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 research and document how YANG push works, what metadata it provides today, propose and document how it can be applied to Data Mesh and what needs to be developed to enable proper semantics integration into Data Mesh messaging. You apply this to the Pmacct [8] open-source and a Huawei closed source YANG push data collection by developing a new library to parse the meta data and use an existing open-source library to obtain a YANG module from a YANG data store.

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 (working group and/or hackathon to be decided later) between November 4-10th 2023 to other network operators, vendors and universities.

Requirements

Good understanding in data engineering, C development, Linux network TCP/IP stack and the Network Telemetry framework. Some basic understanding in network automation with 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.	1 week
Data Collection and SAIN Baseline: Learn what NETCONF, YANG push and SAIN is, how it is being used for network automation, monitoring and closed loop operation	1 week
Research and Propose YANG integration: Research and document how YANG push messaging and semantics needs to be preserved and integrated into Apache Kafka message broker. Research how a YANG schema can be obtained from a YANG data store. Propose results in IETF draft document.	3 weeks
Obtain YANG schema reference: Develop a new library where YANG push metadata is being parsed and stored.	4 weeks
Integrate YANG push library into pmacct: Integrate developed library into pmacct YANG push data collection.	3 weeks
Obtain YANG schema: Research which opensource library could be used to obtain YANG semantics from a YANG data store and integrate into pmacct.	3 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.	6 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

[5] B. Claise, Data Mesh Principles in the Networking World, https://www.claise.be/data-mesh-in-the-networking-world/

[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

[7] B. Claise, J. Quilbeuf, P. Lucente, P. Fasano, T. Arumugam, YANG Modules for Service Assurance, https://datatracker.ietf.org/doc/draft-ietf-opsawg-service-assurance-yang/

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