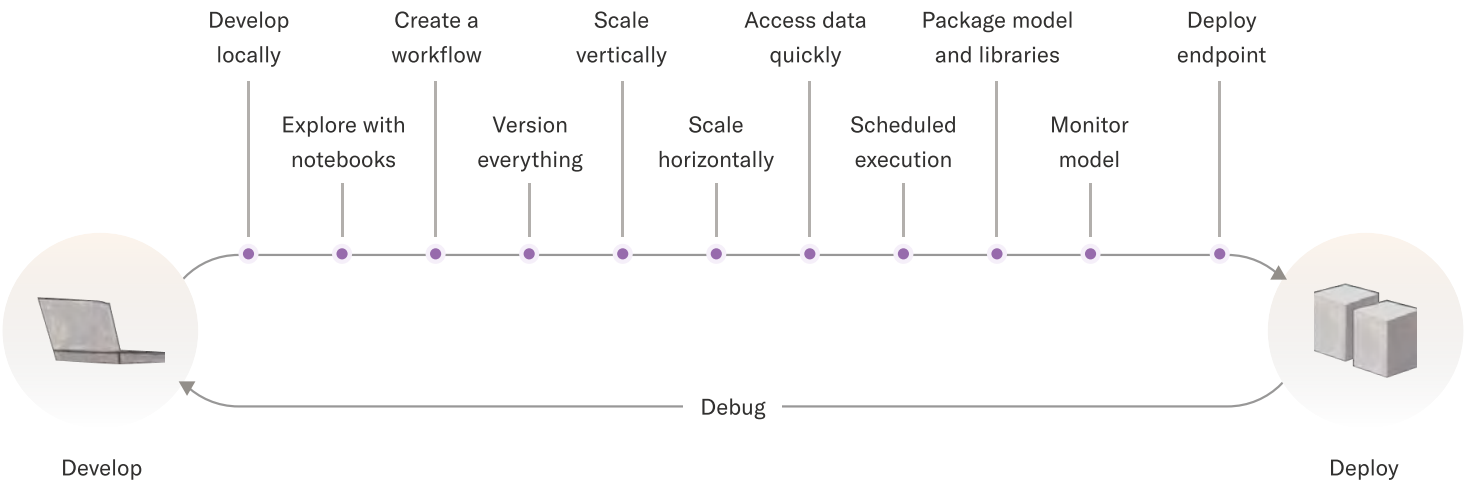
```python
from metaflow import FlowSpec, step. conda_base,\
                     kubernetes, schedule

@conda_base(libraries={'scikit-learn':'1.1.2'})
@schedule(daily=True)
class HelloFlow(FlowSpec):
    @step
    def start(self):
        self.x = 1
        self.next(self.end)

    @kubernetes(memory=64000)
    @step
    def end(self):
        self.x += 1
        print("Hello world! The value of x is", self.x)

if __name__ == '__main__':
    HelloFlow()
```

Stop writing boilerplate and stitching systems together with YAML, Docker, Python, and clunky cloud APIs. Developed by us at Netflix, open-source Metaflow has set the standard for human-centric, Python-first, ML/AI workflows for years.

Productive ML/AI developers

Technical debt of ML and AI systems is often blatantly visible, not hidden

Best ML/AI systems are designed coherently end-to-end, not stitched together from second-hand parts

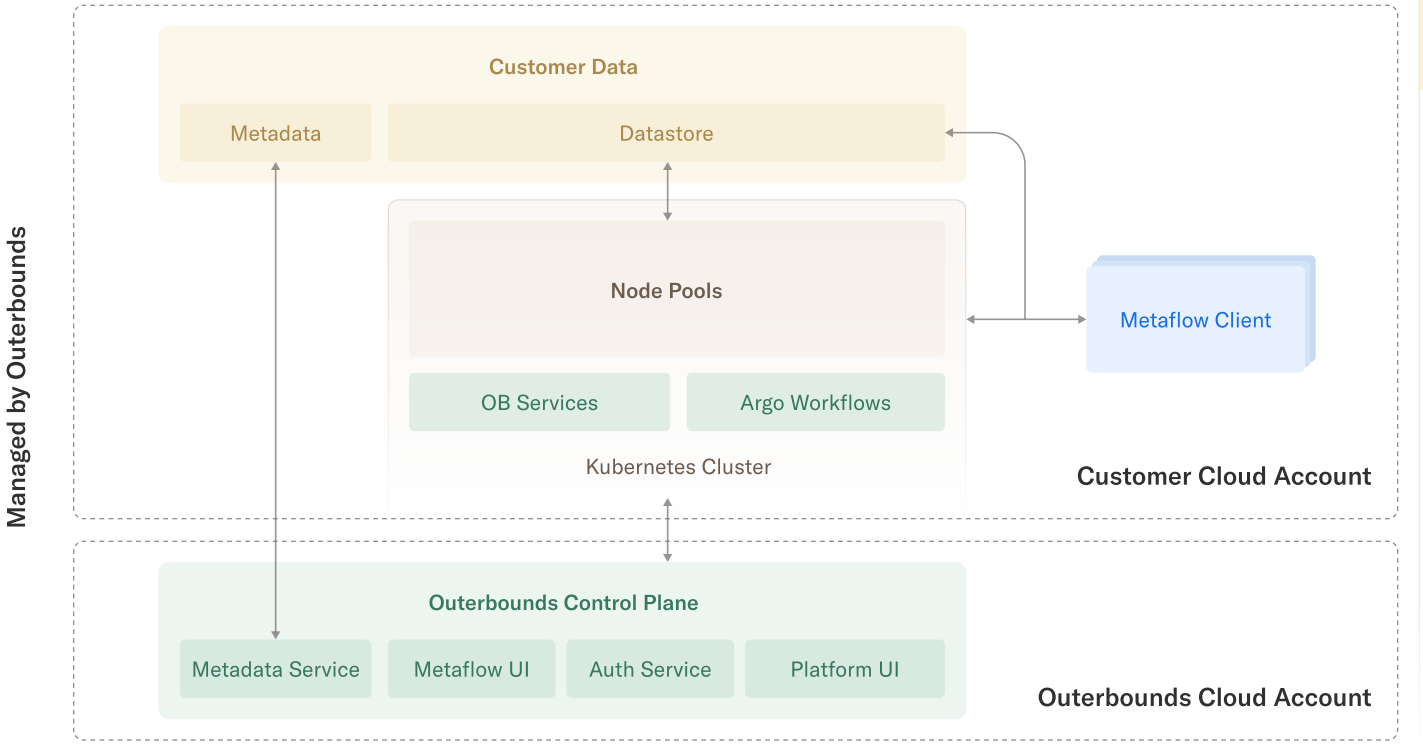


carsales

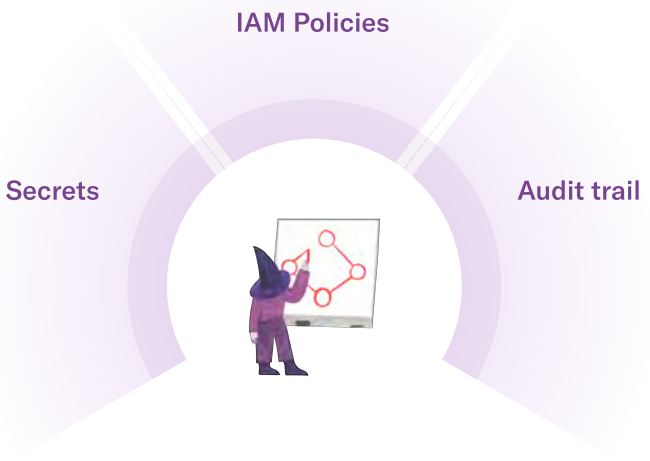
Metaflow has been a fantastic tool for Carsales. Built with the core idea of allowing ML engineers and data scientists to use Python as a native way to work, this friendly approach has tremendously boosted Carsales' productivity in ML related projects.

Samuel Than | Senior MLOps Engineer

Built on robust infrastructure



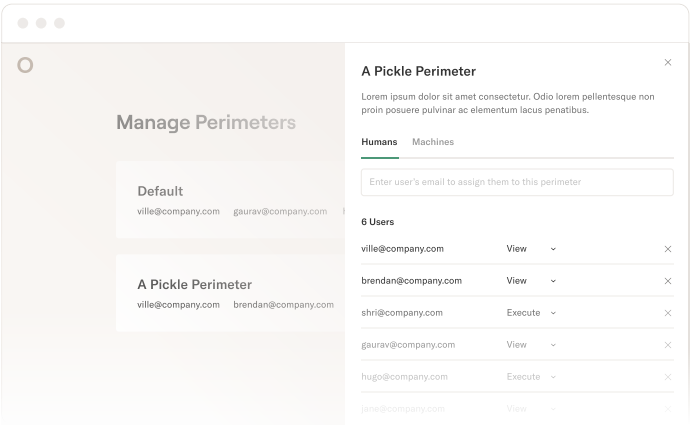
With a few clicks, Outerbounds deploys as a fully managed Kubernetes cluster in your AWS, GCP, or Azure account, fenced inside a strict permission boundary. **No data or compute ever leaves your premises**, so you can integrate ML/AI into your existing systems seamlessly, leveraging existing security policies. Outerbounds manages the cluster 24/7, taking care of upgrades and other maintenance with zero downtime.

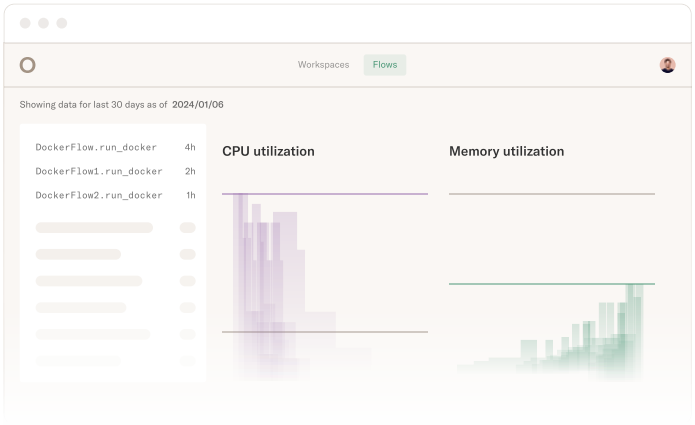


All access to the platform is authorized through your SSO provider, while security perimeters and role based access control allows you to delineate permissions by team and environment.

Built on robust infrastructure

As an administrator, you can configure data access, compute resources, and deployment settings centrally for ML and AI teams. And, you can observe the status of the system conveniently in the same UI.





TRADE REPUBLIC

We are a bank, everything we do needs to be auditable. This means we need to be able to reproduce everything that has been in production. Outerbounds gives us that for free, as all models and metadata are versioned. I sleep much more comfortably knowing this.

Thanasis Noulas | VP of Engineering

Develop ML/AI systems, not just models

- Seamless data integrations
- Production-grade orchestration
- Cost-efficient compute at scale
- Built-in tracking and versioning

WADHWANI AI

Outerbounds has proven to be transformative for us

The ability to scale our operations both vertically and horizontally, and launch parallel experiments and pipelines, has made Metaflow and Outerbounds an appealing choice for our ML team, significantly enhancing our productivity, without the burden of worrying about infra.

Soma S Dhavala | PhD, Director, Wadhwani Institute for AI

Get started with free 30 day trial

Appendix

Outerbounds vs. Open-Source Metaflow

How does the managed Outerbounds platform differ from open-source Metaflow?

	Outerbounds	Open-source Metwflow
Developer-friendly API	Same open-source Metaflow	
No lock-in, build apps with open-source APIs	Same open-source Metaflow	
Version and track everything	Same open-source Metaflow	
Simple access to scalable compute	Same open-source Metaflow	
Deploy to production with a single click	Same open-source Metaflow	
Deploys securely in your cloud account	Yes	Yes
Unlimited compute at no extra cost	Yes	Yes
Secure data integrations	Included	Basic version in OSS
Scalable compute backend	Included w/ additional features	Basic version in OSS
Highly-available production orchestration	Managed and optimized	Basic version in OSS
Durable metadata	Managed and optimized	Basic version in OSS
Cloud workstations	Included	
Comprehensive UI	Included	
Multi-cloud compute	Included	
Platform- and task-level performance metrics	Included	
Cost tracking and optimization	Included	N/A
Auth via SSO and machine tokens	Included	
Role-Based Access Control	Included	
Multiple isolated environments	Included	
Audit logs	Included	
Fully managed with 24/7 dedicated support	Included	

Appendix

Outerbounds vs. Amazon Sagemaker

Amazon Sagemaker is a portfolio of over 20 AWS products targeting ML and AI use cases. True to Amazon's philosophy, these products are built by separate teams in a loosely coupled manner, leaving it to the customer to learn and integrate the parts, some of which are more immature than others, into a working system.

Differences	Outerbounds	Sagemaker
Developer Experience	Cohesively designed, integrated platform with a particular focus on clean Python APIs and easy operations. **Result: Develop and deploy end-to-end systems quicker.**	A portfolio of loosely coupled products with inconsistent APIs and varying levels of maturity. **Result: Building and operating systems takes significantly more effort.**
Cloud Agnosticity	The same APIs work across clouds and on-prem. No changes needed to migrate to other clouds. **Result: Leverage multiple clouds for cost-optimization and flexibility.**	The APIs are specific to AWS. Complete rewrite needed to migrate to other clouds. **Result: Complete lock-in to AWS.**
Cost	Utilize the lowest-cost cloud instances without extra margin. Outerbounds is incentivized to minimize your AWS bill. **Result: Lower, transparent, predictable costs, higher ROI.**	Must use ml. instances which are normal EC2 instances with an extra margin. AWS is incentivized to maximize your AWS bill. **Result: Higher, opaque, unpredictable costs, lower ROI.**
Support	Dedicated Slack channel with experienced engineers, supporting both ML/AI developers and infrastructure. **Result: Faster time to market, quicker time to resolution.**	Standard AWS support - access to knowledgeable engineers is very expensive. Very limited support for end-user developers. **Result: Slower development time, more surprises.**

Similarities	Outerbounds	Sagemaker
Adopt easily in your AWS account	Outerbounds deploys in your AWS account through AWS marketplace, becomes a line-item in your AWS invoice.	Sagemakers runs in your AWS account, becomes a line-item in your AWS invoice.
Interoperability	Works seamlessly with other AWS services, authorized via IAM policies. You may leverage any Sagemaker services with Outerbounds.	Works seamlessly with other AWS services, authorized via IAM policies. You must mix-and-match Sagemaker services you want to utilize.
Fully Managed	The platform relies on foundational AWS services like EC2 and S3. It is fully managed by Outerbounds 24/7 with a guaranteed SLA.	The platform relies on foundational AWS services like EC2 and S3. It is fully managed by AWS 24/7 with a guaranteed SLA.

Case Studies

Amazon uses Metaflow to power recommendations for Prime Video, thanks to its excellent developer experience.

Autodesk a marquee Sagemaker customer, uses Metaflow with Sagemaker to enable their developers to be productive with demanding Generative AI use cases, and to run large-scale workloads on AWS cost-efficiently.

Outerbounds

Smarter machines, built by happier humans

outerbounds.com | sales@outerbounds.co