

УДК 004.75

АНАЛИТИКА IOT ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СЕРВИСА AWS IOT ANALYTICS ПРИ ИССЛЕДОВАНИИ ЗАГАЗОВАННОСТИ ОКРУЖАЮЩЕЙ СРЕДЫ



К.О. Климов
Магистрант БГУИР,
специалист по
сопровождению ПО
EPAM Systems



Г. А. Пискун
Заместитель декана
факультета компьютерного
проектирования по научной
работе, кандидат
технических наук, доцент



***Д.В. Лихачевский**
декан факультета
компьютерного
проектирования, кандидат
технических наук, доцент



В.Ф. Алексеев
Доцент кафедры проектирования
информационных компьютерных систем,
кандидат технических наук, доцент



В.В. Шаталова
Заместитель декана факультета
компьютерного проектирования по учебно-
методической работе, кандидат технических
наук, доцент

* Приглашённый профессор Евразийский национальный университет им. Л. Н. Гумилева, г. Нур-Султан, Республика Казахстан
Белорусского государственного университета информатики и радиоэлектроники, Республика Беларусь.
ИООО «ЭПАМ СИСТЕМЗ», Республика Беларусь.
E-mail: klimovkostya5@gmail.com.

К. О. Климов

Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистрант БГУИР. Работает в EPAM Systems в должности специалиста по сопровождению ПО. Проводит научные исследования систем Интернета-вещей.

Г. А. Пискун

Заместитель декана факультета компьютерного проектирования по научной работе, кандидат технических наук, доцент.

Д. В. Лихачевский

Декан факультета компьютерного проектирования, кандидат технических наук, доцент.

В. Ф. Алексеев

Доцент кафедры проектирования информационных компьютерных систем, кандидат технических наук, доцент.

В. В. Шаталова

Заместитель декана факультета компьютерного проектирования по учебно-методической работе, кандидат технических наук, доцент.

Аннотация. Количество устройств Интернета-вещей и данных, производимых ими, растёт каждый день: некоторые решения могут производить до петабайта данных и поддерживать подключение миллиона устройств одновременно. Под устройствами Интернета-вещей принято понимать совокупность датчиков и платформы, к которой они подключены для отправки собранных данных облачному провайдеру. Для обработки такого объёма данных требуются огромные вычислительные ресурсы, размещение и обслуживание которых в собственном дата-центре не всегда предоставляется возможным. В таких случаях имеет смысл воспользоваться услугами облачных провайдеров, которые предоставляют вычислительные ресурсы, берут на себя обслуживание оборудования и позволяют сэкономить значительные средства.

Ключевые слова: Интернет-вещей, IoT, Internet of Things, AWS IoT Analytics, аналитика больших данных.

Введение.

Рост автомобилизации и развитие промышленных предприятий пагубно сказывается на окружающей среде. Содержание вредных веществ и углекислого газа выросло на 10 % за последние 10 лет [1]. По этой причине отслеживание и мониторинг этих показателей в реальном времени с разных локаций может существенно повлиять на экологическую обстановку в конкретной области.

Для решения этой проблемы требуется разработать систему для сбора и анализа данных с устройств Интернета-вещей из различных локаций в городе в режиме реального времени. Чтобы обработать такой объём данных, требуются огромные вычислительные ресурсы, которые не всегда возможно создать на базе локального дата-центра. В таких случаях можно обратиться к облачным операторам, которые предоставляют в аренду сервера.

Алгоритмы обработки IoT данных в облаке AWS. В последнее время операторы, помимо вычислительных ресурсов, начали предоставлять и готовые сервисы, которые позволяют решать задачи клиентов быстрее и без необходимости разработки собственных продуктов. Одним из таких сервисов для обработки и аналитики IoT-данных является AWS IoT.

AWS IoT это набор сервисов, которые позволяют IoT устройствам подключаться к AWS и позволяют приложениям в облаке взаимодействовать с подключёнными устройствами. Для аналитики поступающих данных используется сервис AWS IoT Analytics – это полностью управляемый сервис аналитики, разработанный специально для систем Интернета-вещей, который собирает, обрабатывает, хранит и анализирует данные с этих устройств [2].

Аналитика с помощью AWS IoT Analytics – не единственный способ анализа данных Интернета-вещей (IoT) в облаке AWS. Для анализа IoT-данных может использоваться связка множества сервисов: Amazon S3, Amazon Kinesis, AWS Glue, Amazon Athena, Amazon Redshift и Redshift Spectrum. Преимущество AWS IoT Analytics заключается в том, что сервис был специально разработан для обработки данных IoT в петабайтном масштабе. При обработке IoT-данных часто возникают проблемы с целостностью сообщений, а также ложные показания датчиков, которые необходимо исключить перед выполнением анализа. Данный сервис может исключить аномалии, в отличие от связки сервисов, где исключение данных ложится на клиента. Кроме того, данные Интернета-вещей часто должны быть обогащены и преобразованы для получения лучших результатов аналитики. IoT Analytics может фильтровать и преобразовывать данные Интернета-вещей перед их складированием в хранилище данных.

Алгоритм обработки данных. В качестве данных для AWS IoT Analytics будут использованы данные датчиков окружающей среды, собранные практически в реальном времени с ряда устройств Интернета вещей. Сенсоры подключаются к одноплатному компьютеру Raspberry Pi, для наглядности устройства были специально размещены в физических местах, которые различаются по температуре, влажности и другим условиям окружающей среды.

Каждое устройство включает в себя следующие датчики: датчик качества воздуха, датчик обнаружения опасных газов, цифровой датчик температуры и влажности. Устройство вместе с датчиками представлено на рисунке 1.

На каждом устройстве Raspberry Pi запущен Python скрипт, который использует пакет разработки AWS SDK для Интернета-вещей. Скрипт собирает в общей сложности семь различных показаний от трёх датчиков с регулярным интервалом, далее скрипт отправляет показания

датчиков вместе с идентификатором устройства и меткой времени в виде единого сообщения в AWS, используя стандартный сетевой протокол Интернета-вещей MQTT.

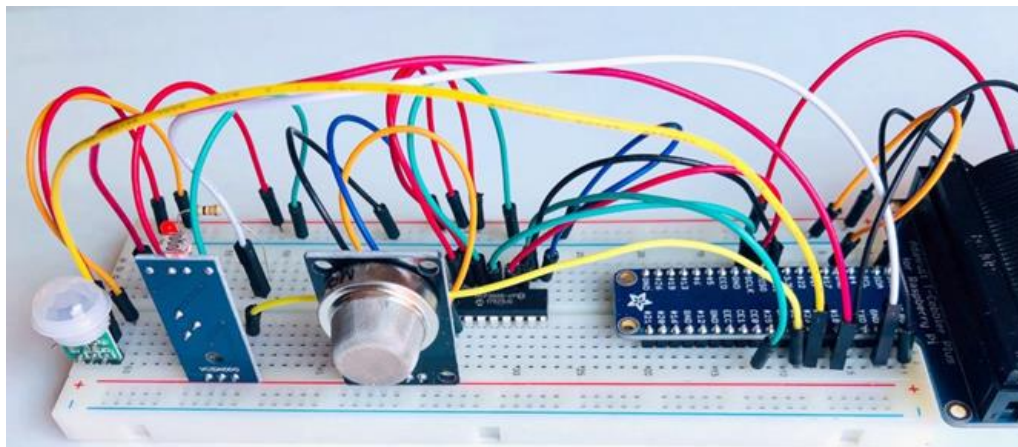


Рисунок 1. Устройство для сбора и обработки данных

Каждое MQTT сообщение содержит данные о содержании углекислого газа в воздухе, влажности и температуры окружающей среды (Рисунок 2). Устройства Raspberry Pi заранее зарегистрированы в сервисе AWS IoT Core. Этот сервис позволяет пользователям быстро и безопасно подключать устройства и управлять ими в AWS. Сервис может масштабироваться до миллиарда устройств и триллиона сообщений. В сервисе AWS IoT Core зарегистрированные устройства называются вещами. Вещь – это представление конкретного устройства или логического объекта, а также информация о его местоположении, времени добавления и другие метаданные. Информация о вещи хранится в реестре в виде JSON данных.

```
1 {
2   "data": {
3     "co": 0.0061044480269226063,
4     "humidity": 55.099998474121094,
5     "light": true,
6     "lpg": 0.008895956948783413,
7     "motion": false,
8     "smoke": 0.023978358312270912,
9     "temp": 31.799999237060547
10  },
11  "device_id": "6e:81:c9:d4:9e:58",
12  "ts": 1594419195.292461
13 }
```

Рисунок 2. Содержание MQTT сообщения

AWS IoT Core предоставляет шлюз устройств, который управляет всеми активными подключениями. В настоящее время он поддерживает протоколы MQTT, WebSockets и HTTP 1.1. За обработку сообщений отвечает Message Gateway – высокопроизводительный pub/sub Message Broker, который надёжно передаёт сообщения на все устройства и приложения Интернета-вещей с низкой задержкой. AWS IoT Core также обеспечивает аутентификацию и шифрование, обеспечивая зашифрованный обмен данными между AWS и устройствами. Все данные отправляются с использованием протокола шифрования TLS 1.2 с X.509 и цифровыми сертификатами через HTTPS порт 443. Авторизация устройства для доступа к любому ресурсу на AWS контролируется отдельными политиками AWS IoT аналогично политикам AWS IAM.

Устройства отправляют данные раз в 5 секунд, генерируя более 50 тыс. сообщений за 24 часа (рисунок 3). На рисунке 4 проиллюстрирована архитектура AWS IoT Core. После получения MQTT сообщения от устройства Интернета-вещей сервис перенаправляет его на один из настроенных каналов AWS IoT Analytics Channel. Каналы AWS IoT Analytics – это логическое представление потока обработки данных, который позволяет отследить объём данных, переданных ему, передать их другому сервису аналитики или архивировать.



Рисунок 3. Данные генерируемые устройствами

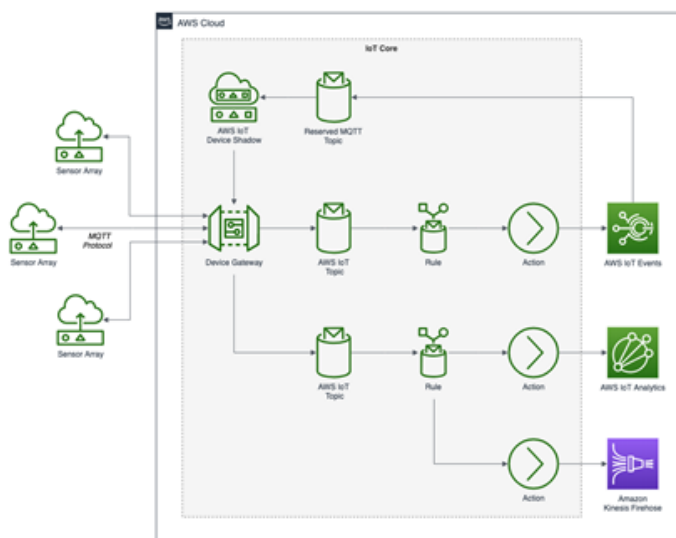


Рисунок 4. Архитектура AWS IoT Core для обработки данных с устройств

Далее сообщения фильтруются с помощью языка SQL и пересылаются на другие сервисы AWS или используются AWS IoT Analytics для последующей аналитики. На рисунке 5 представлена архитектура AWS IoT Analytics [3].

AWS IoT Analytics Pipeline использует сообщения от одного или нескольких каналов, а затем преобразует, фильтрует и обогащает сообщения перед их хранением в IoT Analytics Data stores. Сервис позволяет производить до 23 различных действий над сообщениями [4].

После первичной обработки сообщений в Analytics Pipeline данные загружаются в AWS IoT Analytics Data Set для хранения и последующей обработки в других сервисах AWS. Просмотреть данные можно с помощью выполнения запросов на языке SQL.

Для выполнения статистического анализа полученных данных может использоваться сервис AWS IoT Analytics Notebook, который использует Jupyter Notebooks и язык Python для обработки и представления данных. Сервис IoT Analytics Notebook включает набор шаблонов, которые содержат разработанные AWS модели машинного обучения и визуализации (рисунок 6). Шаблоны могут храниться в хранилище исходного кода.

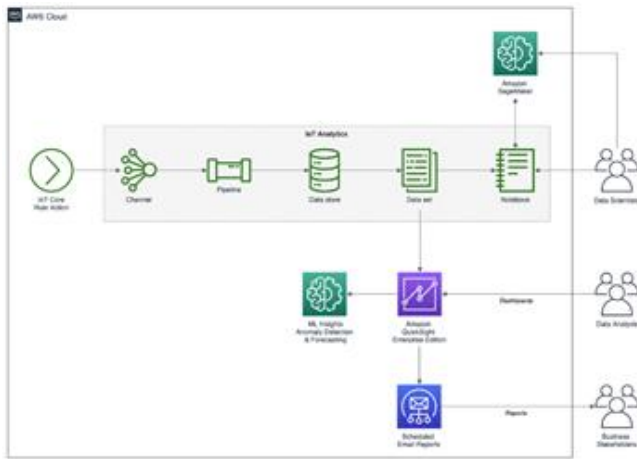


Рисунок 5. Архитектура AWS IoT Analytics

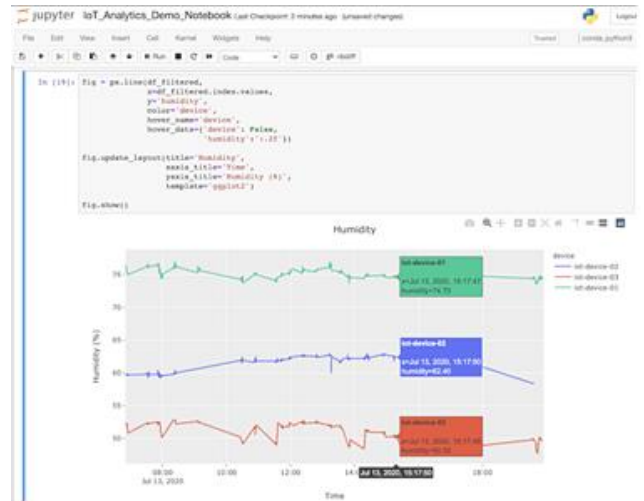


Рисунок 6. Шаблоны IoT Analytics Notebook

Для визуализации полученных данных воспользуемся сервисом Amazon QuickSight. Данный сервис полностью интегрируется и позволяет загружать данные в режиме реального времени напрямую из Analytics Data Set (рисунок 7). QuickSight предоставляет широкий спектр функций, позволяющих выполнять динамические вычисления значений полей.

Преимущество сервиса заключается в том, что для создания систем мониторинга не требуется использование языков программирования и разметки. QuickSight содержит ряд предустановленных визуализаций, позволяет наносить данные на карту (рисунок 8).

Для отслеживания полученных данных можно создать дэшборд с различными визуализациями данных, которые будут поступать в режиме реального времени к пользователю (рисунок 9,10).

The screenshot shows the Amazon QuickSight Data Set interface. It displays a table of data points with the following columns: device, humidity, temp, light, motion, sig, vs, and battery. The data points are filtered to show records from May 15, 2020, between 08:00 and 18:00. The table contains 20 rows of data.

device	humidity	temp	light	motion	sig	vs	battery
iot-device-01	67.9	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	49.0	70.02	false	false	0.0005	0.0009	0.0171
iot-device-01	49.0	70.02	false	false	0.0005	0.0009	0.0171
iot-device-01	78.7	62.24	false	false	0.0004	0.0004	0.0226
iot-device-01	47.1	70.7	false	false	0.0005	0.0009	0.0171
iot-device-01	78.5	62.96	false	false	0.0004	0.0004	0.0226
iot-device-01	67.9	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	49.0	70.02	false	false	0.0005	0.0004	0.0172
iot-device-01	67.9	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	78.6	62.96	false	false	0.0004	0.0004	0.0226
iot-device-01	49.0	70.02	false	false	0.0005	0.0004	0.0172
iot-device-01	67.9	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	78.7	62.24	false	false	0.0004	0.0004	0.0226
iot-device-01	49.0	70.02	false	false	0.0005	0.0004	0.0172
iot-device-01	78.5	62.96	false	false	0.0004	0.0004	0.0226
iot-device-01	47.1	71.06	false	false	0.0005	0.0009	0.0166
iot-device-01	78.7	62.24	false	false	0.0005	0.0004	0.0219
iot-device-01	47.0	71.06	false	false	0.0005	0.0009	0.0166
iot-device-01	62.1	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	47.0	71.06	false	false	0.0005	0.0009	0.0166
iot-device-01	78.7	62.24	false	false	0.0005	0.0004	0.0219
iot-device-01	47.1	71.06	false	false	0.0005	0.0009	0.0166
iot-device-01	68.0	63.74	true	false	0.0004	0.0001	0.0175
iot-device-01	47.0	71.06	false	false	0.0005	0.0009	0.0166
iot-device-01	78.7	62.24	false	false	0.0005	0.0004	0.0219
iot-device-01	85.0	63.68	false	false	0.0004	0.0005	0.0203
iot-device-01	67.0	66.56	true	false	0.0004	0.0002	0.0143
iot-device-01	44.8	70.74	false	false	0.0004	0.0004	0.0202

Рисунок 7. QuickSight Data Set



Рисунок 8. Amazon QuickSight визуализация на карте

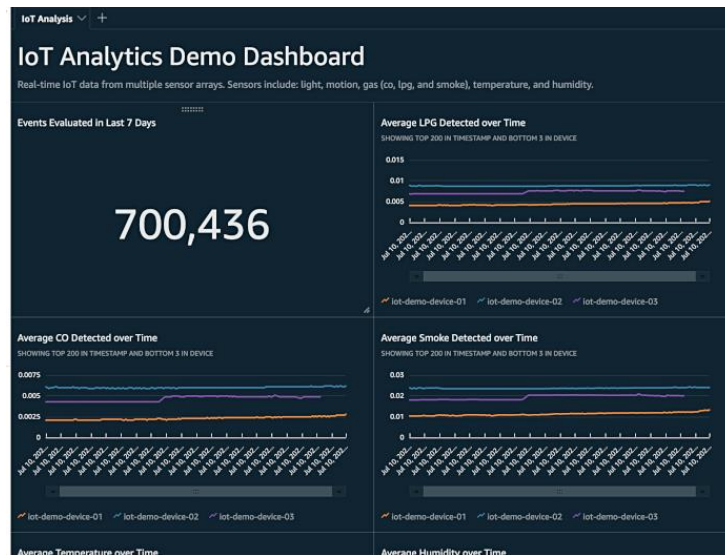


Рисунок 9. Дашборд QuickSight



Рисунок 10. Дашборд влажности окружающей среды в QuickSight

Для обработки аномальных событий использован QuickSight ML Anomaly Detection,

который непрерывно анализирует данные для обнаружения аномалий в поступающих данных. ML-прогнозирование на базе технологии QuickSight может использоваться для точного прогнозирования ситуации состояния окружающей среды в будущем. Встроенные алгоритмы QuickSight упрощают использование ML, который обучается на поступающих данных, чтобы предоставить точные прогнозы, основанные на исторических тенденциях.

Заключение.

Использование сервисов AWS IoT позволяет в кратчайшие сроки построить инфраструктуру для анализа и обработки данных Интернета-вещей. Готовые сервисы позволяют клиентам сэкономить значительные средства на разработку продуктов и обслуживание инфраструктуры. Другим плюсом таких систем является их масштабируемость: данные сервисы могут масштабироваться с одного подключённого устройства до миллиарда в течение нескольких минут, что позволяет клиентам не задумываться о возможных проблемах в случае роста нагрузки на их систему.

Список литературы

- [1] Atmospheric CO2 Concentration [Электронный ресурс]. – Режим доступа: <https://www.co2.earth/>
- [2] AWS IoT Core [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/iot-core/>
- [3] AWS IoT Analytics [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/iot-analytics/>
- [4] Getting started with AWS IoT Analytics [Электронный ресурс]. – Режим доступа: <https://docs.aws.amazon.com/iotanalytics/latest/userguide/getting-started.html>

IOT DATA ANALYTICS USING AWS IOT ANALYTICS FOR ENVIRONMENTAL GAS ANALYSIS

K.O. KLIMOV
*BSUIR Master,
DevOps Engineer
EPAM Systems*

G. A. PISKUN
*Deputy Dean of the Faculty of
Computer Design for Scientific
Work, Candidate of Engineering
of Sciences, Associate Professor*

D.M. LIKHACHEVSKY
*Dean of the Faculty of Computer
Design, candidate of Technical
Sciences, Associate Professor*

V.F. ALEKSEEV
*Associate Professor, Department of Information
Computer Systems Design, Candidate of Technical
sciences, Associate Professor*

V.V. SHATALOVA
*Deputy Dean of the Faculty of Computer Design for
Educational and Methodological Work, Associate
Professor*

*Department of Information and Computer Systems Design
Faculty of Computer Engineering
Belarusian State University of computer science and Radio Electronics, Republic of Belarus
EPAM Systems, Republic of Belarus
E-mail: klimovkostya5@gmail.com*

Abstract. Number of IoT devices and the data they produce grows every day. Some solutions can produce up to a petabyte of data and maintain the connection of a million devices at the same time. Usually an IoT device is a set of sensors and the platform to which they are connected to send the collected data to the cloud provider. Processing this amount of data requires huge computing resources, the placement and maintenance of which in your data center is not always possible. In such cases, it makes sense to use the services of cloud providers that provide computing resources, take over the maintenance of equipment and save significant money.

Keywords: IoT, Internet of Things, AWS IoT Analytics, Big Data Analytics.