

# Variety in NoSQL databases

Anikó Vágner<sup>a</sup>, Mustafa Alzaidi<sup>b</sup>

<sup>a</sup>Department of Information Technology, Faculty of Informatics, University of Debrecen  
[vagner.aniko@gmail.com](mailto:vagner.aniko@gmail.com)

<sup>b</sup>Department of Information Technology, Faculty of Informatics, University of Debrecen  
[mustafa.alzaidi@inf.unideb.hu](mailto:mustafa.alzaidi@inf.unideb.hu)

## Abstract

There were two main reasons why the NoSQL databases were born [6]: to make the development easier and to spread the data across many computers. If we store the data on many computers, it supports the reliability of the system, since if a few computers fail the system can work and answer.

Many research papers state that NoSQL databases support big data. [4] [5] [8] [10] The big data has many features, like the 3 V [2]: volume, variety, velocity. But these 3V-s are increased to 5V-s [3] by veracity and value, and finally, there are 42 Vs. [7]

NoSQL databases can be classified into many categories, where the well-known categories are the key-value, the document, the column-family and the graph databases, however, the XML and object-oriented databases can also be considered NoSQL databases. [6] [9]

The DB-Engines Ranking website [1] lists the most used SQL and NoSQL databases, where the first databases of each NoSQL category are the MongoDB for the document category, Redis for the key-value category, Cassandra for the column-family category and Neo4j for graph category. Of course, each category has many other databases, in the first 40 databases we can find Amazon DynamoDB, HBase, Couchbase, Memcached, Firebase and CouchDB databases.

In this paper, our goal is to examine how the NoSQL databases solve the variety of big data features. We consider only the first database from each category based on the DB-Engines Ranking website [1].

Variety as a feature of big data refers to the multiple data formats and types. On a high level, you can categorise data types into structured, semi-structured and unstructured data types. [3]

Structured data are often stored in tabular form in most cases in relational databases. [3]

Unstructured data can be binary or textual. The binary unstructured data are often media files, like image, audio and video files. The textual unstructured data can be blog postings or tweets. The unstructured data cannot be directly processed using the statements of a database management system, since usually it does not contain tools to play a video. [3]

Semi-structured data usually is hierarchical or graph-based. XML and JSON are common forms of semi-structured data. Semi-structured data can be easier processed even using the statements of the database management systems compared to unstructured data. [3]

In this paper, we are interested how these NoSQL databases (Redis, MongoDB, Cassandra and Neo4j) can store unstructured data, namely video, audio and image. We can consider media files as binary values, and if the database management systems do not offer any other tools to store media files, we will store them with the help of binary data type.

There is another common solution in which the media files are stored in the file system, and the database contains only their path. In this case, the NoSQL databases cannot offer support belonging to tools and normal working of the database for these data, like sharding, replication, and backup. We will not examine this solution.

## References

- [1] *DB-Engines Ranking*, 2023, URL: <https://db-engines.com/en/ranking>.
- [2] E. DUMBILL: *Volume, Velocity, Variety: What You Need to Know About Big Data*, 2012, URL: <http://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/>.
- [3] T. ERL, W. KHATTAK, P. BUHLER: *Big data fundamentals: concepts, drivers & techniques*, The Prentice Hall service technology series from Thomas Erl, OCLC: ocn918932107, Boston : [Vancouver, BC]: Prentice Hall ; Service Tech Press, 2016, ISBN: 978-0-13-429107-9.
- [4] Z. LI: *NoSQL Databases*, preprint, Open Science Framework, July 2019, DOI: [10.31219/osf.io/u5q3y](https://doi.org/10.31219/osf.io/u5q3y), URL: <https://osf.io/u5q3y> (visited on 01/05/2023).
- [5] A. B. M. MONIRUZZAMAN, S. A. HOSSAIN: *NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison*, arXiv:1307.0191 [cs] (June 2013), arXiv: 1307.0191, URL: <http://arxiv.org/abs/1307.0191> (visited on 10/16/2021).
- [6] P. J. SADALAGE, M. FOWLER: *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*, Upper Saddle River, NJ: Addison-Wesley, 2013, ISBN: 978-0-321-82662-6.
- [7] T. SHAFER: *The 42 V's of Big Data and Data Science*, 2017, URL: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>.
- [8] *SQL & NoSQL databases*, New York, NY: Springer Berlin Heidelberg, 2019, ISBN: 978-3-658-24548-1.
- [9] D. SULLIVAN: *NoSQL for mere mortals*, For mere mortals, OCLC: ocn889525203, Hoboken, NJ: Addison-Wesley, 2015, ISBN: 978-0-13-402321-2.

- [10] M. YOUNAS: *Research challenges of big data*, en, Service Oriented Computing and Applications 13.2 (June 2019), pp. 105–107, ISSN: 1863-2386, 1863-2394, DOI: [10.1007/s11761-019-00265-x](https://doi.org/10.1007/s11761-019-00265-x), URL: <http://link.springer.com/10.1007/s11761-019-00265-x> (visited on 01/05/2023).