

GRAPH DATABASE APPLICATION FOR REAL-TIME SENSOR SYSTEMS

Eng. Georgi Georgiev, PhD Student

Department of Computer Systems & Technologies University of Ruse

Tel.: 0035989-883-1934

E-mail: ggeorgiev@uni-ruse.bg

Abstract: Every day more and more devices are connected to the Internet simultaneously. The ability to connect and exchange information over many devices is essential nowadays. Constantly the data transferred over devices are heterogeneous and real-time. Many devices transmit data over the internet in vast quantities. Devices such as cars operate with live sensor data, GPS coordinates and many more. The collected data must be processed in a fast manner and represented to users in a more natural form. Graph databases remain a viable alternative over the old database systems. The dominance of Graph databases in real-time applications forces developers to operate with such systems due to the fact that graphs have a high expressive power to process complicated structures of sensor data. In this paper, we study a graph database modelling for storing live sensor readings and offer a new way to process real-time data. The main objective of the survey is to provide an architectural approach for storage, representation and processing of large amounts of live streaming data.

Keywords: Graph Database, real-time, sensors, data storage, GPS coordinates

ВЪВЕДЕНИЕ

Базите от данни са важна и неразделна част от едно приложение, независимо от характера и сферата на реализация. Съществуват много и най-разнообразни видове бази от данни. Те се разделят на няколко основни типа според спецификата на работа и използването на определен вид данни. За различните видове данни съществуват различни решения за запис на данни, (Pereira, D., Morais, W., & Freitas, E., 2017) описват и сравняват основните типове бази от данни и тяхното характерно приложение според вида на данните, които използват. Данните са в основата на една система. Според вида на данните строго зависи и подбора на компонентите, изграждащи една система: избора на хардуерните елементи, безжичната комуникация, приложението, сървърната част и не на последно място базите от данни.

Данните, предавани в реално време са често подценяван ресурс, но правилната манипулация с подобен вид данни води до ефективни решения в системите с разпределена обработка на информация. Те са хетерогенни, предават се в реално време, големи по обем и често се предават с висок интензитет.

Сензорите ни предоставят важна информация за света около нас. Благодарение на данните от тях можем да следим метеорологичните условия, разстояние между обекти, местоположение на определени обекти и много други. Те намират приложение във всички сфери пряко свързани с обзор и анализ на данни.

ИЗЛОЖЕНИЕ

Обзор на съществуващите решения

В статията си (Zappi, P., et al., 2008), авторите прилагат кратък обзор на възможностите на подобни системи. Базите от данни, базирани графов модел на практика са приложими за всякакви сфери на индустрията.

Патентът приложен от (Aasman, J., 2011) представя методите за реализиране на разширяема система на базата на графов модел. Авторът добре сравнява и описва основните методи за разработване на подобна система.

Колегиумът разработил системата (Hilsenbeck, S., et al., 2014) демонстрира как използването на сензори от общ характер позволяват изследването на нови територии. Употребата на мобилни телефони и средства за безжична комуникация осъществяват точното позициониране на обекти в закрити пространства.

Системата, разработена от (Isikdaga, U., Piloukb, M., & Fine, M., 2016), позволява използването на микрокомпютърен набор от устройства и сензори за отмерването и изпращането на данни в реално време на отдалечен сървър. Географските данни, получени от микрокомпютърната система биват записани за обработка и анализ на по-късен етап. Поради спецификата на данните съхранението им в бази от данни, базирани на графовия модел е от съществено значение.

Системите, описани по-горе, демонстрират как употребата на правилния вид бази от данни играе важна роля в производителността на една система, нейната мащабируемост и обхват на дейност. Базите от данни базирани на графовия модел намират широко приложение във всякакви предметни области (Miller, J., (2013). Употребата им помага за премахването на редица проблеми пред разработчиците. Обединяването на положителните качества от двата основни типа бази от данни (релационните и нерелационните) води до революционни промени в света на базите от данни.

Бази от данни базирани на графов модел

Релационните бази от данни обикновено записват структуриран тип данни в редове и колони на таблици. Таблиците и връзките между тях са предварително дефинирани. Промяната на таблиците и връзките в една работеща система за управление на бази от данни е трудна и скъпа за изпълнение манипулация. Работата на една релационна система се основава на ACID свойствата (Atomicity: Атомарност, Consistency: Постоянност, Isolation: Изолираност, Durability: Стабилност).

Нерелационните бази от данни от друга страна работят с неструктурирани данни. При тях схемите се променят динамично, при някои бази от данни дори автономно. Те работят в разпределен режим и връзките между данните са лесни за създаване и управление. Нерелационните бази от данни се основават на BASE свойствата (Basic Availability: Постоянна наличност, Soft State: Неустановено състояние, Eventual Consistency: Евентуална съгласуваност).

Базите от данни базирани на графов модел (Angles, R., & Gutierrez, C., 2005) обединяват двата основни типа бази от данни. Те едновременно изпълняват ACID свойствата и имат свободата на нерелационните бази от данни. Графовия модел значително разширява релационните бази от данни като представя информацията под формата на граф, където данните представляват възли и листа, а връзките между тях представляват релациите. Базите от данни, базирани на графовия модел, позволяват моделирането на данните да се извършва визуално, което е близко до представянето на данните в реалния свят. Представянето на данните в граф се свежда до набор от обекти, наречени още възли или върхове (листа), които са свързани помежду с помощта на линии, като линиите представляват връзките между възлите.

Базите от данни базирани на графовия модел (Williams, D., Huan, J., & Wang, W., 2007) се използват в широк спектър от приложения категоризирани в два основни типа. Първият е базиран на системи, основани на знанието, които използват голям набор от колекции от класове. Използват се главно за приложения, свързани с биологичните и живите науки. Втория тип е базиран на приложения, свързани с търсенето на данни в транзакциите. Използват се за приложения в областта на телекомуникациите, социалните услуги, защиты против измами, телеметрията и за обработка на сензорни данни.

Графовия модел на данните е предпочитано решение за запис на свързани данни. В един граф връзките между възлите и листата са добре описани и съществуват редица вътрешни алгоритми за търсене в граф. Поради тази специфика независимо от големината на графа откриването на дадена информация в графа става моментално (Cheng, J., et al., 2008). Употребата на специализирания език за заявки- Cypher и запазването на поддръжката на ACID свойствата, което води до сигурност на данните измества базите от данни, базирани на графов модел в челните позиции. Сравнителната характеристика на базите от данни е представена на таблица 1.

Таблица 1. Сравнителна характеристика между основните видове бази от данни

	Бази от данни базирани на графов модел (Graph DB)	Релационни бази от данни (Relational DB)	Нерелационни бази от данни (NoSQL DB)
Съхранение на данните	Структура на запис на данните базирана на графов модел	Фиксирани, предварително дефинирани таблици с редове и колони	Свързани данни от хетерогенен тип, неподходящи за запис в таблици
Модел на данните	Гъвкав модел на данните	Фиксиран модел на данните произлизащ от логическия модел	Неподходящ за корпоративни архитектури
Производителност при заявките	Страхотна производителност независимо от броя на заявките и дълбочината на търсенето	Производителността намалява при увеличаването на операциите за обединение	Връзките трябва да се създадат в приложния слой
Език на заявките	Cypher: интуитивен език лесен за разбиране	SQL: сложността на заявката се увеличава с броя на операциите за обединение	Използват се различни езици за програмиране, но не са приспособени за заявки
Транзакции	Запазва ACID свойствата	Основава се на ACID свойствата	Основава се на BASE свойствата
Разширяемост	Висока разширяемост и сигурност на данните	Разширяване чрез реплики, което коства доста ресурси	Висока разширяемост но няма сигурност на данните

Реализиране на системата

Реализирането на системата се осъществява с помощта на предварително изграден микрокомпютърен модул, блок схемата на който е представена на фигура 1. С помощта на модула можем да запишем сензорни данни и да ги предадем в реално време до отдалечен сървър. Модулът е инсталиран в реален автомобил, като записва информация както за заобикалящата го среда и състояние на автомобила, така и за неговото текущо местоположение и скорост. За целите на изследването функционалността на модула е сведена до минимум, тъй като имаме за цел за обработване само данни свързани със скоростта, местоположението и температурата на околната среда. Щом данните са записани, микропроцесора ги преформатира и изпраща на сървъра.

За комуникационен път е избран GSM модул с интернет свързаност. Чрез него данните биват изпратени до отдалечен сървър в обикновен XML документ. Примерен XML документ е представен на фигура 2. След като документът е получен от сървъра данните биват обработени и записани. Отчитането на времето за запис на извършените манипулации в бази от данни базирани на графов модел е главна задача на проекта.



Фиг. 1. Блок схема на микропроцесорната система

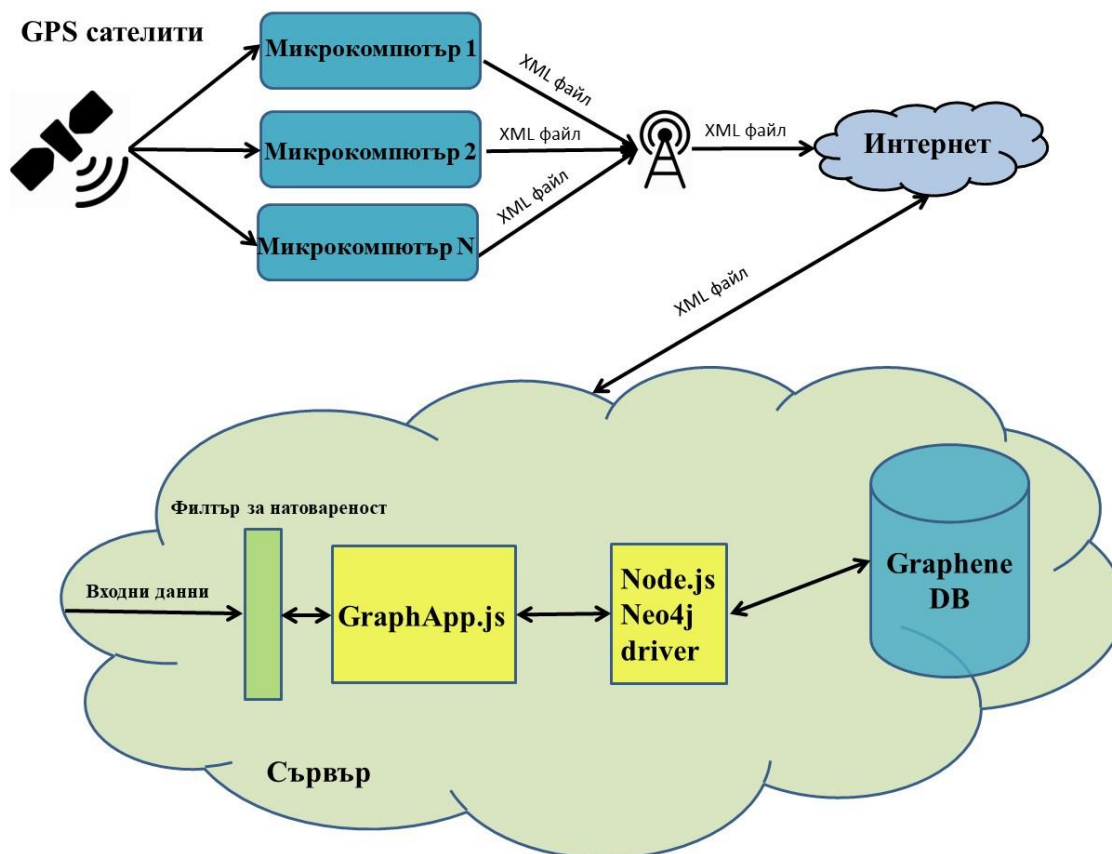
```

<?xml version= "1.0"?>
<xml>
  <id>
    0785
  </id>
  <sensors>
    <temperature>
      22
    </temperature>
    <latitude>
      43.849949
    </latitude>
    <longitude>
      25.954359
    </longitude>
    <speed>
      54
    </speed>
  </sensors>
</xml>

```

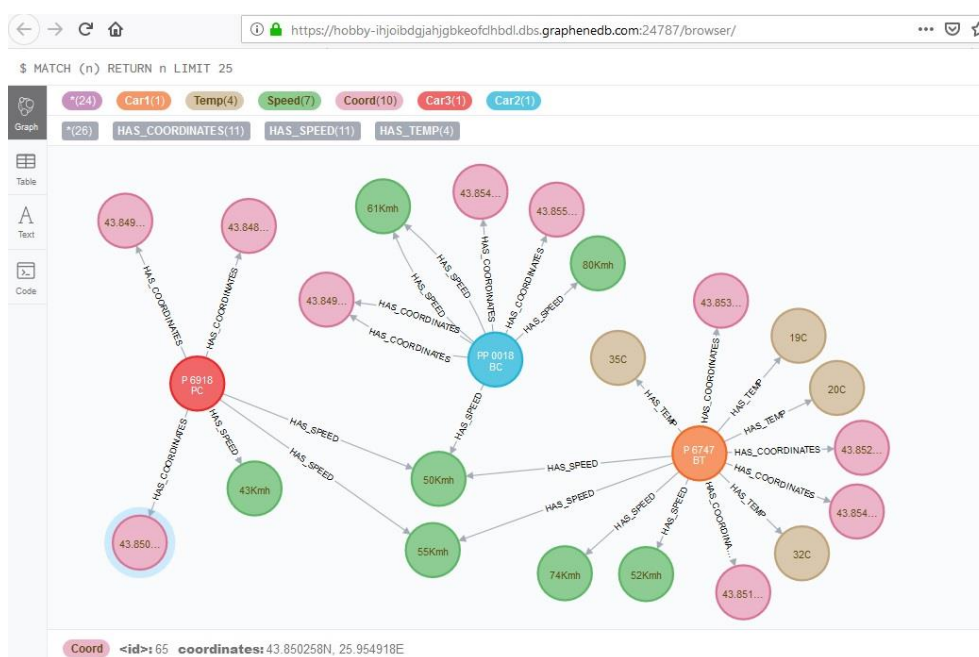
Фиг. 2. XML документ

Друга основна част на реализираната система е сървърната част. Тя е съставена от следните компоненти: Сървърната част е реализирана с помощта на Node.js платформа, GrapheneDB база от данни и Neo4j Bolt driver (Holzschuher, F., & Peinl, R., 2013). Сървърът е качен в облачното пространство на Heroku (повече на: <https://www.heroku.com>). Така реализираната система е достъпна от всяка точка на света. С помощта на изграждащите компоненти на сървърната част е възможно лесното манипулиране на входящите данни, благодарение на използването на код с отворен достъп. На фигура 3 е представена архитектурата на изградената система.



Фиг. 3. Архитектура на реализираната система

Реализираната система може да приема данни едновременно от няколко устройства. Както е показано на фигура 3 потока от данни може да постъпи от много места. За това данните първо се сортират за излишна информация и обработват. Анализът на данни се основава само на постъпилите данни от мобилните устройства. На този етап от разработка са реализирани само две микропроцесорни устройства предаващи данни, като за в бъдеще могат да бъдат добавени и още. На фигура 4 е представен графовия модел на данните.



Фиг. 4. Графов модел на данните (изграден чрез Neo4j)

Тестови резултати

Тестовите за отчитане на времето за запис на сензорни данни бяха проведени в различни условия. Микропроцесорните устройства бяха поставени в различни автомобили и бяха проведени редица тестове. Изпълнението на командите за запис бяха изпълнени на извън населени места, както и в градски условия. Повечето различни сценарии бяха изпълнени над 1000 пъти, като бе взето под внимание и натовареността на мобилната мрежа. Тестовите бяха проведени в продължение на няколко дена в различно време от денонощието. Средните стойности от проведените резултати от изследванията са изобразени под формата на диаграма представени на фигура 5.



Фиг. 5. Стойности на отчетените резултати за операция запис.

ИЗВОДИ

Изградената система доказва, че базите от данни, базирани на графов модел, са подходящ избор за запис на сензорни данни, предавани в реално време. Графовия модел отлично описва данните във вид удобен за разбиране. Силата на ACID свойствата в тази нерелационна база от данни добавя значителна сигурност на данните. Проведените практически изследвания заключват, че микропроцесорните устройства успешно предават информация до сървъра и може да се разчита на тяхната надеждност. Сървърната част отлично обработва данните, въпреки реализирането на системата в облачно пространство с отворен достъп. Записаните данни отчитат минимално време за запис на информация в базите от данни. Като недостатък на системата се отчита технологичното време за началното свързване на микропроцесорното устройство и времето от порядъка на 200 милисекунди породено от силата на сигнала (GRPS) и натовареността на мрежата.

REFERENCES

- Aasman, J. (2011). Method for creating a scalable graph database. URL: <https://patentswarm.com/patents/US7890518B2> (Accessed on 16.09.2019).
- Angles, R., & Gutierrez, C. (2005). *Querying RDF Data from a Graph Database Perspective*. Paper presented at the European Semantic Web Conference, 2005, Berlin, 346-360.
- Cheng, J., Ding, Y., Yu, B., & Wang, H. (2008). *Fast Graph Pattern Matching*. Paper presented at the IEEE 24th International Conference on Data Engineering, 7th-12th April 2008, Cancun.
- Hilsenbeck, S., Bobkov, D., Schroth, G., Huitl, R., & Steinbach, E. (2014). *Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning*. Paper presented at the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 13th-17th September 2014, Seattle.

Holzschuher, F., & Peinl, R. (2013). *Performance of graph query languages: comparison of cypher, gremlin and native access in Neo4j*. Paper presented at the EDBT '13 Proceedings of the Joint EDBT/ICDT 2013 Workshops, 18th- 19th March 2013, Genoa, 195-204.

Isikdaga, U., Piloukb, M., & Fine, M. (2016). *Integration of Geo-Sensor Feeds and Event Consumer Services for Real-Time Representation of Iot Nodes*. Paper presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 12th-19th July 2016, Prague.

Miller, J. (2013). *Graph Database Applications and Concepts with Neo4j*. Paper presented at the Proceedings of the Southern Association for Information Systems Conference, 23rd-24th March 2013, Atlanta.

Pereira, D., Morais, W., & Freitas, E. (2017). NoSQL real-time database performance comparison, *International Journal of Parallel, Emergent and Distributed Systems*, 33(2).

Williams, D., Huan, J., & Wang, W. (2007). *Graph Database Indexing Using Structured Graph Decomposition*. Paper presented at the IEEE 23rd International Conference on Data Engineering, 15th-20th April 2007, Istanbul.

Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., & Tröster, G. (2008). *Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection*. Paper presented at the 5th European Conference, EWSN 2008, 28th-30th January 2008, Italy.