

В. Ю. Ромасько^{1}*

Геоинформационный сервис «Метео-Сибирь». Архитектура и реализация

¹ Сибирский центр Федерального государственного бюджетного учреждения «Научно-исследовательский центр космической гидрометеорологии «Планета», г. Новосибирск, Российская Федерация
*e-mail: rvu@rcpod.siberia.net

Аннотация. В работе рассмотрена архитектура, принципы построения и детали реализации, используемое программное обеспечение (ПО) и аппаратное обеспечение геопортала «Метео-Сибирь», предназначенного для предоставления гидрометеорологической информации, полученной как по данным дистанционного зондирования, так и по данным наземных измерений. Приводится описание архитектуры геопортала на основе комбинирования свободного и специально разработанного ПО для сбора и обработки данных, подготовки и представления растровой информации большого объёма в формате прямоугольных тайловых сеток. Рассматриваются принципы построения системы и сервисов в её составе, обеспечивающие полную автоматизацию обработки данных, автоматическое восстановление после сбоев, управление большими объёмами и, одновременно, реализацию быстрого доступа к растровой информации. Отмечаются функциональные преимущества разработанных для геопортала специализированных сервисов, отсутствующие в доступном свободном ПО. Описано используемое аппаратное обеспечение, приводятся оценки производительности и объёмы пространственно-временного покрытия оперативно обновляемыми данными. Продемонстрирована возможность построения регионального гидрометеорологического геопортала на ограниченных аппаратных средствах.

Ключевые слова: геопортал, дистанционное зондирование, гидрометеорология

V. Y. Romasko^{1}*

Geoinformation service “Meteo-Siberia”. Architecture and realization

¹ Siberian Center of the Federal State Budgetary Institution «Scientific Research Center of Space Hydrometeorology «Planeta», Novosibirsk, Russian Federation
*e-mail: rvu@rcpod.siberia.net

Abstract. Article is devoted to discussion of architecture, common principles and implementation details, used software and hardware of geoportal «Meteo-Siberia» developed for presentation of hydrometeorological information based on remotely sensed and ground data. The architecture is based on combination of free and specially developed software for data collecting and processing, preparing and presenting a large volume of raster in form of quadrilateral tile grid. Common principles laid to the base of built system provide full data processing automation, automated restore after failures, managing large volume and simultaneously fast access to raster data. Advantages of specially developed services are remarked over available free those. Applied hardware together with its productivity estimations as well as special and temporal coverages of operatively renewed data are given. Thus, the possibility of building the regional hydrometeorological geoportal using modest hardware is shown.

Keywords: geoportal, remote sensing, hydrometeorology

Введение

Первая фотография облачности со спутника была сделана 14.08.1959 г. с Explorer-6 [1], первое телевизионное изображение в реальном времени, которое только и применимо в метеорологии, было получено с первого метеорологического спутника TIROS-A 01.04.1960 г. [2]. К настоящему времени на орбитах находится множество космических аппаратов метеорологического назначения, в том числе находящихся на геостационарных орбитах, а периодичность съёмки одного и того же района уже достигла 2,5 минут, полушария – 10 минут [3]. И специалисты используют для работы с такими спутниковыми изображениями современные геопортальные технологии. Инструменты на основе веб-технологий, с одной стороны, не требуют установки специального программного обеспечения и, с другой стороны, привносят в обработку данных вычислительную мощность серверной стороны.

На сегодняшний день на территории РФ действует множество региональных геопорталов различного назначения. Ряд информационных ресурсов подобран на странице Росгидромета [4]. Следует отметить геопортал ГИС «Метео-ДВ» [5], действующий с 2015 г. и покрывающий информацией всю территорию ДФО. Для целей обеспечения специалистов Росгидромета на территории СФО данными дистанционного зондирования, а также результатами их тематической обработки в 2016 г. в СЦ ФГБУ «НИЦ Планета» была начата разработка геопортала «Метео-Сибирь».

Ориентация на синоптиков, агрометеорологов и гидрологов как главных пользователей ресурса, а на растровые данные дистанционного зондирования как на основной вид информации, изначально накладывало ряд требований на архитектуру и реализацию геопортала, связанных с разнообразием и непрерывным оперативным обновлением данных. Поэтому был задан ряд ключевых принципов построения системы и алгоритмов обработки данных, связанный с особенностями самих данных:

1. Полная автоматизация сбора и подготовки данных, отображаемых на геопортале.
2. Распределенная параллельная обработка данных для достижения максимальной производительности и равномерного распределения нагрузки на вычислительные ресурсы.
3. Планирование обработки с приоритетом для самых свежих данных при перегрузке системы (например, после сбоя в сетях передачи данных).
4. Организация управления всеми хранилищами данных в режиме окна данных по времени для исключения переполнения хранилищ, всегда имеющих ограниченную емкость.
5. Возможность неограниченного масштабирования в ширину серверной подсистемы для повышения производительности.
6. Растровая природа основных видов данных и необходимость их наложения друг на друга требует широкого применения работы прозрачными пиксе-

лями или других, незаслоняющих подложку видов представления данных, таких как изолинии, сетки числовых значений данных и т. п.

7. Наличие доступа к полному исходному коду всех частей системы для возможности поиска и исправления проблемных с точки зрения производительности системы.

Методы и материалы

На настоящее время свободных программных средств для организации геопортала с растровым наполнением существует немного. Одним из самых известных и популярных является GeoServer [6], разрабатывающийся на платформе Java с 2001 г. и являющийся наиболее полной эталонной реализацией множества стандартов OGC. Альтернативным подходом к реализации серверных подсистем геоинформационных веб-приложений является CGI приложение MapServer, разрабатывающееся на платформе C/C++ с 1994 г. в Университете Миннесоты и финансировавшееся в начале по программе ForNet NASA [7]. В любом случае, это программное обеспечение реализует хоть и существенную, но лишь часть полной информационной системы. Так, оба пакета требуют внешних СУБД для управления и поиска данных, среди безусловных лидеров из свободного ПО являются СУБД PostgreSQL совместно с расширением для обработки пространственной и растровой информации PostGIS [8].

Для реализации клиентской части веб-приложения существуют и непрерывно совершенствуются библиотеки на основе платформы ECMAScript для обозревателей интернета. Из свободных следует отметить две наиболее распространённых Leaflet и Open Layers. Библиотека Leaflet, впервые выпущенная в 2011 г. [9], является относительно простой и легкой в применении. Библиотека Open Layers [10] разрабатывается с 1997 г. и является, возможно, самой многофункциональной среди свободных. Также существуют и проприетарные, но бесплатные при определённом шаблоне использования библиотеки, например, от Google Maps API [11] или ArcGIS Maps SDK [12].

Из всего перечисленного в геопортале «Метео-Сибирь» применяются:

- MapServer – как основа графической подсистемы работы с векторными объектами;
- PostgreSQL и PostGIS – как система управления векторными данными;
- ESRI API for JavaScript версии 3.37 – как основа веб-интерфейса пользователя.

Растровая подсистема серверной части геопортала является полностью оригинальной. Это связано с большим объемом подготовительной обработки данных, которая в нашем случае интегрирована в процесс подготовки тайловых покрытий, а также с требованием высокой производительности.

Сервисы подготовки и представления тайловых покрытий написаны как параллельные программы, взаимодействующие для передачи данных, в основном, через общие каталоги. В них применяются различные парадигмы параллельного программирования [13], такие как: «читатели-писатели», «портфель задач», «производители и потребители», – реализованные на C++ с использованием

стандартной библиотеки C++. Сервисы работают под управлением ОС Windows Server x64.

Система в серверной части логически подразделяется на две, практически, никак не связанные между собой части: векторную и растровую. Результаты их работы объединяются только в пределах окна геопортала в виде накладывающихся друг на друга слоёв. В целом, взаимодействие компонентов системы выглядит следующим образом (рис. 1).

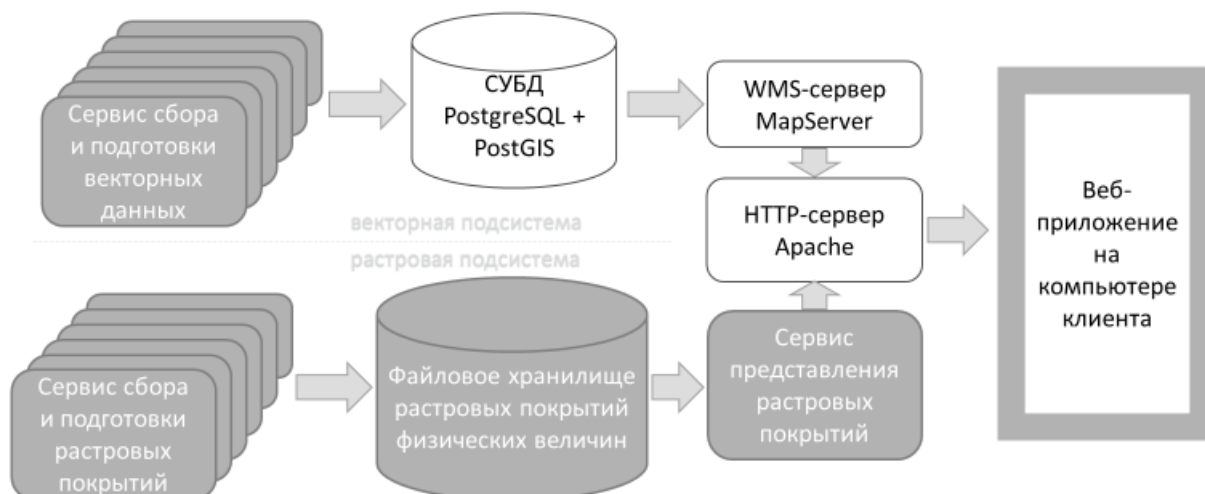


Рис. 1 Общая схема архитектуры геопортала «Метео-Сибирь»

Векторная часть системы включает:

- Сервисы сбора данных и актуализации базы данных.
- Хранилище данных на основе PostgreSQL и PostGIS.
- WMS сервис на основе MapServer.
- Сервис SQL-запросов.

Семь сервисов сбора и актуализации базы данных непрерывно загружают из общедоступных источников, получают по каналам связи или в результате обработки первичной информации данные и наполняют ими базу данных. Информация в БД либо непрерывно накапливается (самый распространённый случай), либо хранится ограниченное время в пределах скользящего окна (в основном очень объёмные или быстро устаревающие данные). Накопленные данные отображаются WMS-сервисом на основе MapServer-а в виде карт, запрашиваемых на весь отображаемый экстенд и накладываемых поверх растровых слоёв. Сервис SQL-запросов форматирует в JSON результаты исполнения SQL-запросов и используется в различных вспомогательных целях, например, отображения календаря наличия данных.

Растровая часть рассчитана на подготовку, хранение и отображения тайлов, представленных в квадратной сетке тайлов со стороной 256 пикселей в проекции Веб-Меркатор и включает:

- Сервисы подготовки тайловых покрытий различных видов данных.
- Сервис управления хранилищем и представления тайловых покрытий.

Сервисы подготовки тайловых покрытий – каждый для своего вида информации (всего 11 сервисов) – выполняют 3 главные функции: загрузка и управлением кэшем исходных данных из источника; подготовка данных (входное преобразование, коррекция, нормализация и т. п.) и, собственно, проектирование и нарезка тайлов. Растровые данные в системе существуют преимущественно в виде тайловых покрытий физических величин (альбедо, радиационной температуры, интенсивности осадков и т.п.), но встречаются и непосредственно в готовом виде цветных картосхем. Один файл в хранилище тайлов соответствует одной физической величине одного канала спутникового радиометра или другого продукта и содержит, помимо самих тайлов, словарь пар ключ-значение метаданных покрытия и карту исходных и пирамидальных тайлов. Формат файла собственный.

Кэши загруженных из исходных источников растровых данных, в основном, применяются, если данные находятся далеко, и процесс загрузки длительный, а также для исключения параллельной загрузки для нескольких процессов обработки внутри одной организации. Внутренние для организации данные не кэшируются, и обрабатываются непосредственно из их местонахождения.

Предварительная подготовка растровых данных при нарезке тайловых покрытий обычно включает необходимые операции из следующего списка:

1. Предварительное преобразование (для формата GeoTIFF).
2. Нормализация освещенности (коррекция на высоту Солнца).
3. Атмосферная коррекция.
4. Обрезка по углу Солнца (видимые, ближние ИК каналы).
5. Обрезка по углу зрения (данные геостационарных КА).
6. Расчет дополнительных тематических продуктов.

Предварительное преобразование необходимо для перевода обычно целочисленного значения пикселя в физическую величину в выбранных единицах представления. При нарезке, в зависимости от характеристик данных, пиксели тайлов хранятся в целочисленном, целочисленном масштабированном (с линейным преобразованием) или вещественном виде, что отмечается в метаданных. Практически все файлы содержат данные, сжатые на уровне тайлов по алгоритму GZIP, PNG, JPEG или JPEG с отдельным сжатием канала прозрачности по алгоритму GZIP.

Сервис управления хранилищем и представления тайловых покрытий является самым сложным из всех сервисов системы. Он выполняет одновременно и параллельно следующие функции:

1. Сбор подготовленных тайловых покрытий от сервисов их подготовки.
2. Обнаружение и замену идентичных покрытий новыми версиями.
3. Удержание указанного размера хранилища путём удаления устаревших файлов с вычислением штрафной функции по указанному алгоритму.
4. Перераспределение файлов между томами хранилища.

5. Представление тайлов веб-интерфейсу геопорталов путем сборки цвето-синтезированного композита из последовательности покрытий по указанным в запросе параметрам.

6. Обработка запросов о составе и метаданных покрытий.

Сбор покрытий реализован параллельно, в парадигме «портфеля задач» [13], исполняемых ограниченным числом работников для управления нагрузкой на сети и тома на запись. Функции со второй по четвертую из-за блокирования каталога покрытий на запись на каждом этапе исполняются последовательно с ограничением времени цикла в 1 минуту. Это необходимо для оперативного сбора покрытий и справедливого выделения процессорного времени другим задачам. Последние две функции исполняются параллельно и независимо от других, общими работниками из множества рабочих потоков (по числу логических ядер в системе). Для быстрой обработки клиентских запросов полный каталог метаданных покрытий хранится загруженным в оперативной памяти.

Результаты и обсуждение

Серверная часть реализованной системы развернута на двух серверах (не считая сетевой инфраструктуры предприятия между серверами и выходным шлюзом). Сервера имеют 96 и 128 ГБ ОЗУ, 24 и 32 ядра в двух ЦПУ Intel E5-2650 v 2 и v 4, соответственно. Под хранилище отведено 19 из 22 ТБ на RAID 6-го уровня на первом сервере.

Первый сервер несет основную нагрузку. На второй сервер вынесена самая большая вычислительная нагрузка – подготовка тайлов приборов с большими объемами и построение изолиний метеорологических полей (самая вычислительно затратная процедура), а также основной сервер PostgreSQL. Еще один сервер PostgreSQL работает исключительно на изолинии метеорологических полей на первом сервере. Базы векторных данных включают несколько тематических баз данных, объемом до 70 ГБ.

Хранилище растровых данных функционирует в режиме скользящего окна имеющихся данных. На настоящее время, в нем содержится свыше 470 тыс. покрытий на территорию Сибири (данные геостационарных спутников – преимущественно на восточное полушарие). Примерно, 4 % объема занимают данные накапливаемых исторических серий исследовательских наблюдений, тогда как остальные 96 % непрерывно обновляются. Суточный объем поступающей в хранилище информации составляет около 140 ГБ, и основное место занимают данные семи геостационарных спутников с интервалом съемки от 15 минут до 1 часа. Объем хранилища обеспечивает длину скользящего окна в 114 суток.

Оценки производительности растровой части с помощью утилиты `ab` из поставки сервера Apache показали максимальную среднюю скорость отдачи тайлов около 260 с^{-1} (тайлов в секунду) при измерении в пределах одного сегмента ЛВС. Анализ длительностей различных фаз обработки запросов показал, что 90 % времени тратится на сжатие PNG, около 2 % на чтение тайлов из файла, примерно так же, 2 % – на ожидание исключительного доступа к файлу (поскольку нельзя в одном процессе читать один файла в разных местах одновременно). Переход

на запросы тайлов в формате JPEG радикально повышает производительность до 2200 с^{-1} в тех же условиях, что объясняется как сокращением затрат на сжатие, так и передачей файлов в десятки раз меньшего объема по сети.

В случае наличия интернета как среды передачи данных с многочисленными шлюзами на пути скорость отдачи тайлов, естественно, снижается. Например, для клиента в Красноярске со скоростью подключения к провайдеру 100 МБ/с при сохранении остальных условий скорость отдачи тайлов составляет уже около 65 с^{-1} и не зависит от размера. Это говорит о преимущественном влиянии задержки или ограничений на шлюзах на число одновременных подключений, а не объемов передаваемой информации.

Быстродействие векторной подсистемы в расчете на один запрос существенно ниже, но и запросов в ней в несколько десятков раз меньше, поскольку запрашиваются слои WMS целиком на экран, а не тайлами. Оценить среднее время отрисовки слоев сложно из-за существенной разницы между слоями с различным наполнением. Но, например, для изолиний метеорологических полей оно составляет в районе 1 с. При совместном просмотре растрового слоя с таким же векторным, время их обновления на весь экран показывает примерную соразмерность в производительности векторной и растровой подсистем.

Опыт практического применения геопортала показал, что при подобной архитектуре и производительности на данной конфигурации без существенного ухудшения производительности могут работать около 100 клиентов одновременно, что примерно соответствует числу пользователей. Предельное число одновременных клиентов может достигать до 1000, время ожидания отклика при этом доходит до минуты.

Заключение

Таким образом, используя свободное ПО совместно со специально разработанным, можно на достаточно ограниченном по современным меркам оборудовании развернуть тематический геопортал на основе оперативных растровых данных ДЗЗ. Геопортал способен хранить и отображать данные за несколько месяцев и обслуживать несколько десятков пользователей. При этом путем масштабирования аппаратного обеспечения вширь можно пропорционально увеличивать число обслуживаемых клиентов.

В перспективе, при соблюдении соответствующих организационных мероприятий можно на этом же ПО строить геопорталы, территориально распределенные по нескольким узлам. При этом нагрузка по подготовке покрытий может быть распределена между узлами по принципу близости к источнику данных, а каждый узел при этом будет способен собирать данные из всех остальных узлов. Так что все они будут отображать одинаковую картину, а клиенты смогут заходить на ближайший для себя узел, снижая нагрузку на сети передачи данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Explorer-6: сайт [Электронный ресурс]. / NASA – URL: <https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1959-004A> (дата обращения 13.04.2023).

2. George D. Anderson The first weather satellite picture // *Weather*, – 2010. – 65: 87-87. – DOI: <https://doi.org/10.1002/wea.550>.
3. New geostationary meteorological satellites – Himawari-8/9. [Электронный ресурс] / Japan Meteorological Agency – 2014. URL: https://www.jma.go.jp/jma/jma-eng/satellite/news/himawari89/himawari89_leaflet.pdf (дата обращения 13.04.2023).
4. Основные информационные ресурсы и продукция Росгидромета – Продукция – Росгидромет. сайт [Электронный ресурс] / Росгидромет – URL: <https://www.meteorf.gov.ru/product/info/> (дата обращения 13.04.2023).
5. ГИС МЕТЕО ДВ: сайт [Электронный ресурс] / Дальневосточное управление по гидрометеорологии и мониторингу окружающей среды – URL: <https://meteo-dv.ru/> (дата обращения 13.04.2023).
6. GeoServer: сайт [Электронный ресурс]. / Open Source Geospatial Foundation – URL: <https://geoserver.org/> (дата обращения 17.04.2023).
7. Vatsavai R. R. Shekhar S. Burk T. E. Lime S. UMN-MapServer: A high-performance, interoperable, and open source web mapping and geo-spatial analysis system// *Geographic Information Science* – 2006, Vol. 4197. – с. 400–417. – DOI: 10.1007/11863939_26
8. Obe R. O, Hsu L. S. *PostGIS in Action*. / ред. Sebastian Sterling и Linda Recktenwald. – Geography. Manning. – 2011. 520 с.
9. Leaflet: сайт [Электронный ресурс]. / Володимир Агафонкин – URL: <https://leafletjs.com/> (режим доступа: недоступен из РФ без VPN, дата обращения 17.04.2023).
10. Open Layers – Welcome: сайт. [Электронный ресурс] / Open Source Geospatial Foundation – URL: <https://openlayers.org/> (дата обращения 20.04.2023).
11. Платформа Google Карт | Google Developers: сайт. [Электронный ресурс] / Google – URL: <https://developers.google.com/maps?hl=ru> (дата обращения 20.04.2023).
12. Overview | ArcGIS Maps SDK for JavaScript 4.26 | ArcGIS Developers: сайт. [Электронный ресурс] / ESRI – URL: <https://developers.arcgis.com/javascript/latest/> (дата обращения 20.04.2023).
13. Эндрюс Г. Р. Основы многопоточного, параллельного и распределённого программирования / Г. Р. Эндрюс. – Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 512 с.

© В. Ю. Ромасько, 2023