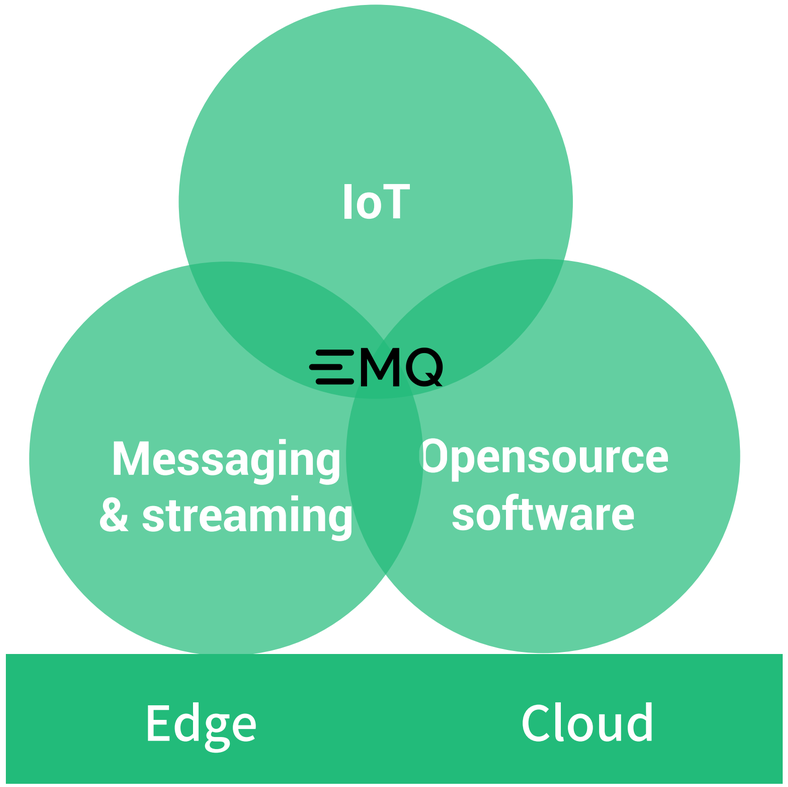
EMQ X Kuiper introduction

Rocky Jin / rocky@emqx.io

Apr, 2021

EMQ Technologies - OSS leader for IoT messaging and streaming



EMQ Technologies Co., Ltd.

1. COSS
2. For IoT in 5G era
3. Messaging and streaming
4. 10000+ global users
5. Globally operation: China, US & EU

The world #1 MQTT opensource message broker

World #1

The leader in open source MQTT broker

100M+

100M+ connected IoT devices

5M+

5M+ downloads and continuing to grow rapidly

20K+

20,000+ cluster deployments globally

10K+

10,000+ enterprise users globally

7K+

7K+ GitHub Stars

50+

Users in 50+ countries and regions

Streaming analytics

- Streaming analytics
  - A software or framework for stateful computations over unbounded data streams.
  - Allows management, monitoring, and real-time analytics of live streaming data.
  - Now typically running at cloud data center.

- Apache Flink & Spark are not fit for edge streaming analytics
  - Latency
  - Data security
  - Bandwidth costs

- The challenges for edge streaming analytics
  - Lightweight & high efficiency: restricted resource (CPU & Memory) in edge side
  - Agile & flexible: need more agile approach to update the biz logic
  - Deployment & management: not centralized, distributed deployed and weak network access

Kuiper - OSS IoT Edge analytics

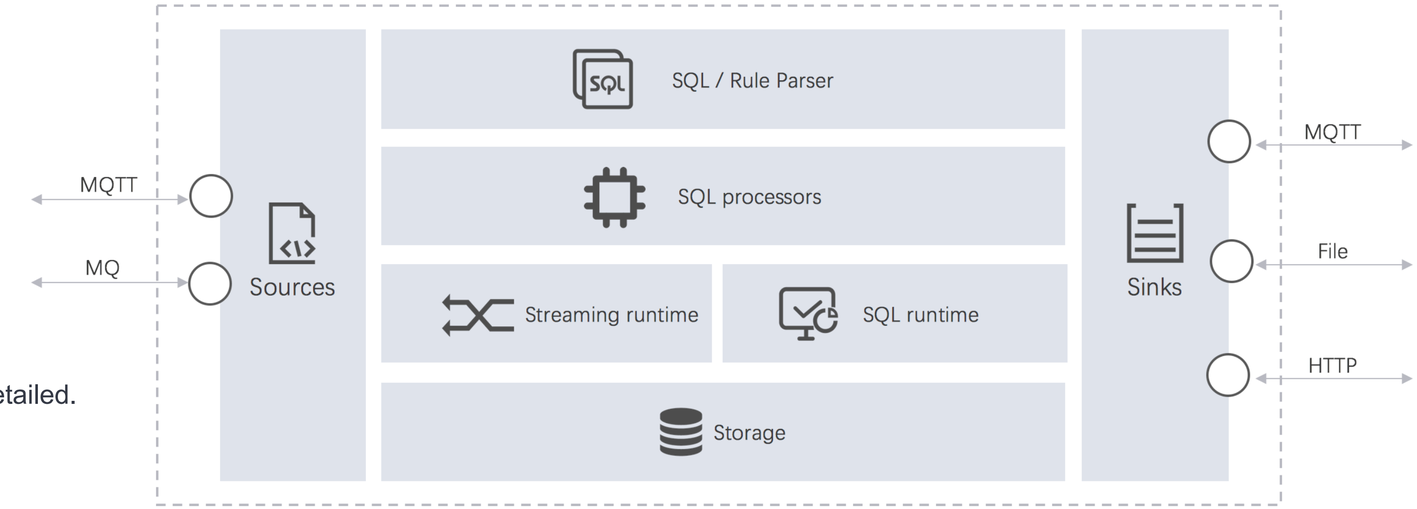
- Kuiper Milestones
  - Apr, 2021: about 1500 download per week
  - Feb, 2021: released 1.1.1, supported binary data processing, and ML/AI function support
  - Oct, 2020: released 1.0.0, the 1st stable major release
  - Jul, 2020: released 0.5.1, and integrated with KubeEdge
  - Apr, 2020: released 0.3.2, and integrated with EdgeX Foundry
  - Oct, 2019: open sourced & released the 1st version

Kuiper overview

- Binary installable & Docker images
  - 8MB install package; 10MB initial mem overhead
  - X86 AMD * 32, X86 AMD * 64; ARM * 32, ARM * 64; PPC
  - Linux distributions, OpenWrt Linux, MacOS and Docker

- Performance
  - Raspberry Pi 3B+
    - TPS: 12k
    - CPU: sys+user 70%
    - Memory: 20M

  * Refer to https://github.com/emqx/kuiper#performance-test-result for more detailed.



- Kuiper - Data ETL
  - Data extraction: sources
  - Data transformation: analytics + transformation with SQL
  - Data loading: Sinks

3 steps to use Kuiper

- Create a stream

```
create stream demo '() WITH (FORMAT="JSON", DATASOURCE="$hw/events/device/+/twin/update")'
```

- Create a rule

```
{
  "sql": "SELECT data->tag1->value AS temperature, data->tag2->value AS humidity FROM demo",
  "actions": [
    {
      "log": {}
    },
    {
      "mqtt": {
        "server": "tcp://broker.emqx.io:1883",
        "topic": "devices/result",
        "qos": 1,
        "clientId": "demo_001"
      }
    }
  ]
}
```

- Submit & run the rule

```
curl -X POST \ http://$kuiper_server:9081/rules \ -H ‘Content-Type: application/json’ \ -d ‘$my_rule’
```

SQL analytics

- Functions
  - Math: sin, cos, abs, log, mod etc; Totally 25 functions
  - String: concat, substring etc; Totally 19 functions
  - Aggregation: avg, count, max, min, sum, collect & deduplicate; Totally 7 functions
  - Conversion/ Encoding & decoding / Hashing / JSON processing / Others; Totally 18 functions
- Filter
  - WHERE / CASE WHEN
- Join (LEFT | RIGHT | FULL | CROSS JOIN)
  - Streams: Dynamic flowing data
  - Tables: Static data, which normally is used for associating additional info. E,g, user has an id, and to get related name.
- Window
  - Tumbling / Hopping / Sliding / Session / Count
- Group By & Order By

Advanced analytics

- Binary data type support
  - Allows user to analyze image, audio etc
- Binary image processing
  - resize – to resize the image before sending to cloud
  - thumbnail – to reduce the image that retains the aspect ratio to the maximum size

```
SELECT resize(avg,width,height) AS r1 FROM test;
```

- Geohash
  - geohashEncode, geohashDecode, geohashNeighbor & geohashBoundingBox etc for processing longitude & latitude, totally 10 functions

```
Input: {"lo" :131.036192,"la":-25.345457}
Output: {"geohashEncode":"qgmpvf18h86e"}

SELECT geohashEncode(la,lo) FROM test
```

- ML/AI streaming processing
  - Encapsulate ML/AI with Kuiper plugin: Better performance, but with higher dev & maintenance effort
  - Call ML/AI services by RPC or Rest-API: Lower dev & deployment effort, sacrifice some performance

```
Input: Image byte array
Output: {"labelImage":"peacock"}

SELECT labelImage(self) FROM tfdemo
```

Extension & plugins

- Extension points
  - Source
  - Sink
  - Function
- Steps
  - Development & debug
  - Compile to *.so file
  - Deploy plugins
- Native plugin development
  - Pros: Better performance
  - Cons: Strict limitations from Golang, includes same go version, strict lib dependencies and GOPATH

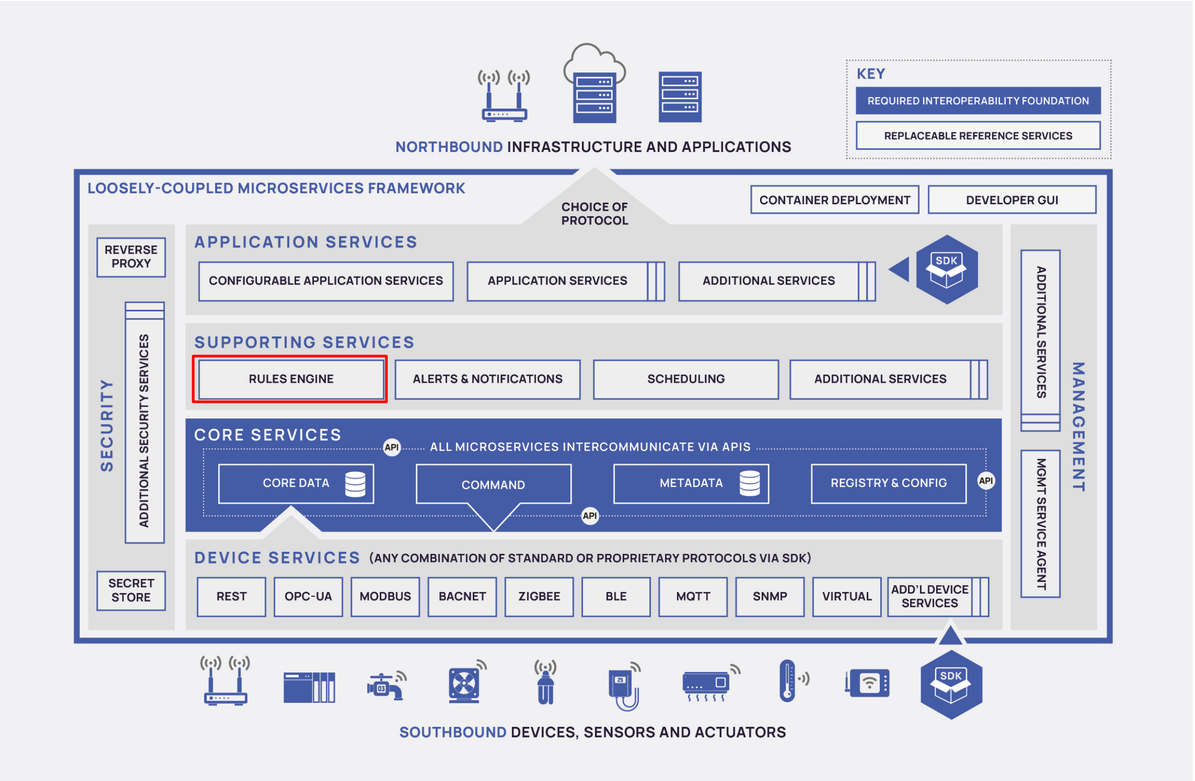
```
//Called during initialization. Configure the source with the data source(e.g. topic for mqtt)
//and the properties read from the yaml
Configure(datasource string, props map[string]interface{}) error

//Should be sync function for normal case. The container will run it in go func
Open(ctx StreamContext, consumer chan<- SourceTuple, errCh chan<- error)

Close(ctx StreamContext) error

function MySource() api.Source{
    return &mySource{}
}
```

Source extension: required interface for source

More user scenarios

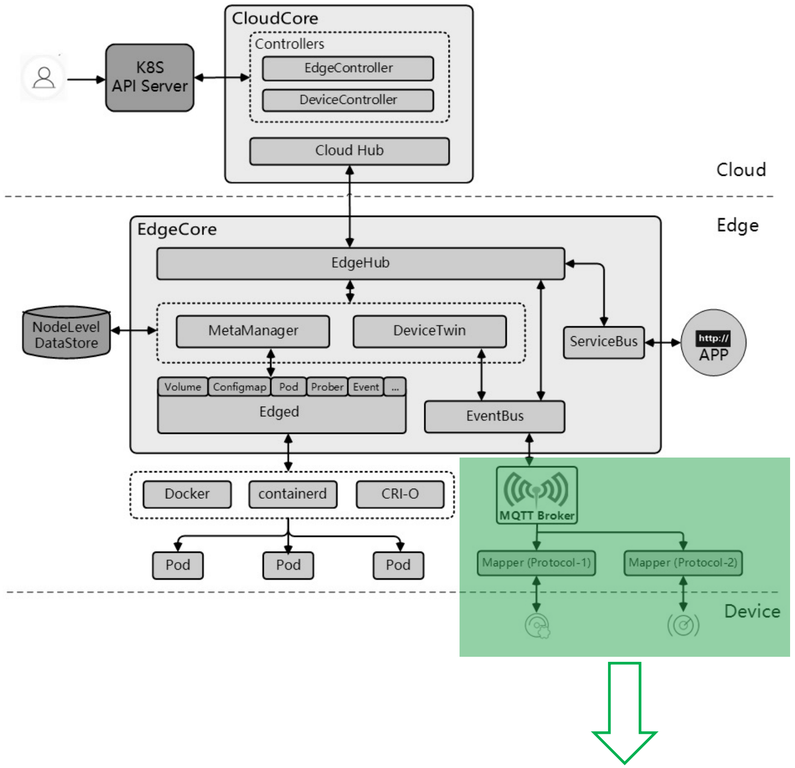
- Rules engine
  - EdgeX Foundry: The referenced rule engine

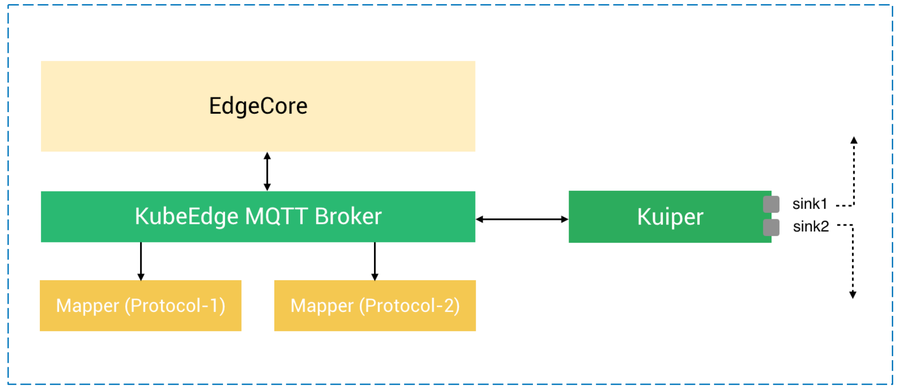
- Data format & protocol conversion
  - Flexible extension capabilities: data can be processed with build-in & extended functions from different source
  - Customize output data format of sinks: result can be customized with data template to adapt to different target systems
  - A customer case is using SAP NetWeaver RFC SDK to extract data from SAP system, and send to other system after processing



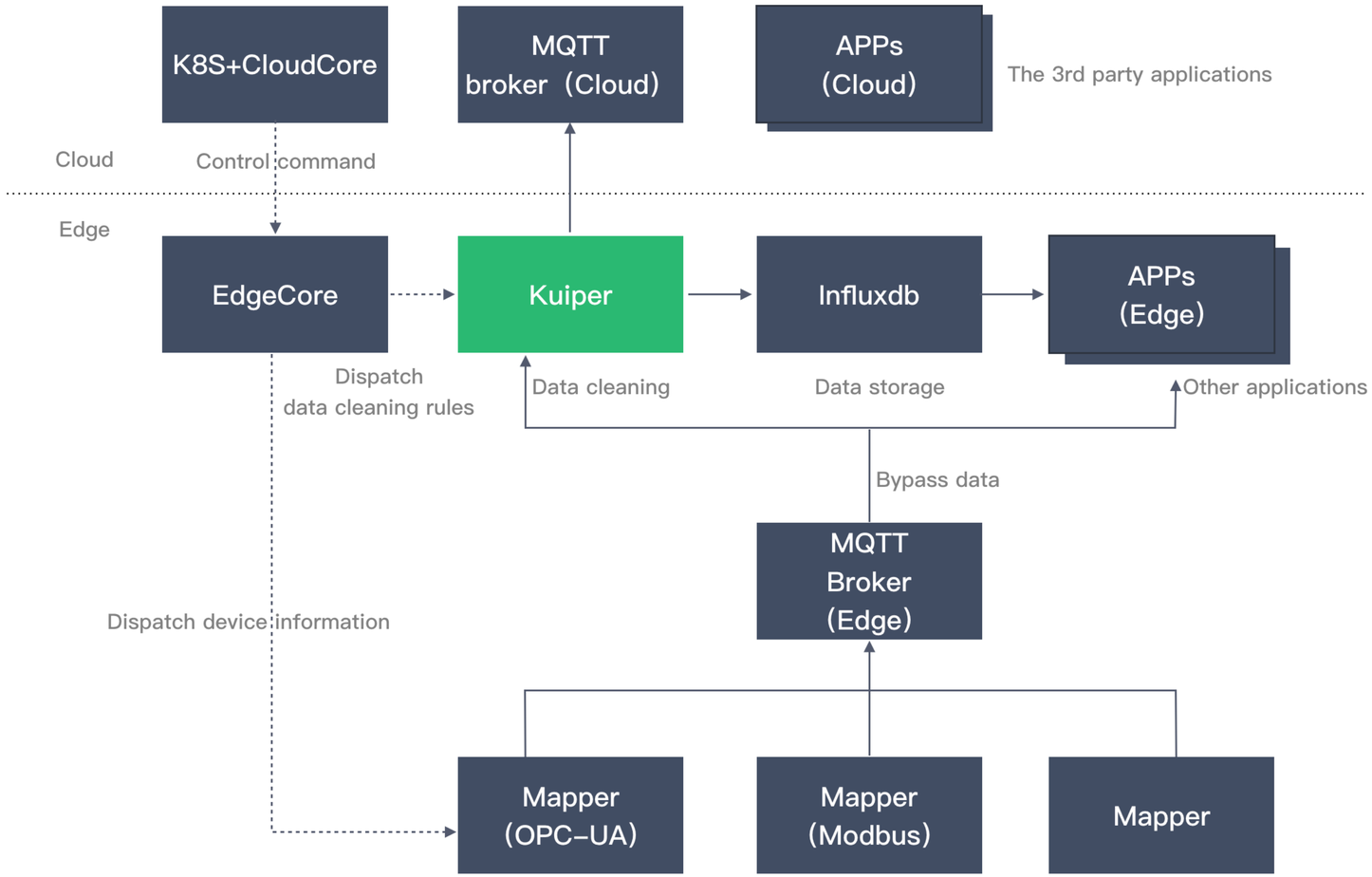
KubeEdge & Kuiper integration

- KubeEdge
  - An open-source project extending native containerized application orchestration capabilities to hosts at Edge
- Kuiper enhanced the edge analytics capabilities
- Benefits – resolved IoT edge computing challenges
  - Lower latency, bandwidth cost saving
  - Easy for user to implement business logic
  - Manage & deploy Kuiper, applications & AI algorithm from cloud





Customer case: KubeEdge + Kuiper



China Mobile – IIoT Big Data Center

EMQ

Next step

- Collaborate with more open-source projects

- More features will be introduced at 2021
  - The 3rd party language plugin development support
  - Persistence support with the 3rd party frameworks, such as Redis
  - More detailed, refer to 2021 roadmap - https://github.com/emqx/kuiper/projects/10

EMQ

THANKS