AWS Pricing Optimization for Asian-Based Data Streaming System

Introduction

Project Background: This project simulates the infrastructure for an Asian-based music streaming system using AWS services such as Amazon Kinesis Data Streams, Amazon Redshift, and Amazon Simple Storage Service (Amazon S3). The goal is to support real-time analytics, cost-effective scalability, and cross-regional performance.

Objective: Analyze the given configuration using the AWS Pricing Calculator and propose an optimized infrastructure without sacrificing required performance or functionality.

Services Overview

AWS Service	Purpose
Amazon Kinesis Data Streams	Ingest real-time music streaming logs
Amazon Redshift	Run analytics on streaming and historical data
Amazon Simple Storage Service (S3)	Store, tier, and retrieve music content and logs

- ✅ Switched Kinesis from On-Demand to Provisioned mode and disabled Enhanced Fan-Out, saving over $800/month
- ✅ Replaced 3 Redshift DS2 nodes with 2 Reserved RA3 nodes, reducing cost by ~$4,400/month
- ✅ Enabled lifecycle tiering and S3 Select in Amazon S3 to bring storage cost down by ~$6/month

These changes ensure that the system remains scalable and real-time capable while dramatically reducing cost.

Amazon Kinesis Data Streams (KDS)

Original Configuration:

- On-Demand mode
- 100 records/sec, 500 KB per record
- 3 consumer applications

Optimized Configuration:

- Provisioned Mode
- 100 KB per record
- 2 shards + buffer
- Enhanced Fan-Out disabled

Metric	Original	Optimized	Type / Reason
Monthly Cost	~$850+	$30.73	Total stream cost using Provisioned Mode
Record Size	500 KB	100 KB	Reduced for realistic streaming metadata size
Mode	On-Demand	Provisioned	Cost-efficient for consistent load
Shards	Auto-scale	2	Manual Controlled throughput and scaling
Fan-Out	Enabled	Disabled	3 consumers only, not needed
vCPU/Memory	N/A	N/A	Serverless, shard-based architecture

Why this configuration was used: Provisioned mode provides predictable throughput while significantly lowering costs. Enhanced Fan-Out was disabled since our workload (3 consumers) didn't require <70 ms latency. Firehose was not used due to its buffering delay, making it unsuitable for real-time streaming.

Amazon Redshift (Data Warehouse)

Original Configuration:

- 1 x ds2.xlarge (On-Demand) - 4 vCPU, 31 GiB RAM, 2 TB HDD
- 2 x ds2.8xlarge (Reserved) - 36 vCPU, 244 GiB RAM, 16 TB HDD each

Optimized Configuration:

- 2 x ra3.xlplus (Reserved, No Upfront) - 4 vCPU, 32 GiB RAM each
- Managed Storage + 10 GB Backup
- 0.05 TB Spectrum Query

Metric	Original	Optimized	Type / Reason
Instance Type	ds2.xlarge + ds2.8xlarge	ra3.xlplus x2	Instance families
vCPUs	4 + 72	8 total	Reduced due to newer architecture
Memory	31 + 488 GiB	64 GiB total	Modern memory efficiency
Storage	34 TB HDD	Managed Elastic managed storage	
Spectrum	50 GB	50 GB	Constant for querying S3
Monthly Cost	~$5,700	$1,306.61	Reserved, efficient instance family

Why this configuration was used: RA3 provides scalable, managed storage and separates compute, making it future-ready. It is also AWS's recommended replacement for DS2 nodes. Reserved pricing helps cut long-term cost without reducing performance.

Amazon S3 (Storage)

Original Configuration:

- 50 GB Standard, 100 GB Intelligent-Tiering
- No lifecycle transitions modeled

Optimized Configuration:

- 50 GB S3 Standard
- 100 GB S3 Intelligent-Tiering with:
  - 50% Frequent, 30% Infrequent, 20% Archive Instant
- 1,000 lifecycle transitions
- 20 GB S3 Select scanned / 5 GB returned
- 50 GB inbound / 100 GB outbound

Metric	Original	Optimized	Type / Reason
Total Storage	150 GB	150 GB	No change
Lifecycle	Not used	1,000 transitions	Cost-efficient archiving
S3 Select	Not used	20/5 GB	Ad hoc queries for savings
Data Transfer	Not modeled	Modeled (50 In/100 Out)	Accurate billing
Monthly Cost	~$20	$14.65	Tiering and usage modeled

Why this configuration was used: S3 Intelligent-Tiering automates transitions without overhead. Adding lifecycle rules and S3 Select reflects real-world data patterns while reducing costs. Glacier wasn’t used as low-latency access was still required.

Cost Comparison Summary

Service	Original Cost	Optimized Cost	Cost Type	What Changed
Kinesis	$850	$30.73	Monthly Stream	On-Demand to Provisioned, removed Fan-Out
Redshift	~$5,700	$1,306.61	Compute + Storage	Switched to RA3 Reserved
S3	~$20	$14.65	Storage + Transfer	Lifecycle, Select, and Tiering modeled
Total	**$6,600+**	**$1,351.99**	**Monthly Total**	All services optimized

Final Estimate Screenshot

aws

Contact your AWS representative: Contact Sales

Export Date: 04/04/2025 **Language:** English

Estimate url

Estimate summary

Upfront cost	Monthly cost	Total 12 months cost
0.00 USD	1,351.99 USD	16,223.88 USD Includes upfront cost

Detailed Estimate

Name	Group	Region	Upfront cost	Monthly cost
Amazon Kinesis Data Streams	-	Asia Pacific (Tokyo)	0.00 USD	30.73 USD

Status -

Description: On demand knesis

Config summary Duration of data retention (1 days), Baseline number of records (10 per second), Peak number of records (10 per second), Number of Consumer Applications (3)

Name	Group	Region	Upfront cost	Monthly cost
Amazon Redshift	-	Asia Pacific (Tokyo)	0.00 USD	1,306.61 USD

Status -

Description: Asian based data streaming

Config summary Nodes (2), Instance type (ra3.xlplus), Utilization (On-Demand only) (100 %Utilized/Month), Pricing strategy (Reserved No Upfront 1yr), Additional backup storage (10 GB), Data scanned (50 GB)

Name	Group	Region	Upfront cost	Monthly cost
Amazon Simple Storage Service (S3)	-	Asia Pacific (Tokyo)	0.00 USD	14.65 USD

Status -

Description: Asian based streaming

Config summary S3 Standard storage (50 GB per month), PUT, COPY, POST, LIST requests to S3 Standard (10000), GET, SELECT, and all other requests from S3 Standard (10000), Data returned by S3 Select (5 GB per month), Data scanned by S3 Select (20 GB per month) S3 INT storage (100 GB per month), S3 INT Average Object Size (16 MB), Percentage of Storage in INT-Frequent Access Tier (0.5), Percentage of Storage in INT-Infrequent Access Tier (% of storage that hasn't been accessed in the last 30 days) (0.3), Percentage of Storage in INT-Archive Instant Access Tier (% of storage that hasn't been accessed in the last 90 days) (0.2), Percentage of Storage in INT-Archive Access Tier (% of storage that hasn't been accessed for a minimum of 90 days) (0), Percentage of Storage in INT-Deep Archive Access Tier (% of storage that hasn't been accessed for a minimum of 180 days) (0), PUT, COPY, POST, LIST requests to S3 INT (10000), GET, SELECT, and all other requests from S3 INT (50000), Lifecycle Transition requests (1000), Data returned by S3 Select (5 GB per month), Data scanned by S3 Select (20 GB per month) DT Inbound: Internet (50 GB per month), DT Outbound: Internet (100 GB per month)

Future Optimization Opportunities

If future usage increases:

- Enabling Enhanced Fan-Out if Kinesis consumers > 5 or latency < 70 ms is needed
- Scaling Redshift to ra3.4xlarge or adding concurrency scaling if query volume increases
- Using S3 Glacier or Deep Archive if older data is no longer accessed
- Adding CloudFront for content caching across Asia
- Integrating Athena for querying large S3 datasets with SQL