



Universidad  
**Tecmilenio**®



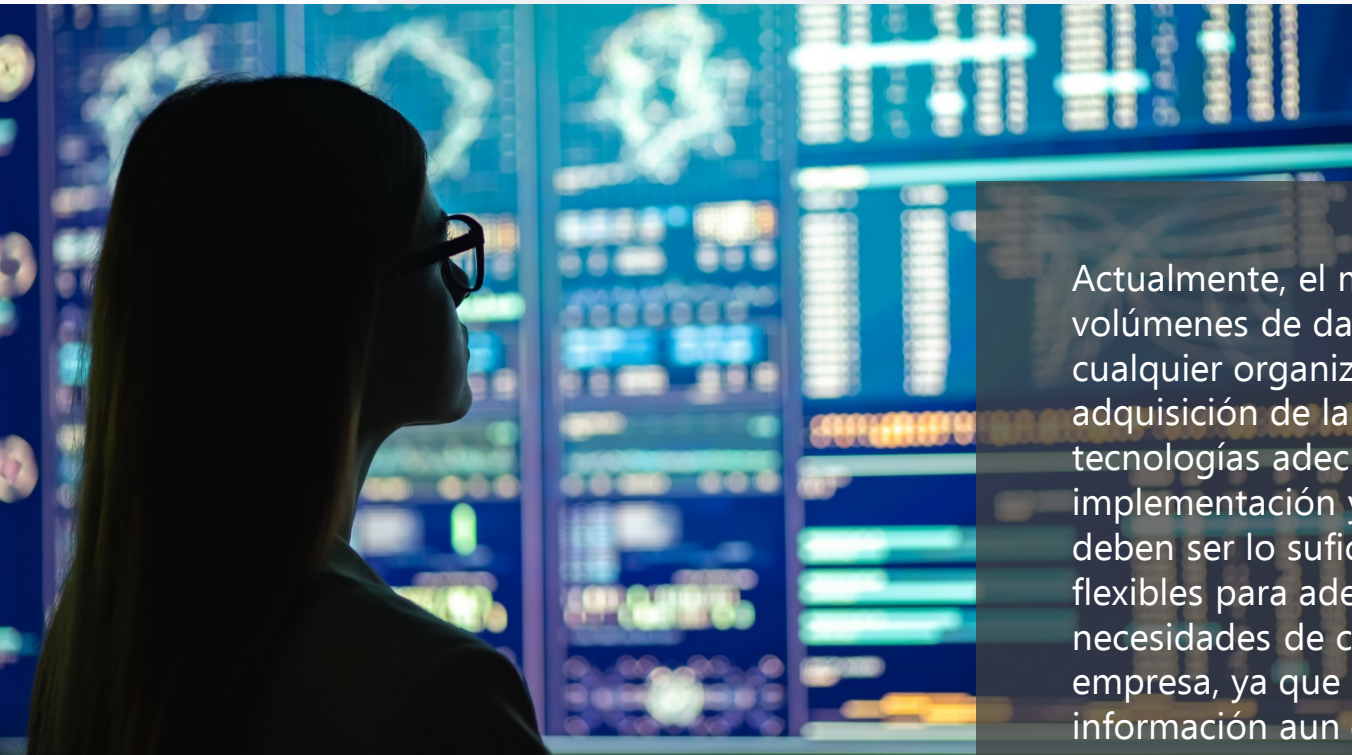


# Fundamentos de programación para Big Data

Fundamentos de Big Data

Semana 1





Actualmente, el manejo de grandes volúmenes de datos es esencial para cualquier organización, así como la adquisición de la infraestructura y las tecnologías adecuadas para su correcta implementación y gestión. Estas últimas deben ser lo suficientemente versátiles y flexibles para adecuarse a los procesos y necesidades de cualquier organización o empresa, ya que se genera un sinnúmero de información aun con las actividades más sencillas.





No existe una definición oficial de Big Data, pero se menciona la siguiente debido a que se toca el concepto del valor de los datos para el negocio, la importación del cambio y el movimiento de la tecnología para el manejo de estos.



Big Data o los “datos masivos” son la capacidad de la sociedad de aprovechar la información de formas novedosas para obtener percepciones útiles o bienes y servicios de valor significativo (Mayer-Schönberger y Cukier, 2015).





Big Data no es solo grandes cantidades de datos "sin pies ni cabeza". Existen cinco lineamientos que se deben cubrir para considerar que se están trabajando o realizando actividades de Big Data (Ishwarappa y Anuradha, 2015):

- **Volumen**
- **Variedad**
- **Veracidad**
- **Velocidad**
- **Valor**





## Las 5 V del Big Data

### Volumen

Cantidad de datos a obtener, procesar y analizar. Los datos en Big Data son muy grandes para ser procesados con un equipo de cómputo convencional como una laptop.

### Variedad

La variedad en Big Data se refiere a las diferentes fuentes y tipos de datos posibles (estructurados, semiestructurados y no estructurados).

### Veracidad

La calidad de los datos es muy importante. Estos son de alta calidad cuando son aptos para analizar y aportan información útil y resultados aplicables. De baja calidad son los que contienen una gran cantidad de datos sin sentido o peso considerable para aportar valor. Y los datos con ruido son aquellos totalmente inútiles para el propósito del procesamiento y análisis.

### Velocidad

Rapidez con la que el dato se obtiene o genera. Desde la extracción del dato hasta que pasa por distintas técnicas de proceso distribuido, por lo tanto, también se considera la velocidad de crecimiento del volumen de los datos y su accesibilidad. Esto ayuda a determinar el crecimiento de determinada área, tendencias o uso de procesos.

### Valor

Valor alcanzado por la empresa al utilizar o consumir los datos obtenidos y analizados, los cuales generan el nacimiento de nuevos productos, servicios o acciones a implementar para mejoras del proceso, incrementos de ventas, captación de usuarios y muchos ejemplos más (Marr, 2015).





Las 5 V del Big Data, su procesamiento y gestión con analítica o algoritmos para obtener información valiosa, además del conocimiento y dominio de su almacenamiento, procesamiento y análisis es lo que hoy en día se conoce como Big Data.

Todo el manejo de datos no tendría ningún objetivo si no se realizaran importantes decisiones y mejoras en las empresas, los servicios, la ciencia y la tecnología.





- Ishwarappa y Anuradha, J. (2015). *A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology*. Recuperado de <https://www.sciencedirect.com/science/article/pii/S1877050915006973>
- Marr, B. (2015). *The 5 V's of Big Data*. Recuperado de <https://www.datasciencecentral.com/profiles/blogs/the-5-v-s-of-big-data-by-bernard-marr>
- Mayer-Schönberger, V., y Cukier, K. (2015). *Big Data. La revolución de los datos masivo*. (2ª ed.). Estados Unidos: Houghton Mifflin Harcourt.







# Fundamentos de programación para Big Data

Terminología y conceptos de  
Big Data

Semana 1





Big Data es una herramienta dentro del universo de recolección de datos para su uso en la inteligencia artificial y en otras herramientas o técnicas, como el *machine learning*.

Además, los diferentes tipos de orígenes se refieren a los datos, ya que el contexto dentro del diseño de ambiente Big Data, su recolección y el tratamiento para su análisis se pueden predefinir para su uso.





**Tipos de datos según su origen o fuente**  
(Carisio, s.f.)

## Tipos de datos



**Web**



**M2M**



**Big Transaction Data**



**Biométricos**



**Generados por humanos**





## Tipos de datos según su origen o fuente

<b>Web</b>	Se refiere a los datos obtenidos desde la navegación web, uso de aplicaciones en línea, comportamiento en redes sociales, etcétera.
<b>M2M</b>	<i>Machine to machine</i> utiliza sensores que capturan un evento en particular (velocidad, temperatura, etc.) y se transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.
<b>Big Transaction Data</b>	Como transacciones bancarias, registros de facturación, registros de llamadas, etc. Estos datos transaccionales están disponibles en formatos semiestructurados y no estructurados.
<b>Biométricos</b>	Estos se obtienen mediante huellas digitales, reconocimiento facial, geometría de la mano, escaneo de retina, entre otros.
<b>Generados por humanos</b>	Estos se obtienen mediante huellas digitales, reconocimiento facial, geometría de la mano, escaneo de retina, entre otros.





## Ventajas del Big Data



1

El uso de datos para el análisis predictivo: así se predicen comportamientos de clientes y/o usuarios, crecimiento de ventas, etc.

2

Captación de usuarios o consumidores en una experiencia que constantemente se esté actualizando a sus necesidades.

3

Eficiencia industrial, ya que permiten un mayor control de las variables clave en un ambiente de producción, por ejemplo, temperatura, disposición de la línea, presión, etc.





Exploraste brevemente cómo los tipos de datos pueden depender de su origen o fuente, pero más adelante abordarás con profundidad los tipos de datos disponibles.





- Carisio, E. (s.f.). *¿Qué es Big Data y para qué sirve? Ejemplos de uso*. Recuperado de <https://blog.mdcloud.es/que-es-big-data-y-para-que-sirve/>

