

**Una herramienta útil, un concepto defectuoso.**



# Y PARA MUESTRA

- **Múltiples V's** : velocidad, variación, volumen, veracidad, valor, .....
- **Mckinsey**: Conjunto de datos que está más allá de las posibilidades de las herramientas de software habituales.
- **The Economist**: La sociedad dispone de más información que en ningún período precedente de la historia. Ese aumento de información nos permite realizar cosas que no podíamos realizar con anterioridad.
- **Wikipedia**: Es el término para un conjunto de datos tan complejo y grande; que son difíciles de procesar siguiendo los procedimientos y con las herramientas habituales.
- **Adam Jacobs**: Tamaño de los datos que nos fuerza a mirar más allá de las metodologías que se emplean habitualmente.
- **Dan Law**: Cualquier conjunto de datos que tienen un valor potencial que no puede extraerse con las técnicas y procedimientos ordinarios.

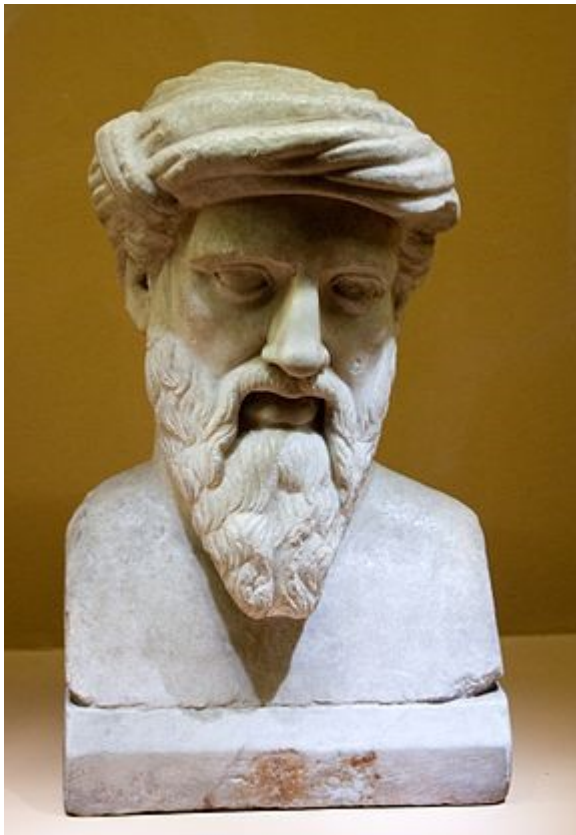


# Y DE QUE DATOS HABLAMOS

- **Datos estructurados:** Tablas, datos relacionales,...
- **Semi estructurados:** Datos en forma híbrida – documentos con tablas –
- **Datos en streaming:** Que se mueven a gran velocidad por redes.
- **Datos geoespaciales:** Incluyen posicionamientos en el espacio: regiones, puntos, caminos, formas...
- **Datos no estructurados:** Texto, fotos, video, audio...
- **Metadata:** datos estructurados entorno a los propios datos.
- **Datos temporales:** incluyen tendencias, actividades temporales.
- **Y muchos mas.....**



# Todo empezó hace casi 2500 años



Pitagoras, en la isla de Samos, en la actual Grecia, pensó que el universo se forma a través de **modelos matemáticos.**



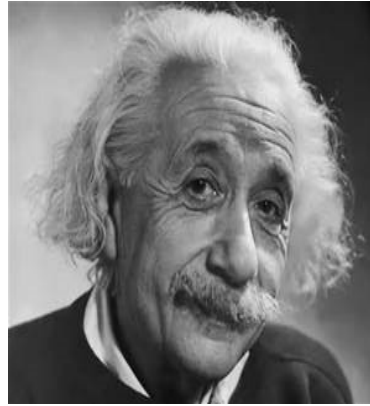
# Y ha seguido hasta hoy



DESCARTES



NEWTON



EINSTEIN



TURING



# Extrayendo conocimiento de los datos (1)

## Con la colaboración de los ordenadores

- Definiendo los conceptos de conocimiento.
- Transformando los datos estructurados en representaciones útiles.
- Clasificando y extrayendo contenido de los datos no estructurados.
- Mediante el análisis de los datos que genere conocimiento relevante que mejore la base decisional.



# Extrayendo conocimiento de los datos (2)

**Y haciendo, entre otras cosas..**

- Dividiendo el problema en trozos y procesando en paralelo.
- Construyendo índices de conocimiento y búsqueda que puedan activarse rápidamente.
- Usando adecuada combinación tecnología y software.
- Una mezcla de todo ello.



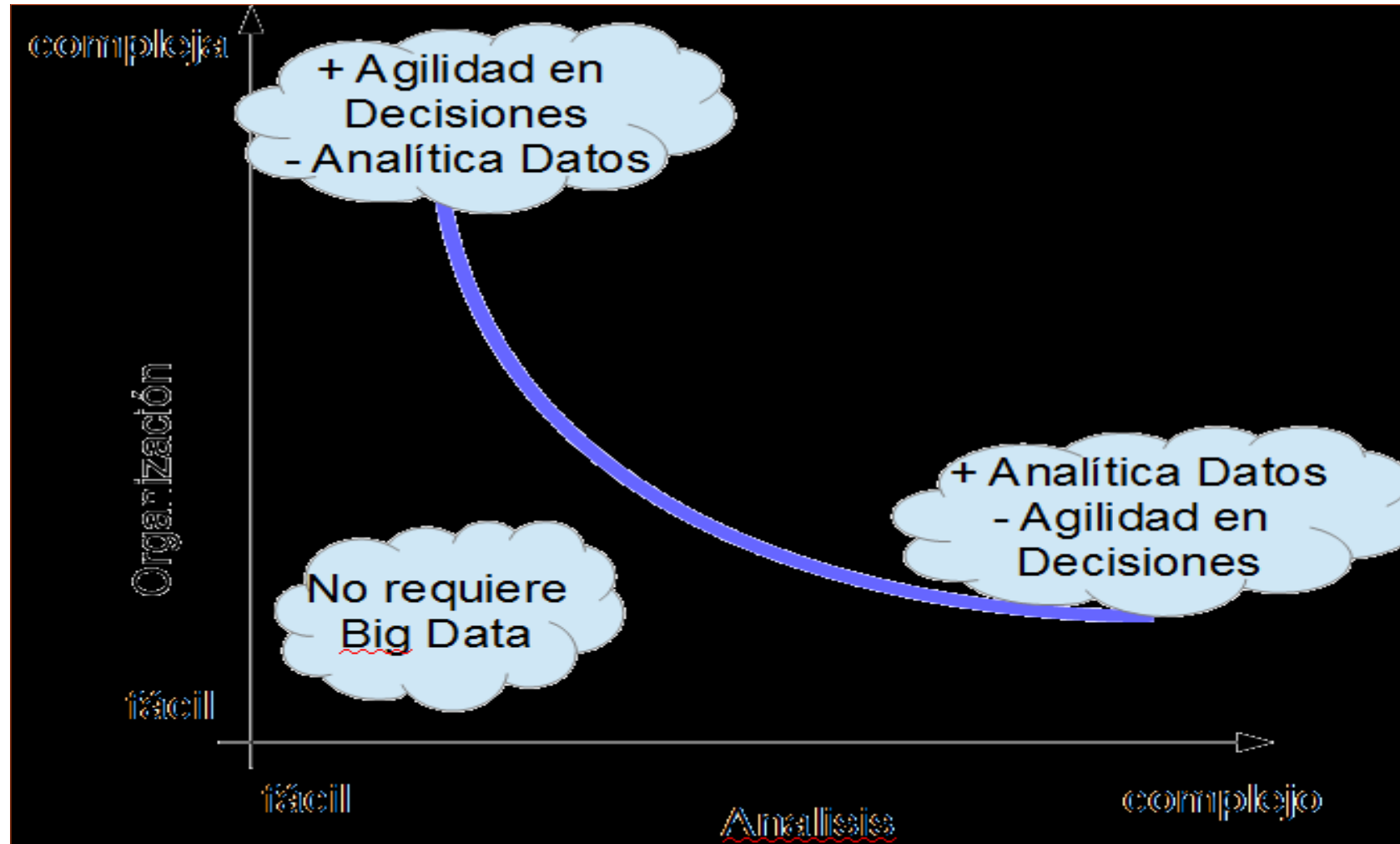
# Moviendo la información

- Usar y estimar parámetros con **aparato estadístico** para:
  - Describir los datos
  - Explorar correlaciones
  - Descubrir modelos
  - Predecir resultados
- Necesitamos prestar mucha **atención** a los siguientes aspectos:
  - Corregir supuestos
  - Cuidado con las variables

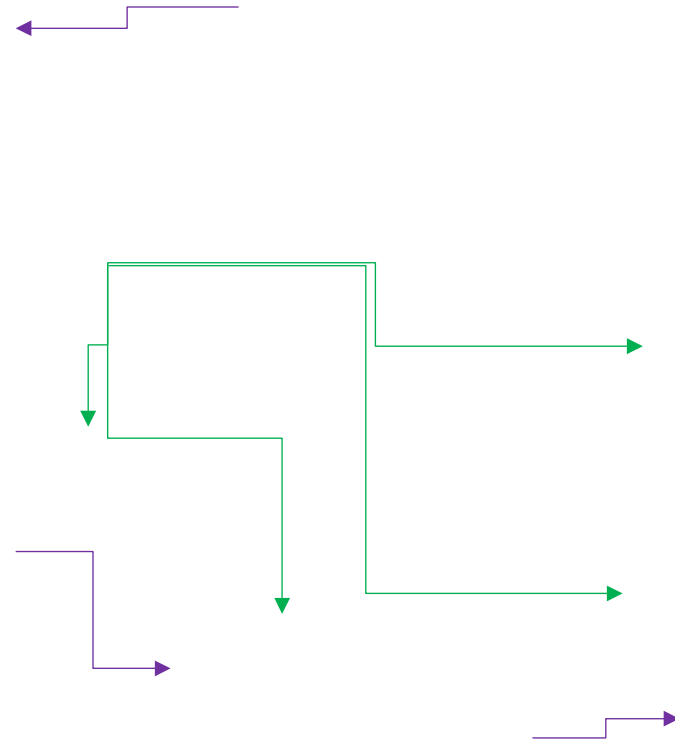




# Tomando decisiones – previas -



# Implementando soluciones en datos



Posibles resultados:

\*

- ANOVA
- MACHINE LEARNING
- REGRESIÓN LINEAL
- ARBOLES DE DECISION
- MODELOS PREDICTIVOS
- MODELOS PRESCRIPTIVOS



# Actores y actividades implicadas



# A modo de conclusión

- Momento clave de un proceso histórico.
- Existencia de gran cantidad de datos y posibilidad de trabajar con ellos
- Mayor importancia cuantitativa de los datos no estructurados
- Oferta de interesantes soluciones tecnológicas
- Experiencias contrastadas
- Impacto en las organizaciones
- Crecimiento en la productividad a partir de una base del 3%
- Costes de no utilizar la gestión basada en datos



# Qué escenario pronostico

- Disminución de costes de productos.
- Flexibilidad de las plataformas
- Soluciones más amigables
- Crecimiento de la oferta de expertos en análisis de datos
- Organizaciones más sensibles hacia el empleo de los datos para la gestión
- Nuevos modelos de negocio
- Un cambio que podemos definirlo como el BIG DATA 2.0

## **LA SUSTITUCIÓN DEL BIG DATA POR EL BIG ANALITICS**

