

## **Módulo 12**

### **Datos masivos**

*Mtro. Omar Mendoza González*



**DGTIC**

**Universidad Nacional Autónoma de México**

**Dirección General de Cómputo y de Tecnologías de Información y Comunicación**

# Contenido

## 1. Introducción al Big Data

### 1.1 Conceptos básicos

#### 1.1.1 Huella digital

#### 1.1.2 Fuentes de datos

### 1.2 Tecnologías para Big Data

### 1.3 Aplicaciones del Big Data

# Huella digital



# Somos seres digitales

Enero  
2021

## POSESIÓN DE DISPOSITIVOS

Porcentaje de usuarios entre 16 a 64 años que poseen cualquier dispositivo



Teléfono móvil  
(cualquier tipo)



98.9%

Teléfono  
inteligente



98.7%

Teléfonos que no  
son inteligentes



11.4%

Laptop o PC



74.2%

Tablet



48.0%

Dispositivo para ver  
televisión en Streaming



24.0%

Consola de  
videojuegos



48.3%

Dispositivo inteligente  
para el hogar



12.2%

Reloj  
inteligente



22.6%

Dispositivo de  
realidad virtual



6.3%

we  
are  
social



Hootsuite

Traducción hecha por:

branch



# Somos seres digitales



# Conceptos Básicos de Big Data

Big Data se refiere a los volúmenes dinámicos, grandes y dispares de datos creados por personas, herramientas y máquinas



# Big Data

“Activos de información de alto **volumen**, alta **velocidad** y/o alta **variedad** que demandan formas rentables e innovadoras de procesamiento de información que permitan una **visión** mejorada, toma de decisiones y automatización de procesos”.

# Big Data

Es más que cantidad

Volumen

Procesos

Enfoques

Herramientas



# Big Data

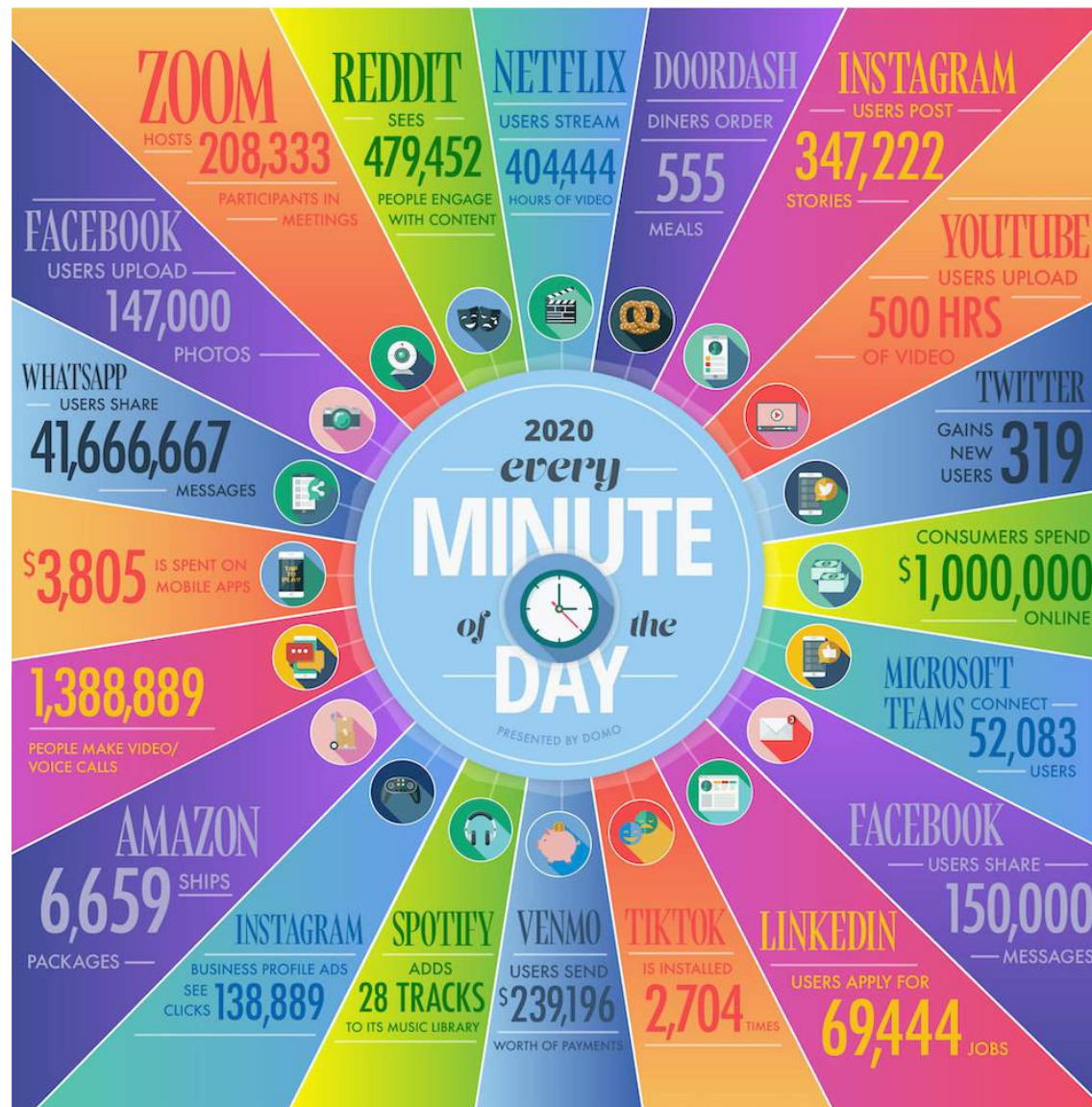
Rastro digital que puede ser analizado

Grande volúmenes de datos creados de forma dinámica por personas, herramientas y máquinas

Alto volumen, alta velocidad y alta variedad que demanda costos efectivos y formas innovadoras de procesamiento

Fuentes tradicionales y novedosas de datos para analizar y generar información

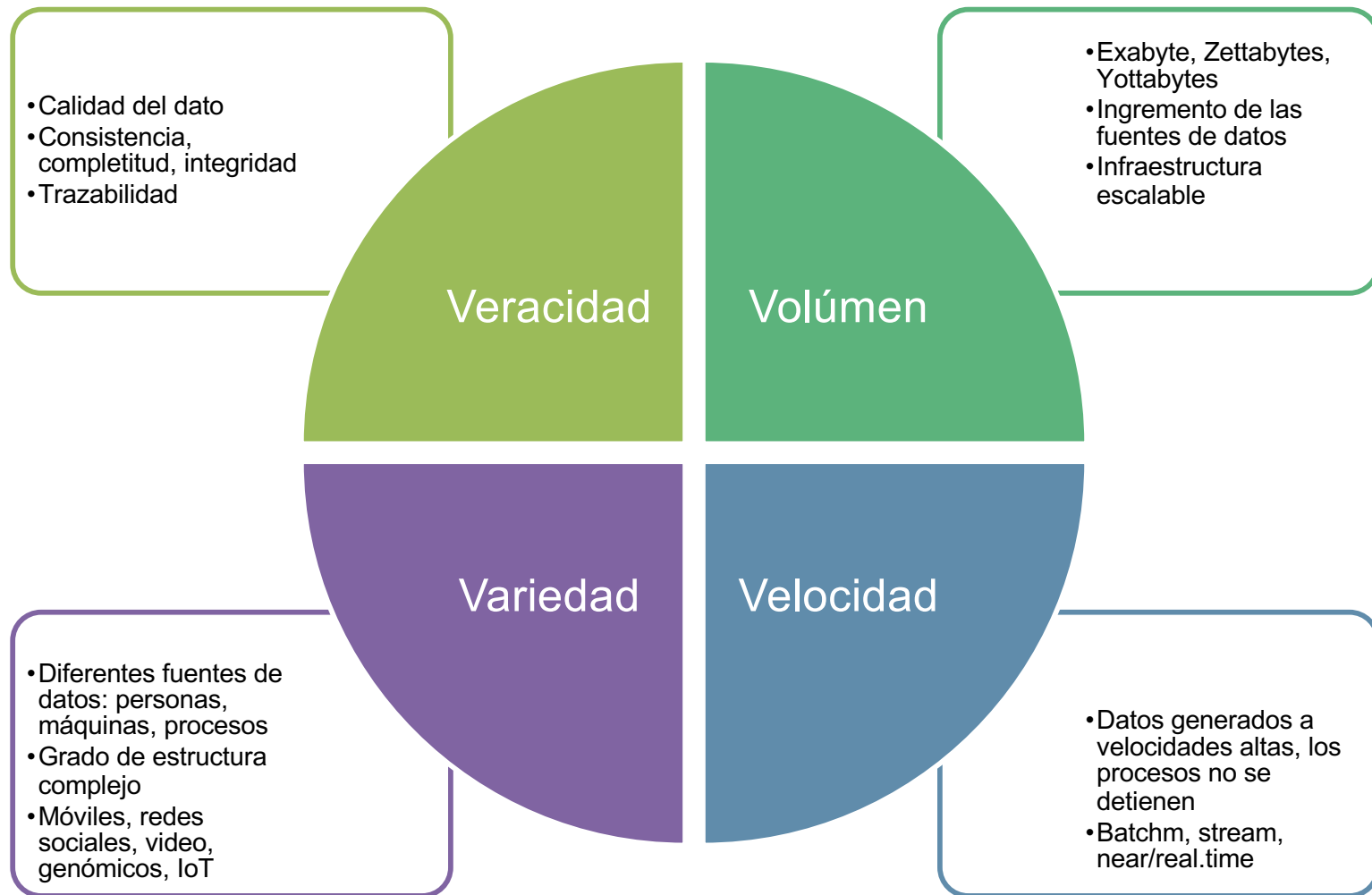
# Big Data



# Las Vs del Big Data

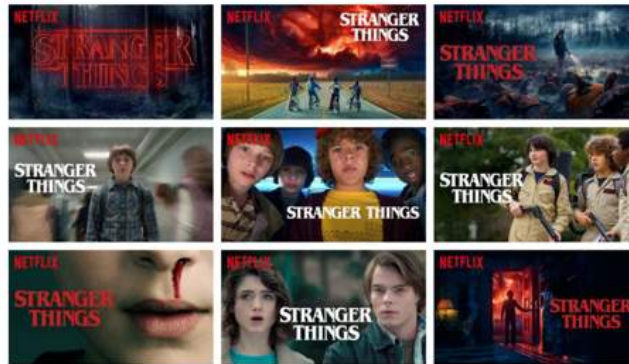


# Las Vs del Big Data





# El Big Data es nuestro día a día





# Aplicaciones del Big Data

**Marketing y Ventas**

**Finanzas**

**Lucha contra el  
crimen**

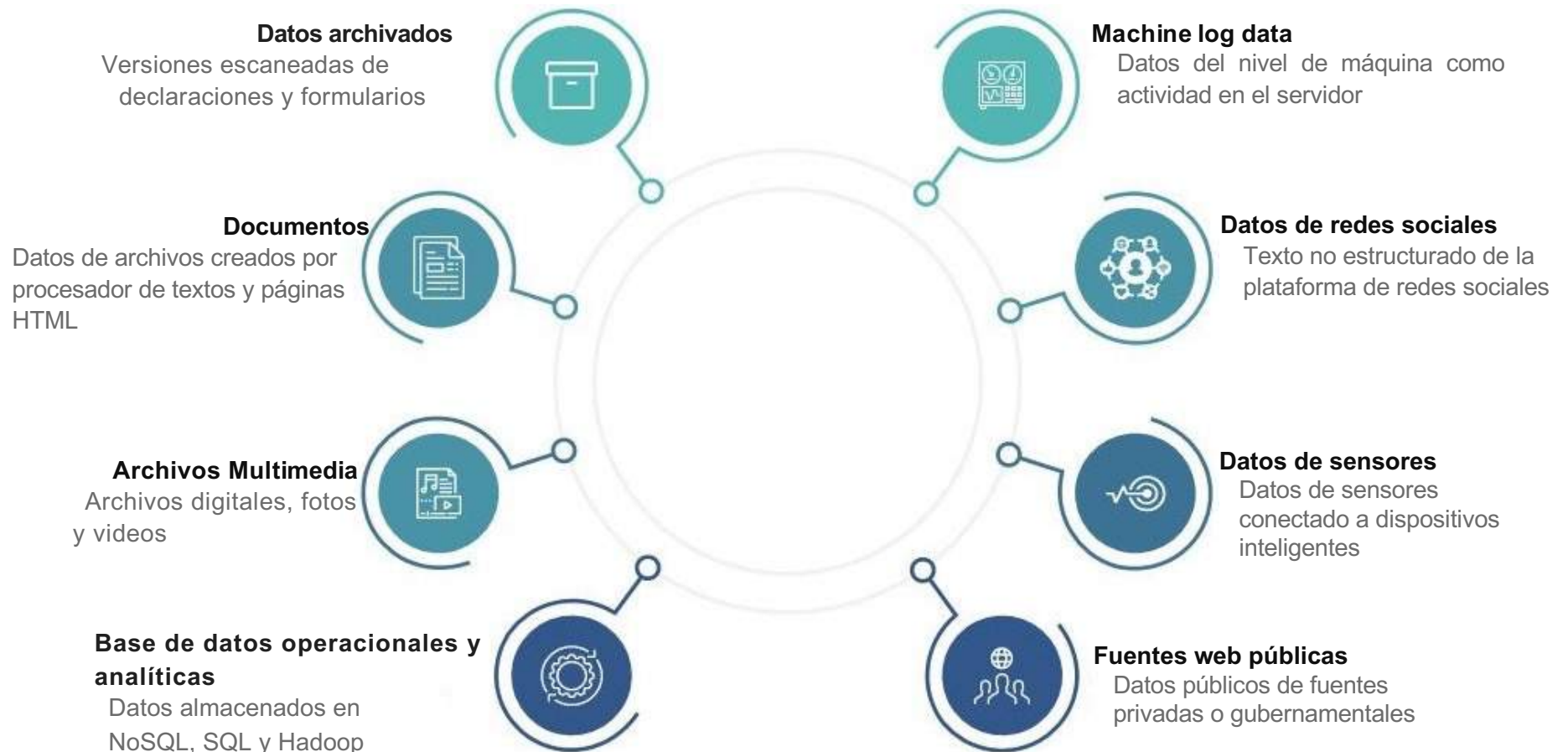
**Deportes**

**Política**

**Telecomunicaciones**

**Salud y Sanidad**

# Fuentes de datos del Big Data

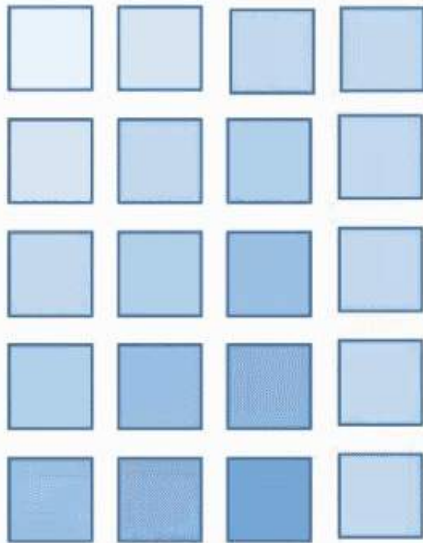


# Que tan grandes son lo datos

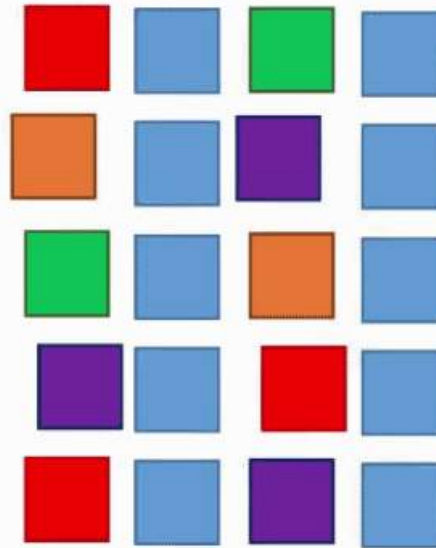


# Tipos de datos por categoría

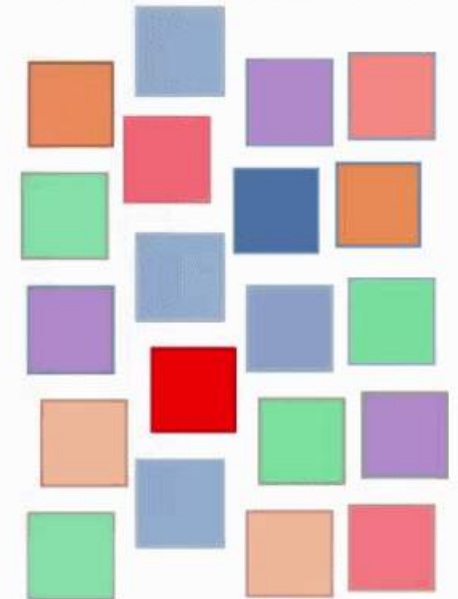
## Datos Estructurados



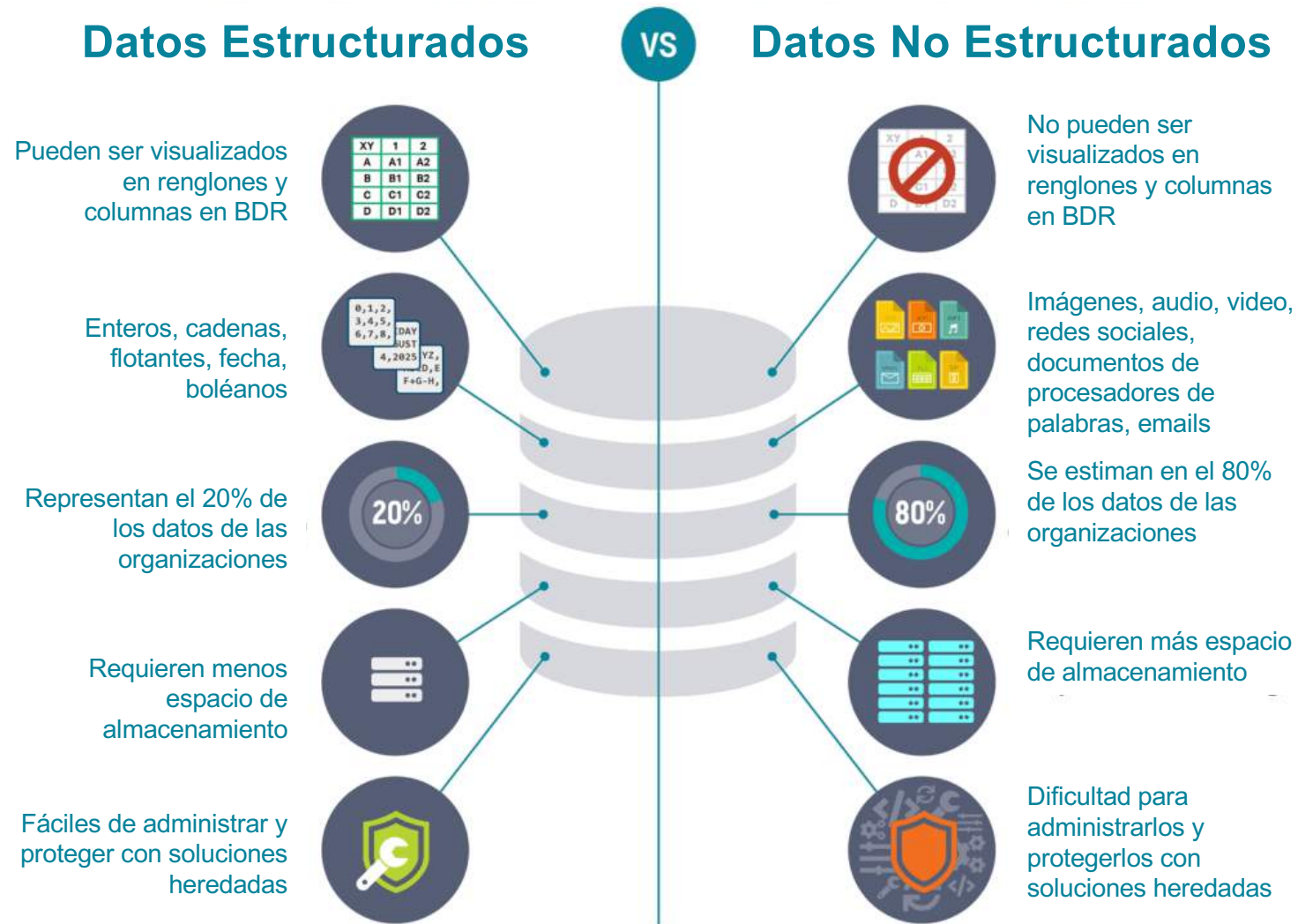
## Datos Semiestructurados



## Datos No Estructurados



# Tipos de datos por categoría





# Tipos de datos por categoría

## Estructurados

- Creados
- Provocados
- Dirigido por transacciones
- Compilados
- Experimentales

## No Estructurados

- Capturados
- Generados por usuarios

## Multi-estructurados o híbridos

- Datos de mercados emergentes
- E-commerce
- Datos meteorológicos

# Tipos de datos por origen

## Web y Redes Sociales

- Información sobre clicks en vínculos y elementos
- Búsquedas
- RRSS
- Contenido Web

## Comunicación entre máquina

- Lecturas RFID
- Señales GPS
- Otros sensores

## Big Transaction

- Registros de comunicaciones
- Registros de pagos

## Biométricos

- Reconocimiento facial
- Información genética (ADN)
- Escaneo de la retina
- Huellas digitales

## Generados por personas

- Llamadas a call centers
- E-mail
- Registros médicos electrónicos

# Identificar las fuentes de datos de la organización

**Fuentes Internas**



**Redes Sociales**

**Fuentes Externas**

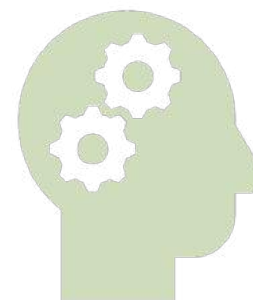


**Open Data**

# Categorías de tecnologías BigData

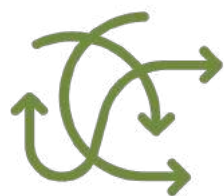


**Tecnologías operativas  
de Big Data**



**Tecnologías analíticas  
de Big Data**

# Categorías de tecnologías BigData



**Integración**



**Gestión**



**Análisis**



# ELT



# Tendencias en tecnologías de Big Data



**Base de datos NoSQL**



**Data Lakes**



**Blockchain**



**Apache Spark**



**Ecosistema Hadoop**



**Inteligencia artificial**



**Análisis predictivo**



**Análisis prescriptivo**



**Programación Python / R**

# Tecnologías básicas de Big Data



Apache Hadoop



HDFS



MapReduce



Apache Spark



RDD



SQL



**Transformaciones**



**Ingesta de datos**



Apache Hive



**NoSQL**



**Programación Python / R**

# Contacto

Omar Mendoza González

*Profesor de carrera ICO FES Aragón*

omarmendoza564@aragon.unam.mx

# Referencias

- **Corea, Francesco, An Introduction to data : everything you need to know about AI, Big data and data science / Francesco Corea -- Cham, Switzerland : Springer, [2019].--** xv, 131 páginas : ilustraciones (Studies in Big data, 2197-6503 ; 50 )
- **Casas Roma, Jordi, Big data : análisis de datos en entornos masivos / Jordi Casas Roma, Jordi Nin Guerrero, Francesc Julbe López -- Barcelona : Editorial UOC, 2019** 287 páginas : ilustraciones (Tecnología ; 623 ).
- **Caballero, Rafael, Big data con Python recolección, almacenamiento y proceso /** Rafael Caballero Adrián Riesco Enrique Martín: Universidad Complutense de Madrid Editorial AlfaOmega, 2019 282 páginas
- **Rioux, Jonathan, Data Analysis with Python and PySpark / Jonathan Rioux: Editorial** Manning Publications, 2020 259 páginas
- **Singh, Pramod, Machine Learning with PySpark: With Natural Language Processing and Recommender Systems / Pramod Singh: Editorial Apress, 2019** 233 páginas