

Распределенный архив изображений дистанционного зондирования земли

Статья опубликована в выпуске журнала № 2 за 2009 год. [17.06.2009]

Авторы: [Московский А.А.](#), [Первин А.Ю.](#), [Тютляева Е.О.](#), [Шевчук Е.В.](#)

Ключевые слова: [распределенный архив](#), [распределенная обработка данных](#), [кластерная файловая система](#), [дистанционное зондирование земли](#), [активное хранилище](#)

Постоянный адрес статьи: <http://swsys.ru/index.php?page=article&id=2197>

Дистанционное зондирование Земли (ДЗЗ) – область человеческой деятельности, тесно связанная с обработкой больших объемов информации при помощи компьютерной техники. При этом накапливаются архивы данных по наблюдениям от одного или нескольких приборов, установленных на искусственных спутниках Земли. Объем архивов может составлять несколько петабайт (10¹⁵ байт) информации.

Эффективное управление возрастающими объемами данных до сих пор остается значительной проблемой. Рост емкости хранилищ только обостряет проблему большой стоимости перемещения данных между обрабатывающими узлами и устройствами хранения. В последние годы были разработаны файловые системы (ФС) для решения проблемы управления данными в контексте высокопроизводительных вычислительных систем. Некоторые из таких ФС, например Lustre [1] и PVFS, используют выпускаемые промышленные серверы в качестве серверов ввода/вывода, то есть мест для хранения данных. Общая вычислительная мощность сотен и тысяч узлов хранения может быть очень значительной, но обычно она не используется, так как этим узлам отводится только роль хранилищ.

Один из подходов по снижению требований к пропускной способности между хранилищами и вычислительными устройствами и задействованию вычислительных мощностей устройств хранения – это приближение вычисления к местам хранения данных. Такой подход позиционируется как активные хранилища в контексте параллельных ФС. Обработка данных непосредственно на узлах хранения существенно снижает объемы передачи данных по сети и, следовательно, общесетевой трафик.

Активные хранилища нацелены на приложения с интенсивной стадией ввода/вывода, входные данные для которых можно разбить на независимые наборы. Они могут использоваться для обработки результатов моделирования в различных научных областях. Задачи обработки и хранения данных ДЗЗ имеют все свойства, позволяющие говорить о том, что реализация для их решения концепции активных хранилищ является прекрасным выбором.

Использование распределенных активных хранилищ позволяет создавать масштабируемые, высокоскоростные, управляемые распределенные системы с высокой пропускной способностью.

Предлагаем один из возможных подходов к организации активного хранилища, реализованный в программной системе для распределенного хранения и обработки

данных ДЗЗ. Описываемая программная система объединяет возможности параллельной ФС Lustre и системы автоматического динамического распараллеливания (библиотеки T-Sim [2]), позволяя создать прототип распределенного архива изображений – активного хранилища, в котором данные обрабатываются на тех же узлах, где хранятся, и обеспечивается автоматическая балансировка нагрузки на узлах кластера.

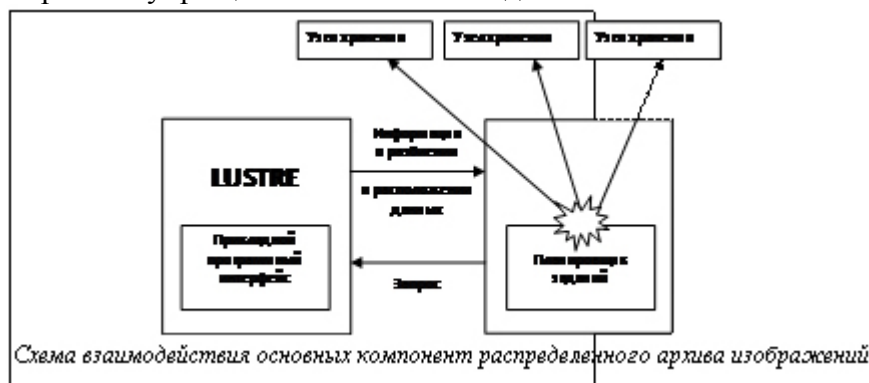
Технология активных хранилищ в последние годы приобрела особую популярность, что свидетельствует о ее актуальности. Существуют схожие по задачам системы, такие как ActiveStorage, Cascading [3], Pig [4], Hadoop [5]. Большинство разработок базируются на какой-либо специализированной кластерной файловой системе, удобной для организации активного хранилища. Предложенный авторами подход не является исключением – данная разработка использует ФС Lustre. Отличительной особенностью подхода является наличие шаблонов задач для различных стратегий распараллеливания, специализированных для обработки данных ДЗЗ, что позволяет повысить эффективность решения задач этого типа.

2) Структура программной системы

Разработанная программная система для хранения и обработки данных ДЗЗ состоит из двух компонент, которые отвечают за выполнение основных функций активного хранилища:

- кластерная ФС Lustre, которая отвечает за хранение данных и распределение их по узлам; при обращении к прикладному программному интерфейсу ФС Lustre можно получить информацию о расположении данных;
- программный комплекс, базирующийся на библиотеке шаблонных классов C++ T-Sim, позволяющий проводить эффективную обработку данных, расположенных в ФС Lustre, путем направления вычислений на узлы, хранящие обрабатываемые данные.

На рисунке изображена упрощенная схема взаимодействия основных компонент системы.



В общем случае обработка данных ДЗЗ в описываемой программной системе происходит по следующей схеме.

- Задаем разбиение для отдельного файла или для отдельной директории в ФС Lustre, наиболее отвечающее особенностям решаемой задачи.
- Планировщик заданий TSim вызывает функцию прикладного программного интерфейса ФС Lustre, которая возвращает, помимо другой информации, число и размер порций (stripes), на которые разбит файл для того, чтобы в соответствии с идеологией активного хранилища произвести обработку данных на узле, где они расположены. Таким образом, определяется схема расположения файла в кластере.

- Определяем, зная стратегию, которая используется Lustre для назначения порций файла (стрипов) на узлы кластера, какой стрип на каком узле кластера будет находиться. IP-адреса всех узлов кластера (по номерам узлов) содержатся в служебной информации Lustre. Планировщик T-Sim направляет задание на обработку стрипа на IP-адрес, который соответствует номеру узла кластера для данного стрипа.

2) Шаблоны хранения и обработки данных

Информация, полученная от спутников ДЗЗ, может храниться в файлах различных форматов и размеров, что влияет на выбираемые способы обработки данных и разбиение для хранения в ФС Lustre. Кроме того, алгоритмы обработки могут сильно отличаться по требовательности к ресурсам (процессорному времени, оперативной памяти). В связи с этим были разработаны два шаблона задач, реализующих различные подходы к хранению и обработке изображений. Для каждого шаблона определены свойства, которыми должна обладать решаемая задача, чтобы применение шаблона было целесообразно. Для каждого шаблона также был реализован пример соответствующей задачи обработки данных ДЗЗ.

Шаблон задачи 1. Данный шаблон предоставляет подход к решению задач по обработке изображений, которые обладают следующими свойствами: входные данные содержатся в файлах достаточно большого размера; формат описания данных – TIFF; вычислительная сложность задачи достаточно высока, общий алгоритм обработки можно представить в виде нескольких независимых потоков, так что затраты на распараллеливание вычислений несутся по сравнению с затратами на обработку одной порции данных.

При реализации этого шаблона задачи были использованы следующие подходы:

- использована библиотека `ltiff` для получения дополнительной информации о разбиении данных в TIFF-файле (разбиение на полоски формата TIFF);
- файл с данными разбивается для хранения на различных узлах кластера с помощью программного интерфейса ФС Lustre;
- обработка данных осуществляется по полосочному – планировщик библиотеки T-Sim направляет задание на обработку данных на тот узел, на котором расположена обрабатываемая в данный момент полоска;
- для обработки изображения используется шаблон параллельного программирования Map (отображение) из библиотеки T-Sim;
- распараллеливание алгоритма обработки изображения происходит автоматически с помощью библиотеки T-Sim (для этого алгоритм обработки должен быть модифицирован – в него необходимо внести изменения, связанные с использованием шаблонов библиотеки T-Sim).

Примером использования данного шаблона является задача контролируемой классификации изображения ДЗЗ при помощи метрики Махаланобиса.

Шаблон задачи 2. Данный шаблон предоставляет подход к решению задач по обработке изображений, которые обладают следующими свойствами: имеется последовательный код обработки изображения, который нужно многократно запускать для различных наборов данных; данные хранятся в файлах небольших размеров, так что обработка файла целиком

на одном узле более эффективна, чем разбиение его на порции; формат данных не имеет значения.

При реализации этого шаблона задачи файл с данными целиком располагается на узле, его обработка происходит там же; последовательный код обработки отдельного изображения не подвергается никакой модификации, реализуется параллелизм по данным. Файлы равномерно распределены по узлам кластера при помощи средств ФС Lustre. При этом планировщик заданий, специально реализованный для данного шаблона с помощью библиотеки T-Sim, используется для направления вычислений на тот узел кластера, на котором расположен обрабатываемый файл.

Отличительной особенностью данного шаблона является то, что последовательный код обработки изображения, как уже было указано, не подвергается никакой модификации.

С применением данного шаблона была реализована задача по перепроецированию данных ДЗЗ.

В заключение можно сделать следующие выводы. Предлагаемая программная система представляет собой распределенный архив данных ДЗЗ, реализованный в соответствии с идеологией активных хранилищ. Эффективность работы системы обеспечивается обработкой данных на тех же узлах распределенного хранилища, на которых они хранятся, что позволяет существенно снизить расходы на передачу больших массивов данных по сети. Балансировка нагрузки на узлах осуществляется планировщиком библиотеки T-Sim на основе информации, предоставляемой интерфейсом кластерной ФС Lustre.

Для решения различных видов задач обработки космических снимков, хранящихся в распределенном архиве, реализованы стратегии распараллеливания, оформленные в виде шаблонов задачи. Применение шаблонов задачи позволяет быстро и эффективно реализовать параллельные вычисления для задач обработки, удовлетворяющих условиям применения шаблона.

В качестве примеров использования предлагаемых шаблонов были реализованы задача контролируемой классификации изображений ДЗЗ при помощи метрики Махаланобиса и задача представления данных ДЗЗ в нужном масштабе и проекции. Тестирование задач, которое проводилось на двух различных установках, продемонстрировало повышение производительности приложения при использовании предлагаемых шаблонов задач, если соблюдались условия применения шаблонов: достаточная вычислительная сложность задачи для шаблона 1 и многократный запуск на различных наборах данных для шаблона 2.

Целесообразность использования описанной системы возрастает при работе с большими объемами исходных данных или при большой вычислительной сложности задач обработки изображения, что справедливо для большинства систем, использующих параллельные вычисления.

3) Литература

1. Sun Microsystems : официальный сайт кластерной файловой системы Lustre. URL: <http://wiki.lustre.org/index.php> (дата обращения: 10.04.2009).

2. Московский А.А. T-Sim – библиотека для параллельных вычислений на основе подхода T-системы // Программные системы: теория и приложения: тр. Междунар. конф. М.: Наука, Физматлит, 2006. Т. 1. С. 183–193.
3. Cascading Home Page : официальная стр. проекта Cascading. URL: <http://www.cascading.org> (дата обращения: 10.04.2009).
4. Welcome to Pig! : официальная стр. проекта Pig. URL: <http://hadoop.apache.org/pig> (дата обращения: 10.04.2009).
5. Welcome to Hadoop! : официальная стр. проекта Hadoop. URL: <http://hadoop.apache.org/core> (дата обращения: 10.04.2009).