

# ETL Pipeline Solution - AWS Cost Analysis Report

## Executive Summary

This comprehensive cost analysis provides detailed pricing estimates for the ETL Pipeline Solution deployed on AWS. The solution transforms Parquet files to JSON format using a serverless architecture with AWS Lambda, API Gateway, AWS Glue, S3, DynamoDB, and CloudWatch.

**Total Estimated Monthly Cost: \$30.19**

## Service Breakdown

### 1. AWS Lambda (API Functions)

- **Usage:** Processing API requests with 512MB memory allocation, 2-second average execution time
- **Monthly Volume:** 100,000 requests/month
- **Unit Pricing:**
  - Requests: \$0.0000002 per request
  - Compute: \$0.0000166667 per GB-second
- **Usage Quantities:**
  - Requests: 100,000 requests/month
  - Compute:  $100,000 \text{ requests} \times 2\text{s} \times 0.5\text{GB} = 100,000 \text{ GB-seconds}$
- **Calculation:**  $\$0.0000002 \times 100,000 + \$0.0000166667 \times 100,000 = \$1.69/\text{month}$
- **Free Tier:** 1M requests and 400,000 GB-seconds per month free
- **Estimated Monthly Cost:** **\$1.69** (within free tier for first year)

### 2. Amazon API Gateway (REST API)

- **Usage:** REST API with 100,000 requests per month for ETL job management
- **Monthly Volume:** 100,000 requests/month
- **Unit Pricing:** \$3.50 per million requests (first 333M)
- **Usage Quantities:** 100,000 requests/month
- **Calculation:**  $\$3.50/1\text{M} \times 0.1\text{M} \text{ requests} = \$0.35/\text{month}$
- **Free Tier:** No free tier for API Gateway
- **Estimated Monthly Cost:** **\$0.35**

### 3. AWS Glue (ETL Jobs)

- **Usage:** G.1X workers (4 vCPU, 16GB memory) running for Parquet to JSON transformation
- **Monthly Volume:** 20 DPU-hours/month ( $1 \text{ DPU} \times 20 \text{ hours}$ )
- **Unit Pricing:** \$0.44 per DPU-Hour
- **Usage Quantities:** 20 DPU-hours/month
- **Calculation:**  $\$0.44 \times 20 \text{ DPU-hours} = \$8.80/\text{month}$
- **Free Tier:** No free tier for Glue ETL jobs
- **Estimated Monthly Cost:** **\$8.80**

### 4. Amazon S3 (Storage)

- **Usage:** Standard storage for input Parquet files and output JSON files

- **Monthly Volume:** 100GB total storage
- **Unit Pricing:**
  - Storage: \$0.023 per GB/month (first 50TB)
  - PUT requests: \$0.0005 per 1,000 requests
  - GET requests: \$0.0004 per 1,000 requests
- **Usage Quantities:**
  - Storage: 100GB/month
  - PUT requests: 1,000 requests/month
  - GET requests: 5,000 requests/month
- **Calculation:** Storage:  $\$0.023 \times 95\text{GB}$  (after free tier) =  $\$2.19$  + PUT:  $\$0.0005 \times 1 = \$0.0005$  + GET:  $\$0.0004 \times 5 = \$0.002 = \$2.19/\text{month}$
- **Free Tier:** 5GB Standard storage free for 12 months
- **Estimated Monthly Cost:** **\$2.30**

## 5. Amazon DynamoDB (Metadata Storage)

- **Usage:** On-demand billing for job metadata storage
- **Monthly Volume:** 10,000 read/write operations per month, 5GB storage
- **Unit Pricing:**
  - Read requests: \$0.125 per million read request units
  - Write requests: \$0.625 per million write request units
  - Storage: \$0.25 per GB/month (after 25GB free)
- **Usage Quantities:**
  - Read requests: 5,000 read request units/month
  - Write requests: 5,000 write request units/month
  - Storage: 5GB/month (within free tier)
- **Calculation:** Reads:  $\$0.125/1\text{M} \times 0.005\text{M} = \$0.000625$  + Writes:  $\$0.625/1\text{M} \times 0.005\text{M} = \$0.003125 = \$0.004/\text{month}$
- **Free Tier:** 25GB storage and 25 RCU/WCU free for 12 months
- **Estimated Monthly Cost:** **\$1.25** (after free tier expires)

## 6. Amazon CloudWatch (Monitoring & Logs)

- **Usage:** Log ingestion and storage for Lambda and Glue job logs, basic monitoring metrics
- **Monthly Volume:** 10GB log ingestion, 50GB log storage, 5 custom metrics
- **Unit Pricing:**
  - Log ingestion: \$0.50 per GB ingested
  - Log storage: \$0.03 per GB/month
  - Custom metrics: \$0.30 per metric/month
- **Usage Quantities:**
  - Log ingestion: 10GB/month
  - Log storage: 50GB/month
  - Custom metrics: 5 metrics
- **Calculation:** Ingestion:  $\$0.50 \times 5\text{GB}$  (after free tier) =  $\$2.50$  + Storage:  $\$0.03 \times 50\text{GB} = \$1.50$  + Metrics:  $\$0$  (within free tier) =  $\$4.00/\text{month}$
- **Free Tier:** 5GB log ingestion and 10 custom metrics free
- **Estimated Monthly Cost:** **\$5.00**

## Usage Scenarios

### Low Usage Scenario (Monthly Cost: \$15.50)

- Lambda: 25,000 requests → \$0.42
- API Gateway: 25,000 requests → \$0.09
- Glue: 5 DPU-hours → \$2.20
- S3: 25GB storage → \$0.58
- DynamoDB: 2,500 operations → \$0.31

- CloudWatch: 5GB logs → \$2.50
- **Total: \$6.10** (with free tier) / **\$15.50** (without free tier)

### **Medium Usage Scenario (Monthly Cost: \$30.19)**

- Current estimates as detailed above
- **Total: \$18.19** (with free tier) / **\$30.19** (without free tier)

### **High Usage Scenario (Monthly Cost: \$85.50)**

- Lambda: 500,000 requests → \$8.35
- API Gateway: 500,000 requests → \$1.75
- Glue: 100 DPU-hours → \$44.00
- S3: 500GB storage → \$11.50
- DynamoDB: 50,000 operations → \$6.25
- CloudWatch: 100GB logs → \$13.65
- **Total: \$85.50**

## **Cost Optimization Recommendations**

### **Immediate Actions**

1. **Optimize Glue job execution time** by using appropriate worker types and parallelization
2. **Implement S3 lifecycle policies** to transition older files to cheaper storage classes
3. **Use DynamoDB on-demand billing** initially, consider provisioned capacity for predictable workloads
4. **Set up CloudWatch log retention policies** to manage storage costs
5. **Monitor Lambda execution time and memory allocation** for cost optimization

### **Best Practices**

1. **Use AWS Cost Explorer** to track spending patterns across all services
2. **Implement tagging strategy** for cost allocation and tracking
3. **Consider using AWS Glue Flex** for cost-effective ETL job execution
4. **Set up billing alerts and budgets** for proactive cost management
5. **Review and optimize data transfer patterns** to minimize cross-region costs
6. **Use S3 Intelligent Tiering** for automatic cost optimization of storage
7. **Implement error handling and retry logic** to avoid unnecessary re-processing costs

### **Assumptions**

- Standard ON DEMAND pricing model for all services
- US East (N. Virginia) region pricing
- No reserved instances or savings plans applied
- Standard tier services without premium features
- Typical ETL workload patterns with moderate usage
- No data transfer costs between services in same region
- Standard security and compliance requirements

### **Exclusions**

- Data transfer costs between regions
- Premium support costs

- Development and testing environment costs
- Third-party integration costs
- Custom security implementations
- Disaster recovery and backup costs beyond standard features
- Network acceleration or dedicated connections

## Free Tier Benefits (First 12 Months)

- **AWS Lambda:** 1M requests and 400,000 GB-seconds per month
- **Amazon S3:** 5GB Standard storage
- **Amazon DynamoDB:** 25GB storage and 25 RCU/WCU
- **Amazon CloudWatch:** 5GB log ingestion and 10 custom metrics
- **Total Monthly Savings:** ~\$12.00 during free tier period

## Pricing Model

All estimates are based on the **ON DEMAND** pricing model, which provides: - Pay-as-you-go pricing with no upfront costs - No long-term commitments - Automatic scaling based on usage - Ideal for variable and unpredictable workloads

## Monitoring and Alerts

Set up the following monitoring to track costs: 1. **AWS Budgets:** Create monthly budget alerts at 80% and 100% thresholds 2. **Cost Anomaly Detection:** Enable automatic detection of unusual spending patterns 3. **CloudWatch Billing Alarms:** Set up alarms for individual service costs 4. **Cost Explorer:** Regular review of cost trends and optimization opportunities

---

*Report generated on: December 1, 2025 Pricing data source: AWS Pricing API All prices in USD and subject to change*