



# Intelligently Collecting Data at the Edge with Apache NiFi and MiNiFi

Andy LoPresto | @yolopey

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

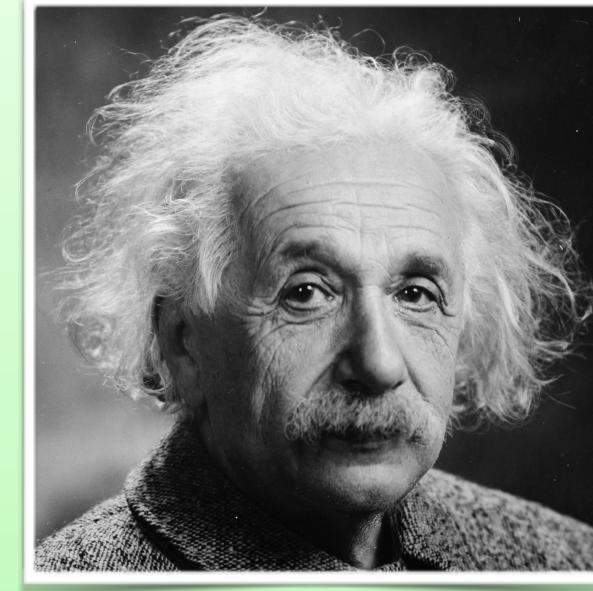
19 April 2018 Dataworks Summit Berlin

# 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

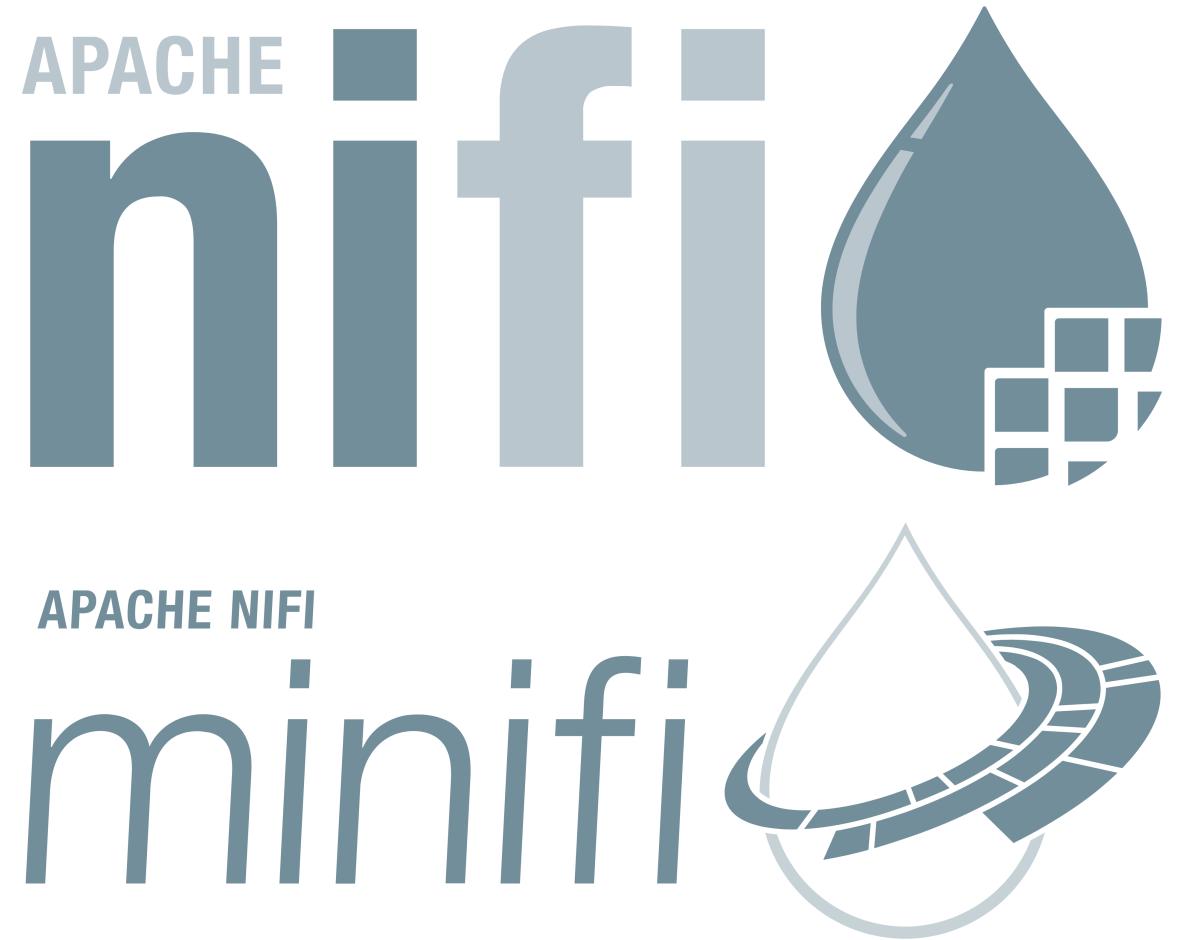


*“I refactored the Ambari integration endpoint to allow for mutual authentication TLS during my coffee break”*

Forgotten more about NiFi than most of us will ever know

# Agenda

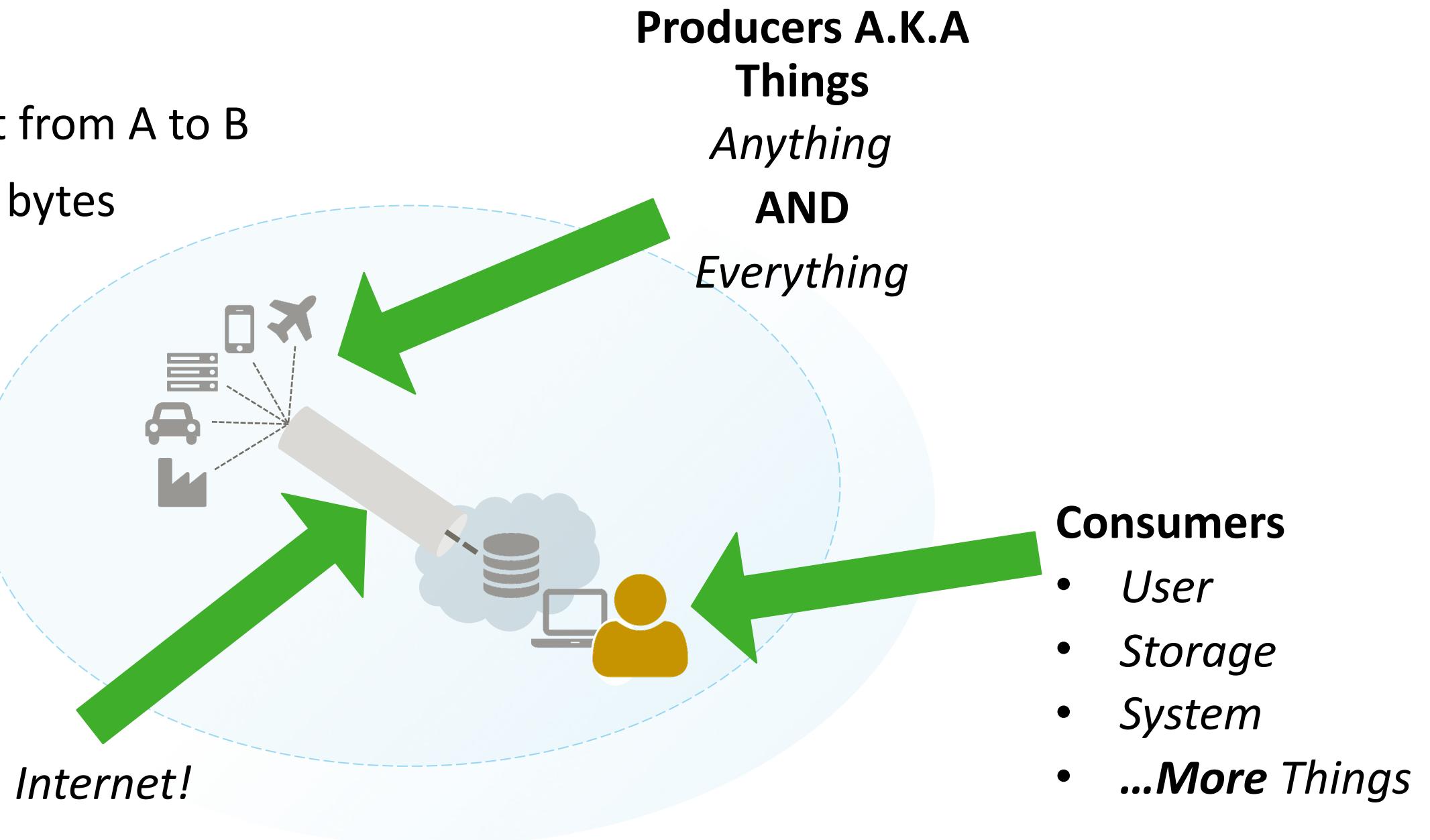
- *What is dataflow and what are the challenges?*
  - *Apache NiFi*
  - IoT Challenges
  - Apache MiNiFi
  - Exploration
  - Community
- 
- All slides provided online, so no need to transcribe



# 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



# Dataflow Challenges In 3 Categories

## Data

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

## Infrastructure

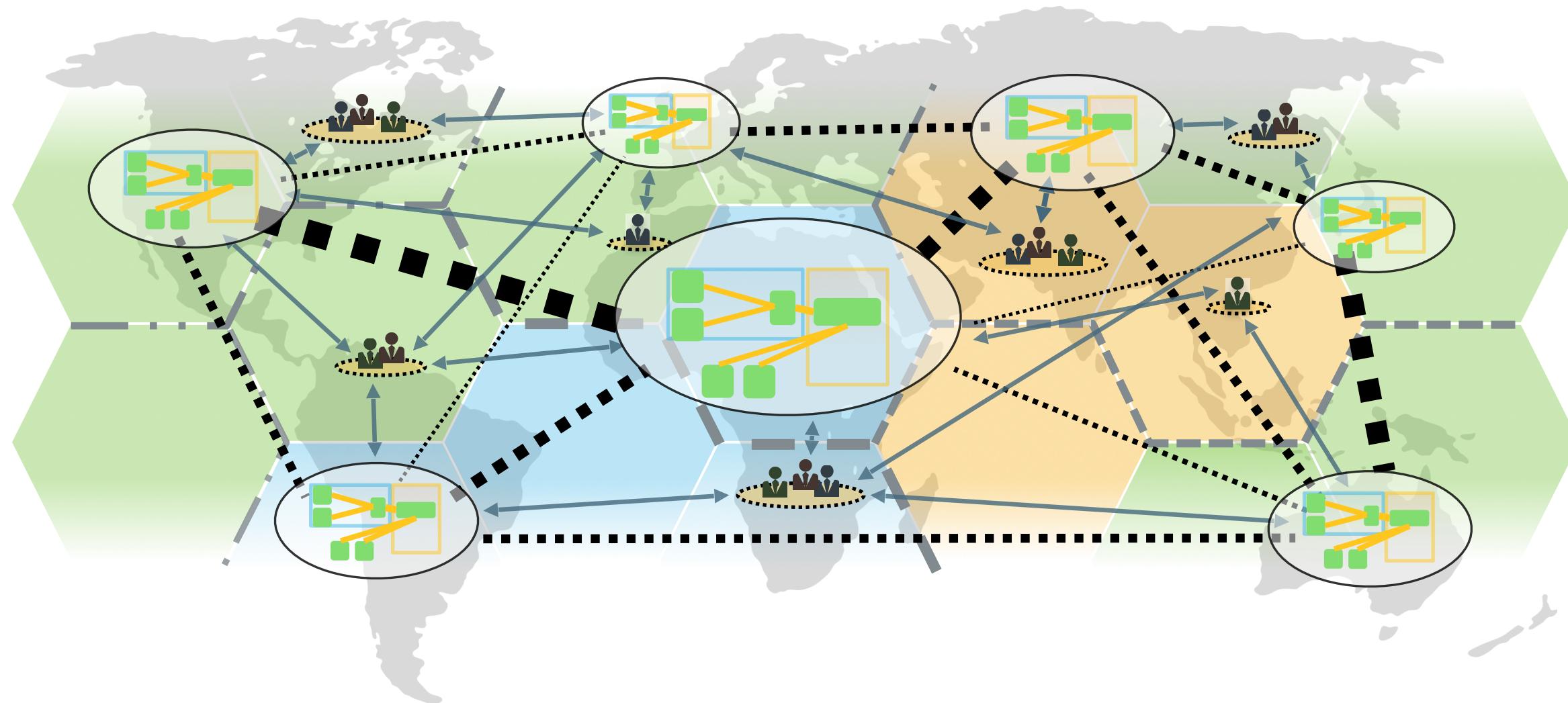
- “Exactly Once”  
Delivery
- Ensuring  
Security
- **Overcoming**  
Security
- Credential  
Management
- Network

## 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 $\Delta$ s to Cs to $\varphi$ s

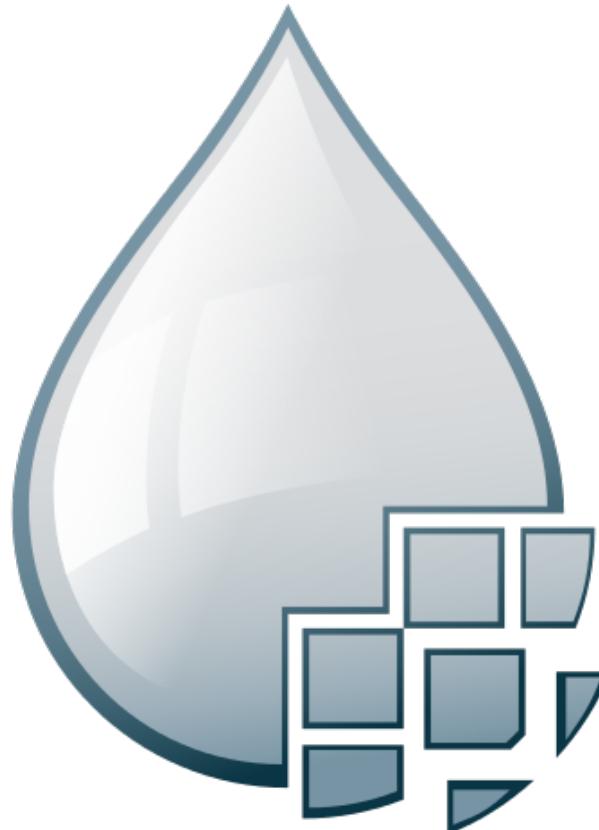
Raise your hand if you want to maintain Python scripts for the rest of your life



# What is Apache NiFi?

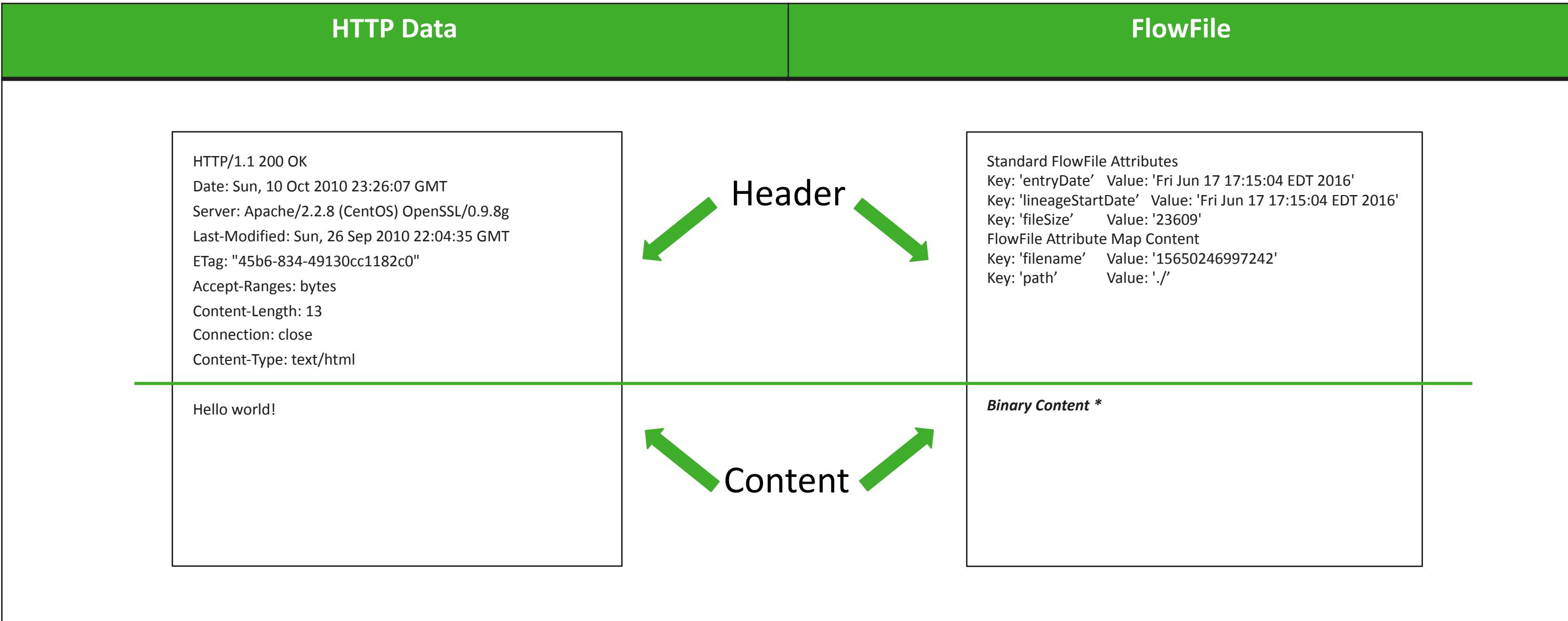
# Apache NiFi

## Key Features



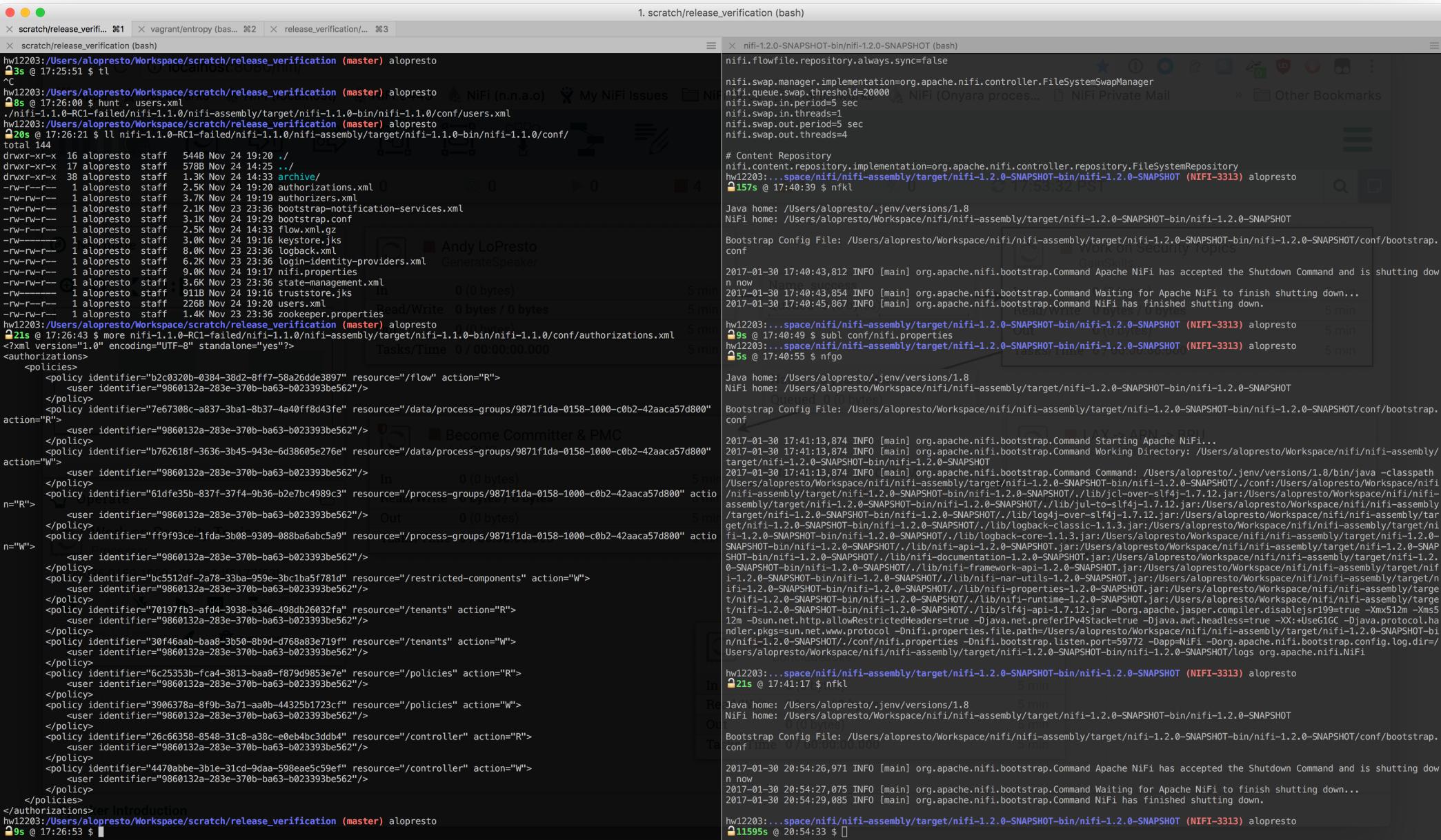
- Guaranteed delivery
  - Data buffering
  - Backpressure
- Pressure release
- Prioritized queuing
- Flow specific QoS
  - Latency vs. throughput
  - Loss tolerance
- Data provenance
- Supports push and pull models
- Recovery/recording a rolling log of fine-grained history
- Visual command and control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

# Flowfiles Are Like HTTP Data



# User Interface

Less of this...



The screenshot displays a Mac OS X desktop environment with three windows open:

- Terminal 1 (Left):** Shows a file listing and command history related to NiFi release verification.
- Terminal 2 (Right):** Shows NiFi logs, including bootstrap configuration and shutdown logs.
- Browser Window:** Displays the Hortonworks logo.

Terminal 1 (Left) Output:

```
hadoop@hadoop-OptiPlex-5090:~/Desktop$ ls -l
total 144
drwxr-xr-x 16 alopresto staff 544B Nov 24 19:20 .
drwxr-xr-x 17 alopresto staff 578B 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-rw-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-rw-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 3.6K 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
hadoop@hadoop-OptiPlex-5090:~/Desktop$ more nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.1.0/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-a837-3ba1-8b37-4a40ff8d43fe" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aac57d800" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="b762618f-3d36-3b45-943e-6d38605e276e" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aac57d800" action="W">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="61dfe35b-837f-4b36-b2e7bc4989c3" resource="/process-groups/9871f1da-0158-1000-c0b2-42aac57d800" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="ff9f93ce-1fda-3b08-9309-088ba6abc5a9" resource="/process-groups/9871f1da-0158-1000-c0b2-42aac57d800" action="W">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="9860132a-283e-370b-ba63-b023393be562" 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-fc4a-3813-baa8-f879d9853e7e" resource="/policies" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="3906378a-8fb8-3a71-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>
    </authorizations>
hadoop@hadoop-OptiPlex-5090:~/Desktop$
```

Terminal 2 (Right) Output:

```
hadoop@hadoop-OptiPlex-5090:~/Desktop$ ./nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT bash
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:26:00,854 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:26:00,854 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:26:00,867 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.

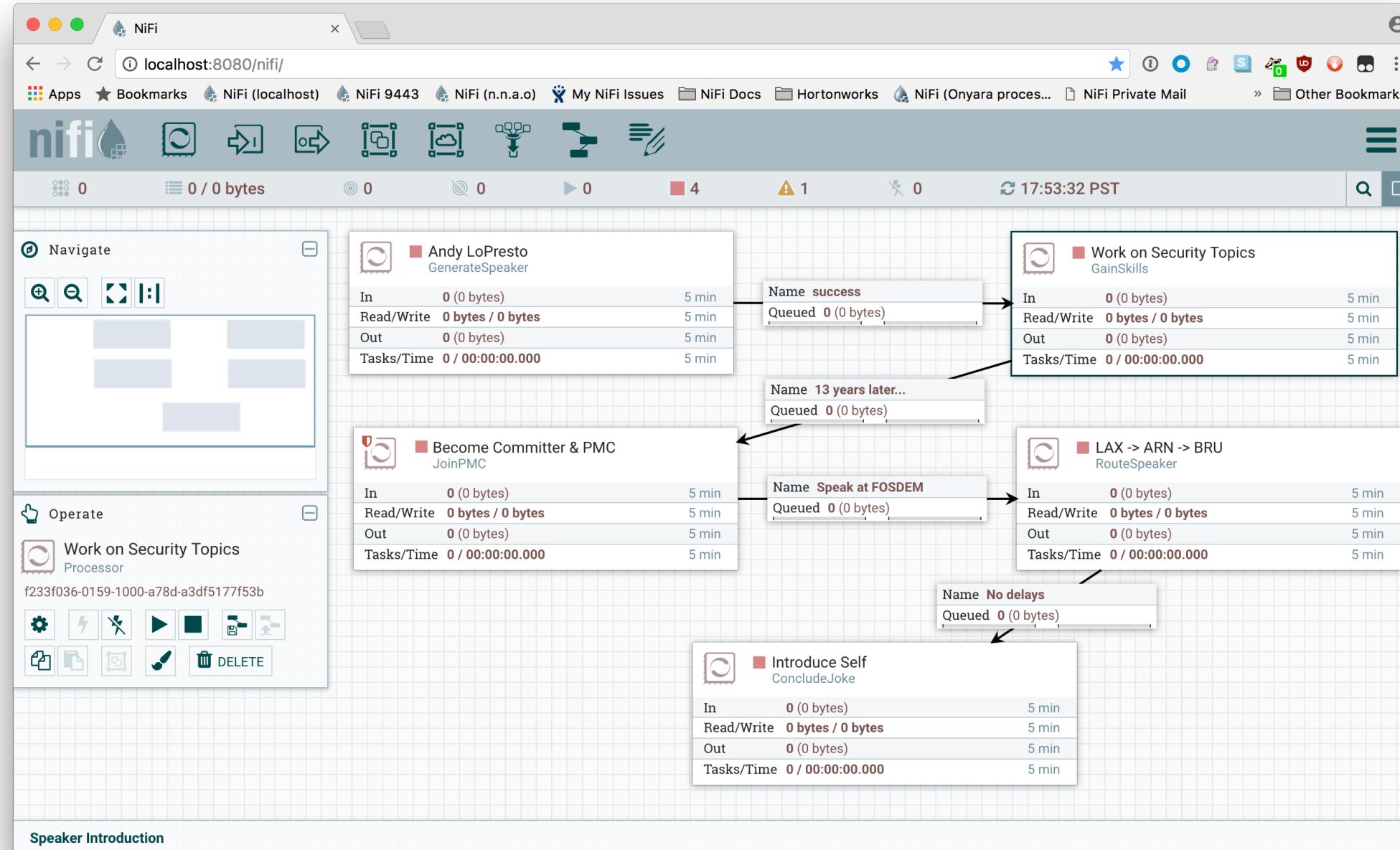
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:39,157s $ nfkl
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:39,157s $ nfkl
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:43,854 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:43,854 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:45,867 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.

[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:49,21s $ subl conf/nifi.properties
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:49,21s $ subl conf/nifi.properties
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:40:55,5s $ nfgo
[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 Starting Apache NiFi...
[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 Working Directory: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
[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/conf:/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-1.2.0-SNAPSHOT/,/lib/nifi-apidocs-1.2.0-SNAPSHOT/,/lib/nifi-documentation-1.2.0-SNAPSHOT/,/lib/nifi-framework-api-1.2.0-SNAPSHOT/,/lib/nifi-nar-utils-1.2.0-SNAPSHOT/,/lib/nifi-properties-1.2.0-SNAPSHOT/,/lib/nifi-runtime-1.2.0-SNAPSHOT/,/lib/nifi-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/,/lib/nifi-compiler-disablejsr199=true-Xmx512m-Xms512m-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.file.path=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT/,/conf/nifi.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/logs org.apache.nifi.NiFi
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:41:17,21s $ nfkl
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 17:41:17,21s $ nfkl
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 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
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 20:54:27,075 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 20:54:29,085 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.

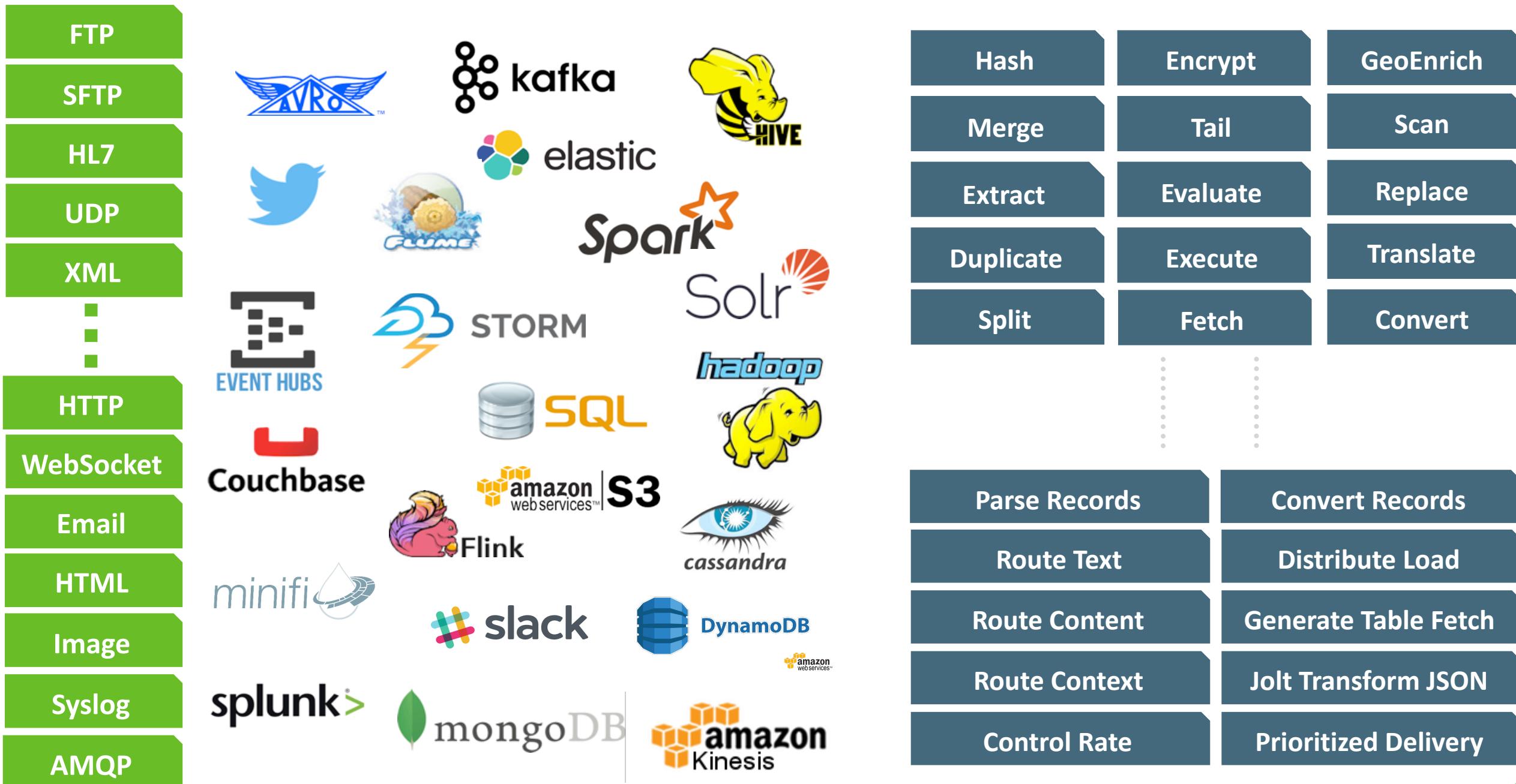
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 20:54:33,159s $ nfkl
[nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT] 2017-01-30 20:54:33,159s $ nfkl
```

# User Interface

Less of this... ... more of this



# Deeper Ecosystem Integration: 260+ Processors, 48 Controller Services



All Apache project logos are trademarks of the ASF and the respective projects.



# 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

# 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<sup>†1</sup>, Shogo Suzuki<sup>†1</sup>, Katsunari Yoshioka<sup>†1</sup>, Tsutomu Matsumoto<sup>†1</sup>,  
Takahiro Kasama<sup>†2</sup>, Christian Rossow<sup>†3</sup>

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

Google search results for "mirai botnet".

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

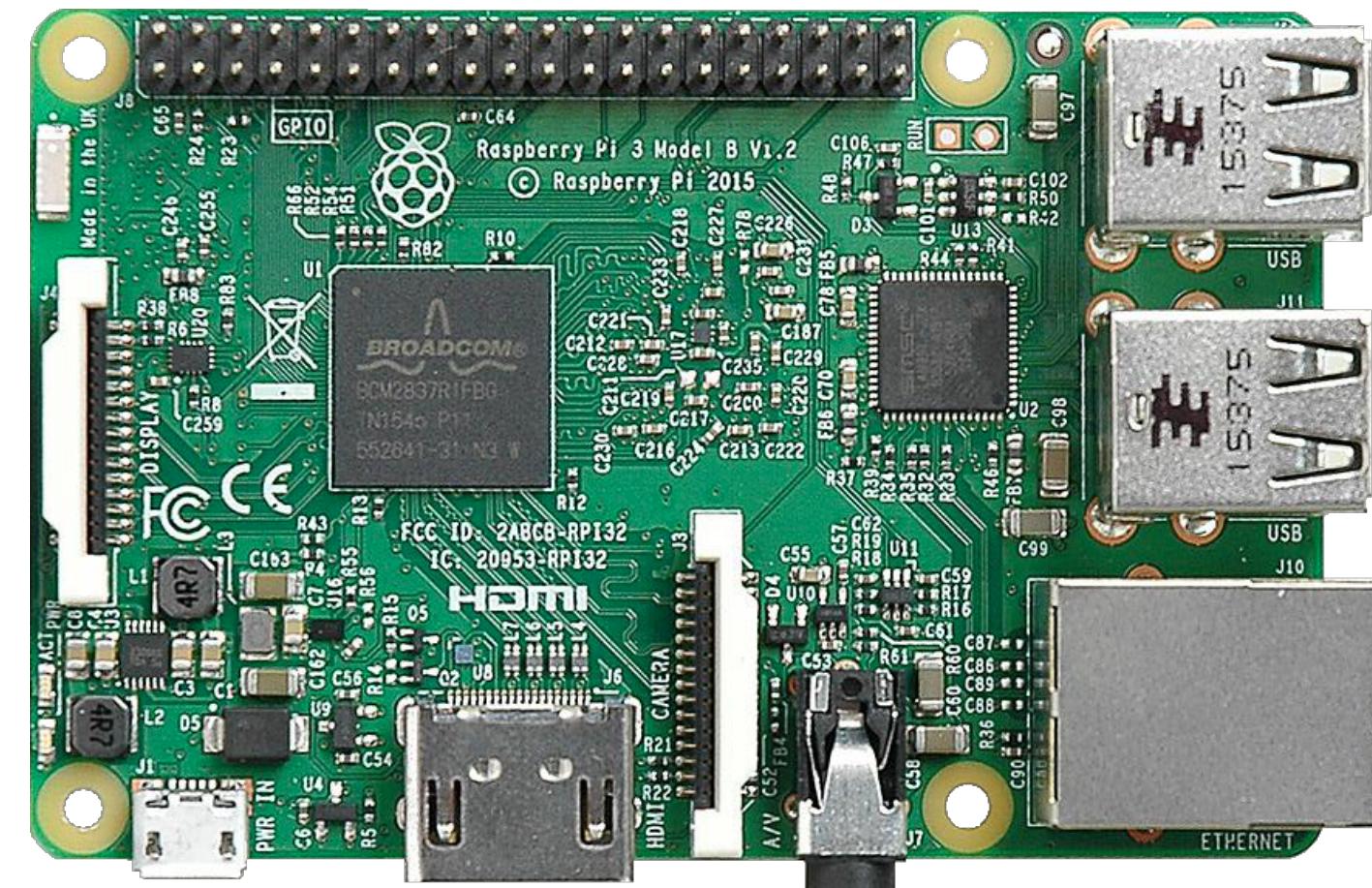
# NiFi Solves Everything\*

- 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

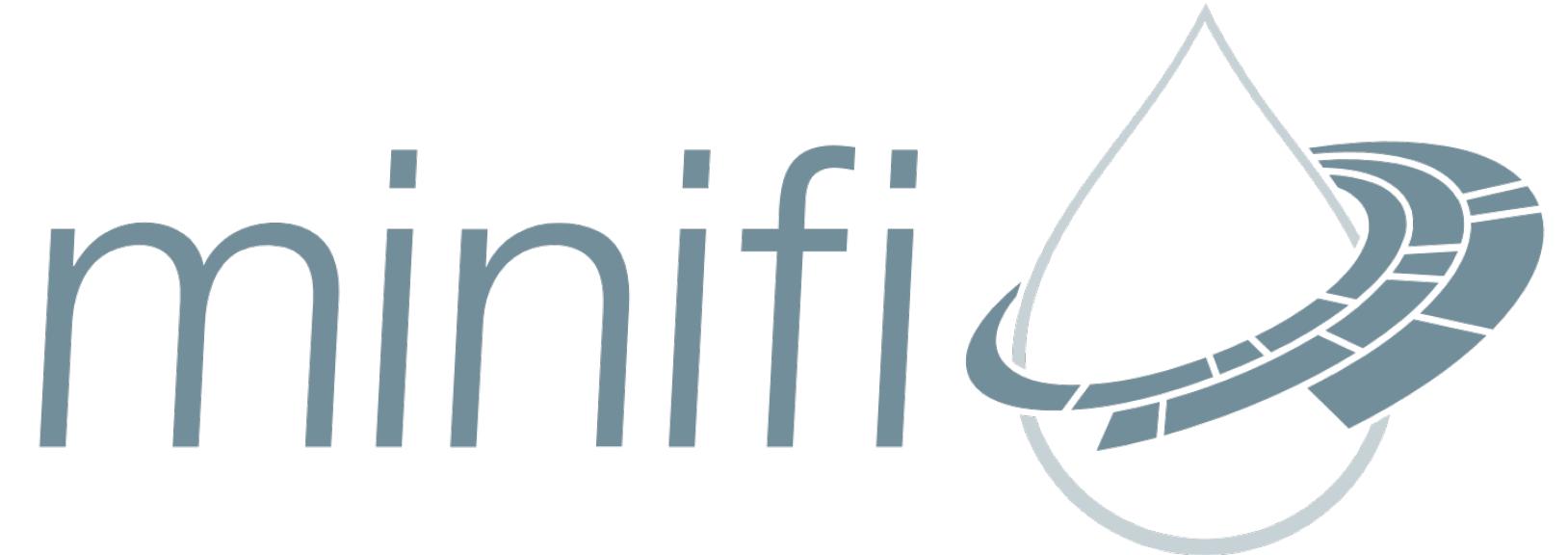
```
▶hw12203:/Users/alopresto/Workspace/scratch/rp3b-demo (master) alopresto
└─ 113s @ 17:09:05 $ ssh pi@my-raspberry-pi
^C
▶hw12203:/Users/alopresto/Workspace/scratch/rp3b-demo (master) alopresto
└─ 145s @ 17:09:37 $ █
```



# 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

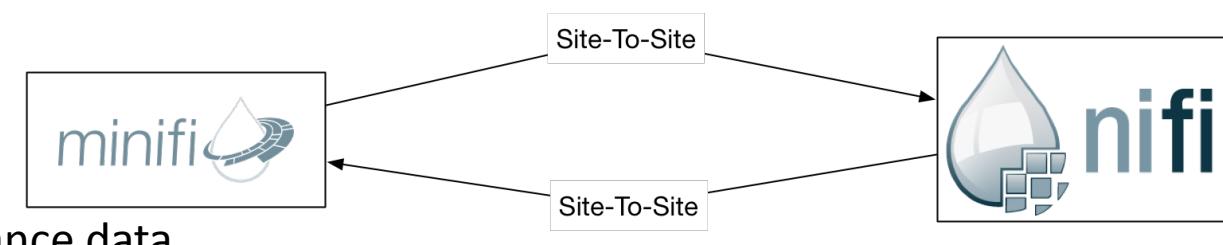
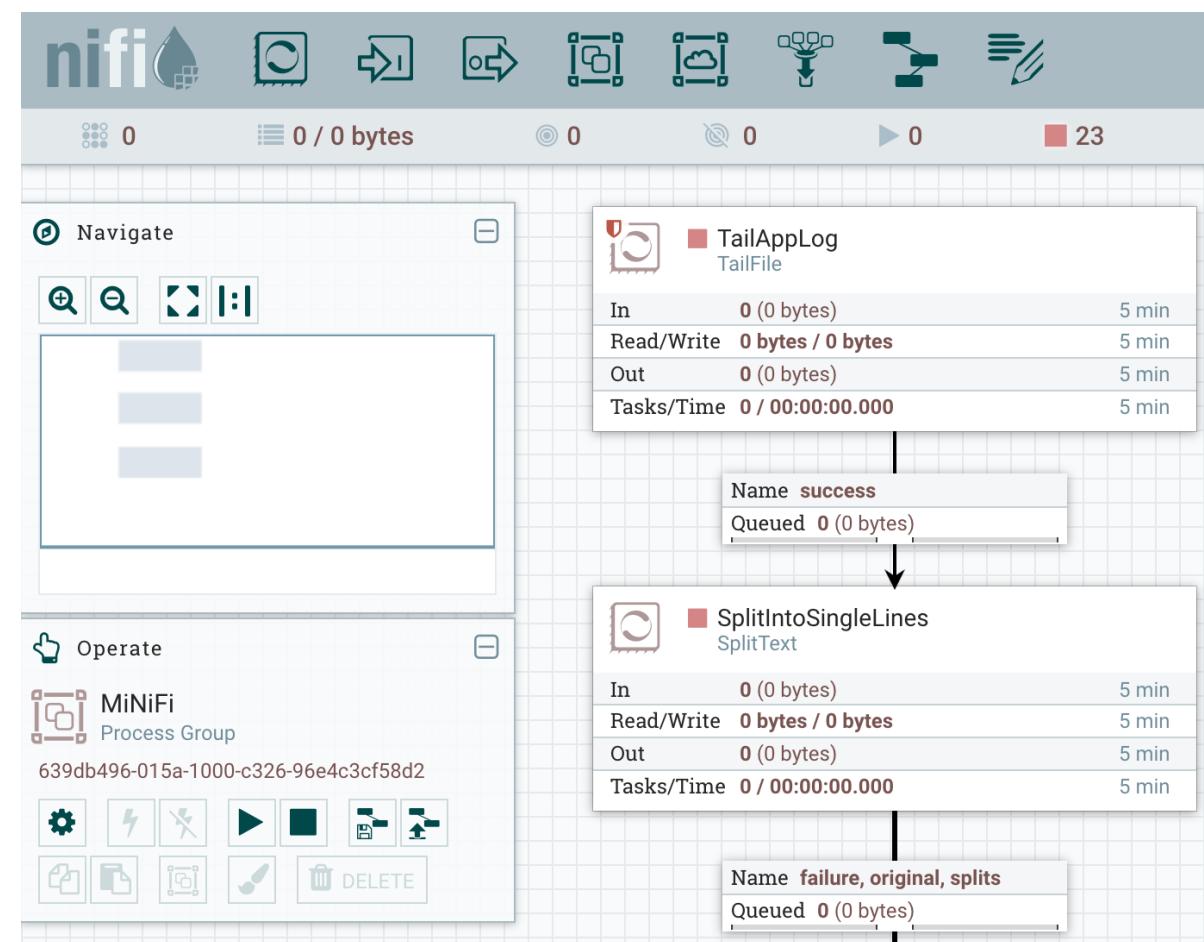


# Why build MiNiFi?

- NiFi is big
  - 1.6.0 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
  - 0.4.0 Java binary is 65 MB, C++ binary is 4.5 MB (**0.2.0 fits on a floppy disk**)
  - “Good guest”

# Flavors of MiNiFi

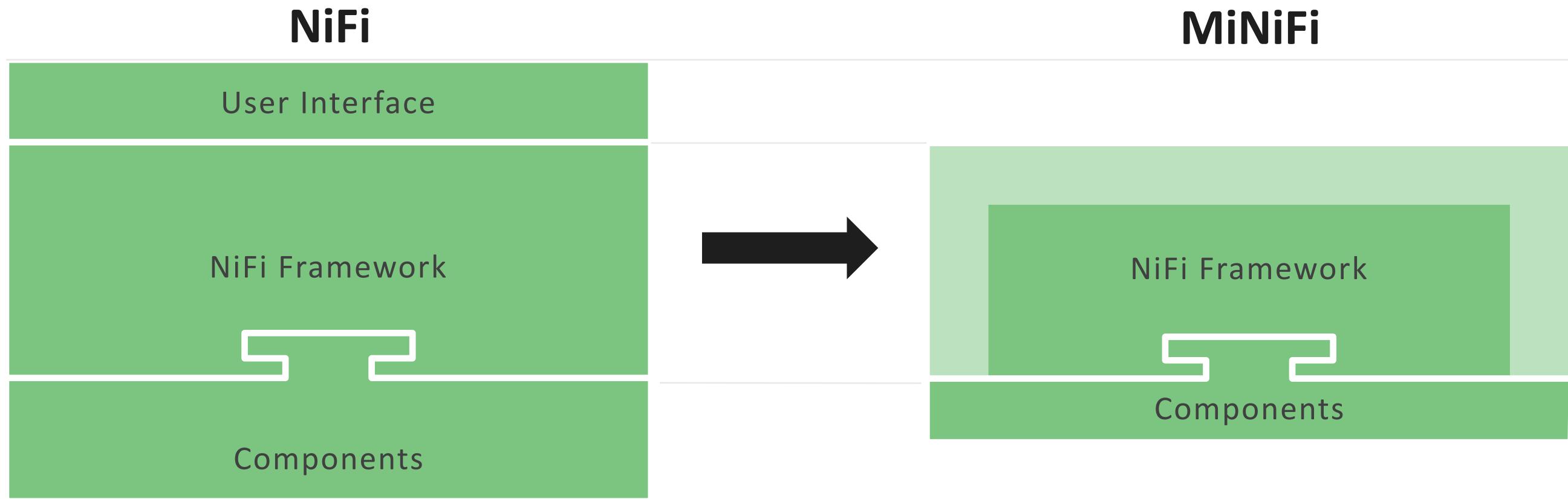
- MiNiFi Java (v0.4.0)
  - Modified version of NiFi
    - No UI
    - YAML configuration
  - Reduced processor count
    - 110+ by default, more available with additional NARs
- MiNiFi C++ (v0.4.0)
  - Written from scratch
  - 28 processors by default
  - Bi-directional site-to-site & provenance data



```
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

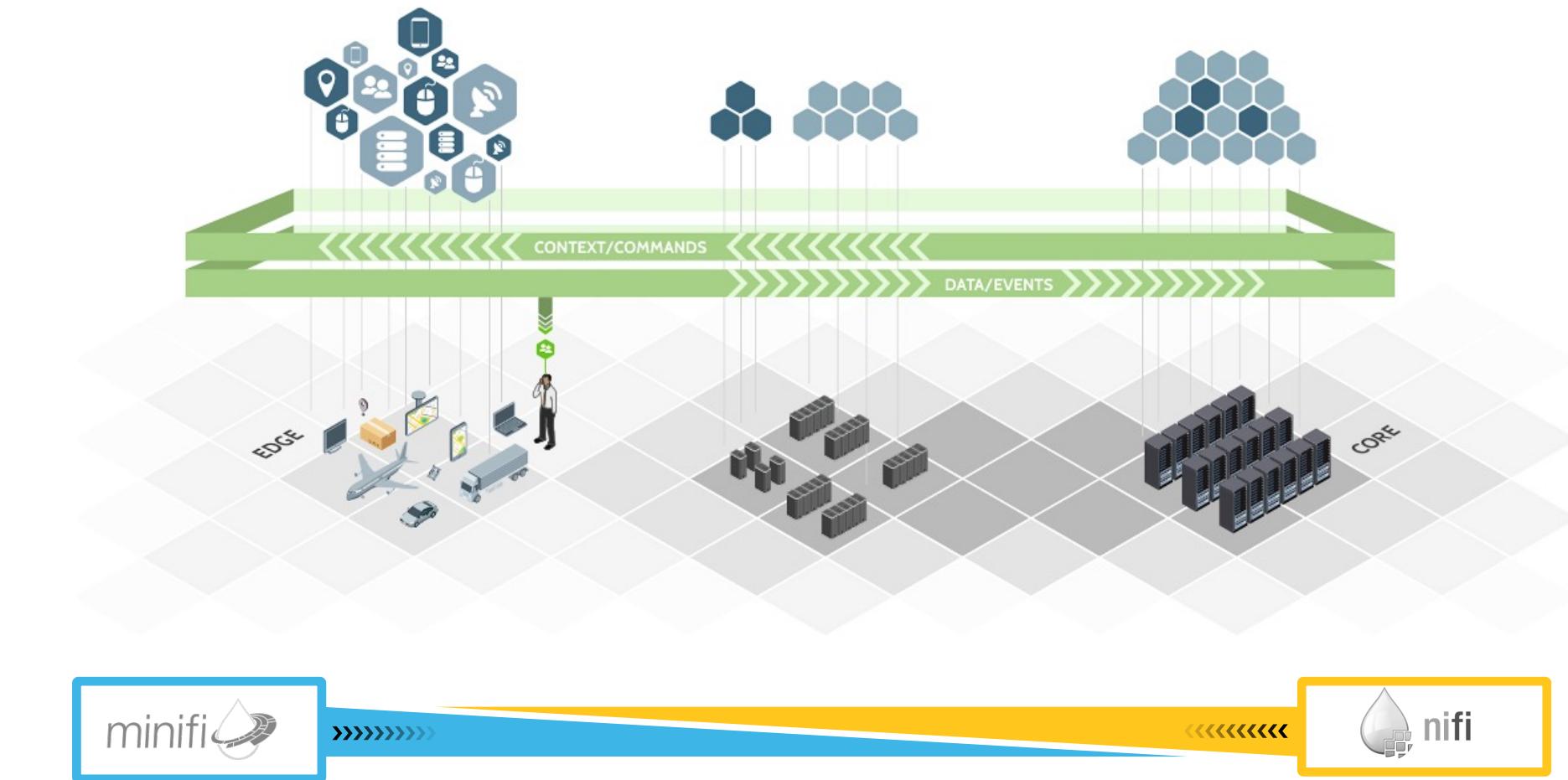
Processors:
- name: TailAppLog
  class: org.apache.nifi.processors.standard.TailFile
  max concurrent tasks: 1
  scheduling strategy: TIMER_DRIVEN
  scheduling period: 10 sec
  penalization period: 30 sec
  yield period: 1 sec
  run duration nanos: 0
  auto-terminated relationships list:
Properties:
  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



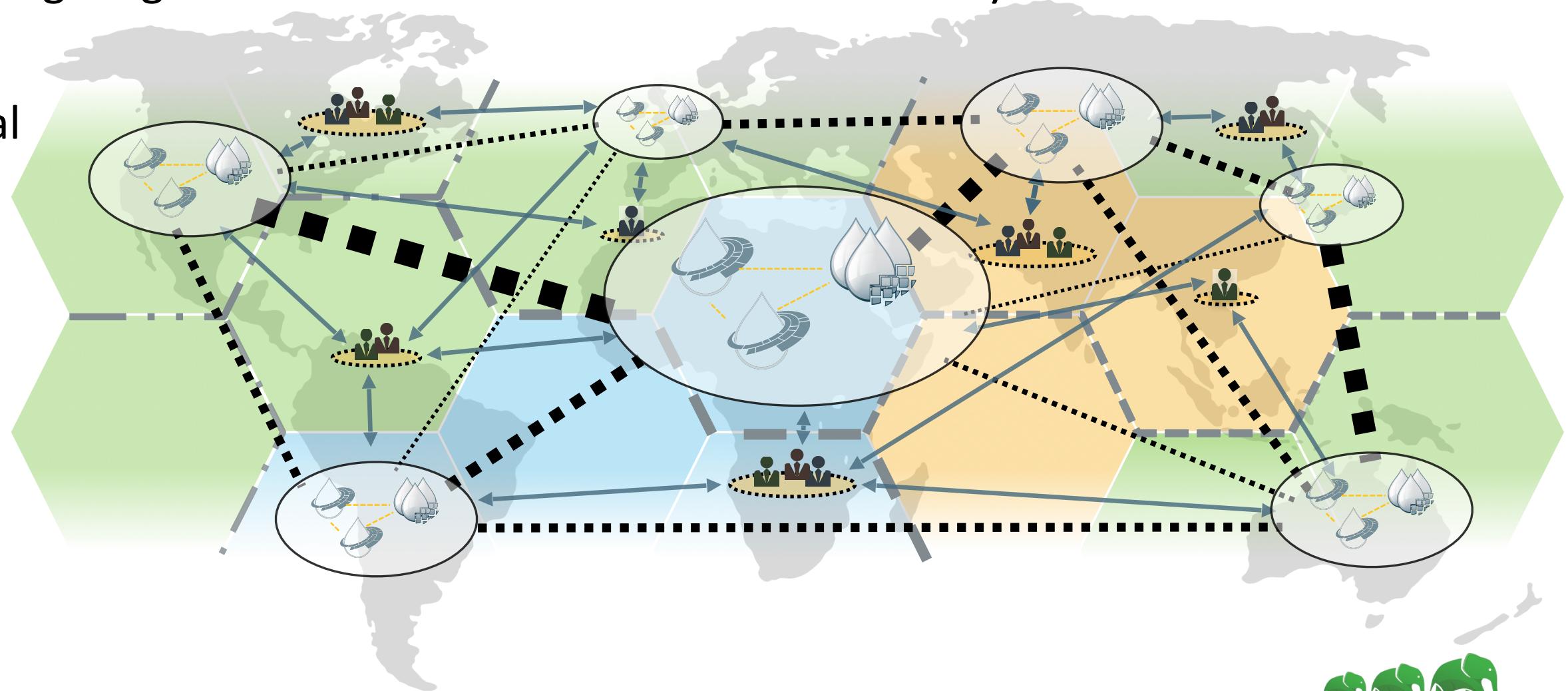
# How Does MiNiFi Interact With NiFi?

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



# Let's Add Dimensionality

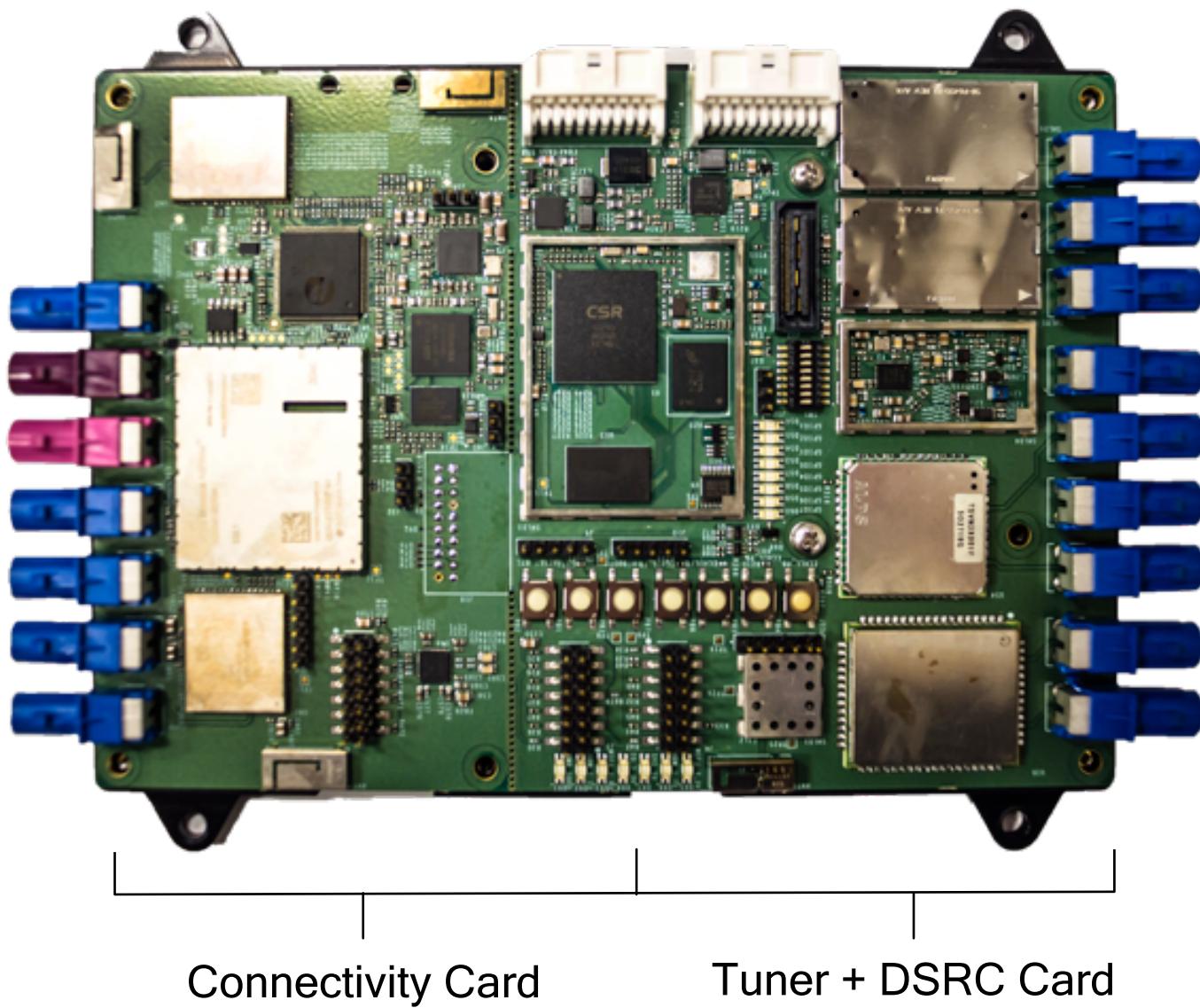
- We've been imagining EDGE to CORE as a bi-directional linear system
- Let's expand that to the real world



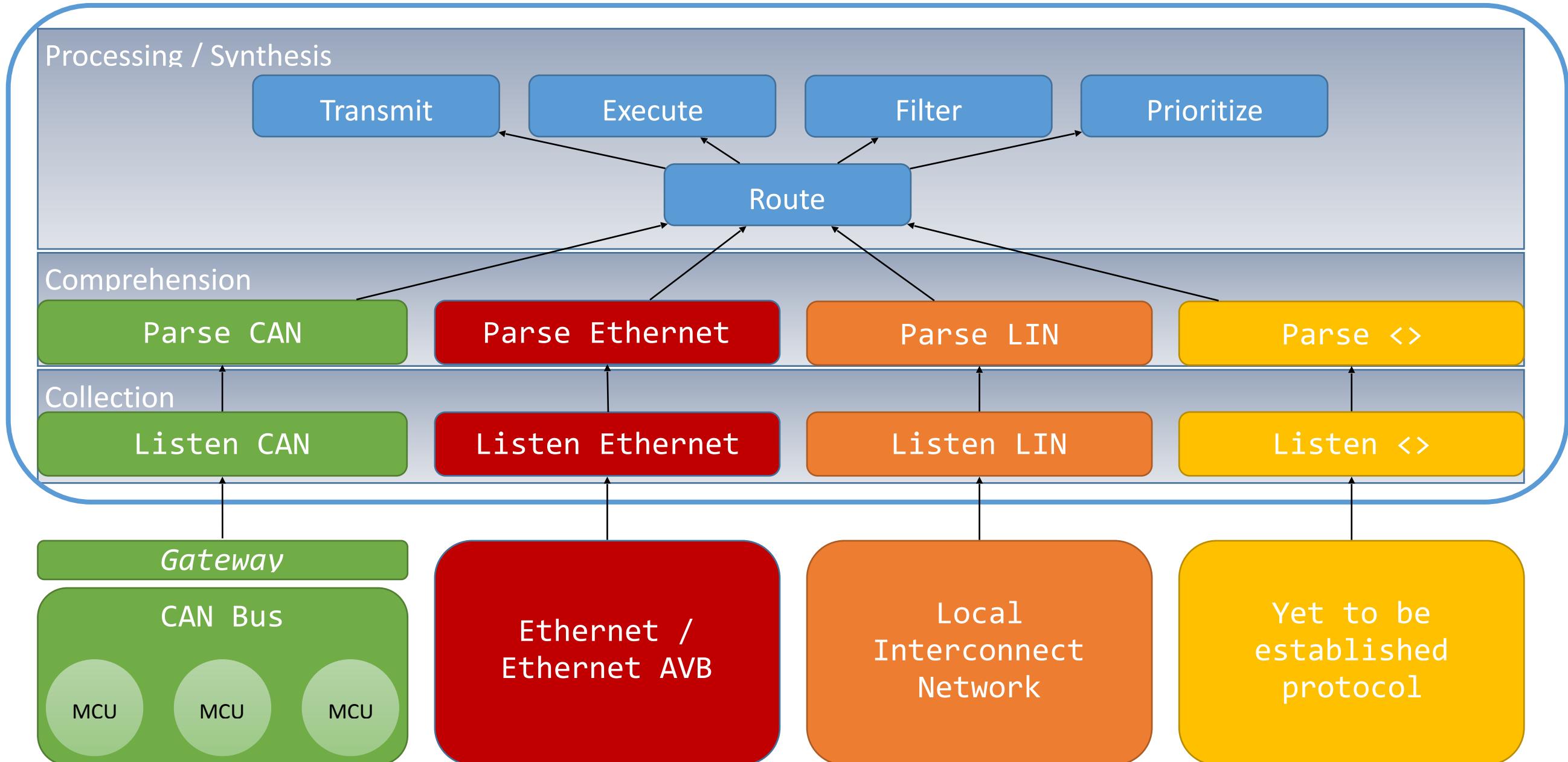
# What does MiNiFi provide?

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

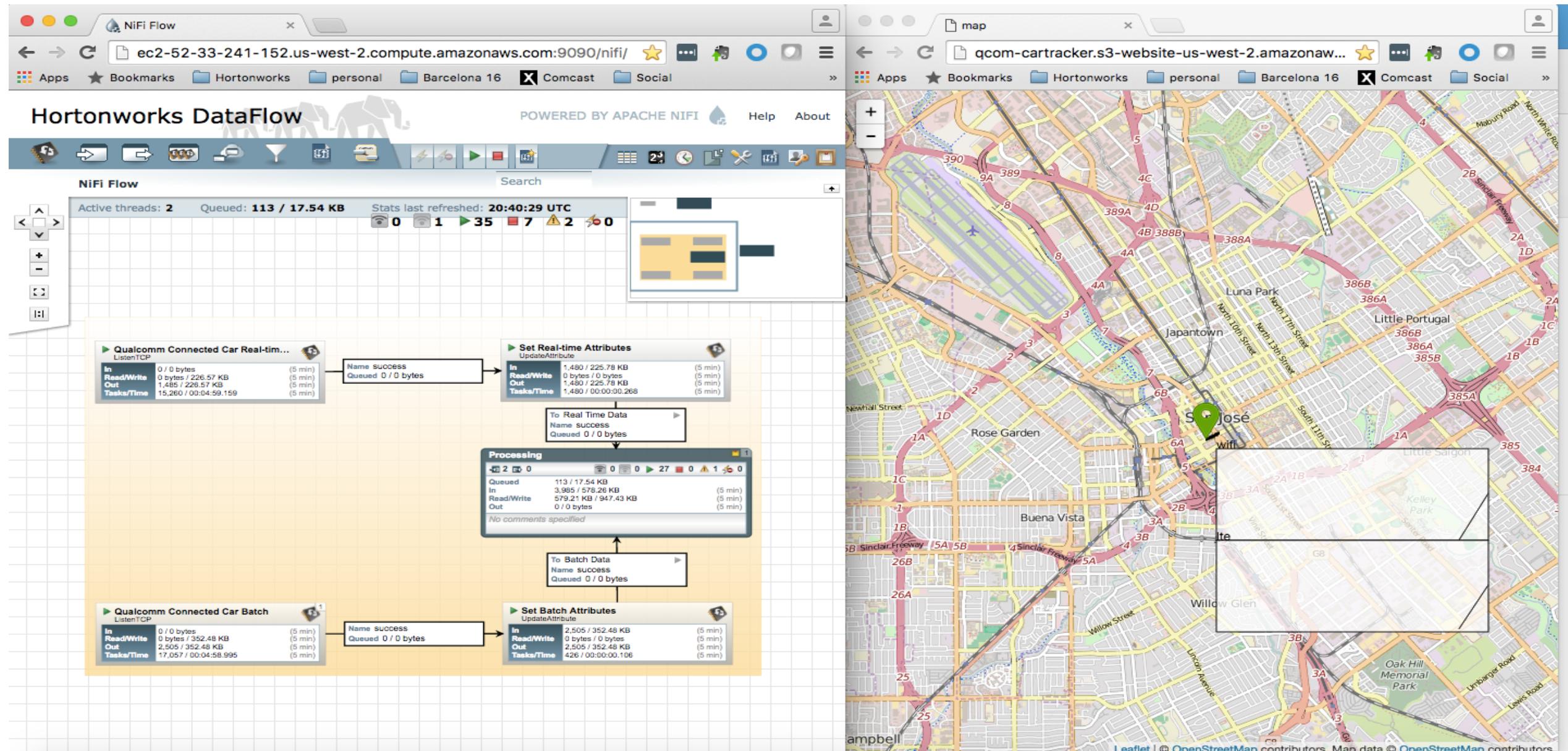
## Connected Car Reference Platform Box



# MiNiFi on a Connected Car



# MiNiFi on a Connected Car



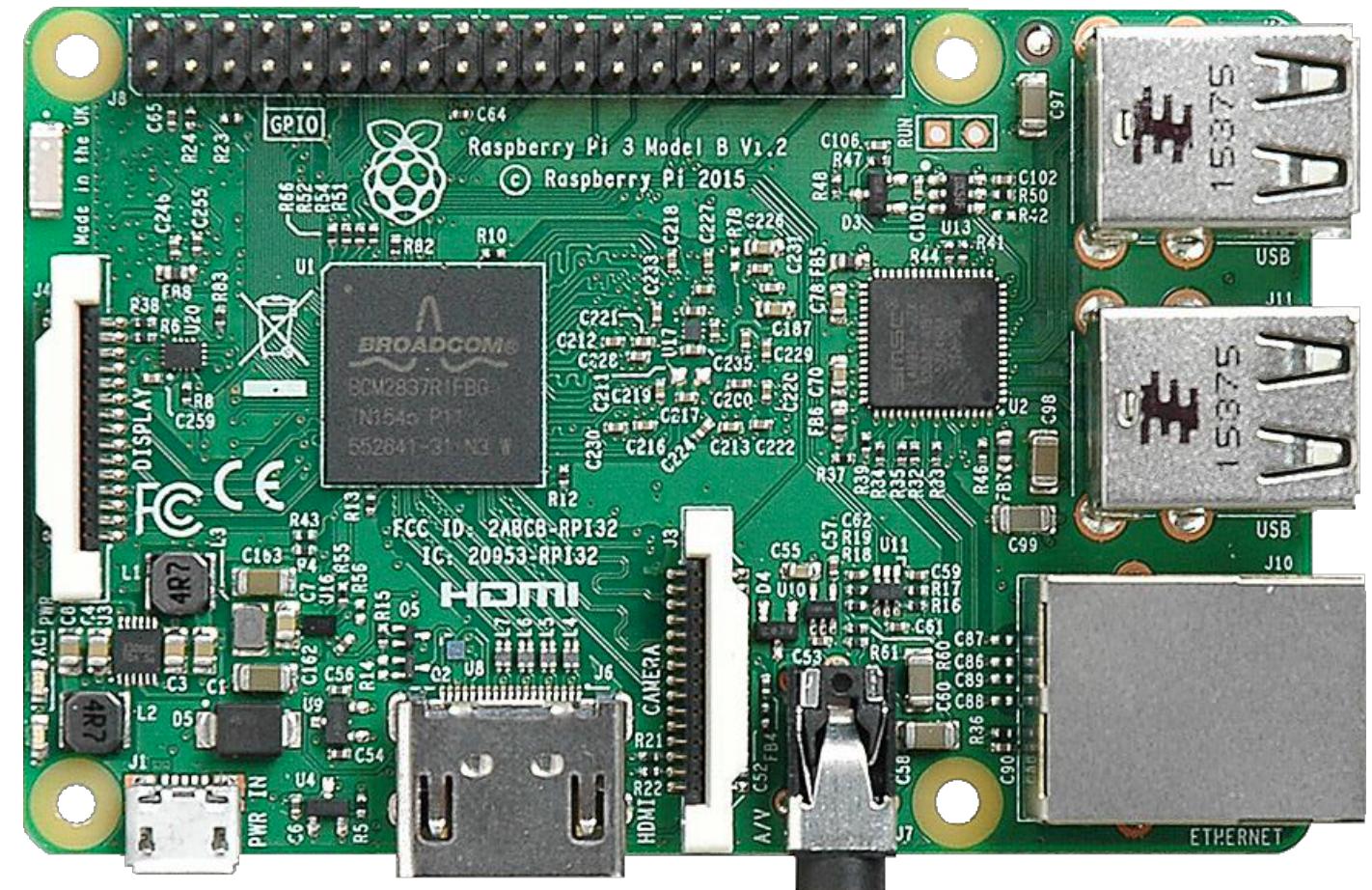
# MiNiFi Exfil

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

# Edge Data Exploration

# Scenario

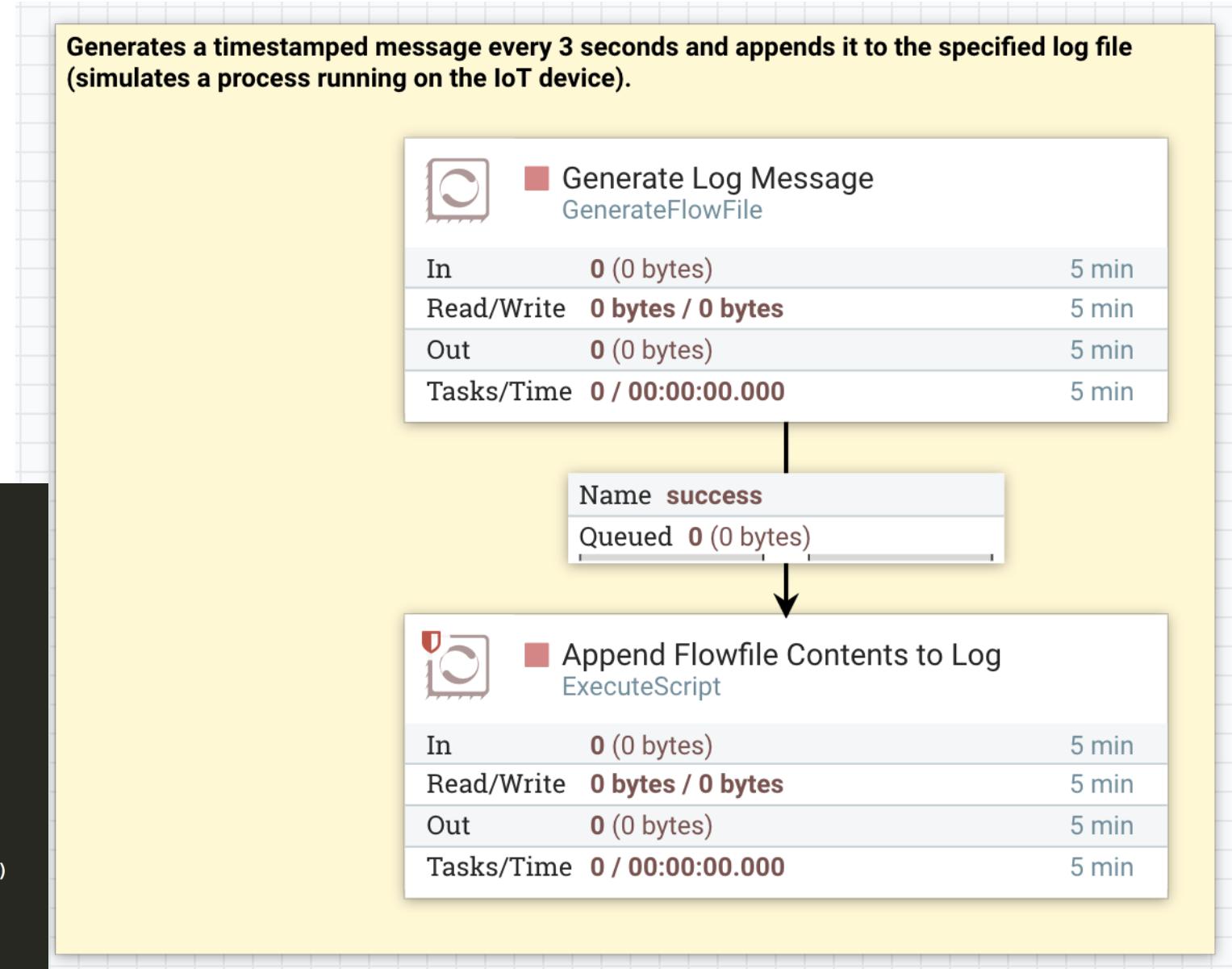
- IoT Device generating log messages
- Need to encrypt data on device
- Need to prioritize some data for unreliable network connectivity
- Transmit data to central node
- Decrypt data and analyze
- *Make determinations and modify live flow*



# NiFi As Test Harness/Environment

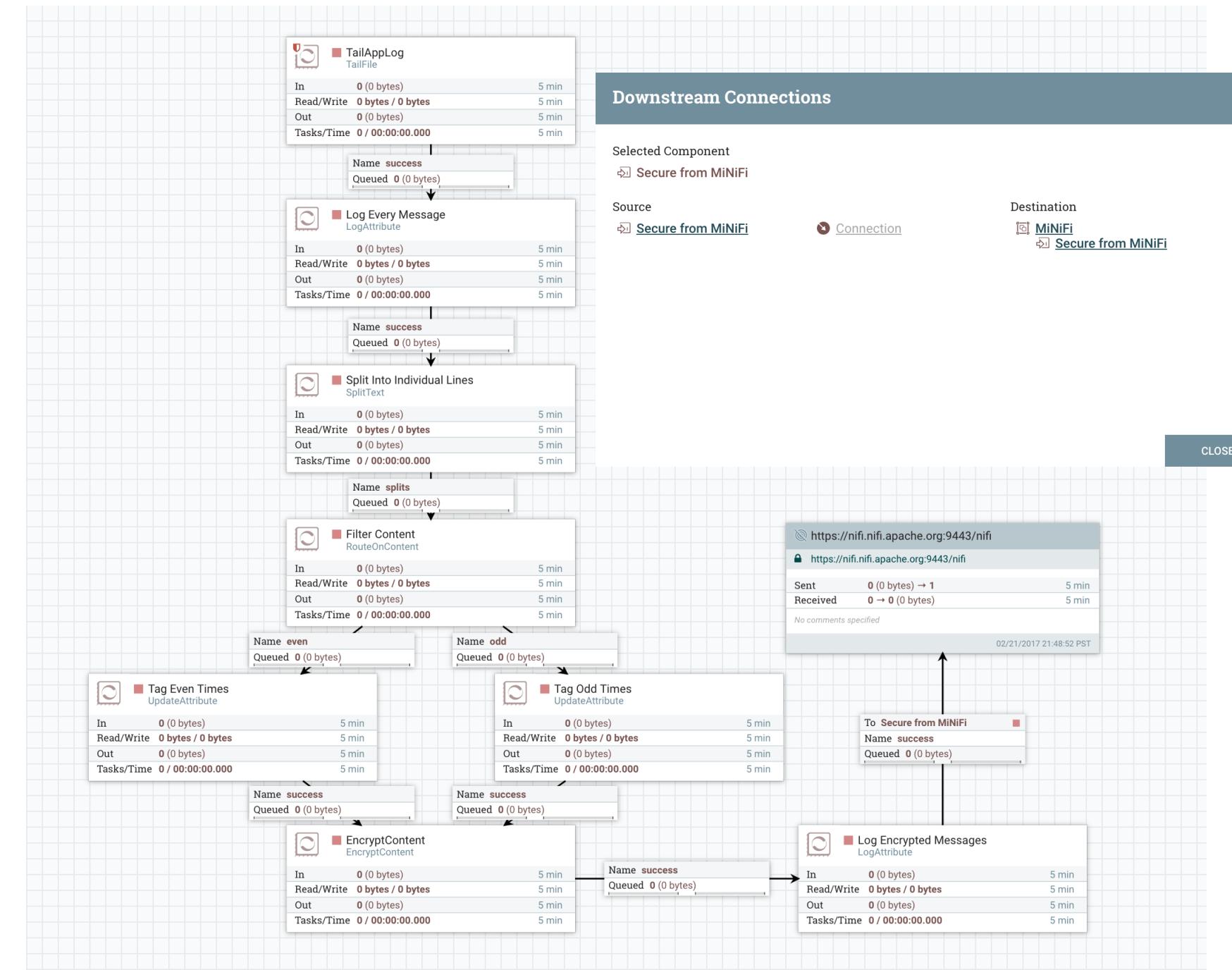
- Simulate the log generation
- Schedule is customizable
- Script can write to dynamic location

```
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{D62B02CFD5556B97EA9B7241F1164C05F392CB81AEAEFF4C564033030055E546A8D6A2E9F7614494
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:
```

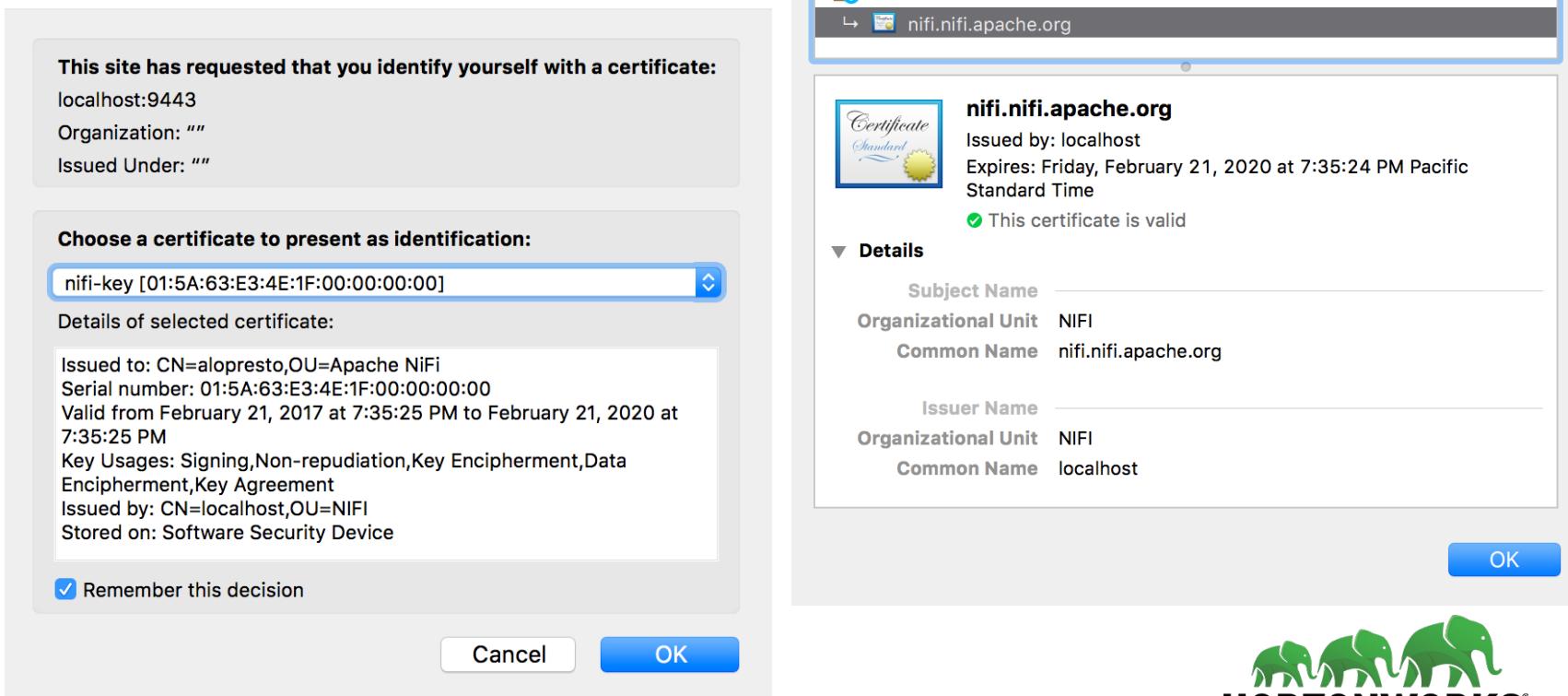
# Setting Up Crypto

- 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

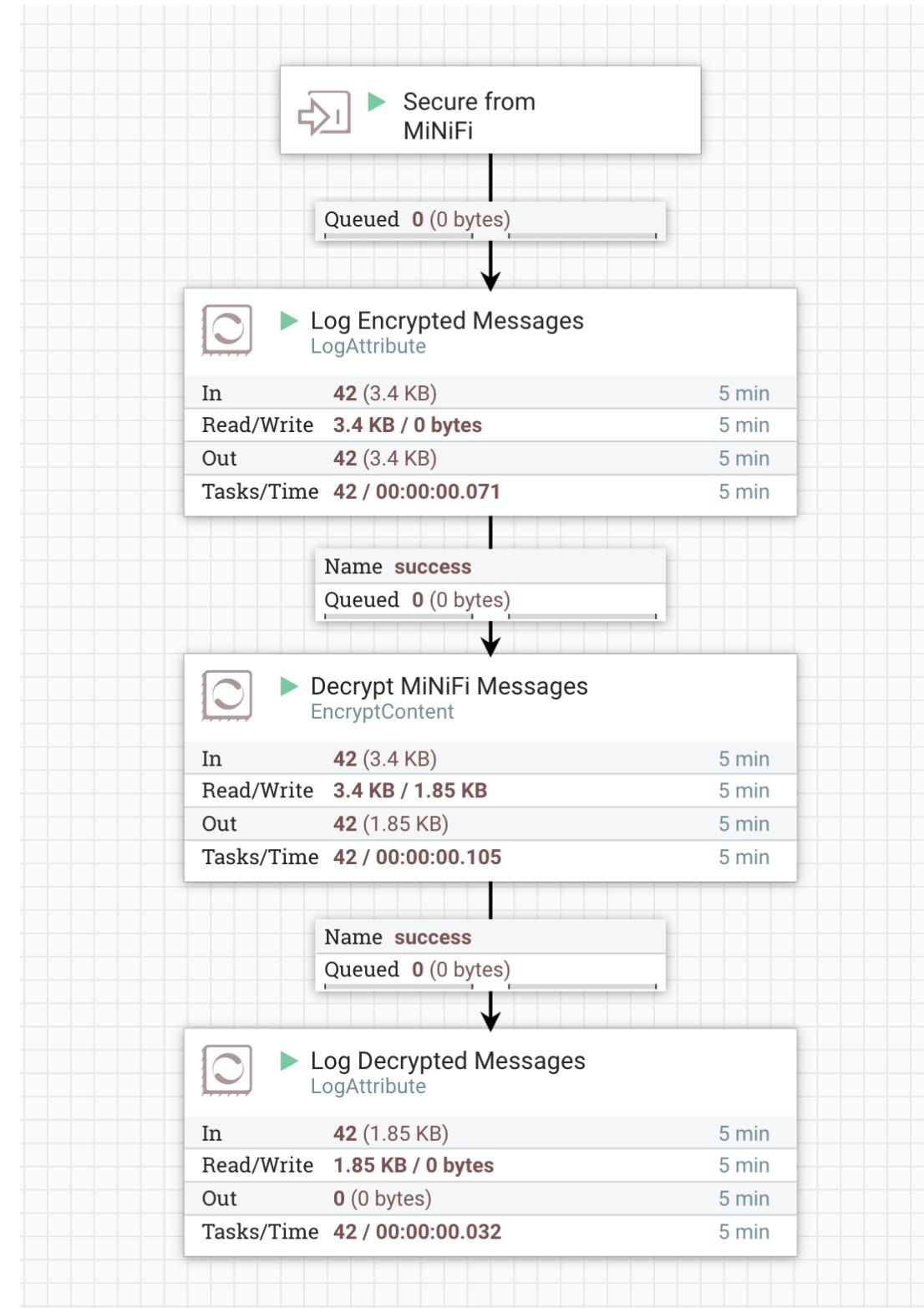
Required field +

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 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옙心中的口擊回回揮櫻噉啖口𢂔口惹十萊鯤邝銓姪旣口口口卵裂𢂔𡇱ㄥ講擘瓊𤁵齡口譖瓈嬖倘
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 B5 ....Sy.....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 =..	
View as:	original
1	This is a message at 2017/02/21 22:16:27.986Z

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 22:22:05.065 PST	CONTENT_MODIFIED
02/21/2017 22:22:05.064 PST	DROP
02/21/2017 22:22:05.064 PST	DROP

Provenance Event

DETAILS	ATTRIBUTES	CONTENT	
<strong>Input Claim</strong>		<strong>Output Claim</strong>	
Container default		Container default	
Section 1		Section 1	
Identifier 1487744083443-1		Identifier 1487744083443-1	
Offset 4259		Offset 6681	
Size 83 bytes		Size 45 bytes	
<a href="#">DOWNLOAD</a>	<a href="#">VIEW</a>	<a href="#">DOWNLOAD</a>	<a href="#">VIEW</a>

Replay

Connection Id 015a1002-3be1-13fc-98e3-f54ee6183280

OK

Last updated: 22:22:49 PST

NiFi Flow » MINIFI

TailFile\_Split\_Prioritize.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-4a5d-8aee-1959605eb86b'
222 This is a message at 2017/02/21 22:27:30.017Z
223 Standard FlowFile Attributes
224 Key: 'entryDate'
225     Value: 'Tue Feb 21 22:27:30 PST 2017'
226 Key: 'lineageStartDate'
227     Value: 'Tue Feb 21 22:27:30 PST 2017'
228 Key: 'fileSize'
229     Value: '45'
230 FlowFile Attribute Map Content
231 Key: 'RouteOnContent.Route'
232     Value: 'even'
233 Key: 'Store.State'
234     Value: 'Do not store state'
235 Key: 'filename'
236     Value: 'example.3083-3175.log'
237 Key: 'fragment.count'
238     Value: '2'
239 Key: 'fragment.identifier'
240     Value: 'c0ffd57f-cc93-49f4-8307-061144324d9b'
241 Key: 'fragment.index'
242     Value: '1'
243 Key: 'fragment.size'
244     Value: '45'
245 Key: 'mime.type'
246     Value: 'text/plain'
247 Key: 'path'
248     Value: './'
249 Key: 'priority'
250     Value: '2'
251 Key: 's2s.address'
252     Value: 'localhost:54726'
253 Key: 's2s.host'
254     Value: 'localhost'
255 Key: 'segment.original.filename'
256     Value: 'example.3083-3175.log'
257 Key: 'tailfile.original.path'
258     Value: '/Users/alopresto/Workspace/scratch/openiot/example.log'
259 Key: 'text.line.count'
260     Value: '1'
261 Key: 'uuid'
262     Value: '917aa10f-ccc0-459b-a315-f8a8ae36ade2'
263 This is a message at 2017/02/21 22:27:29.516Z
264
```

# Next Steps

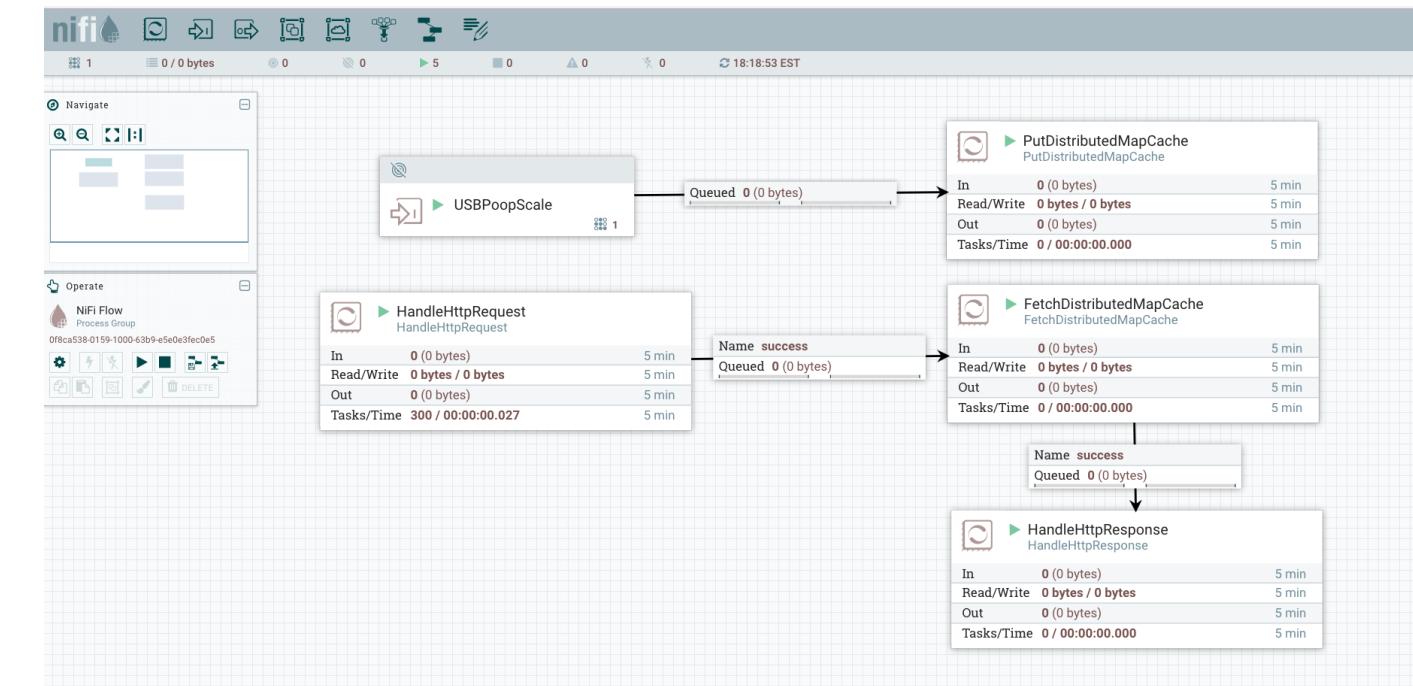
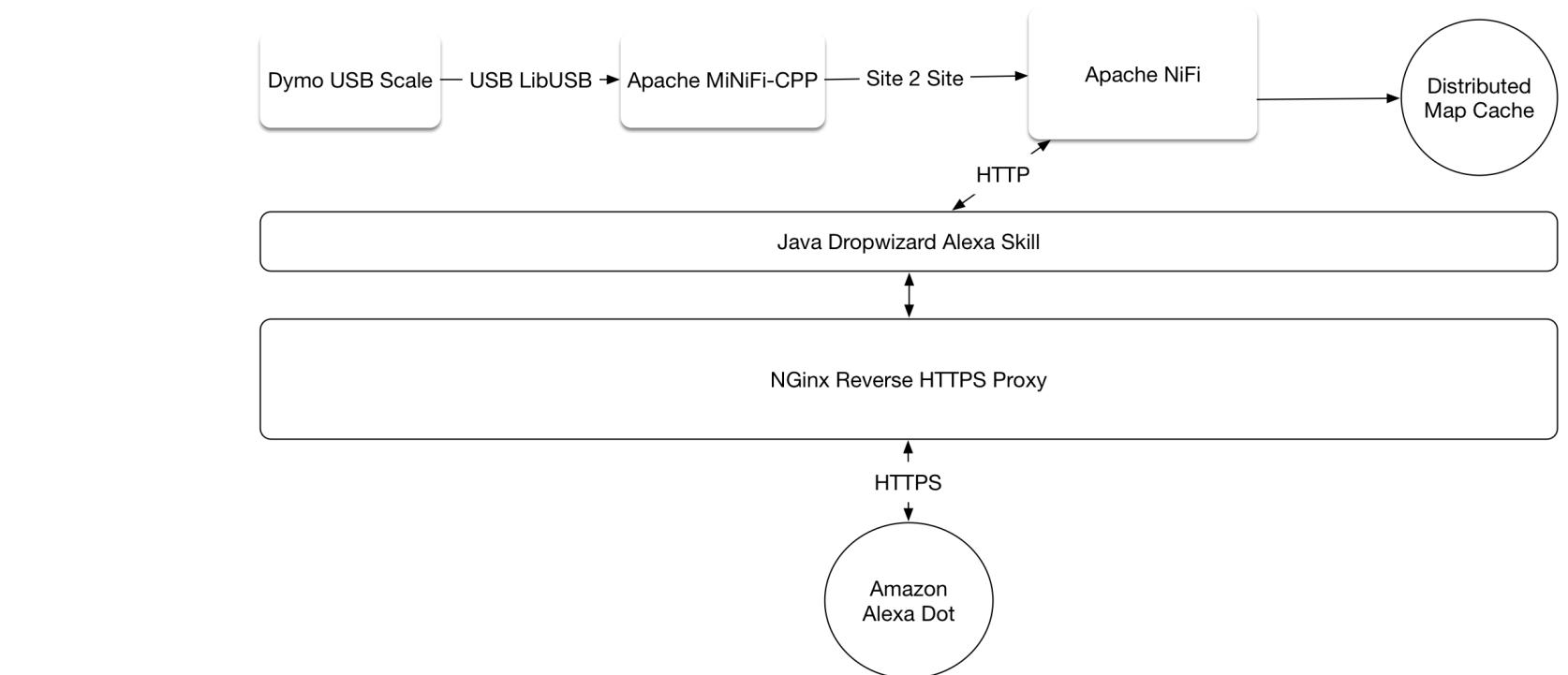
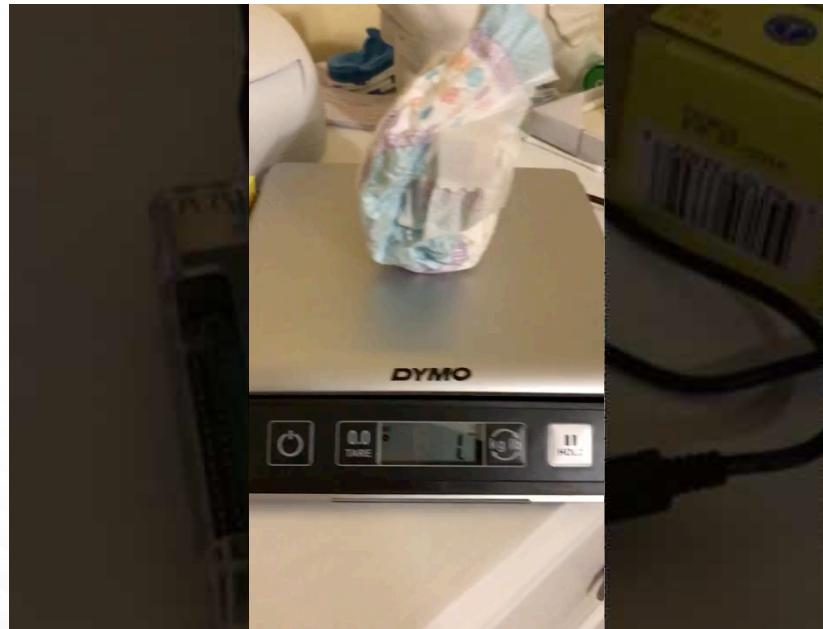
- Window Aggregator
  - If >60% odd in window, switch prioritization
- Encrypt with different keys for different tags & send to different follow-on systems
- Spotty network? Tell MiNiFi to cache low priority and send in batches
- MiNiFi rollover & pruning of monitored log
- Exfil MiNiFi provenance data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

# Community

# Community Example

◆ Jeremy Dyer

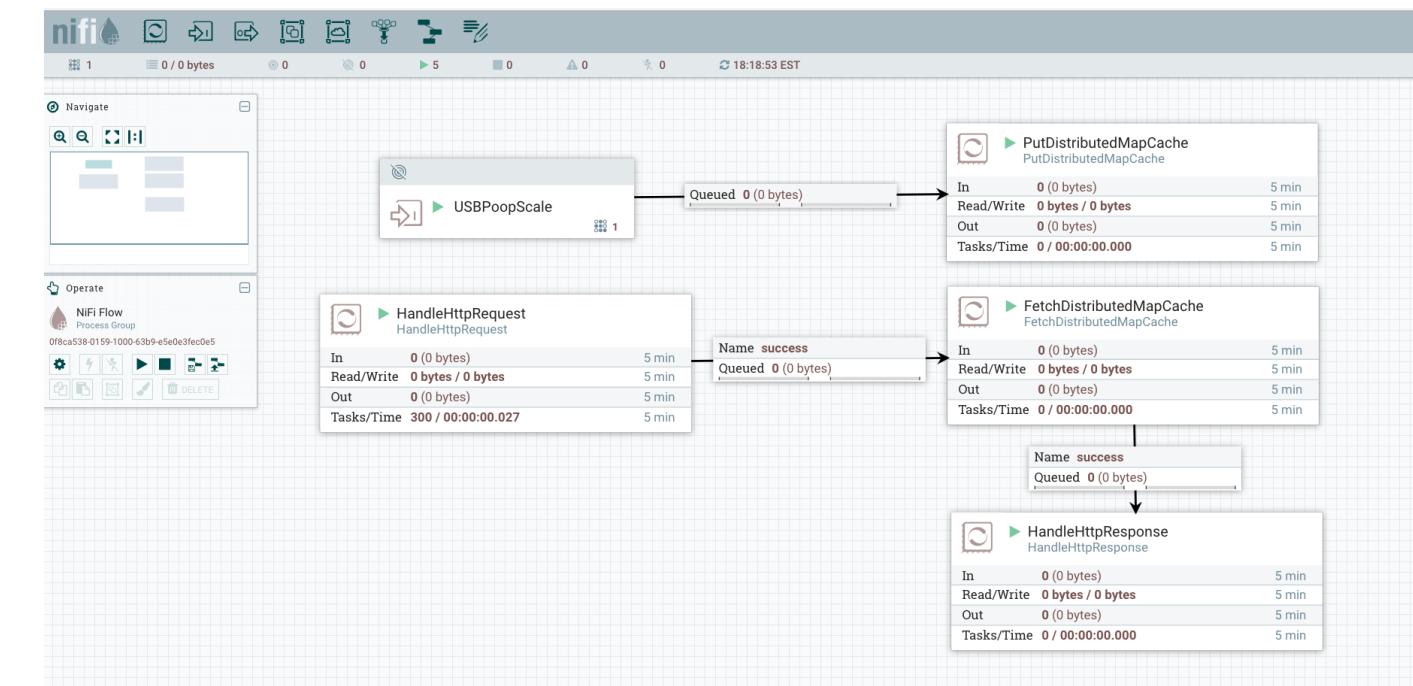
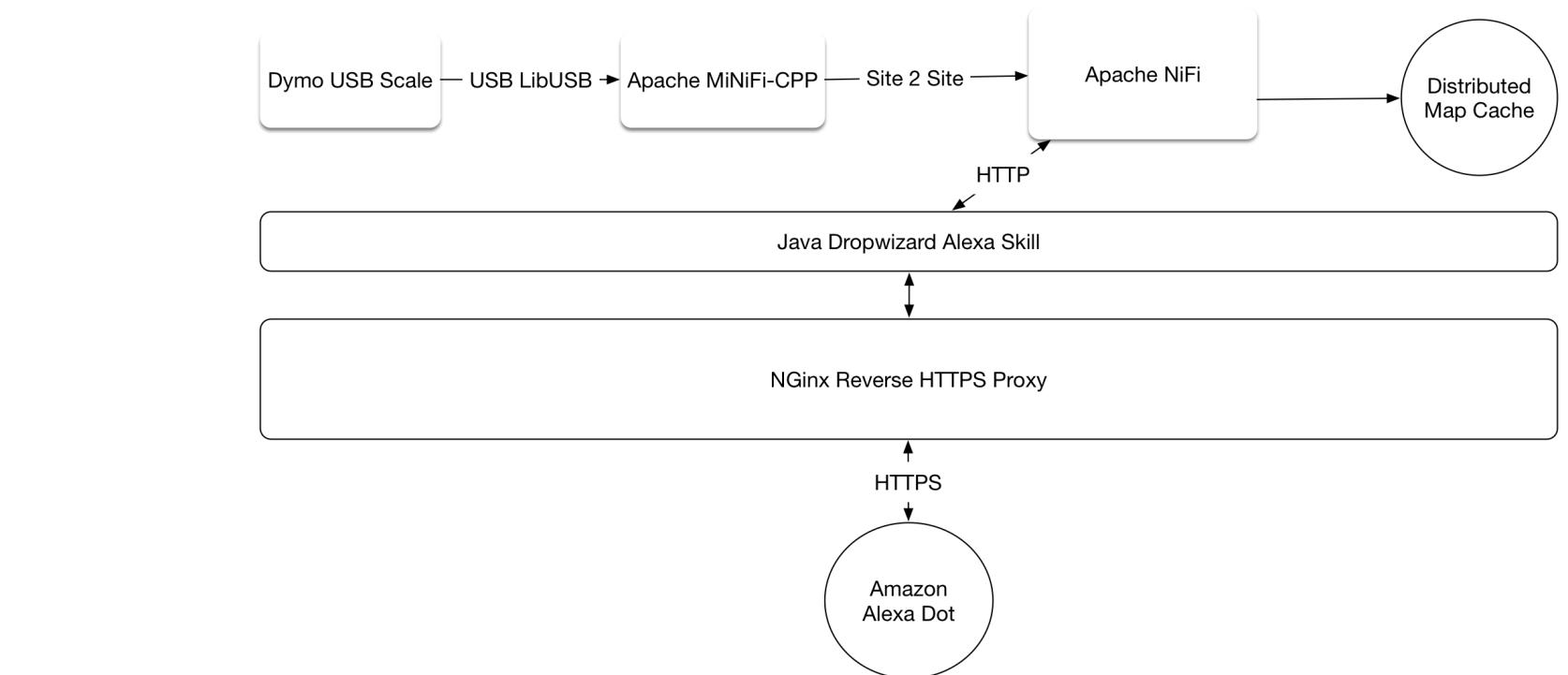
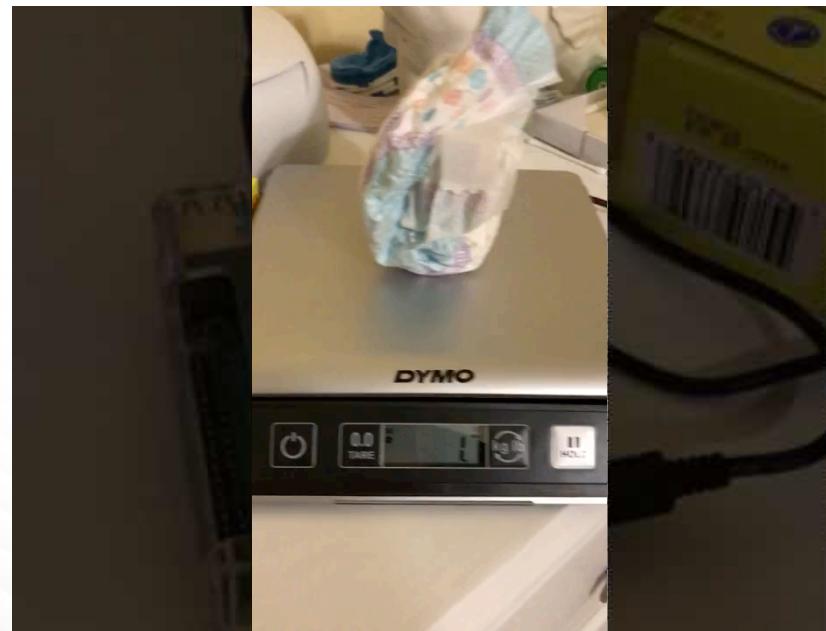
◆ Alexa + MiNiFi + Dyer 2.0



# Community Example

◆ Jeremy Dyer

◆ Alexa + MiNiFi + Dyer 2.0



A wide-angle photograph of a modern airport terminal. The ceiling is a massive glass and steel structure, allowing natural light to flood the space. Several escalators lead up to different levels. In the background, there are directional signs in Chinese and English, one pointing to "3号航站楼 4F" (Terminal 3, 4F) and another to "机场快轨" (Airport Express Train). A few people are visible on the escalators.

What's Next?

# New Announcements

## Introducing Apache NiFi Registry

- NiFi 1.6.0 — 08 April 2018
  - MongoDB, InfluxDB, Druid, HBase components
  - Granular @Restricted components
- MiNiFi C++ 0.4.0 — 27 January 2018
- MiNiFi Java 0.4.0 — 22 January 2018
- NiFi Registry 0.1.0 — 1 January 2018



# NiFi Registry for Dataflows

## Introducing Apache NiFi Registry 0.1.0

- 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

The screenshot shows the Apache NiFi Registry interface. At the top, there's a navigation bar with the title "NiFi Registry / All" and user information "registry\_user LOGOUT". Below the navigation bar, there's a search bar and a sorting dropdown set to "Sort by: Name (a - z)". The main content area displays two flows: "Flow 1 - Bucket 1" (1 version) and "Flow 2 - Bucket 2" (2 versions). For "Flow 2 - Bucket 2", the "DESCRIPTION" field contains "Description 2". The "CHANGE LOG" section shows three entries:

- Version 2 - 40 minutes ago by registry\_user (Details: Add processors, Dec-26-2017 at 11:23 PM)
- Version 1 - 41 minutes ago by registry\_user

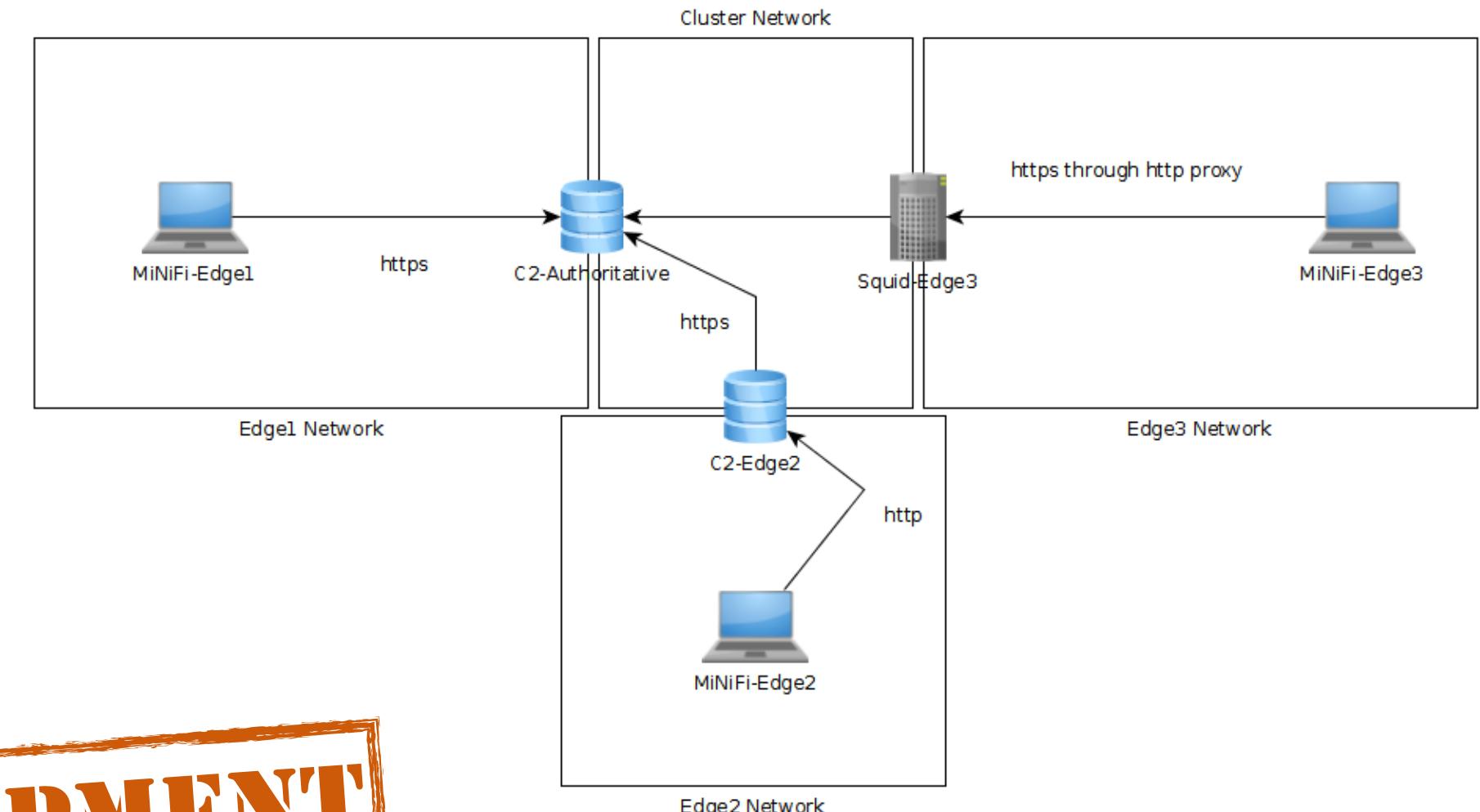
A "DESCRIPTION" field is also visible for "Flow 1 - Bucket 1". A "DESCRIPTION" field is also visible for "Flow 2 - Bucket 2". An "ACTIONS" button is located on the right side of the "Flow 2 - Bucket 2" row.

Learn more at [Forget Duplicating Local Changes: Apache NiFi and the Flow Development Lifecycle \(FDLC\)](#)  
Thursday 19/4 @ 1600, Room II

# MiNiFi C2 Server

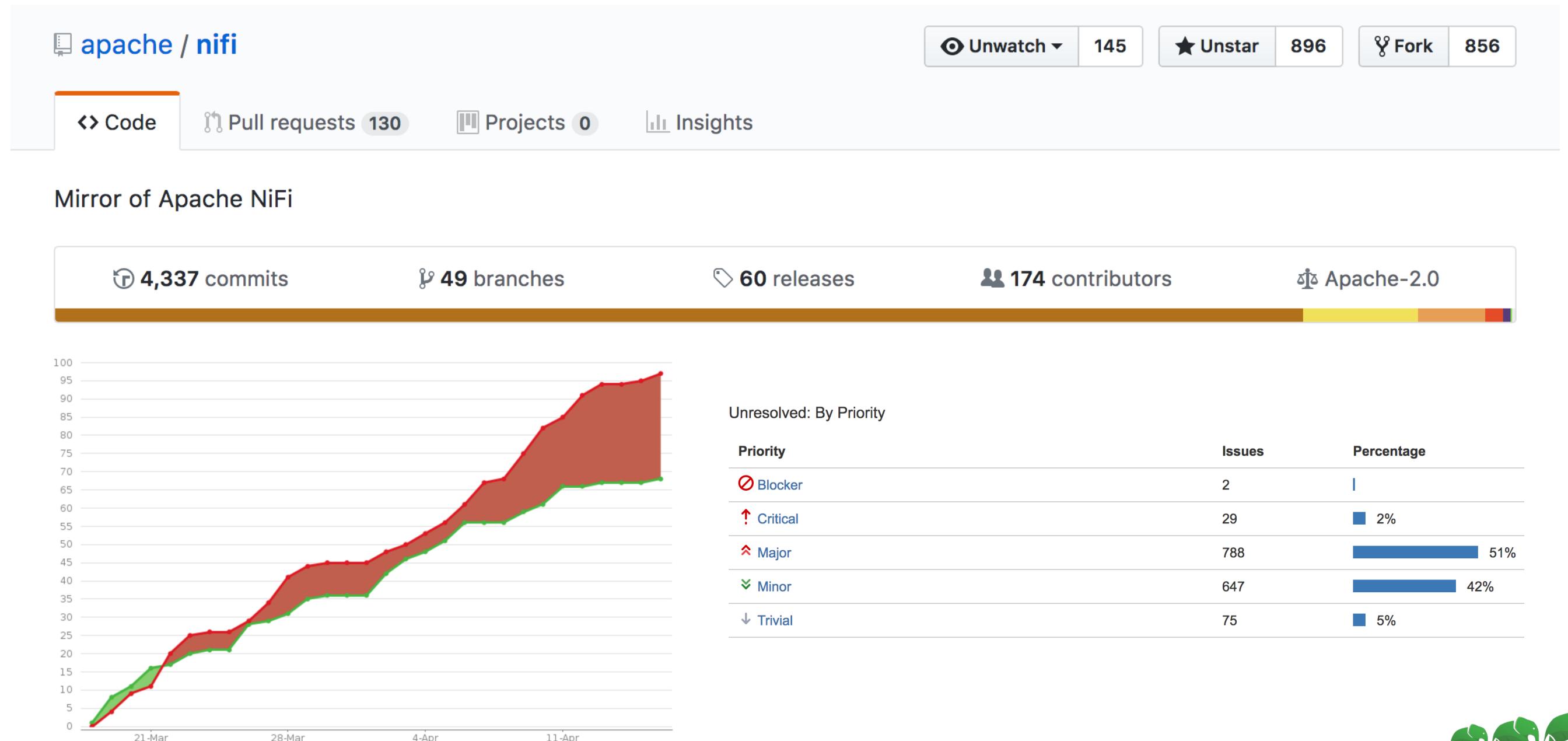
## Command & Control for MiNiFi

- Centralized C2 for **classes** of agents
- Tag devices, platforms, capabilities
- Distribute flows to edge via multiple mechanisms



IN DEVELOPMENT

# Community Health



# 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](mailto:dev@nifi.apache.org)

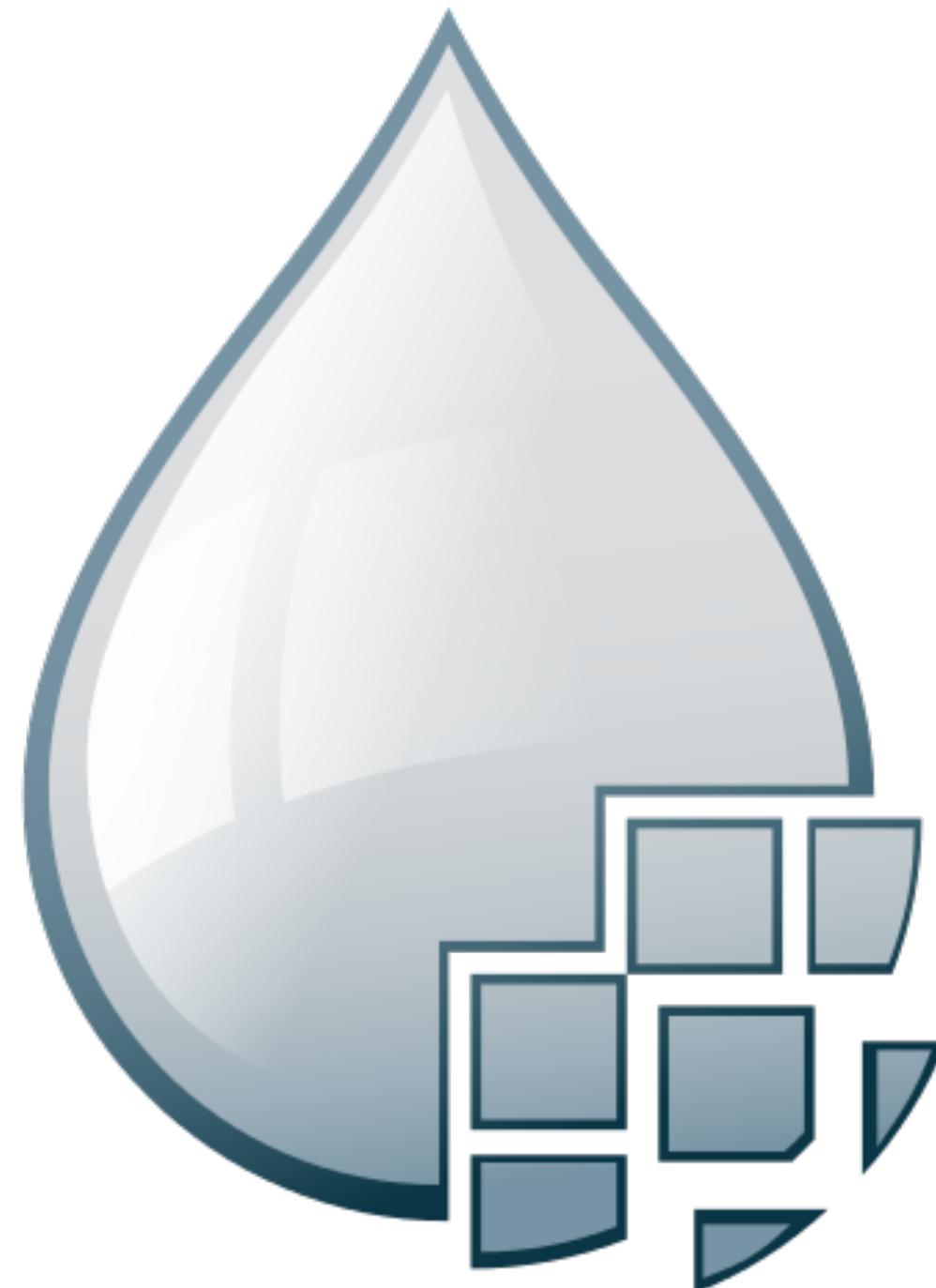
[users@nifi.apache.org](mailto:users@nifi.apache.org)

**Submit Ideas or Issues**

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

**Follow us on Twitter**

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



# More NiFi This Week...

Title	Room	Time	Speaker(s)
Apache NiFi Crash Course	Hall I - D	1115 - 1345	Andy LoPresto, Tim Spann
IoT with Apache MXNet and Apache NiFi and MiNiFi	Hall I - C	1150 - 1230	Tim Spann
Best practices and lessons learnt from Running Apache NiFi at Renault	Europe	1650 - 1730	Adel Gacem, Abdelkrim Hadjidj
From an experiment to a real production environment	Room V	1650 - 1730	Jeroen Wolffensperger, Martijn Groen
IoT, Streaming, and Dataflow Birds of a Feather	Room I	1740 - 1855	George Vetticaden, Davor Bonaci, Andy LoPresto, Stephan Ewen
Intelligently Collecting Data at the Edge — Intro to Apache MiNiFi	Room II	1100 - 1140	Andy LoPresto
The Power of Intelligent Flows: Realtime IoT Botnet Classification with Apache NiFi	Hall I - C	1400 - 1440	Andy LoPresto
Forget Duplicating Local Changes: Apache NiFi and the Flow Development Lifecycle (FDLC)	Room II	1600 - 1640	Andy LoPresto



# Thank you

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