

**FACULTAD DE SISTEMAS
PROGRAMA DE INGENIERIA DE SISTEMAS
BOGOTÁ D.C.**

LICENCIA CREATIVE COMMONS: ATRIBUCION NO COMERCIAL 2.5
COLOMBIA (CC BY-NC 2.5)

AÑO DE ELABORACIÓN: 2015

TÍTULO: ALGORITHMS AND TOOLS OF BIG DATA: A BIBLIOGRAPHIC
REVIEW.

AUTOR (ES):

CARLOS ANDRES CORTES NUÑEZ.

DIRECTOR(ES)/ASESOR(ES):

ING. MARIO MARTINEZ ROJAS.

MODALIDAD:

Trabajo de investigación

PÁGINAS: **TABLAS:** **CUADROS:** **FIGURAS:** **ANEXOS:**

CONTENIDO:

INTRODUCTION
1. PROBLEM STATEMENT
2. OBJECTIVES
2.1 GENERAL OBJECTIVE
2.2 SPECIFICS OBJECTIVES
3. BIG DATA
3.1 THEORETICAL
3.2 UNSTRUCTURED DATA
3.3 METAHEURISTIC

RESUMEN ANALÍTICO EN EDUCACIÓN - RAE -



UNIVERSIDAD CATÓLICA
de Colombia

4. META ANALYSIS

4.1 REASON FOR FULL-TEXT ARTICLES EXCLUDED

4.2 INFORMATION SOURCES

4.3 STUDY SELECTION

4.4 SYNTHESIS OF RESULTS

4.5 RESULTS

5. OPEN FIELDS OF RESEARCH

5.1 BIG DATA SECURITY

5.1.1 Common Techniques for Securing Big Data.
5.1.2 Threats for Big Data.
5.2 BIG DATA INFRASTRUCTURE
5.3. BIG DATA FOR BUSINESS
6. CONCLUSIONS
REFERENCES

DESCRIPCIÓN: Big data is among us, it's present on all parts of our life and like time it's never stops growing. Having in mind that all data will tend to become big data is necessary to change the customs ways of dealing for two simple reasons, the increase in the cost and the decrease of effectiveness. A quick glance in the evolution of data and information shows the leaders and the technologies that have been developing a way to work with big data, all this having as an essential factor the value given to the user.

METODOLOGÍA: BIBLIOGRAPHIC REVIEW WITH STATE OF THE ART OF BIG DATA AND TOOLS.

PALABRAS CLAVE: BIG DATA, ALGORITHMS, TOOLS.

CONCLUSIONES: A vast view of the selected document included in the repository for the Meta-analysis, were all main topics where extract for easier identification when searching for different topics related to big data. Thanks to the use of the PRISMA framework and the guiles lines defined to construct this document and the definition of a state-of-the-art document, the meta-analysis have come out with a up to date research on big data algorithms, identifying and selecting from different sources documents, lectures and articles the investigation have produce a well-organized repository for future investigation on the subject.

FUENTES:

BaseLineMag.Utilizing Cutting-Edge Unstructured Data Analytics [on line]. [August 1 2015]. Available on Internet: < URL: <http://www.baselinemag.com/analytics-big-data/utilizing-cutting-edge-unstructured-data-analytics.html>>

BDISYS. Big Data [on line]. [August 5 2015]. Available on Internet :< URL: <http://www.bdisys.com/27/1/17/BIG%20DATA/HADOOP>>

RESUMEN ANALÍTICO EN EDUCACIÓN - RAE -



Chen, Philip. Zhang, Chun-Yang. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Macau, China: University of Macau. 2013. p. 1.

COMPUTERWEEKLY. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <http://www.computerweekly.com/tip/Securing-NoSQL-applications-Best-practises-for-big-data-security>>

COMPUTERWEEKLY. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <http://www.computerweekly.com/report/Tiered-storage-strategies-and-best-practices>>

DATAECONOMY. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <http://dataconomy.com/understanding-big-data-infrastructure/>>

HADOOP. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <https://hadoop.apache.org/>>

INFOQ. Big Data [on line]. [August 23 2015]. Available on Internet :< URL: <http://www.infoq.com/articles/HadoopSecurityModel> />

INFORMATION-AGE. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <http://www.information-age.com/technology/information-management/123457900/how-build-big-data-infrastructure#sthash.yi36Ozan.dpuf>>

ISACA. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: http://www.isaca.org/groups/professional-english/big-data/groupdocuments/big_data_top_ten_v1.pdf >

JUNIPER. Big Data [on line]. [August 23 2015]. Available on Internet :< URL: <http://www.juniper.net/us/en/local/pdf/whitepapers/2000488-en.pdf>> 66

RESUMEN ANALÍTICO EN EDUCACIÓN - RAE -



McAfee. Big Data [on line]. [August 18 2015]. Available on Internet: < URL: <http://www.mcafee.com/uk/resources/reports/rp-needle-in-a-datastack.pdf> >

McKinsey & Company. Big Data [on line]. [August 1 2015]. Available on Internet: < URL: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation >

MICRON. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: https://www.micron.com/~media/documents/products/technical-marketing-brief/brief_ssds_big_data.pdf >

MICROSOFT. Big Data [on line]. [August 18 2015]. Available on Internet :< URL: <https://msdn.microsoft.com/en-us/library/ff647076.aspx>>

Munawar, Hasan. Data Genetic Algorithm and its application to Big Data Analysis. New Delhi, India: University of Dehradun.2014. p. 1.

OXFORD DICTIONARIES. Big Data [on line]. [August 1 2015]. Available on Internet: < URL: http://www.oxforddictionaries.com/us/definition/american_english/big-data >

Scholarpedia. Metaheuristic Optimization [on line]. [August 1 2015]. Available on Internet: < URL: http://www.scholarpedia.org/article/Metaheuristic_Optimization >

THALES GROUP. Big Data [on line]. [August 1 2015]. Available on Internet: < URL: <https://www.thalesgroup.com/en/worldwide/big-data/four-vs-big-data> >

ThalesGroup.Big Data [on line]. [August 18 2015]. Available on Internet: < URL: <https://www.thalesgroup.com/en/worldwide/big-data/four-vs-big-data> >

LISTA DE ANEXOS:

- Annex 1. 2007-Combinatorial Algorithm for Compressed Sensing
- Annex 2. 2007-Near-Optimal Algorithm for estimating the Entropy of a Stream
- Annex 3. 2009-Data Stream Algorithms200-Barbados Workshop on Computational Complexity
- Annex 4. 2010-Data-Parallel Algorithms and Techniques
- Annex 5. 2011-Introduction to IO Efficient Algorithms-MIT
- Annex 6. 2011-Introduction to Parallel Algorithm Analysis
- Annex 7. 2011-Introduction to Streaming Algorithms
- Annex 8. 2011-IO-Efcient Merge Sort

- Annex 9. 2012-Clustering-Stanford University
- Annex 10. 2012-Graph Streams algorithm
- Annex 11. 2012-IO-Algorithms
- Annex 12. 2012-IO-Algorithms2
- Annex 13. 2012-MapReduce Algorithms.
- Annex 14. 2013-Algorithm_and_approaches_to_handle_large_Data-_A_Survey
- Annex 15. 2013-Algorithmic Techniques for Big Data
- Annex 16. 2014-Genetic-Algorithm-and-its-application-to-Big-Data-Analysis
- Annex 17. 2014-Parallel Algorithms for Geometric Graph Problems
- Annex 18. Date-Clustering-Stanford University
- Annex 19. External-Memory Graph Algorithms
- Annex 20. Fast Parallel GPU-Sorting Using a Hybrid Algorithm
- Annex 21. Fundamental Parallel Algorithms for Private-Cache Chip Multiprocessors
- Annex 22. IO complexity of graph algorithms 10
- Annex 23. K-Center-Algorithm
- Annex 24. K-means Advantages of Careful Seeding
- Annex 25. Lower Bounds in streaming (algorithm)
- Annex 26. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High
- Annex 27. Scalable KMeans++
- Annex 28. Streaming Graph Algorithms-MIT