

# Aplicaciones Inteligentes sobre Internet de las Cosas y Grandes Volúmenes de Datos: Un Enfoque Riguroso

Fernando Asteasuain –Federico D’Angiolo - Manuel Dubinsky – Fernando Pazos – Ivan Kwist - Matias Loiseau-  
David Contreras – Federico Calonge  
Contacto: [fasteasuain@undav.edu.ar](mailto:fasteasuain@undav.edu.ar)  
Ing. en Informática – Dpto. Tecnología y Administración – Universidad Nacional de Avellaneda

## Resumen

Día tras días se generan millones y millones de nuevos datos y la cantidad de información a procesar es un desafío creciente. Entre los más destacados podemos mencionar el crecimiento exponencial de la “Internet de las cosas” o Internet of Things (IoT) en inglés en los sistemas modernos. En este sentido, no sólo el almacenamiento constante de información es un problema a resolver, sino también la extracción inteligente de información y su posterior análisis para introducir mejoras en los sistemas.

Este problema ha sido atacado desde la Inteligencia Artificial y la Optimización Combinatoria, pero se han encontrado algunas debilidades como la falta de modelado y diseño y de la aplicación rigurosa de técnicas de Ingeniería de Software. Esto hace que el problema se ataque de una manera “ad-hoc”, por lo cual es necesario consolidar un enfoque de mayor formalidad en todas las etapas del proceso.

La presente investigación busca aplicar técnicas formales de Ingeniería de Software como Modelado y Model Checking al manejo y análisis de grandes volúmenes de datos.

Como caso de estudio concreto se trabajará con sensores para el Mantenimiento de parámetros del ambiente (como humedad o temperatura) del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT.

**Palabras claves:** *Big Data, Procesamiento de Datos, Inteligencia Artificial, Métodos Rigurosos de Ingeniería de Software.*

## Contexto

El presente proyecto se encuentra enmarcado dentro del proyecto UNDAVCYT denominado “Aplicaciones Inteligentes sobre Internet de las Cosas y Grandes Volúmenes de Datos: Un Enfoque Riguroso”, dirigido por el Dr. Fernando Asteasuain y codirigido por el Ingeniero Federico D’Angiolo. El proyecto formó parte de la convocatoria UNDAVCY2019 y actualmente está en estado de revisión. El grupo de investigación está conformado por los autores del presente trabajo, combinando investigadores consolidados y en formación y estudiantes avanzados. El proyecto tiene una duración de dos años y está comenzando en este año. El proyecto estaría financiado 100% por la Universidad Nacional de Avellaneda en caso de ser otorgado. El proyecto forma parte de un objetivo de investigación más grande dentro de la carrera de Ingeniería en Informática de la UNDAV ya que propone el lanzamiento del Laboratorio de Inteligencia Artificial. En este sentido la carrera ha proporcionado una importante cantidad de recursos para tal fin, destacándose una computadora de alto poder de procesamiento en vistas a consolidar no sólo la investigación de la carrera sino también la posibilidad de transferencia y servicios al sector industrial y productivo. Las líneas de Investigación del proyecto tienen impacto en áreas prioritarias del *Plan Nacional de Ciencia y Técnica 2020* como AgroIndustria, Biodiversidad e Innovación Productiva.

## 1. Introducción

En los últimos años el desarrollo de sistemas para el procesamiento y análisis de grandes

volúmenes de datos (denominado Big Data) se ha transformado en un área trascendente para la Ingeniería de Software [1-8]. Como tal, la comunidad científica busca intensamente desarrollar herramientas y aplicaciones para extraer el máximo valor posible de los datos disponibles. Además, día tras días se generan nuevas fuentes de información, y a esto se le suma el hecho del crecimiento exponencial de la “Internet de las cosas” o Internet of Things (IoT) en inglés, la cual se define como la conexión de distintos equipos a una red para lograr la transmisión y recepción de datos, con el objetivo de gestionar distintos eventos [12]. Ejemplos de esto puede ser la automatización de refrigeración de un ambiente, siendo para esto, muy importante la utilización de sensores que permitan obtener datos de temperatura, humedad, presión o cualquier otra variable que sea necesaria.

No sólo el almacenamiento constante de información es un problema a resolver, sino también la extracción inteligente de información y su posterior análisis para introducir mejoras en los sistemas. Este problema ha sido atacado desde la Inteligencia Artificial [16-18,22] y la Optimización Combinatoria [13-15]. Propuestas como [22] exploran el análisis de Datos Exploratorio (EDA por sus siglas en inglés) basándose fuertemente en la minería de datos. En [16-18] presentan distintos enfoques para Big Data basado en redes neuronales y otras aplicaciones basadas en aprendizaje automático y redes neuronales.

Sin embargo, se han encontrado algunas debilidades como la falta de modelado y diseño y de la aplicación rigurosa de técnicas de Ingeniería de Software [1-8]. En trabajos como [4,6,8] se aplican técnicas formales de validación de software al proceso de BigData. Al estar basados en lógicas temporales hay lugar para introducir mejoras buscando lenguajes de especificación más expresivos y flexibles [11,4]. En trabajos como [1-4] se señalan debilidades en todo el proceso de desarrollo, buscando acercar la brecha entre la Ingeniería de Software y el análisis y

procesamiento de BigData. Si bien estos enfoques han tenido su impacto, la mayoría de estos desafíos continúan abiertos. También desde un punto de vista de arquitectura de software es necesario brindar soluciones que brinden un desarrollo de mayor calidad [9,10]. Por ejemplo, son necesarias soluciones que cuenten con una logística dedicada al procesamiento distribuido y considerando conceptos relacionados al Cloud Computing.

Esto hace que el problema se ataque de una manera “ad-hoc”, por lo cual es necesario consolidar un enfoque de mayor rigurosidad desde todas las etapas del proceso, el cual constituye el objetivo principal de la presente investigación.

El proyecto presentado buscar combinar métodos y técnicas rigurosas y formales de la Ingeniería de Software al procesamiento de grandes volúmenes de datos conocido como Big Data. El desafío involucra una sinergia entre diversas áreas como Inteligencia Artificial, Redes Neuronales, procesamiento dinámico de información incluyendo imágenes, sonido y diversas fuentes multimediales, procesamiento distribuido y Cloud Computing, Validación y Verificación Formal o Arquitecturas de Software.

Como se menciona en [1] el problema esencial de “Big Data” es que se está acumulando todo tipo de información a una velocidad superior a la cual puede ser procesada, y esta tendencia está creciendo día a día. Hay “cuatro V” que caracterizan a Big DATA:

- 1) Volumen: la cantidad creciente de información obtenida.
- 2) Variedad: Diversas y heterogéneas fuentes y tipos de dato.
- 3) Velocidad: para la manipulación y adquisición de datos, streaming en tiempo real, y datos de tiempo variable.
- 4) Veracidad: confianza en los datos obtenidos y su procesamiento.

Una solución a los mencionados problemas tendrá impacto en esta área consolidando métodos, técnicas y herramientas de la

Ingeniería de Software atacando cada una de las “cuatro V” que caracterizan a Big Data.

Se buscará en esta investigación aplicar conceptos de metodologías ágiles [19] al proceso de desarrollo y de investigación [20,21].

Como caso de estudio concreto, se trabajará con sensores para el Mantenimiento de parámetros del ambiente (como humedad o temperatura) del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT. Este caso de estudio permitirá aportar contribuciones en un área cada vez más preponderante denominada “Ciencia Ambiental”, la cual estudia modelos y herramientas de software aplicadas a una mejor conservación de los recursos energéticos de nuestro planeta [25,26]. El grupo de investigación ya viene de obtener resultados en el análisis del mencionado caso de estudio con técnicas de Inteligencia Artificial [27,28, 29].

## 2. Líneas de Investigación y Desarrollo

En el presente trabajo se explorarán las siguientes líneas de investigación:

- Aplicación de métodos formales de Ingeniería de Software al proceso de aprendizaje automático, redes neuronales, Big Data e IoT.
- Mejorar y aplicar nuevas variables (como por ejemplo la presión, además de la temperatura y humedad) a la aplicación del algoritmo *KNN* para el análisis de datos del caso de estudio [27, 28, 29].
- Analizar el problema de *overfitting* y *underfitting* en algoritmos de Regresión. En particular explorar si existe un polinomio que aproxima mejor a los datos obtenidos [27-29].
- Protocolos, modelado y verificación formal en IoT.

- Exploración de Redes Neuronales para la predicción de series temporales. Por ejemplo, redes que puedan realizar una predicción de cómo se puede comportar la temperatura (o humedad) del caso de estudio a utilizar.
- Análisis y Estudio del impacto de la utilización de aires acondicionados en el caso de estudio a utilizar. Dicha información puede ser relevante en diferentes entornos para detectar fallas o mal funcionamiento de los equipos que controlan la temperatura del cuarto refrigerado. También puede ofrecer un informe sobre la energía utilizada y recursos con el fin de estimar la huella ambiental que tiene la actividad desarrollada. Puede también extender la vida útil de los servidores manteniendo un registro controlado de operación dentro del rango de temperatura aconsejado por el fabricante.
- Exploración y utilización de Redes Neuronales Convolucionales.

## 3. Resultados Obtenidos/Esperados

El objetivo principal del presente proyecto de Investigación es aplicar métodos y técnicas rigurosas y formales de la Ingeniería de Software al procesamiento de grandes volúmenes de datos conocido como Big Data.

El desafío involucra combinar diversas áreas como Inteligencia Artificial, Redes Neuronales, procesamiento dinámico de información incluyendo imágenes, sonido y diversas fuentes multimediales, procesamiento distribuido y Cloud Computing, Validación y Verificación Formal o Arquitecturas de Software.

Como objetivos específicos podemos mencionar:

1. Investigar técnicas actuales para el manejo de grandes volúmenes de datos, incluyendo texto, imagen y sonido.

2. Analizar modelos de software y matemáticos que sean aplicables al procesamiento y análisis de información.
3. Relevar técnicas y herramientas utilizadas para el procesamiento y análisis de datos.
4. Aplicar herramientas y técnicas formales de la Ingeniería de Software para el desarrollo de sistemas de Big Data.
5. Consolidar técnicas para el procesamiento de imágenes y sonido.
6. Explorar Arquitecturas de Software que puedan llevar adelante procesos adecuados de BigData, considerando entornos distribuidos y de Cloud Computing.
7. Aplicar los modelos y conceptos teóricos desarrollados para el Mantenimiento de parámetros del ambiente del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT.
8. Incorporar el procesamiento de imágenes obtenidas desde las cámaras dentro del Laboratorio de Redes y Sistemas de Computación.
9. Incorporación de más fuentes de información además de humedad y temperatura.
10. Extrapolar los resultados intermedios y finales para el mencionado Laboratorio a otras áreas del Dpto. de Tecnología y Administración, como otros Laboratorios, aulas y otras dependencias.
11. Extrapolar los resultados obtenidos más allá de la Universidad Nacional de Avellaneda, y buscar que sean aplicables para cualquier contexto y ubicación.
12. Divulgar los resultados de la investigación en congresos y revistas científicas de interés para los temas de la investigación.
13. Consolidar los recursos humanos en inicios de las tareas de investigación como estudiantes avanzados de la carrera, mediante la participación activa en el proyecto o mediante la realización de tesis de final de carrera de grado.

14. Profundizar el área de investigación dentro del Dpto. de Tecnología y Administración.

Los resultados esperados incluyen:

- Lanzamiento del Laboratorio de Inteligencia Artificial de la carrera de Ingeniería en Informática.
- Consolidar y profundizar técnicas formales de la Ingeniería de Software en un área de aplicación de vanguardia como Big Data e Internet de las Cosas. Como caso concreto de estudio se aplicarán los resultados en el Laboratorio de Redes y Sistemas de Computación, buscando optimizar parámetros como humedad, temperatura, y explorando nuevos parámetros a través de imágenes de cámara web y sonidos de ambiente. Esto dará la posibilidad directa de aplicar y validar los contenidos teóricos desarrollados en un caso real de importancia. Luego, es esperable el análisis de extrapolación de los resultados obtenidos a un contexto que trascienda el caso de estudio
- Potenciar la transferencia desde el grupo de investigación al sector productivo e industrial al proveer servicios y experiencia en el manejo de grandes volúmenes de datos, redes neuronales y métodos formales de Ingeniería de Software.
- Dirección de tesis de licenciatura y supervisión de prácticas profesionales (PPS).

#### **4. Formación de Recursos Humanos**

El equipo de investigación cuenta con dos investigadores formados (director y codirector), y dos investigadores con dedicación exclusiva en proceso de formación. Además, incluye 4 estudiantes avanzados que están realizando su tesis de investigación dentro del marco del presente proyecto. Dichas tesis de grado se encuentran en diferentes estadios, desde más avanzadas a algunas todavía en fase de exploración.

#### **5. Bibliografía**

- [1] Embley, David W., and Stephen W. Liddle. "Big data—conceptual modeling to the rescue." *International Conference on Conceptual Modeling*. Springer, Berlin, Heidelberg, 2013.
- 2) van der Aalst, Wil, and Ernesto Damiani. "Processes meet big data: Connecting data science with process science." *IEEE Transactions on Services Computing* 8.6 (2015): 810-819.
- 3) Arndt, Timothy. "Big Data and software engineering: prospects for mutual enrichment." *Iran Journal of Computer Science* 1.1 (2018): 3-10.
- 4) Camilli, Matteo. "Coping with the State Explosion Problem in Formal Methods: Advanced Abstraction Techniques and Big Data Approaches." (2015).
- 5) Hummel, Oliver, et al. "A collection of software engineering challenges for big data system development." *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 2018.
- 6) Bellettini, C., Camilli, M., Capra, L., & Monga, M. (2016). Distributed CTL model checking using MapReduce: theory and practice. *Concurrency and Computation: Practice and Experience*, 28(11), 3025-3041.
- 7) Kumar, Vijay Dipti, and Paulo Alencar. "Software engineering for big data projects: Domains, methodologies and gaps." *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016..
- 8) Camilli, Matteo. "Formal verification problems in a big data world: towards a mighty synergy." *Companion Proceedings of the 36th International Conference on Software Engineering*. ACM, 2014.
- 9) Bakshi, Kapil. "Considerations for big data: Architecture and approach." *2012 IEEE Aerospace Conference*. IEEE, 2012.
- 10) Marz, Nathan, and James Warren. *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co., 2015.
- 11) Asteasuain, Fernando, and Víctor Braberman. "Declaratively building behavior by means of scenario clauses." *Requirements Engineering* 22.2 (2017): 239-274.
- 12) Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey." *Computer networks* 54.15 (2010): 2787-2805.
- 13) Hu, C., Xu, Z., Liu, Y., Mei, L., Chen, L., & Luo, X. (2014). Semantic link network-based model for organizing multimedia big data. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 376-387.;
- 14) Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation.
- 15) *Big Data & Society*, 2(2), 2053951715608406.; Xu, Z., Liu, Y., Yen, N., Mei, L., Luo, X., Wei, X., & Hu, C. (2016). Crowdsourcing based description of urban emergency events using social media big data. *IEEE Transactions on Cloud Computing*.
- 16) Kasun, L. L. C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). Representational learning with extreme learning machine for big data. *IEEE intelligent systems*, 28(6), 31-34;
- 17) Chen, X. W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. *IEEE access*, 2, 514-525; Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015).
- 18) Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1; Aggarwal, C. C. (Ed.). (2014). *Data classification: algorithms and applications*. CRC press.
- 19) Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Kern, J. (2001). *Manifesto for agile software development*.
- 20) ITAS: Highsmith, J. A., & Highsmith, J. (2002). *Agile software development ecosystems*. Addison-Wesley Professional;
- 21) Cockburn, A., & Highsmith, J. (2001). Agile software development: The people factor. *Computer*, (11), 131-133
- 22) Yu, Chong Ho. "Exploratory data analysis." *Methods* 2 (1977): 131-160.
- 23) F. Asteasuain, F. Calonge, F. Diaz, F. Dangiolo, and P. Gamboa. Expressing early behavior specifications with branching visual scenarios. In *CONAIISI*, 2018.
- 24) F. Asteasuain, F. Calonge, and M. Dubinsky. Exploring specification pattern based behavioral synthesis with scenario clauses. In *2018 CACIC ISBN 978-950-658-472-6*, pages 22,34. CACIC, 2018.
- 25) Simm, William Alexander, et al. "Se in es: Opportunities for software engineering and cloud computing in environmental science." *Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Society*. ACM, 2018.
- 26) Hřebíček, Jiří, et al., eds. *Environmental Software Systems*. Computer Science for Environmental Protection: 12th IFIP WG 5.11 International Symposium, ISESS 2017, Zadar, Croatia, May 10-12, 2017, Proceedings. Vol. 507. Springer, 2018
- [27] Algoritmos de Regresión Lineal aplicados al mantenimiento de un Datacenter. Federico Gabriel D'Angiolo, Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Fernando Asteasuain. Universidad Nacional de Avellaneda, Avellaneda, Argentina. Departamento de Tecnología y Administración. Ingeniería en Informática. CACIC 2019.
- [28] Algoritmo de KNN Aplicado al Mantenimiento de un Datacenter. Federico Gabriel D'Angiolo, Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Gregorio Oscar Glas. CONAIISI 2019
- [29] Monitorización de un Datacenter mediante Protocolos de IoT. Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Federico Gabriel D'Angiolo, Roberto Osvaldo Mayer. CONAIISI 2019.