El Big Data actualmente es muy importante en nuestras vidas y principalmente en el sector empresarial, ya que es tanta información que se genera a diario de tantas personas o situaciones que es necesario usar herramientas que nos permitan el análisis de esos grandes volúmenes de datos para extraer un valor que nos permita resolver un problema o hasta “predecir” el futuro.

Todo esto busca cumplir 4 características principales: **Variedad** (en forma y formatos), **Velocidad** (casi humana), **Volumen** (enormes) y **Visibilidad** (bien expresados).

Para lograr el objetivo de manipular, gestionar y analizar los enormes conjuntos de datos es necesario utilizar herramientas que van más alla de las tradicionales, por ejemplo, Scala, Apache Spark, R, Python, etc.

Hoy en día, ya no se busca tener un mega servidor que soporte todas estas operaciones, porque es muy poco eficiente y seguro, ya que el trabajo lo realiza solo una máquina y si tiene problemas, perdemos todos nuestros datos. Por ello para el Big Data se utiliza un nuevo paradigma en el cual los sitemas están ditribuidos en n máquinas que se dividen las tareas y trabajan por un mismo fin; asegurando la eficiencia y la seguridad de nuestros datos, ya que si una falla, las demás suplen su tarea.

Todo esto tiene sus bases en la Estadítisca (descriptiva e inferencial), Minería de Datos, Aprendizaje Automático (Machine Learning), Programación, Cómputo Distribuido y tolerancia a fallas, entre otras.

Finalmente, hoy más que nunca el mundo necesita científicos de datos, es decir, gente que esté especializada en los procesos y sistemas que extraen conocimiento de grandes volúmenes de datos, porque ya no sirven de nada los datos si no se les da un buen tratamiento para convertirlos en información que sirva que sea concisa, precisa, oportuna, verdadera y que sea de ayuda para la toma de decisiones.