

BIG DATA AND DATA ANALYTICS

DIEGO HERRERA MALAMBO



dherrerambo



Soy Administrador de Empresas, Especialista en Gestión de Proyectos, con diplomados en Business Intelligence and Data Mining y Big Data & Business Analytics, actualmente me encuentro cursando la maestría en Estadística Aplicada y Ciencia de Datos. Tengo cerca de 10 años de experiencia en análisis de información. He implementado diferentes soluciones BI con planos/Excel, bases de datos MySQL/Postgre/MSSql además de servicios Cloud serverless de AWS, creando dashboards en PowerBi y Tableau. Manejo los lenguajes para análisis de información como SQL, Python, R y VBA. generando valor para las empresas permitiendo la toma de decisiones basadas en datos. Soy egresado de la primera promoción de DS4A.

DATA ENGINEER
DATA ANALYST
DATA SCIENCE

Big Data

Data Analytics

Business Intelligence

Machine Learning

Data Science

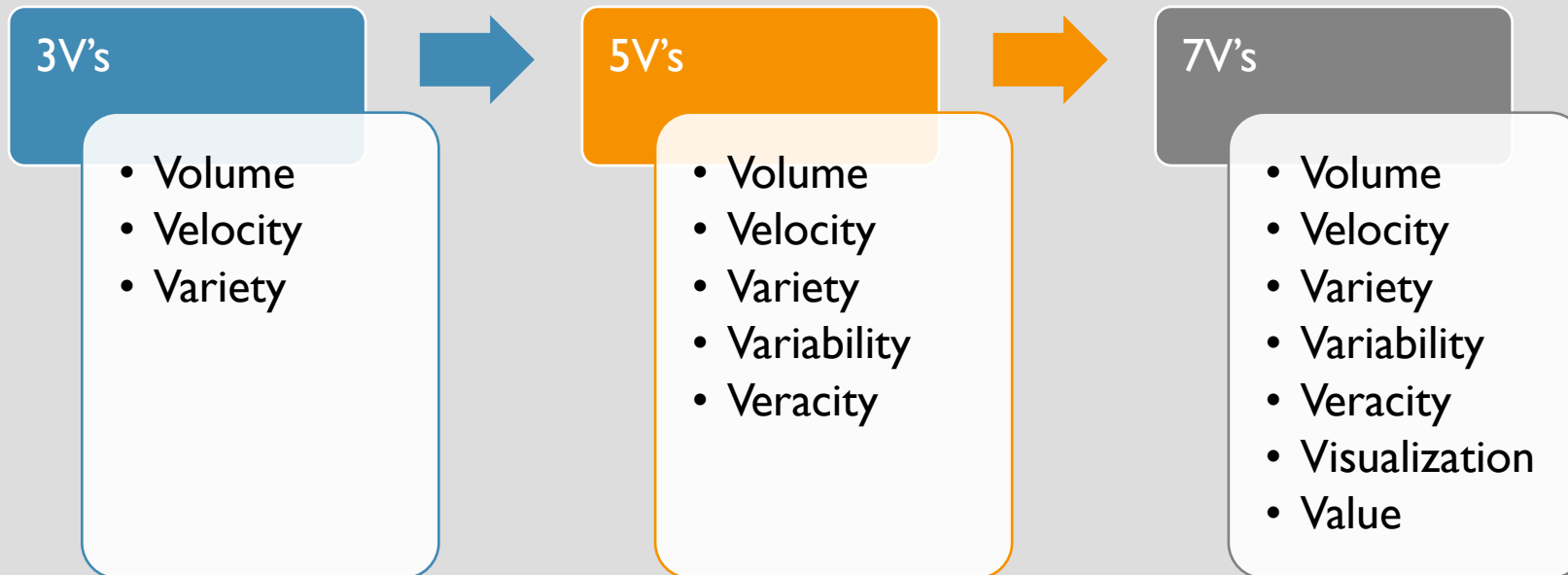
Cloud Computing

Artificial Intelligence

Transformación Digital

Blockchain

LAS “V” DEL BIG DATA



DONDE ENCONTRAMOS “BIG DATA” & “BIG DATA ANALYTICS”

- Redes Sociales
 - Servicios bancarios
 - Deportes
 - Seguridad informática
 - Gestion empresarial
 - Gobierno
 - Salud y medicina
 - Publicidad y Marketing
 - Logística y suministros
 - Desarrollo tecnológicos
 - Telecomunicaciones
 - Comercio electrónico
- y muchas mas....



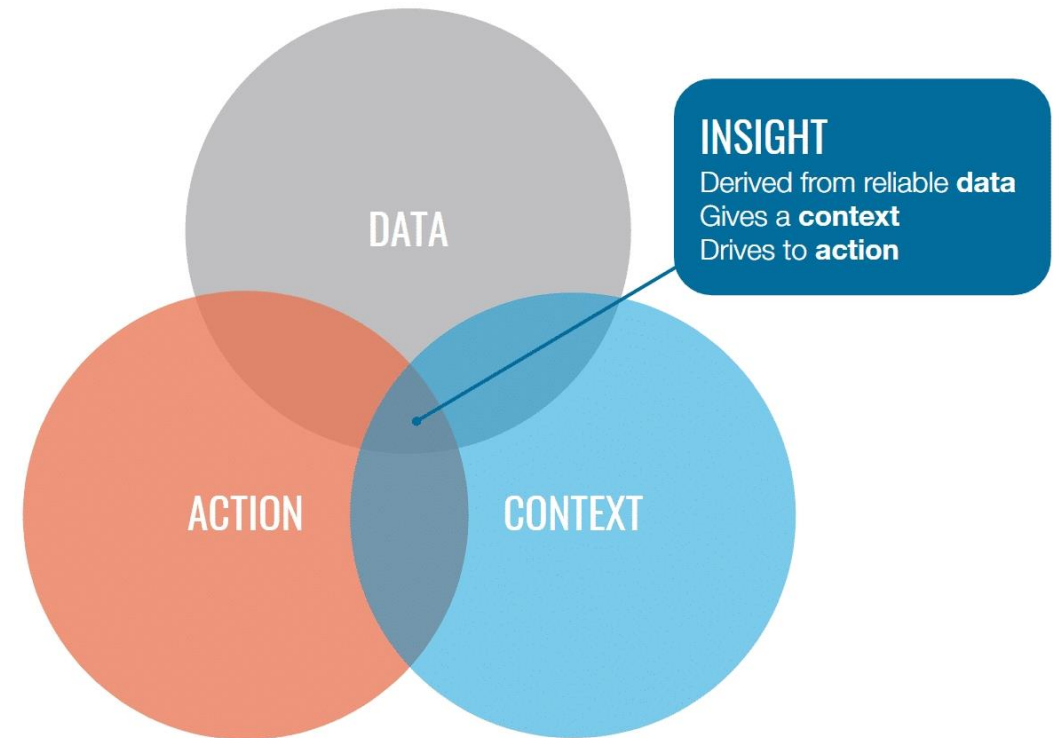
DETECCIÓN DE ANTÍGENOS



BIG DATA EN LOS TIEMPOS DEL COVID-19

DATA ANALYTICS

El Big Data y el Data Analytics han pasado de ser una moda a una necesidad en el quehacer de las empresas. Este ultimo es un proceso que es apoyado por diferentes software que permiten al usuario descubrir tendencias y/o patrones en los datos, permitiendo obtener información y llegar a *insights* del problema de negocio que estemos tratando



Insight: es la capacidad de llegar a un profundo conocimiento de alguien o algo

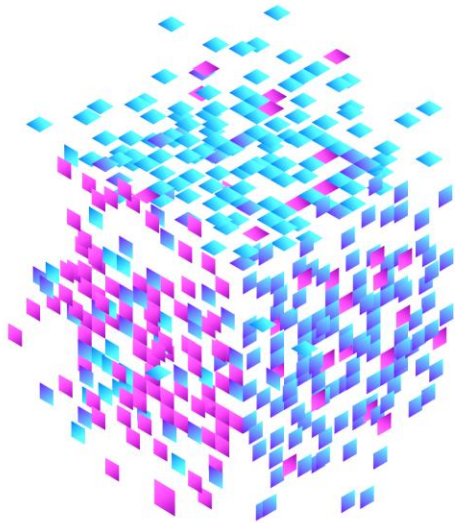


ANTES QUE NADA...

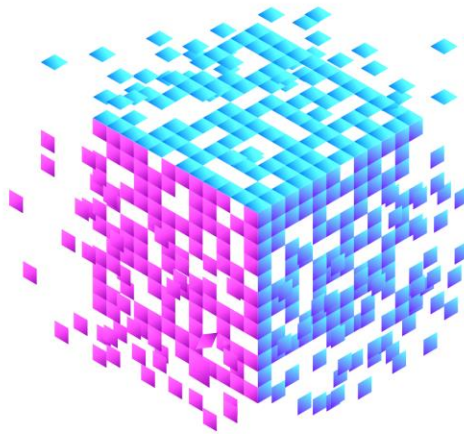
Realizar la pregunta correcta es más importante que realizar un excelente análisis, al hacer la pregunta correcta, podemos: detallar de forma clara y precisa el objetivo del análisis, priorizar variables y definir de forma efectiva nuestro KPI.

“Lo que no se mide, no se puede mejorar, lo que no se mejora, se degrada”

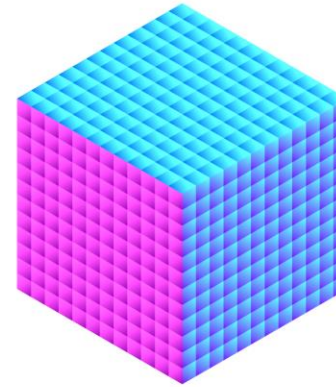
BIG DATA



ANALYTICS



DECISIONS



stargazr

DATA DRIVEN FLOW

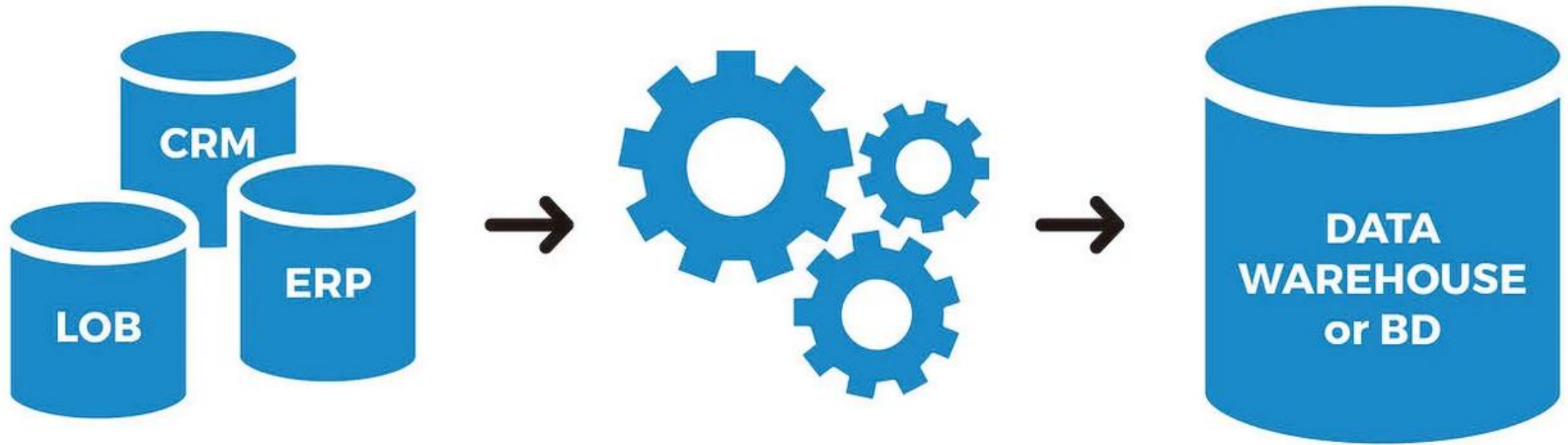
BUSINESS INTELLIGENCE (BI)

Se denomina inteligencia empresarial, inteligencia de negocios, inteligencia comercial o BI (del inglés business intelligence), al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitectura técnicas, los cuales están enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (March 2022)



EXTRACT, TRANSFORM AND LOAD (ETL)

CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)



Comprensión del negocio: Definición de fuentes a usar para la extracción de datos



Comprensión de los datos: exploración de datos para definir las variables a usar



Preparación de los datos: Definición de estructura repositorio y transformaciones.



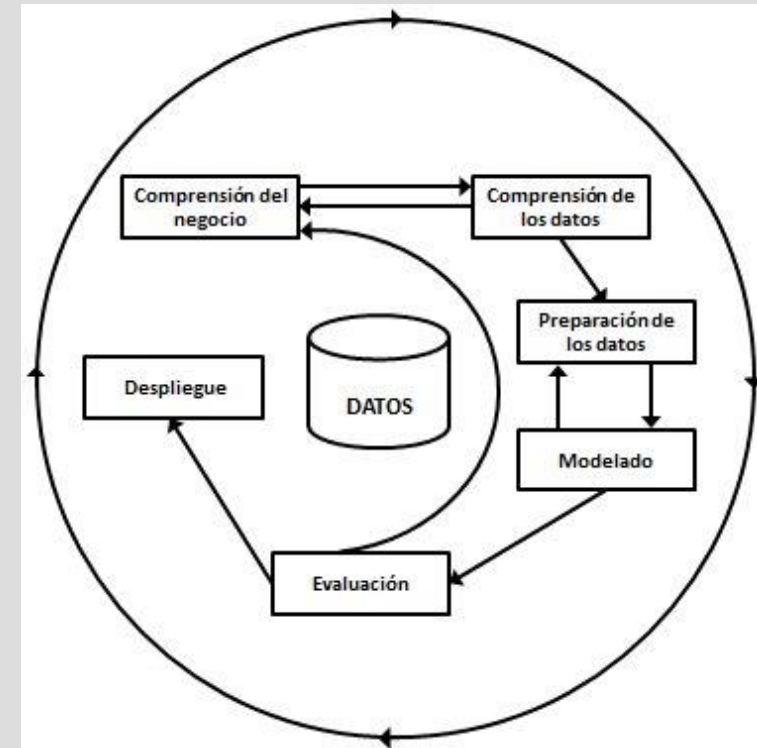
Modelado: Aplicación de modelos de Machine Learning



Evaluación: Toma de decisión en base a el análisis de datos

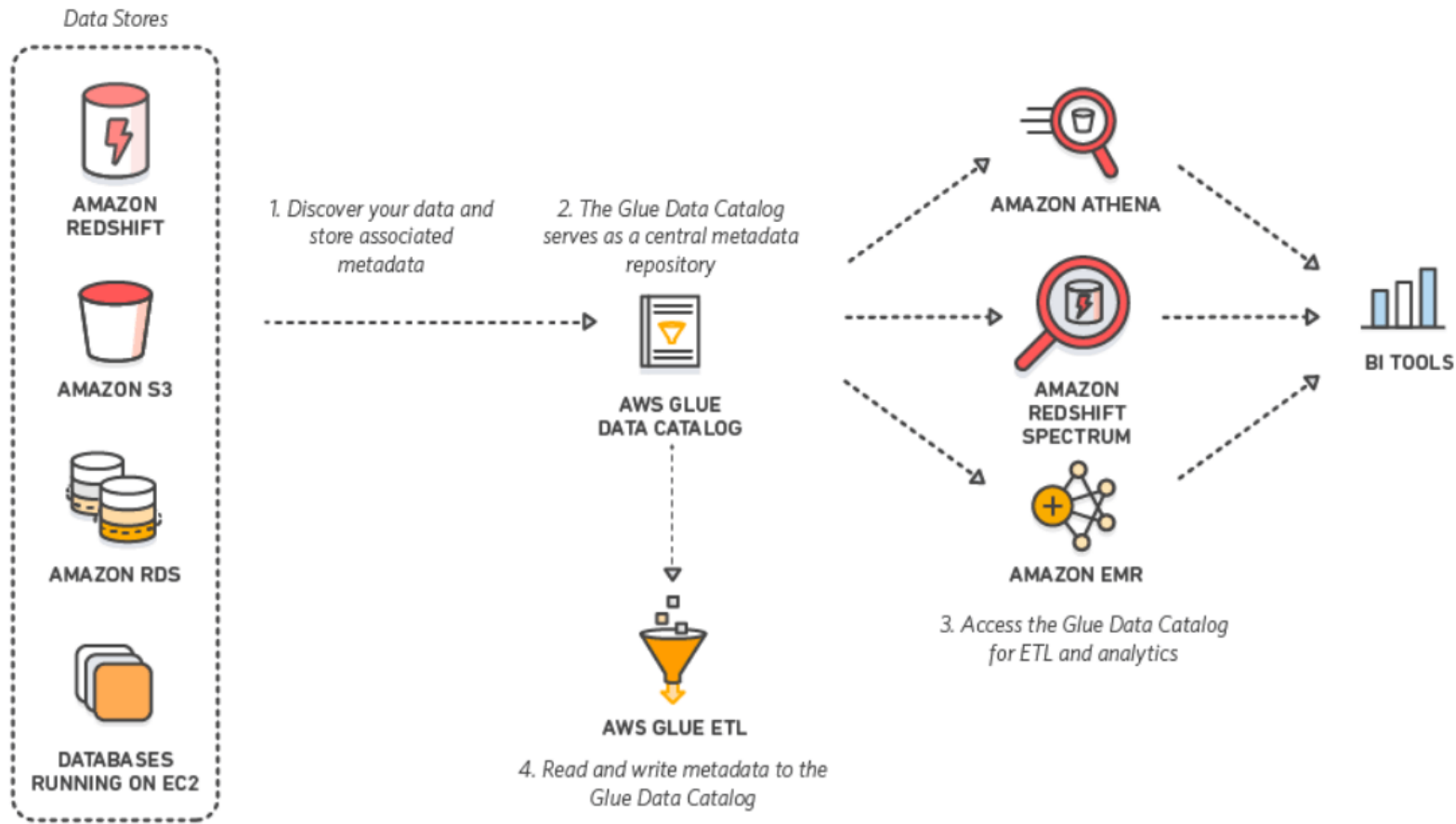


Despliegue: Clasificación, Clusters, Predicciones



PLATAFORMAS CLOUD





EJEMPLO PROCESO BI CON AWS

BIG DATA PROBLEMS?

- Ética
- Escandalo de Facebook
- Elecciones
- Privacidad
- Covid-19
- Data Leakage
- Data Breach

Protección de Datos

<https://www.tendencias.kpmg.es/2018/04/etica-big-data/>
<https://www.bbc.com/mundo/noticias-49093124>
<https://www.forcepoint.com/es/cyber-edu/data-leakage>

Portada / Red Forbes /

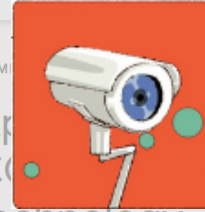
Palmira Tapia Palacios

abril 16, 2020 @ 6:30 am

El derecho a la privacidad en tiempos de COVID-19

¿Qué tan dispuestos estamos de limitar la protección de nuestra información personal a cambio de una mayor protección del Estado ante el COVID-19? La respuesta podría definir el futuro de la democracia.

Google The Keyword



Apple and Google partner on contact tracing technology

Published Apr 10, 2020



Data breach example #4: LinkedIn

How many affected?

165 million accounts.

What happened?

Social media platform, LinkedIn, suffered a data breach that compromised the personal information of 165 million user accounts. The data has since been reported as up for sale on the [dark web marketplace](#).

China's Virus Apps May Outlast the Outbreak, Stirring Privacy Fears

With the disease there mostly under control, officials are looking for new uses for the government software that's now on many phones.

NEWS

Home | US Election | Coronavirus | Video | World | US & Canada | UK | Business | Tech | Science | Stories

Will Big Data lead to Big Brother?

By Gordon Corera
Security correspondent, BBC News

🕒 17 November 2015



RETOS Y OPORTUNIDADES

- Reglamentación
- Talento humano
- Migración a la nube
- Calidad de los datos
- Ciberseguridad
- Transformación digital
- Protección de datos



Ley 1581 de 2012

Ley de Protección de Datos
Personales. Colombia

Documento CONPES

3920

CONSEJO NACIONAL DE POLÍTICA ECONÓMICA Y SOCIAL
REPÚBLICA DE COLOMBIA
DEPARTAMENTO NACIONAL DE PLANEACIÓN

POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS (BIG DATA)

Departamento Nacional de Planeación
Ministerio de Tecnologías de la Información y las Comunicaciones
Superintendencia de Industria y Comercio

Versión aprobada

Bogotá, D.C., 17 de abril de 2018

Colombia primer país en Latinoamérica con política para la explotación de datos

CUARTA REVOLUCIÓN INDUSTRIAL ▶

Intensiva en conocimiento, desarrollo tecnológico, innovación y explotación de los datos



1.^a

1784

Producción mecánica
y energía de vapor



2.^a

1870

Producción en masa
de energía eléctrica



3.^a

1969

Electrónica y TI



4.^a

HOY

Sistemas físicos cibernéticos
(Big Data, IoT, Smart Cities,
inteligencia artificial)

◀ **DATIFICACIÓN:**
transformación del
mundo en datos
procesables y
cuantificables

Crecimiento de datos en el mundo (*billones de gigabytes*)

1,2
billones
2010

16,1
billones
2016

163
billones
2025

2017



1,9 billones
Búsquedas



84,4 billones
Correos enviados



241.000 millones
Tweets



2,2 billones
Videos vistos



25.000 millones
Fotos subidas

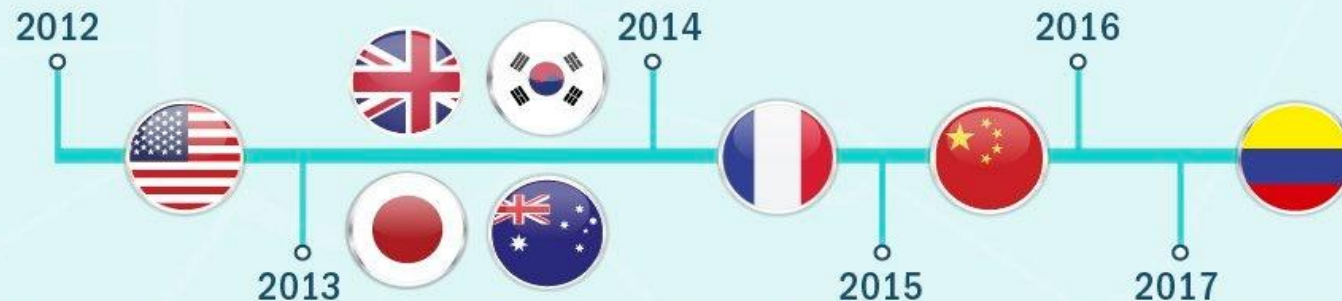


32,2 millones
Páginas hackeadas



89.000 millones
Llamadas

Colombia será la **primera** en América Latina y la **octava** en el mundo con política para la explotación de datos



* En otros países se hace referencia al uso de datos como parte de otras políticas, pero no son una política autónoma e integral.

**EFFECTOS POSITIVOS EN
REDUCCIÓN DE LA CORRUPCIÓN:**

Fuente: Open Data Barometer & Transparency International



78%

La correlación del nivel de datos abiertos de un país y el índice de percepción de corrupción es del 78%, es decir, a medida que hay mayores datos abiertos hay menor percepción de corrupción.

1,62%



2016

2,5%



2020

En 2016 la explotación de datos representó el 1,62% del PIB de la Unión Europea y para 2020 se espera que represente el 2,5%.

El documento CONPES de Big Data aborda cuatro ejes estructurales:

1

Generación de
datos digitales
Insumo

2

Cultura de datos
Demanda

3

Capital
humano
Para la
explotación de
datos (oferta)

4

Marco jurídico,
ético e institucional

Que busca habilitar la generación de valor y reforzar la protección de los individuos en el contexto de disponibilidad masiva de datos.

METAS CONPES BIG DATA A 2022



Pasar del 51% que se tenía en 2017 al 100% el promedio de activos públicos digitalizados y publicados



Impulsar el porcentaje de entidades con proyectos de aprovechamiento de datos, que en 2017 era el 9,3% y para 2022 se espera alcanzar el 90%



Lograr que por lo menos el 50% de las entidades públicas desarrollen proyectos de aprovechamiento de datos para mejorar los servicios prestados a la ciudadanía

Estas se obtendrán mediante la ejecución de **45 acciones** que articulan competencias de **10 entidades públicas**

Las inversiones suman aproximadamente **\$16.728 millones**

Infografía:

<https://protecdatalatam.com/colombia-primer-pais-en-latinoamerica-con-una-politica-publica-para-la-explotacion-de-datos-big-data/>
<https://www.dnp.gov.co/Paginas/Colombia-primer-pa%C3%ADs-en-Latinoam%C3%A9rica-con-una-pol%C3%ADtica-p%C3%BAblica-para-la-explotaci%C3%B3n-de-datos-Big-Data.aspx>

WWW.DATOS.GOV.CO



TENDENCIAS

- <https://addepto.com/big-data-trends-2020/>
- <https://www.infoworks.io/big-data-trends/>
- <https://www.kdnuggets.com/2018/09/10-big-data-trends.html>
- <https://www.geeksforgeeks.org/how-do-companies-use-big-data-analytics-in-real-world/?ref=rp>



EXPERIENCIAS EN EL SECTOR SALUD

CRM + BI

“La capacidad de manipular grandes volúmenes de información y poder interactuar con ellos no es una necesidad para el futuro, lo es para el presente”

DIEGO HERRERA MALAMBO

Administrados de Empresas – Especialista en Gestión de Proyectos / Consultor BI

<https://www.linkedin.com/in/dherrerambo> - dherrerambo@gmail.com



“If you do not know how to ask the right questions, you discover nothing”

W. E. Deming