

Capítulo XXX. Redefiniendo Big Data: una propuesta desde la academia

Julio Lopez-Nunez (Universidad Gabriela Mistral - Chile -)

ORCID ID: 0000-0002-7920-1563

Gerardo Cerda-Neumann (Universidad Gabriela Mistral - Chile -)

I. Introducción

La evolución de la Web, pasando desde un producto totalmente estático de carácter unidireccional para derivar en una plataforma dinámica y bidireccional, dio origen a una infraestructura considerada como una rica fuente en generación de datos, con una gran cantidad de personas de diferentes estratos sociales que participan democráticamente en ella (Leung *et al.*, 2019).

En este contexto, el desarrollo de técnicas capaces de procesar estos grandes volúmenes de datos encuentra cabida en diferentes actividades económicas y sociales. En particular, un sector atraído por descubrir qué dicen los datos sería el económico. Con esto, el procesamiento de datos a gran escala se plantea con un enfoque comercial, logrando establecer una ventaja para aquellas empresas que son capaces de conocer en profundidad lo que sus clientes demandan (Neves y Bernardino, 2015).

En este orden de ideas, la irrupción del *Big Data*, en complemento con las actividades que dicen relación con el análisis de estos conjuntos de datos, surge a raíz de la necesidad de almacenar y procesar Petabytes de datos (Zhou *et al.*, 2012). Así, las técnicas desarrolladas bajo el rótulo de *Big Data* deben considerar una arquitectura escalable, distribuida, de alto rendimiento y tolerante a errores (Neves y Bernardino, 2015).

Dado este contexto, el presente artículo pretende situar en el centro de la discusión el futuro que vemos respecto al concepto de *Big Data*. Para lo anterior, este manuscrito está organizado en tres secciones. La primera de ellas se debe entender como un marco de referencia en torno al concepto de *Big Data*. En una segunda sección se discute sobre las aplicaciones del *Big Data*, esto para establecer su transversalidad en diferentes contextos. Finalmente, la última sección presenta las conclusiones de los autores en torno a lo que hoy se entiende por *Big*

Data, la justificación de su aplicación y, cuál sería el futuro para el objeto de estudio que plantea este trabajo.

II. Objetivos

La acción de reflexionar sobre las distintas definiciones que existen en torno a *Big Data* se enmarca dentro de tres objetivos nucleares que este trabajo declara. Así, esta investigación busca analizar distintas definiciones de *Big Data*, esto con tal de encontrar puntos en común y también diferencias. Lo anterior permite lograr el primer objetivo, esto es, proponer una nueva definición de *Big Data*.

Con todo, este trabajo se enfoca en definir las características que debe cumplir un proyecto informático para ser considerado dentro de la dimensión *Big Data*. Lo anterior, que corresponde al segundo objetivo, permitirá analizar la aplicabilidad de la definición del concepto *Big Data* que proponen los autores.

Finalmente, se vuelve central desarrollar las líneas argumentativas que permitan señalar cuales serían las características que debe poseer un proyecto de tratamiento de datos. En esta línea, un apartado relevante es el estudio de los proyectos que dicen relación con el comportamiento de los estudiantes en ambientes virtuales. De esta forma, el tercer objetivo sería aclarar cuales son las características esenciales que se deben considerar en un proyecto de Learning Analytics.

III. Metodología

Esta investigación, de tipo exploratorio-descriptivo, se ajusta a la revisión de la literatura científica publicada en revistas seriadas de alto impacto (Garfield, 1972) e indexadas en la base de datos *ACM Digital Library*. En esta línea, este repositorio cuenta con opciones de búsqueda de documentos, por autor, por afiliación y, una opción avanzada diseñada para usuarios independientes capaces de estructurar cadenas complejas de búsqueda.

En tanto, este trabajo define tres etapas (García *et al.*, 2017), comenzando con el desarrollo de la estrategia de búsqueda. Así, en la realización de dicha búsqueda se consideró el término “Big Data”, cuya presencia debía ser en el título de la publicación. Posteriormente, una vez obtenidos los resultados se aplicaron criterios de inclusión y exclusión en línea a los planteamientos bibliométricos de Lotka (1926), Bradford (1934) y Price (1963), los cuales buscan establecer un conjunto de investigaciones representativas del fenómeno estudiado. De esta forma, los artículos seleccionados fueron aquellos que declaraban como objeto de estudio *Big Data*, conformando finalmente un subconjunto de 11

publicaciones. Adicionalmente, se incluyeron investigaciones que giraban en torno a la arquitectura necesaria para un proyecto *Big Data*, las herramientas propias del mismo y sus áreas de aplicación. Esta búsqueda se realizó ocupando los conceptos que se consideraron relevantes a partir del estudio inicial de la definición de *Big Data*. Lo anterior derivó en un subconjunto compuesto por 14 investigaciones.

En epítome, los artículos científicos utilizados para el desarrollo de esta publicación, subdivididos en dos dimensiones (*i.e.* Definición del concepto *Big Data* y, Arquitectura y Herramientas para el *Big Data*), consideraron un total de 25 publicaciones científicas incluidas en *ACM Digital Library hasta el mes de enero de 2021*.

Finalmente, establecido el conjunto de artículos propuestos para el desarrollo de esta investigación, se procedió con el procesamiento de las evidencias desde la óptica que entrega el análisis crítico (Greenhalgh, 1997; Carvajal, 2004), tanto de su validez interna como externa, continuando en la última instancia con el análisis sobre cada texto estudiado.

IV. Resultados

4.1. Big Data

El campo tecnológico asociado al *Big Data* ha concentrado el interés de la comunidad científica en los últimos 15 años. Este hecho encuentra asidero cuando se revisa la producción científica en torno a este campo. En una búsqueda realizada en Association for Computer Machinery (ACM) Digital Library a fines de abril de 2021, solo considerando como criterio de inclusión el término *Big Data* en el título de la investigación, los resultados obtenidos fueron de 51,086 artículos. Con esto, se ha desarrollado una definición para el término la cual gira en torno a la propuesta de Laney (2001), el que, sin hablar del término *Big Data*, introduce por primera vez el uso de las “3V”, a saber, Volumen, Velocidad y Variedad. En esta línea, *Big Data* se populariza gracias al explosivo crecimiento en la generación de datos (Yu *et al.* 2017; Manogaran y López, 2018; Akram *et al.*, 2019), entre los cuales podemos encontrar aquellos estructurados (organización tabular) y no estructurados (e. g. comentarios en redes sociales).

En esta línea, las definiciones del término *Big Data* que giran en torno a las tres V, son ampliamente aceptadas como base para la descripción del concepto (Uddin y Gupta, 2014). Pese a esto, es necesario señalar que esta definición inicial de las tres V fue ampliada a lo largo del tiempo, con tal de resaltar no solo las características relativas a la generación de datos, sino más bien para realzar la calidad de los mismos que se debe considerar en las tareas propias del *Big Data*. Así, nuevas propuestas para definirlo consideran en forma adicional a las tres V, las

siguientes características de los datos: valor, veracidad, variabilidad e incluso complejidad (Bedi *et al.*, 2014).

Además, resulta necesario señalar que el volumen de los datos a procesar no se refiere exclusivamente a que exista una gran cantidad de ellos, sino que estos deben ser todos los que están relacionados con el fenómeno en estudio (Mayer-Schönberger y Cukier, 2013). Así, esta característica asociada al volumen permite descubrir detalles que no son apreciables si no se utilizan todos los datos. Este hecho permite centrar la mirada en lo que sucede y no en el por qué. De esta manera, las tareas del *Big Data* deben considerar el universo de los datos cuando estos estén disponibles y no una muestra de ellos.

Sin embargo, también es posible encontrar propuestas que definen el concepto de forma minimalista, enfocando su eje central en las capacidades de procesamiento de la arquitectura que soporta a *Big Data* (National Institute of Standards and Technology [NIST], 2015). Así, todo conjunto de datos que no pueda ser procesado, almacenado y analizado por las tecnologías tradicionales de gestión de datos, se debe considerar como un problema perteneciente al ámbito del *Big Data*. Esto resulta de vital importancia, toda vez que el avance en desarrollos tecnológicos, por transitividad llevaría a una permanente mejora en la capacidad de cómputo y almacenamiento. Entonces, el concepto *Big Data* tendría una presencia volátil e intrascendente, considerando que la capacidad de cómputo y almacenamiento crecen a tasas más aceleradas que las mostradas por la generación de datos de calidad (Manyika *et al.*, 2011).

Dado lo anterior, es posible señalar que el término *Big Data* logra describir un entorno de grandes volúmenes de datos en el que las arquitecturas escalables logran soportar los requisitos de aplicaciones analíticas que los procesan. Además, dicha arquitectura debe poseer la capacidad de soportar una alta velocidad en la generación de datos, incorporando la posibilidad de procesarlos en una gran variedad de formatos (Emmanuel y Stanier, 2016).

Entonces, el desafío de la comunidad científica gira en torno a la arquitectura propicia para el procesamiento de grandes volúmenes de datos, generados en un corto periodo de tiempo, a veces unos pocos segundos (Laney, 2001) y, en una multiplicidad de formatos, ya que la tecnología existente no sería capaz de cumplir con estas exigencias. Con esto, el uso de la computación distribuida dio a luz en 2005 a Hadoop (Wang *et al.*, 2016; Sogodekar *et al.*, 2016; Oussous *et al.*, 2018), propuesta de Cutting y Cafarella que se transformó en un sistema de software de código abierto y escalable (Raza y Xujain, 2020), capaz de asumir los desafíos que presentaban los proyectos que hoy son catalogados como *Big Data*.

Lo recién mencionado permite concluir que no todo proyecto que procese un gran volumen de datos podrá ser considerado como de *Big Data*. Las características que debería cumplir este tipo de proyecto son: incluir todos los datos existentes en el contexto analizado y no una muestra de ellos, una velocidad de generación de datos que impide procesarlos con la arquitectura tradicional y un alto volumen de datos creados en un breve tiempo.

Por su parte, la propuesta de Hadoop permitió realizar cálculos sobre grandes conjuntos de datos gracias a que ellos no serían gestionados de forma tradicional. Por el contrario, Hadoop tiene múltiples réplicas de los datos, esto con la finalidad de tener una alta fiabilidad (Shafer *et al.*, 2010). Finalmente, el reconocimiento al trabajo de Cutting y Cafarella se traduce en una multiplicidad de otras propuestas de arquitectura para su uso en *Big Data*, todas ellas basadas en el corazón de Hadoop (Uzunkaya *et al.*, 2015), tales como Cloudera, Hortonworks, MapR e InfoSphere BigInsights entre otras.

Con todo lo expuesto, es posible presentar una definición de *Big Data* que permita tener un punto de partida para su estudio y posterior aplicación. En esta línea, la propuesta de Isitor Emmanuel (Emmanuel y Stanier, 2016), quien revisa de manera crítica las definiciones existentes, congrega una serie de elementos que la transforman en una definición robusta, consiste y permanente en el tiempo: “El término Big Data describe un entorno de datos en el que las arquitecturas escalables respaldan los requisitos de aplicaciones analíticas y de otro tipo que procesan, con alta velocidad, datos de gran volumen que pueden tener una variedad de formatos de datos y que pueden incluir la adquisición de datos a alta velocidad” (Emmanuel y Stanier, 2016, p. 6).

Esta definición nos parece valiosa dado que se basa en las características de los datos, sus aplicaciones y la arquitectura que se necesita para realizar *Big Data*. Todo esto permite utilizarla, sin depender de características transitorias del *Big Data* y, disponer de una definición permanente en el tiempo. De esta forma, si las organizaciones tienen un conjunto de datos que no son posibles de procesar con la infraestructura tradicional (arquitectura cliente servidor y bases de datos relacionales, por ejemplo), ahí nos encontramos frente a un proyecto *Big Data*.

4.2. Áreas de aplicación

Evocando la propuesta de Nolan (1973), quien desarrolló un modelo de etapas de crecimiento de los recursos informáticos al interior de la empresa, es posible ilustrar que la administración y valoración de los datos sería una de las últimas etapas de maduración que las

organizaciones experimentan en torno a los recursos informáticos que poseen. Dado esto, la figura 1 representa gráficamente la curva de aprendizaje que propone el Modelo de Nolan, el cual está en función del tiempo y considerando la inversión en tecnologías de la información que exhiben las empresas. En este sentido, comprender que los datos representan uno de los principales activos que una empresa posee (Al-Qirim *et al.*, 2017), permite reconocer la importancia que reviste la gestión y análisis de estos.

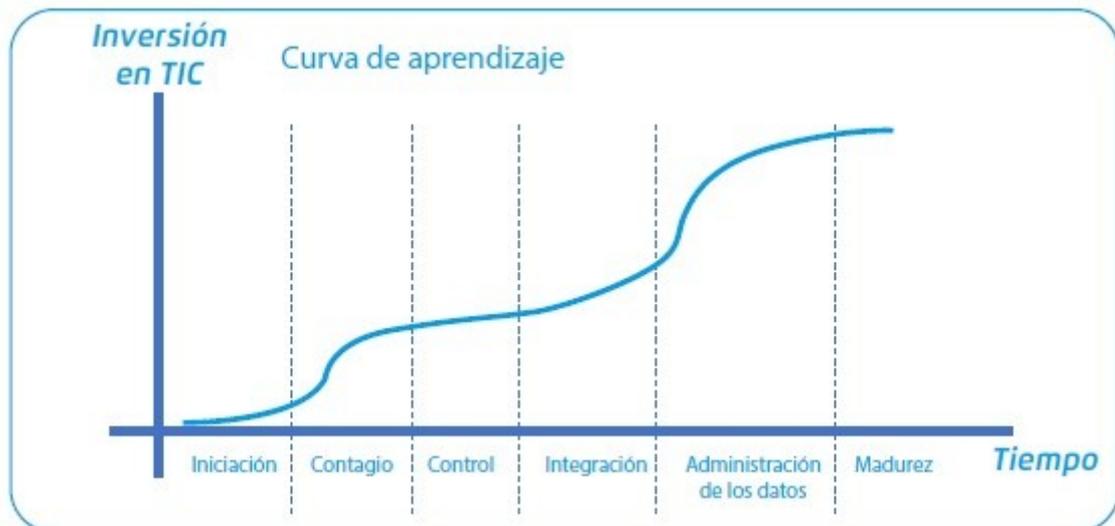


Figura 1. Modelo de Madurez de Richard Nolan. **Fuente:** Elaboración propia, adaptado de (Saaksjarvi, 1985)

Con esto, el análisis del *Big Data* encuentra cabida en áreas tan disimiles como datos de salud, genomas, medio ambiente, redes sociales, educación, logística y actividades comerciales en general (Rana y Mugavero, 2019).

Sin embargo, no debe causar sorpresa que las aplicaciones de *Big Data* contemplan una fuerte componente comercial, enfocados en maximizar el retorno sobre la inversión en los proyectos de tecnología de la información (Gopalkrishnan *et al.*, 2012). Por esto, no es extraño que se destaquen como casos de éxitos proyectos de analítica *Big Data*, los cuales apuntarían hacia procesos de auditoría en temas financieros, entre los que se pueden citar aquellos que buscan explicar los orígenes de los errores, sesgos, duplicidad y omisiones en los datos contables.

Así también, se pueden destacar los casos de Oil-for-Food (proyecto que permitió a Irak cambiar petróleo por alimentos y medicinas luego de la guerra del Golfo) (Gopalkrishnan *et al.*, 2012), Netflix (proyecto *Big Data* que permite recomendar a sus usuarios series, películas y/o

documentales en base a las preferencias de millones de usuarios) (Amatriain, 2013; Chew, Haw, y Subramaniam, 2020) o Google Flu Trends (proyecto capaz de hacer una predicción de la incidencia de la gripe en cada estado de EEUU vía el análisis de las consultas en su motor de búsqueda) (Ginsberg *et al.*, 2009; Culotta, 2010). Todos estos forman una muestra de las diversas aplicaciones que tiene el *Big Data*.

Dado este contexto, un caso de uso especial del *Big Data* corresponde a la aplicación en el ámbito de la educación. Con esto, nuestra intención no es limitarnos a describir aquellos proyectos *Big Data* enfocados en análisis de resultados académicos. Tampoco nos referimos a los análisis predictivos en cuanto a la persistencia estudiantil, todos ellos basados en el comportamiento mostrado por anteriores cohortes de estudiantes. En particular, los proyectos *Big Data* en el ámbito de la educación que queremos resaltar son aquellos posibles de ser catalogados como procesos de analítica del comportamiento de los estudiantes (Learning Analytics), especialmente dado el gran volumen de datos generados producto de la educación virtual.

En este orden de ideas, la primera conferencia de Learning Analytics y conocimiento (Learning Analytics and Knowledge) realizada en 2011 definió el concepto como: “la medición, recopilación, análisis y difusión de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce” (Ferguson & Clow, 2017). A partir de esta definición se puede afirmar que Learning Analytics busca lograr mejoras en la práctica del aprendizaje.

Por tanto, una forma de comprender las características esenciales que debe poseer un proyecto de Learning Analytics es basarse en la propuesta de Chatti, Dyckhoff, Schroeder y Thüs (Chatti *et al.*, 2012) quienes definen un modelo que considera cuatro dimensiones a saber:

- a. ¿Qué? ¿Qué tipo de datos recopila, gestiona y utiliza el sistema para el análisis?
- b. ¿Quién? ¿A quién se dirige el análisis?
- c. ¿Por qué? ¿Por qué el sistema analiza los datos recopilados?
- d. ¿Cómo? ¿Cómo realiza el sistema el análisis de los datos recopilados?

Este modelo, mostrado en la figura 2, permite identificar los elementos que se deben incluir en un proyecto de Learning Analytics. Al estar en línea con la definición de Learning Analytics ya presentada permite enfocar este tipo de proyectos en mejorar el aprendizaje que se produce considerando el contexto donde ocurre, sin limitarse a un análisis estadístico descriptivo.

Dado lo anterior, el Learning Analytics tiene como misión el análisis del desempeño de cada estudiante. Este desempeño se considera tanto de forma individual, como también en relación con el entorno social al que está expuesto dicho estudiante (Zenteno, 2018; Hasbún, 2019; Reyes *et al.*, 2018; Gasevic, 2017).

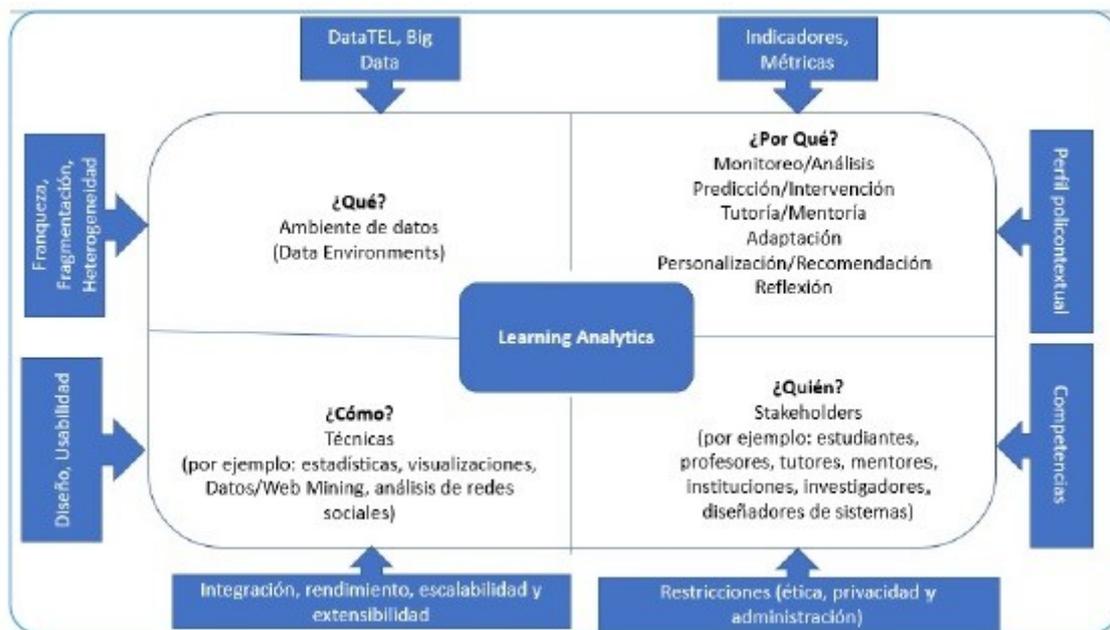


Figura 2. Modelo de Referencia de Learning Analytics. **Fuente:** (Chatti *et al.*, 2012)

Ahora bien, Learning Analytics es útil para resaltar otras características asociadas al *Big Data*, puesto que no nos limitamos a considerar datos estructurados, como pueden ser las calificaciones de los estudiantes, la asistencia, etc. Por el contrario, también incluimos aquellos no estructurados, como serían los comentarios en tableros de discusión que se puedan realizar o la publicación de anuncios, entre otros. Así, poder analizar la interacción que tiene el alumno con sus compañeros, la amplitud del vocabulario que utiliza y el uso correcto del idioma, son una vía para comprender a los estudiantes logrando determinar cuáles son sus carencias y fortalezas, entre otras. Finalmente, este tipo de proyecto también logra resaltar la velocidad en la que se generan datos en contexto de pandemia, ya que los estudiantes pasaron más tiempo en las plataformas virtuales de sus escuelas o universidades, algo que nunca había ocurrido (Ehrlich *et al.*, 2020; Chatziralli *et al.*, 2020; Adnan y Anwar, 2020). Junto a lo anterior, la calidad de la información que considera este tipo de proyecto sin duda aporta valor, puesto que se trata de información totalmente veraz y libre de sesgo dado que no se trabaja con muestras poblacionales sino que con el universo de los datos.

Cerrando este apartado, cabe destacar el aspecto ético en el tratamiento de los datos, ya que estamos trabajando con información relativa al comportamiento de las personas. Así, el marco de trabajo para el *Big Data* deberá cautelar el desarrollo de dimensiones que incorporen estrategias institucionales, adopción de herramientas tecnológicas y, la generación de procesos bien definidos para el tratamiento ético de los datos personales de nuestros estudiantes (Pérez-Sanagustín *et al.*, 2019). Expresado lo anterior, resulta de toda lógica reflexionar respecto a la preparación de las instituciones de educación para utilizar los grandes conjuntos de datos que generan los estudiantes (Maldonado-Mahauad *et al.*, 2018).

V. Discusión

Luego de analizar las distintas definiciones en torno a Big Data, se destaca que la característica asociada al tamaño del conjunto de datos es uno de los puntos en tensión. En esta línea, el concepto de volumen (implícito en *Big Data*) se transforma en un campo de batalla, encontrando autores que exigen para el concepto *Big Data* volúmenes de datos con órdenes de magnitud distintos. Así, Punam Bedi (Bedi *et al.*, 2014) indica que *Big Data* se refiere al uso de datos en el rango de zettabytes (1.000.000 de Petabytes) y yottabytes (1.000.000.000.000 de Petabytes). Sin embargo, con una distancia de 12 meses, Pedro Neves (Neves y Bernardino, 2015) señala que *Big Data* incluye el almacenamiento y procesamiento de conjuntos de datos en el orden de Petabytes. Pese a las diferencias de magnitud, ambas investigaciones señalan la necesidad de utilizar procesamiento distribuido, esto debido a la incapacidad de los sistemas disponibles para procesar esos volúmenes de información.

Lo anterior no es trivial, ya que no solo se trata de tener una métrica o indicador único para definir *Big Data*, más bien la discusión logra develar que la definición del concepto estaría asociado a la capacidad de la infraestructura tecnológica de cada proyecto. Si se posee poca capacidad de procesamiento en función de los datos disponibles, se asume la necesidad de utilizar la tecnología de *Big Data*, lo que resulta sesgado para concluir al respecto.

Adicionando a lo ya dicho y, con relación a la capacidad de procesamiento, (Joyanes, 2013) señala claramente que *Big Data* es un concepto que se asocia a los conjuntos de datos cuyo volumen lo aleja de las capacidades de procesamiento "... de las herramientas informáticas (computación) de uso común para, capturar, gestionar y procesar datos en un lapso razonable" (Joyanes, 2013, p. 3). En esta mención se puede apreciar un equilibrio entre las características del conjunto de datos y la infraestructura necesaria para procesarlo.

Transcurrida toda esta exposición, es dable cuestionar el grado de solidez que presenta el concepto de *Big Data*, en particular considerando su definición sustentada en la arquitectura que define un proyecto de esta índole.

Así, la propuesta del NIST (2015) para categorizar un proyecto como *Big Data*, gira en torno a una arquitectura basada en la computación distribuida. Esta definición está sustentada en la capacidad de cómputo de este tipo de infraestructura, la que estaría más allá del almacenamiento y procesamiento que entregaría la computación clásica. Dicho esto, y basándose exclusivamente en el procesamiento de los datos, es posible considerar que este tipo de definición sería temporal y poco robusta.

VI. Conclusiones

6.1. Conceptos iniciales

Luego de todos los antecedentes expuestos, es posible señalar que el concepto de *Big Data* es un término multivariable y difuso. Es aquí donde encuentra asidero la existencia de variadas definiciones para el concepto. Algunas de ellas basadas en la infraestructura y otras tantas en las características de los datos a procesar. Sin perjuicio de lo señalado, es posible afirmar que un proyecto *Big Data* deberá cumplir con tres premisas para ser catalogado como tal; cada una de ellas enlazada con la anterior.

- a. La velocidad con que se generan los datos es un factor que debe llamar la atención.
- b. Lo anterior está motivado por el volumen de datos que se generará en un corto periodo de tiempo.
- c. A partir de los dos puntos precedentes, la infraestructura tradicional no será suficiente para su correcto procesamiento.

Con todo esto, nuestra definición de *Big Data* es “un concepto que permite describir un conjunto de datos que, dada su velocidad de generación, logra alcanzar una magnitud tal que solo es posible de ser procesado vía el uso de la computación distribuida”. A partir de esta propuesta de definición para *Big Data* es necesario destacar que la variedad de los datos procesados no es un requisito obligatorio para que un proyecto sea considerado de este tipo. Esto porque la capacidad de cómputo no está directamente relacionada con la variedad de los datos sino más bien por un volumen que no es posible de procesar con una arquitectura tradicional. En otras palabras, un proyecto de *Big Data* lo sigue siendo aunque los datos procesados sean todos del mismo formato.

Respecto a Learning Analytics es importante destacar que este tipo de proyectos no se deben centrar en el análisis descriptivo estadístico sino considerar el contexto donde se realiza el aprendizaje de los

estudiantes. El enfoque principal debería estar en la inclusión del comportamiento de los estudiantes en el entorno de aprendizaje, especialmente en relación con sus pares y profesores.

6.2. Trabajos futuros

En cuanto a las líneas de trabajo a futuro, el desarrollo del *Big Data* plantea enormes desafíos relativos a la formación de los profesionales capaces de enfrentarlos. Este contexto permite visualizar tres líneas de investigación: 1) La pertinencia de incluir esta temática dentro de la formación de los ingenieros de datos. 2) Las dificultades que presenta la enseñanza de las técnicas asociadas al *Big Data*. Y, 3) La infraestructura pertinente para lograr los aprendizajes y desarrollo de competencias, propicias para habilitar a los futuros profesionales de esta área.

VII. Referencias

- Adnan, M., y Anwar, K. (2020). Online learning amid the COVID-19 pandemic: Students' perspectives. *Journal of Pedagogical Sociology and Psychology*, 2(1), 45-51. <https://doi.org/10.33902/JSPS.2020261309>
- Akram, W., Varalakshmi, M., Sudeepthi, J. (2019) Big data analytics current status, challenges and connection of unbounded data processing platforms. *International Journal of Innovative Technology and Exploring Engineering*. 8(9S2),698-700. <https://doi.org/10.35940/ijitee.I1144.0789S219>
- Al-Qirim, N., Tarhini, A., y Rouibah, K. (2017). Determinants of Big Data Adoption and Success. *ICACS '17: Proceedings of the International Conference on Algorithms, Computing and Systems* August 2017. 88-92. <https://doi.org/10.1145/3127942.3127961>
- Amatriain, X. (2013). Big & personal: data and models behind Netflix recommendations. En *Proceedings of the 2nd international workshop on big data, streams and heterogeneous source Mining: Algorithms, systems, programming models and applications*, 1-6.
- Ambigavathi, M., y Sridharan, D. (2020). A Survey on Big Data in Healthcare Applications. En *Intelligent communication, control and devices*. Springer. https://doi.org/10.1007/978-981-13-8618-3_77
- Bedi, P., Jindal, V., y Gautam, A. (2014). Beginning with Big Data Simplified. *International Conference on Data Mining and Intelligent Computing (ICDMIC)*, 1-7.
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering: An Illustrated Weekly Journal*, 137, 85-86. <https://doi.org/10.1177/016555158501000407>
- Carvajal, J. A. (2004). Guía para el análisis crítico de publicaciones científicas. *Revista chilena de obstetricia y ginecología*, 69(1), 67-72.
- Chatti, M., Dyckhoff, A., Schroeder, U., & Thüs, H. (2012). A Reference Model for Learning Analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 1-22.

- Chatziralli, I., Ventura, C. V., Touhami, S., Reynolds, R., Nassisi, M., Weinberg, T., Pakzad-Vaezi, K., Anaya, D., Mustapha, M., Plant, A., Yuan, M. y Loewenstein, A. (2020). Transforming ophthalmic education into virtual learning during COVID-19 pandemic: a global perspective. *Eye*, 35, 1-8.
- Chew, L.-J., Haw, S.-C., y Subramaniam, S. (2020). Recommender System for Retail Domain: An Insight on Techniques and Evaluations. *ICCMS '20*, 9-13.
- Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. *1st Workshop on Social Media Analytics (SOMA '10)*, 115-122.
- Cuzzocrea, A. (2020). Uncertainty and Imprecision in Big Data Management: Models, Issues, Paradigms, and Future Research Directions. *ICCBDC'20*, 4-9.
- Ehrlich, H., McKenney, M., y Elkbuli, A. (2020). We asked the experts: Virtual learning in surgical education during the COVID-19 pandemic— Shaping the future of surgical education and training. *World journal of surgery*, 44, 2053-2055.
- Emmanuel, I., y Stanier, C. (2016). Defining Big Data. *BDAW '16: Proceedings of the International Conference on Big Data and Advanced Wireless Technologies* 1-6. <https://doi.org/10.1145/3010089.3010090>
- Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. *LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 56-65.
- García Montero, A. C., Aguaded, I., & Ferrés, J. (2017). Competencia mediática organizacional: una revisión sistemática de literatura científica en Web of Science. *Dixit*, 27, 61-74. <https://doi.org/10.22235/d.v0i27.1497>
- Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science*, 178(1), 471-479. <https://doi.org/10.1126/science.178.4060.471>
- Gasevic, D. (2018). Directions for adoption of learning analytics in the global south. In C. Ping Lim y V. L. Tinio (Eds.), *Learning analytics for the global south*. Quezon City, Philippines: Foundation for Information Technology Education and Development. <https://digital.fundacionceibal.edu.uy/jspui/handle/123456789/243>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., y Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012-1014. <https://doi.org/10.1038/nature07634>
- Gopalkrishnan, V., Steier, D., Lewis, H., y Guszczka, J. (2012). Big Data, Big Business: Bridging the Gap. *BigMine'12*, 7-11. <https://doi.org/10.1145/2351316.2351318>
- Greenhalgh, T. (1997). How to read a paper: Assessing the methodological quality of published papers. *Bmj*, 315, 305-308. <https://doi.org/10.1136/bmj.315.7103.305>
- Hasbún, J. C., Torres-Toukoumidis, A., Bello, R., y Flores, Á. (2019). Analíticas de aprendizaje en la territorialidad virtual de la enseñanza: caso República Dominicana. *HAMUT'AY*, 6(3), 98-113.

- Joyanes Aguilar, L. (2013). Big Data: análisis de grandes volúmenes de datos en organizaciones. Alfaomega.
- Khan, N., Naim, A., Rashid Hussain, M., Noorulhasan Naveed, Q., Ahmad, N., y Qamar, S. (2019). The 51 V's of Big Data. ACM Omni-layer Intelligent Systems Conference, 1-6.
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. Application Delivery Strategies. META Group. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Leung, C., Jiang, F., y Zhang, Y. (2019). Flexible Compression of Big Data. ASONAM'19, 741-748. <https://doi.org/10.1145/3341161.3343512>
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323. <https://doi.org/10.2307/24529203>
- Maldonado-Mahauad, J., Hilliger, I., De Laet, T., Millecamp, M., Verbert, K., Ochoa, X., & Pérez-Sanagustín, M. (2018, March). The LALA project: Building capacity to use learning analytics to improve higher education in Latin America. In companion proceedings of the 8th international learning analytics & knowledge conference, 630-637.
- Manogaran, G., y Lopez, D. (2018). Spatial cumulative sum algorithm with big data analytics for climate change detection. *Computers y Electrical Engineering*, 65, 207-221.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., y Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- Mayer-Schönberger, V., y Cukier, K. (2013). Big Data, la revolución de los datos masivos. Turner.
- National Institute of Standards and Technology NIST. (2015). NIST Big Data Interoperability Framework: Volume 1, Definitions. Gaithersburg: NIST.
- Neves, P., y Bernardino, J. (2015). Big Data Issues. IDEAS '15: Proceedings of the 19th International Database Engineering & Applications Symposium. 200-201. <https://doi.org/10.1145/2790755.2790785>
- Nolan, R. L. (1973). Managing the computer resource: a stage hypothesis. *Communications of the ACM*, 16(7), 399-405.
- Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4), 431-448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Pérez-Sanagustín, M., Hilliger, I., Maldonado-Mahauad, J., Pérez-Álvarez, R., Ramirez, L. N., Muñoz-Merino, P. J., Tsai, Y., Ortiz-Rojas, M., Broos, T., Zúñiga-Prieto, M. A., Sheihing, E. & Whitelock-Wainwright, A. (2019). Building capacity to use learning analytics to improve higher education in Latin America: LALA Framework.
- Rana, A. I., & Mugavero, M. J. (2019). How Big Data Science Can Improve Linkage and Retention in Care. *Infectious Disease Clinics of North America*, 33(3), 807-815. <https://doi.org/10.1016/j.idc.2019.05.009>

- Raza, M. U., & Xujian, Z. (2020, May). A Comprehensive Overview of BIG DATA Technologies: A Survey. In Proceedings of the 2020 5th International Conference on Big Data and Computing, 23-31.
- Reyes, N. S., Morales, J. B., Moya, J. G., Teran, C. E., Rodriguez, D. N., y Altamirano, G. C. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologías de Información*, 17, 258-266.
- Saaksjarvi, M. (1985). End-user participation and the evolution of organizational information systems: an empirical assessment of Nolan's stage model. *SIGCPR '85: Proceedings of the twenty-first annual conference on Computer personnel research*, 181-190.
- Shafer, J., Rixner, S., & Cox, A. L. (2010). The Hadoop Distributed Filesystem: Balancing Portability and Performance. *IEEE International Symposium on Performance Analysis of Systems y Software (ISPASS)*, 122-133.
- Sogodekar, M., Pandey, S., Tupkari, I., & Manekar, A. (2016). Big data analytics: Hadoop and tools. *2016 IEEE Bombay Section Symposium (IBSS)*, 1-6. <https://doi.org/10.1109/IBSS.2016.7940204>
- Uddin, M. F., y Gupta, N. (2014, April). Seven V's of Big Data understanding Big Data to extract value. En Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education, 1-5. <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Uzunkaya, C., Ensari, T., & Kavurucu, Y. (2015). Hadoop Ecosystem and Its Analysis on Tweets. *Procedia - Social and Behavioral Sciences*, 195, 1890-1897. <https://doi.org/10.1016/j.sbspro.2015.06.429>
- Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, 367, 747-765. <https://doi.org/10.1016/j.ins.2016.07.007>
- Yu S, Liu M, Dou W, Liu X, Zhou S (2017) Networking for big data: a survey. *IEEE Communications Surveys & Tutorials* 19(1), 531-549. <https://doi.org/10.1109/COMST.2016.2610963>
- Zenteno, J. A. C. (2018). Big Data-Analítica del aprendizaje y minería de datos aplicados en la Universidad. *Pro Sciences: Revista de Producción, Ciencias e Investigación*, 2(8), 39-54. <https://doi.org/10.29018/issn.2588-1000vol2iss8.2018pp29-54>
- Zhou, J., Bruno, N., Wu, M.-C., Larson, P.-A., Chaiken, R., y Shakib, D. (2012). SCOPE: parallel databases meet MapReduce. *The VLDB Journal*, 21, 611-636. <https://doi.org/10.1007/s00778-012-0280-z>