

XVII CONGRESO INTERNACIONAL DE COSTOS del IIC

Sevilla 2021

**BIG DATA: REFLEXIONES SOBRE LA RELACIÓN ENTRE LAS
CUALIDADES DE LOS DATOS Y SU RELEVANCIA EN LA
GESTIÓN**

Categoría propuesta: Ensayo

Autores:

**Marcelo Podmoguilnye – socio IAPUCo - Universidad de Buenos
Aires (FCE) – Argentina**

**Daniel Farré – socio IAPUCo - Universidad de Buenos Aires (FCE) –
Argentina**

Datos de contacto:

Daniel Farré

Número telefónico: 541149471933

**Palabras clave: Big Data, Gestión, Ciencia de Datos, Data Lake,
Lakehouses**

Big Data: Reflexiones sobre la relación entre las cualidades de los datos y su relevancia en la gestión
Categoría propuesta: Ensayo

Resumen

Los datos aumentaron su valor y disminuyeron su costo exponencialmente. ¿Cuál es el nuevo equilibrio en la ecuación?

A dos décadas de la aparición del estudio que pronosticaba la tendencia de incremento exponencial de tres características de los datos: Volumen, Variedad y Velocidad (Frecuencia), fueron muchos los autores que siguieron el “juego” de la aliteración, a medida que cobraba mayor relevancia el concepto de “*Big Data*”, subiendo de las 3V hacia 4V, 7V y 42V a lo largo de 16 años.

La ponencia reflexiona sobre cada una de las cualidades catalogadas por estos autores, relacionándolas con la relevancia que adquieren cada una de estas cuando se las emplea para la Toma de Decisiones (dejando fuera su relevancia para otros fines más operativos) con el objetivo de evaluar el universo de datos (internos y externos a la Organización) que deben formar parte de los “Lagos de Datos” si queremos que estos sean rentables.

Así mismo, se aborda el tema del rol que deberían desempeñar los profesionales en Ciencias Económicas, específicamente los vinculados con la especialidad Costos y Gestión. Para poder cumplir este objetivo relacionamos la evolución de las plataformas de sustento de información directiva, desde la aparición del “*Datawarehouse*” hasta los actuales “*Datalakes*” y “*Lakehouses*”, con nuestro aporte de valor en cada una de las etapas del ciclo de agregado de valor en la pirámide del conocimiento, también conocida como jerarquía *DIKW* por sus iniciales en inglés.

1. Objetivos

A veinte años de la aparición del artículo de Doug Laney (Laney 2001), el cual versaba sobre que la *Big Data* tenía que ver con las tres “V”, volumen, velocidad y variedad; fueron muchos autores los que siguieron este “juego” de la aliteración. A medida que cobraba mayor relevancia este concepto, iban subiendo el número de “V”.

Es objetivo del presente trabajo, reflexionar sobre cada una de las cualidades catalogadas por estos autores, relacionándolas con la relevancia que adquieren cada una de estas cuando se las emplea para la Gestión y el Desarrollo de modelos de negocio organizacionales, a los efectos de evaluar el universo de datos que deben formar parte de los “Lagos de Datos” y las “Casas del lago”, si pretendemos que estos sean rentables y agreguen valor al desarrollo de las organizaciones.

Así mismo se distingue el concepto de *Big Data* y de *Data Science*, ambos utilizados de manera conjunta en estos artículos, pero sobre los cuales reflexionamos, proponiendo nuestra visión sobre ambos recursos, sus principales diferencias y el rol del profesional de Contabilidad Directiva en este sentido.

Se trata de encontrar respuesta a preguntas muy comunes en el ámbito de utilización de estos recursos como ser: ¿Por qué fallan los proyectos de datos? o ¿Cuáles son los principales casos de uso de la Big Data y la Ciencia de Datos?

Y finalmente, destacar el impacto que tienen estas disrupciones tecnológicas en nuestro ejercicio profesional, como especialistas en costos y gestión empresarial.

2. ¿Qué es el Big Data y qué es la Ciencia de datos?

Big Data es un término que describe un importante volumen de datos, los cuales pueden estar estructurados o no y que “inundan” las organizaciones y los negocios día a día. Sin embargo, la mayoría de los autores coinciden en que lo relevante no es el volumen de datos, sino lo que las organizaciones hacen con los datos. Es esencial entender al Big Data como un fenómeno social y tecnológico, provocado principalmente por dos grandes movimientos:

- El comportamiento de las personas en cuanto a:
 - La utilización constante de dispositivos dotados de internet y de aplicativos que recopilan, de manera permanente, información sobre sus comportamientos, tanto en sus movimientos económicos cuanto en información personal y social.
 - La cultura de compartir públicamente en distintas redes sociales distinto tipo de información personal y de conocimiento (Fenómeno Wiki 2.0)
- El avance tecnológico en dos frentes:
 - La computación pervasiva o ubicua: que acompaña al cambio comportamental ut supra con todo tipo de canales, tecnologías, sistemas, dispositivos y artefactos, propagándose con efecto rizomático.

- El fenómeno de Internet de las cosas (*IOT*), potenciado por la reducción de los costos de originación, conservación y publicación de una cantidad inmensa de datos que logra superar a la generada por el ingreso humano

La ciencia de datos entra en acción creando beneficios transformadores desde la interpretación de la información recopilada en las bases y lagos de datos. Revela tendencias y genera información que las empresas pueden utilizar para mejorar el entendimiento de sus negocios y la toma de decisiones para la creación y reformulación de sus modelos de negocios (sustentando el diseño de productos y servicios innovadores) y para la conducción que permita una ejecución alineada con las decisiones optimizadas. La ciencia de datos, bien orientada por sus analistas y gestores, debe darle utilidad a la inagotable materia prima que conforma el universo de Big Data.

Los datos son la base de la innovación y el conocimiento, pero su valor agregado proviene de la información que los especialistas pueden filtrar, extraer, traducir, procesar, agregar, clasificar y luego aplicar a partir de esos datos. Sin su aporte (etiquetado actualmente como “curaduría de contenidos”) se produce un efecto negativo sobre la Toma de Decisiones, conocido como “Síndrome de Fatiga por Información” de acuerdo con la definición que el psicólogo David Lewis (1996) le dio al efecto inmovilizador producido por la sobrecarga de datos no focalizados en los objetivos directivos.

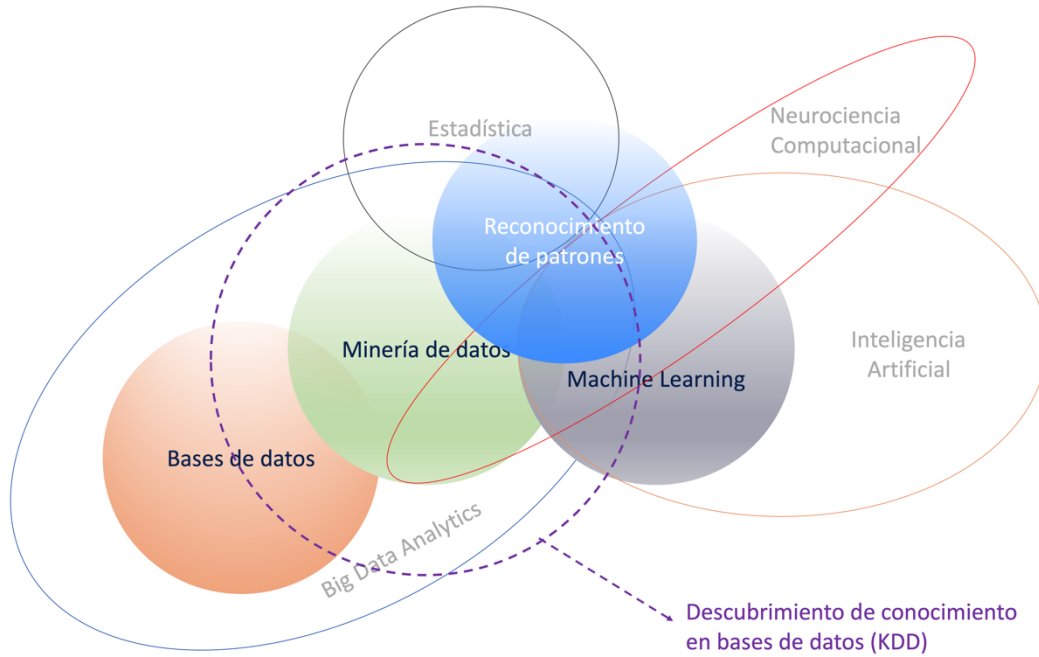
Para una mejor comprensión de su concepto y su aprovechamiento, se hace necesario diferenciar términos relacionados con el campo de la Inteligencia Artificial (IA) y el aprendizaje automático. En muchos artículos e incluso en la aplicación profesional de estas herramientas se suelen confundir semánticamente estos términos indistintamente, situación por la cual explayamos nuestra posición sobre los mismos.

Inteligencia artificial significa hacer que un ordenador imite de alguna manera el comportamiento humano. La ciencia de datos es una rama de la IA que hace referencia a sectores superpuestos de las estadísticas, a la aplicación de métodos científicos y el fijar herramientas de análisis de datos; los cuales son utilizados para obtener significado y lograr conocimiento de dichas fuentes. Convierte la información que inunda las organizaciones en datos de alto valor agregado para las organizaciones.

El aprendizaje automático refiere a un conjunto de técnicas que permiten a los ordenadores descubrir patrones a través de los datos y posibilita la realización de aplicaciones de IA.

Diferenciados estos conceptos, ponemos nuestro foco reflexivo en el “*Big Data Analytics*” (BDA), el cual es un método de trabajo y un conjunto de tecnologías que interactúan con el propósito de encontrar los patrones e interrelaciones ocultos en los grandes volúmenes de datos, y poder así extraer y producir información útil para su gestión y análisis.

De esta forma, se hace necesario entender la naturaleza multidisciplinaria del BDA, ya que incorpora subconjuntos de una amplia variedad de materias. Tradicionalmente ha habido muchas disciplinas relacionadas que se enfocaban centralmente a: encontrar patrones claves en los datos, pero cada una con una especialidad diferente. En el siguiente cuadro #1, se muestra cartográficamente la relevancia y ubicación de estos campos multidisciplinarios del BDA.



Cuadro 1: Subconjuntos de los campos multidisciplinares del BDA (Reelaboración propia a partir de la Fuente: SAS Institute Inc.)

Sintetizando, podríamos afirmar que la potencialidad de la Big Data, la Inteligencia artificial con sus ramas o subconjuntos, la ciencia de datos, el aprendizaje automático y el aprendizaje profundo, en su combinación orientada teleológicamente por los especialistas, nos brindan la posibilidad de agregar valor a la información y provocar que sea útil para su utilización en distintos procesos Directivos (*Data driven Learning*, Toma de Decisiones y Conducción alineada).

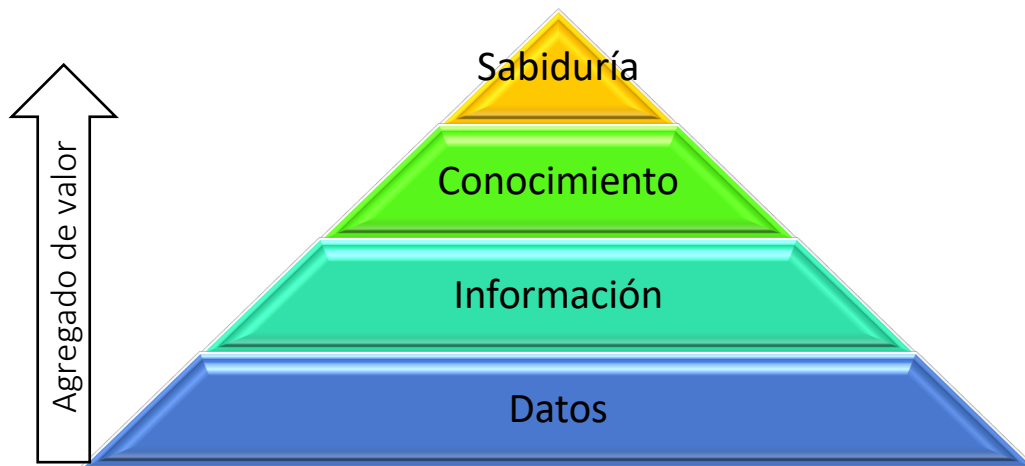
La tecnología debe soportar el negocio y no el negocio a la tecnología. Ninguna empresa demanda tecnología porque sí. Siempre hay un objetivo y necesidades por los cuales se termina incorporando tecnología. La “compra” de datos, no servirá de nada si no hacemos ciencia de datos teleológica con los mismos.

3. ¿Cómo se relacionan estas tecnologías con la jerarquía DIKW (Datos- Información-Conocimiento-Sabiduría)?

La pirámide del conocimiento, también conocida como jerarquía *DIKW* por sus iniciales en inglés, esquematiza la relación entre los Datos, la Información, Conocimiento y Sabiduría, cada uno como un escalón jerárquico (de abajo hacia arriba) que va agregando valor.

Si bien no existe un límite claro entre cada nivel, podemos definir convencionalmente las fronteras por el nivel de entendimiento necesario para ascender en la pirámide:

- Se necesita entender las relaciones para pasar de Dato a Información
- Se necesita entender los patrones para pasar de Información a Conocimiento
- Se necesita entender los principios para pasar de Conocimiento a Sabiduría



Cuadro 2 – Pirámide del Conocimiento o Jerarquía DIKW – (Elaboración propia)

Por lo expuesto ut supra es evidente que es correcta la clasificación de la BD en el nivel inferior, de Datos (*Data* es el latín de Datos), mientras que la Ciencia de Datos es la Ciencia que permite ir pasando de escalones por el aporte en el entendimiento de las relaciones y los patrones. Dejamos para un debate posterior si, en el estado del arte actual podemos o no hablar del pase al nivel de Sabiduría.

4. Reflexiones sobre las cualidades de la Big Data y la Ciencia de Datos

Desde que en 2001 Doug Laney, analista de la consultora Gartner, resumió las características esenciales de la revolución del Big Data con tres palabras que comenzaban con la letra “V”, velocidad, volumen y variedad, muchos fueron los autores que lo sucedieron agregando caracterizaciones, intentando conservar la letra “V” como consonante inicial. Desde nuestra atalaya, este juego de aliteración confunde características propias del Big Data, con las de la Ciencia de Datos y la inteligencia artificial aplicadas al análisis y a la generación de información.

En el cuadro 1 podemos observar las tres “V” que dispararon el artículo de Doug Laney, velocidad, volumen y variedad.



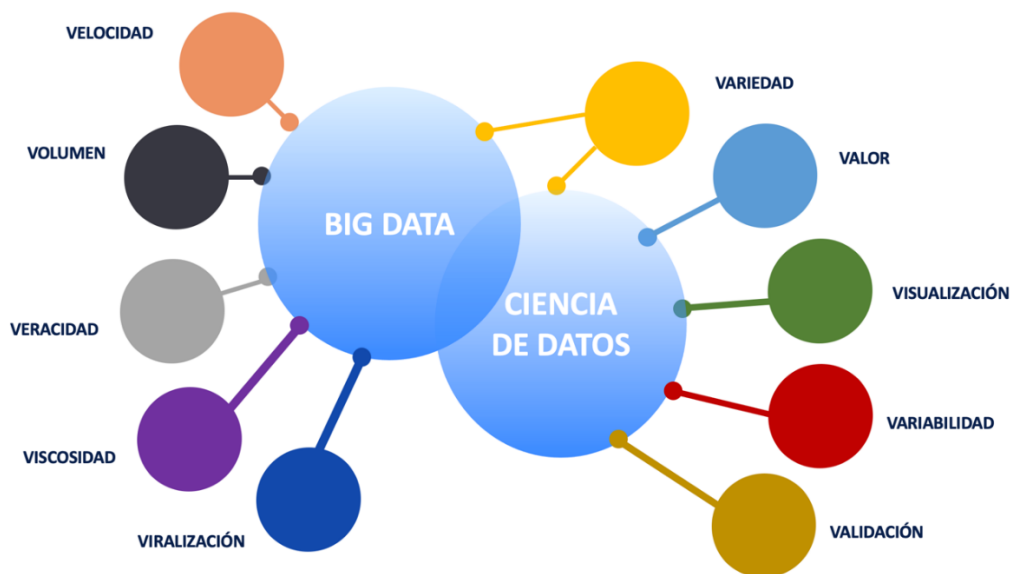
Cuadro 3 – Las tres “V” de la Big Data. (Elaboración propia)

Del análisis realizado en el cuadro precedente, podríamos concluir que la velocidad y el volumen son características propias de la Big Data, mientras que la variedad tiene una interpretación polisémica, ya que tiene que ver con la combinación de diversos formatos de datos (archivos planos, bases de datos relacionales, datos no estructurados conservados en contenedores de organización virtual, redes de gráficos, etc) propios de la clasificación en datos estructurados y no estructurados de la Big Data, pero también con diferentes niveles de complejidad de la información que se asocian a las problemáticas encaradas por la ciencia de datos,

Luego se agregaron dos “V” más, la veracidad y el valor. La primera es propia de la Big Data, y está relacionada con el grado de validez de la información y su confiabilidad, dado que en el proceso de ingesta de datos al Big Data no tiene que pasar por protocolos de veracidad y calidad al origen. Siguiendo un principio husserliano (de clara influencia del método *Epokhé* de los escepticistas), se realiza un primer paso en donde se apagan temporalmente los juicios de valor al momento de ingesta, para que los datos otorguen mayor libertad a la Ciencia de Datos para identificar correlaciones y esencias sin los filtros de sesgos o prejuicios. El valor, entonces, es propia de la Ciencia de Datos aplicada a la Big Data, proporcionando un valor cada vez mayor para sus usuarios a medida que se desarrollan más técnicas de análisis de datos, alineadas con los objetivos de los Directivos.

En el año 2013 se agregaron la visualización y la variabilidad. Ambas son una consecuencia de la aplicación de la ciencia de datos a la Big Data. La variabilidad tiene que ver con la modelización de las fuentes de datos y su aplicabilidad en las organizaciones. Mientras que la visualización se centra en la interacción de los usuarios con la información, lo que sería imposible sin el desarrollo de algoritmos que permitan interpretar información, con su respectivo impacto en el agregado de valor a la misma.

En el 2014 se agregaron las características de viscosidad, validación y viralización. La viscosidad y la viralización son propias de la Big Data. La primera tiene que ver con las dificultades para trabajar con los datos y se relaciona con la velocidad, mientras que la viralización tiene que ver con la fuente de datos, lo cual se vincula a la vez con las aplicaciones y dispositivos que utilizan los usuarios y el universo de la IOT. La validación, en cambio, está vinculada con la ciencia de datos, ya que tiene que ver con el rigor del análisis de la información y con la validez de las predicciones que se puede lograr con los datos.



Cuadro 4 : Las 5 “V” de la *Big Data* y la Ciencia de Datos. (Elaboración propia)

Llegando al artículo de Tom Shafer, se denota al enunciar las 42 V de la *Big Data*, una mezcla de características propias de cada una de las disciplinas que intervienen en el BDA. Muchas de ellas son repetitivas y abarcativas, y probablemente distraigan y confundan al lector, en lugar de clarificar. Ya el sólo hecho de enunciarlas como características del “*Big Data and Data Science*” compromete la comprensión de ambos conceptos.

5. Ejemplificando el aporte del BDA en la transformación de los modelos de negocio

Cada vez más organizaciones están utilizando el BDA para convertir información en una ventaja competitiva la transformación de sus productos y servicios. De recopilaciones realizadas sobre entrevistas a profesionales referentes de la especialidad, surgen algunos casos de uso del BDA:

- Dimensionar y entender los motivos de la fuga de clientes, cruzando y analizando los datos demográficos de los clientes con aquellos que se recopilan en los distintos canales de interacción cliente-negocio: solicitudes de productos y servicios, compras, formas de pago, uso de la omnicanalidad en la posventa. Con dicha información, se definen distintos segmentos de clientes para realizarles acciones preventivas a la fuga y ofrecimientos de retención a medida para que los centros de contacto y sucursales tengan habilitados para los momentos de pedidos de baja.
- Mejorar la eficiencia en el análisis de patrones de tráfico, condiciones climáticas, huellas de carbono y otros factores para que las empresas de logística puedan mejorar los tiempos de entrega y reducir los costos (económicos y ambientales).
- Analizar los diagnósticos de los pacientes mediante el estudio de los exámenes médicos y los síntomas informados para que los profesionales de la salud puedan diagnosticar prematuramente las enfermedades y tratarlas con mayor eficacia.
- Optimizar la cadena de suministro al predecir cuándo se producirán las fallas y averías en los equipos de trabajo.
- Detectar y prevenir los fraudes en los servicios financieros mediante el reconocimiento de comportamientos sospechosos y de acciones anómalas.
- Mejorar las ventas al crear recomendaciones para los clientes basadas en compras anteriores, encuestas de satisfacción e interacción en múltiples canales.
- Mejorar y comprender rápidamente el sentimiento del cliente a través de la escucha de contenido de medios sociales para mejorar la interacción con los clientes y el diseño de productos y servicios que les ofrecen.
- Mejorar el entendimiento de la satisfacción de los clientes en servicios difíciles de medir como el turismo, recopilando y procesando gran volumen de datos, e identificar los potenciales problemas antes que sea demasiado tarde.
- Comprender con mayor profundidad las necesidades y su oportunidad, a partir de información de programas de fidelización de clientes, hábitos de compra y otras fuentes virales, y predecir tendencias y recomendar nuevos productos.

- Desplegar sensores en productos de la industria manufacturera para recibir datos de telemetría, revelando patrones de uso y tasas de fracaso que alimentarán procesos de mejoras de reducir costos de desarrollo y montaje.
- Aprovechar el auge de utilización de teléfonos y dispositivos inteligentes, incluyendo información de sus GPS para dirigirse a los consumidores cuando se acercan a un lugar de interés.

Todo ello conduce a negocios más inteligentes, operaciones más eficientes, mejoras en la rentabilidad y clientes más satisfechos. Las empresas que consiguen dar valor a su información han logrado:

- Reducir costos: las tecnologías y la ciencia de datos aplicadas a las información aportan importantes ventajas en términos de costos. Permiten identificar las maneras más eficientes de hacer negocios.
- Rapidez y mejora en las decisiones, ya que las empresas pueden analizar la información inmediatamente y tomar decisiones en función de lo que han aprendido.
- Nuevos productos y servicios: al tener la capacidad de medir las necesidades de los clientes y las posibilidades de mejorar su grado de satisfacción, se crea el poder de dar a los clientes lo que necesitan y desean.

6. Proyectos de Datos

Muchas organizaciones han hecho del BDA una prioridad y están destinando inversiones en este sentido. Ahora bien, ¿Cómo se llevan adelante los proyectos BDA? ¿Están siendo exitosos?

El proceso de analizar y utilizar los datos es iterativo más que lineal, sin embargo podemos definir algunos pasos críticos para el desarrollo de un modelo BDA:

- Orientación: Definiendo claramente los objetivos buscados. En “Metafísica”, Aristóteles explica su principio teleológico o de causa final, de la siguiente manera:

“puesto que el fin es el punto de llegada de las acciones que se llevan a cabo, y puesto que lo que es fin es una causa, si ello no existiese no existiría nada de lo que sucede”;

Las características descritas ut supra de Volumen Velocidad y Viscosidad conforman un bosque impenetrable si no es con la guía de este principio, como lo describe Natali (1999):

“la causa final [...] permite seleccionar, entre las cosas que la causa motriz puede hacer aquí y ahora, aquella que conduce al bien y al fin, y motiva a la causa motriz a hacer específicamente aquella acción y no otra, entre las infinitas cosas que el agente tiene la posibilidad de hacer en un momento dado [...] como una guía segura”

- Planificación: definiendo un proyecto holístico, fijando la responsabilidad de los involucrados, los tiempos e hitos intermedios y finales.
- Definición de la arquitectura modelo de datos, eligiendo el o los *stacks* tecnológicos objetivo, en la nube y/o en Centros de Data propios, con sus componentes de Sistemas Operativos, Contenedores, conjunto de herramientas de Data

Management, Lenguajes de programación para aquello que no se resuelve con las herramientas de Data, esquemas de integración de datos internos y externos.

- Definición del modelo de Gobernanza de datos.
- Construcción del modelo de datos, teniendo en cuenta la visión negocios, las responsabilidades de los dueños de los datos y las características técnicas.
- Implantación del modelo, en olas progresivas, para no desaprovechar el *momentum* generado por el consenso logrado al inicio del proyecto.
- Evaluación y evolución continua del modelo, por medio de un conjunto completo de métricas para comparar proyección con realidad, de modo de alimentar el mismo con prácticas de mejora continua.

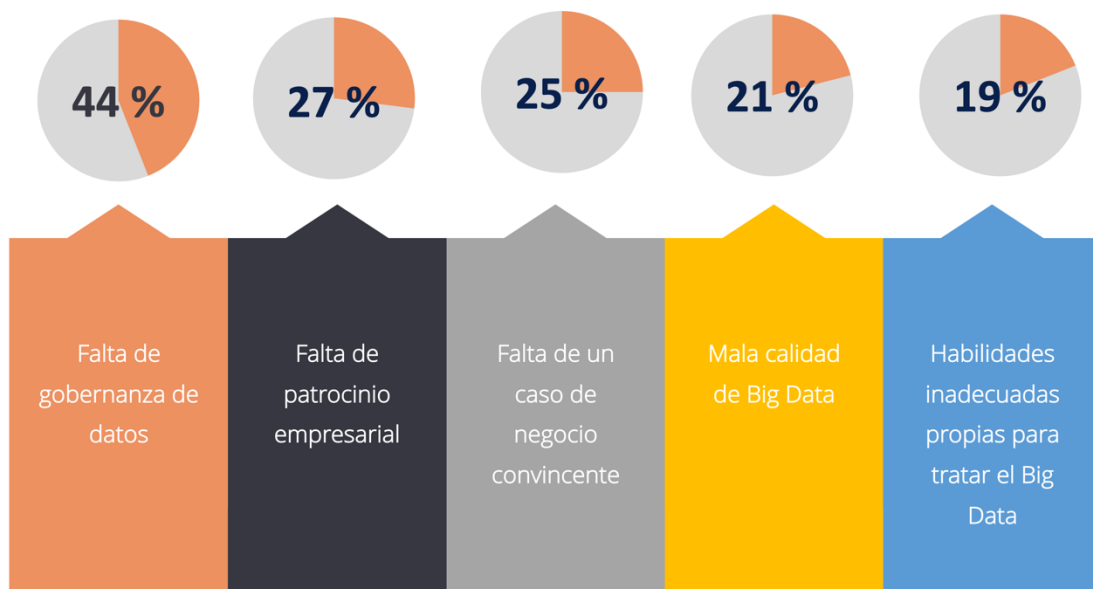
7. Principales motivos de fallas de los proyectos de datos

Múltiples son los motivos de fracasos de los proyectos de Datos. Entre los más comunes, identificamos los siguientes:

- Falta de claridad de los objetivos: En lugar de halar teleológicamente desde una clara definición de objetivos, muchos proyectos se enfocan “empujando” actividad por actividad, imitando proyectos de otras empresas.
- Tratamiento sólo como un proyecto de Tecnología o Sistemas de la Información: La potencia del BDA se desvanece cuando el mismo no está identificado, enlazado y desarrollado con el modelo del negocio. En nuestra experiencia personal hemos encontrado muchos enlaces vinculantes con el proyecto de datos en sí mismos y no con la estrategia de la empresa.
- Falta de involucramiento de la Dirección: Si debemos partir de los objetivos rectores de la compañía, la Dirección debe estar involucrado en dicha definición primaria.
- Procesos de desarrollo muy extendidos en el tiempo, sin hitos intermedios que le den sustentabilidad: La búsqueda de soluciones ideales origina un efecto negativo para la “paciencia organizacional.
- Foco en la historia: La orientación sólo hacia un medición y procesamiento de información histórica para el control ex post limita los beneficios de la proyección de futuro y anticipación de problemáticas propias de un escenario VUCA (volátil, incierto, complejo y ambiguo):
- No se aprovecha el proyecto para desarrollo de consenso ni generación de trabajo en equipo.
- Ausencia de cambios culturales asociados: Los aportes de valor del BDA deben acompañarse por una actitud de humildad de los expertos de negocio en el reconocimiento de relaciones y patrones que no eran de su conocimiento.
- Confusión de los distintos conceptos y sus relaciones jerárquicas: A los ya desarrollados en el presente trabajo se debe incluir la paradójica inversión de relevancia entre los conceptos Data e Información. A diferencia de la pirámide del

conocimiento citada en el capítulo 4, en esta década se fueron creando áreas organizacionales de *Data* separadas y con poco vínculo con las áreas de Información. En muchas instituciones los Gerentes de *Data* (CDO) son pares de los Gerentes de Informática (CIO)

Para dar una idea de dimensiones, en el cuadro #5 seleccionamos los cinco ítems más significativos de un estudio realizado sobre 150 empresas en el 2021, respondiendo por las barreras más representativas de un tipo de proyectos de *Data* orientados al sustento de modelos Directivos:



Cuadro 5 – Barreras más representativas para lograr implementar un lago de datos que se complemente con los Datawarehouses existentes en las organizaciones – (Elaboración propia sobre datos extraídos desde: Best Practices Report 2021 – Q2 Research)

8. El rol del profesional de gestión y costos. Ser “El arquitecto de nuestra casa en el lago”

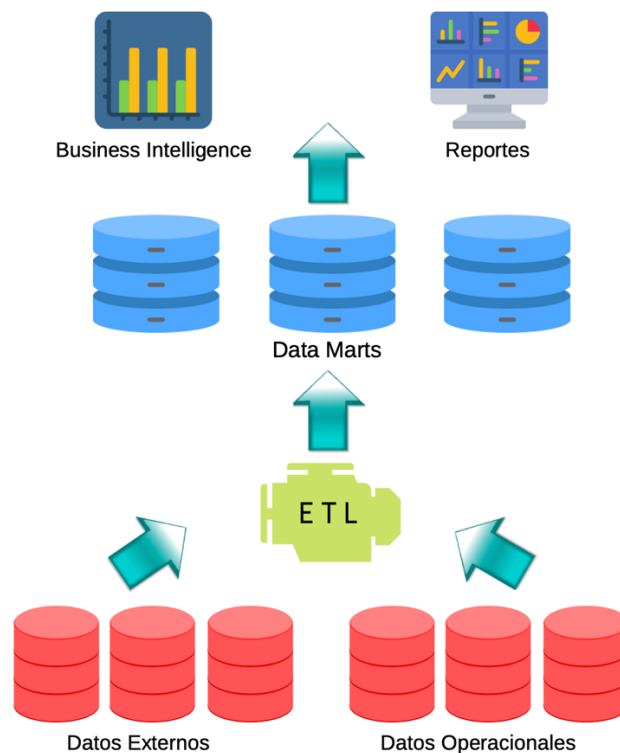
¿Cuál es el rol de los profesionales en ciencias económicas en el trabajo de los datos? Desde las empresas abocadas al BDA, cuando se definen perfiles de profesionales para analizar datos, se da preferencia a aquellos que vienen de los “negocios” y que combinen esos conocimientos con práctica en herramientas de análisis de datos. Conocimiento básico que le permita saber la utilidad de esas herramientas y poder solicitar al área correspondiente su aplicación, explicando el caso para el cual sería desarrollada.

Para poder encuadrar este conocimiento básico, se hace necesario mostrar la evolución de estas tecnologías en el último medio siglo. Hacia los últimos meses de la década del 80, la tecnología nos brindó un salto evolutivo trascendental. La creación de los Datawarehouses y DataMarts nos permitió cubrir necesidades hasta ese momento muy poco asistidas. La estructuración de la información para fines directivos de inteligencia de negocios, separada de las plataformas operativos. De esta forma se pudo obtener:

- La visión multidimensional de la información.

- La incorporación de información no monetaria, fundamental para interpretar la información de costos y desarrollar valoraciones adecuadas vinculadas con la eficacia en las decisiones y la eficiencia en la ejecución de las mismas.
- La inclusión de fuentes “externas” y no sólo “de nuestro ombligo”.
- La mejora en performance tanto en los procesamientos de información multidimensional como en la no injerencia de la baja de performance de los sistemas operativos cuando ejecutábamos nuestros procesos de agregación de la información compartida.

Técnicamente estas plataformas requerían de una carga de información que significaban cierta duplicidad de la misma, pero con otra forma de almacenaje y posterior acceso. En estos tiempos aprendimos a trabajar colaborativamente con profesiones de IT para determinar de antemano la arquitectura de los datos con fuerte orientación teleológica, definiendo previamente los objetivos y frecuencias de uso que posibilitaron distintos niveles de almacenamiento para mejorar el tiempo de respuesta. En estos tiempos tomamos contacto con los almacenamientos originales y preagregados, como así también con las estructuras multidimensionales conocidas como “cubos”.



Cuadro 5 : Tecnología de los *Datawarehouses* y de los *DataMarts* (Finales de la década del 80`) – (Elaboración propia)

Por otra parte, participamos activamente en los procesos de aseguramiento de la calidad y consistencia de la información que se “subía” a las nuevas plataformas para mitigar el efecto dañino de distintos resultados para las mismas preguntas y cuestionamientos, efecto provocado por la falta de holisticidad, sincronismo e integración de los sistemas transaccionales y de la propia contabilidad financiera.

Lamentablemente, el alto costo de estas plataformas operó como una fuerte restricción ante los requerimientos de mayor volumen y diversidad de datos.

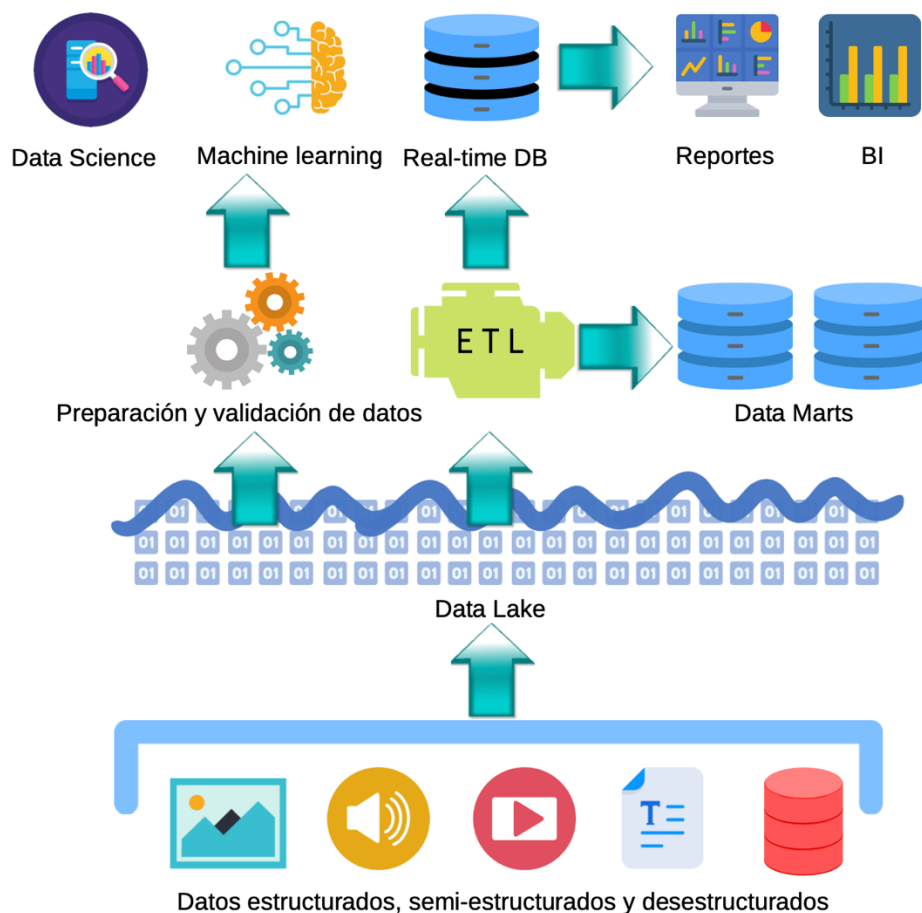
A partir del año 2011 un nuevo salto tecnológico y cultural nos benefició, sobre todo en la apertura de oportunidades de los profesionales de los costos y la gestión para involucrarse y encontrar un campo de acción diferencial y muy bien valorizado por las organizaciones que demandan este tipo de servicios.

La creación de los “*Data Lakes*” (lagos de datos), combinándose con múltiples fenómenos concomitantes:

- El Cloud Computing
- El cambio actitudinal del concepto Wiki 2.0
- El cambio cultural de la digitalización, provocados mayoritariamente por la incorporación de las redes sociales y los hábitos digitales en nuestro quehacer diario.
- La IOT (Internet de las cosas)
- La evolución de la Ciencia de Datos

A diferencia de la plataforma anterior, la ingesta de información tiene una filosofía radicalmente diferente. La misma se carga en un primer paso en forma de “Raw data” y será el aporte de la Ciencia de Datos el que encontrará la riqueza “oculta” de los datos y sus relaciones entre sí. El enfoque “pull” teleológico debe dar paso a un enfoque “push”, a la espera de la depuración. Este avance tecnológico nos quita, a la vez, restricciones a nuestros requerimientos por el bajo en los costos, y nos plantea un fuerte desafío en nuestro rol.

En el primer paso de recopilación y registro, que en épocas de Fray Luca Pacioli fuera el diseño de técnicas de registro manual controlado y estructurado en transacciones monetarias; y en tiempos del DW (Datawarehouse) fue la recopilación controlada y estructurada de información multidimensional, es reemplazado por nuestra ausencia de procesos de ingestas masivas de información sin filtros en cuanto a juicios de valor, y su conservación desestructurada en “lagos de datos”. La información alojada en estos “Lakehouses” agregarán valor una vez que la minería de datos nos revele las relaciones y patrones ocultos existentes entre la diversidad de elementos dispersos y escondidos en estos lagos de información.

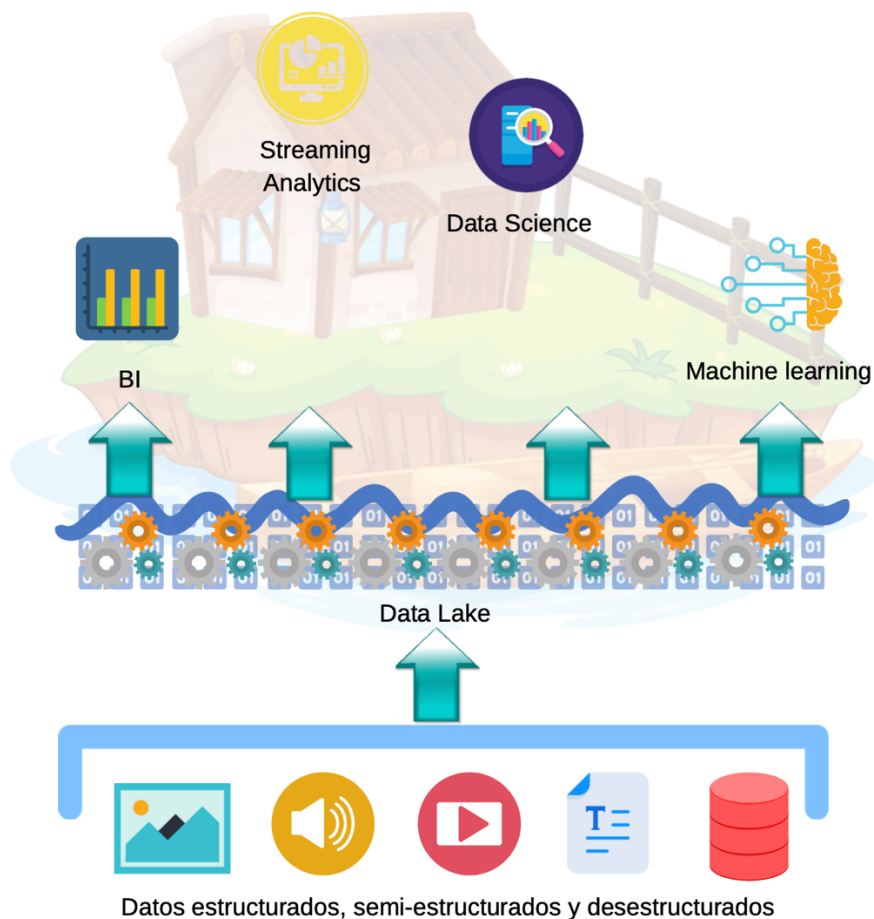


Cuadro 6: La tecnología de los “Data Lakes” – A partir de año 2011 – (Elaboración propia)

A partir del 2020, se comienza a gestar las tecnologías aplicadas al concepto de “Lakehouses”. El objetivo es que todas las relaciones y patrones existentes entre todos los elementos diversos y ocultos en los lagos de datos puedan ser explicadas desde el punto de vista del negocio. Una vez conocidas aprovechar al máximo el potencial de las plataformas y aplicaciones que nos dan ventajas a la hora de modelizar la receptividad de la información dentro de un receptáculo modelizado de acuerdo a las características de cada modelo de negocio en particular. A su vez nos permitirán reducir los costos de la información, controles de calidad, taxonomías y ordenamientos teleológicos en función de los objetivos estratégicos del negocio a los efectos de poder mitigar los dos síndromes ya planteados en trabajos e investigaciones de la especialidad (autores omitidos):

- El síndrome de fatiga por información, originado por la sobrecarga de datos no focalizados en los objetivos estratégicos de los negocios.
- El síndrome de la erosión de la información, manifestado en la pérdida de cantidad, calidad y multidimensionalidad de los datos durante el ciclo que va desde su captación a su mutación de la información, y finalmente la generación de conocimiento para la toma de decisiones y la conducción empresarial.

Será necesario entonces, asumir el rol de ser “los arquitectos de nuestra casa en el lago”



Cuadro 7: La tecnología del “Lakehouses” – Año 2021 – (Elaboración propia)

Desde el punto de vista de la contabilidad directiva, las “casas del lago” deben orientarse al cumplimiento de las siguientes características:

- Integridad y fiabilidad de la información
- Facilitación de la visión del negocio
- Pulsión de procesos de toma de decisiones
 - Con procesos “Data Driven” automatizados de punta a punta, y disparados desde las reglas del negocio que se autoaprenden a través del “machine learning”
 - Con sugerencias de “momentum” y opciones seleccionadas para aportar a la discrecionalidad del directivo en aquellas decisiones de mayor complejidad y/o sensibilidad
- Claridad en la comunicación de objetivos para asegurar ejecuciones alineadas a las decisiones.
- Disponibilidad inmediata de alertas ante la inminencia de desvíos en la ejecución
- Eficiencia en la productividad en cuanto a los costos vinculados con el almacenamiento, clasificación, procesamiento, exposición y difusión de la información.

Finalmente, podríamos concluir que dentro del universo del BDA, los profesionales de la Contabilidad de Gestión podemos aportar valor en tres procesos directivos distintos:

a) Entendimiento del negocio: Desde sus inicios como ciencia aplicada, la Contabilidad aplicó metodologías de hermenéutica que permiten la conceptualización contextualizada, siguiendo el proceso triádico mimético aristotélico a saber:

- Observación sistémica, identificando los elementos esenciales de una realidad económica y su registro estructurado.
- Taxonomía, ordenamiento, cálculo e identificación de relaciones y consistencia entre los datos recolectados, ascendiendo la jerarquía DIKW hacia el nivel de información
- Identificación de patrones y reinterpretación en la concepción del “*Story telling*” para lograr la difusión del conocimiento adquirido hacia los usuarios del mismo

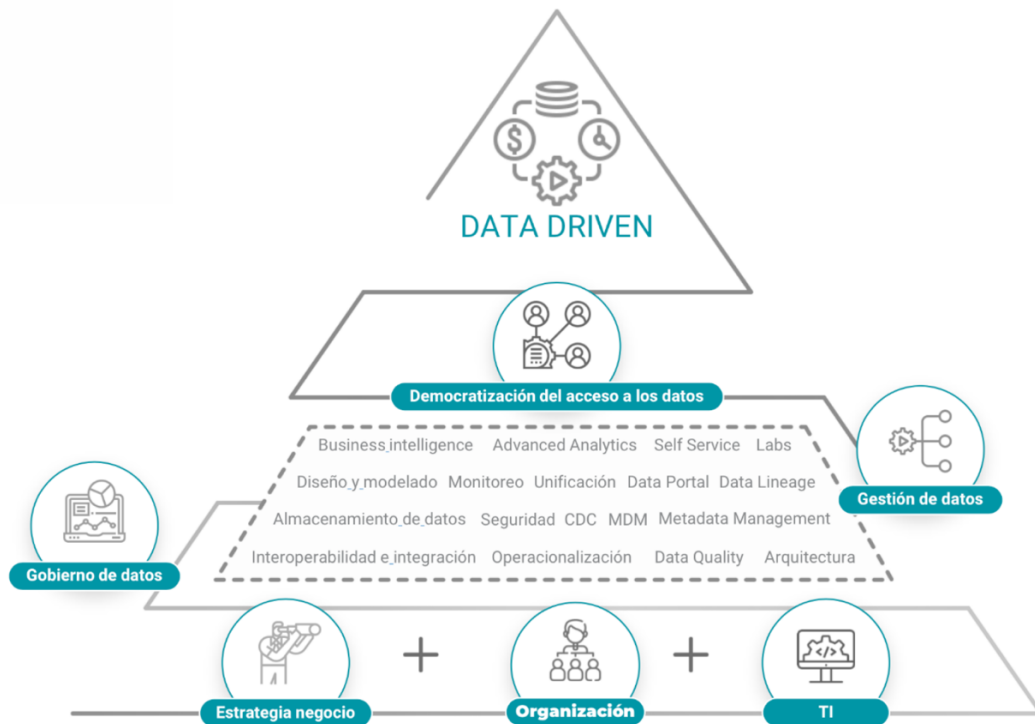
Lo expuesto ut supra plantea un fuerte desafío en nuestro método: El primer paso de recopilación por registro manual controlado y estructurado es reemplazado por ingestas masivas de información sin filtros de juicio de valor y conservación desestructurada en Lagos de Data.

La taxonomía y ordenamiento será apalancado por métodos de segmentación (con y sin ayuda de hipótesis preliminares).

La identificación de patrones tendrá que ser sustentado con herramientas de machine Learning

b) Toma de Decisiones: se hace trascendente la comprensión del enfoque planteado por el “Data Driven Process”, en el cual nosotros ayudamos a definir las reglas de negocio que disparan las decisiones automatizadas. Los procesos de toma de decisiones en la actualidad se ven impulsados por los datos en combinación con las distintas aplicaciones de análisis, seguridad, unificación, monitoreo, diseño y modelado, comprobaciones de calidad, entre otros. Para poder actuar adecuadamente en este entorno se hace necesaria nuestra intervención como consultores y asistentes a las organizaciones en temas claves que favorecen que los procesos sean exitosos:

- Estrategia del negocio
- Organización basada en los procesos y actividades
- Adopción de tecnologías de información adecuadas
- Gestión de datos
- Gobierno de datos
- Democratización del acceso a los datos



Cuadro 8 – El enfoque Data Driven – (Elaboración propia)

- c) Conducción: Desarrolla el concepto de Asistencia al alineamiento entre la ejecución y los objetivos definidos por dos grandes caminos:
- La claridad de la comunicación de los objetivos.
 - La diagramación de alertas inmediatas ante la inminencia de desvíos entre la ejecución y los objetivos en aplicativos de reglas de negocio complementados con IA / machine learning.

9. Conclusiones finales

De lo desarrollado en el presente trabajo, y en consonancia con los objetivos propuestos, se pueden enunciar las siguientes conclusiones a considerar:

- Es frecuente confundir los conceptos de Big Data, Ciencia de Datos, Inteligencia Artificial y Aprendizaje automático. Podríamos afirmar que la potencialidad de la Big Data, la Inteligencia artificial con sus ramas o subconjuntos, la ciencia de datos, el aprendizaje automático y el aprendizaje profundo, en su combinación, nos brindan la posibilidad de agregar valor a la información y provocar que sea útil para su utilización en procesos directivos críticos. En síntesis, el BDA ofrece posibilidades basadas en la combinación del Big Data (datos en volumen), la ciencia de datos, la inteligencia artificial y el aprendizaje automático.
- Sobre los artículos desarrollados por especialistas sobre las “V” de la Big Data y la Ciencia de Datos, podemos concluir que: muchas de ellas, son repetitivas y abarcativas, y probablemente distraigan y confundan al lector, en lugar de clarificar. Ya el sólo hecho de enunciarlas como características del “Big Data and Data Science” compromete la comprensión de ambos conceptos. En el punto 2 del

presente se replantean las diez “V” de la Big Data y de la Ciencia de datos, separándolas en las características propias de cada una de ellas.

- Para poder relacionar adecuadamente estas tecnologías con la jerarquía DIKW se hace necesario entender las relaciones para poder pasar del dato a la información, comprender los patrones para poder convertir la información en conocimiento y significado; y finalmente comprender estos principios para pasar del conocimiento a la sabiduría. La ciencia de datos nos permite develar relaciones ocultas que, luego de nuestro proceso de curaduría del contenido, dota de agregado de valor al conocimiento organizacional para los procesos de conducción y de toma de decisiones.
- Para poder desarrollar estos proyectos de datos y no caer en fracasos o esfuerzos sin los resultados esperados, se hace necesario tener en cuenta los pasos críticos para el desarrollo de un modelo BDA. Orientación, planificación, diseño y arquitectura, modelización de datos, definir los modelos de gobernanza de datos y la evaluación continua de estas modelizaciones son vitales para poder “arrancar con el pie derecho” en el desarrollo e implementación de estas herramientas.
- Sin embargo, muchos de ellos fracasan. La falta de claridad de los objetivos, el dar prioridad a la tecnología por sobre el modelo de negocio, la falta de involucramiento de la gerencia y la ausencia de cambios culturales en la forma de pensar la información son algunas de las causales por las cuales no se cumplen los objetivos al desarrollar estos proyectos.
- Nuestro ensayo identifica un campo de acción importantísimo para la actuación profesional, en un terreno en que la demanda es creciente y no existe la oferta adecuada para satisfacerla. Así como la aparición de los *Datawarehouses* y *DataMarts* nos abrió un nuevo panorama hace algo más de tres décadas, la aparición de la Big Data y los Lagos de Datos hace una, combinado con el comienzo de la era de las “Lakehouses” y del efecto *Datadriven* vuelven a presentarnos desafíos profesionales de innovación y creatividad. Es sobre estas plataformas donde los profesionales de la Información de Gestión podemos aportar mayor valor en el entendimiento de la información y el asesoramiento en los procesos de toma de decisiones y conducción estratégica de las organizaciones.
- La “casa del lago”, debería constituirse en el un espacio receptor de la información curada de los lagos de datos, cuyos cimientos serán el *data Science*, el *machine learning* y el *business Intelligence*; de manera que podamos, en este ámbito, y asistidos con el *streaming analytics* poder conducir nuestro negocio. Será necesario entonces, asumir el rol de ser los arquitectos de casas en el lago para las organizaciones del futuro.

10. Bibliografía

Chung,J (2018). ¿Por qué cada empresa tiene un gran problema de Big Data?. Análisis & Opinión. <https://www.americaeconomia.com/analisis-opinion/por-que-cada-empresa-tiene-un-gran-problema-de-big-data>

Halper, F (2021). Building the Unified Data Warehouse and Data Lake – TDWI Research – A Division of 1105 Media.

Laney, D. (2001). Tres V del Big Data. The Gartner group.

Lewis, D (1996) – Dying for information? – A report on the effects of information overload In the UK and Worldwide - Reuters

Oñate, J. (2020). Los 11 elementos de una solución Big Data. Ediciones PowerData.

Oracle. (2020). ¿Qué es la ciencia de datos?. <https://www.oracle.com/ar/data-science/what-is-data-science/>

Natali, C. (1999). Problemas de la noción de causa final en Aristóteles”. Anuario Filosófico

Rodriguez, M.I., Maffioli, M., de Rossetti, R. (2021). Utilización de la “Ciencia de datos” para obtener conocimientos claves en el desarrollo de modelos de negocios. Ciclo de Conversatorios 2021 – Instituto Argentino de Profesores Universitarios de Costos. https://www.youtube.com/watch?v=pP2pg_MLqZ0

SAS Institute Inc. (2019). Big Data, Data Mining and Machine Learning. Value Creation for Business Leaders and Practitioners Jared Dean. Editorial Willey.

Shafer, T. (2017). The 42 V's of Big Data and Data Science. <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>

Se omite la Bibliografía de los autores