UDC 004.3:614.2

Uladzimir A. Vishniakou[1], Xia YiWei[2], Yu ChuYue[3]
[1] vish2002@mail.ru; [2] xiayiwei4@gmail.com
Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus

EDUCATIONAL IOT NETWORK FOR IT DIAGNOSTICS

This report presents a comprehensive solution for IoT data collection, storage, analysis, and visualization using EMQX MQTT message middleware, ClickHouse OLAP database, and Grafana data visualization. The proposed stack enables seamless integration and efficient processing of large-scale IoT data, facilitating real-time monitoring and valuable insights.

Keywords: IoT, EMQX, ClickHouse, Grafana, data collection, storage, analysis, visualization.

В. А. Вишняков[1], С. Ивэй[2], Ю. Чуйэ[3]
[1] vish2002@mail.ru; [2] xiayiwei4@gmail.com
Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь

УЧЕБНАЯ СЕТЬ ИНТЕРНЕТА ВЕЩЕЙ ДЛЯ ИТ-ДИАГНОСТИКИ

В докладе представлено комплексное решение для сбора, хранения, анализа и визуализации данных Интернета вещей с использованием промежуточного программного обеспечения EMQX MQTT для обмена сообщениями, базы данных OLAP ClickHouse и визуализации данных Grafana. Предлагаемый стек обеспечивает плавную интеграцию и эффективную обработку крупномасштабных данных Интернета вещей, облегчая мониторинг в режиме реального времени и получение ценной информации.

Ключевые слова: IoT, EMQX, ClickHouse, Grafana, сбор данных, хранение, анализ, визуализация.

Introduction

The Internet of Things (IoT) is a technology that connects physical devices, sensors, software and networks, enabling interaction between devices and facilitating data collection, transmission and processing. With the development of the Internet of Things, its applications have expanded in various fields, including smart cities, transportation, smart homes and industrial automation. IT diagnostics, as the most important method of data analysis, plays a vital role in the ecosystem of the Internet of Things. In work [1], the basics of building networks of IoT for product quality analysis are considered.

The concept of IT diagnostics

IT diagnostics [2] refers to the process of analyzing and modeling data uploaded in the Internet of Things (IoT) to monitor and diagnose devices and objects. In IoT, a vast amount of data is generated by sensors and devices, and IT diagnostics helps extract useful information from this data for real-time decision-making. Specifically, IT diagnostics can be viewed as a monitoring and control problem, which can be addressed through mathematical modeling.

Firstly, we define a monitoring system in IoT as a collection of sensor measurements and control decisions. Let the sensor measurements be represented by the vector $f(y) = [y_1, y_2, \ldots, y_n]^T$, where y_i represents the measurement from the i-th sensor. Let the control decisions be represented by the vector $g(u) = [u_1, u_2, \ldots, u_m]^T$, where u_i represents the i-th control decision. The objective is to select the optimal control decisions $g(u)$ based on the sensor measurements $f(y)$ to achieve specific monitoring goals.

To achieve this objective, mathematical modeling can be used to describe the monitoring system. Assuming the state of the monitoring system can be represented by the vector $h(x)= [x_1, x_2, \ldots, x_k]^T$, where x_i represents the i-th state variable. The evolution of state variables can be described by state equations, i.e., $I(x)_{k+1} = f(f)(h\{x\}_k, g\{u\}_k)$, where $f(f)$ represents the state equations. The sensor measurements can be described by measurement equations, i.e., $f(y) = f(h)(h(x), g(v))$, where $h(h)$ represents the measurement equations, and $g(v)$ represents measurement errors. These equations can be used to establish the mathematical model of the monitoring system.

The Structure of the IoT Diagnostics Network

The IoT diagnostic network consists of two main components: the data collection and transmission layer and the analysis and visualization layer. At the data collection and transmission layer, EMQX IoT message middleware facilitates the connection and communication of a large number of devices, enabling efficient data collection and transmission. ClickHouse OLAP database handles the storage and management of the massive amount of IoT data.

After data is stored in ClickHouse, the analysis and visualization layer come into play. ClickHouse, in combination with the open-source software Grafana, allows for rapid development of a comprehensive IoT data analysis and visualization platform. The analysis and visualization layer enables rule-based statistics, monitoring, and metric display, unleashing the full potential of IoT data and providing valuable insights for real-time decision-making.

The Design of the IoT Diagnostic Network

EMQX [3] is developed on the high-concurrency Erlang/OTP language platform, supporting millions of connections and a distributed cluster architecture. It facilitates the storage of IoT message data into ClickHouse through a rule engine.

ClickHouse [4] is a columnar-oriented database management system (DBMS) used for data analysis (OLAP), and it is open-sourced by the Russian search giant Yandex.

Grafana [5] is a cross-platform, open-source tool for metric analytics and visualization. It allows querying and visualizing data from various data sources, offering a wide range of visualizations such as heatmaps, line charts, and graphs. It supports data sources like InfluxDB, OpenTSDB, Prometheus, Elasticsearch, CloudWatch, KairosDB, and more, enabling the creation of custom alert rules and notification to other message processing services or components.

This article simulates an IoT environment data collection scenario, assuming the existence of environmental data collection points with pre-existing data. All data from these collection points is transmitted to the collection platform (MQTT Publish) through the MQTT protocol, with the following topic design:

```
Topic: sensor/data
```

The sensor sends data in JSON format, including temperature, humidity, noise volume, PM10, PM2.5, sulfur dioxide, nitrogen dioxide, carbon monoxide, sensor ID, area, and collection time.

```
{
    "temperature": 30,
    "humidity": 20,
    "volume": 44.5,
    "PM10": 23,
    "pm25": 61,
    "SO2": 14,
    "NO2": 4,
    "CO": 5,
    "id": "10-c6-1 f-1 a-1 f-47",
    "area": 1,
    "ts": 1596157444170
}
```

The following requirements are proposed for real-time storage and subsequent data analysis:

• Each device reports data every 5 seconds, and the database should store each data point for future retrospective analysis.

• Raw data should be stored in ClickHouse and visualized using Grafana for data analysis and visualization.

All the components used in this article have Docker images available for quick setup and deployment. Follow the steps below to deploy the EMQX server locally, start ClickHouse with local listening enabled, and launch Grafana.

```
docker run -d – name emqx -p 1883:1883-p
8083:8083-p 8883:8883-p 8084:8084-p 18083:18083
emqx/emqx
```

Configuring EMQX to store data in ClickHouse:

1.Create the 'test' database:

```
create database test;
use test;
```

2.Create the 'sensor_data' table with the following columns:

```
CREATE TABLE sensor_data (
    temperature Float,
    humidity Float,
    volume Float,
    PM10 Float,
    pm25 Float,
    SO2 Float,
```

```
        NO2 Float,
        CO Float,
        sensor_id String,
        area Int,
        coll_time DateTime,
        coll_date Date
    ) engine = Log;
```

These commands will set up the necessary database and table structures in ClickHouse to store the sensor data from EMQX.

Configuring EMQX Rule Engine:

1.The Rule SQL is used for filtering EMQX messages and events. The following SQL statement filters payload data from the "sensor/data" topic:

```
SELECT
 payload
FROM
 "sensor/data"
```

2.Response Action:

Configuring the response action requires two pieces of data: a associated resource and an SQL template.

a. Associated Resource: Create a ClickHouse resource and configure the connection parameters.

b. SQL Template: In this case, it is an INSERT SQL statement that carries the data. Make sure to specify the database name in the SQL.

```
INSERT INTO test.sensor_data VALUES (
    ${payload.temperature},
    ${payload.humidity},
    ${payload.volume},
    ${payload.PM10},
    ${payload.pm25},
    ${payload.SO2},
    ${payload.NO2},
    ${payload.CO},
    '${payload.id}',
    ${payload.area},
    ${payload.ts}/1000,
    ${payload.ts}/1000
)
```

Generating Simulated Data:

The simulation involves 10 devices, each reporting simulated data every 5 seconds to EMQX over the past 24 hours. With the installation of all components completed and successful simulation data written, now we proceed to add a data source in Grafana and display the data source information. We select the ClickHouse type data source. After adding the data source, we proceed to add the information for the dashboards to be displayed. In this article, average value panels and maximum value panels are added.

```
SELECT
    $timeSeries as t,
```

```
    max (temperature) as temperature,
    max (humidity) as humidity
FROM $table
WHERE $timeFilter
GROUP BY t
ORDER BY t
SELECT
    $timeSeries as t,
    avg (temperature) as temperature,
    avg (humidity) as humidity
FROM $table
WHERE $timeFilter
GROUP BY t
ORDER BY t
```

Conclusion

Our study IoT network demonstrates the successful implementation of an integrated system for IoT data transmission, storage, analysis, and visualization using EMQX and ClickHouse. The combination of EMQX's robust capabilities and ClickHouse's advanced data processing and analysis functionalities proves to be highly effective for IoT data collection. The described network is used in the discipline "Fundamentals of IoT networks" for students of BSUIR specialty "Infocommunication systems".

References

1. Vishniakou, U. A. Specialized IoT systems: Models, Structures, Algorithms, Hardware, Software Tools. Minsk: BSUIR, 2023. 184 p.

2. Lu, S. Edge computing on IoT for machine signal processing and fault diagnosis: A review / Lu S, Lu J, An K, He Q // IEEE Internet of Things Journal. 2023. № 10. P. 11093–11116.

3. Bender, M. Open-source mqtt evaluation / Bender M, Kirdan E, Pahl M O, Georg Carle // 2021 IEEE 18 th Annual Consumer Communications & Networking Conference (CCNC). 2021. P. 1–4.

4. Imasheva, B. The practice of moving to big data on the case of the nosql database, clickhouse / Imasheva B. Azamat N, Sidelkovskiy S, Kopmann, A. // Optimization of Complex Systems: Theory, Models, Algorithms and Applications. Springer International Publishing. 2022. № 12. P. 1–11.

5. Chan, N. A resource utilization analytics platform using grafana and telegraf for the savio supercluster/ Chan N. // Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning). 2019. № 31. C. 1–6.