Inventory Demand Forecasting System Cost Analysis Report

Executive Summary

This comprehensive cost analysis provides detailed pricing estimates for the Inventory Demand Forecasting System, a serverless architecture built on AWS. The system leverages Lambda, API Gateway, DynamoDB, S3, and Amazon Bedrock to deliver scalable demand forecasting capabilities.

Total Estimated Monthly Cost: $16.71

Architecture Overview

The system uses the following AWS services: - **AWS Lambda**: Serverless compute for business logic and data processing - **Amazon API Gateway**: RESTful API endpoints for frontend integration - **Amazon DynamoDB**: NoSQL database for inventory, sales, and forecast data - **Amazon S3**: Object storage for CSV uploads and static assets - **Amazon Bedrock**: Machine learning for demand forecasting models

Pricing Model

This analysis is based on **ON DEMAND** pricing with the following assumptions: - US East (N. Virginia) region pricing - Standard service configurations without reserved capacity - Average usage patterns with moderate seasonal peaks - No cross-region replication or advanced features

Detailed Cost Breakdown

AWS Lambda

- **Usage**: 100,000 API requests per month with 512MB memory, 1s average execution
- **Unit Pricing**:
  - Requests: $0.0000002 per request
  - Compute: $0.0000166667 per GB-second
- **Calculation**: $0.0000002 × 100,000 + $0.0000166667 × 50,000 GB-seconds = $0.85
- **Monthly Cost**: $3.34
- **Free Tier**: First 12 months: 1M requests/month and 400,000 GB-seconds/month free

Amazon API Gateway

- **Usage**: 100,000 REST API requests per month

- **Unit Pricing**: $0.0000035 per request (first 333M requests)
- **Calculation**: $0.0000035 × 100,000 = $0.35
- **Monthly Cost**: $0.35
- **Free Tier**: First 12 months: 1M API calls/month free

Amazon DynamoDB

- **Usage**: 500,000 read requests and 100,000 write requests per month across 4 tables
- **Unit Pricing**:
  - Read Request Units: $0.125 per million units
  - Write Request Units: $0.625 per million units
- **Calculation**: $0.125/1M × 0.5M + $0.625/1M × 0.1M = $0.125
- **Monthly Cost**: $12.56
- **Free Tier**: First 12 months: 25 GB storage, 25 WCU, 25 RCU free

Amazon S3

- **Usage**: 10GB storage for CSV uploads, 50,000 GET requests
- **Unit Pricing**:
  - Storage: $0.023 per GB per month
  - GET Requests: $0.0004 per 1,000 requests
- **Calculation**: $0.023 × 5GB (beyond free tier) + $0.0004 × 30 = $0.127
- **Monthly Cost**: $0.46
- **Free Tier**: First 12 months: 5GB storage, 20,000 GET requests, 2,000 PUT requests free

Cost Scaling Projections

Growth Pattern	Month 1	Month 3	Month 6	Month 12
Steady (1.0x)	$16/mo	$16/mo	$16/mo	$16/mo
Moderate (1.05x)	$16/mo	$18/mo	$21/mo	$28/mo
Rapid (1.1x)	$16/mo	$20/mo	$26/mo	$47/mo

Free Tier Benefits

During the first 12 months, AWS Free Tier significantly reduces costs: - **Lambda**: Up to 1M requests and 400,000 GB-seconds free monthly - **API Gateway**: Up to 1M API calls free monthly - **DynamoDB**: 25 GB storage and capacity units free - **S3**: 5GB storage and request allowances free

Estimated First-Year Savings: ~$150-200 annually

Cost Optimization Recommendations

Immediate Actions

1. **Leverage AWS Free Tier** for first 12 months to minimize initial costs
2. **Use DynamoDB On-Demand pricing** for unpredictable

workloads
3. **Optimize Lambda memory allocation** based on actual performance metrics
4. **Implement API Gateway caching** for frequently accessed forecast data
5. **Use HTTP APIs instead of REST APIs** where WebSocket features aren't needed

Best Practices

1. **Monitor DynamoDB patterns** and consider Provisioned mode for consistent workloads
2. **Implement Lambda warming strategies** to reduce cold start costs
3. **Use S3 Intelligent Tiering** for long-term data storage optimization
4. **Set up CloudWatch alarms** for cost monitoring and budget alerts
5. **Consider Reserved Capacity** for DynamoDB if usage patterns become predictable

Advanced Optimizations

1. **Batch API requests** where possible to reduce per-request costs
2. **Implement data compression** for DynamoDB items to reduce storage costs
3. **Use Lambda Provisioned Concurrency** only for critical low-latency endpoints
4. **Archive old sales data** to S3 Glacier for long-term cost savings
5. **Optimize Bedrock model calls** by caching frequent forecast requests

Exclusions and Limitations

This analysis excludes the following costs: - Data transfer costs between regions - Amazon Bedrock model inference costs (varies by model) - CloudWatch logging and monitoring costs - Development and maintenance costs - Third-party integrations or external APIs - Custom domain and SSL certificate costs - Backup and disaster recovery costs

Risk Factors

Cost Variability Factors

- **Seasonal demand spikes** may increase Lambda and DynamoDB usage
- **Amazon Bedrock costs** depend on model selection and inference frequency
- **Data growth** will impact S3 storage and DynamoDB costs over time
- **API usage patterns** may vary significantly from estimates

Mitigation Strategies

- Implement cost monitoring and alerting

- Use AWS Budgets to track spending
- Regular review of usage patterns and optimization opportunities
- Consider Reserved Instances for predictable workloads

Conclusion

The Inventory Demand Forecasting System offers a cost-effective serverless solution with an estimated monthly cost of $16.71. The architecture leverages AWS Free Tier benefits and pay-as-you-go pricing to minimize initial investment while providing scalability for future growth.

Key benefits: - **Low initial costs** with Free Tier benefits - **Scalable pricing** that grows with usage - **No infrastructure management** overhead - **Built-in high availability** and disaster recovery

By following the optimization recommendations, organizations can maintain cost efficiency while delivering robust demand forecasting capabilities.

This analysis is based on AWS pricing as of December 2024 and actual costs may vary based on usage patterns, regional pricing differences, and service updates.