

Вариант 13.

№ 1

Интернет магазин цифровой дистрибуции имеет сервер, который используется покупателями для загрузки купленного ими ПО в регионе “Россия”.

Всего у магазина 20 000 000 клиентов, ежедневно сервер обслуживает число транзакций, которое равно 2% от числа клиентов. В среднем, длительность одного соединения составляет 30 минут. Сервер может одновременно поддерживать не более 30 000 соединений. Допустимое среднее время ожидания составляет 3 секунды.

Перед новым годом, 10 декабря выходит игра “Biorunk 3077”, которую предзаказали 25% клиентов. Предполагается, что в день выхода её начнут загружать 85% из них (моменты начала загрузки равномерно распределены в течение дня).

Помимо этого аудитория магазина возрастает на 2% каждые 3 месяца.

Нагрузку на сервер можно классифицировать следующим образом:

1. Запросы на загрузку файлов в обычных условиях.
2. Запросы на загрузку файлов при выходе нового ПО.

Первую группу составляют запросы, которые приходят каждый день вне зависимости от происходящих событий. Вторую группу составляют запросы, которые вызывают резкое повышение нагрузки на один день при выходе нового ПО.

Требуется определить:

1. Выдержит ли сервер увеличившуюся нагрузку в день выхода игры?
2. Будет ли справляться сервер с возрастающей в течение года нагрузкой?

Если ответ на первый вопрос “Нет”, то требуется провести модификацию сервера. Возможны следующие ей варианты:

1. Аренда дополнительных серверов и распределение части нагрузки на них.
2. Усиление сжатия установочных файлов, что сократит время загрузки на 15%.

Если ответ на второй вопрос “Нет”, то необходимо определить момент времени, когда сервер потребует усовершенствования. Возможны следующие варианты модификации сервера:

1. Покупка дополнительного сервера.
2. Усовершенствование существующего сервера, что увеличит максимальное число соединений в 2 раза.
3. Модернизация канала связи, что сократит время загрузки на 10%.

№ 2

Информационную систему можно описать как центральный сервер, обрабатывающий запросы пользователей, очередь клиентов, ожидающих обработки и клиентов, находящихся в обработке. Активным устройством является центральный сервер, его основная характеристика – число одновременных подключений, равное 30 000.

Нагрузка на сервер характеризуется следующими параметрам:

1. Интенсивность поступления заявок:
 - a. Обычная: 4,63 запроса в секунду.
 - b. Пиковая: 53.82 запроса в секунду.
2. Время обслуживания заявок, в среднем требует 1800 секунд.

Главная учитываемая метрика производительности – время ожидания, которое не должно превышать 3 секунд.

Если предположить, что время между поступлениями запросов независимо и распределено по экспоненциальному закону распределения, времена обслуживания независимы и распределены по экспоненциальному закону распределения, то для описания этой системы можно использовать модель вида М/М/с, с параметрами $\lambda = 4.63, \mu = \frac{1}{1800}, c = 3000$.

Административный план развития системы будет направлен на увеличении производительности системы с течением времени. С учётом увеличения количества клиентов с течение времени, интенсивность будет расти на примерно 2% каждые 3 месяца. Поэтому необходимо проводить модернизацию сервера одни из следующих способов (или их комбинацией):

1. Покупка дополнительного сервера.
2. Усовершенствование существующего сервера.
3. Модернизация канала связи.

Следует учесть, что без модернизации сервер не сможет справиться с пиковой нагрузкой, следовательно, перед 10 декабря его необходимо модернизировать. Сделать это можно комбинацией следующих способов:

1. Аренда дополнительных серверов и распределение части нагрузки на них.
2. Усиление сжатия установочных файлов.

№ 3

Основными параметрами модели производительности будут являться:

1. Параметры нагрузки:
 - a. Интенсивность поступления заявок в систему.
 - b. Среднее время ожидания.
 - c. Загрузка сервера (при приближении к 100% время ожидания может увеличиваться практически неограниченно).
 - d. Среднее время обработки заявки.
2. Параметры ПО:
 - a. Максимальное количество обрабатываемых одновременно клиентов (необходимо при построении модели, можем увеличить при модернизации, чтобы уменьшить время ожидания и загрузку).

Для сбора статистики можно использовать программный монитор (можем замерить все необходимые характеристики, не используя специальных аппаратных средств) в режиме слежения (поскольку интенсивность прихода заявок невелика).

№ 4

Данные	Решение								
Исходная задача									
Модель М/М/1 $\lambda = 30 \text{ tps}$ $\mu = \frac{1}{0.02} = 50 \text{ tps}$	$U = \rho = \frac{\lambda}{\mu} = \frac{30}{50} = 0.6 = 60\%$ - загрузка $T = \mu\rho = \lambda = 30 \text{ tps}$ – пропускная способность $N_q = \frac{\lambda}{\mu-\lambda} = \frac{30}{20} = 1.5 \text{ t}$ – средняя длина очереди $R = \frac{1}{\mu-\lambda} = \frac{1}{20} = 0.05 \text{ с.}$ – среднее время отклика								
а									
$N = 13$ $\lambda_a = \lambda \frac{N}{10} = 30 \frac{13}{10} = 39 \text{ tps}$ $\mu_a = \mu \left(\frac{N}{10} + 1 \right) = 50 \frac{23}{10} = 115 \text{ tps}$	$U_a = \frac{\lambda_a}{\mu_a} = \frac{39}{115} \approx 0.33913 = 33.913\%$ $T_a = \lambda_a = 39 \text{ tps}$ $N_{qa} = \frac{\lambda_a}{\mu_a - \lambda_a} = \frac{39}{115 - 39} = \frac{39}{76} \approx 0.513 \text{ t}$ $R_a = \frac{1}{\mu_a - \lambda_a} = \frac{1}{115 - 39} = \frac{1}{76} \approx 0.013 \text{ с.}$ Ответ: <table><tr><td>ΔU</td><td>ΔT</td><td>ΔN_q</td><td>ΔR</td></tr><tr><td>-43.478%</td><td>+30%</td><td>-65.8%</td><td>-74%</td></tr></table>	ΔU	ΔT	ΔN_q	ΔR	-43.478%	+30%	-65.8%	-74%
ΔU	ΔT	ΔN_q	ΔR						
-43.478%	+30%	-65.8%	-74%						
б									
$N = 13$ $R_b \leq \frac{1}{20} \text{ с.}$ $\lambda_b = \lambda \left(1 - \frac{3N}{100} \right) = 30 \frac{61}{100} = 18.3 \text{ tps}$	$R_b = \frac{1}{\mu_b - \lambda_b} = \frac{1}{\mu_b - 18.3} \leq \frac{1}{20}$ $\mu_b - 18.3 \geq 20$ $\mu_b \geq 38.3$ Ответ: Мощность процессора должна быть не ниже 38.3 tps.								
с									
$N = 13$ $R_c > \frac{1}{20} \text{ с.}$ $\mu_c = \mu \left(1 + \frac{4N}{100} \right) = 50 \frac{152}{100} = 76 \text{ tps}$	$R_c = \frac{1}{\mu_c - \lambda_c} = \frac{1}{76 - \lambda_c} > \frac{1}{20}$ $76 - \lambda_c < 20$ $\lambda_c > 56$ Ответ: Среднее время отклика превысит 0.05 с., если интенсивность входного потока превысит 56 tps (или увеличится более чем на 86.6(6)%).								

№ 5.

Данные	Решение
Исходная задача	
$\mu_c = 6 \text{ tpm}$ $\mu_s = 2 \text{ tpm}$ $\mu_f = 3 \text{ tpm}$ $p = \frac{1}{2}$	Система уравнений локального баланса: $P_{110} = U_f P_{200}$ $P_{101} = U_s P_{200}$ $P_{020} = U_f U_f P_{200}$ $P_{011} = U_s U_f P_{200}$ $P_{002} = U_s U_s P_{200}$ $P_{200} = (1 + U_f + U_s + U_f U_f + U_f U_s + U_s U_s)^{-1}$ где: $U_f = \frac{\mu_c p}{\mu_f}, U_s = \frac{\mu_c (1 - p)}{\mu_s}$ Найдём распределение состояний системы:

	$U_f = 1, U_s = 1.5$ Распределение:					
	P_{110}	P_{101}	P_{020}	P_{011}	P_{002}	P_{200}
	0.1212	0.1818	0.1212	0.1818	0.2727	0.1212
	Метрики производительности:					
		Процессор $i = c$	Быстрый диск $i = f$	Медленный диск $i = s$		
	Загрузка U^i	$P_{200} + P_{110} + P_{101}$	$P_{110} + P_{020} + P_{011}$	$P_{101} + P_{011} + P_{002}$		
	Пропускная способность T^i	$\mu_c U^c$	$\mu_f U^f$	$\mu_s U^s$		
	Средняя длина очереди N_q^i	$2P_{200} + 1P_{110} + 1P_{101}$	$1P_{110} + 2P_{020} + 1P_{011}$	$1P_{101} + 1P_{011} + 2P_{002}$		
	Среднее время оклика R^i	$\frac{N_q^c}{T^c}$	$\frac{N_q^f}{T^f}$	$\frac{N_q^s}{T^s}$		
	Числовые значения:					
	Процессор $i = c$	Быстрый диск $i = f$	Медленный диск $i = s$			
Загрузка U^i	42.42%	42.42%	63.63%			
Пропускная способность T^i	2.545 tpm	1.273 tpm	1.273 tpm			
Средняя длина очереди N_q^i	0.5454 t.	0.5454 t.	0.909 t.			
Среднее время оклика R^i	0.2143 min.	0.4284 min.	0.7141 min.			
В						
$N = 13$ $\mu_c = 6 \left(1 + \frac{4N}{100} \right) = 9.12 \text{ tpm}$ $\mu_s = 2 \text{ tpm}$ $\mu_f = 3 \text{ tpm}$ $p = \frac{1}{2} + \frac{2N}{100} = 0.76$ Замечание: не до конца понятно условие о перемещении файлов с медленного диска на быстрый. Я предположил, что имеется в виду процент от всех файлов на обоих дисках.	Система уравнений локального баланса:					
	$P_{110} = U_f P_{200}$ $P_{101} = U_s P_{200}$ $P_{020} = U_f U_f P_{200}$ $P_{011} = U_s U_f P_{200}$ $P_{002} = U_s U_s P_{200}$ $P_{200} = (1 + U_f + U_s + U_f U_f + U_f U_s + U_s U_s)^{-1}$ где:					
	$U_f = \frac{\mu_c p}{\mu_f}, U_s = \frac{\mu_c (1 - p)}{\mu_s}$					
	Найдём распределение состояний системы:					
	$U_f = 2.3104, U_s = 1.0944$					
	$P_{200} = (1 + 2.3104 + 1.0944 + 5.3379482 + 2.5285018 + 1.1977114)^{-1} = \frac{1}{13.468961} = 0.0742448$					
	Распределение:					
	P_{110}	P_{101}	P_{020}	P_{011}	P_{002}	P_{200}
	0.1715352	0.0812535	0.3963149	0.1877281	0.0889238	0.0742448
	Метрики производительности:					
	Процессор $i = c$	Быстрый диск $i = f$	Медленный диск $i = s$			
Загрузка U^i	$P_{200} + P_{110} + P_{101}$	$P_{110} + P_{020} + P_{011}$	$P_{101} + P_{011} + P_{002}$			
Пропускная способность T^i	$\mu_c U^c$	$\mu_f U^f$	$\mu_s U^s$			

Средняя длина очереди N_q^i	$2P_{200} + 1P_{110} + 1P_{101}$	$1P_{110} + 2P_{020} + 1P_{011}$	$1P_{101} + 1P_{011} + 2P_{002}$
Среднее время отклика R^i	$\frac{N_q^c}{T^c}$	$\frac{N_q^f}{T^f}$	$\frac{N_q^s}{T^s}$
Числовые значения:			
	Процессор $i = c$	Быстрый диск $i = f$	Медленный диск $i = s$
Загрузка U^i	32.70335%	75.55782%	35.79054%
Пропускная способность T^i	2.98254552 tpm	2.2667346 tpm	0.7158108 tpm
Средняя длина очереди N_q^i	0.4012783 t.	1.1518931 t.	0.4468292 t.
Среднее время отклика R^i	0.134542222 min.	0.508172902 min.	0.624228078 min.
Процентные изменения по сравнению с исходной системой:			
	Процессор $i = c$	Быстрый диск $i = f$	Медленный диск $i = s$
Загрузка U^i	-22,90582273 %	+78,11838755 %	-43,75209807 %
Пропускная способность T^i	+17,19235835 %	+78,06241948 %	-43,76977219 %
Средняя длина очереди N_q^i	-26,42495416 %	+111,2015218 %	-50,84387239 %
Среднее время отклика R^i	-37,21781544 %	+18,62112561 %	-12,58534125 %
Вывод: такая модификация имеет мало смысла. При ней узким местом становится быстрый диск, что выражается в значительном повышении размера очереди и его загрузки. Стоит, однако, заметить, что разница во времени отклика между быстрым и медленным диском почти сравнялась, а у процессора время отклика заметно уменьшилось.			