Research Journal of Applied Sciences, Engineering and Technology 6(6): 1040-1047, 2013
DOI:10.19026/rjaset.6.4010
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 22, 2012 Accepted: November 23, 2012 Published: June 30, 2013

Research Article

Research on Performance Evaluation of Biological Database based on Layered Queuing Network Model under the Cloud Computing Environment

Zhengbin Luo and Dongmei Sun
Department of Computer and Information Technology, Beijing Jiaotong University,
Beijing 100044, China

Abstract: To evaluate the performance of biological database based on layered queuing network model and under cloud computing environment is a premise, as well as an important step for biological database optimization. Based on predecessors' researches concerning computer software and hardware performance evaluation under cloud environment, the study has further constructed a model system to evaluate the performance of biological database based on layered queuing network model and under cloud environment. Moreover, traditional layered queuing network model is also optimized and upgraded in this process. After having constructed the performance evaluation system, the study applies laboratory experiment method to test the validity of the constructed performance model. Shown by the test result, this model is effective in evaluating the performance of biological system under cloud environment and the predicted result is quite close to the tested result. This has demonstrated the validity of the model in evaluating the performance of biological database.

Keywords: Biological database system, cloud computing, layered queuing network model, performance evaluation

INTRODUCTION

With the development of public resource service technology, cloud computing plays an important role in the development of computer network service (Daniele and Enrico, 2012). As a new technology integrating all effective resources to provide users with data service, on the one hand, cloud computing solves the problem of data resource redundancy; on the other hand, it has improve the efficiency of data service via cluster optimization (Mezmaz *et al*., 2011). No matter private cloud serving local area network or public cloud serving whole network users, the most significant similarity and nature of cloud computing lies in its strong cluster integration. By integrating data resources deployed in different computers and different storage equipments, cloud computing is able to form a data resource pool, so as to provide data service according to users' request. Providing on demand means cloud computing has to break through the bottle neck of low efficiency in traditional data service (Niroshinie *et al*., 2013). However, users' demand and desire are unlimited. On the contrary, the capability of underlying data system for supporting cloud computing is limited. As for this, how to effectively improve the service efficiency and response speed of biological database system under cloud environment has become a research hotspot at present (Tanzim *et al*., 2012). In order to solve this problem, an important step is to evaluate the performance of currently existed or designed biological database system. So far, with regard to the problem, the most frequently applied method in the world is to construct a biological database efficiency evaluation model based on a certain mature theoretical system and under cloud environment (Subhas and Arka, 2011). On this basis, we will be able to further optimize the structure and layout of biological database through efficiency evaluation, so as to further perfect the core algorithm of biological database service and to achieve the goal of improving system service efficiency. Among all these mature theoretical systems, the most widely applied method is queuing network model or layered queuing network model. By investigating computer hardware resources (CPU and storage device), Bao-Chyuan *et al*. (1988) constructed a distributed and layered-queuing-based biological database platform. On this basis, under the experimental conditions, they tested the validity of queuing network model in evaluating biological database efficiency. Zhao Jianguang systematically studied the construction and solving method for layered queuing network model and had constructed a biological database system under cloud environment. In addition, through experiment testing, comparison and analysis, the validity of layered queuing network model in biological database system performance evaluation was also explored and discussed. Drawing support from layered queuing network model, Bian Xuegong *et al*.

Corresponding Author: Zhengbin Luo, Department of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

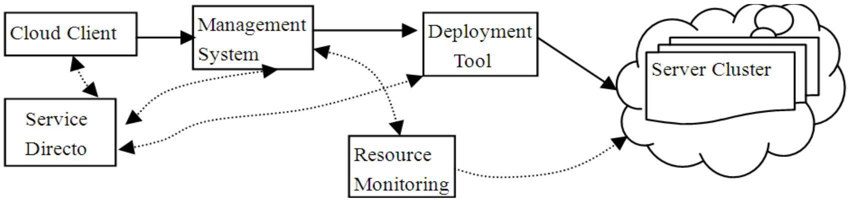

Fig. 1: Cloud computing structure

(2004) predicted the performance of MCU. Based on the performance prediction, optimization of MCU system was research, with very good effect achieved. Based on existing research achievements domestic and overseas and by using for reference traditional layered queuing network model, construction and solving of this model are optimized in this study. Furthermore, under laboratory conditions, the Hbase cloud biological database under cloud environment is established (Vijindra and Sudhir, 2012). By evaluating and predicting the performance of the biological database with optimized layered queuing network model, we may see that, the predicted result has slight difference with the tested result, showing that the model is effective in evaluating the performance of biological database system.

CLOUD COMPUTING AND PERFORMANCE EVALUATION

Cloud computing: Cloud computing refers to a computing model offering cloud service (Andrzej and Michael, 2010). The so-called cloud service functions in this way: service providers combine and pool resources according to certain protocols and rules. As for this, users will easily get their demanded network service, without having to install, upgrade or maintain their own hardware and software. According to actual service mode, cloud computing can be divided as infrastructure cloud, public cloud and private cloud (Aslı, 2011).

The creation of cloud computing liberates computers from the restrictions of hardware, software and data, so that people may share cloud resources for better configuration and more liberty via the network at any place. With a powerful cloud network, cloud computing integrates large amounts of parallel software and hardware resources in the network, so as to offer service to network clients (Aslı, 2011). Normally, its system structure is comprised by cloud client, management system, deployment tool, service directory and resource monitoring. Service directory is a service resource list provided by the server to clients, so that users may request resources in the server cluster according to the directory. Resource monitoring is designed to supervise resource flow and port access in the serve cluster, so as to better distribute service thread. Universal cloud computing structure is shown in the following Fig. 1.

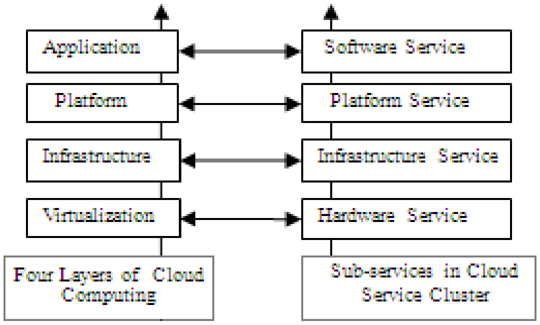

Fig. 2: Cloud computing service layers

According to the aforementioned system structure, cloud computing can generalized into four layers, from bottom to top, separately to be: virtualization layer, infrastructure layer, platform layer and application layer (Wei *et al.*, 2011). The four layers separately correspond to a sub-service. The layers of cloud computing service are shown in Fig. 2.

In the above structure of cloud computing, service is the core concept. It breaks through traditional structural concept of hardware and software facilities, regarding all sharable hardware and software as service. As for this, cloud provider becomes hardware, software and resource provider (Nabil, 2010). Thus, in cloud computing, the kernel is not client deployment, but server resource pool deployment and planning. Biological database is the core to organize, distribute and manage hardware, software and data in the resource pool. A well-structured cloud solution shall have perfect and efficient data system.

Cloud computing: Performance usually refers to the capability of software and hardware and the so-called machine performance often refers to the speed of device, i.e., countdown of program execution time (Chien-Yuan *et al.*, 1995). What's more, program execution time refers to the time from initiating a task until obtaining the required result, including disk and storage access, CPU running, I/O action and operating system operation. However, in multi-task systems, when waiting for I/O actions, CPU may turn to process another task. As for this, the analysis becomes quite complicated. For this reason, when discussing performance, we may as well use CPU time to represent CPU working time, excluding waiting time and time for executing other tasks (Rasha and William, 2012). Obviously, the execution time noticed by users is total time for executing programs and is more than CPU time. Performance evaluation is designed to test a machine's indexes related with performance. The frequently used methods include constructing model, filtering situation, situation analysis, sensitivity analysis, uncertainty analysis, consequence analysis, model test, evaluation result, etc.

The so-called performance evaluation mainly performs test with time as the measurement. Normally, we assume that the clock period of machine is TC, the total number of commands in the program is IN and the average clock period for executing each command is CPI, so that the time of a program running in CPU, i.e., TCPU shall be:

$$T_{CPU} = I_N \times CPI \times T_C \tag{1}$$

in the function:

$$CPI = \frac{\text{Clock Period for Executing the Whole Program}}{\text{Total Number of Command in the Program}} = \frac{\sum_{i=1}^{n}(CPI_i \times I_i)}{I_N} \tag{2}$$

In the above formula, n stands for the number of command types in the program. Assuming that I_i/I_n represents the proportion of the i^{th} command in the program, so that the above formula can be transformed as:

$$CPI = \sum_{i=1}^{n}\left(CPI_i \times \frac{I_i}{I_N}\right) \tag{3}$$

As for biological database performance evaluation under cloud environment, the most frequently used method to compute CPU running time is evaluation measurement and evaluation model. Measurement takes long time and higher consumption. By contract, in modeling method, an effective model should be established firstly according to the nature of the system to be evaluated, so as to figure out the performance indexes of the model (Wei *et al*., 2002). On this basis, the system will be evaluated according to the performance indexes. Currently, the universally used and effective performance model is queuing network model. In this study, this model will be applied to evaluate the performance of cloud biological database.

Cloud computing queuing network model: Queuing network model is designed to evaluate the performance of the system under the situation that multiple users share a single resource at the same time, with network request competition aroused (Wong and Lam, 1982). Its major research objects are service center and users. Users communicate with service center via request, while service center establishes a one-to-many logic relation with users. The number of users may be fixed or uncertain. Thus, queuing network model is comprised by two modes: opened and closed. Opened queuing model is mainly applied to evaluate the performance of transaction-based network load, in which, user data changes all the time (Wong and Lam, 1982). By contrast, close network model is often applied in client or batch processing load, in which, the number of users is uncertain. The two parties take normative parameters as indexes to evaluate system performance. Major parameters are shown in Table 1.

Table 1: Queuing network model parameters

Load type	Input parameter	Output parameter
Transaction	λ	R,X
Batch processing	N	R,X
Client	T	R,X

In the table, λ is arrival rate, N is the average number of users' requesting service in the system, T is clients' reflection time, R is system response speed, X is system throughput and U refers to system resource use rate. On this basis, the following correlation is established:

$$\begin{aligned} N &= X * D \\ U &= X * R \\ R &= \frac{N}{X} - Z \end{aligned} \tag{4}$$

Layered queuing network model is a typical mode of queuing network model, as well as a major model form studied in this study. Layered queuing network model can be regarded comprised by multiple tasks, while each task contains an entrance, a service queue, as well as a multi-thread parallel structure. Moreover, entrance and service are of one-to-one correspondence and are the main body to accept clients' service request. On the other hand, the system distributes threads according to service request. Each service thread may serve a single request, or serve another request at the same time (Pietro *et al*., 2008). With this structure, users are enabled to request different services in different layers.

OPTIMAL BIOLOGICAL DATABASE SYSTEM LAYERED QUEUING NETWORK MODEL UNDER CLOUD ENVIRONMENT

Biological database system under cloud environment: Under cloud computing environment, biological database offers large volumes of users with

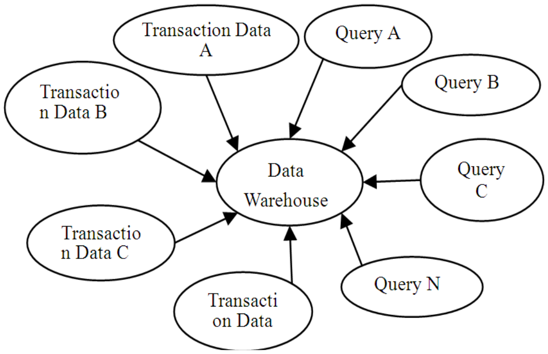

Fig. 3: Structure of OLTP load

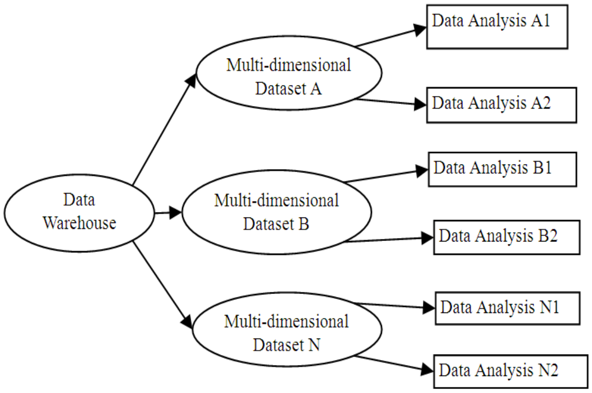

Fig. 4: Structure of OLAP load

Table 2: Comparison of OLAP and OLTP

Indexes	OLTP	OLAP
User	Operator, bottom management personnel	Decision-makers, senior management personnel
Function	Daily operation and processing	Decision analysis
DB design	Application-oriented	Theme-oriented
Data	Latest, detailed, two-dimensional, discrete	Historical, collective, multi-dimensional, integrated
Accessing scale	Reading/writing several (or even hundreds of) records	Reading millions (or billions) of records
Operation frequency	Very frequent (counting in seconds)	Low frequency (counting in hours or weeks)
Working unit	Strict transaction	Complex query
User volume	Hundreds – thousands	Several-hundreds
DB size	100MB-GB	100GB-TB

Under cloud conditions, OLAP can best reflect the loading capacity of biological database

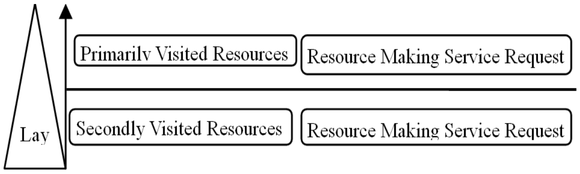

Fig. 5: Biological database service request layers under cloud environment

assessing service with service-oriented mode (Jennifer, 2008). Apart from providing data storage and evacuation function, biological database systems have two fixed loads that can best reflect the high running load and efficiency when multiple users are accessing the data base, i.e., the normal and stable running of Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP). Online transaction process reflects the running status of current biological database storage unit and is mainly used to complete routine task related biological database applications managed by the unit. Normally, this load has no complicated query and analytical processing. In online transaction processing, transactions are executed promptly (Hansen *et al*., 2006). This is opposite to batch processing. In batch processing, a batch of transactions will be stored for a period and then be executed. Online analytical processing is a concept about multi-dimensional biological database and multi-dimensional analysis. It realizes fast, stable, consistent and interactive data assess via several possible observation forms on information. It also allows management personnel and decision-makers to observe data more profoundly. Decision-making data is multi-dimensional data, while multi-dimensional data is the major content for decision-making. Its result can provide decision-makers with lucid result. The concept of OLTP and OLAP loads is shown in Fig. 3 and 4.

Compared with OLTP, OLAP has stronger data processing capability, analytical decision-making function and data accessing efficiency. Comparison of major indexes of OLAP and OLTP is shown in Table 2. Its performance under multi-user and multi-task condition has become an important part in performance evaluation.

Construction of layered queuing network model:

- **Analysis on service request layer under cloud environment:** Under cloud environment, biological database' major load comes from OLAP query request. According to its query rules, multi-layered query language should be executed in sequence, with regard to the same request. Such layered structure can be described in two aspects, as is shown in Fig. 5.

It can be seen from the diagram that, in biological database service request, resource primarily visited and making service request is in a higher layer. Normally,

Table 3: Model components

Hardware	Type	Amount	Normative outlet
CPU	Entrance	1	Software task of CPU making service request
Disk	Entrance	1	Software task of disk making service request
Query agent process	Entrance	1	Query program
Data management process	Entrance	2	Data management program, including data reading and upgrade

in biological database service request, client end is in a higher hierarchy, secondly data management process and thirdly the server end.

- **Construction of layered queuing network model for biological database:** Under cloud environment, the connection between resource and resource, client and service cannot go without processing system, processing process and storage structure. Layered queuing network model can performance efficient modeling and normative description to different software and hardware. As for this, the constituent parts of layered queuing network model for biological database are shown in Table 3:

  If dividing components in Tables according to their hierarchies, we will be able to construct a layered queuing network model with two nodes and three layers, as is shown in Fig. 6.

- **Reading of model parameters:** Layered queuing network model parameters are comprised by input parameter and output parameter. Input parameter refers to given conditions for solving the model, which can be obtained by analyzing the load execution flow of the system, or measuring in laboratory. Output parameter of model refers to the performance index of the system. In layered queuing model, we may utilize input parameter and solving algorithm to figure out the required output parameter. After having established the model frame, we have to figure out the load profile of the model as the basis for the subsequent performance test. In the model, input parameters of all layers are shown in Table 4.

Table 4: Defined input parameters in different layers

Task type	Parameter
Client task	T, N
Query agent process task	N
Data management process task	M
CPU and disk hardware task	K, t

In the Table 4, T is the reflection time, N is the average requesting number for invoking tasks of lower layer, M is the average requesting number for invoking hardware equipment task of lower layer, K is the average visiting number for invoking of upper software task and t is the service time of each visit.

Defining output as the average response time for biological database query request; the major object to be investigated is the average system response time for executing the 1st, 10th, 20th, 40th, 60th, 80th and 100th query, when multiple clients requesting to write 100 records to the biological database.

Time computing model is described as follows, in which T is the time, P is the use ratio and M is the number of query execution:

$$T_{CUP} = \frac{P_{CUP} * T}{M} \qquad T_{DISK} = \frac{P_{DISK} * T}{M} \tag{5}$$

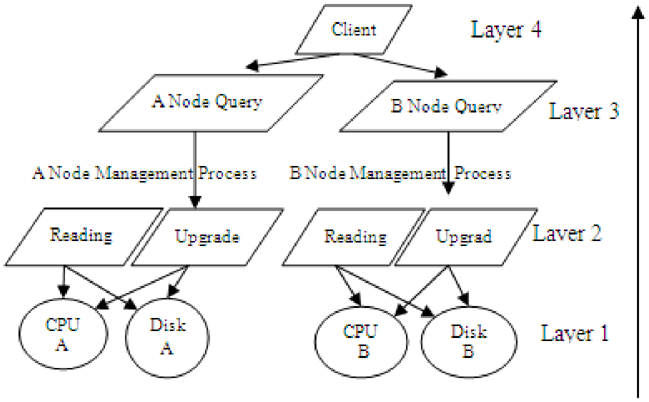

Fig. 6: Layered queuing network model for cloud biological database

Solving of layered queuing network model: The layered queuing network model constructed belongs to multi-load network model, which solves the constructed model via optimizing single-load network model. The solving process is shown as follows:

Step 1: Defining the performance parameter as R, U and initializing the intermediate result

Step 2: Assuming that there is no competition between hardware resources, iterating the initialized parameter from bottom layer to top layer and the operations executed shall be:

$$\begin{aligned} &R_{g,k} \leftarrow D_{g,k} \\ &R_g^k \leftarrow \sum_{k \in K} D_{g,k} \\ &R_g^{GRP} \leftarrow \sum_{k \in G_{i-1}} Vg,h*(R_h^{DEV} + R_h^{GRP}) \\ &R_g^{IDL} \leftarrow Z_g \end{aligned} \tag{6}$$

Step 3: Performing iteration operation separately to the sub-models of hardware and software:

$$\begin{aligned} &\mathit{Hardware}: \mathit{iter} \leftarrow 0 \\ &\mathit{Software}: \mathit{iter}=\mathit{iter}+1 \end{aligned} \tag{7}$$

Step 4: Judging if the sub-model iteration has been completed; judging of current computing result $R_g$is larger than the previous computing result R_gi. If so, finishing this iteration and update R_gi, or else, continue to execute Step 3.

TESTING OF PERFORMANCE EVALUATION BASED ON LAYERED QUEUING NETWORK MODEL

Testing plan: Constructing a layered queuing network model with C language; installing there computers separately of 2G 4G and 256MB RAM with H base experiment environment under Linux; creating tables with SQL statement compiled with Hive. Information in the cloud biological database is comprised by over 75 million records, including train number, departure station, arrival station, intermediate station, expected arrival time for each station, remaining train ticket, price, etc. Multiple users are divided into 4 groups (10, 20, 40 and 60) to visit the biological database at the same time, so as the test the loading capacity of the system by querying related information.

Comparison and analysis: Multiple users are assigned to submit visiting request at the same time, while the requesting language is to query stations passed by trains from Changsha to Beijing, over 20 train numbers, names of stations and ticket prices. The average system response time predicted with layered queuing network model is shown in Table 5.

Table 5: Predicted response time

User volume (person)	Predicted response time (second)
10	45.8
20	105.3
40	240.1
60	358.7

Comparison between response time measured via experiment and prediction is shown in Fig. 7.

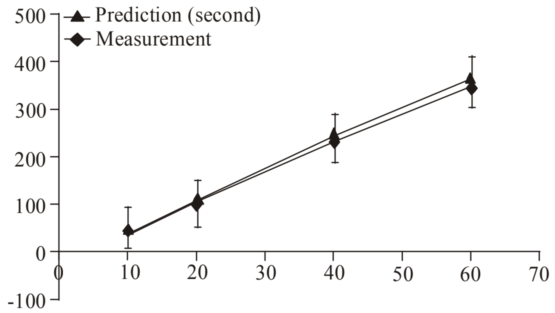

Fig. 7: Comparison between response time measured via experiment and prediction

According to the function:

$$\text{Error}=\frac{\text{Predicted Value - Measured Value}}{\text{Measured Value}} \tag{8}$$

We may draw the error curve for the prediction in Fig. 8.

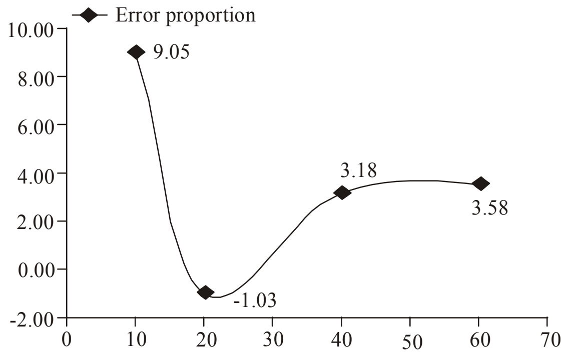

Fig. 8: Error curve

It can be seen from the above comparative analysis that, the layered queuing network model constructed has predicted the response time of railway tick query system under multiple user state. The prediction time of this model is quite close to the average system response time measured, with no obvious error value. This has proved that, this model can perfectly predict the performance of biological database under cloud environment. In the meanwhile, by comparing the error of this prediction, we may find that, the prediction error decreases along with the user number increases. This demonstrates that, when there is more cloud service users, the model has lower error rate in performance evaluation. However, when the number of users is low, the error rate could reach up to 9%, which is still quite considerable. This shows that, the model still needs to be improved. Even so, generally speaking, this test has proved that, under cloud environment, layered network model can perfectly evaluate the performance of biological database, which is in a qualified sense helpful to the improvement of biological database performance, as well as the stability and efficiency of biological database.

CONCLUSION

Cloud computing is the major mode for online service. Under cloud environment, the performance of biological database has significant influence on cloud service efficiency. What's more, performance evaluation is an effective measure to upgrade and optimize the efficiency of biological database. For a long time, layered queuing network model has been proved to be an effective method to evaluate biological database performance. Through optimal layered queuing network model and its solution, biological database performance evaluation model system under cloud environment was established. On this basis, this model system was tested under experimental conditions. By comparing the predicted result with practically measured result, we have found that, the error rate of H base cloud biological database performance predicted by layered queuing network model is within acceptable range, proving that, in performance evaluation, layered queuing network model is of high effectiveness.

REFERENCES

Andrzej, G. and B. Michael, 2010. Toward dynamic and attribute based publication, discovery and selection for cloud computing. Future Gen. Comp. Syst., 26(7): 947-970.

Aslı, D.H.K., 2011. The playground of cloud computing in Turkey. Procedia Comput. Sci., 3: 459-463.

Bao-Chyuan, J., W.H. Kohler and D. Towsley, 1988. A queueing network model for a distributed database testbed system. Softw. Engn., 4(7): 908-921.

Bian Xuegong, Hu Rui-Min, Chen Jun, *et al*, 2004. LON Based Performance Model for MCU Performance Prediction. Chinese J. Comput., 27(2): 209-215.

Chien-Yuan, H., C. Shi-Chung and C. Chern-Lin, 1995. Performance evaluation of a cache-coherent multi-processor by iterative mean value analysis. Perform. Evaluat., 23(1): 31-52.

Daniele, A. and M. Enrico, 2012. A cost-effective cloud computing framework for accelerating multimedia communication simulations. J. Parall. Distrib. Comp., 72(10): 1373-1385.

Hansen, A.J., B.T. Simonsen, C. Børsting, *et al.*, 2006. Semi-automatic preparation of biological database samples for STR typing. International Congress Series, 1288: 663-665.

Jennifer, M.F., 2008. Towards standards for data exchange and integration and their impact on a public database such as CEBS (Chemical Effects in Biological Systems). Toxicol. Appl. Pharmacol., 233(1): 54-62.

Mezmaz, M., N. Melab, Y. Kessaci, Y.C. Lee, E.G. Talbi, A.Y. Zomaya and D. Tuyttens, 2011. A parallel bi-objective hybrid metaheuristic for energy-aware scheduling for cloud computing systems. J. Parall. Distribut. Comp., 71(11): 1497-1508.

Nabil, S., 2010. Cloud computing for education: A new dawn? Int. J. Inform. Manag., 30(2): 109-116.

Niroshinie, F., W.L. Seng and R. Wenny, 2013. Mobile cloud computing: A survey. Future Gen. Comp. Syst., 29(1): 84-106.

Pietro, C., L. Pasquale, M.M. Rina and M. Roberto, 2008. A queuing network model for the management of berth crane operations. Comp. Oper. Res., 35(8): 2432-2446.

Rasha, O. and J.K. William, 2012. Database system performance evaluation models: A survey. Perform. Evaluat., 69(10): 471-493.

Subhas, C.M. and M. Arka, 2011. Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding return on investment. Math. Comp. Modell., 53(3-4): 504-521.

Tanzim, K., A.B.M. Shawkat Ali and S.A. Wasimi, 2012. A survey on gaps: Threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Gen. Comp. Syst., 28(6): 833-851.

Vijindra and S. Sudhir, 2012. Survey on scheduling issues in cloud computing. Proc. Eng., 38: 2881-2888.

Wei, W., W. Bing and T. Don, 2002. Continuous-time hidden Markov models for network performance evaluation. Perform. Evaluat., 49(1-4): 129-146.

Wei, R., Y. Linchen, G. Ren and X. Feng, 2011. Lightweight and compromise resilient storage outsourcing with distributed secure accessibility in mobile cloud computing. Tsinghua Sci. Technol., 16(5): 520-528.

Wong, J.W. and S.S. Lam, 1982. Queuing network models of packet switching networks part 1: Open networks. Perform. Evaluat., 2(1): 9-21.