Brasilia's Database Administrators

Jane Adriana* and Maristela Holanda*

University of Brasilia, BRAZIL

*Correspondence to: Janeadriana88@gmail.com, mholanda@unb.br

ABSTRACT

Database administration has gained an essential role in the management of new database technologies. Different data models are being created for supporting the enormous data volume, from the traditional relational database. These new models are called NoSQL (Not only SQL) databases. The adoption of best practices and procedures, has become essential for the operation of database management systems. Thus, this paper investigates some of the techniques and tools used by database administrators. The study highlights features and particularities in databases within the area of Brasilia, the Capital of Brazil. The results point to which new technologies regarding database management are currently the most relevant, as well as the central issues in this area.

1. INTRODUCTION

Database administration is defined as a set of management tasks and database maintenance, from database projects, analysis, privilege granting, backup, retrieval and recovery. The database management activities are performed by database administrators (DBA) (Elmasri and Navathe 2010). Database management requires particular DBA expertise to run adequate practices and procedures on databases. There are many complex tasks related to the database community, and, Aiken *et al.* (2011) suggest that "data management is still evolving". Large enterprises, such as public institutions, multinational corporations, financial and telecommunications companies have exceptional database administrators for managing database systems, to ensure availability and reliability, wherever it is requested.

With advances on the internet, the world is becoming increasingly connected, and, subsequently, the volume of information continues to grow exponentially. In a review of the International Data Corporation, Tauro *et al.* (2012) showed an increase of nearly 25 times the volume of data between 2007 and 2010. In Abiteboul *et al.* (2003), a summary of a group of academic database researchers affirms: "Database needs are changing, driven by the internet and increasing amounts of scientific and sensor data". New technologies enable the sharing of resources, but also add to the complexity of data management. Thus, with the growing volume of data and the use of NoSQL databases (Adriana and Holanda 2015), the adoption of practices and procedures for carrying out database administration prove to be a new and central challenge.

This article presents the results of a survey, which investigated the environment and procedures used by database administrators, and special NoSQL database administration in Brasilia, the capital of Brazil. Brasilia represents a technological headquarters because of the various Government Systems. The main objective of this study is to investigate the methods and practices of database administrators in managing database systems. Further aims were to investigate the database administrators with new technologies, such as NoSQL databases.

The survey presents a portrait of the situation in Brasilia, the current work scenario, techniques, procedures and the practices used by the DBA, throughout 2015-2016. The results reported in this article highlight the database needs, which will help database communities and enterprises to identify relevant questions in the area of databases, and improvements that they could adopt. This research was based on the work of Holt *et al.* (2015). Their study included interviews of a diverse database community to identify what practices and procedures were utilized in the database lifecycle.

The remainder of this article is organized in 5 sections. In Section 2, the survey is presented. In Section 3, the results obtained in each question are described. In Section 4, the results analysis is presented. Finally, in Section 5, the conclusions are presented.

2. The SURVEY

For this study, a survey was used because it generates information from a brief interview from respondents about a specific topic. The survey reported here uses an elaborated questionnaire intended to cover aspects of the database area, exploring topics such as architecture, design, security, cloud database, database management and training.

This survey was web-based and was carried out via e-mail and also social networks, such as Twitter, Facebook and WhatsApp. We collected data as broadly as possible across the database community in Brasilia. The questionnaire was comprised of 20 different types of questions, including dichotomous, multiple choice and Likert scale (Albaum 1997). In an attempt to obtain the richest survey data and a comparative analysis of survey data by period, the survey was administered among 44 respondents in two different periods: between May 25 and June 20, 2015; and April 13 and April 18, 2016. The aim was to report the findings relevant to the target audience in 2015-2016.

The survey instrument used was Google Forms, with the following link https://docs.google.com/forms/d/1nlpgcWVjW4oWnkH96JdTe15ZXqoaHZ2no-Jp2ZWH3_s/viewform?usp=send_form.

The goal of the survey was to identify the database management methods, practices and procedures that are used in the database lifecycle. We also aimed to collect demographic data on the database administration, and investigate their level of expertise in new technologies in the field of database management.

The research investigated practices and techniques used by the DBA together with database architecture, demographics and training, and also DBA knowledge about new technologies in the field of database management. The survey covered different class organizations, servers, databases, change management, and new technologies, such as NoSQL databases.

3. RESULTS

In this section, the results of the questionnaire are presented. We divided the results into sections according to the subjects: database administrator expertise; enterprise and database demography; query language and training.

3.1 Database Administrator Expertise

Figure 1 illustrates the results of the question regarding years of experience in the field of database administration. Results show that 40% (2015) and 27.3% (2016) of respondents have over 11 years of experience in the field, and 33.3% (2015) and 36.4% (2016) have between 2 (two) and five (5) years of experience in the database field.
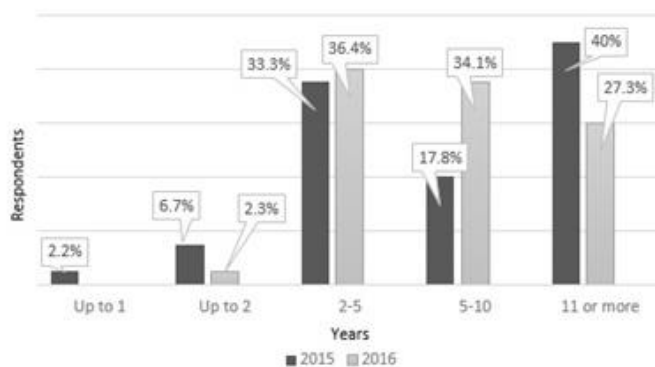

Figure 1: DBA Expertise (in years).

Other results show that the majority are familiar with relational databases, as shown in Figure 2, where 100% of the respondents report some knowledge of relational databases. However, only 11.1% (2015) and 13.6% (2016)

have experience with NoSQL databases, indicating that this is a new concept for most respondents. With these results, it is possible to conclude that over the period of one year there were no changes.

Results of the survey also indicated information regarding the companies where the research subjects work. Table 1 shows, 64.4% (2015) and 63.6% (2016) of respondents worked in organizations with 500 employees or more. In Brasilia, those are considered large organizations.
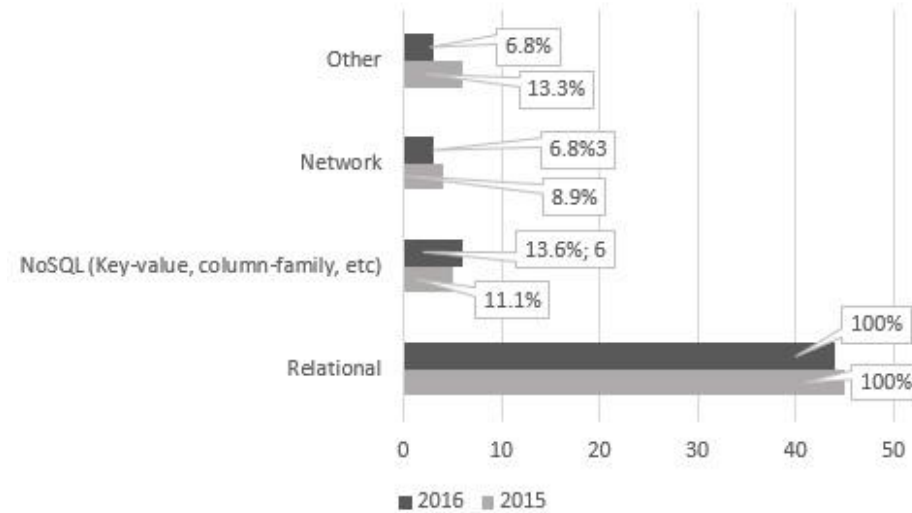

Figure 2: Database Models.

TABLE 1.   ENTERPRISE'S NUMBERS

Enterprise	% (2015)	% (2016)
Up to 100	4.4	4.5
100-300	13.3	18.2
301-500	8.9	11.4
501 or more	64.4	63.6

Armbrust et al. (2010) defined "cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services.". The companies pay for use of computing resources, the cost is on demand (Haihong et al. 2011). In this context, we asked about this new technology in Brasilia's community DBA. In our survey (Figure 3), 97.8% (2015) and 100% (2016) use the owner platform, and 15.6% (2015) and 18.2% (2016) that adopt cloud infrastructure. We attribute these values to the fact that the IT community in Brasilia works mainly within government institutions, which have proprietary infrastructure.
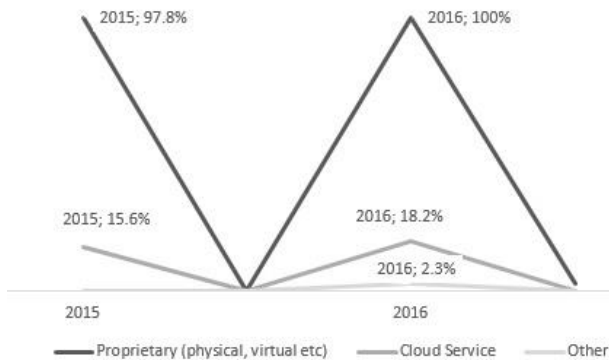

Figure 3: Platforms.

In addition, Table 2 shows, 33.3% (2015) and 47.7% (2016) have between 11-50 databases, with 46.7% (2015) and 31.8% (2016) reporting 100 databases or more. Thus, half of the answers report large numbers of databases.

TABLE 2.   DATABASE DEMOGRAPHIC IN ENTERPRISES

Demographic	% (2015)	% (2016)
Up to 10	11.1	4.5
11-50	33.3	47.7
51-100	8.9	15.9
101 or more	46.7	31.8

For 46.7% (2015) and 45.5% (2016), the database size is between 1TB-20TB, and 28.9% (2015) and 22.7% (2016) is above 20TB, as shown in Table 3.

TABLE 3.   DATABASE SIZE

Size	% (2015)	% (2016)
Up to 100GB	4.4	0
101-500GB	13.3	20.5
501GB-1TB	4.4	11.4
1-20TB	46.7	45.5
20TB or more	28.9	22.7

3.2 Enterprises and Database Demography

As shown in Figure 4, large enterprises usually use relational databases, but new models are also on the rise.

For database changes, 42.2% (2015) and 38.6% (2016) reported some structural change (command execution, schema changes, object creation) and were carried out weekly (Table 4). Holt et al. (2015) make aware of "changes to the database engine, structure and hardware all require practices and procedures to ensure that data is protected".
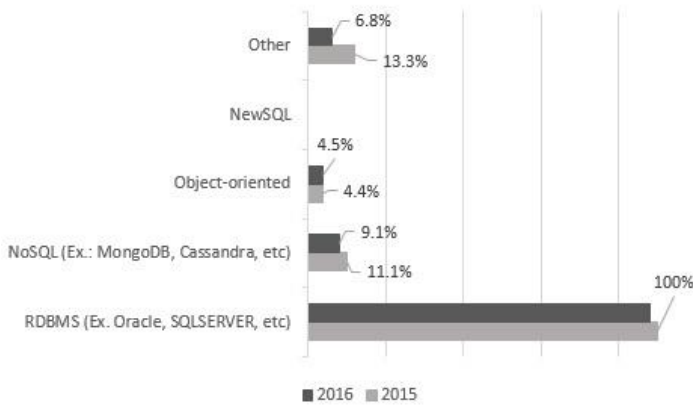

Figure 4 . Enterprises Database Models.

TABLE 4.   STRUCTURAL CHANGES

Changes	% (2015)	% (2016)
Up to 2	11.1	18.2
1-10	42.2	38.6
11-20	17.8	20.5

Changes	% (2015)	% (2016)
21-40	13.3	2.3
40 or more	13.3	20.5

Figure 5 shows that organizations are greatly concerned about the security requirements, availability, performance and backup. This profile has not changed within the enterprises in the last two years. Database storage is becoming a greater concern of many enterprises: 46.7% (2015) and 52.3% (2016).
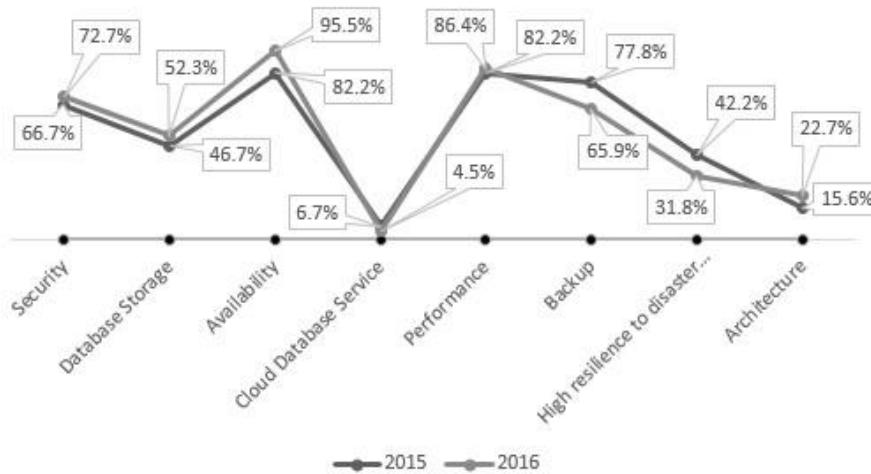

Figure 5 . Concerns.

3.3 Query Language

The SQL language is one of the main languages used for database queries, data manipulation and data retrieval in traditional databases. In addition, the SQL language provides a declarative language interface level so that users can easily declare what they wish (Adriana and Holanda 2015).

New database languages are being created to support the NoSQL database models. Thus, we would like to know how important the SQL language is for database administrators at the present time. Our findings show that 59.1% (2015) and 79.5% (2016) of DBAs use the SQL language as a tool for database administration (Figure 6). In short, a query language should make data manipulation possible, facilitating the administration of the database system.
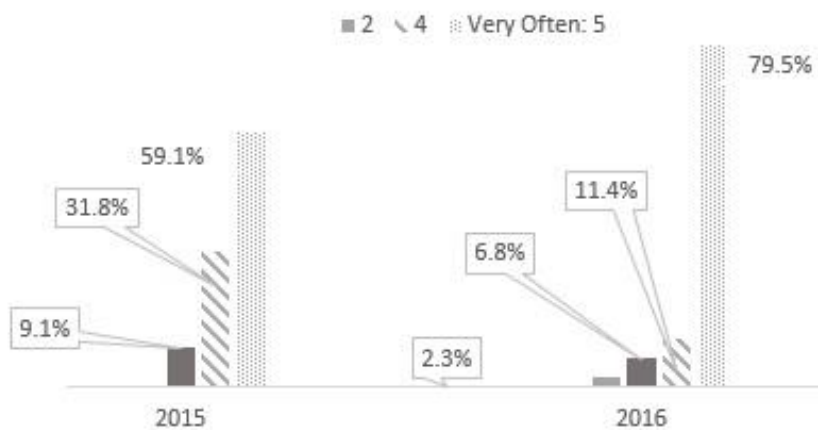

Figure 6. SQL Language.

A high-level query language is considered essential for 95.5% (2015) and 93.2% (2016) of respondents (Figure7).

153

Figure 7. High Level Language.

The database administration should be able to rely on adequate tools to detect problems, improve performance, report generation, upgrade, and manipulate data. If there are no tools available, thereby compromising the database administration, this could cause problems in maintaining it. As seen in Figure 8, the use of administrative tools for distinct database models is encouraging for more than 80% (2015, 2016) of respondents.
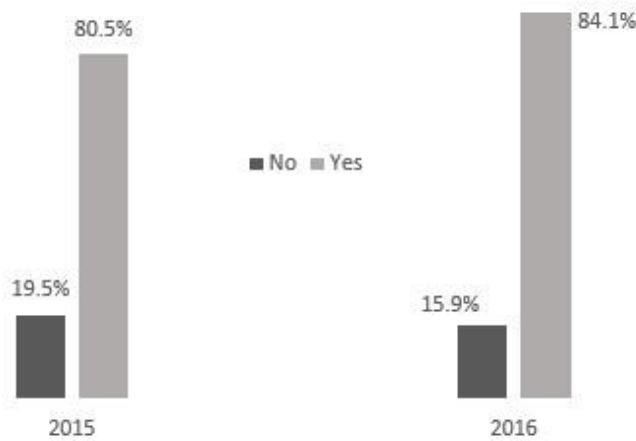

Figure 8. Tools.

3.4 Training

In the professional training category, the following statistics present percentages of the essential elements: 95.6% (2015) and 95.5% (2016) had engaged in self-study, and 82.2% (2015) and 77.3% (2016) read articles and research on forums, 51.1% (2015) and 43.2% (2016) were trained in a provider course, and 28.9% (2015) and 13.6 (2016) have the opportunity to attend external database conferences and workshops (Figure 9).
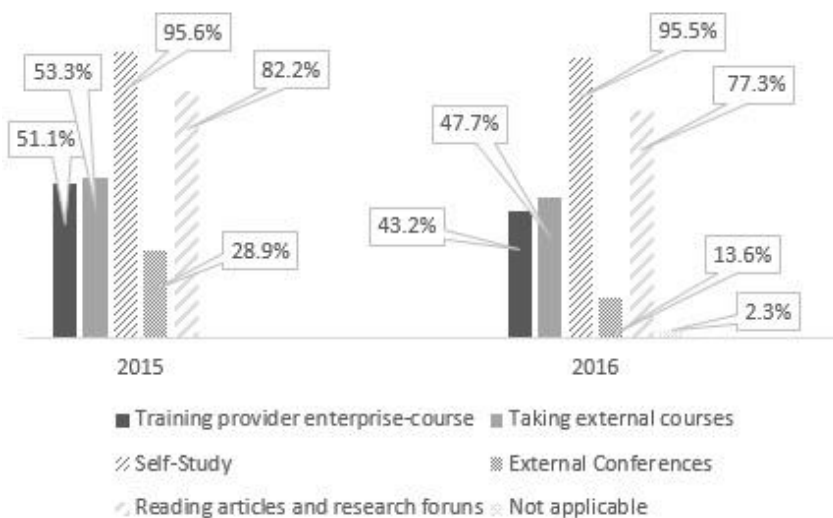

Figure 9. Professional Training.

Another issue in professional training deals with Big Data. Big Data is a collection of very huge sets of data with a great diversity of types, which is difficult to process using traditional data processing platforms (Assunção 2015; Chen and Zhang 2014). In this context, we asked Brasilia's DBA community about how much knowledge they had on this subject. Only 4.4% (2015) and 2.3% (2016) of the respondents reported working on or having knowledge of Big Data (Figure 10).
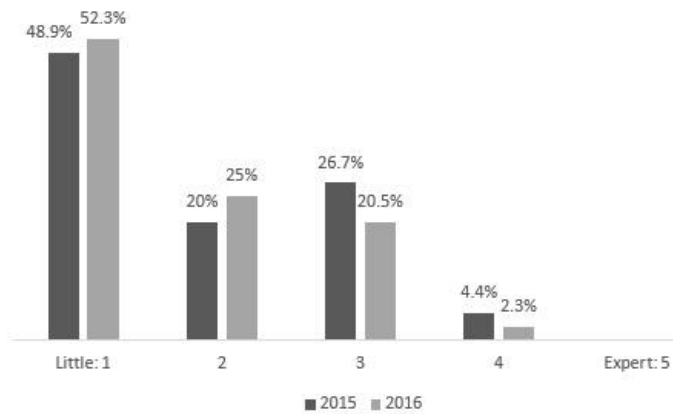

Figure 10. Big Data.

NoSQL database has emerged promising high performance, the ability to handle modern workloads, and ability to process large volumes of unstructured data. As presented in Figure 11, NoSQL databases are new for 46.7% (2015) and 38.6% (2016) of the respondents. It should be noted that an increase in interest about the subject in two phases of research (2015 and 2016), suggest the importance of this theme.

The NoSQL databases demand a high level of expertise by users/developers to write complex queries. The NoSQL databases have a very limited number of querying support and when there is one, no common query language is available in cases where every database differs in its supported query feature set (Adriana and Holanda 2015). These results provide evidence for a considerable need for training in this field, therefore, we asked about this in next question.
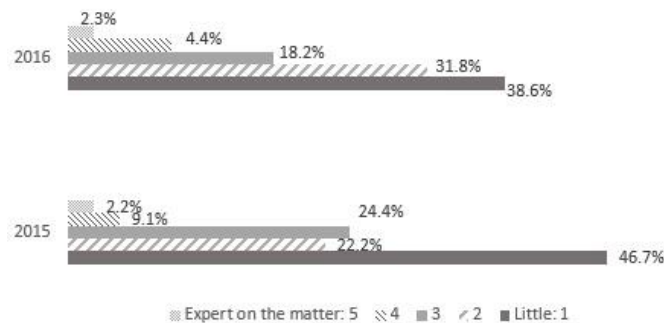

Figure 11. NoSQL Knowledge

Respondents, at the rate of 80% (2015) and 81.8% (2016), also consider that distinct database systems, together with some peculiarities, require a high level (4 and 5 levels) of expertise, for complex database administration, which creates the need for frequent training/studies (Figure 12).
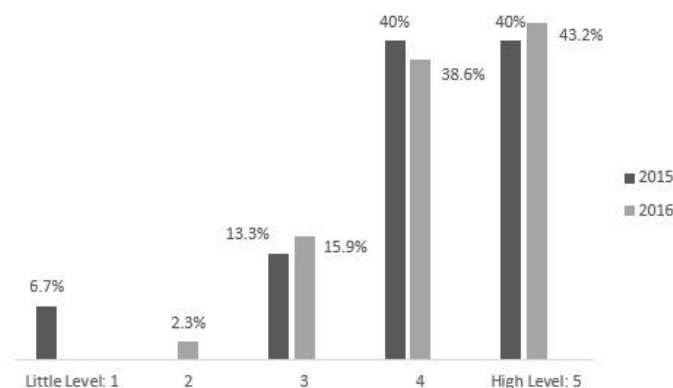

Figure 12. Specialization Level.

4. RESULTS ANALYSIS

Analyzing the findings from the surveys some interesting points emerge:

- The questions were answered by 44 Brasilia's DBAs respondents who had indicated that they and their organizations are involved with database management systems and associated tools.

- The survey aimed to characterize the population of respondents, performing a comparison of answers throughout 2015-2016, in general terms, asking for information about the situation of the database administrators in Brasilia, and the current work scenario, techniques, and practices used by this community.

- The results, which are displayed in tables 2, 3, 4 and Figure 5, illustrate that the members of the survey population are engaged in a variety of activities with regard to different stages of the lifecycle as well as on various institutional levels.

- The cloud database is becoming a choice for a broad range of enterprises, such that the results show 11.1% (2015) and 9.1% (2016) use the NoSQL databases, and 6.7% (2015) and 4.5% (2016) are concerned about the cloud database services. However, the survey highlights that only 2.2% (2015) and 2.3% (2016) of respondents has NoSQL knowledge, and 4.4% (2015) and 2.3% (2016) know Big Data concepts.

- Overall the majority of DBAs consider that a high-level query language is essential to database management, and a large proportion, 59.1% (2015) and 79.5% (2016), use SQL language as an essential database tool. These findings correspond with our expectation of responses within our target audience.

- The survey results suggest that there is a substantial need for training in new database models of the DBA community.

- The respondents consider that database systems require a high level of expertise, 80% (2015) - 81.8% (2016), and the use of administrative tools for distinct database models is encouraging for 80.5% (2015) - 84.1% (2016) of them.

- There are various tools for database management, and the survey suggests that the SQL language is the most applicable, providing user-friendly query languages, which have been indicated as an interesting option for the majority of DBAs.

5. CONCLUSIONS

The findings of our research raise relevant questions in the area of databases, and identify what practices and procedures were utilized in the database at two different moments in time: in 2015 and 2016. The questions in the survey aimed at shedding light on the conditions of DBA with regard to database administration.

Database systems must provide data structures and advanced techniques to improve access and retrieval activities for data management. For the management of large volumes of data, new models are being implemented, such as NoSQL. However, the analysis of the findings from the survey suggests the respondents (database administrators) are not familiar with these new systems. There are many components in database administration so practices and procedures may be affected by these.

In terms of the latter, it is particularly difficult to find database administrations with both subject or domain knowledge and technical expertise in many heterogeneous database models. This requires that staff constantly keep up to date with new databases systems and technologies. In summary, the findings described suggest a great demand for training to acquire the skills and competences needed to maintain database models.

In conclusion, the results of the research suggest a great demand for knowledge in new technologies like Big Data and NoSQL, which arises from a lack of qualified staff in the field. The survey detected, during the 2015 - 2016 period, that the possibility of providing user-friendly query languages and exhibiting the results efficiently is an interesting option for DBAs.

6. REFERENCES

Abiteboul, S., Agrawal, R., Bernstein, P., M., Ceri, S., Croft, B., Naughton, J., et al (2003). The Lowell database research self-assessment, *Published as Microsoft Technical Report MSR-TR*-2003-69, pp. 1-10.

Adriana, J. and Holanda, M. (2015). Query Languages in NoSQL Databases. *15th Chapter of Handbook of Research on Innovative Database Query Processing Techniques, in IGI Global book series Advances in Data Mining and Database Management (ADMDM) IGI Global*, pp. 415-437.

Aiken, P., Gillenson , M., Zhang, X. and Rafner, D. (2011). Data management and data administration: assessing 25 years of practice, *in J. Database Manag.* Vol. 22, pp. 24-44.

Albaum, G. (1997). The Likert scale revisited: an alternative version. Avaliable at: *http://repository.ust.hk/ir/Record/1783.1-17657* (Accessed Apr, 2016).

Armbrust, M., Fox, A., Griffith, R., Joseph, A, D., Katz R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). *A view of cloud computing: Communications of the ACM*, Vol. 53, pp. 50-58.

Assunção, M, D., Calheiros, R. N., Bianchic, S, Nettoc and M. A. S., Buyya, R. (2015). Big Data computing and clouds: Trends and future directions, in J. Parallel Distrib. pp. 79–80.

Chen, P. and Zhang, C. (2014). Data-intensive applications, challenges, techniques and technologies: *A survey on Big Data, in Information Sciences*, pp. 314–347.

Elmasri, R. and Navathe, S. B. (2010). *Fundamentals of Database Systems (6th Edition). Pearson,* ISBN-10: 0136086209.

Haihong, J. H. , E , Le, G., and Du, Ju. (2011). Survey on NoSQL database: *proceedings of the 6th International Conference on Pervasive Computing and Applications (ICPCA)*, pp. 363-366.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Khan, S. U. (2015). The rise of "big data" on cloud computing: *Review and open research issues. Information Systems*, pp. 98–115.

Holt, V., Ramage, M., Kear , K. and Heap, N. (2015). The usage of best practices and procedures in the database community. *Department of Computing & Communications, Faculty of Mathematics, Computing and Technology, The Open University, UK*, pp.163-181.

Tauro, A. and Shreeharsha, A. (2012). Comparative study of the new generation, agile, scalable, high performance nosql databases: *proceedings of the International Journal of Computer Applications*, vol. 48, pp. 1- 4.

Yan, L. (2015). Handbook of Research on Innovative Database Query Processing Techniques. *Nanjing University of Aeronautics and Astronautics*, China.