Microsoft Azure + Pinecone

Company overview

Pinecone is the first natively built vector database, offered as a managed SaaS application built for the enterprise. With a few clicks, developers and business users alike can create an account and load in unstructured data as vector embeddings ready to be queried in minutes.

Pinecone is purpose built for the enterprise with first-class performance, cost and scalability to power the largest GenAI applications in production.

Solution benefits

Pinecone helps accelerate GenAI into production

- **Scalability** - Start small and grow seamlessly with your application
- **Ease of use -** Eliminate worry of managing resources or scalability
- **Performance** - Maximize cost/performance ratio by paying for what you use

Target customers

- Enterprise customers
- AI native startups
- Machine Learning teams

Use cases

- RAG, chatbots
- Recommendation systems
- Semantic search
- Classification

Customer References

- Adobe
- Cisco
- Gong
- CVS Health

Key differentiators

Scalability: Scale up to massive data volume (billions of vectors) while maintaining first-class performance for you application.

Time–to-value: Managed SaaS, ready to query in a few clicks.

Cost/Performance: Start small and scale up with your workloads only paying for what you use.

Average deal size

$80K-$500k
Target deals

1-2 months
Deal cycles

Start for free with our free trial

Consumption/pricing

- Pinecone hosted
- ML/AI workloads in customer account
- Usage based pricing / pay for consumption and storage

Availability

- Azure Marketplace
- AMER, EMEA

Solution integrations

Azure Open AI, Azure Blob Storage

Discovery questions

- Do you have complex and/or multi-modal data?
- Does your customer have data volumes over 10m embeddings?
- Does the customer plan on using multiple models in production?
- How many customers do they have, and is managing multi-tenancy a burden?

> Pinecone saves us time and money by eliminating the need for a full-time data scientist to manage our vector database. It's simple, fast, and retrieves information in just one millisecond, so we never worry about latency"

Kazuki Nakayashiki, CEO at Glasp

Learn more: Pinecone Learning Center
Partner contact: bryan@pinecone.io
Azure contact: v-bcordes@microsoft.com