Intelligent Document Processing (IDP) - AWS Cost Analysis Report

Executive Summary

This comprehensive cost analysis provides detailed pricing estimates for an Intelligent Document Processing (IDP) solution built on AWS serverless architecture. The solution processes 1,000 documents per month using a combination of AWS services including Bedrock, Lambda, Textract, Comprehend, and supporting infrastructure.

Total Estimated Monthly Cost: $12.39

Architecture Overview

The IDP solution consists of: - **Frontend**: React application hosted on S3/CloudFront - **API Layer**: API Gateway with Lambda functions - **Processing Pipeline**: 5 Lambda functions for document processing - **AI/ML Services**: Bedrock (Claude 3.5 Haiku), Textract, Comprehend - **Storage**: S3 for documents, DynamoDB for metadata

Pricing Model

This analysis is based on **ON DEMAND** pricing with the following assumptions: - Processing 1,000 documents per month (baseline scenario) - Average document size: 2-5 pages - Claude 3.5 Haiku model for summarization and classification - Standard configurations without reserved capacity

Detailed Cost Breakdown

Amazon Bedrock Foundation Models

- **Input Tokens**: $0.30 per 1M tokens
- **Output Tokens**: $15.00 per 1M tokens
- **Monthly Usage**: 2M input tokens, 500K output tokens
- **Calculation**: $0.30 × 2 + $15.00 × 0.5 = $8.10
- **Monthly Cost**: **$8.10**

AWS Lambda

- **Requests**: $0.20 per 1M requests
- **Compute**: $0.0000166667 per GB-second
- **Monthly Usage**: 5,000 requests, 75,000 GB-seconds
- **Free Tier**: 1M requests/month, 400K GB-seconds/month (first 12 months)
- **Calculation**: Mostly covered by free tier
- **Monthly Cost**: **$0.42**

Amazon Textract

- **DetectDocumentText**: $0.0015 per page
- **Monthly Usage**: 2,000 pages (1,000 docs × 2 pages avg)
- **Free Tier**: 1,000 pages/month (first 3 months)
- **Calculation**: $0.0015 × 2,000 = $3.00
- **Monthly Cost**: **$3.00**

Amazon Comprehend

- **Custom Classification**: $0.0005 per unit (100 characters)
- **Monthly Usage**: 1M units
- **Free Tier**: 50K units/month (first 12 months)
- **Calculation**: $0.0005 × 950K units = $0.475

- **Monthly Cost**: **$0.50**

Amazon S3

- **Storage**: $0.023 per GB-month
- **Requests**: $0.0004 per 1K PUT, $0.0004 per 10K GET
- **Monthly Usage**: 2.5GB storage, 1K PUTs, 2K GETs
- **Free Tier**: 5GB storage (first 12 months)
- **Monthly Cost**: **$0.06** (covered by free tier)

Amazon DynamoDB

- **Storage**: $0.25 per GB-month
- **Read/Write**: $0.125/$0.625 per 1M request units
- **Monthly Usage**: 0.1GB, 5K reads, 1K writes
- **Free Tier**: 25GB, 25 RCU, 25 WCU (always free)
- **Monthly Cost**: **$0.31** (mostly covered by free tier)

Amazon API Gateway

- **Requests**: $3.50 per 1M requests
- **Monthly Usage**: 1,000 requests
- **Free Tier**: 1M requests/month (first 12 months)
- **Monthly Cost**: **$0.0035** (covered by free tier)

Cost Scaling Analysis

Usage Scenarios

Scenario	Documents/Month	Monthly Cost	Key Cost Drivers
Low	500	$6.20	Bedrock: $4.05, Textract: $1.50
Baseline	1,000	$12.39	Bedrock: $8.10, Textract: $3.00
Medium	2,500	$30.98	Bedrock: $20.25, Textract: $7.50
High	5,000	$61.95	Bedrock: $40.50, Textract: $15.00

Growth Projections (12 Months)

Growth Pattern	Month 1	Month 6	Month 12	Annual Cost
Steady	$12	$12	$12	$144
Moderate (5%/mo)	$12	$15	$21	$204
Rapid (10%/mo)	$12	$19	$35	$312

Key Insights

Cost Distribution

1. **Bedrock (65%)**: Largest cost component due to token-based pricing
2. **Textract (24%)**: Second highest due to per-page processing
3. **Lambda (3%)**: Minimal cost due to efficient serverless architecture
4. **Other Services (8%)**: Storage and API costs are negligible

Free Tier Benefits

- **First Year Savings**: ~$2.50/month from Lambda, API Gateway, S3 free tiers
- **Ongoing Savings**: DynamoDB always-free tier provides continuous value
- **Textract Benefit**: 3-month free tier reduces initial costs

Cost Optimization Recommendations

Immediate Actions

1. **Document Caching**: Implement caching to reduce repeated Bedrock API calls
2. **Batch Processing**: Process multiple documents in single Textract calls
3. **Prompt Optimization**: Reduce token usage through efficient prompt engineering
4. **Lambda Right-sizing**: Optimize memory allocation based on actual usage

Best Practices

1. **Monitor Token Usage**: Track Bedrock token consumption patterns
2. **Implement Error Handling**: Reduce failed API calls and retries
3. **Use CloudWatch**: Set up cost monitoring and alerts
4. **Consider Reserved Capacity**: For predictable high-volume workloads

Advanced Optimizations

1. **S3 Intelligent Tiering**: Automatic cost optimization for long-term storage
2. **Document Preprocessing**: Reduce Textract costs through image optimization
3. **Model Selection**: Evaluate different Bedrock models for cost/performance balance
4. **Comprehend Alternatives**: Use built-in models before custom classification

Risk Factors

Cost Variability

- **Token Usage**: Bedrock costs can vary significantly based on document complexity
- **Document Size**: Larger documents increase both Textract and Bedrock costs
- **Processing Failures**: Failed operations can lead to unnecessary retries

Scaling Considerations

- **Linear Scaling**: Most costs scale linearly with document volume
- **Free Tier Expiration**: Costs will increase after free tier periods end
- **Regional Variations**: Pricing may vary across AWS regions

Assumptions and Limitations

Included in Analysis

- Standard ON DEMAND pricing for all services
- Baseline processing of 1,000 documents/month
- Average document complexity and size
- Claude 3.5 Haiku model for Bedrock operations

Excluded from Analysis

- Data transfer costs between regions
- Custom model training costs
- Development and maintenance expenses

- Monitoring and logging costs (CloudWatch)
- Enterprise support costs
- Third-party integration costs

Conclusion

The IDP solution provides a cost-effective approach to document processing with an estimated monthly cost of $12.39 for 1,000 documents. The serverless architecture ensures costs scale with usage, making it suitable for both small and large-scale deployments.

Key Takeaways: - Bedrock represents the largest cost component but provides high-value AI capabilities - Free tier benefits significantly reduce first-year costs - The solution scales cost-effectively with document volume - Multiple optimization opportunities exist to reduce operational costs

Next Steps: 1. Implement the recommended optimizations 2. Set up cost monitoring and alerts 3. Conduct regular cost reviews as usage patterns evolve 4. Consider reserved capacity for predictable workloads

---

Report Generated: November 2025
Pricing Data Source: AWS Pricing API
Analysis Period: Monthly recurring costs