

АРХИТЕКТУРА ТЕЛЕМЕДИЦИНСКОЙ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ ГЕТЕРОГЕННОГО ВЫЧИСЛИТЕЛЬНОГО КЛАСТЕРА В КАЧЕСТВЕ СЕРВЕРНОЙ ЧАСТИ

В.А. Сапрыкин

Нижегородский госуниверситет им. Н.И.Лобачевского

В настоящей работе описывается подход к созданию общей архитектуры телемедицинской системы на базе гетерогенного вычислительного кластера. Рассматривается задача обеспечения доступа к данным врачам, пациентам, обучающимся и исследователям. Предлагается механизм интегрирования в систему модулей решения вычислительно-емких задач в виде облачных сервисов. Затрагиваются вопросы разграничения доступа для разных категорий пользователей.

Введение

Целью работы является создание единой телемедицинской системы для различных медицинских учреждений и исследователей. Первый шаг, который нужно сделать – это распределенное хранение и обеспечение централизованного доступа к большим объемам изображений (в первую очередь томограмм) одного диагностического учреждения. Ввиду достаточно малой распространенности томографов, актуально предоставить врачам других учреждений доступ к медицинским изображениям конкретного пациента. При этом у диагностического учреждения должна быть возможность контролировать такой доступ. Обратно, несколько диагностических учреждений могут решить разместить свои данные в едином хранилище для облегчения обмена снимками. Пациент так же должен иметь возможность удаленно просматривать свои снимки со своего компьютера, предварительно запросив учетные данные у диагностического учреждения. Для выполнения исследований, создания алгоритмов предварительной диагностики и обучения персонала полезно иметь централизованный доступ к большой коллекции медицинских изображений от нескольких диагностических учреждений. В этом смысле необходимо, чтобы исследователям были доступны обезличенные изображения, а учреждению – полный объем информации по своим снимкам.

Большой объем трехмерных медицинских изображений и вычислительная сложность их обработки определяют высокие требования к аппаратному обеспечению для серверной и клиентской частей телемедицинской системы. Использование гетерогенного вычислительного кластера в качестве серверной части системы дает: ускорение обработки ресурсоемких запросов; уменьшение требований к аппаратному обеспечению на стороне клиента за счет переноса большей части вычислений на сервер; возможность встраивать в телемедицинскую систему модули высокопроизводительных вычислений и моделирования, которые могут быть выполнены за приемлемое время лишь на кластере. Реализация запуска на выполнение таких вычислений по команде с клиента открывает возможность создания облачных сервисов.

Постановка задачи

Для создания системы хранения медицинских изображений, способной решать описанные выше задачи, необходимо кроме самих изображений хранить метainформацию по каждому изображению и информацию о взаимосвязи отдельных изображений.

Эти данные должны включать как минимум сведения об учреждении, о пациенте, дату снимка, тип патологии или признак ее отсутствия и часть тела на изображении.

Ввиду большого объема медицинских изображений (в частности томограмм) требуется способность системы работать в условиях, когда файлы изображений распределены по различным серверам в медицинских учреждениях и специально выделенным серверам. Жестким требованием является поддержка формата DICOM [4] – промышленного стандарта создания, хранения, передачи и визуализации медицинских изображений и документов обследованных пациентов. Нужно максимально ускорить передачу данных между клиентом и сервером.

Предусмотреть возможность отдельного хранения персональных данных, чтобы при различных вариантах развертывания системы иметь возможность оставить персональные данные в стенах медицинского учреждения. Отсюда вытекает необходимость создания подсистемы анонимизации и восстановления персональных данных DICOM изображений.

Для обеспечения конфиденциальности информации нужно предоставлять доступ к системе только уполномоченным пользователям, посредством запроса имени пользователя и пароля. На основе имени пользователя требуется разграничивать доступ к данным для разных категорий пользователей. Поскольку система работает через сеть, для обмена данных между узлами можно пользоваться протоколами с шифрованием.

Требуется разработать механизм подключения к системе модулей высокопроизводительных параллельных расчетов и моделирования на вычисленном кластере с использованием центральных и графических процессоров. Такие модули должны запускаться по команде с удаленного клиента. Входные данные должны извлекаться из системы хранения, а выходные файлы должны помещаться в хранилище по окончании расчетов.

Архитектура системы

Общая архитектура системы изображена на рис. 1. Она отражает ситуацию, когда несколько медицинских учреждений передали свои обезличенные изображения в единое хранилище, оставив себе заботу о персональных данных. Так же эти учреждения по запросу выдают пациентам учетные данные, позволяющие последним получить доступ на чтение к своим снимкам или передать такой доступ другому врачу. Наконец, группа исследователей может работать с совокупным объемом обезличенных медицинских изображений.

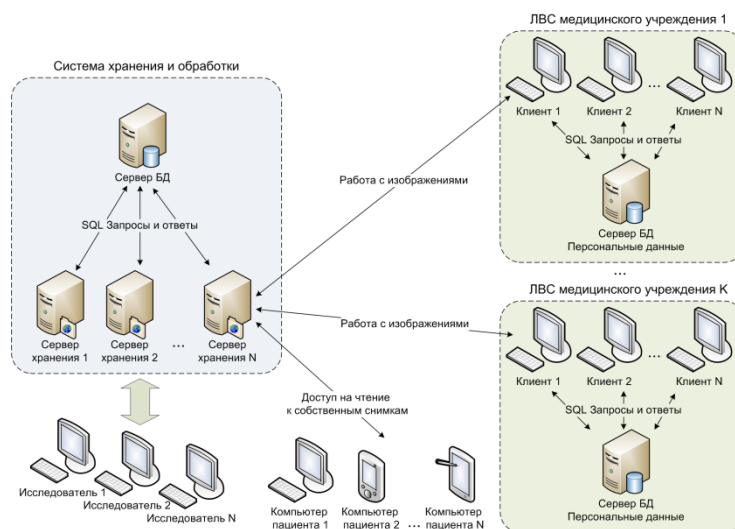


Рис. 1. Общая архитектура системы хранения и обработки медицинских изображений

Серверная часть хранилища представлена центральной СУБД и несколькими серверами хранения, которые непосредственно хранят и обрабатывают данные. Кроме того, в локальной сети каждого медицинского учреждения есть база данных, хранящая персональную информацию пациентов. Запросы к серверной части со стороны любого из клиентов могут быть выполнены удаленно.

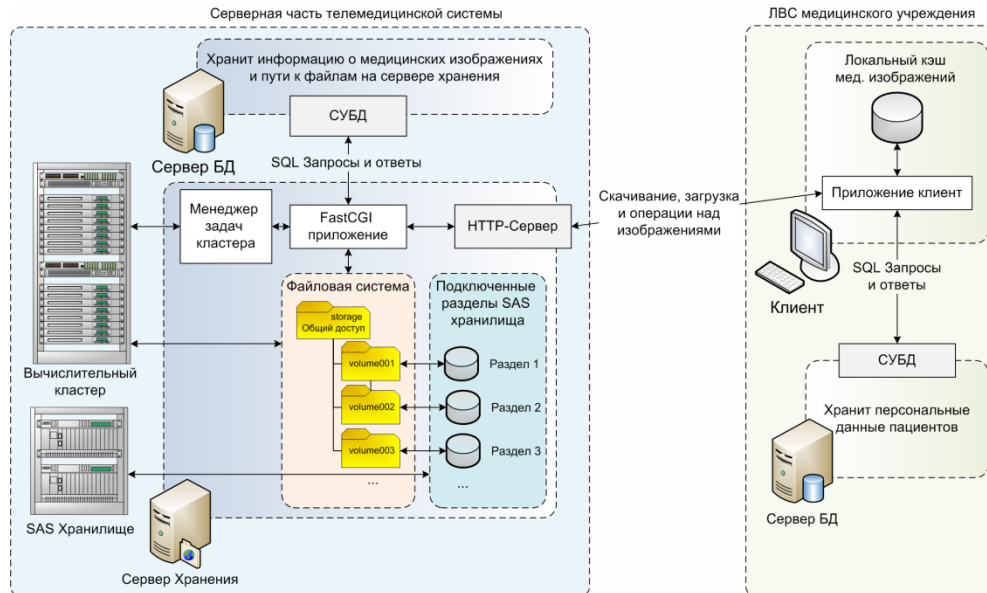


Рис. 2. Устройство компонентов системы хранения и обработки медицинских изображений

Детальное устройство различных компонентов описанной архитектуры показано на рис. 2. Основной программный модуль на сервере хранения – это сервер приложений, выполняющий запросы клиентов. Сервер приложений реализован как FastCGI приложение для web-сервера и взаимодействует с клиентским приложением по протоколу HTTP. Это существенно упрощает дальнейшую разработку, позволяя сосредоточиться на реализации нужной функциональности, абстрагируясь от решения технических проблем принятия запросов и передачи ответов.

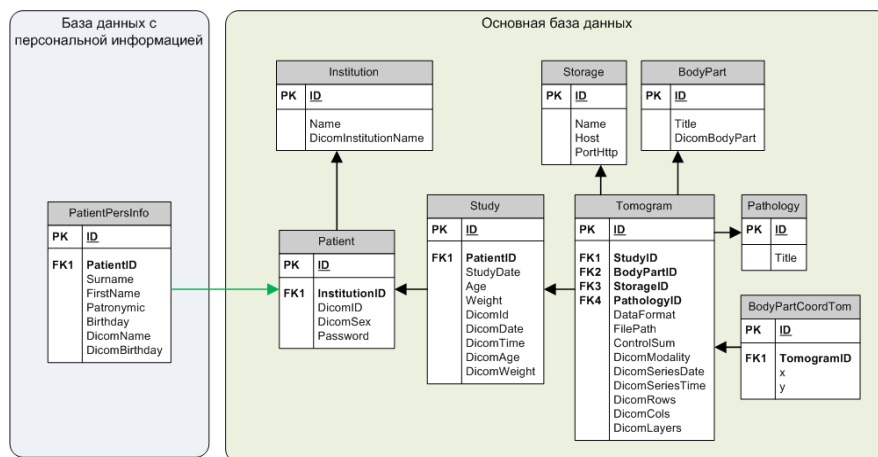


Рис. 3. Схема данных системы хранения и обработки медицинских изображений

Медицинские изображения не хранятся в базе данных, а размещаются в виде файлов на жестких дисках. Центральная база данных лишь хранит различные метаданные, касающиеся медицинских изображений. Сервер приложений имеет доступ к центральной базе данных и к каталогу storage, в который в виде подкаталогов смонтированы

разделы жестких дисков хранилища. FastCGI приложение при импорте снимков упаковывает серии изображений в формате DICOM в единый файл и без сжатия размещает в подкаталогах “volume001”, “volume002” и т.д. в зависимости от наличия свободного пространства (см. рис. 2). Затем для томограммы в центральной базе данных указывается сервер хранения, на котором она находится, и путь к ней на этом сервере относительно каталога storage, например «/volume001/tom245».

Второй программный модуль под названием «Менеджер задач кластера» предназначен для управления запуском вычислительно тяжелых задач на кластере. Предполагается, что модуль постоянно запущен и находится в состоянии ожидания команды от сервера приложений.

Схема данных (рис. 3) помимо отражения присутствующих в предметной области сущностей и их атрибутов содержит ряд специальных атрибутов. В их числе поля с префиксом Dicom, содержащие точную копию данных из DICOM-файлов, помещенных в хранилище. Эти поля используются для установления соответствия между объектами БД и DICOM-объектами, а так же для хранения персональных данных в отдельной базе данных с целью последующего их восстановления в DICOM-файлах при загрузке изображении из хранилища в исходном виде.

Развертывание серверной части

Разработанная первая версия серверной части была развернута на базе вычислительного кластера ННГУ под управлением MS Windows Server 2008 R2 Edition. Серверная часть системы состоит из сервера базы данных PostgreSQL (хранящего метаданные об изображениях) и одного сервера хранения (на который непосредственно загружаются изображения). В качестве HTTP-сервера используется Microsoft IIS. Для непосредственного хранения данных используется SAS (Serial Attached SCSI) хранилище. В целях масштабируемости его разделы монтируются как подкаталоги в каталог storage, вместо того, чтобы присваивать латинскую букву каждому разделу. В настоящий момент используется один раздел объемом 2 Tb, смонтированный в каталог /storage/volume001. В хранилище помещен опытный набор из 10000 MPT томограмм общим объемом 70 Гб.

Особенности реализации передачи DICOM-изображения

Реализована возможность пакетного импорта большого количества медицинских изображений с функцией их обезличивания. При этом кроме записей о самих изображениях в базе данных автоматически создаются и все связанные записи (пациент, исследование, часть тела и др.) на основе информации из DICOM-файлов. По ходу всего процесса импорта отображается полоса прогресса и доступна кнопка прерывания операции. Все передачи изображений проверяются на корректность с помощью подсчета контрольных сумм по обеим сторонам канала и их сверки между собой. Перед загрузкой в хранилище каждое многослойное изображение в формате DICOM подвергается упаковке в единый файл специально разработанного формата. Для ускорения передачи данных по медленным каналам связи используется сжатие «на лету» упакованного DICOM-изображения. Это позволяет примерно вдвое уменьшить время передачи данных. Для ускорения повторного доступа к недавно созданным/запрошенным изображениям, реализована подсистема локального дискового кэширования. Она позволяет избежать повторного скачивания недавно использованных изображений с сервера.

В клиентском приложении реализована функция предварительного просмотра слоев DICOM-изображения без его полного скачивания из хранилища. При передаче такого одиночного слоя, производится его кодирование в формат PNG (сжатие без потерь алгоритмом DEFLATE) с уменьшением глубины цвета до 8 бит. Наряду с обеспечением предварительного просмотра слоев, а так же, с уменьшением объема передаваемых

данных, формат PNG дает возможность удаленного просмотра слоев томограммы через браузер за счет использования протокола HTTP.

Интеграция с подсистемами ресурсоемких вычислений и визуализации

На стороне клиента интеграция с различными программными модулями заключается в запуске приложения с передачей ему параметров командной строки. При необходимости выполняется подготовка входных и анализ выходных данных; отображение результатов работы приложения; загрузка входных данных из хранилища и сохранение результатов работы в хранилище и др.

Клиентская программа интегрирована с просмотрщиком InVols [6], позволяющим просматривать томограммы в различных режимах визуализации. Запуск InVols осуществляется нажатием одной кнопки после выбора томограммы в списке. Если нужного изображения нет в кэше, оно будет автоматически загружено.

Кроме того интегрированы два ресурсоемких консольных приложения: XMCML [7] – моделирование распространения зондирующего излучения в многослойных биологических тканях с произвольной геометрией границ слоев в задаче оптической диффузионной томографии и Picador [1] – реализация моделирования динамики плазмы. Для обеих программ создан пользовательский интерфейс задания параметров моделирования и запуска процесса моделирования. По окончании вычислений запускается программа визуализации результатов. Весь процесс моделирования описывается в автоматически генерируемом командном (BAT) файле.

Заключение

В результате проделанной работы был создан и развернут на вычислительном кластере ННГУ макет телемедицинской системы. Хранилище наполнено первой опытной порцией МРТ-томограмм различных частей тела. Доступная через интернет система для долговременного хранения и дистанционного использования данных медицинской диагностики обеспечена необходимым функционалом для ее практической эксплуатации и построения на ее основе телемедицинских и других систем исследовательской и практической медицины.

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России», госконтракт № 02.740.11.0839.

Литература

1. Bastrakov S., Donchenko R., Gonoskov A., Efimenko E., Malyshev A., Meyerov I., Surmin I., Particle-in-cell plasma simulation on heterogeneous cluster systems // *Journal of Computational Science*. 2012. 3. P. 474–479.
2. Breton V., Dean K., Solomonides T. From Grid to Healthgrid. *Proceedings of Healthgrid 2005 // Studies in Health Technology and Informatics*. IOS Press, 2005. Vol 112. P. 249–321.
3. Isaac N. Bankman *Handbook of medical imaging. Processing and analysis*. – USA: Academic Press, 2000. – 911 p.
4. Pianykh O.S. *Digital Imaging and Communication in Medicine (DICOM)*. – Berlin, Germany: Springer-Verlag, 2008. – 383 p.
5. Standart «Digital Imaging and Communications in Medicine (DICOM)» – USA: National Electrical Manufacturers Association, 2011.
6. Гаврилов Н.И., Белокаменская А. А. Стереовизуализация научных и медицинских объемных данных трассировкой лучей в реальном времени. *Материалы XX конференции GraphiCon'2010, 20-24 сентября, Санкт-Петербург, Россия*. С. 350-352.

7. Горшков А.В., Кириллин М.Ю. Monte Carlo simulation of brain sensing by optical diffuse spectroscopy // Journal of Computational Science, Elsevier, 2012.
8. Грабер М. Введение в SQL. – М.: ЛОРИ, 1996. – 479 с.
9. Сапрыкин В.А. Архитектура системы хранения и обработки медицинских изображений. Материалы XI Всероссийской конференции «Высокопроизводительные параллельные вычисления на кластерных системах», 17-21 ноября, 2011, Нижний Новгород. С. 279–283.