КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2015 Т. 7 № 3 С. 493−498

Ки&М

СЕКЦИОННЫЕ ДОКЛАДЫ

УДК: 004.65

Storage database in cloud processing

A. V. Bogdanov[a], Thurein Kyaw Lwin[b]

Saint Petersburg State University, University ave. 35, Peterhof, St. Petersburg, 198504, Russia

E-mail: [a] bogdanov@csa.ru, [b] trkl.mm@mail.ru

Received October 27, 2014

Storage is the essential and expensive part of cloud computation both from the point of view of network requirements and data access organization. So the choice of storage architecture can be crucial for any application. In this article we can look at the types of cloud architectures for data processing and data storage based on the proven technology of enterprise storage. The advantage of cloud computing is the ability to virtualize and share resources among different applications for better server utilization. We are discussing and evaluating distributed data processing, database architectures for cloud computing and database query in the local network and for real time conditions.

Keywords: Storage database, cloud processing, storage architecture, cloud architecture, data processing

Хранилища баз данных в обработке в облаке

А. В. Богданов, Тхурейн Киав Лвин

[1] *Санкт-Петербургский государственный университет,*
Россия, 198504, г. Санкт-Петербург, Петергоф, Университетский просп., д. 35

Хранение — это существенная и дорогая часть облачных вычислений как с точки зрения требований сети, так и организации доступа к данным, поэтому выбор архитектуры хранения может быть критическим для любого приложения. В этой работе мы сможем посмотреть на типы облачных архитектур для обработки и хранения данных, основанных на доказанной технологии хранения в сети масштаба предприятия. Преимущество облачных вычислений — это способность визуализировать и разделять ресурсы среди различных приложений для наилучшего использования сервера. Мы обсуждаем и оцениваем распределенную обработку данных, архитектуры баз данных для облачных вычислений и очередь баз данных в локальной сети и для условий реального времени.

Ключевые слова: Хранилища баз данных, обработка в облаке, архитектуры хранения, облачная архитектура, обработка и хранение данных

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 3, pp. 493–498.

Introduction

The cloud computing and cloud data stores have been a precursor and facilitator to the emergence of big data. Cloud computing is the commoditization of computing time and data storage by means of standardized technologies. Some cloud services provide consumers with space for storage and use of data for free, others charge a particular payment for services provided to subscribers. There are also private clouds, which are owned and operated by organization. In fact, it is protected network for storing and sharing critical data and programs. To create a private cloud requires hardware, software and other tools from different vendors and managing of the physical servers with the external and internal layers. Hybrid cloud, as is clear from the title, combined resources of different public and private clouds in a single service or a decision [Technical Details…]. The basis of all cloud services, products and solutions are software tools that functionality can be divided into three types, means for processing data and running applications (computing servers) to move data (network) and for storage (SAN).

The structure of Cloud and capabilities

Cloud computing processing and storage quickly gained popularity because they are not only provide a solution to the most complex and pervasive problems in the IT sector, but also open up a number of new features. In some environments, these technologies help to reduce costs and often necessary to expand the range of goals and objectives or services ensure compliance, reaching the necessary indicators of availability, performance, security and data protection [Armbrust, 2010].

Cloud Solutions

Cloud Solutions means to create and store content, as well as strategies that determine where and how the content is used. These solutions are used to create virtual infrastructures, organizations of all sizes which can accommodate the necessary applications and tools, as well as development environments and test new features [Technical Details…].

Some of common terms and phrases that characterize cloud solutions are.

• Optimized and cost-effective: the expansion of the range of services and ensure an appropriate level of service with the resources available

• Ability to create a variety of service options: resource allocation levels for different budgets and requirements for quality of service

• Flexible, scalable continuously: the possibility of expanding without complications

• Reliable, flexible and dynamic: adaptation to the ever-changing needs and availability

• Fast or automatic allocation of resources: quick access to resources and services

• Secure and supporting multi-client architecture: a safe separation of user data integrity

• Measure and manage: to provide metrics for reporting, analysis and management of services

• Scaling by increasing density: taking advantage of multi-client architecture to reduce costs

Distributed data processing

Distributed data processing is an opportunity to integrate fragmented data resources. One approach to centralizing data is to simply decommission existing database systems and to build a new integrated database. An alternative approach is to build an integration layer on top of pre-existing systems [Bobrovsky, 2013]. Building an integration layer on top of existing database systems is a challenge in complexity and performance, but this option sometimes makes the most business and engineering sense. In a data-sharing environment, there is no single best architecture that will solve all problems. Large installations of database systems may be accessed by hundreds of thousands of times a minute. The irreducible latency present even in a fully optical network is not capable of supporting such a performance requirement. Indeed, local disks are also too slow, and most of this sort of information is cached off disk and into memory. In some organizations, if critical data is unavailable for even a matter of minutes, it could affect millions of dollars of revenue. This is why remote data access is not used in such large-scale situations where high availability is critical [Nikulchev, 2008]. There are many small and medium weight applications with modest performance requirements for data. Often, such applications are designed to work with a copy of data because getting a copy and loading it on a local database seems like the easiest solution. Such design does not factor in the cost of maintaining a separate copy of the data. When the applications are put into production and begin having problems keeping their data in sync, these costs become all too apparent. Such applications would probably do better to remotely reference their data. In such cases, it is a good architecture to remotely reference application databases for shared data. Such “distributed” databases need to incorporate some high availability design, depending on the weight of the applications served and their availability requirements. Each application should be analyzed to determine its performance and reliability requirements.

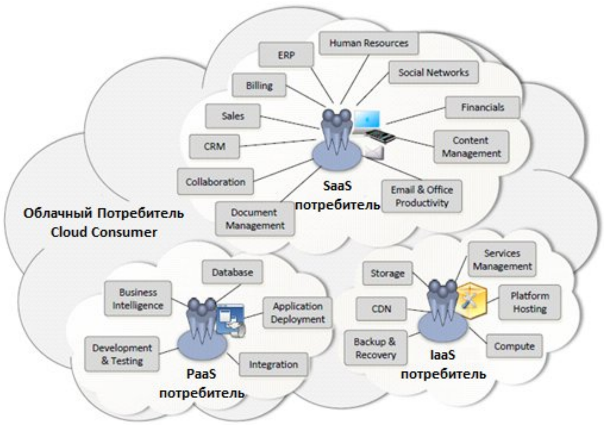

Fig. 1. Structure of cloud services

To choose a proper solution one must carefully look at existing tools and make a proper choice both of instruments and distributed system architecture.

DBMS application as a service

DBMS application is a layer between models DbaaS, SaaS, PaaS and cloud database. As for the DbaaS, it is a managed service in the mode of payment for use and providing access to a database of external programs [Babu, 2009]. Actually cloud DBMS scales almost to the size of PaaS and DbaaS volume and manages the database and the number of client connections to the database. Unlike the classical DbaaS, PaaS model is fairly obvious and clear. PaaS provides a ready hosting, where there is pre tuned and working database, but the interaction with the developed software it is necessary to organize manually, through the local interfaces, management and updating the system needs to be done separately, and payment is taken monthly, for hosting as a whole (this is one of the fundamental differences). When this is not excluded, than the servicing of DbaaS PaaS is implemented via predevelopment tools. The large uncertainties remain only in the ratio of the cloud model with DbaaS. First, the difference is mainly on the technical details of the implementation by the service provider (and the market is divided in the same way), but the end user in most cases can not apply to a single universal concept of "cloud DBMS."

What is the difference of Cloud DBMS and DbaaS

Cloud SMS — is a fully automated multi-user and unlimited scalable service that provides database functionality, but operated and administered "unnoticed" by the service provider [Pluzhnik, Nikulchev]. It should not be confused cloud DBMS and database running on a virtual machine. Cloud model provides high flexibility and scalability of service, quick access to programming interfaces and settings. The user can connect to the system at any moment — for example, one hour — set the desired settings to load data to generate queries and get results without worrying about versions DBMS, its administration and configuration. Database as a Service — providing a simple but functional profile of saturated solutions "database in the cloud" for the needs of medium and small businesses and IT departments of large corporations. It usually does not occur directly in the provider's own data center, and functions as an add-on classical cloud services [Nikulchev, 2008]. Almost always specific DbaaS is one particular database provided in the cloud directly to the developer. Typical examples: Caspio, EnterpriseDB, Heroku Postgres, Xeround [Rittinghouse, 2010]. It is quite evident and displays the difference in business models: cloud DBMS suitable is for large-scale standard tasks and DbaaS is such for specialized application, using a particular brand of database engine, with the possibility of direct communication with its developers [Haak, 2011]. Furthermore, DbaaS allows much more accurate system to pick a right load, in particular, by controlling the amount of client connections.

Analysis of the Database and Hybrid Local Time Condition

We analyze the local database, build over a relational database MS SQL Server and occupied almost 26683.48 MB of memory on the database server. In the hybrid database information about the project and clients are moved from the local database to MS SQL Server in the cloud storage. In this process the memory on the database server of about 46.12 MB, and in the cloud server about 25 GB were used, thus producing the testing of queries. The results of two experiments are shown in Figure 2.

Table 1. The experiment result table

Record of Seconds		The number of records in a query	Time Extraction		Average query time		The number of all Project
Local	Hybrid		Local	Hybrid	Local	Hybrid	
0,534101	0,523230	1	177,533	183,334	1,7975	1,833	100
0,232501	0,220650	2	406,403	411,739	4,0862	4,117	200
0,140102	0,157200	3	621,604	627,270	6,238	6,272	300
0,101307	0,110017	4	880,467	883,8712	8,8168	8,838	400
0,090013	0,090732	5	1075,044	1099,451	10,960	10,994	500
0,073105	0,073070	6	1305,506	1329,938	13,055	13,077	600
0,060137	0,060032	7	1600,401	1604,110	16,014	16,041	700
0,050722	0,050643	8	1940,710	1943,777	19,407	19,437	800
0,04155	0,041502	9	2262,453	2265,402	22,624	22,653	900
0,03565	0,037819	10	2620,356	2622,060	26,203	26,220	1000

In Fig. 2. it is clearly seen, that for large databases with complex inquiries for the semi structured data there is no substantial loss of the time for data transmission in this experiment [Shokin, 2010]. Cloud storage is effectively a boundless data tank. It is important for effective performance in computations that while many solutions scale horizontally, when data is copied in parallel by cluster or parallel computing processes the throughput scales linear with the number of nodes for reading or writing. This includes products and solutions, that are used to deploy public, private and hybrid clouds.

Conclusion

We summarized above the database requirements for cloud databases and compared the suitability of database architectures for cloud computing. We also tested the queries in the local server and in the cloud, used for the construction of information systems and we can make conclusions about its

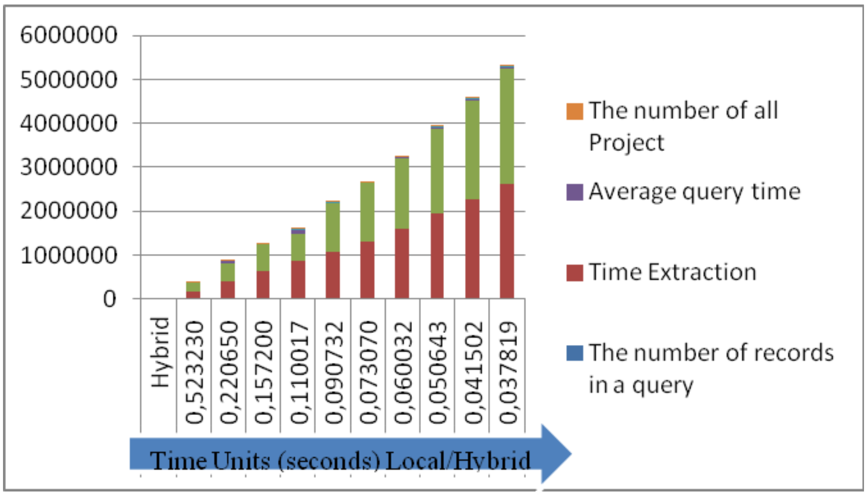

Fig. 2. The experiment results Graph

effectiveness with technology, developed for semi structured data. The steps for organization of such a system are:

1. Estimation of total system parameters (maximum number of users for simultaneous operation, the ability to scale services, the availability of personalized access).

2. Validation of the project (having own server capacity, cost comparison with the cost of launching of rental services).

3. Evaluation the time for data access, query performance evaluation for cloud infrastructures.

4. Constructing of automatic allocation system and sending requests in a distributed database.

To solve the first stage uses multi-criteria and decision-making methods, the second stage is realized on the basis of economic- mathematical methods of evaluation; third and fourth - based on optimization techniques and effective evaluation of search queries. Cloud storage systems are the easiest and most affordable way to solving the problem of highly loaded database. Easily run into work, they will not require re-deploy the application or any modification in the configuration files. The end goal of Database-as-a-Service is to automate database administration tasks and resolve its issues and challenges so that the end user doesn't have to think about anything on the database level. And just like the other "as-a-Service" models, it should also be on demand and easily deployable model. The database is requirements for the cloud databases and compares the suitability of different database architectures to cloud computing. Based on our performance results, we believe that the Cloud Database vision can be made a reality, and we look forward to demonstrating an integrated prototype of next Big Data Solution. Whether we need to assembling, managing or developing on a cloud computing platform and need a cloud-compatible database.

References

Armbrust M. A view of cloud computing / M. Armbrust, A. Fox, R. Griffith etc. // Communications of the ACM. — 2010. — Vol. 53, No. 4. — P. 50–58.

Babu S. Automated control in cloud computing: challenges and opportunities / S. Babu, J. Chase, S. Parekh // 1st workshop on Automated control for datacenters and clouds. — 2009. — P. 13–18 (DOI: 10.1145/1555271.1555275).

Beloglazov A. Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints / A. Beloglazov, R. Buyya // IEEE Transactions on Parallel and Distributed Systems. — 2013. — Vol. 24, No. 7. — P. 1366–1379.

Bobrovsky S. MS. Week / Re № 30-31 (850-851) 19 2013 Noryabrya.

Haak S. Autonomic benchmarking for cloud infrastructures: an economic optimization model / S. Haak, M. Menzel // 1st ACM / IEEE Workshop on Autonomic Computing for Economics. — 2011. — P. 27–32 (DOI: 10.1145/1998561.1998569).

Nikulchev E. V. Construction of the model download channels in data networks based on geometric approach / E.V. Nikulchev, S.V. Payain // News of higher educational institutions. Problems printing and publishing. — 2008. — № 6. — S. 91–95.

Pluzhnik E. V., Nikulchev E. V. // Semistructured DATABASE hybrid cloud infrastructure.

Rittinghouse J. W. Cloud computing-implementation, management, and security / J.W. Rittinghouse, Ransome J.F. — NY: Taylor and Francis Group, 2010.

Shokin Y. I. Technology development of software systems information support of scientific activities, dealing with semistructured documents / Y. Shokin, A. M. Fedotov, V. B. Barakhnin // Computational technologies. — 2010. — T. 15, No. 6. — S. 111–125.

Technical Details Architecture cloud data processing and storage. http://www.seagate.com/ru/ru/tech-insights/cloud-compute-and-cloud-storage-architecture-master-ti/