Reduce data processing time in NoSQL databases based on Grover's algorithm

H.Amellal
Department of computer sciences
Faculty of Sciences
Rabat, Morocco
amellal@yandex.ru

A.Meslouhi
Department of Physics
Faculty of Sciences
Rabat, Morocco
jemjid@gmail.com

A.EL Allati
Department of Physics
Faculty of Sciences and Techniques
Al-Hoceima, Morocco
abdouallati@gmail.com

ABSTRACT

In order to reduce data processing time in NoSQL databases, we propose in this paper a quantum approach to extracting information from unstructured databases. In fact, we apply Grover's algorithm instead of classical algorithms to search in NoSQL databases.

KEYWORDS

NoSQL databases, data mining, quantum computing, Grover's algorithm

ACM Reference format:

H.Amellal, A.Meslouhi, and A.EL Allati.2018. Reduce data processing time in NoSql databases based on Grover's algorithm. In *3rd International Conference on Smart City Applications (SCA '18), October 10–11, 2018, Tetouan, Morocco.* ACM, New York, NY, USA. https://doi.org/10.1145/3286606.3286806

1 INTRODUCTION

The big explosion of digital data has forced the scientists to find new methods to study, search, share, storage, analysis and presentation of data. Thus was born the "Big Data"[1]. It is a concept for storing an unspeakable number of information on a digital basis. The Big data represent literally a massive data, they refer to a very voluminous set of information that no conventional database management or information management tool can really work. Therefore, we can generate trillions bytes of data every day. Generally, the information coming from everywhere: Emails, social media, phone messages, GPS signals and other sources. In fact, the Big Data represents a promising solution designed to allow everyone real-time access to big databases. It aims to propose a solution to the traditional databases such as Platform of Business Intelligence in server SQL and much other. Accordingly, to optimize the processing times on Big Data, there are different methods that developed, such as NoSQL databases. Classically, big data based on several data meaning algorithms for the data processing, however this solutions was founded on principles of classical computing, which make the operations process take a long time. Therefore, we propose in this study to profit advantages of quantum computing to reduce processing time in big data by using Grover's algorithm [2, 3].

The paper is organized as follows: In Sec. II, NoSQL databases. In Sec.III, our proposed Grover's algorithm review to reduce time search in NoSQL databases. Finally, conclusion is drawn in the last section.

2 NOSQL DATABASES

2.1 Relational databases

In simplest terms, a database is a means of storing data in such a way that information can be retrieved from it. A relational database is one that presents data in tables with rows and columns. Relational databases are known as a relational database management system (RDBMS). Moreover, virtually all relational database systems use structured query language (SQL) as a language for querying and maintaining the database. Accordingly, when we talk about high-traffic sites and databases, we rarely hear about relational databases. Indeed, guaranteeing the consistency of data costs expensive in time and is often incompatible with performance. Since, the relational model does not seem appropriate in environments requiring large architectures. In fact, using RDBMS it has many advantages this includes: the technology is mature, so that today SQL is a standardized language, the ability to implement complex queries and wide support is available and strong communities. Unfortunately, the relational database management system (RDBMS) suffers from a combination of drawbacks such as: the modification of the established model can be costly, on a very large volume of data, the model performance can be limited, for some companies, the license price is very high. In order to achieve the requirements of customers, a new movement was born from the initiative of cloud computing architects and social sites: NoSQL (Not Only SQL) was going to enter the world of data representation [1].

2.2 Unstructured databases

NoSQL refers a new family of database management systems that deviates from the classical paradigm of relational databases (RDBMS). It consists of many databases, which are characterized by non-relational logic in data representation. Therefore, no query interfaces in SQL. Unstructured databases are designed to solve the problems of volume, multi-source and multi-format data processing in big data environments. However, no distinction is made in terms of volume or diversity of data when talking about NoSQL technologies. Accordingly, no relational storage encompasses a wide variety of different database technologies have evolved in order to overcome those scalability limits in building modern applications. In general, there are different types of NoSQL databases:

Key-Value stores: The data was organized in tables, there is no schema and the value of the data is opaque. Values are identified and accessed using a key. Each key has to be unique to provide non-ambiguous identification of values. The stored values can be numbers, strings, or counters. The key-value bases are for example adapted to the collection of events (online games), the management of traces (measurement of audience) and user profiles management of high audience sites (e-commerce).

Wide column store: The wide-column store is a type of key-value database, has some passing similarity to principle of RDBMS in its use of rows, columns and tables. The important difference here is the names and format of the columns can change from row to row in the same table.

Graph databases: is a type of database that uses graph theory to store, map, and query relations. It consists essentially of a set of nodes and edges. Each node represents an entity and each edge, connection, or relation between two nodes. Each edge is defined in turn by a unique identifier, a starting node and / or an arrival node, as well as a set of properties. Graph databases are particularly suited to the analysis of interconnections, which explains their great interest in exploring data from social networks. They are also useful for manipulating data in disciplines involving complex relation and dynamic patterns, such as supply chain management, identification of the origin of an IP telephony problem.

Document stores: is a database for applications that manage documents. It is a compilation of documents or other information presented in the form of a database in which the full text of each referenced document can be viewed online, printed or downloaded. For example, this type of databases is used by university libraries and is intended for students and faculty. They are particularly suitable for online courses that can be downloaded from the Internet for distance learning.

From the above we conclude that the importance of NoSQL databases is indisputable. Therefore, increasing data processing speed is inevitable specially that the quantity of information circulating increases every day. Accordingly, we propose to profit quantum computing to reduce the data time processing in NoSQL databases by applying Grover's algorithm [4-10].

3 GROVER'S ALGORITHM REVIEW

Grover's algorithm proposes a smart search through a massive data in order to find the intended information. We consider a search space of size N with no prior knowledge about the data presentation in it, where we want to find an element of that search space satisfying a known property. Classically, this problem requires approximately N operations. Grover's database search algorithm enables a dramatic reduction in the computational complexity of seeking in NoSQL database, where this algorithm allows it to be solved using approximately $\sqrt{N}$ operations [2, 3, 6].

We consider a function $f(y)$ on bits with n inputs where $y \in \{0,1\}^n$. The output of the function is a single bit, so we can have $f(y)=1$ if y is solution to the search problem and $f(y)=0$ otherwise (no such y exists).

Classically, to find the value of y that realize $f(y)=1$ in the generating outputs we should have $2^n - 1$ tries to solve the problem. By using Grover's algorithm, we can solve the problem with only $\sqrt{2^n}$ tries. For example, if we consider 4bits, the classical computer will be need $2^4 - 1$ operation to find the correct y .By using Grover's algorithm we need $\sqrt{2^4 = 4}$ operation to find the correct y .Therefore, Grover's algorithm can be described as a smart quantum database searching algorithm, which reduces the number of operations necessary to solve the problem as compared to a classical algorithms.

Application:

Considering a user's site web database "UD" with N different elements where $id \in \{0, N-1\}$, and we are searching for id_0 where $UD[id_0] = "Grover"$. We give this search problem a general mathematical structure as follows:

Input: A black box U_f for computing an unknown function $f: 0, n^n \to 0, n$.

Output: Find an input $id \in 0, N-1^{nn}$, where $f(id_0)=1$

We consider a quantum state jid0i, which represent the correct value searching, where $f(id_0)=1$, and $f(id)=0$ represent the error values of $|id\rangle$. To solve the problem by using Grover's algorithm we create an input superposition state and rotate it into $|id_0\rangle$ by applying a Grover's operator G.

Then, we consider n bit input state $|0\rangle \otimes n$, and applying the Hadamard gate $H \otimes n$ on the state $|0\rangle$ generate a superposition of states. We denote the state $|\phi\rangle$ is the superposition of all possible states $|id\rangle$ where:

$$|\phi\rangle = \frac{1}{\sqrt{2}} \sum_{id \in o,1^n} |id\rangle \quad (1)$$

We have $|id_0\rangle \in |\phi\rangle$, then

$$\langle id_0 | \phi \rangle \frac{1}{\sqrt{2}} \sum |id \langle id_0 | id \rangle\rangle = \frac{1}{\sqrt{2^2}} \quad (2)$$

To eliminate $|id_0\rangle$, we consider the orthonormal basis of $|id_0\rangle$ where:

$$\left|\phi'\right\rangle = \frac{1}{\sqrt{2^n - 1}} \sum_{id \in 0,1^n, id_0 \neq id} \left|id\right\rangle \qquad (3)$$

Then, we consider two operators. The first given by:

$$\left|\phi\right\rangle = \sqrt{\frac{2^n - 1}{2^n}\left|\phi'\right\rangle + \frac{1}{\sqrt{2^n}}\left|id_0\right\rangle} \qquad (4)$$

Then:

$$\left\langle\phi\middle|\phi'\right\rangle = \sqrt{\frac{2^n + 1}{2^n}} \qquad (5)$$

From the above we can write:

$$\left|id_0\right\rangle = \sqrt{2^2}\left|\phi\right\rangle - \sqrt{2^2 - 1}\left|\phi'\right\rangle \qquad (6)$$

Therefore

$$U_s\left|id_0\right\rangle = \sqrt{2^n}\left|\phi\right\rangle - 2\sqrt{2^2 - 1}\frac{2\sqrt{2^n - 1}}{2^n}\left|\phi\right\rangle + \sqrt{2^n - 1}\left|\phi'\right\rangle \qquad (7)$$

We substitute $\left|\phi\right\rangle$ as following:

$$U_s\left|id_0\right\rangle = 2\sqrt{2^2 - 1}\frac{2\sqrt{2^n - 1}}{2^n}\left|\phi'\right\rangle + (\frac{2}{2^n} - 1)\left|id_0\right\rangle \qquad (8)$$

Then:

$$U_s\left|\phi'\right\rangle = \frac{2\sqrt{2^n - 1}}{2^n} \qquad (9)$$

Accordingly, we consider an angle θ, where:

$\sin\theta = \frac{2\sqrt{2^n - 1}}{2^n}$ .And we notice that the system in a rotation, therefore:

$$U_s\left|id\right\rangle = \cos\theta\left|id_0\right\rangle - \sin\theta\left|\phi'\right\rangle \qquad (10)$$

And

$$U_s\left|\phi'\right\rangle = \sin\theta\left|id_0\right\rangle - \cos\theta\left|\phi'\right\rangle \qquad (11)$$

Know we apply a Grover operator defined by: $G = U_s U_f$.

Then, we obtain a more familiar form of a rotation, namely:

$$G\left|id_0\right\rangle = \cos\theta\left|id_0\right\rangle - \sin\theta\left|\phi'\right\rangle \qquad (12)$$

And

$$G\left|\phi'\right\rangle = \sin\theta\left|id_0\right\rangle - \cos\theta\left|\phi'\right\rangle \qquad (13)$$

The principle is that the Grover operator rotates the state $\left|\phi\right\rangle$ into the state we are searching for $\left|id_0\right\rangle$. We have apply this rotation multiple times t. Then:

$$G^t\left|id_0\right\rangle = \cos\theta\left|id_0\right\rangle - \sin\theta\left|\phi'\right\rangle \qquad (14)$$

And

$$G^t\left|\phi'\right\rangle = \sin\theta\left|id_0\right\rangle - \cos\theta\left|\phi'\right\rangle \qquad (15)$$

4 CONCLUSION

In this paper, we have proposed to profit advantages of quantum computing to reduce data processing time in unsorted databases by applying the Grover's algorithm. The theoretical results are very encouraging, where the quantum search will be reducing the search time in minimum 4 times compared with the classical one. However, with today's technology, it's complicated to integrate the quantum algorithm's in the actual classical infrastructure, but the implementation of quantum key distribution in modern cyber security it makes quantum algorithms closer than ever before to the implementation phase.

REFERENCES

[1] P.Zadrozny and R.Kodali Big Data Analytics Using Splunk: Deriving Operational Intelligence from Social Media, Machine Data, Existing Data Warehouses, and Other Real-Time Streaming Sources ISBN: 143025761X ,9781430257615 (2013).

[2] L.Grover, (1996). Fast quantum mechanical algorithm for database search. In Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC '96), pp.212âĂŞ219.

[3] L.Grover,(1997). Quantum mechanics helps in searching for a needle in a haystack.Physical Review Letters 79(2), 325âĂŞ328.

[4] P.A.M.Dirac, The Principles of Quantum Mechanics, 3rd ed.Oxford: Clarendon Press (1947).

[5] C.Kollmitzer, M.Pivk, Applied quantum cryptogtraphy, Lect.Not.Phys 797, ISBN 978-3-642-04829-6, Springer(2010).

[6] D.McMahon: Quantum computing explained wile interscience A John Wiley Sons, Inc, Publication, Computer society IEEE (2007).

[7] Phillip Kaye,Raymond Laflamme,Michele Mosca: An Introduction to Quantum Computing, Oxford university press (2007).

[8] Sandor Imre, Ferenc Balazs: Quantum Computing and Communications An Engineering Approach, John Wiley Sons, Ltd (2005).

[9] D.Aharonov (1998): Quantum computationâĂŤa review. Annual Review of Computational Physics VI, World Scientifc, 259âĂŞ346.

[10] A.Ambainis (2007): Quantum walk algorithm for element distinctness. SIAM J.Comput.37,210âĂŞ239