



UNIVERSIDAD DE CIENCIAS  
EMPRESARIALES Y SOCIALES

Jornadas de Investigación de la Facultad de Ciencias de la Comunicación 2016

VII° Encuentro de Investigación en Periodismo

IV° Foro de Investigadores en Diseño, Publicidad, Comunicación Social  
y Relaciones Públicas

***“Los Procesos y los Productos de Investigación en Comunicación”***

Big Data y Comunicación

Mg. M. Micaela Bazzano

*Noviembre 2016*

# BIG

# DATA

# STORAGE

## ANALYTICS

### PARALLEL

## TECHNOLOGIES

### SOFTWARE

### BUSINESS

### USE

### ONE

### SINCE

### HUNDREDS

### TIME

### NETWORKS

## INFORMATION

## COMPLEX

### DIFFICULTY

### TARGET

### SENSOR

### ARCHIVES

### ABILITY

### SETS

### EXAMPLES

### APPLIED

### ELAPSED

### CURRENT

### THOUGHT

### AMOUNT

### RECORDING

### PERFORMANCE

### RELATED

### ALSO

### GARTNER

### MASSIVELY

### BIOLOGICAL

### USING

### THREE

### WORKING

### RELATED

### PERFORMANCE

### DISK

### RELATIONAL

### BACKUP

### RECENTLY

### COMPLETED

### TO

### CONVERT

### RESEARCH

### TERABYTES

### MPP

### DATABASES

### RESEARCH

### FOR

### CAPACITY

### PRACTITIONERS

### FOR

### RECORDS

### CONNECTING

### COMPUTING

### TOOLS

### SET

### GENOMICS

### DETAILED

### SHARED

### CONSULT

### SIGNIFICANT

### INCLUDE

### SYSTEMS

### DEFINITION

### SEARCH

### OPPORTUNITIES

### WITHIN

### PROCESS

### RESEARCHING

### MOVING

### REGULATING

### INDUSTRIES

### SAN

### SOLID

### WIRELESS

### RADIO FREQUENCY

### COMPLEXITY

### TOLERABLE

### SIZE

### QUALITY

### NEEDED

### PROCESSING

### MEMBER

### LOGS

### DISTRIBUTED

### CAPTURE

### MANAGE

### GROW

### MAY

### SOCIAL

### MANAGEMENT

### LARGE

### EVERY

### LARGER

### STORE

### CASE

# ¿Qué es “Big data”?

- ▶ El concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.
- ▶ Sin embargo, Big Data no se refiere a alguna cantidad en específico, ya que es usualmente utilizado cuando se habla en términos de petabytes y exabytes de datos.
- ▶ Entonces ¿Cuánto es demasiada información de manera que sea elegible para ser procesada y analizada utilizando Big Data? Analicemos primeramente en términos de bytes:
  - ▶ *Gigabyte* =  $10^9 = 1,000,000,000$
  - ▶ *Terabyte* =  $10^{12} = 1,000,000,000,000$
  - ▶ *Petabyte* =  $10^{15} = 1,000,000,000,000,000$
  - ▶ *Exabyte* =  $10^{18} = 1,000,000,000,000,000,000$

# ¿Qué es “Big data”?

- Volumen
- Variedad
  - dispositivos móviles
  - Audio
  - Video
  - sistemas GPS
  - sensores digitales en equipos industriales
  - medidores eléctricos
- Velocidad → información correcta en el momento preciso



# ¿De dónde surgen todos esos datos?

## ➔ *Acumulación masiva*

Sector público

Censos  
Registros médicos  
Impuestos

Sector privado

Clientes  
Proveedores  
Transacciones

Actividades online

GPS  
Tweets (12 teras al día)  
Facebook (100 petas)

# Tipos de datos en Big Data

## Big Data Types

### Web and Social Media

- Clickstream Data
- Twitter Feeds
- Facebook Postings
- Web Content

### Machine-to-Machine

- Utility Smart Meter Readings
- RFID Readings
- Oil Rig Sensor Readings
- GPS Signals

### Big Transaction Data

- Healthcare Claims
- Telecommunications Call Detail Records
- Utility Billing Records

### Biometrics

- Facial Recognition
- Genetics

### Human Generated

- Call Center Voice Recordings
- Email
- Electronic Medical Records

# WHERE IS DATA COMING FROM?

Twitter users send out

**277,000**  
tweets

Google processes more than

**2 million**  
search queries

Facebook processes almost

**350 GB** of data

**72 hours**

of new video are uploaded to YouTube

**EVERY MINUTE...**



Individuals and organizations launch

**571**

new websites

More than

**100 million**

new emails are generated

Walmart processes almost

**17,000**

transactions

Sprint processes more than

**250,000**

phone calls



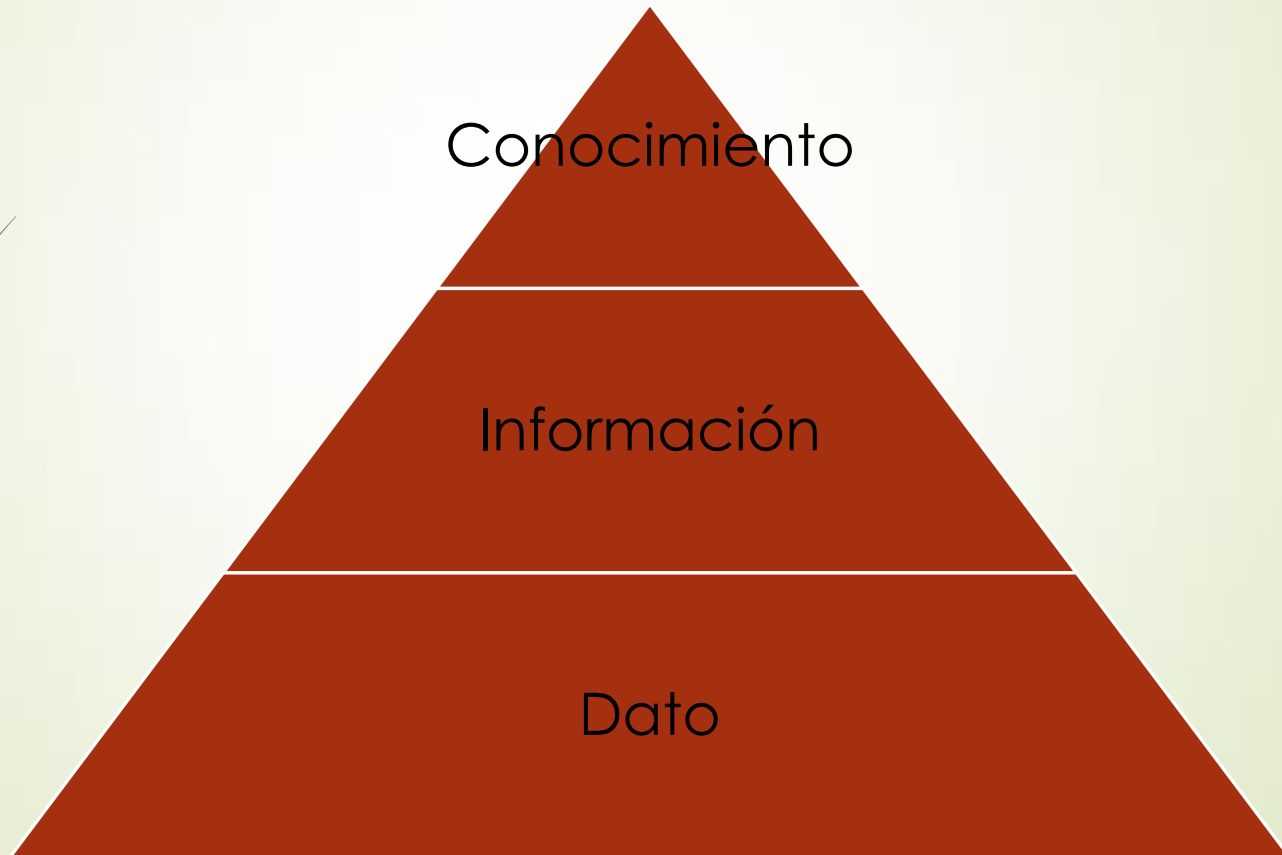
# Tipos de datos en Big Data

- ▶ **Web and Social Media:** Incluye contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, etc, blogs.
- ▶ **Machine-to-Machine (M2M):** M2M se refiere a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.) los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.
- ▶ **Big Transaction Data:** Incluye registros de facturación, en telecomunicaciones registros detallados de las llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.
- ▶ **Biometrics:** Información biométrica en la que se incluye huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación.
- ▶ **Human Generated:** Las personas generamos diversas cantidades de datos como la información que guarda un call center al establecer una llamada telefónica, notas de voz, correos electrónicos, documentos electrónicos, estudios médicos, etc.






# Descubrimiento de Conocimiento en Bases de Datos (KDD)





# Metas del KDD

- Procesar automáticamente grandes cantidades de datos crudos.
  - Identificar los patrones más significativos y relevantes.
  - Presentarlos como conocimiento apropiado para satisfacer las metas del usuario.
- 



# Proceso del KDD

- Diseñar el esquema de un almacén de datos (**Data Warehouse**): que consiga unificar de manera operativa toda la información recogida.
- Implantación del almacén de datos: que permita la navegación y visualización previa de sus datos, para discernir qué aspectos puede interesar que sean estudiados. Esta es la etapa que puede llegar a consumir el mayor tiempo.
- Selección, limpieza y transformación de los datos que se van a analizar: la selección incluye tanto una criba o fusión horizontal (filas) como vertical (atributos).
- Seleccionar y aplicar el método de **minería de datos** apropiado: esto incluye la selección de la tarea de descubrimiento a realizar, por ejemplo, clasificación, agrupamiento o clustering, regresión, etc.
- Evaluación, interpretación, transformación y representación de los patrones extraídos.

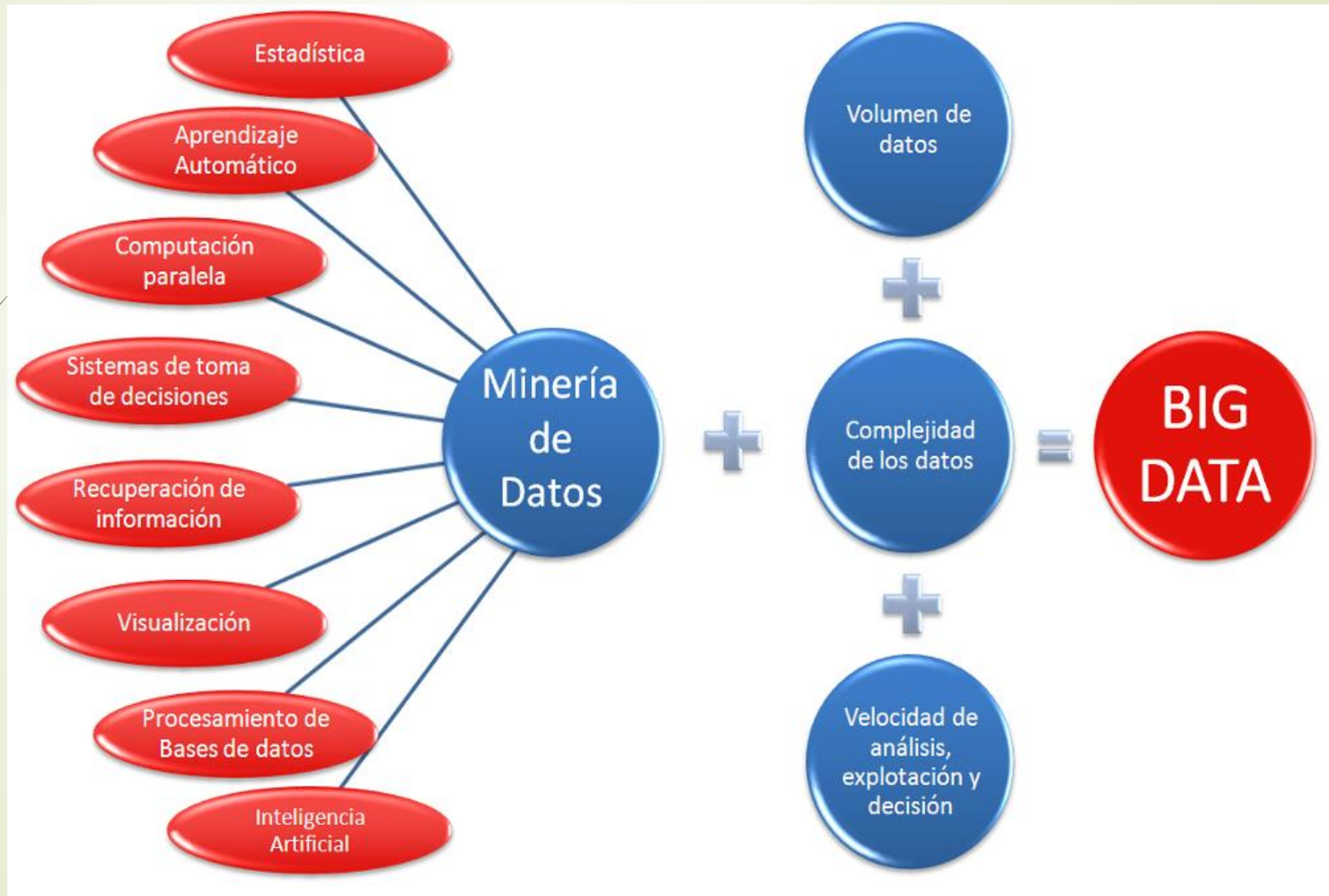


# Data mining o Minería de datos

Reúne las ventajas de varias áreas:

- Estadística
- Inteligencia Artificial
- Computación Gráfica
- Bases de Datos
- Procesamiento Masivo

# ¿Cómo se trabaja con Big data?



# Ejemplo

- ▶ Una franquicia de hoteles a nivel nacional que utiliza aplicaciones de BI para llevar un registro estadístico del porcentaje promedio de ocupación del hotel, así como los días promedio de estancia de cada huésped, considerando las diferencias entre temporadas.
- ▶ Con esta información ellos pueden:
  - ▶ Calcular la rentabilidad de cada hotel en cada temporada del año.
  - ▶ Determinar quién es su segmento de mercado.
  - ▶ Calcular la participación de mercado de la franquicia y de cada hotel.
  - ▶ Identificar oportunidades y amenazas.



## Ejemplos de aplicaciones de Big Data en el sector Retail







# Ejemplo en el ámbito de la salud

- ▶ El *Instituto de Tecnología de la Universidad de Ontario (UOIT)* junto con el Hospital de Toronto utilizan una plataforma de big data para análisis en tiempo real de IBM (*IBM InfoSphere Streams*), la cual permite **monitorear bebés prematuros** en las salas de neonatología para determinar cualquier cambio en la presión arterial, temperatura, alteraciones en los registros del electrocardiograma y electroencefalograma, etc., y así **detectar hasta 24 horas antes** aquellas condiciones que puedan ser una amenaza en la vida de los recién nacidos.




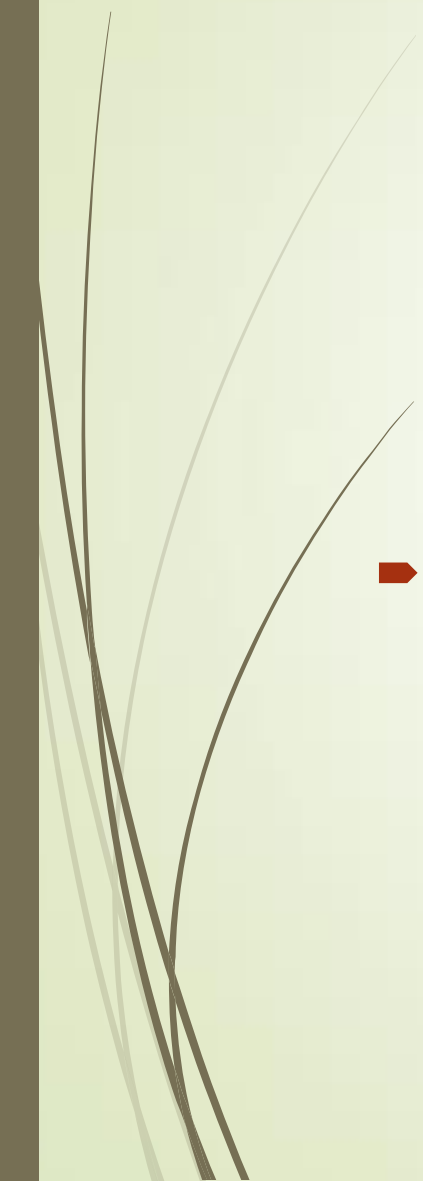
# Ejemplo en el ámbito de la salud

- ▶ La esclerosis múltiple es una enfermedad del sistema nervioso que afecta al cerebro y la médula espinal. La comunidad de investigación biomédica y la *Universidad del Estado de Nueva York (SUNY)* están aplicando análisis con big data para contribuir en la progresión de la investigación, diagnóstico, tratamiento, y quizás hasta la posible cura de la esclerosis múltiple.



# Ejemplo en el ámbito de servicios

- ▶ Los laboratorios *Pacific Northwest National Labs(PNNL)* utilizan de igual manera IBM InfoSphere Streams para analizar eventos de medidores de su red eléctrica y **en tiempo real verificar aquellas excepciones o fallas en los componentes de la red**, logrando comunicar casi de manera inmediata a los consumidores sobre el problema para ayudarlos en administrar su consumo de energía eléctrica.

- 
- 
- Analizar los datos para dos propósitos importantes:
    - ser capaces de detectar y responder a los acontecimientos actuales de una manera **oportuna**
    - y para poder utilizar las predicciones del **aprendizaje histórico**.
  
  - Esta situación requiere del análisis tanto de
    - **datos en movimiento** (datos actuales)
    - **datos en reposo** (datos históricos), que son representados a diferentes y enormes volúmenes, variedades y velocidades.

Una técnica de Google trabaja en el diagnóstico de un procesador sobrecalentado en el centro de datos en The Dalles, Oregon, EEUU.

