Cost Estimate

Region: us-east-1 **Scale**: ~30 papers/month (1 per day) **Last Updated**: November 2025

Monthly Breakdown

Service	What It Does	Cost
RDS PostgreSQL (db.t4g.micro)	Knowledge graph database	$12.41
RDS Storage (20 GB)	Database storage + backups	$4.30
VPC Endpoints (ECR, SQS)	Private network access	$21.91
CloudWatch Logs	Application logging	$2.53
Lambda + ECR	ML processing (mostly free tier)	$1.05
S3 Storage	Papers, UI, exports	$0.13
Route53 (optional)	Custom domain	$0.50
Total		**~$42-43/month**

Most other services (API Gateway, Cognito, SQS, CloudFront) fall under AWS free tier.

Why This Cheap?

- **Serverless**: Lambda only runs when processing papers, not 24/7
- **VPC Endpoints**: $22/month vs $35/month for NAT Gateway
- **Single-AZ RDS**: Half the cost of Multi-AZ (fine for MVP)
- **Small instance**: db.t4g.micro is plenty for <1M graph nodes
- **Free tiers**: API Gateway, Cognito, SQS, CloudFront all free

If we used EC2 instead of serverless: **~$120/month** (3x more expensive)

Cost Per Paper

Processing 30 papers costs about $42/month, which breaks down to:

- Fixed costs: ~$40/month (RDS, VPC endpoints, monitoring)
- Variable costs: ~$2/month for 30 papers

Per paper: About $0.002 in variable costs (basically free to process more)

This means you could process 10x more papers (~300/month) for only ~$45/month total.

Scaling Costs

Papers/Day	Papers/Month	Est. Monthly Cost	Notes
1	30	$42	Current plan
10	300	$45	Same infrastructure
30	900	$50	Might upgrade RDS
100	3,000	$90	Need db.t4g.small

The architecture handles 10x growth with minimal cost increase because most costs are fixed (database, networking).

At Production Scale (1,000 papers/day)

You'd need different infrastructure:

Service	Config	Cost
SageMaker	GPU inference	$250-400
Neptune	Graph database	$350
ElastiCache	Query caching	$15
RDS	Metadata	$150
Lambda	Orchestration	$50
S3 + Networking	~1 TB storage	$125
Total		**~$1,000/month**

That's $0.033 per paper at scale.