

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.622

Денисов
Вадим Александрович

Система хранения и обработки картографических изображений

АВТОРЕФЕРАТ

на соискание степени магистра

по специальности 1-40 80 04 – Информатика и технологии программирования

Научный руководитель
Анисимов В.Я.
к.ф-м.н., доцент

Минск 2021

КРАТКОЕ ВВЕДЕНИЕ

Геопространственные технологии включают географическую информационную систему (GIS) — пространственную аналитику, глобальную навигационную спутниковую систему (GNSS) и определение местоположения, наблюдение Земли и сканирование.

Тренды 2020 года в геопространственной индустрии — использование облачных вычислений и Mapping-as-a-Service (MaaS). В течение 2019 года многие компании добились прогресса в переносе своих вычислительных мощностей в облако. Amazon, Google или другие облачные провайдеры, предоставляют пользователям и разработчикам сложные платформы для организации бессерверных облачных вычислений. Эти платформы позволяют разработчикам запускать приложения и службы без необходимости управлять дорогостоящей и сложной серверной инфраструктурой. Что касается MaaS, то он станет наиболее доступным способом получения геопространственных данных, особенно для компаний, которым необходимо регулярно обновлять картографические снимки.

Последние технологические тенденции позволяют нам лучше управлять нашими объектами, дорогами и инфраструктурой и умные технологии будут продолжать развиваться во многих направлениях и для многих приложений. Концепция умных городов и интеллектуальных транспортных систем (ИТС) становится более популярной в коммунальных и транспортных сообществах. Этот рост будет набирать обороты с распространением больших данных, машинного обучения и глубокого обучения. Также автономное вождение продолжает развиваться, и все больше производителей автомобилей выбирают этот путь, возрастет потребность в технологиях для понимания деталей и условий инфраструктуры с использованием искусственного интеллекта и геопространственных данных.

Таким образом, геопространственное сообщество, а также отрасли архитектуры и машиностроения готовы заработать значительную долю рынка услуг, необходимых для поддержки интеллектуальной революции, и это будет способствовать развитию этих технологий.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Постановка задачи

В рамках магистерской работы поставлена цель разработать программный комплекс для хранения и обработки картографических данных, полученных с камер аэросъемки, предназначенной для упрощения предоставления и визуализации данных.

Минимальными требованиями к реализации программного продукта являются:

1) Использовать облачные технологии для обеспечения доступности и отказоустойчивость разрабатываемой системы.

2) Использование бессерверных вычислений с целью уменьшения затрат на эксплуатацию системы.

3) Интеграция с пользовательским интерфейсом для отслеживания состояния обработки изображений.

4) Обеспечить горизонтальную масштабируемость системы в зависимости от рабочей нагрузки.

5) Обеспечить защиту данных. Разграничить права доступа к данным в зависимости от роли пользователя.

6) Выделить систему ролей и прав доступа в отдельный модуль. Это позволит пользователям платформы использовать стороннюю реализацию системы прав и ролей.

7) Платформа должна иметь модульную архитектуру, позволяющую для программиста быстро и просто подключиться к платформе и работать с данными, не задумываясь о их получении и хранении.

Для достижения поставленной цели необходимо решить следующие задачи:

- Изучить существующие облачные решения.
- Изучить механизмы интеграции с пользовательским интерфейсом.
- Спроектировать гибкое архитектурное решение платформы.

Обзор платформ облачных вычислений

Преимущества, которых может ожидать предприятие при внедрении облачной инфраструктуры.

- эффективность / снижение затрат;
- безопасность данных;
- масштабируемость;
- мобильность ;
- аварийное восстановление;
- контроль.

На сегодняшний день наиболее популярными поставщиками услуг облачных вычислений являются Amazon Web Services (AWS), Microsoft Azure, Google Cloud.

AWS — один из самых популярных поставщиков облачных услуг. Они занимают наибольшую долю на рынке общедоступных облаков. В своем отчете Magic Quadrant for Cloud Infrastructure as a Service, Worldwide, Gartner отмечает: «AWS является лидером рынка облачных IaaS более 10 лет». AWS располагает огромным и постоянно растущим набором доступных сервисов, а также самой разветвленной сетью мировых центров обработки данных. Однако, сервисы от AWS являются наиболее дорогостоящими среди конкурентов.

Microsoft Azure — близкий конкурент AWS с исключительно мощной облачной инфраструктурой. Microsoft поздно вышла на рынок облачных вычислений, но дала себе толчок, по сути взяв свое локальное программное обеспечение - Windows Server, Office, SQL Server, Sharepoint, Dynamics Active Directory, .Net и другие - и перепрофилировав его для облака. Важная причина успеха Azure — операционная система Windows и другое программное обеспечение Microsoft. Поскольку Azure тесно интегрирована с этими другими приложениями, предприятия, которые используют большое количество программного обеспечения Microsoft, часто обнаруживают, что им также имеет смысл использовать Azure.

Google Cloud — хорошо финансируемый аутсайдер в конкурентной борьбе, Google вышел на облачный рынок позже и не имеет корпоративной ориентации, которая помогает привлечь корпоративных клиентов. У Google есть сильные предложения в области контейнеров, поскольку Google разработал стандарт Kubernetes, который теперь предлагают AWS и Azure. GCP специализируется на предложениях с высокой вычислительной мощностью, таких как большие данные, аналитика и машинное обучение. Он также предлагает значительное масштабирование и балансировку нагрузки — Google знает центры обработки данных и быстрое время отклика. С другой стороны, Google занимает третье место на рынке, возможно, потому, что у него нет традиционных отношений с корпоративными клиентами. Тем не менее, он быстро расширяет как свои предложения, так и свое присутствие в глобальных центрах обработки данных.

Резюмируя, в данной работе была использована платформа AWS. Во первых, AWS предоставляет несравнимо больше сервисов и их функций, чем любой другой поставщик облачных услуг: от инфраструктурных технологий, таких как инструменты для вычисления, хранилища и базы данных, до инноваций, например машинного обучения и искусственного интеллекта, озер данных и аналитики, а также Интернета вещей.

Во вторых, на платформе AWS создано крупнейшее и самое динамичное сообщество с миллионами активных клиентов и десятками тысяч партнеров по

всему миру. Клиенты разного масштаба и практически любых отраслей, в том числе стартапы, крупные корпорации и государственные организации, используют AWS для самых разнообразных задач. В партнерскую сеть AWS (APN) входят тысячи системных интеграторов, специализирующихся на сервисах AWS, и десятки тысяч независимых поставщиков ПО (ISV), адаптирующих свои технологии для работы на AWS.

Библиотека БГУИР

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** диссертации определена область и указаны основные направления проведенной работы, показана актуальность выбранной темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** проводится подробный анализ предметной области задачи, поставленной в проекте, кратко описываются различные поставщики услуг облачных вычислений и производится выбор поставщика для развертывания будущего комплекса приложений.

Облачные вычисления предлагают преимущества современному бизнесу, в том числе позволяют нескольким пользователям просматривать данные в реальном времени и легко обмениваться ими. Облачное программное обеспечение и приложения SaaS (программное обеспечение как сервисная платформа) делают работу в синхронизированной команде проще и доступнее.

В прошлом люди запускали приложения или программы из программного обеспечения, загруженного на физический компьютер или сервер в их здании. Теперь, с облачными вычислениями, больше не нужно беспокоиться о неудобных загрузках. Вместо этого вы можете использовать все те же услуги в Интернете из любого места и видеть обновления в режиме реального времени.

Вторая глава посвящена описанию формата и стандартов геоданных, которые используются в данной работе. Описывается формат хранения изображений tiff, отмечая преимущества и недостатки. Далее производится обзор метаданных: *exterior-orientation*, *footprint* и *tile-index*, которые необходимо поставлять с изображениями для определения местоположения проекций на карте.

Третья глава посвящена архитектуре и модулям системы. Проведение анализа начинается с построения архитектуры конвейера обработки данных. Затем приводится краткое описание разрабатываемой системы, используемые технологии, а также результаты работы разработанных компонентов.

На первом этапе обработки данных происходит фильтрация *tile index* — задача, состоящая из нескольких последовательных этапов, результат которых есть объект задающий сетку для изображений на карте. Целью платформы *Filtered Tile Index Platform* является создание высококачественного *Filtered Tile Index* и сокращение ручной работы, необходимой для создания *tile index filtered*.

На втором этапе была интегрирована система обработки ошибок с доской *Jira*. Каждый набор данных для обработки представляет собой некоторую отснятую территорию. Для каждого набора данных создана задача в *Jira*, где колонка отображает этап обработки набора данных. Все задачи по имеющимся территориям хранятся на доске *DATA*. Во время работы лямбда функций могут

возникать два рода ошибок: ошибки валидации и ошибки системы. В случае валидации, необходимо добавить комментарий к задаче с понятным для пользователя текстом ошибки. В случае ошибки системы, необходимо создать отдельную задачу со всей информацией об ошибке.

На третьем этапе был реализован шаг перепроектирования EO файла. Файл EO представляет собой данные внешней ориентации. Внешняя ориентация (EO) — это положение и ориентация камеры при съемке изображения. Данный шаг был реализован с использованием лямбда-функции.

На четвертом этапе проекта была необходимость разработки сервиса для пакетной обработки данных Batch Service. Он предоставляет инфраструктуру для реализации масштабируемого шага в рабочем процессе / оркестровке AWS Step Functions, когда несколько задач должны выполнять одну и ту же операцию на разных входах параллельно.

На пятом этапе были созданы компоненты для генерации Cloud Optimized GeoTIFF. Cloud Optimized GeoTIFF — оптимизированный для облака GeoTIFF. Данная логика была построена поверх написанного ранее Batch Service. В разделе приведены технические детали COGGenerationAgent: подход работы in-memory при загрузке и выгрузке изображений в хранилище S3, а также продемонстрирована работа обработчика изображений.

На шестом этапе разработанные ранее компоненты были собраны в единый конвейер Imagery Processing Step Function. Для этого мы воспользовались AWS Step Functions — средство оркестрации бессерверных функций, которое облегчает создание последовательностей функций AWS Lambda.

На седьмом этапе инфраструктура была сохранена как код. Данный подход позволит развертывать компоненты в AWS автоматически, используя командную строку, и иметь единое место хранения кода инфраструктуры для производственного окружения и окружения разработки.

ЗАКЛЮЧЕНИЕ

В данном диссертационном проекте рассматривается процесс автоматизации обработки и хранения изображений фотоаэросъемки. В рамках проекта был разработан программно-аппаратный модуль для автоматической обработки изображений и набор инструментов для автоматического развертывания системы, а также модуль был интегрирован с пользовательским интерфейсом Jira. Разработанный программный модуль использует несколько шагов для обработки изображений: генерация сетки области, перепроектирование метаданных и генерация COG изображений. Также рассматриваются различные подходы для обработки больших объемов данных, а также производится разработанный набор инструментов для обработки больших объемов данных на базе сервиса облачных вычислений AWS.

В ходе работы была изучена работа с сервисами AWS: создания модулей, написание и развёртывание лямбда функций и интеграция с системой мониторинга.

В целом, разработанная система позволяет обрабатывать набор данных величиной в 20 тысяч за три часа, при ограничении в 500 одновременно работающих лямбда функций. Время работы линейно зависит от количества одновременно работающих лямбда функций, при увеличении их параметра время работы уменьшается.

В результате разработки цель проекта достигнута. Создан и улучшен программно-аппаратный модуль, который сейчас активно используется внутри компании и позволяет предоставлять данные клиентам в течении трех часов после выгрузки в хранилище данных S3.