

ОПТИМАЛЬНЫЙ АЛГОРИТМ ПРОЕКТИРОВАНИЯ УСТОЙЧИВЫХ К СБОЯМ ИНФОРМАЦИОННЫХ СИСТЕМ ОБРАБОТКИ ЭЛЕКТРОННЫХ МЕДИЦИНСКИХ ЗАПИСЕЙ

Канд. экон. наук, доц. МЕЛЮШИН П. В.¹⁾, асп. КАЗАЧИНСКАЯ Е. А.¹⁾, ХМЕЛЬ С. А.²⁾

¹⁾Белорусский национальный технический университет,

²⁾Белорусский государственный университет

E-mail: pmelushin@gmail.com

OPTIMAL DESIGN ALGORITHM FOR FAULT TOLERANT INFORMATION SYSTEMS USED FOR PROCESSING ELECTRONIC MEDICAL RECORDS

MELYUSHIN P. V.¹⁾, KAZACHINSKAYA E. A.¹⁾, KHMEL S. A.²⁾

¹⁾Belarusian National Technical University,

²⁾Belarusian State University

Рассмотрены проблемы проектирования медицинских информационных систем. Предлагается подход по созданию высоконадежной автоматизированной системы обработки электронных медицинских записей на основе оптимизации размещения файлов в узлах сети. Разработана математическая модель оптимального распределения файлов в узлах сети и проведено экспериментальное исследование двух схем медицинских информационных систем.

Ключевые слова: оптимизация, электронные медицинские записи, надежность информационной системы.

Ил. 2. Табл. 1. Библиогр.: 4 назв.

The paper considers problems on designing of medical information systems and proposes an approach to creation of a highly reliable automated system for processing electronic medical records on the basis of file allocation optimization in the network nodes. A mathematical model has been developed for optimal distribution of the files in the network nodes and an experimental investigation of two schemes of medical information systems has been executed in the paper.

Keywords: optimization, electronic medical records, reliability of information system.

Fig. 2. Tab. 1. Ref.: 4 titles.

Введение. Важный элемент медицинских информационных систем – уникальность, безопасность и конфиденциальность данных пациента. Реализация медицинских информационных систем выделилась в направление ЕМР (эффективное медицинское планирование) и приводит к созданию не просто программы, а целой технологии идентификации, сохранения уникальности пациента и доставки данных в любую точку по запросу и разрешению пациента. Разрабатываемая технология должна обеспечить высокую безопасность электронных медицинских записей и оперативность перемещения данных пациента на любой компьютер, а также располагать потенциалом для чрезвычайных ситуаций. Особенности обработки электронных медицинских записей [1], несмотря на огромный рост компьютерных технологий в медицине, затрудняют автоматизацию. Большинство

медицинских учреждений продолжает использовать бумажные носители данных.

Врачи сталкиваются с многочисленным давлением со всех сторон:

- долгие ожидания пациента и государственное регулирование;
- сложные правила страхования и увеличенные накладных расходов;
- снижение прибыли. Усложнение требований к медицинскому обслуживанию означает, что врачи должны работать значительно эффективнее.

Точная документация и оперативная информация являются одними из наиболее важных вопросов для врачей. Растущая обеспокоенность по поводу стоимости и качества медицинской помощи служит основой для дискуссий по оценке выгод от автоматизации в организациях здравоохранения, страховых компаниях и федеральных правительствах.

Анализ подходов по созданию высоконадежной автоматизированной ЕМР-системы обработки электронных медицинских записей. Предлагается подход по созданию высоконадежной автоматизированной ЕМР-системы обработки электронных медицинских записей на основе оптимизации размещения файлов в узлах сети. На качественные (цена и эффективность) и количественные (объемы обрабатываемой информации и скорость работы) характеристики ЕМР-систем оказывает влияние решение следующих проблем.

Оперативная производительность (ОР), связанная с реакцией ЕМР-системы от запроса данных до их доставки потребителю (пациенту, врачу, страховой компании), при которой обеспечивается оперативная производительность за счет сочетания нескольких взаимосвязанных решений, таких как применение узкоспециализированных операционных систем, ориентированных на поддержку баз данных, параллелизм, оптимизация, баланс загрузки сервера. Одним из способов повышения производительности ЕМР-систем является распараллеливание: параллельные операции ввода-вывода с диском, параллельные утилиты и параллельная обработка запросов. Параллельные операции ввода-вывода с диском позволяют серверу эффективно использовать многомерные таблицы и их разбиение. Параллельные операции-утилиты (сортировка, построение индексов, загрузка, резервирование, восстановление) используют как параллельную обработку, так и параллельные обмены с диском.

Надежность (Н) – важнейшая характеристика ЕМР-систем, которая требует хранения более 10 лет медицинских записей и доступа к ним по запросам. Обеспечивается надежность и защищенность данных от сбоев через способность восстанавливаться как по чтению, так и по записи независимо от обстоятельств, включая выход из строя отдельных компонент. Реализуются отказоустойчивые системы на основе аппаратной избыточности и избыточности по данным. В первом случае используются отдельные платформы, процессоры, сдвоенные диски, а также зеркалирование аппаратуры, при котором один диск – копия другого, защищающая его от сбоев. Избыточность по данным возможна в двух формах – программного зер-

калирования и копирования. Благодаря им исключаются одиночные точки отказа (сбой на линии связи не приводит к выходу из строя всей системы).

Безопасность и защита (БЗ) данных связаны с обеспечением врачебной тайны и требованиями неприкосновенности личной жизни. Эта проблема существует в системах бумажного хранения и обостряется в ЕМР-системах. Медицинские бумажные данные дублируются и копируются по факсу через медицинских работников и страховых агентов. Обеспечить надлежащую безопасность в ЕМР-системах еще сложнее, так как компьютерная запись требует особых алгоритмов по защите данных.

Наиболее популярны два основных подхода к защите данных:

- применение к каждому защищаемому объекту набора допустимых привилегий;
- определение для каждого пользователя некоторого набора прав доступа, например основанного на использовании методов криптографии.

Основной характеристикой оперативной доступности (ОД) баз данных (БД) является сопровождение их в реальном режиме времени. В ЕМР-системах утилиты сопровождения должны поддерживать непрерывные операции, с помощью которых система обеспечивает доставку данных и исключает плановые и unplanned неполадки на линии связи и сервере данных. Обеспечивается оперативная доступность через физическое положение БД, реорганизацию, управление памятью, архивирование протоколов и перезапуск системы с минимальным воздействием на приложения.

Уровень взаимодействия (УВ) с пользователем определяется наличием интерфейса, обеспечивающего высокую эффективность общения пользователей с системой при реализации ее основных функций и высоких показателей пользовательских характеристик системы в целом.

Модернизация (М) связана с ростом производительности, пропускной способности путем добавления вычислительных ресурсов без изменения приложений и административного сопровождения. Существуют два способа расширения БД: горизонтальный и вертикальный. Горизонтальное расширение может быть полу-

чено, когда несколько серверов БД действуют независимо и делят рабочее пространство. Вертикальное расширение достигается путем добавления к платформе дополнительных ресурсов с целью уменьшения времени отклика или увеличения пропускной способности. Поддержка сервера БД для вертикального расширения не должна требовать включения дополнительных программных модулей, поскольку это увеличивает административные издержки.

Новые требования к архитектуре корпоративных ЕМР-систем. Комплексная модель проектирования автоматизированной системы обработки электронных медицинских записей представляет собой набор параметров и зависит от доступных для разработчиков технологий работы с данными

$$F = f(OP, H, БЗ, ОД, УВ, М).$$

Совокупность перечисленных проблем определяет важный набор параметров автоматизированной системы обработки электронных медицинских записей. Решение рассмотренных выше проблем для различных типов распределенных ЕМР-систем существенно сказывается на количественных и качественных показателях. Оценка параметров ЕМР-системы конкретного типа осуществляется для конкретного проектного решения, выполняющего набор функций, и содержит не только БД (запоминающее устройство со всеми атрибутами аппаратной и программной поддержки и его внутренним наполнением данными), но и сложную систему управления.

Проектирование комплексной модели по предметной направленности, интегрированной в большую по размеру распределенную ЕМР-систему, – сложная задача, поскольку при этом явно наблюдаются интеграция различных структур данных (неформатированные элементы, мультимедийные, гипертекстовые, распределенные данные и др.) и новые операции над данными (распределенная обработка, мультиплатформенность, многопротокольность, полнотекстовая обработка). Возникли новые требования к архитектуре корпоративных ЕМР-систем, основные из которых следующие:

1) обеспечение доступа ко всем необходимым данным распределенных БД;

2) применение наиболее типовых и перспективных архитектур БД и программных средств: хранилищ данных, оперативной аналитической обработки данных (OLAP), быстрой разработки приложений (RAD), средств поддержки принятия решений (DSS) на основе хранилищ данных;

3) применение методов логического вывода, нейронных сетей, нейрокомпьютеров;

4) применение единого интерфейса пользователя для работы с разными компонентами данных и приложений;

5) постоянная актуализация понятийной модели для учета новых понятий, возникающих при изменении прикладных задач, и динамическое администрирование распределенной корпоративной БД.

Все это привело к модификации известных технологий проектирования ЕМР-систем, а также к поиску новых, сочетающих в себе перспективные направления и методы построения средств искусственного (машинного) интеллекта. Как показал анализ современных и перспективных ЕМР-систем, параметры БД и ЕМР-систем взаимосвязаны: конкретные значения параметров ЕМР-систем могут быть получены только при наличии определенного типа системы управления данными.

В технологии ЕМР-систем выделились три направления совершенствования систем [2]:

Postgres – характеризуется максимальным следованием известным принципам организации системы управления базами данных (СУБД) (если не считать коренной переделки системы управления внешней памятью);

Exodus/Genesis – основная характеристика: создание собственно не системы, а генератора систем, наиболее полно соответствующих потребностям приложений. Решение достигается путем создания наборов модулей со стандартизованными интерфейсами, причем идея распространяется вплоть до самых базисовых слоев системы;

Starburst – характеризуется приспособляемостью к нуждам конкретных приложений путем использования стандартного механизма управления правилами. Системы представляют некоторый интерпретатор системы правил и набор модулей-действий, вызываемых в соответствии с этими правилами. Можно изменять наборы правил (совершенствуется специаль-

ный язык их задания) или изменять действия, подставляя другие модули с тем же интерфейсом.

В работе исследуются два подхода к проектированию ЕМР-систем [3]:

1) централизованная ЕМР-система – обработка данных реализуется на центральном компьютере-сервере, где осуществляется обработка запроса пользователя (технология направления Postgres). Дублируются данные, подверженные частым изменениям, и хранятся децентрализованно. Каждый набор данных размещается на том узле, где он чаще всего используется, что обеспечивает высокий уровень оперативности и снижает нагрузку на каналы передачи. Централизованная ЕМР-система наиболее проста в реализации (рис. 1);

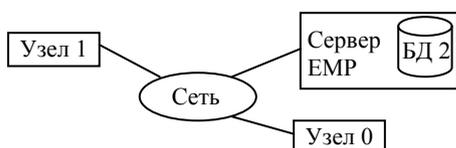


Рис. 1. Централизованная ЕМР-система

2) система с двукратным дублированием – обработка данных реализуется на центральном компьютере-сервере и дублируется на втором сервере, где осуществляется обработка запроса пользователя (технология направления Exodus/Genesis). Дублируются все данные и хранятся в распределенных узлах. Каждый набор данных размещается на двух серверах; это обеспечивает высокий уровень надежности, но снижает оперативность. Система с двукратным дублированием сложна в реализации и эксплуатации (рис. 2).

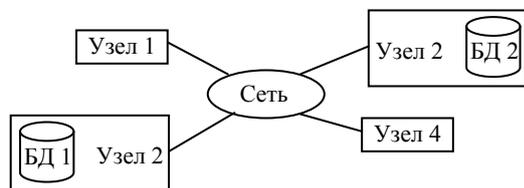


Рис. 2. Система с двукратным дублированием

Рациональное размещение информации существенно влияет на эффективность функционирования предлагаемых схем вычислительной сети ЕМР-систем. Поэтому необходимо определить наилучшее размещение файлов в рас-

пределенной базе данных (РБД) по узлам ЕМР-системы. Один из способов повышения производительности и надежности ЕМР-систем – это распараллеливание операции ввода-вывода данных и дублирование данных. Сложной проблемой выполнения распределенных запросов является оптимизация, т. е. поиск оптимального плана динамично изменяющегося объема хранимых данных. Информация, которая требуется для оптимизации запроса, распределена по серверам и узлам сети.

В данной работе предлагается алгоритм перераспределения файлов для реализации ЕМР-системы, обеспечивающий высокую сохранность данных за счет дублирования и оптимизации хранения распределенных данных. Сформулирована оптимальная задача распределения файлов в функционирующей ЕМР-системе на основе стандартной задачи линейного программирования. Используется типовой симплекс-алгоритм.

Оптимизационный алгоритм функционирования ЕМР-системы. Предлагается подход к построению математической модели оптимального распределения файлов для двух схем (централизованная ЕМР-система и система с двукратным дублированием). Особенностью предложенного оптимального алгоритма является оперативная процедура перераспределения файлов, обеспечивающая 100%-е восстановление при сбое на сервере.

Анализируется вычислительная сеть: все узлы сети связаны между собой каналом передачи данных. Файлы содержатся на сервере и в локальных базах узлов таким образом, что имеется заданное администратором число копий каждого файла. Запрос, инициированный в любом узле, предполагает доступ к центральному серверу, на котором фиксируются последние корректировки запрашиваемого файла.

Постановка задачи проектируемой ЕМР-системы включает:

n – количество узлов сети (в ЕМР-системе общее число пользователей – около 10 тыс., включая пациентов, врачей и служащих страховых компаний);

m – количество файлов РБД (ориентировочно $n \cdot 10$);

F_{ij} – объем файла j в i -м узле сети (колеблется от 1 МБ текстовых данных до 0,1 Гб видеодокументов);

L_{ij} – интенсивность запроса к файлу F_{ij} , инициированного в узле K_i ;

A_{ij} – стоимость запроса к файлу F_{ij} , инициированного в узле K_i ;

B_{ij} – объем запрашиваемых данных при выполнении запроса к файлу F_i , поступившего в узел K_j ;

D_{ij} – стоимость хранения j -го файла в i -м узле сети;

X – матрица размещения файлов: $X_{ij} = 1$, если файл F_{ij} находится в узле K_i , иначе $X_{ij} = 0$;

V_j – объем памяти узла K_i (примерно 1 Гб);

$Q_{ij} = L_{ij}A_{ij} + B_{ij}D_{ij}$ – затраты на передачу и хранение j -го файла в i -м узле сети.

Необходимо найти матрицу размещения файлов X , обеспечивающую минимум затрат S на передачу и хранение файлов в узлах сети:

$$S = \sum_{i=1}^m \sum_{j=1}^n (Q_{ij} X_{ij}) \rightarrow \min.$$

Ограничения в оптимизационной модели следующие.

Общее количество копий файла j – не менее W (задается администратором сети в зависимости от требования надежности данных)

$$\sum_{i=1}^n X_{ik} = W.$$

Суммарный объем памяти в i -м узле ($i = 1$ до n) достаточен для размещения файлов при их перераспределении

$$\sum_{j=1}^m F_{ij} X_{ij} < V_i, \quad i = 1, n.$$

Решение задачи выполняется стандартным симплекс-алгоритмом.

Описание алгоритма функционирования ЕМР-системы. 1. Формируем матрицу X и задаем нулевые значения матрице размещения файлов.

2. На выбранном интервале времени (час, сутки, неделя) решаем оптимизационную задачу по размещению файлов в матрице X .

3. Формируем матрицу запрещенных узлов $E_{ij} = 1$.

4. Реализуем занесение файла F_{ij} среди запрещенных узлов $X_{ij} = 1$.

Если файл F_{ij} помещается в узел K_i , то размещаем его в узле K_p : $E_{ij} = 1$.

Если файл F_i не помещается в узел K_i , то:

а) запрещаем этот узел: $E_{ij} = 0$;

б) среди всех узлов находим первый запрещенный узел $E_{ij} = 1$, в который помещается файл F_i .

5. Если на выбранном интервале времени сбоя в системе не произошло, то переходим к пункту 1.

6. Если произошел сбой в работе сервера ЕМР-системы, то, основываясь на матрице $X_{ij} = 1$, организуем полное восстановление всей системы.

Экспериментальные результаты оптимального алгоритма. Для экспериментов использовали 64-разрядную ЭВМ с частотой 2,1 ГГц. Программа составлена на языке C++, операционная система LINUX (Suse 10). В качестве сервера принята стабильная версия MySQL 3.2 с 60 Гб хранимых данных. Создавали модель ЕМР-системы для 10 тыс. пользователей, 10 оперативных администраторов и одного системного администратора. В системном журнале фиксировали информацию по эффективности предлагаемого варианта.

Производительность по каждой из схем ЕМР-систем определяли как количество запросов (для одного запроса обрабатывали 1 Мб по доставке на сервер, выборке и выдаче результатов), осуществляемых за допустимое время реакции системы в 6 с. Надежность ЕМР-систем определяли как время в минутах до полного восстановления всех данных при выходе из строя центрального и дублирующего серверов (данные на сервере удаляли, имитируя физическую потерю информации). Стоимость обработки одного запроса рассчитывали на основе данных [4], учитывающих средние начальные затраты на ЕМР-систему в размере 5183 дол. в год (31100 дол. за 6 лет) и текущие годовые затраты на эксплуатацию в размере 44600 дол. Результаты работы алгоритма сравнительных характеристик различных схем ЕМР-системы приведены в табл. 1.

Как видно из табл. 1, централизованная ЕМР-система имеет высокие показатели по производительности и стоимости при значи-

тельной величине времени на полное восстановление данных. Рекомендуется для медицинских учреждений и страховых организаций при низких требованиях к надежности.

Таблица 1

Сравнительные характеристики двух схем ЕМР-системы

Показатель	Централизованная система	Система с двукратным дублированием
Производительность	180 запросов за 6 с	130 запросов за 6 с
Надежность	40 мин (полное восстановление данных)	8 мин (при потере данных на сервере), 24 мин (сбой двух серверов)
Стоимость за 1 год	5183 дол. + + 44600 дол. = = 49783 дол.	2 · 5183 дол. + + 44600 дол. = = 54966 дол.

Система ЕМР с двукратным дублированием имеет высокие показатели при полном восстановлении данных. Производительность и стоимость незначительно возрастают. Целесообразно создавать эти системы для медицинских учреждений при повышенных требованиях к надежности и качеству проектирования (скорая помощь и поликлиники). Перспективной является дальнейшая разработка методов и алгоритмов по повышению показателя стоимость/надежность.

ВЫВОД

На основе оптимизационной задачи предложен подход по созданию высоконадежной автоматизированной системы обработки электронных медицинских записей. Разработана математическая модель оптимального распределения файлов в узлах сети и проведено экспериментальное исследование двух схем медицинских информационных систем. Оценена

эффективность двух схем ЕМР-систем при разных требованиях по надежности обрабатываемых данных в медицинских информационных системах.

ЛИТЕРАТУРА

1. Мелюшин, В. П. Тенденции автоматизации приборостроительных предприятий / В. П. Мелюшин, С. А. Хмель // Материалы 5-й Междунар. конф. «Приборостроение-2012». – Минск: БНТУ, 2012.

2. Пятибратов, А. П. Вычислительные системы, сети и телекоммуникации / А. П. Пятибратов, Л. П. Гудыно, А. А. Кириченко. – М.: Финансы и статистика, 2008. – 480 с.

3. Таненбаум, А. С. Компьютерные сети / А. С. Таненбаум; пер. с англ. В. Шрага. – 4-е изд. – СПб.: Питер, 2012. – 596 с.

4. Cost-Benefit Analysis of Electronic Medical Records in Primary Care / J. S. Wang [et al.] // The American Journal of Medicine. – 2003. – Vol. 114 (5). – С. 397–403.

REFERENCES

1. Melyushin, V. P., & Khmel, S. A. (2012) Tendencies in Automation of Instrument-Making Enterprises. *Priborostroenie-2012: Materialy 5-i Mezhdunarodnoi Nauchno-Tekhnicheskoi Konferentsii [Instrument-2012: Proceedings of the 5th International Scientific and Technical Conference]*. Minsk: BNTU.

2. Pyatibratov, A. P., Gudyno, L. P., & Kirichenko, A. A. (2008) *Computer Systems, Networks and Telecommunications*. Moscow: Finnasy i Statistika.

3. Tanenbaum, A. S. (2012) *Computer Networks*. 4th ed. St. Petersburg: Piter.

4. Wang, S. J., Middleton, B., Prosser, L. A., Barndon, C. G., Spurr, C. D., Carchidi, P. J., Kittler, A. F., Goldszer, R. C., Fairchild, D. G., Sussman, A. J., Kuperman, G. J., & Bates, D. W. (2003) Cost-Benefit Analysis of Electronic Medical Records in Primary Care. *American Journal of Medicine*, 114 (5), 397–403. doi:10.1016/S0002-9343(03)00057-3

Поступила 09.10.2013