УДК 681.324

АНАЛИЗ ЭФФЕКТИВНОСТИ КЛАСТЕРНЫХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ

Фельдман Л.П., Михайлова Т.В.
Кафедра ПМиИ ДонНТУ
E-mail: feldman@r5.dgtu.donetsk.ua

Фельдман Л.П., Михайлова Т.В. Аналіз эфективності кластерних систем за застосуванням ймовірних моделей. В статті розглянуто ймовірна модель кластерної системи з загальним застосуванням дискового простору.

Упрощенная модель кластера с совместным использованием дискового пространства [1,2] приведена на рис.1. Входной узел (маршрутизатор или управляющий модуль) равномерно распределяет между серверами приложений (выполняющими одинаковые приложения) задачи. Количество серверов – приложений - *N1*. Каждый из них может обратиться к данным, распложенным на дисках, количество которых *N2*. Ввиду ограниченных вычислительных возможностей будем считать, что количество задач, обрабатываемое такой вычислительной системой не более M [3].

Допустим, задачи, обрабатываемые на таком кластере, однородные и имеют следующие характеристики: p_{12} – вероятность запроса к одному из *N1* серверов, p_{23} - вероятность запроса к одному из *N2* дисков, p_{21} – вероятность завершения обслуживания одним из *N1* серверов, p_{10} – вероятность завершения обслуживания задачи, q_0 – вероятность появления новой задачи.

Для вычислительных ресурсов введем такие показатели: q_i – вероятность завершения обслуживания задачи на входном узле, сервере приложений или дисковом массиве, соответственно, $(i=\overline{1,3})$, r_i – вероятность продолжения обслуживания задачи на входном узле, сервере приложений или дисковом массиве, соответственно, $(i=\overline{1,3})$, $r_i=1-q_i$.

Для построения дискретной модели [4] вычислительной системы, приведенной на рис.1, определим все возможные состояния. За состояние системы примем размещение М заявок по трем узлам $\overline{m}=(m_1,m_2,m_3)$, где m_i – количество задач в i-ом узле. Обозначим множество состояний через $S=\{(m_1,m_2,m_3)\,|\,\sum_{i=1}^{N} m_i = M\}$.

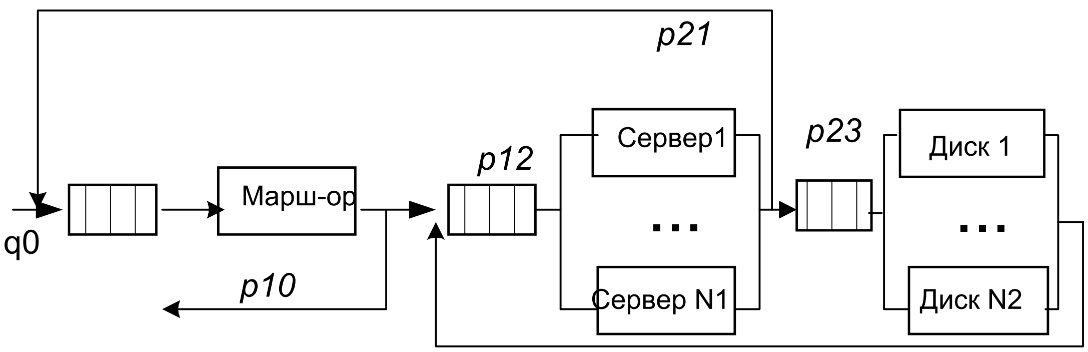

Рисунок 1 – Структурная схема Марковской модели кластера с совместным использованием дискового пространства

Число состояний системы вычисляется по формуле:

$$L = \sum_{j=0}^{M} C_{j+N-1}^{N-1}.$$

Вектор $\overline{k} = (k_1, k_2, k_3)$ определяет количество устройств в каждом узле ВС. Определим матрицу переходных вероятностей P .

Рассмотрим переход из произвольного состояния $\overline{m}$ (в котором обрабатывается $j = \sum_{s=1}^{N} m_S$ задач) в состояние $\overline{m'}$ (обрабатывается $j' = \sum_{s=1}^{N} m'_S$ задач). Так как за один такт может поступить на обработку в ВС или покинуть ее одна задача, то разница $\Delta = j - j'$ должна удовлетворять условию

$$\Delta \in \{1, 0, -1\}. \tag{1}$$

Введем вектор $\overline{\alpha}$, s-я компонента которого определяет число

$$\alpha_S = \min(m_S, k_S), \quad s = \overline{1, N}, \tag{2}$$

загруженных устройств в s-м узле.

Вычислим вектор $\bar{i}$

$$\bar{i} = \overline{m} - \overline{m'} = (i_1, i_2, \ldots, i_N), \tag{3}$$

каждая компонента i_S которого представляет изменение числа задач в s-ом узле. При любом переходе из $\overline{m}$ число задач, обслуженных s-м узлом обработки, не может быть больше α_s, т.е.

$$i_S \le \alpha_S, \quad s = \overline{1, N}. \tag{4}$$

Определим множество J номеров узлов обработки:

$$J = \{S \big|_{i_S < 0}\}, \quad s = \overline{1, N}, s \ne 2. \tag{5}$$

Если $J \neq \varnothing$, то величины $\gamma_S = | \ i_S | \ (S \in J)$ определяют минимально возможное число программ, которые поступают к тем узлам обработки, номера s которых принадлежат множеству *J*. Число программ, поступивших в один из таких узлов *s* равно $|i_S|$. Количество поступивших задач к распределительному узлу от серверов не может превышать количества работающих серверов (за исключением случая, когда появляется новая задача или покидает ВС обслуженная задача, о чем свидетельствует величина δ)

$$|i_1| + \delta \leq \alpha_2, \tag{6}$$

где $\delta = \begin{cases} -1, & если\, \Delta = \pm 1, \\ 0, & если\, \Delta = 0. \end{cases}$

Аналогично, количество поступивших задач к дискам от серверов не может превышать количества работающих серверов

$$|i_3| \leq \alpha_2. \tag{7}$$

От серверного узла задачи поступают и к пользователям и к дискам, поэтому общее количество ушедших задач не превышает количества работающих серверов

$$|i_1| + \delta + |i_3| \leq \alpha_2. \tag{8}$$

Количество поступивших задач к серверам не превышает количества работающих клиентов и дисков

$$|i_2| \leq \alpha_1 + \alpha_3. \tag{9}$$

Чтобы в общем виде описать алгоритм построения матрицы переходных вероятностей, надо откорректировать вектор $\bar{i}$ следующим образом: $i_1 = i_1 + \delta$, а остальные компоненты оставить без изменения. Эта операция позволит не учитывать пришедшую и покинувшую в рассматриваемом такте задачи при распределении задач по узлам ВС.

Переход $\overline{m} \to \overline{m'}$ возможен при выполнении (1),(8) и (9).

Распределение программ, обслуженных в серверном узле, по узлам системы определим *N*-мерным вектором $\overline{\mathrm{Z}}$, компонента которого равна количеству задач, обслуженных серверами и распределенных по остальным узлам системы: $\mathrm{Z}_2 = \sum_{s=1, s\neq 2}^{N} Z_S$.

Назовем этот вектор индикатором распределения задач, обслуженных серверным узлом по остальным узлам системы. Распределение задач, обслуженных узлами системы, зададим *N*-мерным вектором $\overline{Y}$, компонента y_s $(s = \overline{1, N}, s \neq 2)$ которого равна числу задач, обслуженных *s*-м узлом, а компонента $y_2 = \sum_{s=1, s\neq 2}^{N} y_s$ равна общему числу задач, поступивших в серверный узел от всех остальных. Вектор $\overline{Z}$ назовем индикатором распределения задач, обслуженных узлами системы.

Найдем множества допустимых индикаторов распределения задач $\Im = \{\overline{Z}\}$ и $\Re = \{\overline{Y}\}$ для произвольного перехода $\overline{m} \to \overline{m'}$. Число программ, которые могут поступить от серверного узла, заключено в следующих границах:

$$\gamma_{2} \le z_2 \le \alpha_2 . \qquad (10)$$

Для определения возможности распределения этих задач по каналам введем вектор $\overline{\beta}$, компоненты которого определяются формулами $\beta_2 = \alpha_2 - \gamma_2$;

$$\beta_S = \begin{cases} \alpha_S, & \text{если } i_S \le 0; \\ \alpha_S - i_S, & \text{если } i_S > 0, \quad s = \overline{1, N} \quad, s \ne 2. \end{cases}$$

Компонента β_2, равна максимально возможному числу программ, которые дополнительно, сверх γ_2 могут быть обслужены серверным узлом. Аналогично, β_s равна максимально возможному числу программ, которые дополнительно сверх обязательно поступивших в s-й узел обработки, могут еще поступить в него, если допустить, что в s-й узел поступает больше задач, тогда при переходе $\overline{m} \to \overline{m'}$ узел s должен обслужить более α_s задач, что невозможно.

$$\gamma_S = \begin{cases} |i_S|, & \text{если } i_S \le 0; \\ 0, & \text{если } i_S > 0, \quad s = \overline{1, N}. \end{cases} \qquad (11)$$

Вектор $\overline{\gamma}$ определяет для данного перехода распределение по каналам минимального числа γ_2 программ, обслуженных процессорным узлом. Отсюда следует, что индикатор распределения программ $\overline{Z}$ будет допустим для перехода $\overline{m} \to \overline{m'}$, если выполняются условия

$$\gamma_{S} \le z_S \le \gamma_S + \beta_S \qquad (12)$$

Чтобы сделать дальнейшие выводы надо в векторах $\overline{\gamma}$, $\overline{\beta}$, $\overline{i}$, $\overline{\alpha}$, $\overline{Z}$, $\overline{Y}$ переставить местами первую и вторую компоненты.

Определим множество $\Im$ допустимых для перехода $\overline{m} \to \overline{m'}$ индикаторов распределения программ для общего случая. Эта задача сведется к задаче размещения l программ, где $0 \le l \le \beta_s$,по $N - 1$ каналу, если в любом s-м из них не может быть более β_s программ. Множество $\Im$ является объединением всех таких размещении, полученных для каждого целого $l = \overline{0, \beta_S}$.

Обозначим через $B(l,n)$ множество размещений l задач по n узлам, если в любом из s узлов может быть не более β_s программ и соблюдается условие $0 \le l \le \sum_{i=1}^{n} \beta_{s+1}$.

Элементы множества *B* могут быть получены по формуле :

$$B(l,n) = \bigcup_{j \in \{j_0, j_M\}} B(l-j, n-1) \times \{j\}, \qquad (13)$$

где $0 \le l \le \sum_{s=1}^{n} \beta_{s+1}, \quad j_0 = \max(0, l - \sum_{s=1}^{n-1} \beta_{s+1}), \quad j_M = \min(l, \beta_{n+1}).$

Множество всевозможных размещений $l = \overline{0, \beta_1}$ задач по N-1 каналу системы с учетом ограничений β_s получим как объединение множеств, определяемых формулой

$$D(\beta, N) = \bigcup_{l = \{0, \beta_1\}} \{l\} \times B(l, N-1). \qquad (14)$$

Элемент $\overline{\partial} \in D(\beta, N)$ является N-мерным вектором, первая компонента которого равна $d_1 = \sum_{s=2}^{N} d_s = l$ числу задач, распределенных по N - 1 узлу системы с учетом ограничений, определяемых (12). Все допустимые для перехода индикаторы распределения программ $\overline{z}$ равны : $\overline{z} = \overline{\partial} + \overline{\gamma}.$

Для каждого допустимого на переходе $\overline{m} \to \overline{m'}$ размещения $\overline{\partial} \in D(\beta_1, N)$, определяющего распределение по каналам системы программ, поступивших от процессоров, найдем соответствующий допустимый на том переходе индикатор $\underline{y}$ числа задач, обслуженных каналами по формуле $\overline{y} = \overline{\partial} + \overline{e}$,

где $\overline{e_s} = \begin{cases} i_s, & если\, i_s > 0, \\ 0, & если\, i_s \le 0, \end{cases} \qquad e_1 = \sum_{s=2}^{N} e_s, s = \overline{2, N}.$

Для того, чтобы отобразить реальную структуру моделируемой ВС, переставим опять первую и вторую компоненты векторов $\overline{\gamma}$, $\overline{\beta}$, $\overline{i}$, $\overline{\alpha}$, $\overline{Z}$, $\overline{Y}$. Рассчитаем матрицу переходных вероятностей.

Рассмотрим случай $i = \overline{1,k}$, т.е. количество задач, обрабатываемых в ВС при переходе $\overline{m} \to \overline{m'}$, остается неизменным.

Найдем вероятность для события, определяемого произвольным допустимым для перехода $\overline{m} \to \overline{m'}$ индикатором распределения программ $\overline{z} = (z_1, z_2, ..., z_N)$. Значение компоненты z_2 равно числу задач, поступивших в узлы от α_2 серверов, работающих в состоянии $\overline{m}$. Число различных комбинаций, когда из α_2 работающих процессоров, z_2 завершают обслуживание своих программ, равно $C_{\alpha_2}^{z_2}$. Для любой такой комбинации, содержащей номера z_2-го процессора, число различных размещений z_S, обслуженных программ, когда в s-й канал поступают программы от z_s процессоров, равно числу перестановок из z_2, с повторениями, т.е.

$$P(z_1,...,z_N) = \frac{P(z_2)}{P(z_1)P(z_3)...P(z_N)}, \qquad (15)$$

где $z_2 = \sum_{s=1,s\neq 2}^{N} z_s$. Любой из остальных серверов, обслуживающих программы, может продолжать ее обслуживание. Вероятность такого исхода равна величине, определяемой формулой

$$P(\overline{z}) = \prod_{s=1,s\neq 2}^{N} (q_2 p_{2s})^{z_s} r_2^{\alpha_2 - z_2}. \qquad (16)$$

где $r_i = 1 - q_i$, и одинакова для всех таких исходов, то вероятность наступления события, определяемого индикатором $\overline{z}$, будет равна

$$P_r(\overline{z}) = C_{\alpha_2}^{z_2} P(\overline{z}) P(z_1,...,z_N). \qquad (17)$$

Аналогично находится вероятность события, определяемого индикатором $\overline{y}$ распределения обслуженных задач первым и третьим узлами. Надо учесть, что первый узел может покинуть задача и вместо нее поступить новая в случае, когда $y_1 = 0$. Вероятность такого события $(q_1 p_{10} q_0 + r_1 r_0)(1 - y_1)$. Если $y_1 = 1$, задача завершает обслуживание в первом узле и переходит во второй. Вероятность такого события $(q_1 p_{12} r_0) y_1$. Общая вероятность события, определяемого индикатором $\overline{y}$ равна

$$P_r(\overline{y}) = \begin{cases} ((q_1 p_{10} q_0 + r_1 r_0)(1 - y_1) + q_1 p_{12} r_0 y_1) \prod_{s=3}^{N} C_{\alpha_S}^{y_S} (q_S p_{S2})^{y_S} r_S^{\alpha_S - y_S}, \\ \qquad\qquad j < M, \\ ((q_1 p_{10} + r_1)(1 - y_1) + q_1 p_{12} y_1 \prod_{s=3}^{N} C_{\alpha_S}^{y_S} (q_S p_{S2})^{y_S} r_S^{\alpha_S - y_S}, \quad j = M. \end{cases}$$

(18)

Таким образом, вероятность события, определяемого индикатором распределения программ $\overline{z}$ и соответствующим ему индикатором обслуженных каналами программ $\overline{y}$ равна произведению вероятностей

$$P_r(\overline{z},\overline{y}) = P_r(\overline{z})P_r(\overline{y}). \tag{19}$$

Вероятность перехода $\overline{m} \to \overline{m'}$ определяется теперь как сумма вероятностей событий для всех допустимых на этом переходе индикаторов $\overline{z}$ и $\overline{y}$, вычисляемых по формуле (19), т.е.

$$P_r(\overline{m} \to \overline{m'}) = \sum P_r(\overline{z},\overline{y}) \tag{20}$$

для всех $\overline{z} \in \Im(\overline{m} \to \overline{m'})$ и соответствующих $\overline{y} \in \Re(\overline{m} \to \overline{m'})$.

Рассмотрим случай $\Delta = -1$, т.е. количество задач, обрабатываемых в ВС при переходе $\overline{m} \to \overline{m'}$, увеличивается. Если m_1 =0, то вероятность перехода $\overline{m} \to \overline{m'}$ равна нулю, если m_1 >0, то вероятность – не равна нулю. Рассчитаем ее. Вероятности $P(\overline{z})$ и $P_r(\overline{z})$ вычисляются по формулам (17) и (18). А вероятность $P_r(\overline{y})$ вычисляется по следующей формуле

$$P_r(\overline{y}) = \begin{cases} q_0(r_1(1-y_1)+(q_1p_{12})y_1)\prod\limits_{s=3}^{N} C_{\alpha_S}^{y_S}(q_Sp_{S2})^{y_S}\, r_S^{\alpha_S - y_S}, & j < M, \\ (r_1(1-y_1)+(q_1p_{12})y_1)\prod\limits_{s=3}^{N} C_{\alpha_S}^{y_S}(q_Sp_{S2})^{y_S}\, r_S^{\alpha_S - y_S}, & j = M. \end{cases} \tag{21}$$

Рассмотрим случай $\Delta = 1$. В этом случае количество задач, обрабатываемых в ВС при переходе $\overline{m} \to \overline{m'}$, уменьшается. Если m_1 =0, то вероятность перехода $\overline{m} \to \overline{m'}$ равна нулю, если m_1 >0, то вероятность – ненулевая, вычислим ее. Вероятности $P(\overline{z})$ и $P_r(\overline{z})$ вычисляются по формулам (17) и (18). А вероятность $P_r(\overline{y})$ вычисляется по следующей формуле

$$P_r(\overline{y}) = \begin{cases} r_0q_1p_{10}\prod\limits_{s=3}^{N} C_{\alpha_S}^{y_S}(q_Sp_{S2})^{y_S}\, r_S^{\alpha_S - y_S}, & j < M, \\ q_1p_{10}\prod\limits_{s=3}^{N} C_{\alpha_S}^{y_S}(q_Sp_{S2})^{y_S}\, r_S^{\alpha_S - y_S}, & j = M. \end{cases} \tag{22}$$

Для определения вектора стационарных вероятностей состояний $\overline{\pi}$ необходимо решить систему линейных алгебраических уравнений (СЛАУ) $\overline{\pi} = \overline{\pi}P$, соответствующую рассмотренной модели ВС, с помощью которых вычисляются основные характеристики [4] (среднее число занятых устройств в s-м узле, среднего числа задач, находящихся в s-м узле и в очереди к s-му узлу, загрузки устройств).

Выводы

Объем статьи не позволяет привести численные результаты, позволяющие определить эффективность эксплуатации кластерных систем. С помощью этой модели вычисляются основные характеристики [5] для каждого класса задач при различных M и $\overline{k}=(k_1,k_2,k_3)$, на основании которых производится оценуа вычислительной среды.

Эти модели можно использовать как на стадии оптимизации ВС так и на стадии проектирования и эксплуатации для выработки рекомендаций для рационального использования вычислительных ресурсов.

Литература

1. Шнитман В. Современные высокопроизводительные компьютеры. Информационно-аналитические материалы центра информационных технологий, 1996: http//hardware/app_kis.
2. Корнеев В.В. Параллельные вычислительные системы. –М., 1999, 312с.
3. Столингс У. Структурная организация и архитектура компьютерных систем.- М.: Вильямс, 2002, 893с.
4. Авен О. И. и др. Оценка качества и оптимизация вычислительных систем. – М.: Наука, 1982, 464с.
5. Менаске Д., Алмейда В. Производительность WEB-служб.-СПб: ООО “ДиаСофтЮП”, 2003, 480с.

Поступила в редакцию 11 января 2004 года