物联网应用的MySQL与NoSQL数据库比较

摘要

物联网（IoT）概念在技术界已经存在了近几年。IOT专注于智能设备数量的连接。在不久的将来，物联网将在各个领域有应用，这些应用将产生巨大的数据量。随着异构数据的不断生成，数据的存储、传输和管理问题越来越突出。传统的数据库系统使用结构化查询语言（SQL）数据库，支持简单、鲁棒性、灵活性、可扩展性和性能的所有用户需求。但他们面临的主要限制是静态模式，这使得RDBMS不适合于物联网应用。另一方面，市场上出现的NoSQL数据库具有比SQL数据库更好的性能。NoSQL数据库是非关系的、无模式的、无连接、易于复制支持、水平可伸缩等。NoSQL在所有应用场景中都比SQL更好吗？本文对此进行了尝试。本文将SQL和NoSQL数据库应用于小型喷水灭火系统的物联网应用，并研究了NoSQL在不同场景下的性能是否优于SQL。

关键词: 数据库管理系统（DBMS）；物联网（IOT）；NoSQL；SQL；

1. 说明

物联网（IOT）领域正在进行大量的研究工作。物联网指的是我们周围任何与互联网相连的事物。这个对象还能够通过互联网进行通信。传感器连接到这些物体上，比如连接到烤箱、冰箱的温度传感器或者仅仅测量室温。物联网是这样的事物的网络，它们可以彼此通信或与用户通信。物联网在智能城市、工业、医疗服务等领域有着广泛的应用，由于应用领域广，这些数据所产生的数据类型是异构的。在不久的将来，这些物联网应用所产生的数据将是大数据，可以进一步分析以使资源得到最佳利用。此外，使用物联网应用将增加实时数据的生成，使得存储和管理数据变得具有挑战性。一些新的技术和数据库管理系统（DBMS）正在出现，以处理数据增长和提高系统性能。到目前为止，SQL数据库一直是最流行的数据库系统，问题是传统的关系数据库是否会按标准执行。

随着用户需求的不断增加。一种新的数据库正在出现并被称为NoSQL数据库。NoSQL数据库与传统的关系数据库系统的方法相比，以不同的方式存储数据。它们是用于无模式结构的数据。NoSQL数据库可以很容易地分发，因此它们提供了高的可扩展性和可用性。这些是管理大数据或物联网数据所需的属性。

1. 相关工作

为了实现物联网技术，在物联网的不同领域进行了大量的研究。IOT数据将是异构的，因为在不同的领域中有各种各样的物联网应用。这些IOT应用程序生成的数据需要进一步处理和分析，以便进一步优化应用程序。因此，数据采集、存储、数据处理、报表生成等任务都由数据库管理系统来管理。为了处理海量数据，使数据检索速度更快，作者3提出了三种RFID标签数据检索方法的设计、实现和评价。基于时延、功耗、通信距离和读取速率等参数，对交替、顺序、并行传感器采样调度方案进行了比较。数据的有限使用和传感器的数量是研究的局限性。作者4专注于面向应用的工作。采用紧急医疗服务相关的物联网应用来探讨数据访问方法。对于各种IOT应用程序，SQL或NoSQL数据库是否被证明是更好的问题，仍然没有答案。这样的比较研究是用云存储的数据进行的。对于特定的物联网应用，SQL和NoSQL数据库的分析是我们的兴趣所在。

1. 数据库管理系统
2. 数据库概述

数据库可以定义为结构化的数据集合。处理数据库的数据、事务、问题或任何其他方面的系统是数据库管理系统（DBMS）。使用结构化查询语言（SQL）的关系数据库管理系统是传统的数据库系统。市场的最新趋势是称为“非关系数据库”NOSQL。这两个数据库是彼此平等的潜在竞争对手。本研究旨在找出适合物联网应用的资料库。在对物联网应用数据库进行比较研究之前，对数据库的基本特性进行了探讨。数据库的这些特性由CAP定理给出，解释一致性、可用性和划分。

一致性：意味着一旦更新操作完成，每个人都可以从数据库中读取最新版本的数据，并且这样的系统是一致的系统。同时，一个系统，其中更新的数据不能被所有的用户一次被称为最终一致。可用性：如果系统总是提供连续操作，则实现。可用性是通过使用多个节点上的复制或分区数据将数据库部署为一组节点来实现的。在这种情况下，如果一个节点崩溃，其他节点仍然可以继续工作。分区容限：一个数据库系统，即使其中一个节点失效或不能访问，它也可以运行。这是通过将所有查询重定向到失败的节点到该系统的某个其他活动节点来完成的。

传统的数据库系统，如SQL，侧重于一致性，支持以下的酸性属性。

Atomicity：部分完成的事务被丢弃。事务只能成功或失败。一致性：在事务失败的情况下，系统恢复事务并返回到先前的稳定状态。因此，系统始终保持稳定。隔离：事务完成无干扰，独立处理。持久性：所有提交的事务都保存在日志中，不会丢失。这有助于在异常终止的情况下恢复系统。

另一方面，NoSQL数据库系统更多地关注可用性和分区，并给出最终的一致性。这些系统遵循基本属性。

基本可用性：NoSQL数据库关注CAP定理要求的数据可用性。软状态：数据库系统的状态是动态的，并且可能由于最终的一致性而随时间变化。数据库的所有副本都不必一直保持一致。最终一致性：在任何写入或更新或删除操作之后，系统可能无法立即反映所做的修改。但是，最终它会变得一致地显示所有复制品中的修改数据。

遵循基本属性的系统不是严格一致的。但是，所有的更新或修改最终都是可用的。在这样的系统中，当正在进行更新或复制过程时，客户端可能遇到数据不一致。但是，在复制完成后，数据将达到预期的一致状态。在物联网应用领域，用户数量、速度需求和用户需求不断增加，因此数据库的划分容限成为需要考虑的重要特征。NoSQL数据库优先考虑可用性超过一致性。另一方面，SQL遵循相反的顺序。目的是找出这两个数据库中的哪一个将更好地为小规模的物联网应用喷水灭火系统。

1. 从物联网角度看SQL vs NOSQL

SQL数据库：SQL数据库遵循关系数据模型来存储数据。在这个模型中，数据以表格形式存储在行和列中。相关表可以相互连接在一起。可用的各种关系数据库是MySQL、Oracle、SQLServer等。

NoSQL数据库：NoSQL遵循非关系数据模型。非关系模型支持数据的无模式存储，如文档、图形等。具有水平可伸缩性、无模式存储、支持非结构化数据的特性，NoSQL成为存储IOT数据的能手。NoSQL数据库的流行是因为它提供的特性包括高扩展性、易访问性和分布式体系结构。MangGDB、ReDIS、CouCHDB、HBASE是一些流行的NoSQL数据库。

从物联网的观点来看，必须从这两种数据库类型中选择正确的数据库。下面两个数据库类型之间的差异用IOT透视图来讨论。

* + 可扩展性：SQL数据库支持垂直可扩展性，而NoSQL支持水平可伸缩性。垂直可伸缩性指的是通过向同一节点添加诸如存储器或处理器等资源来提高单个节点的性能的能力。在水平可伸缩性中，增加节点（服务器）的数量，以便共享系统负载。对于正在开发阶段的新的物联网应用，使用具有未来扩展范围的数据库将是一个实际的选择。这将允许在必要时由IOT应用扩展资源，并且不需要高初始投资。
  + 数据检索：当用户必须从数据库中获取数据以进行进一步处理时，需要更快的数据检索功能。在SQL中，各种表链接在一起。为了查找来自不同表的数据，用户必须使用创建视图的连接语句。这是一个耗时的过程。另一方面，在NoSQL中，数据以包含所有相关数据的对象的形式存储。这消除了合并和显示数据的过程，从而节省了响应时间。
  + 系统成熟度：SQL是一种有经验的技术，因此大部分问题都已经解决了。安全特性如身份验证、数据机密性和完整性被结合在SQL中。另一方面，在NoSQL中还需要解决这样的安全特性。NoSQL可能会因为系统不够成熟而产生更多的问题或安全漏洞。在一些IOT应用中，需要一个安全的通信信道来传输敏感数据。对于这样的应用，最好使用安全的数据存储系统以及安全的通信信道。

结论

比较了MySQL和MunGDB数据库对由物联网应用生成的数据的性能。MySQL是一个传统的SQL数据库，MunGDB是一个NoSQL数据库，用于存储IOT数据。比较研究基于执行对不同数量的记录和线程的选择和插入查询所花费的时间。随着表的记录大小的增加，系统上的负载增加，进一步增加等待时间或响应时间。每个数据库都有它自己的优点和缺点。从该研究中，观察到在某些场景中，MyGDB与MySQL相比需要较少的响应时间。但是，MySQL响应与MangGDB相比稳定。因此，选择一个更好的IOT数据库取决于使用何种查询和应用需求。这项工作的未来研究运行包括与其他先进的负载测试工具可用的其他数据库管理系统进行的调查。通过使用云、分布式数据库可以进一步扩展研究，这将使工作更加逼真。

英文原文

MySQL and NoSQL database comparison for

IoT application

***Abstract*— Internet of Things (IoT) concept has been around in tech world for few years now. IoT focuses on connection of number of smart devices. In near future, IoT will have applications in various domains and these applications are going to produce tremendous amount of data. With the continuous generation of heterogeneous data, problem arises to store, transfer & manage the data efficiently. Traditional database systems used Structured Query Language (SQL) database which has supported all the user requirements along with simplicity, robustness, flexibility, scalability, performance. But the main limitation they are facing is their static schema which is making RDBMS not suitable for IoT applications. On the other hand, NoSQL databases emerging in market have claimed to perform better than SQL database. The NoSQL databases are nonrelational, schema free, no joins, easy replication support, horizontally scalable, etc. Does NoSQL perform better than SQL in all application scenarios? An effort to answer the same has been made in this paper. This paper compares SQL and NoSQL databases for a small scale IoT application of water sprinkler system and investigates whether NoSQL performs better than SQL in different scenarios.**

***Keywords— Database Management System (DBMS); Internet of Things (IoT); NoSQL; SQL;***

I. INTRODUCTION

A lot of research work is going on in Internet of Things(IoT) domain. IoT refers to any object/thing around us which is connected to internet. This object is also able to communicate through internet [1]. Sensors are attached to these objects/things, like temperature sensor connected to oven, fridge or just to measure a room temperature. IoT is network of such things which may communicate with each other or with user. IoT is having application in various domains like smart city, industrial, medical services, etc. As, the application area is wide, the type of data generated by these things is heterogeneous. In near future, data generated by these IoT applications will be big data which can be further analyzed to make the optimum use of resources. Also, the use of IoT applications will increase the generation of real-time data, making it challenging to store and manage the data. Several new technologies and database management systems (DBMS) are emerging to handle the data growth and to improve the system performance. SQL databases have been the most popular database systems so far, question arises whether the traditional relational databases will perform at par with ever increasing user requirements. A new class of database is emerging and is referred to as NoSQL database [2]. NoSQL databases store data in a different manner as compared to the traditional methods of relational database systems. They are meant for data of schema-less structure. NoSQL databases can be easily distributed and hence they provide high scalability and availability. These are the properties which are actually required to manage big data or Internet of Things data.

II. RELATED WORK

To realize the Internet of Things technology, a lot research is going on in different fields related to IoT. IoT data is going to be heterogeneous as there are various applications of IoT in varied domains. The data generated by these IoT applications is further supposed to be processed and analyze so as to optimize the application further. So, tasks like Data collection, storage, data processing, generating reports are managed by the database management system. To handle the huge amount of data and make the data retrieval faster, authors [3] have proposed the design, implementation and evaluation of three data retrieval approaches for RFID Tags. Alternate, Sequential & parallel sensor sampling scheduling schemes proposed were compared based on parameters such as time delay, power consumption, communication range & read rate. Limited use of data and number of sensors is the limitation of the study. Authors [4] are focusing on application oriented work. Emergency medical services related IoT application is used to explore data accessing methods. For various IoT applications, question of whether SQL or NoSQL databases prove to be better, still remains unanswered. Such comparative study are done [5] [6] with data stored in cloud. Analysis of SQL and NoSQL database for specific IoT application is of our interest.

III. DATABASE MANAGEMENT SYSTEM

*A. Database Overview*

Database can be defined as an structured collection of data. The system which handles the data, transactions, problems or any other aspect of the database is the Database Management System (DBMS). Relational DBMS which use structured query language (SQL) was traditional database system. The latest trend in market is the non-relational database known as NoSQL [7]. Both these databases are equal potential competitors of each other. This study is to find out the suitable database for IoT application. Before approaching to the comparative study of databases for IoT application, basic characteristics of database are explored. These characteristics of database are given by the CAP theorem which explains Consistency, Availability and Partitioning.

Consistency: means that once an update operation is finished, everyone can read that latest version of the data from the database and such system is consistent system. While, a system in which, updated data cannot be seen by all users at once is known as eventually consistent. Availability: is achieved if the system always provides continuous operation. Availability is achieved by deploying the database as a cluster of nodes, using replication or partitioning data across multiple nodes. In such case, if one node crashes, the other nodes can still continue to work. Partition Tolerance: A database system which can operate even if one of the nodes fail or is inaccessible. This is done by redirecting all queries to the failed node to some other active node of that system.

The traditional database systems like SQL, focused on consistency and supports following ACID properties.

Atomicity: Partially accomplished transactions are discarded. Transaction can only be successful or unsuccessful. Consistency: In the event of transaction failure, system reverts the transaction and goes back to the previous stable state. And hence, system always remains stable. Isolation: Transactions are completed without any interference and are processed independently. Durability: All the committed transactions are saved in logs and will not be lost. This helps to recover the system in case of abnormal terminations.

On the other hand, NoSQL database systems focus more on Availability and Partitioning and give eventual consistency. These systems follow BASE properties.

Basic Availability: NoSQL database focuses on availability of data as per CAP theorem requirements. Soft State: State of the database system is dynamic and may change over time due to eventual consistency. All the replicas of database do not have to be consistent all the time. Eventual Consistency: After any Write or update or delete operation, system may not immediately reflect the modifications done. But, eventually it will become consistent showing the modified data in all replicas.

The system following BASE properties is not strictly consistent. But, all the updates or modifications will be available eventually. In such systems, clients may encounter an inconsistency in data when updation or replication process is in progress. But, the data will reach the expected consistent state after the completion of replication. In IoT application domain, Number of users, speed requirements and user demands are ever increasing and hence partition tolerance of database becomes the important characteristic to be considered. NoSQL database gives priority to availability over consistency. On the other side, SQL follows the reverse order. The aim is to find out which of these two databases will perform better for a small scale IoT application of water sprinkler system.

236

*B. SQL Vs NoSQL from IoT perspective*

SQL Database: SQL Database follows relational data model to store the data. In this model, data is stored in rows and columns in a tabular form. Related tables can be interlinked together. Various relational databases available are MySQL, Oracle, SQLServer, etc.

NoSQL Databases: NoSQL follows non-relational data model. Non relational model supports schema-free storage of data in various forms such as document, graph. With features like horizontal scalability, schema less storage, support for unstructured data, NoSQL becomes competent for storing IoT data. Popularity for NoSQL databases spur because of the features it provides including high scalability, easy access, and distributed architecture. MongoDB, Redis, CouchDB, Hbase are some of the popular NoSQL databases available.

From IoT point of view, choosing the right database from these two database types is must. Differences between the two database types are discussed below with IoT perspective.

* *Scalability:* SQL database supports vertical scalability while NoSQL supports horizontal scalability. Vertical scalability refers to the ability to increase the performance of single node with adding resources such as memory or processors to the same node. In horizontal scalability, number of nodes (servers) is increased so as to share the system load. For new IoT applications which are in developing stage, using a database that has future expansion scope will be a practical choice. This will allow expansion of resources as and when required by the IoT application and no need for high initial investment.
* *Data Retrieval:* Faster data retrieval feature will be required when user has to fetch data from database for further processing. In SQL, various tables are linked together. To lookup data from different tables, user has to use JOIN statements which creates Views. This is a time consuming process. On the other hand, in NoSQL, data is stored in form of objects which will contain all the related data. This eliminates process of combining and then displaying the data, hence saving response time.
* *System Maturity:* SQL is an experienced technology and hence most of the issues have taken care of. Security features like authentication, data confidentiality & integrity are incorporated in SQL. On other hand, such security features are yet to be addressed in NoSQL. NoSQL might generate more issues or security breaches because of lack of maturity of the system. In some IoT applications, a secured communication channel is required to transfer sensitive data. For such applications, it is better to use a

secured data storage system along with secured communication channel.

CONCLUSIONS

Performance of MySQL and MongoDB databases is compared for the data generated by an Internet of Things application. MySQL which is a traditional SQL database and MongoDB which is a NoSQL database are used to store the IoT data. The comparison study is based on the time taken to execute Select and Insert queries against varying number of records and threads. As, the number of records / size of the table increases, load on the system increases, further increasing the latency or response time. Each database has its own pros and cons. From the study, it was observed that in some scenario, MongoDB required less response time compared to MySQL. But, MySQL responses were stable as compared to MongoDB. Therefore, choosing a better database for IoT depends on which query is mostly used and the requirements of application. Future research run for this work include investigation performed with several other database management systems available with advanced load testing tools. Research can be further extended by using cloud, distributed databases which will make the work more realistic.