

Intelligently Collecting Data at the Edge

Andy LoPresto - [@yolopey](#)

Dataflow with Apache NiFi/MiNiFi

Apache NiFi PMC

DataWorks Summit 2017 - Sydney

21 Sep 2017



Agenda

What is dataflow and what are the challenges?

Apache NiFi

IoT Challenges

Apache MiNiFi

Exploration

Community

Agenda



What is dataflow and what are the challenges?

Apache NiFi

IoT Challenges

Apache MiNiFi

Exploration

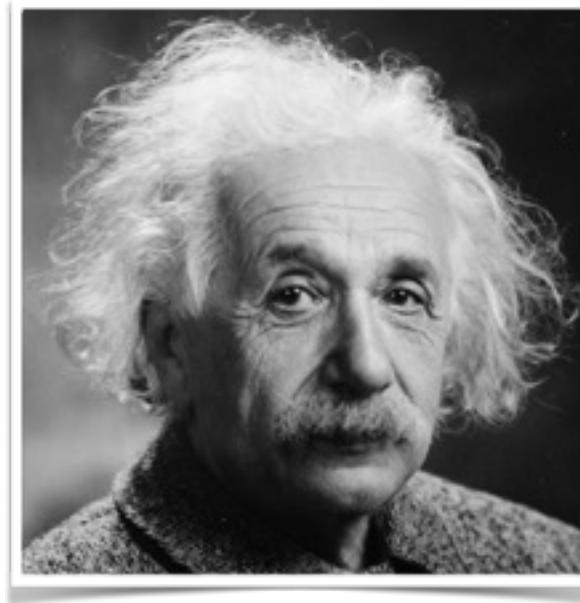
Community

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

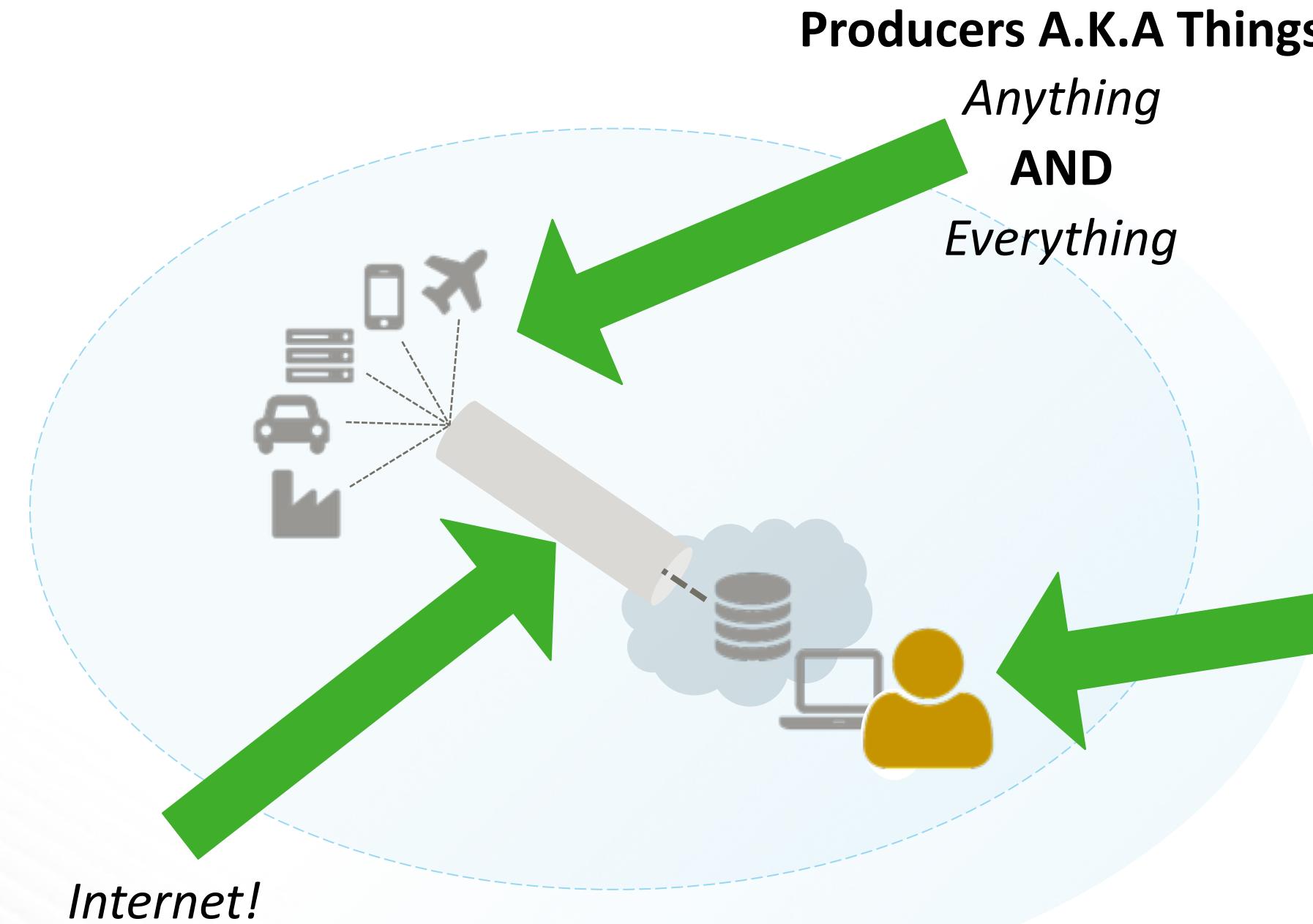


"I refactored the Ambari integration endpoint to allow for mutual authentication TLS during my coffee break"

Forgotten more about NiFi
than most of us will ever
know



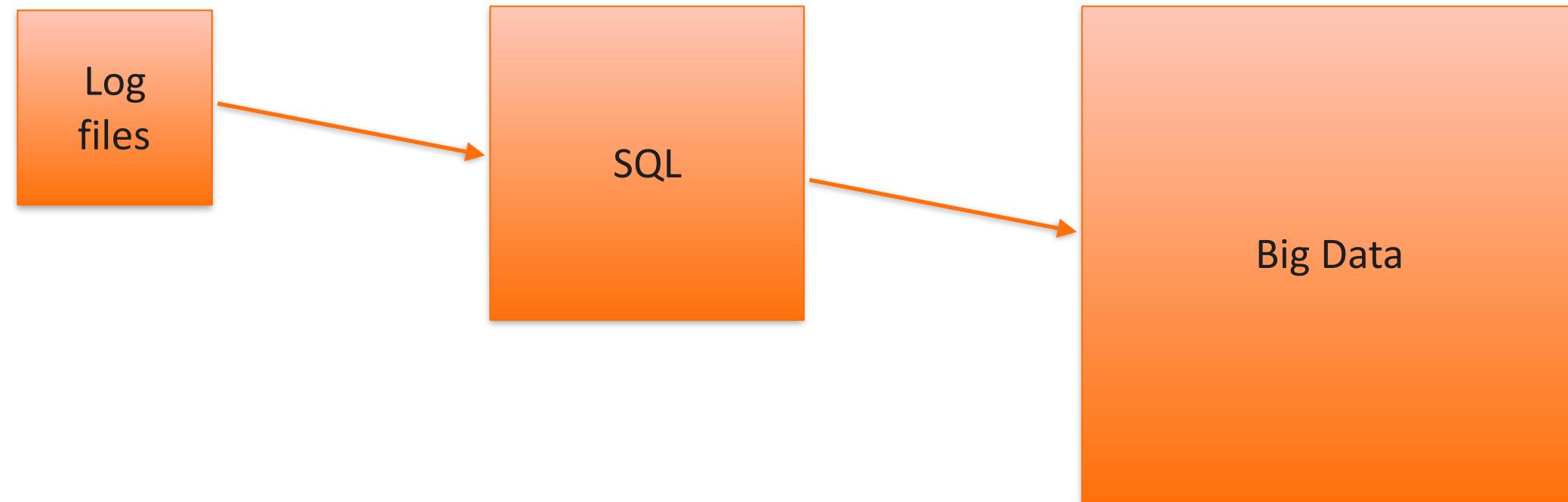
Let's Connect A to B



- Consumers**
- User
 - Storage
 - System
 - ...**More Things**

Connecting A to B to C

Easy enough with Bash scripts, Ruby/Python/Groovy, etc.

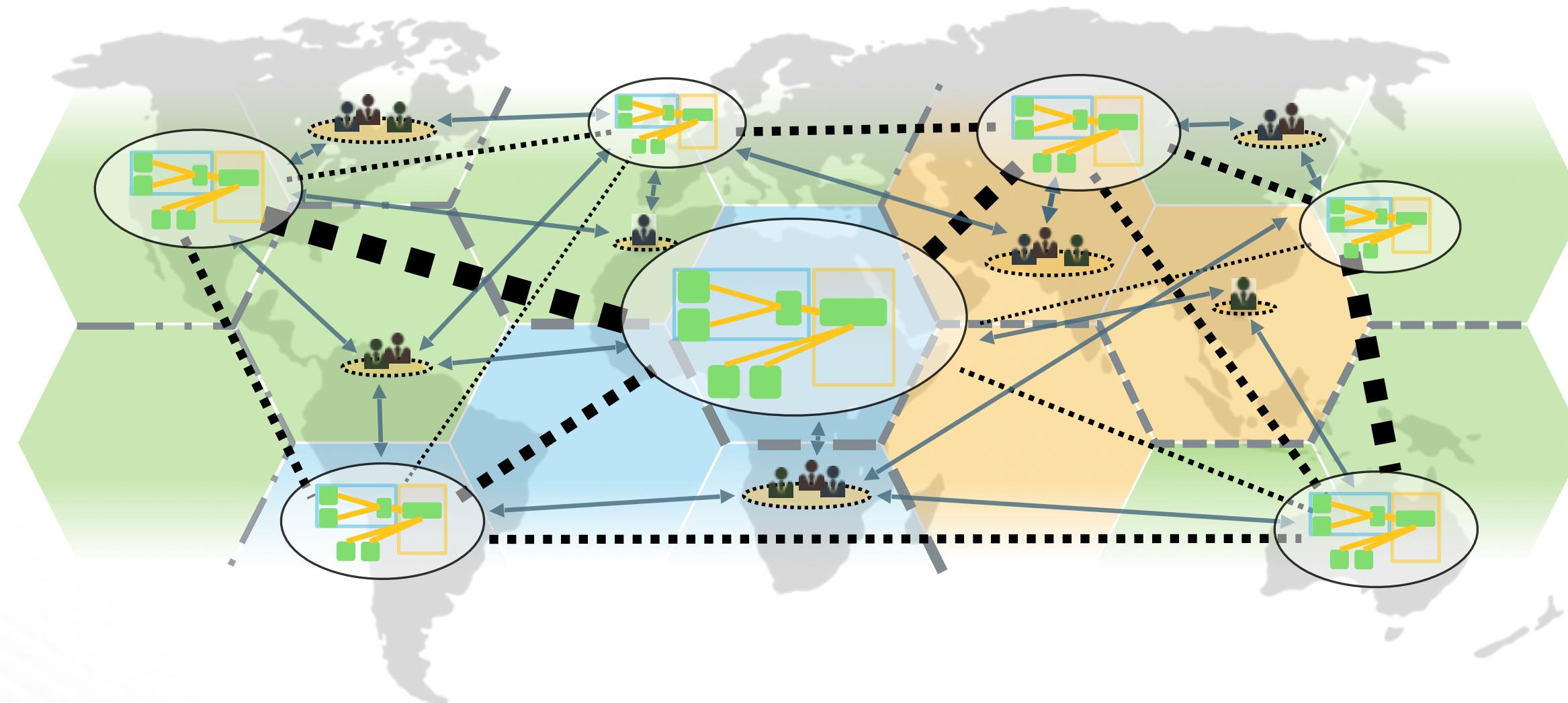


Why is moving data *effectively* hard?

- ◆ Standards
- ◆ Formats
- ◆ “Exactly Once” Delivery
- ◆ Protocols
- ◆ Veracity of Information
- ◆ Validity of Information
- ◆ Ensuring Security
- ◆ Overcoming Security
- ◆ Compliance
- ◆ Schemas
- ◆ Consumers Change
- ◆ Credential Management
- ◆ “*That [person | team | group]*”
- ◆ Network*
- ◆ “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



Agenda



What is dataflow and what are the challenges?

Apache NiFi

IoT Challenges

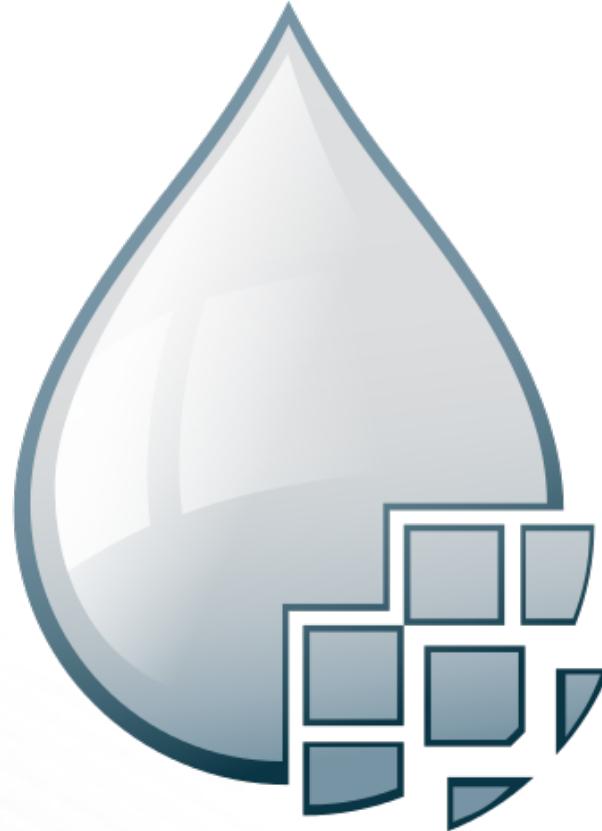
Apache MiNiFi

Exploration

Community

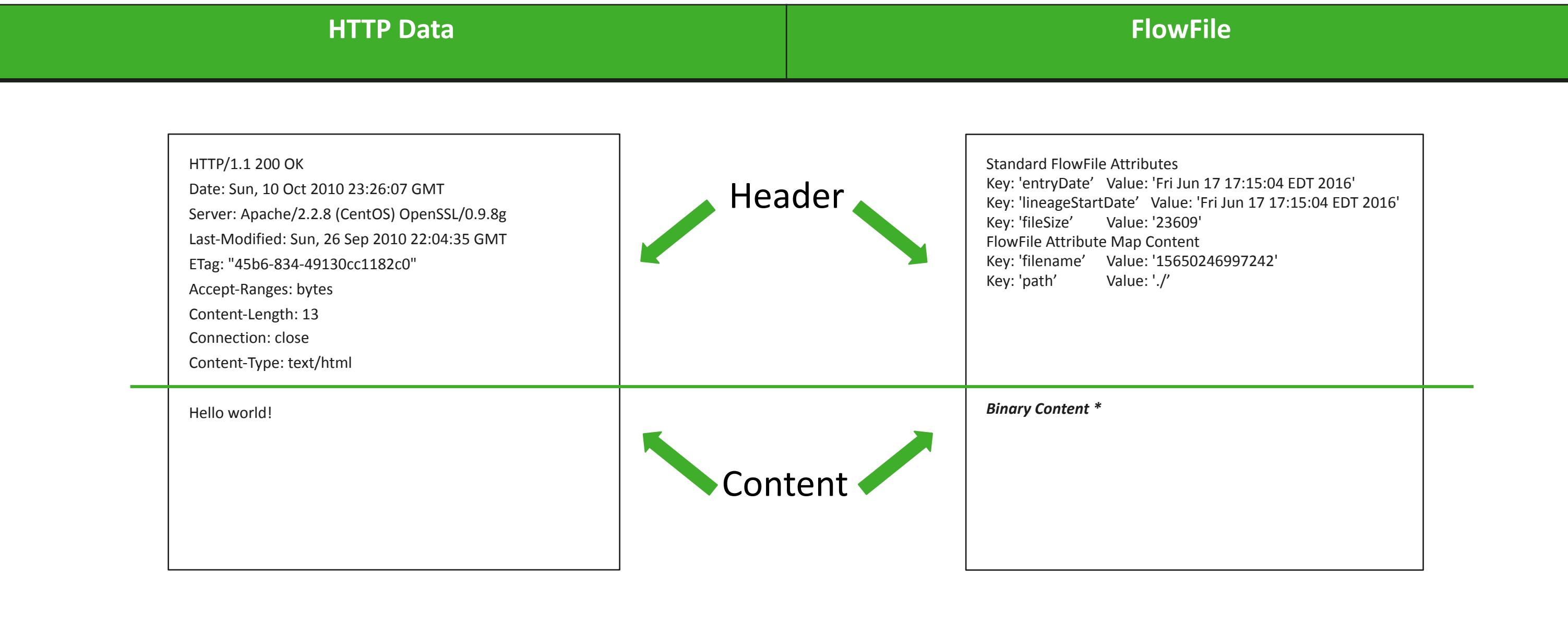
Apache NiFi

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

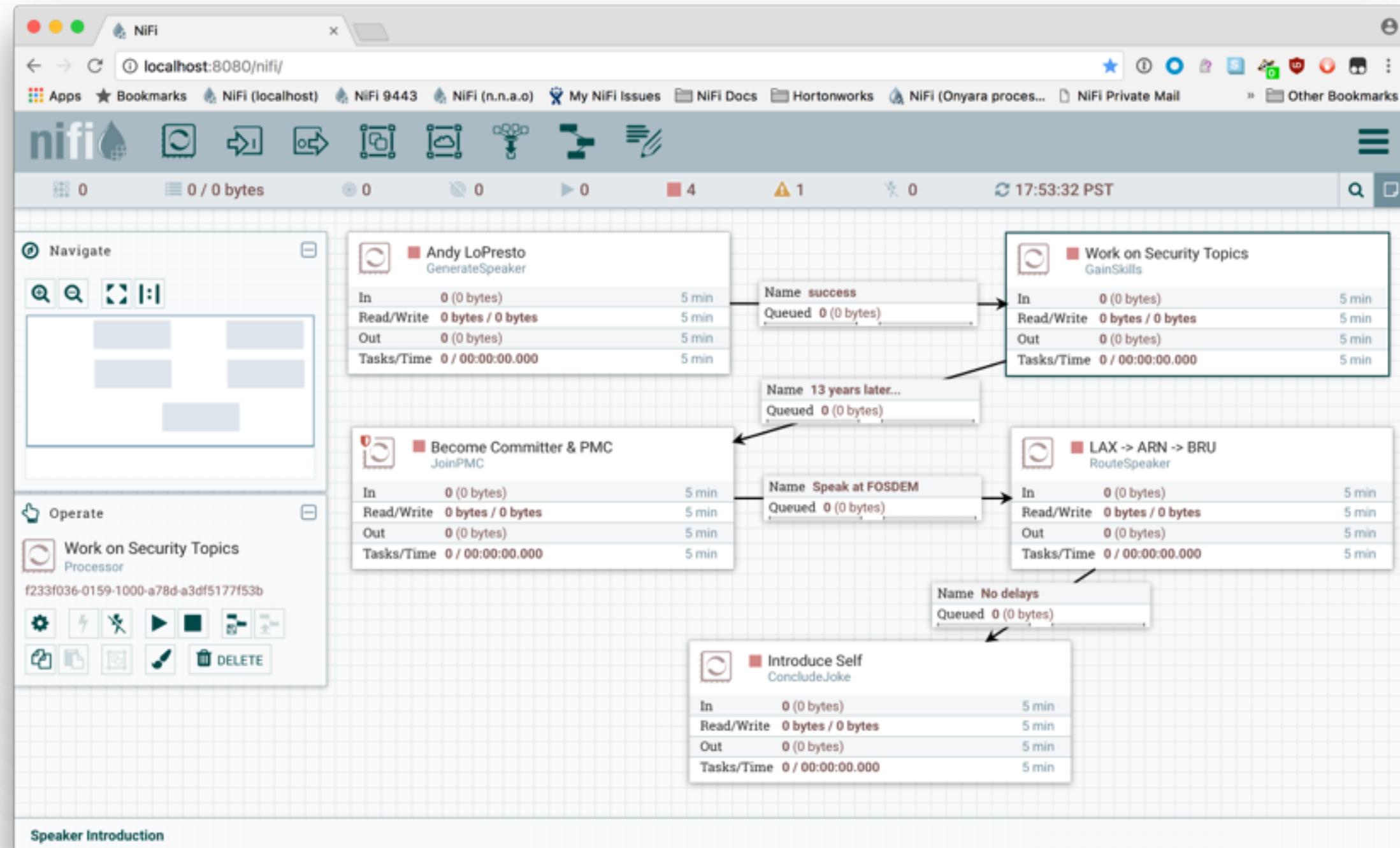


User Interface

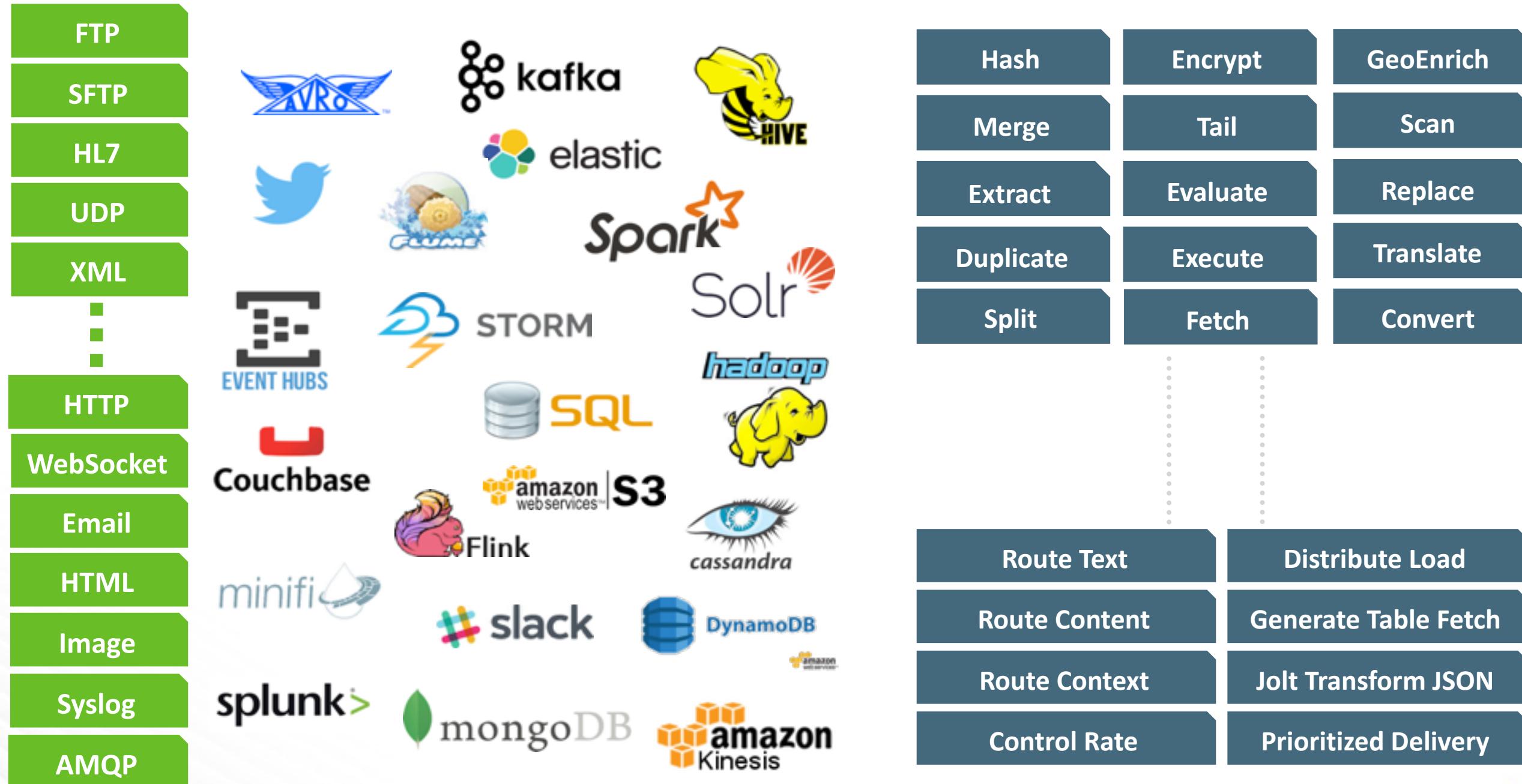
Less of this... ... more of this



User Interface



Deeper Ecosystem Integration: 220+ Processors



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

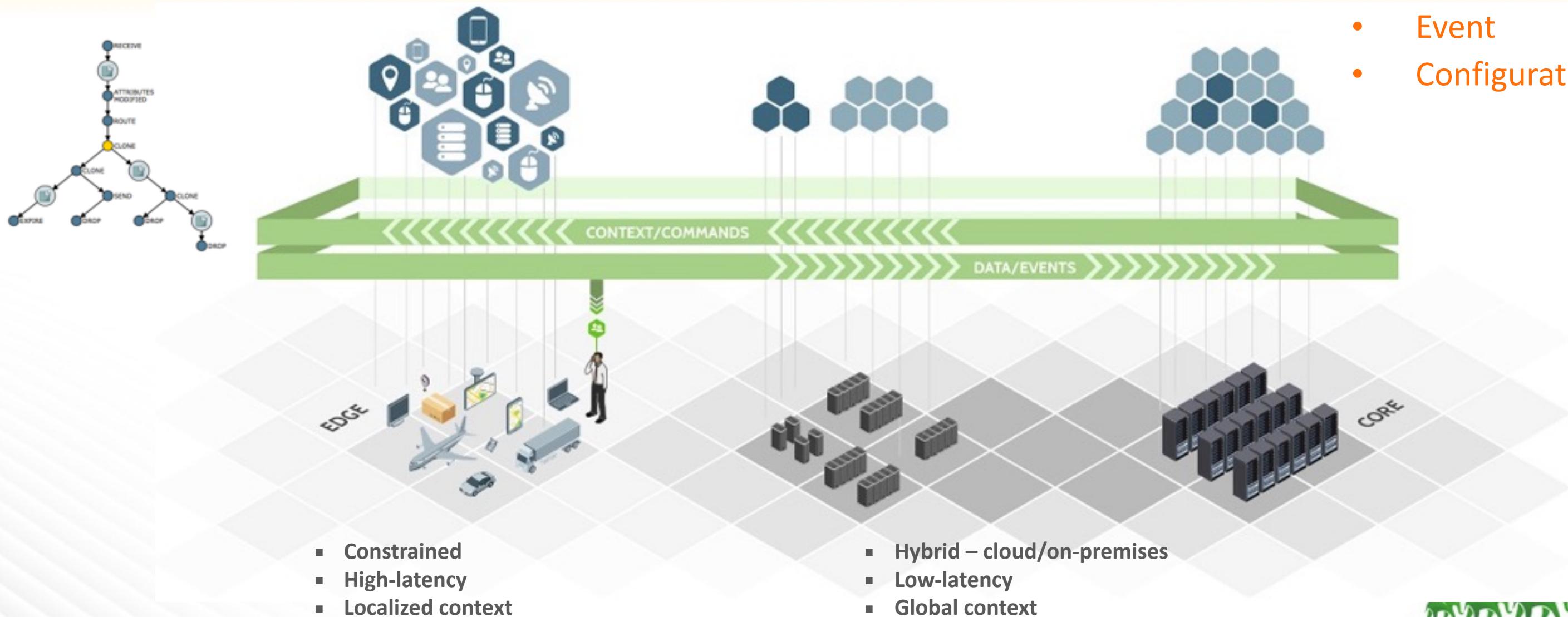
Data Provenance

Origin – attribution
Replay – recovery

Evolution of topologies
Long retention

Types of Lineage

- Event
- Configuration



Agenda



What is dataflow and what are the challenges?

Apache NiFi

IoT Challenges

Apache MiNiFi

Exploration

Community

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

Dyn Analysis Summary Of Friday October 21 Attack
Company News // Oct 26, 2016 // Scott Hilton

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}

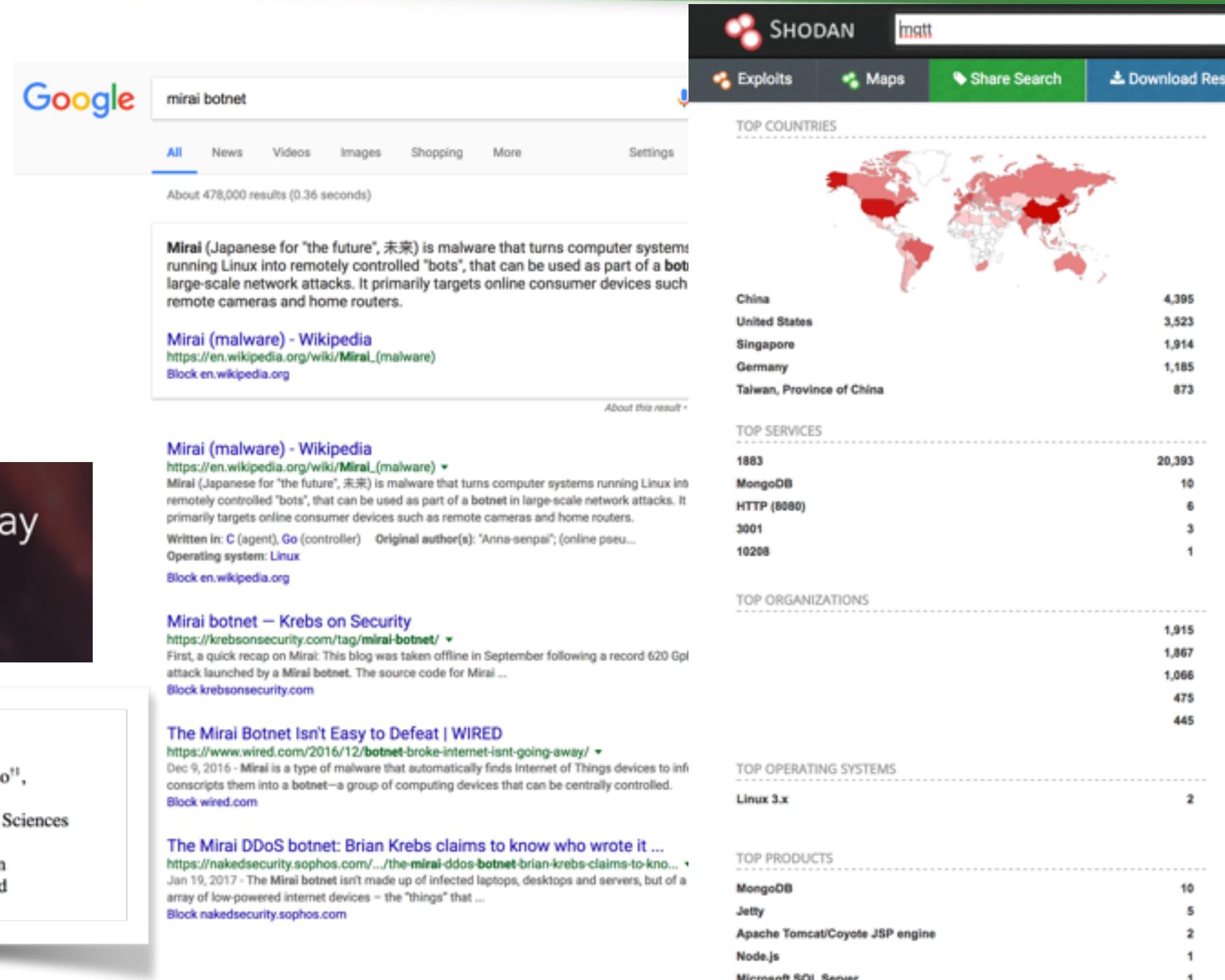
^{†1}Graduate School of Environment and Information Sciences/Institute of Advanced Sciences

^{†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



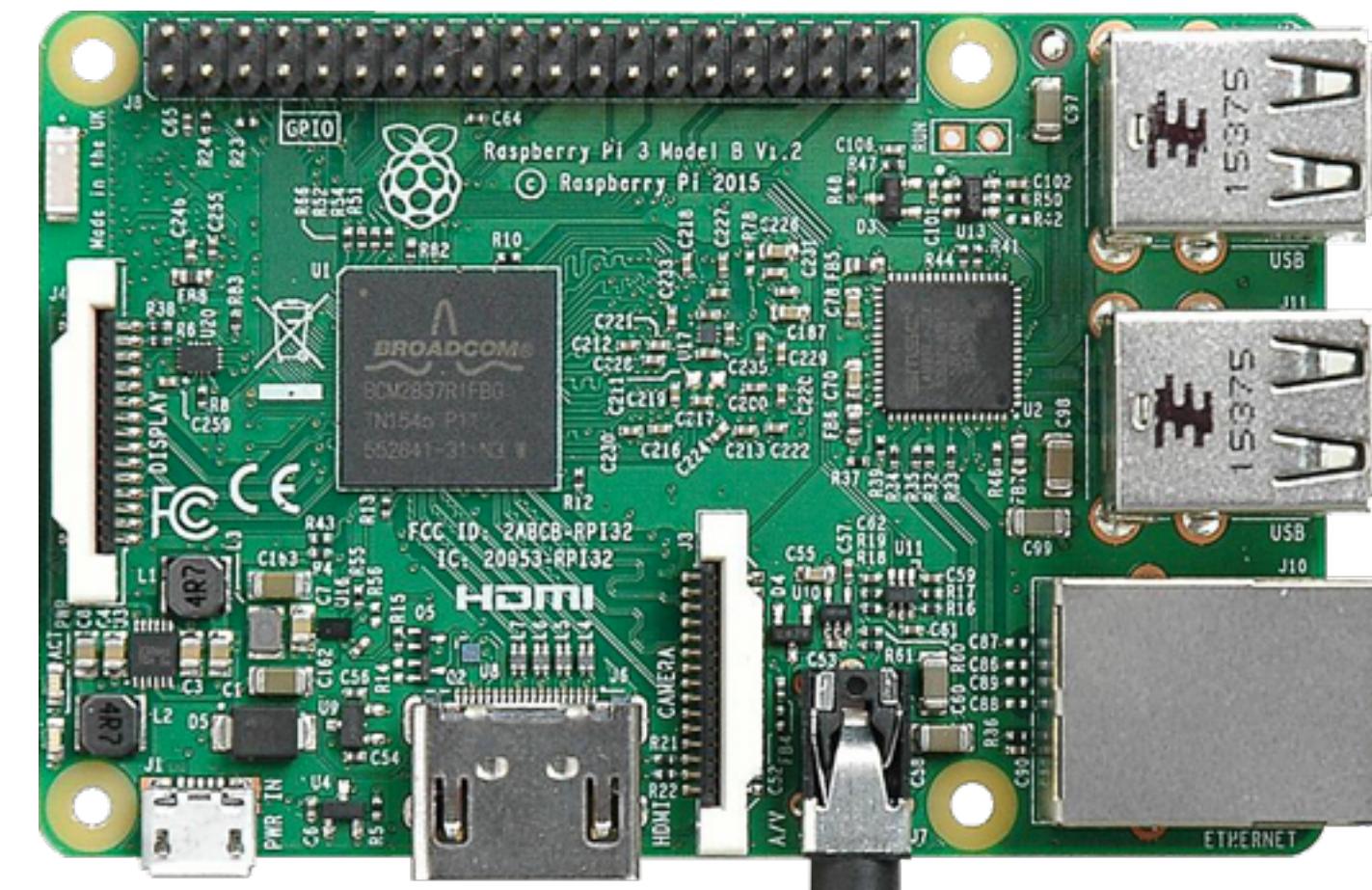
NiFi Solves Everything*

- ◆ Runs on JVM
- ◆ Provides UI for flow design & monitoring
- ◆ Security built-in
 - ◆ TLS, authn/authz, 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 $ █
```

Agenda



- What is dataflow and what are the challenges?*
- Apache NiFi*
- IoT Challenges
- Apache MiNiFi
- Exploration
- Community

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

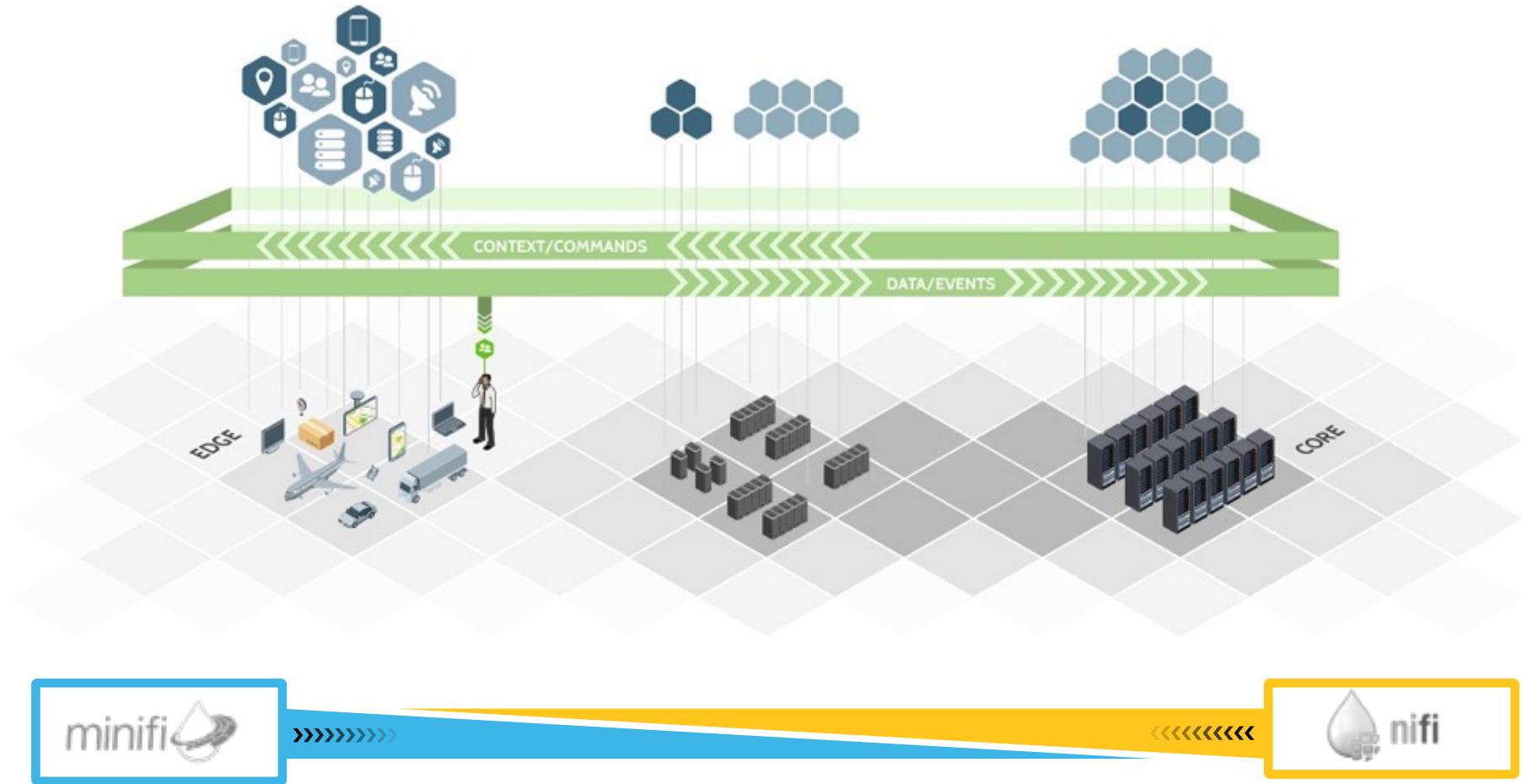


Why build MiNiFi?

- ◆ NiFi is big
 - ◆ 1.3.0 release is 933 MB 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
 - ◆ 0.2.0 Java binary is 59 MB, C++ binary is 1193 KB
 - ◆ “Good guest”

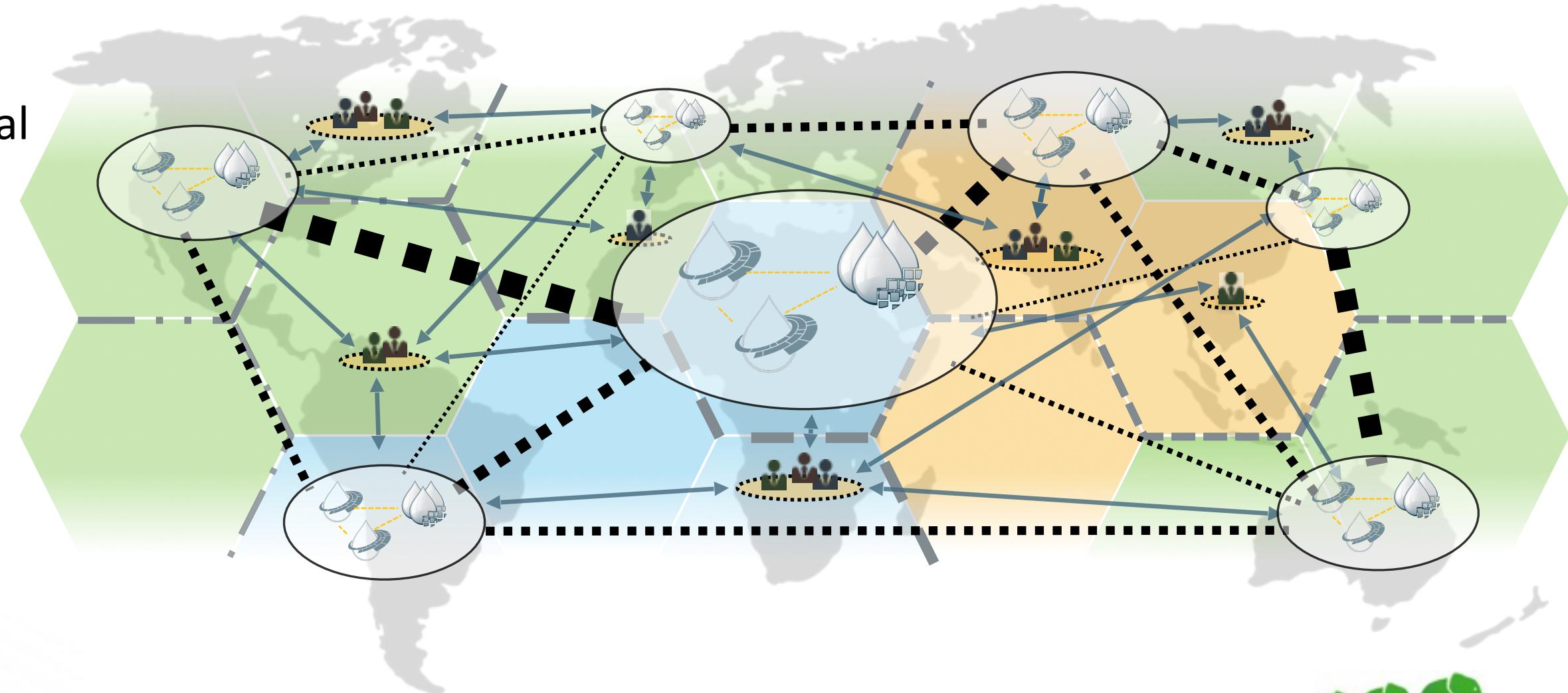
How Does MiNiFi Interact With NiFi?

- ◆ 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



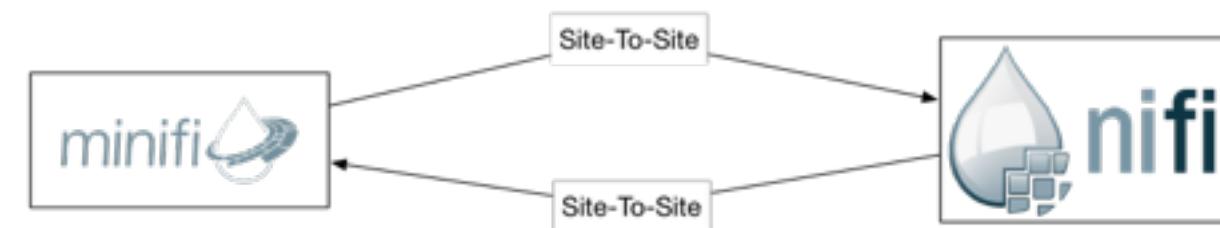
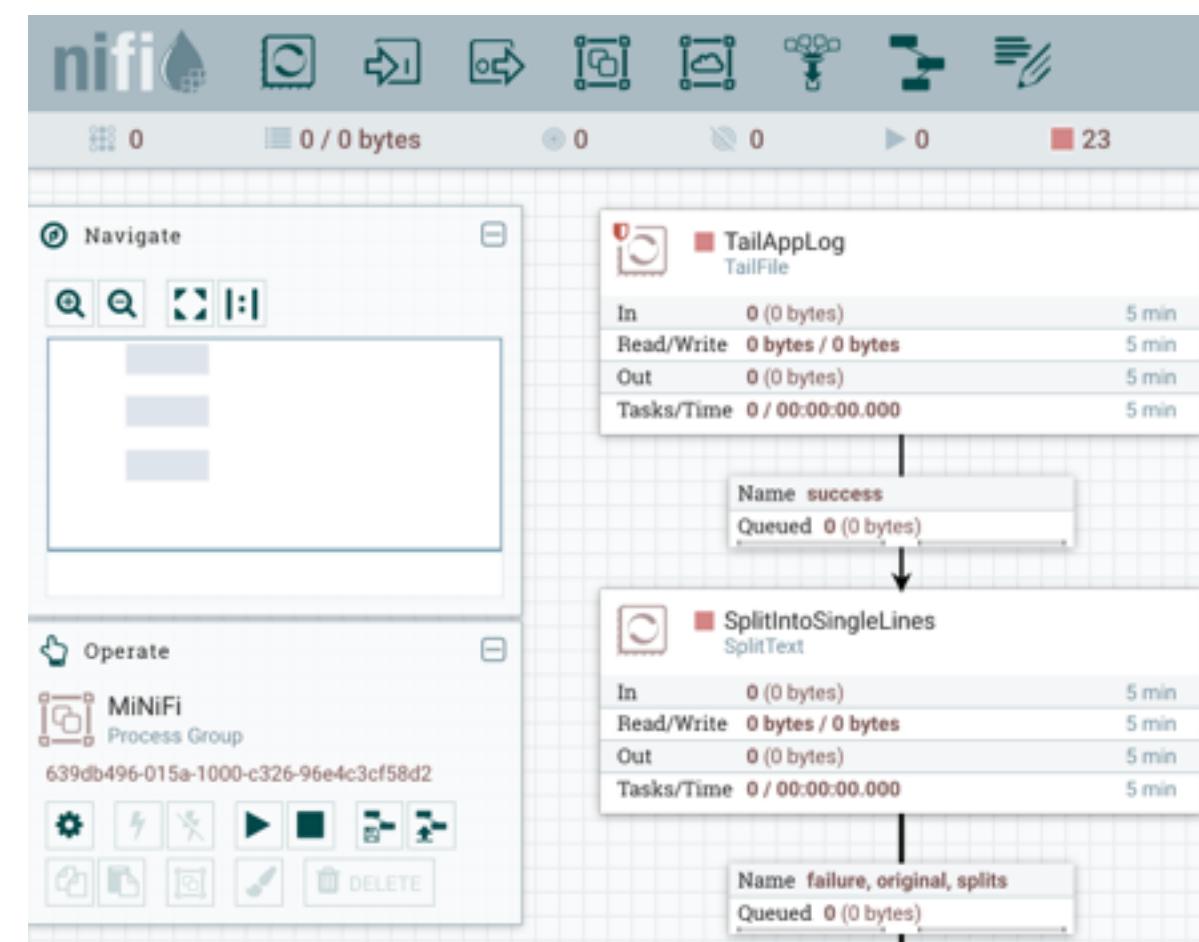
Flavors of MiNiFi

◆ MiNiFi Java (v0.2.0)

- ◆ Modified version of NiFi
 - ◆ No UI
 - ◆ YAML configuration
- ◆ Reduced processor count
 - ◆ 110 by default, more available with additional NARs

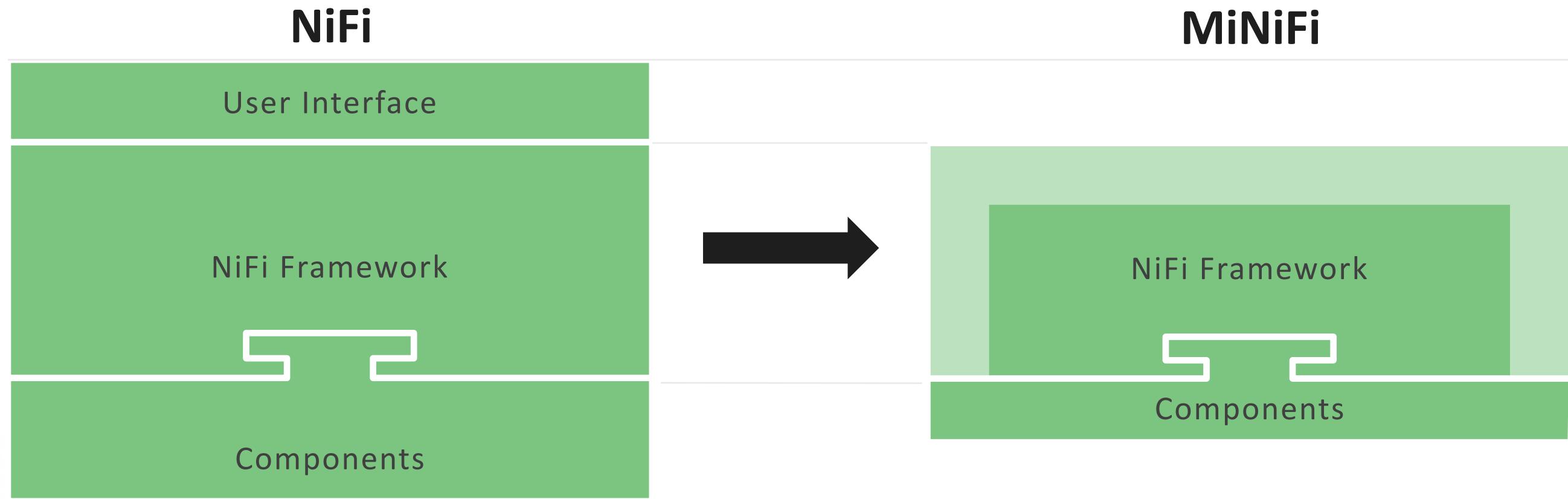
◆ MiNiFi C++ (v0.2.0)

- ◆ Written from scratch
- ◆ 10 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
```

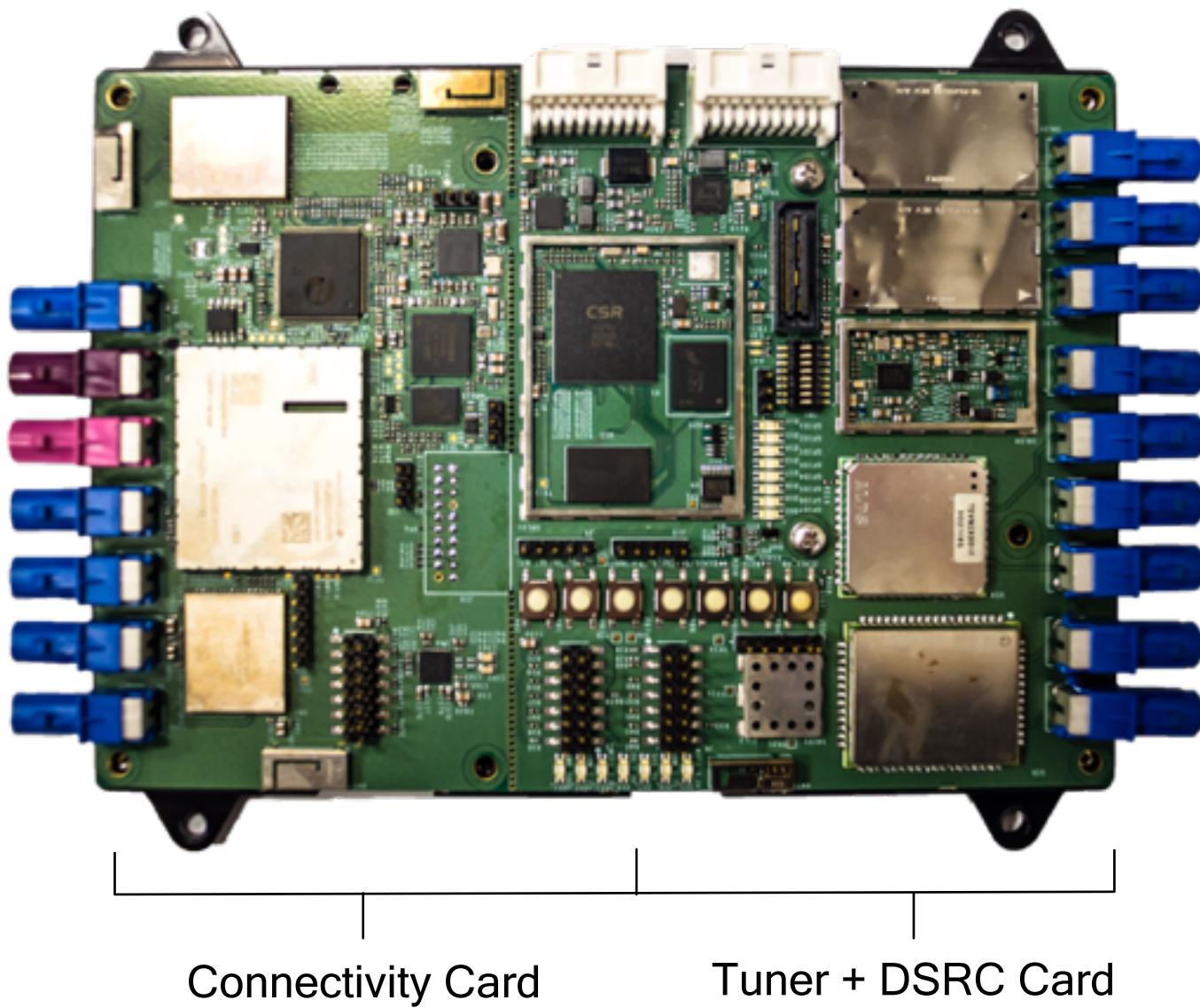
NiFi vs MiNiFi Java Processes



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 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*)

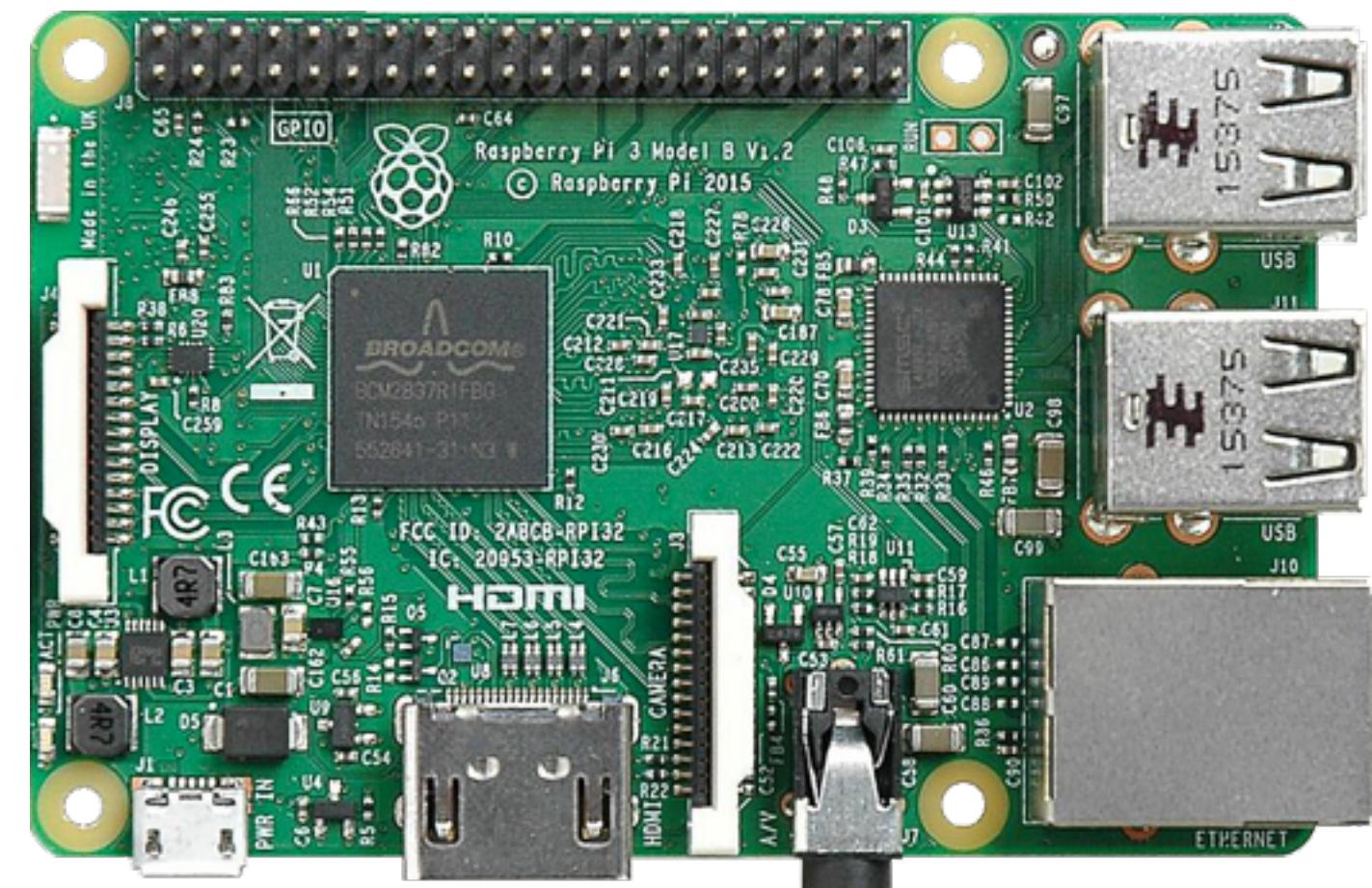
Agenda



- What is dataflow and what are the challenges?*
- Apache NiFi*
- IoT Challenges
- Apache MiNiFi
- Exploration
- Community

Scenario

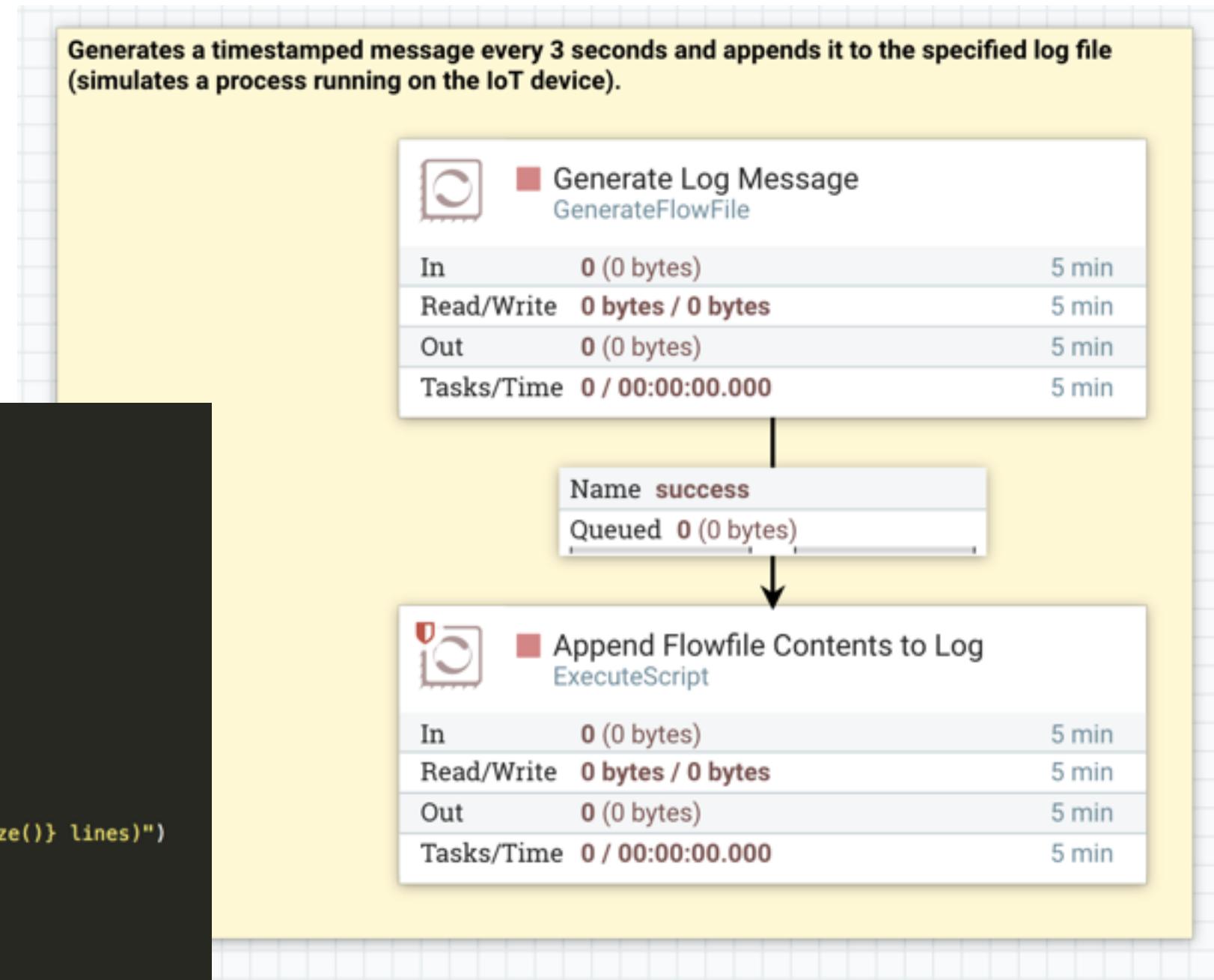
- IoT Device generating log messages
- Need to encrypt data on device
- Need to prioritize some data for unreliable network connectivity
- Transmit data to central node
- Decrypt data and analyze
- Make determinations and modify live flow*



NiFi As Test Harness/Environment

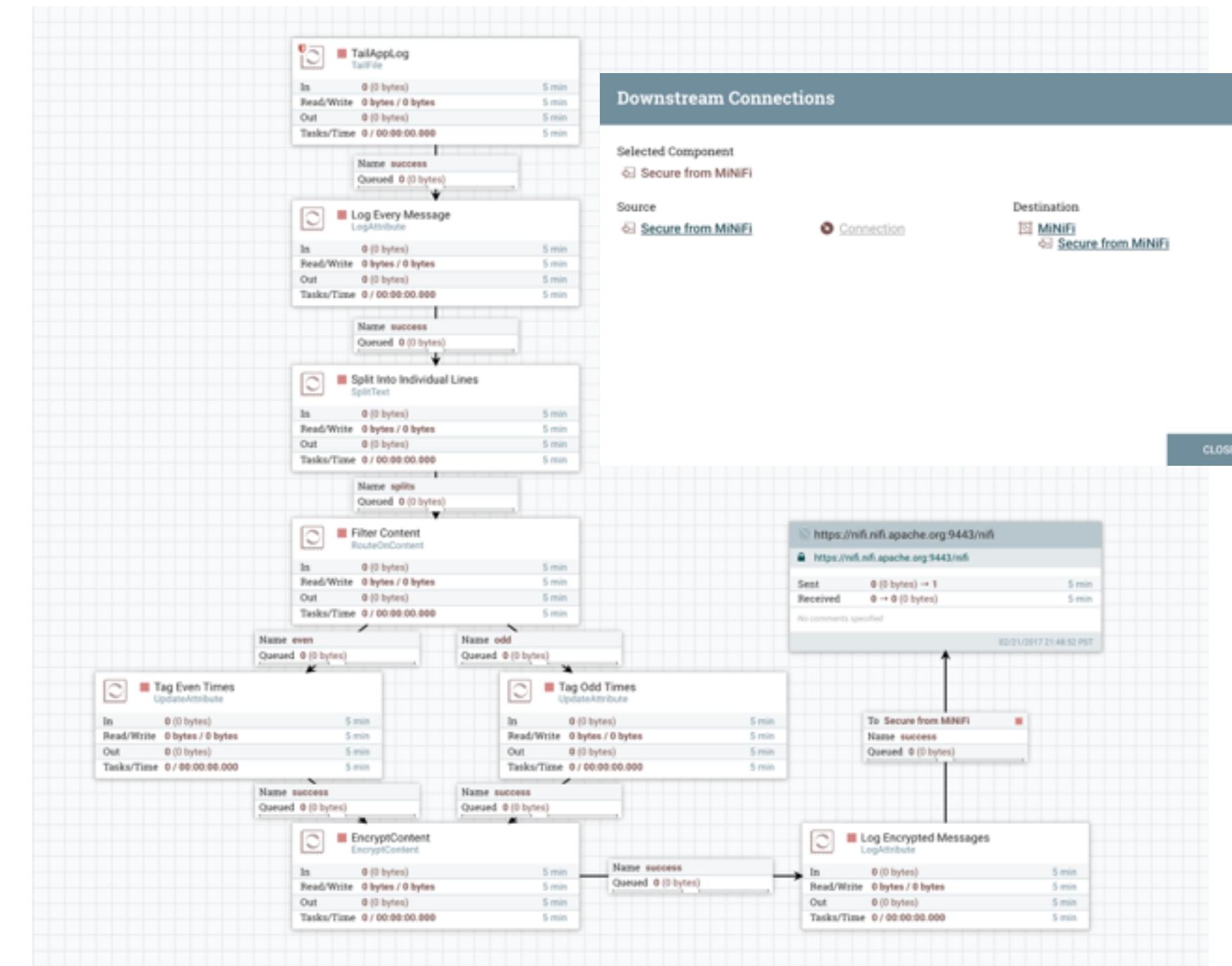
- Simulate the log generation
- Schedule is customizable
- Script can write to dynamic location

```
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*

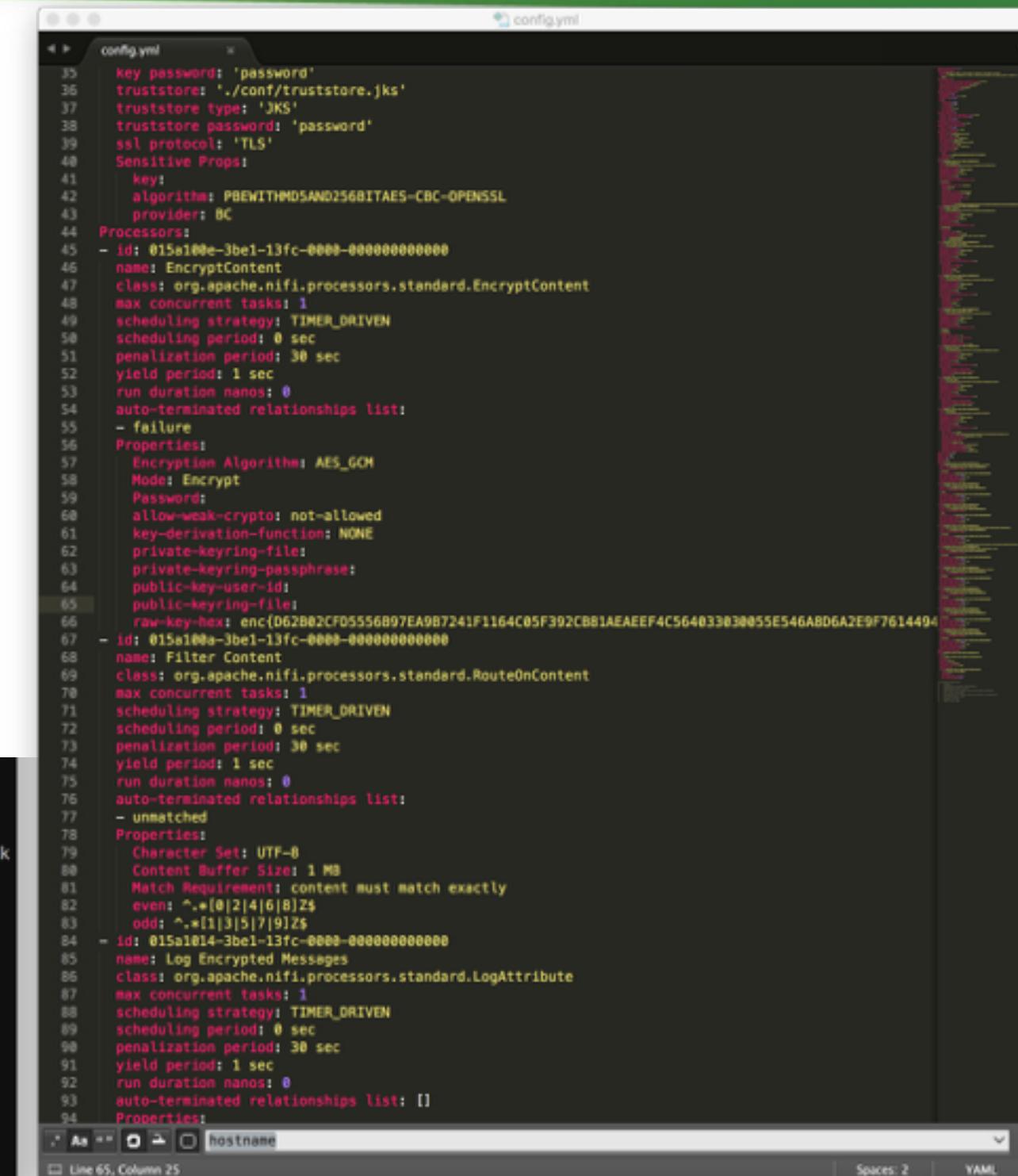
*Still need to set up TLS & encrypted properties

```
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.

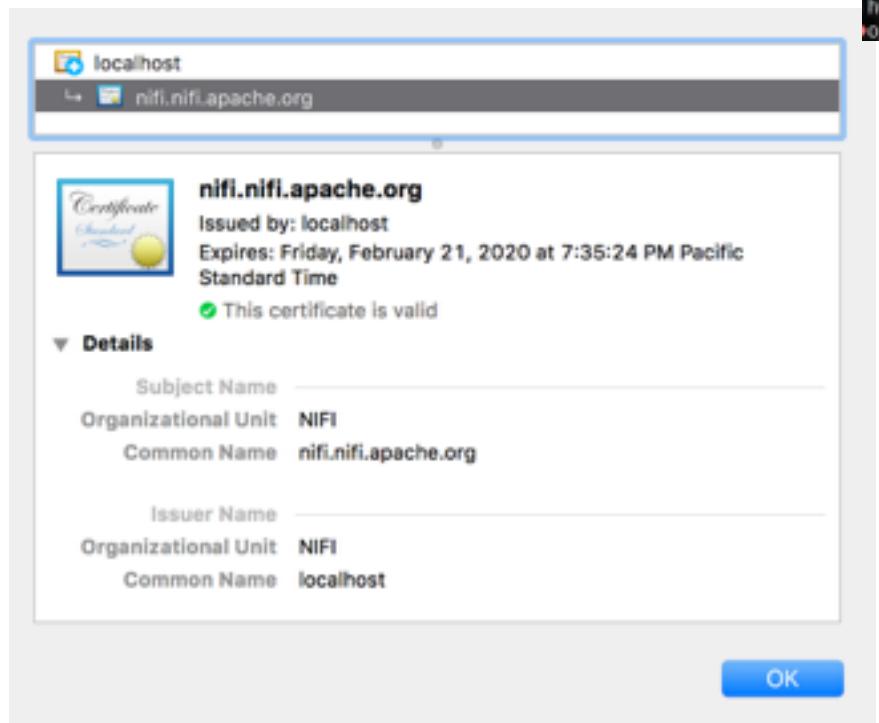
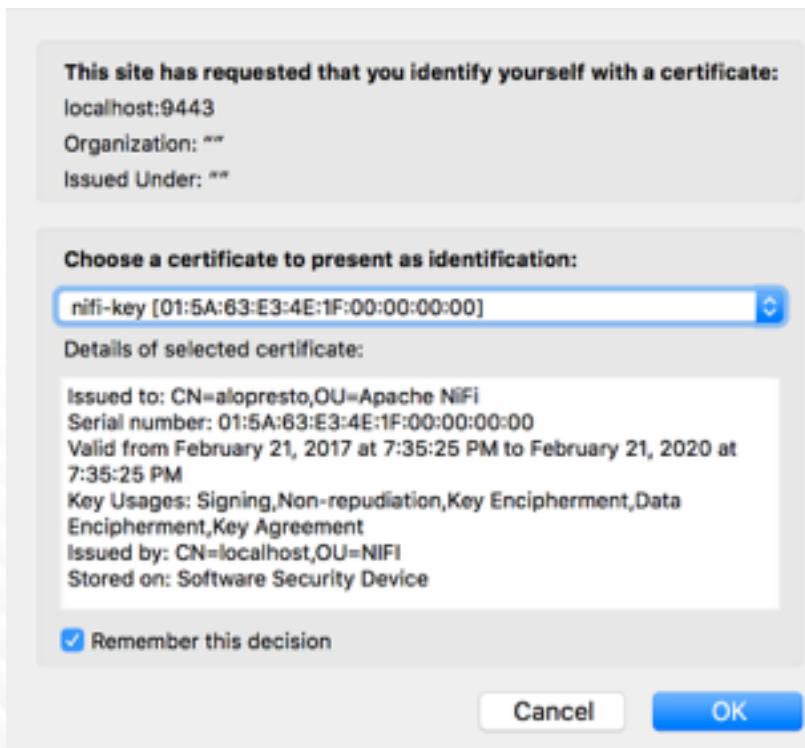
hw12203:...inifi-toolkit-assembly/target/minifi-toolkit-0.1.0-bin/minifi-toolkit-0.1.0 (master) alopresto
4ls @ 21:55:37 $ ll
total 104
drwxr-xr-x  8 alopresto  staff  272B Feb 21 21:55 .
drwxr-xr-x  3 alopresto  staff  102B Dec  1 11:13 ..
-rw-r--r--  1 alopresto  staff  18K Nov 30 22:55 LICENSE
-rw-r--r--  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
key password: 'password'
truststore: './conf/truststore.jks'
truststore type: 'JKS'
truststore password: 'password'
ssl protocol: 'TLS'
Sensitive Props:
  key:
    algorithm: PBEWITHMD5AND256BITAES-CBC-OPENSSL
    provider: BC
Processors:
  - id: 015a100e-3be1-13fc-0000-000000000000
    name: EncryptContent
    class: org.apache.nifi.processors.standard.EncryptContent
    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
    Properties:
      Encryption Algorithm: AES_GCM
      Mode: Encrypt
      Password:
        allow-weak-crypto: not-allowed
        key-derivation-function: NONE
        private-keyring-file:
        private-keyring-passphrase:
        public-key-user-id:
        public-keyring-file:
        raw-key-hex: enc(D62B02CFD5556897EA9B7241F1164C05F392C881AEAEFF4C564B33030055E546A8D6A2E9F7614494
  - id: 015a100a-3be1-13fc-0000-000000000000
    name: Filter Content
    class: org.apache.nifi.processors.standard.RouteOnContent
    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:
      - unmatched
    Properties:
      Character Set: UTF-8
      Content Buffer Size: 1 MB
      Match Requirement: content must match exactly
      even: ^.[0|2|4|6|8]$
      odd: ^.[1|3|5|7|9]$
  - id: 015a1014-3be1-13fc-0000-000000000000
    name: Log Encrypted Messages
    class: org.apache.nifi.processors.standard.LogAttribute
    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: []
    Properties:
```

Setting Up Crypto

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



```
hw12203:...assembly/target/nifi-toolkit-1.2.0-SNAPSHOT-bin/nifi-toolkit-1.2.0-SNAPSHOT (NIFI-3486-RC1) all
+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 -t ../../../../../../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: Using ../../../../../../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 standalone certificate generation with output directory ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf
2017/02/21 19:35:24 INFO [main] org.apache.nifi.toolkit.tls.standalone.TlsToolkitStandalone: Generated new CA certificate ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/nifi-cert.pem and key ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/nifi-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 new client certificate ../../../../../../nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/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-SNAPSHOT/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) all
+opresto
```

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).



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

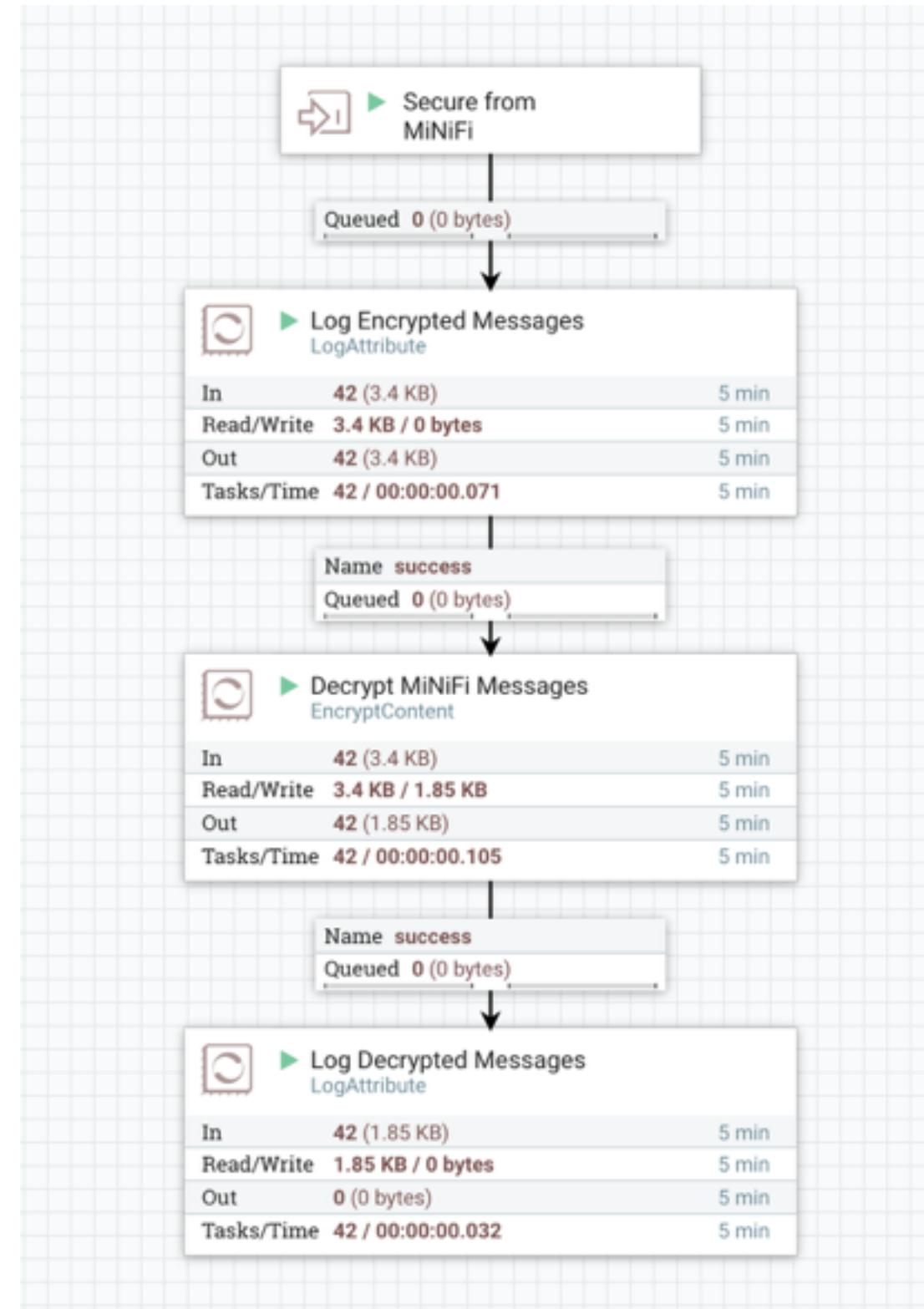
Property	Value
Mode	Encrypt
Key Derivation Function	None
Encryption Algorithm	AES_GCM
Allow insecure cryptographic modes	Not Allowed
Password	No value set
Raw Key (hexadecimal)	Sensitive value set
Public Keyring File	In keyed encryption, this is the raw key, encoded in hexadecimal
Public Key User Id	Supports expression language: false
Private Keyring File	No value set
Private Keyring Passphrase	No value set

In keyed encryption, this is the raw key, encoded in hexadecimal
Supports expression language: false
History:
• ***** - 02/21/2017 20:30:46 PST (CN=alopresto, OU=Apache NiFi)
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?

View as: original

1 脚業レ複口擊? 梶棲敵號口留口惹+ 菜體邝蛙姪既□□□□型范塔∠ 調擊環啓鈞口體璇斐倘◆

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 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 =..



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

Date/Time	Type	FlowFile Unit	Size	Component Name	Component Type
03/21/2017 22:22:05.079 PST	DROP			Messages	LogAttribute
03/21/2017 22:22:05.078 PST	DROP			Messages	LogAttribute
03/21/2017 22:22:05.077 PST	CONTENT_MODIFIED			EncryptContent	EncryptContent
03/21/2017 22:22:05.076 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.075 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.074 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.073 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.072 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.071 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.070 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.069 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.068 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.067 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.066 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.065 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.064 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.063 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.062 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.061 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.060 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.059 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.058 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.057 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.056 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.055 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.054 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.053 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.052 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.051 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.050 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.049 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.048 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.047 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.046 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.045 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.044 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.043 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.042 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.041 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.040 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.039 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.038 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.037 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.036 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.035 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.034 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.033 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.032 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.031 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.030 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.029 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.028 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.027 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.026 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.025 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.024 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.023 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.022 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.021 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.020 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.019 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.018 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.017 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.016 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.015 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.014 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.013 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.012 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.011 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.010 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.009 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.008 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.007 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.006 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.005 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.004 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.003 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.002 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.001 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.000 PST	CONTENT_MODIFIED			Messages	EncryptContent
03/21/2017 22:22:05.000 PST	CONTENT_MODIFIED			Log Decrypted Messages	LogAttribute
03/21/2017 22:22:05.000 PST	CONTENT_MODIFIED			Log Decrypted Messages	LogAttribute
03/21/2017 22:22:05.000 PST	CONTENT_MODIFIED			Decrypt MNifi Messages	EncryptContent
03/21/2017 22:22:05.000 PST	CONTENT_MODIFIED			Log Decrypted Messages	LogAttribute

Provenance Event

Input Claim
Container: default
Section: 1
Identifier: 1487744083443-1
Offset: 4259
Size: 6681

Output Claim
Container: default
Section: 1
Identifier: 1487744083443-1
Offset: 6681
Size: 45 bytes

Replay
Connection Id: 015a1002-3b71-13fc-98e3-f54ee6183280

OK



Prioritization?

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

```
201 Value: './'
204 Key: 'priority'
205 Value: '1'
206 Key: 's2s.address'
207 Value: 'localhost:54726'
208 Key: 's2s.host'
209 Value: 'localhost'
210 Key: 'segment.original.filename'
211 Value: 'example.3083-3175.log'
212 Key: 'tailfile.original.path'
213 Value: '/Users/aloipresto/Workspace/scratch/openiot/example.log'
214 Key: 'text.line.count'
215 Value: '1'
216 Key: 'uuid'
217 Value: 'a6b0f431-f7bd-4c1e-8aee-1959605eb86b'
218 This is a message at 2017/02/21 22:27:30.017Z
219 Standard Flowfile Attributes
220 Key: 'entryDate'
221 Value: 'Tue Feb 21 22:27:30 PST 2017'
222 Key: 'lineageStartDate'
223 Value: 'Tue Feb 21 22:27:30 PST 2017'
224 Key: 'fileSize'
225 Value: '45'
226 FlowFile Attribute Map Content
227 Key: 'RouteOnContent.Route'
228 Value: 'even'
229 Key: 'Store.State'
230 Value: 'Do not store state'
231 Key: 'filename'
232 Value: 'example.3083-3175.log'
233 Key: 'fragment.count'
234 Value: '2'
235 Key: 'fragment.identifier'
236 Value: 'c0ffd57f-cc93-49f4-8387-061144324d9b'
237 Key: 'fragment.index'
238 Value: '1'
239 Key: 'fragment.size'
240 Value: '45'
241 Key: 'mime.type'
242 Value: 'text/plain'
243 Key: 'path'
244 Value: './'
245 Key: 'priority'
246 Value: '2'
247 Key: 's2s.address'
248 Value: 'localhost:54726'
249 Key: 's2s.host'
250 Value: 'localhost'
251 Key: 'segment.original.filename'
252 Value: 'example.3083-3175.log'
253 Key: 'tailfile.original.path'
254 Value: '/Users/aloipresto/Workspace/scratch/openiot/example.log'
255 Key: 'text.line.count'
256 Value: '1'
257 Key: 'uuid'
258 Value: '917aa18f-ccc0-459b-a315-f8a8ae36ade2'
259 This is a message at 2017/02/21 22:27:29.516Z
260
261
262
263
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

Agenda

What is dataflow and what are the challenges?

Apache NiFi

IoT Challenges

Apache MiNiFi

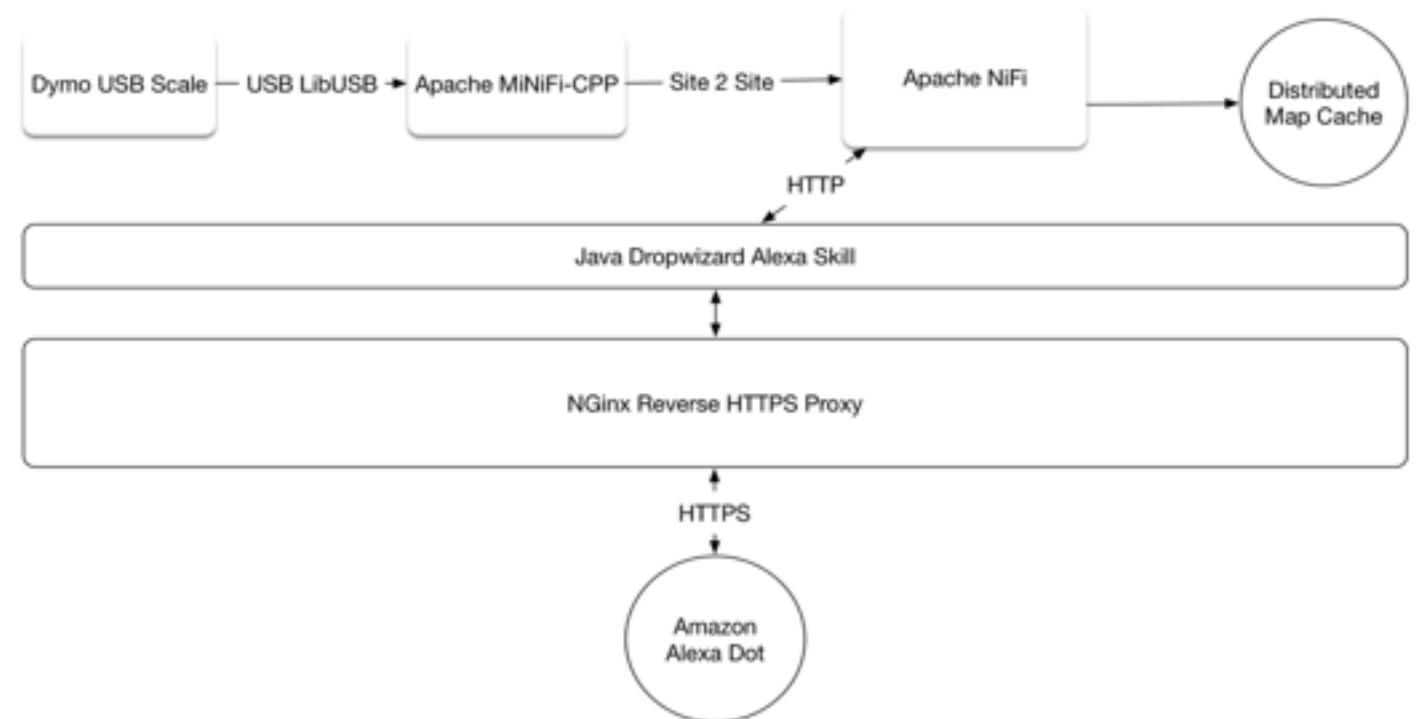
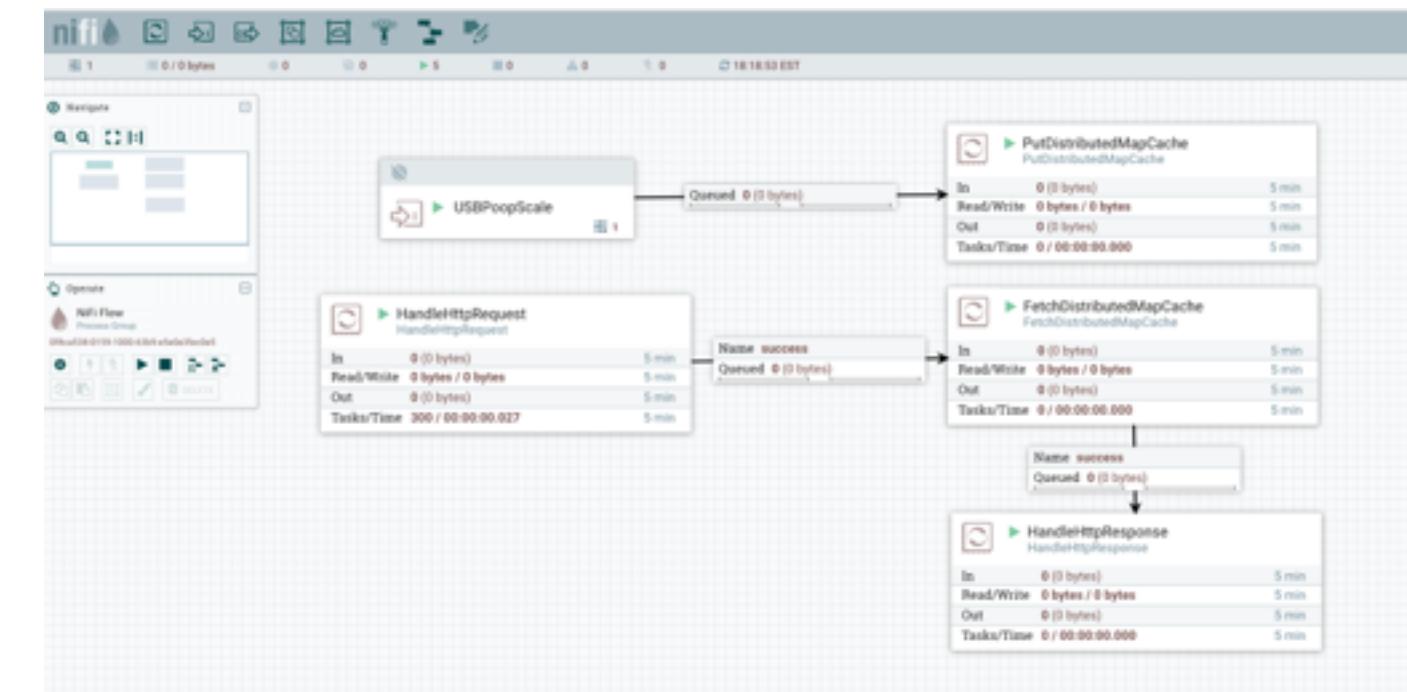
Exploration

Community



Community Example

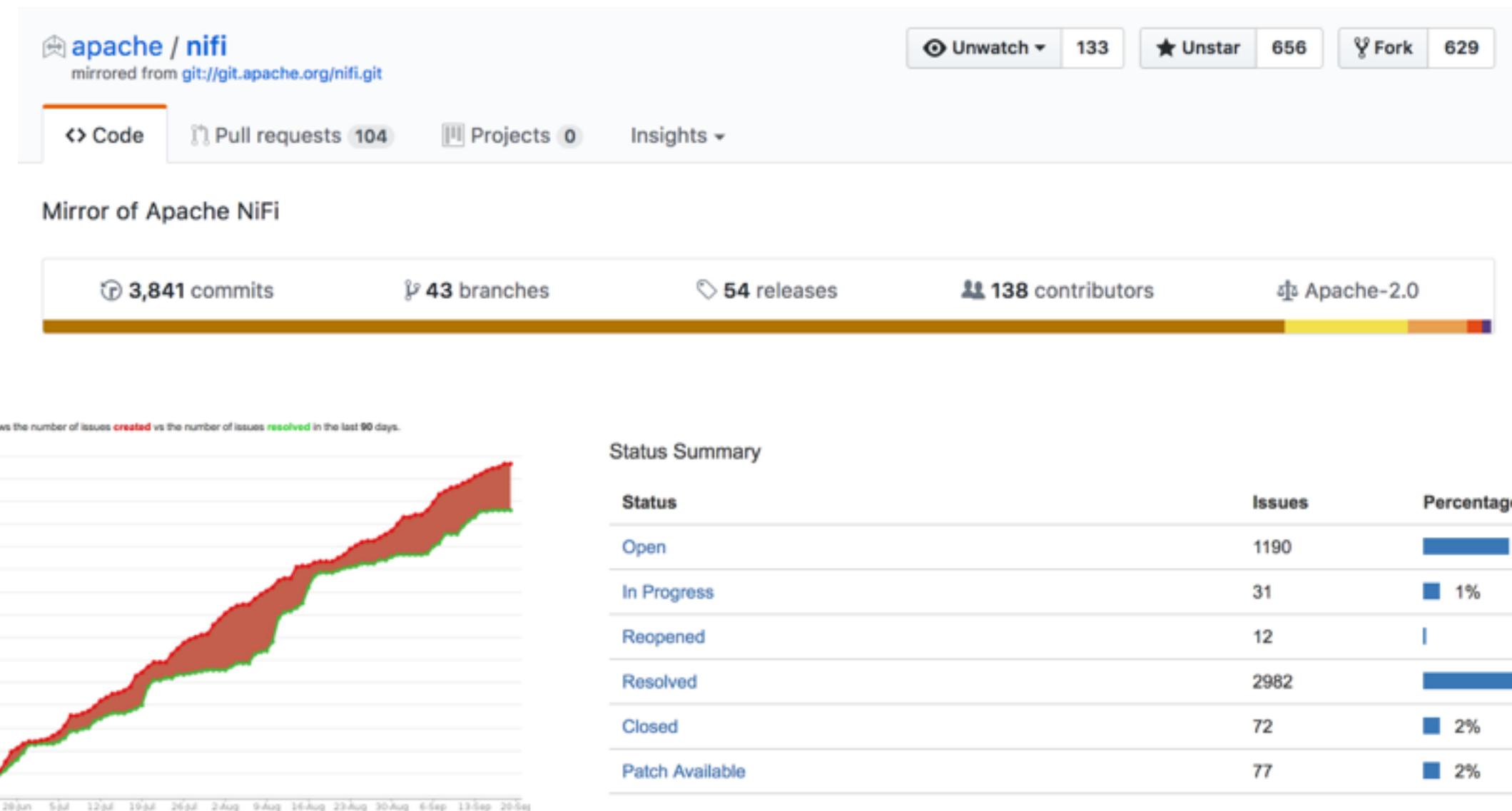
- ◆ Jeremy Dyer
- ◆ Alexa + MiNiFi + Dyer 2.0



Why NiFi & MiNiFi?

- ◆ Moving data is multifaceted in its challenges and these are present in different contexts at varying scopes
 - Inter vs intra, domestically, internationally
- ◆ Provide common tooling and extensions that are needed but be flexible for extension
 - Leverage existing libraries and expansive Java ecosystem for functionality
 - Allow organizations to integrate with their existing infrastructure
- ◆ Empower folks managing your infrastructure to make changes and reason about issues that are occurring
 - Data Provenance to show context and data's journey
 - User Interface/Experience a key component

Healthy Community



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

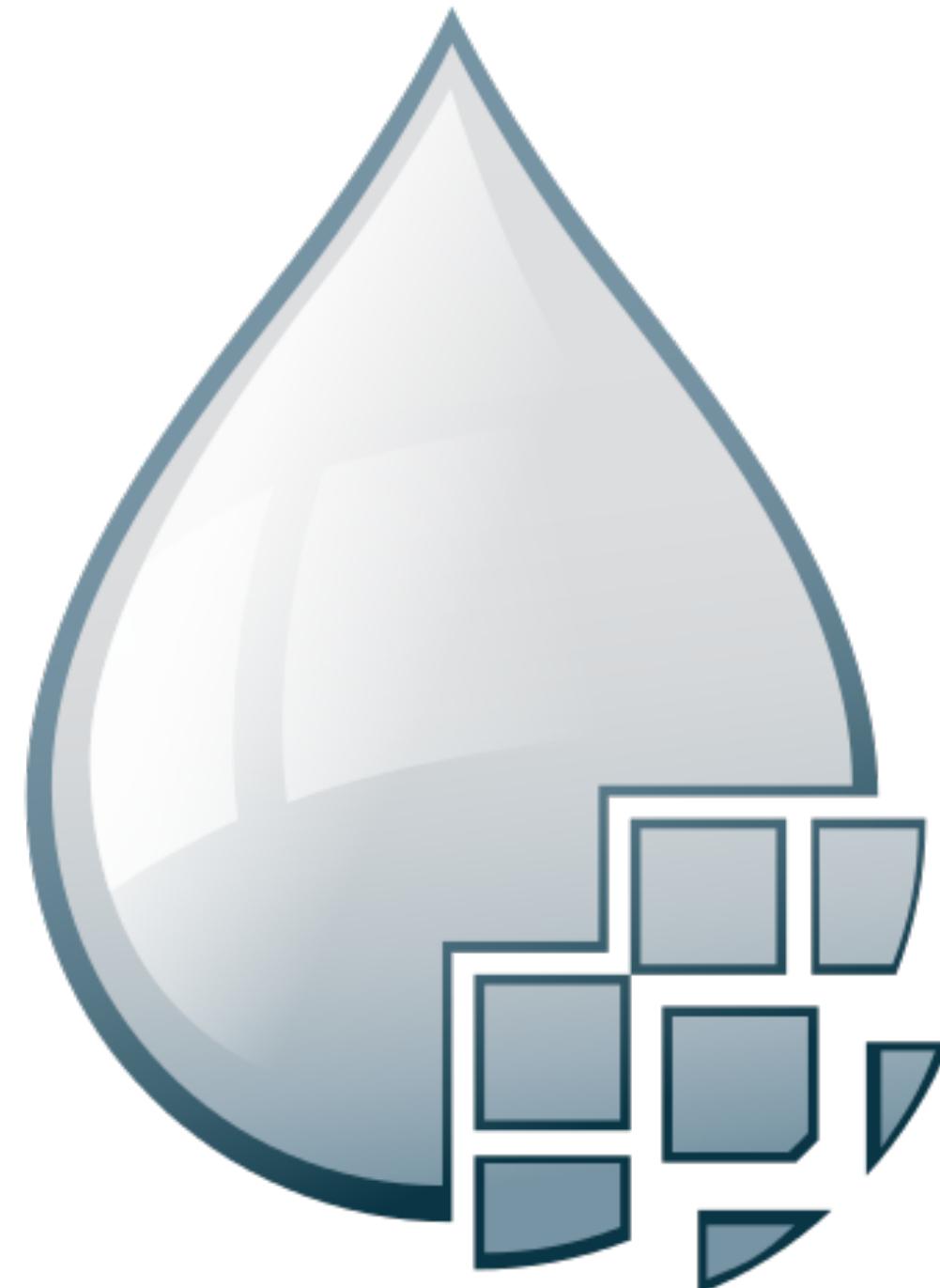
users@nifi.apache.org

Submit Ideas or Issues

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

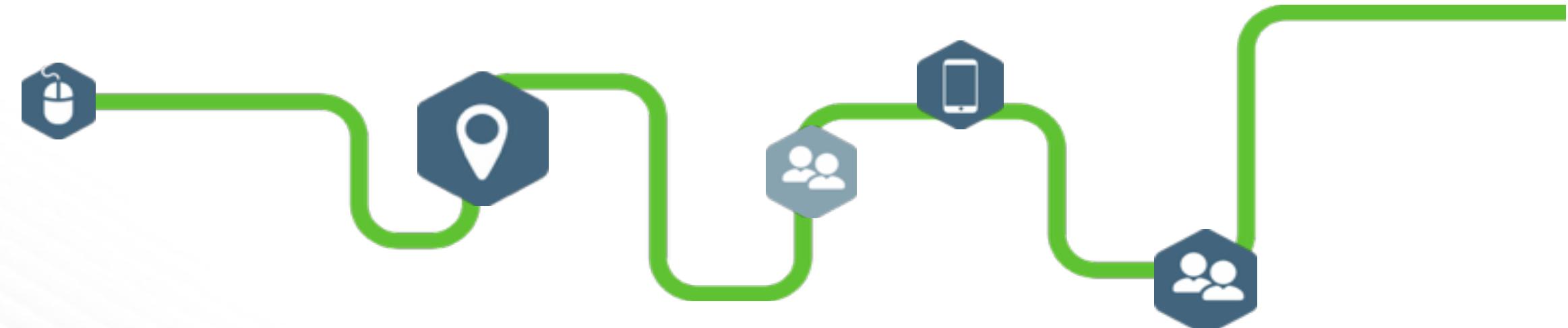
Follow us on Twitter

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



Learn and share at Birds of a Feather IOT, STREAMING & DATA FLOW

Thursday September 21
6:00 pm, C4.6



Thank You

I'm sticking around for discussions/questions



@yolopey / @apachennifi

alopresto@apache.org

PGP: 70EC B3E5 98A6 5A3F D3C4 BACE 3C6E F65B **2F7D EF69**