

An Analysis of NoSQL Database Systems with Regards to Strengths, Weaknesses and Utility

COS 326 essay submitted by

Group Members:

Natasha Schoeman u14009562

Emilio Singh u14006512

Renton McIntyre u14312710

Presentation date: 13 September

Neil Leavitt(2010) presents the question of whether NoSQL databases will become viable alternatives to traditional DMBS paradigms. In the paper, the author highlights some of the limitations of existing RDBMS and examines the elements of a NoSQL database as well as the challenges in its adoption. He also presents some of the advantages and disadvantages to the paradigm in depth. Finally he posits the utility of the NoSQL database for specialised applications. The objective of this essay is to analyse the arguments presented and evaluate the value of the NoSQL paradigm both in the context of academia (as is depicted by COS326) and industry by examining the arguments presented in a variety of research papers and technical compositions.

According to Moniruzzaman & Hossain (2013),NoSQL,"Not only SQL", refers to nonrelational data management systems that do not use tables as representation mechanisms and avoid the use of SQL to function unlike the current dominant database paradigm (Leavitt 2010), the Relational Database. This model has a number of particular limitations that makes it potentially unsuitable for specific database applications. In particular a major drawback of the language SQL and its derivatives is poor distribution and scaling over machines/networks and a reliance on data that is structured such that it conforms easily into tables (Leavitt 2010). This makes them unsuitable for dynamic, unstructured applications whose implementation spans multiple scalable devices McCreary (2014). NoSQL database systems, on the other hand, make use of various methods (Leavitt 2010) that overcome these issues, such as the use of key-value and document-based stores for data and the use of column-oriented databases for scalable, complex data that can be stored by indices in the form of unstructured document types or using top-down traversal models. (Leavitt 2010) posits that NoSQL databases will generally have faster data processing times due to no strict or mandated enforcement of the ACID (Atomicity, Consistency, Isolation, Durability) restraints which is only possible by making use of a non-SQL-based approach. According to (Leavitt 2010), NoSQL systems have drawbacks associated with the lack of reliability that comes from more efficient processing and increased complexity and cost in the face of existing large investments in SQL systems.

The critical point made about NoSQL is that it is not meant to replace traditional SQL systems but rather offer an alternative database for specific use. This is supported by Moniruzzaman & Hossain (2013) who analysed NoSQL for Big Data applications. The scalability McCreary (2014), of NoSQL systems in both

virtual and real computer architectures makes it very attractive for data-fluid enterprises. In a whitepaper, (CouchBase 2013) posited the exponential growth of ‘unstructured data’ to vastly exceed structured data. If this prediction is to be believed, then conventional RDBMS models will prove insufficient for future data needs as already indicated in their limited capacity to work well with unstructured data. The increasing prevalence of unstructured data has a contiguous effect on the data-processing performance requirements. Consider that SQL-languages, generally, make use of JOIN statements in order to serve complex or compound data requests. Joins are expensive to perform and if there is an increased need for complex data, will adversely affect data servicing if done often. In contrast, NoSQL systems avoid join operations entirely and can process complex data requests using alternative methods to deliver information faster than traditional Relational Database systems. Furthermore, the rise of Cloud computing and distributive computing architectures and models means that traditional table-based database systems will be insufficient to meet performance needs whereas many NoSQL systems can be designed with intent for distributed implementation from the onset CouchBase (2013).

The relevance of NoSQL to academia is great as it represents a new field of database study. As the model develops and is adopted more, increasingly it will be important for research into the capacities, failings and opportunities that using NoSQL can provide. Already, we see this impetus as both BaseX and Neo4J, XML document and graph databases respectively are covered by the course syllabus. These additions provide us, the students, with additional database tools with which to draw upon in creating applications to process data. This allows us to choose the more appropriate tools to accomplish a task in the face of increasingly difficult and diverse problems.

The importance of the NoSQL database system is large in industry. We often see companies like Facebook use proprietary NoSQL systems for their storage services, Yahoo created SandStorm for Cloud services, Google’s has their BigTable system and Amazon’s has Dyanamo Chandra (2015). These companies invested heavily into NoSQL alternatives because of the changing nature of data, from traditional financial transaction data that could be encapsulated easily in tables to complex, composite data without a clear structure Moniruzzaman & Hossain (2013). For businesses, investments into database systems represent large amounts of resources in both capital and man hours. For companies like Amazon, Google and

Facebook to be investing their resources into NoSQL solutions for their data needs gives some validation to the NoSQL paradigm in terms of its suitability for unstructured, scalable data processing.

In conclusion, there are two primary merits to the NoSQL paradigm, being scalability and suitability for unstructured data. These merits facilitate the use of NoSQL based systems for use in commercial industry-grade applications and for smaller applications without any loss of functionality or efficiency. NoSQL systems still have their weaknesses in both access to skills, additional complexity and reliability issues compounded by an existing market dominance by traditional RDBMS models. Despite this, NoSQL is a valid alternative database paradigm that will likely see a greater share of the market in the future.

References

- Chandra, D. G. (2015), 'Base analysis of nosql database', *Elsevier* **13**(21).
- CouchBase (2013), Why nosql?, Technical report, CouchBase.
- Leavitt, N. (2010), 'Will nosql databases live up to their promise?', *IEEE* **43**(2), 12–14.
- McCreary, D. (2014), Making sense of nosql, Technical report, Midwest Architecture Community Collaboration.
- Moniruzzaman, A. & Hossain, S. A. (2013), 'Nosql database: New era of databases for big data analytics- classification, characteristics and comparison', *Internation Journal of Database Theory and Application* **6**(4).