Beyond Messaging

Enterprise Dataflow with Apache NiFi

Presented by: Joe Witt
Apache NiFi PMC Member
July 15, 2015 – Achieved Top Level Project status at the Apache Software Foundation!
The problem space: Enterprise Dataflow

Acquire Data

Process / Analyze Data

Store Data
Buzzword level use cases to consider

- Remote sensor delivery (IoT)
- Intra-site / Inter-site / global distribution (Enterprise)
- Ingest (Big Data)
- Data Processing (Simple Event Processing)
The challenges we faced

- Messaging addresses a tiny subset of the problem
- Needed to understand the big picture
- Needed the ability to make *immediate* changes
- Must maintain chain of custody for data
- Rigorous security and compliance requirements
There are many solid options (both open and closed):
• Kafka, ActiveMQ, Tibco, … a lot more

But for this talk let’s invent the Chuck Norris of messaging systems

It’s features:
• It has zero latency
• It has perfect data durability
• It supports unlimited consumers and producers
When dataflow is viewed as a messaging problem

“But my system needs….”

- A different format and/or schema
- To use a different protocol
- The highest priority information first
- Large objects (event batches) / Small Objects (streams)
- Authorization to the data level (topic level is too broad)
- Only interested in a subset of data on a topic
- Data needs to be enriched/sanitized before it arrives
When dataflow is viewed as a messaging problem

“The solution to those issues…”

• Add new systems to handle the protocol differences
• Add new systems to convert the data
• Add new systems to reorder the data
• Add new systems to filter the unauthorized data
• Add new topics to represent ‘stages of the flow’


Who/What manages the bulk of the complexity?
• The producers and consumers.
• The operations teams :: Hunting down log files…
Introducing Apache NiFi

NiFi is built to tackle enterprise dataflow problems. Key concepts:

- Interactive command and control
- Data lineage
- Platform built for extension (FBP)
- High performance content and metadata repositories
NiFi Architecture
Live Flow
Things the NiFi Community is Discussing

Roadmap Highlights

High Availability Cluster Manager

Live / Rolling Upgrades

Provenance Query Language / Reporting

A complete user experience enabled by provenance

Extension and Template Registry
Learn more about Apache NiFi

Apache NiFi site
http://nifi.apache.org

Subscribe to and collaborate at
dev@nifi.apache.org
users@nifi.apache.org

Submit Ideas or Issues
https://issues.apache.org/jira/browse/NIFI

@apachenifi