

DATOS MASIVOS I

CONCEPTOS BÁSICOS

Gibran Fuentes-Pineda
Enero 2020

MUCHOS MÁS DATOS Y A MAYOR VELOCIDAD

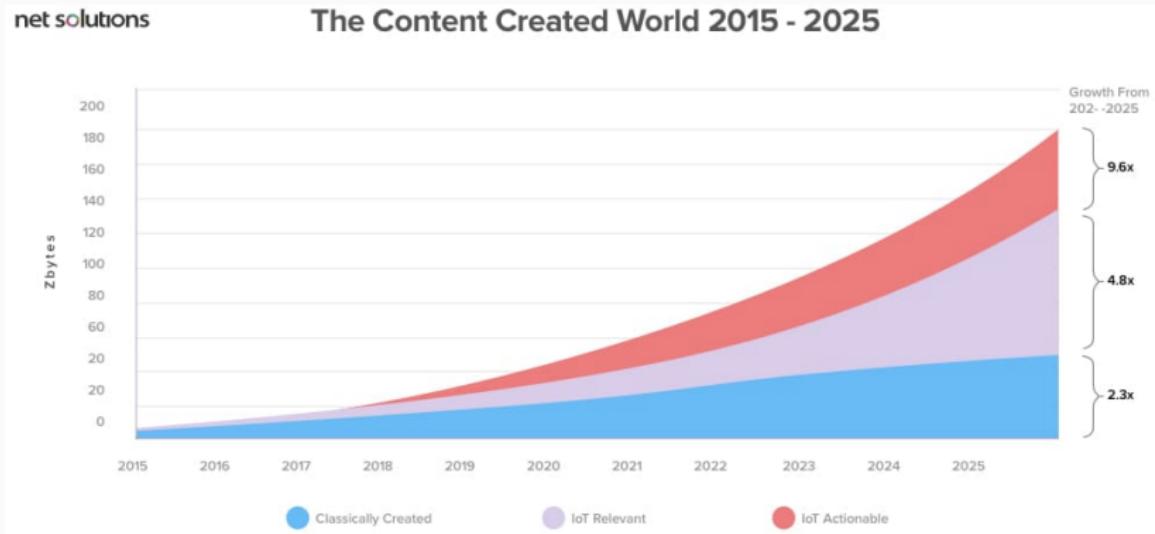


Imagen tomada de <http://www.tech-dynamics.com/wp-content/uploads/2014/02/BigDataChart.png>

MÁS PARÁMETROS Y MÁS DATOS NO ESTRUCTURADOS

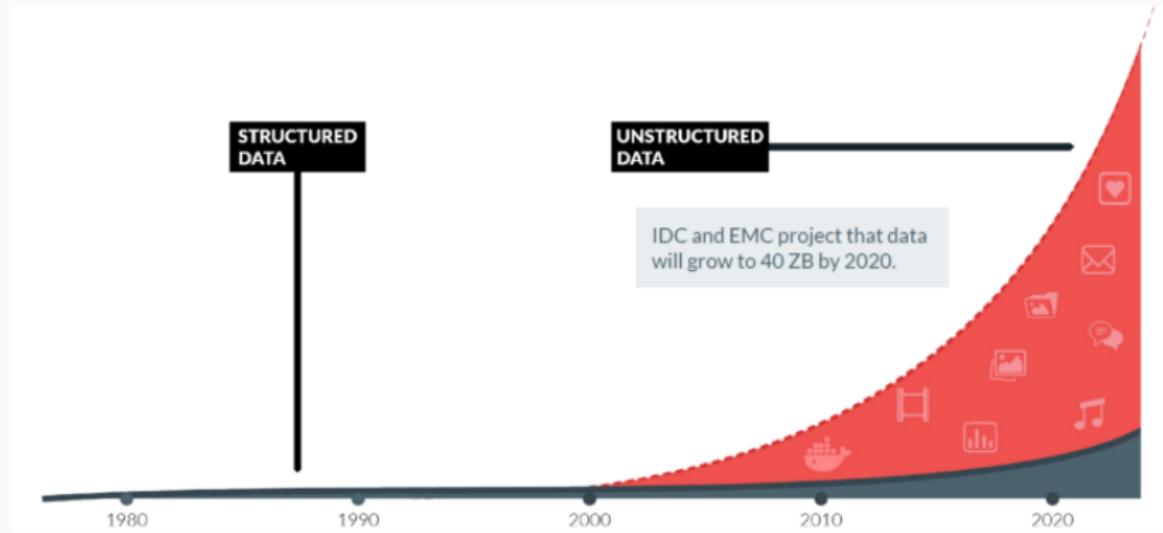
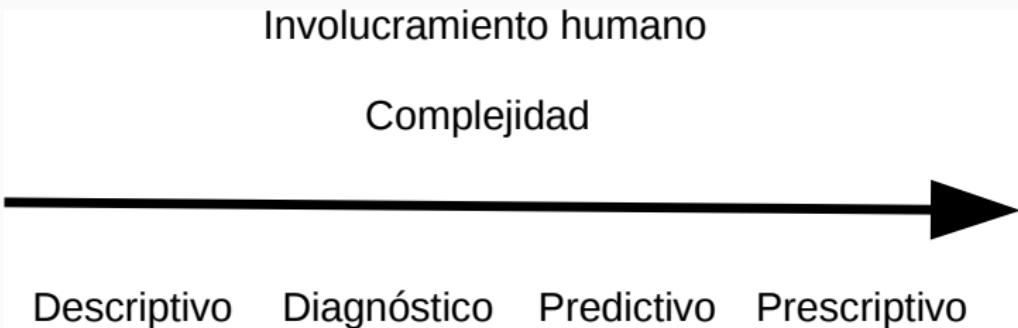


Imagen tomada de <https://www.datanami.com/2017/02/01/solving-storage-just-beginning-minio-ceo-periasamy/>

ANÁLISIS MENOS DESCRIPTIVO, MÁS PREDICTIVO/PRESCRIPTIVO



MÁS APLICACIONES (1)



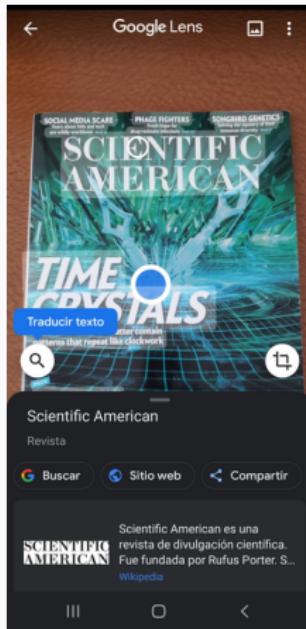
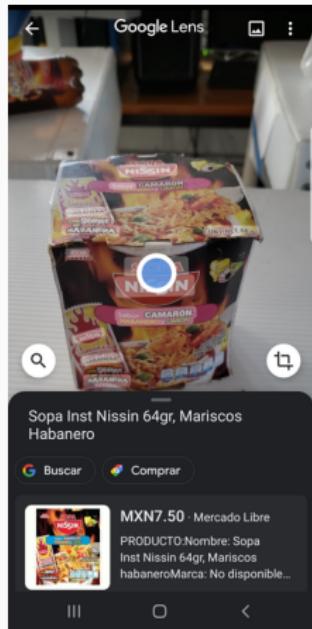
Imagen tomada de *The Washington Post*

MÁS APLICACIONES (2)



Imagen tomada de Snavely et al. Photo Tourism: Exploring Photo Collections in 3D, *ACM Transactions on Graphics* , 2006.

MÁS APLICACIONES (3)



MEJORES MODELOS (1)

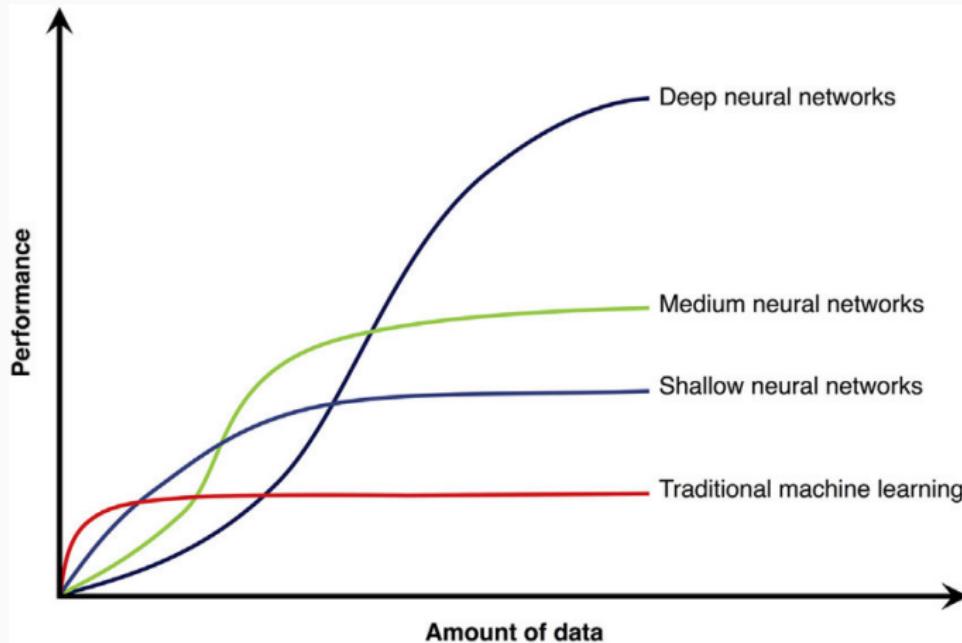


Imagen tomada de Tang et al., Canadian Association of Radiologists Journal 69(2), 2018

MEJORES MODELOS (2)

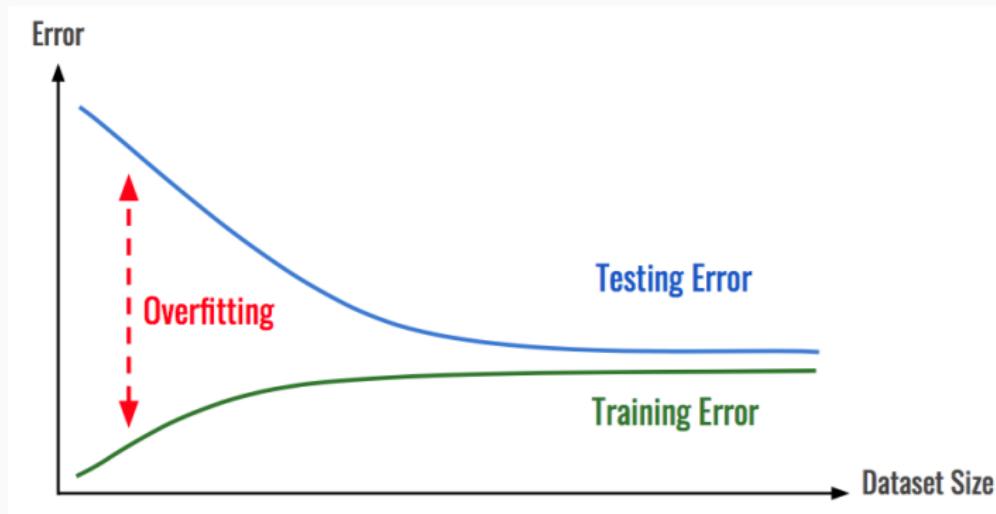
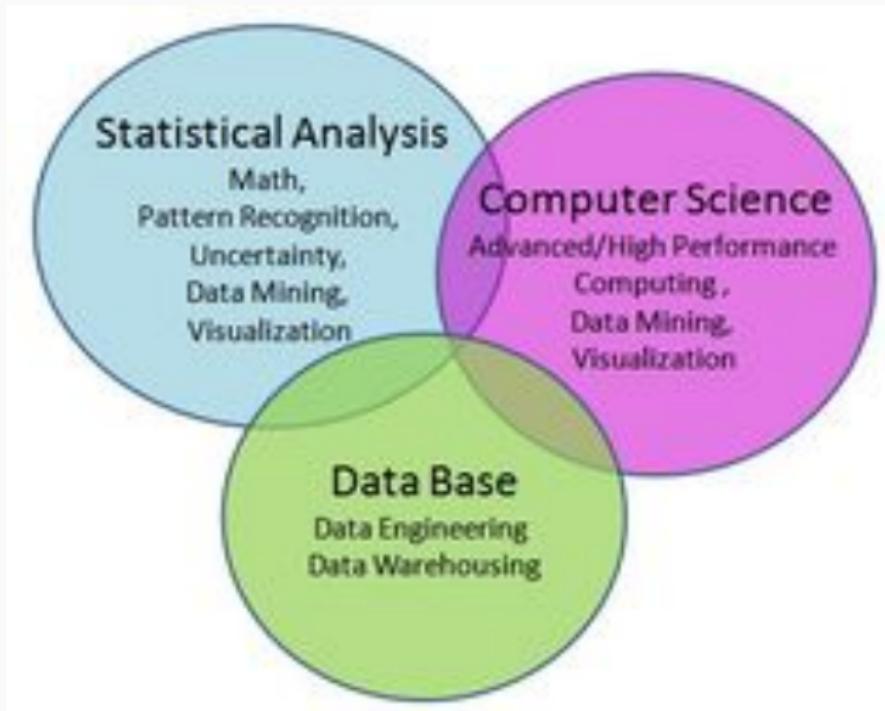


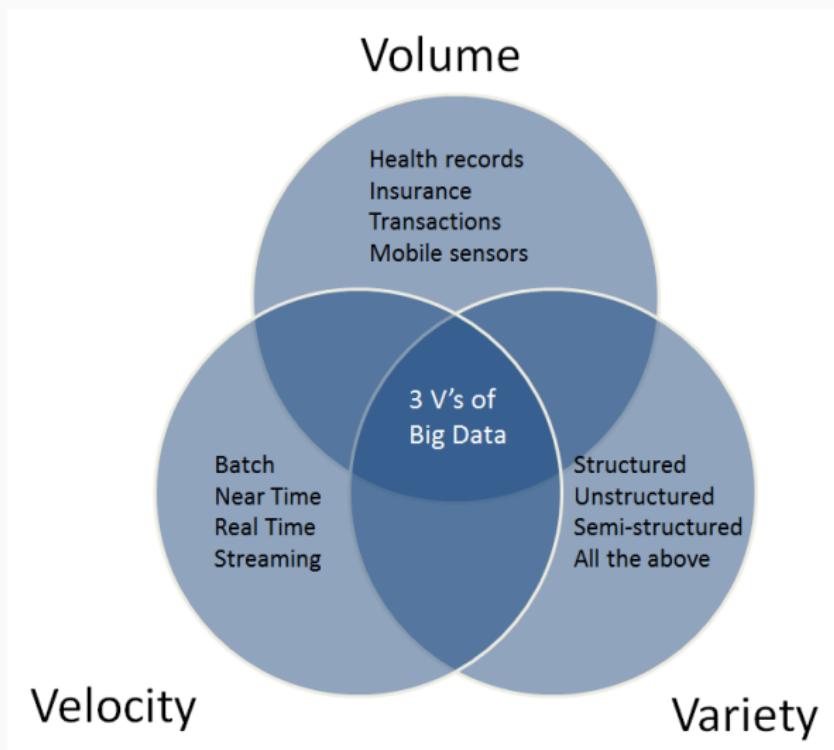
Imagen tomada de

<https://hackernoon.com/memorizing-is-not-learning-6-tricks-to-prevent-overfitting-in-machine-learning-820b091dc42>

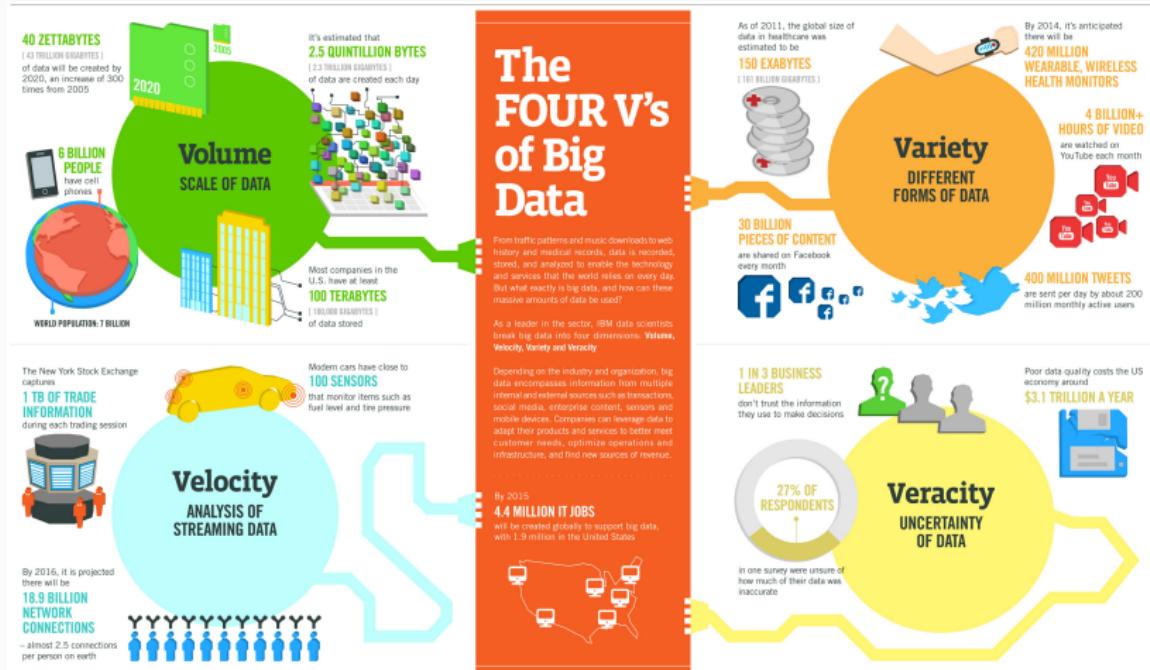
DEFINICIÓN



CARACTERÍSTICAS: 3Vs



CARACTERÍSTICAS: 4Vs



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MPTPEC, GAI



QUIZÁS MUCHAS Vs

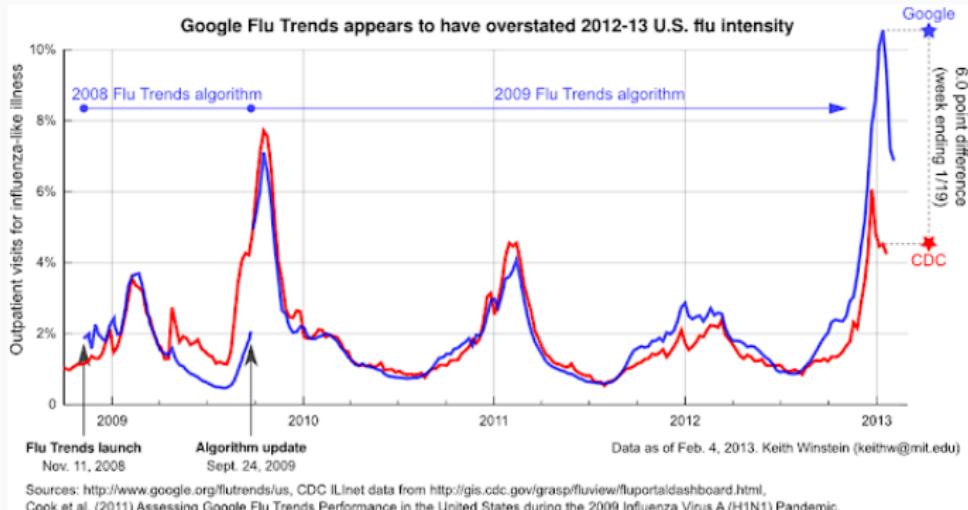


- Concepciones: cuestión simplemente implementacional, lo mismo pero con más datos
- Muchos retos
 - Algoritmos, estructuras de datos y modelos escalables
 - Representaciones eficientes
 - Dimensionalidad, diversidad y variabilidad extremas
 - Respuesta en línea o tiempo real
 - Paralelización
 - Modelado
 - Búsqueda, organización y exploración

EL PRINCIPIO DE BONFERRONI

Calcula el número esperado de ocurrencias de un evento bajo la suposición que es aleatorio. Si el número es mucho mayor al de las ocurrencias reales, entonces las conclusiones que puedas sacar a partir de estos eventos probablemente sean falsas.

EL CASO DE GOOGLE FLU TRENDS¹



¹Ginsberg et al. Detecting influenza epidemics using search engine query data, Nature, 2009.

- Vigilancia
- Sesgos
- Baja exploración