Implantación de Sistemas Canalización de Datos para análisis de Información

Contenido

[1. Introducción 3](#__RefHeading___Toc302_3309208292)

[2. Objetivos y Requisitos 4](#__RefHeading___Toc304_3309208292)

[- OBJETIVO DEL PROYECTO 4](#__RefHeading___Toc306_3309208292)

[- REQUISITOS DE LA SOLUCIÓN 5](#__RefHeading___Toc308_3309208292)

[3. Estudio Previo 6](#__RefHeading___Toc310_3309208292)

[- ESTADO ACTUAL 6](#__RefHeading___Toc312_3309208292)

[- POSIBLES SOLUCIONES 6](#__RefHeading___Toc314_3309208292)

[4. Plan de Trabajo 7](#__RefHeading___Toc316_3309208292)

[5. Diseño 8](#__RefHeading___Toc318_3309208292)

[- DISEÑO GENERAL 8](#__RefHeading___Toc320_3309208292)

[- DISEÑO DETALLADO 8](#__RefHeading___Toc322_3309208292)

[- PRESUPUESTO 10](#__RefHeading___Toc507_319201592)

# 1. Introducción

La información sobre las preferencias de los clientes es uno de los recursos más necesarios e importante a la hora de emprender prácticamente cualquier proyecto de negocio, para obtener dicha información, a lo largo del tiempo se han llevado a cabo diferentes técnicas para atraer la atención de los clientes a posibles productos de su interés, recogiendo los resultados a posteriori y analizándolos con intención de desarrollar tácticas de mercado sobre dicha información.

Hoy en día las técnicas de recolección de información se han adaptado a las nuevas tecnologías, aprovechando la popularidad de las páginas web y, la compra online de productos y servicios.

En este proyecto se trabajara sobre un escenario ficticio, en el cual se llevara a cabo la implementación de un sistema de canalización de datos sobre los clientes que acceden a la página web de una franquicia de cafeterías llamada AnyCompany Café, con intención de que la empresa pueda utilizar dicha información para tomar decisiones sobre dónde invertir a futuro.

# 2. Objetivos y Requisitos

## - OBJETIVO DEL PROYECTO

La situación planteada sería la siguiente:

La empresa AnyCompany Café vende postres y artículos de café a través de su página web. Esta posee múltiples cafeterías en varias ciudades alrededor del mundo.

Y la empresa está interesada en obtener información sobre su negocio mediante el uso de datos sobre como las personas interactúan con su sitio web. Con ello la empresa planea analizar las tendencias de los datos del flujo de clicks, con intención de tomar decisiones inteligentes sobre dónde invertir en el futuro.

Por ello AnyCompany Café contrataría a mi empresa de consultoría para que llevara a cabo este proyecto.

Los datos de registro del flujo de clics del sitio web de la cafetería incluyen una entrada por cada clic que hace un cliente potencial mientras navega por el sitio web. El objetivo del proyecto por ende consiste en diseñar y crear un canal de análisis de datos para capturar la información del flujo de clics. Además se creará un panel de análisis de manera que el dueño de las cafeterías pueda observar rápidamente el comportamiento de los clientes.

De esta manera se busca utilizar la información obtenida sobre el comportamiento de los usuarios en la pagina web para centrar los esfuerzos publicitarios eficientemente, e incluso decidir donde abrir nuevos locales.

## - REQUISITOS DE LA SOLUCIÓN

Conociendo el objetivo, los requisitos necesarios para la solución son los siguientes:

* Diseñar y optimizar los costos de la solución antes de la implantación de la misma
* Asegurarnos que la solución sea funcional y rentable
* Transformar los datos recogidos para su posterior analisis
* Registrar los datos del Access\_log del servidor web de la cafetería
* Usar datos de registro simulados
* Analizar y visualizar los datos
* Generar conocimiento e información a partir de los datos recogidos

# 3. Estudio Previo

## - ESTADO ACTUAL

En la situación desde la que partimos tenemos que contamos con una pagina web implementada por la propia empresa de AnyCompany Café, así como un nodo EC3, el cual en este caso supuesto se usará como host de la misma página web, además que podrá acceder a los servicios de AWS Systems Manager, Amazon S3, and CloudWatch para enviar los datos.

En el equipo utilizado se utilizará Apache2 y el servicio de Cloud9 para la manipulación de archivos.

## - POSIBLES SOLUCIONES

Una de las soluciones más viables que podemos encontrar para el análisis de datos de esta clase, es la implantación y utilización del servicio de cloudwatch proporcionado por AWS para la canalización, almacenamiento y análisis de datos mediante consultas programadas en el tiempo, de manera que se pueda llevar una monitorización clara y concisa de las visitas a la web y las ventas realizadas a lo largo del tiempo.

También convendría permitir el almacenamiento de los logs recogidos, en un servicio s3 para su posterior utilización mediante diferentes posibles servicios, esto otorgaría más flexibilidad para posibles cambios que puedan producirse como vista a futuro para la empresa.

# 4. Plan de Trabajo

Conociendo los objetivos de la solución tenemos que las fases del proyecto se pueden dividir en las siguientes:

| **Fases del Proyecto** | **Detalles** |
| --- | --- |
| 1 | Creación de un diagrama de arquitectura y coste estimado para la solución |
| 2 | Usar AWS Cloud9 para acceder al server web. Observar los registros de acceso que se generan cuando se navega en la web. |
| 3 | Instalar y configurar el agente de Amazon CloudWatch y el archivo de configuracion (httpd.conf) del servidor web de apache de manera que los logs de acceso y error puedan ser recogidos y enviados a CloudWatch. |
| 4 | Confirmar que el servidor web genera logs que el agente de CloudWatch puede recoger y enviar a CloudWatch. |
| 5 | Usar log simulados para verificar que CloudWatch recibe los logs correctamente. |
| 6 | Usar CloudWatch Logs Insights para hacer consultas al grupo de logs de acceso y generar visualizaciones. |
| 7 | Ajustar el proceso de analisis de datos para ofrecer nueva información con los datos de geolocalizacion de los visitantes. Crear un panel de visualizacion para que el dueño del Cafe pueda tomar decisiones en funcion de la información recogida |

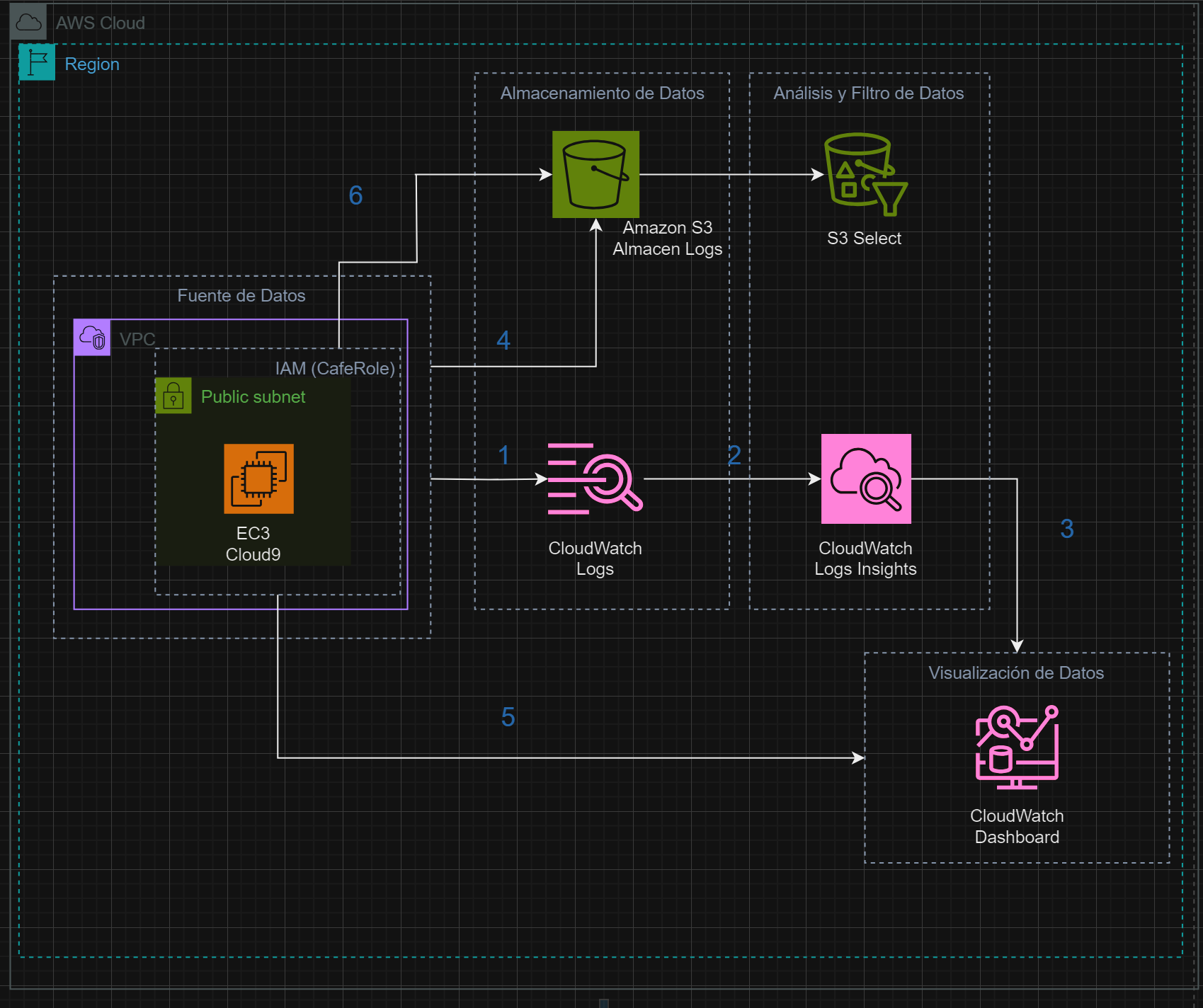
# 5. Diseño

## - DISEÑO GENERAL

La solución propuesta consistiría en la implantación y utilización del servicio de cloudwatch proporcionado por AWS para la canalización, almacenamiento y análisis de datos, de manera que se pueda llevar una monitorización clara y concisa de las visitas a la web y las ventas realizadas a lo largo del tiempo.

## - DISEÑO DETALLADO

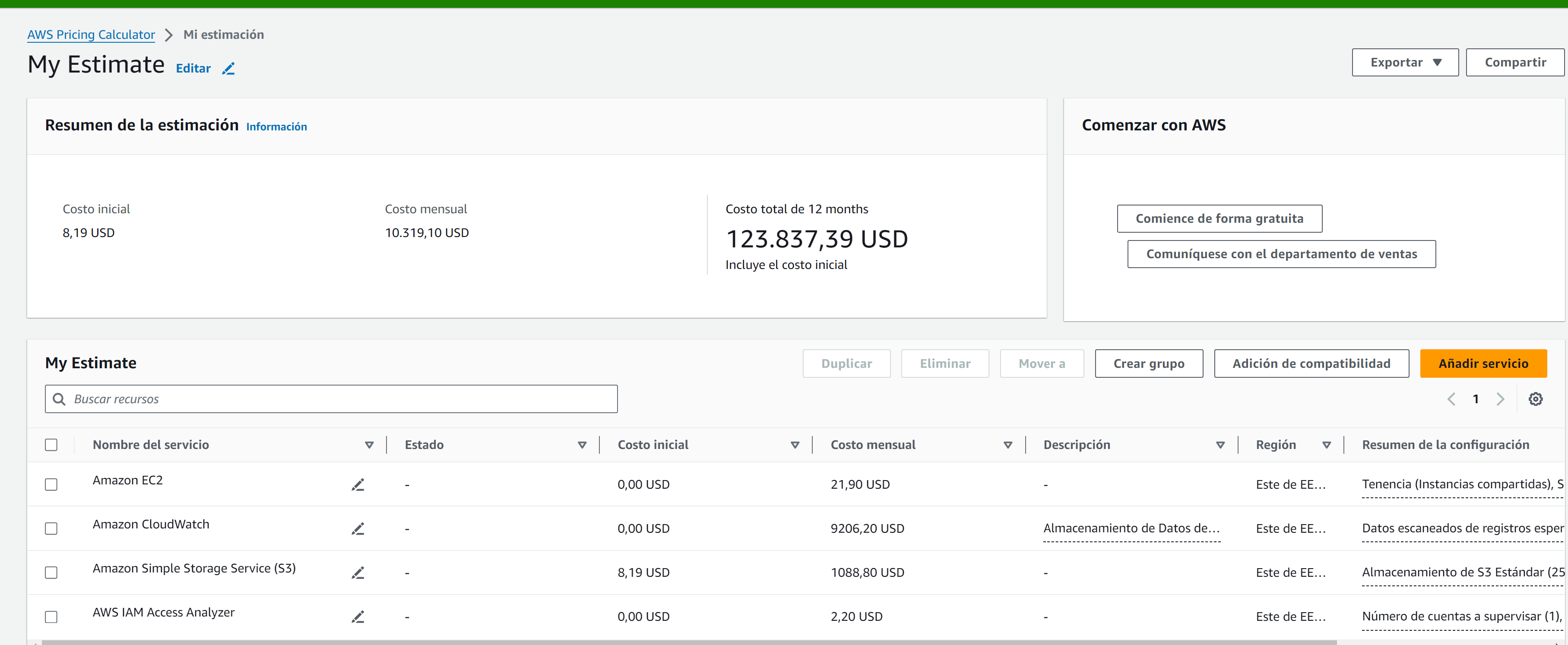
La solución propuesta se puede ver ilustrada de mejor manera una vez finalizada la primera fase del proyecto, una vez creado el diagrama de la arquitectura del sistema que se planteará:



1. Los registros se recogen desde los logs de los servidores web y se envían a Cloudwatch almacenándose en el servicio
2. Posteriormente dichos registros almacenados son procesados por el servicio de CloudWatch Logs Insights donde podrán filtrados mediante consultas.
3. Se obtiene información sobre los movimientos a lo largo del tiempo en la web mediante consultas con los parámetros requeridos utilizando el servicio de CloudWatch Logs Insights, posteriormente dicha información se muestra actualizada en el panel de muestra de CloudWatch Dashboard.
4. Los registros de igual manera se recogen desde los logs del servidor web para ser enviados a un almacenamiento Bucket S3, de manera que se pueda acceder a ellos utilizando consultas SQL con el servicio S3 Select, de esta forma nos aseguramos que en el futuro se pueda acceder y filtrar los datos utilizando otra clase de servicios, con visión a posibles expansiones de la estructura.
5. Se contara con un rol IAM CafeRole el cual estará vinculado a una EC3 y que permitirá el envío de datos de datos a CloudWatch
6. Igualmente el rol IAM CafeRole permitirá el envió de datos a instancias S3

## - PRESUPUESTO

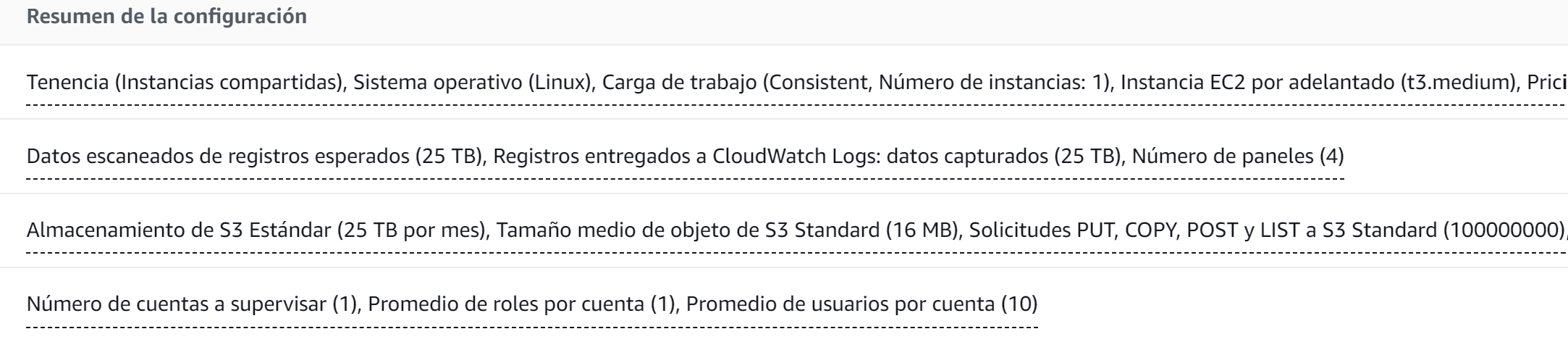
El presupuesto necesario cumpliendo todos los requerimientos de servicios ascendería a 10.319 $ al més lo que haría unos 123.837 $ al año.



El precio asciende de tal manera debido a las dimensiones de la empresa, tenemos en cuenta que se trata de una empresa a escala internacional, que cuenta con múltiples cafeterías múltiples ciudades a lo largo de todo el globo, esto hace que la capacidad de almacenamiento de registros de clientes potenciales ascienda exponencialmente

Como resultado, la capacidad media inicial de almacenamiento de datos tanto en Cloudwatch como S3 por la que se ha optado para realizar el aproximado, es de unos 25 TB de almacenamiento los cuales se refrescaran mensualmente.

Sin embargo esta capacidad puede variar si es que esta se expande aún más.



# 6. Implantación

En este paso vamos a proceder a la instalación e implementación de la solución que hemos propuesto, para ello utilizaremos las herramientas software de aws mencionadas anteriormente, instancias EC3, Cloud9, AIM, CloudWatch y Amazon S3.

La implementación llevada acabo, con sus pasos explicados más a profundidad son documentados en el manual de implementación que presento a continuación: