

| Pricing Component | AWS OpenSearch Serverless | Pinecone (Standard) |
|--------------------------------------|---|---|
| Compute unit price | \$0.24 per OCU-hour(1 OC U = 6 GB RAM + vCPU) (Instaclustr) | \$0.0012 per pod-minute(\$0.072 per pod-hour) (Pinecone Docs) |
| Storage price | \$0.024 per GB-month | \$0.00000056 per GB-minute(≈\$0.025 per GB-month) (Pinecone Docs) |
| Minimum required compute | 2 OCUs (one per AZ) for indexing + 2 OCUs for search = 4 OCUs running | 1 pod (no redundancy by default) |
| Example compute cost (720 h) | 4 OCUs × 720 h × \$0.24 = \$691.20 | 1 pod × 720 h × \$0.072 = \$51.84 |
| Example storage cost (100 GB) | 100 GB × \$0.024 = \$2.40 | 100 GB × \$0.025 = \$2.50 |
| Total example monthly | \$693.60 | \$54.34 |

| Pricing Component | Pinecone Serverless | AWS OpenSearch Serverless |
|------------------------------|---|---|
| Compute pricing | • Read Units (RUs): \$16 per 1 M reads• Write Units (WUs): \$4 per 1 M writes (SaaSwothy) | • Indexing OCUs: \$0.24 per OCU-hour• Search OCUs: \$0.24 per OCU-hour (Instaclustr) |
| Storage pricing | \$0.33 per GB-month (SaaSwothy) | \$0.024 per GB-month (S3-backed) (Instaclustr) |
| Minimum configuration | • None—scales to zero when idle | • Production: 2 OCUs for indexing + 2 OCUs for search = 4 OCUs minimum (Amazon Web Services, Inc.) |

**Example
small
workload**

Assumptions: 1 M reads,
1 M writes, 100 GB stored,
idle otherwise

- Reads: $1 \times \$16 = \16
- Writes: $1 \times \$4 = \4
- Storage: $100 \times \$0.33 = \33

Total \approx \$53

Assumptions: index always-on at 4 OCUs;
100 GB stored

- Compute: $4 \text{ OCUs} \times 720 \text{ h} \times \$0.24 = \$691.20$
- Storage: $100 \times \$0.024 = \2.40

Total \approx \$693.60

Key takeaways:

- **Pinecone Serverless**

- No “always-on” units—you only pay for the vector operations you actually perform (reads + writes) and for the GB-months you store.
- Scales to zero when idle.
- Example small-scale bill (1 M reads + 1 M writes + 100 GB storage): **\approx \$53 / month.**

- **AWS OpenSearch Serverless**

- Billed by OpenSearch Compute Units (OCUs), each OCU = 1 vCPU + 6 GiB RAM, at \$0.24/hour for both indexing and search.
- Requires a minimum of 4 OCUs to span AZs in production (2 for indexing + 2 for search).
- Example small-scale bill (4 OCUs always-on + 100 GB storage): **\approx \$693.60 / month.**

If your vector-search workload is light or highly variable, **Pinecone’s serverless model** can be an order of magnitude cheaper—especially at low scale—because you’re not paying for idle capacity. Conversely, if you need AWS-native search with predictable OCU-based billing and tight integration with other AWS services, **OpenSearch Serverless** may be the better fit.