

А. В. Скибунов, Е. П. Кечко
(ГГУ им. Ф. Скорины, Гомель)

РАЗРАБОТКА DATA LAKE И DATA WAREHOUSE С ИСПОЛЬЗОВАНИЕМ СЕРВИСОВ AMAZON WEB SERVICES

В наше время каждая крупная компания генерирует десятки терабайт данных. Начиная с данных о продажах товара клиентам, заканчивая многоструктурированными данными о работе спутников. Эти данные нужно собирать, обрабатывать, находить ошибки и анализировать для конечного пользователя.

Целью данной работы является автоматизация описанного процесса. В качестве технологии разработки использовались язык программирования Python, который имеет обширную коллекцию библиотек для числовых вычислений и обработки данных, и облачные сервисы платформы Amazon Web Service (AWS): сервис хранения объектов Amazon Simple Storage Service (Amazon S3); бессерверный сервис данных Amazon Glue Service, который упрощает поиск, подготовку, перемещение и интеграцию данных из множества источников для анализа, машинного обучения и разработки приложений; облачная СУБД Amazon Redshift.

Рассмотрим следующую ситуацию: мы собственники большой сети книжных магазинов, но каждый магазин использует свою систему учета, будь то различные базы данных, файлы (json, csv) или системы CRM. Возникает проблема, как эти все данные собрать вместе, обработать и агрегировать для аналитики.

Необходимо произвести следующий порядок действий по созданию Big Data: на AWS S3 создаются папки, которые мы назвали raw - zone, stage - zone, analytics - zone; все данные из разных систем сохраняются в raw - zone; на AWS Glue Jobs выполняются скрипты, которые читают данные, приводят их к общему формату и удаляют дубликаты. Результаты действий сохраняются в stage - zone и analytics - zone; обработанные данные загружаются в хранилище данных в AWS Redshift; перечисленные действия повторяются по заданному расписанию с помощью AWS Glue Jobs, тем самым данные обновляются и остаются актуальными.

В конечном итоге мы получим хранилище обработанных данных со всей информацией по книжным магазинам. Теперь мы можем

анализировать и извлекать полезные метрики из данных, которые помогут сделать бизнес более успешным.

Н. И. Уваров
(БрГТУ, Брест)

СИСТЕМА ПРИЛОЖЕНИЙ «ТРАНСПОРТ ПО ЗАПРОСУ» ДЛЯ УЛУЧШЕНИЯ И ОПТИМИЗАЦИИ РАБОТЫ ОБЩЕСТВЕННОГО ТРАНСПОРТА И МАРШРУТНОГО ТАКСИ

В наши дни повсеместно активно развивается инфраструктура городского общественного транспорта в городах по всему миру. Городской общественный транспорт представляет собой одну из крупнейших отраслей народного хозяйства со сложной и многообразной техникой, а также специфической организацией и системой управления. В последнее время получает распространение автоматический городской интеллектуальный транспорт по перевозке пассажиров.

Значимая часть перевозки пассажиров приходится на систему маршрутного такси: микроавтобусов, осуществляющих перемещение жителей городов (пассажиров) по установленным маршрутам, забор и высадку – из пункта отправления в пункт назначения.

Периодичность движения транспорта не изменяется в течение дня, что не соответствует изменениям интенсивности прибывающего пассажиропотока. Нужно отметить, что количество транспортных средств (маршрутных такси) не связано с тем, сколько пассажиров ожидают на остановочных пунктах из-за недостатка информации.

В связи с вышеописанными проблемами была предложена оптимизированная модель городского транспорта, известная как «Транспорт по запросу». Поэтому нами начата разработка автоматизированной системы с целью оптимизации и улучшения работы городского транспорта, в частности маршрутного такси.

При проектировании данной системы можно выделить основные задачи:

- 1 Пассажир, находясь на остановочном пункте, создаёт заявку, через мобильное приложение клиента, в которой отмечает пункты отправления и назначения для маршрутного такси;