http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (16): 3571-3575, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3571.3575

Performance Analysis and Evaluation of Database Connection Pool Technology Based on Queueing Model

Huo Zhan-Qiang, Zhang Jin-Cheng and Wang Zhi-Heng
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000, Henan, China

Abstract: In order to reveal the dependency relationships between configuration parameters and performance measures of database connection pool technology, the Geom/Geom/c/c discrete-time queuing model with multi-server is built according to its working principles. Using the embedded Markov chain method, the analysis of the model is completed and the mathematical expressions of system performance measures are derived, such as request blocking probability, average number of connections, system utilization, and system throughput etc. Finally, numerical examples are given to discuss the impact of these four performance measures on request arrival rate and maxConn (the maximum number of connections) intuitively. The analysis results of this study provide theoretical bases for the parameter configuration of database connection pool.

Key words: Database connection pool, performance analysis, queueing model

INTRODUCTION

Database connection, as a key resource to interact with the database, is particularly important in the applications, especially in web application of multi-user. Connection reuse can improve the utilization rate of connection and improve the performance of the system about database operations and reduce the system overhead. The core of the connection pool is connection reuse.

Wang *et al*. (2010) gives a quantitative data about the improvement of the performance of the database access based on the connection pool through performance contrast tests under three different kinds of database access modes. Li Bingzhang and other people propose the optimal allocation method of database connection pool based on the log file records. Zhu Changsheng *et al*. give us the configuration method in detail for database connection pool parameter configuration based on XML under J2EE framework. Liang *et al*. (2006), Meng *et al*. (2013), Zeng and Fu (2011), Lv *et al*. (2012) and Liu (2012) *et al*. make a study and have obtained some significant results from database connection pool technology.

The performance analysis methods of database connection pool technology can be divided into three categories: Experimental data analysis, simulation research method and mathematical model analysis methods. In this study, the mathematical model analysis method is used for the database connection pool technology to establish a discrete-time multi-server queuing model. The analysis of classical discrete-time queuing model is shown by Tian and Zhang (2006).

WORKING PRINCIPLE

Database connection pool takes the database connection as the management entity. It well solved too much overhead and low response speed of the system and so on caused by frequent database access by the system. The process of the connection pool management is shown in Fig. 1.

- When a client requests a database connection, the first is to see if there are free connections in the pool. If there is a free connection, then take out a free connection to assign to the customers to use and at the same time mark the connection as busy connection, as shown in (1) in Fig. 1
- If there is no free connection, check whether the total number of current free connections and the busy connections have reached the maximum number of connections set by the system. If it is not up to the maximum number of connections, create a connection to the request of the customer, as shown in (2) in Fig. 1. If it has reached the maximum number of connections, wait according to the maximum waiting time set by the system, if it exceeds the maximum waiting time, an exception will be thrown to the customer, as shown in (3) in Fig. 1

Corresponding Author: Wang Zhi-Heng, School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000, Henan, China

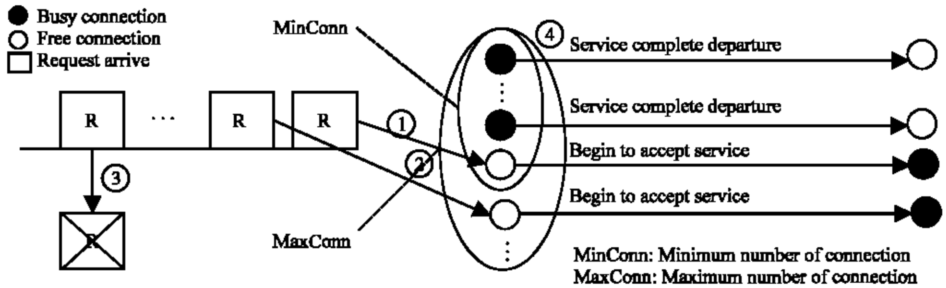

Fig. 1: Diagram of management process of database connection pool

- In the end of the use of connection for customers, put the connection back into the connection pool to improve the reuse rate of connection while the connection is marked as free connections, as shown in (4) in Fig. 1

DESCRIPTION AND ANALYSIS

Description of model: Timeline is divided into fixed interval of small pieces, called time slot. Arrival can only occur at the end of each time slot $t = n^-$ ($n = 1,2,...$), known as late arrival system. In order to clearly define the number of customers in system at a discrete time $t = n$, all the start and end agreed upon in the service will be at $t = n$. If customers arrive at n^- and someone left at $t = n$, he can immediately begin in the service at $t = n$. This system is called immediate access. Customer arrival and departure behaviors are shown in Fig. 2.

Each end of time slot shall have an arrival with probability p and no arrival with probability $\bar{p} = 1 - p$, the arrival on different time slots are independent. In other words, the arrival interval T obeys geometric distribution. The customer service time marked S also obeys the geometric distribution. Arrival interval and service time are independent of each other. There is C front desks in the system, if the customer arrival is in the situation that C front desks are occupied and no customers left in his arrival time, the customer will be refused to enter into the system.

Model analysis: Assume that $L_n = L(n^+)$ indicates the number of customers at the time $t = n^+$ in the system. According to the arrival and departure behavior of the agreement, the arriving customers at $t = n^-$ will be included in L_n and the customers who left at $t = n$ shall not be included. $\{L_n, n \geq 1\}$ is a Markov chain with state space $\Omega = \{0, 1, \ldots, c\}$, the transition probabilities formation is:

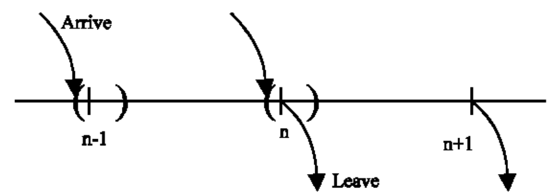

Fig. 2: Customer arrival and departure

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & & & & \\ p_{1,0} & p_{1,1} & p_{1,2} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ p_{c-1,0} & p_{c-1,1} & p_{c-1,2} & \cdots & \cdots & p_{c-1,c} \\ p_{c,0} & p_{c,1} & p_{c,2} & \cdots & \cdots & p_{c,c} \end{pmatrix}$$

In which:

$$p_{i,j} = P\{L_{n+1} = j \mid L_n = i\}, \quad i = 0,1,\cdots,c; j = 0,1,\cdots,\min(i+1,c)$$

And:

$$p_{i,j} = \begin{cases} \bar{p}\mu^i, & i = 0,1,\cdots,c; j = 0, \\ \bar{p}\binom{i}{i-j}\mu^{i-j}\bar{\mu}^j + p\binom{i}{i-j+1}\mu^{i-j+1}\bar{\mu}^{j-1}, & \\ & i = 0,1,\cdots,c-1; j = 1,\cdots,i, \\ p\bar{\mu}^i, & i = 0,1,\cdots,c-1; j = i+1, \\ \bar{p}\binom{c}{c-j}\mu^{c-j}\bar{\mu}^j + p\binom{c}{c-j+1}\mu^{c-j+1}\bar{\mu}^{j-1}, & \\ & i = c; j = 1,\cdots,c-1, \\ \bar{\mu}^c + p\binom{c}{1}\mu\bar{\mu}^{c-1}, & i = c; j = c. \end{cases}$$

If:

$$\binom{n}{k} = 0$$

is agreed, when k<0 or k>n, except for the above non- zero transition probability that does not included $p_{c,c}$ can be expressed as:

$$p_{i,j} = \bar{p}\binom{i}{i-j}\mu^{i-j}\bar{\mu}^{j} + p\binom{i}{i-j+1}\mu^{i-j+1}\bar{\mu}^{j-1} \tag{1}$$

Assume that the limit distribution of Markov chain is Π, take $\Pi = \{\pi_0, \pi_1,..., \pi_c\}$ to express the limit distribution of $\{L_n, n\geq 0\}$. Introduce:

$$\psi_j^{(k)} = 1-\sum_{i=0}^{j}\binom{k}{i}\mu^{i}\mu^{k-j} = \sum_{i=j+1}^{k}\binom{k}{i}\mu^{i}\mu^{k-j} \tag{2}$$

By the theorem (Tian and Zhang, 2006), the probability of steady state team length $p_0, p_1,\ldots, p_c$ can meet stable distribution and the normalization condition of recursive relations for long probability which can get the following equation:

$$p\bar{\mu}^{k}\pi_k = \sum_{j=k+1}^{c}\pi_j(\bar{p}\psi_{j-k-1}^{(j)} + p\psi_{j-k}^{(j)}) \tag{3}$$

$$p\bar{\mu}^{c-1}\pi_{c-1} = \pi_c(\bar{p}\psi_0^{(c)} + p\psi_1^{(c)}) \tag{4}$$

in which, k is equal to 0, 1, 2,..., c-2.

PERFORMANCE MEARURES

Request blocking probability: The request blocking probability, namely the possibility that the customer requests to establish a connection but to be rejected in steady state to be recorded as p_1, then $p_1 = \pi_c\bar{\mu}^{c}$.

Average number of connections: The average number of connections in the system, namely the average value of number of connections established in the database connection pool in steady state. To be recorded as $\bar{k}$, then:

$$\bar{k} = \sum_{i=0}^{c} i\pi_i$$

Utilization ratio: The utilization ratio of the system, can be established in connection pool. To be recorded as η, then:

$$\eta = \frac{\bar{k}}{c} = \frac{\sum_{i=0}^{c} i\pi_i}{c}$$

System throughput: System throughput, namely the number of customers who are completed the service per unit of time. To be recorded as θ, then:

$$\theta = \sum_{i=0}^{c} i\mu\pi_i = \mu\sum_{i=0}^{c} i\pi_i = \mu\bar{k}$$

NUMERICAL EXAMPLES

The relevant parameters in this section are set as follows: connection request arrival probability p in each time slot is set to 0.1-0.9. Serve ratio μ in each time slot is set to 0.03. The maximum number of connections C is set to 1-32.

Figure 3 shows when the service rate is set to 0.03, the maximum number of connections is set to 4,8,16 and 32 separately and the curve of request blocking probability complies with the arrival rate. When the service rate μ and the maximum number of connections c are under certain conditions. Request blocking probability increases with the increase of request arrival rate p and decreases with the increase of the maximum number of connections c.

Figure 4 shows when the service rate is set to 0.03, the request arrival rate is set to 0.2, 0.4, 0.6 and 0.8

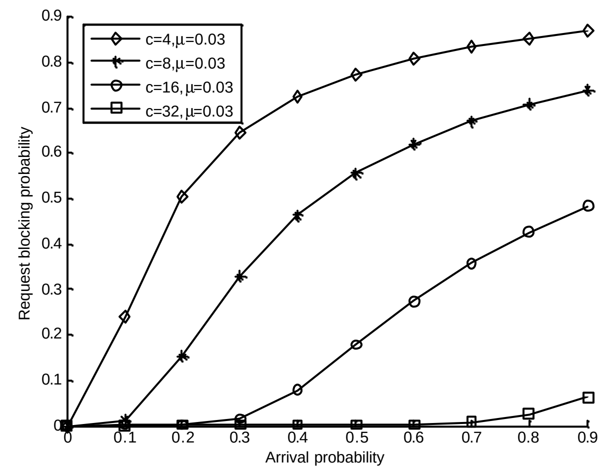

Fig 3: Changes of request blocking probability comply with the arrival rate

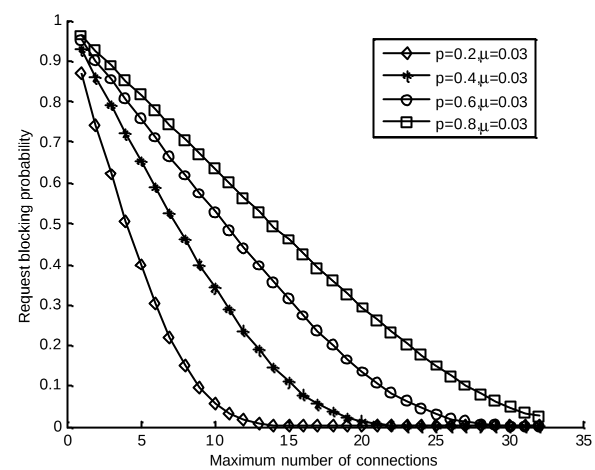

Fig. 4: Changes of the request blocking

separately, the curve of request blocking probability with the maximum number of connections. When the service rate μ and the request arrival θ are under certain conditions, the request blocking probability decreases with the maximum number of connections c. When the service rate μ and the maximum number of connections c are under certain conditions, the request blocking probability increases with the increase of request arrival rate p.

Probability comply with the maximum number of connections: Figure 5 shows when the service rate is set to 0.03, the maximum number of connections is set to 4,8,16 and 32 separately, the curve of system average number of connections with the arrival rate.

While the service rate μ and the maximum number of connections are fixed, the system average number of connections increases with request arrival rate p and closes to the fixed value of c. When the service rateÌand the request arrival rate are under certain condition, the system average number of connections increases with the increase of maximum number of connections c.

Figure 6 shows when the service rate is set to 0.03; the request arrival rate is set to 0.2, 0.4, 0.6 and 0.8 separately, the curve of system average number of connections complies with the maximum number of connections.

Figure 7 shows when the service rate is set to 0.03 and the maximum number of connections is set to 4,8,16 and 32 separately, the curve of system utilization rate complies with the arrival rate.

Figure 8 shows when the service rate is set to 0.03 and the request arrival rate is set to 0.2, 0.4, 0.6 and 0.8 separately, the curve of system utilization rate complies with the maximum number of connections. When the

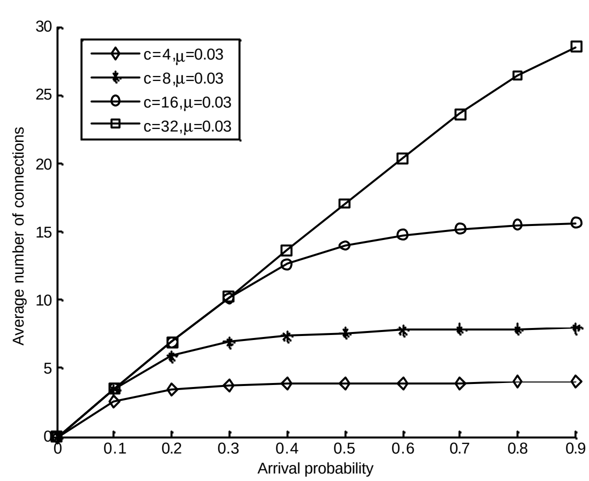

Fig. 5: Changes of system average number of connections comply with arrival rate

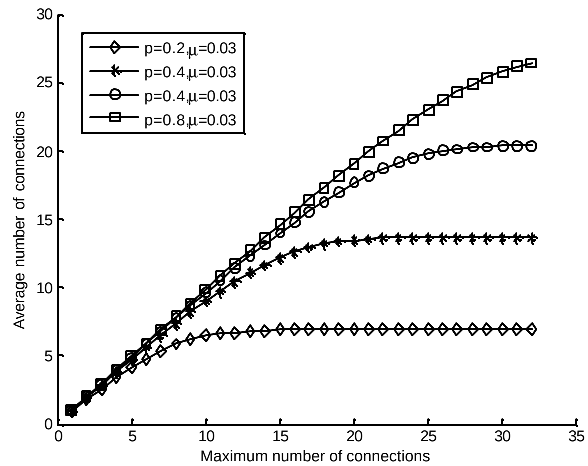

Fig. 6: Changes of system average number of connections comply with maximum number of connections

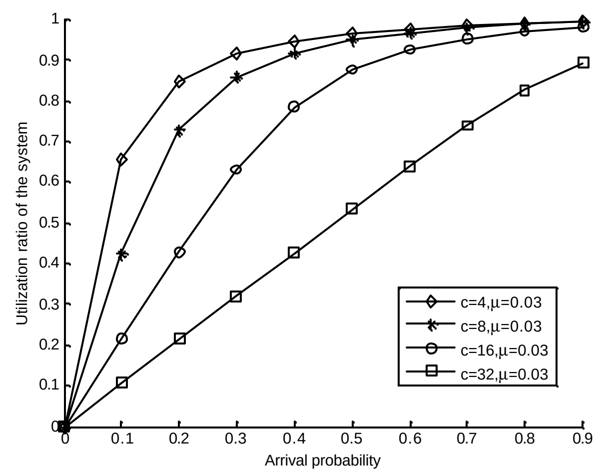

Fig. 7: Changes of system utilization rate comply with the arrival rate

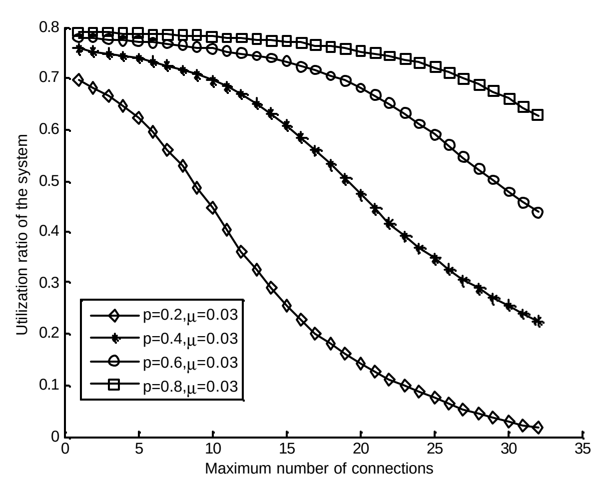

Fig 8: Changes of system utilization rate

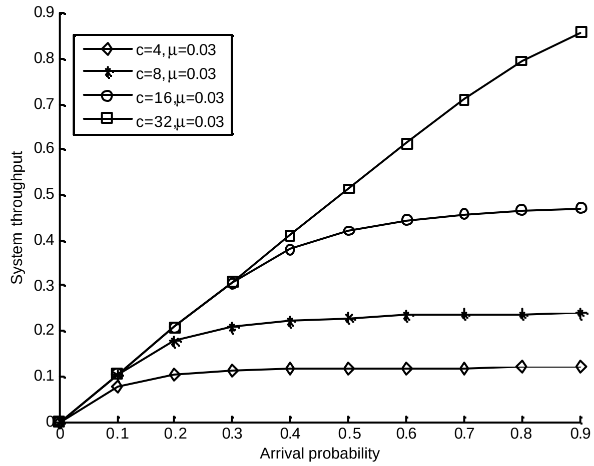

Fig. 9: Changes of system throughput comply with the arrival rate

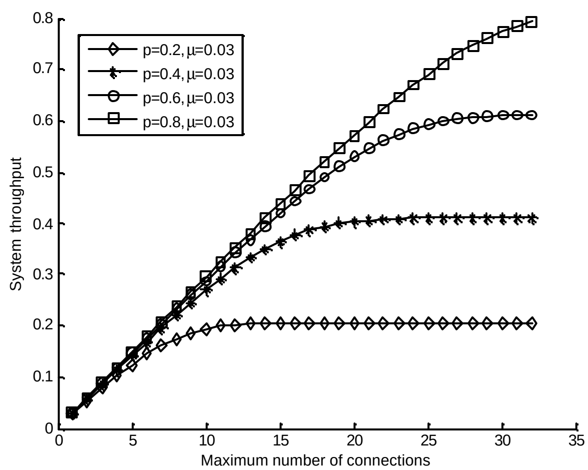

Fig. 10: Changes of system throughput comply with the maximum number of connections

service rate μ and the request arrival rate p are under certain conditions, the system utilization rate decreases with the increase of maximum number of connections c. When the service rate|Ìand the maximum number of connections c are under certain conditions, the system utilization rate increases with the increase of request arrival rate p.

Comply with the maximum number of connections: Figure 9 shows when the service rate is set to 0.03 and the maximum number of connections is set to 4,8,16 and 32 separately, the curve of system throughput complies with the arrival rate.

Figure 10 shows when the service rate is set to 0.03 and the request arrival rate is set to 0.2, 0.4, 0.6 and 0.8 separately, the curve of system throughput complies with the maximum number of connections. When the service rate|Ìand the request arrival rate p are under certain conditions, the number of system throughput increases with the increase of maximum number of connections c. When the service rate|Ìand the maximum number of connections c are under certain conditions, the system throughput increases with the increase of request arrival rate.

ACKNOWLEDGMENT

This study was supportedby Natural Science Foundation of Hebei Province (No. F2009000475) and Doctorial Foundation of HPU (No. B2010-26).

REFERENCES

Liu, F., 2012. A method of design and optim-ization of database connection pool. Proceedings of the 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, Volume 2, August 26-27, 2012, Nanchang, Jiangxi, pp: 272-274.

Lv, J., G. Dai and D. Mu, 2012. Implementation of absolute delay guarantee in database connection pool in Web application servers. Appl. Res. Comput., 29: 1838-1841.

Liang, Q., Z. Shen, J. Luo, H. Fan, D. Ming and J. Li, 2006. Study of database connection pool in LBS platform. Comput. Eng., 18: 39-41.

Meng, P., S. Hu and H. Shu, 2013. Optimization strategy of connection pool based on ADO. Comput. Eng. Design, 5: 1706-1710.

Tian, N. and Z.G. Zhang, 2006. Vacation Queueing Models: Theory and Applications. Springer, New York.

Wang, Y., F. Wang and W. Yang, 2010. Performance comparison test of database access based on database connection pool. Ind. Control Comput., 6: 92-94.

Zeng, G. and X. Fu, 2011. A new model of database connection pool. Comput. Digital Eng., 39: 163-166.