



The First Mile – Edge and IoT Data Collection with Apache NiFi and MiNiFi

Andy LoPresto | @yolopey

Sr. Member of Technical Staff at Hortonworks, Apache NiFi PMC & Committer

06 February 2019 Dataworks Summit Melbourne

Acknowledgement of Country

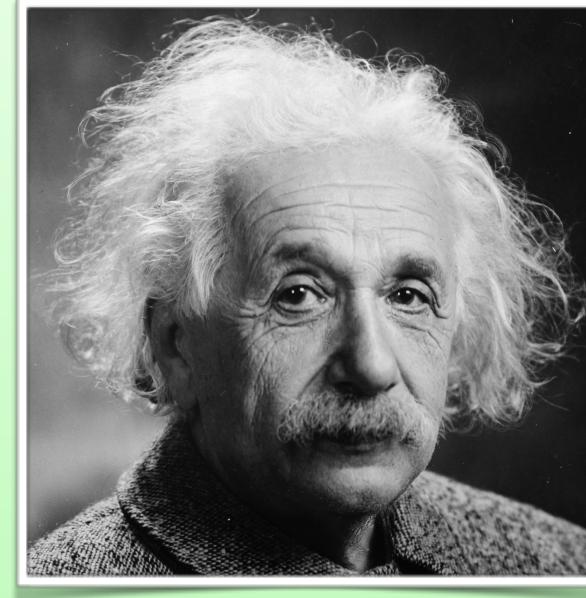
I acknowledge the Traditional Owners of the land on which we are meeting. I pay my respects to their Elders, past and present, and the Aboriginal Elders of other communities who may be here today.

Gauging Audience Familiarity With NiFi



“What’s a NeeFee?”

No experience with dataflow
No experience with NiFi



“I can pick this up pretty quickly”

Some experience with dataflow
Some experience with NiFi

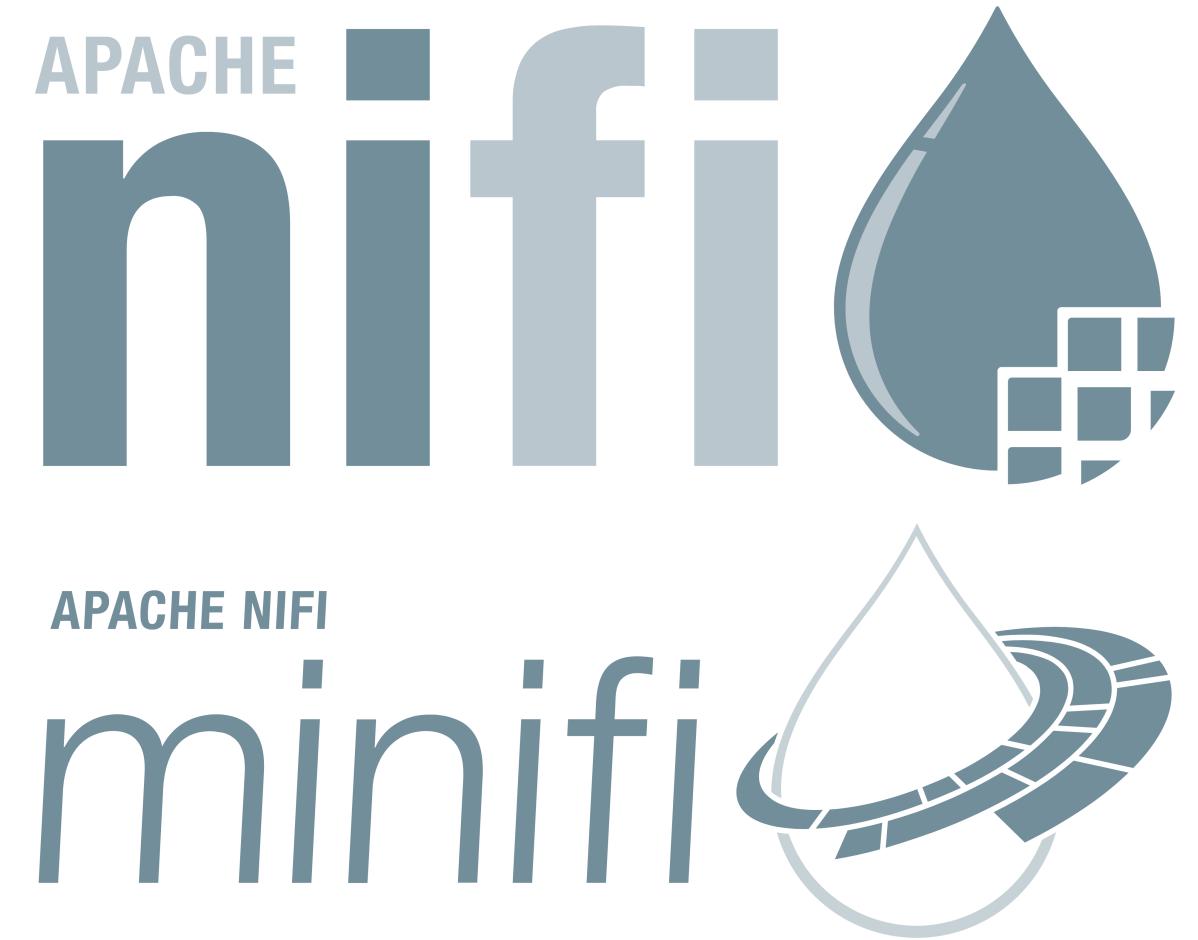


“I refactored the Ambari integration endpoint to allow for mutual authentication TLS during my coffee break”

Forgotten more about NiFi
than most of us will ever
know

Agenda

- *What is dataflow and what are the challenges?*
 - *Apache NiFi*
 - IoT Challenges
 - Apache MiNiFi
 - Exploration
 - Community
-
- All slides provided online, so no need to transcribe

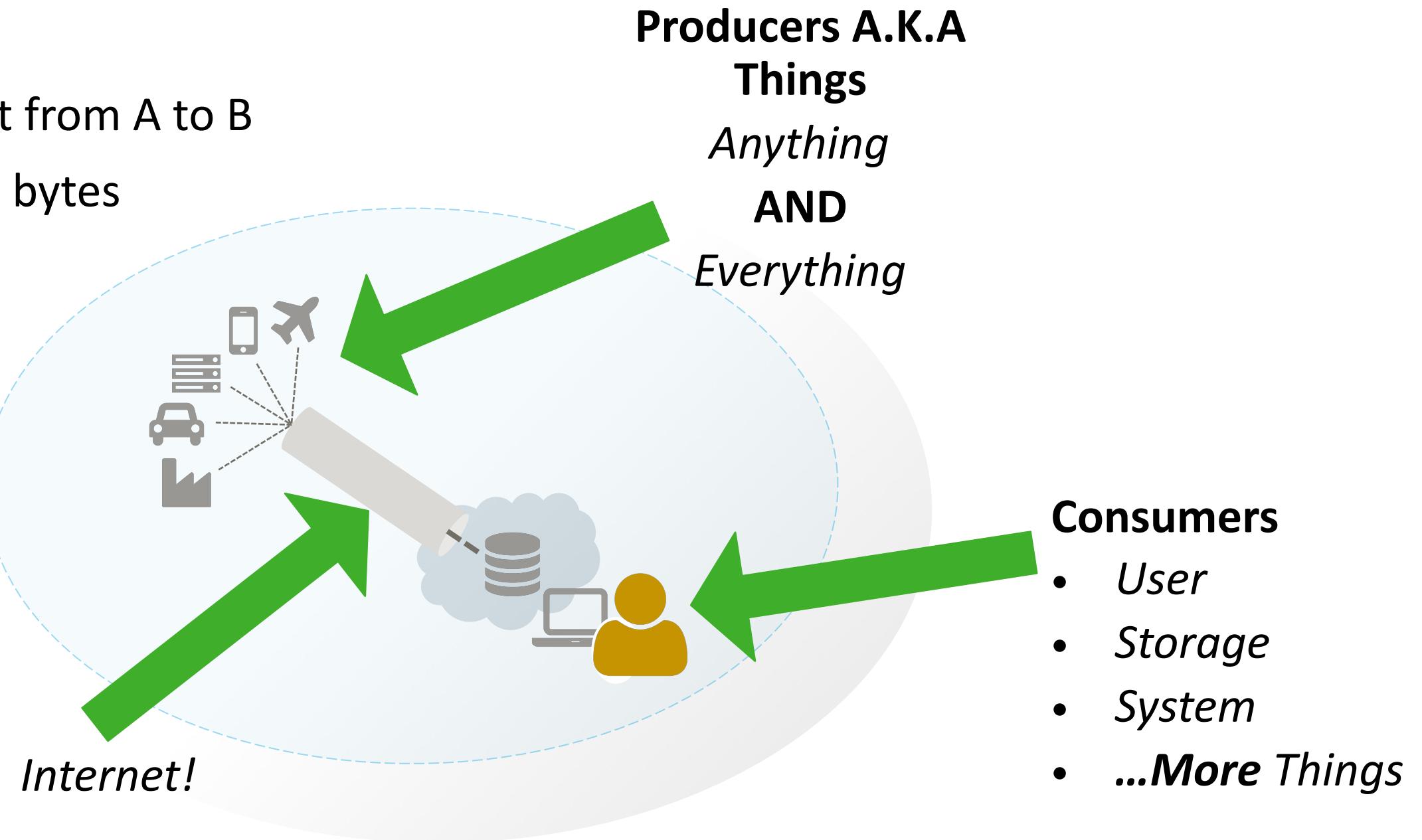




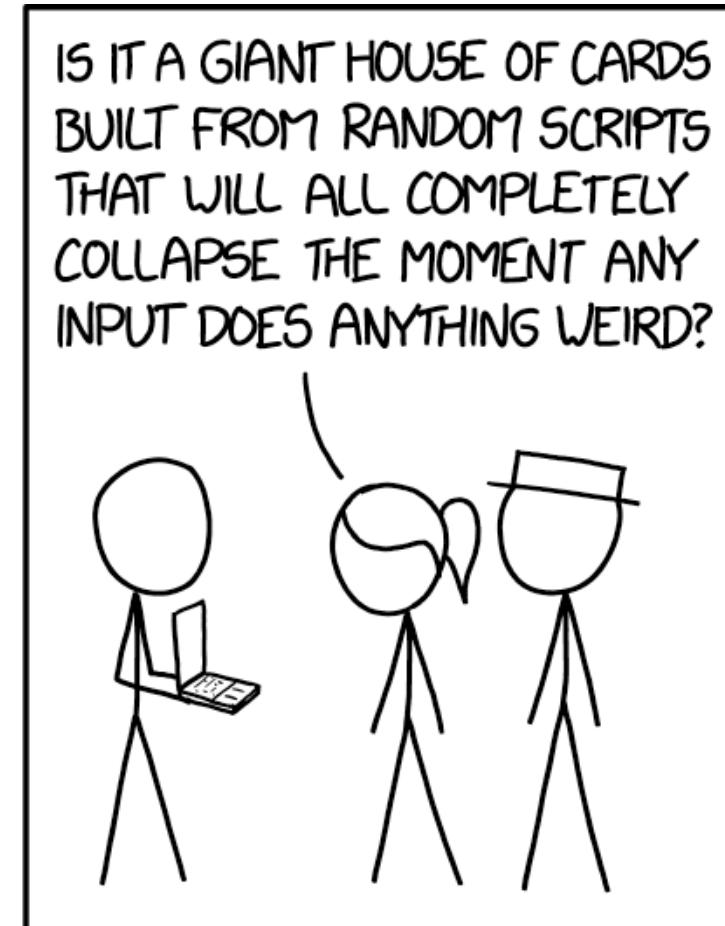
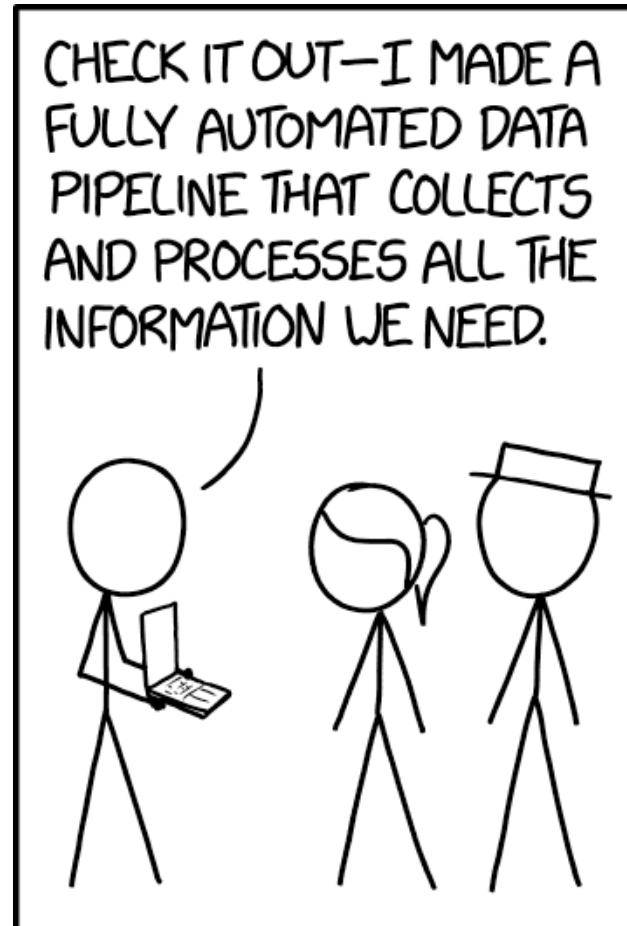
What is dataflow?

What is dataflow?

- Moving some content from A to B
- Content could be any bytes
 - Logs
 - HTTP
 - XML
 - CSV
 - Images
 - Video
 - Telemetry



Moving data *effectively* is hard



“Data Pipeline” <https://xkcd.com/2054/>

Dataflow Challenges In 3 Categories

Data

- Standards
- **Formats**
- Protocols
- Veracity
- Validity
- Schemas
- Partitioning/
Bundling

Infrastructure

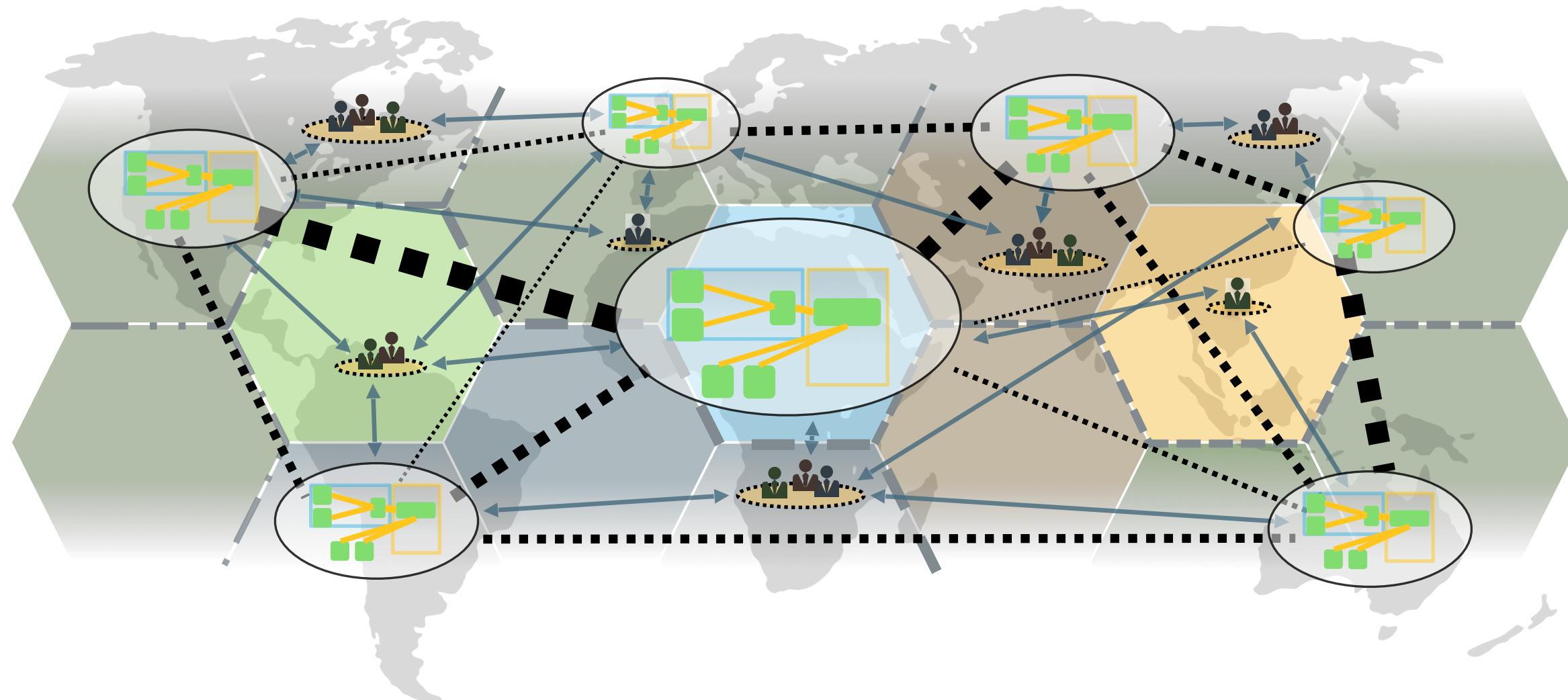
- “Exactly Once”
Delivery
- Ensuring
Security
- **Overcoming**
Security
- Credential
Management
- Network

People

- Compliance
- “**That** [person |
team | group]”
- **Consumers
Change**
- **Requirements
Change**
- “Exactly Once”
Delivery

Let's Connect Lots of As to Bs to As to Cs to Bs to Δ s to Cs to φ s

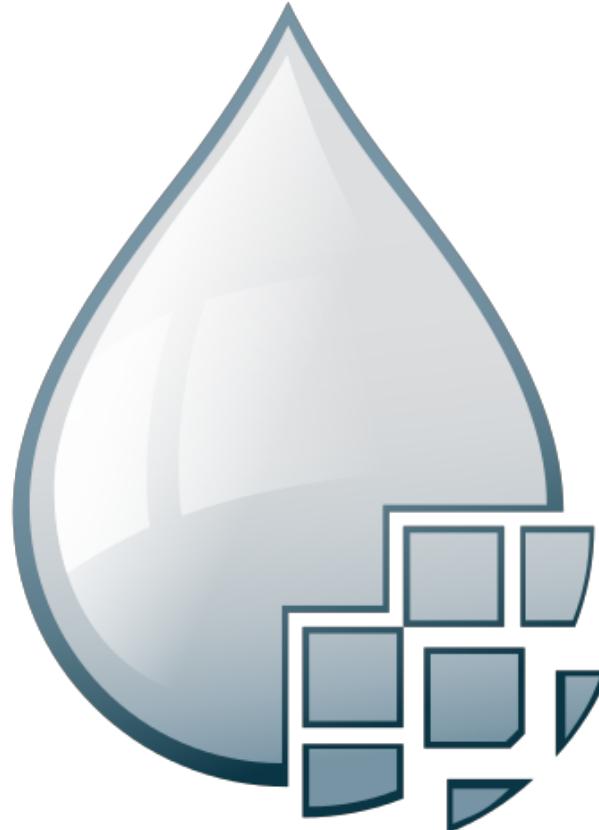
Raise your hand if you want to maintain Python scripts for the rest of your life



What is Apache NiFi?

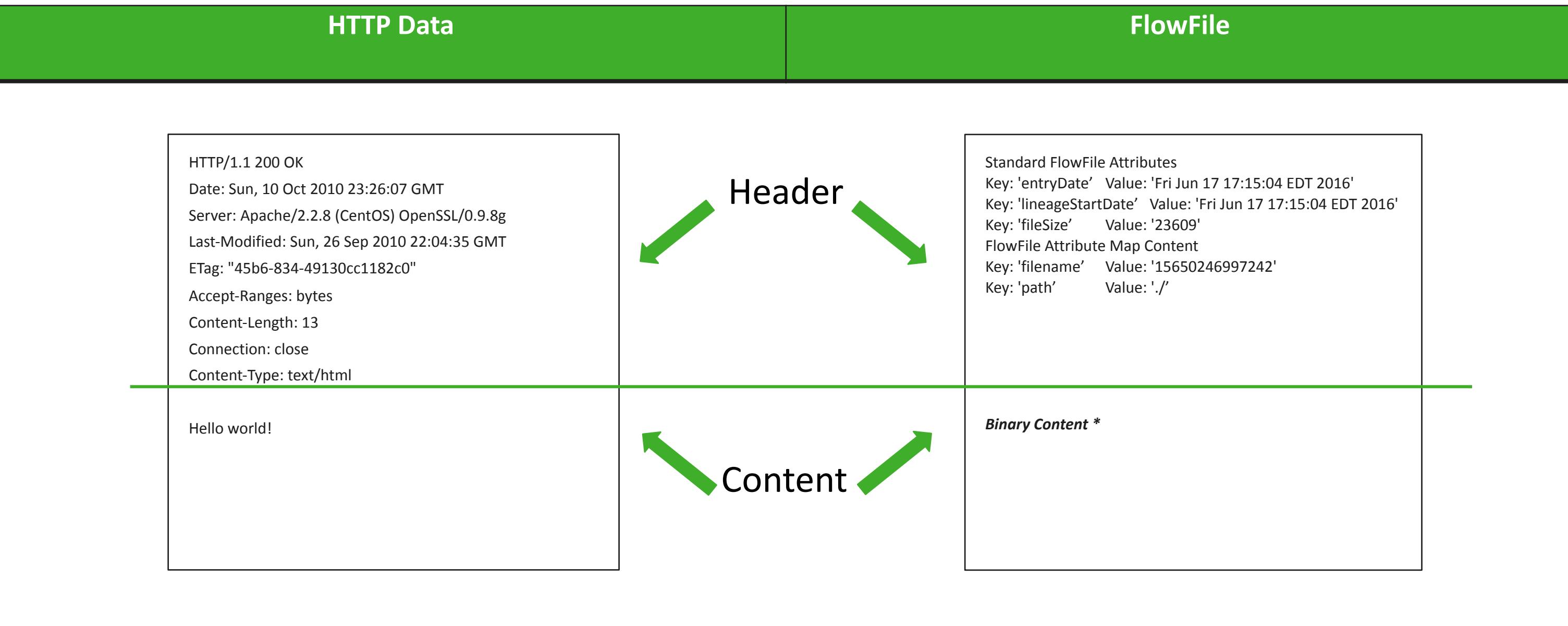
Apache NiFi

Key Features



- Guaranteed delivery
 - Data buffering
 - Backpressure
- Pressure release
- Prioritized queuing
- Flow specific QoS
 - Latency vs. throughput
 - Loss tolerance
- Data provenance
- Supports push and pull models
- Recovery/recording a rolling log of fine-grained history
- Visual command and control
- Flow templates
- Pluggable, multi-tenant security
- Designed for extension
- Clustering

Flowfiles Are Like HTTP Data



User Interface

Less of this..... more of this

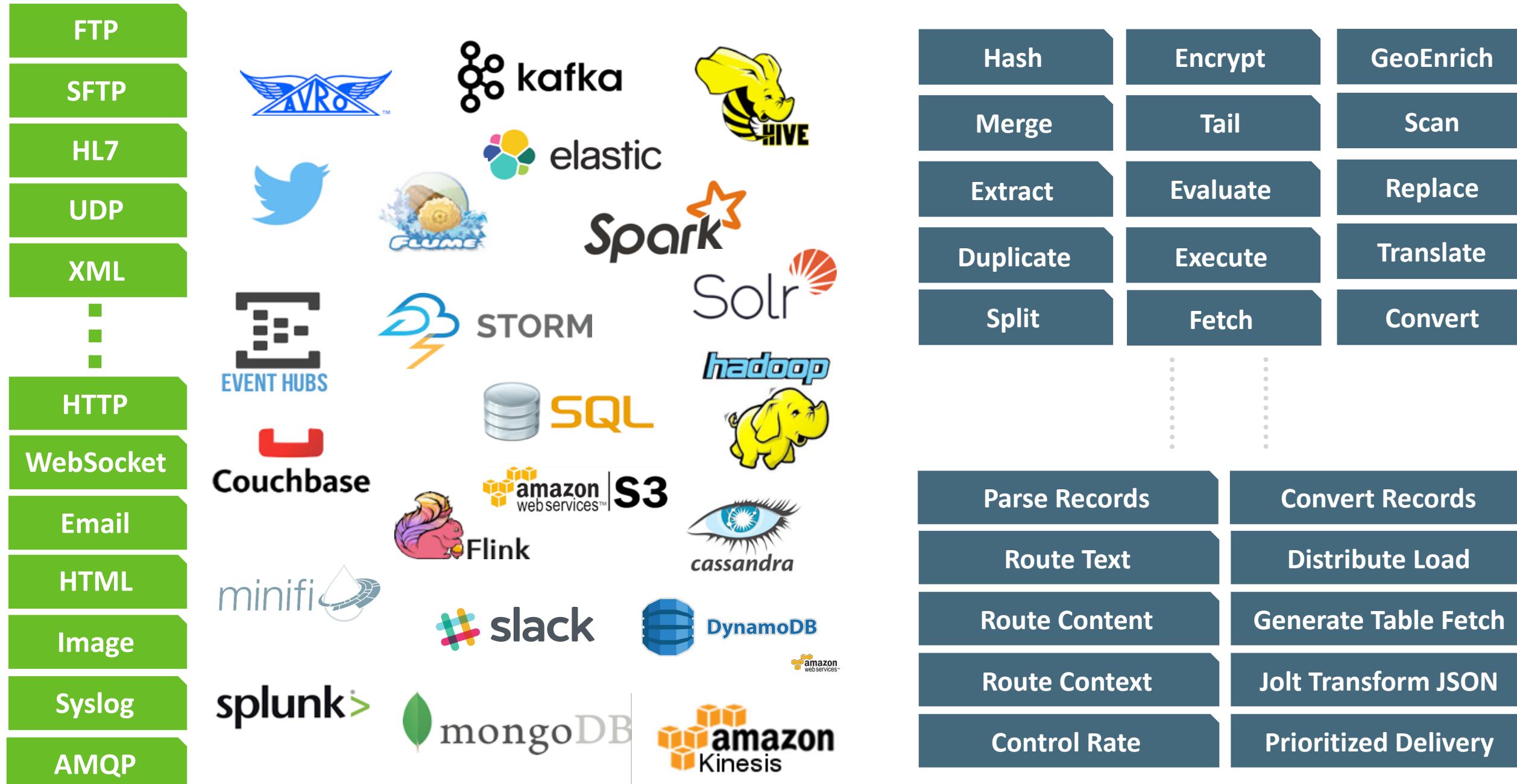
```
1. scratch/release_verification (bash)
scratch/release_verifi... #1 x vagrant/entropy (bas... #2 x release_verification/... #3
scratch/release_verification (bash)
hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto
 3s @ 17:25:51 $ t
^C
hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto NiFi (n.n.a.o) My NiFi Issues NiFi
 8s @ 17:26:00 $ hunt users.xml
./nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.0-bin/nifi-1.0/conf/users.xml
hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto
 20s @ 17:26:21 $ ll nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.0-bin/nifi-1.0/conf/
total 144
drwxr-xr-x 16 alopresto staff 544B Nov 24 19:20 ..
drwxr-xr-x 17 alopresto staff 5788 Nov 24 14:25 ..
drwxr-xr-x 38 alopresto staff 1.3K Nov 24 14:33 archive/
-rw-r--r-- 1 alopresto staff 2.5K Nov 24 19:20 authorizations.xml
-rw-r--r-- 1 alopresto staff 3.7K Nov 24 19:19 authorizers.xml
-rw-r--r-- 1 alopresto staff 2.1K Nov 23 23:36 bootstrap-notification-services.xml
-rw-r--r-- 1 alopresto staff 3.1K Nov 24 19:29 bootstrap.conf
-rw-r--r-- 1 alopresto staff 2.5K Nov 24 14:33 flow.xml.gz
-rw-r--r-- 1 alopresto staff 3.0K Nov 24 19:16 keystore.jks
-rw-r--r-- 1 alopresto staff 8.0K Nov 23 23:36 logback.xml
-rw-r--r-- 1 alopresto staff 6.2K Nov 23 23:36 login-identity-providers.xml
-rw-r--r-- 1 alopresto staff 9.0K Nov 24 19:17 nifi.properties
-rw-r--r-- 1 alopresto staff 2.3K Nov 23 23:36 state-management.xml
-rw-r--r-- 1 alopresto staff 911B Nov 24 19:16 truststore.jks
-rw-r--r-- 1 alopresto staff 226B Nov 24 19:20 users.xml
-rw-r--r-- 1 alopresto staff 1.4K Nov 23 23:36 zookeeper.properties
hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto
 21s @ 17:26:43 $ more nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.0-bin/nifi-1.0/conf/authorizations.xml
<xml version="1.0" encoding="UTF-8" standalone="yes"?>
<authorizations>
  <policies>
    <policy identifier="b2c0320b-0384-38d2-8ff7-58a26dde3897" resource="/flow" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="7e67308c-a837-3ba1-b837-4a40ff8d43fe" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="b762618f-3636-3b45-943e-6d38605e276e" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="61dfe35b-837f-37f4-9b36-b2e7bc4989c3" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="ff9f93c3-1fda-3b08-9309-088ba6abc5a9" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="bc5512df-2a78-33ba-950e-3bc1ba5f781d" resource="/restricted-components" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="70197fb3-af4d-3938-b346-498db26032fa" resource="/tenants" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="30f46aab-baa8-3b50-8b9d-d768a83e719f" resource="/tenants" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="6c25353b-fca4-3813-baa8-f879d9853e7e" resource="/policies" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="3006378a-8f9b-3a71-aa0b-44325b1723cf" resource="/policies" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="26c66358-8548-31c8-a38c-e0eb4bc3ddb4" resource="/controller" action="R">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
    <policy identifier="4470abbe-3b1e-31cd-9daa-598eae5c59ef" resource="/controller" action="W">
      <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
    </policy>
  </policies>
</authorizations>
hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto
 9s @ 17:26:53 $ 

nifi.flowfile.repository.always.sync=false
nifi.queue.swap.threshold=20000
nifi.swap.in.period=5 sec
nifi.swap.in.threads=1
nifi.swap.out.period=5 sec
nifi.swap.out.threads=4

# Content Repository
nifi.content.repository.implementation=org.apache.nifi.controller.repository.FileSystemRepository
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
 157s @ 17:40:39 $ nfk
Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf
2017-01-30 17:40:43,812 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now
  Name: success
2017-01-30 17:40:43,854 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
2017-01-30 17:40:45,867 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
 9s @ 17:40:49 $ subl conf/nifi.properties
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
 5s @ 17:40:55 $ nfg
Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf
2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Starting Apache NiFi...
2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Working Directory: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Command: /Users/alopresto/.jenv/versions/1.8/bin/java -classpath /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/jcl-over-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/jul-to-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/log4j-over-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/logback-classic-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/logback-core-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-api-1.2-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-documentation-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-framework-api-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-nar-utils-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-properties-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-runtime-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/:/lib/nifi-slf4j-apl-1.7.12.jar:/Dorg.apache.jasper.compiler.disablejsr199=true-Xms512m-Xm512m-Dsun.net.http.allowRestrictedHeaders=true-Djava.net.preferIPv4Stack=true-Djava.awt.headless=true-XX:+UseG1GC-Djava.protocol.handler.pkgs=sun.net.www.protocol -Dnifi.properties -Dnifi.bootstrap.listen.port=59772-Dapp=NiFi -Dorg.apache.nifi.bootstrap.config.log.dir=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/logs org.apache.nifi.NiFi
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
 21s @ 17:41:17 $ nfk
Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf
2017-01-30 20:54:26,971 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now
2017-01-30 20:54:27,075 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
2017-01-30 20:54:29,085 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
 11595s @ 20:54:33 $ 
```



Deeper Ecosystem Integration: 286+ Processors, 61 Controller Services

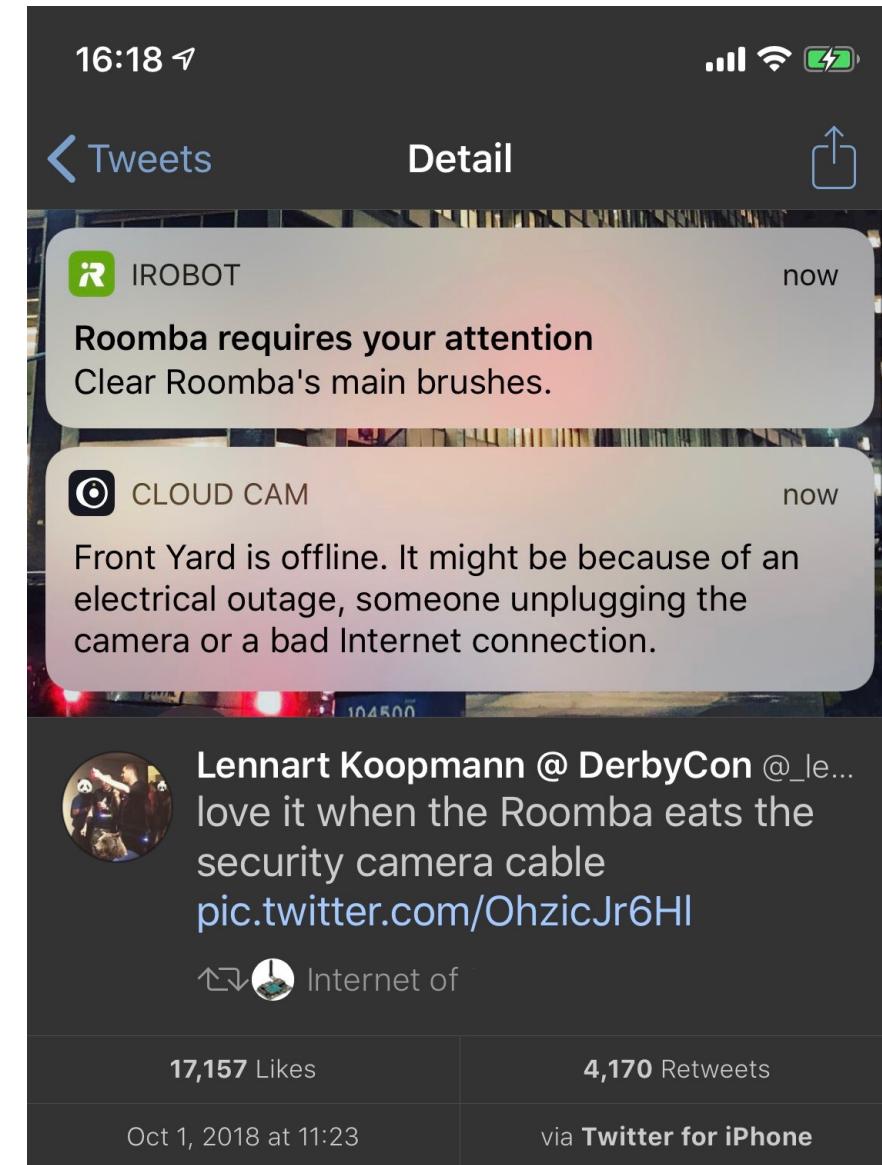




What are the IoT challenges?

IoT Challenges

- Limited computing capability
- Limited power/network
- **Restricted software library/platform availability**
- **No UI**
- Physically inaccessible
- Not frequently updated
- **Competing standards/protocols**
- Scalability
- **Privacy & Security**



@_lennart

Recent Examples

- When the Mirai attack has its own Wikipedia page, that's not good
- Hackers stole high-roller database from casino via aquarium thermometer connected to internet (04/2018)



IoTPOT: Analysing the Rise of IoT Compromises

Yin Minn Pa Pa^{†1}, Shogo Suzuki^{†1}, Katsunari Yoshioka^{†1}, Tsutomu Matsumoto^{†1},
Takahiro Kasama^{†2}, Christian Rossow^{†3}

^{†1}Graduate School of Environment and Information Sciences/Institute of Advanced Sciences
^{†1}Yokohama National University, Japan
^{†2}National Institute of Information and Communications Technology, Japan
^{†3}Institute of Advanced Sciences, Yokohama National University, Japan and
^{†3}Cluster of Excellence, MMCI, Saarland University, Germany

Google

mirai botnet

All News Videos Images Shopping More Settings Tools

About 478,000 results (0.36 seconds)

Mirai (Japanese for "the future", 未来) is malware that turns computer systems running Linux into remotely controlled "bots", that can be used as part of a **botnet** in large-scale network attacks. It primarily targets online consumer devices such as remote cameras and home routers.

Mirai (malware) - Wikipedia
[https://en.wikipedia.org/wiki/Mirai_\(malware\)](https://en.wikipedia.org/wiki/Mirai_(malware))
[Block en.wikipedia.org](#)

About this result • Feedback

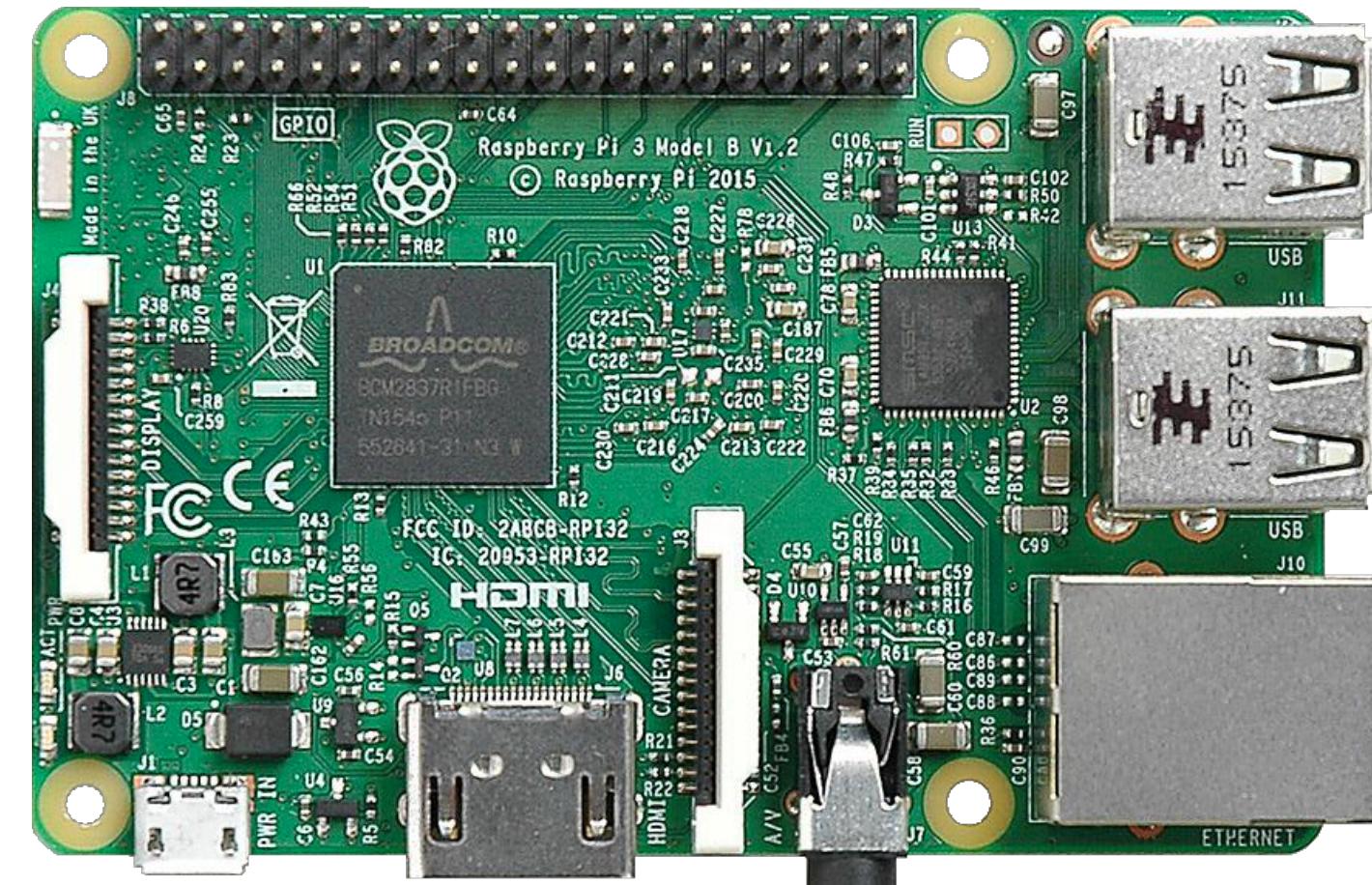
NiFi Solves Everything*

- Runs on JVM
- Provides UI for flow design & monitoring
- Security built-in
 - TLS, authentication/authorization, encrypted data
- Handles practically any format/protocol

NiFi for IoT

- NiFi supports AMQP, MQTT, UDP, TCP, HTTP(S), CEF, JMS, (S)FTP, *AWSIoT*
- With a little pruning, NiFi can run on a Raspberry Pi

```
bootstrap
jcl-over-slf4j-1.7.12.jar
jul-to-slf4j-1.7.12.jar
log4j-over-slf4j-1.7.12.jar
logback-classic-1.1.3.jar
logback-core-1.1.3.jar
nifi-api-0.6.1.jar
nifi-documentation-0.6.1.jar
nifi-framework-nar-0.6.1.nar
nifi-html-nar-0.6.1.nar
nifi-http-context-map-nar-0.6.1.nar
nifi-jetty-bundle-0.6.1.nar
nifi-kerberos-iaa-providers-nar-0.6.1.nar
nifi-ldap-iaa-providers-nar-0.6.1.nar
nifi-nar-utils-0.6.1.jar
nifi-properties-0.6.1.jar
nifi-provenance-repository-nar-0.6.1.nar
nifi-runtime-0.6.1.jar
nifi-scripting-nar-0.6.1.nar
nifi-ssl-context-service-nar-0.6.1.nar
nifi-standard-nar-0.6.1.nar
nifi-standard-services-api-nar-0.6.1.nar
nifi-update-attribute-nar-0.6.1.nar
slf4j-api-1.7.12.jar
```



So Why Do We Need A Different Solution?

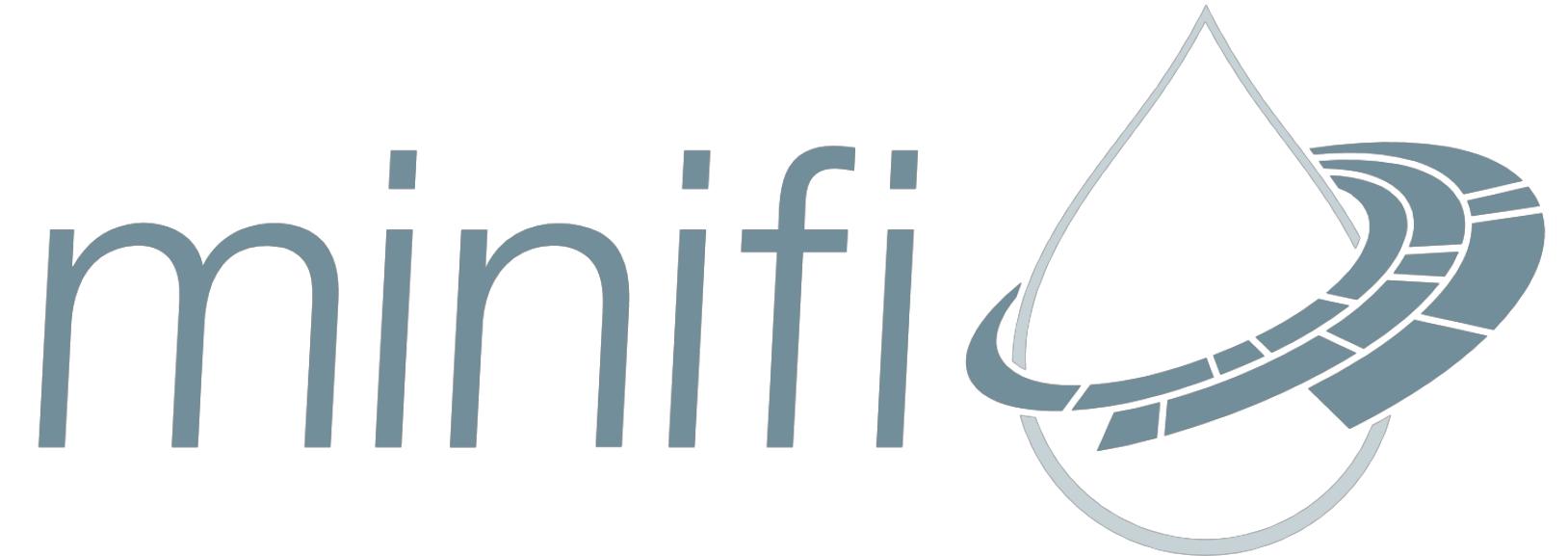
- NiFi is designed to “own the box”
- NiFi 0.7.x started up in about 10-15 minutes on RP3 (593 MB)
- NiFi 1.x started up in about 30 minutes on RP3 (760 MB)
 - 33 new processors
 - Rewrite for multi tenant authorization
 - Complete UI overhaul

```
▶hw12203:/Users/alopresto/Workspace/scratch/rp3b-demo (master) alopresto
 ━ 113s @ 17:09:05 $ ssh pi@my-raspberry-pi
 ^C
▶hw12203:/Users/alopresto/Workspace/scratch/rp3b-demo (master) alopresto
 ━ 145s @ 17:09:37 $ █
```

Enter Apache MiNiFi

Apache NiFi Subproject: MiNiFi

- Get the key parts of NiFi close to where data begins and provide bidirectional communication
- NiFi lives in the data center — give it an enterprise server or a cluster of them
- MiNiFi lives as close to where data is born and is a guest on that device or system
 - IoT
 - Connected car
 - Legacy hardware

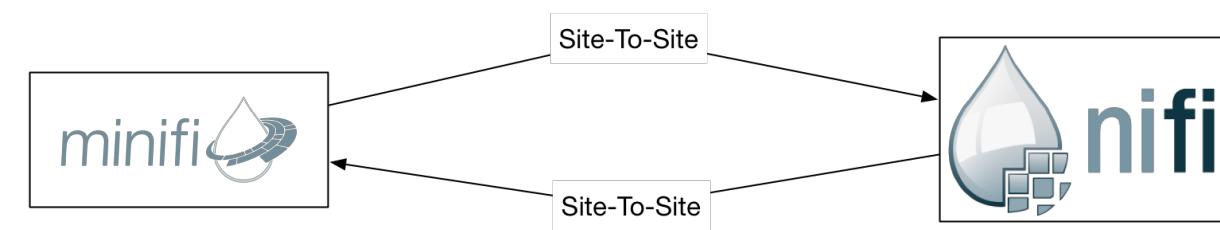
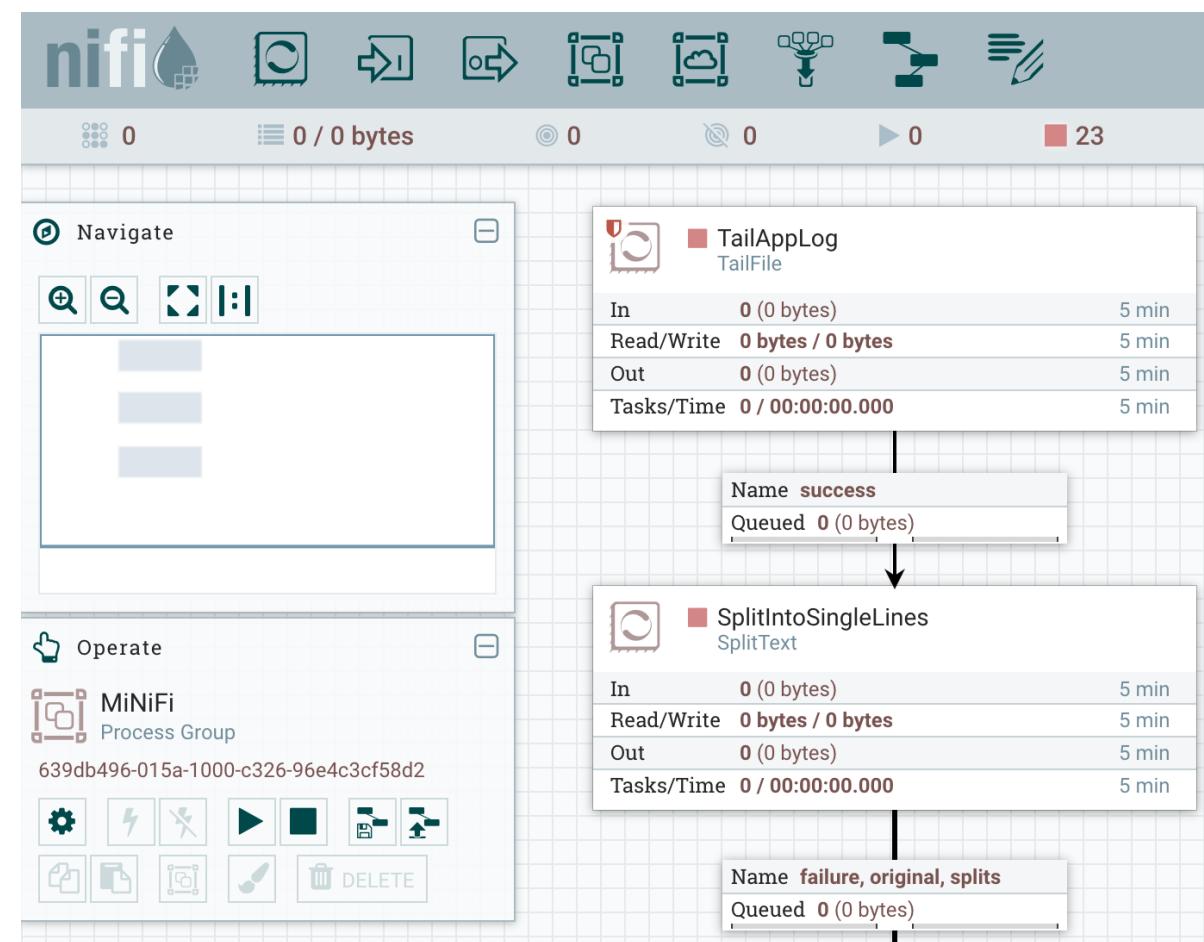


Why build MiNiFi?

- NiFi is big
 - 1.8.0 release is 1.2 GB compressed
 - Can be modified to run in restricted environments, but requires manual surgery
 - Provides UI, provenance query, etc.
 - Runs on dedicated machines/clusters — “owns the box”
- MiNiFi lives at the edge
 - No UI
 - 0.5.0 Java release is 67 MB, C++ release is 6.1 MB (**0.2.0 fits on a floppy disk**)
 - “Good guest”

Flavors of MiNiFi

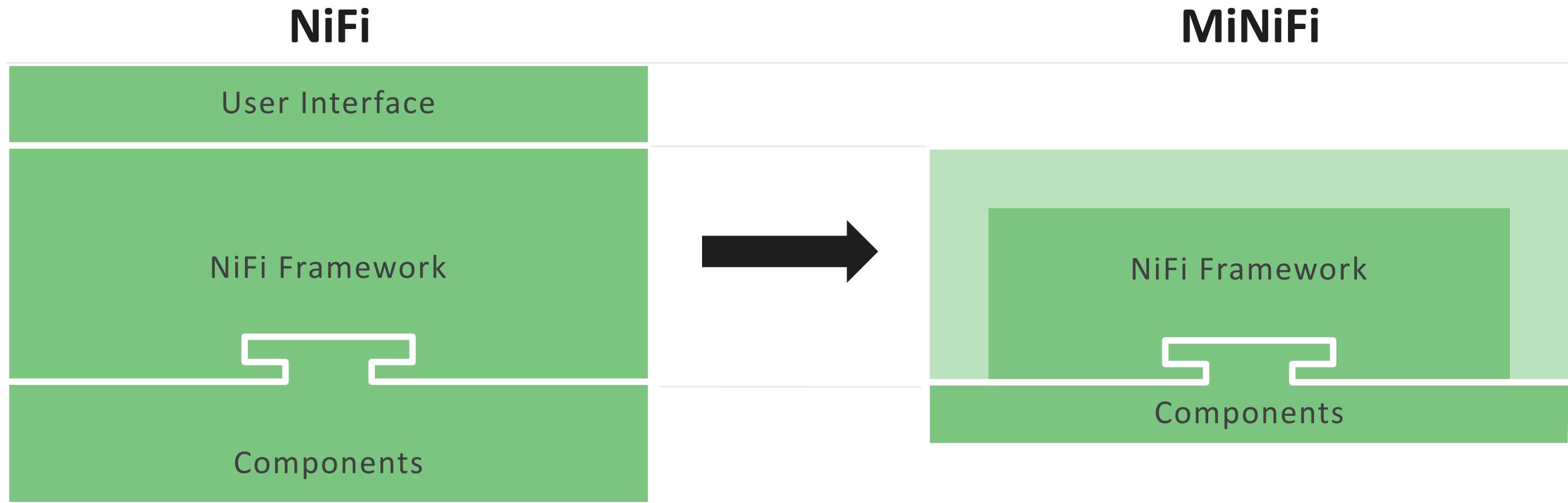
- MiNiFi Java (v0.5.0)
 - Modified version of NiFi
 - No UI
 - YAML configuration
 - Reduced processor count
 - 63+ by default, more available with additional NARs
- MiNiFi C++ (v0.5.0)
 - Written from scratch
 - 33 processors by default
 - Bi-directional site-to-site & provenance data



```
Security Properties:
  keystore: /tmp/ssl/localhost-ks.jks
  keystore type: JKS
  keystore password: localtest
  key password: localtest
  truststore: /tmp/ssl/localhost-ts.jks
  truststore type: JKS
  truststore password: localtest
  ssl protocol: TLS
  Sensitive Props:
    key:
      algorithm: PBWEWITHMD5AND256BITAES-CBC-OPENSSL
      provider: BC

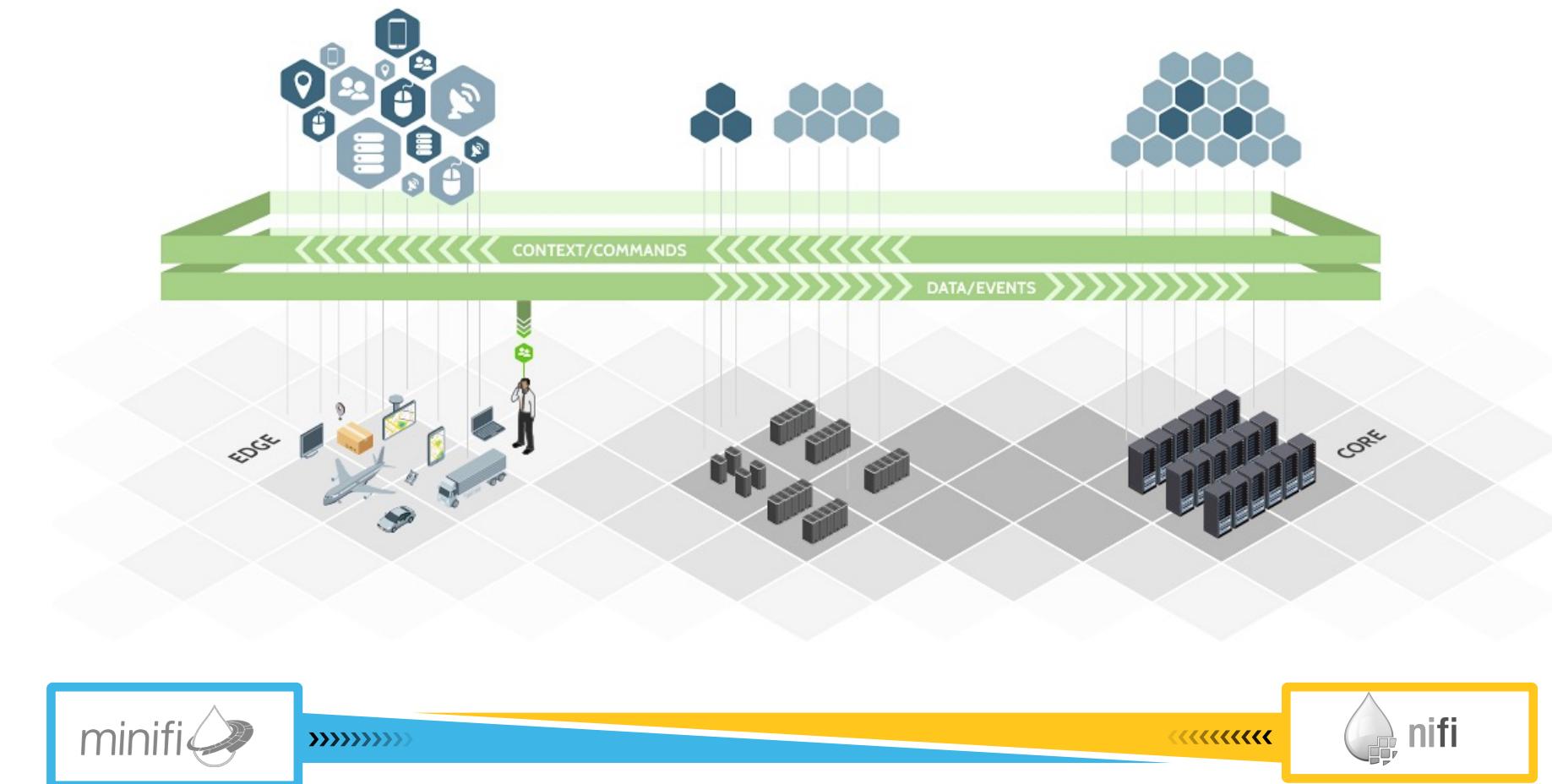
Processors:
  - name: TailAppLog
    class: org.apache.nifi.processors.standard.TailFile
    max concurrent tasks: 1
    scheduling strategy: TIMER_DRIVEN
    scheduling period: 10 sec
    penalization period: 30 sec
    yield period: 1 sec
    run duration nanos: 0
    auto-terminated relationships list:
      Properties:
        File to Tail: logs/minifi-app.log
        Rolling Filename Pattern: minifi-app*
        Initial Start Position: Beginning of File
  - name: SplitIntoSingleLines
    class: org.apache.nifi.processors.standard.SplitText
    max concurrent tasks: 1
    scheduling strategy: TIMER_DRIVEN
    scheduling period: 0 sec
    penalization period: 30 sec
    yield period: 1 sec
    run duration nanos: 0
    auto-terminated relationships list:
      - failure
      - original
    Properties:
      Line Split Count: 1
      Header Line Count: 0
      Remove Trailing Newlines: true
```

NiFi vs MiNiFi Java Processes



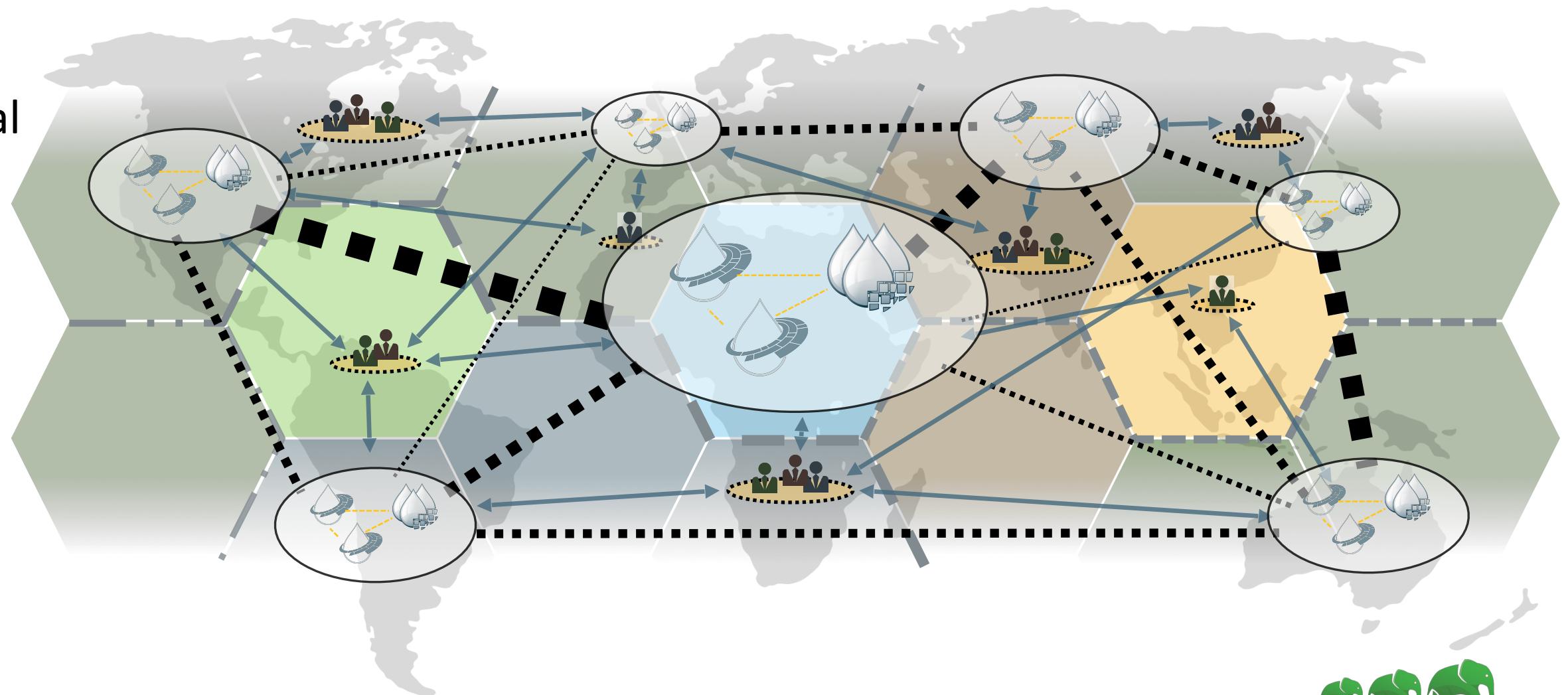
How Does MiNiFi Interact With NiFi?

- NiFi
 - Design flows
 - Aggregate data from many sources
 - Perform routing/analysis/SEP
- MiNiFi
 - Receive flows
 - Collect data
 - Send for processing



Let's Add Dimensionality

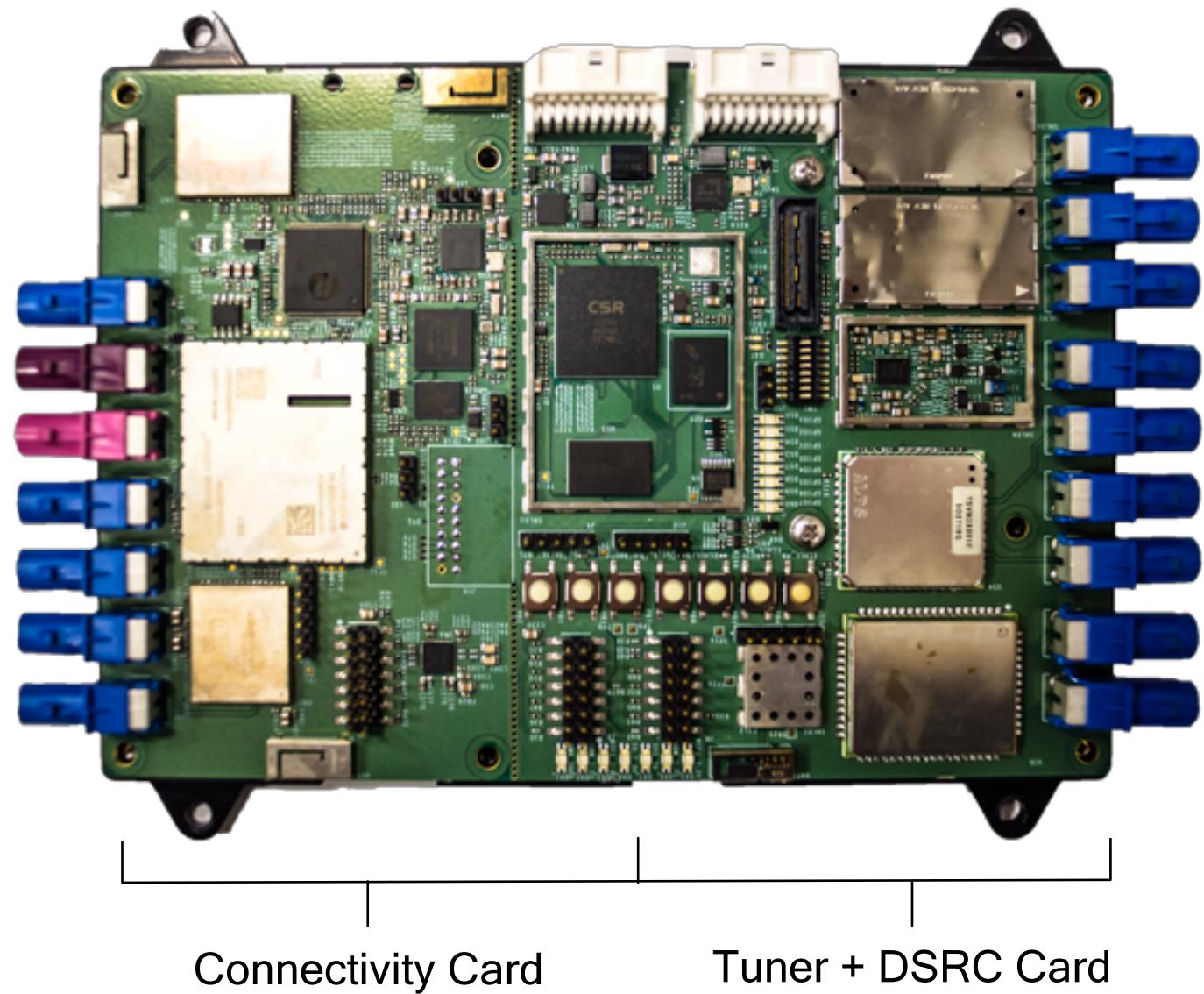
- We've been imagining EDGE to CORE as a bi-directional linear system
- Let's expand that to the real world



What does MiNiFi provide?

- Data tagging/provenance
- Governance from edge (geopolitical restrictions)
- Security (encryption, certificate-based authentication)
- Low latency (immediate reactions & decision-making)

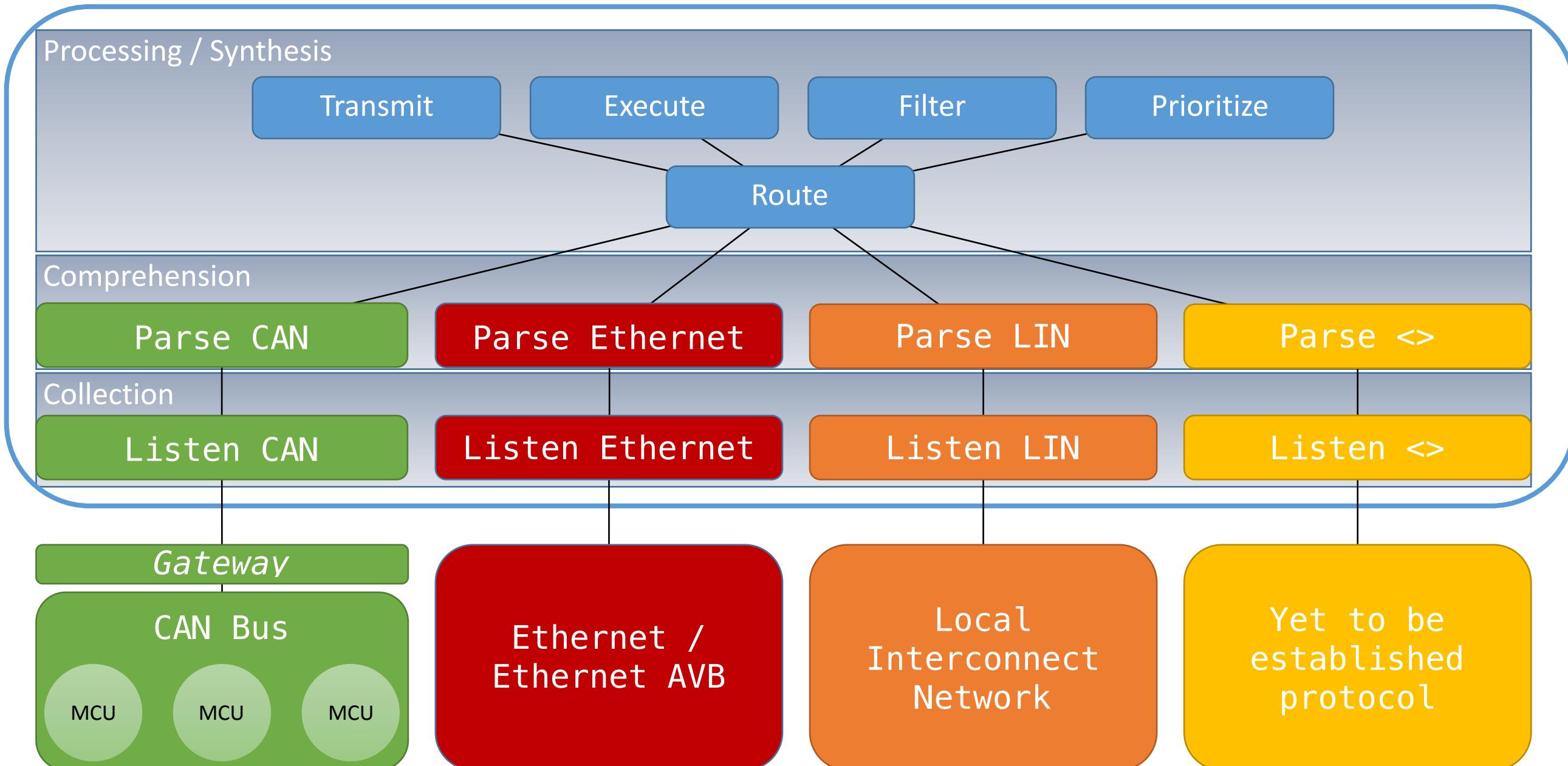
Connected Car Reference Platform Box



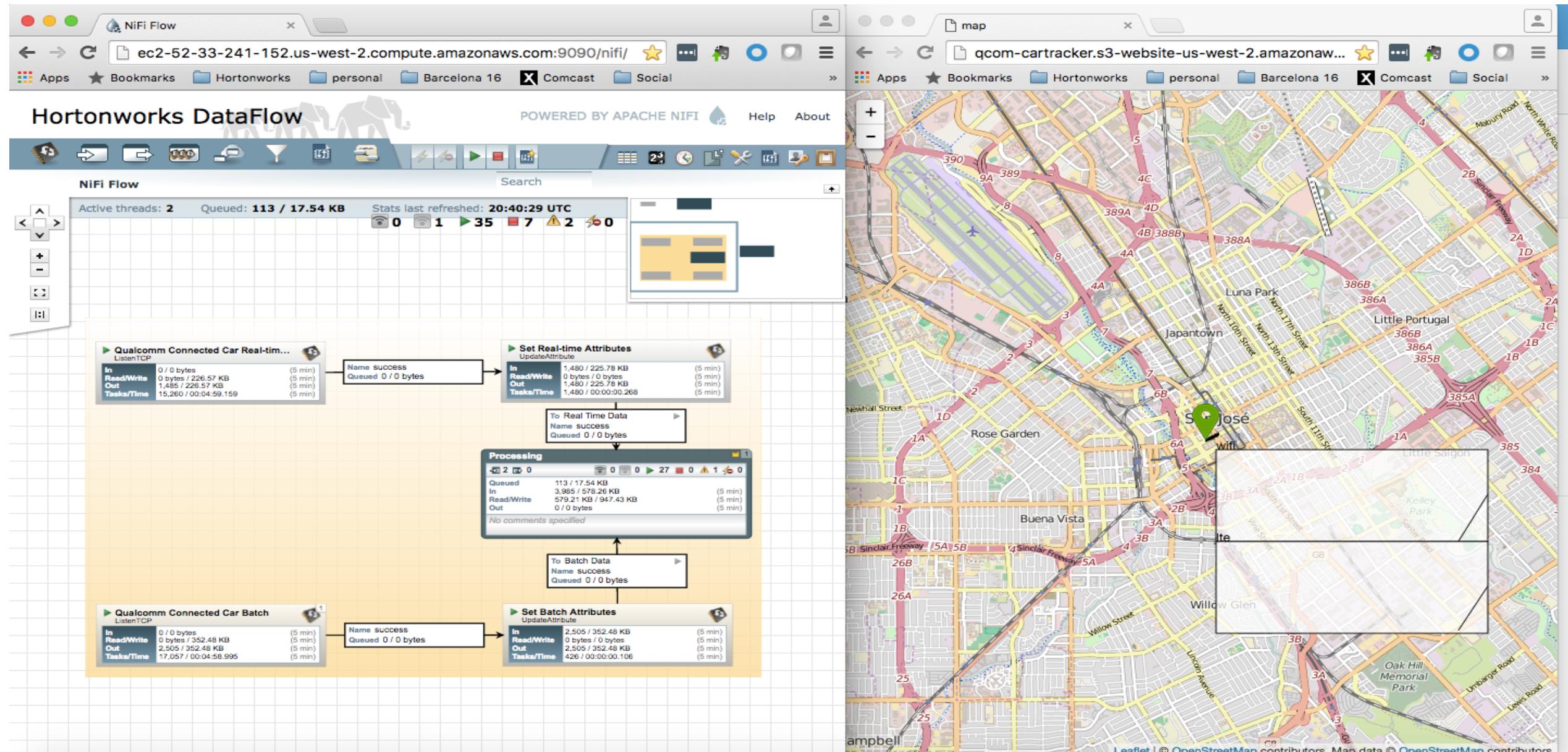
Connectivity Card

Tuner + DSRC Card

MiNiFi on a Connected Car



MiNiFi on a Connected Car



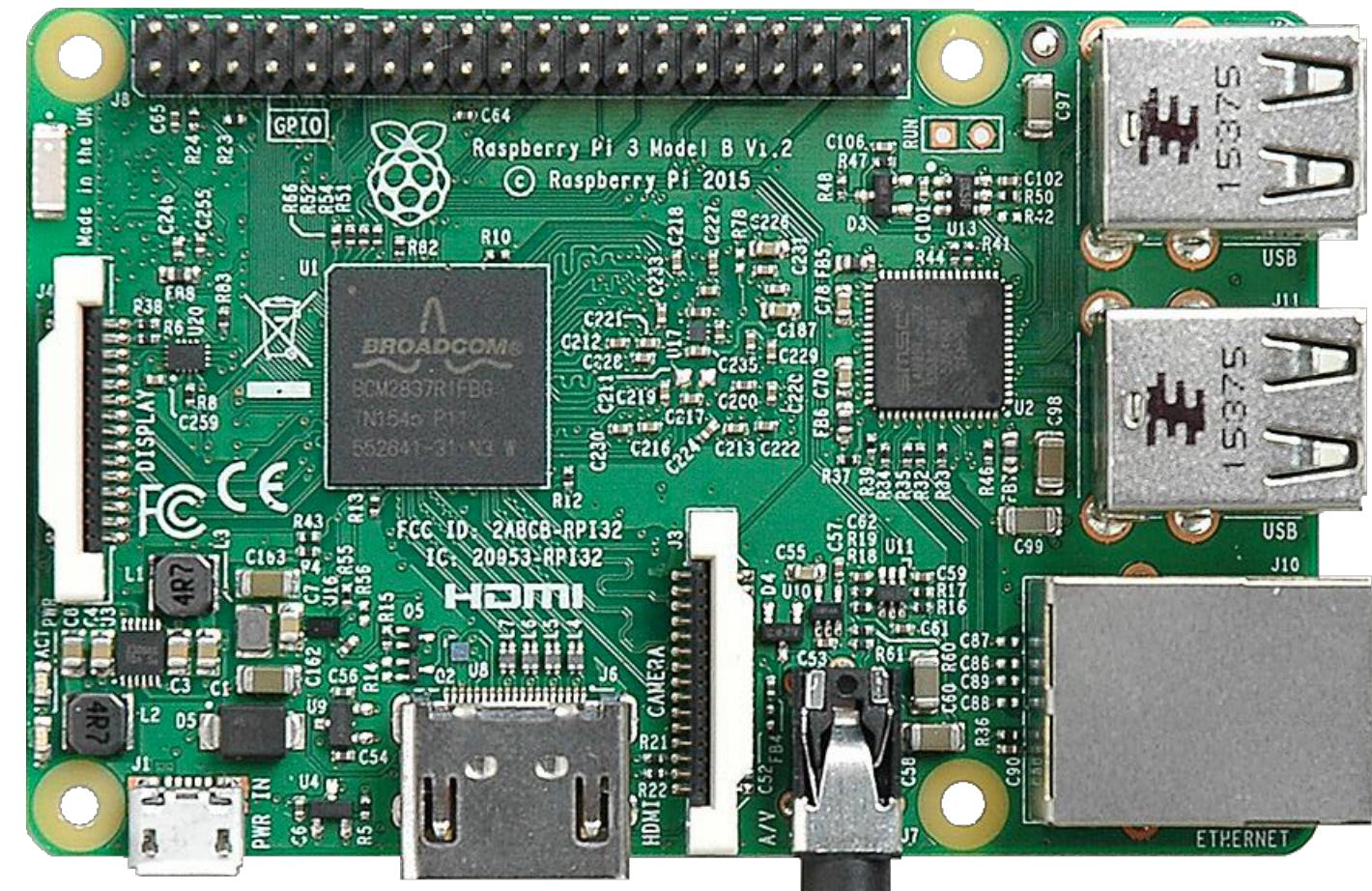
MiNiFi Exfil

- Site-to-Site
 - NiFi protocol
 - Two implementations
 - Raw socket
 - HTTP(S)
- Secured with mutual authentication TLS
 - HTTP(S), (S)FTP, JMS, Syslog, File, Email, Process

Edge Data Exploration

Scenario

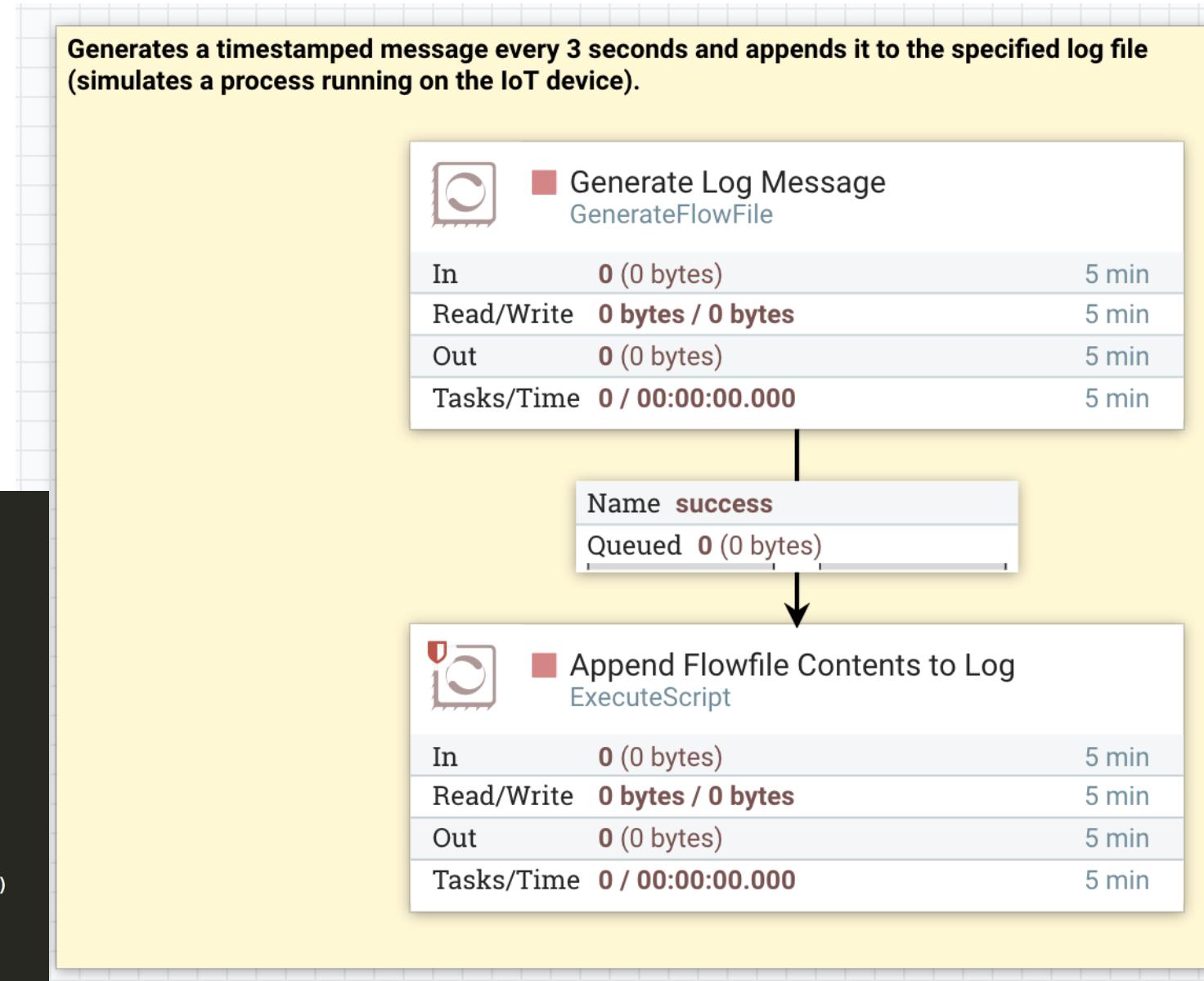
- IoT Device generating log messages
- Need to encrypt data on device
- Need to prioritize some data for unreliable network connectivity
- Transmit data to central node
- Decrypt data and analyze
- *Make determinations and modify live flow*



NiFi As Test Harness/Environment

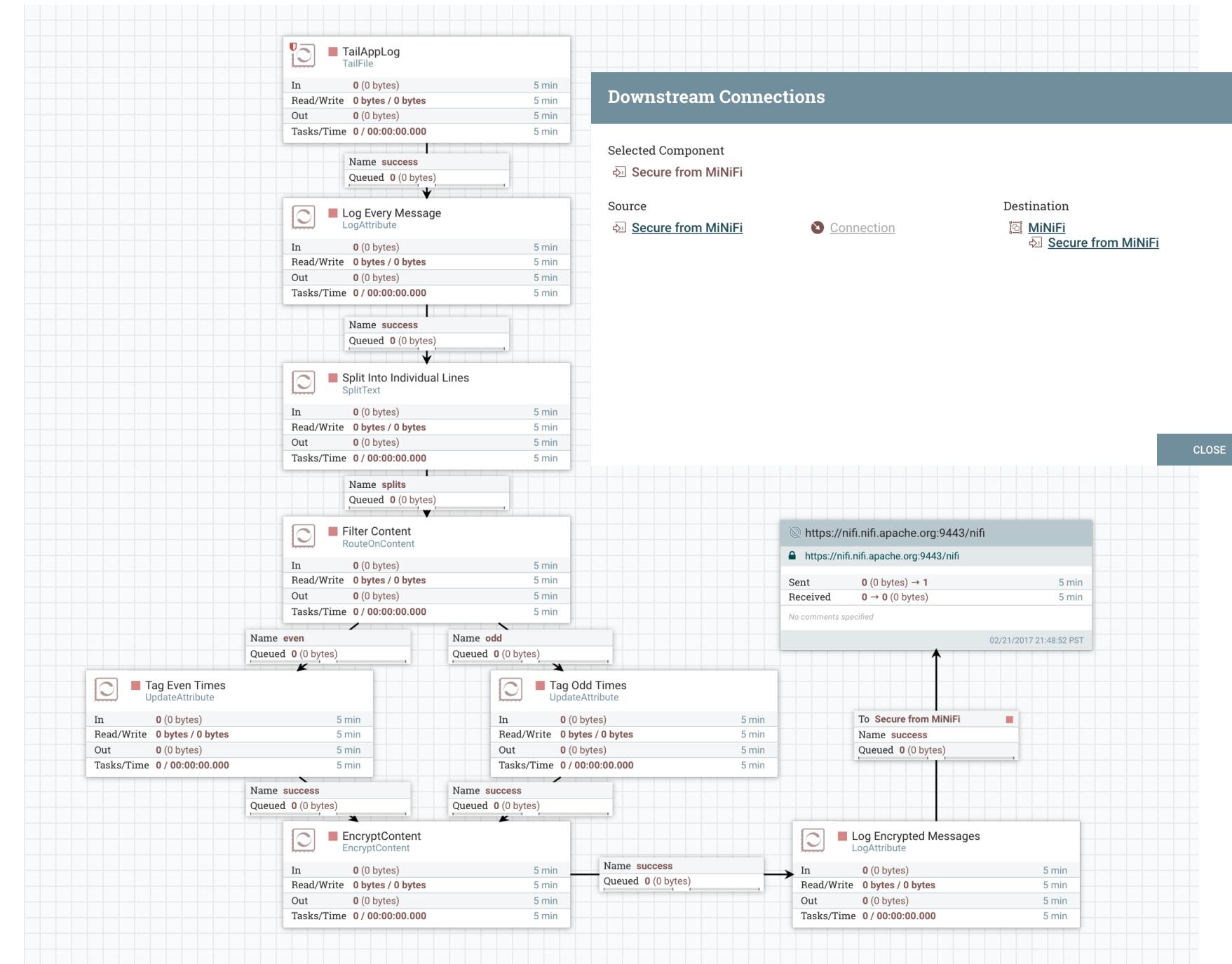
- Simulate the log generation
- Schedule is customizable
- Script can write to dynamic location

```
1 import org.apache.commons.io.IOUtils
2 import java.nio.charset.*
3
4 def flowfile = session.get()
5 if (!flowfile) return
6
7 try {
8     log.info("Received log message from generator")
9
10    def inputStream = session.read(flowfile)
11    def logMessage = IOUtils.toString(inputStream, StandardCharsets.UTF_8)
12    inputStream.close()
13
14    def logFile = new File(logFilePath as String)
15    logFile.append(System.getProperty("line.separator") + logMessage)
16
17    log.info("Wrote message to log file (total: ${logFile.text.split("\n").size()} lines)")
18
19    session.transfer(flowfile, REL_SUCCESS)
20 } catch (Exception e) {
21     log.error(e)
22     session.transfer(flowfile, REL_FAILURE)
23 }
```



Build the MiNiFi Flow

- Tails a log file
- Logs the raw contents (can be multiple lines in time window)
- Splits into individual lines
- Filters the content
 - Using parity of the timestamp
- Prioritizes
- Encrypts using AES/GCM
- Exfils to remote NiFi



Export from NiFi to MiNiFi

- Save as template from NiFi
- Run `$./bin/config.sh transform template.xml config.yml`
- MiNiFi flow ready to run*

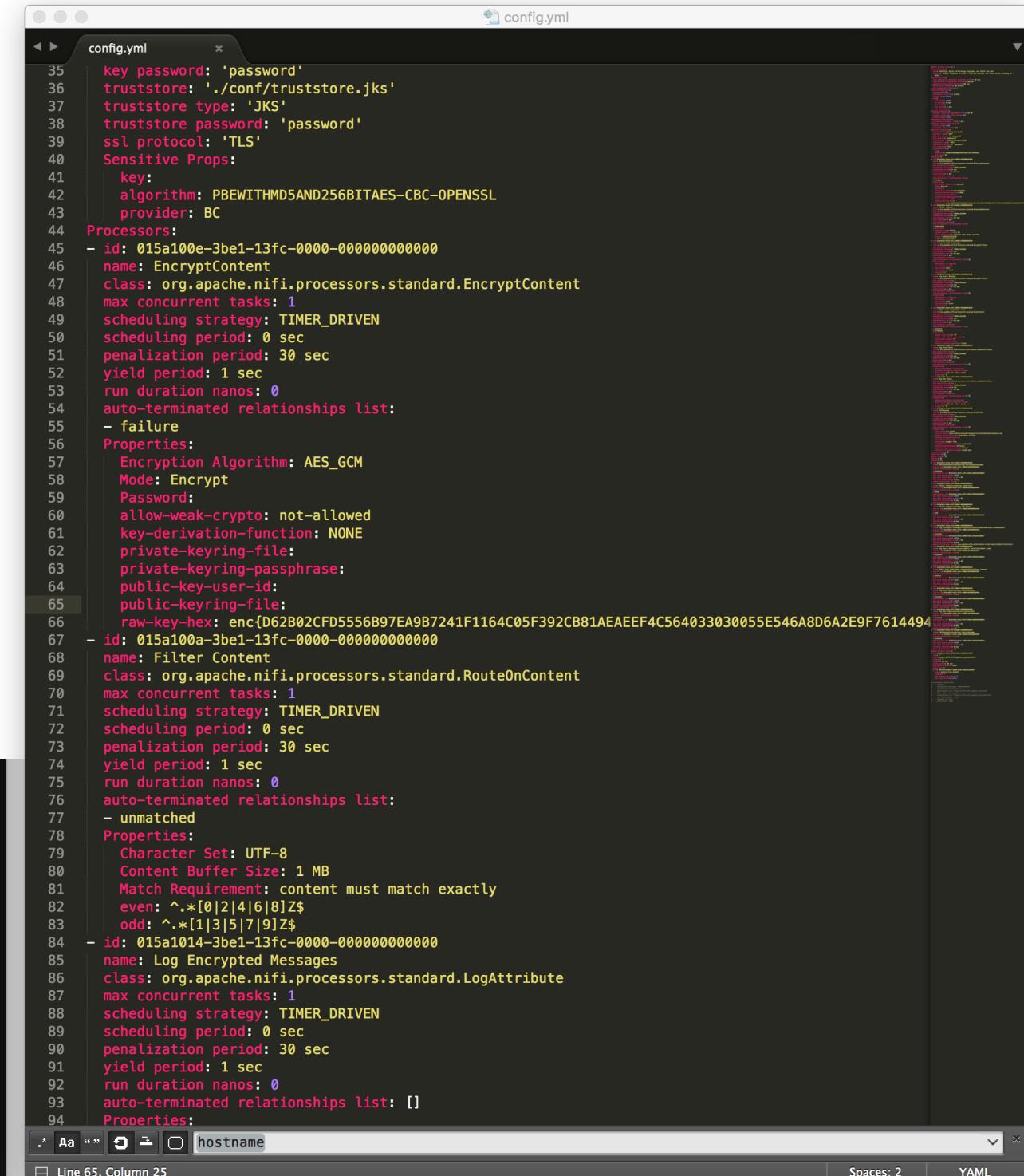
*Still need to set up TLS & encrypted properties

```
hw12203:...inifi-toolkit-assembly/target/minifi-toolkit-0.1.0-bin/minifi-toolkit-0.1.0 (master) alopresto
2s @ 21:54:55 $ ./bin/config.sh transform ~/Downloads/TailFile\,_Split\,_Prioritize\,_Encrypt\,_and_Exfil_via_S2S.xml config.yml

Java home: /Users/alopresto/.jenv/versions/1.8
MiNiFi Toolkit home: /Users/alopresto/Workspace/scratch/release_verification/minifi-java-0.1.0/minifi-0.1.0/minifi-toolkit/minifi-toolkit-assembly/target/minifi-toolkit-0.1.0-bin/minifi-toolkit-0.1.0

No validation errors found in converted configuration.

hw12203:...inifi-toolkit-assembly/target/minifi-toolkit-0.1.0-bin/minifi-toolkit-0.1.0 (master) alopresto
41s @ 21:55:37 $ ll
total 104
drwxr-xr-x  8 alopresto  staff  272B Feb 21 21:55 .
drwxr-xr-x  3 alopresto  staff  102B Dec  1 11:13 ..
-rw-r--r--  1 alopresto  staff  18K Nov 30 22:55 LICENSE
-rw-r--r--  1 alopresto  staff  11K Nov 30 22:55 NOTICE
-rw-r--r--  1 alopresto  staff  4.6K Nov 30 22:55 README
drwxr-xr-x  4 alopresto  staff  136B Dec  1 11:13 bin/
-rw-r--r--  1 alopresto  staff  8.9K Feb 21 21:55 config.yml
drwxr-xr-x 83 alopresto  staff  2.8K Dec  1 11:13 lib/
hw12203:...inifi-toolkit-assembly/target/minifi-toolkit-0.1.0-bin/minifi-toolkit-0.1.0 (master) alopresto
10s @ 21:55:47 $
```



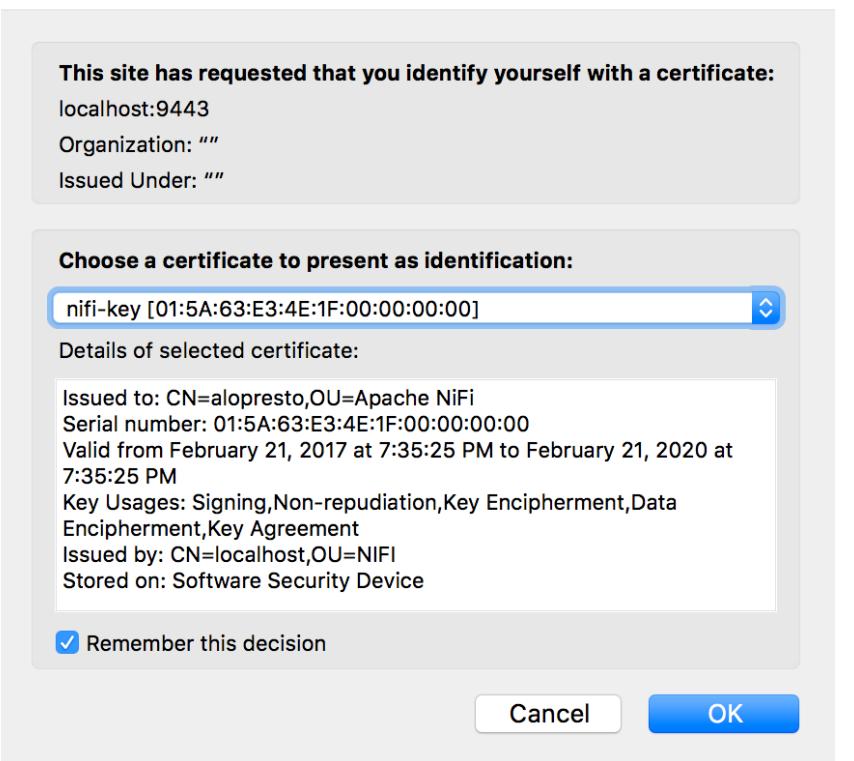
```
config.yml
35 key password: 'password'
36 truststore: './conf/truststore.jks'
37 truststore type: 'JKS'
38 truststore password: 'password'
39 ssl protocol: 'TLS'
40 Sensitive Props:
41   key:
42     algorithm: PBEWITHMD5AND256BITAES-CBC-OPENSSL
43     provider: BC
44 Processors:
45 - id: 015a100e-3be1-13fc-0000-000000000000
46   name: EncryptContent
47   class: org.apache.nifi.processors.standard.EncryptContent
48   max concurrent tasks: 1
49   scheduling strategy: TIMER_DRIVEN
50   scheduling period: 0 sec
51   penalization period: 30 sec
52   yield period: 1 sec
53   run duration nanos: 0
54   auto-terminated relationships list:
55   - failure
56   Properties:
57     Encryption Algorithm: AES_GCM
58     Mode: Encrypt
59     Password:
60     allow-weak-crypto: not-allowed
61     key-derivation-function: NONE
62     private-keyring-file:
63     private-keyring-passphrase:
64     public-key-user-id:
65     public-keyring-file:
66     raw-key-hex: enc{D62B02CFD5556897EA9B7241F1164C05F392CB81AEAEF4C564033030055E546A8D6A2E9F7614494
67 - id: 015a100a-3be1-13fc-0000-000000000000
68   name: Filter Content
69   class: org.apache.nifi.processors.standard.RouteOnContent
70   max concurrent tasks: 1
71   scheduling strategy: TIMER_DRIVEN
72   scheduling period: 0 sec
73   penalization period: 30 sec
74   yield period: 1 sec
75   run duration nanos: 0
76   auto-terminated relationships list:
77   - unmatched
78   Properties:
79     Character Set: UTF-8
80     Content Buffer Size: 1 MB
81     Match Requirement: content must match exactly
82     even: ^.*[0|2|4|6|8]Z$
83     odd: ^.*[1|3|5|7|9]Z$
84 - id: 015a1014-3be1-13fc-0000-000000000000
85   name: Log Encrypted Messages
86   class: org.apache.nifi.processors.standard.LogAttribute
87   max concurrent tasks: 1
88   scheduling strategy: TIMER_DRIVEN
89   scheduling period: 0 sec
90   penalization period: 30 sec
91   yield period: 1 sec
92   run duration nanos: 0
93   auto-terminated relationships list: []
94   Properties:
```

Setting Up Crypto

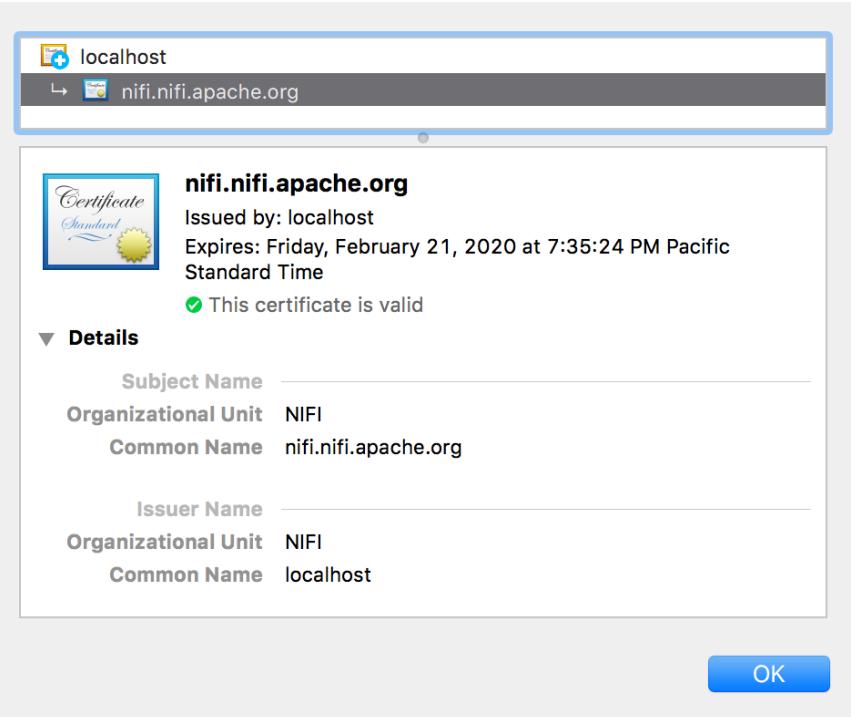
- NiFi TLS Toolkit makes certificates & keystores simple (and secure)
- Copy encrypted property value from flow.xml.gz to config.yml (*flow repo*)

Secure Connection

The connection to this site is encrypted and authenticated using a strong protocol (TLS 1.2), a strong key exchange (ECDHE_RSA with P-256), and a strong cipher (AES_256_GCM).



```
hw12203: ...assembly/target/nifi-toolkit-1.2.0-SNAPSHOT-bin/nifi-toolkit-1.2.0-SNAPSHOT (NIFI-3486-RC1) al
opresto
173s @ 19:34:24 $ ./bin/tls-toolkit.sh standalone -n 'nifi.nifi.apache.org' -C 'CN=alopresto, OU=Apache
NiFi' -P password -S password -B password -f ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin
/nifi-1.2.0-SNAPSHOT/conf/nifi.properties -o ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/
nifi-1.2.0-SNAPSHOT/conf/
2017/02/21 19:35:23 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandaloneCommandLine: U
sing ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/nifi.properties
as template.
2017/02/21 19:35:23 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Running stan
dalone certificate generation with output directory ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSH
OT-bin/nifi-1.2.0-SNAPSHOT/conf
2017/02/21 19:35:24 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Generated ne
w CA certificate ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/n
ifi-cert.pem and key ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/n
ifi-key.key
2017/02/21 19:35:24 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Writing new
ssl configuration to ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf
/nifi.nifi.apache.org
2017/02/21 19:35:24 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Successfully
generated TLS configuration for nifi.nifi.apache.org 1 in ../../../../../../nifi-assembly/target/nifi-1.2.0
-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/nifi.nifi.apache.org
2017/02/21 19:35:24 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Generating n
ew client certificate ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/con
f/CN=alopresto_OU=Apache_NiFi.p12
2017/02/21 19:35:25 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Successfully
generated client certificate ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAP
SHOT/conf/CN=alopresto_OU=Apache_NiFi.p12
2017/02/21 19:35:25 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: tls-toolkit
standalone completed successfully
hw12203: ...assembly/target/nifi-toolkit-1.2.0-SNAPSHOT-bin/nifi-toolkit-1.2.0-SNAPSHOT (NIFI-3486-RC1) al
opresto
```



If We Really Have TLS, Why Encrypt?

- All data transmitted over TLS is encrypted
- On NiFi, automatically decrypted
- Attributes visible
- Content still encrypted because of EncryptContent processor
- Can serve as secure route for follow-on systems

Configure Processor

SETTINGS SCHEDULING PROPERTIES COMMENTS

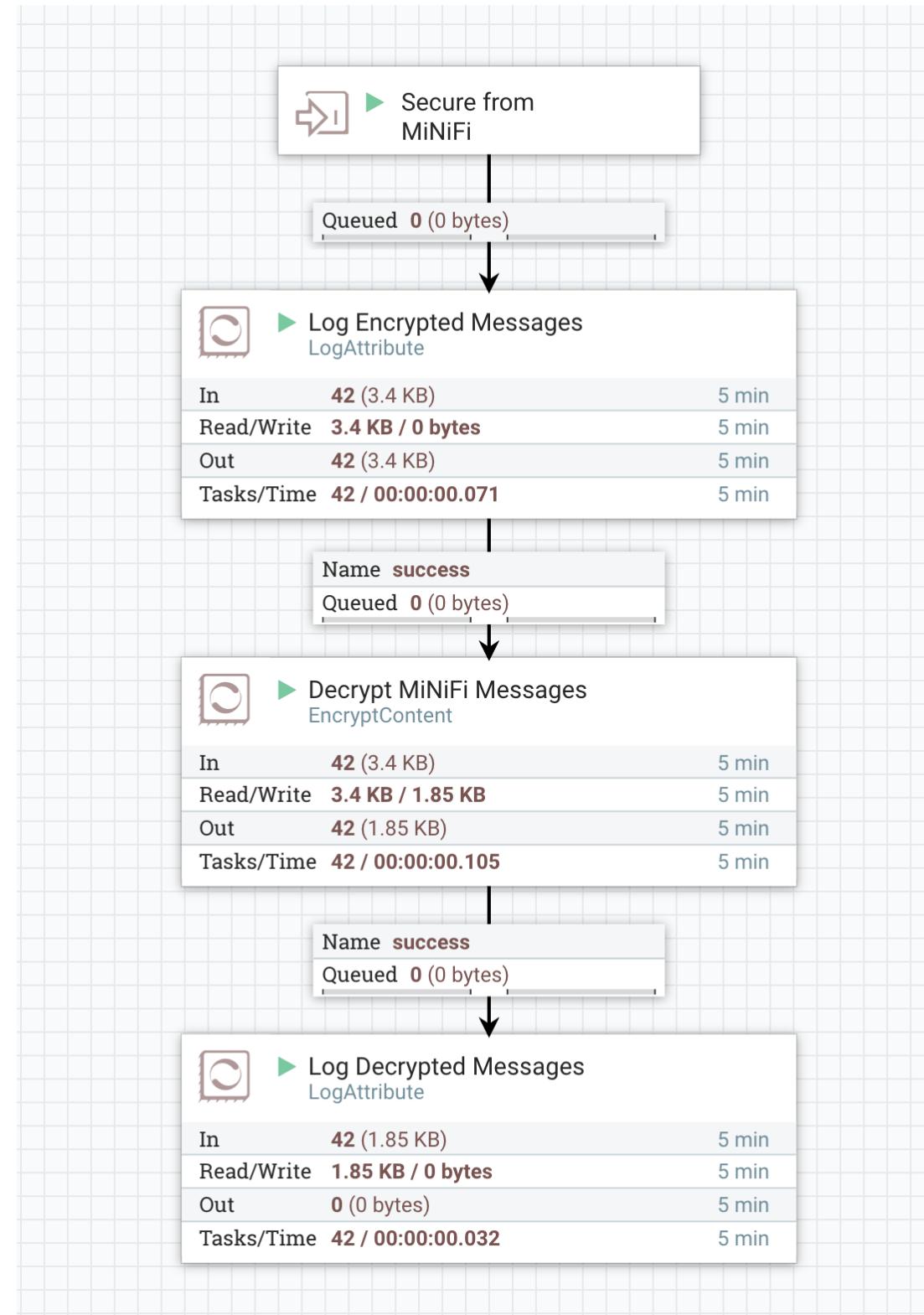
Required field

Property	Value
Mode	Encrypt
Key Derivation Function	None
Encryption Algorithm	AES_GCM
Allow insecure cryptographic modes	Not Allowed
Password	No value set
Raw Key (hexadecimal)	Sensitive value set
Public Keyring File	In keyed encryption, this is the raw key, encoded in hexadecimal
Public Key User Id	Supports expression language: false
Private Keyring File	History: • ***** - 02/21/2017 20:30:46 PST (CN=alopresto, OU=Apache NiFi)
Private Keyring Passphrase	No value set

CANCEL APPLY

Process Data In NiFi

- Receive the data over S2S
- Log the incoming messages
- Decrypt content
- Log again



Does It Work?

1	駒嶽△稜口擊? ? 樟櫻瞰岱口留口惹, 莱駿邝銅既口口口卅豎韻塔, 講擘瓊瑩齡口贈璣髮僑
View as:	hex
0x00000000	CB 54 29 E4 BA 64 3B 1B 55 85 84 70 22 9B 77 03 .T)...d;.U...p".w.
0x00000010	4E 69 46 69 49 56 2C 19 3C B1 DA AF 45 65 F9 A9 NiFiIV,.<...Ee..
0x00000020	E8 D6 F7 B1 53 79 A1 11 81 83 96 FA E6 48 20 B5Sy.....H .
0x00000030	CD AE 71 66 03 C7 20 11 C0 DD 65 9E 4A 54 59 4B ..qf.. ...e.JTYK
0x00000040	4E 6A CD D7 5D 85 7D 68 C6 AF EF 70 F8 23 E0 D6 Nj...].}h...p.#...
0x00000050	3D 95 8E =..

NiFi

Not Secure https://nifi.nifi.apache.org:9443/nifi/

Bookmarks NiFi (localhost) NiFi 9443 NiFi (n.n.a.o) My NiFi Issues ASF Git NiFi NiFi Docs NiFi (Onyara proces... NiFi Private Mail dev@nifi.apache.org users@nifi.apache.org NiFi Feature Propos... Hortonworks

NiFi Data Provenance

Displaying 213 of 213
Oldest event available: 02/21/2017 22:14:43 PST

Filter	by component name
Date/Time	Type
02/21/2017 22:22:05.079 PST	DROP
02/21/2017 22:22:05.078 PST	DROP
02/21/2017 22:22:05.077 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.076 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.076 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.075 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.074 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.073 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.073 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.072 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.071 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.070 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.069 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.069 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.068 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.068 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.067 PST	DROP
02/21/2017 22:22:05.067 PST	DROP
02/21/2017 22:22:05.067 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.066 PST	DROP
02/21/2017 22:22:05.066 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.066 PST	DROP
02/21/2017 22:22:05.065 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.065 PST	DROP
02/21/2017 22:22:05.065 PST	DROP
02/21/2017 22:22:05.064 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.064 PST	DROP

Provenance Event

DETAILS	ATTRIBUTES	CONTENT
Input Claim Container default Section 1 Identifier 1487744083443-1 Offset 4259 Size 83 bytes	Output Claim Container default Section 1 Identifier 1487744083443-1 Offset 6681 Size 45 bytes	Download View
Replay Connection Id 015a1002-3be1-13fc-98e3-f54ee6183280		
OK		

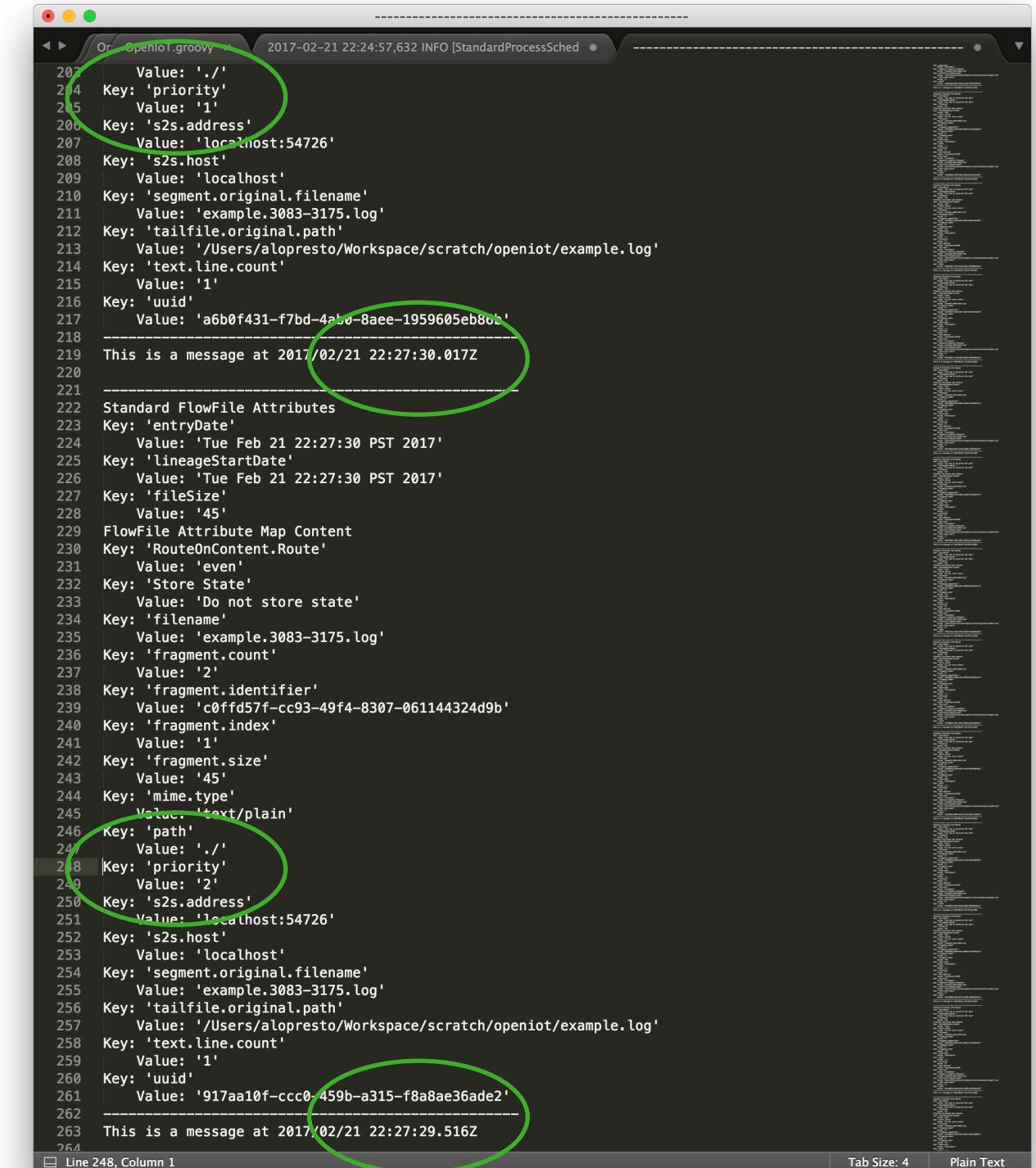
Last updated: 22:22:49 PST

NiFi Flow » MINIFI

TailFile_Split_Priorit...xml

Prioritization?

- Increase the write frequency
- Check that newer records (within tail window) with higher priority arrive first



```
207     Value: './'
208     Key: 'priority'
209     Value: '1'
210     Key: 's2s.address'
211     Value: 'localhost:54726'
212     Key: 's2s.host'
213     Value: 'localhost'
214     Key: 'segment.original.filename'
215     Value: 'example.3083-3175.log'
216     Key: 'tailfile.original.path'
217     Value: '/Users/alopresto/Workspace/scratch/openiot/example.log'
218     Key: 'text.line.count'
219     Value: '1'
220     Key: 'uuid'
221     Value: 'a6b0f431-f7bd-4a50-8aee-1959605eb80b'
222
223 This is a message at 2017/02/21 22:27:30.017Z
224
225 Standard FlowFile Attributes
226 Key: 'entryDate'
227     Value: 'Tue Feb 21 22:27:30 PST 2017'
228 Key: 'lineageStartDate'
229     Value: 'Tue Feb 21 22:27:30 PST 2017'
230 Key: 'fileSize'
231     Value: '45'
232 FlowFile Attribute Map Content
233 Key: 'RouteOnContent.Route'
234     Value: 'even'
235 Key: 'Store State'
236     Value: 'Do not store state'
237 Key: 'filename'
238     Value: 'example.3083-3175.log'
239 Key: 'fragment.count'
240     Value: '2'
241 Key: 'fragment.identifier'
242     Value: 'c0ffd57f-cc93-49f4-8307-061144324d9b'
243 Key: 'fragment.index'
244     Value: '1'
245 Key: 'fragment.size'
246     Value: '45'
247 Key: 'mime.type'
248     Value: 'text/plain'
249 Key: 'path'
250     Value: './'
251     Key: 'priority'
252     Value: '2'
253     Key: 's2s.address'
254     Value: 'localhost:54726'
255     Key: 's2s.host'
256     Value: 'localhost'
257     Key: 'segment.original.filename'
258     Value: 'example.3083-3175.log'
259     Key: 'tailfile.original.path'
260     Value: '/Users/alopresto/Workspace/scratch/openiot/example.log'
261     Key: 'text.line.count'
262     Value: '1'
263     Key: 'uuid'
264     Value: '917aa10f-ccc0-459b-a315-f8a8ae36ade2'
265
266 This is a message at 2017/02/21 22:27:29.516Z
267
```

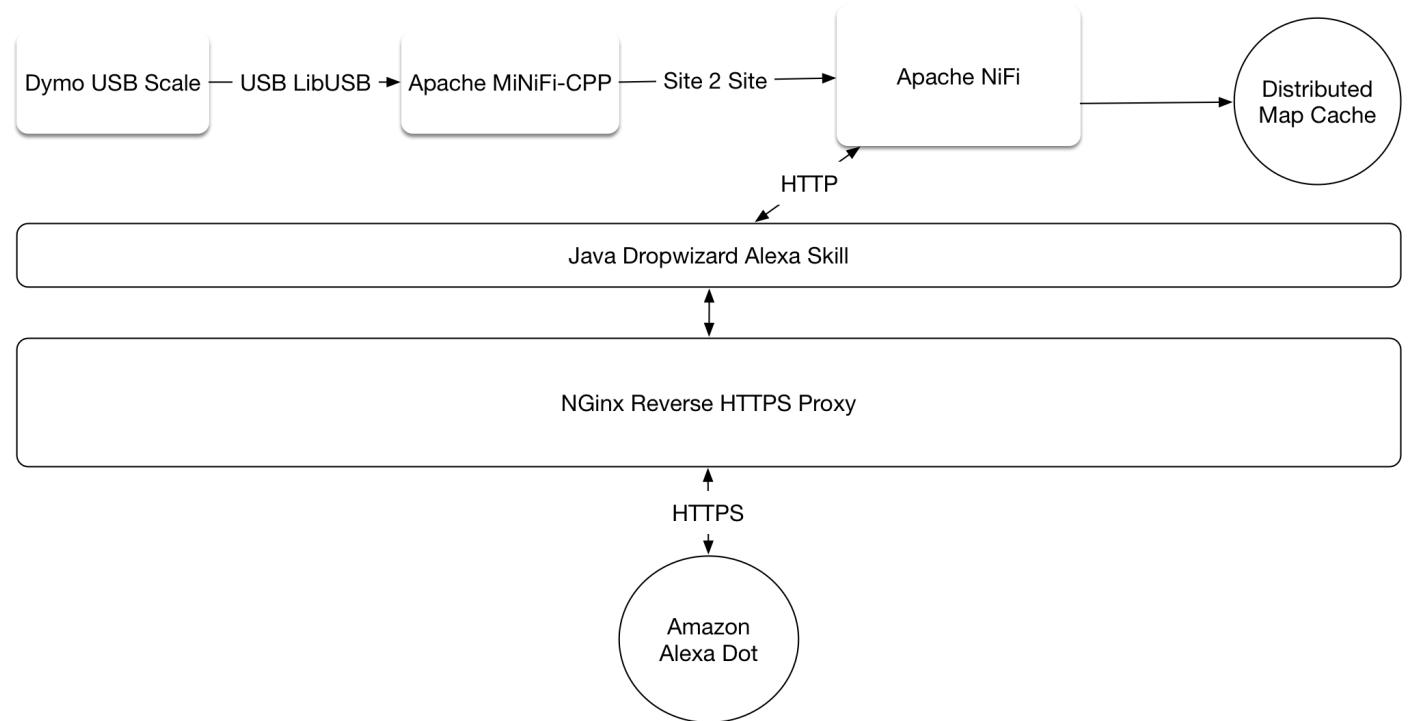
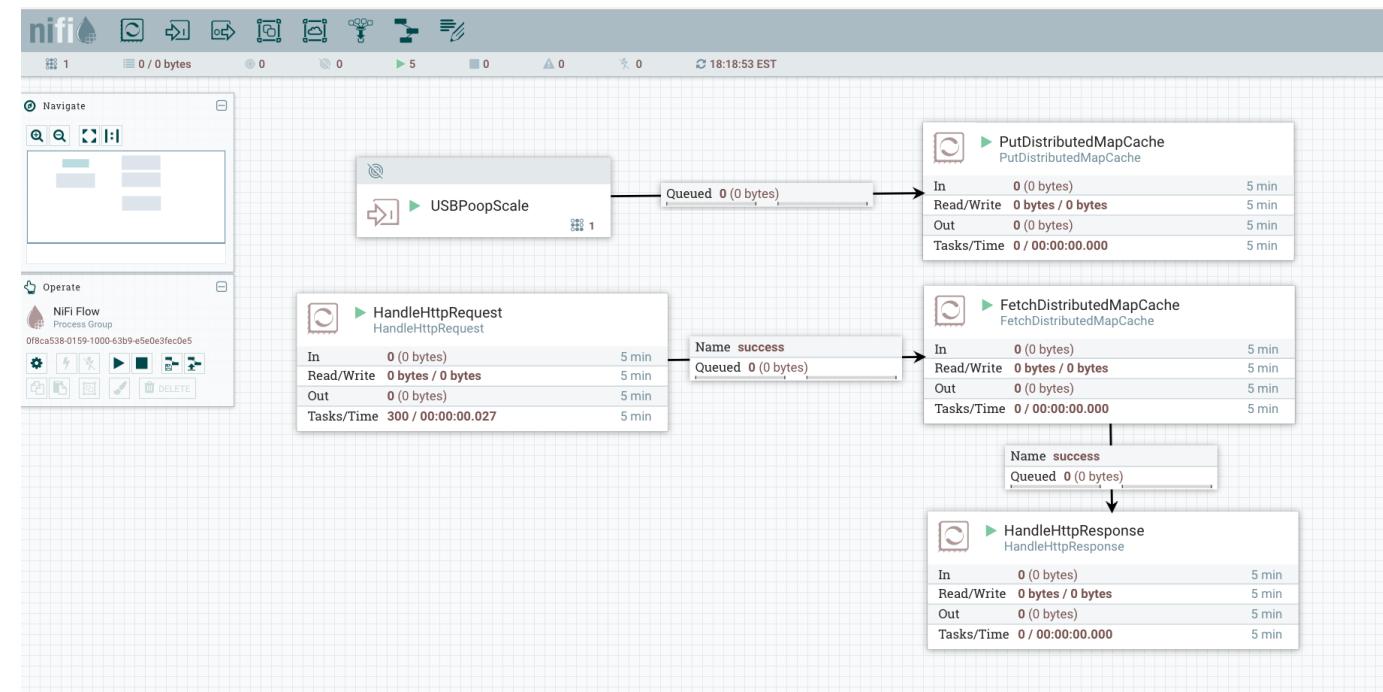
Next Steps

- Window Aggregator
 - If >60% odd in window, switch prioritization
- Encrypt with different keys for different tags & send to different follow-on systems
- Spotty network? Tell MiNiFi to cache low priority and send in batches
- MiNiFi rollover & pruning of monitored log
- Exfil MiNiFi provenance data to NiFi

Community

Community Example

- Jeremy Dyer
- Alexa + MiNiFi + Dyer 2.0





What's Next?

New Announcements

Introducing Apache NiFi Registry

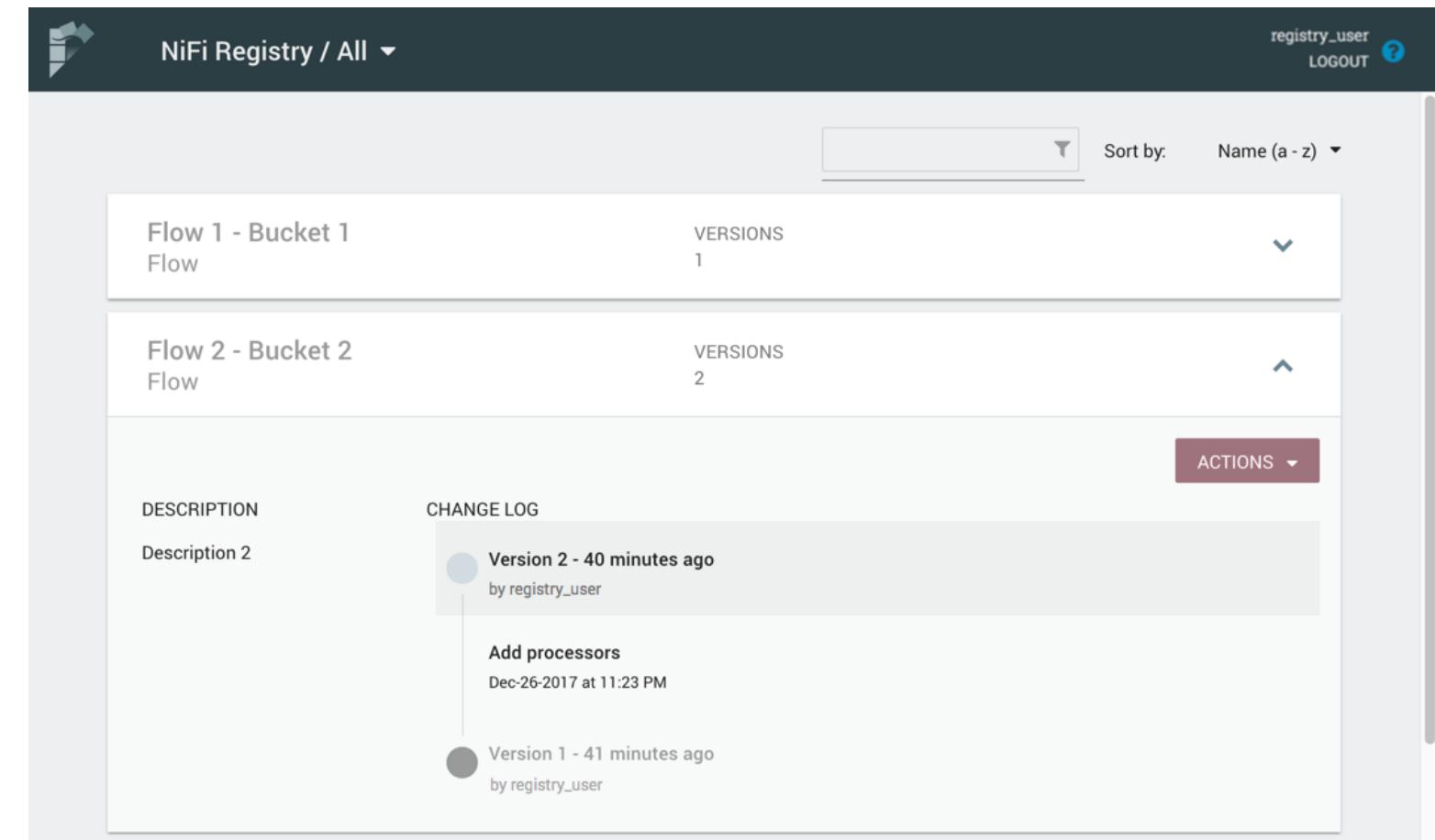
- NiFi 1.8.0 — 26 Oct 2018 (212+ Jiras)
 - Jetty, DB improvements
 - Auto load-balancing queues
 - TLS Toolkit w/ external CA
 - Record processor improvements
- MiNiFi C++ 0.5.0 — 6 June 2018
- MiNiFi Java 0.5.0 — 7 July 2018
- NiFi Registry 0.3.0 — 25 Sept 2018



NiFi Registry for Dataflows

Introducing Apache NiFi Registry 0.3.0

- Previously, flows were exported via XML templates
 - Didn't contain sensitive values
 - Couldn't be updated in-place
 - No tracking system
- NiFi Registry brings asset management as first-class citizen to NiFi
- Flows can be versioned
- Flows can be promoted between environments



NiFi Registry / All ▾

registry_user LOGOUT ?

Sort by: Name (a - z) ▾

Flow 1 - Bucket 1
Flow

VERSIONS 1

Flow 2 - Bucket 2
Flow

VERSIONS 2

DESCRIPTION Description 2

CHANGE LOG

- Version 2 - 40 minutes ago by registry_user
- Add processors Dec-26-2017 at 11:23 PM
- Version 1 - 41 minutes ago by registry_user

ACTIONS ▾

Community Health

apache / nifi

Unwatch 154 Unstar 1,103 Fork 1,041

Code Pull requests 145 Projects 0 Insights

Mirror of Apache NiFi

4,720 commits

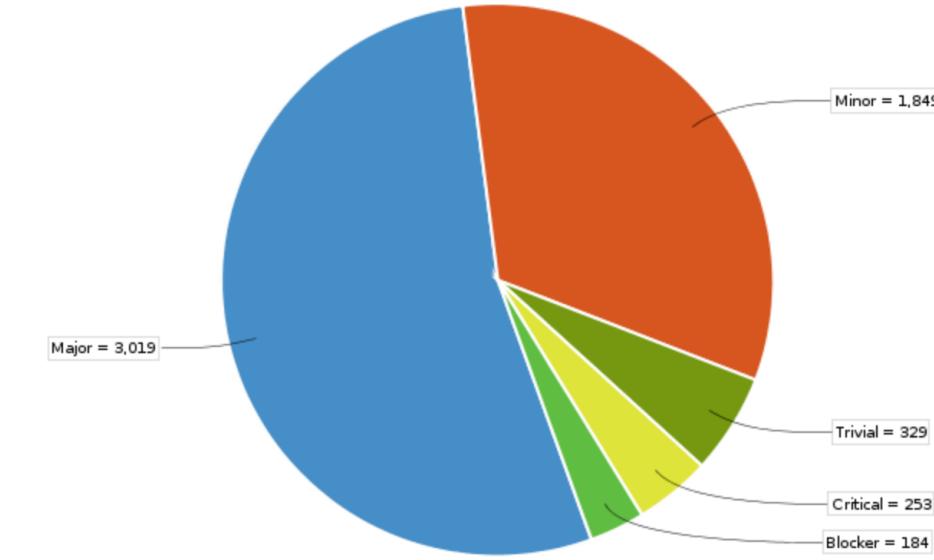
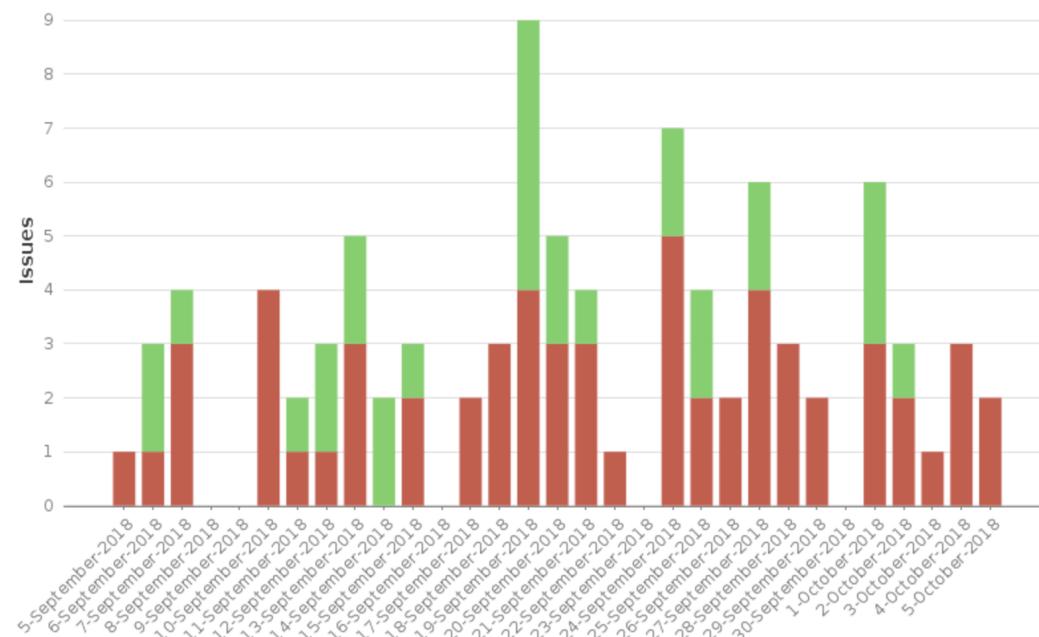
52 branches

63 releases

200 contributors

Apache-2.0

This chart shows the issues created in the last 30 days



Learn more and join us

Apache NiFi site

<https://nifi.apache.org>

Subproject MiNiFi site

<https://nifi.apache.org/minifi/>

Subscribe to and collaborate at

dev@nifi.apache.org

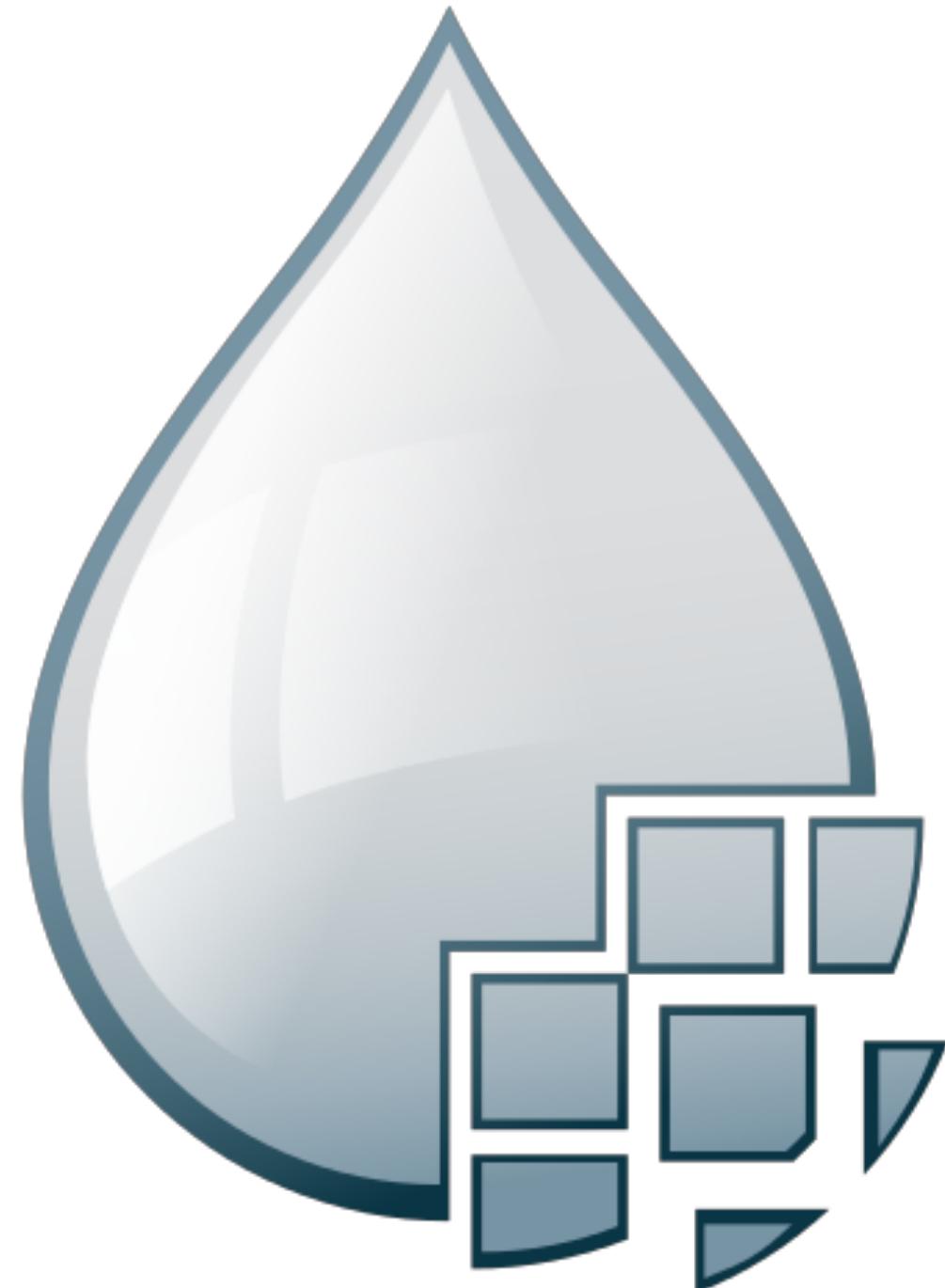
users@nifi.apache.org

Submit Ideas or Issues

<https://issues.apache.org/jira/browse/NIFI>

Follow us on Twitter

[@apachennifi](https://twitter.com/apachennifi)



More NiFi Today

Title	Time	Room
The First Mile – Edge and IoT Data Collection with Apache NiFi and MiNiFi	1100 - 1140	Room 103
Apache NiFi Crash Course	1400 - 1600	Room 109
Dataflow Management From Edge to Core with Apache NiFi	1650 - 1730	Room 112
Using Spark Streaming and NiFi for the Next Generation of ETL in the Enterprise	1650 - 1730	Room 103



Thank you

alopresto@hortonworks.com | alopresto@apache.org | [@yolopey](https://twitter.com/yolopey)
github.com/alopresto/slides