

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

**FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
BOGOTÁ D.C.**

LICENCIA CREATIVE COMMONS: Atribución-NoComercial-SinDerivadas 2.5 Colombia (CC BY-NC-ND 2.5)

AÑO DE ELABORACIÓN: 2018

TÍTULO: IMPLEMENTACIÓN DE UNA HERRAMIENTA DE AUTOGESTIÓN Y AUTOCONFIGURACIÓN PARA LA IMPLEMENTACIÓN DE SERVICIOS EN PROYECTOS DE BIG DATA.

AUTOR (ES):

Torres Robles, Edwin Andres y Rincón Saavedra, William Alonso.

DIRECTOR(ES)/ASESOR(ES):

Rincón Yáñez , Diego Alberto.

MODALIDAD:

Trabajo de investigación.

PÁGINAS: 156 **TABLAS:** 10 **CUADROS:** 0 **FIGURAS:** 24 **ANEXOS:** 3

CONTENIDO: Se refiere a los capítulos que se desarrollaron. Sólo los grandes capítulos. Ejemplo:

1. INTRODUCCIÓN.
2. GENERALIDADES.
3. OBJETIVOS DEL PROYECTO.
4. MARCO DE REFERENCIA.
5. METODOLOGÍA
6. DESARROLLO DEL PROYECTO.
7. ENTREGABLES DEL PROYECTO

RESUMEN ANALÍTICO EN EDUCACIÓN - RAE -



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

8. CONCLUSIONES
9. RECOMENDACIONES
10. TRABAJOS FUTUROS
11. BIBLIOGRAFÍA
12. ANEXOS

DESCRIPCIÓN:

Este proyecto busca analizar, investigar y facilitar información a los usuarios sobre las actuales herramientas gratuitas, que permiten la autogestión, autoconfiguración de la infraestructura y el manejo de datos en proyectos orientados a Big Data. Para llevar a cabo este proyecto se establece una metodología en cuatro fases y posteriormente se diseña el mecanismo a implementar, basándose en las arquitecturas Zeta, Kappa y Lambda; haciendo uso de ocho herramientas seleccionadas que se administran a través de Chef Server.

METODOLOGÍA: Bajo los Proyectos de Big Data , aun no hay un Estándar o Metodología aplicable, de acuerdo al enfoque del proyecto , se decidió realizar por etapas teniendo en cuenta las pautas de un proyecto de investigación, las cuales se enumeran a continuación, cabe resaltar que las etapas se basan , en la forma como la arquitectura de Big Data es concedida.

- Investigación y análisis de herramientas existentes.
- Diseño del modelo a implementar.
- Implementación de la herramienta seleccionada y Construcción de manuales de usuario y configuración.
- Pruebas y Resultado a la herramienta implementada.

PALABRAS CLAVE:

AUTOCONFIGURACIÓN, AUTOGESTIÓN, BIG DATA, CHEF SERVER, TI.



CONCLUSIONES:

- Al realizar esta investigación, se encontró que existen herramientas que facilitan la administración en las áreas TI, lo cual permite la optimización de los recursos y permite que las áreas de TI contribuyan al plan estratégico de una compañía u organización.
- Al hacer uso de herramientas de automatización, se disminuye el desgaste y el impacto que tiene implementar una infraestructura en Big Data, tanto en tiempo como en costos, lo cual permite a las compañías el escalamiento dinámico de las áreas de TI.
- El uso de Big Data es una tendencia a nivel Global y de manera especial en la actualidad en Colombia, por tanto adquirir conocimiento para este tipo de herramientas, permite explotar las cualidades que tiene este nuevo concepto de la ingeniería, maximizando el uso de los datos a nivel estratégico en las organizaciones.
- El uso de herramientas de Big Data, permite diseñar estrategias apoyadas en el estudio de los datos para la toma de decisiones.

FUENTES:

- [1] C. R. Blanco, "SQL básico, Página 2 Qué es SQL."
- [2] D. El Mundo, "Buscar en el sitio," pp. 1–7, 2014.
- [3] L. miguel Armendariz, "SOBRE EL TM CÓDIGO ABIERTO TM (OPEN SOURCE)," no. c. pp. 2003–2006, 2006.
- [4] "¿Qué es Business Intelligence?," 2012. [Online]. Available: http://www.sinnexus.com/business_intelligence/. [Accessed: 15-Nov-2017].
- [5] D. Evans, "The Internet of Things - How the Next Evolution of the Internet is Changing Everything," *CISCO white Pap.*, no. April, pp. 1–11, 2011.
- [6] "Sistema de SaaS (Software as a Service) para centros educativos."
- [7] J. C. C. Romani, "El concepto de tecnologías de la información."



- Benchmarking sobre las definiciones de las TIC en la sociedad del conocimiento,” *Zer - Rev. Estud. Comun.*, vol. 14, no. 27, pp. 285–318, 2009.
- [8] E. Alegría, A. Adrián, C. Demartini, P. E. Catrilef, and E. P. Zuleta, “NAT y su relación con IPv6.”
- [9] M. -Castilla León -Galicia -Levante, “Hacemos que las piezas encajen www.trc.es.”
- [10] A. L. Guillen, “Una introducción al ecosistema Hadoop.”
- [11] L. Al, “Bases de datos NoSQL,” 2011.
- [12] “Curso librerías Web 2.0.”
- [13] J. Gutiérrez, “Qué es un framework web?,” p. 1, 2006.
- [14] “INTRODUCCIÓN AL LENGUAJE HTML.”
- [15] G. Lehey and <grog@freebsd Org>, “Qué es BSD Tabla de contenidos.”
- [16] J. A. A. M. V. Universitat de Barcelona. Facultat de Biblioteconomia i Documentació., *BiD : textos universitaris de biblioteconomia i documentació*. Facultat de Biblioteconomia i Documentació, Universitat de Barcelona, 1999.
- [17] “IBM Almacenamiento en cloud | Cloud Business | IBM España.” [Online]. Available: https://www.ibm.com/cloud-computing/es-es/infrastructure/object-storage/?S_PKG=AW&cm_mmc=Search_Google_-_Leadership+Agenda_DIGITAL+INBOUND_-_ES_-_+ibm++services_Broad_AW&cm_mmca1=000009PY&cm_mmca2=10002507&mkwid=7995ce9a-6448-43f8-ad17-245d7cd64589%257C594%25. [Accessed: 19-Oct-2017].
- [18] J. Schroeck, Michael; Shockley, Rebecca; Smart, “Analytics: el uso de big data en el mundo real,” *IBM. Inf. Ejec.*, p. 22, 2012.
- [19] “Historia De Las Bases De Datos Timeline | Preceden.” [Online]. Available: <https://www.preceden.com/timelines/48236-historia-de--las-bases-de-datos>. [Accessed: 19-Oct-2017].
- [20] A. Muñoz Chaparro, *Oracle 11g SQL curso práctico de formación*. RC Libros, 2011.
- [21] W. R. Neuman and Y. J. I. N. Park, “Tracking the Flow of Information into the Home : An Empirical Assessment of the Digital Revolution in the United States , 1960 – 2005 University of Michigan,” vol. 6, pp. 1022–1041, 2012.
- [22] D. Boyd and K. Crawford, “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon,” *Inf. Commun. Soc.*, vol. 15, no. 5, pp. 662–679, Jun. 2012.
- [23] D. J. Power, “A Brief History of Decision Support Systems,” 2014.
- [24] T. Especial De Grado, P. Alfonso, and P. Muñoz, “Desarrollo de una Arquitectura Big Data para Registros Mercantiles.”
- [25] “Big Data – Historia cronológica | Winshuttle.” [Online]. Available:



- <https://www.winshuttle.es/big-data-historia-cronologica/>. [Accessed: 19-Oct-2017].
- [26] M. Cox and D. Ellsworth, "Application-controlled demand paging for out-of-core visualization," *Proceedings. Vis. '97 (Cat. No. 97CB36155)*, no. July, p. 235–244, 1997.
- [27] "Cómo el Internet de las cosas transformará su cadena de suministro - Winshuttle Spanish." [Online]. Available: <https://www.winshuttle.es/blog/como-el-internet-de-las-cosas-transformara-su-cadena-de-suministro/>. [Accessed: 19-Oct-2017].
- [28] "How Much Information Is 'Too Much Information'? - NYTimes.com." [Online]. Available: <https://mobile.nytimes.com/blogs/learning/2012/02/17/how-much-information-is-too-much-information/?referer=>. [Accessed: 19-Oct-2017].
- [29] J. A. Carrillo Ruiz *et al.*, "Big Data en los entornos de Defensa y Seguridad," *Inst. Español Estud. Estratégicos*, vol. 1, p. 124, 2013.
- [30] M. Olson, "HADOOP: Scalable, Flexible Data Storage and Analysis," *IQT Q.*, vol. 1, no. 3, pp. 14–18, 2010.
- [31] D. Reinsel *et al.*, "John F. Gantz, Project Director A Forecast of Worldwide Information Growth Through 2010," 2007.
- [32] "BIG DATA Y OPEN DATA: EL UNIVERSO DIGITAL DE DATOS timeline | Timetoast timelines." [Online]. Available: <https://www.timetoast.com/timelines/big-data-y-open-data-el-universo-digital-de-datos>. [Accessed: 21-Oct-2017].
- [33] R. Bryant, R. Katz, and E. Lazowska, "Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society," *Comput. Res. Assoc.*, pp. 1–15, 2008.
- [34] W. R. Neuman, Y. J. Park, and E. Panek, "Info Capacity| Tracking the Flow of Information into the Home: An Empirical Assessment of the Digital Revolution in the U.S. from 1960–2005," *Int. J. Commun.*, vol. 6, no. January, p. 20, 2012.
- [35] "A special report on managing information I," 2010.
- [36] S. Rogers, "Top 10 Trends in Business Intelligence and Analytics for 2011," pp. 1–3, 2011.
- [37] J. Rivera and R. van der Meulen, "Gartner Says 4.9 Billion Connected 'Things' Will Be in Use in 2015," *Gart. - Newsroom*, pp. 9–10, 2014.
- [38] D. Borthakur, "HDFS architecture guide," *Hadoop Apache Proj. http://hadoop.apache ...*, pp. 1–13, 2008.
- [39] R. Barranco, "¿Qué es Big Data?," 18-06-2012, 2012. [Online]. Available: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>.



- [Accessed: 21-Oct-2017].
- [40] A. McAfee and E. Brynjolfsson, “Big Data. The management revolution,” *Harvard Business Rev.*, vol. 90, no. 10, pp. 61–68, 2012.
- [41] Y. Demchenko, C. De Laat, and P. Membrey, “Defining architecture components of the Big Data Ecosystem,” *2014 Int. Conf. Collab. Technol. Syst. CTS 2014*, no. March 2015, pp. 104–112, 2014.
- [42] P. C. Science, “Spark,” *INNS Conference on Big Data*, pp. 121–131, 2015.
- [43] “Encuentro mundial de NATs.” .
- [44] “¿Qué es CAOBA? - Alianza Caoba MinTic, Conciencias, universidades.” [Online]. Available: <http://alianzacaoba.co/que-es-caoba/>. [Accessed: 21-Oct-2017].
- [45] Y. Demchenko, C. De Laat, and P. Membrey, “Defining Architecture Components of the Big Data Ecosystem-Reviewed .pdf.” .
- [46] Chukwa, “Chukwa - Welcome to Apache Chukwa,” 2016. [Online]. Available: <http://chukwa.apache.org/>. [Accessed: 21-Oct-2017].
- [47] “¿Qué es Kibana? – Un poco de Java.” [Online]. Available: <https://unpocodejava.com/2012/10/25/que-es-apache-flume/>. [Accessed: 21-Oct-2017].
- [48] R. Serrat Morros, “Big Data : análisis de herramientas y soluciones,” pp. 10–18, 2013.
- [49] “IBM Knowledge Center - Rangos de direcciones privadas.” [Online]. Available: https://www.ibm.com/support/knowledgecenter/es/SSPT3X_4.1.0/com.ibm.swg.im.infosphere.biginsights.analyze.doc/doc/hbaseConcepts.html. [Accessed: 21-Oct-2017].
- [50] L. Joyanes, *Big data - análisis de grandes volúmenes de datos en organizaciones*. 2013.
- [51] MongoDB, “Reinventando la gestión de datos | MongoDB,” 2016. [Online]. Available: <https://www.mongodb.com/es>. [Accessed: 21-Oct-2017].
- [52] Neo4j, “Neo4j, the world’s leading graph database - Neo4j Graph Database.” [Online]. Available: <https://neo4j.com/>. [Accessed: 21-Oct-2017].
- [53] N. Jiménez Barquín, “Big Data: Hadoop,” 2014.
- [54] M. K. Islam and A. Srinivasan, *Apache Oozie*. 2015.
- [55] M. Marquis, “RedisConf 2017 is near,” 2017. [Online]. Available: <https://redis.io/>.
- [56] J. L. Reyes-Ortiz, L. Oneto, and D. Anguita, “Big data analytics in the cloud: Spark on Hadoop vs MPI/OpenMP on Beowulf,” in *Procedia Computer Science*, 2015, vol. 53, no. 1, pp. 121–130.
- [57] M. (University of C. Al-Fares, A. (University of C. Loukissas, and A.



- (University of C. Vahdat, "A scalable, commodity data center network architecture," *Sigcomm*, pp. 63–74, 2008.
- [58] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Proc. 6th Symp. Oper. Syst. Des. Implement.*, pp. 137–149, 2004.
- [59] B. Y. J. Dean and S. Ghemawat, "MapReduce: a flexible data processing tool," *Commun. ACM*, vol. 53, no. 1, pp. 72–77, 2010.
- [60] R. Wesley, M. Eldridge, and P. T. Terlecki, "An analytic data engine for visualization in tableau," *Proc. 2011 Int. Conf. Manag. data - SIGMOD '11*, p. 1185, 2011.
- [61] J. Pérez Díaz and C. P. Solà, "Plataforma para analizar la red Bitcoin," 2017.
- [62] "Graphite." [Online]. Available: <http://graphiteapp.org/>. [Accessed: 23-Oct-2017].
- [63] G. Hernández Coca, "Tipos de Modelos en Investigación de Operaciones," 2011.
- [64] C. Espino, "Trabajo de Fin de Grado ' Análisis predictivo : técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso ,'" p. 65, 2017.
- [65] "Modelos predictivos: reforzando el valor de una buena decisión." [Online]. Available: <https://blog.es.logicalis.com/analytics/modelos-predictivos-reforzando-el-valor-de-una-buena-decision>. [Accessed: 21-Oct-2017].
- [66] "Aprovecha los modelos predictivos de Big Data: tres de cada cuatro predicciones son correctas | Centro de Innovación BBVA." [Online]. Available: <http://www.centrodeinnovacionbbva.com/noticias/aprovecha-los-modelos-predictivos-de-big-data-tres-de-cada-cuatro-predicciones-son>. [Accessed: 21-Oct-2017].
- [67] "full-text."
- [68] J. Ignacio and M. Alberdi, "La piedra angular del Análisis Predictivo."
- [69] "Machine Learning, ¿qué aporta al Big Data? - CICE." [Online]. Available: https://www.cice.es/noticia/machine-learning-big-data/?gclid=CjwKCAjwo4jOBRBmEiwABWNaMXjJgzeyB83aPbfM5GhjLXxMg8dZ4PHAX_YwFyUOKjvGOIz9U97NxxoCoHAQAvD_BwE. [Accessed: 21-Oct-2017].
- [70] "Asir 14-15," 2015.
- [71] L. M. Gracia, "Arquitectura Lambda: Principios de Arquitectura para Sistemas Big Data en Tiempo Real," p. 3, 2013.
- [72] V. Astakhov and M. Chayel, "Lambda Architecture for Batch and Real- Time Processing on AWS with Spark Streaming and Spark SQL," no. May, pp. 1–12, 2015.



- [73] J. Careaga, “Arquitectura Lambda vs Arquitectura Kappa ¿Cuál es el mejor enfoque para implementar un ambiente de trabajo para procesar big data?”
- [74] J. Forgeat, “Data processing architectures – Lambda and Kappa,” *Ericsson Research Blog*, 2015. [Online]. Available: <https://www.ericsson.com/research-blog/data-processing-architectures-lambda-and-kappa/>. [Accessed: 21-Oct-2017].
- [75] J. Scott, “Zeta Architecture: Hexagon is the new circle,” 205AD. [Online]. Available: <https://www.oreilly.com/ideas/zeta-architecture-hexagon-is-the-new-circle>.
- [76] “NEXT GENERATION ENTERPRISE MODELLING,” p. 2016, 2016.
- [77] J. R. C. Rodríguez, “Base de datos distribuidos,” pp. 1–14, 2014.
- [78] Departamento de ciencias de la computación e I.A, “Introducción a las bases de datos,” *Univ. Granada*, 2013.
- [79] V. R. M. O. Carlos Enrique Rodas Gálvez, Álvaro Daniel Castillo Carrera, Miguel Enrique Guerra Connor, “CLUSTER Curso: Sistemas Operativos II Plataforma: Linux - OpenSuse CLUSTER,” *Univ. San Carlos Guatemala Fac. Ing. Esc. Ciencias y Sist.*
- [80] L. Joyanes Aguilar, “COMPUTACIÓN EN LA NUBE. Notas para una estrategia española en cloud computing,” *Rev. del Inst. Español Estud. Estratégicos*, vol. 1, no. 1, pp. 89–112, Apr. 2012.
- [81] “Qué es un ERP | Encuentra tu solución ERP en España.” [Online]. Available: <https://www.elegirerp.com/erp/que-es-un-erp>. [Accessed: 21-Oct-2017].
- [82] “Zettabyte.” [Online]. Available: <http://www.tecnologiahechapalabra.com/datos/eventos/articulo.asp?i=5896>. [Accessed: 21-Oct-2017].
- [83] L. Euler, G. Kirchhoff, F. Guthrie, K. Appel, and W. Haken, “Teoría de grafos w w w i b r o s Z . c.”
- [84] “Integración de datos: Concepto e importancia en la empresa actual.” [Online]. Available: <https://www.powerdata.es/integracion-de-datos>. [Accessed: 23-Oct-2017].
- [85] Ernst&Young, “Inteligencia de Negocio (BI),” p. 2, 2013.
- [86] M. Internet and Q. O. Cosa, “EL ‘INTERNET DE LAS COSAS.’”
- [87] “IPV4-IPV6 | CERTIFICACIÓN DE SISTEMAS OPERATIVOS.” [Online]. Available: <https://gcvvc09.wordpress.com/ipv4-ipv6/>. [Accessed: 23-Oct-2017].
- [88] “Definiciones de metadatos.”
- [89] “Teoría de Lenguajes de Programación: Paradigmas.” [Online]. Available: <http://tlp-lcc-umt.blogspot.com.co/2017/04/paradigmas.html>. [Accessed: 23-

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

Oct-2017].

[90] “PCPI Cheste - Procesamiento en paralelo (SLI, Crossfire).” [Online].

Available:

[https://informaticapcpicheste.wikispaces.com/Procesamiento+en+paralelo+\(SLI,+Crossfire\)](https://informaticapcpicheste.wikispaces.com/Procesamiento+en+paralelo+(SLI,+Crossfire)). [Accessed: 23-Oct-2017].

[91] “¿Qué es un zettabyte?” [Online]. Available:

http://www.parentesis.com/noticias/ciencias/Sabes_que_es_un_zettabyte.

[Accessed: 23-Oct-2017].

LISTA DE ANEXOS:

- Documento Implementación de Herramientas de Autogestión para el levantamiento de servicios en proyectos de Big Data
- Manuales de Instalación y configuración de las herramientas usadas para la implementación de los servicios.