

## Proyecto final – Cloud Computing

**Integrantes:** -Eduar Alonso Ramírez Morales

-Vanessa Montoya Restrepo

-Manuela Londoño González

### 1. Descripción del problema:

- **Explicación del problema o necesidad:** La dificultad de ofrecer recomendaciones personalizadas de películas o series a los usuarios basadas en sus preferencias y reacciones a las reseñas es el problema que buscamos resolver. El desafío de las plataformas de streaming y los portales de reseñas es proporcionar recomendaciones pertinentes que se alineen con los gustos del usuario, especialmente en lo que respecta a la comprensión de la percepción subjetiva del usuario del contenido, como si una película les resulta positiva o negativa.
- **Justificación de la relevancia y su importancia:** Este problema es importante porque mejorar la experiencia de los usuarios en las plataformas de contenido puede aumentar la retención de clientes y su satisfacción. Se pueden recomendar películas basadas en las reseñas para ofrecer un enfoque más personalizado, ajustando las recomendaciones de contenido a las preferencias emocionales de los usuarios. Esto mejora el tiempo de interacción en la plataforma y ayuda a los servicios de streaming a diferenciarse de los competidores. Además, una recomendación precisa puede ser vital para los usuarios que buscan contenido que se ajuste a su estado de ánimo, creando un vínculo más fuerte con la plataforma.

### 2. Investigación de herramientas cloud:

**Uso de servicios de nube gratuitos (AWS Free Tier):** La computación en la nube, especialmente con servicios gratuitos como AWS Free Tier, permite crear soluciones escalables y eficientes sin incurrir en costos iniciales elevados. En este caso, los servicios de nube nos ayudan a administrar el almacenamiento de datos, realizar análisis automáticos de sentimientos y proporcionar recomendaciones personalizadas. Los siguientes servicios se utilizan:

**Servicios de almacenamiento en la nube:**

- **Amazon S3 (Simple Storage Service):**

**Propósito:** Utilizamos S3 para almacenar las reseñas de películas y series. Las reseñas se almacenan en formatos como CSV o JSON, organizadas en carpetas, por ejemplo, /reviews/movies/2024/.

**Ventajas:** S3 es ideal para almacenar grandes volúmenes de datos de manera segura y permite fácil acceso y organización. Además, dentro del AWS Free Tier, se ofrece hasta 5 GB de almacenamiento gratuito, lo que es suficiente para almacenar una cantidad considerable de reseñas durante las fases iniciales del proyecto.

**Escalabilidad:** S3 puede crecer a medida que el volumen de reseñas aumenta, sin necesidad de gestionar manualmente la infraestructura de almacenamiento.

### **Computación serverless:**

- **Amazon Lambda:**

**Propósito:** AWS Lambda nos permite ejecutar código sin necesidad de gestionar servidores, lo que facilita la automatización de tareas como el procesamiento de las reseñas y el análisis de sentimientos.

**Uso en el proyecto:** Lambda se activa cada vez que se sube un nuevo archivo de reseñas a S3, desencadenando un análisis de sentimiento utilizando Amazon Comprehend. También puede actualizar los registros en DynamoDB con los resultados de este análisis de forma automática.

**Ventajas:** Lambda es altamente escalable y sólo cobra por el tiempo de ejecución del código, lo cual es perfecto para tareas que ocurren de forma intermitente como el procesamiento de nuevas reseñas. Dentro del AWS Free Tier, se ofrece 1 millón de solicitudes y 400,000 GB-segundos de computación por mes, lo cual es ideal para las primeras etapas de desarrollo.

**Eliminación de Gestión de Infraestructura:** Al ser serverless, no se necesita preocuparse por la configuración, el mantenimiento o la escalabilidad de servidores, lo cual reduce la carga operativa y permite concentrarse en el desarrollo de la lógica de negocio.

- **Análisis de Sentimientos con Amazon Comprehend:**

**Propósito:** Este servicio analiza el contenido de texto de las reseñas y determina el sentimiento (positivo, negativo, neutro o mixto).

**Uso en el proyecto:** Amazon Comprehend se utiliza para analizar el sentimiento de cada reseña, lo que nos permite almacenar el resultado junto con la película o serie correspondiente en DynamoDB, formando la base para un sistema de recomendaciones más personalizado.

**Ventajas:** Amazon Comprehend proporciona 50,000 unidades de texto de forma gratuita durante los primeros 12 meses, lo que es útil para procesar y analizar una cantidad significativa de reseñas sin costos adicionales.

**Integración Sencilla:** La integración con otros servicios de AWS como S3 y DynamoDB facilita el flujo de datos, desde el almacenamiento de reseñas hasta la generación de recomendaciones basadas en sentimientos.

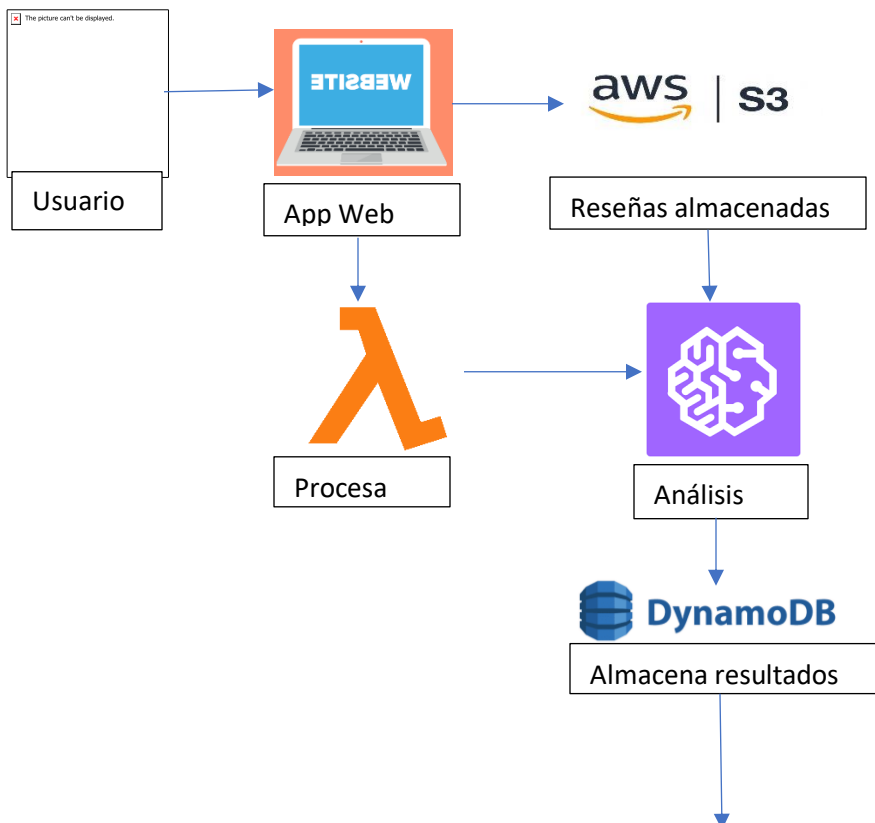
### 3. Documento de diseño del proyecto:

#### 1. Arquitectura de alto nivel:

La arquitectura del sistema se construirá utilizando los siguientes servicios principales de AWS:

- **Amazon S3:** Se utilizará para almacenar reseñas de películas/series en formato CSV o JSON.
- **Amazon Comprehend:** Para analizar los sentimientos en las reseñas.
- **Amazon DynamoDB:** Para almacenar los resultados del análisis de sentimientos y otros datos relacionados (como el ID de la película, categorías, etc.).
- **Amazon Lambda:** Para administrar el procesamiento serverless de los datos, la invocación de análisis con Comprehend y la actualización de DynamoDB.
- **Interfaz Web:** Los usuarios recibirán recomendaciones basadas en sus preferencias para interactuar con el sistema.

#### Diagrama de arquitectura:





## 2. Descripción del flujo de trabajo en la nube:

### 1) Recopilación y almacenamiento de datos:

- **Datos:** Se obtienen comentarios de sitios como Kaggle.
- **Almacenamiento en S3:** Las reseñas se almacenan en archivos JSON o CSV que se organizan en un bucket S3, lo que facilita el acceso y el procesamiento.

### 2) Análisis de sentimientos:

- **Invocación Lambda:** Una función lambda se activa cuando se encuentra un nuevo archivo de reseñas en S3.
- **Análisis con Comprehend:** Lambda procesa el archivo y envía cada comentario a Amazon Comprehend para que realice un análisis de sentimientos que clasificará cada comentario como positivo, negativo, neutral o mixto.

### 3) Almacenamiento en DynamoDB: Cada reseña se almacenará en DynamoDB con la siguiente estructura:

- ID de la película/serie
- Texto de la reseña
- Resultado del análisis de sentimientos como por ejemplo positivo, negativo, mixto o neutral.
- Datos adicionales que podrían ser los años de lanzamiento, el género de la película, etc.

### 4) Recomendaciones personalizadas:

- **Consulta a DynamoDB:** La base de datos consulta las preferencias del usuario que se encontrarán almacenadas en DynamoDB, para encontrar películas o series con sentimientos más positivos dentro de los géneros que el usuario prefiere.

- **Lógica de recomendación:** Se pueden hacer recomendaciones personalizadas como por ejemplo “Series de romance con reseñas mixtas” o “Películas de terror con reseñas positivas”.

### **3. Uso de servicios Serverless y de almacenamiento:**

#### **1) Amazon S3:**

- Funciona como una solución de almacenamiento de reseñas central. S3 es muy rentable dentro del plan gratuito y es ideal para manejar grandes cantidades de datos no estructurados.
- Los datos almacenados aquí serán procesados a intervalos regulares o según la necesidad.

#### **2) Amazon Lambda:**

- Lambda permite que el sistema funcione sin servidores específicos. Un evento de notificación activa una función Lambda en el momento en que se sube un archivo de reseñas a S3.
- Lambda también será responsable de enviar los textos a las reseñas a Amazon Comprehend para los análisis de sentimientos y actualizar DynamoDB con los resultados.

#### **3) Amazon Comprehend:**

- Es el servicio responsable de analizar los sentimientos de cada reseña, donde las funciones Lambda lo invocan automáticamente.

#### **4) Amazon DynamoDB:**

- Este servicio guardará los resultados del análisis de sentimientos, así como cualquier otra información relacionada, como los géneros de películas o las preferencias del usuario. DynamoDB es escalable, rápido y adecuado para consultas rápidas, lo que permite obtener los datos personalizados de una forma eficiente.

### **4. Cronograma y Presupuesto:**

#### **Cronograma:**

| <b>Semana</b>   | <b>Tarea</b>  | <b>Responsable</b>     |
|-----------------|---|------------------------|
| <b>Semana 1</b> | Redacción de la descripción del problema, importancia y justificación | <b>Todo el equipo</b>  |
|                 | Configuración de la cuenta AWS Free Tier y bucket en S3               | <b>Manuela Londoño</b> |

|                 |  |                 |
|-----------------|--|-----------------|
|                 | Investigación fuentes de información   | Eduar Ramirez   |
|                 | <b>Entrega etapa 1:</b> Planteamiento del problema y diseño del proyecto       | Vanessa Montoya |
| <b>Semana 2</b> | Descargar fuentes de información (Reseñas películas)                           | Eduar Ramirez   |
|                 | Configuración Amazon Comprehend para el análisis de sentimientos               | Manuela Londoño |
| <b>Semana 3</b> | Crear tabla en DynamoDB para almacenar resultados                              | Vanessa Montoya |
|                 | Implementar AWS Lambda para procesar las reseñas de forma automática           | Eduar Ramirez   |
| <b>Semana 4</b> | Desarrollar lógica para la recomendación de películas en base a los resultados | Manuela Londoño |
|                 | Integración de una interfaz de usuario <b>(Opcional)</b>                       | Vanessa Montoya |
| <b>Semana 5</b> | Pruebas al sistema de forma general  | Todo el equipo  |
|                 | Optimización del sistema AWS Free Tier   | Manuela Londoño |
|                 | Documentación y preparación para entrega del proyecto                          | Todo el equipo  |
|                 | <b>Entrega final del proyecto</b>  | Todo el equipo  |

| Semana          | Descripción  |
|-----------------|--|
| <b>Semana 1</b> | Nos enfocamos en la definición del problema, la configuración de AWS y la investigación de nuestras fuentes de información                   |
| <b>Semana 2</b> | Nos enfocamos en la descarga de las reseñas que serán utilizadas y en la configuración de Amazon Comprehend para el análisis de sentimientos |
| <b>Semana 3</b> | Semana clave para la configuración de DynamoDB y AWS Lambda, herramientas claves para la automatización del procesamiento de las reseñas     |
| <b>Semana 4</b> | Desarrollo de nuestra lógica para el sistema de recomendación y de forma opcional, creación de una interfaz de usuario                       |
| <b>Semana 5</b> | Realización de pruebas, optimización y documentación final para nuestra entrega final  |

### Estrategia – Presupuesto:

| Servicio  | Descripción  | Estrategia/Acciones   |
|---|--|---|
| <b>Amazon S3<br/>(Almacenamiento)</b>                   | Límite del AWS Free Tier: 5 GB de almacenamiento estándar, 20,000 GET requests, 2,000 PUT requests     | <b>Usar archivos comprimidos y filtrar datos innecesarios</b>   |
| <b>Amazon Comprehend<br/>(Análisis de sentimientos)</b> | Límite del AWS Free Tier: 50,000 unidades de texto por mes (100 caracteres = 1 unidad)                 | <b>Utilizar textos relevantes para el análisis</b>  |
| <b>Amazon DynamoDB<br/>(Base NoSQL)</b>                 | Límite del AWS Free Tier: 25 GB de almacenamiento, 25 unidades de capacidad de lectura/escritura       | <b>Guardar solo los resultados más relevantes de nuestro análisis de sentimientos</b>                   |
| <b>AWS Lambda<br/>(Computación Serverless)</b>          | Límite del AWS Free Tier: 1 millón de invocaciones por mes, 400,000 GB-segundos de tiempo de ejecución | <b>Mantener funciones ligeras y eficaces para minimizar el tiempo de ejecución</b>                      |
| <b>Monitoreo general con AWS CloudWatch</b>             | Herramienta gratuita dentro del Free Tier que nos permitirá monitorizar el uso de los servicios        | <b>Configurar alertas para recibir notificaciones cuando nos acerquemos a los límites del Free Tier</b> |

### Permisos – Docente:

| Permiso/Recurso | Descripción  |
|-----------------|--|
| AWS Free Tier   | Acceso a la cuenta compartida de AWS   |
| Dataset         | Acceso a la base de datos que se utilizará   |
| Feedback        | Revisión periódica por parte de nuestro docente para asegurar el cumplimiento de los objetivos |