YANG push Data Collection integration into Apache Kafka Message Broker

Internship timeline with Huawei

Description

Network operators collect Network Telemetry [1] metrics with BMP [2], IPFIX [3] and YANG push [4] to gain analytical insights in their networks. The Big Data architecture has going through several evolution steps toward a decentralized approach called Data Mesh [5] in the last decade. Data Mesh allows data to be shared among different organizations and scales therefore in terms how many data scientists can work on the same data set compared to a centralized Data Lake approach. From surveys among different network operators, we know that Apache Kafka is the message broker of choice for Data Mesh.

In Data Mesh the standardization of operational data models plays a central role to allow interoperability among organizations. In Network Analytics, IETF and the YANG data modelling language [6] is going to play evermore important role but imposes a challenge to Apache Kafka and its connected systems today.

During this internship you will first learn what YANG is, how it is being used for network automation and for collecting operational metrics. You will be introduced to the SAIN architecture [7] and learn how it is enabling closed loop operation.

You will research and document how YANG push works, what metadata it provides today and propose with IETF authors together how this needs to be extended to enable proper YANG schema integration into Data Mesh schema registry. You will extend Pmacct [8] open-source and a Huawei closed source YANG push data collection by developing a library to recognize and obtain the matching YANG schema, register it to the Confluent Kafka schema registry and hand over the new schema id for data serialization into Apache Kafka message broker.

You will be part of an industry expert group from Confluent, INSA Lyon, Pmacct, Huawei, Swisscom and Imply who will develop other parts such as, Confluent Schema Registry and Apache Kafka serializer extension in parallel. Most of the experts are active at the IETF NETCONF and NETMOD working groups where YANG is being standardized and versioning is currently being extended.

Finally, you can present your internship results at the IETF 118 NETCONF/NETMOD working group between November 4-10th 2023 to other network operators, vendors and universities.

Requirements

Good understanding in C development, Linux network TCP/IP stack and the Network Telemetry framework. Some basic understanding in Netconf and YANG are an advantage. Don't be scared about the application and implementation parts.

Timetable

Table 1: Suggested schedule for 6 months (26 weeks)

Milestone	Estimated Effort
Onboarding: Setting up and getting to know the IETF interoperability lab and the peers.	2 weeks
Data Collection Baseline: Learn what NETCONF and YANG push is, how it is being used for network automation and monitoring.	2 weeks
SAIN Baseline: Learn what SAIN is, what the current development state is and how it provides closed loop operation.	1 week
Obtain YANG schema reference: Extend code so that YANG push metadata is parsed to identify xpath.	2 weeks
Obtain YANG schema: Extend code so that YANG schema is obtained through NETCONF.	5 weeks
Register YANG schema: Extend code so that obtained YANG schema is registered and cached for data serialization.	2 weeks
Wrap up open-source project: Test and document developed open-source code and submit pull request.	3 weeks
Huawei closed-source project: Extend code so that same logic developed in open-source can also be applied in closed-source.	7 weeks
Wrap up closed-source project: Test and document developed closed-source code and submit pull request.	2 weeks

References

[1] H. Song, F. Qin, P. Martinez-Julia, L. Ciavaglia, A. Wang, Network Telemetry Framework, https://datatracker.ietf.org/doc/html/rfc9232

[2] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", https://www.rfc-editor.org/info/rfc7854

[3] B. Claise, Ed., B. Trammell, Ed., P. Aitken, Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, https://tools.ietf.org/html/rfc7011

[4] E. Voit, A. Clemm, A. Gonzalez Prieto, E. Nilsen-Nygaard, A. Tripathy, Subscription to YANG Notifications, https://datatracker.ietf.org/doc/html/rfc8639

[4] E. Voit, A. Clemm, A. Gonzalez Prieto, E. Nilsen-Nygaard, A. Tripathy, Dynamic Subscription to YANG Events and Datastores over NETCONF, https://datatracker.ietf.org/doc/html/rfc8640

[5] Z. Dehghani, Data Mesh Principles and Architecture, https://martinfowler.com/articles/data-mesh-principles.html

[5] Z. Dehghani, How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh, https://martinfowler.com/articles/data-monolith-to-mesh.html

[6] M. Bjorklund, Ed., The YANG 1.1 Data Modeling Language, https://datatracker.ietf.org/doc/html/rfc7950

[7] B. Claise, J. Quilbeuf, D. Lopez, D. Voyer, T. Arumugam, Service Assurance for Intent-based Networking Architecture, https://datatracker.ietf.org/doc/html/draft-ietf-opsawg-service-assurance-architecture

[8] P. Lucente, A small set of multi-purpose passive network monitoring tools, https://github.com/pmacct/pmacct