

Home

About me

What I do

My experience

My work



Las 5 Vs de Big Data En Netflix



Katherine Jimenez Vivanco
Mauricio Carazas Segovia
Pio Huisa



Los servicio de streaming son cada vez más utilizados por las personas

- Al tener más usuarios estos servicios tienen más datos que analizar para poder brindar un servicio de calidad a sus usuarios.
- En este trabajo analizaremos cómo se aplican las 5v's de Big Data en la empresa Netflix
 - Volumen
 - Variedad
 - Velocidad
 - Veracidad
 - Valor

¿Cuál es el volumen de la data de Netflix?

- ¿Qué aspecto tienen los servidores de almacenamiento que albergan el contenido de Netflix y lo transmiten? Una mezcla de discos duros empacados juntos en una computadora se compone de ellos. Utilizan 36 unidades que pueden transportar aproximadamente 100 TB de datos. Estos servidores son capaces de almacenar y cargar continuamente entre 10 000 y 20 000 películas.

A mayo de 2019, Netflix cuenta con alrededor de 13.612 títulos (Gaël, 2019). Solo su biblioteca estadounidense consta de 5087 títulos.

A partir de 2016, Netflix completó su migración a Amazon Web Services. Sus datos de decenas de petabytes de datos se trasladaron a AWS (Brodkin et al., 2016). Consiste en sus datos de ingeniería, datos corporativos y otra documentación.

Se puede calcular aproximadamente que Netflix almacena aproximadamente 105 TB de datos solo con respecto a los videos. Sin embargo, se espera que su conjunto de datos para los algoritmos de recomendación sea muy grande, ya que necesita incorporar toda la información mencionada anteriormente

¿Dónde se ejecutan todos los microservicios de netflix?



Netflix tenía bajo su poder una amplia red de servidores, sin embargo se dieron cuenta que la velocidad de su crecimiento haría que fuese imposible mantener estos a largo plazo y significa tiempo valioso que podrían invertir en otras áreas de la empresa.

Por lo cual decidieron trasladar todo su material a la nube, para poder ejecutarlo en servidores de otras empresas en este caso Amazon Web Services.

Para lo cual los programadores de Netflix escribieron cientos de programas y los instalaron rápidamente en los servidores de Amazon Web Services.

¿Cual es la variedad de la data de netflix?

- Netflix ofrece una gran variedad de contenido para sus usuarios
- Se ofrece: programas de televisión, películas, programas, y contenidos en diferentes idiomas
- Netflix recopila la mayoría de los datos en un formato estructurado, como la hora del día, la duración de la visualización, la popularidad, los datos sociales, la información relacionada con la búsqueda, los datos relacionados con la transmisión, etc.
- Netflix también podría estar utilizando datos no estructurados.

¿Cual es la velocidad de la data de netflix?

- Según Netflix, usa aproximadamente 1 GB de datos por hora para transmitir un programa de TV o una película en definición estándar y hasta 3 GB de datos por hora cuando transmite videos HD.
- **Netflix** usa un DDBMS (**sistemas de administración de bases de datos distribuida**) para que los datos se puedan almacenar localmente en las ubicaciones con mayor demanda. Esto mejora el tiempo de acceso. Además, al acceder a los datos, que es lo que sucede cuando selecciona una película en Netflix, la consulta se puede procesar en varias ubicaciones de la base de datos, aprovechando la potencia de procesamiento de la CPU paralela disponible en un DDBMS. Esto reduce las demandas de hardware y mejora la eficiencia.

Veracidad

- Se observó una gran variación en los datos.
 - No todas las películas fueron calificadas por igual por un individuo.
 - Una película tenía solo 3 calificaciones, mientras que un solo usuario calificó más de 17.000 películas.
 - Con el tipo y la cantidad de información, los datos de Netflix definitivamente contendrían muchas anomalías, sesgos y ruido.

Valor

- Netflix ofrece a sus usuarios una amplia selección de contenido de calidad a un precio razonable. La plataforma ofrece una variedad de contenido para satisfacer a todos sus usuarios, desde películas de estreno hasta series de televisión y programas originales.
- Netflix ofrece algunas funciones especiales, como recomendaciones personalizadas y perfiles de usuario, que permiten a los usuarios encontrar contenido según sus gustos y preferencias.

Referencias

- AMAN SINGH, CLOUD STORAGE: HOW DOES NETFLIX STORE THEIR CONTENT?, Enero, 2021
disponible en:
<https://www.globaltechcouncil.org/big-data/cloud-storage-how-does-netflix-store-their-content/>
- Eric Walz, How Netflix Uses a Distributed Database Management System to Deliver Your Movies, 6 de julio de 2016 disponible en :
<https://www.linkedin.com/pulse/how-netflix-uses-distributed-database-management-system-eric-walz/>
- Chaithanya Pramodh Kasula, Netflix Recommender System — A Big Data Case Study, Junio 2020
Disponible en:
<https://towardsdatascience.com/netflix-recommender-system-a-big-data-case-study-19cfa6d56ff5>