# PII Redaction Flow Cost Analysis

This document outlines the cost analysis for a PII redaction process using AWS services. The flow involves storing documents in Amazon S3, triggering AWS Lambda to call Amazon Comprehend for PII detection and redaction, and storing the redacted data back to Amazon S3. The cost analysis is based on current AWS pricing and the provided assumptions.

## Assumptions

- Document distribution for 10,000 documents:  
 - Large: 4,000 documents @ 20 MB each → Total: 80,000 MB (78.125 GB).  
 - Medium: 3,000 documents @ 12 MB each → Total: 36,000 MB (35.156 GB).  
 - Small: 3,000 documents @ 4 MB each → Total: 12,000 MB (11.719 GB).  
- Total document size: 126,000 MB (123.047 GB).  
- Amazon Comprehend pricing: $0.0001 per unit (1 unit = 100 characters).

## AWS Services in Use

### 1. Amazon S3 Storage

Input Storage:  
- Pricing: $0.023/GB for the first 50TB.  
- Cost for 123.047 GB: $2.83/month.  
  
Output Storage (PII-Redacted Data):  
- Assume same size as input.  
- Pricing: $0.023/GB for the first 50TB.  
- Cost for 123.047 GB: $2.83/month.

### 2. AWS Lambda Execution

Processing:  
- Average execution time: 3 seconds/document.  
- Total compute time: 30,000 seconds (8.33 hours).  
- Memory: 512 MB allocated per invocation.  
- Cost: $0.0000166667 per GB-second.  
- Compute cost: $0.25.  
  
Requests:  
- Total: 10,000 requests.  
- Cost: $0.20 per 1 million requests.  
- Request cost: $0.002.

### 3. Amazon Comprehend PII Detection

Pricing:  
- $0.0001 per unit (1 unit = 100 characters).  
  
Units:  
- Large: 20 MB/document → 200,000 units/document.  
- Medium: 12 MB/document → 120,000 units/document.  
- Small: 4 MB/document → 40,000 units/document.  
  
Total Units:  
- Large: 4,000 documents → 800,000,000 units.  
- Medium: 3,000 documents → 360,000,000 units.  
- Small: 3,000 documents → 120,000,000 units.  
- Total: 1,280,000,000 units.  
  
Cost:  
- 1,280,000,000 units × $0.0001/unit = $128,000.00.

## Total Monthly Cost

Service Costs:  
- S3 Input Storage: $2.83.  
- S3 Output Storage: $2.83.  
- Lambda Compute: $0.25.  
- Lambda Requests: $0.002.  
- Comprehend PII Detection: $128,000.00.  
  
Total: $128,005.91.