

A Review of JavaScript Object Notation in Data Analysis

Miaochan Zhou

Queen's University, School of Computing, Kingston, Canada, K7L 3N6
Shilishuang@cas-harbour.org

ABSTRACT

JSON is an extremely popular data format in the 2000s. It is used for transforming the data type. It is not different for reading and writing, and it is simple for machines. As time goes through the age of big data, there appears a group of applications, for example, the two that will be discussed in this essay: NoSQL and NewSQL, for data analysis and big data. The SQL database is always a popular database since the 1980s. In recent years, key-value storage, which always means NoSQL, became more and more popular. Also, the NewSQL database type was developed to solve a similar challenge more efficiently. This essay aims to find the application of these data formats in the real world. According to the research results, JSON could be used as the data format when the data are shifting between the Android application and the Web server.

Keywords: data format, database, JSON, NoSQL, NewSQL

1. INTRODUCTION

JSON (JavaScript Object Notation) is a lightweight data exchange format. It is easy to read and write. It is also easy for machines to parse and generate. JSON is based on JavaScript programming language, which is a subset of standard ECMA-262 3rd Edition – December 1999. JSON is a data format promoted and used by Douglas Crockford in 2001. It officially became the main data format from 2005 to 2006, since Yahoo and Google began to widely use JSON format [1].

The relational database or Structured Query Language (SQL) database, such as PostgreSQL, MySQL, and Oracle, is always the real interface of the database since the 1980s and is the base of electronic trade in the entire world. In recent years, for the purpose of helping to analyze the data used for Internet searching, key-value storage for representing big sparse tables has been developed, which always refers to NoSQL databases, such as Google BigTable, Apache Accumulo, and MongoDB, for the purpose of helping to analyze the data used for Internet searching. Thus, most of the data on the Internet are now analyzed by using key-value pairs storage to do analysis. In order to manage similar challenges, a relational database developed a new kind of database, which is NewSQL, including C-Store, H-Store, SciDB, VoltDB, and Graphulo, to sustain the new analytical functions in the

database. The concept of SQL, NoSQL, and NewSQL is also mixed into the hybrid processing system, like Apache Pig, Apache Spark, and HaLoop [2].

The purpose of this paper is to introduce and discuss JSON, NoSQL, and NewSQL when people are dealing with database and data analysis problems. Therefore, this paper will give a brief introduction and then state the area in which they are used.

2. ANALYSIS

2.1. JSON

JavaScript Object Notation (JSON) is a lightweight, text-based, language-independent data representation format. It comes from the ECMAScript Programming Language Standard, and it set a definition of a small set of stylization rules for the portable representation of structured data. The purpose of the design of JSON is to minimize the data and make them transplantable and text-based and to make JSON a subset of JavaScript. It can be parsed quickly and efficiently. JSON uses text to represent the information of JavaScript objects, supports string, number, object, array, and other types, and provides a method to quickly read data. It is widely used in data collection and data mining. JSON is used to interchange data which is written in these programming languages: ActionScript、C、C#、ColdFusion、

Common Lisp, E, Erlang, Java, JavaScript, Lua, Objective CAML, Perl, PHP, Python, Rebol, Ruby, and Scheme [3].

JASON model has two structures. The first one is called the object (Figure 1). One object contains a sequence of non-sorted key-value pairs, and it begins with "{" and end with "}". Each key-value pair is separated by a comma. The second structure is an array. An array is a set of a number of values (Figure 2). It begins with and ends with a pair of square brackets, and every two elements of the array are separated by a comma. The terminology "array" and the terminology "object" are from JavaScript [1].

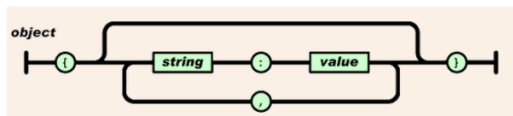


Figure 1. one of the JASON model structures: object

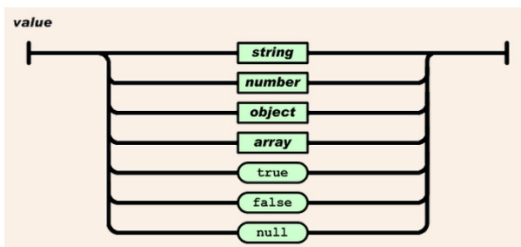


Figure 2. one of the JASON model structures: value

JSON servers for the Android system. In the Android system, the application can have direct interaction with the database server only when interacting with a web server. It is essential to add, delete, and change data by using a Web server. Then the data are exchanged between the Android application and Web server in order to make the Android application handles the database. During this process, the data is in the form of JSON [4].

The specific steps are described as follows. First of all, the application of Android gets some data and instructions, which are provided by the users from the interface. It uses available JSON tools to assemble the data about the operating database to RequestJSON and then sends it to the Web server. Next, when the webserver finishes accepting RequestJSON, it can receive real request data by JSON analysis tools. Then, people can use these data to operate the database and get the results at the same time. The data for results will be connected to ResponseJSON and will return to the Android application. Finally, when ResponseJSON is received by the Android application, it can use JSON analysis tools to gain result data, and show the results to users for analysis mode. Therefore, the whole work is composed of two parts which are the phone client-side and the Web server [4].

2.2. NoSQL

NoSQL is short of "not only SQL", which means a group of data controlling systems that is eclectic, not relational, and gradually familiar. In this data controlling system, the database is not established most by tables. Meanwhile, SQL is usually not used for data operation. The database controlling system of NoSQL is essential when managing a large number of data since the data's property does not need relation models [5].

The systems of NoSQL are databases that are distributive and not related to each other. It is designed for saving extensive data and for managing a large quantity of parallel data which crosses a mass of commodity servers. They also use languages and mechanisms which are not SQL languages to interact with data. The database system of NoSQL appears together with dominating internet companies such as Google, Amazon, and Facebook, as the solution of traditional relational database management system cannot resolve the challenge of handling a mass of data. They can support a variety of activities including analysis about exploration and prediction, ETL-styled data transition, and non-task important OLTP (for instance, controlling time-cost or inter-organization tasks). These systems are initially operated by Web 2.0, and the design could expand to thousands or millions of users for updating and reading, compared with traditional DMBS and DW (data warehouse) [5].

The thought of relational database development is to offer a kind of method of data storage that uses structured searching or SQL. The history of these databases could trace back to the 1970s when the data type was not as complicated as today. Besides, at that time, the cost of storage and data archiving was huge. As social media become popular, the events, objects, and people to be stored were increasing rapidly. Nowadays, the use of data is not only limited to storing data, but it also expands to frequent data searching and managing to serve the function of generating real-time summaries and ordering advertisements [6].

Since the message to be managed is complicated and it needs to deal with many databases demand in order to answer individual API requests or to show web pages, the requirement for present database systems is always increasing. Some crucial driving elements of this area are the requirement for interaction, incremental complexity, and continuing developing user web. In order to satisfy these increasing requirements, complicated strategies and modified infrastructure for calculation are used. However, deploying individual servers is expensive and highly complex, which results in the tendency of using cloud hardware. Moreover, the use of swift methods shortens the time of developing and deploying, which allows it to answer the users' requirements more quickly [6].

There are two methods in NoSQL to solve the problem that the relational database has. The first one is called manual sharding. For the purpose to utilize distributed paradigm, it is needed to be split into smaller units, and then these units must be stored on different machines. This process is called manual sharding. However, this process could only be realized by the developing people since this process cannot be used in a traditional database. In addition, the data stored in each instance is executed anonymous. Application codes are responsible for cutting data into segments, storing data in the way of distribution, performing searching management, and aggregating results to users. The other method is called distributed cache. Caching is the procedure that is often used. It is mainly used to enhance the reading ability of the system. Moreover, the use of caching does not have any effects on writing performance and could have a large growth on the complexity of the whole system, as figure 3 shows. Hence, if the demands from the system are read-intensive, then it is necessary to consider distributed caching. Moreover, write-intensive or read/write-intensive applications do not need the distributed cache [7].

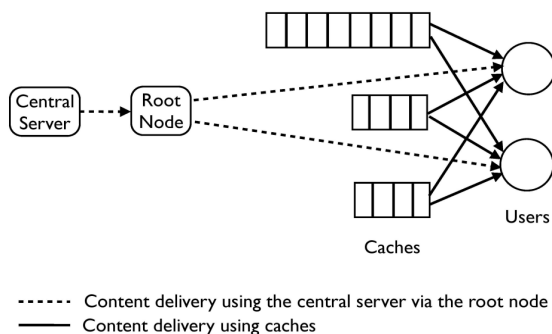


Figure 3. A cache set consisting of two users of three cache services

2.3. NewSQL

NewSQL, which is used for Online Transaction Processing (OLTP), is the next generation of extensible relational database management systems (RDBMS). It offers extensible properties to read and write workload of NoSQL system and maintains the atomicity, consistency, isolation, and durability (ACID) warranties of the traditional database system. Through the methods of using functions of NoSQL style or using technologies such as in-memory processing, massively parallel processing (MPP), or symmetric multiprocessing (SMP), these systems make a breakthrough to the limitation of traditional RDBMS performance and gather NoSQL or searching components, in order to deal with the challenge of the volume, variety, velocity, and variability of big data [7].

Some of the NewSQL databases are NuoDB which is a distributional database that uses the design of SQL.

The design of SQL refers to all of the ACID properties and standard SQL language supportive and related logic. NuoDB is a web-scale distributional database, which offers abundant SQL realization and real ACID events. It was designed as a distributional system at the beginning, which allows it to expand its service to cloud service and increase usability and elasticity. As an extensible cloud database, NuoDB is designed for the contemporary data center. It also is the NewSQL solution for simplifying the deployment of applications [8].

NewSQL movement presents a shift of big data. Using NewSQL-style data analysis in BigTable, which is designed for key-value storage and is created by Google, has an extensive meaning. First of all, it increases the locality by saving the computation and storing data in the same address. The saving from the aspect of data communication enhances the performance of some computers. In addition, in order to accelerate the reusing of the fundamental structure, it can avoid the spending of configuring an extra system, because organizations prefer solutions that already exist to solutions that are unfamiliar and un-tested from new systems. Finally, it could use the database functions, such as visiting the subsets of data quickly selectional can be used. In the distributional database, in return for cooperation with the database accessing path, the algorithms running in the database gain distributional executing for free [8].

3. CONCLUSION

In conclusion, JSON is used for the Android system. When the data is managed between the Android system and the server, the data form is JSON. The whole process is divided into three parts. The first step is to get the information and send it to the Web server. The second step is to edit the accepted data and return the data to the Android application. The last step is to show the result to users.

NoSQL is used to deal with non-relational databases. It is used by the most popular internet companies to manage data. It is different from SQL in that the language NoSQL uses is not SQL language. NoSQL can support various activities of data management because it has the ability to deal with abundant data. NoSQL could solve the problems that happened in recent years of a relational database. The first method is manual sharding. Manual sharding is the process that separates the task into smaller parts and stores each part on different machines. The second method is distributed cache. Caching could make the system more readable and could increase the system's complexity, so it is suitable for read-intensive systems.

NewSQL is a new generation of a database management system that is used for OLTP. It extends

the functions of NoSQL and traditional database systems and also has some progress to deal with new challenges. A special example of NewSQL is NuoDB that is a distributional database that uses the design of SQL. It can expand the ability of NewSQL to could service. NewSQL can be used in Google's BigTable which is for key-value storage. The usage of NewSQL

in BigTable has a very important meaning because it increases the locality of the data, avoids the spending of configuring an extra system, and has the ability to use database functions quickly selectional.

During the comparesion between JSON and NoSQL, they are referenced in different data systems. As the table 1 show below:

Table 1. RDBMS vs NoSQL vs NewSQL databases

	Traditional RDBMS	NoSQL	NewSQL
SQL	Supported	Not supported	Supported
Machine dependency	Singe machine	Multi-machine/Distributed	Multi-machine/Distributed
DBMS type	Relational	Non- relational	Relational
Schema	Table	Key-value, column-store, document store	Both
Storage	On disk + cache	On disk + cache	On disk + cache
Properties support	ACID	CAP through BASE	ACID
Horizontal scalability	Not supported	Supported	Supported
Query Complexity	Low	High	Very High
Security concern	Very high	Low	Low
Big volume	Less performance	Fully supported	Fully supported
OLTP	Not fully supported	Supported	Fully supported
Cloud support	Not fully supported	Supported	Fully supported

This article only briefly introduces NewSQL and JSON through research and qualitative analysis, as well as the elaboration of their application methods, and lacks the verification of actual operation. And the way of comparison has not been verified by specific data and experiments. In the follow-up, the two data processing methods will be compared and tested through experiments to obtain more accurate comparisons and application differences.

ACKNOWLEDGMENT

First of all, I would like to express my deepest gratitude to the teachers and professors of my university for their inspiring guidance on the knowledge points of the dissertation, and also very grateful to them for providing me with invaluable guidance at every stage of the dissertation writing. Guidance and encouragement.

In addition, I would like to thank my parents for their selfless encouragement and support during the preparation and completion of the thesis, as well as their support for my academic and professional development.

Finally, I would like to thank my friends for their impressive kindness when I needed help, without whom I would not have been able to successfully complete my dissertation.

REFERENCES

- [1] M. Wickham, Introduction to JSON [J]. Practical Android, 2018.
- [2] J. Kepner, V. Gadepally, D. Hutchison, H. Jananthan, A. Reuther, Associative array model of SQL,

NoSQL, and NewSQL databases, 2016 IEEE High Performance Extreme Computing Conference (HPEC), IEEE, 2016.

- [3] D. Crockford, The application/json Media Type for JavaScript Object Notation (JSON)[J], RFC 4627, 2006.
- [4] H. Fang, J. Chen, B. Xu, The Interaction Mechanism based on JSON for Android Database Application[J]. Information Technology Journal, 2013, 12(1), pp. 224-228.
- [5] A. Moniruzzaman, S.A. Hossain, NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison[J]. International Journal of Database Theory & Application, 2013, 6.
- [6] S. Khan,X. Liu, S.A. Ali, M. Alam Bivariate, Cluster and Suitability Analysis of NoSQL Solutions for Different Application Areas[J]. 2019.
- [7] A. Moniruzzaman, NewSQL: Towards Next-Generation Scalable RDBMS for Online Transaction Processing (OLTP) for Big Data Management[J]. Computer Science, 2014, 7.
- [8] D. Hutchison, J. Kepner, V. Gadepally, B. Howe, From NoSQL Accumulo to NewSQL Graphulo: Design and Utility of Graph Algorithms inside a BigTable Database. In 2016 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2016, pp. 1-9.