

Emiliano Consoli

GESTIÓN DE DATOS EN ORGANIZACIONES

CENTRO DE INVESTIGACIONES
EN CIENCIAS ECONÓMICAS (CICE)

GESTIÓN DE DATOS EN ORGANIZACIONES

Lic. Emiliano Consoli

Consoli, Emiliano

Gestión de datos en organizaciones / Emiliano Consoli. - 1a ed. - Ciudad Autónoma de Buenos Aires : Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires, 2022.

Libro digital, PDF

Archivo Digital: descarga

ISBN 978-987-660-352-2

1. Sistemas de Gestión. 2. Sistemas de Gestión de Bases de Datos. I. Título.

CDD 658.4038

Hecho el depósito que marca la Ley 11.723.

Prohibida su reproducción total o parcial por cualquier medio sin autorización previa del CPCECABA.

EDICON

Fondo Editorial Consejo

Consejo Profesional de Ciencias Económicas de la

Ciudad Autónoma de Buenos Aires

Viamonte 1549 - CABA

Tel. 5382-9200

www.consejo.org.ar

www.edicon.org.ar

Tabla de contenido

Gestión de datos en organizaciones	5
1. Introducción	5
2. Transformación digital	6
2.1 <i>Big Data</i>	7
2.2 Modelo de negocios basados en plataformas	8
3. Roles que intervienen en un proyecto de <i>Big Data</i>	8
3.1 Roles directivos	9
3.1.1 <i>Chief Data Officer</i> (CDO)	9
3.1.2 <i>Chief Analytics Officer</i> (CAO)	9
3.2 Roles que conforman equipos de <i>Big Data</i>	9
3.2.1 Arquitecto <i>Big Data</i>	9
3.2.2 Ingeniero <i>Big Data</i>	9
3.2.3 Científico de datos	10
3.2.4 Consultor de negocios <i>Big Data</i>	10
4. Gobierno del dato	11
4.1 Marcos de trabajo	12
4.1.1 COBIT	12
4.1.2 Normas ISO 38500	12
5. Arquitectura empresarial	13
5.1 Marcos de trabajo	13
5.1.1 TOGAF	13
5.1.2 Zachman	13
6. Arquitectura de datos	13
7. Ingesta de datos	14
7.1 Origen de datos	15
7.1.1 Datos internos	15
7.1.2 Datos externos	15
7.1.3 Datos alternativos	15
7.2 Tipos de datos	15
7.2.1 Datos estructurados	15
7.2.2 Datos semiestructurados	15
7.2.3 Datos no estructurados	16
7.3 Fuentes de datos	16

7.4 Periodicidad de ingesta	17
7.4.1 Ingesta única	17
7.4.2 Ingesta periódica	17
7.4.3 Ingesta en tiempo real	17
8. Almacenamiento	18
8.1 Arquitectura de almacenamiento	18
8.1.1 Data Warehouse	18
8.1.2 Data Lake	18
8.1.3 Lakehouse	19
8.2 Bases de datos	19
8.2.1 Bases de datos relacionales	19
8.2.2 Bases de datos no relacionales	19
8.3 Calidad del dato	20
9. Procesamiento de datos	20
9.1 Etapas para el análisis de datos	20
9.1.1 Importar	20
9.1.2 Ordenar y limpiar	21
9.1.3 Transformar	21
9.1.4 Visualizar	21
9.1.5 Modelar	22
9.1.6 Comunicar	22
9.2 Metodologías de análisis de datos más utilizadas	22
9.2.1 Métodos multivariados	22
9.2.2 Investigación operativa	23
9.2.3 Métodos de análisis predictivo	23
9.2.4 Aprendizaje automático	23
9.2.5 Procesamiento del lenguaje natural	23
10. Explotación	24
10.1 Reporte	24
10.2 Utilización en procesos	24
11. A modo de cierre	25
Referencias bibliográficas	27

Gestión de datos en organizaciones¹

Lic. Emiliano Consoli

1. Introducción

En un mundo cambiante donde las tecnologías digitales tomaron protagonismo en el último tiempo, se están produciendo cambios en todos los ámbitos de la vida. Organizaciones de todo tipo son percutidas por la transformación digital, provocando no solo el surgimiento de nuevos modelos de negocios o nuevas maneras de trabajar, sino cambios generalizados en las estrategias de todas las organizaciones hacia la extracción y explotación de datos como método privilegiado de crear valor.

Este trabajo plantea el objetivo de facilitar la transición de las organizaciones hacia la transformación digital introduciendo a los profesionales en Ciencias Económicas en la gestión de datos en contextos organizacionales. Para ello, se recorren los principales tópicos relacionados buscando esclarecer los elementos conceptuales necesarios para tener una buena preparación que permita acompañar este cambio inminente en las organizaciones.

Como es esperable, se comienza estudiando el alcance del fenómeno de la transformación digital, las principales características de *Big Data*, y su consecuente impacto en las organizaciones en relación con los modelos de negocios, cuya principal herramienta resultan ser las plataformas. Estas últimas permiten mediar relaciones y recopilar datos que, procesados, harán posible obtener valor agregado.

Como paso siguiente se presentan los roles surgidos en el contexto separándolos entre directivos y quienes específicamente cubren funciones en el procesamiento y análisis de datos. Para los primeros se consideran el *Chief Data Officer* (CDO) y el *Chief Analytics Officer* (CAO). Para el segundo grupo se describen cuatro roles: arquitecto *Big Data*, ingeniero *Big Data*, científico de datos y consultor de negocios *Big Data*.

A continuación, se conceptualiza el gobierno del dato, lo que permite establecer criterios unificados acerca de cómo serán tratados dentro de la organización. De manera complementaria se describen los marcos de trabajo más utilizados: COBIT y Normas ISO 38500.

De la misma manera, se exponen las principales características de la arquitectura empresarial que busca alinear la estrategia con la infraestructura tecnológica. En este apartado también se introduce en los principales marcos de trabajo: TOGAF y Zachman.

Como dimensión de la arquitectura empresarial, a los fines de este trabajo, se destaca la arquitectura de datos. Su implementación busca establecer estándares en el manejo de datos que permitan alinear su uso con la estrategia organizacional. Por ello, se presentan los distintos elementos que la

¹ Este trabajo es el resultado del proyecto de investigación “Roles y procedimientos de proyectos de *Big Data* en organizaciones argentinas. Un enfoque desde la administración”, del Centro de Investigaciones en Ciencias Económicas (CICE) del Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires.

componen, lo que da lugar a los apartados siguientes, donde se describen las distintas fases asociadas a proyectos de *Big Data*.

Comenzando por la ingesta de datos, se exhiben los componentes más importantes para la primera fase de la gestión de datos: la captura. Pasando por el origen, tipos y fuentes de datos, se finaliza precisando los aspectos a considerar para precisar la periodicidad de ingesta.

El almacenamiento implica conocer la actualidad y la evolución reciente en torno a sus arquitecturas. Por tal motivo, se explican tres lógicas de almacenamiento que se desarrollaron en los últimos años al calor de la transformación digital: *Datawarehouse*, *Data Lake* y *Lakehouse*. Adicionalmente, se define y diferencia entre bases de datos relacionales y no relacionales. Por último, se precisan los objetivos de calidad de datos.

A continuación, se da paso a la fase de procesamiento de datos. Se enumeran las distintas etapas de este proceso desde el punto de vista de la ciencia de datos. De la misma manera, se enumeran los principales métodos analíticos utilizados para extraer valor del dato.

Por último, se incluye la fase de explotación, dado que, si no se logra aprendizaje organizacional efectivo, de nada servirá haber analizado los datos. Es decir que en esta fase, finalmente, se crea valor a partir de los datos.

En el plano organizacional, cada una de las nociones mencionadas implica tomar decisiones acerca de cómo gestionar los datos. Se espera que el recorrido propuesto permita a profesionales de Ciencias Económicas comprender la naturaleza de la transformación digital en el plano organizacional para luego visualizar en conjunto el proceso de gestión de datos con sus actores específicos dentro de la organización. Con un panorama completo será sencillo para cada lector o lectora encontrar el camino para profundizar cada concepto.

2. Transformación digital

Aunque el término es ampliamente utilizado para referir al profundo cambio que estamos atravesando a nivel mundial por el avance de distintas tecnologías digitales, también es cierto que prima cierta ausencia de claridad conceptual en su uso. En el campo académico, existen revisiones de la literatura que concluyen que aún no se ha llegado a consolidar un marco teórico común que permita una operacionalización compartida para el tratamiento del fenómeno (Henriette *et al.* 2015; Morakanyane *et al.*, 2017; Halpern y Valderrama, 2018).

De la misma manera, tampoco hay un criterio unificado para citarlo. Prueba de ello son los variados términos utilizados para tratar de describir este contexto de cambio y tendencia hacia la digitalización de los distintos espacios sociales y organizacionales: cuarta revolución digital, economía digital, economía de plataformas, economía colaborativa, organizaciones exponenciales, economía de vigilancia, industria 4.0, data tsunami, entre otros (Ismail *et al.*, 2014; Srnicek, 2018; Lassalle, 2019).

A fin de definir el criterio del presente trabajo, se sigue el recorrido propuesto por Halpern y Valderrama (2018) para definir la transformación digital. Para ello es necesario distinguir entre los términos “digitación”, “digitalización” y “transformación digital”.

Desde este punto de vista, al mencionar la digitación, se hace referencia a la conversión de información originalmente en formato analógico a información digital, es decir, dígitos binarios legibles por computadoras. Esto implica el pasaje de información desde el plano físico al plano digital.

Este proceso reporta una serie de beneficios incluyendo la posibilidad de almacenar, conservar, replicar, transmitir información a gran escala. Luego se podrá procesar, filtrar o visualizar para generar conocimiento con la particularidad de que, a diferencia de otros soportes, la información digitalizada no sufre desgaste por su uso ni degradación por el paso del tiempo.

La aparición y difusión de Internet de la nube, que implica la existencia de servidores deslocalizados en todo el mundo, llevaron a que la transmisión y el almacenamiento de datos tengan costo marginal decreciente cada vez más insignificante con relación a la potencialidad y volumen de los datos; además, resultan procesos casi instantáneos (Knoll y Viola, 2018; Ismail *et al.*, 2014).

La penetración de estas tecnologías en la sociedad generó un profundo cambio tanto en la conducta de personas y organizaciones como en su manera de relacionarse. Así se llega a que la digitalización de la sociedad va más allá de la simple conversión de información analógica hacia información digital, sino que implica un cambio de hábitos generalizado que, incluso, implica el desarrollo de nuevas habilidades en toda la sociedad (Halpern y Valderrama, 2018).

El fenómeno de la transformación digital, que en la actualidad estamos atravesando en todo el mundo, dio lugar a la aparición de “nuevas tecnologías y formas novedosas de percibir el mundo que desencadenaron un cambio profundo en los sistemas económicos y las estructuras sociales” (Schwab, 2016, p.12).

Atendiendo específicamente al plano organizacional, en una primera etapa, las grandes organizaciones han comenzado por la modificación de los procesos internos y el modo de trabajo en el marco de una transformación gradual y constante que buscaba mejorar la eficiencia productiva. En los últimos años, la digitalización se generalizó y penetró decididamente tanto en los procesos y en el interior de las organizaciones transversalmente como en las redes comerciales propias de la cadena de valor agregado, desde el aprovisionamiento de materias primas hasta la relación con clientes. Ello provocó el surgimiento de nuevos productos, nuevos modelos de negocios y, sobre todo, una manera distinta de vincularse con el entorno (Schreckling y Steiger, 2017).

Frente a este nuevo contexto social, económico y cultural, la digitalización ya dejó de ser una herramienta útil para el aumento de la productividad para transformarse en una necesidad de supervivencia de las organizaciones. Es decir que las organizaciones, en su caracterización como fenómenos adaptativos, deberán modificar su operatoria y su lógica de funcionamiento para lograr sobrevivir al entorno. Adicionalmente, resulta vital que las organizaciones gestionen el cambio de manera controlada y alineado a su estrategia (Halpern y Valderrama, 2018).

Entonces, la transformación digital implica la integración digital transversal en todas las áreas de las organizaciones, modificando drásticamente su operatoria y la conducta de sus miembros hacia escenarios donde las decisiones de todos los niveles (operativo, táctico y estratégico) son guiadas y facilitadas por datos digitalizados y procesados (OCDE, 2019).

La transformación digital ya no se trata de una oportunidad sino de un imperativo. Su alcance nos lleva a concluir que “la pregunta para todas las industrias y empresas, sin excepción, ya no es ‘¿voy a experimentar alguna disrupción?’’, sino ‘¿cuándo llegará la disrupción, qué forma adoptará y cómo nos afectará a mí y a mi organización?’” (Schwab, 2016, p.17).

2.1 *Big Data*

En el recorrido propuesto, toma especial importancia el concepto de *Big Data*. La digitalización ha llevado a que la actividad humana sea mediada por tecnologías digitales acrecentando la capacidad de generación de datos en volúmenes sin precedentes, cuya acumulación aumenta a ritmo exponencial: datos cuantitativos y cualitativos; estructurados y no estructurados; productivos; comerciales y sociales. Este fenómeno, asociado a la disponibilidad de grandes volúmenes de datos, es conocido como *Big Data*. Acompañado por el incremento en la capacidad de procesamiento de las tecnologías actuales y el desarrollo de técnicas estadísticas acordes, *Big Data* permite analizar y procesar los datos al posibilitar y facilitar la creación de valor a partir de estos (Knoll y Viola, 2018).

En un esfuerzo por caracterizar el término, Douglas Laney (2001) conceptualizó las tres “V” de *Big Data*: (1) Volumen, refiriéndose a las cantidades masivas de datos; (2) Velocidad, referida al flujo y generación de información en tiempo real; (3) Variedad, distintos tipos de datos tales como cuantitativos, cualitativos, estructurados, no estructurados, semiestructurados.

Otras características se agregaron a la conceptualización conforme al desarrollo del fenómeno: (4) Variabilidad, en referencia a los flujos de datos con picos periódicos inconsistentes; (5) Complejidad, pues son dificultosas la vinculación, igualación, limpieza y transformación de datos (BID, 2018).

Este fenómeno, alimentado por Internet de la nube, tiene gran potencia analítica para contribuir a la comprensión acerca de la dinámica social y organizacional a un nivel elevado (Knoll y Viola,

2018). Incluso en el campo de la ciencia, de la generación de conocimiento y el desarrollo de teoría (especialmente en las ciencias sociales), se observa un salto cualitativo importante, dado que, hasta la actualidad, las muestras sobre las que se viene trabajando son generalmente pequeñas. Así, *Big Data* ofrece la posibilidad de trabajar con muestras casi exhaustivas y de mayor diversidad (Rudder, 2016).

2.2 Modelo de negocios basados en plataformas

Los modelos de negocios dominantes en general están girando de manera acelerada, y en todo el mundo, hacia estrategias centradas en la extracción, el procesamiento y el uso de los datos como materia prima (Srniczek, 2016).

Esto significa que existe una fuente natural de generación de datos y, además, existe una forma de extraerlos. Las propias relaciones e interacciones humanas son las generadoras de los datos en cuestión. La transformación digital nos lleva a que la comunicación y la actividad humana sean canalizadas por medio de las nuevas tecnologías, generando de esta manera una enorme facilidad, sin precedentes, para almacenar, grabar y, por ende, extraer datos que, al ser procesados, son fuente inagotable de conocimiento.

Ni las interacciones mencionadas, ni el uso de los datos son algo nuevo. Históricamente, el uso de los datos tuvo un rol destacado en el desarrollo de la economía, aunque la disponibilidad hacía que fuera un aspecto secundario. Los negocios estaban orientados hacia la escalada en la eficiencia de los procesos y la reducción de costos. El innovador aporte de las tecnologías, que en los últimos años se abarataron significativamente, redundan entonces en la disponibilidad de los datos (Srniczek, 2016).

Este contexto da lugar al surgimiento de nuevas industrias orientadas a la extracción y procesamiento de datos para ser utilizados como manera de optimizar los procesos de producción, aumentar el conocimiento acerca de las preferencias y conductas de los consumidores, extraer información útil para la generación de nuevos productos, vender servicios a anunciantes, entre otras cosas.

Los datos han llegado a servir para varias funciones capitalistas clave: educan y dan ventajas competitivas a los algoritmos; habilitan la coordinación y deslocalización de trabajadores; permiten la optimización y la flexibilidad de los procesos productivos; hacen posible la transformación de productos de bajo margen en servicios de alto margen; y el análisis de datos, en sí mismo generador de datos, en un círculo virtuoso (Srniczek, 2016, p.44).

La manera privilegiada de extracción de datos en la actualidad resulta ser la plataforma, entendida como la infraestructura digital que permite la interacción entre personas y grupos. Se posiciona, entonces, como intermediaria entre clientes, anunciantes, proveedores de servicios, productores, distribuidores, incluso objetos físicos.

Srniczek (2016) conceptualiza cinco tipos de plataformas: (1) plataformas publicitarias que apuntan a brindar un servicio gratuito para extraer datos y luego vender publicidad a anunciantes orientada sobre la base de los datos de los usuarios, como Facebook o Google; (2) plataformas de la nube, empresas propietarias de *hardware* y *software* con el fin de alquilarlo a medida a otras empresas, como Amazon Web Services o Mercado Libre; (3) plataformas industriales, orientadas a optimizar los procesos productivos tradicionales, tal como se propone el proyecto alemán: industria 4.0; (4) plataformas de productos que montan su negocio en torno a la transformación de un bien tradicional en un servicio, como Spotify, Netflix o Rolls Royce; y (5) plataformas austeras, cuya naturaleza radica en reducir al mínimo la propiedad de activos y los costos a través de transferencia y tercerización, como el modelo de negocios de Uber o Airbnb.

3. Roles que intervienen en un proyecto de *Big Data*

Como es esperable, la profundidad de la transformación digital y el rol central que ocupa el fenómeno de *Big Data* llevaron a la conformación de equipos de trabajo específicos que contribuyan al aprovechamiento del contexto hacia el interior de las organizaciones. A continuación, se presentan los principales perfiles necesarios para la implementación de proyectos de *Big Data*.

3.1 Roles directivos

3.1.1 Chief Data Officer (CDO)

El *Chief Data Officer* es el responsable de los datos a nivel organizacional. Es decir que se trata de un cargo directivo responsable de orientar a la organización hacia estrategias basadas en la toma de decisiones a partir de la lectura de datos. Será quien defina los principios de la gobernanza del dato y, por lo tanto, será quien coordine los criterios y las metodologías asociados.

Debe ser responsable por asegurar la calidad, trazabilidad y accesibilidad de los datos en la organización. Además, debe posicionar los datos dentro de la estrategia organizacional. No se trata de un perfil técnico, sino de un perfil con liderazgo y conocimiento del negocio y con visión estratégica que le permita impulsar y buscar nuevas iniciativas de negocios (Elfenbaum, 2019).

3.1.2 Chief Analytics Officer (CAO)

El *Chief Analytics Officer* es el responsable de la analítica a nivel organizacional. Al igual que el CDO, es un rol directivo, aunque su responsabilidad principal es generar valor estratégico a partir de la disponibilidad de datos y de las distintas iniciativas de analíticas en la organización.

Así, deberá liderar, coordinar y supervisar, con foco en la estrategia de la organización, los equipos analíticos de la organización. Es su función principal la conexión entre los equipos analíticos y los negocios. Tampoco se trata de un perfil técnico, sino de un perfil con liderazgo y conocimiento del negocio principalmente, aunque es importante que cuente con nociones de analítica de datos.

Ambos roles buscan motorizar transversalmente (a nivel directivo) la transformación digital de las organizaciones. Es habitual encontrar que las organizaciones tiendan a concentrar en un solo cargo ambos roles (Elfenbaum, 2019).

3.2 Roles que conforman equipos de *Big Data*

3.2.1 Arquitecto Big Data

Al referirnos al rol del arquitecto de datos, hablamos de un perfil cuya responsabilidad se centra en garantizar las bases en la organización para la gestión y análisis de datos. Es así como diseñará y explicitará los requerimientos necesarios para la ingesta, almacenamiento, procesamiento, visualización y explotación de los datos.

Para ello deberá diseñar la arquitectura de datos, que implica precisar las fuentes de datos (tanto internas como externas), especificar los métodos de ingesta de datos (incluyendo periodicidad), escoger el almacenamiento adecuado que permita la disponibilidad de los datos, definir las herramientas y los requerimientos analíticos que posibiliten el procesamiento de los datos, definir el modo de reportar los hallazgos. Para ello deberá seleccionar la infraestructura necesaria considerando el *hardware* y el *software*, que además deberá adecuarse a las necesidades de la organización.

Para el correcto desempeño, deberá tener conocimientos y habilidades relacionados con las bases de datos disponibles, la infraestructura de soporte, lenguajes de programación, sistemas operativos, explotación, visualización de datos y en uso de API (Interfaz de Programación de Aplicaciones) (Alamán, 2019).

3.2.2 Ingeniero Big Data

Por su similitud, se tiende a la confusión entre el rol del ingeniero de datos y el del arquitecto de datos. Incluso, es habitual encontrar organizaciones donde ambas funciones son ejercidas por una sola persona. Para diferenciar claramente los roles, es preciso mencionar que, mientras el arquitecto *Big Data* se ocupa de las tareas relacionadas con la plataforma de soporte para la implementación de *Big Data*, los ingenieros *Big Data* se dedican a la preparación de datos mediante el diseño y la optimización de procesos relacionados.

La principal responsabilidad del rol está asociada a la adaptación del modelo de datos a las políticas definidas según la gobernanza de datos determinada por la organización. Es así como se ocupa de las tareas ligadas a la definición de procedimientos que implican la determinación de reglas de seguridad, reglas de calidad, movimiento de datos entre bases de datos, incluyendo las transformaciones

necesarias, automatización de procesos de gestión de datos, modelado, carga y monitoreo de datos.

Deberá tener experiencia y conocimiento acerca de bases de datos, gestión de metadatos, implementación de procesos en ecosistemas de datos; lenguajes de programación (Alamán, 2019).

3.2.3 Científico de datos

Este perfil se encuentra dedicado a la extracción de valor de los datos a partir de la aplicación de analítica avanzada y de presentarlos de manera que la organización pueda aprovechar los hallazgos para guiar la toma de decisiones.

Es decir que su responsabilidad es extraer conocimientos de los datos a partir de la detección de relaciones ocultas y patrones en los datos de manera que se genere valor para los procesos y productos de una organización. En su labor, analiza datos provenientes de fuentes internas y externas. Para ello se sirve de herramientas de analítica avanzada, potenciada en los últimos años por los avances de la capacidad de cómputo de la informática (Alamán, 2019).

Este perfil deberá contar con conocimientos avanzados en estadística y matemática. Deberá tener habilidades de programación; capacidad de comprensión del valor del dato; conocimiento sobre el proceso de transformación y limpieza de datos. Tendrá que manejar saberes acerca de métricas avanzadas vinculadas a la medición del rendimiento organizacional.

Adicionalmente, el perfil del científico de datos aglutina el rol de visualización. Es esperable que en el futuro la tarea tienda a separarse al comenzar a incorporar a los equipos perfiles especialistas en visualización de datos (Paliotta, 2018).

Así, deberíamos agregarle al perfil la responsabilidad de diseño e implementación de técnicas de visualización interactiva para que los usuarios puedan extraer información y conocimiento que permitan aportar valor a la organización. Esta tarea debe buscar la explotación autónoma de los datos por los usuarios de la organización.

Es necesario agregar al perfil conocimientos en diseño de herramientas de visualización, *dashboard* o tableros interactivos, informes y reportes. También, nociones sobre arquitectura web y conocimientos sobre desarrollo de API (Interfaz de Programación de Aplicaciones). Se sumarán habilidades de consulta de bases de datos (Paliotta, 2018).

3.2.4 Consultor de negocios Big Data

La principal preocupación del rol se centra en la alineación de la estrategia de la organización con las iniciativas de *Big Data*.

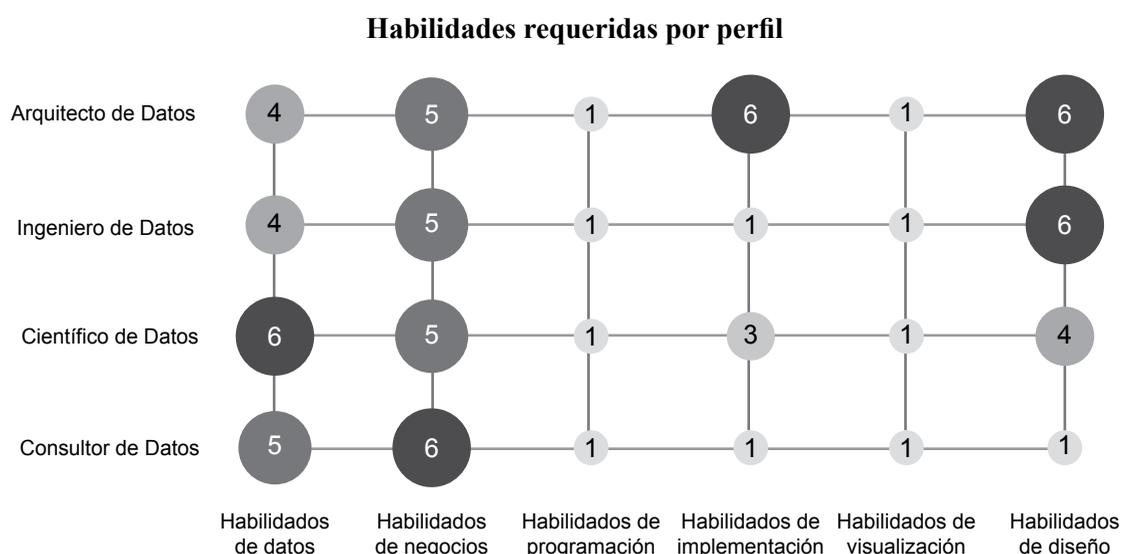
Su responsabilidad se basa en la identificación de necesidades de negocios susceptibles de ser cubiertas mediante la implementación de proyectos de analítica avanzada. De la misma manera, es responsable de identificar oportunidades de mejoras en los procesos y productos de la organización a partir del uso de datos (Alamán, 2019).

Por tratarse de un perfil transversal, será necesario que tenga nociones básicas de todas las disciplinas del equipo, aunque es necesario que el perfil cuente con visión estratégica que permita alinear las acciones con los objetivos de la organización. Es el perfil con mayores requerimientos de habilidades blandas; es decir, negociación, comunicación. Debe tener gran capacidad de adaptación dado el entorno cambiante. Y debe tener habilidades en la gestión y elaboración de proyectos. Es habitual encontrar que este rol es cubierto por profesionales en Ciencias Económicas (Consoli, 2019)

A partir de la descripción de cada uno de los perfiles se puede conseguir seis grupos de conocimientos necesarios para conformar equipos de *Big data*: (1) habilidades de datos, referidas a la habilidad de procesar y extraer conocimiento de datos; (2) habilidades de negocios, referidas a la capacidad de alinear los proyectos y procesos con los objetivos organizacionales; (3) habilidades de programación para agrupar todo conocimiento en relación con el armado y la lectura de códigos; (4) habilidades de implementación, que incluyen los conocimientos necesarios para implementar las arquitecturas de datos; (5) habilidades de visualización; capacidad de diseño de tableros visuales para facilitar al usuario de la organización la lectura y comprensión de los datos; (6) habilidades de diseño, referidas al diseño de arquitecturas de datos. Cada perfil necesita los seis grupos de habilidades,

aunque con distinta intensidad. La Figura 1 resume la importancia de cada conjunto de habilidades para cada perfil (Consoli, 2021).

Figura 1.



Nota. Su utilidad radica en representar la importancia de cada tipo de conocimiento necesario para el desarrollo de un proyecto de Big Data, y la conformación del equipo, discriminando cada uno de los perfiles que lo componen. La manera correcta de leer este gráfico es a partir de comprender que cada uno de estos perfiles necesita todas las habilidades mencionadas, aunque no con la misma intensidad. Fuente: Elaboración propia sobre la base de Consoli (2021).

4. Gobierno del dato

La comprensión de las organizaciones como fenómenos sociales complejos nos lleva a concluir que su gestión necesita mecanismos sofisticados que garanticen tanto la articulación entre todas las partes interesadas así como también el cumplimiento de sus objetivos.

En relación con el concepto de gobierno corporativo, se hace referencia al conjunto de normas, principios y procedimientos que regulan la estructura organizacional. El gobierno corporativo determina tanto la estrategia organizacional como la relación entre los distintos órganos de gestión estableciendo pautas que reglan y guían el proceso de toma de decisiones en sus distintos niveles de manera que la organización se oriente hacia la generación de valor.

Por lo tanto, se trata de la observación y teorización de los mecanismos e instituciones que permiten la coordinación entre las unidades organizacionales, lo que facilita el arbitraje de divergencias, contribuye a la cohesión del conjunto y aporta transparencia a la gestión (Hufty, 2008).

En el año 1999, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) ha aprobado los Principios de Gobierno Corporativo. Desde entonces se convirtieron en una alternativa normativa de aplicación internacional de gran influencia para responsables políticos, inversores, empresas, etc. De esta manera se ha marcado la agenda del gobierno corporativo, con orientaciones específicas para la puesta en marcha de iniciativas de carácter legislativo y reglamentario tanto en países de la OCDE como en otros no miembros (OCDE, 2004).

Las nuevas tecnologías agregan complejidad al entramado social y, por consiguiente, al organizacional. Por ello, en los últimos años, se han desarrollado, como parte del gobierno corporativo, mecanismos específicos orientados a la gobernanza de las tecnologías de la información y la comunicación.

Es posible definir al Gobierno de la Tecnología de la Información y la Comunicación como el proceso de administración o gestión que busca asegurar que la implementación de estas tecnologías esté alineada a la estrategia organizacional de manera que se logre la generación de valor y los beneficios esperados de manera controlada y sostenida en el tiempo (Piorun, 2016).

Por último, dentro del Gobierno de la Tecnología de la Información y la Comunicación es fundamental considerar el gobierno del dato. Siguiendo la misma lógica, se trata de un conjunto de procesos que buscan asegurar que los datos, considerados activos de la organización, sirvan a la estrategia organizacional y aporten el valor esperado. Es decir, se trata de la formalización de la gestión de datos en relación con su planificación, control y uso (Piorun, 2016).

Existen marcos de trabajo que apuntan al desarrollo de políticas y prácticas, cuya ejecución permite adquirir, controlar, proteger y mejorar el valor de los datos y de la información en las organizaciones. A continuación, se presentan los dos marcos de trabajo más difundidos relacionados con el gobierno del dato: el marco COBIT (*Control Objectives for Information and related Technology*) y las Normas ISO 38500.

4.1 Marcos de trabajo

4.1.1 COBIT

COBIT (*Control Objectives for Information and related Technology*) es un marco de trabajo creado por la organización ISACA y gestionado en la actualidad por el Instituto de Gobernanza de Tecnologías de la Información. Está ligado al Gobierno de la Tecnología de la Información y de los datos. Busca generar buenas prácticas mediante su implementación, basada en procesos y en la planificación estructurada de actividades, que se enfocan decididamente en el control y en menor medida en la ejecución.

La característica destacada para este marco de trabajo resulta ser su orientación hacia los negocios buscando vincular los objetivos organizacionales con las metas de incorporación de tecnologías de la información y la comunicación. De esta manera aporta métricas y modelos que permiten medir resultados e identificar y delimitar responsabilidades.

COBIT es, entonces, un modelo de referencia que establece 37 procesos relacionados con las tecnologías de la información y la comunicación. Cada uno de los cuales se describe en detalle indicando entradas, salidas, actividades, objetivos e indicadores de desempeño. Los procesos se distribuyen considerando un área global de gobierno corporativo (5 procesos) y cuatro áreas de responsabilidad de gestión: Planear (13 procesos), Construir (10 procesos), Ejecutar (6 procesos) y Monitorear (3 procesos) (ISACA, 2022).

4.1.2 Normas ISO 38500

Publicadas en junio de 2008 y basadas en la norma australiana AS8015:2005, se reconoce a las normas ISO 38500:2008 como las primeras normas internacionales ligadas al gobierno de la tecnología de la información para las organizaciones (Piorun, 2016).

Estas normas establecen estándares para el gobierno de procesos y para la toma de decisiones en relación con las tecnologías de la información y la comunicación proponiendo un modelo global que permita a la dirección hacer efectivo el gobierno de la tecnología de la información. Para ello define 6 principios: (1) Responsabilidad (determinar niveles de responsabilidad para todos los miembros de la organización); (2) Estrategia (considerando las capacidades tecnológicas actuales y futuras, se busca que los planes estratégicos de la tecnología de la información y comunicación satisfagan las necesidades de la estrategia organizacional); (3) Adquisición (busca el equilibrio entre beneficios, oportunidades, costos y riesgos al incorporar tecnologías); (4) Desempeño (las tecnologías deben proporcionar servicios de calidad adecuados para las necesidades estratégicas de la organización); (5) Conformidad (cumplimiento de normativas y legislaciones vigentes); y (6) Comportamiento humano (busca que la gestión de las tecnologías respeten el factor humano en la organización) (ISO, 2022).

Para completar el modelo, las normas ISO 38500 proponen conseguir la implementación de los principios a partir de la ejecución de tres actividades: (1) Evaluar; (2) Dirigir; (3) Monitorear.

5. Arquitectura empresarial

Como se ha explicado en el apartado 2, la transformación digital implica un importante cambio en las organizaciones que involucra una definición diferente de cómo se hacen las cosas. Conlleva un proceso inminente de traspaso de su operatoria desde el mundo físico hacia el mundo digital, lo que modifica fuertemente la experiencia organizacional cotidiana (Piorun, 2016).

La arquitectura empresarial plantea como objetivo la alineación de la infraestructura tecnológica y la estrategia de la organización. Se formula como un conjunto de principios, métodos y modelos que, utilizados en el diseño, elaboración y ejecución de los sistemas de información e infraestructura de una organización, permiten alinearlos a sus procesos de negocios. De esta manera se tiende a integrar la dirección de la organización en el proceso de innovación a partir de considerar cuál es la infraestructura disponible y cuál, la infraestructura deseada. En paralelo se busca que exista una visión única e integrada de quienes tienen a cargo la toma de decisiones.

La estructura empresarial considera cuatro dimensiones: (1) Arquitectura de Negocio; (2) Arquitectura de Aplicaciones; (3) Arquitectura de Datos; y (4) Arquitectura de Infraestructura (Piorun, 2016).

Al igual que en lo referente al gobierno del dato, existen marcos de trabajo más o menos generalizados en su uso que facilitan el proceso de diseño y gestión de la arquitectura empresarial. Se presentan a continuación dos de los marcos de trabajo más difundidos: TOGAF y Zachman.

5.1 Marcos de trabajo

5.1.1 TOGAF

TOGAF (*The Open Group Architecture Framework*) es un desarrollo de *The Open Group*, un consorcio internacional conformado por más de 450 organizaciones miembro. Se trata de un marco de trabajo que establece un enfoque para el diseño, planificación, implementación y gobierno de una arquitectura empresarial centrándose en la información.

TOGAF determina cuatro dimensiones a considerar para el diseño de una arquitectura empresarial: (1) Negocios; (2) Tecnologías; (3) Datos; y (4) Aplicaciones. A partir de ellas establece un conjunto de arquitecturas base que buscan facilitar el diseño a las organizaciones. A través de TOGAF es posible definir el estado actual de la infraestructura tecnológica y su relación con el negocio, así como también definir y planificar el futuro de la arquitectura empresarial (*The Open Group*, 2022).

5.1.2 Zachman

Creado por John A. Zachman en 1984, este marco de trabajo es utilizado para el diseño e implementación de arquitecturas empresariales. Este modelo establece dos grandes clasificaciones cuya intersección permite diseñar, explicar y visualizar arquitecturas empresariales.

La primera clasificación viene dada por una serie de preguntas básicas: (1) ¿Qué?; (2) ¿Cómo?; (3) ¿Dónde?; (4) ¿Quién?; (5) ¿Cuándo?; (6) ¿Por qué? Mientras que la segunda clasificación responde a seis visiones diferentes involucradas: (1) identificación del alcance, según la perspectiva de la dirección; (2) requerimientos, según la perspectiva del negocio; (3) diseño lógico, según la perspectiva del arquitecto; (4) diseño físico, según la perspectiva del ingeniero; (5) configuración, según la perspectiva del desarrollador; (6) implementación, según la perspectiva del producto.

Según indica el marco Zachman, el cruce de las respuestas a estas preguntas básicas y visiones involucradas puede explicar en forma simple ideas y entornos complejos, de manera que sea posible transitar desde una abstracción hacia una intervención concreta y tangible (Zachman, 2022).

6. Arquitectura de datos

Se plantea como objetivo de la arquitectura de datos ser el nexo que permita alinear la estrategia organizacional y el uso de los datos provenientes tanto de la ejecución y utilización de las tecnologías

de la información como de los datos surgidos de fuentes clásicas. De tal manera, la gestión de los datos se considera parte integrante de la gestión de la organización; por lo tanto, la arquitectura de datos se corresponde con una dimensión de la arquitectura empresarial (Piorun, 2016).

Específicamente, se ocupa de la definición y construcción de soluciones basadas en datos. Por ello, su función se centra en la disposición de elementos físicos, elementos lógicos y procedimientos en torno a la extracción, almacenamiento, procesamiento y explotación de los datos. Estos elementos deben estar orientados a mejorar la función de los datos y su rendimiento en torno a la creación de valor.

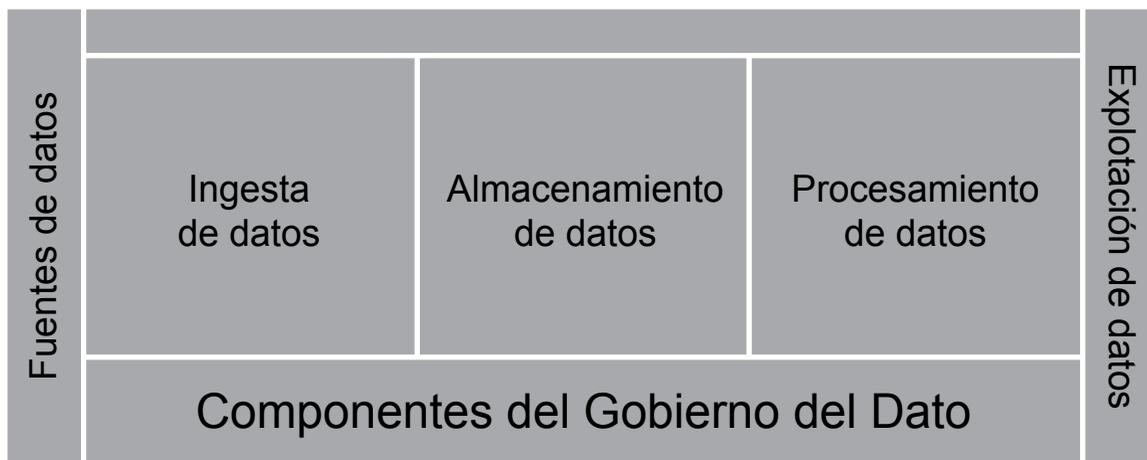
Para ello, es necesario crear en las organizaciones modelos que establezcan políticas, reglas y estándares que indiquen de qué manera se debe tratar los datos a fin de garantizar que se almacenen y utilicen de manera organizada e integrada. Así, los datos estarán disponibles, serán aprovechables y útiles, cuidando de que su calidad esté resguardada.

Como se viene desarrollando, en la actualidad, los datos tienen un rol decisivo en cualquier organización. Considerando su importancia, la gestión de los datos es una función clave. Por lo tanto, es importante que la arquitectura de datos esté alineada con los procesos de la organización y que acompañe la evolución de las tecnologías considerando el crecimiento de la organización.

De esta manera se hace posible definir las distintas fases de una arquitectura de datos típica, tal como se expone en la Figura 2. A continuación, se recorrerán estas diferentes fases a fin de introducirse en cada una ellas buscando facilitar las decisiones que deben tomar aquellas organizaciones que se encuentren atravesando procesos de transformación digital, incorporando proyectos o procesos ligados a la explotación de datos.

Figura 2.

Arquitectura de datos típica de una organización



Nota: el esquema grafica las fases típicas de una arquitectura de datos. Fuente: elaboración propia.

7. Ingesta de datos

Se reconoce esta fase como el momento en que se capturan los datos que serán utilizados para los distintos procesos analíticos, para lo que es necesario reconocer qué tipos de datos existen y cuáles son las fuentes de estos. Ello permitirá seleccionar tanto el método analítico adecuado, así como el almacenamiento más conveniente.

7.1 Origen de datos

7.1.1 Datos internos

Una primera distinción que debemos hacer es de dónde provienen los datos, ya que ello dará lugar a estrategias diferentes de explotación. Para aquellos datos provenientes de la operación interna de la organización se corresponde la definición de datos internos. Es decir, todos aquellos datos generados por la operación diaria de la organización. En general nos referimos a datos transaccionales (Joyanes Aguiar, 2013).

7.1.2 Datos externos

Por otra parte, cuando los datos provienen del exterior de la organización, se los cataloga como datos externos. En esta categoría se consideran datos provenientes del mercado, de clientes, de competidores, etc. Pueden incluirse encuestas, estadísticas oficiales, investigaciones, entre otras cosas. La inclusión de datos externos debe contribuir a la creación de valor y a la búsqueda de optimización en la toma de decisiones (Joyanes Aguiar, 2013).

7.1.3 Datos alternativos

Una mención particular merece el concepto de datos alternativos. Hasta aquí se han descrito fuentes que pueden ser consideradas tradicionales. El avance de la transformación digital da lugar a la incorporación del concepto.

Se trata de datos extraídos de canales no tradicionales que son útiles en las organizaciones para ayudar a comprender elementos claves de su operatoria; por ejemplo, el comportamiento del consumidor. Del mismo modo, el uso de este tipo de datos debe contribuir a tomar mejores decisiones mejorando los productos y servicios ofrecidos.

Como ejemplo de este tipo de datos se puede considerar la información de dispositivos móviles, de búsquedas en línea, de iniciativas de *open data*, de *redes sociales*, entre otros *origenes* (Joyanes Aguiar, 2013).

7.2 Tipos de datos

Partiendo de las características fundamentales del *Big Data* (las 3 V), se observa que la variedad se manifiesta a partir de la existencia de distintos tipos de datos que deben ser gestionados de manera apropiada para generar valor y cuyo ritmo de producción hace necesaria una capacidad de respuesta cada vez más instantánea.

Se hace necesario trabajar con diferentes tipos de datos, lo que conduce a distintas estrategias para su tratamiento. Es posible encontrar datos estructurados, como bases de datos tradicionales; semiestructurados, como información de máquinas o sensores; y datos no estructurados, como imágenes, videos, comentarios, textos, entre otros.

7.2.1 Datos estructurados

Más allá del crecimiento de los datos no estructurados y semiestructurados disponibles, los datos estructurados siguen siendo los de mayor explotación en las organizaciones. Principalmente, son tablas y otras estructuras de datos de bases relacionales (Joyanes Aguiar, 2013).

Corresponden a aquellos datos cuya recopilación implica una definición previa acerca de cómo serán estructurados. Son datos que guardan un orden estricto y están generalmente almacenados en tablas donde las filas representan observaciones y las columnas, variables. Son el tipo de datos utilizados por la estadística tradicional y dan lugar a bases de datos relacionales.

7.2.2 Datos semiestructurados

Se consideran datos semiestructurados aquellos datos cuya disposición no responde a un orden estricto predefinido, como en el caso de los datos estructurados, aunque sí existe un esquema implícito que puede ser reproducido a través de un procedimiento (Joyanes Aguiar, 2013).

Los datos semiestructurados pueden provenir de distintos lugares, como maquinarias, sensores, páginas web, bases de datos jerárquicas o heredadas, entre otros.

7.2.3 Datos no estructurados

Al mencionar datos no estructurados se hace referencia a aquellos datos que no están contenidos por ninguna estructura predefinida, ni responden a ningún esquema que sea posible reproducir. Si bien los datos de mayor explotación en la actualidad son los estructurados, el tipo de datos más abundante son sin duda los no estructurados, ya que provienen de toda actividad humana. La mediación tecnológica de las relaciones humanas y la capacidad informática creciente facilitan el almacenamiento de este tipo de datos haciendo que su importancia en términos analíticos se muestre en aumento y sin perspectiva de freno en su progreso (Joyanes Aguiar, 2013).

Como ejemplo de este tipo de datos podemos citar todo tipo de texto (documentos, correos electrónicos), datos provenientes de las redes sociales (comentarios sobre una empresa, sobre un producto o sobre la competencia), imágenes, videos, audios, flujo de personas, datos de geolocalización, entre otros.

7.3 Fuentes de datos

Otra de las características centrales de *Big Data*, conceptualizadas a partir de las 3 V, es el Volumen. Dada la gran cantidad de recursos disponibles para la extracción y el procesamiento de datos como producto de la transformación digital, esta característica guarda estrecha relación con las fuentes de donde provienen los datos.

Frente a tal caudal de datos, las organizaciones deben generar habilidades para seleccionar las fuentes que mayor beneficio reporten a menor costo económico y operativo. Por ello resulta importante comprender cuáles son las categorías de datos existentes y cuáles, las fuentes disponibles a fin de identificar cuáles satisfacen las necesidades de la organización (Joyanes Aguiar, 2013).

A continuación, se mencionan y explican brevemente las fuentes de datos más comúnmente utilizadas.

- Bases de datos: las bases de datos relacionales y no relacionales (ver Almacenamiento) son consideradas una de las fuentes de datos más utilizadas.
- Máquina a Máquina (M2M): el vínculo Máquina a Máquina se da a través de sensores; es también conocido como Internet de las Cosas (IoT). Refiere a la interconexión digital de objetos (tanto de la vida cotidiana como de la industria) con el soporte de Internet. A través de este vínculo es posible conseguir datos que, al ser procesados, aportan valor (Knoll y Viola, 2018).
- Datos de transacciones: se consideran todos los datos que se generan a partir de la operatoria diaria de cada organización. Entre los más habituales es posible encontrar datos de facturación, de ventas, llamadas, mensajerías, reclamos, pagos, consultas, o cualquier tipo de registro de transacciones. Generalmente, tales datos se encuentran almacenados por sistemas de información, como ERP (*Enterprise Resource Planning*) o CRM (*Customer Relationship Management*).
- *Web y Social Media*: son datos provenientes de Internet y de las redes sociales. Por su masividad, esta fuente de datos constituye una porción importante del fenómeno *Big Data*. Tal es así que resulta ser la principal fuente de datos alternativos para las organizaciones en la actualidad. Así, es posible extraer información de clics en vínculos o elementos, búsquedas realizadas, publicaciones de redes sociales, contenido de páginas web, imágenes, videos, blogs, wikis, audio, fotografías, libros, etc.
 - » *Web Scraping*: es la técnica utilizada para extraer información en forma semiestructurada o no estructurada de manera semiautomática de cualquier página web previamente establecida (Paliotta, 2018).
 - » *API (Application Programming Interface)*: es el conjunto de protocolos y procedimientos útiles para integrar y desarrollar *software* de aplicaciones. A los fines de este trabajo,

son útiles para el intercambio de información entre aplicaciones. Esto significa que facilitan la extracción de datos; por ejemplo, de redes sociales.

- Datos biométricos: en general, son aquellos empleados para la identificación automática de personas basada en sus características anatómicas. Sin embargo, también pueden ser usados para el análisis de táctica deportiva o en la investigación veterinaria, por ejemplo.
- Plataforma: es la manera privilegiada de extracción de datos en la actualidad. Es entendida como la infraestructura digital que permite la interacción entre personas y grupos. Se posiciona, entonces, como intermediaria entre clientes, anunciantes, proveedores de servicios, productores, distribuidores, incluso objetos físicos (Srnicek, 2016).
- *Open Data*: se consideran datos abiertos a aquellos datos accesibles y reutilizables que no necesitan permiso ni exigencias para su uso. Para definir datos como abiertos deben estar disponibles de forma conveniente y en un formato modificable. Deben poder ser reutilizados y debe ser posible su redistribución con requisitos mínimos, como la correspondiente referencia a la fuente. Existen iniciativas de datos abiertos tanto en el sector público como en el sector privado (Bron, 2015).
- Datos producidos por humanos: las personas, de forma cotidiana, generan, permanentemente, datos que son pasibles de ser almacenados siempre que la actividad sea mediada de alguna manera por tecnologías. Se pueden considerar muchos aspectos en este punto, como datos de posición geográfica y de tiempo (geolocalización), operadoras de telecomunicaciones, entre otros.
- Ciudades inteligentes: se trata de aquellas ciudades donde, mediante el uso de la infraestructura de las telecomunicaciones y de la información basada en Internet, se brinda a sus ciudadanos una serie de servicios digitales enfocados a mejorar el nivel de desarrollo humano, económico y cultural de la comunidad en su conjunto. Como es esperable, es posible generar y extraer gran cantidad de datos; entre ellos, información de transporte público, flujo de personas, utilización de servicios (Finquelievich y Prince, 2008).

7.4 Periodicidad de ingesta

Todos los procesos de gestión de datos implican decidir la periodicidad en que se capturan los datos que luego permiten el análisis.

7.4.1 *Ingesta única*

Se trata de proyectos útiles para tomar decisiones puntuales. Se hace una única vez la ingesta de la información requerida para el análisis; así, se genera el conocimiento necesario para dar soporte a la toma de decisiones y finaliza la etapa de ingesta.

7.4.2 *Ingesta periódica*

El carácter recurrente de ciertos procesos hace necesaria la ingesta periódica de datos. Es decir que, en función de las necesidades del proyecto, se define una frecuencia de extracción de datos que irá nutriendo al proceso analítico de la información que se necesita de manera diaria, semanal, mensual, trimestral, anual, etc.

7.4.3 *Ingesta en tiempo real*

Se han mencionado en varias oportunidades las 3 V de *Big Data*. En este punto toma especial relevancia la velocidad. Como producto de la disponibilidad de fuentes de información que surgen de la intermediación de la actividad humana, la generación de datos es permanente y constante.

En la actualidad, la presencia generalizada de dispositivos móviles interconectados (especialmente, los teléfonos inteligentes) demanda para ciertos rubros que la toma de decisiones deba ser inmediata; serán decisiones rápidas en tiempo real o con breves demoras.

El flujo continuo de datos, que contribuye al aumento creciente de su volumen almacenado y generado, da lugar a una oportunidad de creación de valor a partir de los datos. El procesamiento de

datos en tiempo real implica que los datos se capturan y almacenan en el mismo momento en que son producidos. El análisis de este flujo continuo permite encontrar patrones significativos (Joyanes Aguiar, 2013).

8. Almacenamiento

8.1 Arquitectura de almacenamiento

El almacenamiento de los datos es una fase muy importante en el proceso de gestión de datos para cualquier organización. Al considerar la arquitectura de almacenamiento de datos, es momento de definir la disposición física y lógica de los datos que se capturan; tanto el lugar donde se almacenan como la disposición y relación entre bases de datos o reservorios. De ello depende la facilidad o la dificultad para extraer datos sin procesar que, más adelante, serán fuente de valor. Es decir, la elección en torno al almacenamiento deberá estar alineada a las necesidades del proceso de gestión de datos de la organización en términos operativos, y, en términos estratégicos, según sus necesidades.

Para ser procesados eficazmente, los datos necesitan ser organizados y limpiados. La fuente y el origen de los datos son determinante a la hora de definir tanto el método bajo el que serán procesados como el criterio de organización y limpieza. Es decir que el tipo de almacenamiento deberá ser cuidadosamente definido en función del tipo, origen y fuente de los datos, por un lado, y, por el otro, por el proceso analítico, dado que es en función de este que se definen las necesidades de limpieza y organización de los datos. Según estos condicionantes, existen diferentes soluciones de almacenamiento.

8.1.1 Data Warehouse

La primera generación de plataformas de almacenamiento es asociada a almacenes centralizados. Bajo esta lógica, las organizaciones obtienen información principalmente de bases de datos operativas que registran perfiles y datos transaccionales de la organización. Estos datos son utilizados como soporte para la toma de decisiones y para procesos de inteligencia de negocios (*business intelligence*).

A medida que avanza la transformación digital, se presentan desafíos para las lógicas de almacenamiento. Este tipo de almacenamiento consiste en un esquema donde los datos son almacenados en un dispositivo local a cargo de cada organización. En tanto se incrementan el volumen y la velocidad de datos generados, aumenta la exigencia sobre los sistemas de almacenamiento con su consecuente aumento de costos en términos económicos y operativos.

Adicionalmente, no solo los conjuntos de datos disponibles crecen exponencialmente, sino cambian su forma. Ya no se trata de datos almacenados bajo una estructura predefinida, sino que coexisten con datos no estructurados cada vez con mayor masividad; por ejemplo, documentos de video, audio y texto. La existencia y la necesidad de procesamiento de datos no estructurados contradicen en muchos aspectos la lógica de funcionamiento de este tipo de almacenamiento. Por ese motivo surgen nuevos sistemas, dado que este tipo de almacenamiento está diseñado para dar soporte a procesos de datos esencialmente estructurados (Armbrust *et al.*, 2021).

8.1.2 Data Lake

Como solución para el problema planteado, surge una segunda generación de almacenamiento de datos conocida como lago de datos y *Data Lake*. Este sistema permite almacenar a bajo costo datos sin procesar en diversos formatos, que luego serán extraídos mediante procesos ETL (*Extract, Transform and Load*) hacia almacenes de datos descendentes para estar disponibles y ser utilizados en procesos analíticos de creación de valor y dar soporte a la toma de decisiones.

La implementación de este tipo de soluciones de almacenamiento facilitó el uso de métodos emergentes, como el aprendizaje automático. Al mismo tiempo favoreció el almacenamiento de datos no estructurados, que luego estarán disponibles para su ordenamiento y limpieza, lo que permitirá su procesamiento.

Si bien este tipo de arquitectura de datos aún se muestra dominante en el mercado con estructuras de costos accesibles para las organizaciones, su diseño en dos niveles resulta demasiado complejo para los usuarios, lo que incrementa el costo operativo (Armbrust *et al.*, 2021).

8.1.3 Lakehouse

Una tercera generación de almacenamiento de datos se muestra como superadora de las tecnologías anteriores. *Lakehouse* se define como un sistema de datos basado en almacenamiento de bajo costo y de acceso directo a los datos. Admite tanto datos no estructurados como datos estructurados. Es posible el empleo de métodos tradicionales asociados a almacenes o *Data Warehouse*, como métodos analíticos tradicionales, indexación, auditoría; y el uso de métodos asociados a la segunda generación de almacenamientos, como el aprendizaje automático y la ciencia de datos. Por lo tanto, esta tecnología combina los beneficios de *Data Lake* y de *Data Warehouse*.

Es habitual encontrar que los datos transaccionales que dan lugar a bases de datos estructuradas sean más fácilmente almacenados en *Data Warehouse*, mientras que los datos semiestructurados y no estructurados presentan mayor facilidad de almacenamiento en *Data Lakes*. Dado que la preferencia es hacia el uso de almacenamientos unificados de los datos que faciliten el análisis integrado de distintas fuentes que incluyen datos estructurados, semiestructurados y no estructurados, se observa tendencia hacia la utilización de almacenamientos tipo *Lakehouse*. Esta arquitectura combina elementos de arquitecturas *Data Warehouse* y de arquitecturas *Data Lake*, lo cual permite optimizar la integración de datos de distintos tipos, distintas fuentes y distintas velocidades de procesamiento (información histórica, en tiempo real, etc.).

8.2 Bases de datos

Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático. Existen distintos tipos de bases de datos. A los efectos de este trabajo, se definirán las bases de datos relacionales y las no relacionales.

8.2.1 Bases de datos relacionales

Este tipo de bases de datos almacena y permite el acceso a los datos mediante relaciones existentes entre sí. Esta manera de almacenar está asociada a la estadística tradicional o rectangular donde los datos se organizan en tablas. Cada fila se corresponde con una observación y cada columna, con un atributo que las describe, es decir que estamos frente a datos estructurados. Cada fila en una tabla se identifica con un código de identificación individual (conocido como ID o clave), por lo cual cada registro es único. Para cada atributo, identificado en las columnas, los registros tienen un solo valor, lo que hace sencillo establecer relaciones entre distintas tablas, contribuyendo, además, a optimizar el almacenamiento en términos de espacio físico o capacidad de almacenamiento (Joyanes Aguiar, 2013).

8.2.2 Bases de datos no relacionales

Dado que para la gestión de bases de datos relacionales el lenguaje más difundido es SQL, a las bases de datos no relacionales también se las conoce como bases NoSQL. Este tipo de bases de datos fue diseñado para la manipulación de grandes volúmenes de datos, lo que implica considerar datos estructurados y no estructurados. No están basadas en el modelo entidad-relación, por lo cual no se conciben con las bases de datos tradicionales.

Las bases de datos no relacionales muestran mayor flexibilidad, dado que no requieren que el almacenamiento venga dado por tablas ni responden a estructuras de datos prefijadas, lo que da lugar a distintas maneras de mapear los datos. Difieren del modelo clásico de almacenamiento y su ausencia de relaciones obligatorias entre conjuntos de datos hace que sea posible almacenar mayor cantidad de datos dando escalabilidad en los recursos que se podrán utilizar.

8.3 Calidad del dato

El volumen de datos que en la actualidad manejan las organizaciones hace que el mantenimiento y fomento de la calidad de los datos sean uno de los grandes retos a los que se enfrentan. Al considerar el almacenamiento será importante tener en cuenta la calidad, dado que de ello dependerá la utilidad de estos.

El concepto de calidad del dato está asociado principalmente a datos almacenados en bases relacionales, es decir que resulta aplicable solo a datos estructurados. Existen marcos de trabajo que pueden ser utilizados para medir y promover el uso de datos de calidad en las organizaciones. El más difundido es la norma ISO/IEC 25012, según la cual los datos de calidad son aquellos que responden a seis objetivos de calidad: completitud, unicidad, oportunidad, validez, precisión y consistencia. Para facilitar la comprensión se plantean preguntas y para facilitar su medición o seguimiento se establecen métricas que aportan información acerca de cada objetivo de calidad. Adicionalmente, se establecen niveles deseados de calidad, asociados a dichas métricas, que establecen la tolerancia que cada organización tendrá frente a desvíos en cada objetivo (Calabrese *et al.*, 2019).

Objetivos de calidad:

- Completitud: ¿Existen datos faltantes? Medición habitual: cantidades o porcentajes de datos faltantes o nulos.
- Unicidad: ¿Hay valores registrados más de una vez? Medición habitual: cantidad o porcentajes de registros duplicados.
- Oportunidad: ¿Representan la realidad en el momento requerido? Medición usual: diferencias temporales.
- Validez: ¿La sintaxis de los datos se condice con su definición (tipo de dato, formato y rango de valores)? Medición usual: porcentaje de datos inválidos.
- Precisión: ¿En qué medida el dato describe adecuadamente lo que representa? Medición usual: porcentaje de datos con diferencias en relación con un patrón (comparados contra un conjunto de datos validado).
- Consistencia: ¿Existe alguna diferencia entre datos que representan un mismo aspecto de un tema? Medición usual: porcentaje de datos con diferencias.

9. Procesamiento de datos

En esta instancia, se llega al momento de extraer valor de los datos. En concreto, en esta fase se trabaja sobre los datos a partir de la aplicación de técnicas de analítica que permiten detectar patrones y generar conocimiento, lo que se debe traducir en que los datos aporten valor agregado a la organización. Esto es algo que sólo será posible sabiendo gestionar tanto la información que nos ofrecen los datos estructurados, así como los no estructurados de la organización, con la ayuda de las herramientas de analítica adecuadas.

El científico de datos está especialmente vinculado a esta etapa, pues su función lo relaciona directamente. Es por ello por lo que se toma el proceso típico para la ciencia de datos a fin de tener una visión general acerca de cuáles son las etapas del procesamiento de datos.

9.1 Etapas para el análisis de datos

9.1.1 Importar

En un primer momento, el responsable del proceso de análisis deberá tener acceso a los datos; es decir, debe tener los datos disponibles para trabajar. Esto implica que deberá acceder, por sí o a través de un tercero, al almacenamiento de datos necesario. Típicamente, en esta instancia, los sets de datos se extraen de archivos guardados (en formato .csv), base de datos o utilizando una API (Wickham y Grolemund, 2017).

9.1.2 Ordenar y limpiar

Además de estar disponibles, es importante que los sets de datos estén en el formato necesario según las técnicas de análisis que se utilicen. Ordenar hace referencia a la necesidad de estructurar los datos según la forma necesaria para diferentes funciones. En general, se considera que los datos están ordenados cuando cada fila responde a una observación y cada columna, a una variable que la describe. Es decir que se busca que los datos respondan a la forma de tablas rectangulares propuesta por la estadística tradicional.

Por otra parte, la limpieza de datos implica encontrar y subsanar aquellos valores para los cuales existe un problema que hace que no sean aplicables las técnicas de análisis de datos o que sus resultados se vean comprometidos por la posibilidad de estar distorsionados. Con frecuencia se trata de errores de carga u operativos en el momento de traspaso de datos a distintos soportes; por ejemplo, que no se respete el formato de fecha, por lo que el sistema no sería capaz de reconocerlos.

También es posible encontrar valores perdidos cuando hallamos registros para los que una o más variables no presentan ningún valor. De la misma forma, los valores atípicos son observaciones donde uno o más registros son significativamente distantes, en términos numéricos, del conjunto de observaciones considerado; por ejemplo, si el atributo edad, para una observación que representa a una persona cualquiera, contiene el número 250. Por último, es frecuente que existan problemas de formatos cuando no se sigue el mismo criterio para todas las cargas.

Para subsanar estos problemas se recurre a la imputación o a la eliminación de registros. La imputación implica que se establezca un criterio para reemplazar los valores en conflicto por un valor estimado. Es habitual que se utilice la moda (valor más frecuente) cuando se trata de variables categóricas y se use la media aritmética cuando se trata de variables continuas; existen también otros métodos de asignación. Es posible también que sea necesario eliminar el registro en conflicto. Para seleccionar cómo proceder frente a estos datos, se prioriza evitar pérdida de información y se busca minimizar la distorsión en los resultados (Wickham y Grolemond, 2017).

9.1.3 Transformar

Una vez superados la limpieza y el orden de datos, se da paso a la transformación, la que tiene varias implicancias. En primer lugar, se busca reducir la cantidad de observaciones hasta conseguir una muestra que responda al interés del proyecto; por ejemplo, personas de una ciudad, clientes que respondieron a cierta campaña de *marketing*, etc.

En segundo lugar, se considera la creación de variables derivadas. Es decir, se calculan nuevas variables en función de las variables existentes; por ejemplo, el tiempo transcurrido desde la última compra a partir de las fechas.

En tercer lugar, se explora la posibilidad de incorporar datos externos que ayuden a mejorar la capacidad predictiva o analítica de los modelos. Por ejemplo, si se estuviera estimando el valor de propiedades de una ciudad, sería posible incorporar datos del valor del metro cuadrado o la distancia a los principales medios de transporte según la localización del inmueble.

Una vez que los datos están ordenados, limpios y transformados, existen dos herramientas complementarias que facilitan la generación de conocimiento y, con ello, valor para la organización: la visualización y el modelado. Ambas etapas son poseedoras de fortalezas y debilidades, aunque su carácter complementario lleva a que en los procesos reales se iteren varias veces entre ellas (Wickham y Grolemond, 2017).

9.1.4 Visualizar

Se trata de la representación gráfica de un set de datos o de un subgrupo de sus variables u observaciones. La finalidad es facilitar la detección de tendencias y patrones que ayuden a extraer conocimiento. La visualización ayudará a observar tendencias no esperadas que puedan dar lugar a nuevas preguntas de investigación sobre el conjunto de datos analizado o sobre el tema trabajado.

También puede dar información acerca de la orientación del proyecto, si se están haciendo las preguntas adecuadas o si es necesario cambiarlas, así como también podrá mostrar necesidad de recolección de datos con criterios diferentes o específicos.

Esta etapa hace fundamental la participación de personas que conozcan la organización y su operatoria. La interpretación se dará a partir de estrategias de visualización, lo que hace necesaria la participación humana. Así, el hecho de conocer la organización facilita las conclusiones (Wickham y Grolemond, 2017).

9.1.5 Modelar

El término “modelar” implica representar la realidad a partir de un esquema simplificado. Como estamos frente a modelos matemáticos, tal representación se realizará utilizando fórmulas, ecuaciones o funciones. Son útiles para mostrar relaciones entre distintas variables, parámetros y restricciones. Por ello, siempre que haya un modelo, habrá al menos una variable dependiente y al menos una variable independiente.

Tal como se mencionó, los modelos sirven como complemento de las visualizaciones iterando entre sí en el proceso. Esto es así, pues, si la visualización permite detectar un patrón de comportamiento o una relación entre variables, el modelado permite cuantificarlo haciendo comparables elementos del fenómeno estudiado, como observaciones o casos.

La utilidad de un modelo puede sintetizarse en dos propósitos: predecir y explicar. Cuando se busca predecir, se intenta estimar el valor de una variable sobre la base de los valores conocidos del pasado o de otras variables conocidas a las que se las suele llamar predictores. Al modelar con la finalidad de predecir, no es importante entender cómo o por qué interactúan entre sí las variables, sino que interesa que la estimación sea lo más acertada posible.

Por otra parte, si el uso del modelo busca explicar, se buscará la comprensión de la relación entre la variable dependiente y las independientes. En este tipo de modelos es importante determinar si la relación es significativa y cuantificarla a fin de poder explicar el porqué de la relación, lo que contribuiría a entender el fenómeno estudiado.

Una última cuestión para considerar es la evaluación del modelo. Generalmente, se prueban más de un modelo antes de seleccionar el que se utilizará. Para ello existen métricas que exponen la *performance* o rendimiento de los modelos. Son indicadores que permiten comparar y seleccionar el modelo que mejores resultados consiga para los datos analizados (Wickham y Grolemond, 2017).

9.1.6 Comunicar

Tan importante como todos los pasos anteriores, comunicar es crítico para que el trabajo sobre los datos sea útil para la organización. No importa qué tan bueno haya sido el proceso si no se logra que la organización se apropie de los resultados para mejorar sus procesos y para tomar mejores decisiones.

Para que la comunicación sea útil deberá presentar los hallazgos de manera clara e intuitiva. Generalmente, los soportes visuales ayudan a comprender y concluir con mayor facilidad. Esta etapa deberá ser diseñada de manera que se favorezca la explotación autónoma de los datos. Es deseable que los usuarios de la organización puedan comprender por sí solos la dinámica de los datos.

9.2 Metodologías de análisis de datos más utilizadas

9.2.1 Métodos multivariados

La rama del análisis estadístico centrada en el estudio simultáneo de dos o más variables (características de un fenómeno) se conoce como análisis multivariado. Se considera la medición de las variables que explican un conjunto de objetos donde la importancia del análisis radica en la relación simultánea que estas manifiestan o no. En el análisis multivariante toma importancia el estudio de la covarianza y las correlaciones reflejadas en la relación entre tres o más variables (Aldás y Uriel, 2017).

Los métodos estadísticos multivariados más usados para el análisis de datos se enumeran a continuación: Análisis de componentes principales; Análisis factorial exploratorio; Regresión lineal

múltiple; Análisis de correspondencia; Análisis de conglomerados o *cluster* jerárquico y no jerárquico; Correlación canónica; Análisis ANOVA y MANOVA; Modelo de ecuaciones estructurales; Análisis discriminante; Análisis de desviaciones; Escalamiento multidimensional (Consoli, 2020).

9.2.2 Investigación operativa

La investigación operativa utiliza métodos analíticos avanzados buscando construir soluciones óptimas para problemas complejos de toma de decisiones; y este último es su objetivo final como disciplina. Surgida como resultado del estudio de operaciones militares durante la Segunda Guerra Mundial, el uso de este enfoque se ha incrementado para tratar problemas de asignación de recursos en distintas industrias y rubros comerciales.

Emplea técnicas matemáticas, modelado, análisis estadístico y optimización. Frecuentemente se ocupa de determinar valores extremos (mínimos o máximos) según un objetivo planteado que, sujeto a restricciones, se busca optimizar. Como ejemplo es posible citar el armado de una cartera de inversión sujeto a restricciones económicas y de resguardo de capital buscando maximizar la ganancia. También es una técnica muy utilizada para programar la logística de mercaderías o el traslado de personas en busca de minimizar el costo operativo (Hamdy, 2012).

Algunas de las técnicas más usadas en este campo son: programación lineal y entera; programación lineal en redes; programación de proyectos dinámica y de metas; procesos estocásticos; cadenas de Markov; simulación; Modelo General de Filas de Espera; redes neuronales (Consoli, 2020).

9.2.3 Métodos de análisis predictivo

Como puede deducirse por su nombre, este grupo de técnicas busca realizar predicciones del futuro o sobre acontecimientos no conocidos. Para lograrlo analiza datos actuales e históricos reales a partir de técnicas estadísticas, modelización, aprendizaje automático y minería de datos.

De esta manera, los modelos predictivos detectan patrones de comportamiento de datos históricos construyendo tendencias que permitirán predecir valores futuros. Para cada sujeto u observación se genera una puntuación con la finalidad de estimar el valor futuro o desconocido de una observación.

En general se utiliza un set de datos de entrenamiento que aportará la información histórica necesaria para la detección de patrones o tendencias con que se construirá el modelo que luego será utilizado para generar las estimaciones. Adicionalmente, se reserva un conjunto de datos con resultados conocidos que se emplearán para evaluar el modelo. En concreto, se estiman los valores que se espera predecir para el set de datos de validación y se comparan con los datos conocidos para cuantificar los aciertos y evaluar el modelo (Fitz-Enz y Mattox, 2014).

Las técnicas más utilizadas son los siguientes: métodos predictivos de clasificación y regresión; modelo bayesiano *naïve*; k-vecinos más cercanos; árboles de clasificación y regresión; máquinas de soporte vectorial; redes neuronales; regresión lineal; regresión logística. También se encuentran: ensambles por promedios, por votación, *boosting* y *bagging*; *Random Forests* y *Gradient Boosting*; análisis de conglomerados; *Cross validation* (Consoli, 2020).

9.2.4 Aprendizaje automático

Si bien está basado en técnicas de análisis predictivo, merece una mención especial por su difusión. La inteligencia artificial parte de la premisa de que es posible programar una computadora de manera que pueda emular o reproducir funciones cognitivas humanas (Ganascia, 2018). El aprendizaje automático o *machine learning* es considerado como una rama de la inteligencia artificial que se plantea como objetivo que las computadoras aprendan. Como evidencia del aprendizaje, estas técnicas plantean que una computadora es capaz de aprender siempre que logra mejorar su desempeño a partir de la experiencia. Es decir, los algoritmos de aprendizaje automático mejoran su rendimiento con su implementación a partir del uso de datos (Joyanes Aguiar, 2013).

9.2.5 Procesamiento del lenguaje natural

Con raíces en la lingüística computacional, el procesamiento de lenguaje natural (PNL) se ocupa

principalmente de diseñar y construir aplicaciones y sistemas que faciliten la interacción entre máquinas y los lenguajes naturales; es decir, los creados por humanos. Así, es posible emplear técnicas que permitan que una computadora procese y comprenda el lenguaje humano que podrá ser utilizado tanto para generar respuestas, acciones o conversaciones, así como también para generar resultados útiles, como resumir el contenido de un texto (Sarkar, 2019).

Algunas de las soluciones logradas a partir del procesamiento del lenguaje natural son: sistemas de traducción y transliteración; sistemas de reconocimiento de voz; sistemas de respuesta a preguntas; extracción de contenido semántico; categorización o catalogación de textos; análisis de texto; análisis de sentimientos (Consoli, 2020).

10. Explotación

Los datos pueden definirse como conjunto de valores que no tienen importancia por sí mismos, por sí solos resultan irrelevantes, no reportan beneficios. Tras ser procesados y al entablar relaciones, los datos pasan a tener significado, por lo que se transforman en información. Por último, cuando la información es integrada con la experiencia organizacional acumulada y anclada en sus creencias, es posible definirla como conocimiento. En esta última instancia es cuando una organización logra el aprendizaje efectivo, se apropia del conocimiento y modifica sus procesos internos junto con el comportamiento de sus miembros. Solo de esta manera es posible crear valor a partir de los datos (Gore, 2012).

Por este motivo, nada de lo anterior tiene sentido si no se le da paso a la explotación. En la etapa de explotación se utilizan los resultados del proceso de análisis y procesamiento de datos para intervenir en los procesos organizacionales, mejorarlos y crear valor.

10.1 Reporte

El reporte es un soporte típico usado en diferentes ámbitos para transmitir información de interés para quien la recibe. Este tipo de documentos puede adoptar formatos diferentes (escrito, oral, presentación, gráficos, etc.). Sin importar de qué formato se trate, es importante que al realizarlo se tenga en cuenta el público al que se dirige para ajustar el diálogo de manera que la comunicación sea efectiva.

Generalmente, los usuarios de la organización no son especialistas en estadística, en matemáticas o en informática como quienes intervienen de manera directa en el proceso de análisis de datos. Por ello, el momento de la comunicación tiene importancia crítica para lograr que los datos sean explotados de la mejor manera. Es necesario que al escribir informes se lo haga seriamente en un lenguaje claro, directo y coloquial. Es importante que se comuniquen las conclusiones de manera explícita persiguiendo la finalidad de que cada miembro involucrado pueda comprender más allá de no tener conocimientos en estadísticas. Así se logra mejorar la toma de decisiones en distintos niveles, lo que redundará en creación de valor a partir de los datos.

10.2 Utilización en procesos

En ocasiones, cuando nos encontramos frente a procesos automatizados donde las máquinas deben tomar decisiones con rutinas más o menos rígidas, el procesamiento de datos contribuye a mejorar la eficiencia en ellos. Un ejemplo típico es la estimación del valor de un automotor desde el momento en que un usuario cualquiera puede acceder a ese servicio aportando al sistema información que le es requerida y conocer de antemano una estimación del valor de su vehículo. Otro ejemplo podría ser la personalización de productos ofrecidos sobre la base del perfil y el comportamiento del usuario que está utilizando una plataforma. En ambos casos se usa información histórica para tomar decisiones instantáneas, sin intervención directa de personas, buscando mejorar la experiencia del usuario.

11. A modo de cierre

Se han recorrido todos los tópicos necesarios para tener una visión global del proceso de gestión de datos en una organización. Cada uno de los campos enunciados implica decisiones que se deben tomar para su definición y mejora. Para que estas decisiones lleven a acciones que produzcan mejoras en la organización es importante que el criterio de selección de tecnologías parta de una correcta definición del problema que se busca resolver y que sean orientadas a atacar las necesidades concretas de la organización.

Existe una falsa creencia que lleva a percibir la tecnología como la solución eficaz a todos los problemas de las organizaciones. Lleva a creer que la compra de tecnología vendría, en todos los casos, acompañada de aumentos en la productividad. Así, se cree que mayor tecnología es siempre mejor y, cuanto más nueva, más eficaz. Bajo esta lógica, la inversión debería estar orientada hacia la incorporación de las tecnologías más modernas.

Esta creencia opera como una presión a la hora de planificar la incorporación de tecnología y la inversión. En consecuencia, resulta sencillo e incluso habitual perder de vista el problema y quedar atrapados en la solución forzando su aplicación.

En lugar de definir de manera precisa el problema, se examinan las soluciones tecnológicas disponibles. De esta manera queda limitada la flexibilidad y la creatividad para la resolución del problema planteado. Además, pensándolo de esta manera, se incurre en el riesgo de ejecutar grandes desembolsos de dinero en la adquisición de tecnologías de punta, que no serán útiles, que se aplicarán parcialmente o que, simplemente, no son necesarias.

Si bien la implementación de soluciones tecnológicas es un buen camino para enfrentar distintos problemas de la vida organizacional, no siempre estas funcionan para ello y no todas las tecnologías son aplicables a todos los casos.

Partir de una buena definición del problema permite poner el foco en resolverlo. Conocer los elementos del proceso de gestión de datos ayuda a explorar las soluciones posibles aunque en todos los casos el éxito dependerá de seleccionar la mejor solución en función de las necesidades de la organización y no en función de la mejor tecnología disponible.

Referencias bibliográficas

- Alamán, P. (2019), “Nuevos roles y perfiles. Big Data sin misterios”, Banco Interamericano de Desarrollo.
- Aldas, J. y Uriel, E. (2017), *Análisis Multivariante aplicado con R, Ediciones Paraninfo*.
- Armbrust, M.; Ghodsi, A.; Xin, R.; Zaharia, M. (11-15 de enero de 2021), “Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics”, 11th Annual Conference on Innovative Data Systems Research (CIDR '21), Virtual. http://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf
- Banco Interamericano de Desarrollo [BID] (2018), “Disrupción exponencial en la economía digital (1)”. Recuperado de <https://publications.iadb.org/publications/spanish/document/Disrupci%C3%B3n-exponencial-en-la-econom%C3%ADa-digital.pdf>
- Bron, M. (2015), *Open data : miradas y perspectivas de los datos abiertos*, Consejo Latinoamericano de Ciencias Sociales. http://biblioteca.clacso.edu.ar/Argentina/unlar/20171117050559/pdf_1513.pdf
- Calabrese, J.; Esponda, S.; Pasini, A.; Boracchia, M.; Pesado, P. (14-19 de octubre de 2019), “Guía para evaluar calidad de datos basada en ISO/IEC 25012”, XXV Congreso Argentino de Ciencias de la Computación, Universidad Nacional de Río Cuarto, Córdoba. <http://sedici.unlp.edu.ar/handle/10915/91086>
- Consoli, E. (2019), “Desarrollo, riesgos y problemática de la Economía Digital”, Tercer puesto Premio anual Dr. Manuel Belgrano, CPCECABA.
- Consoli, E. (2021), “Roles y procedimientos de proyectos de Big Data en organizaciones argentinas para los años 2020 y 2021. Un enfoque desde la administración” (Informe Final Proyectos de Investigación 2020). Centro de Investigación en Ciencias Económicas – Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires.
- Consoli, Emiliano (2020), “Ciencia de datos y estudios organizacionales: estudio exploratorio y transversal de posgrados orientados a la ciencia de datos en el plano mundial” (Trabajo Final de Posgrado. Universidad de Buenos Aires). http://bibliotecadigital.econ.uba.ar/download/tpos/1502-1835_ConsoliE.pdf
- Douglas, L. (2001), *3D data management: controlling data volume, velocity, and variety*, Meta-group.
- Elfenbaum, M. (2019), “El aprendizaje automático como herramienta innovadora para crear valor en las organizaciones” [Tesis de doctorado, Universidad de Buenos Aires]. Repositorio de la Biblioteca Digital de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires.
- Finqueliévich, S. y Prince A. (14 y 15 de abril de 2008), “Gobiernos locales y ciudades digitales”, En: Seminario La Conectividad y las Políticas de Gobierno Electrónico en los Gobiernos Locales de Iberoamérica. <http://municipios.unq.edu.ar/modules/mislibros/archivos/gob-locales-ciuddig.pdf>
- Fitz-Enz, J. y Mattox J. (2014), *Predictive analytics for human resources* (1a ed.), Wiley.
- Gore, E. (2012), *El próximo management, acción, práctica y aprendizaje*, Buenos Aires, Ediciones Granica.

- Halpern, D. y Valderrama, M. (2018), "Adaptando la empresa a la sociedad digital", Santiago de Chile, Think Tank Trendigital.
- Hamdy, T. (2012), *Investigación de Operaciones* (9ª ed.), Pearson.
- Henriette, E., Feki, M. y Boughzala, I. (2015), "The Shape of Digital Transformation: A Systematic Literature Review", *MCIS 2015 Proceedings*. Recuperado de: <http://aisel.aisnet.org/mcis2015/10>
- Hufty, M. (2008), "Una propuesta para concretizar el concepto de gobernanza: El Marco Analítico de la Gobernanza", *Actes & Mémoires de l'Institut Français d'Études Andines* 25. 77-100.
- International Organization for Standardization [ISO] (17 de febrero de 2022), "ISO/IEC 38500:2008 - Corporate governance of information technology". <https://www.iso.org/standard/51639.html>
- ISACA (17 de febrero de 2022). "COBIT an ISACA framework", <https://www.isaca.org/resources/cobit>
- Ismail, S., Malone, M. y Van Geest, Y. (2014), *Exponential Organizations*, Bubuk Publishing.
- Joyanes Aguiar, L. (2013), *Big Data, análisis de grandes volúmenes de datos en organizaciones*, Alfaomega Grupo Editor.
- Knoll P. y Viola, A. (2018), "Economía digital: acelerado avance y desafíos que presenta", Observatorio de la Economía Mundial, Universidad de San Martín. Recuperado de: <http://www.unsam.edu.ar/escuelas/economia/investigacionpublicaciones/economia-internacional/economia-digital-acelerado-avance-y-desafios-que-presenta-1/>
- Lassalle, J. (2019), *Ciberleviatán, el colapso de la democracia liberal frente a la revolución digital*, Arpa & Alfil Editores.
- Morakanyane, R., Grace, A., & O'Reilly, P. (2017), "Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature", Bled eConference.
- Organisation for Economic Co-operation and Development [OECD] (2004), *Principios de Gobierno Corporativo de la OCDE*, Paris, OECD.
- Organisation for Economic Co-operation and Development [OECD] (2019), *OECD Reviews of Digital Transformation: Going Digital in Colombia*, OECD Publishing. <https://doi.org/10.1787/781185b1-en>
- Paliotta, A. P. (2018), "Nuevas profesiones y técnicas de web data mining en Argentina: el caso del data scientist", *Revista del Centro de Estudios de Sociología del Trabajo* (10), 97-113. Recuperado de: <http://www.economicas.uba.ar/revista-cesot>
- Piorun, D. (2016), "Normas y Marcos relacionados con la implementación de esquemas de Gobierno de la Tecnología de la Información" [Tesis de maestría, Universidad de Buenos Aires - Facultad de Ciencias Económicas]. Biblioteca Digital de FCE.
- Sarkar, D. (2019), *Text analytics with Python: a practitioner's guide to natural language processing*, Bangalore: Apress.
- Schreckling, E. & Steiger, C. (2017), "Digitalize or Drown". En *Shaping the Digital Enterprise* (pp. 3-27), Cham, Suiza, Springer.
- Schwab, K. (2016), *The Fourth Industrial Revolution*, Ginebra, Suiza, World Economic Forum.
- Srinicek, N. (2018), *Platforms capitalism*, Cambridge, UK, Polity Press Ltd.
- The Open Group (17 de febrero de 2022), "About the TOGAF Standard, Version 9.2". <https://www.opengroup.org/togaf>
- Wickham, H. y Grolemund, G. (2017), *R para Ciencia de Datos*, Oreily.
- Zachman, J. (17 de febrero de 2022), "The Concise Definition of The Zachman Framework". <https://www.zachman.com/about-the-zachman-framework>