

Organization of Functioning of the Space Monitoring System of Emercom of Russia Regarding Using Information Technologies of Distributed Storage and Multithreaded Data Processing to Improve Quality and Efficiency of Decision Making in Forecasting and Eliminating Emergency Situations

Yaroslav V. Alekseenko^{1, 2}

¹ Russian Agency for Disaster and Emergency Management, Moscow, Russia
alex.zik@mail.ru

² Saint-Petersburg University of State Fire Service of EMERCOM of Russia
Saint-Petersburg, Russia

Abstract

Emercom of Russia actively applies results of remote sensing to ensuring protection of the population and territories from emergency situations. The observed growth of volume of the obtained data demands to provide guaranteed data storage and also to provide operational data processing. During the forecasting and emergency response the population, objects of economy and the environment need to take in the shortest possible time measures for prevention of loss of human life and causing material damage. It can be reached only by reduction of time from the moment of detection of threat prior to holding necessary actions. For a solution of this problem in article the model of creation of a distributed file system and the distributed multithreaded data processing received by the System of space monitoring of Emercom of Russia is offered. The offered model allows to provide guaranteed data storage, to reduce time for data processing and also to provide flexible scaling of hardware.

Keywords: data processing, data storage, processing of space pictures, the distributed data storage, multithreaded processing, the system of space monitoring, Emercom of Russia

ОРГАНИЗАЦИЯ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ КОСМИЧЕСКОГО МОНИТОРИНГА ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ МЧС РОССИИ В АСПЕКТЕ ПРИМЕНЕНИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ РАСПРЕДЕЛЁННОГО ХРАНЕНИЯ И МНОГОПОТОЧНОЙ ОБРАБОТКИ ДАННЫХ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА И ЭФФЕКТИВНОСТИ РЕШЕНИЙ, ПРИНИМАЕМЫХ ПРИ ПРОГНОЗИРОВАНИИ И ЛИКВИДАЦИИ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ

Я.В. Алексеенко^{1,2}

¹ Национальный центр управления в кризисных ситуациях, Москва, Россия
alex.zik@mail.ru

² Санкт-Петербургский университет государственной противопожарной службы МЧС России
Санкт-Петербург, Россия

МЧС России активно применяет результаты дистанционного зондирования для обеспечения защиты населения и территорий от чрезвычайных ситуаций. Наблюдающийся рост объёма получаемых данных требует обеспечения гарантированного хранения данных, а также обеспечить оперативную обработку данных. При прогнозировании и реагировании на чрезвычайные ситуации требуется в кратчайшее время принять меры по недопущению человеческих жертв и нанесения материального ущерба населению, объектам экономики и окружающей среде. Это может быть достигнуто только сокращением времени от момента обнаружения угрозы до начала проведения необходимых мероприятий. Для решения данной проблемы в статье предлагается модель построения распределённой файловой системы и распределённой многопоточной обработки данных, получаемых Системой космического мониторинга МЧС России. Предлагаемая модель позволяет обеспечить гарантированное хранения данных, сократить время на обработку данных, а также обеспечить гибкое масштабирование аппаратных средств.

Ключевые слова: обработка данных, хранение данных, обработка космических снимков, распределенное хранение данных, многопоточная обработка, система космического мониторинга, МЧС России.

Введение

Применение результатов дистанционного зондирования (РДЗ) применяются во многих областях науки и экономики (сельское хозяйство, лесное хозяйство и т.д.). РДЗ также активно и успешно применяются для прогнозирования чрезвычайных ситуаций (ЧС), а также оценки масштабов ЧС [1-3]. В МЧС России более 20 лет функционирует Система космического мониторинга ЧС (СКМ ЧС) [4]. СКМ ЧС представлена 4 центрами приёма и обработки РДЗ МЧС России, расположенных в городах Москва, Красноярск, Владивосток и Вологда. Также СКМ ЧС интегрирована с совместными центрами приёма и обработки РДЗ МЧС России и Госкорпорации «Роскосмос». Для обеспечения эффективного прогнозирования и реагирования на ЧС требуется получение данных о ранних предвестниках ЧС, а также оперативной информации с наибольшей

оперативностью. РДЗ являются одним из главных источников оперативной и достоверной информации при реагировании на ЧС [5-8]. Для повышения достоверности и оперативности СКМ ЧС использует следующие источники информации:

- данные, получаемые центрами приёма и обработки РДЗ МЧС России;
- данные, получаемые от отечественной группировки космических аппаратов дистанционного зондирования Земли (КА ДЗЗ);
- данные, получаемый с борта Международной космической станции в рамках проведения космического эксперимента «Сценарий»;
- данные из открытых источников информации.

В последнее время наблюдается рост количества функционирующих КА ДЗЗ, помимо этого улучшается и качество получаемых с КА ДЗЗ данных. Эти факторы приводят к значительному увеличению объёма получаемой информации. Учитывая необходимость обеспечения максимальной оперативности получения и обработки РДЗ о ЧС, необходимо решение проблем гарантированного хранения получаемых данных (с минимизацией времени при выходе из строя устройств и/или узлов хранения данных) и обеспечения оперативной обработки данных.

В связи с вышеизложенным, для оперативного обеспечения потребителей РДЗ необходимо создание распределённой файловой системы (РФС), а также реализовать распределённую многопоточную обработку данных (РМОД).

Структура и задачи СКМ ЧС

Структура СКМ ЧС представлена на рисунке 1.

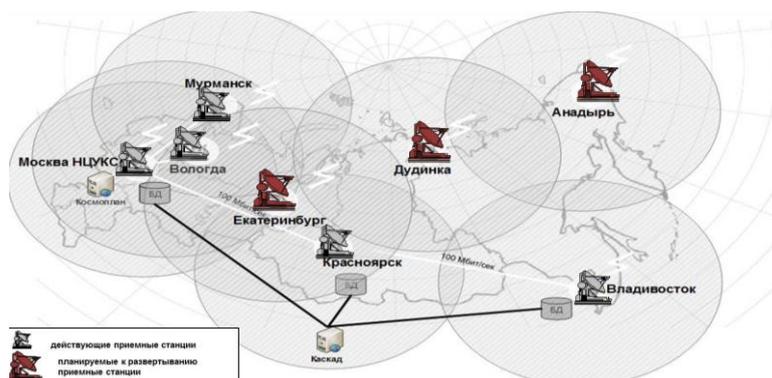


Рис. 1. Структура СКМ ЧС

На рисунке 2 представлены источники получения информации, используемые в СКМ ЧС.



Рис. 2. Источники информации для СКМ ЧС

Основными, по оперативности и объёму получения данных, являются станции СКМ ЧС, а также орбитальная группировка КА ДЗЗ Российской Федерации. При реагировании на

возникающие ЧС, а также для своевременного реагирования на прогнозируемые ЧС необходимо в кратчайшие сроки довести информацию до потребителей – территориальных органов МЧС России, а также органов власти субъектов Российской Федерации и муниципальных образований. Для решения этой задачи в СКМ ЧС применяются геоинформационные порталы «Космоплан» и «Каскад».

СКМ ЧС решает следующие задачи:

- мониторинг природных пожаров;
- мониторинг паводковой обстановки и наводнений;
- мониторинг потенциально-опасных территорий и объектов;
- мониторинг разливов нефтепродуктов на суше и акваториях;
- поиск аварийных объектов на суше и акваториях.

В настоящее время вся получаемая СКМ ЧС информация аккумулируется в базах данных и хранилищах данных. Анализ функционирования СКМ ЧС показал, что существует проблема обеспечения гарантированного хранения данных, а также наличие больших временных затрат на обработку получаемых данных и на передачу информации о складывающейся оперативной обстановки потребителям. Для эффективного функционирования СКМ ЧС необходимо обеспечить интеграцию РФС и РМОД. Общая схема реализации данного подхода представлена ниже

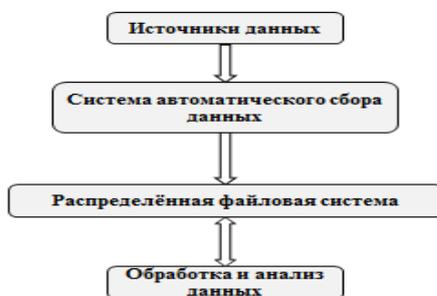


Рис. 3. Общая схема реализации сбора, хранения и обработки данных

Распределённая файловая система СКМ ЧС

Для обеспечения хранения данных в СКМ ЧС применяются хранилища данных, в которых для гарантированного хранения применяются серверное оборудование с расширенным объёмом дискового пространства, которые оснащены специализированными аппаратными средствами (контроллерами), обеспечивающими функционирование избыточных массивов независимых дисков (ИМНД). Узлы хранения данных распределены по территории Российской Федерации (находятся в центрах приёма и обработки СКМ ЧС) [9]. Такая структура имеет ряд недостатков.

Во-первых, при выходе из строя устройств хранения данных резко уменьшается производительность узла хранения данных. Это связано со спецификой функционирования ИМДН. Так как требуется время на перестроение структуры ИМДН и пока выполняется этот процесс - снижается скорость чтения и записи на данный ИМДН. Также для обеспечения доступа к данным при выходе из строя одного из узлов хранения данных необходимо обеспечить доступ хранящихся на нём данных потребителям [10, 11].

Во-вторых, такой подход к построению хранилищ данных требует применение дорогостоящего оборудования, иметь в наличии запасные устройства хранения данных а также обеспечить функционирование гетерогенных устройств хранения данных. Также необходим постоянный контроль за состоянием ИМДН, так как при выходе из строя критического количества устройств хранения данных ИМДН разрушается и данные теряются [12, 13].

Для решения данных проблем предлагается создание РФС, схема функционирования которой представлена на рисунке 4.

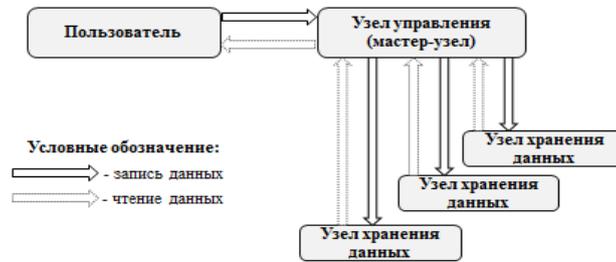


Рис. 4. Общая схема функционирования РФС

РФС функционирует поверх файловой системы операционной системы. За сохранность и гарантированное хранения данных при аппаратных сбоях отвечает сама РФС. РФС состоит из узла управления (УУ), который хранит метаданные о хранящихся данных (имя файла, место хранения, тип репликации и т.п.), а также узлов хранения данных (УХД), на которых непосредственного хранятся данные.

Поступающие на запись данные разбиваются на блоки одинакового размера (размер блока фиксирован и может устанавливаться администратором РФС), за исключением последнего блока. После этого, в соответствии с выбранным коэффициентом репликации (по умолчанию он равен 3), блоки записываются на УХД. При этом УУ ведёт не только учёт места записи блоков, но и старается распределять блоки так, чтобы они находились либо на разных физических средствах (серверах), либо на УХД, которые территориально размещены в разных местах (серверных стойках,

дата-центрах). Метаданные УУ хранит в оперативной памяти. Это необходимо для оперативного получения пользователями информации о запрашиваемых данных. Анализируя данную структуру РФС, несложно заметить, что «узким горлышком» является УУ. При выходе его из строя РФС не сможет функционировать. Для решения данной проблемы необходимо применять дублирующий УУ, который в случае выхода из строя основного УУ обеспечит функционирование РФС. Для чтения данных пользователь обращается к УУ и запрашивает список УХД, которые хранят блоки запрашиваемых данных и версию реплик необходимых блоков данных. После этого пользователь обращается к УХД для получения требуемых данных. При этом УУ предоставляет пользователю УХД, расположенных максимально близко к нему. Если один из УХД не доступен либо хранит устаревшую реплику блока (это может произойти, если УХД был недоступен и УУ не успел провести обновление репликации блоков) пользователь вновь обращается к УУ для получения актуальной версии реплики блока.

Данный подход позволяет решить проблемы гарантированного хранения данных; использования не дорогостоящего оборудования; обеспечения реализации РМОД.

Таким образом, предлагаемая модель РФС имеет следующие преимущества:

- функционирует на кластере серверов (недорогих, без использования дорогостоящих аппаратных средств);
- для пользователя РФС представлена в виде «одного большого диска»;
- представляет собой приложение пользовательского уровня;
- обеспечивает устойчивость к ошибкам (данные не теряются при выходе из строя устройств хранения данных или серверов).

РФС хорошо подходит для хранения файлов, имеющих большой объём (сотни мегабайт, гигабайты, терабайты, петабайты). В РФС реализован паттерн «один раз записал/много раз прочитал», процесс чтения данных выполняется последовательно. Это позволяет сократить накладные расходы на время, необходимое для перестановки головки чтения устройств хранения данных.

Однако РФС имеет ряд недостатков.

Во-первых, РФС не позволяет быстро получить большой объём данных в режиме «онлайн». РФС больше подходит для «офлайн» получения больших объёмов данных. Для решения данной проблемы возможно применять системы управления базами данных класса NoSQL.

Во-вторых, хранить в РФС большое количество файлов с объёмом менее 100 мегабайт. Это связано с самой структурой РФС. Так как УУ хранит всю метainформацию о хранящихся файлах в оперативной памяти, то при большом количестве файлов может быть достигнут предел выделенной оперативной памяти и РФС не сможет выполнять записать новых файлы. Для решения данной проблемы целесообразно применять метод объединения нескольких файлов в один блок. Таким образом, один блок РФС будет содержать несколько файлов и УУ сможет функционировать в «штатном» режиме.

В-третьих, РФС не подходит для многопоточной записи. В РФС существует только один процесс записи на файл, а также данные дописываются в конец файла, т.е. нет записи по смещению.

Несмотря на наличие недостатков РФС позволяет эффективно организовать хранение данных и доступ пользователей к ним. Также немаловажен тот факт, что РФС позволяет интегрировать РФС с РМОД.

Распределённая многопоточная обработка данных в СКМ ЧС

Развитие современных аппаратных средств и информационных технологий позволяет выполнять параллельную обработку данных. Большая часть программного обеспечения для параллельной обработки данных использует возможности центрального процессора ЭВМ, также в узконаправленных областях науки для параллельной обработки применяют графические процессоры. Параллельная обработка позволяет значительно сократить время на обработку данных и получения необходимой информации [14-17]. Однако при реализации параллельной обработки существуют определённые проблемы:

- наличие квалифицированных программистов;
- наличие соответствующих аппаратных средств;
- сложность масштабирования.

Для выполнения эффективной параллельной обработки необходимо наличие значительных вычислительных мощностей, однако, зачастую данные хранятся в хранилищах данных, которые не имеют необходимых вычислительных мощностей. Поэтому данные, которые необходимо обработать, передают от места хранения к месту обработки. После обработки данных необходимо передать результат обработки в хранилище данных. Подобная схема организации обработки данных приводит к значительным накладным расходам на время передачу данных между узлами.

Ввиду этого, необходима интеграция файловой системы и процесса обработки данных, для минимизации времени на передачу данных [18]. Для решения этой проблемы предлагается РМОД, главной особенностью которой является подход «передача кода к данным», а не «передача данных к коду» и интеграция с РФС. Общая схема функционирования РМОД представлена на рисунке 5.

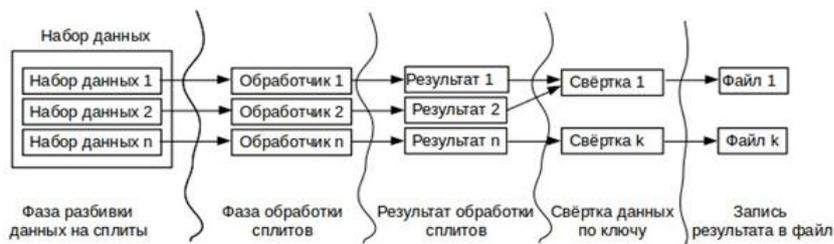


Рис. 5. Распределённая многопоточная обработка данных

Входные данные представлены набором данных, которые разбиваются на сплиты (независимые друг от друга части) и передаются на этап обработки. Каждый обработчик передаётся к месту хранения данных. При обработке результаты формируются по кортежу <ключ, значение>. После обработки всех сплитов полученные кортежи передаются на этап

свёртки. В процессе свёртки все полученные кортежи объединяются по соответствующему ключу. После окончания этапа свёртки полученный результат записывается в РФС.

Преимуществом предложенной РМОД является то, что для обработки данных требуется определить лишь две функции, которые будут выполняться на этапе обработки и свёртки соответственно. Ввиду того, что входные данные разбиваются на сплиты, возможно организовать многопоточную обработку. А интеграция с РФС (файлы, которые необходимо обработать, размещены поблочно и распределены между УХД) позволяет осуществлять распределённую обработку – достаточно лишь передать код (скрипт) к данным. Такой подход также позволяет легко проводить масштабирование аппаратных средств - достаточно добавить дополнительные устройства хранения данных или же дополнительный УХД, что не требует изменения функций обработки и свёртки.

Оценка применения РФС и РМОД

Для проведения оценки применения РФС и РМОД в СКМ ЧС были созданы 4 УХД и 1 УУ, коэффициент репликации - 3. В качестве входных данных были выбраны космические снимки (формат GEOTIFF), полученные с КА ДЗЗ типа «Ресурс-П» и «Канопус-В» (общий объём – 140 Гбайт). УХД и УУ имели двухядерный центральный процессор с тактовой частотой 2,2 ГГц, объём оперативной памяти – 8 Гбайт, объём хранилища данных – 600 Гбайт. Оценка предлагаемой модели построения РФС и РМОД осуществлялась с помощью реализации алгоритма детектирования изменений на местности. При этом выполнялось поэтапное увеличение количества УХД. Результаты проведённой оценки представлены на рисунке 6.

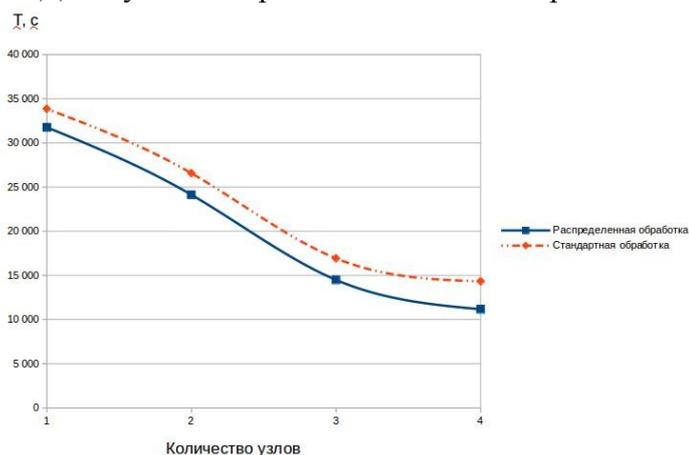


Рис. 6. Результаты оценки применения РФС и РМОД

Заключение

В последнее время наблюдается взрывной рост объёмов получаемой и обрабатываемой информации. Этот тренд наблюдается и в области космического мониторинга. В МЧС России активно применяется СКМ ЧС, позволяющая снижать риски возникновения угроз населению и территориям за счёт проведения своевременных превентивных мероприятий. В тоже самое время, с ростом получаемых данных, существуют проблемы обеспечения максимальной оперативности доведения результатов обработки РДЗ до потребителей, а также гарантированного хранения получаемых данных. Для решения данных проблем в СКМ ЧС была создана РФС и РМОД. Оценка применения РФС и РМОД в СКМ ЧС показала актуальность применения данного подхода.

В тоже время, при размещении данных в РФС применяется коэффициент репликации равный 3 (по умолчанию). Это приводит к увеличению требуемого объёма хранилища данных. При такой организации хранения данных неизбежные накладные расходы (ННР) составляют 200%. Для сокращения ННР необходимо изменить подход к хранению данных. Одним из путей

снижения ННР и обеспечения такого же уровня сохранности данных может выступать применение кодов избыточности (к примеру, основанных на кодах Рида-Соломона), что по предварительным расчётам снизит ННР до 40%. Автор проводит исследования для решения данной проблемы.

References

- [1] Rozenberg I.N. Space Monitoring. *Slavyanskii forum*, 2016, № 2, pp. 216-222. (In Russian).
- [2] Toth C., Józków G. Remote Sensing Platforms and Sensors: A Survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, T. 115, pp. 22-36. DOI: 10.1016/j.isprsjprs.2015.10.004.
- [3] Giri C. P. Remote Sensing of Land Use and Land Cover: Principles and Applications. *CRC Press*, 2016. DOI: 10.1201/b11964.
- [4] Fakhmi Sh.S., Kryukova M.S., Alekseenko Ya.V., Salem Ali, Video System of Space Monitoring of Emercom of Russia for Managerial Decision Making. *Technologies for Constructing Transport systems*, 2018, pp. 236-243. (In Russian).
- [5] Yu.I. Shokin et al. An information system for acquisition, processing and access to satellite data and its applications in environmental monitoring. *Computational Technologies*, 2015, 20 (5). (In Russian).
- [6] Kolesenkov A. N. Modern Approaches to the Processing of Data in the Construction of Information Systems of Environmental Monitoring. *Izvestiya Tula State University*, 2016, №9. (In Russian).
- [7] Hoque M. A. A. et al. Tropical cyclone disaster management using remote sensing and spatial analysis: A review. *International journal of disaster risk reduction*, 2017, T. 22, pp. 345-354. DOI: 10.1016/j.ijdr.2017.02.008.
- [8] Liu Y., Wu L. Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning. *Procedia Computer Science*, 2016, T. 91, pp. 566-575. DOI: 10.1016/j.procs.2016.07.144.
- [9] Fakhmi Sh.S., Almahrouk Muhib Muhamed, Kryukova M.S., Alekseenko Ya.V. Model of Distributed Processing and Storage of Space Pictures. *Technologies for Constructing Transport systems*, 2018, pp. 231-236. (In Russian).
- [10] Shvachko K. V. et al. Distributed File System Using Consensus Nodes : patent. 9424272 USA, 2016.
- [11] Ramakrishnan R. et al. Azure Data Lake Store: a Hyperscale Distributed File Service for Big Data Analytics. *Proceedings of the 2017 ACM International Conference on Management of Data*. – ACM, 2017, pp. 51-63. DOI: 10.1145/3035918.3056100.
- [12] Kakoulli E., Herodotou H. OctopusFS: A Distributed File System With Tiered Storage Management. *Proceedings of the 2017 ACM International Conference on Management of Data*. – ACM, 2017, pp. 65-78. DOI: 10.1145/3035918.3064023.
- [13] Sherstnev V. S. et al. Development of Distributed File System for Storing Weather Data. *22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. – *International Society for Optics and Photonics*, 2016, T. 10035, pp. 100350G. DOI: 10.1117/12.2249248.
- [14] Mori K. I. et al. Design of a Local Parallel Pattern Processor for Image Processing. *Special Computer Architectures for Pattern Processing*. – *CRC Press*, 2018, pp. 197-210. DOI: 10.1109/AFIPS.1978.71.
- [15] Ma Y. et al. Parallel Programming Templates for Remote Sensing Image Processing on GPU Architectures: Design and Implementation. *Computing*, 2016, T. 98, №. 1-2, pp. 7-33. DOI: 10.1007/s00607-014-0392-y.
- [16] Tan K. et al. GPU Parallel Implementation of Support Vector Machines for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, T. 8, №. 10, pp. 4647-4656. DOI: 10.1109/JSTARS.2015.2453411.
- [17] Li W. et al. GPU Parallel Implementation of Isometric Mapping for Hyperspectral Classification. *IEEE Geoscience and Remote Sensing Letters*, 2017, T. 14, №. 9, pp. 1532-1536. DOI: 10.1109/LGRS.2017.2720778.
- [18] Rathore M. M. U. et al. Real-Time Big Data Analytical Architecture for Remote Sensing Application. *IEEE journal of selected topics in applied earth observations and remote sensing*, 2015, T. 8, № 10, pp. 4610-4621. DOI: 10.1109/JSTARS.2015.2424683.