

AWS Community Fest Perú 2025



Integrando LLMs en AWS

PONENTE: SERGIO INFANTE

CARGO/ROL: SOFTWARE DEVELOPER

¿Quién Soy?

- Soy un desarrollador de software con 15+ años de experiencia especializado en Python, AWS y Arquitecturas Serverless. He contribuido al desarrollo de GNOME y OpenOffice durante varios años como desarrollador open source.
- Además de liderar equipos de desarrollo en múltiples países, soy creador de contenido técnico, youtube, podcasts e instructor en Udemy.

SERGIO INFANTE
SOFTWARE DEVELOPER

AGENDA

- 01 El desafío: Análisis Crítico de Pruebas Clínicas
- 02 Solución con AWS
- 03 Caso de Exito
- 04 Resultados

Promesa: Al final de esta charla, habrán observado un caso de éxito de como usar diversos servicios AWS para optimizar un proceso crítico en la industria farmacéutica, y que pueden aplicar en sus propios proyectos.

aws

COMMUNITY FEST

PERU

EL DESAFÍO: ANÁLISIS CRÍTICO DE PRUEBAS CLÍNICAS

- Las farmacéuticas deben procesar miles de reportes de efectos secundarios para validar la seguridad de nuevos medicamentos antes de su lanzamiento.

Evaluar criticidad y riesgo vital de cada reporte

Determinar ajustes y dosificaciones

Acelerar la llegada segura de medicamentos al mercado



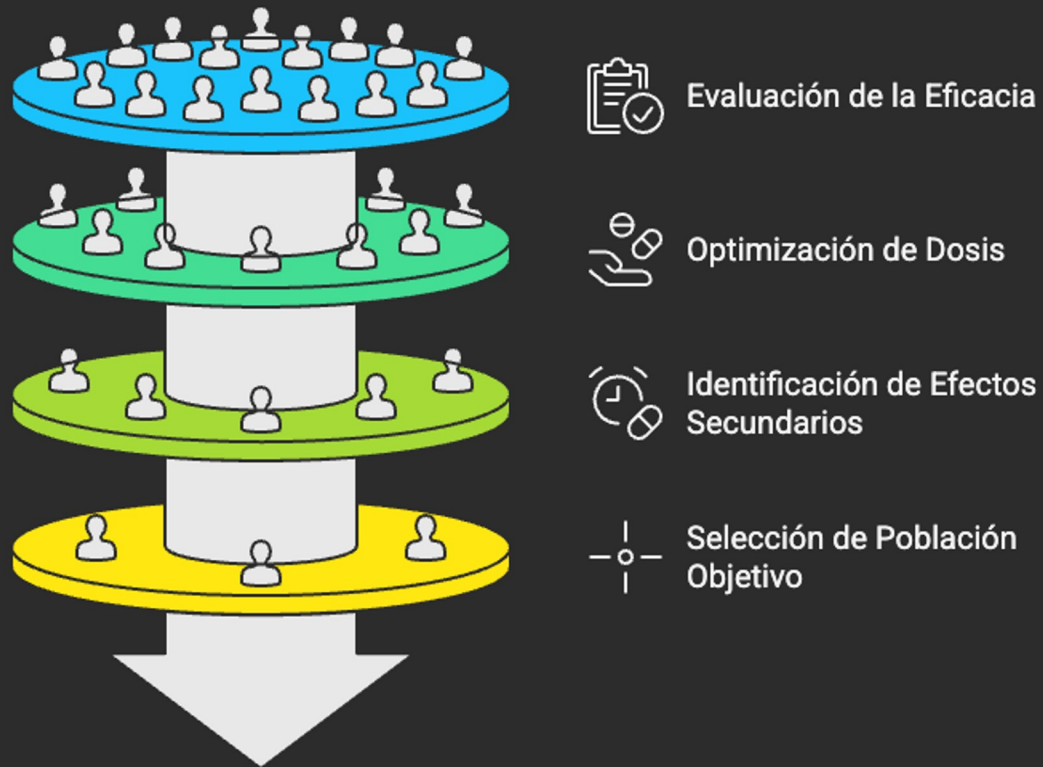
El viaje de una década en el desarrollo de medicamentos



Explorando los Resultados de la Fase I de Pruebas Clínicas



Refinando la Evaluación del Tratamiento



Datos de ensayos clínicos



Eficacia comparativa

Comparación de la eficacia con los tratamientos existentes.

Perfil completo de seguridad del tratamiento.

Perfil de seguridad



Datos de etiquetado

Datos para el etiquetado del tratamiento.

Análisis de subgrupos de pacientes.

Análisis de subgrupos



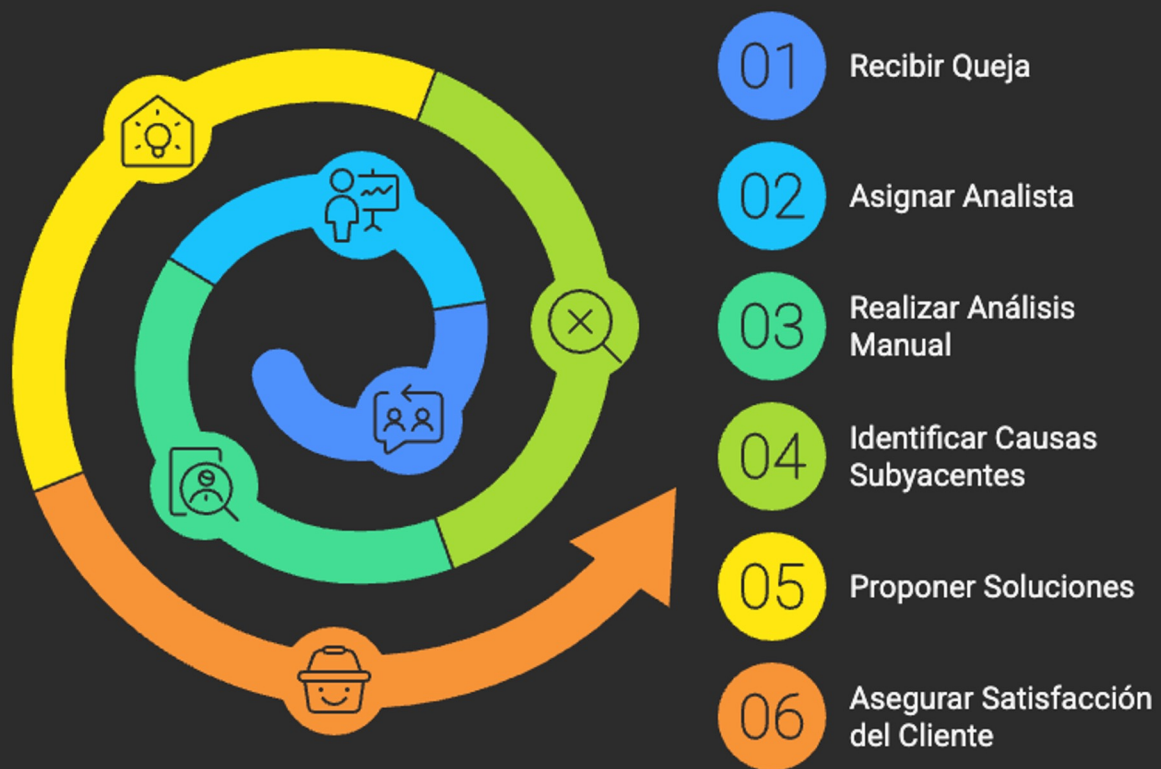
Problema central: El análisis manual de cada complaint toma semanas, retrasando decisiones críticas que pueden salvar vidas y afectar el desarrollo de tratamientos esenciales.

Un solo medicamento puede generar +10,000 reportes durante pruebas clínicas - el procesamiento manual puede retrasar la aprobación hasta 6 meses.

Análisis de Quejas Farmacéuticas



Proceso de Análisis de Quejas



El antiinflamatorio Vioxx pudo haber causado entre 88.000 y 140.000 infartos sólo en EE UU

El medicamento de MSD, retirado en 2004, fue el mayor éxito de ventas de la historia

EL PAÍS

Barcelona - 31 ENE 2005 - 18:00 PET



El rofecoxib, el antiinflamatorio no esteroideo comercializado por MSD como Vioxx y retirado del mercado mundial el 30 de septiembre de 2004, pudo haber causado sólo en Estados Unidos entre 88.000 y 140.000 infartos o enfermedades coronarias graves desde su lanzamiento en 1999. Estas estimaciones han sido realizadas por la Food and Drug Administration (FDA), la agencia del medicamento estadounidense, y se publican en la edición digital de la revista médica *The Lancet*. En España, cuando se retiró el medicamento, más de 277.000 personas lo estaban

ÚLTIMAS NOTICIAS

09:10 | Baleares, Murcia y Comunidad Valenciana en alerta naranja por lluvia de 50 litros por hora

09:00 | La ventaja de ser hijo de universitario: tres de cada cuatro alcanzan educación superior, lo que les proporciona mejor salario y salud

08:28 | Última hora de la actualidad



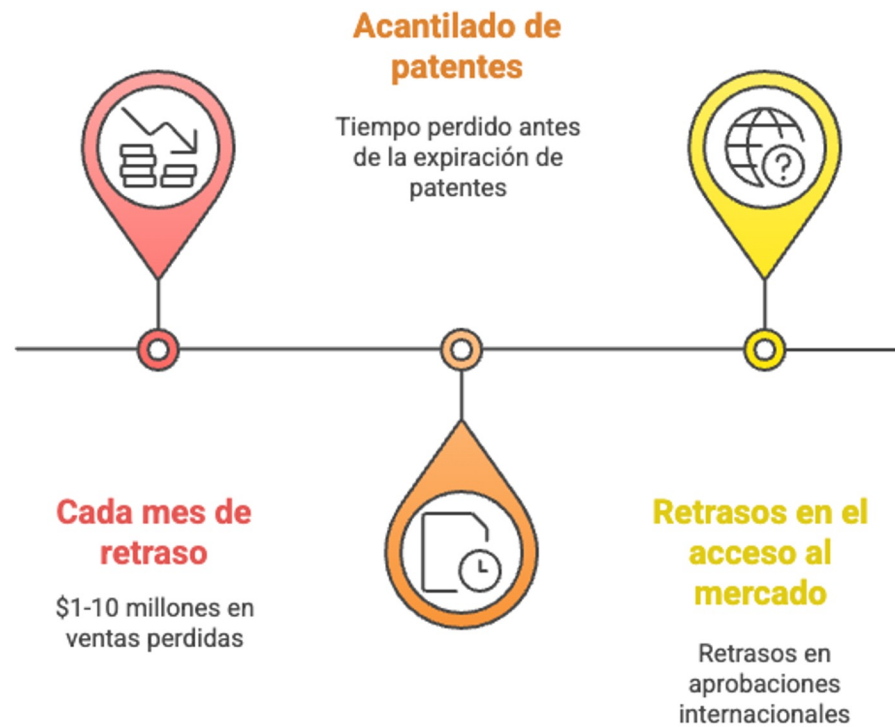
Las familias de las víctimas de la heparina contaminada declaran ante el Congreso

La heparina se utiliza en medicina como anticoagulante, es decir, impide la coagulación de la sangre. Debido a este efecto, el fármaco se utiliza a menudo para evitar la coagulación de la sangre o el plasma en dispositivos médicos como las máquinas de diálisis renal.

En marzo de 2008, la FDA anunció una importante retirada de heparina importada de China por contaminación. Hoy, las familias de las víctimas que murieron por exposición a heparina contaminada mientras se sometían a diálisis han testificado ante un subcomité del Congreso que investiga la retirada. El subcomité también escuchó a la Dra. Janet Woodcock, directora del Centro de Evaluación e Investigación de Medicamentos de la FDA, quien indicó que las pruebas iniciales demuestran que la contaminación fue intencionada y no un error de fabricación.

Los [abogados especializados en errores de medicación](#) de [Washington y Oregón](#) de D'Amore & Associates ayudan a los clientes y a sus familias a obtener una indemnización por lesiones derivadas de la prescripción de un medicamento equivocado o de sus efectos secundarios.

El Impacto Financiero de los Retrasos en el Desarrollo de Medicamentos

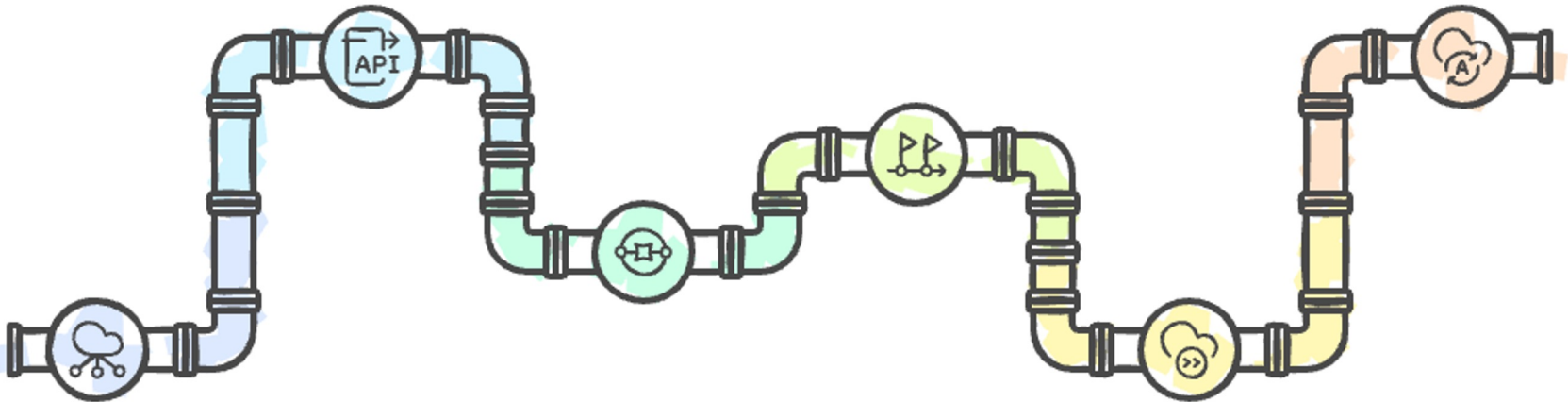


SOLUCIÓN

La Solución con AI QMS Complaints Platform:

- **Arquitectura serverless completa** - React (CloudFront)
→ API Gateway → SQS → Step Functions → Lambda →
DynamoDB
- **Procesamiento paralelo inteligente** - Step Functions
orquesta modelos SageMaker + LLMs Bedrock para
análisis multi-dimensional

Solución Serverless propuesta



01

Entrega de Contenido

React y CloudFront entregan contenido al usuario

02

Gestión de Solicitudes

API Gateway gestiona las solicitudes entrantes

03

Gestión de Colas

SQS gestiona las colas de mensajes

04

Orquestación de Procesos

Step Functions orquesta los procesos

05

Ejecución de Funciones

Lambda ejecuta funciones sin servidor

06

Almacenamiento de Datos

DynamoDB almacena los datos de manera escalable

Beneficios Clave:

- **Automatización del 90% del flujo de análisis** con validación humana final
- **Escalabilidad automática** para procesar miles de complaints simultáneamente
- **Mejora continua mediante ML Ops** - Auditoría completa permite optimización de prompts y reentrenamiento

SOLUCIÓN

Profundizando en el Tema

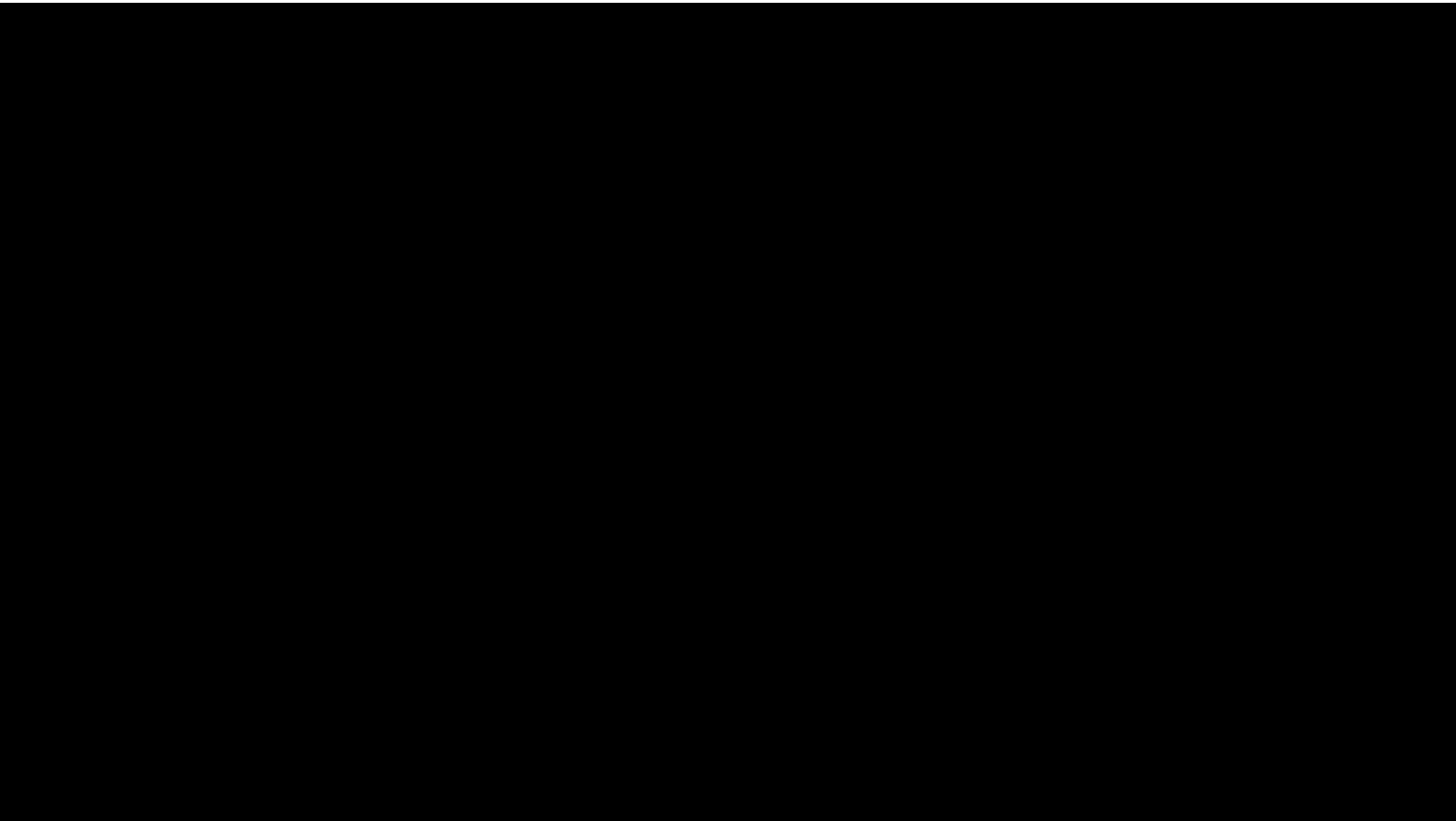
Descripción de los componentes (razones de uso)

Caso de Uso: Diagrama de arquitectura (recurso externo)

Code snippets

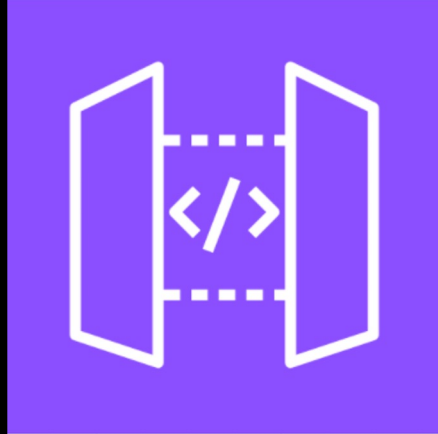


**Amazon
Cloudfront**





```
# CloudFront Distribution for React App
CloudFrontDistribution:
  Type: AWS::CloudFront::Distribution
  Properties:
    DistributionConfig:
      Origins:
        - Id: S3Origin
          DomainName: !GetAtt S3Bucket.DomainName
          S3OriginConfig:
            OriginAccessIdentity: !Sub "origin-access-identity/cloudfront/${OriginAccessIdentity}"
      DefaultCacheBehavior:
        TargetOriginId: S3Origin
        ViewerProtocolPolicy: redirect-to-https
        Compress: true
        DefaultTTL: 86400
      DefaultRootObject: index.html
      CustomErrorResponses:
        - ErrorCode: 404
          ResponseCode: 200
          ResponsePagePath: /index.html # SPA routing support
    Enabled: true
```



Amazon API Gateway



Producer
Sends messages to Amazon SQS



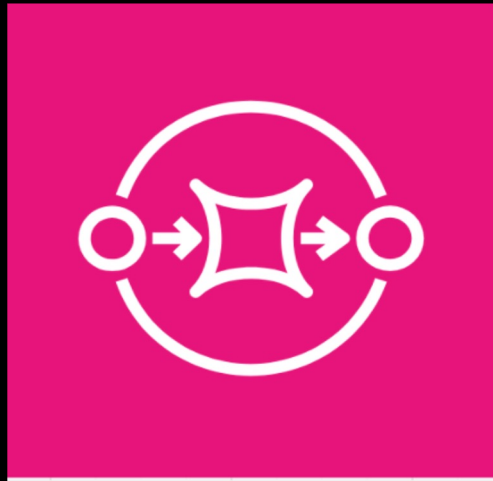
Amazon SQS
Fully managed message queuing service to reliably and continually exchange any volume of messages from anywhere



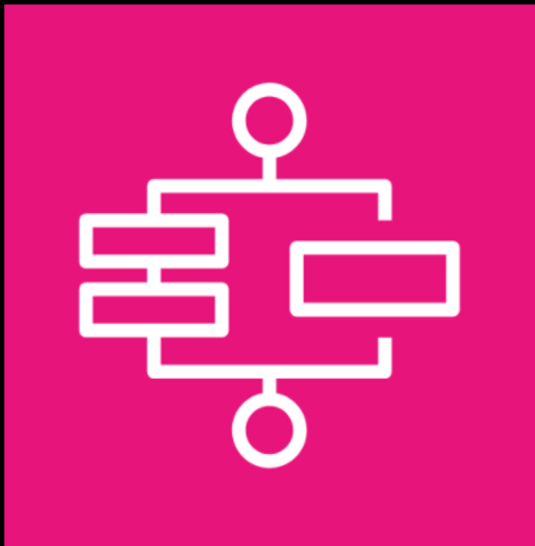
Encryption
Messages are encrypted at rest (AWS KMS) and in flight with HTTPS/TLS



Consumers
Nearly infinite scalability and ability to increase message throughput to consumers without the need to preprovision capacity



Amazon Simple Queue Service



Amazon Step Functions



AWS Lambda



**Amazon
SageMaker**



```
# Ejemplo de inferencia con RoBERTa en SageMaker
import boto3

def classify_complaint_priority(narrative_text):
    """
    Usa modelo RoBERTa fine-tuned para clasificar prioridad
    """
    sagemaker_runtime = boto3.client('sagemaker-runtime')

    # Modelo entrenado específicamente para complaints clínicos
    endpoint_name = 'roberta-complaint-classifier-endpoint'

    payload = {
        'inputs': narrative_text,
        'parameters': {
            'max_length': 512,
            'return_all_scores': True
        }
    }

    response = sagemaker_runtime.invoke_endpoint(
        EndpointName=endpoint_name,
        ContentType='application/json',
        Body=json.dumps(payload)
    )

    # Retorna: CRITICAL, HIGH, MEDIUM, LOW con probabilidades
    return json.loads(response['Body']).read()
```



**Amazon
Bedrock**



**Amazon
DynamoDB**

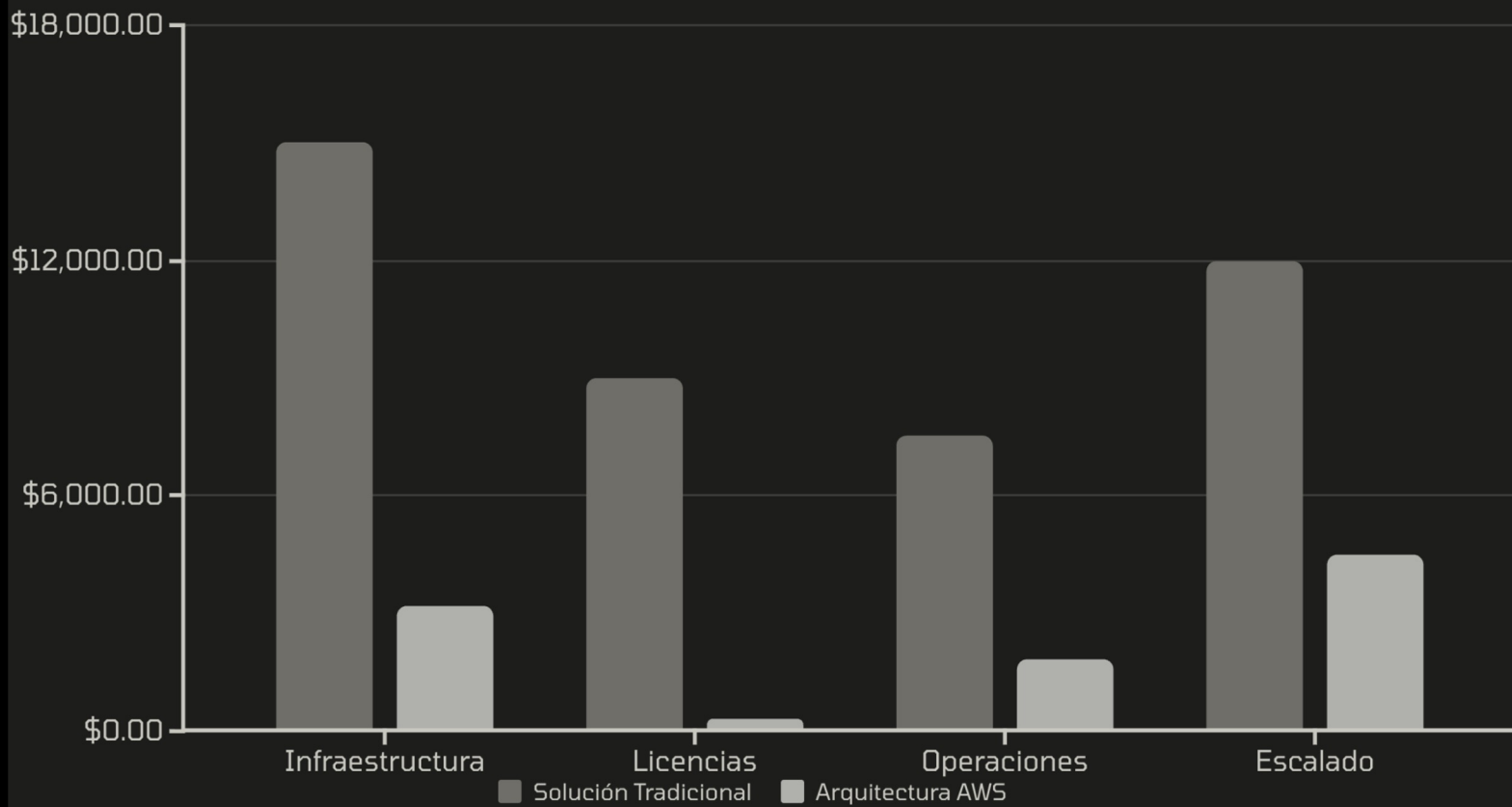
CASO DE EXITO

Diagrama de arquitectura



RESULTADOS

Análisis Comparativo de Costes



La arquitectura serverless en AWS reduce costes totales en un 74% respecto a soluciones tradicionales on-premise, con mayor flexibilidad para ajustar capacidad según demanda real.

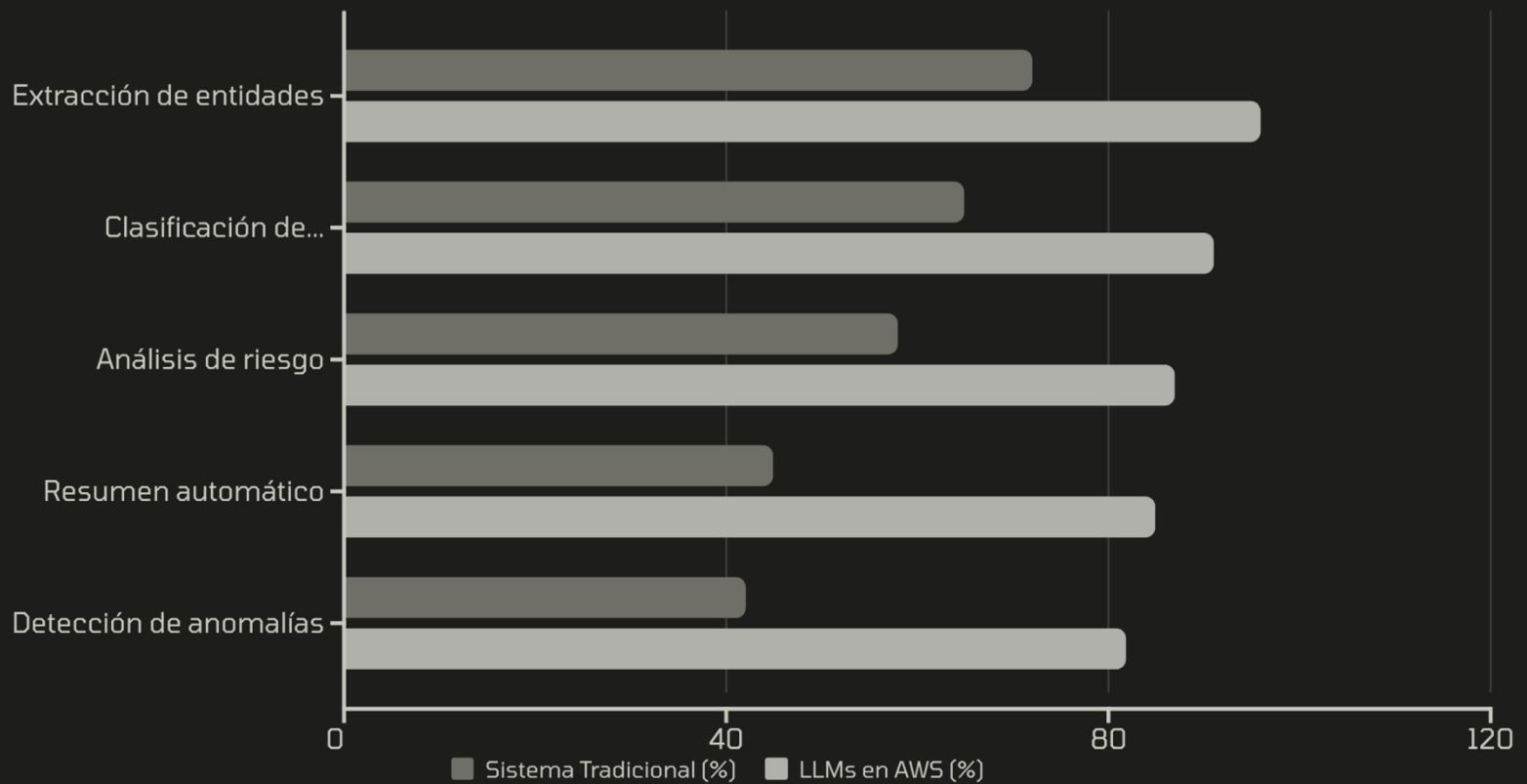
MÉTRICAS DE RENDIMIENTO

Métricas de Rendimiento: Antes vs Después



La arquitectura AWS con LLMs no solo mejoró significativamente los tiempos de procesamiento y precisión, sino que además redujo costes y aumentó dramáticamente la capacidad máxima de procesamiento.

Análisis de Precisión por Tipo de Tarea



La mejora en precisión fue especialmente significativa en tareas complejas como resúmenes y análisis de riesgo, donde los modelos tradicionales presentaban mayores limitaciones.

Lecciones Aprendidas: Mejores Prácticas



Prompt Engineering

Invertir tiempo en optimizar prompts sistemáticamente. Diferencias del 15-30% en precisión con el mismo modelo mediante refinamiento iterativo de instrucciones.



Testing Exhaustivo

Crear corpus de evaluación representativo. Implementar evaluaciones automáticas contra ground truth para detectar regresiones en calidad.



Monitorización Proactiva

Supervisar no solo métricas técnicas sino también calidad de outputs. Implementar evaluación continua con métricas como ROUGE, BERTScore, etc.



Sistemas de Respaldo

Diseñar arquitectura con modelos de fallback. Implementar degradación graceful manteniendo funcionalidad básica incluso ante fallos.

Conclusiones: Impacto en el Negocio

82%

Reducción tiempo

Disminución en el tiempo medio de procesamiento de documentos legales complejos.

65%

Ahorro costes

Reducción en costes operativos totales respecto al sistema anterior.

94%

Precisión

Tasa de acierto en extracción de entidades y cláusulas críticas.

8x

Capacidad

Aumento en capacidad máxima de procesamiento en períodos pico.

El impacto real va más allá de las métricas técnicas: la implementación ha transformado procesos de negocio completos, permitiendo a los profesionales legales centrarse en tareas de alto valor añadido mientras el sistema gestiona el análisis documental rutinario.

Preguntas y Contacto

- **Correo Electrónico:** sergio@infante.io
- **LinkedIn/Twitter:** <https://www.linkedin.com/in/sergioinfante/>



COMMUNITY FEST
PERÚ

GRACIAS