

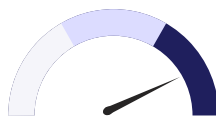
Anyscale Compute Platform

Develop AI with unmatched scale, performance, and efficiency



Save >50% on costs

[Handshake](#) saved over 50% on LLM GPU cloud compute costs with Anyscale.



100% resource utilization

Companies like [Canva](#) and [SewerAI](#) achieve more efficient GPU utilization and reduced bottlenecks on data ingestion.



Upto 12x faster iteration

With Anyscale, [Canva](#) iterates 12X quicker while reducing costs by 50%, leading to cheaper and faster product delivery.

PRODUCT OVERVIEW

The Compute Platform for Any AI/ML Workload

The Anyscale Platform is a first-of-its-kind compute platform for any AI and GenAI workload. With Anyscale, you can develop, scale, and deploy AI/ML workloads in one place.

From data processing to state-of-the-art GenAI model training to LLM fine-tuning and inference and beyond, Anyscale is made for how you work today and how you want to scale tomorrow.

Trusted by industry giants like OpenAI, Uber, and Netflix, Ray defines the standard for AI infrastructure. Anyscale builds on the foundation of Ray, with proprietary enhancements that bolster performance, reliability, efficiency, and scale while speeding up development and productivity.

FEATURES



Smart Infrastructure Management

Simplify operations, reduce infrastructure management time, and enhance AI workload performance with a fully managed compute platform on AWS, GCP, and on-prem. Plus, get optimized autoscaling and automated instance selection.



Optimized Ray Compute Engine

Get enhanced performance, reliability, scale, and cost efficiency with Manta, the compute engine on the Anyscale Platform. Designed for a variety of AI/ML workloads including data processing, model training, and serving, Manta is fully compatible with Ray, the AI industry standard.



Powerful Developer Tooling

Scale development work with Anyscale Workspaces, which integrates with IDEs like VSCode and Jupyter. With built-in observability, log viewing, and distributed debugging, developers can process more data, train more models, and deploy up to 12X faster.



Enterprise Governance

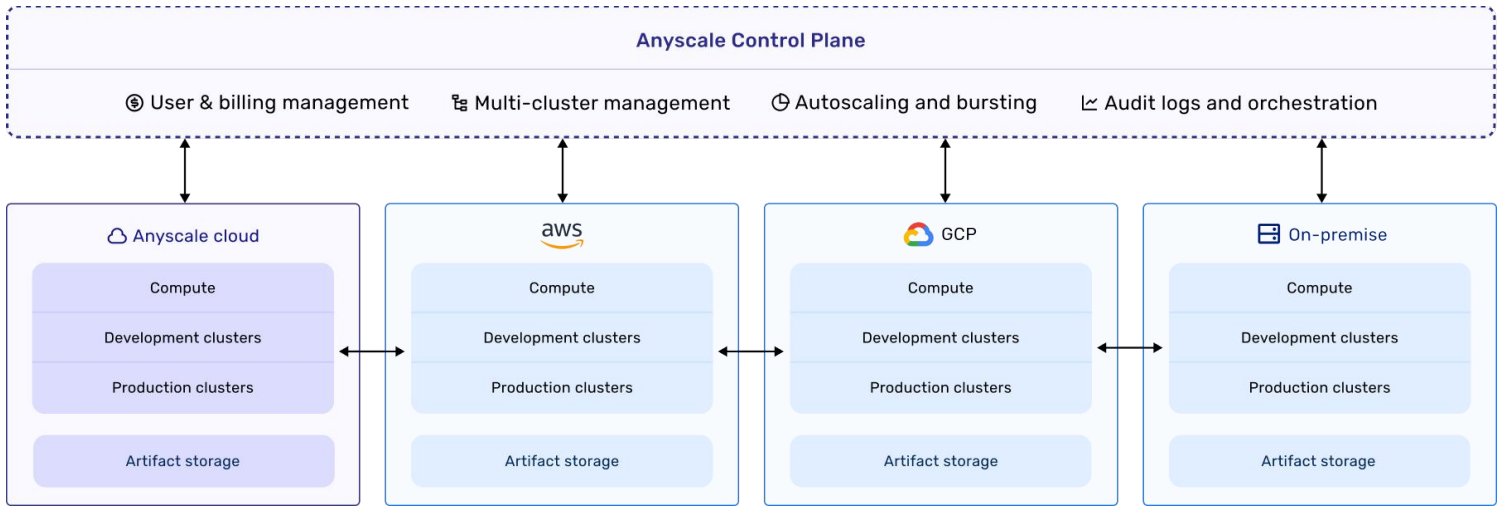
Manage compute across teams with Quota Management and ensure resource allocation with Priority Scheduling. Enjoy built-in security features like user isolation, access controls, private networking, and secure deployment options.

THE COMPUTE PLATFORM FOR ANY AI / ML WORKLOAD

DATA	TRAINING	SERVING	Gen AI
Data preprocessing Feature Engineering Image processing Video processing	Model training Time series forecasting Many model training Traditional AI Hyperparameter tuning	Multi-model inference Model endpoints Model multiplexing Pipelines Business Logic Batch inference	LLM Fine-tuning and Inference LoRA Serving Stable Diffusion RAG Application hosting Evaluations Embeddings

HOW DOES IT WORK?

Anyscale, the AI compute platform powered by Ray, combines the scalability and performance of advanced computing with the flexibility of cloud-native infrastructure, solving for any AI workload. Anyscale's Compute Platform can serve as an end-to-end AI Platform with integrations to your broader ML Ops ecosystem or serve as the universal compute foundation for existing AI Platforms via powerful APIs.



BENEFITS

- ✔ **BEST-IN-CLASS PERFORMANCE & EFFICIENCY**

Anyscale's compute platform leverages fractional heterogeneous resources and optimized infrastructure management to deliver unmatched efficiency, scale, reliability, and cost.
- ✔ **ENHANCED DEVELOPER PRODUCTIVITY**

Anyscale simplifies distributed systems with powerful APIs and tooling, helping developers reach production up to 12X faster.

- ✔ **TRUSTED BY TOP AI COMPANIES**

Anyscale is built by the creators of Ray, the framework trusted by OpenAI, Spotify, Uber and more for their AI workloads. Our team of experts is constantly pushing the boundaries of AI innovation.
- ✔ **FUTURE PROOF & ENTERPRISE READY**

Built on Ray, Anyscale's cloud-native platform provides first-class support for GPUs and is designed for distributed computing at scale.