

В. И. Терехов, Н. Г. Треногин

## ОПТИМАЛЬНОЕ РАЗМЕЩЕНИЕ ИНФОРМАЦИИ В ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ С УЧЕТОМ СТРУКТУРНОЙ НАДЕЖНОСТИ КОМПОНЕНТОВ\*

Существует довольно большой класс вычислительных систем, построенных на базе серверов СУБД. При этом данные (таблицы) могут располагаться либо на одном сервере СУБД (централизованная система), либо на нескольких серверах, удаленных друг от друга (распределенная система).

В данной статье описан метод для оптимального распределения данных по серверам СУБД с учетом надежности системы.

Постановка задачи: рассматривается распределенная информационная система, узлы которой связаны между собой каналами глобальной сети. Каждый узел подключен к этой сети на определенной скорости. Узел состоит из локальной сети, к которой подключены пользователи, и сервера, хранящего данные пользователей.

Для моделирования описанной выше информационной системы предлагается использовать аналитическую модель, построенную на основе теории очередей.

Введем следующие обозначения:

$N$  — количество узлов;

$i, k$  — номер узла  $I = 1..N, k = 1..N$  (причем, пользователи  $i$ -го узла обращаются к узлу  $k$ );

$n_i$  — число пользователей, подключенных к  $i$ -му узлу;

$M$  — количество таблиц, которые предстоит оптимально распределить по серверам сети;

$j$  — номер таблицы,  $j = 1..M$ ;

$V_j$  — объем  $j$ -й таблицы;

$\lambda_j$  — интенсивность запросов пользователя к  $j$ -й таблице;

$Y_{ijk} = \begin{cases} 1, & \text{если пользователи } i\text{-го узла обращаются к } j\text{-й таблице;} \\ 0 & \text{иначе;} \end{cases}$

$X_{ij} = \begin{cases} 1, & \text{если } j\text{-я таблица имеется на } i\text{-м узле,} \\ 0 & \text{иначе.} \end{cases}$

---

\* Работа выполнена при финансовой поддержке Министерства образования РФ.

Тогда  $X_{ij} = Y_{ijk, l=k}$ , т.е. если таблица  $j$  имеется на  $i$ -м узле, то пользователи обращаются к ней.

Кроме того, таблица должна быть хотя бы на одном из узлов, и пользователи каждого узла должны иметь доступ ко всем таблицам. В установившемся режиме пользователи  $i$ -го узла обращаются к одному конкретному  $k$ -му узлу за  $j$ -й таблицей.

Тогда для всех  $j$  и  $i$  имеем:

$$\sum_k Y_{ijk} = 1.$$

Интенсивность нагрузки, поступающей на сервер  $k$ -го узла, определяется как

$$\Lambda_{k \text{ вх}} = \sum_j \sum_i Y_{ijk} \lambda_j n_i.$$

Часть этой нагрузки приходит от пользователей этого же узла  $k$ , а другая часть — по сети. Определим эту часть:

$$\Lambda'_{k \text{ вх}} = \Lambda_{k \text{ вх}} - n_k \sum_j X_{kj} \lambda_j.$$

Для каждого  $i$ -го узла интенсивность запросов к другим узлам сети  $\Lambda'_{i \text{ исх}}$  определяется как суммарная интенсивность от пользователей  $i$ -го узла  $\Lambda_{i \text{ исх}}$  за вычетом той доли, что замыкается в  $i$ -м узле:

$$\Lambda_{i \text{ исх}} = \sum_j \sum_k Y_{ijk} \lambda_j n_i;$$

$$\Lambda'_{i \text{ исх}} = \Lambda_{i \text{ исх}} - n_i \sum_j X_{ij} \lambda_j.$$

Нагрузка от  $i$ -го узла к  $k$ -му находится как

$$\Lambda_{ik} = n_i \sum_j Y_{ijk} \lambda_j.$$

Кроме того, могут быть заданы ограничения на объем хранимых данных в каждом узле

$$\sum_j X_{kj} V_j < V_{\text{max}}.$$

Время обслуживания на различных элементах СеМО зависит от интенсивности обслуживания в канале  $k$ -го узла и на сервере  $k$ -го узла ( $\mu'_i$  и  $\mu_i$  соответственно).

Нагрузка, входящая в  $k$ -й узел, складывается из ответов на запросы пользователей  $k$ -го узла к остальным узлам системы и запросов пользователей к серверу  $k$ -го узла. Считаем, что интенсивность ответов однозначно связана с интенсивностью запросов:

$$\Lambda'_{k \text{ вх } \Sigma} = \Lambda'_{k \text{ вх}} + f(\Lambda'_{k \text{ исх}}).$$

Нагрузка, исходящая из узла  $k$ , складывается из запросов пользователей  $k$ -го узла к остальным узлам системы и ответов на запросы пользователям других узлов. Считаем, что интенсивность ответов однозначно связана с интенсивностью запросов:

$$\Lambda'_{k \text{ исх } \Sigma} = \Lambda'_{k \text{ исх}} + f(\Lambda'_{k \text{ вх}}).$$

В простейшем случае:

$$\Lambda'_{k \text{ исх } \Sigma} = \Lambda'_{k \text{ вх } \Sigma} = \Lambda'_{k \text{ вх}} + \Lambda'_{k \text{ исх}}.$$

Для СМО М/М/1 время обслуживания на сервере и на канале в обоих направлениях равно

$$\begin{aligned} T_{k \text{ сеп}} &= 1/(\mu_k - \Lambda'_{k \text{ вх}}); \\ T_{k \text{ вх}} &= 1/(\mu'_k - \Lambda'_{k \text{ вх } \Sigma}); \\ T_{k \text{ исх}} &= 1/(\mu'_k - \Lambda'_{k \text{ исх } \Sigma}). \end{aligned}$$

Среднее время задержки для пользователей каждого из узлов может быть найдено как

$$T_k = ((\Lambda'_{k \text{ исх}} - \Lambda'_{k \text{ исх}}) / \Lambda'_{k \text{ исх}}) T_{k \text{ сеп}} + (\Lambda'_{k \text{ исх}} / \Lambda'_{k \text{ исх}}) T_{k \text{ исх}} + \sum (\Lambda'_{ik} / \Lambda'_{k \text{ исх}})(T_{i \text{ вх}} + T_{i \text{ сеп}} + T_{i \text{ исх}} + T_{k \text{ вх}}).$$

Это время не должно быть больше заданного.

В приложении к расчетным системам операторов связи, данный подход освещен в [1].

Введем в данную модель дополнительные ограничения, учитывающие показатели надежности системы.

Пусть  $K_{\text{сер } i}$  — коэффициент готовности сервера  $i$ -го узла,  $K_{\text{кан } i}$  — коэффициент готовности канала  $i$ -го узла.

Рассматривая данную информационную систему с точки зрения надежности, можно выделить два случая.

1. В случае критического сбоя клиентское приложение имеет возможность переключаться и работать с другими копиями таблиц. Таким образом, в системе параллельно существуют несколько копий каждой таблицы. Для такого варианта правильнее рассматривать коэффициент готовности каждой  $j$ -й таблицы

$$K_{ij} = 1 - \prod_i (1 - K_{\text{сер } i} K_{\text{кан } i} X_{ij}).$$

Таким образом, чем больше копий таблицы будет размещено на различных узлах, тем выше коэффициент готовности. Данное соотношение можно рассматривать в виде дополнительного ограничения при поиске оптимального решения, если установить  $K_{ij} > 0.98$ .

2. Архитектура клиентского приложения неизменна, и переключение на другие копии невозможно. В данном случае надежность можно рассматривать с точки зрения отказа в обслуживании пользователям узла  $i$ .

$$K_{гi} = (1 - (1 - K_{г\text{сер } i})(1 - \prod_j (1 - X_{ij}))) \cdot \prod_{k, k \neq i} (1 - (1 - K_{г\text{сер } k} K_{г\text{кан } k})(1 - \prod_j (1 - Y_{ijk}))) \times (1 - (1 - K_{г\text{кан } i})(1 - \prod_j X_{ij})).$$

В данной формуле коэффициент готовности системы для пользователей  $i$ -ого узла определятся как произведение коэффициента готовности сервера на узле  $i$  (при условии, что на этом узле используется хотя одна таблица), коэффициентов готовности серверов и каналов узлов  $k$  (если используются таблицы с узла  $k$ ) и коэффициента готовности канала узла  $i$  (если используется хотя бы одна таблица вне  $i$ -ого узла). В данном случае надежность тем ниже, чем больше количество узлов, к которым необходимо обращаться.

Приведенное соотношение можно рассматривать в виде дополнительного ограничения при поиске оптимального решения, если установить  $K_{гi} > 0.98$ .

## СПИСОК ЛИТЕРАТУРЫ

1. **Треногин Н. Г., Терехов В. И.** Методика размещения данных при проектировании биллинговых систем // Электросвязь. — 2002. — № 3. — С. 32–34.