ETL Pipeline System Cost Analysis Estimate Report

Service Overview

ETL Pipeline System is a fully managed, serverless service that allows you to This project uses multiple AWS services.. This service follows a pay-as-you-go pricing model, making it cost-effective for various workloads.

Pricing Model

This cost analysis estimate is based on the following pricing model: - **ON DEMAND** pricing (pay-as-you-go) unless otherwise specified - Standard service configurations without reserved capacity or savings plans - No caching or optimization techniques applied

Assumptions

- Standard ON DEMAND pricing model for all services
- US East (N. Virginia) region pricing
- Average Parquet file size of 50MB
- Processing 100 files per month
- Lambda functions with 512MB memory allocation
- Glue jobs using 2 DPU workers
- DynamoDB on-demand billing mode
- S3 Standard storage class

Limitations and Exclusions

- Data transfer costs between regions
- CloudWatch detailed monitoring costs
- Development and testing environment costs
- Network acceleration costs
- Reserved capacity pricing options

Cost Breakdown

Unit Pricing Details

Service	Resource Type	Unit	Price	Free Tier
AWS Lambda	Requests	request	$0.0000002000	First 12 months: 1M requests/month and 400,000 GB-seconds/month

				free
AWS Lambda	Compute	GB-second	$0.0000166667	First 12 months: 1M requests/month and 400,000 GB-seconds/month free
AWS Glue ETL Jobs	Standard Etl	DPU-Hour	$0.44	No free tier for AWS Glue services
AWS Glue ETL Jobs	Flex Etl	DPU-Hour (34% savings)	$0.29	No free tier for AWS Glue services
AWS Glue Data Catalog	Storage	object per month	$0.0000100000	First 1M requests per month free for first 12 months
AWS Glue Data Catalog	Requests	request	$0.0000010000	First 1M requests per month free for first 12 months
Amazon S3	Storage	GB per month	$0.023	First 12 months: 5GB storage, 20,000 GET requests, 2,000 PUT requests free
Amazon S3	Put Requests	1,000 requests	$0.0005	First 12 months: 5GB storage, 20,000 GET requests, 2,000 PUT requests free
Amazon S3	Get Requests	1,000 requests	$0.0004	First 12 months: 5GB storage, 20,000 GET requests, 2,000 PUT requests free
Amazon DynamoDB	Read Requests	million read request read requests	$0.25	25GB storage and 25 read/write capacity units free permanently
Amazon DynamoDB	Write Requests	million write request write requests	$1.25	25GB storage and 25 read/write capacity units free permanently
				25GB storage

Amazon DynamoDB	Storage	GB per month	$0.25	and 25 read/write capacity units free permanently
Amazon API Gateway	Requests	million requests	$3.50	First 12 months: 1M API calls per month free

Cost Calculation

Service	Usage	Calculation	Monthly Cost
AWS Lambda	File upload, job trigger, status checker, and output handler functions - 6,000 requests per month with 512MB memory (Requests: 6,000 requests, Compute: 6,000 requests × 1s × 0.5GB = 3,000 GB-seconds)	$0.0000002000 × 6,000 requests + $0.0000166667 × 3,000 GB-seconds = $0.38	$0.38
AWS Glue ETL Jobs	Processing 100 Parquet files per month using standard ETL jobs with 2 DPU workers, 10 minutes average runtime (Dpu Hours: 100 jobs × 2 DPU ×	$0.44 × 33.4 DPU-Hours = $14.67 (Standard ETL) or $0.29 × 33.4 DPU-Hours = $9.69 (Flex ETL)	$14.67

	0.167 hours = 33.4 DPU-Hours)		
AWS Glue Data Catalog	Storing metadata for 1,000 objects and 10,000 API requests per month (Objects: 1,000 objects, Requests: 10,000 requests)	$0.0000100000 × 1,000 objects + $0.0000010000 × 10,000 requests = $0.02	$0.02
Amazon S3	Storing 5GB of input Parquet files and 3GB of output JSON files with 1,000 PUT/GET requests (Storage: 8GB total storage, Put Requests: 500 PUT requests, Get Requests: 500 GET requests)	$0.023 × 8GB + $0.0005 × 0.5K PUT + $0.0004 × 0.5K GET = $0.20	$0.20
Amazon DynamoDB	On-demand billing for job metadata storage - 1,000 read/write requests per month (Read Requests: 500 read requests, Write Requests: 500 write requests,	$0.25 × 0.0005M reads + $1.25 × 0.0005M writes + $0.25 × 0.1GB = $0.25	$0.25

	Storage: 0.1GB)		
Amazon API Gateway	REST API with 6,000 requests per month (Requests: 6,000 requests)	$3.50 × 0.006M requests = $0.02	$0.02
Total	**All services**	**Sum of all calculations**	**$15.54/month**

Free Tier

Free tier information by service: - **AWS Lambda**: First 12 months: 1M requests/month and 400,000 GB-seconds/month free - **AWS Glue ETL Jobs**: No free tier for AWS Glue services - **AWS Glue Data Catalog**: First 1M requests per month free for first 12 months - **Amazon S3**: First 12 months: 5GB storage, 20,000 GET requests, 2,000 PUT requests free - **Amazon DynamoDB**: 25GB storage and 25 read/write capacity units free permanently - **Amazon API Gateway**: First 12 months: 1M API calls per month free

Cost Scaling with Usage

The following table illustrates how cost estimates scale with different usage levels:

Service	Low Usage	Medium Usage	High Usage
AWS Lambda	$0/month	$0/month	$0/month
AWS Glue ETL Jobs	$7/month	$14/month	$29/month
AWS Glue Data Catalog	$0/month	$0/month	$0/month
Amazon S3	$0/month	$0/month	$0/month
Amazon DynamoDB	$0/month	$0/month	$0/month
Amazon API Gateway	$0/month	$0/month	$0/month

Key Cost Factors

- **AWS Lambda**: File upload, job trigger, status checker, and output handler functions - 6,000 requests per month with 512MB memory
- **AWS Glue ETL Jobs**: Processing 100 Parquet files per month using standard ETL jobs with 2 DPU workers, 10 minutes average runtime
- **AWS Glue Data Catalog**: Storing metadata for 1,000 objects and 10,000 API requests per month
- **Amazon S3**: Storing 5GB of input Parquet files and 3GB of output JSON files with 1,000 PUT/GET requests
- **Amazon DynamoDB**: On-demand billing for job metadata storage - 1,000 read/write requests per month
- **Amazon API Gateway**: REST API with 6,000 requests per month

Projected Costs Over Time

The following projections show estimated monthly costs over a 12-month period based on different growth patterns:

Base monthly cost calculation:

Service	Monthly Cost
AWS Lambda	$0.38
AWS Glue ETL Jobs	$14.67
AWS Glue Data Catalog	$0.02
Amazon S3	$0.20
Amazon DynamoDB	$0.25
Amazon API Gateway	$0.02
Total Monthly Cost	**$15**

Growth Pattern	Month 1	Month 3	Month 6	Month 12
Steady	$15/mo	$15/mo	$15/mo	$15/mo
Moderate	$15/mo	$17/mo	$19/mo	$26/mo
Rapid	$15/mo	$18/mo	$25/mo	$44/mo

- Steady: No monthly growth (1.0x)
- Moderate: 5% monthly growth (1.05x)
- Rapid: 10% monthly growth (1.1x)

Detailed Cost Analysis

Pricing Model

ON DEMAND

Exclusions

- Data transfer costs between regions
- CloudWatch detailed monitoring costs
- Development and testing environment costs
- Network acceleration costs
- Reserved capacity pricing options

Recommendations

Immediate Actions

- Use AWS Glue Flex ETL jobs instead of standard ETL jobs for 34% cost savings ($9.69 vs $14.67)
- Implement file size-based DPU allocation to optimize Glue job costs
- Leverage AWS Free Tier benefits for the first 12 months to reduce initial costs
- Consider S3 Intelligent Tiering for long-term storage cost optimization #### Best Practices

- Monitor Glue job execution times and optimize worker allocation
- Use DynamoDB on-demand billing for variable workloads
- Implement CloudWatch cost monitoring and alerts
- Consider Reserved Capacity for predictable high-volume workloads
- Use S3 lifecycle policies to transition old files to cheaper storage classes

Cost Optimization Recommendations

Immediate Actions

- Use AWS Glue Flex ETL jobs instead of standard ETL jobs for 34% cost savings ($9.69 vs $14.67)
- Implement file size-based DPU allocation to optimize Glue job costs
- Leverage AWS Free Tier benefits for the first 12 months to reduce initial costs

Best Practices

- Monitor Glue job execution times and optimize worker allocation
- Use DynamoDB on-demand billing for variable workloads
- Implement CloudWatch cost monitoring and alerts

Conclusion

By following the recommendations in this report, you can optimize your ETL Pipeline System costs while maintaining performance and reliability. Regular monitoring and adjustment of your usage patterns will help ensure cost efficiency as your workload evolves.