

EL BIG DATA; OPORTUNIDADES, CASOS DE USO Y RETOS

BIG DATA; OPPORTUNITIES, USE CASES AND CHALLENGES

Eugenia PÉREZ-PONS^{1,2}, Javier PARRA-DOMÍNGUEZ^{1,2}, Sergio MARQUEZ^{1,2}, Sergio MANZANO¹, y Jorge HERRERA-SANTOS^{1,2}

¹ *BISITE Digital Innovation Hub, University of Salamanca, I+D+i Building, Calle Espejo, 2 - 37007, Salamanca, Spain*

{eugenia.perez, javierparra, smarquez, smanzano, jorgehsmp} @usal.es

² *IoT Digital Innovation Hub, Edificio Parque Científico, Módulo 305, Paseo de Belén, 11, Campus Miguel Delibes, 47011, Valladolid, Spain*

RESUMEN: Cada año aumentan el número de artículos e investigaciones realizadas en el campo del Big Data, así como el impacto que ha tenido en las diferentes industrias. Además de todos los avances que ha permitido esta nueva forma de utilizar los datos, también se recogen los diferentes retos que plantea esta nueva forma de captura y uso de los datos y de la información.

PALABRAS CLAVE: big data; captura de datos; procesamiento de datos.

ABSTRACT: Every year, the number of articles and research in the field of Big Data is increasing, as is the impact it has had in the area of Big Data, as well as the impact it has had on different industries, are growing every year. In addition to all the advances in this new way of using data, the various challenges this new way of capturing and using data are also this new way of capturing and using data and information.

KEYWORDS: big data; captura de datos; minería de datos.

1 Introducción

En el siguiente documento se describe cómo el big data y otros factores que intervienen en el crecimiento de los determinados sectores de la economía. Jin et al [1] recogieron los principales retos que afronta el big data en diferentes ámbitos, tanto en mercados e industrias que fomentan el desarrollo nacional como a nivel investigación. O'Neil [2] incluso afirmó que «el futuro pertenece a las empresas y personas que convierten los datos en productos». En el presente documento, a partir de la cuestión principal de identificar si el big data es útil para todos los sectores, se analizan diferentes palancas que convergen en este nuevo paradigma. Una posible aplicación de los datos del sector privado es la creación de estadísticas sobre la actividad económica agregada que pueden utilizarse para hacer un seguimiento de la economía o como insumos para otras investigaciones [3].

Como se recogió en el documento de IBM ya en 2013 [4], se han registrado más datos en los últimos dos años que en toda la historia humana anterior. Desde la perspectiva de la industria de la información, el Big Data constituye un fuerte impulso para la próxima generación de la industria de la tecnología de la información, que se basa esencialmente en la tercera plataforma, y se refiere principalmente a los grandes datos, la computación en nube, la Internet móvil y los negocios sociales.

El big data y las diferentes arquitecturas, así como posibilidades de tratar la información ha evolucionado desde sus inicios hasta ahora como se muestra en la figura 1.



Fig 1. Evolución Big Data.

Hay diferentes sectores que han sufrido muchos cambios y han aumentado su potencial de crecimiento debido al big data [5]. Entre ellos, en este artículo se destacan sectores que de forma diferente hacen uso del big data; como el

supply chain [6], el Healthcare [7], el Retail [8] y casos de uso concretos en el financiero y construcción [9].

El presente documento se organiza de la siguiente forma; en el apartado 2 se recogen diferentes casos de uso y estudios realizados en sectores de actividad concretos, el punto 3 los desafíos y retos generales que plantea el Big data y finalmente el último apartado recoge las conclusiones derivadas de los apartados anteriores.

2 Principales Industrias

El presente documento se organiza de la siguiente forma; en el apartado 2 se recogen diferentes casos de uso y estudios realizados en sectores de actividad concretos, el punto 3 los desafíos y retos generales que plantea el Big data y finalmente el último apartado recoge las conclusiones derivadas de los apartados anteriores.

2.1 Supply Chain

El *supply chain*, también denominado cadena de suministro, ha progresado hacia un entendimiento común de la Resiliencia [10]. MacCarthy et al, definieron el ciclo de vida de la cadena de suministro e identificaron seis factores que interactúan y pueden afectar a una cadena de suministro a lo largo de su ciclo de vida: la tecnología y la innovación, la economía, los mercados y la competencia, la política y la reglamentación, la adquisición y el abastecimiento, las estrategias de la cadena de suministro y la reingeniería [11]. Por otro lado, autores como Waller et al [12] sugieren que; El análisis predictivo de la logística utiliza métodos cuantitativos y cualitativos para estimar el comportamiento pasado y futuro del flujo y el almacenamiento del inventario, así como los costos asociados y los niveles de servicio. El análisis predictivo de Supply Chain Management utiliza métodos tanto cuantitativos como cualitativos para mejorar el diseño de la cadena de suministro y la competitividad mediante la estimación de los niveles pasados y futuros de integración de los procesos empresariales entre las funciones o empresas, así como los costos asociados y los niveles de servicio.

2.2 Healthcare

En una revisión del estado del arte e investigación en el campo del big data en healthcare llevada a cabo en 2017, se concluyó que; «a) *En la etapa inicial, los investigadores de los Estados Unidos, la República Popular China, el Reino Unido y Alemania fueron los que más contribuyeron a la bibliografía relacionada con la investigación de grandes datos sobre la atención de la salud y la vía de la innovación en este ámbito.* b) *La vía de la innovación en los grandes datos de la atención de la salud consta de tres etapas: la fase de detección temprana, diagnóstico, tratamiento y pronóstico de la enfermedad, la fase de promoción de la vida y la salud, y la fase de enfermería.* c) *Los puntos calientes de la investigación se concentran principalmente en tres dimensiones: la dimensión de la enfermedad»* [13].

Por otro lado, [14] desarrollaron un estudio en el que se elaboró un marco que pudiera utilizarse para orientar el análisis que puede traducir y transformar grandes datos en un recurso más útil y útil para los profesionales de la salud, mejorando así los servicios que prestan a los pacientes. Como Margolis definió «Los grandes datos no son sólo una nueva realidad para el científico biomédico, sino un imperativo que debe ser comprendido y utilizado eficazmente en la búsqueda de nuevos conocimientos» [15].

2.3 Retail

La venta al por menor es una de las actividades comerciales más destacadas y diversificadas del mundo, que ha transformado considerablemente las estrategias comerciales para obtener más beneficios. Hoy en día, la definición de venta al por menor es sinónimo de tiendas de mercancías atractivas y adecuadamente gestionadas con una comodidad y un ambiente increíbles, en lugar de las tiendas tradicionales apiladas al azar. Además, el cliente moderno está enfocado hacia la calidad/marcas y espera que los servicios que le entregan los diferentes vendedores le lleguen a su casa con un solo clic. Como resultado, los clientes prefieren comprar en varios sitios web de compras en línea en lugar de trasladarse físicamente a una tienda minorista, lo que a su vez lleva a la caída de las ventas de los minoristas que se ha convertido en una importante amenaza para ellos.

McArthur et al [16] realizaron una revisión de la literatura en cuanto a las diferentes aplicaciones del big data en retail. Esta revisión de la literatura aplica un enfoque macro y teoría de sistemas a la literatura multidisciplinaria, y vincula los cuerpos de trabajo que, hasta ahora, han permanecido conceptualmente desconectados. Esto proporciona una meta tipología de seis factores que podrían explicar el cambio en la venta al por menor: eficiencias económicas, patrones cíclicos, desigualdades de poder, comportamiento innovador, influencias ambientales y partes interdependientes del sistema en co-evolución.

3 Retos del Big Data

El paradigma del big data aún tiene muchas incógnitas y retos a los que sobreponerse. Sivarajah et al [17] condujeron un estudio en el que se ponían de manifiesto los diferentes retos y oportunidades que presentaba el big data en la actualidad. Así mismo, Hairi et al [18] recoge los temas relacionados con las cinco V de los grandes datos, sin embargo, existen muchas otras V. En términos de la investigación existente, se ha prestado mucha atención al volumen, variedad, velocidad y veracidad de los datos, con menos trabajo disponible en valor (por ejemplo, los datos relacionados con los intereses corporativos y la toma de decisiones en dominios específicos).

3.1 Seguridad y Privacidad

A pesar de las grandes ganancias en materia de datos, existen también numerosos desafíos y entre ellos el de mantener la privacidad de los datos es la preocupación más importante en las grandes aplicaciones de extracción de datos, ya que el procesamiento en gran escala de conjuntos de datos delicados, como los registros de salud y los registros de transacciones bancarias, debe mantenerse de manera que los datos privados no se revelen a ninguna persona no autorizada. Sharma et al [19] recogen y examinan los problemas que plantea el big data así como los problemas que plantea la extracción de datos en gran escala y la preocupación por la privacidad de los grandes datos.

4. Conclusiones

Como se muestra a lo largo del documento, hay diferentes factores que contribuyen al uso del big data, así como el uso óptimo de los recursos ofrecidos. Siendo una nueva forma de entender los datos, y habiendo aparecido hace más de 10 años, todavía se encuentra en una fase muy temprana en la que sigue habiendo muchos retos y desafíos por solventar, así como diferentes formas para poder alcanzar máximos rendimientos.

Agradecimientos

This publication has been produced within the framework of the DIH World project which is co-funded by Horizon 2020 Framework Programme of the European Union Under grant agreement N.º 952176

Referencias

1. Jin, X., Wah, B.W., Cheng, X., Wang, Y.: Significance and challenges of big data research. *Big Data Research*, 2(2), 59-64 (2015)
2. O'Neil, C., Schutt, R.: *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc. (2013)
3. Einav, L., Levin, J.: Economics in the age of big data. *Science*, 346(6210).Source: IBM, <http://www-01.ibm.com/software/data/bigdata/> accessed March 27, 2013 (2014)
4. Lee, I.: Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303 (2017)
5. Tiwari, S., Wee, H.M., Daryanto, Y.: Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers Industrial Engineering*, 115, 319-330 (2018)
6. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2(1), 3 (2014)
7. Lovelace, R., Birkin, M., Cross, P., Clarke, M.: From big noise to big data: toward the verification of large data sets for understanding regional retail flows. *Geographical Analysis*, 48(1), 59-81 (2016)

8. Parra, J., Pérez-Pons, M.E., González, J.: Technology as a Lever for the Evolution and Recovery of the Financial and Construction Sectors in Spain. In *International Symposium on Distributed Computing and Artificial Intelligence*, pp. 169-174. Springer, Cham (2020)
9. Pettit, T.J., Croxton, K. L., Fiksel, J.: The evolution of resilience in supply chain management: a retrospective on ensuring supply chain resilience. *Journal of Business Logistics*, 40(1), 56-65 (2019)
10. MacCarthy, B.L., Blome, C., Olhager, J., Srari, J.S., Zhao, X.: Supply chain evolution-theory, concepts and science. *International Journal of Operations Production Management* (2016)
11. Waller, M.A., Fawcett, S.E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84 (2013)
12. Gu, D., Li, J., Li, X., Liang, C.: Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International journal of medical informatics*, 98, 22-32 (2017)
13. Iyamu, T., Mgudlwa, S.: Transformation of healthcare big data through the lens of actor network theory. *International Journal of Healthcare Management*, 11(3), 182-192 (2018)
14. R. Margolis, L. Derr, M. Dunn et al.: «The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data,» *Journal of the American Medical Informatics Association*, vol. 21, n.º 6, pp. 957-958 (2014)
15. McArthur, E., Weaven, S., Dant, R.: The evolution of retailing: A meta review of the literature. *Journal of Macromarketing*, 36(3), 272-286 (2016)
16. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286 (2017)
17. Hariri, R.H., Fredericks, E.M., Bowers, K.M.: Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, 6(1), 44 (2019)
18. Sharma, A., Singh, G., Rehman, S.: A review of big data challenges and preserving privacy in big data. In *Advances in Data and Information Sciences*, pp. 57-65. Springer, Singapore (2020)

