

УПРАВЛЕНИЕ ДОСТУПОМ ПОЛЬЗОВАТЕЛЕЙ К РЕСУРСАМ ЦЕНТРА ОБРАБОТКИ ДАННЫХ (ЦОД)

Медведникова Ю.С., студентка; Турупалов В.В., к.т.н., проф.

(Донецкий национальный технический университет, г.Донецк, Украина)

Поскольку центр обработки данных это целый комплекс инженерных и IT-систем, который является неотъемлемой частью множества телекоммуникационных структур, он должен обеспечить единый информационный ресурс с гарантированными уровнями достоверности, доступности и безопасности данных.

Цель статьи – предложить один из способов контроля объемов запросов, которые поступают в ЦОД, и управления распределением их на сервера.

Пусть существуют разные типы запросов, за которыми стоит определенный ресурс. Запросы имеют свою последовательность, время обдумывания, то есть время между ответом web-сервера и новым запросом в пределах одной сессии.

Здесь главной целью контроллера будет предоставление необходимых ресурсов для принятия запросов, на период времени достаточный для их обработки; при этом необходимо учитывать запросы от других групп пользователей и соответственно перераспределение между ними сетевого ресурса (в нашем случае серверов). Каждый новый запрос сессии будет поступать в очередь на виртуальный сервер, где после паузы снова обращается к системе в качестве нового запроса. Таким образом имеем $k = 1, \dots, K$ классов запросов, J сессий и ресурсов $j = 1, \dots, J$. Користувацькі сесії типу j надходять у систему з інтенсивністю l_j й починаються із запиту класу k , де $j = 1, \dots, J$, пауза t_j . [1].

Вероятность поступления повторного запроса от класса k (k') $P'_{kk'}$, а выхода из системы с завершением сессии:

$$1 - \sum_{k'=1}^K P'_{kk'} \quad (1)$$

Матрица размерностью $K \times K$ $P^j = [p_{kk'}^j]$ - это матрица вероятностей переходов запросов по ресурсам ЦОД (рис.).

Пусть время обслуживания для класса k одним сервером кластера i это независимая случайная величина T_i^k с математическим ожиданием:

$$t_i^k = M[T_i^k] \quad (2)$$

Все серверы кластера i вместе обрабатывают запросы как единый сервер, времяобработки t_i^k / N_i . Тогда номинальная нагрузка кластера i равна:

$$q_i = \sum_{k=1}^K \lambda_i^k t_i^k \quad (3)$$

Тогда при количестве серверов N_i нагрузка на один сервер кластера i будет:

$$p_i = q_i / N_i. \quad (4)$$

Для того, чтобы система была стационарной, необходимо выполнение условия $N_i > q_i$, где $\vec{N} = (N_1, N_2, \dots, N_n)$ это вектор распределения ресурсов.

Пусть m_i^k это количество запросов класса k , обрабатываемых кластером i в стационарном режиме, тогда:

$$m_i = \sum_{k=1}^K m_i^k \quad (5)$$

количество классов запросов соответственно. Общее количество запросов k в системе t^k . Запрос принадлежит классу k с вероятностью:

$$p_i^k / p_i. \quad (6)$$

Где:

$$p_i^k = \frac{\lambda_i^k t_i^k}{N_i}, \quad (7)$$

$$p_i = \sum_{k=1}^K p_i^k. \quad (8)$$

Находим:

$$M[m_i] = \frac{p_i}{1 - p_i}; \quad (9)$$

$$M[m_i^k] = \frac{p_i^k}{1 - p_i}. \quad (10)$$

Теперь можно найти время ответа на запрос класса k :

$$t^k = M[t^k] = \sum_{i=1}^n \frac{\lambda_i^k}{\lambda^k} \frac{t_i^k}{N_i - q_i} \quad (11)$$

где n – количество кластеров серверного комплекса;

t_i^k – среднее время ответа на запрос класса k сервером кластера i . [3]

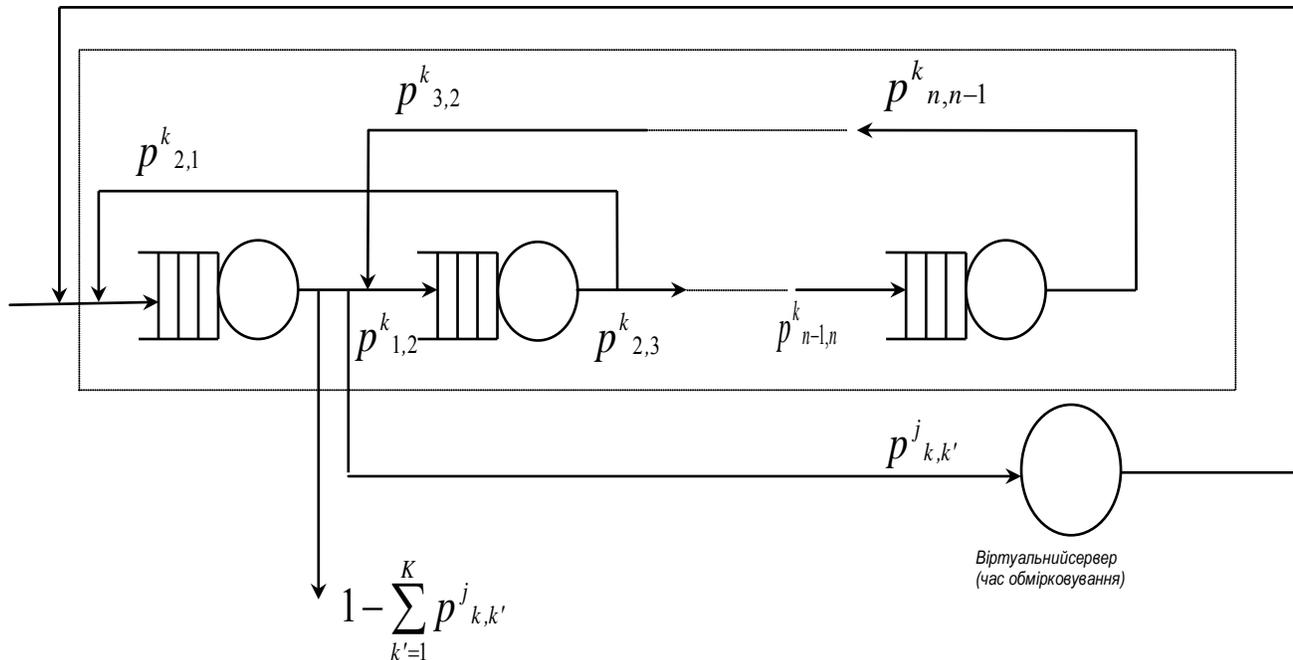


Рисунок 1 - Модель серверного комплексу ЦОДс учетом классов запросов как системы массового обслуживания.

Заключение. Таким образом найден ключевой параметр – время ответа на запрос, исходя из которого главный контроллер будет принимать решение об оптимальности выделения системных ресурсов и о необходимости их перераспределения. Главным образом это будет зависеть от важности поступающих запросов и необходимой срочности их обслуживания.

Перечень ссылок

1. Стаття «Побудова ЦОД»: <http://www.sterling.zp.ua/serv/net/cod.htm>
2. Центри обробки даних: стандарти в дії: <http://citcity.ru/14587/>
3. Стаття «Транспортні технології сучасного ЦОД»: <http://www.publish.ru/lan/240740/text/5577358.html>