

# Técnicas de Extracción y Almacenamiento de datos masivos

Febrero 2023





**ALFONSO AYALA PALOMA**

**MAGISTER EN INGENIERÍA – AREA SISTEMAS Y COMPUTACIÓN**

### **PERFIL PROFESIONAL**

Magíster en Ingeniería en el área de Sistemas y computación. Especialista en Seguridad de la Información, Especialista en Docencia Universitaria, Profesional en Ingeniería de Sistemas. Catedrático en las Universidades Cooperativa y del Tolima. Amplia experiencia en proyectos de desarrollo de sistemas de información, herramientas de soporte a toma de decisiones, procesos de Transformación Digital y coaching de Innovación.



*Temas*

- 1 Intro
- 2 Modelo RCPT
- 3 Discusión
- 4 Preguntas y respuestas
- 5



compensar

fundación  
universitaria

## PRODUCT GOAL

Conocer ETL, almacenamiento y seguridad de data

- con el propósito de aplicarlo profesionalmente
- Por medio del trabajo individual y en equipo

a.k.a.: “Intro a DATA ENGINEERING” / ETL

## Criterios de aceptación Producto

CR 1: Comprender Extracción

CR 2: Conocer fundamentos de seguridad

CR 3: Aprender a usar extracción, transformación y aseguramiento

CR 4: Simulación de integridad de data,  
Almacenamiento

CR 5: Socialización de ETL

## Sprint Goal

Introducción al curso

Modelo RCPT

Discusión

Preguntas y respuestas



## Historia 1

Como un estudiante de la especialización quiero conocer el modelo Reconocimiento, Contextualización, Profundización y Transferencia.

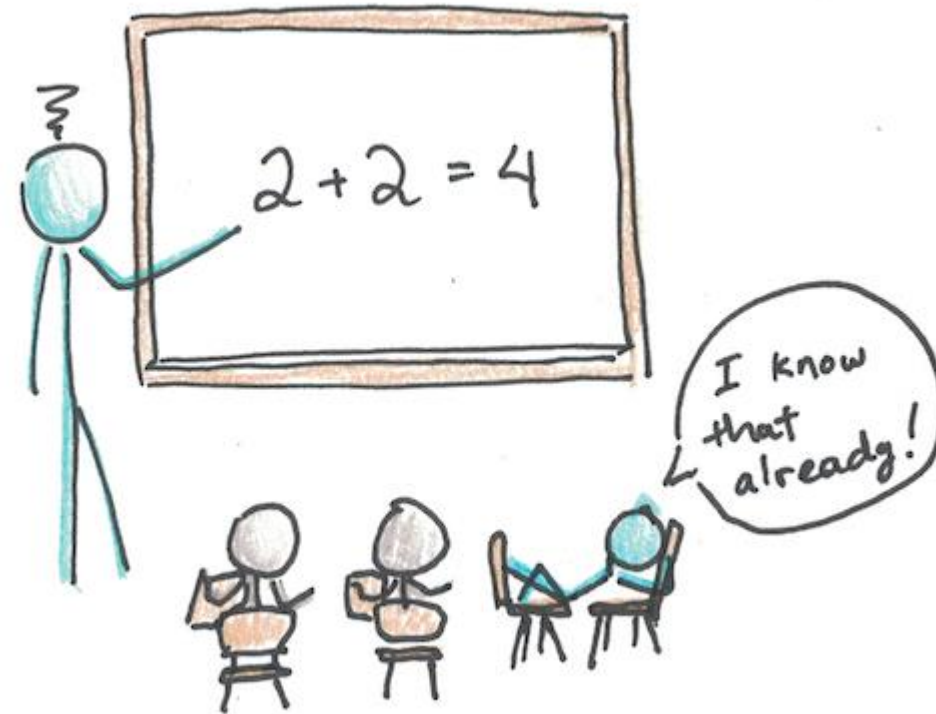
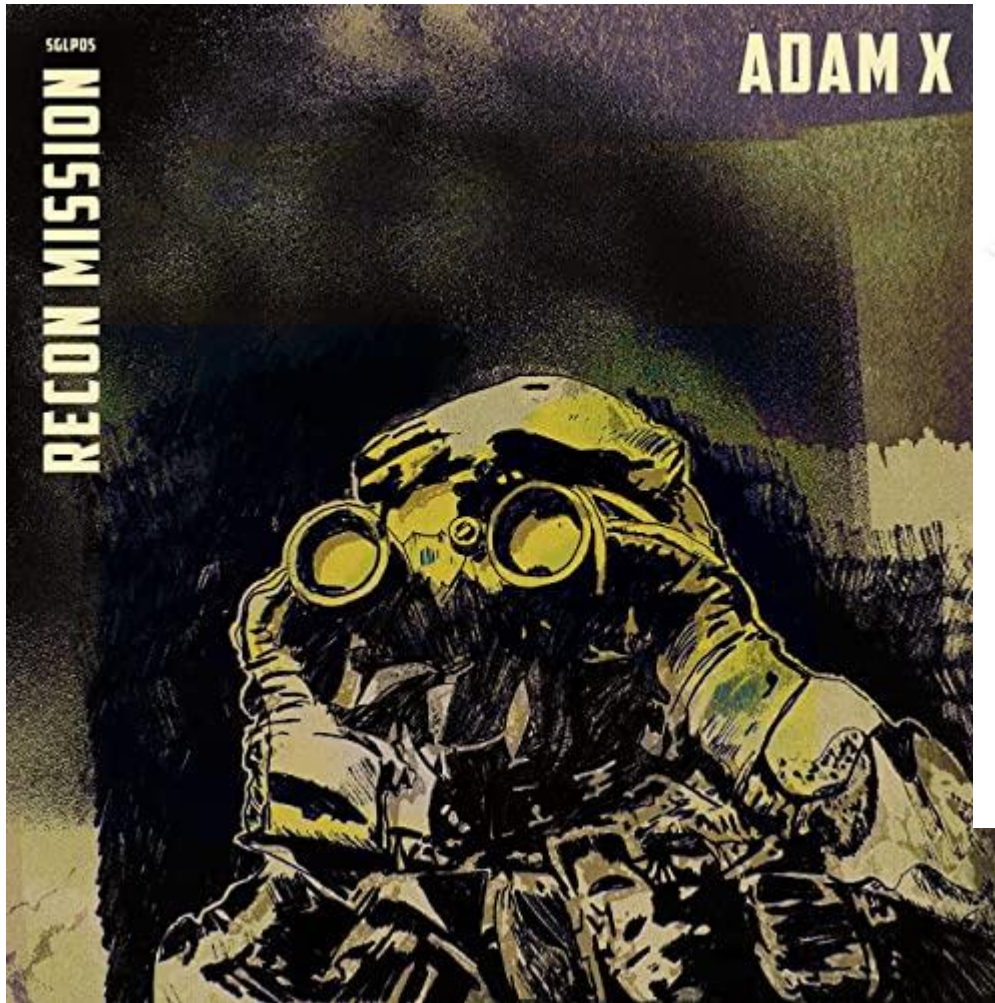


# Reconocimiento



compensar

fundación  
universitaria



What do we already know about this topic?

3 things I already know about this topic...

- 1.
- 2.
- 3.

3 things my partner already knows about this topic...

- 1.
- 2.
- 3.

3 things the class already knows about this topic...

- 1.
- 2.
- 3.

Tomado de: [https://m.media-amazon.com/images/I/A1zGFfENqKL.\\_SS500\\_.jpg](https://m.media-amazon.com/images/I/A1zGFfENqKL._SS500_.jpg)

© TODOS LOS DERECHOS RESERVADOS POR  
FUNDACIÓN UNIVERSITARIA COMPENSAR



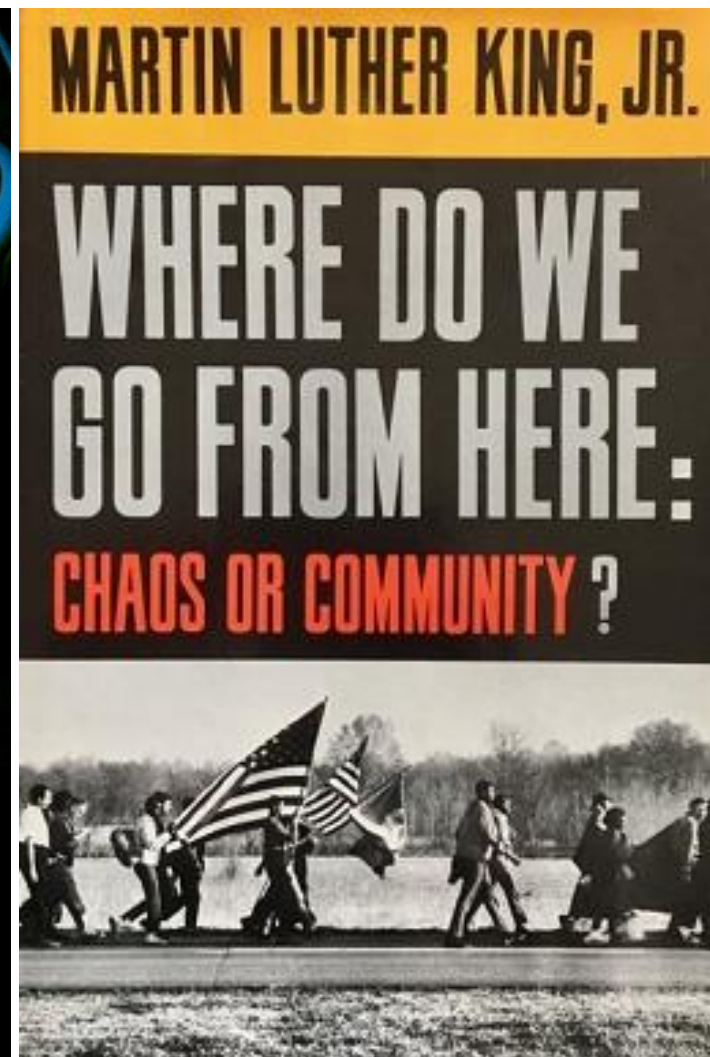
Lo que ya sabemos

Por qué es importante la Data?

Cómo podemos usar la Data?

Quién usa la data para su beneficio?

# Contextualización



Tomado de: <https://es.fanpop.com/clubs/alice-in-wonderland/images/25961950/title/alice-cheshire-cat-fanart>

© TODOS LOS DERECHOS RESERVADOS POR  
FUNDACIÓN UNIVERSITARIA COMPENSAR

# Cuál es la meta?





# Cuál es la meta?



# Syllabus

## **Etapla 0. Reconocimiento**

Concepto de dato y seguridad de dato

Aspectos éticos en el almacenamiento de dato

## **Etapla 1. Contextualización**

- El dato: Orígenes, fuentes, naturaleza y tipos de los datos
- Generalidades del proceso de extracción y almacenamiento de Datos Masivos

## **Etapla 2. Profundización**

- Importancia de la seguridad de los datos en el proceso de extracción y almacenamiento de datos masivos
- Políticas y legislación en seguridad de los datos masivos
- Técnicas y procedimientos de seguridad de datos masivos
- Técnicas de extracción
- Técnicas de almacenamiento y carga de datos masivos con seguridad

## **Etapla 3: Transferencia**

- Efectividad en la interpretación de las técnicas y procedimientos de extracción de datos (de acuerdo con el origen, fuente, naturaleza y tipos de los datos masivos)
- Cómo sustentar los procedimientos de seguridad
- Simulación de la integridad de los datos en la organización
- Socialización de las técnicas de extracción y almacenamiento aplicados: uso de modelos computacionales y realización de interpretación de los datos masivos.





# Qué necesitamos para llegar?

Aprendizaje:

E

T

L

Seguridad

4 unidades

En resumen:

Conceptos

Técnicas ETL

Herramientas: Python / PowerQuery

Aseguramiento

Ejecución de proyectos de software (devops, scrum)

Uso de herramientas de Cloud (Azure)

Problema a solucionar. (???)

# Profundización



compensar

fundación  
universitaria

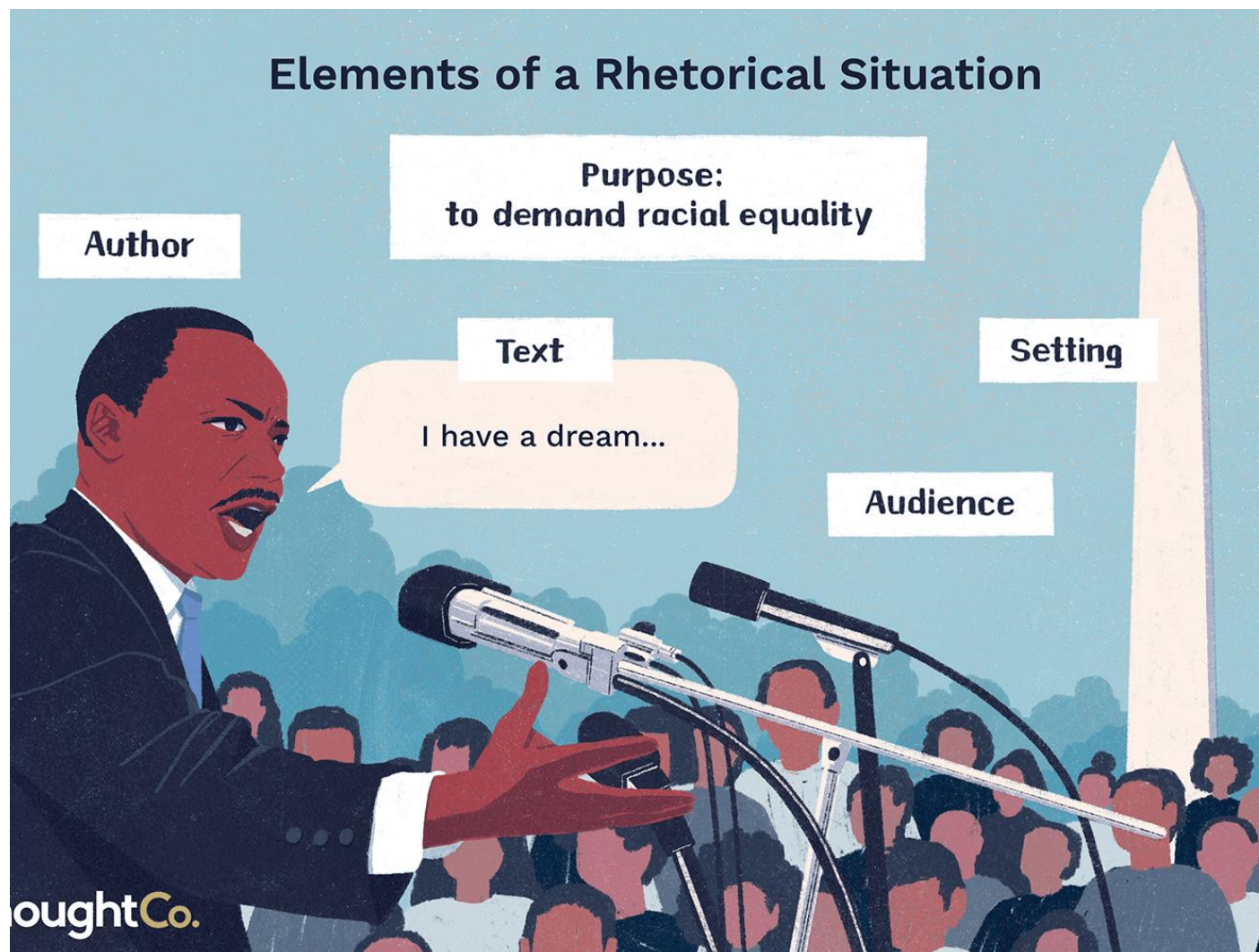


Empecemos por el comienzo.

Qué vamos a hacer?



# Transferencia



## ETHOS

USING CHARACTER, CREDIBILITY AND ETHICS TO PERSUADE



## PATHOS

USING EMOTIONS AND PASSION TO PERSUADE



## LOGOS

USING LOGICAL REASONING AND EVIDENCE TO PERSUADE

Tomado de: <https://www.thoughtco.com/rhetorical-situation-1692062>

© TODOS LOS DERECHOS RESERVADOS POR  
FUNDACIÓN UNIVERSITARIA COMPENSAR



## Compartir en comunidad

Publicar las iniciativas en el tablero.

Revisar las iniciativas e “inspirarse”

Formar equipos de 3

Presentar la iniciativa escogida de cada uno.

Seleccionar la iniciativa a trabajar en equipo.

Compartirla con el resto de equipos.

# Discusión



compensar

fundación  
universitaria



# book outage: what went wrong and did it take so long to fix after social form went down?

of users were unable to access Facebook, Instagram and up for hours while the social media giant scrambled to services



## Mark Zuckerberg apologises for Facebook, WhatsApp disruption

Facebook cited faulty configuration changes on its routers as the root cause of the nearly six-hour outage that prevented the company's 3.5 billion users from accessing its social media and messaging services.

## Jennifer Lawrence leaked nude photos: Apple iCloud password hack could be 'responsible for security breach'

Tech experts are now saying weak passwords gave hackers the chance to access private pictures thanks to a software glitch

By **David Raven**

13:41, 1 Sep 2014 | UPDATED 10:37, 23 Sep 2014





# What Google Knows

Google compiles enough data to build comprehensive portfolios of most users—who they are, where they go and what they do—and the information is all available at [google.com/dashboard](http://google.com/dashboard). Here are just a few things WSJ reporter Tom Gara found out about himself.

## GOOGLE SEARCH 64,019

Google thinks Tom performs most of his searches around 8 a.m. ET, but this is probably skewed by years spent outside the U.S.

## ANDROID DEVICES 3

Google knows all of Tom's synched Android phones, including the old Nexus S phone that he gave to his mom.

## WALLET 3

Credit cards (two expired) saved in Google Wallet, plus two shipping addresses and 13 itemized purchases since June 2009.

## DOCS 855

Documents Tom has created, plus the 115 he has opened that belong to other people.

## GMAIL 134,966

All of Tom's emails since he first got a Gmail account in 2004. Google also stores his 6,147 chats.

## CONTACTS 2,702

Google knows the people that Tom emails the most. At the top is a friend in Egypt.

## YOUTUBE 9,220

Videos Tom has watched, listed in chronological order, including a series viewed in June about canoes.

## GOOGLE PLAY 117

That's how many apps Tom has downloaded from Google's store.

## PASSWORDS 35

Number of website passwords saved in Google's Chrome browser.

## LOCATION

Willunga,  
South Australia

Due to an unknown glitch, Google bases Tom's location from one of his old Android phones, which he gave to his mother in Australia.

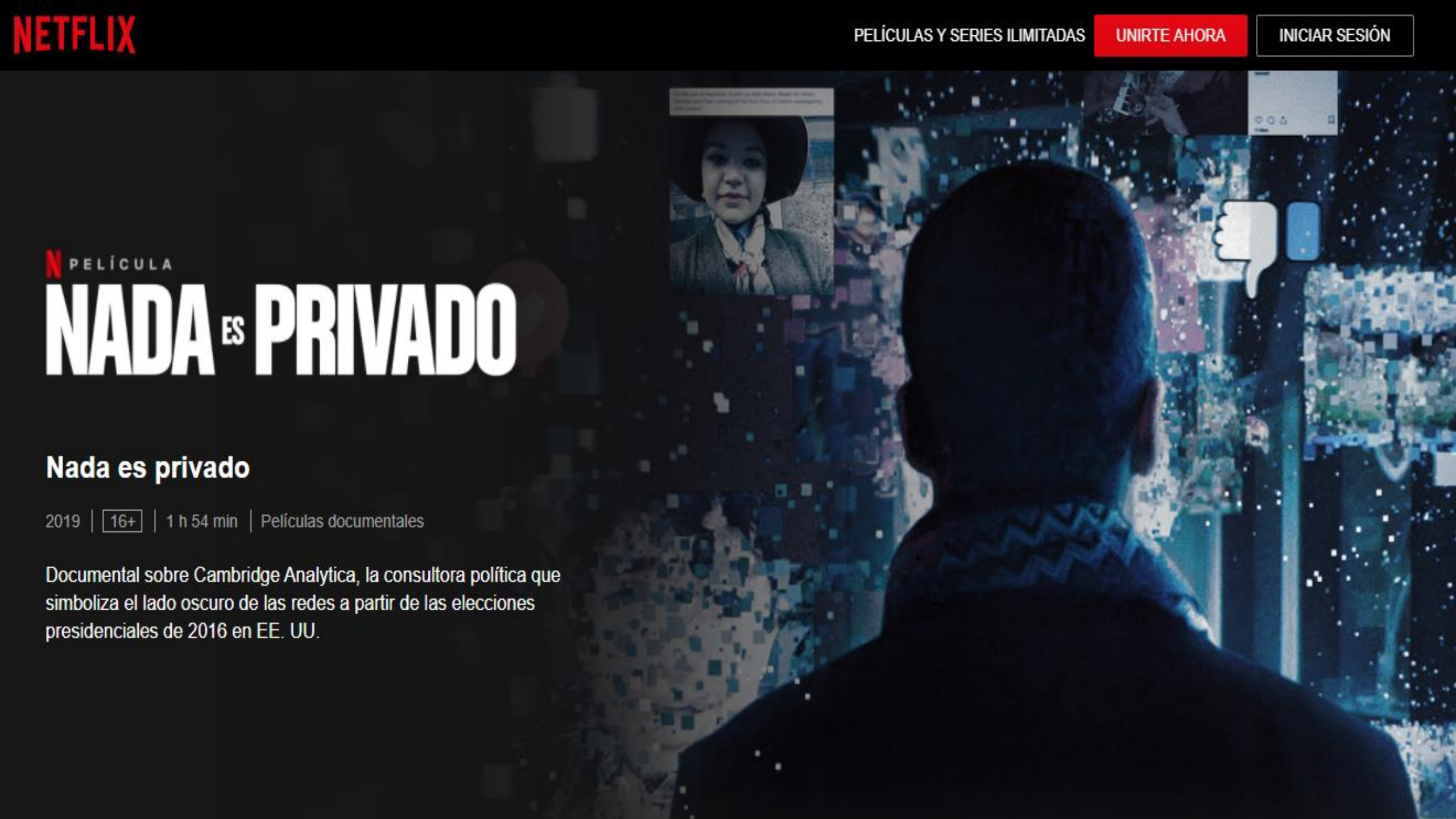
# Qué tanto sabe Google de nosotros?

**BIG BROTHER**



**IS WATCHING YOU**





N PELÍCULA

# NADA ES PRIVADO

## Nada es privado

2019 | 16+ | 1 h 54 min | Películas documentales

Documental sobre Cambridge Analytica, la consultora política que simboliza el lado oscuro de las redes a partir de las elecciones presidenciales de 2016 en EE. UU.

Si ya eres suscriptor, [actívalo sin costo](#)

## Suscríbete a EL TIEMPO DIGITAL

El medio colombiano informativo más consultado en Internet.

### PLAN DIGITAL ANUAL

**\$ 9.992 / MES**

Oferta especial de  
lanzamiento 35% de  
descuento

Valor anual con descuento

\$ 119.900

Valor anual sin descuento

~~\$ 184.900~~

SUSCRÍBETE

En este portal utilizamos "cookies" propias y de terceros para gestionar el portal, elaborar información estadística, optimizar la funcionalidad del sitio y mostrar publicidad relacionada con sus preferencias a través del análisis de la navegación. Puede conocer cómo deshabilitarlas u obtener más información [aquí](#)

ACEPTO

# Preguntas y Respuestas



## SPRINT REVIEW

Intro

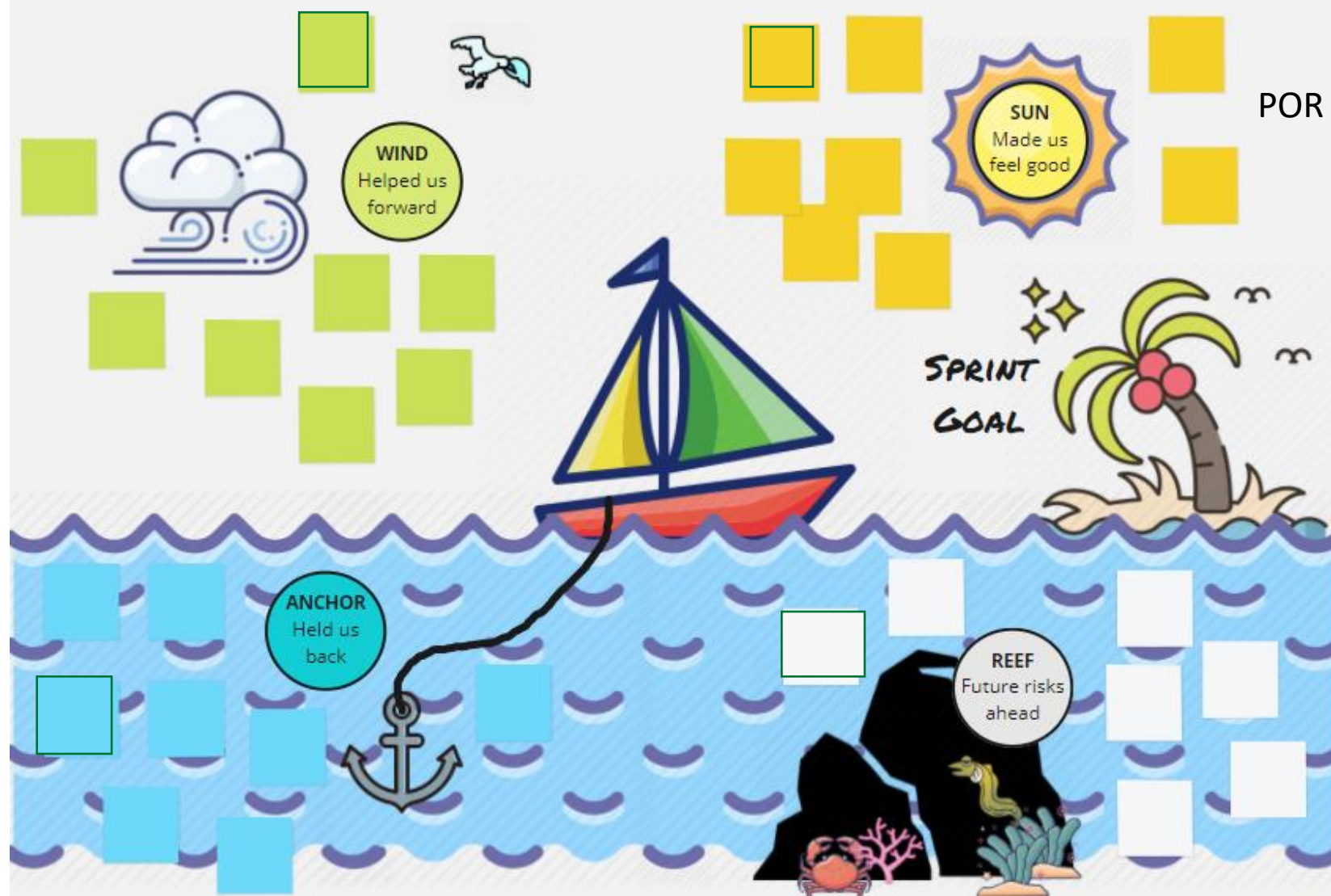
Discusión

Preguntas y respuestas





## SPRINT RETROSPECTIVE







fundación  
universitaria

VIGILADA MINEDUCACIÓN



Av. Calle 32 No. 17 - 30  
Pbx: 555 82 10  
**ucompensar.edu.co**  
Bogotá, D.C. - Colombia

ucompensar

