

# UEA Datos a Gran Escala

## Introducción a los Datos a Gran Escala



**Dr. Pedro Pablo González Pérez**

e-mail: [pgonzalez@correo.cua.uam.mx](mailto:pgonzalez@correo.cua.uam.mx)

<http://dcni.cua.uam.mx/division/usuario?p=31#>

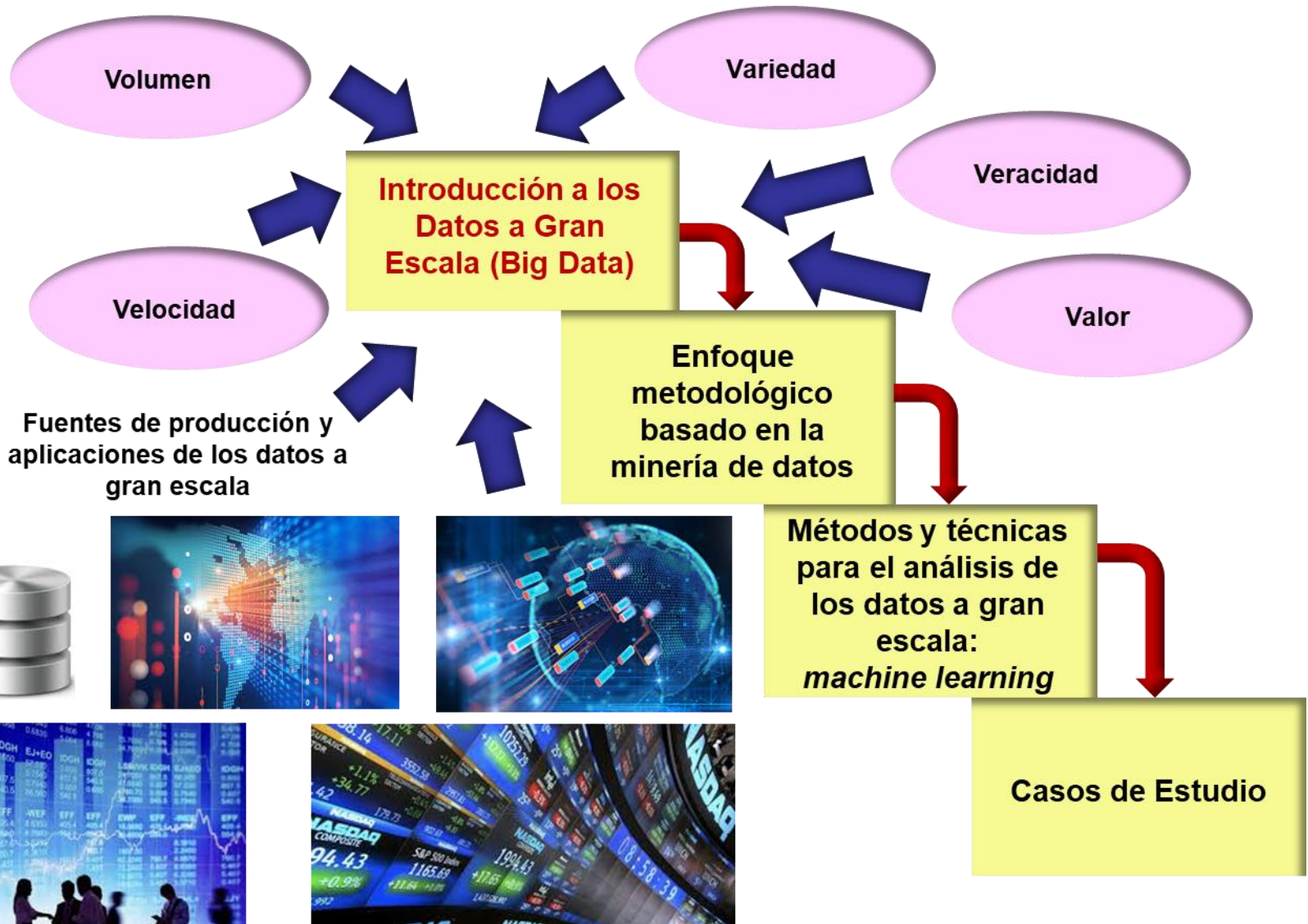
**Departamento de Matemáticas Aplicadas y Sistemas**



**UNIVERSIDAD  
AUTÓNOMA  
METROPOLITANA**  
Unidad Cuajimalpa

# UEA Datos a Gran Escala

## Introducción a los Datos a Gran Escala



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): aproximaciones al concepto de datos a gran escala

### ➤ ¿A qué nos referimos con Datos a Gran Escala?

- ❑ Resulta difícil encontrar una definición ampliamente aceptada del término **big data**, ya que este último se ha utilizado de manera ubicua y, en la gran mayoría de los casos, las aproximaciones proporcionadas a este término dependen fuertemente del dominio en el que se ha utilizado esta tecnología.
- ❑ Sin embargo, cuando nos referimos a **big data** seguramente estamos describiendo problemas que surgen con relación a:
  - ❖ **Datos cuyo volumen y complejidad requieren métodos más sofisticados de almacenamiento, recuperación, interacción, preparación, análisis-inferencia y presentación.**



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): aproximaciones al concepto de datos a gran escala

### ➤ ¿A qué nos referimos con Datos a Gran Escala?

- ❖ Sistemas de software cuya funcionalidad ya resulta inadecuada para lidiar con el enorme volumen y la gran complejidad de los datos que deben procesar.
- ❖ Grandes volúmenes de datos que implican tanto datos estructurados como no estructurados, lo que hace que su tratamiento sea mucho más complejo.
- ❖ La aplicación de un potente procesamiento computacional a conjuntos de datos altamente masivos y complejos.





# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): aproximaciones al concepto de datos a gran escala

### ➤ ¿A qué nos referimos con Datos a Gran Escala?

- ❑ El término **Big data** se pone de manifiesto siempre que nuestros datos resultan demasiado voluminosos para ser procesados por una aplicación o sistema de cómputo.
- ❑ El concepto **Big data** también se manifiesta cuando los sistemas de gestión o servidores de bases de datos no son capaces de proporcionar en tiempo razonable los datos requeridos, debido a problemas con la carga, búsqueda, selección, etc.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): aproximaciones al concepto de datos a gran escala

### ➤ Aproximaciones al concepto de Datos a Gran Escala

- ❑ “...*data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time*” [Tolk, 2015; Jacobs, 2009].
- ❑ Oracle: “*Big data is the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data.*” [Tolk, 2015; Ward and Baker, 2013].
- ❑ Intel: “*Big data opportunities emerge in organizations generating a median of 300 terabytes of data a week. The most common forms of data analyzed in this way are business transactions stored in relational databases, followed by documents, e-mail, sensor data, blogs, and social media.*” [Tolk, 2015; Ward and Baker, 2013].



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): aproximaciones al concepto de datos a gran escala

### ➤ Aproximaciones al concepto de Datos a Gran Escala

- ❑ Microsoft: “*Big data is the term increasingly used to describe the process of applying serious computing power — the latest in machine learning and artificial intelligence — to seriously massive and often highly complex sets of information.*” [Tolk, 2015; Ward and Baker, 2013].
- ❑ National Institute of Standards and Technology (NIST): “*Big data is data which exceed the capacity or capability of current or conventional methods and systems. In other words, the notion of big is relative to the current standard of computation, such is observed in.*” [Tolk, 2015; Jacobs, 2009].



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Principales características de los Datos a Gran Escala

De las anteriores definiciones emergen las principales características de los **datos a gran escala** (*big data*):

**Volumen**

**Velocidad**

**Variedad**

**Veracidad**

**Valor**



Comúnmente, los datos masivos contienen incertidumbre, debido a la naturaleza de las diferentes fuentes de las cuales proceden.





# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Volumen.

- ❑ Se refiere a los grandes volúmenes de datos que se generan, recopilan y, en muchas ocasiones, se analizan constantemente.
- ❑ El volumen es precisamente la característica que más asociamos a los datos masivos; es imposible no pensar en el volumen cuando nos referimos a *big data*.
- ❑ Volumen medido en *gigabytes*, *terabytes* y *petabytes*.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Velocidad.

- ❑ Se refiere a la rapidez con la cual los datos son generados, recopilados y procesados. Pensemos en la gran velocidad con la que se generan datos en aplicaciones tales como los motores de búsqueda, el mercado de valores, las plataformas *e-commerce*, y en las redes sociales, por mencionar algunos ejemplos.
- ❑ En una gran variedad de dominios el tiempo de respuesta se convierte en una variable esencial. Este es el caso de los sistemas de cómputo que deben ofrecer respuesta en tiempo real.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Variedad.

- ❑ Se refiere a la no homogeneidad o diversidad de los datos, ya que éstos provienen de orígenes o fuentes muy diferentes, lo que implica que los datos sean de tipos muy diferentes, tales como numéricos, booleanos, categóricos, nominales, texto estructurado, texto no estructurado, presentaciones en algún visualizador particular, imágenes, videos, etc.
- ❑ Datos no estructurados:
  - Archivos de texto tales como emails, PDF, blogs, etc.
  - Fotos e imágenes.
  - Videos.
  - Sitios Web.
  - Presentaciones (por ejemplo, en *Power Point*).



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Veracidad.

- ❑ Se refiere a la necesidad de contender con la incertidumbre contenida en los datos, debido principalmente a la gran variedad que caracteriza a los mismos, derivado de las diferentes fuentes que los generan.
- ❑ El problema de la veracidad en los datos a gran escala, se presenta comúnmente en los datos de texto no estructurado, debido a la libertad que caracteriza su creación.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Valor.

- ❑ El término valor se refiere a la importancia y el significado que los datos a gran escala pueden proporcionar a empresas, compañías e instituciones en la toma de decisiones, lo que conduce a éstas a ser mucho más rentables y exitosas.





# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Big data: volumen, velocidad, variedad, veracidad y valor

- ❑ Por lo tanto, los métodos y tecnologías de los **datos a gran escala** (*big data*) nos permiten procesar y analizar datos estructurados y no estructurados –fuertemente caracterizados por su volumen, velocidad, variedad, comúnmente, gran incertidumbre en su veracidad, y valor– y producir nuevos datos derivados, los cuales contienen información de gran valor para la toma de decisiones en el dominio en cuestión.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): características de los datos a gran escala

### ➤ Big data: volumen, velocidad, variedad, veracidad y valor

- ❑ La toma de decisiones se base en los resultados de los modelos aplicados durante el análisis del volumen de datos. Comúnmente, modelos de correlación, predicción, clasificación, agrupamiento, reconocimiento de patrones, entre otros.
- ❑ Entre los métodos, técnicas y algoritmos comúnmente utilizados para el análisis de los datos a gran escala se encuentran las siguientes familias:
  - ❖ **Métodos estadísticos.**
  - ❖ **Aprendizaje automatizado (*Machine learning*).**
  - ❖ **Aprendizaje profundo (*Deep learning*).**



# UEA Datos a Gran Escala

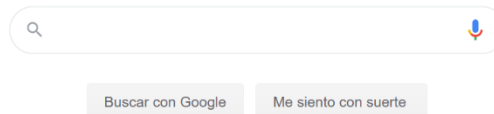
## Introducción a los datos a gran escala (*big data*): fuentes de producción y uso

### ➤ Fuentes de producción y uso de los datos a gran escala



Connected Wearable Devices  
in Healthcare: Wearables in  
Medical, Wellness, and Fitness  
Markets 2019 – 2024

February 2019



Ofrecido por Google en: [English](#) [Español \(Latinoamérica\)](#)

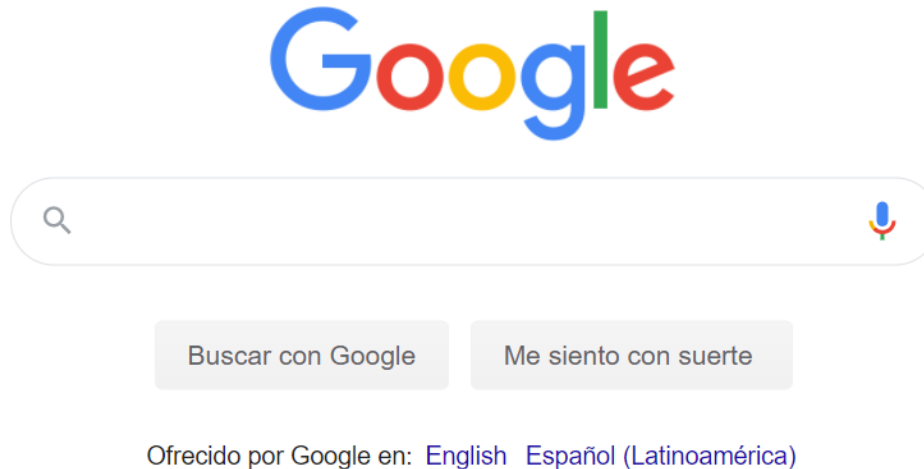


# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción y uso de los datos a gran escala

#### ❑ Motores de búsqueda.



- ❑ No sólo a través del motor de búsqueda #1 en el mundo, **Google**, sino también a través del motor de búsqueda local de sitios especializados, tales como plataformas *e-commerce*, repositorios de documentación, repositorios de recursos multimedia, etc.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción y uso de los datos a gran escala

- ❑ **En la práctica deportiva.** En deportes tales como el tenis, basquetbol, golf, béisbol, boxeo, esgrima, y otros muchos más, donde, **tanto los jugadores como los entrenadores pueden ser retroalimentados sobre el desempeño de los primeros**, a partir de una **gran cantidad de sensores y cámaras incorporados**. Estos sensores generan de forma continua una inmensa cantidad de datos.
- ❑ Por ejemplo, en el tenis, se utilizan **sistemas de sensores capaces de registrar el rendimiento de un jugador**, **proporcionando estadísticas en tiempo real y análisis completos de los partidos**.



# UEA Datos a Gran Escala

## ➤ Fuentes de producción y uso de los datos a gran escala

### En la práctica deportiva



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción y uso de los datos a gran escala

❑ **Monitoreo de la salud y bienestar de una persona, sobre la marcha.** Existen una amplia gama de **dispositivos portátiles (por ejemplo, del tipo *Smart band*)** que las personas pueden llevar consigo mismas, en forma de bandas, pulseras, relojes, etc., para el monitoreo de aspectos tales como:

- **Nivel de glucosa.**
- **Ritmo cardíaco.**
- **Presión arterial.**
- **Número de pasos caminados en un intervalo de tiempo.**
- **Calorías quemadas.**
- **Calidad del sueño durante la noche.**
- **...**

# UEA Datos a Gran Escala

## ➤ Fuentes de producción y uso de los datos a gran escala

**Monitoreo de la salud y bienestar de una persona, sobre la marcha**



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Salud más inteligente (*Smarter health*)**. Monitoreo y transmisión en tiempo real de datos personales generados durante la actividad deportiva y de fitness, así como relacionados con el estado de salud y bienestar.
  - Monitoreo de pacientes a través de aplicaciones basadas en sensores instaladas en el teléfono celular.

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

Salud más inteligente  
(*Smarter health*)



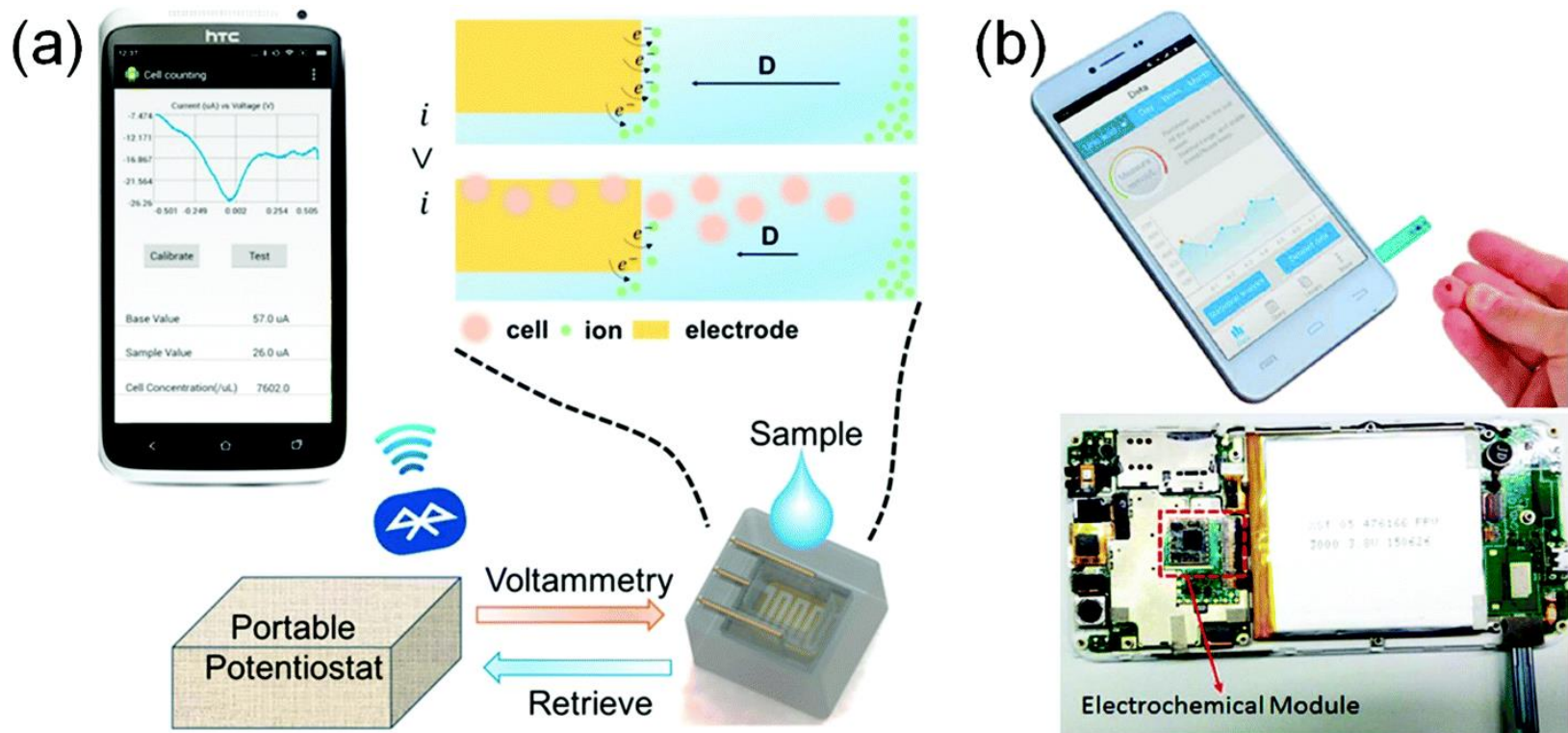
<https://www.compassintelligence.com>



# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

Salud más inteligente (*Smarter health*)



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

❑ **Hogar inteligente – Internet de las cosas.** El uso de Internet para la interconexión de sensores, actuadores, dispositivos y equipos electrodomésticos, de forma tal que el comportamiento resultante se traduzca en un uso mucho más eficiente de los últimos, a la vez que proporciona mayores beneficios a la persona.

- Regulación automática de la temperatura de las habitaciones, en dependencia de si están siendo usadas, a partir de hábitos de las personas aprendidos por el sistema de control, horarios en que comúnmente se usan, etc.
- Monitoreo y control a través de Internet de dispositivos y equipos electrodomésticos.
- Smart TV.
- ...

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Hogar inteligente – Internet de las cosas



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Automóviles inteligentes.** Gran soporte a la conducción basado en tableros muy sofisticados, cámaras y sensores distribuidos a través del automóvil, recepción, transmisión e integración de información en tiempo real, lo cual permite al conductor:
  - Recibir alertas al ejecutar operaciones de cambios de carril.
  - Recibir alertas sobre cuando el automóvil se encuentra muy cercano a algún tipo de obstáculo o a otro automóvil.
  - Facilidad en la operación de estacionamiento.
  - Operación de estacionamiento autónoma por parte del propio automóvil.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

#### ❑ Automóviles inteligentes.

- Sincronización con la información del tráfico en tiempo real y proponer al conductor alternativas de redireccionamiento hacia mejores rutas, evitando rutas en las que ha ocurrido algún incidente o accidente.
- Retroalimentación sobre el desempeño del conductor durante la conducción, errores que ha estado cometiendo, etc.
- ...



# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Automóviles inteligentes



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

❑ **Plataformas e-commerce.** Las plataformas *e-commerce* se refieren a una extensa gama de aplicaciones de comercio electrónico disponibles en Internet, y a las cuales el usuario comúnmente accede a través de una URL o de una App en el dispositivo celular. Una plataforma *e-commerce* es un punto de venta en línea dedicado al comercio de productos y servicios:

- Plataformas dedicadas al comercio electrónico de productos.
- Plataformas dedicadas al comercio electrónico de servicios.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Plataformas e-commerce.** Las plataformas *e-commerce* producen inmensos volúmenes de datos, derivados de toda la interacción producida por el usuario, a través de la navegación, las búsquedas, consultas y compras de los productos o servicios ofrecidos.
- ❑ Los datos generados por estas plataformas son un valioso insumo, que además de contener información personal y financiera de los usuarios, también reflejan las preferencias de éstos hacia determinados productos en particular, la capacidad de consumo, la frecuencia en la que se efectúa el consumo, el nivel socio-económico, etc.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Plataformas e-commerce.** El gran volumen de datos que producen de forma continua las plataformas *e-commerce*, constituyen un valioso insumo para los modelos de minería de datos que analizan dicha información y que, como resultado, producen inferencias, predicciones y estimaciones de gran utilidad para la toma de decisiones en la venta de productos en línea.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

#### ❑ Principales plataformas e-commerce en México:

- Amazon.
- Mercado Libre.
- Walmart.
- Liverpool.
- Coppel.
- Sam's Club.
- Home Depot.
- Sears.

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Plataformas e-commerce

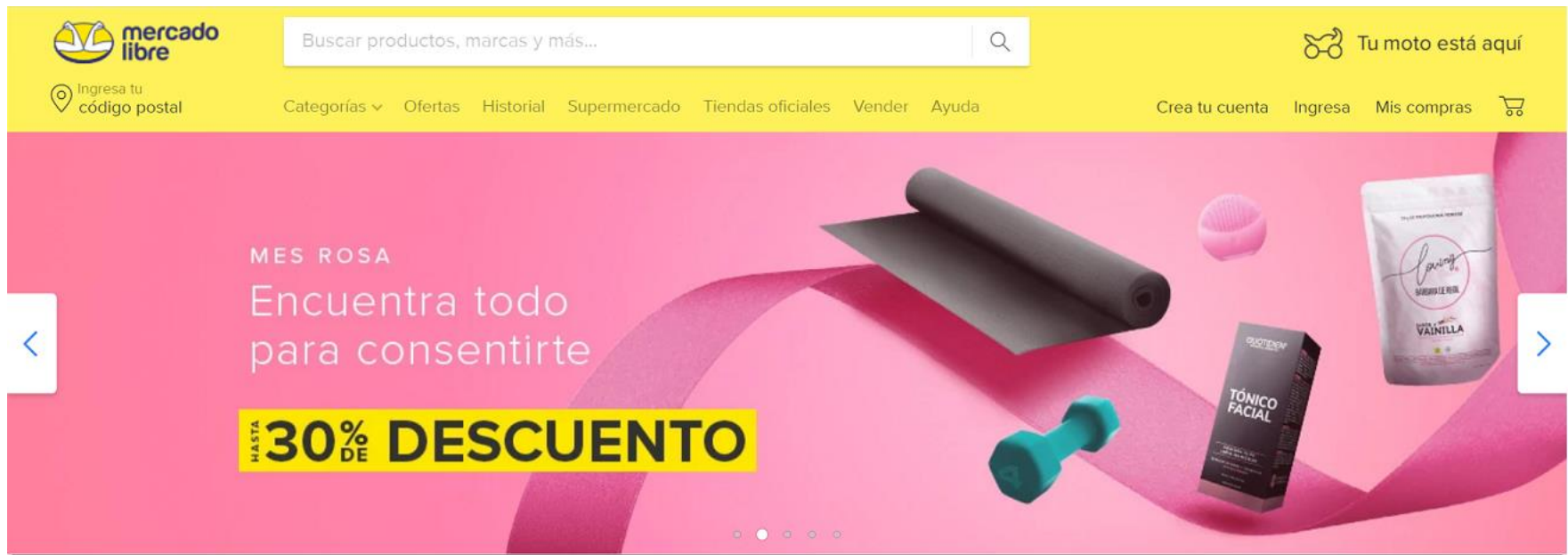




# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

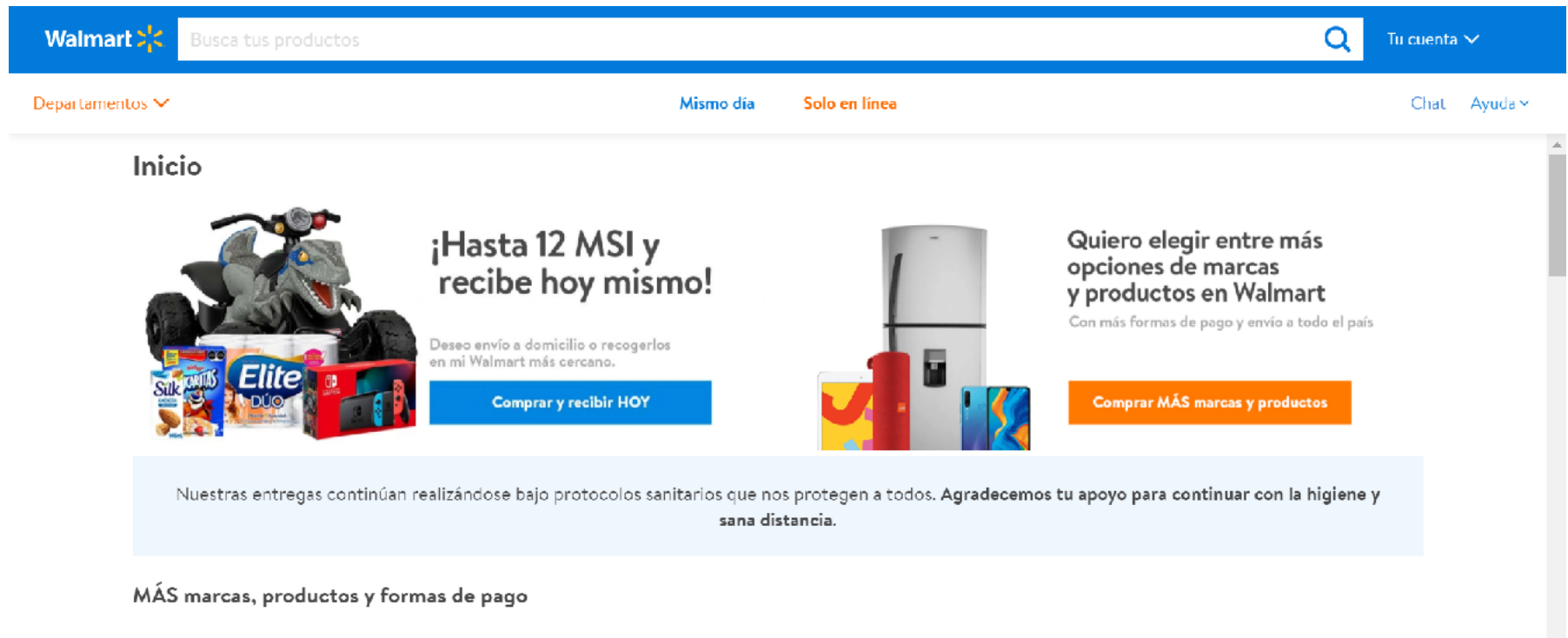
### Plataformas e-commerce






# UEA Datos a Gran Escala



## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Plataformas e-commerce





The image is a screenshot of the Walmart Mexico homepage. At the top, there is a blue header bar containing the Walmart logo, a search bar with the placeholder text "Busca tus productos", a magnifying glass icon, and a link to "Tu cuenta". Below the header, there are navigation links for "Departamentos", "Mismo día", "Solo en línea", "Chat", and "Ayuda". The main content area features a "Inicio" section with two prominent promotions. The first promotion on the left shows a collection of products including a dinosaur toy, Silk Kahlúa, Elite Duo, and a Nintendo Switch, with the headline "¡Hasta 12 MSI y recibe hoy mismo!" and a blue "Comprar y recibir HOY" button. The second promotion on the right shows a refrigerator and a smartphone, with the headline "Quiero elegir entre más opciones de marcas y productos en Walmart" and an orange "Comprar MÁS marcas y productos" button. A light blue banner at the bottom of the main section contains the text: "Nuestras entregas continúan realizándose bajo protocolos sanitarios que nos protegen a todos. Agradecemos tu apoyo para continuar con la higiene y sana distancia." Below this banner, there is a link that says "MÁS marcas, productos y formas de pago".

Walmart  Busca tus productos  Tu cuenta 

Departamentos  Mismo día Solo en línea Chat Ayuda 

Inicio

 ¡Hasta 12 MSI y recibe hoy mismo!  
Deseo envío a domicilio o recogerlos en mi Walmart más cercano.  
[Comprar y recibir HOY](#)

 Quiero elegir entre más opciones de marcas y productos en Walmart  
Con más formas de pago y envío a todo el país  
[Comprar MÁS marcas y productos](#)

Nuestras entregas continúan realizándose bajo protocolos sanitarios que nos protegen a todos. Agradecemos tu apoyo para continuar con la higiene y sana distancia.

[MÁS marcas, productos y formas de pago](#)

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

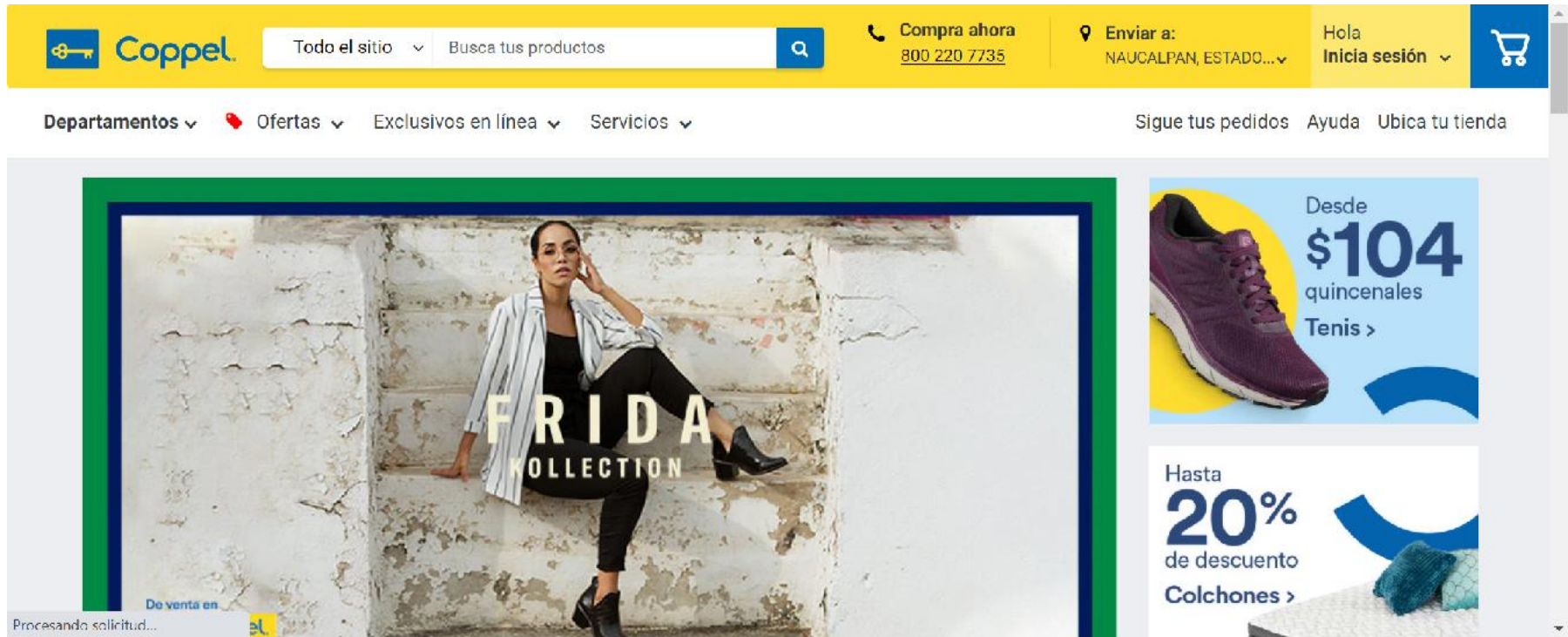
### Plataformas *e-commerce*



# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Plataformas e-commerce





# UEA Datos a Gran Escala

**Introducción a los datos a gran escala (*big data*): ): fuentes de producción y uso**

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

❑ Principales plataformas o apps e-commerce de servicio en México:

- Uber, y otras de su tipo.
- Airbnb, y otras de su tipo.
- Rappi, y otras de su tipo.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Mercado bursátil – Bolsa de valores.** La inmensa cantidad de datos, que de forma continua y a velocidad vertiginosa, proporciona el mercado bursátil, a través de las bolsas de valores, son un insumo muy apreciado para mejorar la toma de decisiones de los inversionistas. Estos grandes volúmenes de datos son preprocesados y analizados a través de técnicas y modelos de minería de datos, para producir inferencias, predicciones y estimaciones muy útiles en la toma de decisiones.
- ❑ De esta forma, los datos masivos que genera el mercado bursátil juegan un papel relevante en la era del *Big Data*.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): fuentes de producción y uso

### ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

- ❑ **Mercado bursátil – Bolsa de valores.** La información (los datos masivos) que produce de forma continua el mercado bursátil es tan importante, que entre el 15% y el 50% de los ingresos de las bolsas de valores en el mundo provienen de la venta de información.
- ❑ Este hecho refleja la importancia de estos inmensos volúmenes de datos como insumo para los modelos de minería de datos que analizan dicha información y que como resultado producen, inferencias, predicciones y estimaciones de gran utilidad para la toma de decisiones en el comercio e inversión.

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala y principales aplicaciones

### Mercado bursátil – Bolsa de valores



# UEA Datos a Gran Escala

## Mercado bursátil – Bolsa de valores



# UEA Datos a Gran Escala

## Mercado bursátil – Bolsa de valores

ACCIONES BMV												BMV/BIVA ↗	
BMV	Ind. Com. y Serv.	Sist. Int. Cot.	Trac's Extr.	Extranjeras	Seg. y Fian.	C. Bolsa	Trac's Nac.	Alta y med. burs.	Baja y mín. burs.	"In" Actinver	Fibras		
Instrumentos ^	Actual/PPP ^	Último Hecho ^	Anterior ^	Variación ^	Variación % ^	Vol. (miles) ^	Compra ^	Vol. Compra ^	Venta ^	Vol. Venta ^	Hora ^		
AC*	130.53	130.52	132.47	-1.94	-1.46	1,779	130.34	15,200	130.57	900	14:59		
ALFAA	14.02	14.00	13.90	0.12	0.86	10,234	14.00	139,573	14.08	90,647	14:59		
ALPEKA	26.03	26.15	25.86	0.17	0.66	1,264	26.07	400	26.17	19,781	14:59		
ALSEA*	44.01	43.97	42.95	1.06	2.47	1,449	43.97	42	44.04	21,100	14:59		
AMXL	20.06	20.05	19.40	0.66	3.40	74,914	20.00	94,500	20.05	27,642	14:59		
ASURB	448.70	449.04	433.56	15.14	3.49	723	446.58	8,645	449.04	813	14:59		
BIMBOA	58.97	59.11	58.43	0.54	0.92	5,247	58.74	22,641	59.11	60,654	14:59		
BSMXB	21.09	21.01	21.17	-0.08	-0.38	129	20.95	570	21.34	600	14:59		
CEMEXCPO	8.89	8.91	8.70	0.19	2.18	26,263	8.91	100	8.92	48,600	14:59		
CUERVO*	46.34	46.28	46.12	0.22	0.48	5,697	46.05	25,242	46.28	14,975	14:59		
ELEKTRA*	1,052.26	1,058.01	1,070.60	-18.34	-1.71	77	1,047.70	250	1,058.01	8	14:59		
FEMSAUBD	153.27	153.00	153.03	0.24	0.16	2,760	153.00	116	153.10	1,478	14:59		
GAPB	305.32	304.01	288.36	16.96	5.88	1,326	304.02	9,600	305.55	1,368	14:59		
GCARSOA1	65.74	65.99	64.96	0.78	1.20	349	65.55	9,880	65.99	300	14:59		
GENTERA*	14.64	14.61	15.13	-0.49	-3.24	2,117	14.61	16,578	14.70	1,929	14:59		

# UEA Datos a Gran Escala

---

## ➤ Fuentes de producción de los datos a gran escala

**Principales compañías que poseen el mayor acceso y uso de los datos personales de los usuarios**

- **Google.**
- **Amazon.**
- **Facebook.**

- ❑ Compañías como Amazon, Google y Facebook poseen una ventaja competitiva considerable respecto a todas las restantes empresas, debido al enorme volumen de datos a los que tienen acceso.
- ❑ Aparte del acceso y control de estos grandes volúmenes de datos, estos gigantes de la información y el comercio también cuentan con grandes presupuestos y recursos humanos para la elaboración y análisis de estos datos.

# UEA Datos a Gran Escala

## ➤ Fuentes de producción de los datos a gran escala

**Principales compañías que poseen el mayor acceso y uso de los datos personales de los usuarios**



Buscar con Google

Me siento con suerte

Ofrecido por Google en: [English](#) [Español \(Latinoamérica\)](#)

The image shows the Facebook "Crea una cuenta" (Create an account) page. It has a blue header with the Facebook logo and a link to "Iniciar sesión en una cuenta existente". The main heading is "Crea una cuenta" followed by the subtext "Es rápido y fácil.". Below this are several input fields: "Nombre" and "Apellido" (first and last name), "Número de celular o correo electrónico" (phone number or email), and "Contraseña nueva" (new password). There's also a "Fecha de nacimiento" (date of birth) section with a calendar icon and a "Sexo" (gender) section with radio buttons for "Mujer", "Hombre", and "Personalizado". At the very bottom, there's a small link: "Al hacer clic en 'Registrarte', aceptas nuestros Condiciones, la".



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

- ❑ En los últimos años el uso de los métodos y técnicas de **big data** ha estado cambiando la naturaleza del comercio y de los negocios. Esto se debe principalmente a la velocidad con la cual los datos generados por los propias personas (clientes, usuarios, visitantes de plataformas *e-commerce* y otros sitios web) pueden ser recolectados, procesados y analizados, y el resultado utilizado en la toma de decisiones que haga que las compañías, empresas e instituciones sean mucho más rentables y exitosas.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

- ❑ Las plataformas de comercio electrónico y sitios web tratan de recopilar la mayor cantidad posible de información sobre los datos personales, financieros, sociales, culturales, etc. de sus clientes y usuarios para satisfacer sus necesidades de cambio (mayor rentabilidad, mejor posicionamiento en el mercado, etc.) de una forma mucho más eficaz.
- ❑ En la manufactura, agricultura, producción industrial, etc., una vez que un producto ha sido elaborado, cultivado o producido, es necesario proceder a la publicidad, colocación, venta y distribución del mismo. Es aquí donde los datos masivos de clientes, usuarios y visitantes recopilados por plataformas *e-commerce* y otros sitios web, resultan de gran utilidad para indicar al comercio y a los negocios quiénes podrían ser los clientes (compradores) potenciales de estos productos.

# UEA Datos a Gran Escala

Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

➤ ***Big data* como la base de los negocios inteligentes**



Transformar los datos en información y la información en conocimiento, con la finalidad de optimizar la toma de decisiones.

**Datos**



**Información**



**Conocimiento**

# UEA Datos a Gran Escala

Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

➤ *Big data* como la base de los negocios inteligentes

Datos

Información

Conocimiento



EMPRESA	Momentum 6 meses *	Precio 30-4-2019	Precio 5-7-2019	Delta total	Rentabilidad
PHARMIA MAR	95.55%	2.125	1.880	0,000	-10,43%
CHATELAIN GASTRA	43.55%	15.990	14.990	0,000	-6,25%
LAN 2016	36.11%	7.800	7.800	0,000	0,00%
AUDIOS PERSONALIZABLES	45.11%	2.800	2.800	0,000	0,00%
ACCIONA	36.51%	105.500	97.250	5,500	-5,47%
COLABIA	36.58%	5.590	5.050	0,000	-9,70%
CONFECCION GENDRARIO	33.51%	4.000	5.000	0,000	-25,00%
REYESOLA	31.89%	8.100	8.700	0,000	-10,12%
CELLNER TELECOM	31.11%	17.400	10.500	0,000	-39,66%
SAIGU	31.10%	5.880	5.400	0,000	-8,16%
SERVICE POINT SOLUTIONS	26.95%	0.800	0.500	0,000	-37,50%
NICOLAS CORREA	26,68%	4,000	4,250	0,250	6,25%
PARQUES REUNIDOS	26,59%	15,800	15,300	0,000	-3,16%
ZAF	26,51%	40,000	40,350	0,000	0,88%
FORNOCOR	26,17%	25,900	23,000	0,000	-11,20%
ACTA	24,98%	40,000	38,500	0,000	-3,75%
INDESA	23,97%	20,200	20,300	0,100	0,50%
RENA	23,05%	155,000	156,800	0,000	1,16%
1 Class	Cost	Promotion	Before	After	
2 Confection	23.99	1467	114957	122762	
3 Drink	79.29	1745	123378	137097	
4 Luxury	81.99	1426	135246	141172	
5 Confection	74.18	1098	231389	244456	
6 Confection	90.09	1968	235648	261940	
7 Meat	69.85	1486	148885	156232	
8 Meat	100.15	1248	123760	128441	
9 Luxury	21.01	1364	251072	268134	
10 Luxury	87.32	1585	287043	310857	
11 Drink	26.58	1835	240805	272863	
12 Drink	65.23	1194	212406	227836	
13 Meat	79.82	1596	174022	181489	
14 Confection	41.39	1161	270631	283189	
15 Meat	36.82	1151	231281	235722	
16 Meat	44.05	1482	178138	185934	
17 Drink	84.62	1623	247885	278031	
18 Confection	51.82	1969	148597	165598	
19 Confection	90.08	1462	215102	228696	
20 Luxury	57.3	1842	246885	270082	
21 Drink	11.02	1370	164984	176802	
22 Confection	95.86	1815	257882	284835	



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

- ❑ Los datos recolectados, procesados y analizados de los clientes, usuarios y visitantes de las plataformas *e-commerce* y sitios web son comúnmente utilizados por el comercio y negocios para:
  - ❖ **Pronosticar las preferencias de los clientes.**
  - ❖ **Efectuar ventas cruzadas.**
  - ❖ **Determinar dónde exhibir las existencias de un determinado producto.**
  - ❖ **Identificar cuáles tiendas serán los mejores puntos de venta de un producto particular.**
  - ❖ **Rastrear los movimientos de los clientes, tanto en tiendas físicas como en tiendas en línea.**
  - ❖ **...**

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

- ❑ Hoy en día, la industria, la manufactura, la agricultura, el comercio y los negocios intentan aprovechar al máximo los grandes beneficios que proporciona la tecnología de datos masivos.
- ❑ Por tal motivo, estos sectores se han dado a la tarea de llevar la innovación a los métodos de recopilación, preparación y análisis de estos grandes volúmenes de datos, de forma tal de ir siempre más adelante con relación a las necesidades, gustos y preferencias de los clientes; de forma tal de predecir con gran certidumbre los mismos de forma muy anticipada.



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ ***Big data* como la base de los negocios inteligentes**

- ❑ Por otra parte, el enfoque de ***big data*** no sólo ha sido adoptado por estos sectores para incrementar sus ventas y beneficios, sino que también está siendo utilizado para la búsqueda de los mejores recursos humanos que posean la experiencia, los conocimientos y las habilidades que se requieren para los diferentes cargos de líderes, ejecutivos, operativos, etc.
- ❑ *Si contratar la persona equivocada en un trabajo en particular podría ser un desastre, contratar la persona equivocada en un cargo de líder o ejecutivo podría ser una catástrofe [Marr, 2015].*

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

- ❑ Para convertir un negocio en negocio inteligente (*smart business*), y por lo tanto, lograr una mayor rentabilidad, posicionamiento en el mercado y éxito, no sólo basta con poseer el acceso a los grandes volúmenes de datos generados por los visitantes, usuarios o clientes del negocio; también es imprescindible basarse en una metodología, modelo o enfoque de minería de datos, que guíe de una forma altamente organizada y estructurada la preparación, análisis y presentación de estos grandes volúmenes de datos.

# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ **Big data como la base de los negocios inteligentes**

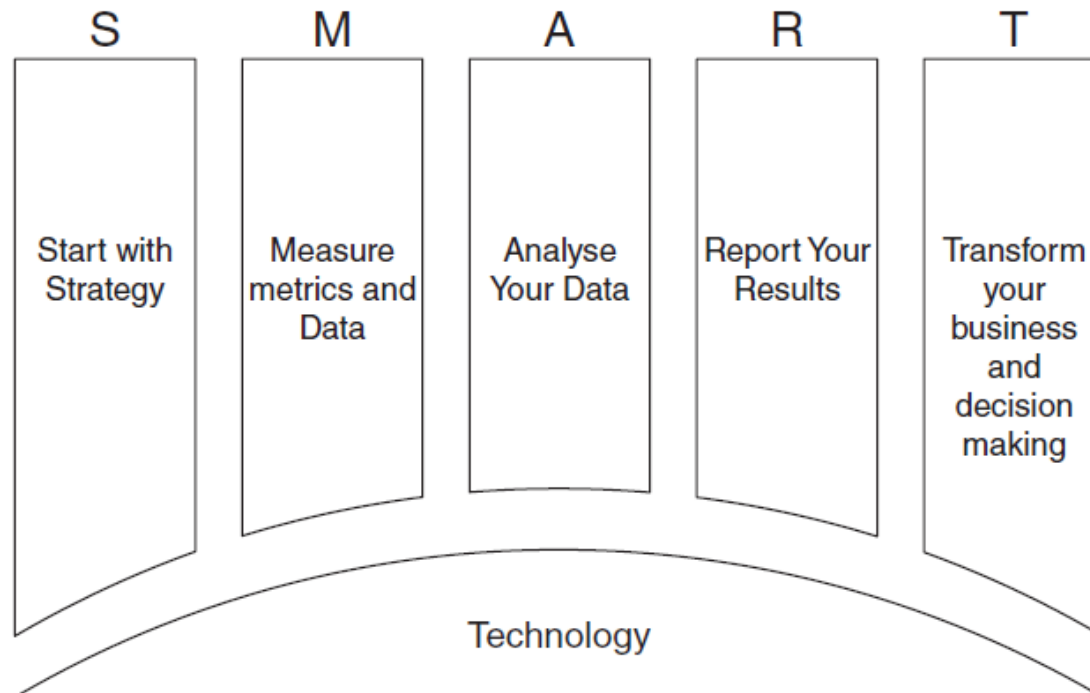
- ❑ Dos modelos, ambos basados en la minería de datos, que han resultando muy exitosos y comúnmente utilizados para sacar el mayor provecho a los **big data** son:
  - **Smart model** [Marr, 2015].
  - **CRISP-DM** (**Cross Industry Standard Process for Data Mining**) [Shearer, 2000; IBM Corporation, 2012].
- ❑ En este curso nos centraremos en la metodología **CRISP-DM**.

# UEA Datos a Gran Escala

Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

➤ **Big data como la base de los negocios inteligentes**

**SMART model [Marr, 2015]**



# UEA Datos a Gran Escala

## Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

### ➤ *Big data* como la base de los negocios inteligentes

#### *SMART model* [Marr, 2015]

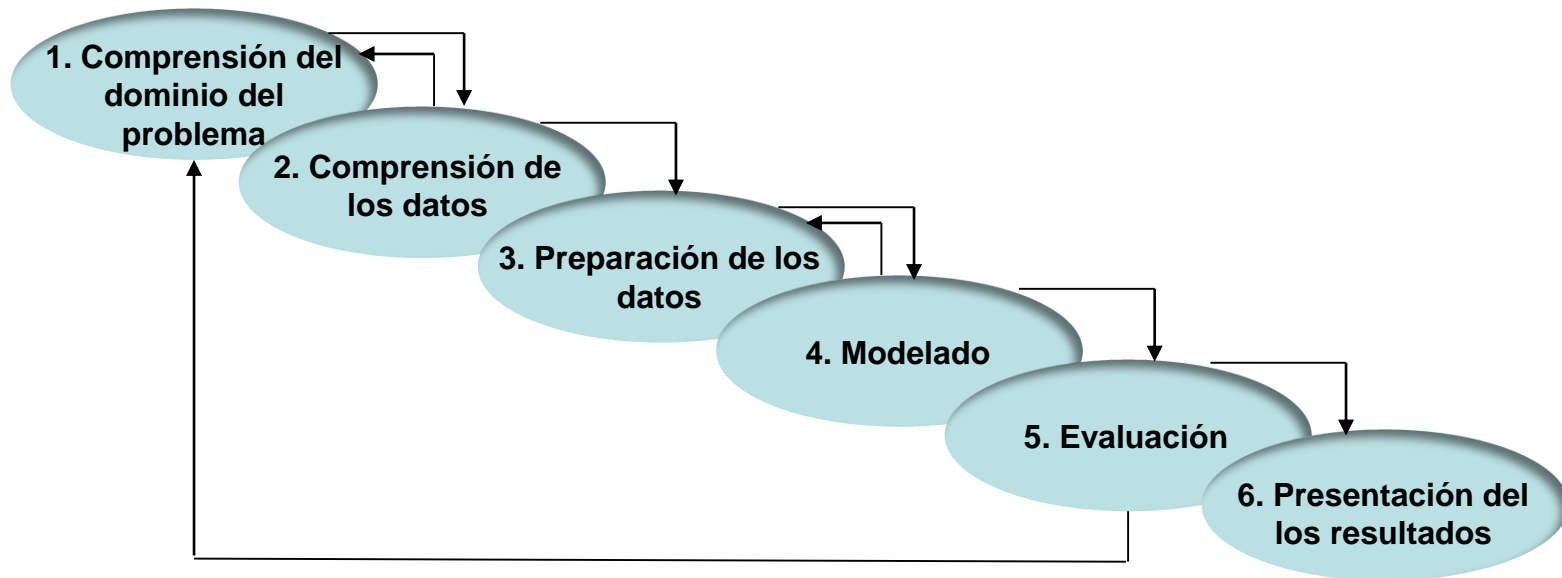
- ❖ ***S: Start with Strategy.*** Definir los objetivos del negocio y lo que realmente se está tratando de alcanzar.
- ❖ ***M: Measure metrics and data.*** Cómo acceder a la información que se requiere para alcanzar los objetivos del negocio y proporcionar respuestas a los cuestionamientos formulados. Conocer qué tipos de datos están disponibles.
- ❖ ***A: Analyse your data.*** Analizar los datos para extraer ideas útiles desde éstos que ayuden a responder las cuestiones estratégicas antes planteadas.
- ❖ ***R: Report your Results.*** Interpretar, discutir y presentar los resultados del análisis.
- ❖ ***T: Transform your business and decisión making.*** A partir de los resultados obtenidos, transformar la toma de decisiones en el negocio, optimizando su rentabilidad y posicionamiento.

# UEA Datos a Gran Escala

Introducción a los datos a gran escala (*big data*): su papel en los negocios inteligentes (*smart business*)

➤ **Big data como la base de los negocios inteligentes**

CRISP-DM [Shearer, 2000; IBM Corporation, 2012]





# UEA Datos a Gran Escala

---

## **Bibliografía necesaria o recomendable:**

1. Joyanes Aguilar, L. Big data. Análisis de grandes volúmenes de datos en organizaciones. Alfaomega, 2013.
2. Marr, B. Big data in practice. How 45 successful companies used big data analytics to deliver extraordinary results. Wiley, 2016.
3. Marz, N., Warren, J. Big data: Principles and best practices of scalable realtime data systems. Manning Publications, 2015.
4. Mayer-Schönberger, V., Cukier, K. Big data. La revolución de los datos masivos. Turner Noema, 2013.
5. Mayer-Schönberger, V., Cukier, K. Big data: A revolution that will transform how we live, work, and think. John Murray, 2017.
6. Sinha, S. Making big data work for your business. Impackt Publishing, 2014.
7. Marr, B. Big data. Using Smart big data analytics and metrics to make better decisions and improve performance. Wiley. 2015.
8. IBM. Manual CRISP-DM de IBM SPSS Modeler. IBM Corporation. 2012.

# UEA Datos a Gran Escala

---

## **Bibliografía necesaria o recomendable:**

9. Tolk, A. The Next Generation of Modeling & Simulation: Integrating Big Data and Deep Learning. SCSC 2015, July 26-29, 2015, Chicago, IL, USA. 2015 Society for Modeling & Simulation International (SCS).
10. Jacobs, A. The pathologies of big data. Communications of the ACM 52, 8 (2009), 36–44.
11. Ward, J. S., and Barker, A. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821(2013).