



# Instituto Tecnológico Autónomo de México

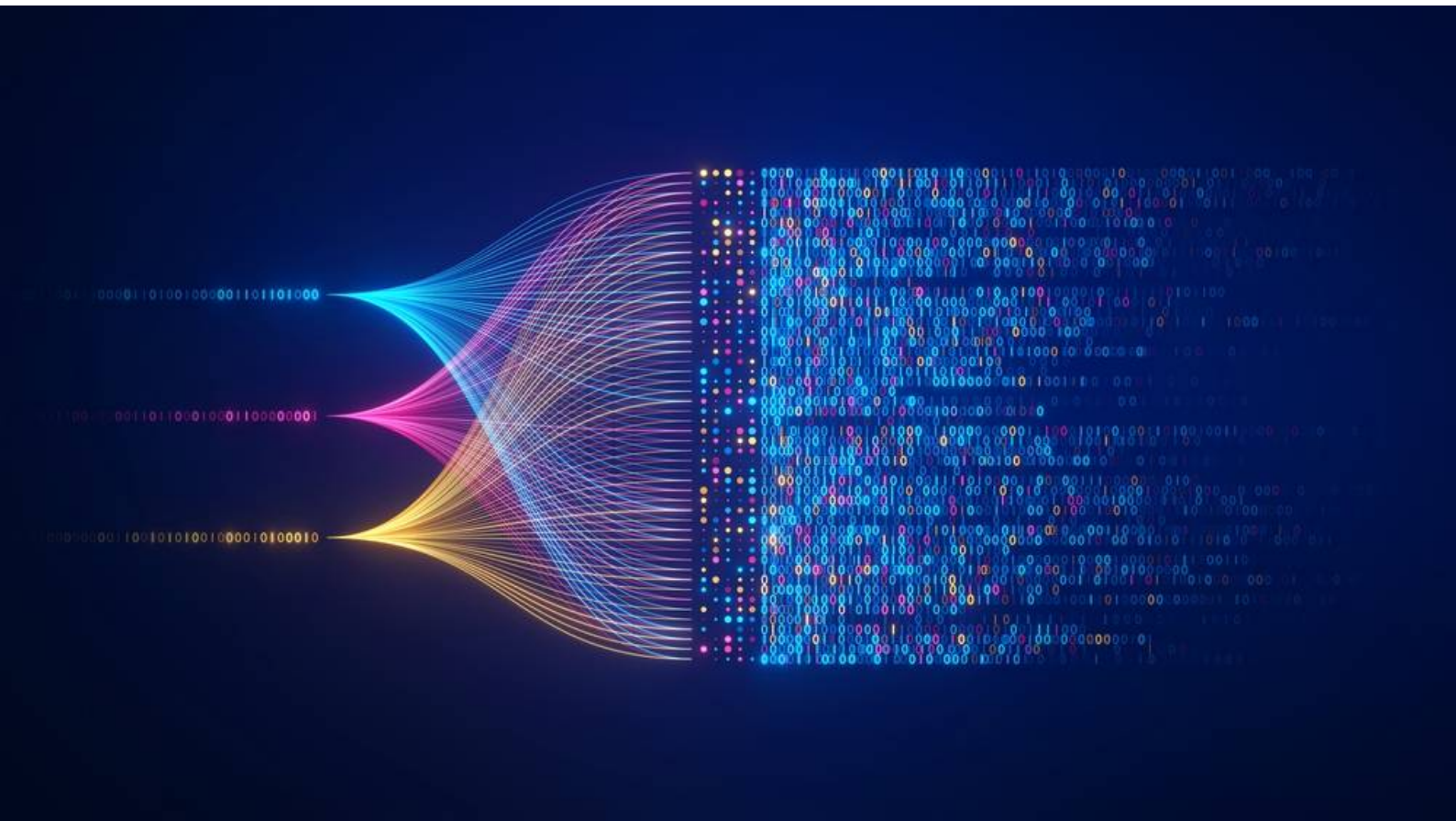
**Diplomado:** Ciencia de Datos y Machine Learning Aplicado a Finanzas

**Módulo 3:** Ciencia de Datos

## Tema

*Big Data*

**Nombre:** Denzel Arik Martinez Maldonado



## Big Data y sus características (V's) en la industria

Cuando la gente se refiere a Big Data, muchas veces no significa solo “muchos datos”, sino que el problema viene por varios lados al mismo tiempo: que hay mucho, que llega rápido o que viene en varios formatos, que a veces viene sucio o dudoso y que además no sirve de nada si no se traduce en decisiones. La idea central del artículo es esa: para entender qué es Big Data en serio, hay que entender sus características (las famosas V's), y cómo han ido creciendo con el tiempo.

También se remarca que no hay un consenso perfecto. Hay investigadores que se quedan con 3 V's, otros con 5, otros llegan a 7 y algunos proponen hasta 10 o más. Esto crea una especie de desorden porque dependiendo del autor o del sector, te cambian qué V's son “las importantes”.

### Las 3 V's clásicas

Estas son las que normalmente se toman como base, porque describen el problema más evidente:

- **Volumen:** hay muchísimos datos. No es lo mismo analizar 10 mil filas que millones o billones.
- **Velocidad:** los datos llegan y cambian rápido. Si te tardas demasiado en procesarlos, ya se volvieron viejos.
- **Variedad:** los datos vienen en distintos formatos como tablas, texto, audio, video, sensores, logs, etc. No todo es Excel bonito.

### La 4ª V: cuando aparece el tema de “¿esto es confiable?”

Con el tiempo se vio que no basta con que los datos sean grandes, rápidos y variados, porque muchas veces vienen con problemas como ruido, sesgos, errores, datos viejos o datos que se contradicen.

- **Veracidad:** se refiere a qué tan confiables son los datos, si están sesgados, si tienen errores, si son consistentes.

### La 5ª V: “¿y esto para qué sirve?”

Aquí entra lo más importante desde el lado negocio: puedes tener el mejor sistema del mundo, pero si no genera algo útil no sirve.

- **Valor:** es la capacidad de convertir datos en valor real, normalmente valor económico, pero también valor social o estratégico.

### Cuando se sube a 7 V's: afinando lo “confiable” y lo “usable”

Aquí agregan dos ideas que suenan parecidas a cosas anteriores, pero en la práctica sí cambian cómo trabajas:

- **Validez:** que el dato sea “correcto” para el uso que tú quieres. O sea, puede tener buena veracidad en general, pero aun así no ser válido para tu objetivo, por ejemplo: medir algo con una variable que no representa bien lo que dices medir.

- **Volatilidad:** qué tan rápido cambian los datos y qué tanto conviene guardar o desechar información vieja. Hay datos que expiran en valor.

### Cuando se llega a 10 V's: ya no es solo guardar datos, es administrarlos bien

Aquí el enfoque ya es más completo: no solo se trata de “tener datos”, sino de poder analizarlos, comunicarlos y protegerlos.

- **Variabilidad:** no solo variedad de formatos, sino que los datos pueden ser inconsistentes o cambiar en comportamiento dependiendo del momento o contexto. Esto te obliga a pensar en anomalías y outliers.
- **Visualización:** cómo presentas los datos importa muchísimo, porque una gráfica mal elegida puede confundir o llevar a conclusiones equivocadas.
- **Vulnerabilidad:** la parte de seguridad y cumplimiento, que los datos se manejen bien legalmente y bajo lo que el usuario permite.

### Cada industria usa distintas V's

Una parte clave del artículo es que no todas las industrias se fijan en lo mismo. En sectores como retail y oil & gas se reporta que ya se trabaja mucho con 7 V's, mientras que otros como telecomunicaciones se quedan en 3 V's. La interpretación práctica es:

- Si tu sector maneja datos muy sensibles o con alto riesgo de error (salud, seguridad, gobierno), te importa más veracidad, valor, validez, etc.
- Si tu sector maneja flujos enormes y constantes (telecom), a veces se enfocan primero en volumen/velocidad/variedad y lo demás lo van incorporando.

Big Data es un conjunto de problemas y responsabilidades. Las 3 V's son el punto de partida, pero en el mundo real, cuando quieres que los datos realmente sirvan, empiezas a necesitar más V's para confiabilidad, utilidad, seguridad, comunicación. Y además, depende del sector: cada industria implementa sus V's según el tipo de datos y las decisiones que necesita tomar.