

Methods and Tools for Determining Delays in the Process of Automated Stream Processing of Satellite Data

Andrey A. Proshin, Mikhail A. Burtsev, Dmitriy A. Kobets, Ilya G. Sychugov

Space Research Institute, Russian Academy of Sciences, Moscow, Russia
Ilya.Sychugov@gmail.com

Abstract

This work is about approaches to detecting and analyzing delays of satellites products obtaining in the process of complex stream processing that were established in the Space Research Institute of the Russian Academy of Sciences. These approaches make it possible to increase the efficiency of information systems that provide access to products obtained on Earth remote sensing data, and also make them more failover in the face of ever-increasing requirements for building such systems. The main reasons for creating specialized tools for detecting delays are considered. It also describes in detail the possible reasons for their occurrence, methods for their detection, and also analyzes the data obtained and notes the reasons why this analysis is necessary. In conclusion, there are examples of the implementation of this system in currently used information systems.

Keywords: Earth remote sensing, satellite data, automated stream processing of satellite data, data processing technologies, analysis of delays in receiving and processing data, error control, BI, OLAP

МЕТОДЫ И ИНСТРУМЕНТЫ ВЫЯВЛЕНИЯ ЗАДЕРЖЕК В ПРОЦЕССЕ АВТОМАТИЗИРОВАННОЙ ПОТОКОВОЙ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ

А.А. Прошин, М.А. Бурцев, Д.А. Кобец, И.Г. Сычугов

Институт космических исследований РАН, Москва, Россия
Ilya.Sychugov@gmail.com

Доклад посвящён описанию методов и инструментов выявления задержек в процессе автоматизированной потоковой обработки спутниковых данных, которые были разработаны за последние несколько лет в ИКИ РАН. Своевременное детектирование задержек в процессе обработки данных их анализ и устранение позволяют повысить оперативность поступления новых данных в информационные системы, предоставляющие доступ к продуктам, получаемых на основе данных дистанционного зондирования Земли. Благодаря решению этих задач повышается эффективность работы информационных систем и достигается большая их отказоустойчивость. В докладе обосновывается необходимость создания специализированных инструментов для детектирования задержек при проведении отдельных этапов обработки спутниковых данных. Описываются наиболее распространенные причины возникновения необоснованных задержек в процессе обработки спутниковых данных и способы их детектирования. Приводятся краткие сведения о реализованных на основе BI-аналитики инструментах для визуального анализа получаемой контрольной информации.

Ключевые слова: дистанционное зондирование, спутниковые данные, потоковая обработка спутниковых данных, технологии работы с данными, анализ задержек получения и обработки данных, контроль ошибок, BI, OLAP

Введение

Современные информационные системы (ИС) работы со спутниковыми данными обеспечивают доступ к большому числу различных информационных продуктов, которые зачастую получают на основе проведения достаточно сложных многоэтапных процедур обработки исходных спутниковых данных [1]. Для решения многих задач мониторинга окружающей среды и природных ресурсов особую ценность представляют оперативные данные, т.е. данные для которых важна минимальная задержка с момента спутниковой съемки до предоставления готовых информационных продуктов заинтересованным в них пользователям. Для этого необходимо своевременно выявлять необоснованные задержки, возникающие в процессах получения данных, их обработки и архивации. Кроме этого необходимо эффективно выявлять те этапы обработки, которые наиболее нуждаются в оптимизации.

Настоящая статья посвящена разработке методов и инструментов, позволяющих своевременно выявлять задержки в процессе сложной автоматизированной обработки спутниковых данных. Эффективная реализация этих методов практически невозможна без детального протоколирования всех этапов проведения реализованных обработок спутниковых данных. Для наглядного анализа такой сложной многомерной информации, на наш взгляд, хорошо подходят различные BI-технологии [2], позволяющие получить доступ к данным

посредством разнообразных интерактивных отчетов [3]. Именно на основе этих технологий были разработаны основные инструменты для анализа временных задержек, описанные в настоящей работе [4].

Автоматизированная потоковая обработка

В настоящее время активно развиваются различные системы дистанционного мониторинга [5]. Одной из особенностей этих систем является то, что они должны обеспечивать оперативное получение и обработку значительных объемов спутниковых данных, используемых для получения различных спутниковых информационных продуктов [6]. Для обеспечения стабильной работы подобных информационных систем необходимо обеспечить полностью автоматизированную потоковую обработку поступающих спутниковых данных, максимально эффективно использующую имеющиеся в наличии вычислительные ресурсы [7]. К сожалению, заранее предусмотреть и предупредить все возможные неполадки в работе таких систем, включая необоснованные задержки в процессе обработки данных, на наш взгляд, практически невозможно. Это приводит к необходимости создания и постоянного совершенствования специальных инструментов для детального контроля над всеми этапами обработки и выявления причин необоснованных задержек при получении требуемых информационных продуктов. На основе многолетнего опыта организации многопотоковой автоматизированной обработки спутниковых данных в ИКИ РАН [8], были созданы и внедрены специальные инструменты, предназначенные для оперативного выявления задержек в процессе обработки спутниковых данных и их дальнейшего анализа с целью устранения неполадок в работе системы и ее оптимизации.

Причины задержек и их выявление

Учитывая объёмы поступающей информации и сложность самого процесса обработки, практически невозможно все операции по обработке данных проводить на одном узле. На практике даже для данных одного спутника в этом процессе могут участвовать несколько серверов обработки и хранения, что существенно повышает как вероятность возникновения сбоя в получении конечных продуктов, так и сложность в оперативном выявлении узла, на котором он возник и причины его появления. Задержки в поступлении продуктов обработки спутниковых данных могут возникнуть по целому ряду различных причин, таких как: сетевые проблемы, аппаратные сбои, несовершенство программного обеспечения, в том числе и стороннего, а также вследствие ошибок в поступающих входных данных. Такие сбои и ошибки могут привести к самым разным негативным последствиям, приводящим как к отсутствию поступления в архивы тех или иных типов информационных продуктов, так и к выходу из строя целых блоков обработки и даже полного прекращения поступления новых данных в архивы.

Для детектирования серьезных сбоев в работе системы обработки спутниковых данных используются достаточно универсальные средства, контролирующие поступление в архивы исходных спутниковых данных и всех информационных продуктов, получаемых на их основе. Настоящая статья посвящена более детальному анализу работы системы обработки, учитывающему специфику реализации конкретных цепочек обработки спутниковых данных, которые зачастую довольно существенно отличаются друг от друга. Для этого необходимо отдельно контролировать каждый из этапов обработки, включающий в себя ограниченный набор операций над спутниковыми данными и представляющий из себя элементарный блок преобразования данных. Для каждого из таких блоков должны быть описаны поступающие на вход данные и данные, которые должны получиться в результате проведения обработки. Для того чтобы выявить необоснованные задержки в процессе обработки необходимо протоколировать для каждого из таких этапов время поступление входных данных и время получения результатов. При этом такая информация должна отслеживаться для каждого экземпляра (сеанса или сцены) исходных спутниковых данных.

Для решения описанных выше задач в ИКИ РАН был разработан программный пакет `data_monitoring`, предназначенный как для сбора всей необходимой контрольной информации, так и для ее дальнейшего анализа. В рамках этого пакета реализованы программные вызовы, позволяющие заносить всю требуемую для выявления задержек информацию в специальную базу данных, состоящую из двух основных таблиц. Первая из них предназначена для хранения информации обо всех поступающих на обработку сеансах спутниковых данных и включает в себя такие атрибуты, как время съемки, спутник, прибор и станция приема. Вторая таблица предназначена для протоколирования для каждого из сеансов всех этапов при проведении их обработки и содержит информацию о типе обработки, времени ее начала и времени ее окончания.

Анализ

Для визуального анализа детальной информации о выполнении всех этапов выполнения реализованных цепочек потоковой обработки спутниковых данных в ИКИ РАН был создан блок статической информации на основе использования BI-аналитики. Данный блок предоставляет пользователю широкий инструментальный интерактивных WEB-интерфейсов, которые позволяют производить выборку интересующих данных, автоматически предоставлять их в удобной для анализа форме (график, гистограмма, таблица, карта), работать с выбранными данными (фильтровать по различным признакам, сортировать) и выгружать полученный результат (Excel, Word, PDF, HTML, PNG).

OLAP-кубы представляют из себя многомерное хранилище анализируемых данных. Процессы формирования многомерных OLAP-кубов переводят данные из реляционной модели хранения в многомерную, в которой реализуются два типа показателей: измерения и факты. Измерения – это дискретный набор характеристик, описывающий исследуемый процесс (например, данные каких спутников обрабатывались на сервере), по которым возможно фильтровать факты. Факты – это численные значения, характеризующие исследуемый процесс (например, загруженность сервера, или продолжительность обработки данных). Процессы формирования интерактивных отчетов визуализируют данные из OLAP-куба в соответствии с функциями свертки в иерархии измерений, в виде интерактивного WEB-интерфейса (таблица, диаграмма, графика, карты).

Реализованные на основе использования BI-аналитики интерфейсы позволяют для каждого из контролируемых центров обработки спутниковых выбрать интересующие оператора типы обрабатываемых данных и визуально в виде гистограмм представить времена выполнения каждого из этапов обработки данных. Пример отчета о задержках на различных этапах обработки сеансов спутниковых данных приведен на рис. 1.

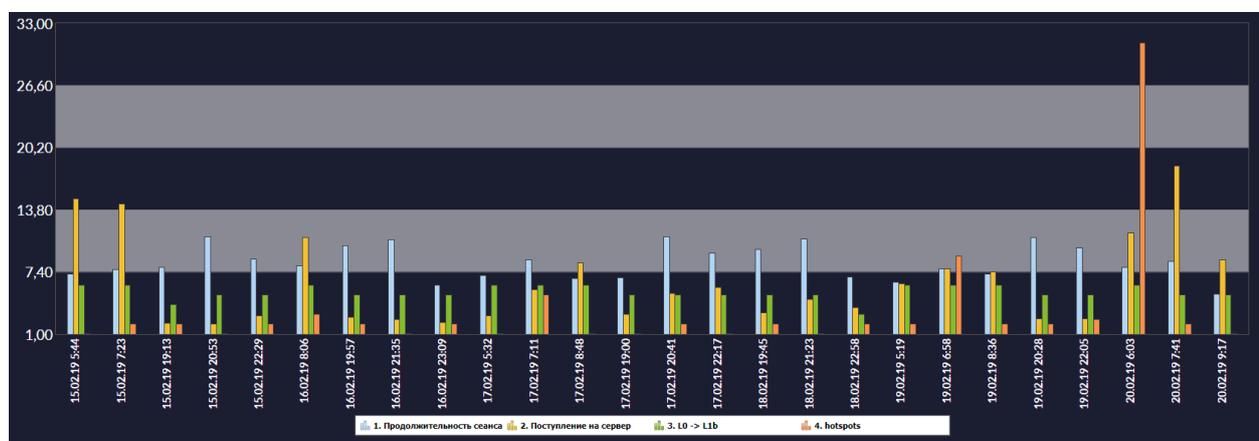


Рис. 1. Пример отчета о задержках на различных этапах обработки сеансов спутниковых данных

Важно отметить, что наряду с описанными инструментами выявления задержек в ИКИ РАН реализован широкий спектр других инструментов, предназначенных для мониторинга потоковой обработки спутниковых данных. Такие инструменты, в частности, позволяют получить статистическую информацию о выполнении обработок на различных станциях обработки, наличии ошибок по каждому из типов обработки, загруженности станций обработки и т.п. Совместное использование всех реализованных инструментов позволяет не только выявить наличие задержек, но и установить причину их возникновения.

Заключение

Реализация описанных методов выявления задержек в процессе потоковой обработки спутниковых данных позволила создать удобные инструменты для детального анализа выполнения всех этапов реализованных цепочек обработки и выявления необоснованных задержек в процессе получения итоговых информационных продуктов. Использование этих инструментов наряду с ранее созданными инструментами позволяет более точно диагностировать возникающие в работе системы обработки неполадки и выявить причины их возникновения. Наряду с этим, реализованные инструменты позволяют нам легче выявить те этапы обработки, которые наиболее нуждаются в оптимизации. Упростилось также детектирование редко возникающих, но при этом регулярных неполадок, которые ранее могли остаться незамеченными.

Как результат, описанные в настоящей статье подходы к выявлению задержек в процессе потоковой обработки спутниковых данных, позволили нам повысить надежность функционирования реализуемых нами информационных систем и сократить время с момента поступления исходных данных, до предоставления готовых информационных продуктов пользователям этих систем.

В реализации представленных инструментов выявления задержек в процессе автоматизированной обработки спутниковых данных использовались технологические возможности Центра коллективного пользования “ИКИ-Мониторинг”, работа которого осуществляется при поддержке темы «Мониторинг», госрегистрация №01.20.0.2.00164.

References

- [1] Loupian E.A., Proshin A.A., Burtsev M.A., Balashov I.V., Bartalev S.A., Efremov V.Yu., Kashnitskiy A.V., Mazurov A.A., Matveev A.M., Sudneva O.A., Sychugov I.G., Tolpin V.A., Uvarov I.A. IKI RAS Center of collective use of systems for archiving, processing and analyzing satellite data for solving problems of studying and monitoring the environment, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, Vol. 12, No 5, pp. 263-284.
- [2] Moss L.T., Atre S.. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. 576 p.
- [3] Kobets D.A., Balashov I.V., Sychugov I.G., Tolpin V.A. Organization of control and analysis of performance of automated satellite data processing systems using BI-technologies, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2017, vol. 14(3). pp. 92-103. DOI: 10.21046/2070-7401-2017-14-3-92-103.
- [4] Proshin A.A., Burtsev M.A., Kobets D.A., Sychugov I.G. Methods and tools for determining delays in the process of automated stream processing of satellite data, *16-th All-Russian open conference "Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa". 2018, November, 12-16. IKI RAS, Moscow. Abstracts*, 2018. p. 108. DOI: 10.21046/2070-16DZZconf-2018a.
- [5] Loupian E.A., Balashov I.V., Burtsev M.A., Efremov V.Yu., Kashnitskiy A.V., Kobets D.A., Krashennnikova Yu.S., Mazurov A.A., Nazirov R.R., Proshin A.A., Sychugov I.G., Tolpin V.A., Uvarov I.A., Flitman E.V. Creating technologies of construction information systems for remote sensing, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, Vol. 12, No 5, pp. 53-75.

- [6] Proshin A.A., Matveev A.M., Kobets D.A., Radchenko M.V., Sychugov I.G. Support of troubleproof work of complex distributed systems for satellite data collection, processing, archiving and access, *15-th All-Russian open conference "Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa"*. 2017, November , 13–17. IKI RAS, Moscow. Abstracts, 2017. p. 106.
- [7] Kobets D.A., Matveev A.M., Mazurov A.A., Proshin A.A. Organization of automated multistream processing of satellite information in remote sensing systems, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, Vol. 12, No 1, pp. 145-155
- [8] Andreev M.V., Galeev A.A., Efremov V.Yu., Ilyin V.O., Krashennnikova Yu.S., Loupian E.A., Mazurov A.A., Nazirov R.R., Proshin A.A., Flitman E.V. Construction of automated systems for collection, storage, processing and representation of satellite data for solving problems of environmental monitoring, *Solar-Terrestrial Physics*, 2004, Issue 5, pp. 8-11.