UbiOps

Run and manage AI at scale

Easily manage, train and run your AI / ML jobs in one place.

UbiOps is an AI infrastructure platform to help teams to quickly run their AI & ML workloads as reliable and secure microservices, without upending their existing workflows.

Whether you are a start-up looking to launch an AI product, or a data science team at a large organization. UbiOps will be there for you as a reliable backbone for any AI or ML service.

Optimize compute with rapid adaptive scaling

Scale your AI workloads dynamically with usage without paying for idle time. Accelerate model training and inference with instant on-demand access to powerful GPUs enhanced with serverless, multi-cloud workload distribution.

Easily handle peak loads for notoriously compute-intensive LLMs, CV, and generative models with automatic scaling and zero-scaling.

Build modular applications

Improve the efficiency and scalability of your applications with Pipelines: UbiOps' unique workflow management system.

Re-use and combine multiple deployments into a workflow with its own unique API.

Multiple compute environments, one interface

UbiOps supports hybrid and multi-cloud workload orchestration. Deploy models on your own infrastructure or private cloud, where data and models never leave your environment.

Or scale out to hybrid and multi-cloud environments to optimize for costs, compliance and guaranteed compute resources.
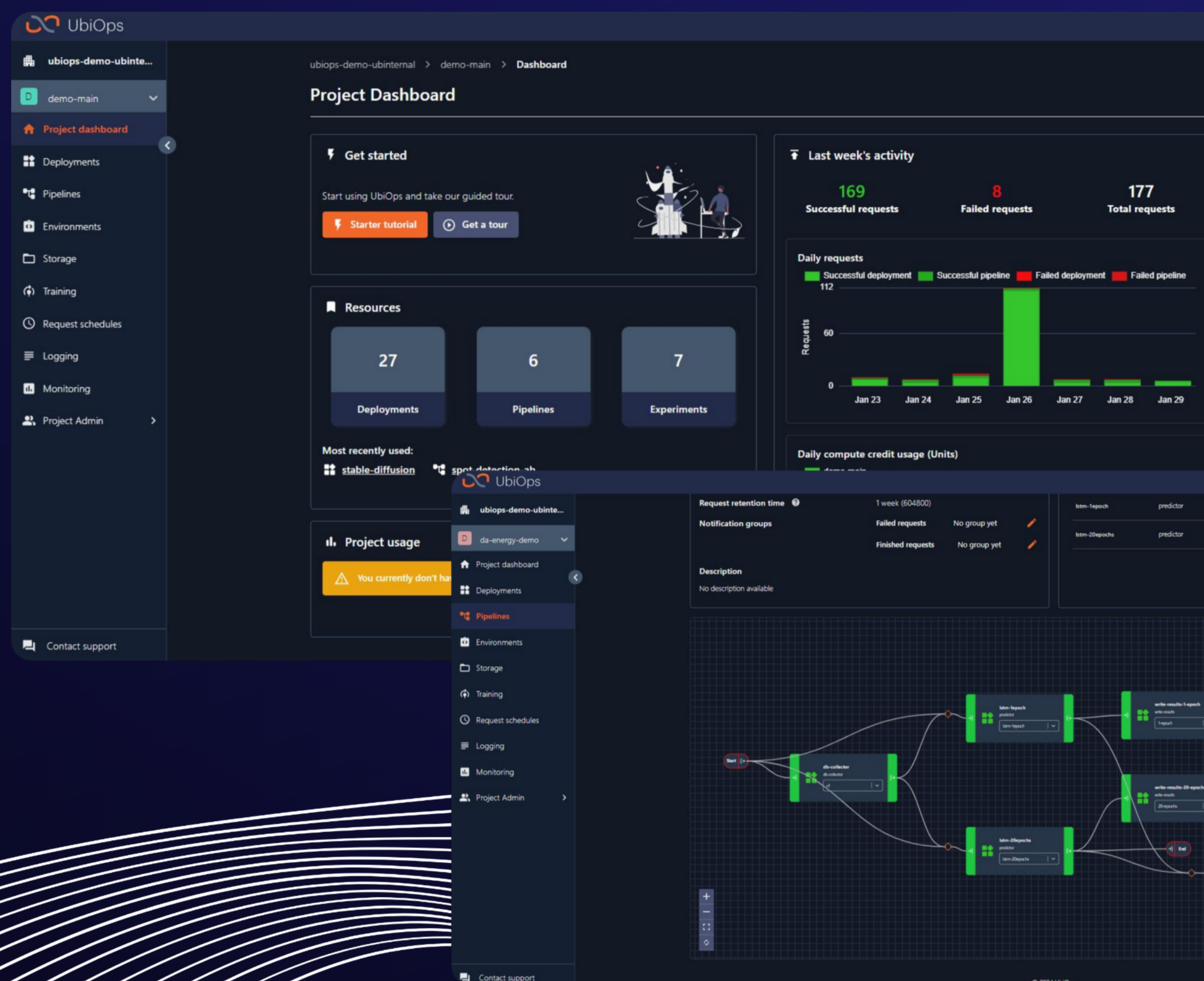
Trusted by the world's premier technology and solution providers. ISO27001 certified.

BAYER | NVIDIA | ASSET Rail | bam | studio automated | DUCKDUCKGOOSE | iv-Groep | EJLOGON.AI | SIEMENS | mendix Go make it | PHENO-INSPECT | Gradyent

The fastest route to production-grade ML / AI workloads

There's so much you can build with UbiOps at your fingertips. Deploy your Python & R code on UbiOps and instantly scale in the cloud.

Use it for real-time serverless inference or long-running jobs. You can build single services as well as large workflows.

Choose between efficient CPU or accelerated GPU instances.

"

"The on demand offering of UbiOps ensures that there's GPU availability, with the option to scale very rapidly.
*Also, with **UbiOps' scale-to-zero functionality we don't need to pay for GPU resources if the application is not being used,** e.g., off-season."*

BAYER Crop Science

Without UbiOps, **we would easily have spent six more months on the go-to-market strategy.** And then we would still have had to find good full-stack developers if we had had to build everything ourselves.

"

BELLOGON.AI

On-demand rapid scaling of **AI and ML workloads on GPU**

Instant scaling

UbiOps ensures that there's on-demand GPU availability with the option to scale instantly.

High throughput

UbiOps is designed for high throughput, real-time workloads and reliable processing with 99.99% uptime.

Pay-as-you-go

UbiOps' scale-to-zero functionality ensures that you don't need to pay for GPU resources if your application is not processing data. Saving your team from under- and over-provisioning of GPUs.

No DevOps

With UbiOps there is no need for an upfront investment in DevOps and IT.

UbiOps

Run and manage AI at scale

Book your demo

Follow us!