



# No SQL Database Management Systems for Big Data

Luciano Caroprese, Ester Zumpano, Eugenio Vocaturo

**Abstract:** It is well known that, at the present, a huge amount of information, often referred as Big Data, is processed by each domain of modern society. Big data are well defined by the seven dimensions: Volume, Velocity, Variety, Variability, Veracity, Visualization and Value. The traditional database management systems cannot handle the requirements of high availability, scalability and reliability emerged with Big Data.

The good news is that we are now in the age of NoSQL databases. NoSQL do not have a fixed structure, they have a flexible structure and are suited for storing unstructured data produced in a large scale in various field.

This work outlines the four main types of NoSQL databases and presents some of their representative solutions.

**Keywords :** About four key words or phrases in alphabetical order, separated by commas.

## I. INTRODUCTION

The IBM in a recent study, reports that in the last two years about the 90% of the data has been created. Each day the worlds outputs about 2.5 quintillion bytes a day.

Domo, a computer software company specialized in business intelligence tools and data visualization, publishes the annual report (<https://www.domo.com/learn/data-never-sleeps-7>) Data Never Sleeps - that illustrates the online world's activities happening within a minute. Figure I reports the data generated every minute in 2019.

The IDC white paper by Seagate [38] forecasts that in 2025 enterprises will create more than 60% of data in the world. Therefore, in the next future the focus will be on the creation, utilization and management of data coming from governs, consumers and businesses.

There is no doubt that the era of Big Data will continue to fast go on. Big data implies a big interest both from industries and academia. Big data research and applications are funded by many government agencies (e.g. the Obama Administration's Big Data Working Group report [39]) and topics related to Big Data are frequently reported by relevant public media such as the Economist [26, 39] and the New York Times [35]. Much effort is devoted by the research community on relevant

challenges in Big Data. Premier conferences and prestigious journals, such as Nature and Science.

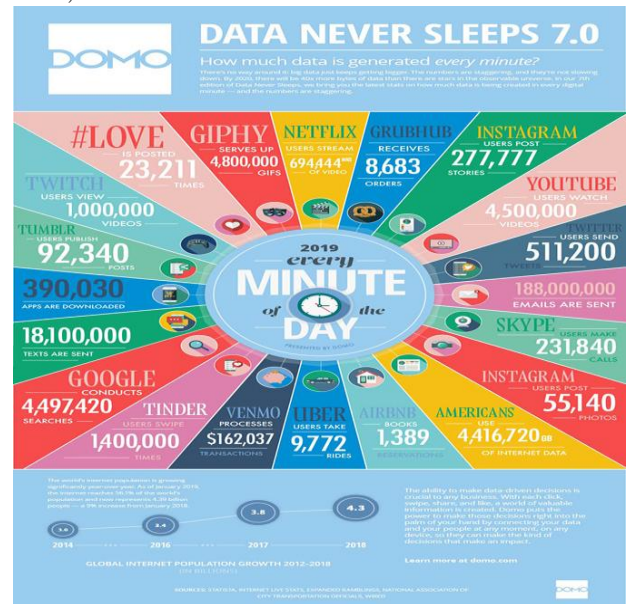


Fig. I - Data generated every minute in 2019

## II. BIG DATA

We are surrounded by Big Data!

Big data are extensively part of all human domains and all sectors of our society [9], even many problems are still to be solved in their management [47] and many challenges are open [49-51].

In the domain of healthcare Big Data are extensively used in medical information tools for data prediction and diagnostic procedures [1, 17-18, 27-28], as well as in computer vision and in machine learning techniques for melanoma lesions characterization and features detection [32-34, 40-42].

We are in the era of Big Data.

But the question is: what is the exact meaning of Big Data? The reference to "Big Data" appeared in a paper written by researchers of NASA [3] in 1997.

Many different definitions have been provided by recent literature [4, 10, 48], each acceptable and reporting a different vision, a different perspective of what is Big Data. In order to catch the real meaning of Big Data, as something different from big size data, many other dimensions have to be considered. In [11] Laney first, explicitly, uses the "3 V's", Volume, Velocity and Variety, to describe Big Data, even if he does not explicitly mention them. The "3 V's", as a mean to characterize Big Data, find a general consensus in the literature [10, 12, 19, 20, 21, 22]. Starting from this, many researchers and institutes such as IEEE focused on additional important features of Big Data such Value, Veracity, Visualization, Variability [4, 22, 29, 30]. Finally, the seven



Revised Manuscript Received on May 27, 2020.

\* Correspondence Author

Ester Zumpano\*, DIMES, University of Calabria, Rende (CS), Italy. Email: e.zumpano@dimes.unical.it l.caroprese@dimes.unical.it

Luciano Caroprese, DIMES, University of Calabria, Rende (CS), Italy. Email: l.caroprese@dimes.unical.it

Eugenio Vocaturo, DIMES, University of Calabria, Rende (CS), Italy. Email: e.vocaturo@dimes.unical.it; CNR-NANOTEC National Research Council, Rende, Italy. E-mail eugenio.vocaturo@cnr.it.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

V's allows to provide a pretty nice definition of Big Data.

- **Volume.** It is one of the feature most associated with Big Data and it represents the total amount of data. Facebook, for example, is a community in which more than 890 million of individuals log each day. People share documents, photos, comments.
- **Velocity.** It is the velocity in which data is generated and managed.
- **Variety.** It describes the variety of data often provided by different sources. This variety implies data cannot fit in structured relational database systems. The feature refers to structured (data from relational database), semistructured (web logs, emails, etc.) or unstructured (video, audio, clicks, etc.) data.
- **Variability.** It differs from variety and has several potential meanings. It can state for: is data available at times? or also, in the case of extreme values are these outliers or just noisy data?
- **Veracity.** It refers to the accuracy of the data. The truthfulness of data is a key point. Therefore, besides, the quality of data, important dimensions are the source, the accuracy, the availability of data. Uncertainty can be caused by inconsistency, ambiguity and incompleteness of the data.
- **Visualization.** It refers to the use of tools allowing to visualize and analyze results on data. The availability of a huge amount of data is useless if you are not able to catch the information it contains. Different data visualization tools, with different insights in the collection of data are now available. Some of these popular tools are: Google Chart, Tableau, D3, Fusion Chart, Highcharts, Microsoft Power BI.
- **Value.** This dimension captures the final aim of Big data. Each organization wants to get *value* from data.

### III. FROM RBMS TO NOSQL DBMS

Traditional Relational Database Management Systems (RDBMS) have been evolving for more than 40 years. Various RDBMS have been successfully implemented and used by different organizations to support the management of different datasets at different scales.

In a relational database management system data are organized in tables, columns and rows, are entered once, can be stored across multiple tables and can be referenced from elsewhere in the database by fixing a relationship between tables. The relational DB model was provided by Codd, of IBM's Research Laboratory in 1969 and published in [2]. RDBMS provide a logical view of the data and the SQL language for querying and fetching data. They have been for long time the standard for storing data, as they guarantee stability, good performance and consistency.

Anyhow, they are not up to the task of handling the requirements of high availability, scalability and reliability brought by Big Data; they reached a bottleneck and new systems have been recently developed. They are known as NoSQL databases.

In contrast to RDBMS, NoSQL do not have a fixed structure. They have a flexible structure and are suited for storing unstructured data produced in a large scale in various

field. NoSQL databases are a good choice for big data. As for a comparison between RDBMS and NoSQL databases, in RDBMS basic principles govern how changes are applied to a database.

Specifically, database transactions should be ACID [46]:

- **Atomicity** a transaction to a database should work or fail as a whole;
- **Consistency** a transaction has to ensure that it will not affect the consistent state of the database;
- **Isolation** a transaction does not interfere with any other;
- **Durability** a completed transaction is persistent.

These qualities are incompatible with availability and performance with very large systems in a distributed environment. More specifically, the CAP theorem [37] states that a distributed system cannot guarantee simultaneously more than two of the following properties: (i) Consistency, (ii) Availability, (iii) Partition tolerance.

In a distributed environment BASE is implemented.

- **Basically available** states that availability is guaranteed by the system;
- **Soft state** reports that the state of the system may change due to the eventual consistency model;
- **Eventual consistency** (or weakly consistent) states that the system will become consistent over time. Therefore, temporary database inconsistencies are accepted.

Roughly speaking, in a distributed system the choice is between consistency and availability.

### IV. PROPOSED METHODOLOGY

In the rest of this section we outline the most popular types of NoSQL databases and present some of their representative solutions.

#### A. Key-value Oriented Databases

Key-value oriented databases are the simplest implementation of NoSQL database.

The key/value approach has been also used in P2P [5-8, 13-16, 23-24] systems such as Tapestry [43], Chord [44] and Kademlia [45].

A key-value database stores (key-value) pairs, which can be a binary object of different type. Every key is unique and queries can be posed according to the keys. Therefore, the key-value structure, known the key allows to retrieve a value.

The concept of key-value is different with respect to the one of relational database, as this last uses fields (maintained in tables) having well defined data types. In a key-value oriented database there is neither relation nor structure.

A generic key-value stores data as a single collection that can have for every record different fields. On the contrary, in a key-value oriented database a data is a single collection and therefore each record can have different fields. This simple solution allows to maintain the complete control over the value and guarantee high expandability and fast query response time. Scalability and availability are supported by partitioning and replicating data over a cluster.

Popular key-value databases are DynamoDB and Voldemort.

#### ▪ DynamoDB

<https://aws.amazon.com/it/dynamodb/> DynamoDB also known as Dynamo Database or DDB, is a NoSQL key-value data storage system offered by Amazon.com within the Amazon Web Services. DynamoDB can manage huge amount of data and is effective on high level of request traffic. In DynamoDB data are stored in tables, and are fetched using read and write operations. Each DynamoDB item is uniquely identified by a primary key that is used to execute queries. DynamoDB automatically spreads the data over a sufficient number of servers and ensures consistency and fast performances. Solid state disks are used for data storing and replication is automatically performed providing high availability and data durability. DynamoDB allows to perform asynchronous update on all copies by read consistency.

#### ▪ Voldemort -

<http://www.project-voldemort.com/voldemort/>

Voldemort is a key-value storage system, used and developed by LinkedIn. Three simple operations are included in its interface: the read, write and delete operation. Multiple servers are used to maintain portion of data. The partition and replication of data is automatically performed. The data copy mechanism in Voldemort is the same as in DynamoDB: data are stored in RAM and inserted into a storage engine. The database does not guarantee data consistency, but provides facilities for asynchronous updating concurrent control of multiple editions. There is no central coordination point for failure as each node is independent. In order to maximize data integrity in the case of a failure scenario it supports data versioning.

#### ▪ REDIS - <https://redis.io/topics/introduction>

Redis, developed in 2009 and written in C, is a NoSQL database offering good performance and speed. It implements the key-value mechanism to store data. In particular, it is considered a data structure server as it allows keys to contain hashes as well as strings, lists and (sorted) sets.

### B. Column Oriented Databases

A *Column-oriented Database* stores data in column(s), contrary to relational database, in which data is stored in row(s) of a table. In addition, differently from RDBMS, each row of the table can specify names and formats of its column. A column-oriented database groups columns. It does not have any prestructured table to work with data. In a column-oriented database a single disk operation is sufficient to retrieve related data, as the only retrieved columns are those associated with the query.

This contrast with what happens in RDBMS in which data of a table are stored in different location of the the disk, therefore in order to retrieve data many different read operation are needed. This mechanism makes this type of database a good choice for managing large unstructured data. Among the most popular column - oriented database we recall Google's Bigtable, Cassandra and HBase.

**Google Bigtable.** [36] is a column oriented distributed database developed by Google Inc. with the final aim of managing huge amounts of data. Bigtable was made available

as public version in 2015 and was designed to be used with petabytes of data and to support applications requiring massive scalability. At the present, Google Bigtable is the database used by many Google applications, such for example YouTube, Gmail, Google Maps, Google Book Search and Google Earth. Google has maintained the propriety of the software, anyhow its organization together with the fact that it is an open source product allow others to easily create BigTable derivatives (examples are the Apache HBase database and Cassandra).

**HBase** - <https://hbase.apache.org/> HBase is a distributed database modeled after Google's Bigtable. It is open source, written in Java and developed under the Apache Hadoop project. HBase follows the BigTable model and has been shown to be best suited for real-time querying of Big Data.

**Cassandra** - <https://cassandra.apache.org/> Cassandra is an open source database developed in 2008 at Facebook. It is considered the right choice if scalability and high availability are needed, without compromising performance. Cassandra integrates the distributed technology of Dynamo and the data model used by BigTable and by doing this it offers the benefit of a column oriented database and the good performance of log-structured updates, ensuring an effective support for denormalization, a powerful built-in caching and materialized views. Cassandra is considered a good choice if, without prejudicing the performance, there is the need of ensuring both scalability and availability. Cassandra is used, among others, at CERN, eBay, Instagram, Comcast and Netflix.

### C. Document Oriented Databases

*Document Oriented Databases* were developed as a consequence of the need of having a database system that, on the contrary of traditional relational database, did not rely on schema.

In such a database a generic record and its associated data is a document that can have a structure identical or completely different with respect to another row.

Each document typically stores self-describing documents, JSON (JavaScript Object Notation), XML (Extensible Markup Language) and BSON (Binary JSON). The structure results to be similar to the key-value storage, but in document oriented database the value consists of a single document. Therefore, more complex data forms can be supported (a document within a document and so on). More specifically, even if the key is not known, it is possible to obtain a fast retrieval if popular fields are indexed.

MongoDB and CouchDB are among the most important document oriented databases.

#### ▪ MongoDB - <https://www.mongodb.com>

MongoDB is a open source database, document-oriented, cross-platform and it supports, natively, JSON format. The software development began in 2007 and in 2017 it became a publicly-traded company, listed on NASDAQ. MongoDB can be used easily and no database administrator is needed for bootstrap. It is robust and allows versioning to guarantee consistency in the case of complex transactions.



It offers dynamic queries and powerful aggregates. Mongo DB features includes the possibility of using dynamic queries and offers a powerful tool of aggregates. These characteristics make of Mongo DB a tool suitable if high volumes of data have to be managed and the write loads are high.

- **CouchDB** - <http://couchdb.apache.org/>

Apache CouchDB is an open source database, implemented in Erlang, easy to be used and with scalable architecture. It has been developed in 2005 and in 2008 became an Apache Software Foundation project. CouchDB, that stores data using JSON and performs queries using JavaScript, is a good choice for the Web. It well addresses redundancy and conflict resolutions and stores every change on the disk in terms of a document revision.

### D. Graph Oriented Databases

A *Graph-oriented Database* completely differs from the previous three NoSQL database type. A graph structure is used not only to store, but also to map and query. The entities in the graph are the nodes. Each node exhibits a number of properties, i.e. attributes each one consisting of a pair key-value. Labels are used to tag nodes so that describing roles and also to associate metadata, as well as constraints and index to a node. Relationships define directed, named and semantically relevant connections between nodes.

A popular graph oriented database is Neo4j.

- **Neo4j** - <https://neo4j.com/product/>

Neo4j, developed by Neo4j Inc, is a graph database management system that fits ACID properties. It is implemented in Java and stores everything in the form

of a node or an edge or an attribute. It is available in three different editions: the Community, the Enterprise and finally the Government edition.

## V. RESULT ANALYSIS

Each NOSQL database offers different solutions to specific context and has different application scenarios. As for an analysis Table I reports the main storage types of NoSQL database, whereas Table II reports a comparison among some NoSQL databases.

**Table- I: Comparison of NoSQL DBMS types**

Database type	application field	Database System
key-value storage	Log System	Dynamo, Redis, Voldemort
column-based storage	Distributed File System	BigTable, Cassandra, HBase
document-storage	Web Application	MongoDB, CouchDB
graph storage	Social Networking	Neo4j, GraphDB

## VI. CONCLUSION

There is a very great number of available NoSQL databases and different way to classify them. The traditional approach of "one fits all" cannot be anymore applies. Far from being exhaustive, this paper reports the basic feature of the main types of NoSQL databases and presents the most popular types of NoSQL databases and some of their representative solutions.

**Table- II: Feature of some NoSQL systems**

Database System	Schema	Supported Data	Architecture	Language	Licence
DynamoDB	Schema free	Structured	Master-Slave	Async	Proprietary
BigTable	Uses Schema	Structured	Master-Slave	Sync and Async	Proprietary
HBase	Uses Schema	Structured	Master-Slave	Async	Open Source
Cassandra	Optional	(semi)structured Unstructured	P2P	Async	Open Source
MongoDB	Dynamic schema	(semi)structured Unstructured	Master-Slave	Async	Open Source

## REFERENCES

1. L. Caroprese, P. Veltri, E. Vocatur, E. Zumpano, Deep Learning Techniques for Electronic Health Record Analysis. IISA 2018: 1-4. DOI: 10.1109/IISA.2018.8633647
2. E. F. Codd. A relational model of data for large shared data banks. Communications of the ACM, 13, 6.
3. M. Cox, D. Ellsworth, Application-controlled demand paging for out-of-core visualization. IEEE Visualization 1997: 235-244
4. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011.
5. L. Caroprese, E. Zumpano, Aggregates and priorities in P2P data management systems. IDEAS 2011: 1-7. DOI: 10.1145/2076623.2076625
6. L. Caroprese, E. Zumpano, Handling Preferences in P2P Systems. FoIKS 2012: 91-106. DOI:10.1007/978-3-642-28472-4\_6
7. L. Caroprese, E. Zumpano, Computing a Deterministic Semantics for P2P Deductive Databases. IDEAS 2017: 184-191. <http://dx.doi.org/10.1145/3105831.3105837>
8. L. Caroprese, E. Zumpano, Generalized Maximal Consistent Answers in P2P Deductive Databases. DEXA (2) 2016: 368-376. <http://dx.doi.org/978-331944405-5>
9. G. Halevi, H. Moed, The evolution of big data as a research and scientific topic: Overview of the literature, Res. Trends (2012) 3-6.
10. K. Davis, D. Patterson, Ethics of Big Data: Balancing Risk and Innovation, O'Reilly Media, 2012.
11. Laney, D. (2001). 3-D Data Management:Controlling Data Volume, Velocity and Variety. META Group Research Note, (February), 1-4.
12. A. Reeve, Managing Data in Motion: Data Integration Best Practice Techniques and Technologies, Morgan Kaufmann, 2013.
13. L. Caroprese, E. Zumpano, A Logic Framework for P2P Deductive Databases. Theory Pract. Log. Program. 20(1): 1-43 (2020). DOI:10.1017/S1471068419000073
14. L. Caroprese, E. Zumpano, Integration of Unsound Data in P2PSystems, ADBIS 2018: 278-290 Systems. ADBIS 2018: 278-290. [http://dx.doi.org/10.1007/978-3-319-98398-1\\_19](http://dx.doi.org/10.1007/978-3-319-98398-1_19)
15. L. Caroprese, E. Zumpano, A Declarative Semantics for P2PSystems. CD-MAKE 2017: 315-329. DOI:10.1007/978-3-319-66808-6\_21
16. L. Caroprese, E. Zumpano, Restoring Consistency in P2P Deductive Databases. SUM 2012: 168-17. DOI: 10.1007/978-3-642-33362-0\_13
17. E. Zumpano, P. Iaquina, L. Caroprese, F. Dattola, G. Tradigo, P. Veltri, E. Vocatur, SIMPATICO 3D Mobile for Diagnostic Procedures. iiWAS 2019: 468-472. DOI: 10.1145/3366030.3366087

18. E. Zumpano, P. Iaquina, L. Caroprese, G.L. Cascini, F. Dattola, P. Franco, M. Iusi, P. Veltri, E. Vocaturo. SIMPATICO 3D: A Medical Information System for Diagnostic Procedures. *BIBM* 2018: 2125-2128. DOI: 10.1109/BIBM.2018.8621090
19. X. Dong, D. Srivastava, Big data integration, in: *Data Engineering (ICDE)*, 2013 IEEE 29th International Conference on, 2013, pp. 1245–1248.
20. P. Zikopoulos, C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Education, 2011.
21. I. O'Reilly Media, *Big Data Now: 2014 Edition*, O'Reilly Media, 2014.
22. P. Hitzler, K. Janowicz, Linked data, big data, and the 4th paradigm, *Semant. web* (2013) 233–235.
23. L. Caroprese, E. Zumpano, P2P deductive databases: a system prototype. *iiWAS* 2017: 258-265. DOI: 10.1145/3151759.3151768
24. L. Caroprese, E. Zumpano, P2P Deductive Databases: Well Founded Semantics and Distributed Computation. *ADBIS (Short Papers and Workshops) 2017*: 91-99. DOI: 10.1007/978-3-319-67162-8\_11
25. <http://bigdata-madesimple.com/a-deep-dive-into-nosql-a-complete-list-of-nosql-databases/>
26. Cukier K (2010) Data, data everywhere: a special report on managing information. *Economist Newspaper*. <https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>
27. P. Iaquina, M. Iusi, L. Caroprese, S. Turano, S. Palazzo, F. Dattola, I. Pellegrino, G. Tradigo, G. Cascini, P. Veltri, and E. Zumpano. eimes 3d mobile: A mobile application for diagnostic procedures. *BIBM*, 1634–1641, 2018 DOI: 10.1109/BIBM.2017.8217904
28. P. Iaquina, M. Iusi, L. Caroprese, S. Turano, S. Palazzo, F. Dattola, I. Pellegrino, P. Veltri, and E. Zumpano. eimes 3d: An innovative medical images analysis tool to support diagnostic and surgical intervention. In *FNC/MobiSPC 2017*, 459–464. DOI: 10.1016/j.procs.2017.06.122
29. Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P. (2012). *Analytics: The real-world use of big data*. New York, NY: IBM Institute for Business Value, Said Business School.
30. Dijkstra, J. (2013). *Oracle: Big data for the enterprise*. Oracle White Paper. Redwood Shores, CA: Oracle Corporation.
31. <https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
32. E. Vocaturo, E. Zumpano, and P. Veltri. Image pre-processing in computer vision systems for melanoma detection. *BIBM* 2018, 2117–2124. DOI: 10.1109/BIBM.2018.8621507
33. E. Vocaturo, E. Zumpano, and P. Veltri, "Features for Melanoma Lesions Characterization in Computer Vision Systems", 9th International Conference on Information, Intelligence, Systems and Applications, (IISA) 2018, pp. 1--8, 2018. DOI: 10.1109/IISA.2018.8633651
34. Vocaturo, E., Perna D., and Zumpano E., *Machine Learning Techniques for Automated Melanoma Detection*, IEEE International Conference on Bioinformatics and Biomedicine, *BIBM* 2019, pp. 2310-17, doi: 10.1109/BIBM47256.2019.8983165.
35. Lohr S. The age of big data. *New York Times*, pp 11., 2012.
36. Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Michael Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber: *Bigtable: A Distributed Storage System for Structured Data*. *ACM Trans. Comput. Syst.* 26(2): 4:1-4:26 (2008)
37. Seth Gilbert and Nancy Lynch, "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services", *ACM SIGACT News*, Volume 33 Issue 2 (2002), pp. 51–59.
38. David Reinsel John Gantz John Rydning. *Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big* IDC White Paper. Fact sheet: Big data across the federal government (2012).
39. Fuel of the future. Data is giving rise to a new economy. <https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy>
40. A. Fuduli, P. Veltri, E. Vocaturo, E. Zumpano, "Melanoma detection using color and texture features in computer vision systems", *Advances in Science, Technology and Engineering Systems Journal*, vol. 4, no. 5, pp. 16-22 (2019). DOI: 10.25046/aj040502
41. E. Vocaturo, E. Zumpano, P. Veltri, On the Usefulness of Pre-Processing Step in Melanoma Detection Using Multiple Instance Learning, *FQAS*, Springer, pp. 374-382, 2019. DOI: 10.1007/978-3-030-27629-4\_34
42. E. Vocaturo, E. Zumpano, Dangerousness of dysplastic nevi: a Multiple Instance Learning Solution for Early Diagnosis. *BIBM* 2019: 2318-2323. DOI: 10.1109/BIBM47256.2019.8983056
43. Zhao, Ben Y.; Huang, Ling; Stribling, Jeremy; Rhea, Sean C.; Joseph, Anthony D.; Kubiawicz, John D. (2004). "Tapestry: A Resilient Global-scale Overlay for Service Deployment" (PDF). *IEEE Journal on Selected Areas in Communications*. 22 (1): 41–53. CiteSeerX 10.1.1.71.2718. doi:10.1109/JSAC.2003.818784.
44. Stoica, I.; Morris, R.; Kaashoek, M. F.; Balakrishnan, H. (2001). "Chord: A scalable peer-to-peer lookup service for internet applications" (PDF). *ACM SIGCOMM Computer Communication Review*. 31 (4): 149. doi:10.1145/964723.383071.
45. Petar Maymounkov, David Mazières: *Kademlia: A Peer-to-Peer Information System Based on the XOR Metric*. *IPTPS* 2002: 53-65.
46. ACID (Atomicity, Consistency, Isolation, and Durability.), in *Dictionary of E-Business*, Hoboken, Wiley, 2003.
47. Jacobs, A., *The Pathologies of Big Data*. *ACMQueue*, 6 luglio 2009.
48. Andrea De Mauro, Marco Greco e Michele Grimaldi, A Formal definition of Big Data based on its essential features, in *Library Review*, vol. 65, n. 3, 2016,
49. Nasser Thabet e Tariq Rahim Soomro, *Big Data Challenges*, in *Journal of Computer Engineering & Information Technology*, 2015, DOI:10.4172/2324-9307.1000133.
50. Xin Luna Dong, Theodoros Rekatsinas: *Data Integration and Machine Learning: A Natural Synergy*. *KDD* 2019: 3193-3194
51. Shazia Wasim Sadiq, Tamraparni Dasu, Xin Luna Dong, Juliana Freire, Ihab F. Ilyas, Sebastian Link, René J. Miller, Felix Naumann, Xiaofang Zhou, Divesh Srivastava: *Data Quality: The Role of Empiricism*. *SIGMOD Rec.* 46(4): 35-43 (2017).

## AUTHORS PROFILE



**Luciano Caroprese** received the Laurea degree in computer science engineering from the University of Calabria, in 2004. He received the Ph.D. in computer science from the University of Calabria in 2008. Currently he is Researcher at the Institute for high performance computing and networking (ICAR-CNR). His area of research includes logic programming, deductive database, database integration, P2P systems,

Data Analytics and Machine Learning. From 2005 to today, he cooperated as Tutor and Professor to the educational activities of the Engineering Faculty of the University of Calabria. In 2019 he has been professor at the Information Engineering, Infrastructure and Sustainable Energy Department (DIIES) of the Mediterranean University of Reggio Calabria. He's author of more than 50 papers presented on National and International Conferences or Published on International Journals.



**Ester Zumpano** is an associate professor of computer engineering at the Department of Computer Engineering, Modeling, Electronics and Systems Science, University of Calabria. She received a Phd in Computer Science from the University of Calabria. Her area of research includes health information systems, data integration, logic programming, view updating, distributed systems, artificial Intelligence, database management. She is member of the

Steering Committee of the ADBIS conference. She has been program chair of the 22nd International Database Engineering & Applications Symposium - IDEAS 2018; organizer of the Special Session Machine Learning in Health@ ICMLA 2019; program chair of the IEEE International Workshop on Artificial Intelligence Techniques for BioMedicine and HealthCare (AIBH) at *BIBM* 2018 and *BIBM* 2019; editor of the special issue on Application of Machine Learning Methods in Bio-medical Informatics for *Mathematical Biosciences and Engineering* journal; editor of the special issue *Mathematical Theories in the Era of Big Data* for the *Mathematical Problems in Engineering* journal. She is author of over 100 publications including journal articles, conference papers and book chapters.

## No SQL Database Management Systems for Big Data



**Eugenio Vocaturo** received a degree in management engineering (2002), a Master in Industrial Engineering Management (2006), a master in Finance issued by SDA Bocconi (2016) and is pursuing a PhD degree in Information and Communication Technologies at the University of Calabria, Rende, Italy. He has decades of experience as company director, production and

logistics manager of business groups. He is currently a contract professor of Computer Science and assistant of the Process Mining course at the Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES) of University of Calabria. His current research interests include machine learning, optimization, issues related to classification problems applied to the medical context, emerging issues related to Cultural Heritage. He is a member of SIBIM (Italian Scientific Society of Biomedical Informatics) and of HL7 Italy (formed in 2003 as part of HL7 International), company responsible for the localization of health standards aiming at promoting the modernization of Italian health IT.