

Каталогизация пространственных данных для облачных вычислений*

УПРАВЛЕНИЕ КОСМИЧЕСКИМИ СНИМКАМИ

Эффективное использование разнообразных «ролей», выполняемых облаком Каталогизация снимков в виде облака предоставляет аналитикам широкий диапазон возможностей. Подобно великому актеру, облако способно играть самые разные «роли». Ниже перечислены некоторые из ролей, исполняемых динамической функцией каталогизации.

Регулировщик движения: Без ущерба для навигации по видеоданным (которые теперь поступают из облака), достигается улучшенная способность к быстрой смене масштаба просмотра или полосы пропускания, поскольку информационный канал теперь может использовать ресурсы облака аналогично информационному каналу центрального процессора с предсказанием переходов.

Историк-консультант: Визуальное представление истории анализа снимков, например, кто и когда получал доступ к снимкам. Также возможно получение других полезных сведений, например, какие части снимков просматривались, с каким разрешением и в каких полосах пропускания, а также любые другие метаданные, касающиеся пользовательских сеансов.

Механизм и структура вычислений

Эксперт: Доступ к дополнительным возможностям обработки в облаке, таким как более эффективное

обнаружение аномалий на основе критериев уведомления, определенных пользователем; удаление искусственных признаков на основе исторической информации или подробной структуры полосы.

Прогнозирование потребностей: Выполнение оверлейных программ в реальном времени для загрузки по требованию актуальных данных, таких как известные искусственные сооружения, данные лазерной системы обнаружения и измерения дальности и данные в других полосах пропускания, и прочую информацию, например, демографические данные или текущий уровень и вместимость водоемов.

Гибкий менеджер метаданных: Интерактивное расширение метаданных за счет включения вложенных данных (например, данных, которые обычно присутствуют во вторичных таблицах системы управления реляционной базы данных), что позволяет сэкономить место и повысить быстродействие.

Помощник: Внесение в каталог сформированных пользователями данных, содержащих ссылки на фундаментальные исходные данные (снимки, видеоданные, векторы и т.д.) с возможностью их последующего извлечения и совместного использования. Поиск ключевых слов в комментариях аналитиков с возможностью подписки и обзора последних проанализированных изображений, содержащих определенные ключевые слова.

* Перевод с английского языка статьи «Cataloging Geospatial Data for the Cloud» (авторы James S. Blundell, Mike West, Betty Davlin, and Brandon Johnson, Geospatial Products and Solutions, Overwatch Systems, Sterling, Va.), опубликованной на сайте <http://eijournal.com/2011/cataloging-geospatial-data-for-the-cloud>. Перевод подготовлен к публикации Б.А. Дворкиным (Компания «Совзонд»)

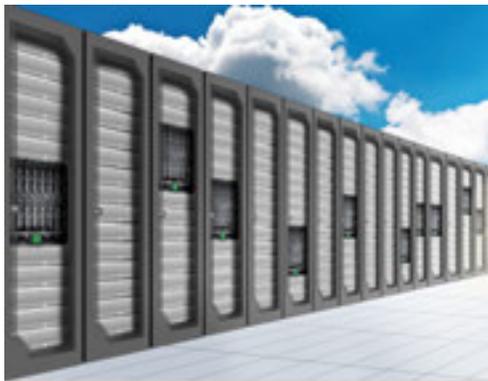


Рис. 1. Использование облачных вычислений в правительственных организациях позволяет снизить стоимость пространственных данных и улучшить доступ для пользователей

Использование облачных вычислений в качестве новой концепции для реализации целей правительства, состоящих в сокращении затрат и улучшении доступа для пользователей, открывает новые пути в области каталогизации пространственных данных.

Практика показывает, что возможности систем дистанционного зондирования Земли (ДЗЗ) по-прежнему превосходят возможности аналитиков по своевременной обработке и расшифровке информации. Например, группировка спутников компании DigitalGlobe — QuickBird, WorldView-1 и WorldView-2 — в общей сложности осуществляет съемку сверхвысокого разрешения около 1,5 млн км в сутки. Этот разрыв стал еще более заметен после введения платформ сбора данных, обеспечивающих получение видео с полноценным представлением движения (full-motion video — FMV) и наблюдение в обширном районе (wide-area surveillance — WAS), что особенно актуально при срочной расшифровке фото- и видеоданных, применяемых в решении задач стратегической и тактической разведки.

В целях ликвидации указанного разрыва организации, занимающиеся разведкой, непрерывно совершенствуют и оттачивают стратегии хранения корпоративных данных и управления данными, включая проектирование больших хранилищ данных, способных обеспечить повсеместный экономически эффективный доступ к широкому диапазону выходных дан-

ных, таких как отчеты, карты и изображения.

Несмотря на существующие трудности, разрабатываемые стратегии и подходы открывают широкий диапазон преимуществ, связанных с каталогизацией геокосмических данных в виде облака для хранения корпоративных данных и управления данными.

ВЫЧИСЛИТЕЛЬНЫЕ СТРАТЕГИИ

Облачные вычисления обеспечивают доступ к функциям Map Reduce (предварительная обработка+свертка данных, MR), Bulk Synchronous Parallelization (массивная синхронная параллелизация, BSP) и Language Integrated Query (язык интегрированных запросов, LINQ), применяемым для стратегий высокоскоростных вычислений (High-Performance Computing — HPC). Например, принцип параллелизации, присущий облачным вычислениям, позволяет повысить производительность при подготовке большого количества мозаичных изображений, настраивая выходные данные в зависимости от текущего местоположения пользователя и его заданных предпочтений. Процессоры, работающие по принципу облака, могут одновременно принимать и отправлять очередной набор мозаичных изображений клиенту для кэширования, подготавливая оверлейные сегменты с искусственными сооружениями для свободного доступа по запросу и формируя наборы мозаичных изображений с более низким разрешением, которые могут быть использованы для восполнения пробелов в случае внезапного снижения пропускной способности канала связи во время роуминга.

Такие программные продукты, как Microsoft LINQ для HPC, могут помочь в определении оптимальной стратегии структурирования, обработки и отправки запросов пользователей. Эти технологии обеспечивают, с одной стороны, совместимость со стандартными SQL-запросами (Structured Query Language — язык структурированных запросов), а с другой стороны — преобразование в эквивалентный запрос на другом языке, который использует соответствующий процессор и кэш данных. Такие стратегии также позволяют достичь более наглядной реализации логики программы. Например, цветовое представление пикселей по SQL-запросу может отличаться от функции, определенной пользователем, но это пред-

ставление будет совершенно естественным и легко воспроизводимым во время выполнения этапа reduce (свертка данных) операции map-reduce (предварительная обработка-свертка данных).

СТРАТЕГИИ УПРАВЛЕНИЯ ДАННЫМИ

Облако также поддерживает функции копирования и совместного использования базы данных, что позволяет создавать надежные, масштабируемые и распределяемые серверы мозаичных изображений, данных и метаданных. Благодаря использованию хранилищ данных NoSQL, таких как HBase, Cassandra, MongoDB и CouchDB, облако обеспечивает высокую доступность данных и метаданных. Бесчисленные способы управления данными, реализуемые в таких хранилищах, позволяют выбирать оптимальные стратегии запросов.

Например, хранилище данных Hbase — или даже файловая система Hadoop (HDFS), используемая для Hbase — предназначены для получения доступа к большим массивам периодически записываемых и часто читаемых данных, таких как мозаичные сегменты карты. Колоночные хранилища данных Cassandra одинаково хорошо поддерживают данные RDF (Resource Description Framework, схема описания ресурсов) и метаданные, структурированные в виде иерархии. Хранилища документов MongoDB поддерживают быстрые обновления записей без необходимости переписывания записи целиком. Другие технологии, такие как Lucene/Solr, позволяют улучшить процесс обработки запросов путем индексации содержания в облаке.

Опыт компании Overwatch в обработке данных изображений позволил достичь высоких результатов в каталогизации и управлении базами данных. В частности, разработанное компанией приложение GeoCatalog предоставляет возможности масштабируемого управления данными и каталогизации данных, позволяя пользователям организовывать, находить и извлекать геокосмические данные с помощью настраиваемых схем, запросов, шаблонов и индексации результатов (рис. 2). Благодаря масштабируемости приложение можно использовать для работы на отдельных портативных компьютерах, в маленьких рабочих группах и на крупных предприятиях, оно поддерживает самые раз-

нообразные данные и наборы данных, включая DPPDB (Data Point Positioning Database, База данных размещения точек привязки), используемую правительственными организациями США.

Приложение GeoCatalog интегрировано в такие средства анализа изображений, как RemoteView и ELT. GeoCatalog включает язык систематизации или определения схем (рис. 3), который может использоваться для формирования искусственных признаков DDL (Data Definition Language, язык описания данных), поддерживающих как SQL-сервер, так и ПО Oracle. GeoCatalog предоставляет широкий диапазон возможностей, позволяющих аналитикам быстро и эффективно осуществлять поиск и управление актуальными пространственными данными и оптимизировать поток операций.

Существует несколько переходных систем, обеспечивающих объединенный доступ к дополнительным источникам данных и стандартным технологиям запроса наследуемых данных на основе использования HTTP-запросов (Hypertext Transfer Protocol, протокол передачи гипертекста) и архитектуры ODBC (Open Database Connectivity, открытые средства связи с базами данных). Эти подключаемые переходные системы можно расширить таким образом, чтобы они включали интерфейс OGC (Open Geospatial Consortium, открытый геокосмический консорциум) и другие стандартные интерфейсы.

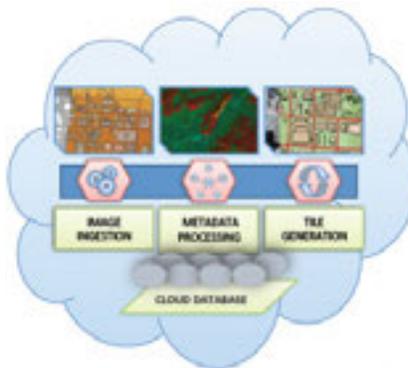


Рис.2. Сохранение данных в виде облака позволяет одновременно осуществлять приём изображений, обработку метаданных и создание мозаики благодаря использованию возможностей массивного масштабирования, присущих «облачным» процессорам.

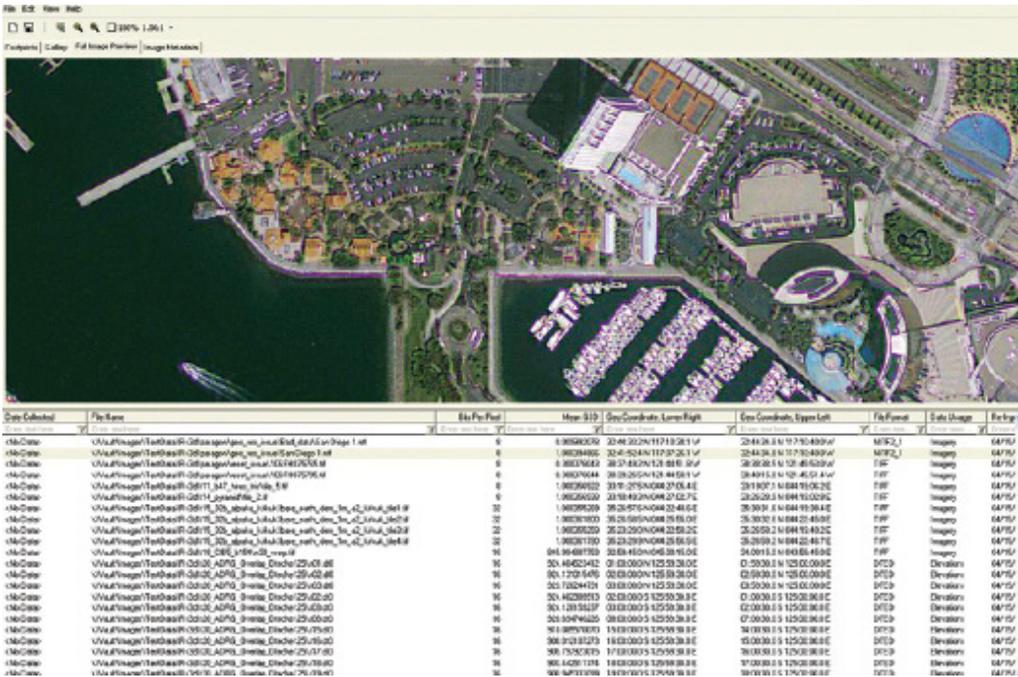


Рис. 3. Приложение GeoCatalog, разработанное компанией Overwatch, предоставляет возможности масштабируемого управления данными и каталогизации данных

Переходная буферная система IEC (Integrated Exploitation Capability, Встроенная функция расшифровки) — это клиентская функция, легко настраиваемая на месте с помощью конфигурационных файлов XML (Extensible Markup Language, расширяемый язык разметки). Переходная система DataMaster Adaptor приложения GeoCatalog включает в себя клиентский компонент, аналогичный IEC, который осуществляет связь с серверным компонентом на сервере DataMaster Solaris. Эти переходные системы выполняют один запрос, используя общий пользовательский интерфейс, и отображают объединенный результат.

ИСПОЛЬЗОВАНИЕ ПРИНЦИПА ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ ДЛЯ КАТАЛОГИЗАЦИИ

Использование принципа облачных вычислений для каталогизации космических снимков позволяет оптимизировать существующие функции и добавить новые возможности. Традиционная каталогизация

основывается на SQL-хранилищах метаданных, предназначенных для определения, создания и поддержки метаданных, связанных с изображениями и картографическими данными. Для осуществления этой цели компания OverWatch объединяет возможности приложения GeoCatalog с гибкостью функции Cloudwave, предоставляющей сетевые виджеты для исследования, запросов и представления данных. GeoCatalog, работающее на основе Cloudwave, позволит улучшить сетевую визуализацию метаданных, в т.ч. интерактивную систематизацию, историю изменений и схемы каналов, связывающих данные аналитиков с результатами работы. Исследования, выполняемые экспертами по дистанционному зондированию в области возможностей каталогизации геокосмических данных в облаке, являются многообещающими и позволяют надеяться на то, что аналитики, нуждающиеся в эффективном инструменте, получат его в ближайшем будущем.