

# No Relational and No SQL Databases

Ladislav Buřita<sup>1</sup>, Karel Zeman<sup>2</sup>,

<sup>1</sup>University of Defence, Kounicova 65, 662 10 Brno, Czech Republic and  
Tomas Bata University in Zlin, Mostní 5139, 760 01 Zlin, Czech Republic,  
[ladislav.burita@unob.cz](mailto:ladislav.burita@unob.cz)

<sup>2</sup>M-line a.s., Pekárenská 77/255, 370 04 České Budějovice, Czech Republic and  
University of Defence, Kounicova 65, 662 10 Brno, Czech Republic,  
[karel.zeman@m-line.cz](mailto:karel.zeman@m-line.cz)

**Abstract.** The content of the paper is a description of the actual available no relational and NoSQL databases. The main topic is to categorize the databases by data storage and data processing technology. There is a list of main representatives with functionality explanation, in any category. Each category and his representatives are evaluated in advantages or disadvantages. This document also contains a short description of InterSystems Caché technology that is a core of M-line information systems. Finally, the paper summarizes the importance of NoSQL technology and estimates the future development.

**Keywords:** Database, NoSQL, No Relational, BigTable, Column Store, Document Store, Graph, Object, XML, Multidimensional, InterSystems Caché.

## 1 Introduction

The paper is a particular result of the University of Defence (UoD) in Brno research project [1]. In the paper are first the basic terms defined. Second, the NoSQL databases are categorized and explained. Than is mentioned the InterSystems Caché, as an example of the no relational commercial system. Finally, the paper summarizes the importance of NoSQL technology and estimates the future development.

### 1.2 Relational Databases

The relation database is a database that has a collection of tables with data items, all of which is formally described and organized according to relational data model (RDM) [2]. The RDM is a data storage method that allows relations between data attributes to create logical data structures. This model uses Structured Query Language (SQL) to query specified data from a database. It is the reason that relational databases are sometimes called SQL databases.

The advantage of relational databases is their data storage versatility, which is given by the strictly defined descriptive model of simple two-dimensional data structures and where another dimension is created by using a relation.

The disadvantage is focus on mentioned strict descriptive model, where each additional nested dimension may cause a problem with the existing data model.

Getting data using SQL across multiple databases may cause the high demands on computing performance of the system. The result is the creation of redundant indexes, which help to speed up the queries and reduce the query complexity. Another problem is modeling of complex systems, where index tables outweigh the data tables.

The facts above mentioned are the consequence why so called NoSQL databases were developed.

## **1.2 NoSQL Databases**

The NoSQL database is a data storage system that uses for data access not only SQL language. These systems primary use user defined retrieval methods across given specialization for maximum performance. This is their greatest advantage. NoSQL databases have excellent capabilities to distribute required power to multiple data sources – that is because there is no strict data integrity need. Every database can be represented by separate logical part – by regions, data quality, data type, etc. [3].

The producer of the principle of NoSQL databases is Google Company with their BigTable technology. It was developed to store the content of web pages and links between web pages. BigTable technology uses simple NoSQL data storage method key-value (row key, column key) with an added timestamp attribute. The application interface provides only data lookup; data insert and data delete methods [4].

## **2 NoSQL Databases Categories**

### **2.1 Column Store**

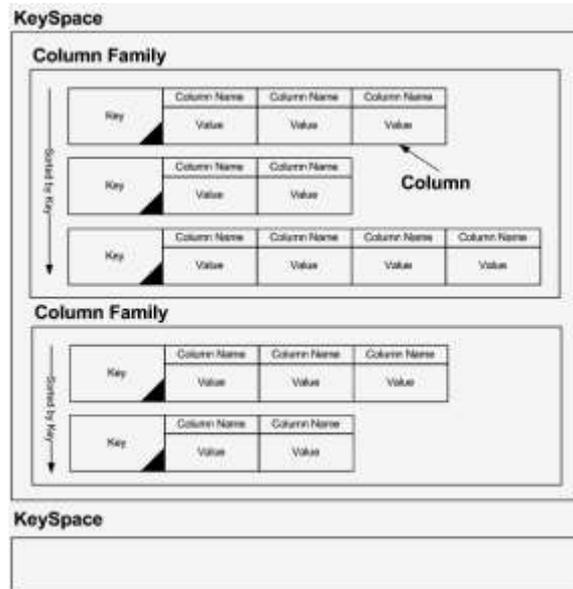
The Column Store databases are designed for distributed storages of large volumes of data. Stored data are represented by set of attributes KeySpace, Column Family, Row, Column and Value. Every Column has assigned extra timestamp, which may also represent the time after the record is deleted, see Fig. 1.

The advantage of this data model is the simplicity for storing and retrieving data, as well as scalability, or the solution ability in a distributed environment. In distributed environment data are stored on a predetermined number of database nodes by replication strategy and gradually are data replicated to other nodes.

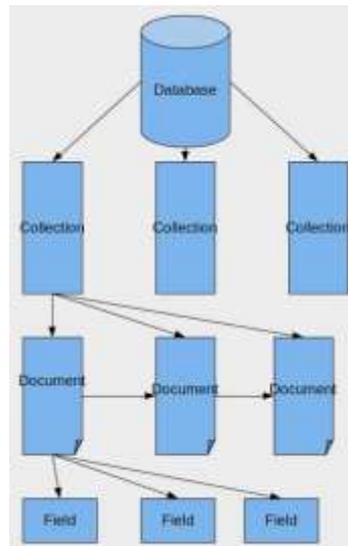
Main representatives of Column Store technology are Hadoop (heavily used database platform by Yahoo!), Cassandra (developed by Facebook, now the most popular NoSQL database), Hypertable and Accumulo (based on Google BigTable).

### **2.2 Document Store**

The Document Store databases are primary designed to store document oriented data called Documents. Each Document may contain data enclosed in XML, JSON or BSON envelope. The content of the document is structured with sets of attributes, see Fig. 2.



**Fig. 1.** Structure of the Column Store database (source [7])



**Fig. 2.** Structure of the Document Store database (source [8])

Every Document is identified by unique key that represents a particular document data. The key can be a simple string, URL or path. Document also can be stored in collections. The advantage of this data model is saving space compared to traditional storage used in DMS (Document management systems) and faster access to the document data with user indexes.

Main representatives of Document Store technology are CouchDB (uses JSON objects), MongoDB (fast database suitable not for only document data, but pictures and videos stored in BSON data model and collections).

### 2.3 Graph Databases

The Graph databases solve the issue of relationship between data elements. They are based on the principle of the existence of nodes and edges that connect them with many other attributes, see Fig. 3. It is more difficult to create new record, but it is much easier work with existing record. The search speed is the main advantage of graph databases. The main representatives of Graph databases are Neo4j [5] (young and very progressive database), InfoGrid (focused on web based applications) or HyperGraphDB.

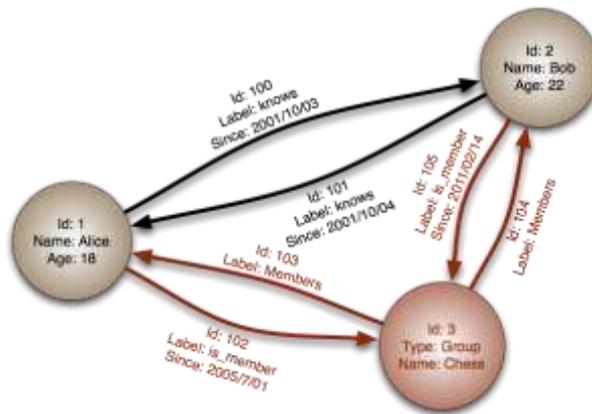


Fig. 3. The scheme of the Graph database (source [10])

### 2.4 Key Value

The Key Value databases have data storage mechanism using simple Key – Value attributes like: color is red, name is John etc. It is required specify Key as unique value to query and get unique answer. The advantages of this technology are speed, simplicity and horizontal scalability. The disadvantage of the model is that not all data can be defined as pair of key – value. Main representatives of Key Value technology are DynamoDB (use Amazon within its cloud solution) or Redis (data primarily stored in operational memory and stored into persistent memory in predetermined intervals; very fast solution).

### 2.5 Multidimensional Databases

Multidimensional databases use to store data multidimensional tree structures called root and subscripted nodes and leaves. Every subscript of defined tree can have its

value. Physically the data structures are stored in binary trees, see Fig. 4. The advantage of the principle of data storage is a very fast access to the target value. In case when system uses balanced trees (B-trees) is the final path length optimized to achieve the value. Main representatives are InterSystems Caché, Globals (free platform of Caché, has no object oriented access to data) and iBase.

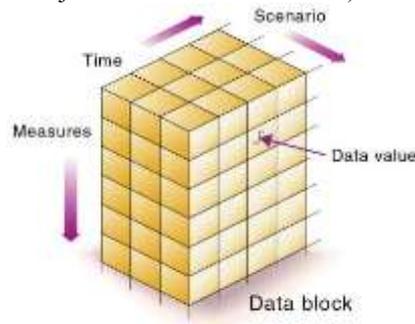


Fig. 4. The Multidimensional database (source [9])

## 2.6 Object Databases

The Object or Object-oriented are databases, where data are represented as an object. It defines classes and methods, and uses principles of inheritance. A typical representative data instances for object databases are multimedia data. The advantages of object-oriented databases are fast operations with objects, attributes maintenance and easy expansion of object-data model. Main representatives of Object databases are db4o (for .NET and Java platform) and Versant (object database enterprise solution).

## 2.7 XML Databases

The XML databases store XML documents. It is based on standard definitions of XML and its validation schemes. Databases allow perform indexing based on existing XML tags and intelligently search the contents of XML documents through XPath or XQuery language. The main representatives are eXist and Berkeley DB XML.

## 2.8 NoSQL Databases Summary

The origin of NoSQL databases was to obtain better performance for databases focused on specific tasks, issues and data processing. It has gained a large group of supporters thanks to leading companies as Google, Facebook, Yahoo!, etc. At the present, modern NoSQL databases are more powerful and scalable than typical relational databases. Modern NoSQL databases include object oriented languages with distributed control of data storage. The main advantages of NoSQL databases are fast access to data, orientation on specific task and distributed processing.

### **3 InterSystems Caché**

The InterSystems Caché database has developed from its original multidimensional technology to the object, multivalued database, so today InterSystems Caché belongs rather into Multimodel databases [6].

#### **3.1 System Description**

Current database model allows connect multidimensional data through database object structures to query data using SQL language. That technology is by InterSystems called like Post-Relational technology. InterSystems Caché is designed for business application, client-server and web applications processing client-side or server-side transactions.

The InterSystems database combines data storage, data and application server and developer tools together. Data in Caché database can be stored directly into their original data structures called Globals, or can be maintained by object with full object-oriented access.

Object structure is automatically converted into Globals by Caché database engine. These structures can be queried by SQL language.

InterSystems Caché database in recent years has been promoted in hospitals and healthcare facilities with application extension InterSystems HealthShare. This is mainly due to its robustness and ability to process huge amount of data.

#### **3.2 Experiences with Database**

More than one decade of experience with Caché technology has author of the paper (Zeman) in design data models for applications based on the large amount data processing.

Original Caché data models use data access direct via Globals. The advantage of this solution is fast design of a data model and implementation of data access methods in database. The disadvantage is that only programmer perfectly familiar with the data model is able to use effectively this approach.

In addition, if the data model is designed incorrectly, it is difficult to update them to the system requirements. For example: if the programmer needs to insert another index to the definition of node of existing data model, it forces to overwrite all existing methods of data structure.

New versions of Caché brought new opportunities to work with data and store them as objects with appropriate methods for operating data. However, database Caché internally maps objects structures into Globals, so that all causes significant performance degradation. Caché therefore introduces the possibility of additional indexation like hyper indexes or binary indexes. All this is at the cost of increased data storage.

Finally, after many personal experiences we return to strict definition of data storage in Globals for maximum performance. However, it is necessary to spend a lot of time defining the correct data model of the specific solution.

## 4 Conclusions

The No-SQL databases have in recent years gained a large group of supporters, mainly due deployment in such companies as Google, Facebook, Yahoo!, Amazon and more. The biggest advantage of No-SQL databases is focus on work with specific types of data, which ensures No-SQL databases, high performance, scalability and adaptability to specific issues.

The InterSystems Caché has been in recent years promoted in hospitals and healthcare facilities, thanks to its robustness and ability to process huge amounts of data with the HealthShare application.

The fact predict a No-SQL databases for the development and further expansion of the business and user companies, especially in sectors dealing with specific storage and large data volume.

This study is at first the preparation stage for the dissertation process (author Zeman) and then provides the basis for consideration of the use of NoSQL databases in the Ministry of Defence and the Army of the Czech Republic. Last but not least will be the study used in the teaching at the UoD (author Buřita).

## References

1. Project of the Institutional support to the Ministry of Defence, Czech Republic, to development of the research organization - University of Defence (UoD), Faculty of military technology, CIS department "Advanced technologies in communications and information systems", subproject: "Information and knowledge management in NEC environment". UoD: Brno, Czech Republic (2011-2015)
2. WIKIPEDIA The Free Encyclopedia, Relational data model, [http://en.wikipedia.org/wiki/Relational\\_Database](http://en.wikipedia.org/wiki/Relational_Database) (2013)
3. NOSQL Databases, <http://nosql-database.org> (2013)
4. Pros and cons of using NoSQL solutions, <http://blog.sphereinc.com/2012/03/pros-and-cons-of-using-nosql-solutions/> (2012)
5. Share PowerPoint presentations and documents, <http://www.slideshare.net/> (2013)
6. InterSystems Caché, <http://www.intersystems.com/> (2013)
7. Cassandra demystified, <http://bigdata-cassandra.blogspot.cz> (2013)
8. NoSQL DB Mongo DB, <http://sjohn4.wordpress.com> (2013)
9. Understanding Multidimensional Databases, <http://docs.oracle.com> (2013)
10. WIKIPEDIA The Free Encyclopedia, Graph database, [http://en.wikipedia.org/wiki/Graph\\_Database](http://en.wikipedia.org/wiki/Graph_Database) (2013)