

Методика анализа характеристик производительности информационных систем с использованием комплекса инструментальных средств анализа моделей доступа к базам данных

77-48211/630574

09, сентябрь 2013

Григорьев Ю. А., Плужникова О.

УДК 004.415.2.043

Россия, МГТУ им. Н.Э. Баумана

grigorev@iu5.bmstu.ru

Введение

Важно дать проектировщику, выполняющему разработку в среде СУБД, инструмент, позволяющий прогнозировать временные показатели ИС и выявлять потенциальные "узкие места" ИС на основе описаний проектных решений. Это особенно важно, если учесть, что основным языком (SQL) доступа к данным относится к классу непроцедурных языков. Неосторожное использование спецификаций этого языка может привести к существенному увеличению времени реакции проектируемой системы. Таким инструментом можно считать программный Комплекс инструментальных Средств Анализа Моделей доступа к базам данных (КСАМ) [1]. В комплексе используются:

- новый класс моделей анализа временных характеристик РСОД;
- метод принятия решений на основе целостного выбора варианта архитектуры;
- реестр сетей различного уровня: локальных, городских, глобальных;
- реестр узлов сетей, тестированных по спецификациям международной организации Transactional Processing Performance Council (TPC); реестр включает описания разных серверов и суперсерверов (более 20 фирм: Sun, Compaq, HP, Digital, IBM и т.

д.), СУБД (Oracle, Informix, Sybase, MS SQL Server и т. д.), сетевых ОС (различных версий UNIX, Windows и т. д.) и серверов приложений (Tuxedo, TopEnd и т. д.).

Комплекс учитывает в процессе анализа моделей доступа к БД следующие параметры (возможно их описание в виде нечётких чисел):

- описание концептуальной (инфологической) схемы базы данных PCOD и параметры наполнения базы данных (прогнозируемое число записей в таблицах и мощности атрибутов),
- описание транзакций PCOD, которые могут обращаться к другим транзакциям распределённой системы, на уровне запросов к базе данных,
- механизмы декомпозиции запросов (SQL-операторов) к распределённой БД, особенности обработки запросов в узлах системы,
- параметры архитектуры PCOD (топологии и характеристик узлов и сетей из реестров результатов тестов TPC и параметров сетей),
- распределение таблиц (с учетом тиражирования) и транзакций по узлам PCOD,
- интенсивности обращений рабочих станций к транзакциям и транзакций к другим транзакциям.

В результате моделирования пользователь получает и может проанализировать:

- загрузки узлов (серверов и клиентов) и сетей и их "разложение" по запросам и транзакциям с целью выявления «узкого места»,
- время выполнения транзакций и его "разложение" по узлам, сетям, запросам и вызываемым транзакциям,
- время доступа из узлов к транзакциям, учитывающее время передачи входных и выходных данных от рабочих станций,
- число блоков, обрабатываемых при выполнении запросов SQL.

Методика применения KСAM.

Методика использования KСAM приведена на рис. 1 (рисунок предоставлен студентом Бирулиным В.М.).

Для достижения адекватного представления PCOD в среде KСAM необходимо выделить основные составляющие предметной области и сформировать тезаурус для каждой составляющей. Данные сведения необходимо получить на этапе изучения и

анализа предметной области, и от их полноты и адекватности представления зависит информационная корректность модели системы.

После формирования тезауруса необходимо выделить сущности инфологической модели системы, которые впоследствии будут использованы в качестве таблиц БД. Для каждой сущности необходимо выбрать набор атрибутов, который должен входить в тезаурус одной из составляющих предметной области, при этом можно формировать сущности с использованием тезауруса разных составляющих предметной области. Такое допущение делает возможным моделирование внешних ключей и других специфических элементов БД. Для каждого атрибута следует указать его тип и другие характеристики, получаемые на этапе логического проектирования. Все указанные данные необходимо в соответствии с руководством пользователя ввести в информационные поля формы КСАМ «ОПИСАНИЕ СХЕМЫ БАЗЫ ДАННЫХ» (см. 1 на рис. 1).

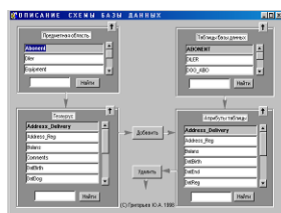
После формирования модели таблиц необходимо определить модель информационных потоков системы, представляемых в системе в виде транзакций и запросов. Вначале необходимо на основе диаграмм потоков данных выделить процессы, которые впоследствии будут реализованы в виде транзакций. Данные процессы необходимо детализировать до уровня элементарных запросов. После этого в систему вводятся модели элементарных запросов, для которых описываются таблицы, с которыми производятся операции, участвующие атрибуты, а также логические условия связи между таблицами для запросов, обращающихся к нескольким таблицам. После формирования моделей запросов они группируются по принципу принадлежности к транзакциям, и это отражается в модели транзакций. КСАМ позволяет использовать однородную структуру запросов, поэтому возможно включение одного запроса в состав различных транзакций. Все полученные данные необходимо ввести в информационные поля формы КСАМ «ОПИСАНИЕ ЗАПРОСОВ И ТРАНЗАКЦИЙ» (см. 2 на рис. 1).

Начало моделирования

Методика моделирования распределённых систем при помощи KСAM - комплекса инструментальных средств анализа моделей доступа к базам данных РСОД

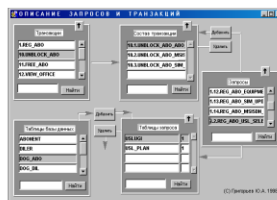
1. Описание схемы БД

- Описание предметной области
- Определение тезауруса для каждой подсистемы
- Определение сущностей (таблиц)
- Включение реквизитов тезауруса в сущности (определение атрибутов)



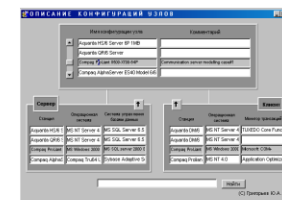
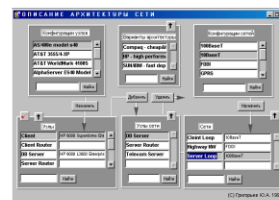
2. Описание запросов и транзакций

- Определение таблиц, входящих в запрос
- и связей между таблицами
- Определение состава транзакций (неразрывный набор запросов)



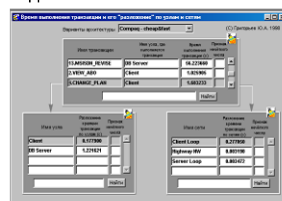
3. Описание архитектуры системы

- Определение вариантов реализации
- Выбор комплекса ТС для каждого варианта
- Описание вычислительных узлов системы
- Описание сетей, используемых в системе



8. Оценка времени выполнения транзакций

- Выявление узлов с критической загрузкой
- Выявление сегментов сети с макс. задержкой
- Определение запросов, требующих макс. время для выполнения



Завершение моделирования

да
Время оптимально?

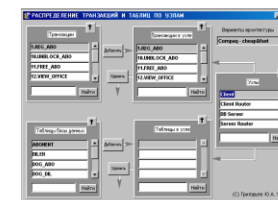
нет

Коррекция исходных данных

Перераспределение, создание копии

4. Распределение транзакций и таблиц

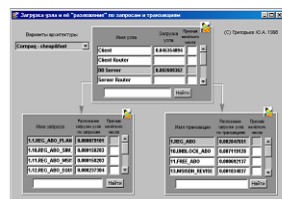
- Определение узлов выполнения транзакций
- Распределение таблиц и реплик по узлам



да

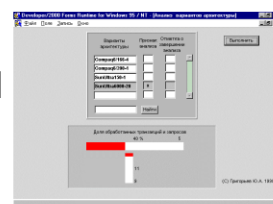
Загрузка оптимальна?

нет



7. Оценка загрузки узлов и сетей

- Выявление узлов с критической загрузкой
- Выявление сегментов сети с критической загрузкой
- Выявление транзакций, вызывающих максимальную загрузку



5. Обращения к транзакциям

- Определение узлов, из которых происходит обращение к транзакциям
- Определение транзакций, из которых происходит обращения к другим транзакциям
- Определение интенсивности обращений

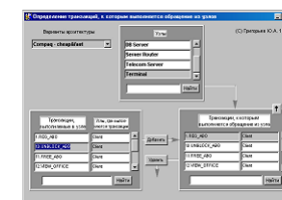


Рис. 1. Методика использования KСAM

После завершения формирования модели архитектуры и узлов системы с использованием формы КСАМ «ОПИСАНИЕ АРХИТЕКТУРЫ СИСТЕМЫ» (см. 3 на рис. 1) необходимо задать параметры информационных потоков, для чего, в первую очередь, необходимо определить физическое расположение таблиц БД и узлов, где выполняются транзакции (форма «РАСПРЕДЕЛЕНИЕ ТАБЛИЦ И ТРАНЗАКЦИЙ ПО УЗЛАМ», см. 4 на рис. 1).

Необходимо учитывать следующее: полностью архитектура системы очерчивается именно на данном этапе моделирования, поскольку от того, как будут размещены таблицы и транзакции, будет зависеть как поведение отдельных узлов, так и всей системы в целом. Приведём некоторые варианты размещения таблиц и транзакций и получаемые архитектуры:

- таблицы БД, выполняемые транзакции, терминалы расположены на разных узлах – 3-х уровневая архитектура «сервер БД – сервер приложений - терминал», наиболее гибкая и производительная структура,

- в одних и тех же узлах размещены таблицы БД и выполняются транзакции, терминалы размещаются на других узлах – 3-х уровневая архитектура, где на клиентские машины приходится минимум вычислений, и максимальная нагрузка ложится на основной вычислительный центр,

- таблицы БД расположены в одном узле, а транзакции выполняются в узлах, где размещены рабочие станции - 2-ух уровневая архитектура «клиент-сервер БД» - наиболее распространённая и легко реализуемая структура.

Пятым этапом является определение узлов системы, которые иницируют обращения к транзакциям (терминалы). КСАМ позволяет моделировать обращения транзакций друг к другу, что делает возможным моделирование систем, построенных с использованием технологий объектно-ориентированного доступа к БД. Данная функция позволяет эмулировать доступ к БД через объектный интерфейс путём создания двух разделённых баз данных, соответственно содержащих прикладную пользовательскую информацию и служебную информацию об объектах-интерфейсах и организации взаимодействия между ними на уровне вложенного вызова транзакций. Для каждого узла или транзакции необходимо задать интенсивность обращения к транзакциям и прочие нагрузочные характеристики (формы "ОБРАЩЕНИЕ К ТРАНЗАКЦИЯМ ИЗ УЗЛОВ" и "ОБРАЩЕНИЕ К ТРАНЗАКЦИЯМ ИЗ ТРАНЗАКЦИЙ", см. 5 на рис. 1).

На шестом этапе проводятся вычислительные эксперименты, а на седьмом и восьмом этапах выполняется анализ их результатов (см. 6-8 на рис.1). Ниже приводится краткое описание результатов моделирования, получаемых КСАМ.

1. Загрузки узлов.

В КСАМ под загрузкой узла понимается загрузка одного диска станции этого узла с учётом наличия общесистемного буфера. В современных информационных системах в качестве серверов баз данных, где хранятся большие объёмы данных, используются станции с одним или несколькими мощными процессорами. Для таких систем потенциальным "узким местом" является внешняя память. Загрузка устройства примерно равна доли времени, когда устройство выполняет полезную работу, т. е. не простаивает.

В блоке "Разложение загрузки узла по запросам" (рис. 2) отображается список запросов к базе данных, при выполнении которых были использованы ресурсы выбранного узла, и составляющие загрузки узла, связанные с выполнением соответствующих запросов. В блоке "Разложение загрузки узла по транзакциям" выводится список транзакций, при выполнении которых были задействованы ресурсы выбранного узла. В этом же блоке отображаются составляющие загрузки узла, связанные с выполнением соответствующих транзакций.

Известно, что для устойчивой работы информационной системы загрузки её ресурсов (в частности узлов) не должны превышать величины 0.6. Если это не так, то с помощью информации, представленной в двух нижних блоках (рис. 2), можно выявить "узкое место". Затем следует скорректировать параметры проектируемой ИС так, чтобы развязать (устранить) это "узкое место".

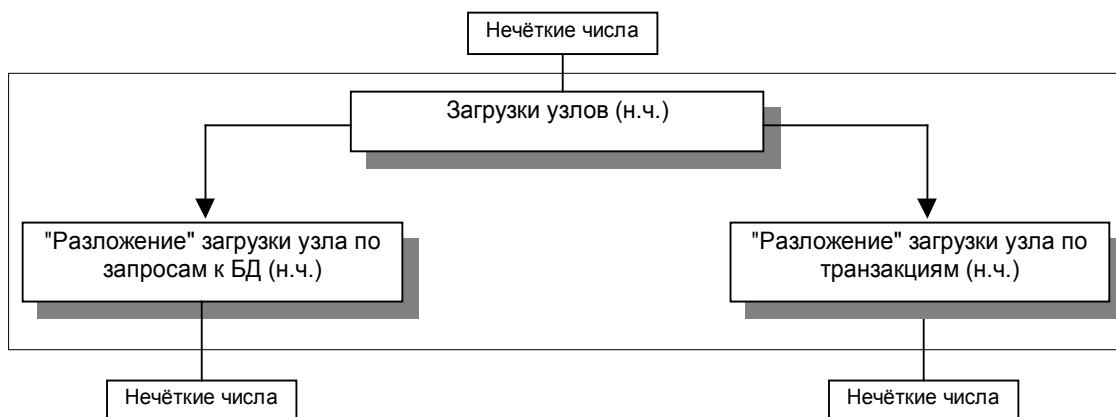


Рис. 2. Схема формы, которая используется для анализа загрузок узлов.

2. Загрузки сетей.

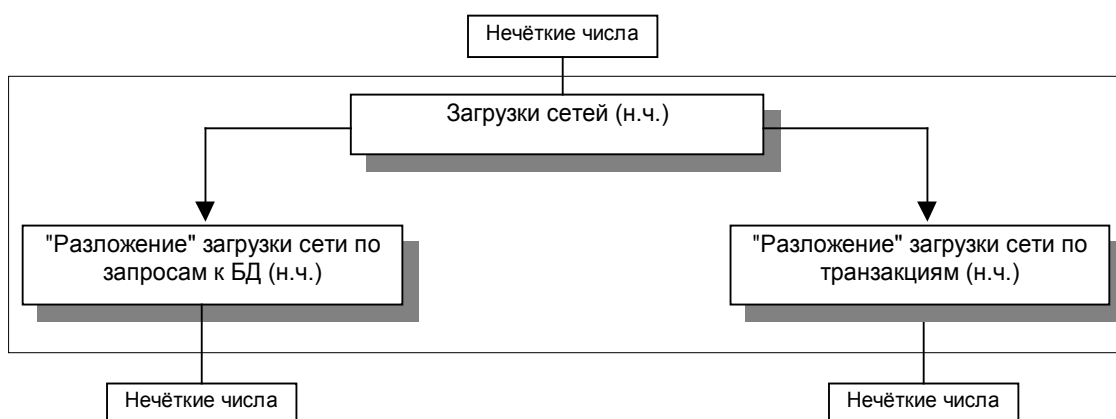


Рис. 3. Схема формы, которая используется для анализа загрузок сетей.

С помощью рассматриваемой формы (рис. 3) можно проанализировать загрузки сетей проектируемой системы. В КСАМ под загрузкой сети понимается нагрузка разделяемого канала сети или одного порта коммутатора. Загрузка сети примерно равна доли времени, когда сеть выполняет передачу кадров, т. е. не простаивает. Сказанное выше для узлов остаётся справедливым и для сетей.

3. Время выполнения транзакций.

Здесь можно проанализировать среднее время выполнения транзакций проектируемой системы.

В блоке "Разложение времени по узлам" (рис. 4) отображается список узлов ИС, ресурсы которых были использованы при выполнении выделенной транзакции, и составляющие времени выполнения транзакции, связанные с использованием ресурсов этих узлов (обработка запросов или их подзапросов на сервере; обработка записей курсора). В блоке "Разложение времени по сетям" (рис. 4) выводится список сетей ИС, ресурсы которых были задействованы при выполнении выделенной транзакции, а также составляющие времени выполнения транзакции, связанные с использованием ресурсов этих сетей (передача входных и выходных данных запросов и их подзапросов на серверы, где они обрабатываются; передача через сеть входных и выходных данных при обращении к вызываемым транзакциям).

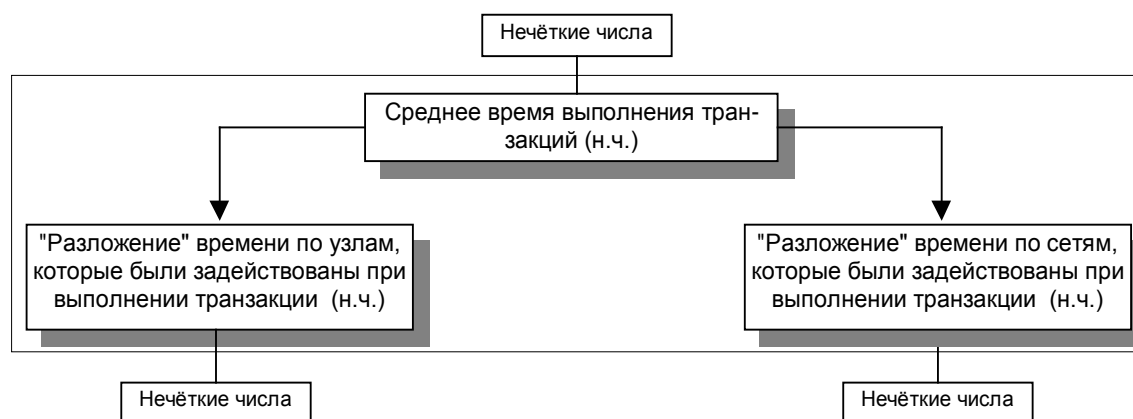


Рис. 4. Схема формы, которая используется для анализа времени выполнения транзакций ("разложение" по узлам и сетям).

Форма, схема которой представлена на рис. 5, может быть использована для анализа "разложения" времени выполнения транзакции на составляющие, связанные с запросами и вызываемыми транзакциями.

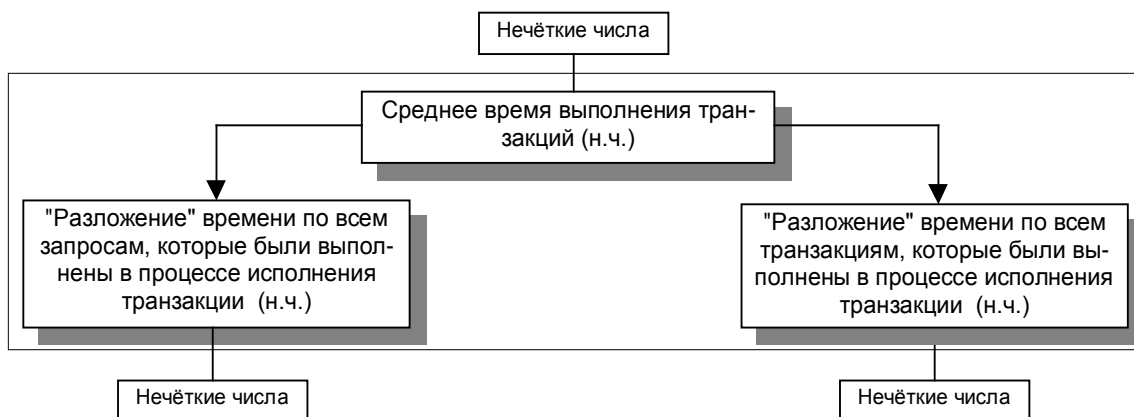


Рис. 5. Схема формы, которая используется для анализа времени выполнения транзакций ("разложение" по запросам и транзакциям).

С помощью КСАМ можно оценить нечёткое время доступа из узла к транзакции, учитывающее время передачи входных и выходных данных от рабочих станций, а также нечёткое среднее число блоков, обрабатываемых при выполнении запроса.

Ниже на схематичном уровне рассматриваются примеры описания некоторых информационных систем, которые поддерживают архитектуру клиент/сервер.

1. Модель сервера базы данных (рис. 6).

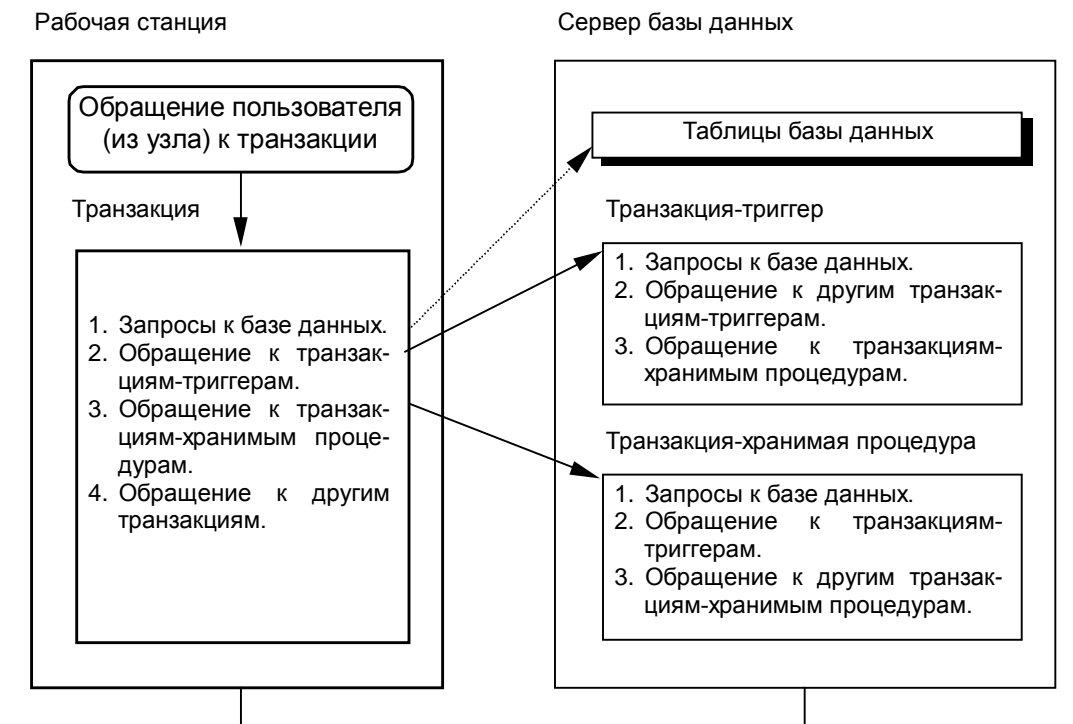


Рис. 6. Схема описания системы, поддерживающей модель сервера базы данных.

Здесь приложения (транзакции) выполняются, в основном, на рабочих станциях. Доступ к транзакции осуществляется из того же узла, где эта транзакция выполняется. Описание транзакции может содержать определения запросов к базе данных (то есть описания SQL-операторов), обращения к триггерам и хранимым процедурам, которые выполняются на сервере базы данных, обращения к другим транзакциям узла. В КСАМ вызов триггера осуществляется из транзакции, хотя в реальной системе к триггеру обращается ядро СУБД при наступлении какого-либо события (например, до или после обновления таблицы, а также при выполнении некоторого логического условия). В данном случае методическая ошибка моделирования связана только с передачей двух кадров (вход и выход) с пустыми пакетами.

В дальнейшем на рисунках отрезком прямой обозначена корпоративная сеть, которая может включать в себя взаимосвязанные LAN-, MAN- и WAN-сети.

Ясно, что при описании этой модели можно определить несколько транзакций различных типов, множество рабочих станций, серверов баз данных, по которым распределены таблицы базы данных.

2. Модель сервера приложений (рис. 7).

Основное отличие от предыдущей модели заключается в том, что транзакции запускаются не на рабочих станциях, а на сервере приложений. Это позволяет разгрузить рабочие станции, то есть перейти к "тонким" клиентам. Транзакции сервера приложений могут обращаться (может быть с некоторой вероятностью) к другим прикладным транзакциям, которые выполняются на других серверах приложений.

Сервер базы данных и сервер приложений могут функционировать на одной станции. В КСАМ этот вариант описывается просто: достаточно таблицы базы данных и транзакции распределить на один узел.

При описании этой модели можно определить несколько транзакций различных типов, множество рабочих станций, серверов транзакций, серверов баз данных.

В рассматриваемой модели допускается выполнение транзакций и на рабочих станциях.

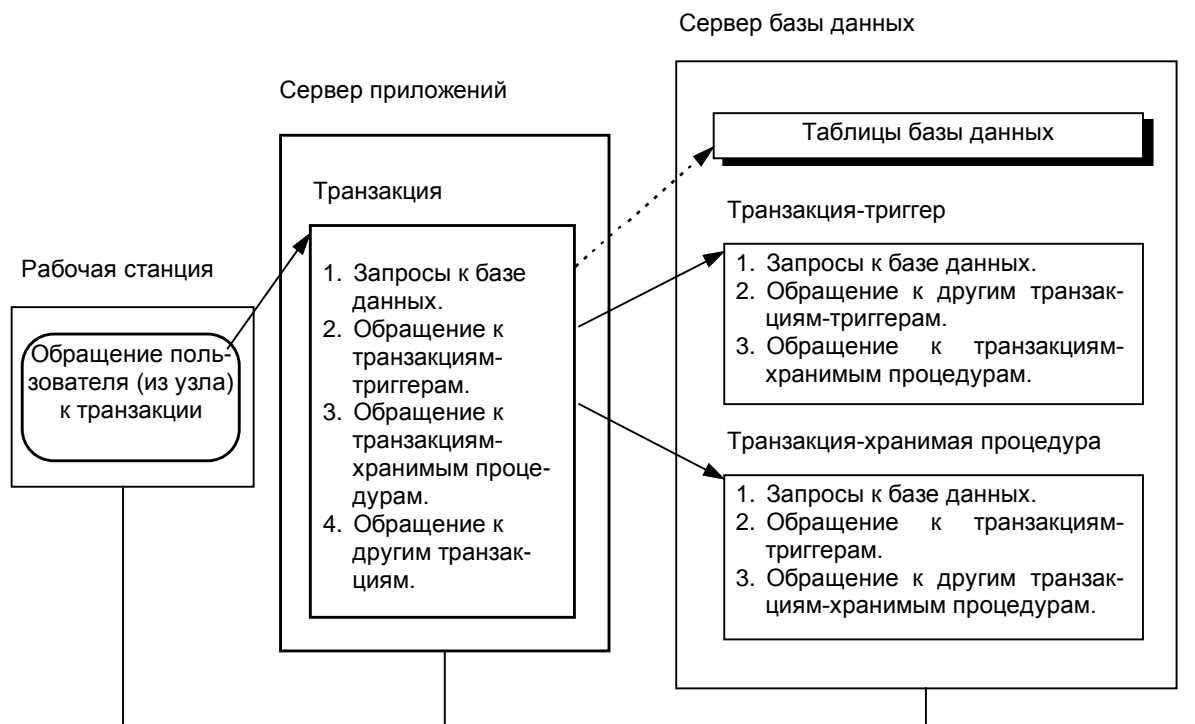


Рис. 7. Схема описания системы, поддерживающей модель сервера приложений

3. Технология Internet/Intranet.

А. Доступ к базе данных из CGI- или ISAPI-программы (рис. 8).

В настоящее время для написания приложений в среде Microsoft Internet Information Server вместо медленного CGI-интерфейса (CGI-программа выполняется как иной, нежели Web-сервер, процесс) часто используют ISAPI-интерфейс, позволяющий создавать dll-программы, выполняющиеся в адресном пространстве Web-сервера. Для доступа к базе данных часто используют уже готовые ISAPI-приложения типа dbWeb, IDC и др.

Здесь можно считать (рис. 8), что все HTML-документы как бы хранятся в некоторой таблице (например, с именем "HTML-документы"). Схема таблицы состоит из двух атрибутов: первый атрибут (индексированный) - это имя документа, его длина равна длине DET-элемента каталога файлов сервера, второй атрибут - это содержимое документа, его нечёткая длина равна размеру соответствующего HTML-файла. При

поиске (Select) одной записи будет читаться блок индекса (элемент каталога) и блок(и) данных (файл с документом).

Сервер базы данных и Web-сервер могут функционировать на одной станции.

Описание этой модели может включать в себя определение большого числа рабочих станций, Web-серверов, серверов баз данных, транзакций различных типов.

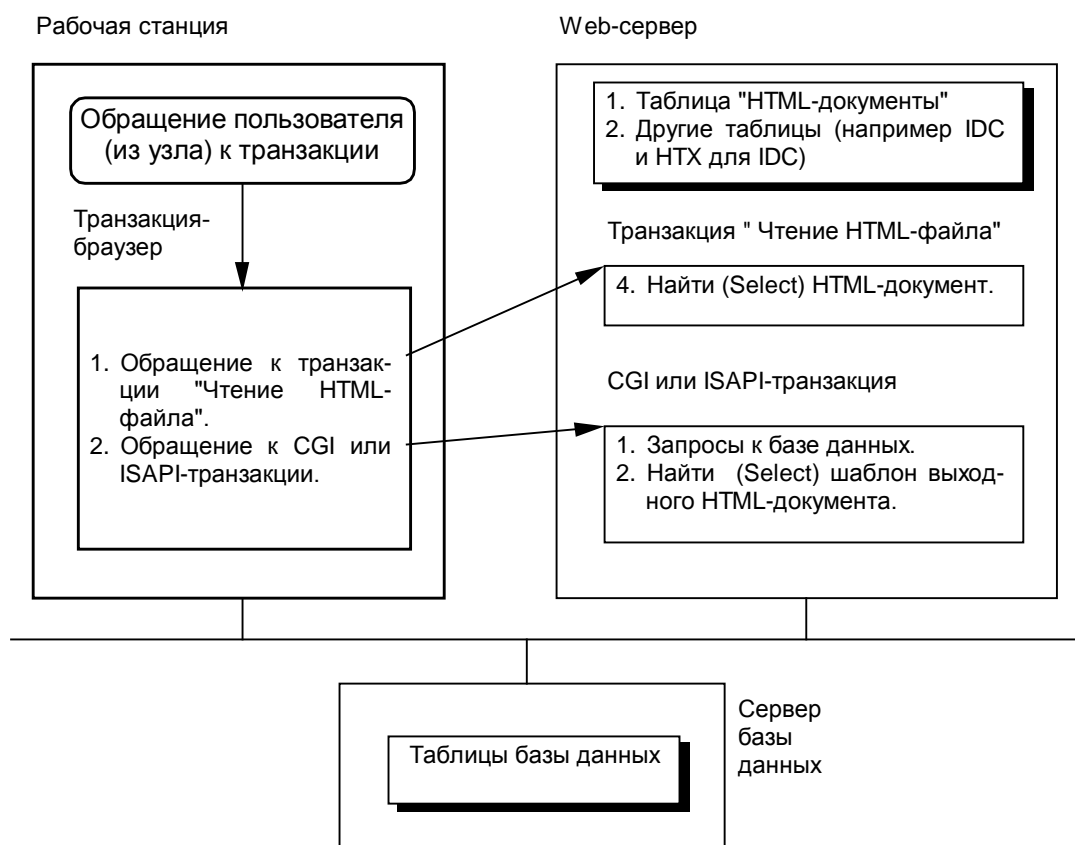


Рис. 8. Схема описания системы, поддерживающей модель доступа к базе данных из CGI- или ISAPI-программы.

Б. Доступ к базе данных из Java- или JavaScript -программы (рис. 9).

В процессе выполнения Java-программы на рабочей станции она может через интерфейс JDBC (Java Data Base Connectivity) обращаться к удалённому серверу базы данных. Для этой цели JDBC использует соответствующий ODBC-драйвер на Web-сервере, откуда была загружена Java-программа. Возможны и другие варианты доступа.

Современные Web-серверы и Web-браузеры одновременно поддерживают HTML- и Java-программы. Более того, в HTML-документ могут быть встроены

фрагменты на языке JavaScript (или VBScript), откуда можно вызывать методы объектов ActiveX и передавать им SQL-операторы для дальнейшей обработки.

Как и в предыдущем случае, здесь также можно считать (рис.9), что все HTML-документы и Java-апплеты хранятся в таблицах. Если обращение к базе данных выполняется из фрагмента на языке JavaScript (или VBScript), встроенного в HTML-документ, то обращение к транзакции "Чтение Java-апплета" не требуется.

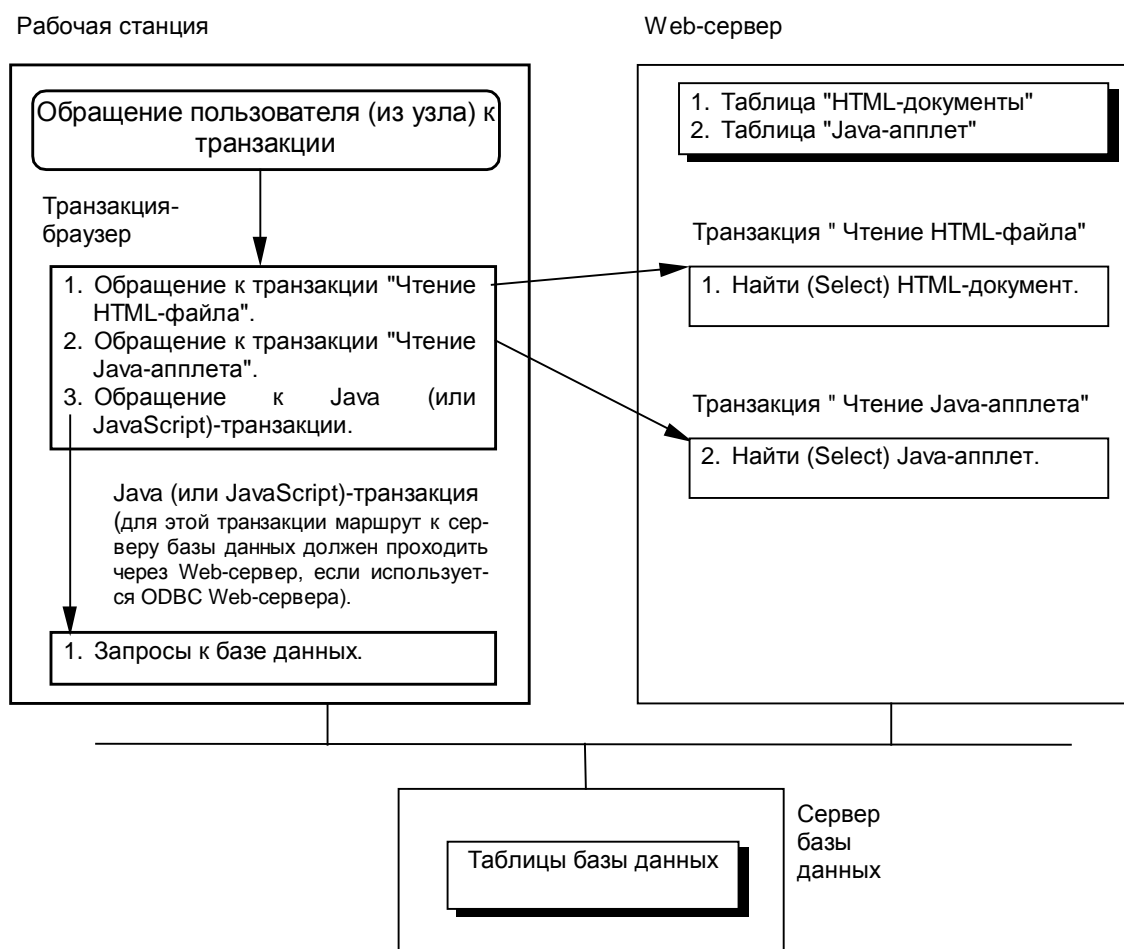


Рис. 9. Схема описания системы, поддерживающей модель доступа к базе данных из Java- или JavaScript -программы.

При описании реальной распределённой системы обработки данных возможны различные сочетания рассмотренных выше моделей доступа к данным.

Заключение

В статье рассмотрена методика использования Комплекса инструментальных Средств Анализа Моделей доступа к базам данных (КСАМ). Показано, что это инструмент проектировщика, позволяющий прогнозировать временные показатели сложных информационных систем (ИС) и выявлять потенциальные "узкие места" ИС на основе описаний проектных решений. Это особенно важно, если учесть, что основным языком (SQL) доступа к данным относится к классу непроецедурных языков.

Список литературы

1. Григорьев Ю.А. Программный комплекс анализа индексов производительности распределённых систем обработки данных // Наука и образование [Электронный ресурс]. - Электрон. журн. - М., 2012, № 2. - URL: <http://technomag.edu.ru/doc/296095.html>, свободный. - (Дата обращения: 15.01.2012)
2. Григорьев Ю.А., Плутенко А.Д. Жизненный цикл проектирования распределённых баз данных. - Благовещенск: Изд-во Амурского гос. ун-та, 1999. - 266 с.
3. Кузнецов С.Д. Стандарты языка реляционных баз данных SQL: краткий обзор // Системы управления базами данных (М.). - 1996. - № 2. - с. 6-36.
4. Григорьев Ю.А. Методические указания к выполнению курсовой работы по дисциплине "Структурное проектирование АСОИУ" //Сборник учебно-методических работ кафедры "Системы обработки информации и управления (бакалавры). Учебное пособие. Вып. 1 / Под ред. В.М. Чёрненко. М.: Изд-во ООО "АртКом", 2013, 280 с.