Volume 5 Issue 1, June 2017
ISSN: 2321-6840

Optimal Processing Time for B2C Electronic Commerce Architecture in a Cloud Computing Environment

Riktesh Srivastava
Associate Professor, Information Systems, Skyline University College, Sharjah, UAE.
E-mail: rsrivastava@skylineuniversity.ac.ae, riktesh.srivastava@gmail.com

Abstract: Hosting electronic commerce applications in a client/server architecture was problematic, and moreover these companies were retail focused, rather than technology organizations. Advent of cloud computing has solved the problems of technological requirements, however, handling massive traffic was still a challenge. Citing this reason, these companies relocated to public cloud environment expecting to resolve the issues. But, failure persisted owing to number of requests being increased tenfold. The numbers augmented due to usage of mobile and tablets and the existing architecture was not prepared for such a substantial growth and often resulted in breakdown. The study is an effort to mathematically evaluating the optimal processing time for an architecture hosted in a public cloud.

Keywords: Electronic commerce architecture, Request time, Response time, Optimal processing time.

I. Introduction

Electronic commerce market will jump from 14% in 2014 to 29% by 2018, making it 1/3rd of all retail sales (Madrigal, 2013). Also, the total market is expected to rise to $27 trillion by 2020 and continues to grow (eMarketer, 2016). However, the growth is slowdown in next few years (eMarketer, 2016). The slowdown is because of poor performance of sites on mobile devices, which accounted for 73% of sales in 2016 (Silver, 2016). Ease to access the site from anywhere and from any device possesses an additional challenge for electronic commerce infrastructure. The infrastructure was not ready to face the enormous user turnout and resulted in major pitfall. To quote an example, Flipkart's "Big Billion Day" in 2014 sale generated unprecedented buzz in the online community but failed drastically (Sarkar, 2014). They received 83% more hits than expected and over 75% of them through mobile devices (Julka, 2015), the server failed desolately (Priyanka Pani, 2014).

Cloud computing moved data processing from servers maintained by companies into large data centers. It refers to applications delivered as services over the Internet as well as to the actual cloud infrastructure (Dikaiakos, 2009). Hosting electronic commerce applications into cloud and accessing it as services enable businesses to rapidly respond to market changes. The major advantages of adoption of Cloud in electronic commerce applications are (Kulkarni, 2015) :

- Allow businesses to respond swiftly to market opportunities and challenges, thereby providing flexibility,
- Capability to hold massive traffic without large upfront investments, thereby reducing capital expenditure.

It is due to this reason Flipkart has now adopted exclusive public cloud platform, Microsoft Azure in 2017 (Vignesh, 2017). The study mathematically calculates the optimal processing time, based on request received and response generated. By doing so, we can easily identify the number of requests to be processed per unit time, thereby avoiding failure.

Rest of paper is organized as follows: Section 2 gives particulars of electronic commerce architecture adopted for study. The section starts with 2-tier electronic commerce till public cloud based architecture. Section 3 organizes the mathematical formulation of optimal processing time for architecture. Grounded on outcomes of section 3, section 4 gives recommendations accordingly.

II. Electronic Commerce Architecture

The early electronic commerce architecture was based on 2-tier configuration (Alex Homer, 2000), as specified in Figure 1.

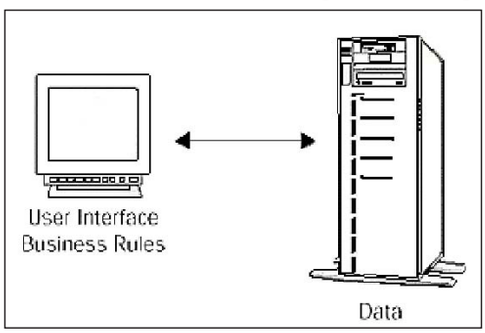

Fig. 1: 2-Tier Electronic Commerce Architecture

As mentioned in Figure 1, the requests are received and processed at the server and then response is generated. The problem arises, when the number of request outburst, responses have to indefinitely wait. To eradicate the problem, 3-tier architecture was proposed as indicated in Figure 2 (Thiru, 2016). In the architecture, the number of an application server(s) can be even be '*n*', contingent to necessity. For simplicity, a hypothesis is assumed to a single request arrival and response departure process is generated at a time.

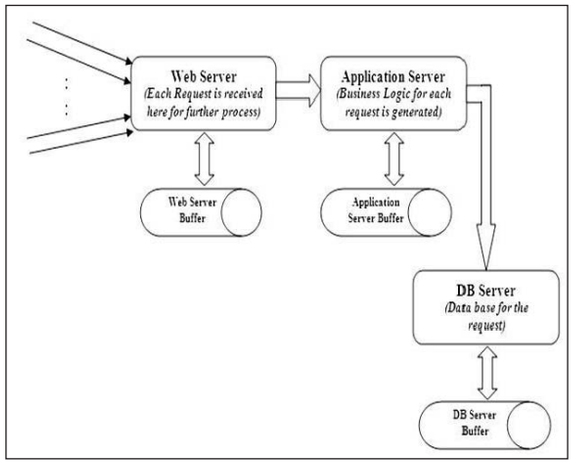

Fig. 2: 3-Tier Electronic Commerce Architecture

There are two main challenges for the above-mentioned architecture are:

1. Maintenance of servers, which is both time consuming and expensive.
2. Request and response generated are random and hence the condition of ergodicity needs to me preserved.

The first challenge is resolved by instigating cloud computing infrastructure as adopted from Figure 3 (Riktesh, 2013). Web server now no longer does entire processing of requests. The network of servers hosted in the cloud does the task accordingly.

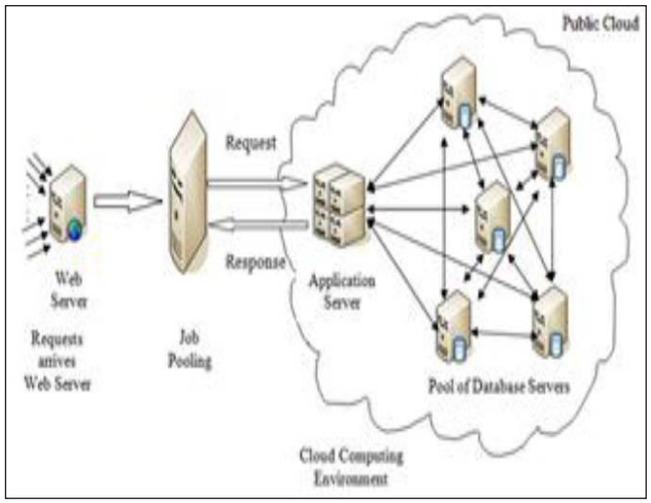

Fig. 3: B2C Electronic Commerce Architecture Implemented in Public Cloud

It is assumed that the number of requests which arrives at web server are $[r_1 + r_2 + r_3 + ... + r_n]$. These requests are random in nature and denoted by λ. At any time *t*, the number of requests are denoted as:

$$\lambda = \frac{[r_1 + r_2 + r_3 + ... + r_n]}{t} \quad (1)$$

These requests are processed at severs (in the cloud computing environment) denoted by μ. At any time *t*, the number of response are denoted as:

$$\mu = \frac{[rp_1 + rp_2 + rp_3 + ... + rp_n]}{t} \quad (2)$$

Figure 4 illustrates three types of processing time at the architecture

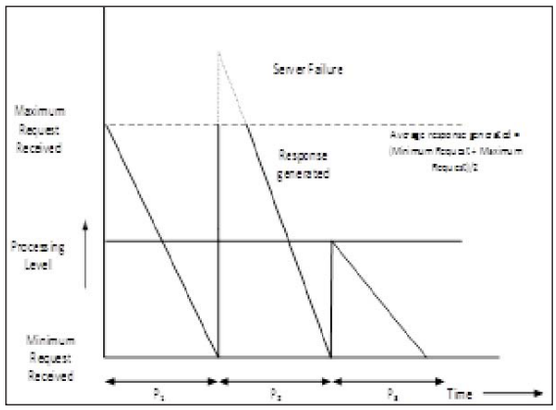

Fig. 4: Types of Processing Time

At P_1, $\lambda = \mu$, called as Null State. This state is practically neither occurring nor desirable.

At P_2, $\lambda > \mu$, called as Burst State. In this state, the number of requests is much higher than responses that can be generated and results is server failure.

At P_3, $\lambda < \mu$, called as Ergodic State. In this state, the number of requests is less than responses, may results is underperformance of architecture.

(LK Singh, Riktesh Srivastava, 2007) studied the condition of ergodicity and proposed that that maximum throughput for ergodicity is achieved, if following condition is maintained, if the following condition (LK Singh, Riktesh Srivastava, 2008) is sustained:

$$\lambda = \mu + 1 \quad (3)$$

The condition as mentioned in equation (3) will be upheld through the mathematical formulation of optimal processing time.

III. Mathematical Formulation-Optimal Processing Time

In order to mathematically evaluate the optimal processing time, the number of requests and responses are evaluated, as described in Figure 5 below:

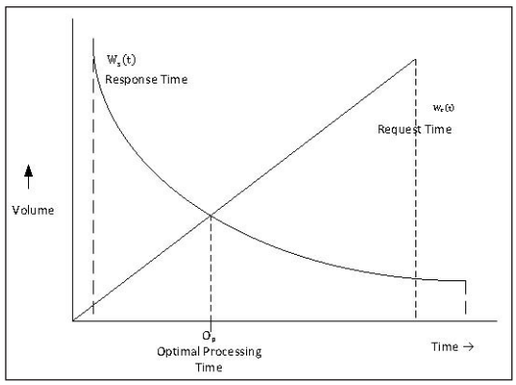

Fig. 5: Mathematical Evaluation of Processing Time

The two observations from Figure 5 are:

1. The response time, $W_r(t)$, is quite high initially however decreases subsequently.
2. Request time $W_s(t)$ is lower initially and increases at later stages.

Based on observations optimal processing time can be mathematically denoted as:

$$\left(\frac{O_p}{2}\right) W_r(t) = \frac{\lambda}{O_p} W_s(t) \tag{5}$$

A. Request Time Formulation $W_r(t)$

For estimation of the probability of n data in the servers, certain assumptions are to be made. This can be given as follows:

1. Δt is a very small time, in which only one process can occur, i.e., either arrival or departure.
2. State of arrival is denoted as λ and state of departure is denoted as μ.

Probability of one arrival $= \lambda \Delta t$

and, the probability of one departure $= \mu \Delta t$

Then, probability of no arrival $= 1 = \lambda \Delta t$

and, the probability of no departure $= 1 - \mu \Delta t$

Considering n data to present at any time t and is denoted by $P_n(t)$. If the time is increased from t to $t + \Delta t$, then

$$P_n(t + \Delta t) = \begin{cases} P_n(t)\,(1 - \lambda \Delta t)\,(1 - \mu \Delta t) \\ \quad P_{n+1}(t)\,(\mu \Delta t) \\ \quad P_{n-1}(t)\,(\lambda \Delta t) \end{cases} \tag{6}$$

$$P_n(t+\Delta t) = P_n(t)\,(1 - \lambda \Delta t)\,(1 - \mu \Delta t) + P_{n-1}(t)\,(\lambda \Delta t) + P_{n+1}(t)\,(\mu \Delta t) \tag{7}$$

or,

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_n(t) - \mu P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t) \tag{8}$$

But,

$$\lim_{\Delta t \to 0} \left\{ \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} \right\} = \frac{d}{dt} [P_n(t) = 0] \text{ for stable condition}$$

Thus, the R.H.S. of equation (8) becomes

$$P_{n-1}(t)\lambda - (\lambda + \mu)\, P_n(t) + P_{n+1}(t)_\mu = 0 \tag{9}$$

To solve equation (9), i is assumed that there were 0 requests at time $(t + \Delta t)$. This can be obtained from the states as given under:

$$P_0(t + \Delta t) = P_0\,(t)\,(1 - \lambda \Delta t)$$

$$= P_1(t)\,\mu \Delta t$$

$$\left(\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} \right) = P_n(t + \Delta t) = -P_0(t)\lambda + P_1(t)\mu \tag{10}$$

Thus, L.H.S. of equation (10) becomes

$$\lim_{\Delta t \to 0} \left(\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} \right)$$

$$= \frac{d}{dt} \{P_0(t) = 0\} \text{ for stable condition} \tag{11}$$

Hence equation (11) becomes

$$P_1(t) = \left(\frac{\lambda}{\mu}\right) P_0(t) \tag{12}$$

From equations (9) and (11), the following can be derived as:

$$P_0(t) = \left(\frac{\lambda}{\mu}\right)^0 P_0(t)$$

$$P_2(t) = \left(\frac{\lambda}{\mu}\right)^1 P_0(t)$$

$$P_2(t) = \left(\frac{\lambda}{\mu}\right)^2 P_0(t)$$

:

:

$$P_n(t) = \left(\frac{\lambda}{\mu}\right)^n P_0(t)$$

all the equations:

$$\sum_{i=0}^{n} P_i(t) = \left\{ \left(\frac{\lambda}{\mu}\right)^0 + \left(\frac{\lambda}{\mu}\right)^1 + \left(\frac{\lambda}{\mu}\right)^2 + \ldots + \left(\frac{\lambda}{\mu}\right)^n \right\} P_0(t) \tag{14}$$

Based on limiting condition, when $n \to \infty$ and $\frac{\lambda}{\mu} < 1$, L.H.S. becomes 1 and R.H.S. becomes

$$\left[\frac{1}{\left(1 - \frac{\lambda}{\mu}\right)} \right] P_0(t)$$

Thus equation (13) becomes

$$1 = \left[\frac{1}{\left(1-\frac{\lambda}{\mu}\right)}\right] P_0(t) \tag{15}$$

Substituting equation (14) in equation (13), we get (Riktesh, 2013)

$$P_n(t) = \left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right) \tag{16}$$

Thus, the average response time can be evaluated as

$$W_r(t) = \sum_{n\to\infty}^{N} nP_n(t)$$

$$= \sum_{n\to\infty}^{N} \left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right)$$

$$= \left(1-\frac{\lambda}{\mu}\right) \sum_{n\to\infty}^{N} \left(\frac{\lambda}{\mu}\right)^n$$

$$W_r(\mathrm{t}) = \frac{\left(\frac{\lambda}{\mu}\right)}{\left(1-\frac{\lambda}{\mu}\right)} \tag{17}$$

B. Response Time Formulation $W_s(t)$

There are certain assumptions to be made:

(1) The requests arrive at Application Servers according to Poisson process with parameter λt.

(2) Response time for each request is μ.

Based on assumptions (1) and (2)

Utilization, ρ A = Average rate X Average response rate

$$= \frac{\lambda}{\mu} \tag{18}$$

Another important assumption is $\rho < 1$, as without this assumption the response queue may grow without limit.

$\therefore$ The probability that there are n responses in the cloud computing system

$$\Rightarrow P_n = \rho^n [1-\rho] \tag{19}$$

Based on figure 6, in the steady state, the expected number of transition from n to $n+1$ should be equal to number of transitions from $n+1$ to n.

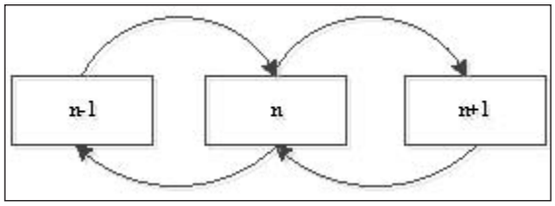

Fig. 6: Transition from n-1 to n+1

$$\therefore \lambda P_n = \mu P_{n+1} \tag{20}$$

For $n = 0$

$$P_1 = \frac{\lambda}{\mu} P_0 = \rho P_0 \tag{21}$$

Using equation (20) repeatedly, we get

$$P_n = \rho^n P_0 \tag{22}$$

It is also observed that

$$\sum_{n=0}^{\infty} P_n = 1 \tag{23}$$

By equation (22)

$$1 = P_0 \sum_{n=0}^{\infty} \rho^n = P_0 \frac{1}{1-\rho} \tag{24}$$

Since $\rho < 1$, equation (24) implies that

$$P_0 = 1-\rho \tag{25}$$

Expected number of responses L_1 is thus

$$L_1 = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n [1-\rho] \tag{26}$$

Simplifying equation (26), we get

$$L_1 = \sum n(1-\rho)\rho^n = (1-\rho)\sum n\rho^n$$

$$= (1-\rho)\rho \sum d\rho^n / d\rho$$

$$= (1-\rho)\rho \frac{d}{d\rho} \sum \rho^n$$

$$= (1-\rho)\rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \tag{27}$$

Also, the collection of servers adopted for B2C Electronic Commerce architecture were assumed as one server, the expected number of responses is denoted by L_2.

$$L_2 = \sum_{n=1}^{\infty} [n-1]P_n = \sum_{n=1}^{\infty} [n-1]\rho^n [1-\rho]$$

$$= \rho[1-\rho] \sum_{n-1=0}^{\infty} [n-1]\rho^{n-1} = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)} \tag{28}$$

Based on equation (28), average time each response stays in the system

$$W_1 = \frac{L_1}{\lambda} \frac{l}{(\mu-\lambda)} \tag{29}$$

and, average time each response stays in the cache

$$W_2 = \frac{L_2}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} \tag{30}$$

Taking equations (29) and (30) for the total response time

$$W_s(t) = W_1 + W_2 \approx \left(\frac{1}{(\mu-\lambda)} + \frac{\lambda}{\mu(\mu-\lambda)}\right) \tag{31}$$

C. Evaluating Optimal Processing Time Op

Substituting equations (17) and (31) in equation (5), we get

$$\left(\frac{O_p}{2}\right)\left[\frac{\frac{\lambda}{\mu}}{\left(1-\frac{\lambda}{\mu}\right)}\right]=\frac{\lambda}{O_p}\left[\frac{\left(\frac{\lambda}{\mu}\right)}{\left(1-\frac{\lambda}{\mu}\right)}\right] \tag{32}$$

Upon solving equation, we get

$$O_p=\sqrt{2(1+\rho)} \tag{33}$$

As discussed, taking the condition of ergodicity from equation (3), we get

$$O_p=\sqrt{2\left(2+\frac{1}{\mu}\right)} \tag{34}$$

where,

$\mu = \mu_{WS} + \mu_{AS} + \mu_{DBS}$

Also, $\mu_{WS} << \mu_{AS} + \mu_{DBS}$ (35)

$\mu = \mu_{AS} + \mu_{DBS}$

As mentioned by (Srivastava, 2012), the optimal performance by Database Server is obtained when

$\mu_{DBS} + 2\,\mu_{DBS}$ (36)

$$O_p=\sqrt{2\left(2+\frac{1}{\mu_{AS}+2.\mu DBS}\right)} \tag{37}$$

IV. Conclusion and Recommendations

As mentioned in equation (37), optimal processing time inversely depends on response generated by application and database server(s). This clearly indicates that overall performance is increased when the requested is available in the cache, called hit, decreases otherwise.

References

[1] A. Homer, "Professional Active Server Pages 3.0.," Wrox Press, 2000.

[2] M. D. Dikaiakos, "Cloud computing: Distributed internet computing for IT and scientific research," *IEEE Internet Computing*, pp. 10-13, 2009.

[3] eMarketer, "*Worldwide Retail E-commerce Sales Will Reach $1.915 Trillion This Year,"* eMarketer, 2016.

[4] H. Julka, "Flipkart revs up engine ahead of 'Big-Billion' sale," *The Economic Times*, 2015. Available: http://economictimes.indiatimes.com/industry/services/retail/flipkart-revs-up-engine-ahead-of-big-billion-sale/articleshow/49220661.cms

[5] V. Kulkarni, "Cloud computing impact on growth of ecommerce applications," *ESDS*, 2015. Available: https://www.esds.co.in/blog/cloud-computing-impact-on-growth-of-ecommerce-applications/#sthash.wdfCmwNe.hYv3NHQx.dpbs

[6] L. K. Singh, and R. Srivastava, "Estimation of buffer size of internet gateway server via G/M/1 queuing model," *International Journal of Applied Science, Engineering and Technology*, vol. 4, no.1, pp. 474-482, January, 2007.

[7] L. K. Singh, and R. Srivastava, "Design and implementation of G/G/1 queuing model algorithm for its applicability in internet gateway server," *The International Arab Journal of Information Technology*, vol. 5, no. 4, pp. 111-119, October, 2008.

[8] A. C. Madrigal, "*There Will Be as Much Mobile Commerce in 2018 as E-Commerce in 2013."* Goldman Sachs, 2013.

[9] P. Pani, and R. Kurup, Flipkart fumbles on the big day as server fails. *The Hindu (BusinessLine)*. Available: http://www.thehindubusinessline.com/info-tech/big-billion-day-sale-flipkart-site-crashes-on-heavy-demand/article6475295.ece

[10] R. Srivastava, 2014. "Evaluation of response time using gang scheduling algorithm for B2C electronic commerce architecture implemented in cloud computing environment by queuing models," *International Journal of Future Computer and Communication*, vol. 2, no. 2, pp. 71-75, 2013.

[11] D. Sarkar, "Big Billion Day' sale gets Flipkart millions of unhappy customers," *The Indian*, 2014. *Express*. Available: http://indianexpress.com/article/technology/technology-others/bigbillionday-gets-flipkart-millions-of-unhappy-customers/

[12] H. Silver, "Top 8 Mobile Shopping Problems," Multichannel Merchant, 2016, Available: http://multichannelmerchant.com/ecommerce/top-8-mobile-shopping-problems-16022016/

[13] Srivastava, R. "Estimation of Data Base (DB) server cache size using GI/G/n/k queuing model," *International Journal of Advancements in Technology*, vol. 3, no. 4, pp. 230-241, December, 2012.

[14] Thiru, "*E-Commerce Architecture,"* 2016. Available: http://www.myreadingroom.co.in/notes-and-studymaterial/66-e-commerce/517-e-commerce-architecture.html

[15] J. Vignesh, "Flipkart to use Microsoft Azure for its cloud infrastructure," *The Times of India,* 2017. Available: http://timesofindia.indiatimes.com/companies/flipkart-to-use-microsoft-azure-for-its-cloud-infrastructure/articleshow/57251465.cms