UNSW School of Computer Systems Engineering

COMP9334
Capacity Planning of Computer Systems and Networks

# ASSIGNMENT – Part B

In this assignment, we analyze proposed configurations for a server complex. The study consists of the following steps:

1) Showing that the proposed system can process the expected load;

2) Calculating the expected response times;

Basic system

The system has several computers connected by a "switched Ethernet" network (actually IEEE 802.3), as shown below. The internal network has such high capacity and is so fast that it can be ignored in the analyses that follow. Each computer consists of a processor and one disk.
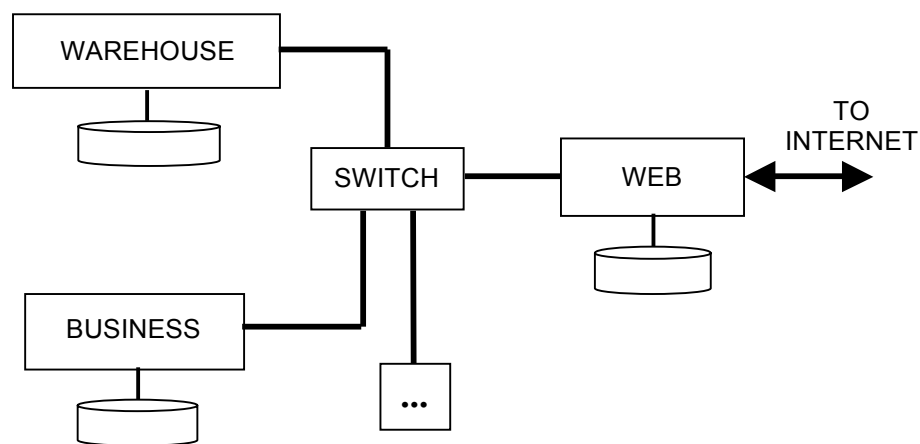
The workload will consist of two distinct classes:

* Business processing: WEB passes requests received from Internet for information, orders, payments and so on ('transactions') to one of several identical computers; we only need to consider one, labelled BUSINESS. Each one processes the request and sends the result to WEB for transmission. Currently there are three of these, each processing one third of all the transactions.
* Data analysis: Business data is extracted from the business computers each night and added to the large database managed by the WAREHOUSE machine. A small group of analysts use this machine to look for trends and customer characteristics. They access the machine via WEB so they can work from home. WEB passes SQL queries to WAREHOUSE, which prepares the required data and returns it to WEB, where it is formatted for transmission and sent.

Note that this configuration is simpler than would be found in real life. All single disks would be replaced by a Storage Area Network (SAN) or Cloud, to simplify maintenance and upgrading.

Transactions (abbreviated to Tx in the table) and queries pass through WEB twice, once when they are received ("up"), and once when the result is returned to the Internet ("down"). Resource requirements are as follows:

Workload component	CPU secs	Disk accesses	Machine
Tx up	0.01	–	WEB
Tx processing	0.12	10	BUSINESS
Tx down	0.015	3	WEB
Queries up	0.008	–	WEB
Queries processing	2.6	2000	WAREHOUSE
Queries down	0.03	10	WEB

Expected transaction arrival rate: 12 per second
Average disk access time for computers WEB and BUSINESS: 11 msec
Average disk access time for WAREHOUSE: 14 msec
Average think time for analysts: 30 sec
Number of analysts: 10
Initial number of BUSINESS machines: 3

Step 1: Feasibility (8 marks)

Use bounds analysis to show that the proposed configuration can process the above workload. To do this, you need to estimate the transaction rate for queries so you can compute the utilisation of WEB. As a first approximation, compute the throughput bound for the WAREHOUSE machine as a closed system, and use that throughput to compute the resource requirements of queries in WEB. Will the resulting estimate of resource usage in WEB be higher or lower than actual? Explain your answer.

Step 2: Simple response times (9 marks)

Estimate the overall response times for both workloads, assuming transactions form an open workload in WEB and BUSINESS, while queries form a closed workload on WAREHOUSE. Do the queries have much effect on the transactions? Explain your answer.

Step 3: Modify WAREHOUSE system (8 marks)

The analysts' representative in the design team complains that the response times for query processing are excessive. The database designers rework the indexes and record sizes, and decide that if main memory is increased so first level indexes can be kept in memory, the number of accesses per query can be reduced to 800, although the time per disk access will increase to 0.016 msec sec, and the CPU time per query will increase to 4.1 sec. Compute the new response time for queries. If these changes are put into the system, what effect will they have on response times for transactions?

Step 4: Coping with growth (5 marks)

Marketing expect the number of transactions to grow, so the present installation of three computers of type BUSINESS will not be adequate for long. For the first year of operation, marketing expect the transaction rate to increase at 7% per month, compound. Using the empirical rule that no device should exceed 75% average utilization, calculate whether extra

BUSINESS computers will be required during the year, and give the approximate times of installation.

Submission Instructions:

1.   The submission deadline is 23:59:59 ~~Friday 6th~~ Tuesday 10th April 2012 .
2.   The total marks for this part is 30 marks. Please note that final part will be released in due course.
3.   You are required to submit a written report detailing the work that you have done. It must be in Acrobat "pdf" format. It must be called Ass1PartB.pdf". : Please note that this is not a programming exercise and you need to present intermediate working results in tabular form. When you are ready to submit, type 9334 at the bash prompt and then the command: give cs9334 Assignment Ass1PartB.pdf (please check notice board for any change in these instructions)
4.   Please note that the system will only accept "Ass1PartB.pdf " as the filename for submission. Also, note that the total size of your submission should be smaller than 2 MBytes. If you still have difficulty, email cs9334@cse.unsw.edu.au for instructions.
5.   You can submit multiple times before the deadline. The latest submission overrides the earlier submissions, so make sure you submit the correct file. Do not leave until the last moment to submit, as there may be technical or communications error and you will not have time to rectify that.

Late Submission Penalty: Late penalty will be applied as follows:
o   1 day after deadline: 10% reduction
o   2 days after deadline: 20% reduction
o   3 days after deadline: 30% reduction
o   4 days after deadline: 40% reduction
o   5 or more days late: NOT accepted

NOTE: The above penalty is applied to your final total. For example, if you submit your assignment 1 day late and your score on the assignment is 20, then your final mark will be 30 – 30*(10% penalty) = 27.

Plagiarism: You are to write all of the code for this assignment and produce the report yourself.   The LIC will decide on appropriate penalty for detected cases of plagiarism. The most likely penalty would be to reduce the assignment mark to ZERO. We are aware that a lot of learning takes place in student conversations, and don't wish to discourage those. However, it is important, for both those helping others and those being helped, not to provide/accept any programming language code in writing, as this is apt to be used exactly as is, and lead to plagiarism penalties for both the supplier and the copier of the codes. Write something on a piece of paper, by all means, but tear it up/take it away when the discussion is over.