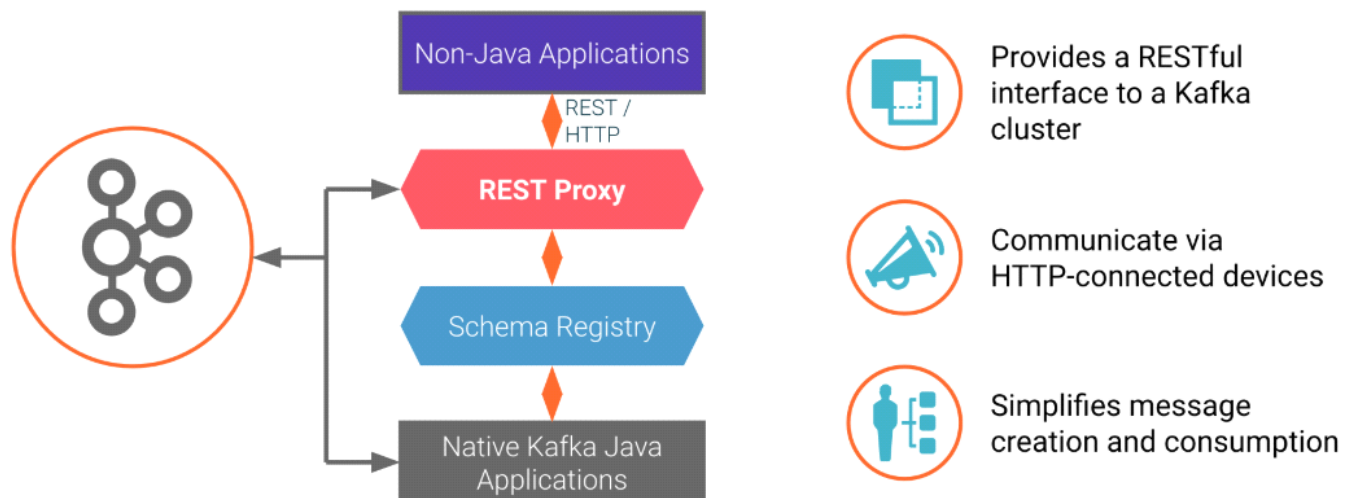


## REST Proxy



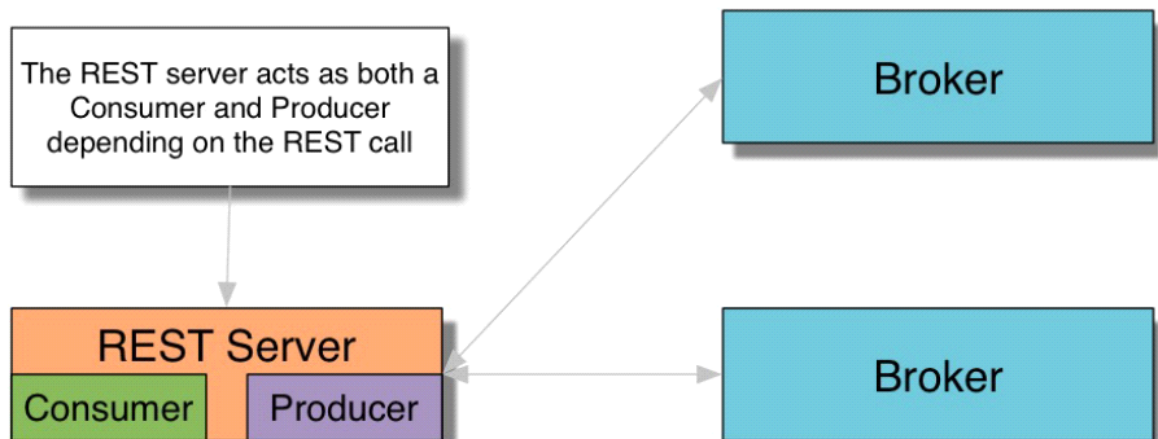
- Confluent Open Source also includes a REST Proxy for Kafka. This allows **any language to access Kafka** via REST. We will discuss REST Proxy in much detail in a later module
- Confluent Enterprise also includes an **MQTT proxy** to allow direct ingestion of IoT data. This will also be discussed in a later module

Three uses for the REST Proxy:

1. For remote clients (such as outside the datacenter including over the public internet)
2. For internal client apps that are written in a language not supported by native Kafka client libraries (i.e. other than Java, C/C++, Python, and Go)
3. For developers or existing apps for which REST is more productive and familiar than Kafka Producer/Consumer API.

## About the Confluent REST Proxy

- The Confluent REST Proxy allows you to use HTTP to perform actions on the Kafka cluster
- The REST calls are translated into native Kafka calls
- This allows virtually any language to access Kafka
- Uses **POST** to send data to Kafka
  - Embedded formats: JSON, base64-encoded JSON, or Avro-encoded JSON
- Uses **GET** to retrieve data from Kafka



The Confluent REST Proxy is part of Confluent Open Source and Confluent Enterprise distributions. The proxy provides a RESTful interface to a Kafka cluster, making it easy to produce and consume messages, view the state of the cluster, and perform administrative actions without using the native Kafka protocol or clients.

Some example use cases are:

- Reporting data to Kafka from any **frontend app** built in any language not supported by official Kafka clients
- Ingesting messages into a stream processing framework that doesn't yet support Kafka
- Scripting administrative actions



The configuration files for the REST proxy are preconfigured in our lab environments. Configuring the REST proxy is not part of this course. Refer students to the documentation for more details if they ask:

<https://docs.confluent.io/current/kafka-rest/docs/deployment.html>

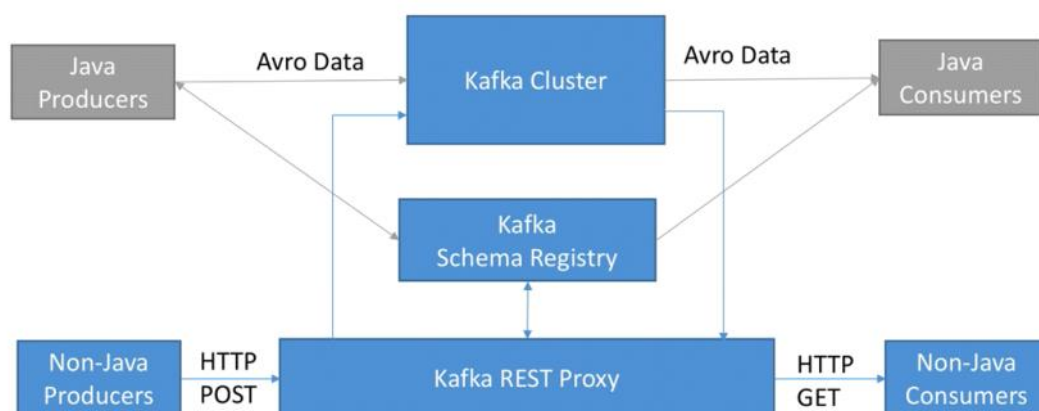
работа с кафкой через HTTP ( **вполовину медленнее** для продьюсера и **в пять раз медленнее** для консьюмера) чем через TCP

- REST это 66% от TCP для продьюсеров
- REST это 20% от TCP для консьюмеров



## Introduction

- Kafka is great for Java based consumers / producers, but sometimes clients are lacking for other languages
- Although things are getting better
- Additionally, sometimes Avro support for some languages isn't great, whereas JSON / HTTP requests are great
- For all these reasons, Confluent created the REST Proxy
- It's an open source project created by Confluent



медленнее работает потому что приходится работать через HTTP а не kafka протоколом

# Making a request to the REST Proxy



- Content Type has to be specified in a Header (plus an Accept header)

application/vnd.kafka[.embedded\_format].[api\_version]+[serialization\_format]

json, binary or avro

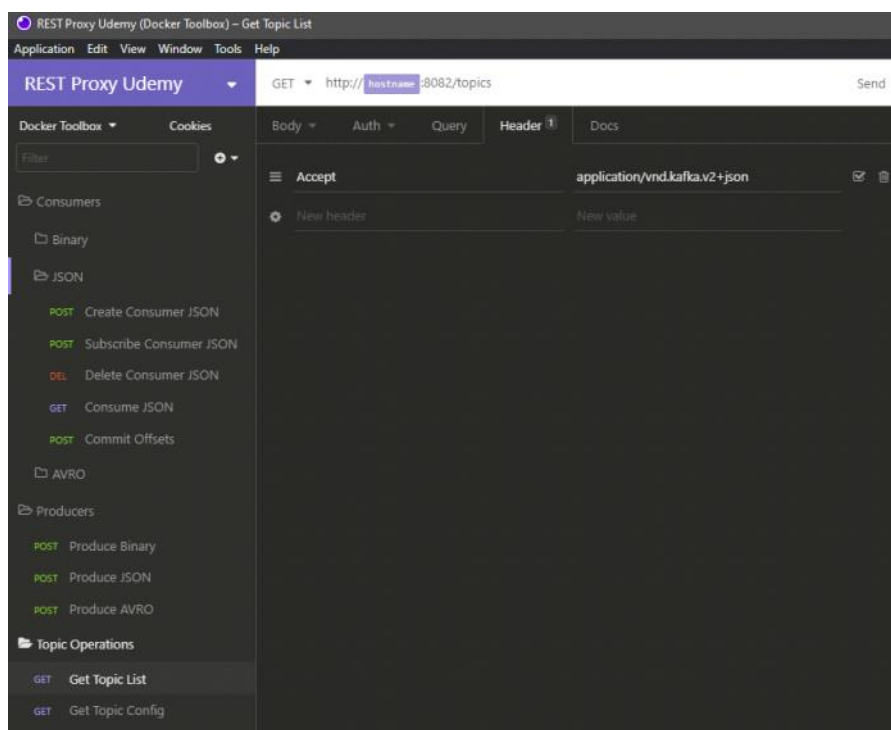
Always v2

Always json

Example:

**Content-Type:** application/vnd.kafka.avro.v2+json  
**Accept:** application/vnd.kafka.v2+json

теперь операции consume/produce можно выполнять через proxy по HTTP-REST



прокси - позволяет с кафкой работать по ресту например чтоб кафку дергать из SAP PI или Informatica

- удобно также чтобы FRONTEND напрямую(через HTTP) мог работать с кафкой

- You have 3 choices with the REST Proxy to produce data:
  - Binary (raw bytes encoded with base64)
  - JSON (plain json)
  - Avro (JSON encoded)
- When in doubt, always refer to the documentation:  
<https://docs.confluent.io/current/kafka-rest/docs/index.html>
- You can do batching in the produce request

особенности консьюмеров

## Consuming with the REST Proxy



- To consume with the REST Proxy, you first need to create a consumer in a specific consumer group.
- Once you open a consumer, the REST Proxy returns a URL to directly hit in order to keep on consuming from the same REST Proxy instance
- If the REST Proxy shuts down, it will try to gracefully close the consumers
- You can set **auto.offset.reset** (latest or earliest)
- You can set **auto.commit.enable** (true or false)

binary формат закодированный BASE64



# Producing with the REST Proxy Binary



- Binary data has to be base64 encoded before sending it off to Kafka.
- Base64 is a smart way invented to safely transfer bytes over the internet, in a way that won't break protocols (only by using 64 different characters)
- This way, any binary array can be sent (image data, string with weird characters, etc).
- There are many libraries for each language to base64 encode your binary data. You can learn about base64 here: <https://en.wikipedia.org/wiki/Base64>

## 1 продюсером запостим сообщение

The screenshot shows the REST Proxy Udemu interface. The top part displays a JSON payload for a Kafka message:

```
1 {
2   "records": [
3     {
4       "key": "azv5",
5       "value": "aGvsbG8gd29ybGhISE="
6     },
7     {
8       "value": "xc3yWskb286j5Qh",
9       "partition": 0
10    },
11    {
12      "value": "bm8gcGFydG10ahBucw=="
13    }
14  ]
15 }
```

The bottom part shows the headers for a POST request to `http://hostname:8082/topics/rest-binary`:

```
Content-Type: application/vnd.kafka.binary.v2+json
Accept: application/vnd.kafka.v2+json, appl
```

## 1 создадим консьюмера

The screenshot shows the REST Proxy Udemu interface. The top part displays the URL `POST http://hostname:8082/consumers/my-consumer-group`. The bottom part shows the JSON payload for creating a consumer:

```
1 {
2   "name": "my_consumer_json",
3   "format": "json",
4   "auto.offset.reset": "earliest",
5   "auto.commit.enable": "false"
6 }
```

## 2 чтобы читать консьюмером сначала нужно подписаться

The screenshot shows the REST Proxy Udemu interface. The top part displays the URL `POST http://hostname:8082/consumers/my_consumer_binary/subscription`. The bottom part shows the JSON payload for subscribing to a topic:

```
1 {
2   "topics": [
3     "rest-binary"
4   ]
5 }
```

## 3 законсыюим сообщения те прочитаем

REST Proxy UdemY

GET records?timeout=3000&max\_bytes=300000 Send

200 OK TIME 3 s SIZE 272 B

Body Auth Query Header 1 Docs

Accept d.kafka.binary.v2+json

New My-Header New Value

Preview Header 4 Cookie Timeline

```

1 [
2 {
3   "topic": "rest-binary",
4   "key": "a2V5",
5   "value": "aGVsbG8gd29ybGQhISE=",
6   "partition": 0,
7   "offset": 0
8 },
9 {
10  "topic": "rest-binary",
11  "key": null,
12  "value": "XCJyYW5kb206JSQh",
13  "partition": 0,
14  "offset": 1
15 },
16 {
17  "topic": "rest-binary",
18  "key": null,
19  "value": "bm8gcGFydGl0aW9ucw==",
20  "partition": 0,
21  "offset": 2
22 }

```

#### 4. закомитим офсет

REST Proxy UdemY

POST p/instances/my\_consumer\_binary/offsets Send

Other Auth Query Header 1 Docs

```

1
2 {
3   "offsets": [
4     {
5       "topic": "rest-binary",
6       "partition": 0,
7       "offset": 2
8     }
9   ]
10 }

```

JSON формат REST сообщения

## Producing with the REST Proxy JSON



- JSON data does not need to be transformed before being sent to the REST Proxy, as our POST request takes JSON as an input
- Any kind of valid JSON can be sent, there are no restrictions!
- Each language has support for JSON so that's a very easy way to send semi structured data.

- **It is the same as before, except the header changes**



- Steps are (for any data format):

1. Create consumer
2. Subscribe to a topic (or topic list)
3. Get records
4. Process records (your app)
5. Commit offsets (once in a while)

JSON	
POST	Create Consumer JSON
POST	Subscribe Consumer JSON
DEL	Delete Consumer JSON
GET	Consume JSON
POST	Commit Offsets

```
POST http://hostname:8082/topics/rest-json
JSON
1 {
2   "records": [
3     {
4       "key": "somekey",
5       "value": {"foo": "bar"}
6     },
7     {
8       "value": [ "foo", "bar" ],
9       "partition": 0
10    },
11    {
12      "value": 53.5
13    }
14  ]
15 }
```

## AVRO формат REST сообщения

# Producing with the REST Proxy Avro



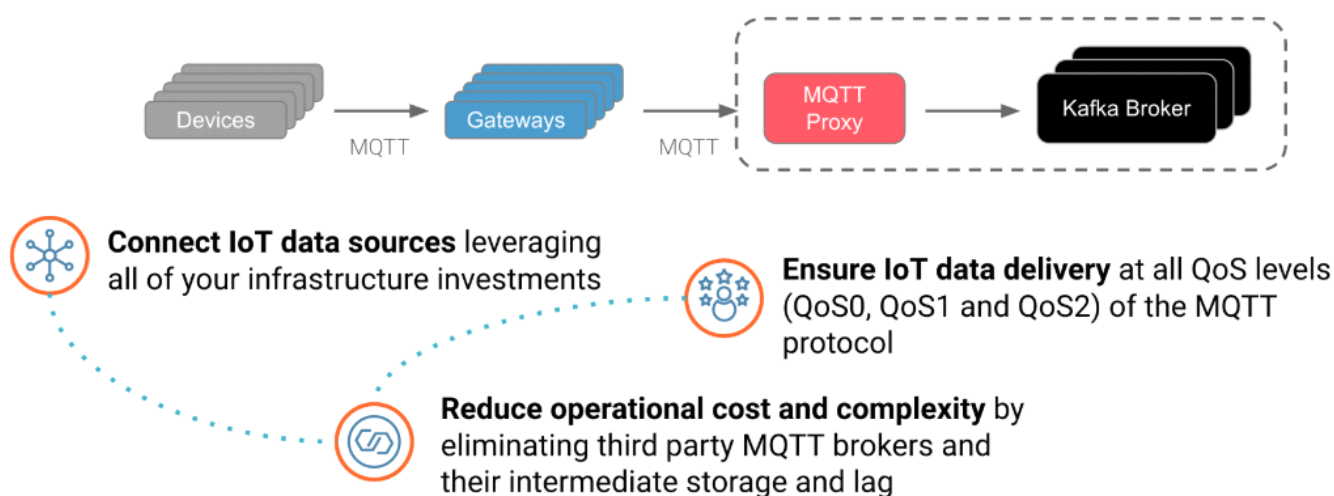
- The REST Proxy has primary support for Avro as it's directly connected to the Schema Registry
- You send the schema in JSON (stringified), and you send the Avro payload encoded in JSON
- After the first produce call, you can get a schema id to re-use in the next requests to make them smaller
- The header changes as well

```
POST http://hostname:8082/topics/rest-avro
Other
1 {
2   "value_schema": "{ \"type\": \"record\", \"name\": \"User\", \"fields\": [{ \"name\": \"name\", \"type\": \"string\" }, { \"name\": \"age\", \"type\": [ \"null\", \"int\" ] } ] }",
3   "records": [
4     {
5       "value": { "name": "testuser", "age": null }
6     },
7     {
8       "value": { "name": "testuser", "age": { "int": 25 } },
9       "partition": 0
10    }
11  ]
12 }
13 }
```

# MQTT protocol (для интернета вещей)

14 декабря 2020 г. 13:51

## MQTT Proxy



- Confluent Enterprise includes an MQTT (**Message Queuing Telemetry Transport**) proxy to allow direct ingestion of IoT data
- The MQTT proxy is intended to allow customers to add Kafka to their IoT architecture to enable stream processing. MQTT is a pub/sub messaging transport protocol (the **de facto standard for IoT devices**) that is designed to support thin lightweight clients that run in a very diverse set of devices and communicate asynchronously with scalable, more heavy weight, protocol aware MQTT brokers. Without this proxy, customers would have to maintain separate MQTT clients and brokers, adding layers and complexity to their architectures. The current implementation of the proxy is read-only (i.e., producer only).

пример продьюсера

```
private final int qos = 1;
private String host = "ssl://mqtt.hsl.fi:8883";
private String clientId = "MQTT-Java-Example";
private String topic = "/hfp/v2/journey/ongoing/vp/#";
private String kafka_topic = "vehicle-positions";
private MqttClient client;
```

We have the following:

- qos defines the **quality of service** of the MQTT service
- host is the URI of the host providing the data, which we're going to connect to
- clientId is just how we identify ourselves to the host
- topic is the **MQTT topic (NOT the Kafka topic!)** we're going to subscribe to
- kafka\_topic is the name of the Kafka topic we're going to write the MQTT data to
- Finally we are defining a variable client for the MqttClient instance



If you have problems with the amount of data that the MQTT source produces, you can adjust the value of topic to say only return vehicle positions for **trams**. The corresponding value would then be: /hfp/v2/journey/ongoing/vp/tram/#  
For more details please consult <https://digitransit.fi/en/developers/apis/4-realtime-api/vehicle-positions/>

9. Let's now define the start method:

```
package clients;
import java.util.Properties;
import org.apache.kafka.clients.producer.KafkaProducer;
import org.apache.kafka.clients.producer.ProducerConfig;
import org.apache.kafka.common.serialization.StringSerializer;
import org.eclipse.paho.client.mqttv3.*;
public class VehiclePositionProducer {
    public static void main(String[] args) throws MqttException {
        System.out.println("*** Starting VP Producer ***");
        Properties settings = new Properties();
        settings.put(ProducerConfig.CLIENT_ID_CONFIG, "vp-producer");
        settings.put(ProducerConfig.BootstrapServers_CONFIG, "kafka:9092");
        settings.put(ProducerConfig.KEY_SERIALIZER_CLASS_CONFIG, StringSerializer.class);
        settings.put(ProducerConfig.VALUE_SERIALIZER_CLASS_CONFIG, StringSerializer.class);
        final KafkaProducer<String, String> producer = new KafkaProducer<>(settings);

        Runtime.getRuntime().addShutdownHook(new Thread(() -> {
            System.out.println("### Stopping VP Producer ###");
            producer.close();
        }));

        Subscriber subscriber = new Subscriber(producer);
        subscriber.start();
    }
}
```

```
package clients;
import java.util.UUID;
import org.apache.kafka.clients.producer.KafkaProducer;
import org.apache.kafka.clients.producer.ProducerRecord;
import org.eclipse.paho.client.mqttv3.*;
import org.eclipse.paho.client.mqttv3.persist.MemoryPersistence;
public class Subscriber implements MqttCallback {
```

```

private final int qos = 1;
private String host = "ssl://mqtt.hsl.fi:8883";
private String clientId = "MQTT-Java-Example";
private String topic = "/hfp/v2/journey/ongoing/vp/#";
private String kafka_topic = "vehicle-positions";
private MqttClient client;
private KafkaProducer<String, String> producer;
public Subscriber(KafkaProducer<String, String> producer) {
    this.producer = producer;
}
public void start() throws MqttException {
    MqttConnectOptions conOpt = new MqttConnectOptions();
    conOpt.setCleanSession(true);
    final String uuid = UUID.randomUUID().toString().replace("-", "");

    String clientId = this.clientId + "-" + uuid;
    this.client = new MqttClient(this.host, clientId, new MemoryPersistence());
    this.client.setCallback(this);
    this.client.connect(conOpt);

    this.client.subscribe(this.topic, this.qos);
}
public void connectionLost(Throwable cause) {
    System.out.println("Connection lost because: " + cause);
    System.exit(1);
}
public void deliveryComplete(IMqttDeliveryToken token) {
}
public void messageArrived(String topic, MqttMessage message) throws MqttException {
    System.out.println(String.format("[%s] %s", topic, new String(message.getPayload())));
    final String key = topic;
    final String value = new String(message.getPayload());
    final ProducerRecord<String, String> record = new ProducerRecord<>
(this.kafka_topic, key, value);
    producer.send(record);
}
}

```

# (var1) MQTT connector

18 декабря 2020 г. 14:41

## Testing the MQTT Data Source

1. Let's install **Mosquitto**, a MQTT client:

```
$ sudo apt-get update && \
  sudo apt-get install -y software-properties-common && \
  sudo apt-add-repository ppa:mosquitto-dev/mosquitto-ppa && \
  sudo apt-get install -y mosquitto-clients
```

2. To subscribe to the MQTT data provider we can use **Mosquitto's** subscription tool as follows:

```
$ mosquitto_sub -h mqtt.hsl.fi -p 1883 -d -t "/hfp/v2/journey/ongoing/vp/#"
```

6. Create the Kafka Connect MQTT Source connector using the Connect REST API:

```
$ curl -s -X POST -H 'Content-Type: application/json' -d '{
  "name" : "mqtt-source",
  "config" : {
    "connector.class" : "io.confluent.connect.mqtt.MqttSourceConnector",
    "tasks.max" : "1",
    "mqtt.server.uri" : "tcp://mqtt.hsl.fi:1883",
    "mqtt.topics" : "/hfp/v2/journey/ongoing/vp/#",
    "kafka.topic" : "vehicle-positions",
    "confluent.topic.bootstrap.servers": "kafka:9092",
    "confluent.topic.replication.factor": "1",
    "confluent.license": "",
    "key.converter": "org.apache.kafka.connect.storage.StringConverter",
    "value.converter": "org.apache.kafka.connect.converters.ByteArrayConverter"
  }
}' http://connect:8083/connectors | jq .
```

7. Check if connector is loaded and status is 'RUNNING':

```
$ curl -s "http://connect:8083/connectors"
["mqtt-source"]
```

# (var2) Confluent MQTT proxy

18 декабря 2020 г. 14:42

1. If you haven't installed **Mosquitto** in the previous lab then follow these instructions:

```
$ sudo apt-get update && \
  sudo apt-get install -y software-properties-common && \
  sudo apt-add-repository ppa:mosquitto-dev/mosquitto-ppa && \
  sudo apt-get install -y mosquitto-clients
```

2. Navigate to the labs folder:

```
$ cd ~/confluent-dev/labs/mqtt-proxy
```

3. Run the Kafka cluster, including Kafka Connect:

```
$ docker-compose up -d
```

4. Open the file `docker-compose.yml` in your project folder and analyze the definition of the `mqtt` service:

```
29 mqtt:
30   image: confluentinc/cp-kafka-mqtt:5.3.0
31   hostname: mqtt
32   networks:
33     - confluent
34   ports:
35     - 1883:1883
36   environment:
37     KAFKA_MQTT_BOOTSTRAP_SERVERS: kafka:9092
38     KAFKA_MQTT_CONFLUENT_TOPIC_REPLICATION_FACTOR: 1
39     KAFKA_MQTT_TOPIC_REGEX_LIST: temperature:.*temperature, brightness:.*brightness
40     KAFKA_MQTT_LISTENERS: 0.0.0.0:1883
```

Specifically have an eye on how the environment variable `KAFKA_MQTT_TOPIC_REGEX_LIST` is defined, which maps MQTT topics to Kafka topics. In our case every MQTT topic that ends in `temperature` is mapped to the Kafka topic `temperature`. Every MQTT topic that ends in `brightness` is mapped to the Kafka topic



## \* python producer example

14 декабря 2020 г. 16:02

для связи с сервером не используется какой-либо специальной библиотеки

### A Python Producer Using the REST Proxy

```
1 #!/usr/bin/python
2
3 import requests
4 import json
5
6 url = "http://restproxy:8082/topics/my_topic"
7 headers = {
8     "Content-Type" : "application/vnd.kafka.json.v2+json"
9 }
10 # Create one or more messages
11 payload = {"records":
12     [{
13         "key": "firstkey",
14         "value": "firstvalue"
15     }]
16 }
17 # Send the message
18 r = requests.post(url, data=json.dumps(payload), headers=headers)
19 if r.status_code != 200:
20     print "Status Code: " + str(r.status_code)
21     print r.text
```

The target topic is specified as part of the URL, after the REST proxy.

This example shows how to use the JSON format. Refer to the docs.confluent.io for examples using the other formats (e.g., Avro-encoded JSON).



A question that often comes up during class wrt REST Proxy - how are all the producer/consumer config settings we talked about used through the REST Proxy?

**Answer:** in almost all cases, the configs have to be defined on the REST server, and cannot be specified as part of the client request.

## \* python consumer example

14 декабря 2020 г. 20:24

### Main Logic:

```
1 import requests
2 import json
3 import sys
4
5 FORMAT = "application/vnd.kafka.v2+json"
6 POST_HEADERS = { "Content-Type": FORMAT }
7 GET_HEADERS = { "Accepts": FORMAT }
8
9 base_uri = create_consumer_instance("group1", "my_consumer")
10 subscribe_to_topic(base_uri, "hello_world_topic")
11 consume_messages(base_uri)
12 delete_consumer(base_uri)
```

Using the REST Proxy as a consumer has a few more steps than the producer.

First we define a few global variables:

1. **FORMAT**: defines Kafka's specific JSON format in version 2, used in HTTP request
2. **POST\_HEADERS**: headers used for **POST** requests
3. **GET\_HEADERS**: headers used for **GET** requests

The actual application logic:

1. Create a consumer instance called **my\_consumer** in consumer group **group1**
2. Subscribe the consumer to the topic **hello\_world\_topic**
3. Now consume messages
4. When done, delete the consumer instance to avoid resource leaks



**base\_uri** is used to identify the consumer instance

## Creating the Consumer Instance:

```
1 def create_consumer_instance(group_name, instance_name):
2     url = f'http://restproxy:8082/consumers/{group_name}'
3     payload = {
4         "name": instance_name,
5         "format": "json"
6     }
7     r = requests.post(url, data=json.dumps(payload), headers=POST_HEADERS)
8
9     if r.status_code != 200:
10        print("Status Code: " + str(r.status_code))
11        print(r.text)
12        sys.exit("Error thrown while creating consumer")
13
14    return r.json()["base_uri"]
```

In this part of the example, we are creating the instance of the consumer. Notice that the `url` is referencing the consumers rather than a topic name as we saw in the producer example.

1. We define the `url` to the desired consumer group
2. The payload defines among other things the instance name we want to use
3. Now we do an HTTP POST request to the `url`
4. In case of an error we stop the application with `sys.exit(...)`
5. As a last step, we return `base_uri`, which is used to identify the consumer instance

## Subscribing to a topic

```
1 def subscribe_to_topic(uri, topic_name):
2     payload = {
3         "topics": [topic_name]
4     }
5
6     r = requests.post(uri + "/subscription",
7                       data=json.dumps(payload),
8                       headers=POST_HEADERS)
9
10    if r.status_code != 204:
11        print("Status Code: " + str(r.status_code))
12        print(r.text)
13        delete_consumer(uri)
14        sys.exit("Error thrown while subscribing the consumer to the topic")
```

In the body of the POST request we send the list of topics (here only a single one) that we want this consumer to subscribe to.

If there is an error, we delete the consumer and then exit the application

## Consuming and processing messages:

```
1 def consume_messages(uri):
2     r = requests.get(base_uri + "/records", headers=GET_HEADERS, timeout=20)
3
4     if r.status_code != 200:
5         print ("Status Code: " + str(r.status_code))
6         print (r.text)
7         sys.exit("Error thrown while getting message")
8
9     for message in r.json():
10        if message["key"] is not None:
11            print ("Message Key:" + message["key"])
12            print ("Message Value:" + message["value"])
```

In a more realistic example you would loop indefinitely, since a topic is an open ended stream of data.

## Deleting the consumer:

```
1 def delete_consumer(uri):
2     r = requests.delete(uri, headers=GET_HEADERS)
3
4     if r.status_code != 204:
5         print ("Status Code: " + str(r.status_code))
6         print (r.text)
```

To verify that the consumer instances are removed properly, explicitly remove them using the **DELETE** call before exiting your code.

**statefull call** - используется четыре вызова

если не сделать deleting consumer то он отвалится по таймауту

## \* python consumer example: manual commit

16 декабря 2020 г. 17:33

### Manually Committing Offsets From the REST Proxy

- It is possible to manually commit offsets from the REST Proxy

```
1 payload = {
2     "format": "binary",
3     # Manually/Programmatically commit offset
4     "auto.commit.enable": "false"
5 }
6
7 headers = {
8     "Content-Type" : "application/vnd.kafka.v2+json"
9 }
10
11 r = requests.post(baseurl, data=json.dumps(payload), headers=headers)
12
13 # Commit the offsets
14 if shouldCommit() == True:
15     r = requests.post(base_uri + "/offsets", headers=headers, timeout=20)
16     if r.status_code != 200:
17         print ("Status Code: " + str(r.status_code))
18         print (r.text)
19         sys.exit("Error thrown while committing")
20     print "Committed"
```



Offset committed == offset of **next record to read**

- The offset committed (whether automatically or manually) is the offset of the next record to be read
  - Not the offset of the last record which was read

## \* python producer+consumer с реестром схем

17 декабря 2020 г. 21:53

### REST API Avro Producer Example

```
1 # Read in the Avro files
2 key_schema = open("my_key.avsc", 'rU').read()
3 value_schema = open("my_value.avsc", 'rU').read()
4
5 producerurl = "http://kafkarest1:8082/topics/my_avro_topic"
6 headers = {
7     "Content-Type" : "application/vnd.kafka.avro.v2+json"
8 }
9 payload = {
10     "key_schema": key_schema,
11     "value_schema": value_schema,
12     "records":
13     [{
14         "key": {"suit": "spades"},
15         "value": {"suit": "spades", "denomination": "ace"}
16     }]
17 }
18 # Send the message
19 r = requests.post(producerurl, data=json.dumps(payload), headers=headers)
20 if r.status_code != 200:
21     print "Status Code: " + str(r.status_code)
22     print r.text
```

Note that the producer code sends its payload as JSON (and **not** as AVRO). The communication between the producer and the REST Proxy is over HTTP. Only the REST Proxy deals with AVRO serialization/deserialization.



The examples on this page and the next - although written in Python - use the REST Proxy, **not** the Python client.

This is just an example on **how to use** the REST API and not a recommendation for a Python client...



## REST API Avro Consumer Example

```
1 # Get the message(s) from the consumer
2 headers = {
3     "Accept" : "application/vnd.kafka.avro.v2+json"
4 }
5 # Request messages for the instance on the topic
6 r = requests.get(base_uri + "/topics/my_avro_topic", headers=headers, timeout=20)
7 if r.status_code != 200:
8     print "Status Code: " + str(r.status_code)
9     print r.text
10    sys.exit("Error thrown while getting message")
11 # Output all messages
12 for message in r.json():
13     keysuit = message["key"]["suit"]
14     valuesuit = message["value"]["suit"]
15     valuecard = message["value"]["denomination"]
16     # Do something with the data
```

Don't be surprised to see the client code deal with JSON and **not** with AVRO. AVRO serialization and deserialization is handled by the REST Proxy. Our client communicates via HTTP REST calls with the Proxy. From a consumer's perspective AVRO is hidden away. Only the producer needs to know about AVRO, as it needs to send the schemas with the POST request to the REST Proxy.