



# Information on Time

## *Remote Data Ingestion and Transformation with Apache NiFi/MiNiFi and IoT*

Andy LoPresto | @yolopey

Security Engineering Manager, Hortonworks, Apache NiFi PMC & Committer

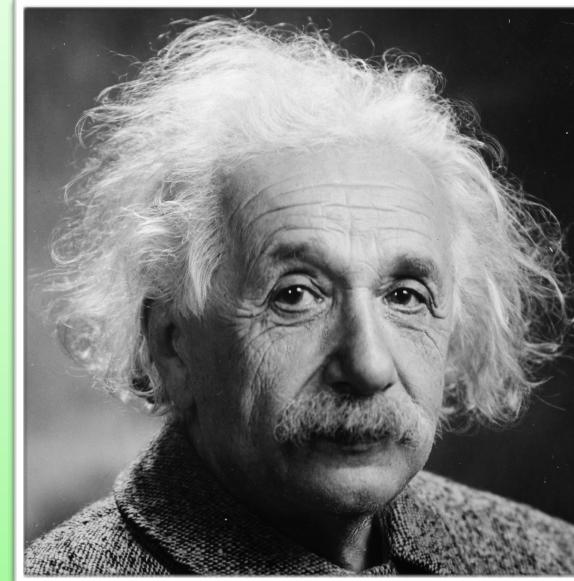
19 March 2019 Future of Data Barcelona

# Gauging Audience Familiarity With NiFi



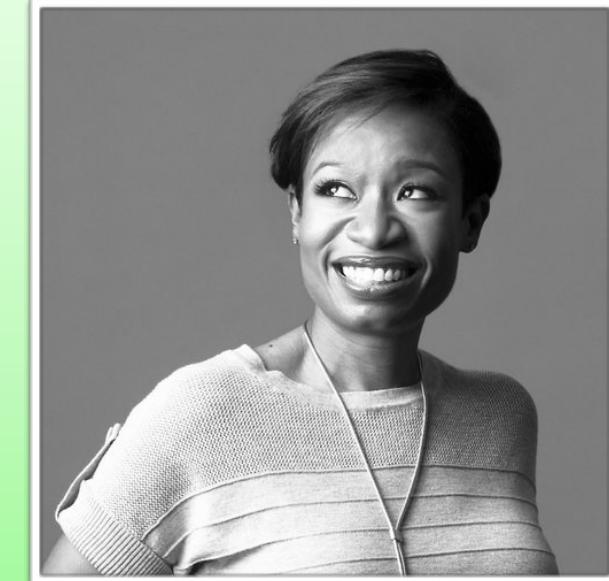
*“What’s a NeeFee?”*

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
- Apache MiNiFi
- Apache NiFi Registry
- IoT
- Demo
- *Roadmap*
- Q&A



*All slides provided online, no need to transcribe*

# Apache NiFi

# One Minute Intro to the NiFi Ecosystem



## NiFi

- Server/DC class
- GUI & REST API
- Interacts with hundreds of services



## MiNiFi (JVM/C++)

- Agent class
- Runs on minimal hardware
- Interacts with NiFi
- Simple event processing

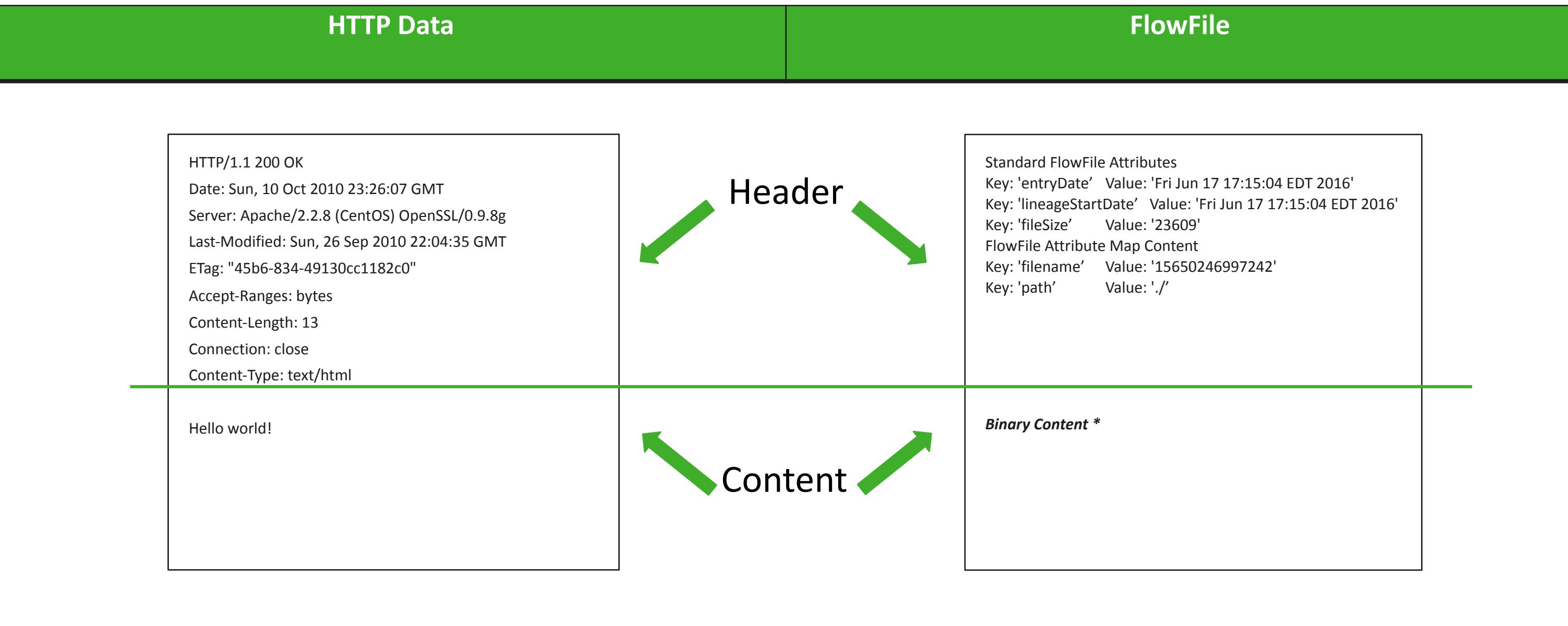


## NiFi Registry

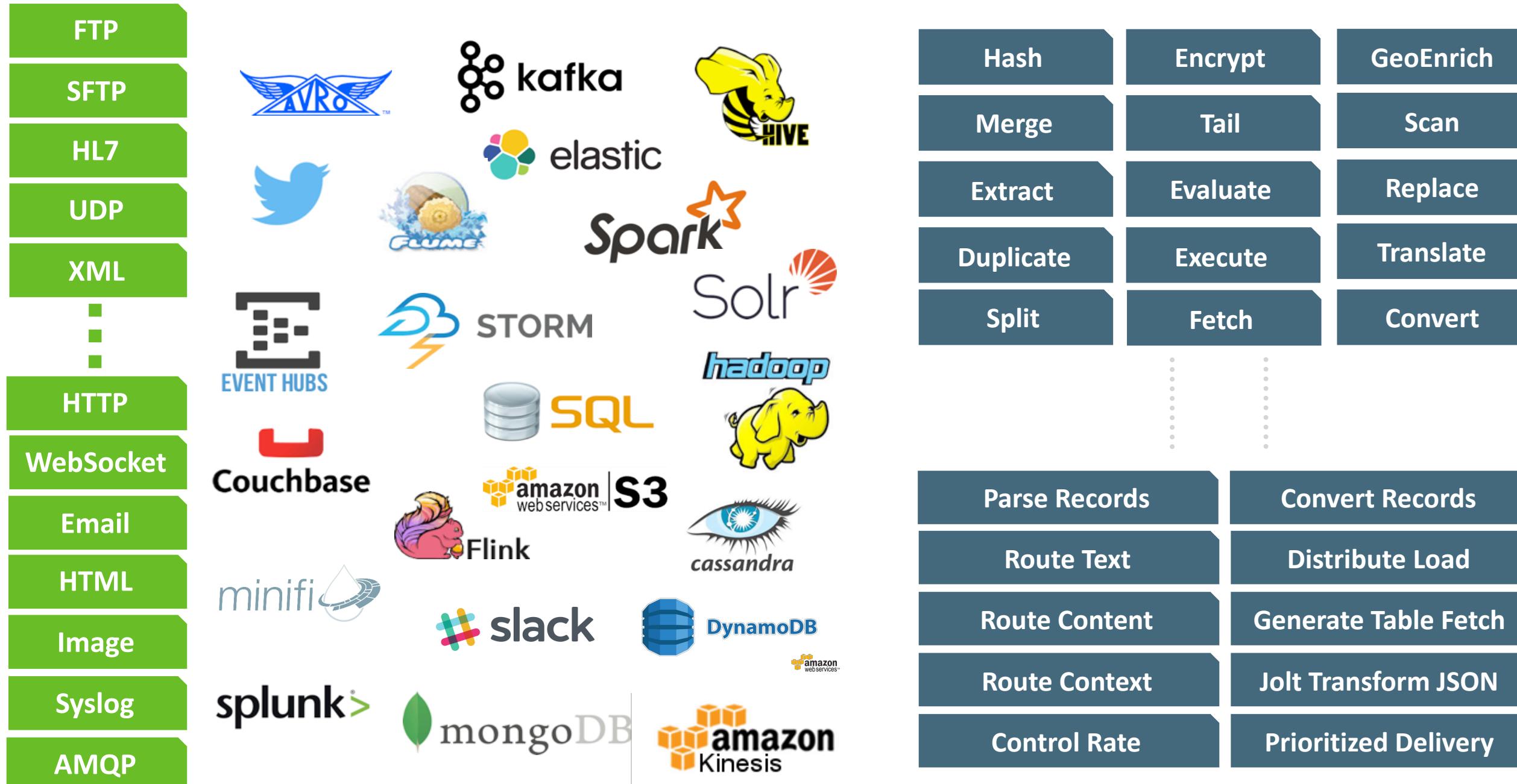
- Standalone application
- Handles asset management
- Flow versioning & deployment
- *Extension registry*



# Flowfiles Are Like HTTP Data



# Deeper Ecosystem Integration: 286+ Processors, 61 Controller Services

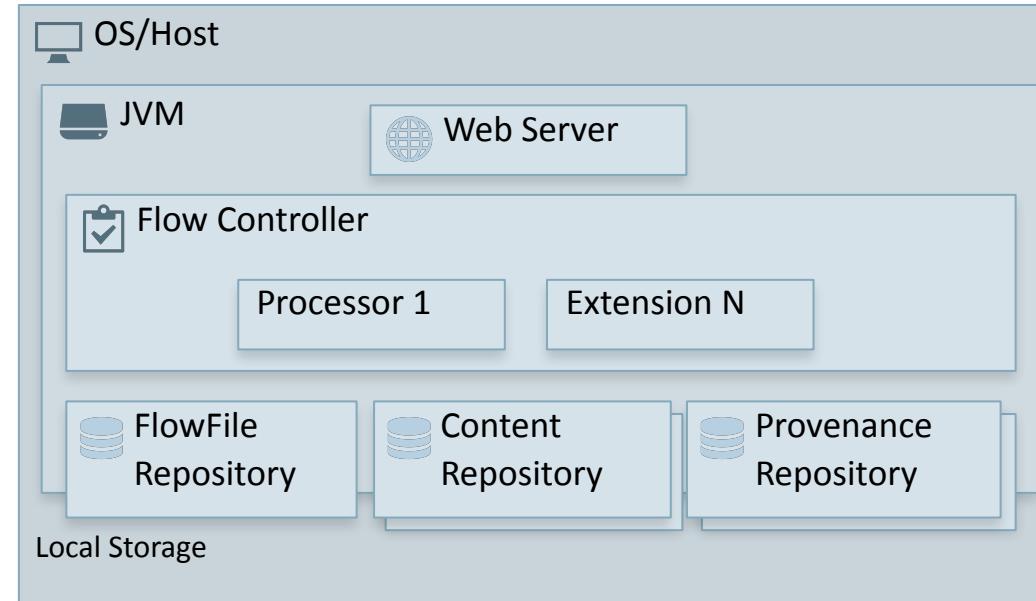


All Apache project logos are trademarks of the ASF and the respective projects.

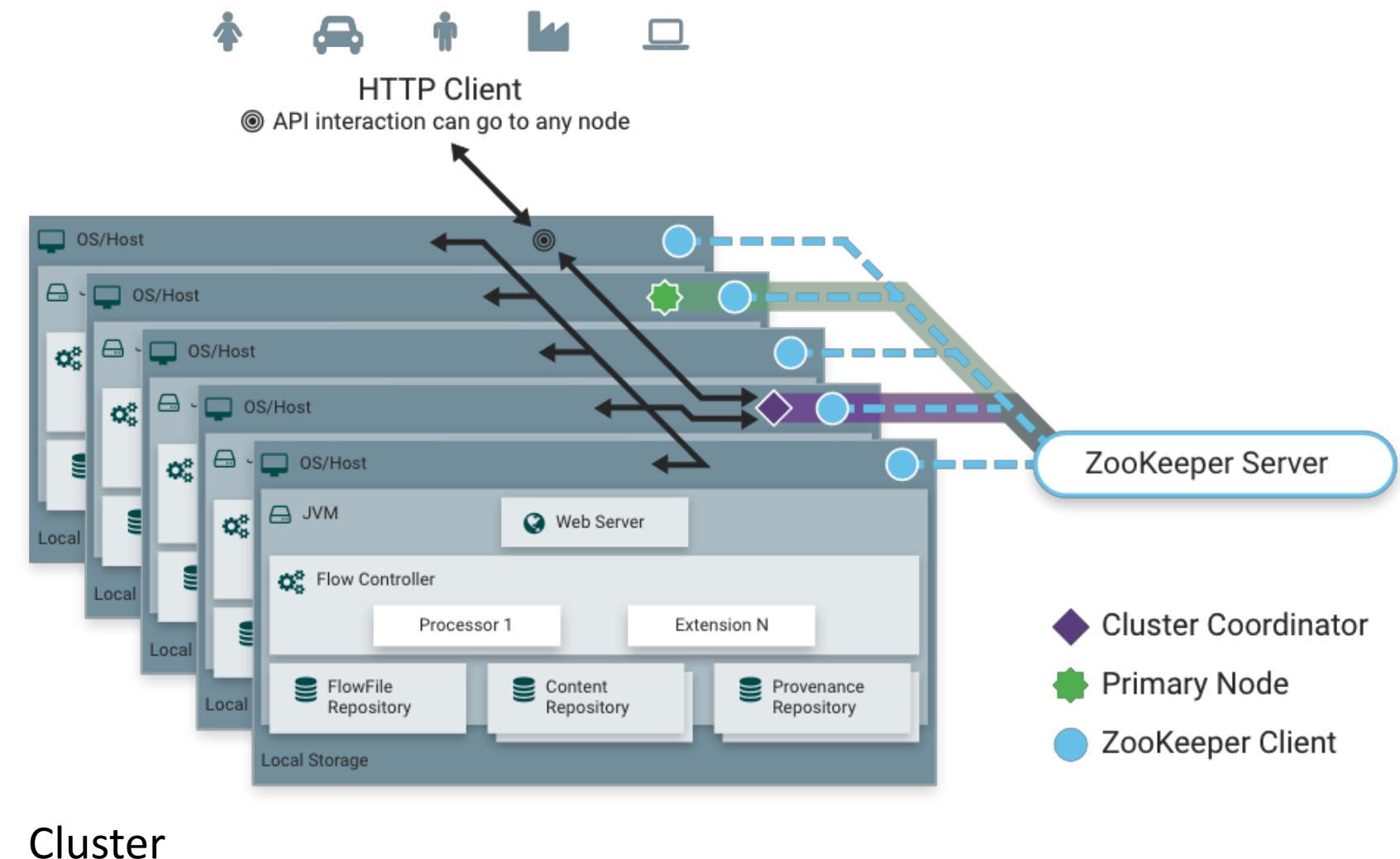
7

© Hortonworks Inc. 2011–2019. All rights reserved

# Architecture



Standalone

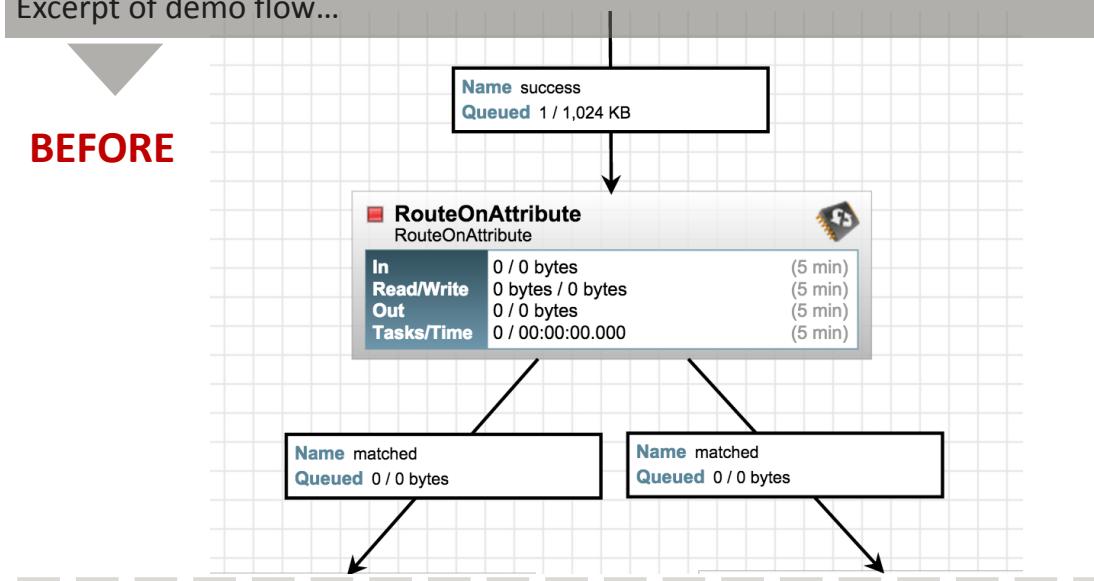


Cluster

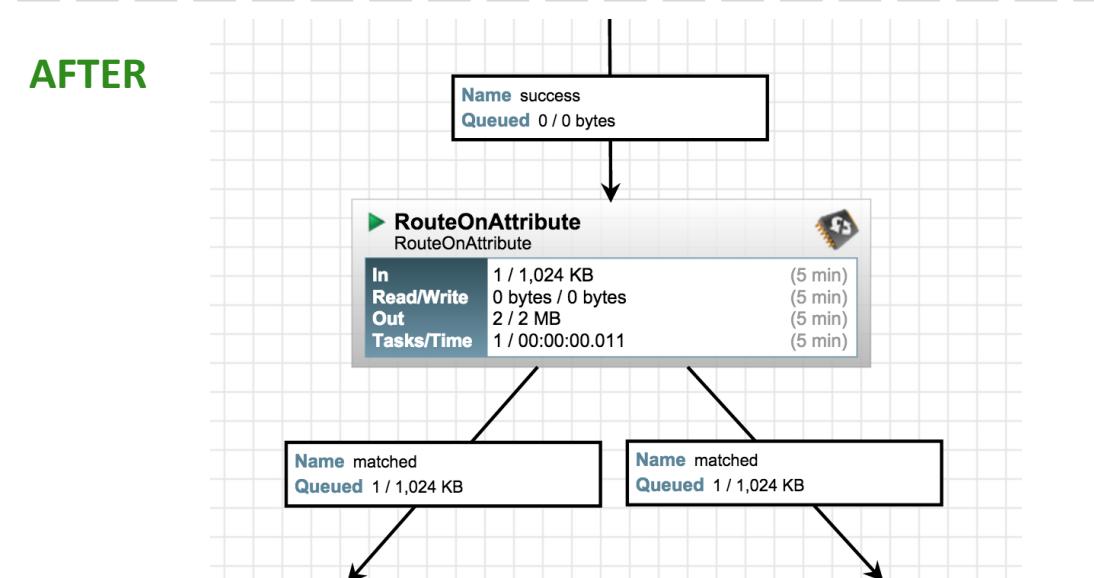
# NiFi Architecture – Repositories - Pass by reference

Excerpt of demo flow...

BEFORE



AFTER



What's happening inside the repositories...

$F_1 \rightarrow C_1$

$C_1$

$P_1 \rightarrow F_1$

$F_1 \rightarrow C_1$

$C_1$

$P_1 \rightarrow F_1 - \text{Create}$

$P_2 \rightarrow F_1 - \text{Route}$

$P_3 \rightarrow F_2 - \text{Clone}(F_1)$

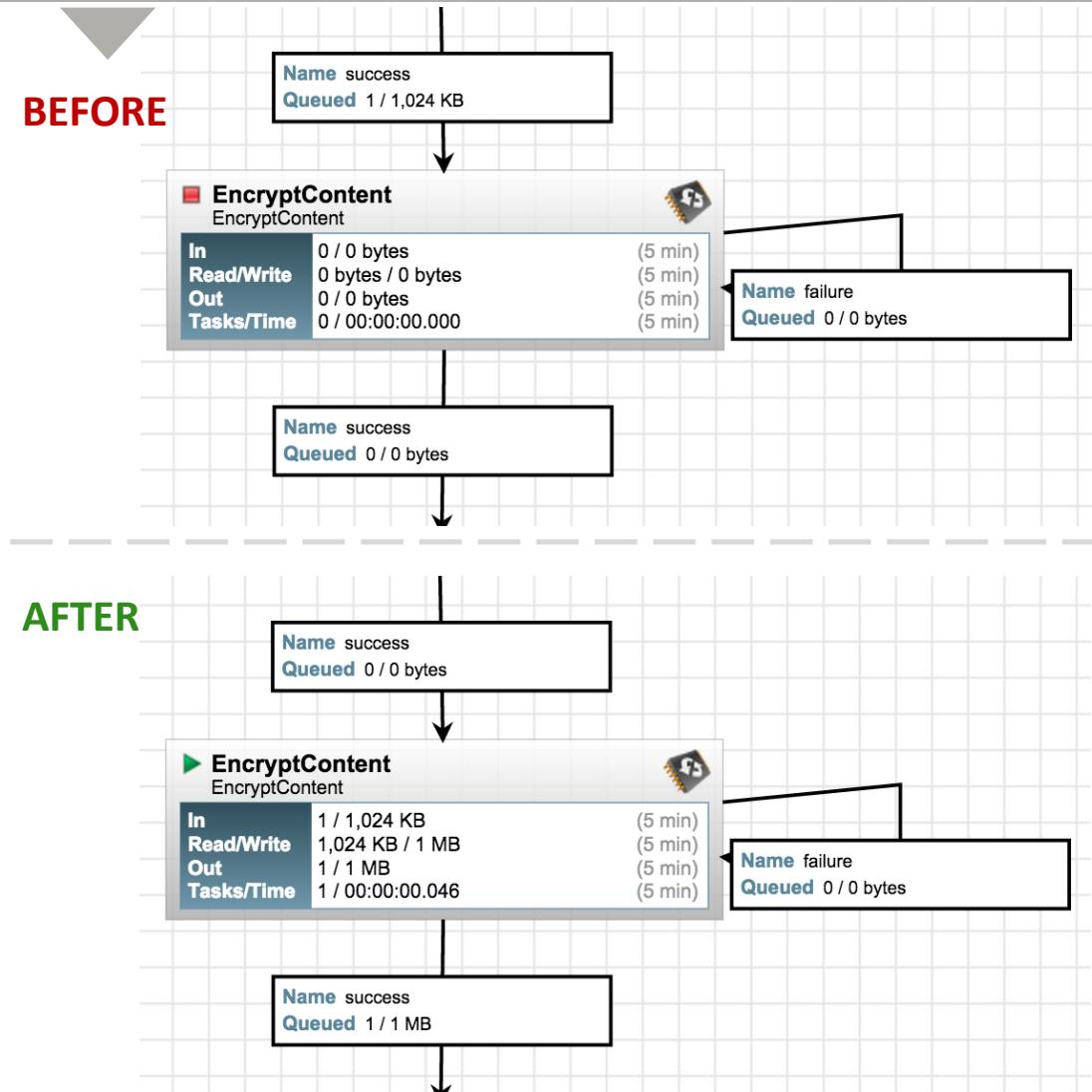
FlowFile

Content

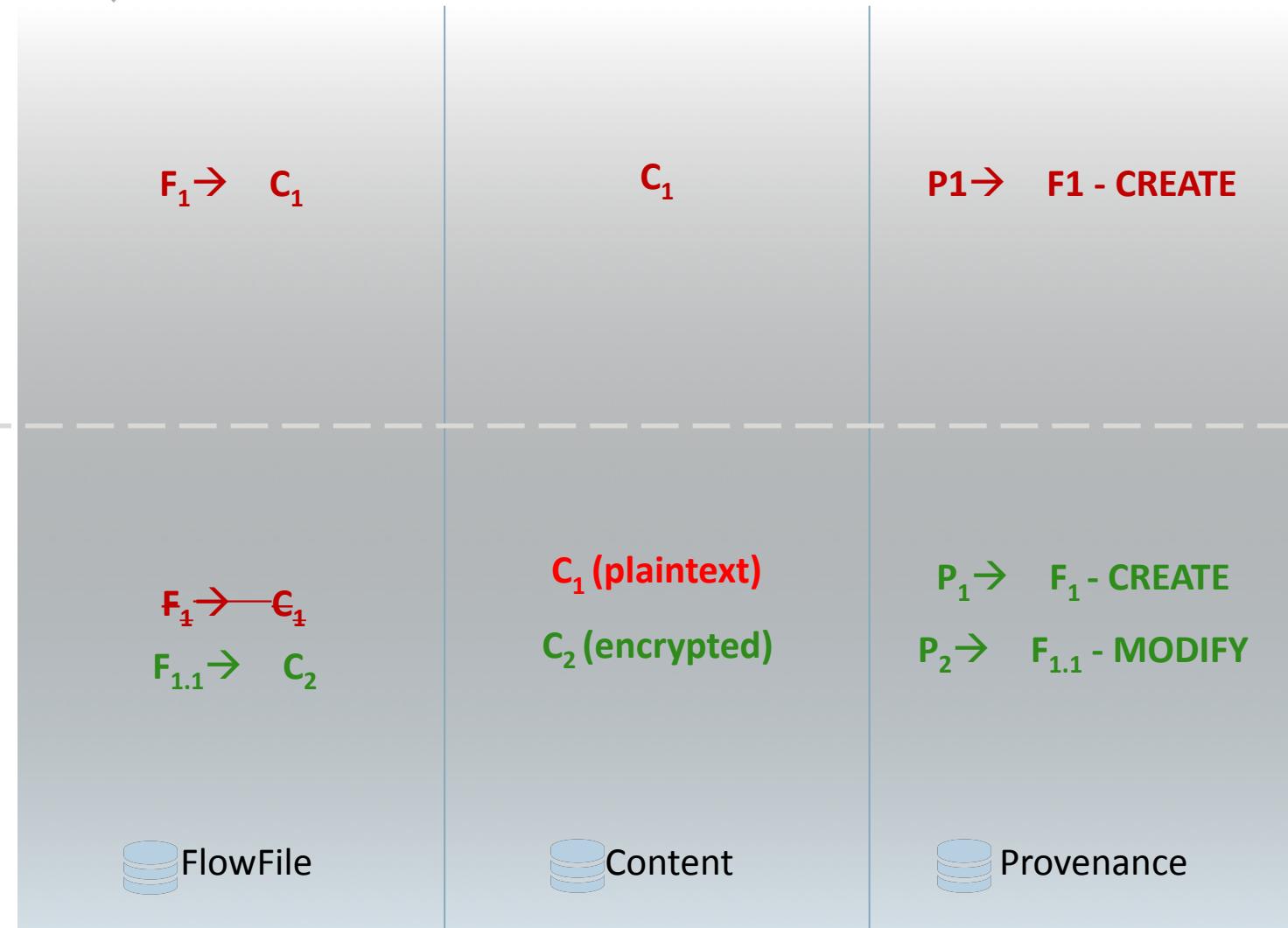
Provenance

# NiFi Architecture – Repositories – Copy on Write

Excerpt of demo flow...



What's happening inside the repositories...



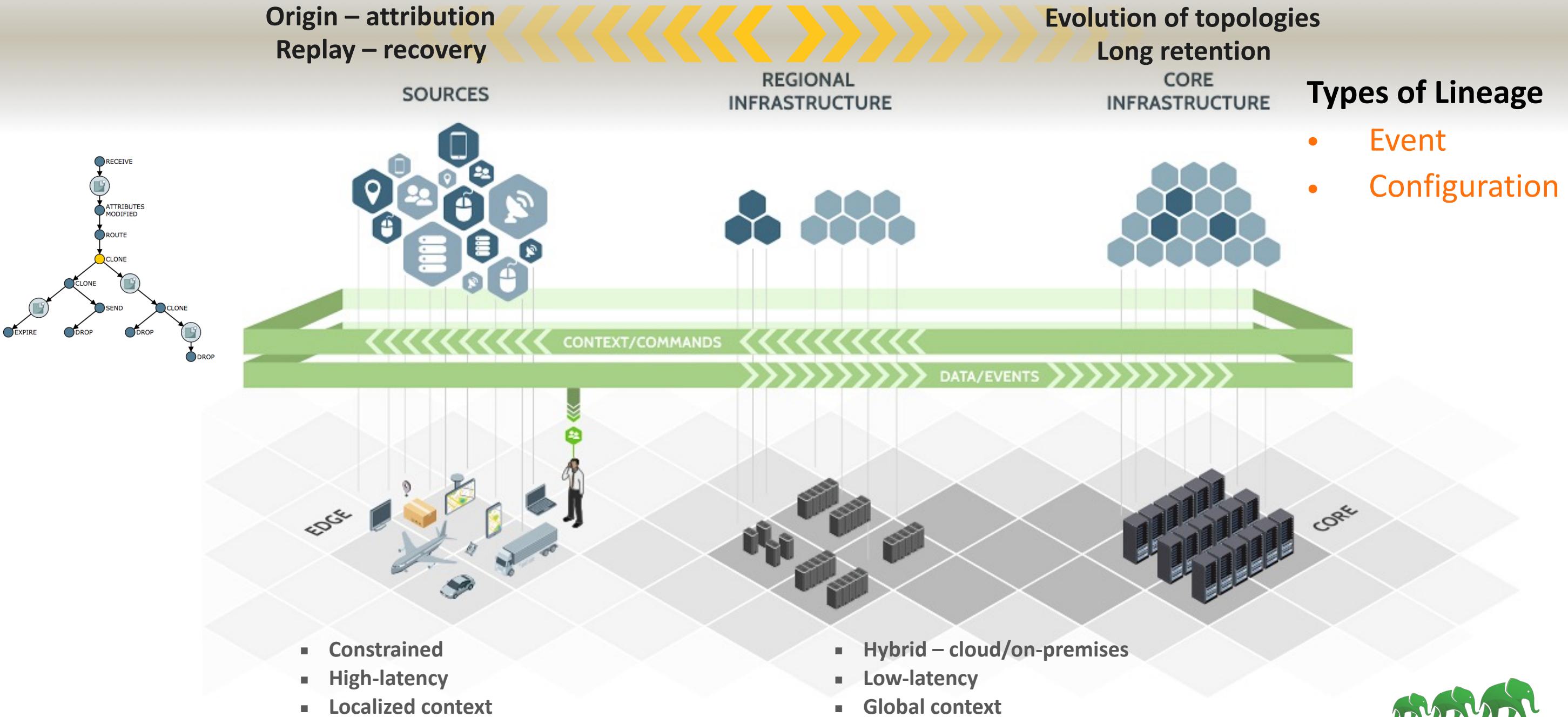
# People want to know about their data



[Colin the Chicken | Portlandia | IFC](#)

# Data Provenance

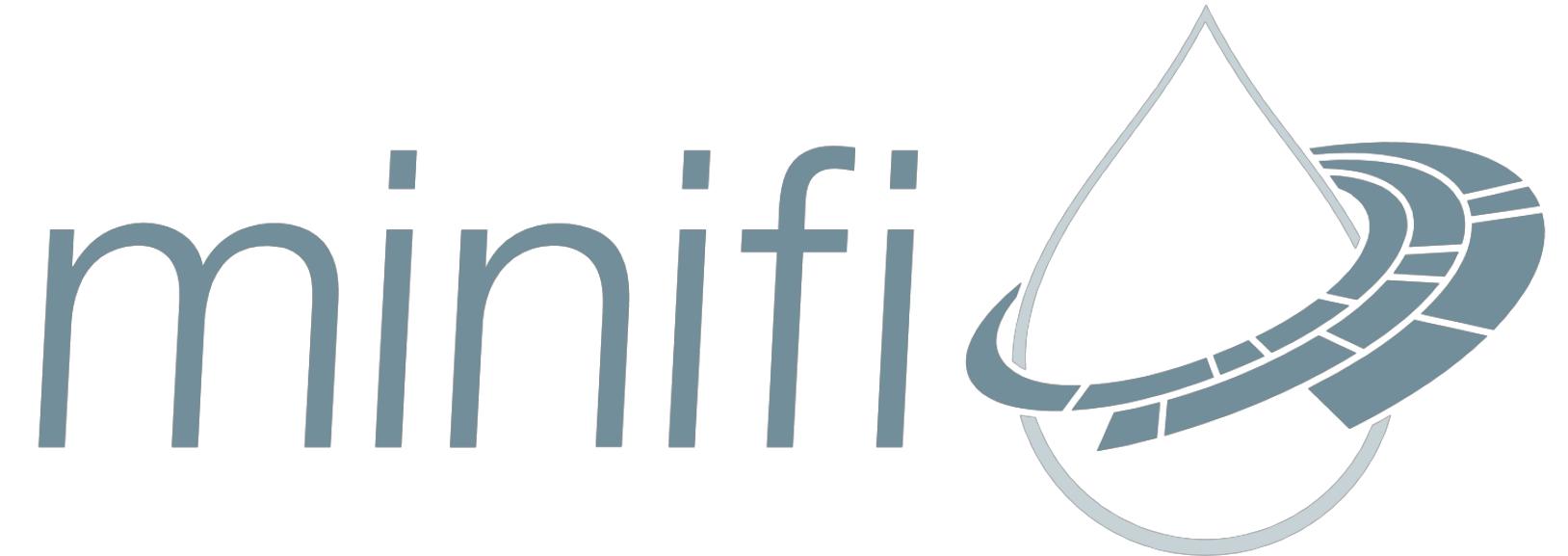
Origin – attribution  
Replay – recovery



# Apache MiNiFi

# Apache NiFi Subproject: MiNiFi

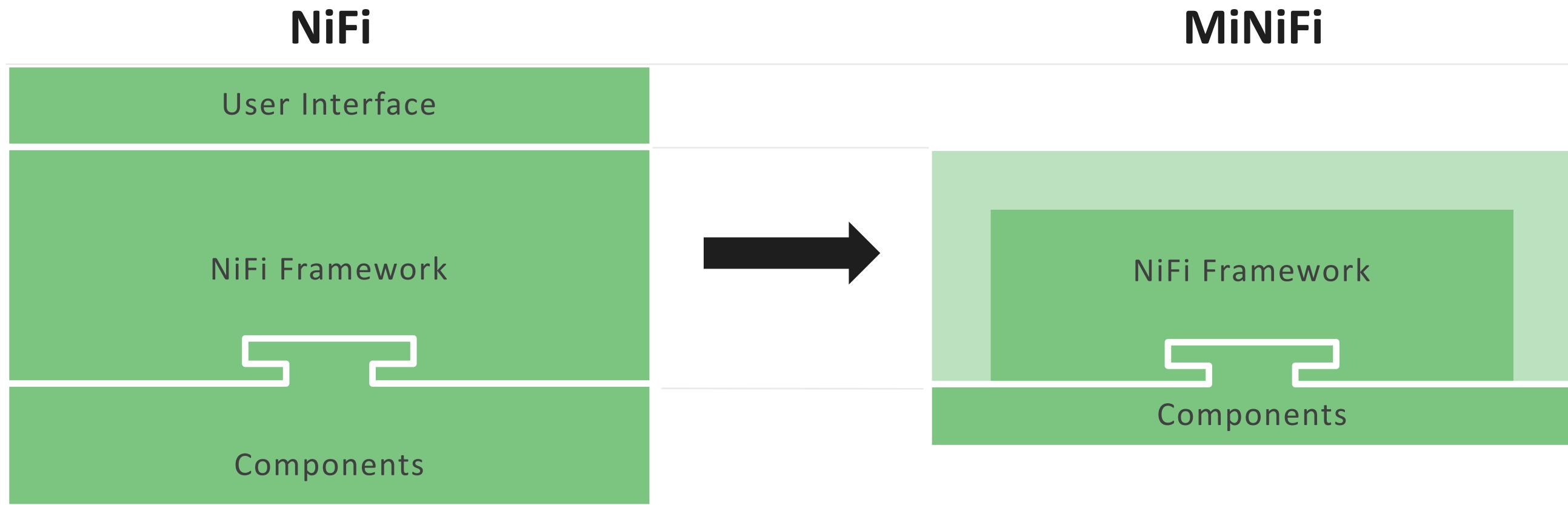
- Get the key parts of NiFi close to where data begins and provide bidirectional communication
- NiFi lives in the data center — give it an enterprise server or a cluster of them
- MiNiFi lives as close to where data is born and is a guest on that device or system
  - IoT
  - Connected car
  - Legacy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



# Why build MiNiFi?

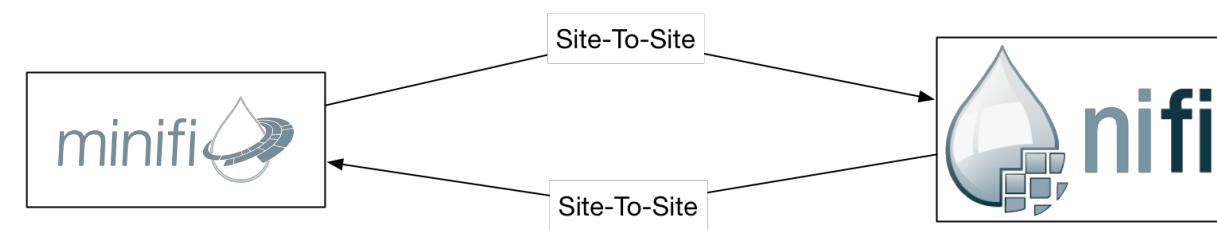
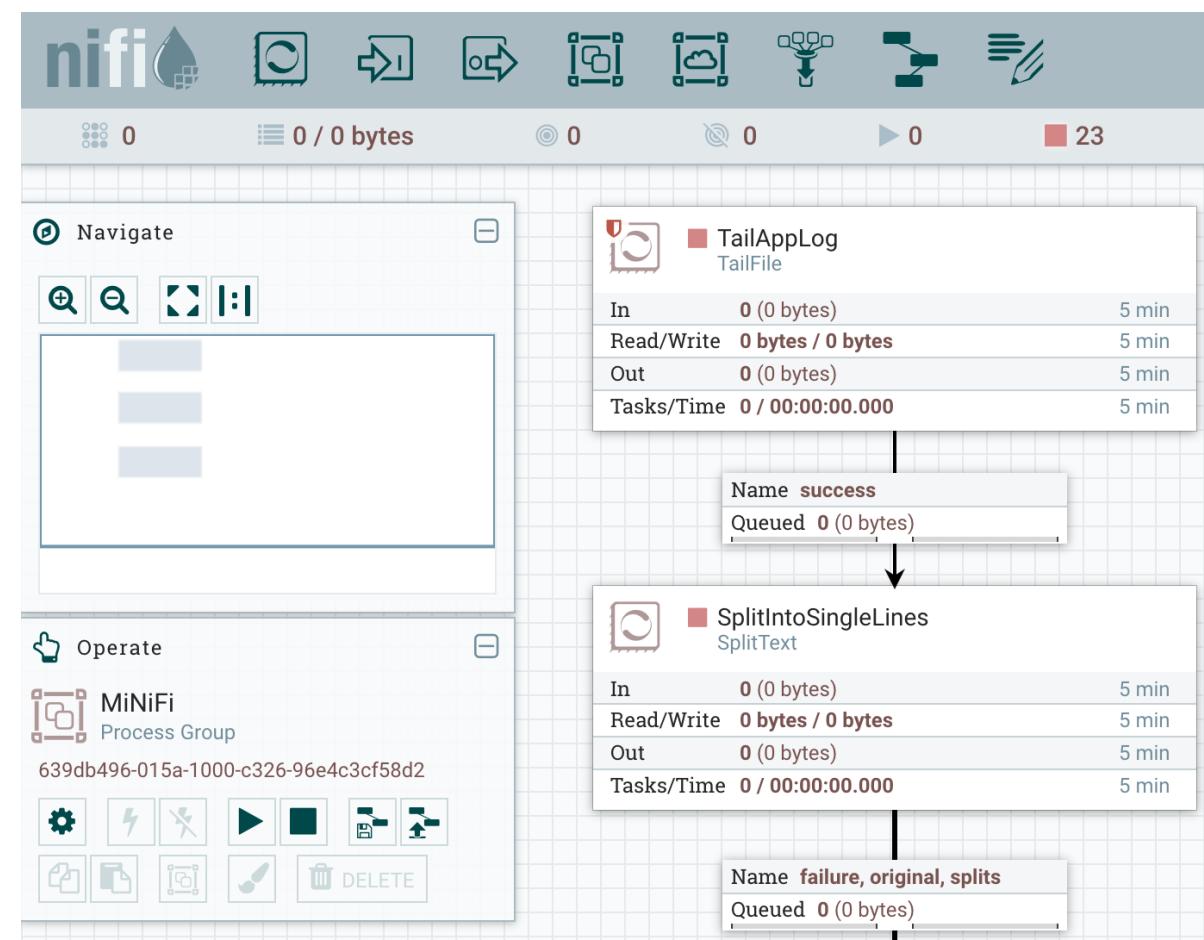
- NiFi is big
  - 1.8.0 release is 1.2 GB compressed
  - Can be modified to run in restricted environments, but requires manual surgery
  - Provides UI, provenance query, etc.
  - Runs on dedicated machines/clusters — “owns the box”
- MiNiFi lives at the edge
  - No UI
  - 0.5.0 Java release is 67 MB,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

# NiFi vs MiNiFi Java Processes



# 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

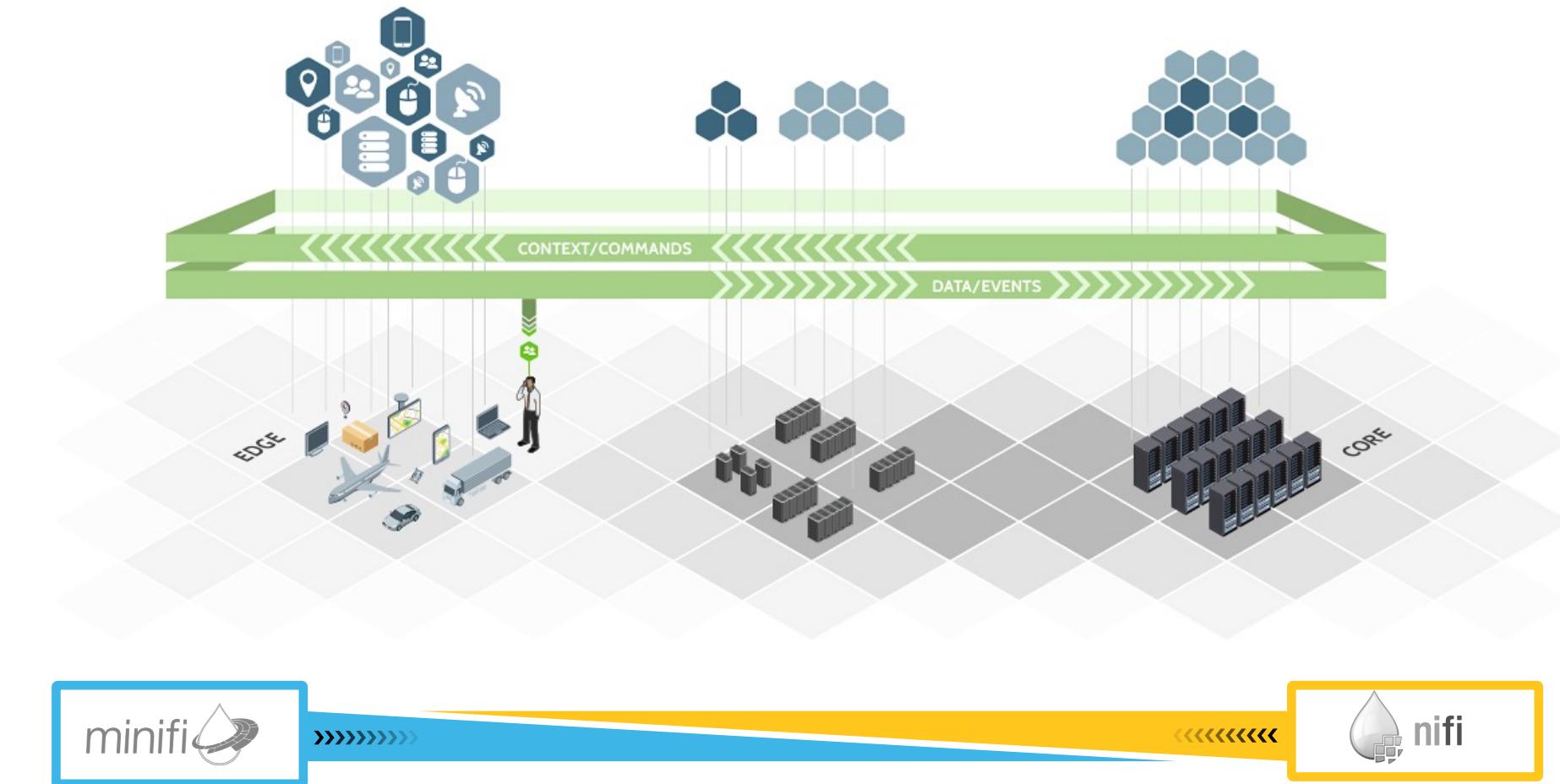


```
Security Properties:
  keystore: /tmp/ssl/localhost-ks.jks
  keystore type: JKS
  keystore password: localtest
  key password: localtest
  truststore: /tmp/ssl/localhost-ts.jks
  truststore type: JKS
  truststore password: localtest
  ssl protocol: TLS
  Sensitive Props:
    key:
      algorithm: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

# How Does MiNiFi Interact With NiFi?

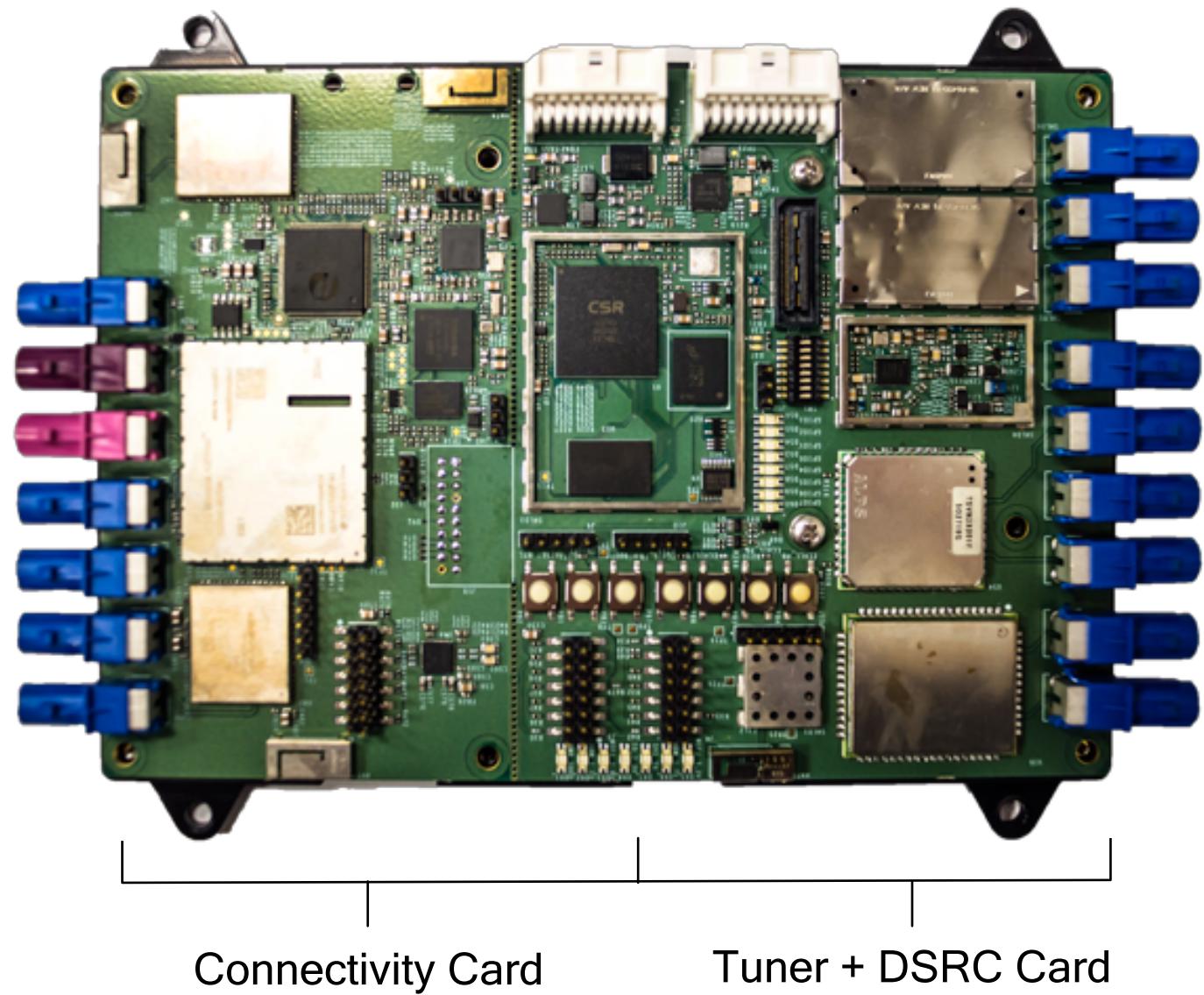
- NiFi
  - Design flows
  - Aggregate data from many sources
  - Perform routing/analysis/SEP
- MiNiFi
  - Receive flows
  - Collect data
  - Send for processing



# 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



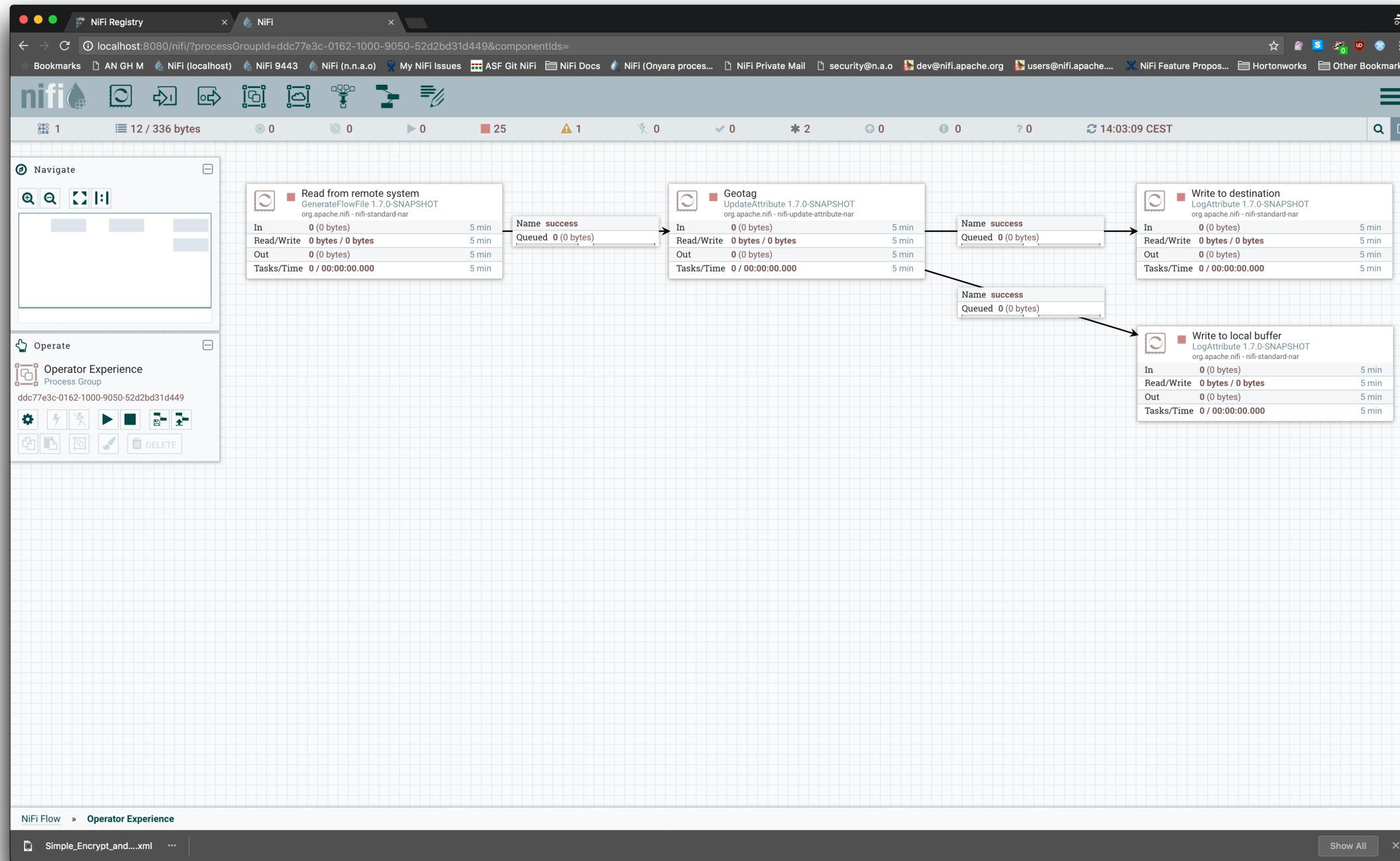
# Apache NiFi Registry

# Flow Development Lifecycle (FDLC)

- Origins of NiFi
- Operator Experience
  - MC data, don't drop, mitigate temporarily
- Version Control
- Environment Promotion



# Operator Experience



# Challenges

- Templates
  - Updates/replacement
  - Sensitive property replacement
- Flow.xml.gz migration
  - Key synchronization
- Environment promotion
  - Approval processes
  - Verifiability

# Template Replacement

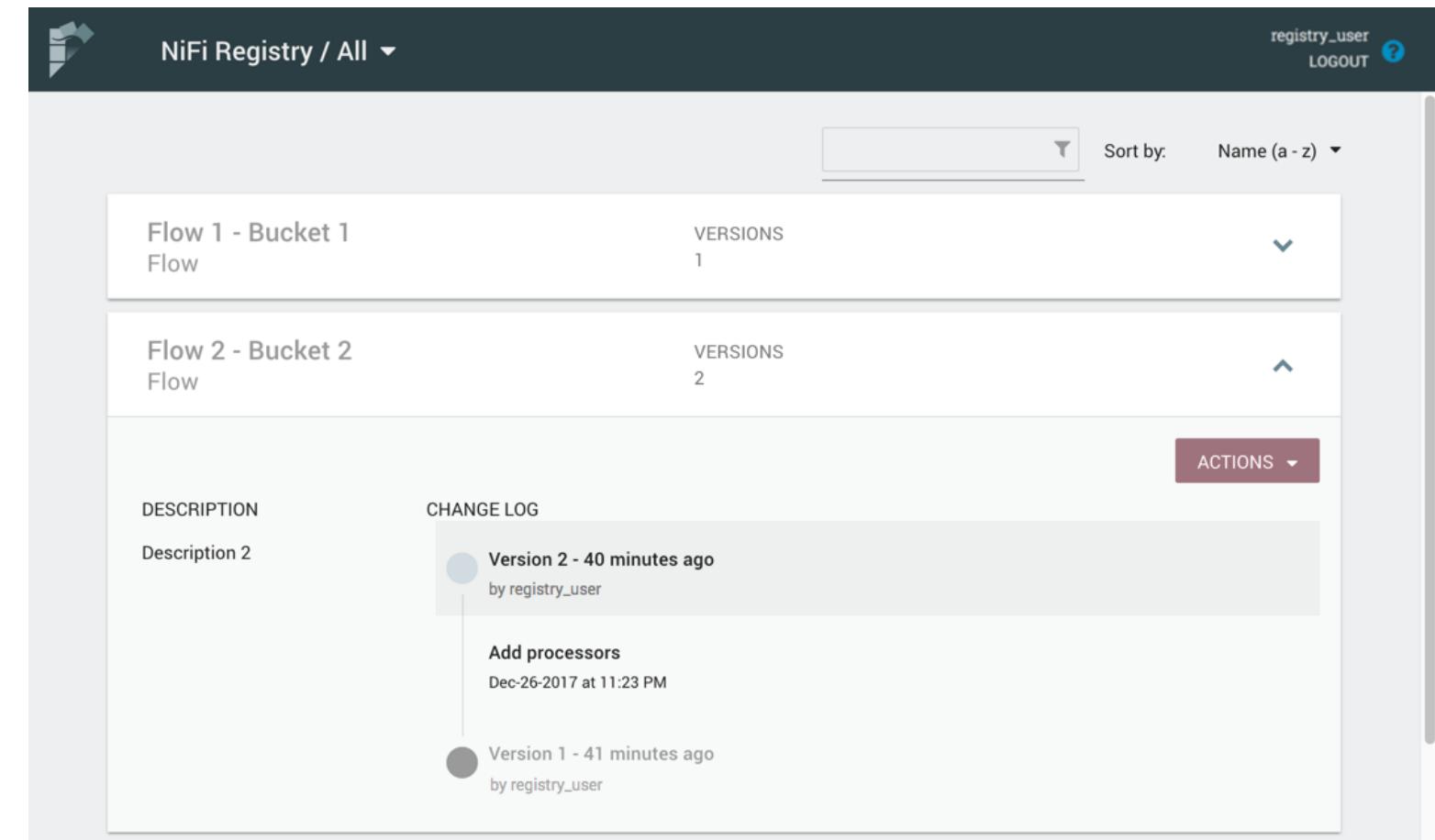
- Export a new version of template
- Transfer (somehow)
- Verify?
- Import onto canvas side-by-side existing flow
- Stop processors
- Empty queues
- Reconnect queues
- Start
- *Pray?*

```
-rw-r--r-- 1 alopresto staff 381B Apr 19 15:02 flow_20170626-171416_label_description.xml
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170628-111620_xxe_transform_xml.xml
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170628-155452_xxe_with_patch.xml
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170629-103436_xxe_password_leak.xml
-rw-r--r-- 1 alopresto staff 1.9K Apr 19 15:02 flow_20170717-192339_trusted_hostname.xml
-rw-r--r-- 1 alopresto staff 2.7K Apr 19 15:02 flow_20170718-101751_evaluate_json_path.xml
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170721-175745_site_to_site_secure.xml
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170724-161539.xml
-rw-r--r-- 1 alopresto staff 5.7K Apr 19 15:02 flow_20170724-181056_listen_http_sslv3.xml
-rw-r--r-- 1 alopresto staff 2.3K Apr 19 15:02 flow_20170726-155905_encrypted_spk.xml
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170802-172619_simple_gen_and_log.xml
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20170804-164011_simple_gen_and_log.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170807-154118_groovy_json_uppercase.xml
-rw-r--r-- 1 alopresto staff 1.8K Apr 19 15:02 flow_20170807-220536_put_mongo_record.xml
-rw-r--r-- 1 alopresto staff 6.1K Apr 19 15:02 flow_20170809-105304_date_difference.xml
-rw-r--r-- 1 alopresto staff 6.4K Apr 19 15:02 flow_20170809-174400_json_record_path.xml
-rw-r--r-- 1 alopresto staff 1.3K Apr 19 15:02 flow_20170809-210516_validate_xml.xml
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20170815-102522_json_to_sql.xml
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20170822-104407_json_formatter.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170823-123729_evaluatexpath.xml
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20170825-115758_encrypt_and_decrypt.xml
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20170825-121351_encrypt_decrypt_clean.xml
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20170825-121656_encrypt_decrypt_aligned.xml
-rw-r--r-- 1 alopresto staff 2.3K Apr 19 15:02 flow_20170825-190225_listen_http_and_invoke_http_with_tls1_2.xml
-rw-r--r-- 1 alopresto staff 5.0K Apr 19 15:02 flow_20170831-153220_psaltis_meetup_template_local.xml
-rw-r--r-- 1 alopresto staff 4.0K Apr 19 15:02 flow_20170905-175343_xxe_template.xml
-rw-r--r-- 1 alopresto staff 4.3K Apr 19 15:02 flow_20170905-185907_funnel_bus.xml
-rw-r--r-- 1 alopresto staff 8.1K Apr 19 15:02 flow_20170906-111801_chained_get_html_element_NIFI-4356.xml
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20170909-121848_sydney_demo.xml
-rw-r--r-- 1 alopresto staff 3.2K Apr 19 15:02 flow_20170919-202215_meetup_sydney.xml
-rw-r--r-- 1 alopresto staff 3.8K Apr 19 15:02 flow_20170920-190328_crash_course.xml
-rw-r--r-- 1 alopresto staff 2.9K Apr 19 15:02 flow_20170926-180548_handle_http_with_tls1_2.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20170929-172333_encrypt_decrypt_1_4_0_verification.xml
-rw-r--r-- 1 alopresto staff 6.3K Apr 19 15:02 flow_20171003-192928_update_record_and_replace_text.xml
-rw-r--r-- 1 alopresto staff 8.9K Apr 19 15:02 flow_20171004-180529_web_server_and_json.xml
-rw-r--r-- 1 alopresto staff 3.8K Apr 19 15:02 flow_20171103-183334_groovy_json_to_sql_srsw.xml
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20171127-190651_local_site_to_site.xml
-rw-r--r-- 1 alopresto staff 2.1K Apr 19 15:02 flow_20180101-145059_registry_rc_verification.xml
-rw-r--r-- 1 alopresto staff 1.4K Apr 19 15:02 flow_20180104-104211_test_count_text.xml
-rw-r--r-- 1 alopresto staff 2.0K Apr 19 15:02 flow_20180104-112138_test_count_text_remote.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-100218_consume_jms_test.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-105514.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-131629_consume_jms_10_threads.xml
-rw-r--r-- 1 alopresto staff 1.7K Apr 19 15:02 flow_20180202-160620_consume_jms_multithread.xml
-rw-r--r-- 1 alopresto staff 1.6K Apr 19 15:02 flow_20180208-200731_simple_demo_route_on_time_parity.xml
-rw-r--r-- 1 alopresto staff 5.2K Apr 19 15:02 flow_20180214-182110_jms_deserialization_test.xml
-rw-r--r-- 1 alopresto staff 2.6K Apr 19 15:02 flow_20180216-205336_attribute_evaluator.xml
-rw-r--r-- 1 alopresto staff 5.7K Apr 19 15:02 flow_20180228-130457_attribute_sqo.xml
-rw-r--r-- 1 alopresto staff 8.3K Apr 19 15:02 flow_20180307-171740_NIFI-4928.xml
-rw-r--r-- 1 alopresto staff 1.5K Apr 19 15:02 flow_20180313-171310_python_process_attribute.xml
-rw-r--r-- 1 alopresto staff 2.4K Apr 19 15:02 flow_20180315-113634_NIFI-4246-0Auth.xml
-rw-r--r-- 1 alopresto staff 1.9K Apr 19 15:02 flow_20180328-115056.xml
-rw-r--r-- 1 alopresto staff 2.5K Apr 19 15:02 flow_20180405-111419_prioritizers.xml
```

# 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



NiFi Registry / All ▾

registry\_user LOGOUT ?

Sort by: Name (a - z) ▾

Flow	VERSIONS
Flow 1 - Bucket 1	1
Flow 2 - Bucket 2	2

DESCRIPTION

Description 2

CHANGE LOG

Version 2 - 40 minutes ago  
by registry\_user

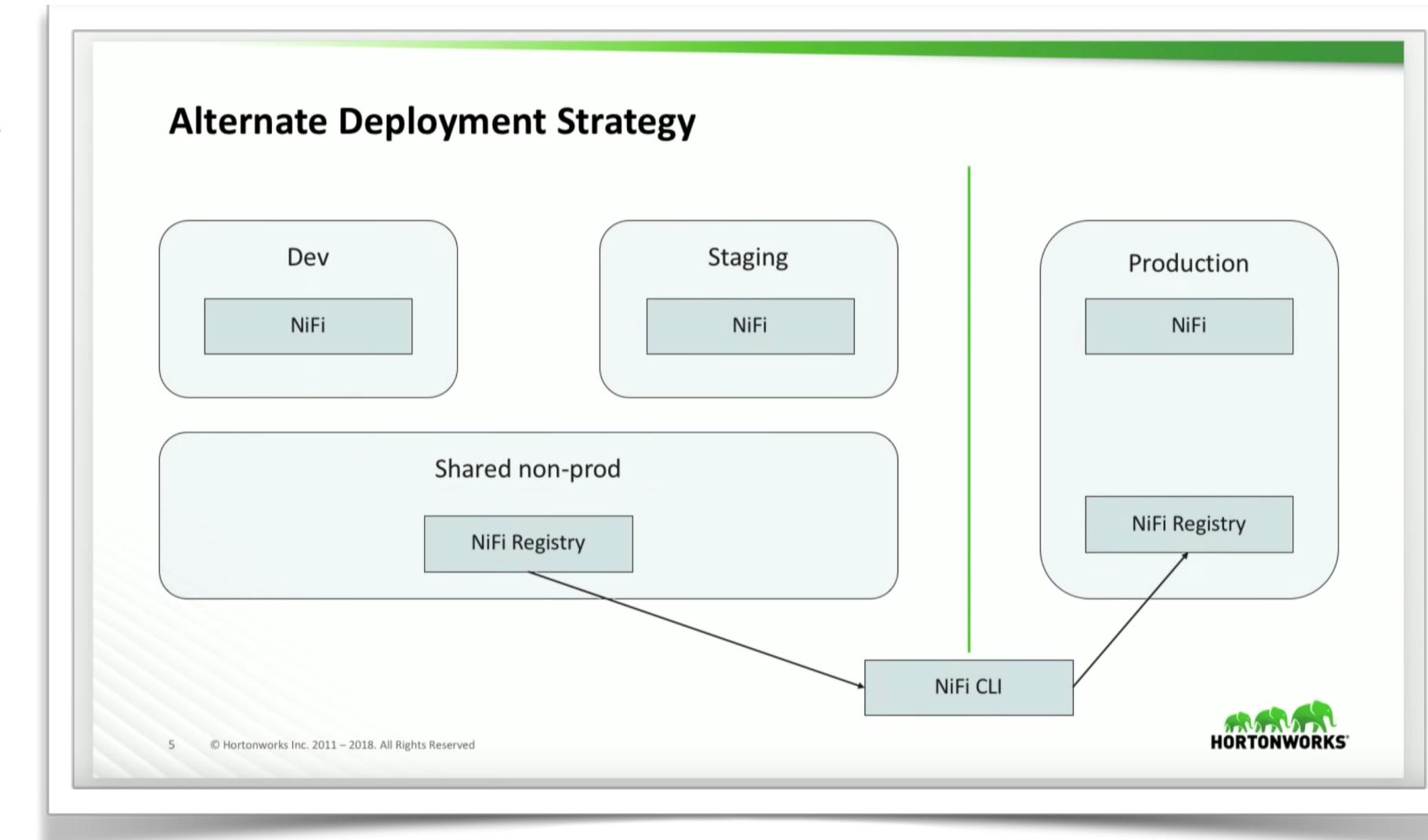
Add processors  
Dec-26-2017 at 11:23 PM

Version 1 - 41 minutes ago  
by registry\_user

ACTIONS ▾

# Flows can be promoted between environments

- Connect multiple NiFi instances to a NiFi Registry instance
  - Communicate between multiple NiFi Registry instances
    - via multiple Registry Clients
    - via *NiFi CLI*



# Extensibility

- Git-backed persistence
  - Share flows via GitHub, etc.
- Commit hooks
  - Register a hook & action
  - “When a new version of the flow is **committed** to QA Registry, **email** the QA team and **post** in the QA Deploy Slack channel”
- Pluggable DB implementations

## Event Hooks

Event hooks are an integration point that allows for custom code to be triggered when NiFi Registry application events occur.

Event Name	Description
CREATE_BUCKET	A new registry bucket is created.
CREATE_FLOW	A new flow is created in a specified bucket. Only triggered on first time creation of a flow with a given name.
CREATE_FLOW_VERSION	A new version for a flow has been saved in the registry.
UPDATE_BUCKET	A bucket has been updated.
UPDATE_FLOW	A flow that exists in a bucket has been updated.
DELETE_BUCKET	An existing bucket in the registry is deleted.
DELETE_FLOW	An existing flow in the registry is deleted.
REGISTRY_START	Invoked once the NiFi Registry application has been successfully started. This is only invoked after a complete and successful start.

# Complementary Tools

# Complementary Tools

- NiFi Toolkit
- NiPyAPI
- MiNiFi Converter Toolkit

# NiFi Toolkit

- TLS Toolkit
  - Generates, signs, and packages keys and certificates for NiFi services (node/cluster, clients)
- Encrypt Config
  - Protects sensitive configuration values like passwords
- CLI
  - Interacts with NiFi & NiFi Registry to operate on flows

The default output type in interactive mode is `simple`, and the default output type in standalone mode is `json`.

Example of simple output for list-buckets:

```
#> registry list-buckets -ot simple
My Bucket - 3c7b7467-0012-4d8f-a918-6aa42b6b9d39
```

Example of json output for list-buckets:

```
#> registry list-buckets -ot json
[ {
  "identifier" : "3c7b7467-0012-4d8f-a918-6aa42b6b9d39",
  "name" : "My Bucket",
  "createdTimestamp" : 1516718733854,
  "permissions" : {
    "canRead" : true,
    "canWrite" : true,
    "canDelete" : true
  },
  "link" : {
    "params" : {
      "rel" : "self"
    },
    "href" : "buckets/3c7b7467-0012-4d8f-a918-6aa42b6b9d39"
  }
}]
```

# NiPyAPI

- Python wrapper around NiFi REST API
- Community-provided by Daniel Chaffelson
- Exposes common operations for automation, batch processing, recursion, etc.

```
dev_bucket = nipyapi.versioning.get_registry_bucket(dev_bucket_name)
dev_ver_flow = nipyapi.versioning.get_flow_in_bucket(
    dev_bucket.identifier,
    identifier=dev_ver_flow_name
)
dev_export = nipyapi.versioning.export_flow_version(
    bucket_id=dev_bucket.identifier,
    flow_id=dev_ver_flow.identifier,
    mode='yaml'
)
```

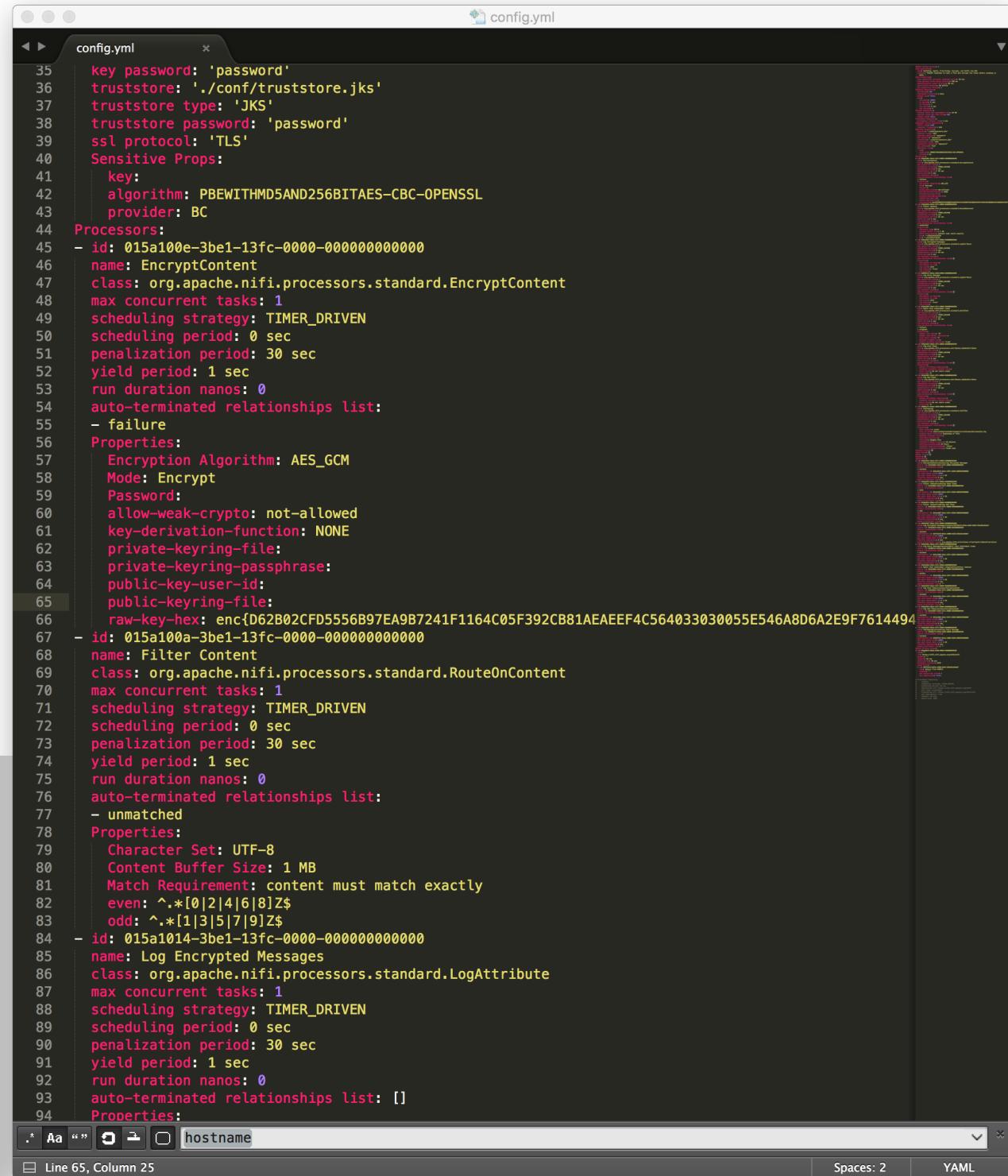
# MiNiFi Converter Toolkit

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

```
▶ 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 $
```

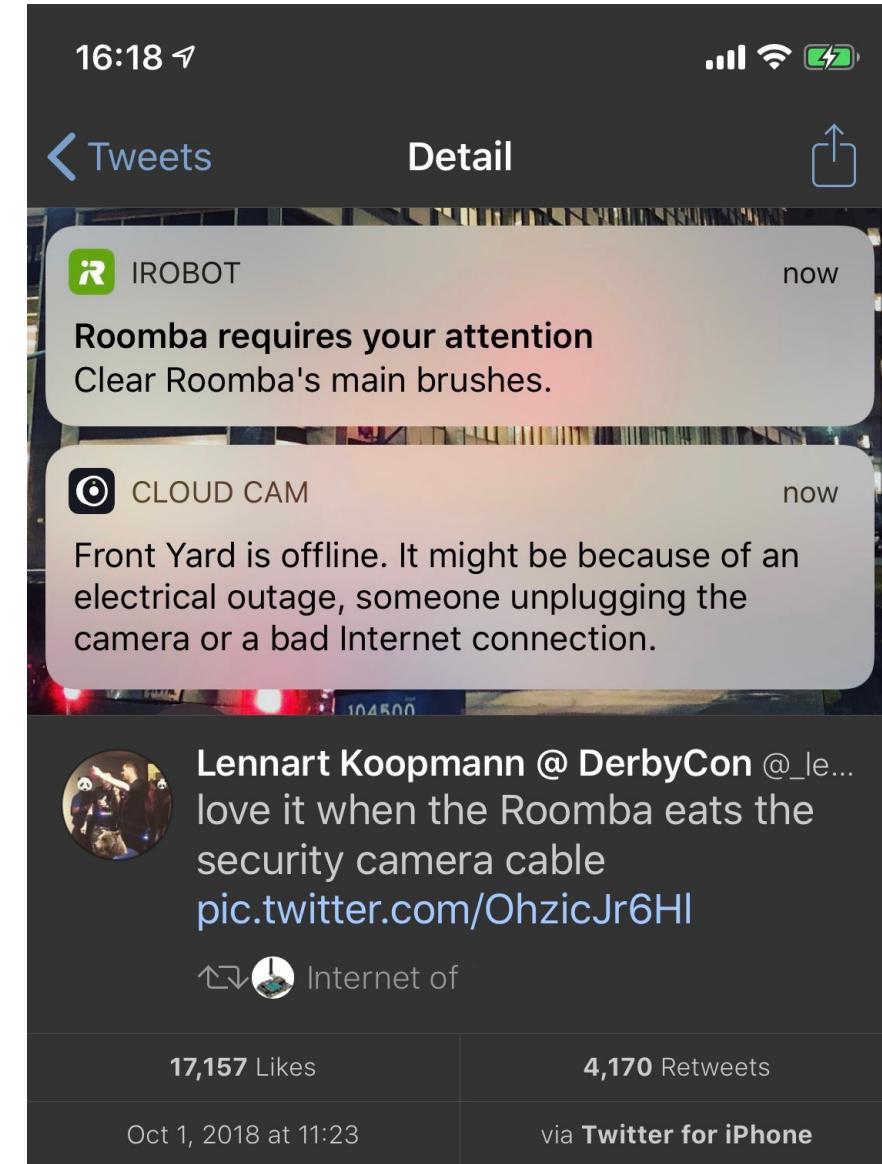


```
config.yml
35 key password: 'password'
36 truststore: './conf/truststore.jks'
37 truststore type: 'JKS'
38 truststore password: 'password'
39 ssl protocol: 'TLS'
40 Sensitive Props:
41   key:
42     algorithm: PBEWITHMD5AND256BITAES-CBC-OPENSSL
43     provider: BC
44 Processors:
45 - id: 015a100e-3be1-13fc-0000-000000000000
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:
```

# Internet of Things

# 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

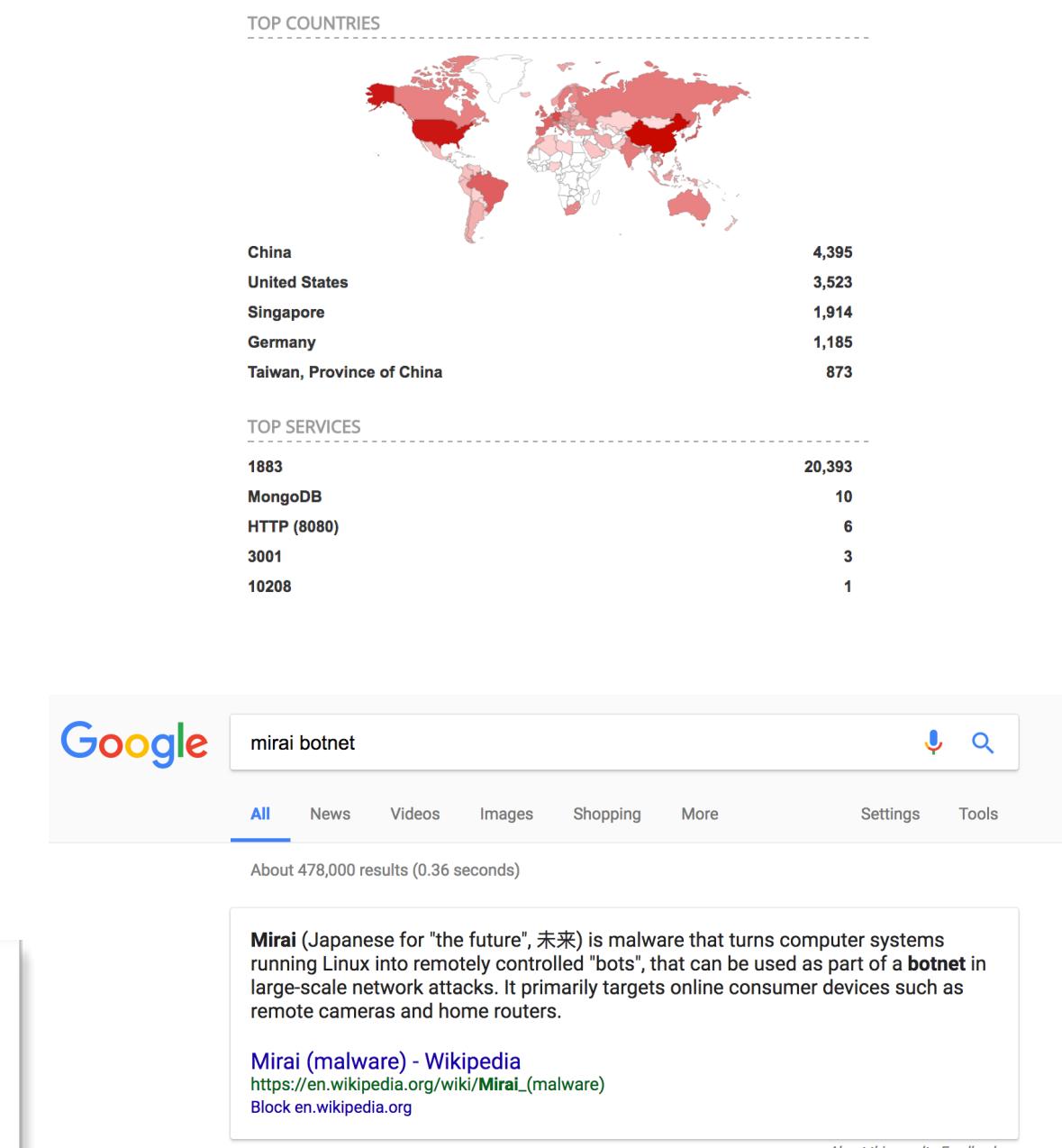
- When the Mirai attack has its own Wikipedia page, that's not good
- Hackers stole high-roller database from casino via aquarium thermometer connected to internet (04/2018)



**IoTPOT: Analysing the Rise of IoT Compromises**

Yin Minn Pa Pa<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





“The S in IoT stands  
for security”

Oleg Šelajev, @shelajev

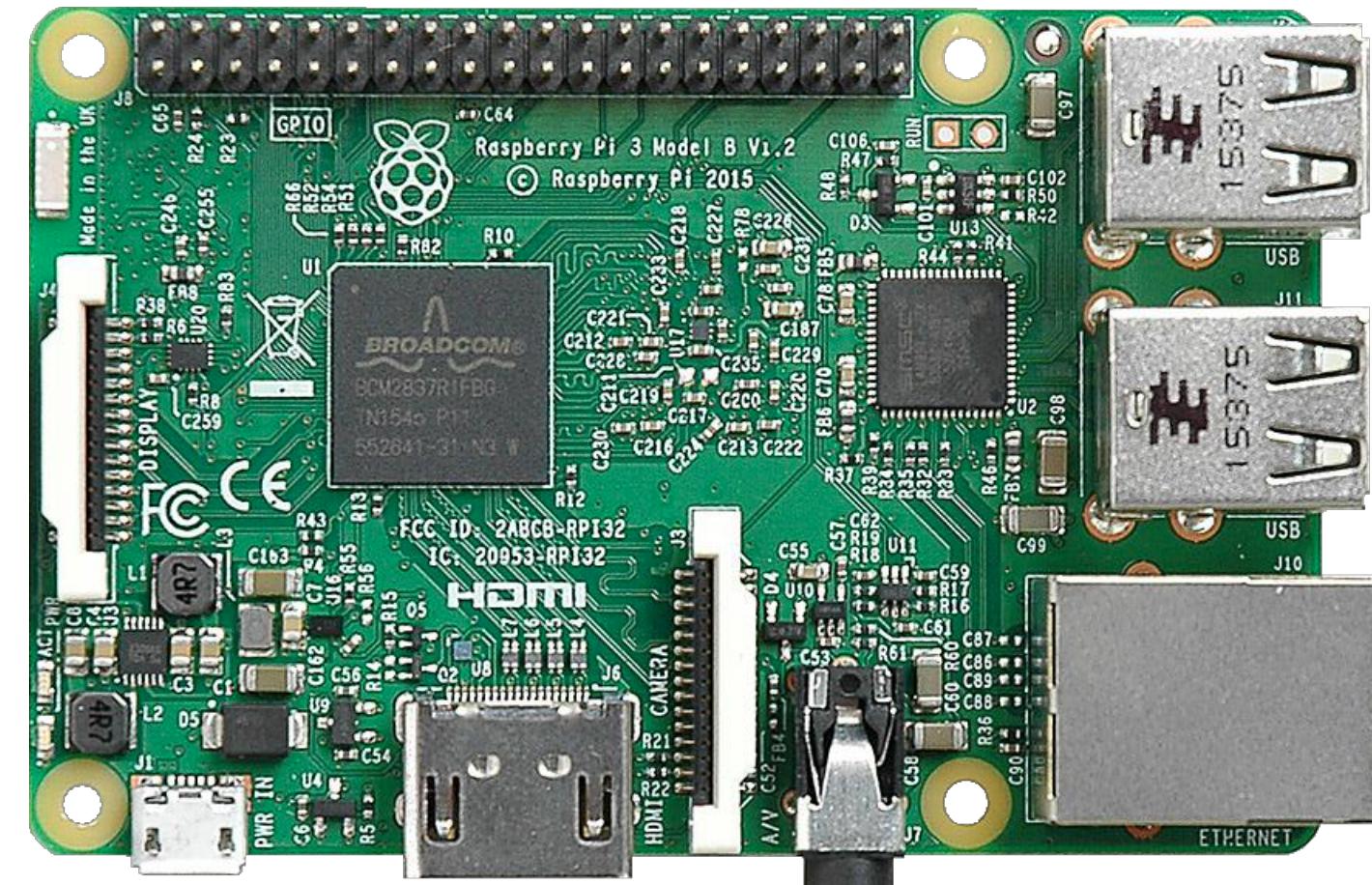
# NiFi Solves Everything\*

- 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
```

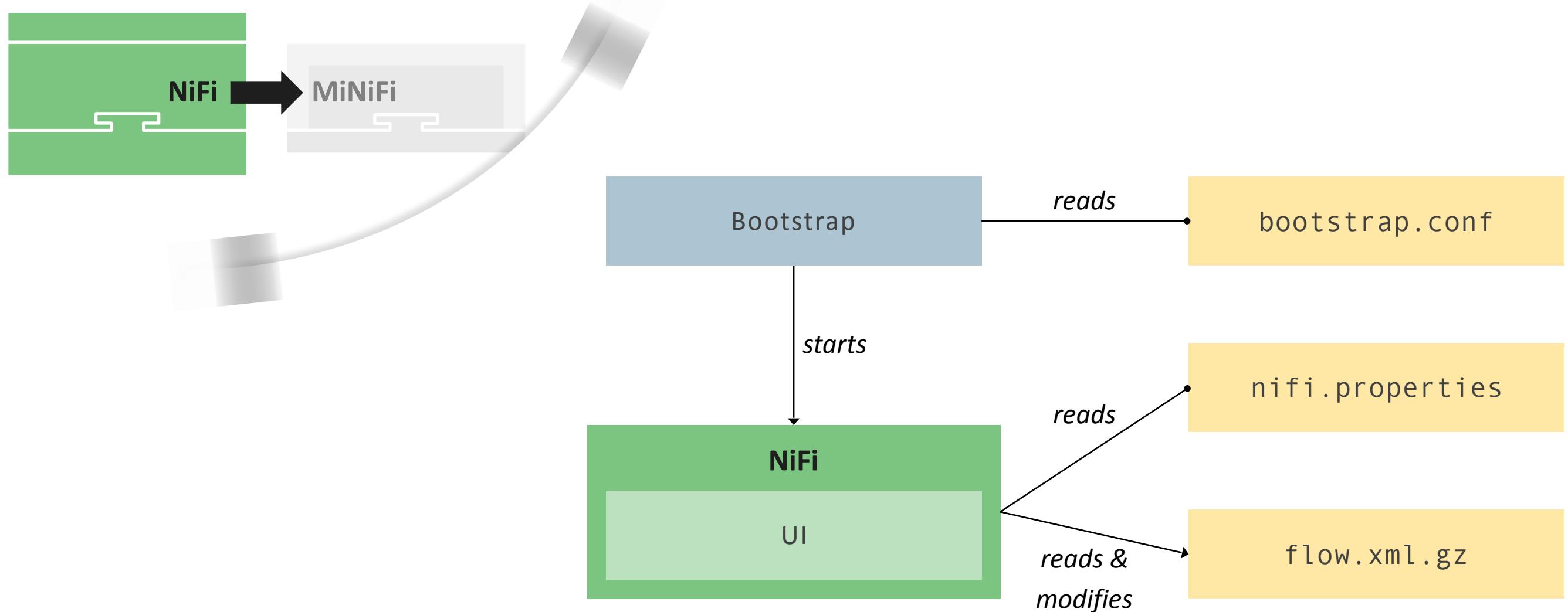


# So Why Do We Need A Different Solution?

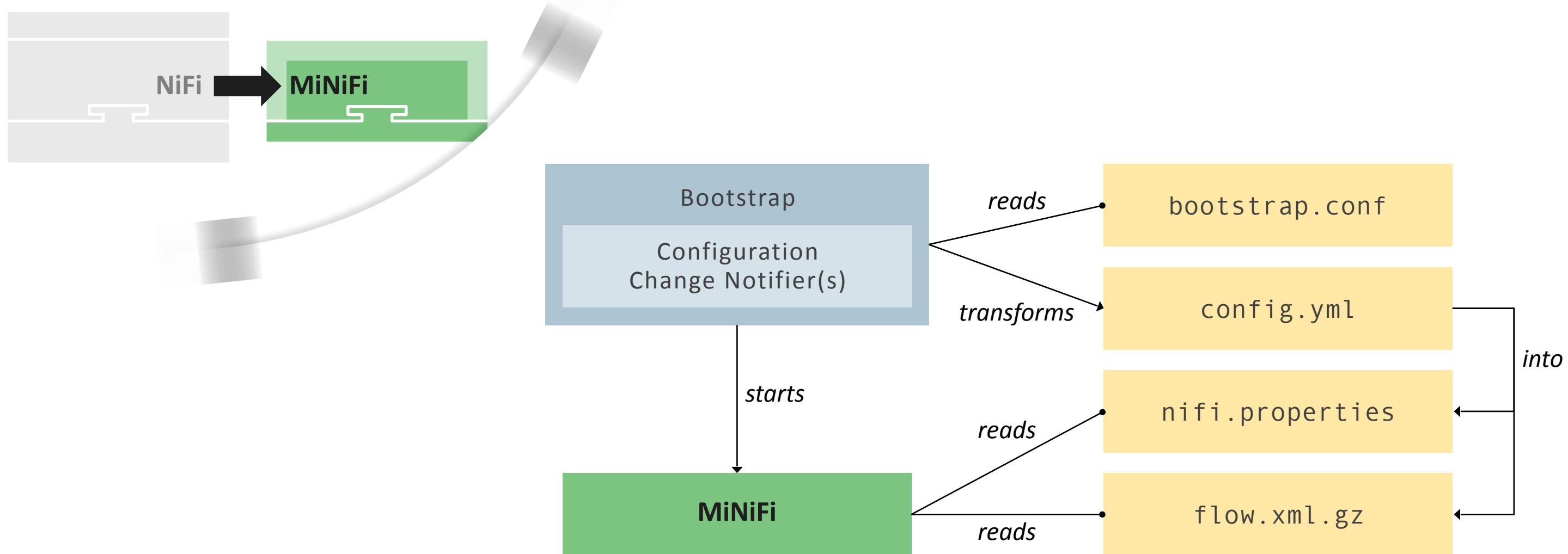
- NiFi is designed to “own the box”
- NiFi 0.7.x started up in about 10-15 minutes on RP3 (593 MB)
- NiFi 1.x started up in about 30 minutes on RP3 (760 MB)
  - 33 new processors
  - Rewrite for multi tenant authorization
  - Complete UI overhaul

```
▶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 $ █
```

# NiFi Java Processes



# MiNiFi Java Processes



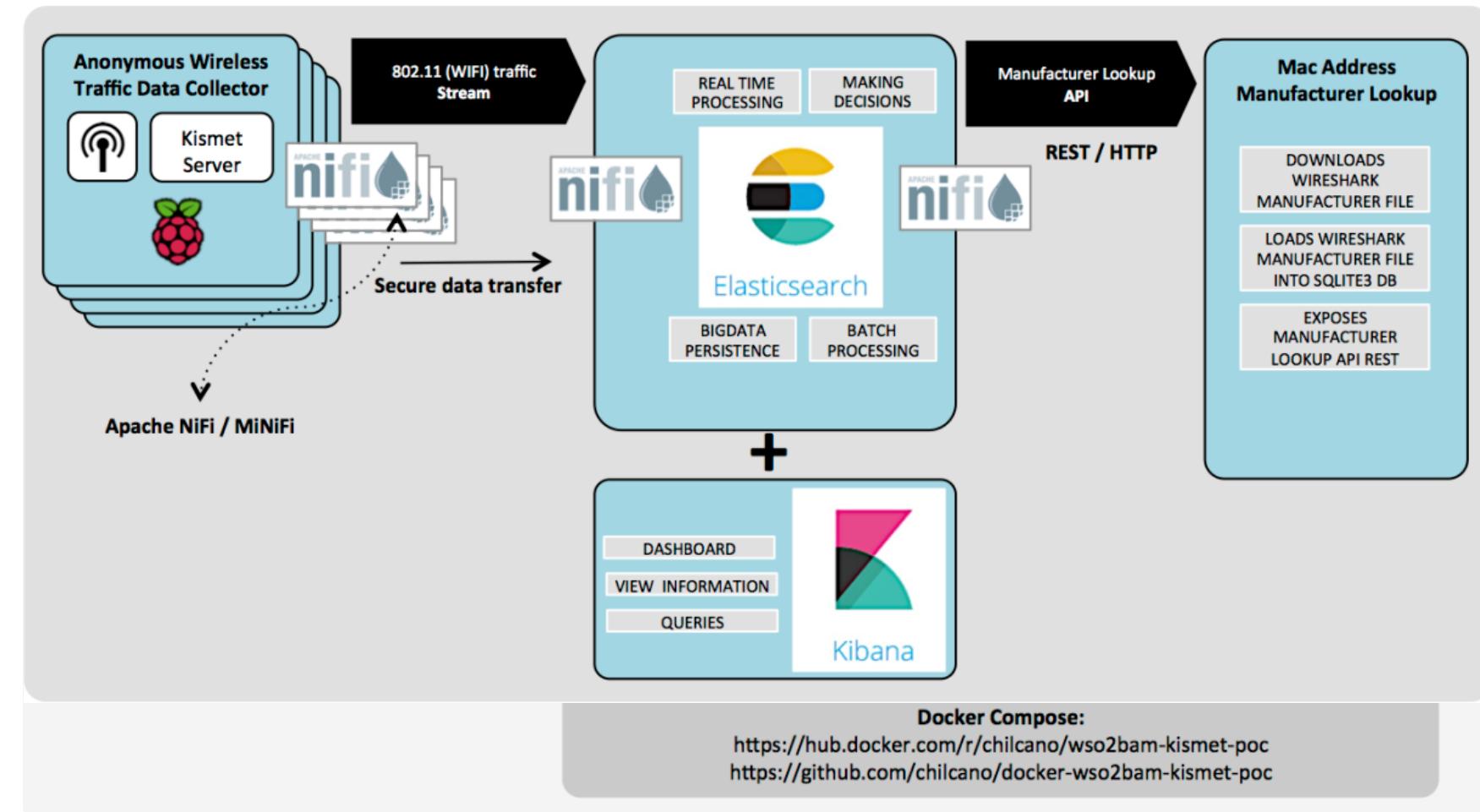
# Community Example

- Roger Carhuatocto
- RP3B, Apache Thrift, Kismet, Apache Cassandra, and more

- Large scaling.
  - I would like to avoid to deploy Logstash in each Raspberry Pi just to transform in JSON the captured 802.11 (WIFI) traffic and send it to Elasticsearch. Other approach what I want to avoid is to deploy Logstash with the UDP/TCP Input Plugin in the Elasticsearch side, because both choices need parse/transform/filter the captured traffic by using [GROK](#) and Elasticsearch Index Templates for each Logstash instance deployed. What if I have 100 or more Raspberry Pi distributed in different locations?.
- Security.
  - I'm using Kismet installed in each Raspberry Pi to capture 802.11 traffic, by default Kismet sends that traffic over UDP, UDP is faster but not secure. The big problem with Logstash listening UDP traffic over a port is that Logstash is susceptible to DoS attacks and the traffic to be spoofed. I have to update UDP to the "secure UDP", UDP over SSL/TLS for example.
- Monitoring/Tracking.
  - How to monitor if Kismet is running in the Raspberry Pi?, How to know if Raspberry Pi is healthy ?.
- Administrable remotely.
  - Definitely I can't do that in a massively distributed Raspberry Pi's.

Then, what can I do ?....

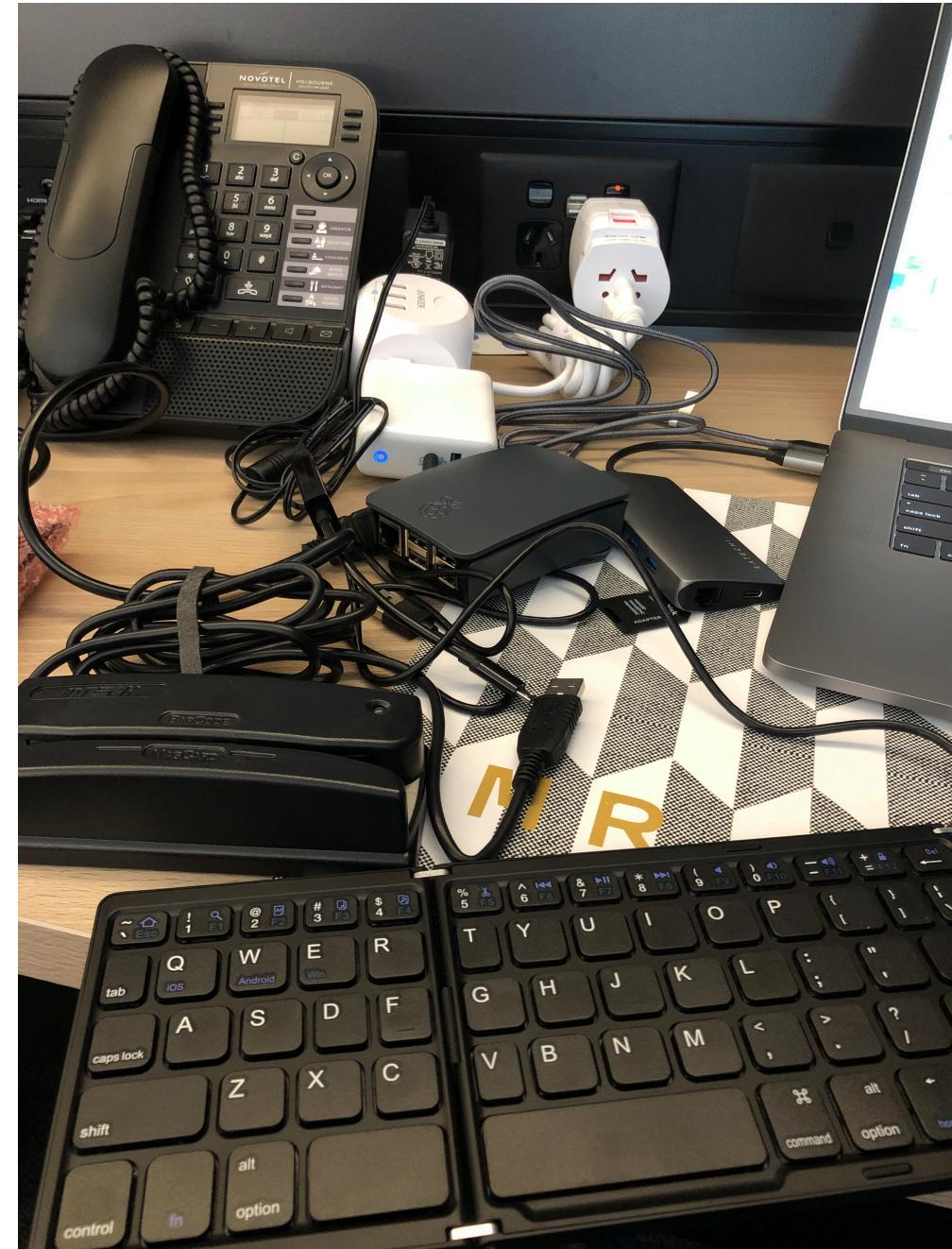
Apache NiFi to the rescue!



# Demo - Credit Cards

# Collect Card Contents

- User swipes magnetic stripe card
- Raspberry Pi collects data
- Sends back to laptop



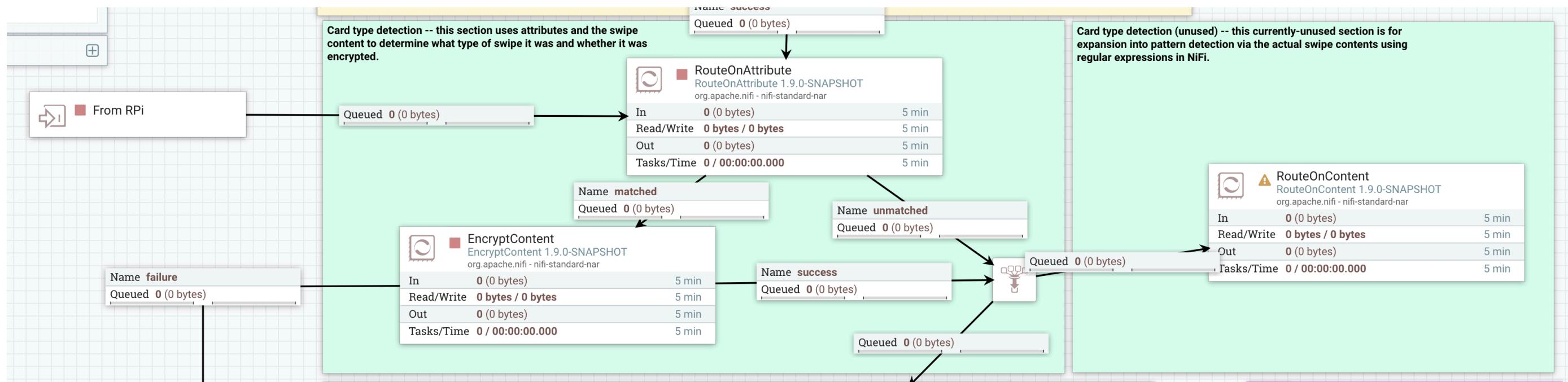
# Component List

- IDTECH WCR3237-733U
- Raspberry Pi 3 B+
- Apache MiNiFi 0.6.0-SNAPSHOT
- Apache NiFi 1.9.0-SNAPSHOT
- Apache NiFi Registry 0.4.0-SNAPSHOT

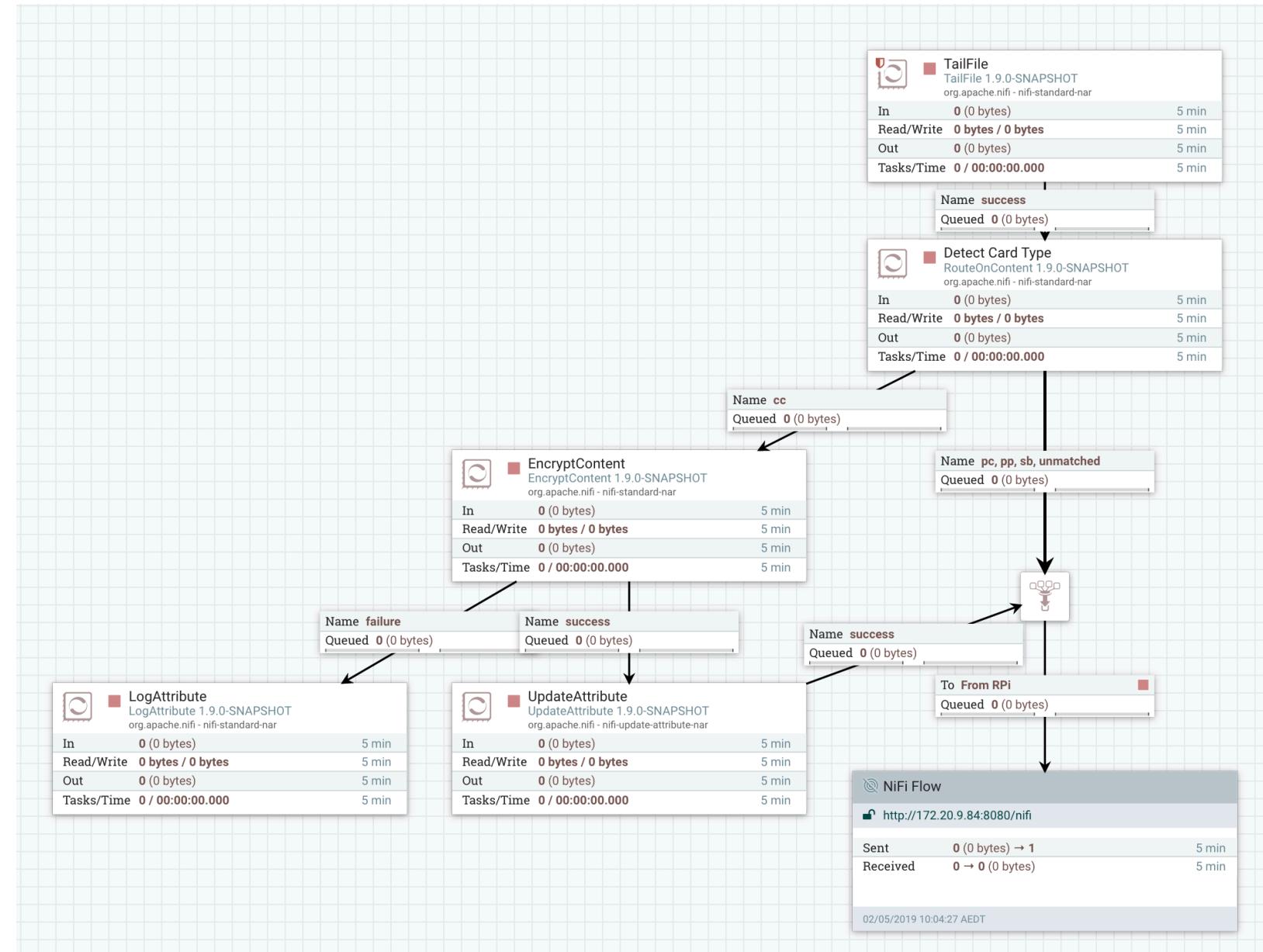
# Flow Design

- NiFi Flow
  - Receive data via S2S
  - Route encrypted CC data to decrypt
  - Route all plaintext swipes through parser
  - Convert to JSON
  - Batch & write to file
- MiNiFi Flow
  - Tail file containing swipe data
  - Detect CC swipes & encrypt
  - Transmit all swipes to NiFi via S2S

# NiFi Flow



# MiNiFi Flow



# Custom MSR Parsing Logic

```
1 import groovy.json.JsonBuilder
2 import groovy.json.JsonOutput
3 import groovy.json.JsonSlurper
4 import org.apache.commons.io.IOUtils
5 import org.apache.nifi.processor.io.StreamCallback
6
7 import java.nio.charset.StandardCharsets
8
9 def flowFile = session.get()
10 if (flowFile == null) {
11     return
12 }
13
14 final String CC_PATTERN = /%B\d{16}\^A-Za-z+/
15 final String PC_PATTERN = /%1w+/
16 final String PP_PATTERN = /%PP\//\w+/
17 final String SBUX_PATTERN = /%B601.*SERVICERECOVERY/
18
19 String determinedType = "unknown"
20
21 flowFile = session.write(flowFile, { inputStream, outputStream ->
22     def content = IOUtils.toString(inputStream, StandardCharsets.UTF_8)
23     Map parsedObject
24
25     // Logic to do various parsing
26     if (content =~ CC_PATTERN) {
27         log.info("Detected CC format")
28         parsedObject = parse(content, detectedType: "CC")
29     } else if (content =~ PP_PATTERN) {
30         log.info("Detected Priority Pass format")
31         parsedObject = parse(content, detectedType: "PP")
32     } else if (content =~ PC_PATTERN) {
33         log.info("Detected Players Club format")
34         parsedObject = parse(content, detectedType: "PC")
35     } else if (content =~ SBUX_PATTERN) {
36         log.info("Detected Starbucks Gift Card format")
37         parsedObject = parse(content, detectedType: "SB")
38     } else {
39         log.info("Could not determine MSR type; using generic parsing heuristics")
40         parsedObject = parse(content, detectedType: "generic")
41     }
42
43     // Write the determined type to an attribute
44     determinedType = parsedObject["type"]
45
46     // Write the timing info to the parsed object
47     parsedObject["time_parsed"] = new Date().format(format: 'YYYY-MM-dd_HH-mm-ss.SSS Z')
48     parsedObject["id"] = flowFile.uuid
49     parsedObject["source"] = flowFile.filename
50
51     // Convert the object into JSON
52     def jsonString = JsonOutput.toJson(parsedObject)
53     log.info("Formed JSON: ${jsonString}")
54
55     outputStream.write(jsonString.getBytes(StandardCharsets.UTF_8))
56
57 } as StreamCallback)
58 flowFile.determinedType = determinedType
```

```
59
60     private Map parse(String content, String detectedType) {
61         Map parsedObject = [:]
62         parsedObject["type"] = detectedType
63
64         def tracks = content.split( regex: /\?\n/ )
65         log.info("Tracks ${tracks.size()}: ${tracks.collect { "\n\t${it}" }}")
66
67         // Put the track 1 data directly
68         parsedObject["track1"] = tracks[0]
69
70         def track1Elements = tracks[0].split( regex: /\^/ )
71         log.info("Track 1 Elements ${track1Elements.size()}: ${track1Elements.collect { "\n\t${it}" }}")
72
73         // The PAN and Name are common elements
74         parsedObject["pan"] = track1Elements[0][2..-1] // Skip the start sentinel and format code (%B)
75         if (track1Elements.size() > 1) {
76             parsedObject["name"] = track1Elements[1].trim()
77         }
78         if (track1Elements.size() > 2) {
79             // Put the remaining track 1 data in a generic field
80             parsedObject["additionalTrack1"] = track1Elements[2..-1].join("^")
81         }
82
83         // Put track 2 data directly
84         parsedObject["track2"] = tracks[1]
85
86         // Return the parsed object
87         parsedObject
88
89
90     session.transfer(flowFile, REL_SUCCESS)
```

# Running on Raspberry Pi

LXTerminal

File Edit Tabs Help

```
pi@gobias:/opt/iot$ ./prompt.sh
Swipe your card : : pi@gobias:/opt/iot$ more swipes.txt
%PP//ANDREW/D/LOPREST0?\n;7357450001927901=102018?
pi@gobias:/opt/iot$
```

0 items

LXTerminal

File Edit Tabs Help

```
459-510.txt,size=51]
-----[Transmitting data]-----
Standard FlowFile Attributes
Key: 'entryDate'
    Value: 'Tue Feb 05 14:43:32 AEDT 2019'
Key: 'lineageStartDate'
    Value: 'Tue Feb 05 14:43:32 AEDT 2019'
Key: 'fileSize'
    Value: '51'
FlowFile Attribute Map Content
Key: 'RouteOnContent.Route'
    Value: 'pp'
Key: 'filename'
    Value: 'swipes.459-510.txt'
Key: 'mime.type'
    Value: 'text/plain'
Key: 'path'
    Value: './'
Key: 'tailfile.original.path'
    Value: '/opt/iot/swipes.txt'
Key: 'uuid'
    Value: '3dc63eb3-a3fd-4e47-bfce-46dc7792d4a3'
-----[Transmitting data]-----
%PP//ANDREW/D/LOPREST0?\n;7357450001927901=102018?
```

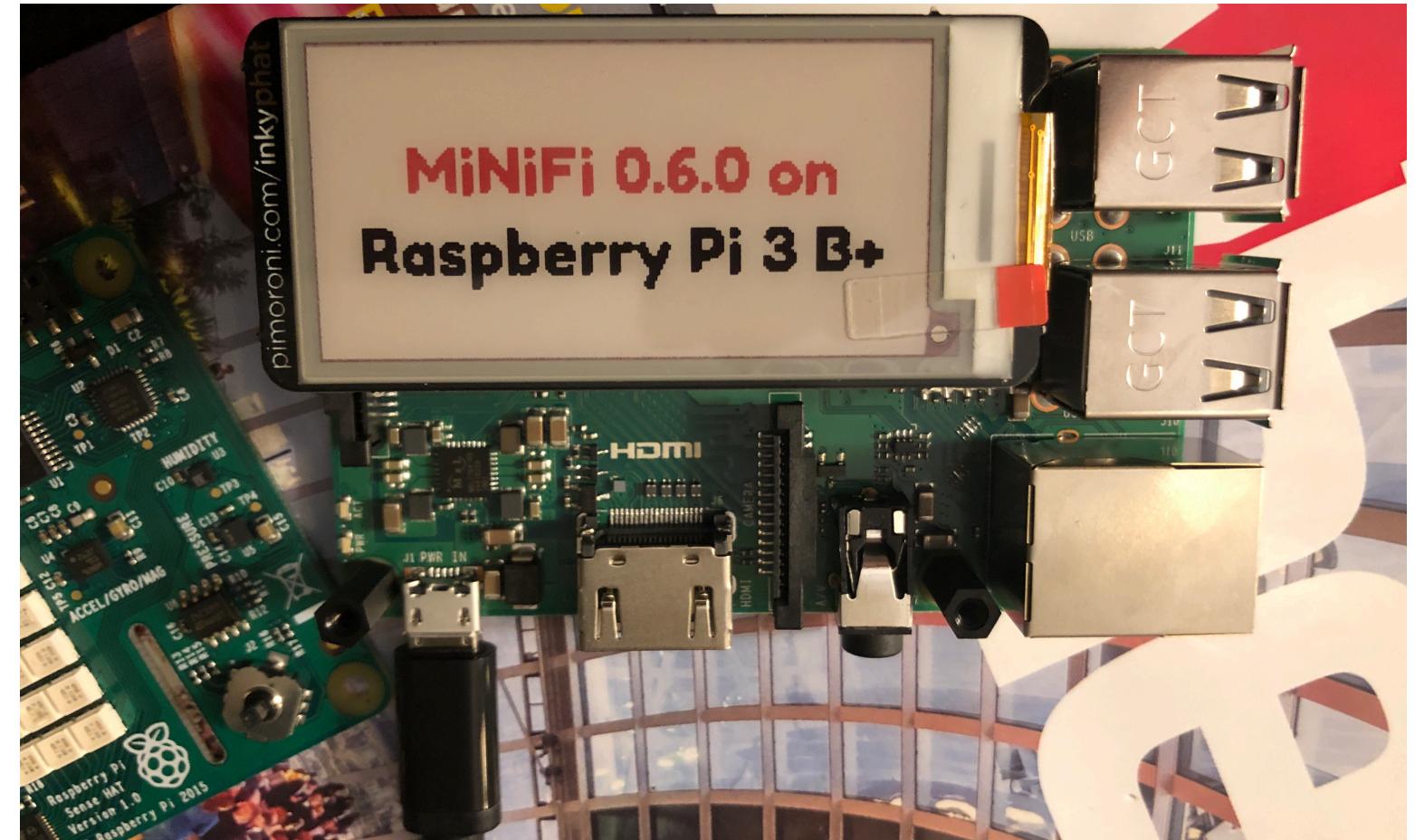
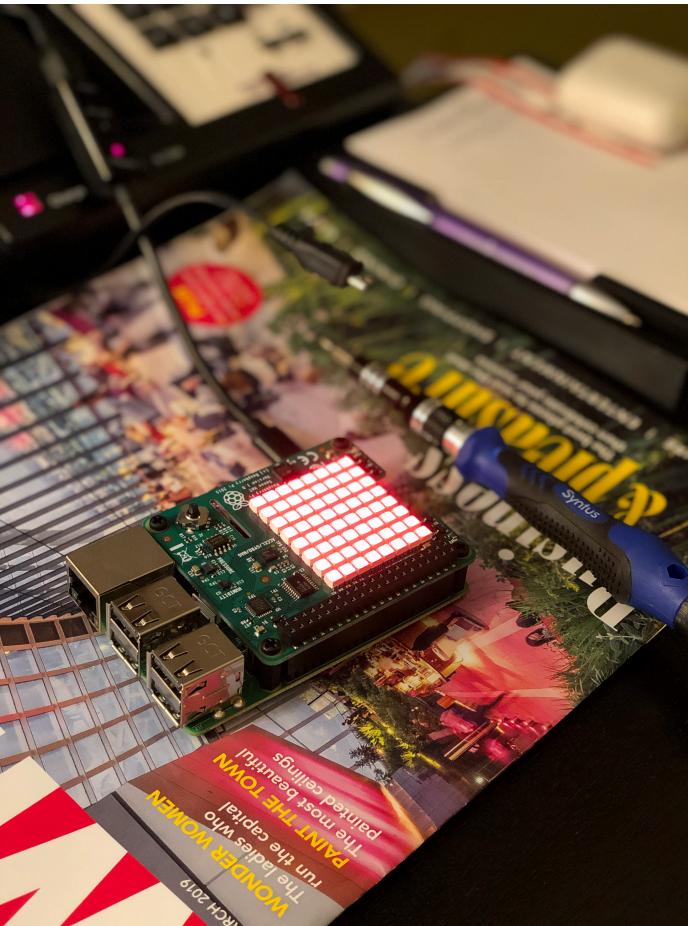
Free spa

# Running on MacBook Pro

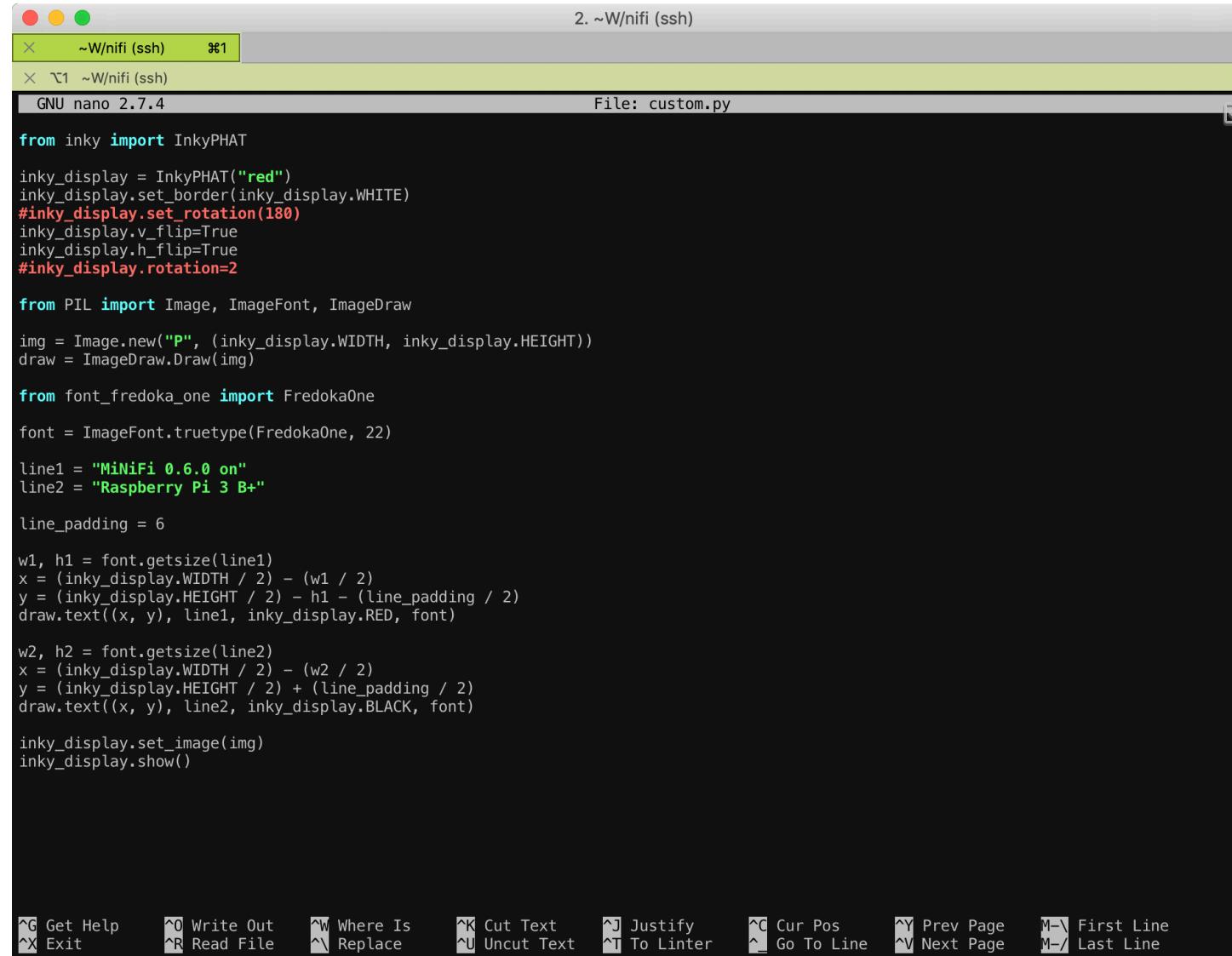
```
× 2 scratch/iot_ms (more)
drwxr-xr-x 8 alopreno staff 256B Feb 5 14:47 .
drwxr-xr-x 180 alopreno staff 5.6K Feb 3 14:17 ..
-rw-r--r-- 1 alopreno staff 2.9K Feb 3 14:19 2019-02-03_14-19-24_824_+1100.json
-rw-r--r-- 1 alopreno staff 2.9K Feb 3 14:19 2019-02-03_14-19-29_828_+1100.json
-rw-r--r-- 1 alopreno staff 3.5K Feb 3 14:19 2019-02-03_14-19-34_836_+1100.json
-rw-r--r-- 1 alopreno staff 3.4K Feb 3 14:19 2019-02-03_14-19-40_837_+1100.json
-rw-r--r-- 1 alopreno staff 3.7K Feb 3 14:19 2019-02-03_14-19-46_849_+1100.json
-rw-r--r-- 1 alopreno staff 3.4K Feb 3 14:19 2019-02-03_14-19-52_853_+1100.json
> . . . sers/alopreno/Workspace/scratch/iot_ms (master) 😊
os @ 14:49:01 $ more 2019-02-03_14-19-52_853_+1100.json
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-46.884 +1100", "id": "47fc007e-aa24-47b5-94aa-7793acb8cbe4", "source": "564a9f2f-74b8-4a47-9cad-df8742e4a02a"}  
{"type": "PP", "track1": "%PP//ANDREW/D/LOPREST0", "pan": "P//ANDREW/D/LOPREST0", "track2": "7357450001927901=102018?", "time_parsed": "2019-02-03_14-19-47.862 +1100", "id": "ffea2add-03e6-49d8-b0ea-570ff7485d0c", "source": "739599e1-2286-440a-8635-623206f589cf"}  
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-47.863 +1100", "id": "948ba9d7-aada-462f-8e37-fb9ab86515c9", "source": "70c3b55e-ce4e-49b6-b818-e3308eee6f80"}  
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-48.877 +1100", "id": "5bb6bead-a1c2-4b90-883d-532873d9204d", "source": "323b7c10-21d1-439f-9ce6-20b7afa233c1"}  
{"type": "CC", "track1": "%B4147098571707173^LOPREST0/ANDREW ^170510100000000014051514000000", "pan": "4147098571707173", "name": "LOPREST0/ANDREW", "additionalTrack1": "170510100000000014051514000000", "track2": "4147098571707173=170510114051514?", "time_parsed": "2019-02-03_14-19-48.899 +1100", "id": "2b30ccfe-14d1-4a34-bd8c-72200d013ff0", "source": "9ceb7d41-d863-4187-87d4-64607471539c"}  
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-49.867 +1100", "id": "5738eab2-8042-4d15-bc4f-fa23171f9f87", "source": "963d9b68-8f8a-4cc2-a51f-d4c8aeb63250"}  
{"type": "PP", "track1": "%PP//ANDREW/D/LOPREST0", "pan": "P//ANDREW/D/LOPREST0", "track2": "7357450001927901=102018?", "time_parsed": "2019-02-03_14-19-49.868 +1100", "id": "dbfa5730-c689-47a3-8583-bd90b6adb0da", "source": "a29256ed-a299-4bb2-8f48-0e1be944a16f"}  
{"type": "SB", "track1": "%B6010569709984040^0254/SERVICERECOVERYUSD^00010004000060144271. ", "pan": "6010569709984040", "name": "0254/SERVICERECOVERYUSD", "additionalTrack1": "00010004000060144271. ", "track2": "6010569709984040=00010004000060144271?", "time_parsed": "2019-02-03_14-19-49.866 +1100", "id": "96e37f83-af28-4b53-a86e-fcb8ba3184ab", "source": "0cd54ca1-7efb-4e42-b600-979c56d0b6b3"}  
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-50.879 +1100", "id": "91bb2200-3b65-4659-a8ad-33943a22f335", "source": "6730a23a-5106-4ebb-b3d9-bae3b896f4c1"}  
{"type": "CC", "track1": "%B4147098571707173^LOPREST0/ANDREW ^170510100000000014051514000000", "pan": "4147098571707173", "name": "LOPREST0/ANDREW", "additionalTrack1": "170510100000000014051514000000", "track2": "4147098571707173=170510114051514?", "time_parsed": "2019-02-03_14-19-51.868 +1100", "id": "882a0255-81c4-41ed-9b34-9856557509ec", "source": "d3bd20ee-f74a-4a01-9676-9dd0384fca86"}  
{"type": "PC", "track1": "%1AANDREW. 01037983102300", "pan": "AANDREW. 01037983102300", "track2": "01037983102300000?", "time_parsed": "2019-02-03_14-19-51.867 +1100", "id": "1dacea93-59e1-4205-b77a-6872907c0013", "source": "dc1e8196-16a4-421f-870a-b369b4bbe099"}  
{"type": "PP", "track1": "%PP//ANDREW/D/LOPREST0", "pan": "P//ANDREW/D/LOPREST0", "track2": "7357450001927901=102018?", "time_parsed": "2019-02-03_14-19-51.869 +1100", "id": "18c23941-7665-49f5-a99a-d157de6d828f", "source": "6b1de415-1578-4be9-b872-ce6216d86cc9"}  
2019-02-03_14-19-52_853_+1100.json (END)
```

# Demo - Sense HAT & Inky pHAT

# Sense Hat & Inky pHAT



# Inky pHAT Python Code & VNC



```
2. ~W/nifi (ssh)
~W/nifi (ssh) 2. ~W/nifi (ssh)
GNU nano 2.7.4
File: custom.py

from inky import InkyPHAT
inky_display = InkyPHAT("red")
inky_display.set_border(inky_display.WHITE)
#inky_display.set_rotation(180)
inky_display.v_flip=True
inky_display.h_flip=True
#inky_display.rotation=2

from PIL import Image, ImageFont, ImageDraw

img = Image.new("P", (inky_display.WIDTH, inky_display.HEIGHT))
draw = ImageDraw.Draw(img)

from font_fredoka_one import FredokaOne

font = ImageFont.truetype(FredokaOne, 22)

line1 = "MiNiFi 0.6.0 on"
line2 = "Raspberry Pi 3 B+"

line_padding = 6

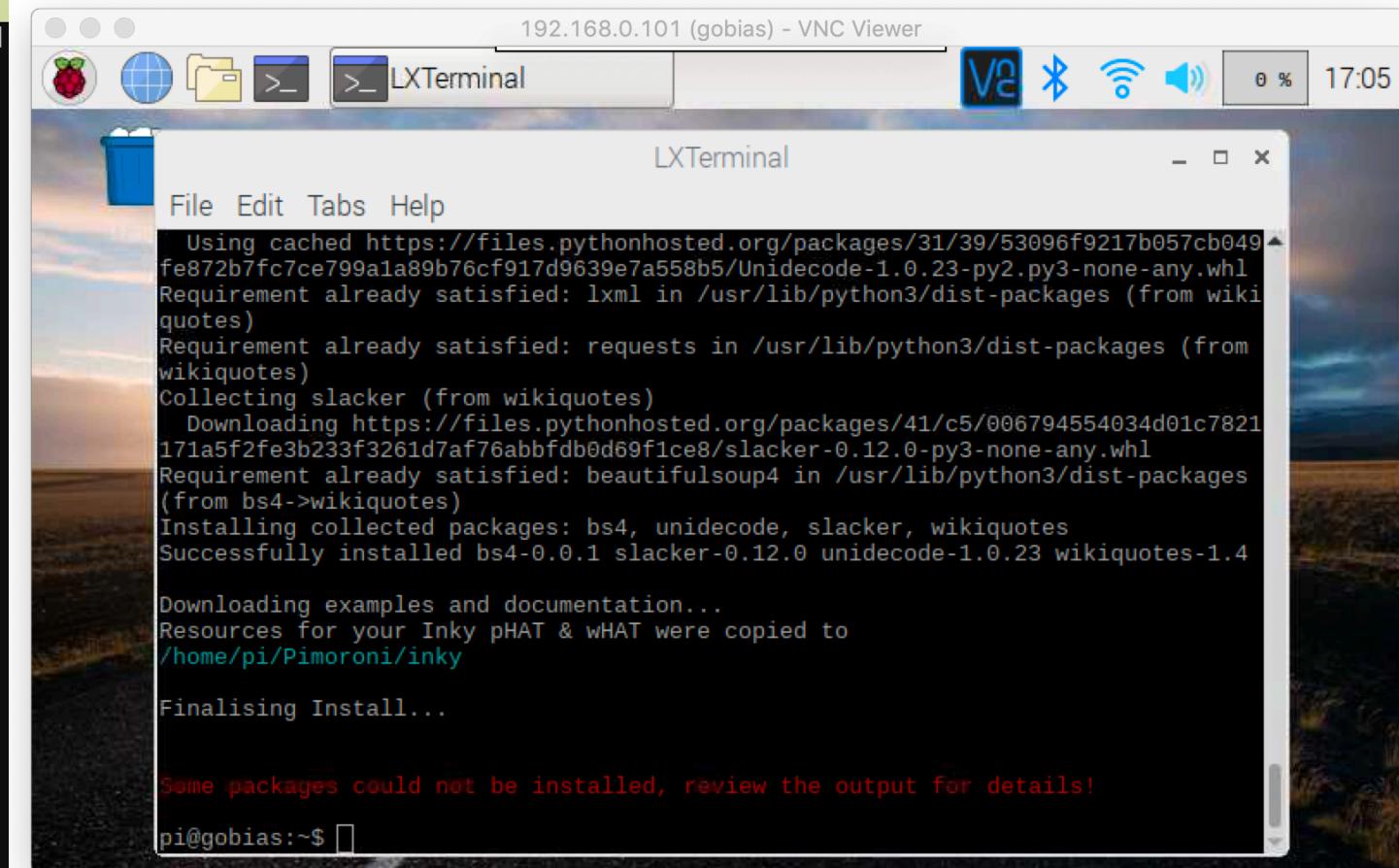
w1, h1 = font.getsize(line1)
x = (inky_display.WIDTH / 2) - (w1 / 2)
y = (inky_display.HEIGHT / 2) - h1 - (line_padding / 2)
draw.text((x, y), line1, inky_display.RED, font)

w2, h2 = font.getsize(line2)
x = (inky_display.WIDTH / 2) - (w2 / 2)
y = (inky_display.HEIGHT / 2) + (line_padding / 2)
draw.text((x, y), line2, inky_display.BLACK, font)

inky_display.set_image(img)
inky_display.show()

inky_display.set_image(img)
inky_display.show()
```

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos ^Y Prev Page M-\ First Line  
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^\_ Go To Line ^V Next Page M-/ Last Line



```
192.168.0.101 (gobias) - VNC Viewer
17:05
File Edit Tabs Help
Using cached https://files.pythonhosted.org/packages/31/39/53096f9217b057cb049fe872b7fc7ce799a1a89b76cf917d9639e7a558b5/Unidecode-1.0.23-py2.py3-none-any.whl
Requirement already satisfied: lxml in /usr/lib/python3/dist-packages (from wiki
quotes)
Requirement already satisfied: requests in /usr/lib/python3/dist-packages (from
wikiquotes)
Collecting slacker (from wikiquotes)
  Downloading https://files.pythonhosted.org/packages/41/c5/006794554034d01c7821
171a5f2fe3b233f3261d7af76abbfdb0d69f1ce8/slacker-0.12.0-py3-none-any.whl
Requirement already satisfied: beautifulsoup4 in /usr/lib/python3/dist-packages
(from bs4->wikiquotes)
Installing collected packages: bs4, unidecode, slacker, wikiquotes
Successfully installed bs4-0.0.1 slacker-0.12.0 unidecode-1.0.23 wikiquotes-1.4

Downloading examples and documentation...
Resources for your Inky pHAT & wHAT were copied to
/home/pi/Pimoroni/inky

Finalising Install...

Some packages could not be installed, review the output for details!
pi@gobias:~$
```

# Sense Hat & Inky pHAT



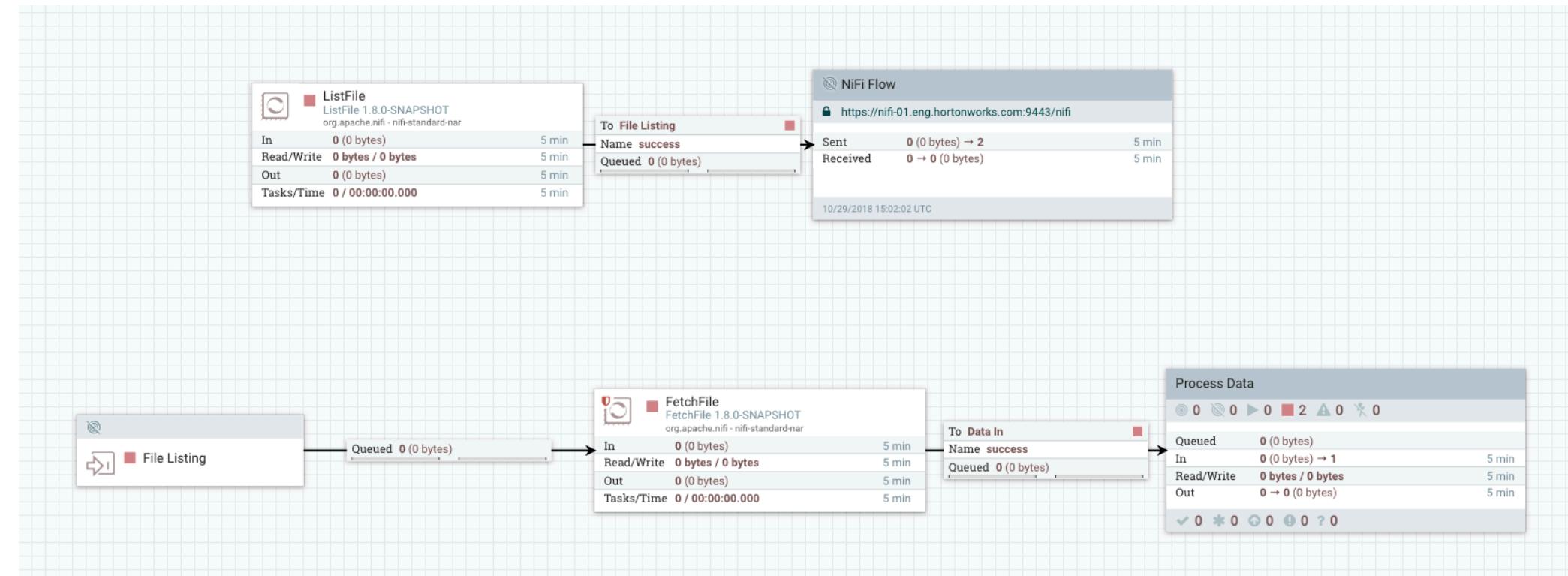
# New Features

# New Features in 1.8.0 - 1.9.1

- Cluster Data Management
  - **Load-Balanced Connections**
  - Node Decommissioning
- SQL results to record format
- Hot-loading NARs
- Auto-infer schema for record processors
- Docker improvements
- TLS Toolkit signing w/ external CA (standalone only)

# Previously, on *NiFi Clusters...*

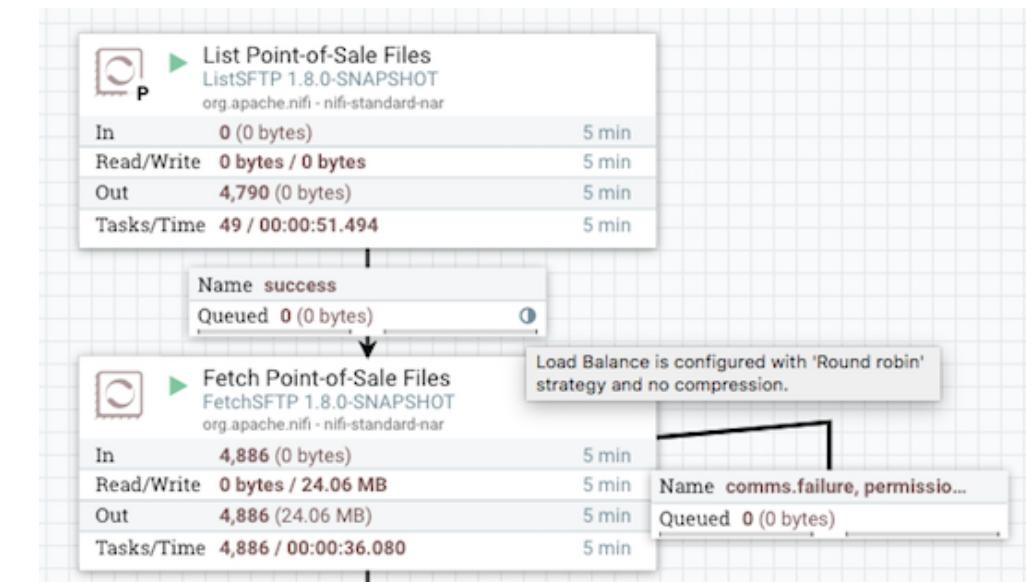
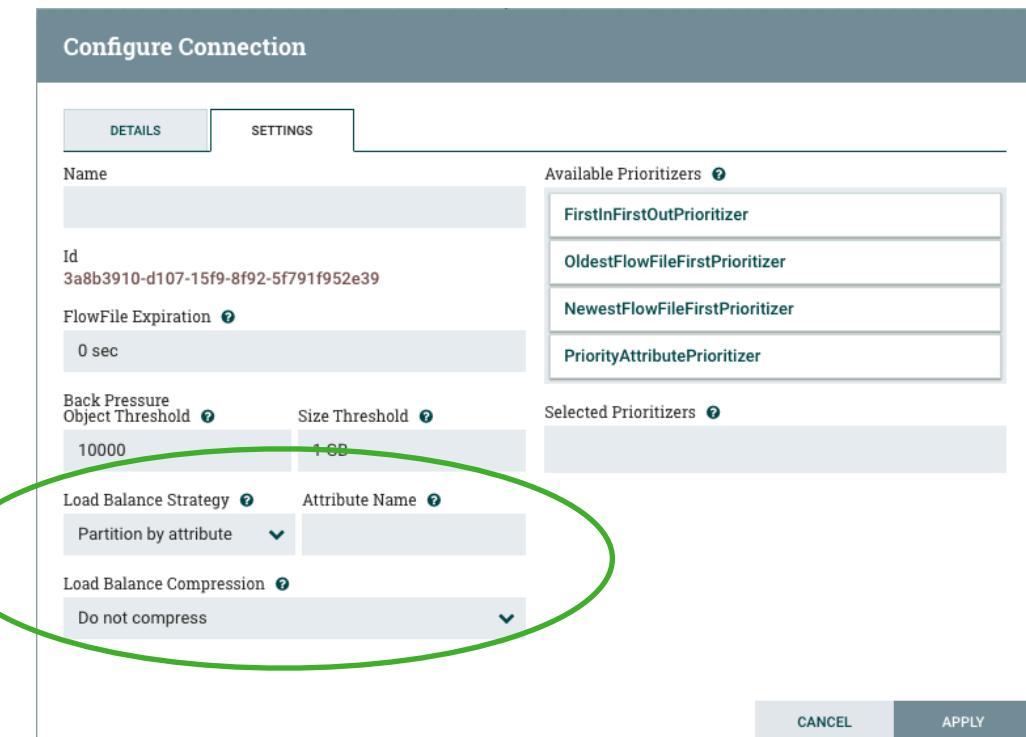
- Ingest activity (FTP, file read, etc.) performed on single node
- Routed via RPG to Input Port in cluster
- Flowfiles distributed across all nodes



[Mark Payne, Load Balancing Data Across a NiFi Cluster](#)

# Load-Balanced Connections

- Individual connections can be load-balanced
  - None
  - Round Robin
  - Attribute node affinity
  - Single node

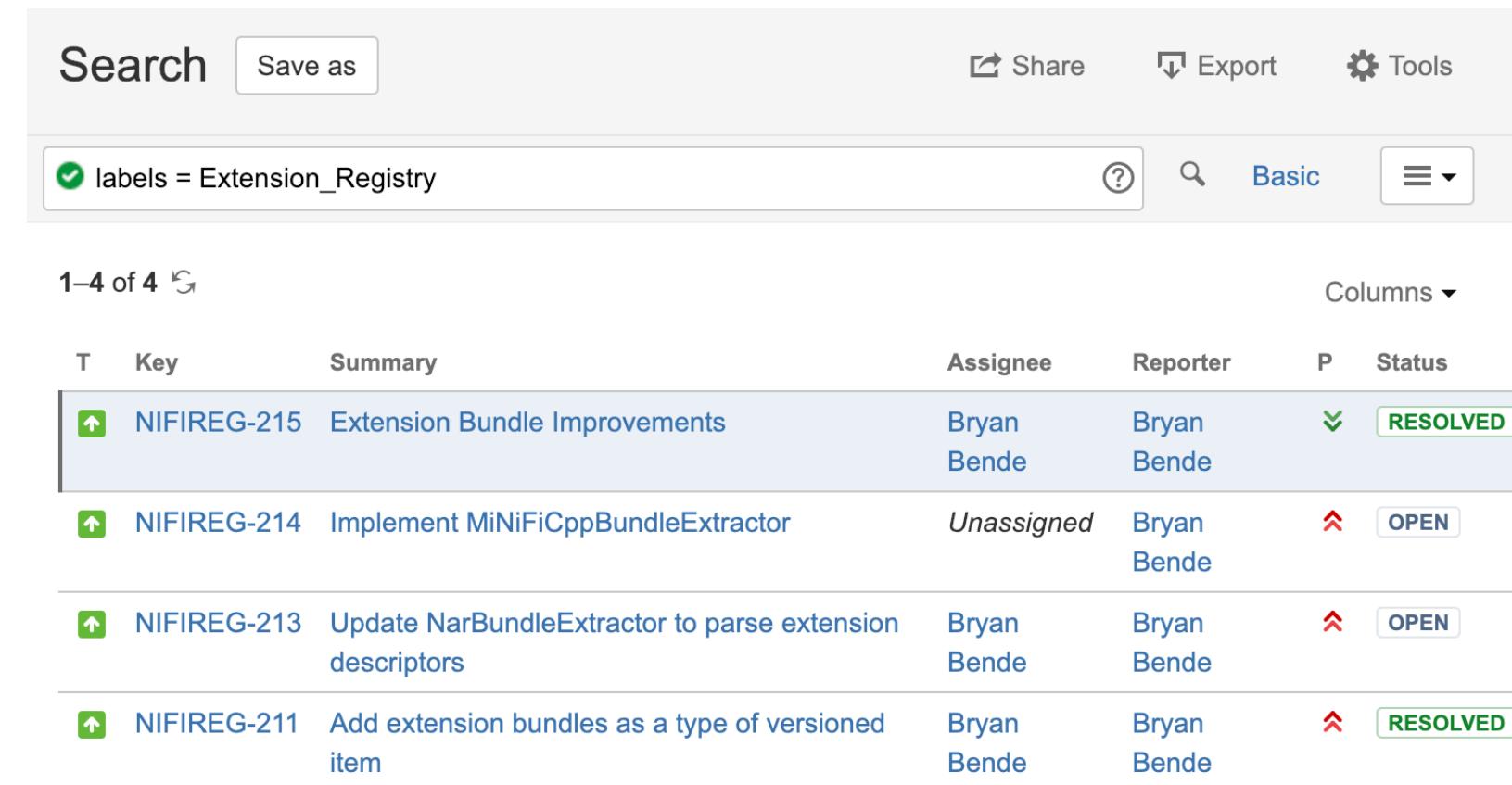


[Mark Payne, Load Balancing Data Across a NiFi Cluster](#)

# Roadmap

# Roadmap

- Command & Control Architecture
- Extension Registry



The screenshot shows a Jira search results page with the following details:

**Search:** labels = Extension\_Registry

**Results:** 1–4 of 4

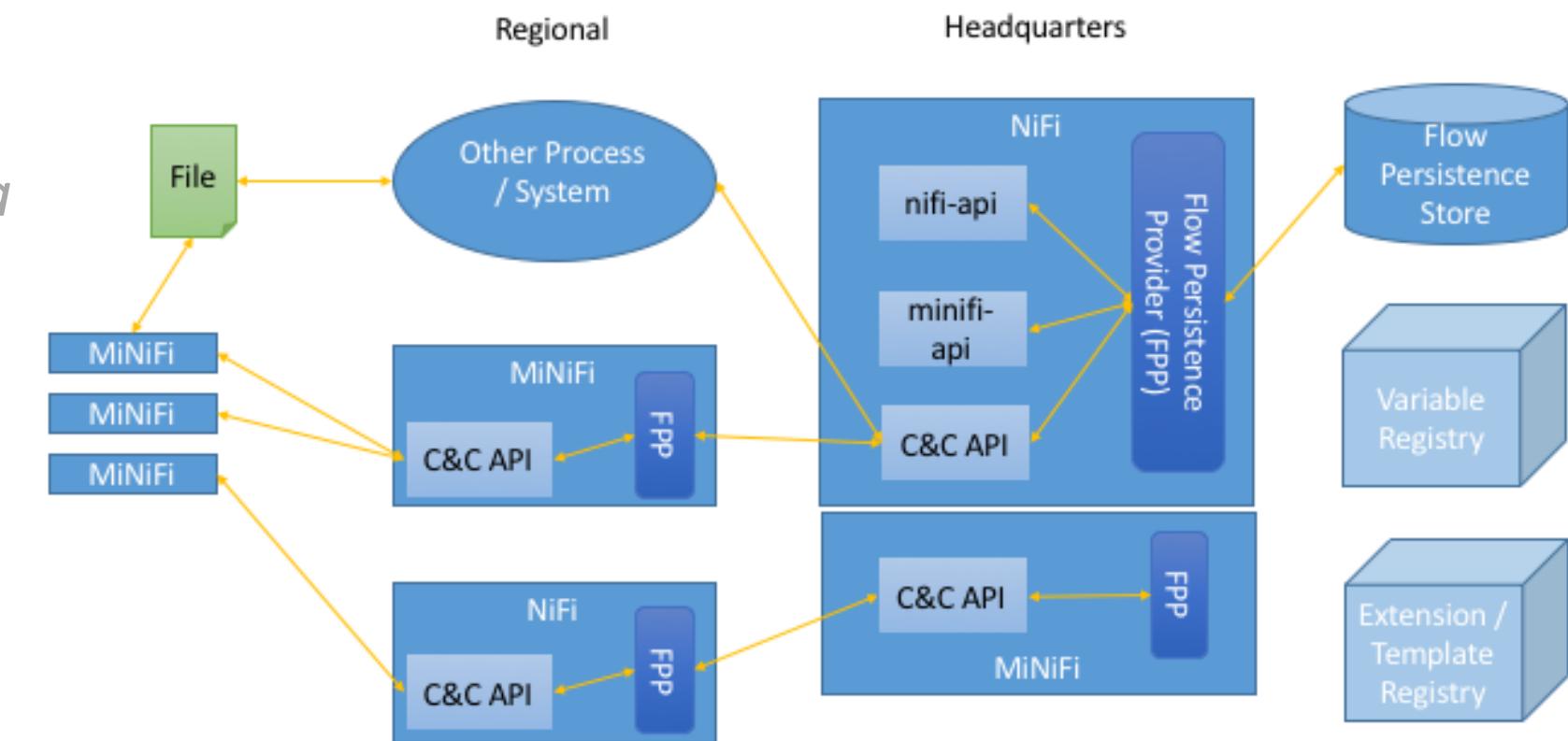
**Table Headers:** T, Key, Summary, Assignee, Reporter, P, Status

**Table Data:**

T	Key	Summary	Assignee	Reporter	P	Status
↑	NIFIREG-215	Extension Bundle Improvements	Bryan Bende	Bryan Bende	✓	RESOLVED
↑	NIFIREG-214	Implement MiNiFiCppBundleExtractor	Unassigned	Bryan Bende	✗	OPEN
↑	NIFIREG-213	Update NarBundleExtractor to parse extension descriptors	Bryan Bende	Bryan Bende	✗	OPEN
↑	NIFIREG-211	Add extension bundles as a type of versioned item	Bryan Bende	Bryan Bende	✗	RESOLVED

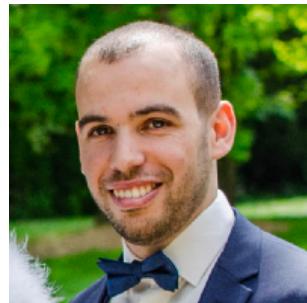
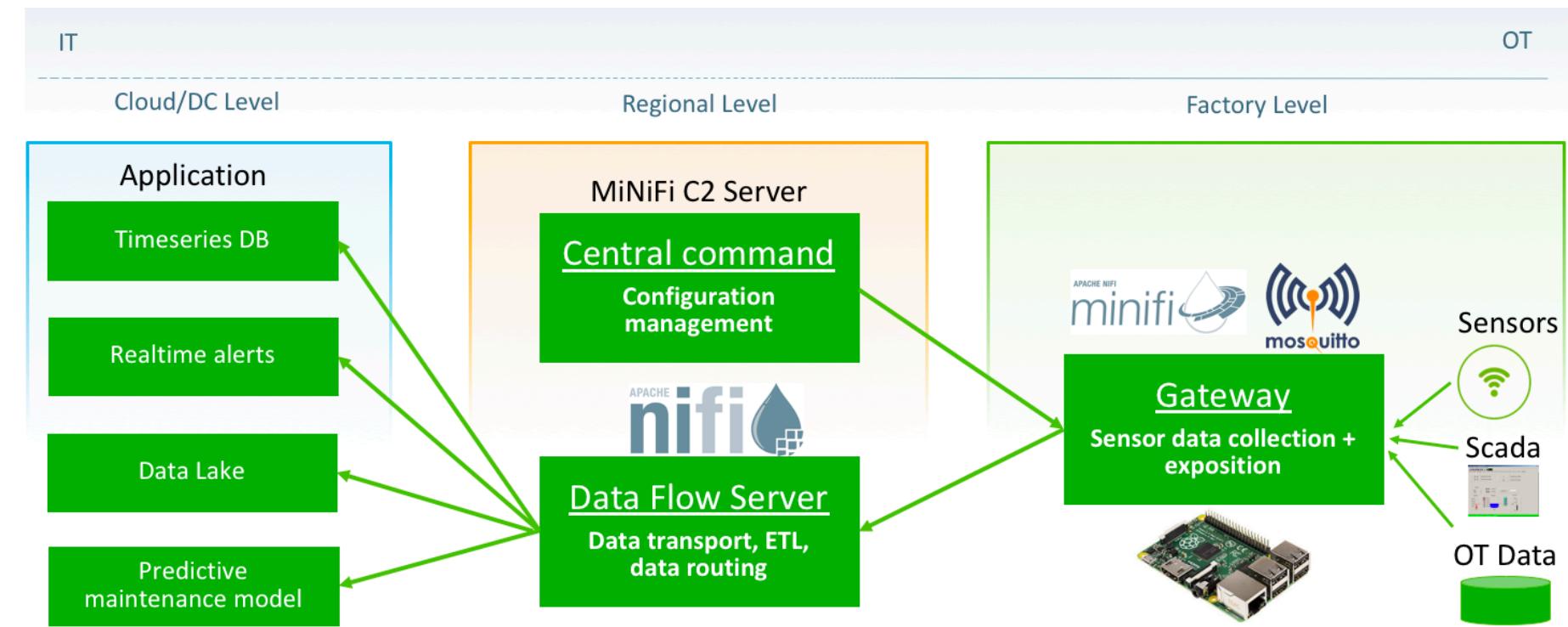
# MiNiFi Command & Control Architecture

- Flow Versioning
  - Develop flows for class of MiNiFi instances
- Command & Control (C2) API (*in Java master*)
  - FileChangeIngestor
  - RestAPIIngestor
  - PullHTTPIngestor



# MiNiFi C2 + Raspberry Pi

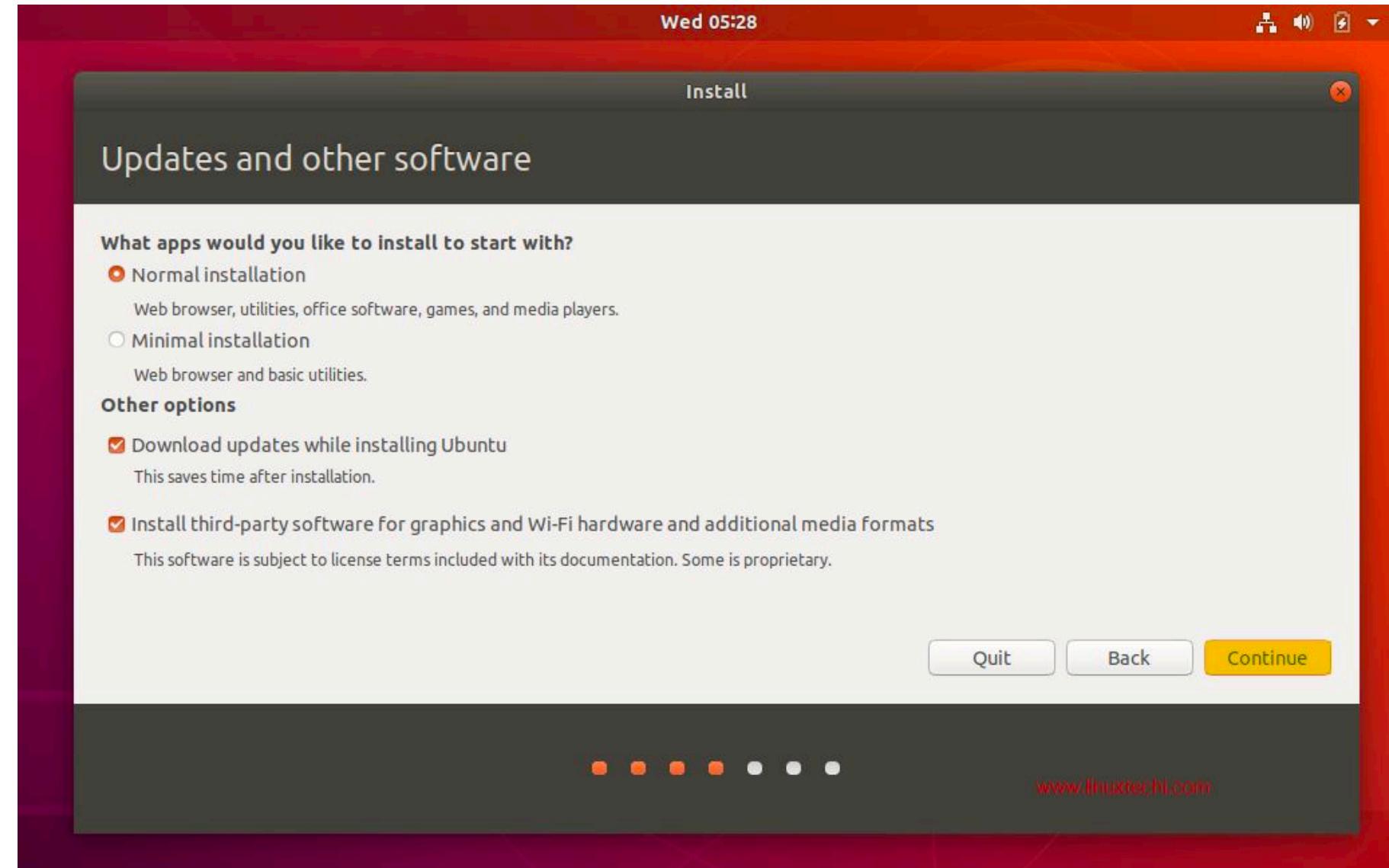
- Abdelkrim Hadidj
- Step-by-step walkthrough
  - deploying services
  - configuring endpoints
  - developing flows



[How to build an IIoT system using Apache NiFi,  
MiNiFi, C2 Server, MQTT and Raspberry Pi](#)

# Extension Registry

- Uses NiFi Registry to manage extensions (“narballs”)
- Allows faster download, build, and deployment
  - *Think Ubuntu Server vs. Desktop*
- Create, upload, and download functional bundles



[LinuxTechi.com](http://LinuxTechi.com)

- [NiFi Wiki - Extension Registry Feature Proposal](#)
- [NiFi Dev Mailing List - Discussion of implementation](#)

# Community

# New Announcements

- NiFi 1.9.1 – 16 Mar 2019 (167+ Jiras)
  - New processors (Kudu, Slack)
  - Hot-loading NARs
  - Security HTTP headers
  - Record processors automatically infer schema
- MiNiFi C++ 0.5.0 – 6 June 2018
- MiNiFi Java 0.5.0 – 7 July 2018
- NiFi Registry 0.3.0 – 25 Sept 2018



# 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)

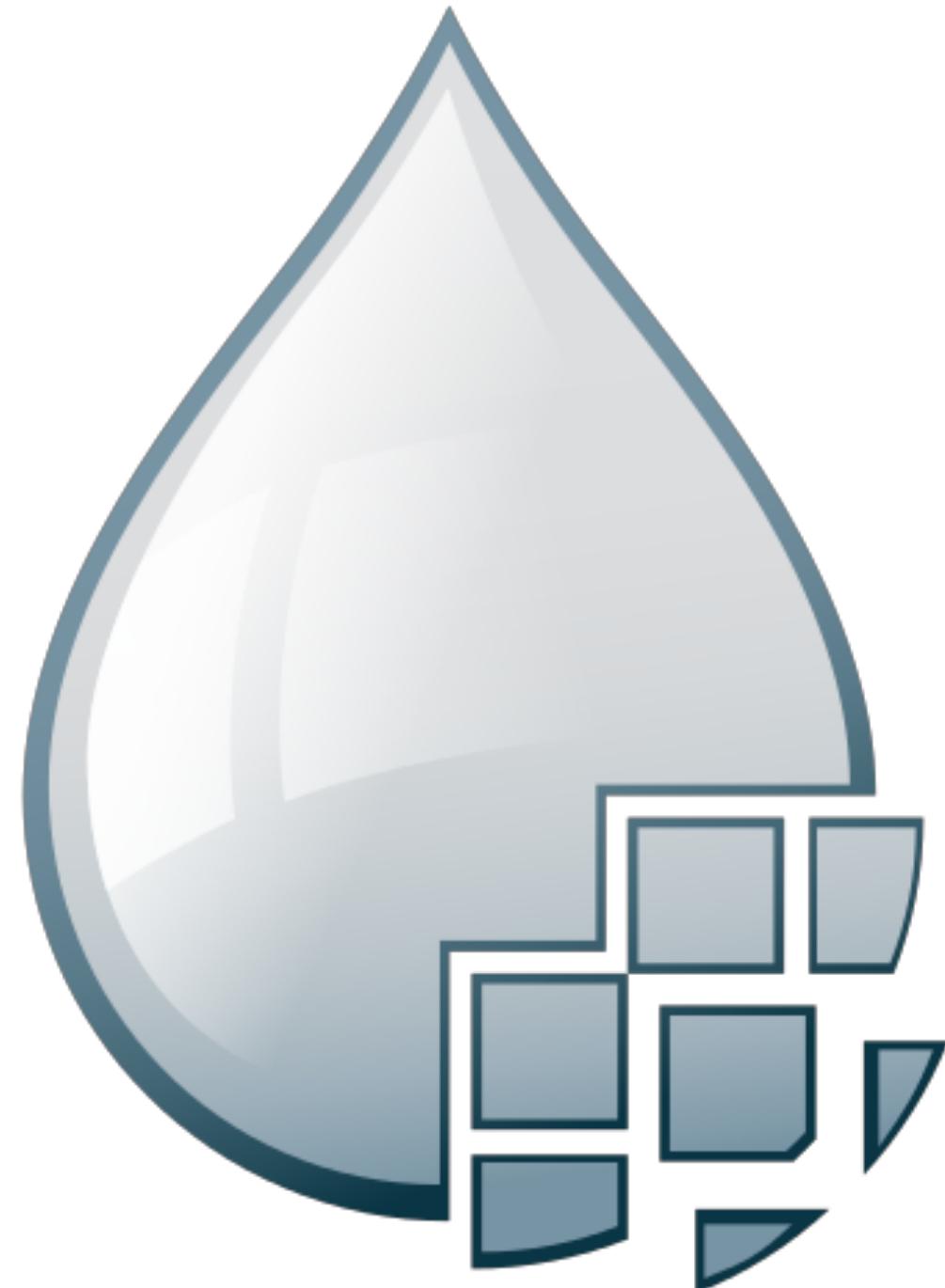
[users@nifi.apache.org](mailto:users@nifi.apache.org)

**Submit Ideas or Issues**

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

**Follow us on Twitter**

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

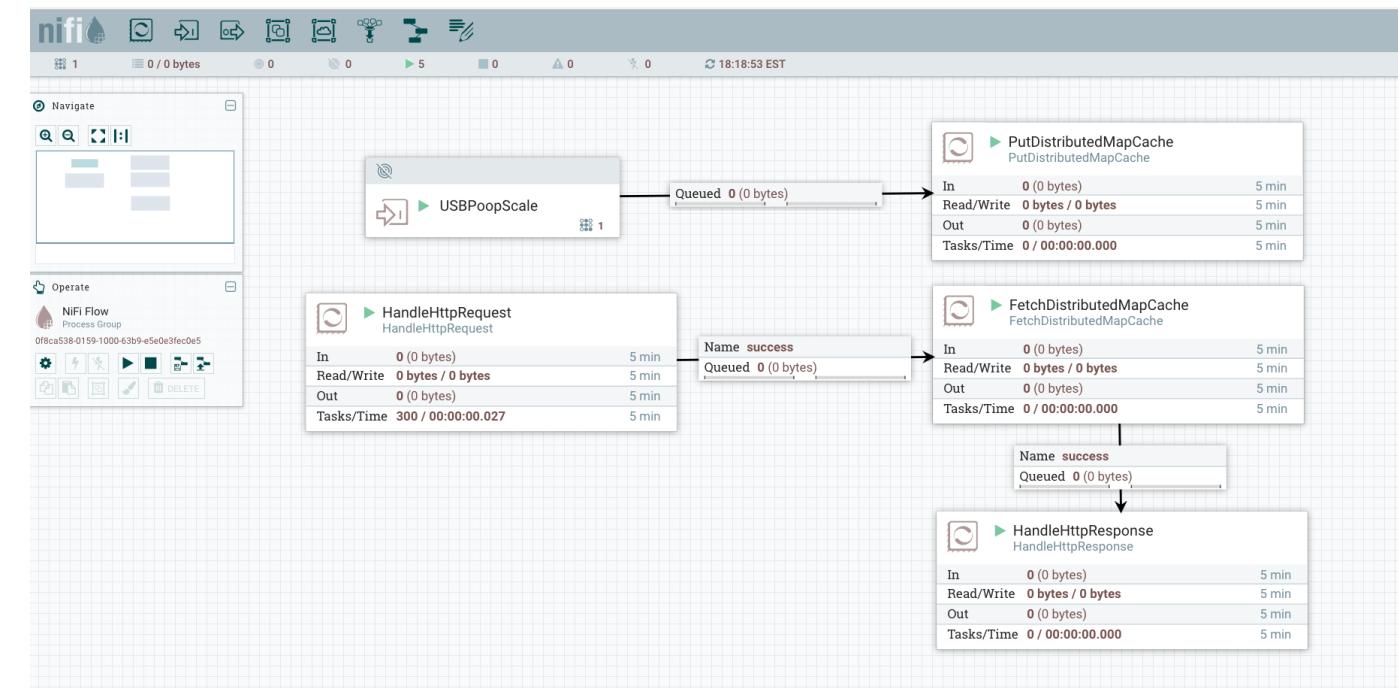
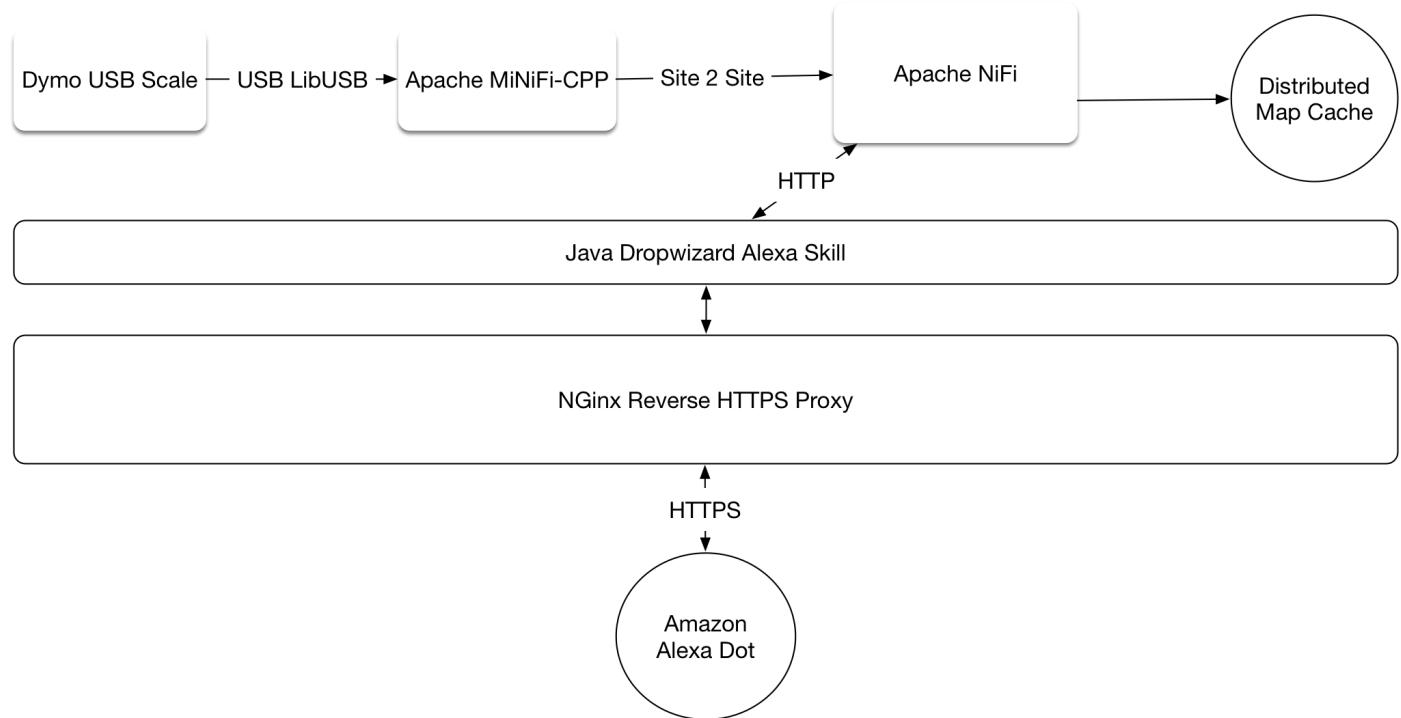


# More NiFi This Week

Title	Date		Speakers	Room
Cloudera Data in Motion Meetup - Future of Data Barcelona (free)	Tuesday 3/19	1800 - 2000	Tim Spann, Dan Chaffelson, Andy LoPresto	127-128
Apache NiFi Crash Course		1400 - 1600	Nathan Gough, Andy LoPresto	111
Data Acquisition Automation for NiFi in a Hybrid Cloud environment – the Path towards DataOps	Wednesday 3/20	1450 - 1530	Arda Basar, Ivan Georgiev	129-130
IoT, Streaming, and Dataflow Birds of a Feather		1600 - 1730	Tim Spann, Dan Chaffelson, Purnima Reddy Kuchikulla, Andy LoPresto	129-130
Addressing Challenges with IoT Edge Management		1100 - 1140	Dinesh Chandrasekhar	127-128
Intelligently Collecting Data at the Edge – Intro to Apache MiNiFi		1150 - 1230	Andy LoPresto	127-128
Edge to AI: Analytics from Edge to Cloud with Efficient Movement of Machine Data	Thursday 3/21	1400 - 1440	Tim Spann	129-130
BYOP: Custom Processor Development with Apache NiFi		1600 - 1640	Andy LoPresto	131-132
Apache Deep Learning 201		1650 - 1730	Tim Spann	122-123
Platform for the Research and Analysis of Cybernetic Threats		1650 - 1730	Monica Franceschini	124-125

# Community Example

- Jeremy Dyer
  - Alexa + MiNiFi + Dyer 2.0





# 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)