



Secure Dataflow From Edge to Core with Apache NiFi and MiNiFi

Andy LoPresto | @yolopey

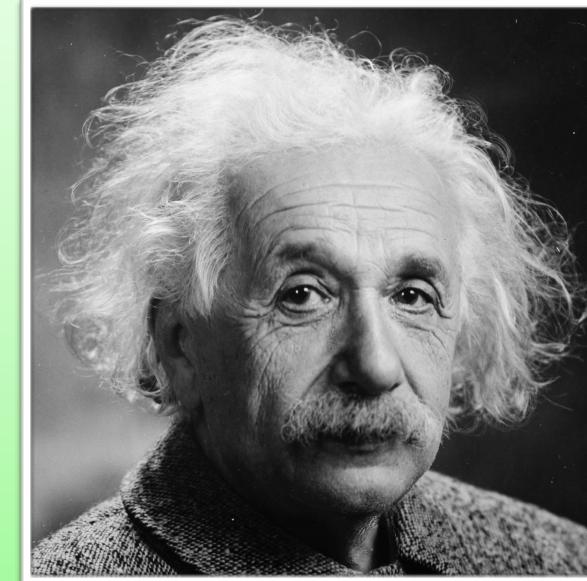
Sr. Member of Technical Staff at Hortonworks, Apache NiFi PMC & Committer

08 November 2018 Future of Data Melbourne

Acknowledgement of Country

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

Gauging Audience Familiarity With NiFi



“What’s a NeeFee?”

No experience with dataflow
No experience with NiFi

“I can pick this up pretty quickly”

Some experience with dataflow
Some experience with NiFi

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

Forgotten more about NiFi than most of us will ever know

Agenda

- *What is dataflow and what are the challenges?*
- Apache NiFi
- Apache MiNiFi
- Apache NiFi Registry
- *Security*
- New Features
- Demo
- *Roadmap*
- Q&A

All slides provided online, no need to transcribe

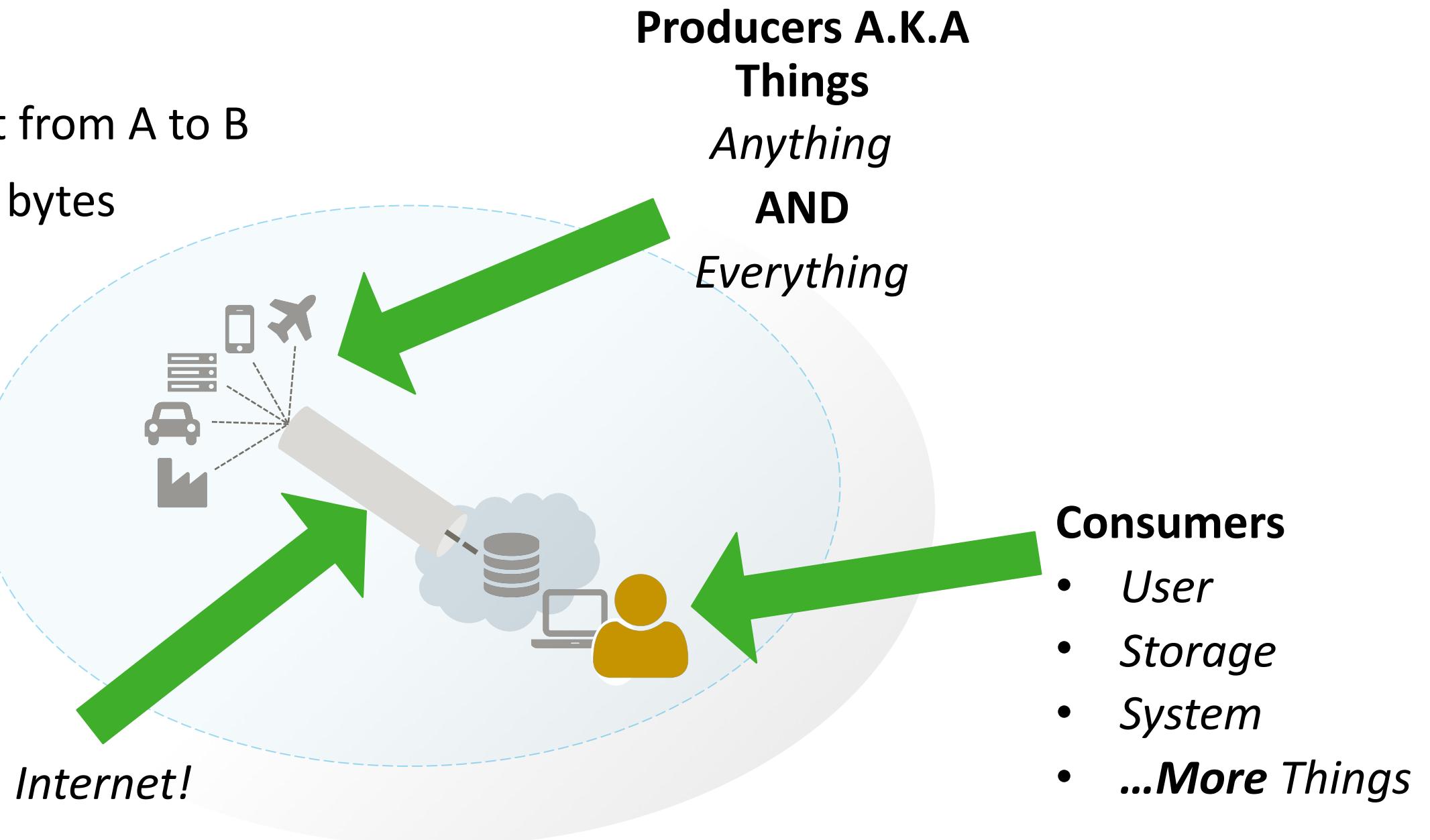




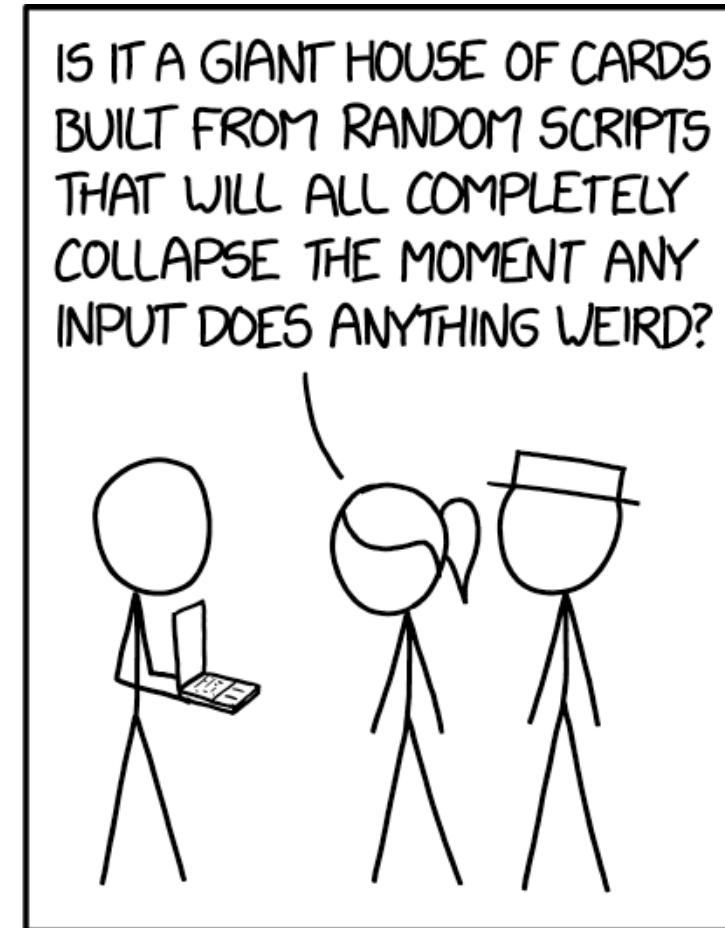
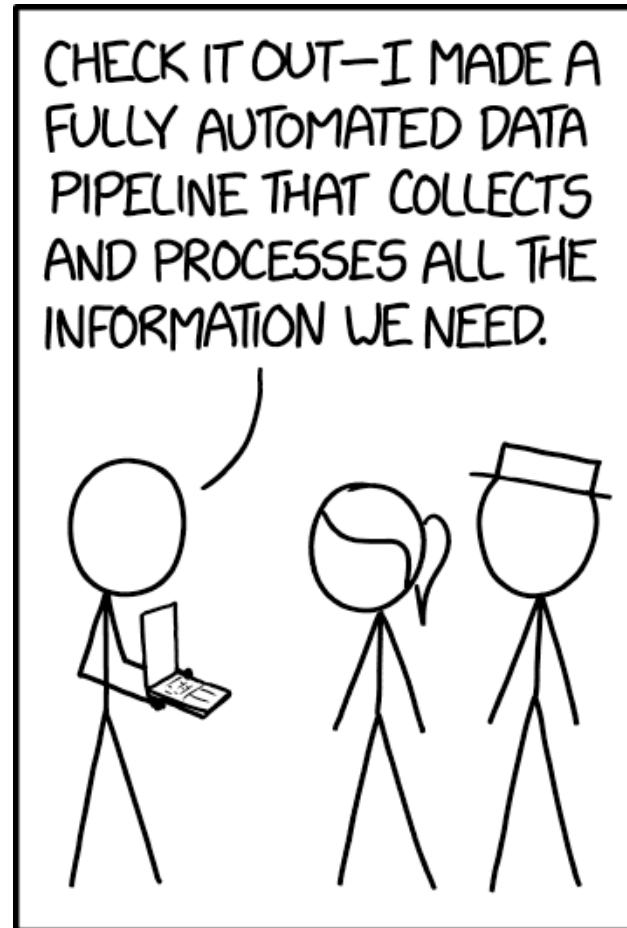
What is dataflow?

What is dataflow?

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



Moving data *effectively* is hard



"Data Pipeline" <https://xkcd.com/2054/>

Dataflow Challenges In 3 Categories

Data

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

Infrastructure

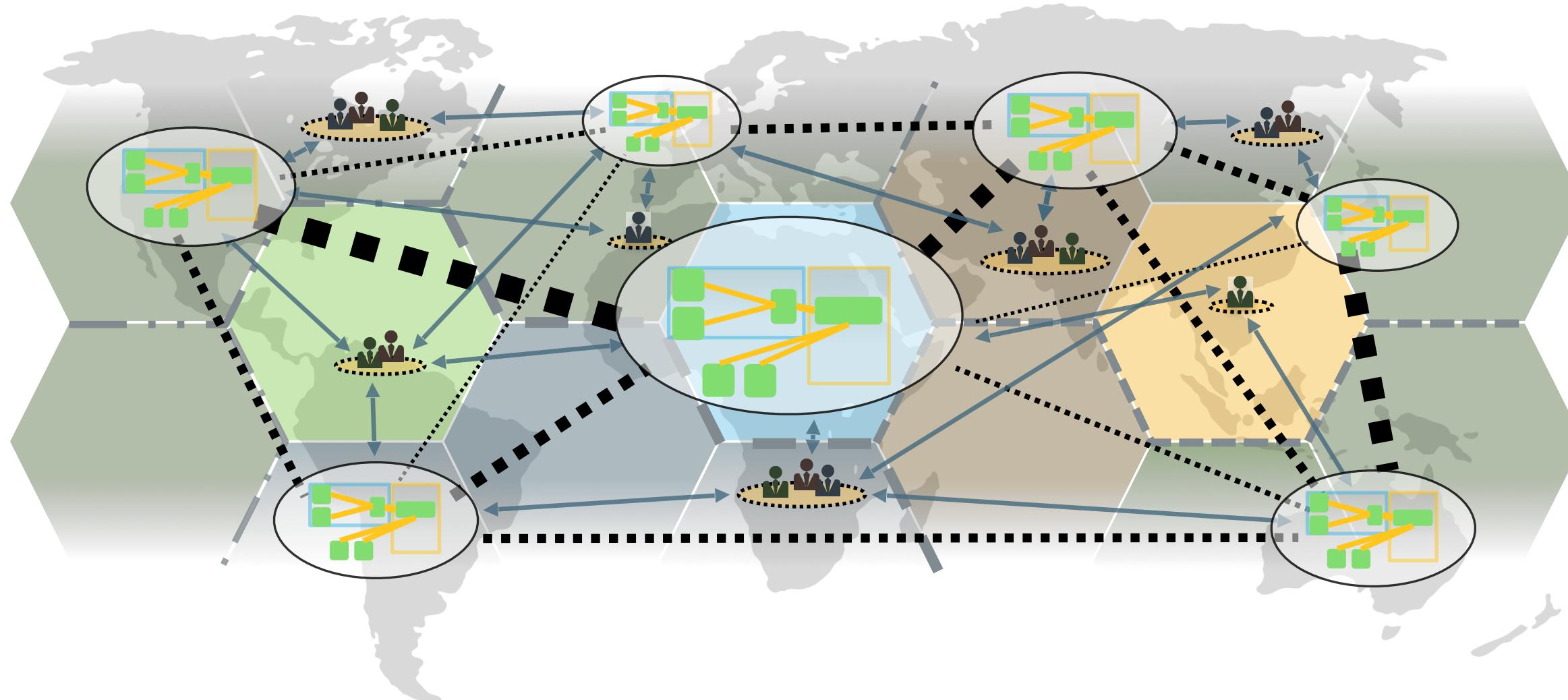
- “Exactly Once”
Delivery
- Ensuring
Security
- **Overcoming**
Security
- Credential
Management
- Network

People

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

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

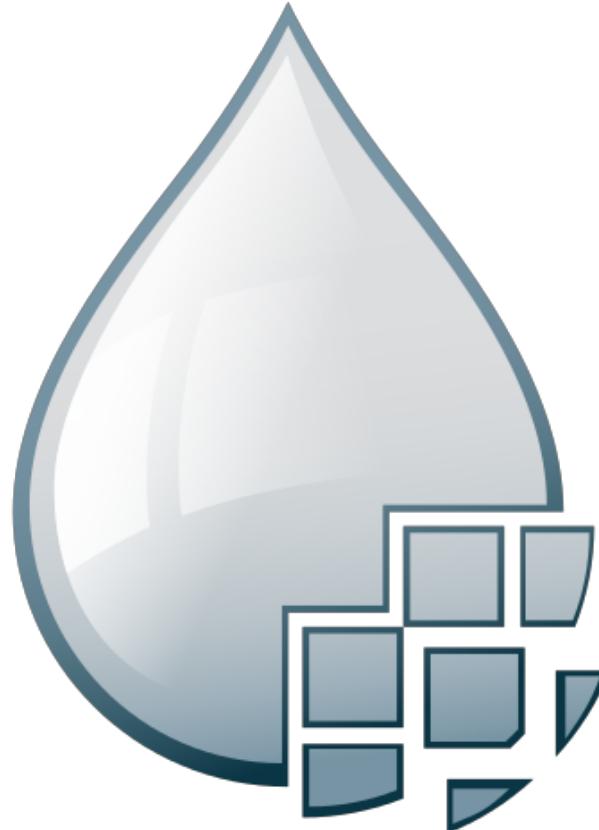
Raise your hand if you want to maintain Python scripts for the rest of your life



Apache NiFi

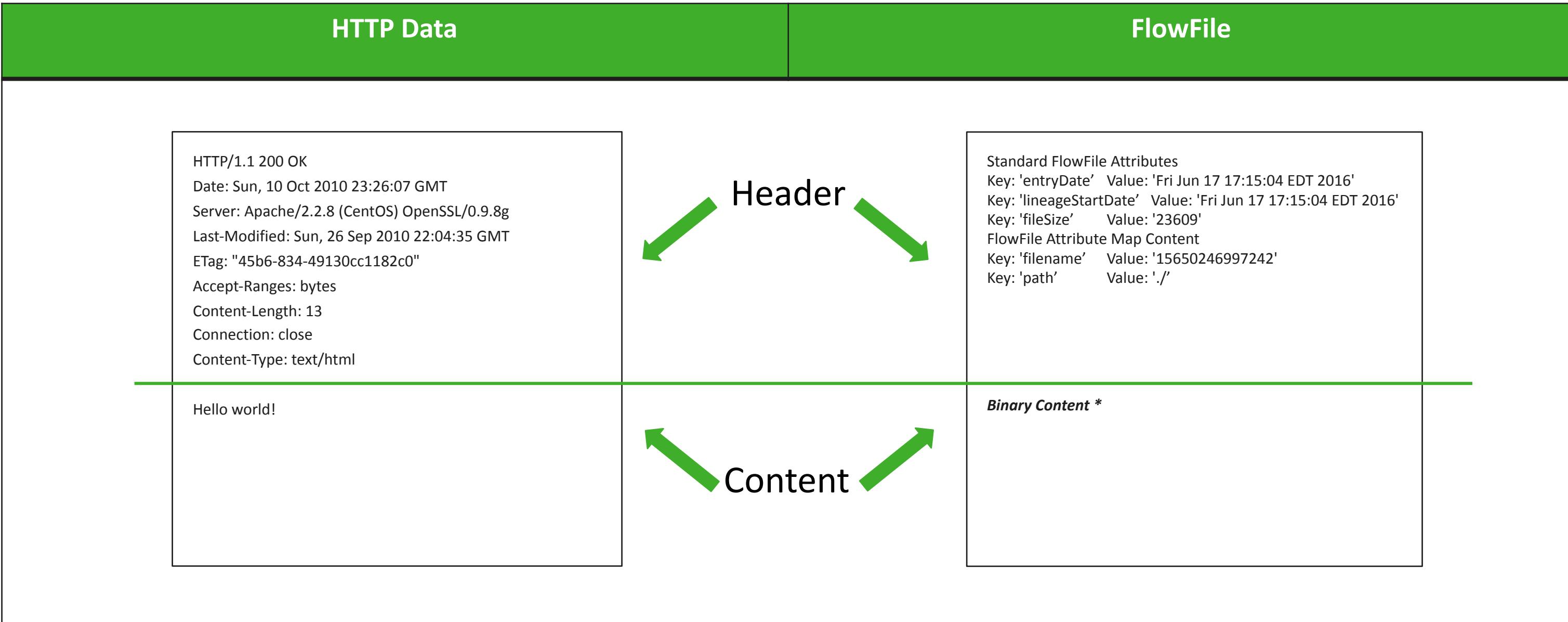
Apache NiFi

Key Features



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

Flowfiles Are Like HTTP Data



User Interface

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

The screenshot shows a terminal window and a web browser side-by-side.

Terminal (Left):

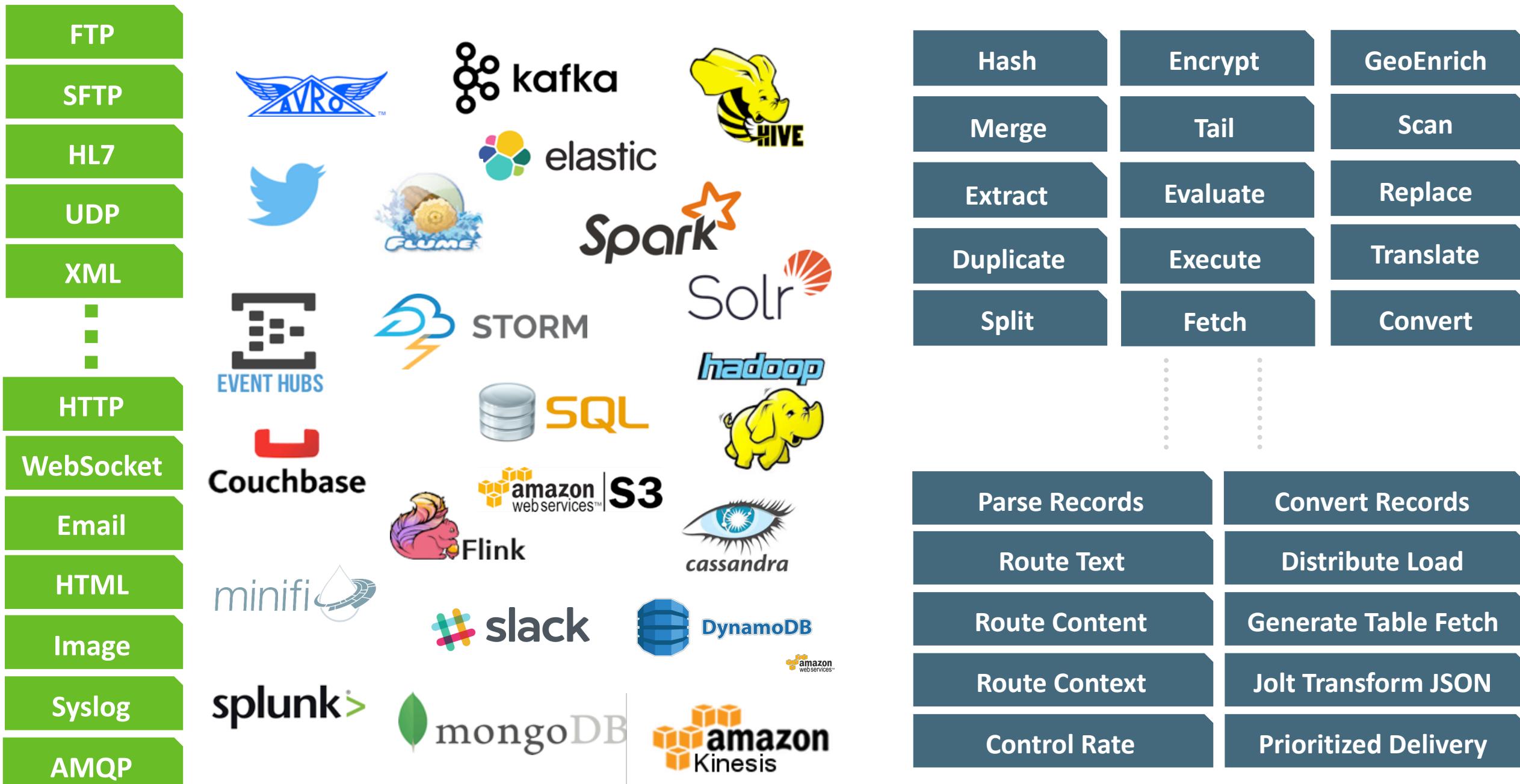
```
scratch/release_verification (master) alopreno
ls -l
total 144
drwxr-xr-x 16 alopreno staff 544B Nov 24 19:20 .
drwxr-xr-x 17 alopreno staff 578B Nov 24 14:25 ..
drwxr-xr-x 38 alopreno staff 1.3K Nov 24 14:33 archive/
-rw-r--r-- 1 alopreno staff 2.5K Nov 24 19:20 authorizations.xml
-rw-r--r-- 1 alopreno staff 3.7K Nov 24 19:19 authorizers.xml
-rw-rw-r-- 1 alopreno staff 2.1K Nov 23 23:36 bootstrap-notification-services.xml
-rw-r--r-- 1 alopreno staff 3.1K Nov 24 19:29 bootstrap.conf
-rw-r--r-- 1 alopreno staff 2.5K Nov 24 14:33 flow.xml.gz
-rw-r----- 1 alopreno staff 3.0K Nov 24 19:16 keystore.jks
-rw-rw-r-- 1 alopreno staff 8.0K Nov 23 23:36 logback.xml
-rw-r--r-- 1 alopreno staff 2.6K Nov 23 23:36 login-identity-providers.xml
-rw-r--r-- 1 alopreno staff 9.0K Nov 24 19:17 nifi.properties
-rw-r--r-- 1 alopreno staff 3.6K Nov 23 23:36 state-management.xml
-rw-r--r-- 1 alopreno staff 911B Nov 24 19:16 truststore.jks
-rw-r--r-- 1 alopreno staff 226B Nov 24 19:20 users.xml
-rw-r--r-- 1 alopreno staff 1.4K Nov 23 23:36 zookeeper.properties
h12203: /Users/alopreno/Workspace/scratch/release_verification (master) alopreno
21s @ 17:26:43 $ 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-b023393be562"/>
        </policy>
        <policy identifier="7e67308c-a837-3ba1-8b37-4a40ff8d43fe" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="b762618f-3d36-3b45-943e-bd38605e276e" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="61dfe35b-b37f-4b36-b2e7bc4989c3" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="ff9f93ce-1fda-3b08-9309-088ba6abc5a9" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="W">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="bc5512df-2a78-3bba-959e-3bc1ba5f781d" resource="/restricted-components" action="W">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="70197fb3-af4d-3938-b246-498db26032fa" 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-8f9b-3a71-aa0b-44325b1723cf" resource="/policies" action="W">
            <user identifier="9860132a-283e-370b-ba63-b023393be562"/>
        </policy>
        <policy identifier="26c66358-8548-31c8-a38c-e0eb4bc3ddba" 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>

```

Browser (Right):

The browser window shows the Apache NiFi user interface. The title bar says "1. scratch/release_verification (bash)". The main content area displays the NiFi configuration file "nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (bash)". The configuration includes settings like "nifi.flowfile.repository.always.sync=false", "nifi.swap.in.period=5 sec", and "nifi.swap.out.period=5 sec". It also lists various NiFi components and their configurations, such as "Content Repository", "Bootstrap Config File", and "Java home". Log entries from the NiFi application are visible at the bottom of the page, including messages about accepting shutdown commands and finishing shutdowns.

Deeper Ecosystem Integration: 286+ Processors, 61 Controller Services



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

“This is local? I’m going to ask you this *one* more time — this is local?”



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

“Is that USDA organic, or Oregon organic, or Portland organic?”



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

“Oh you have this information? This is fantastic!”



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

Data Provenance

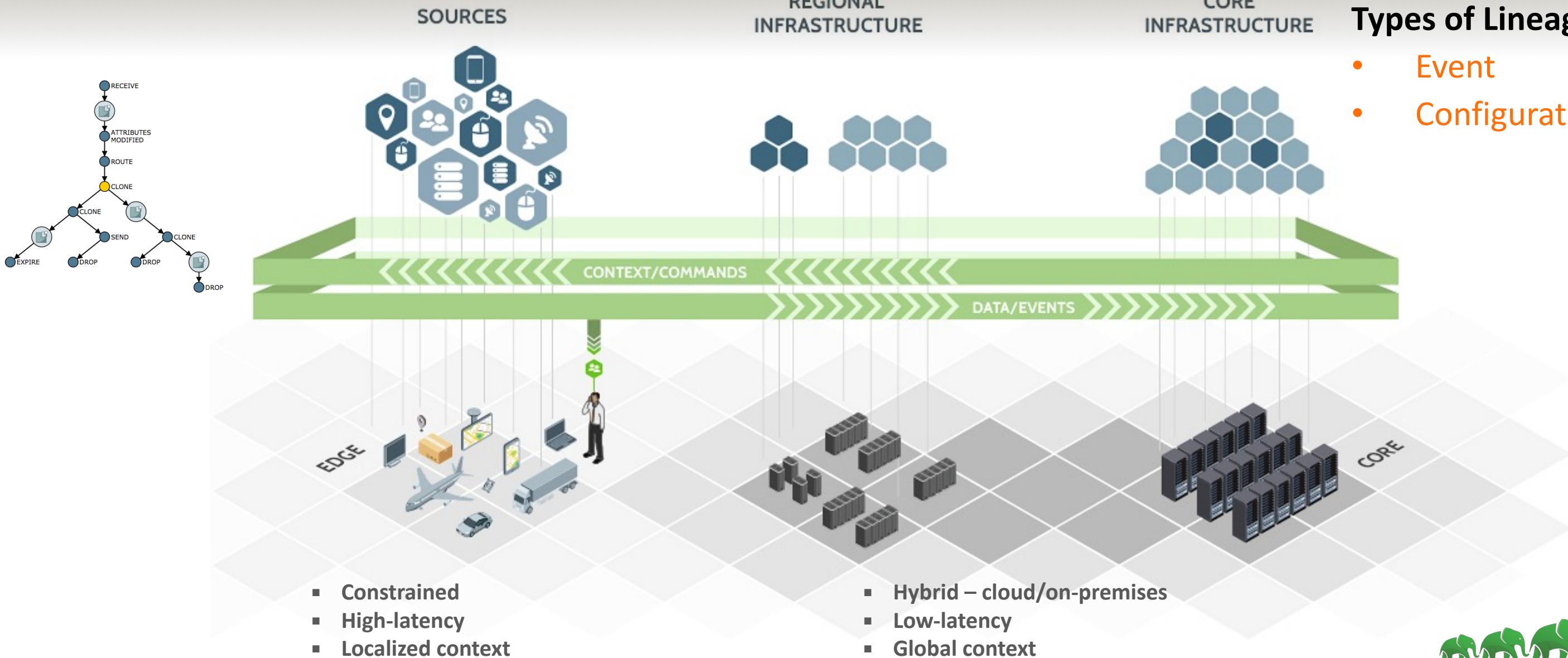
Origin – attribution
Replay – recovery

Evolution of topologies
Long retention

CORE
INFRASTRUCTURE

Types of Lineage

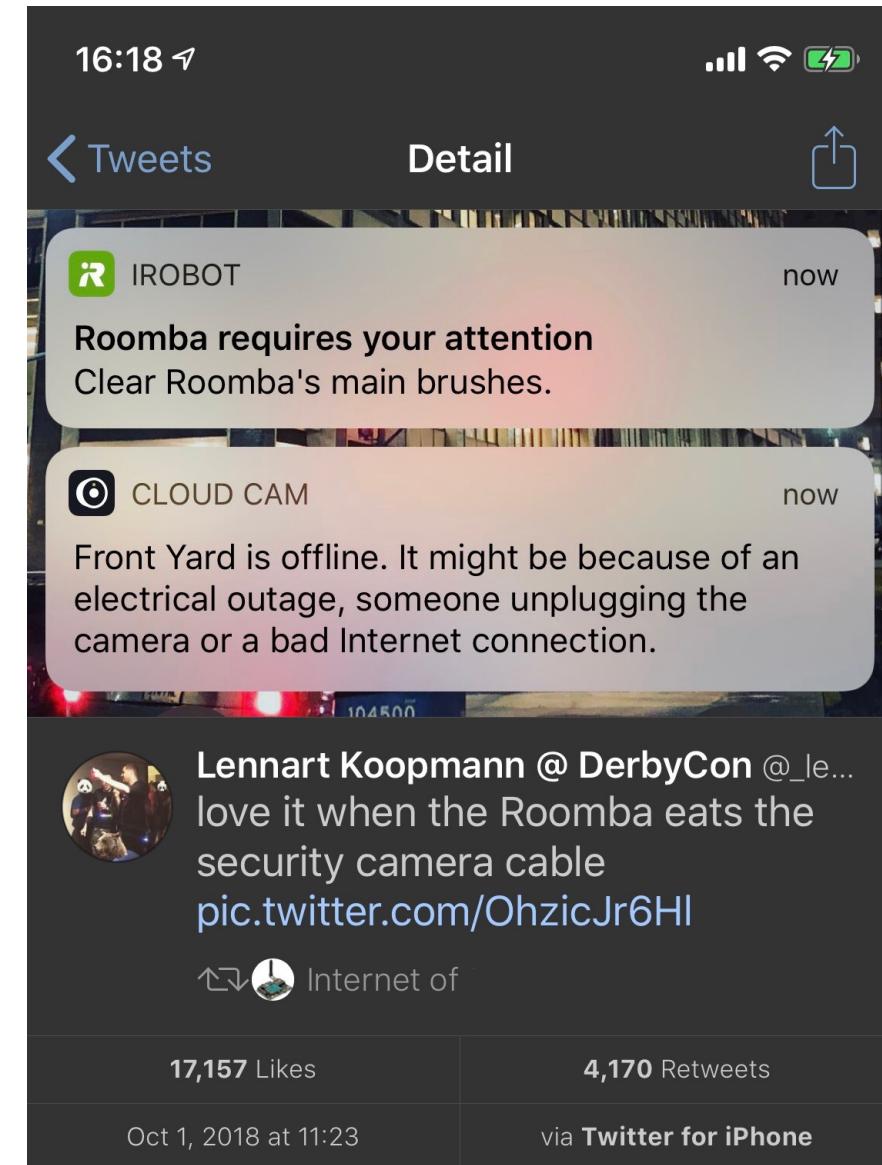
- Event
- Configuration



Apache MiNiFi

IoT Challenges

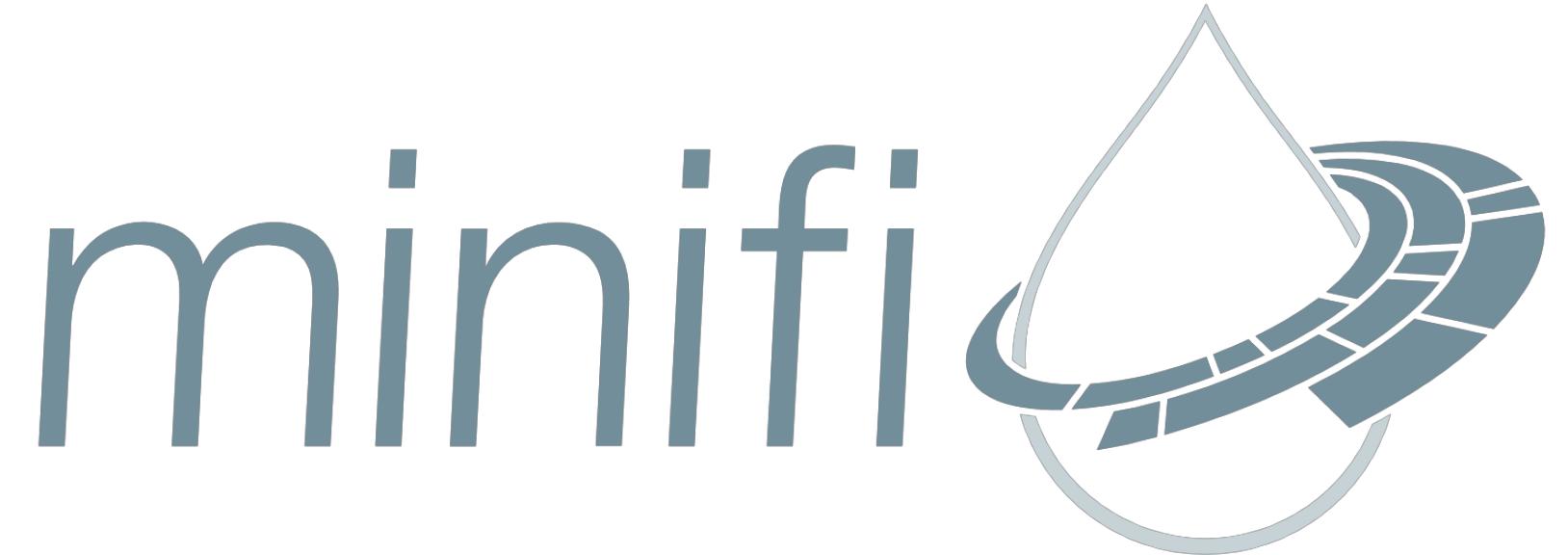
- Limited computing capability
- Limited power/network
- **Restricted software library/platform availability**
- **No UI**
- Physically inaccessible
- Not frequently updated
- **Competing standards/protocols**
- Scalability
- **Privacy & Security**



@_lennart

Apache NiFi Subproject: MiNiFi

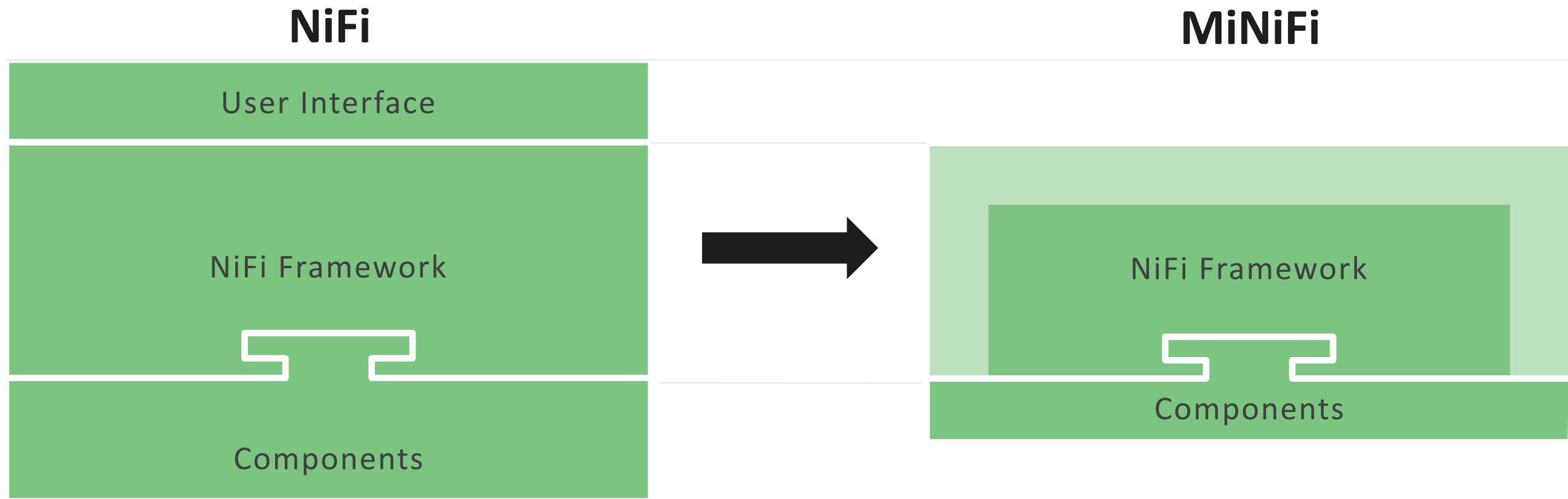
- Get the key parts of NiFi close to where data begins and provide bidirectional communication
- NiFi lives in the data center — give it an enterprise server or a cluster of them
- MiNiFi lives as close to where data is born and is a guest on that device or system
 - IoT
 - Connected car
 - Legacy hardware



Why build MiNiFi?

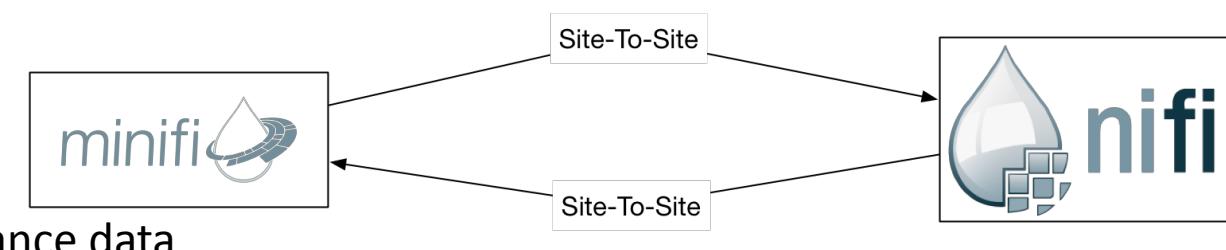
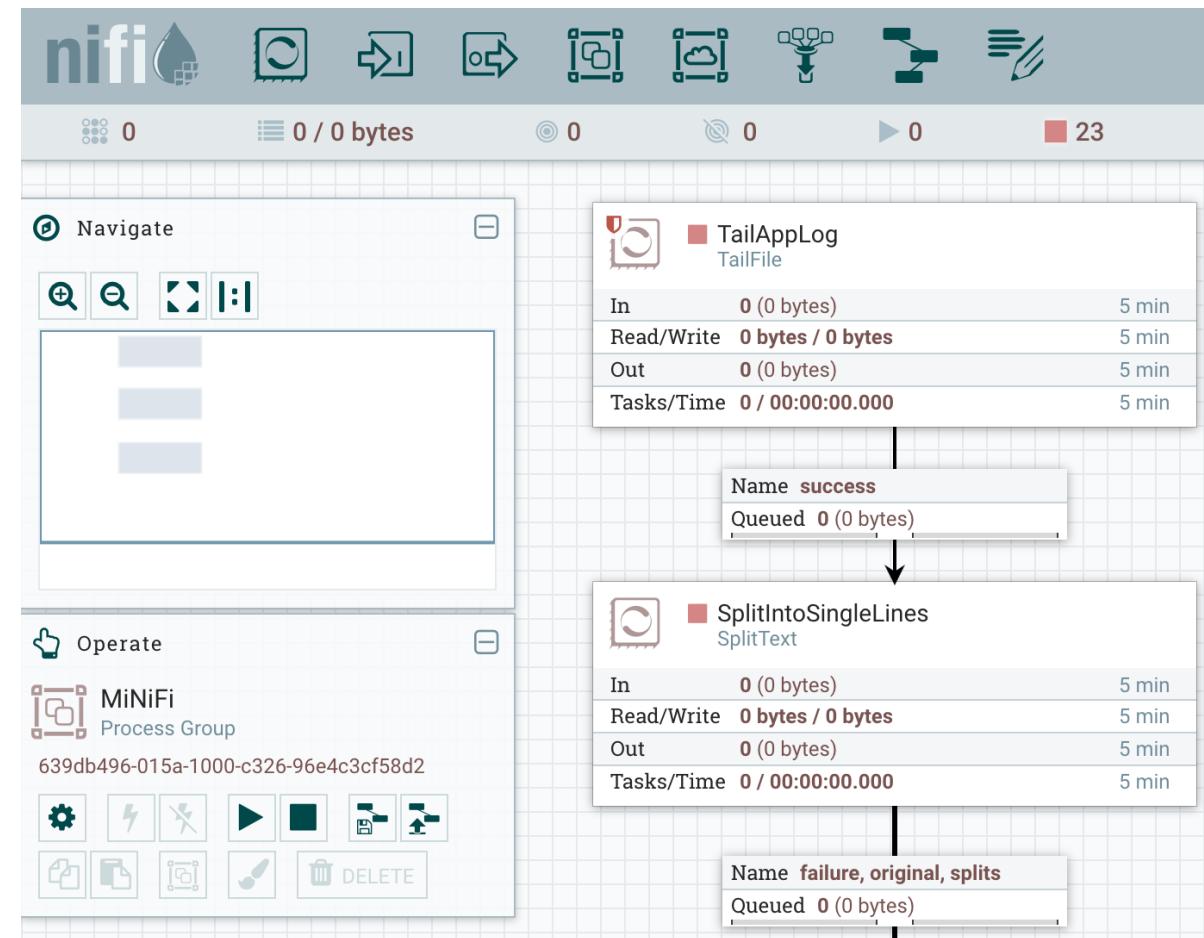
- NiFi is big
 - 1.8.0 release is 1.2 GB compressed
 - Can be modified to run in restricted environments, but requires manual surgery
 - Provides UI, provenance query, etc.
 - Runs on dedicated machines/clusters — “owns the box”
- MiNiFi lives at the edge
 - No UI
 - 0.5.0 Java release is 67 MB, C++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

NiFi vs MiNiFi Java Processes



Flavors of MiNiFi

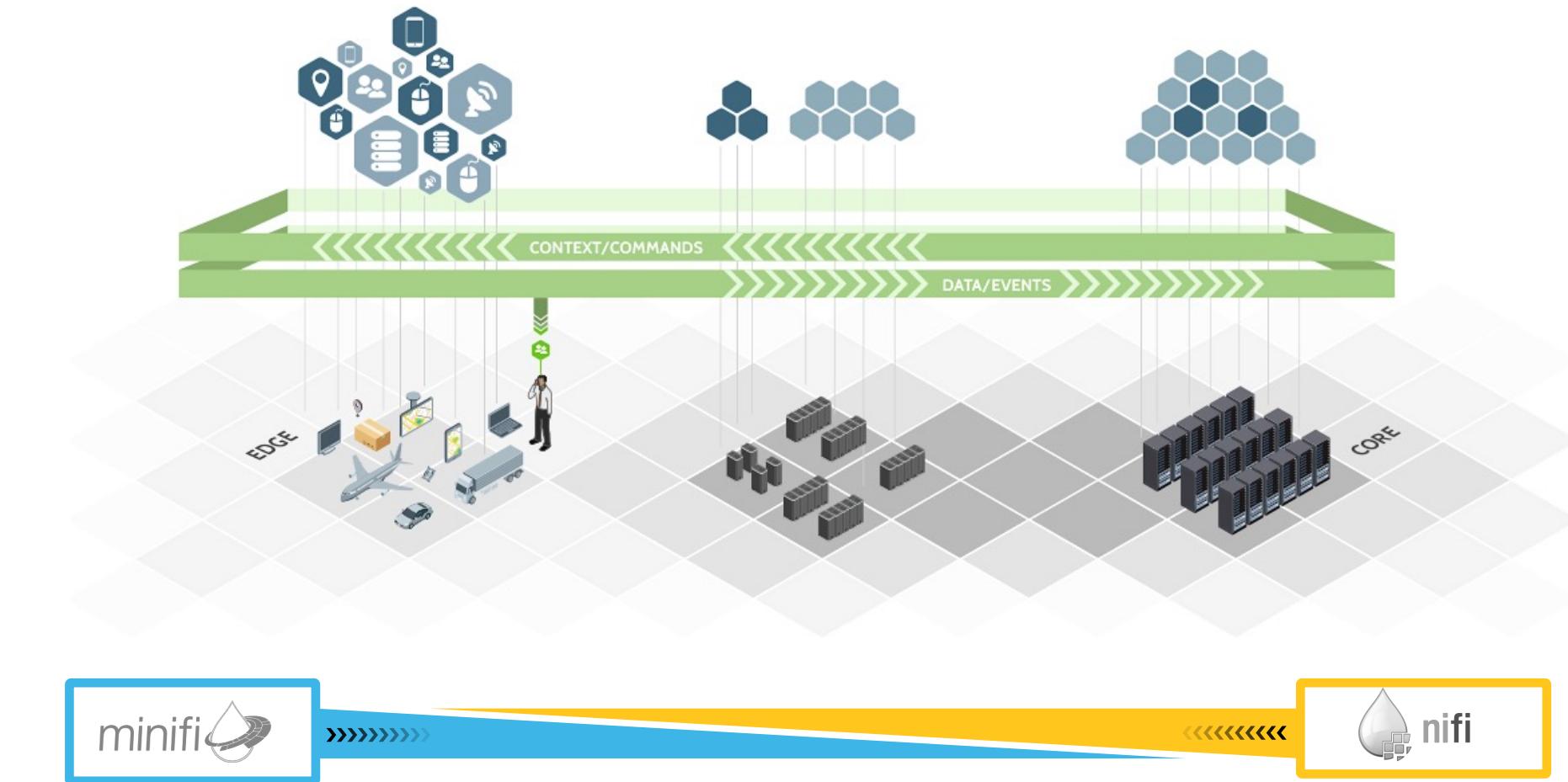
- MiNiFi Java (v0.5.0)
 - Modified version of NiFi
 - No UI
 - YAML configuration
 - Reduced processor count
 - 63+ by default, more available with additional NARs
- MiNiFi C++ (v0.5.0)
 - Written from scratch
 - 33 processors by default
 - Bi-directional site-to-site & provenance data



```
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?

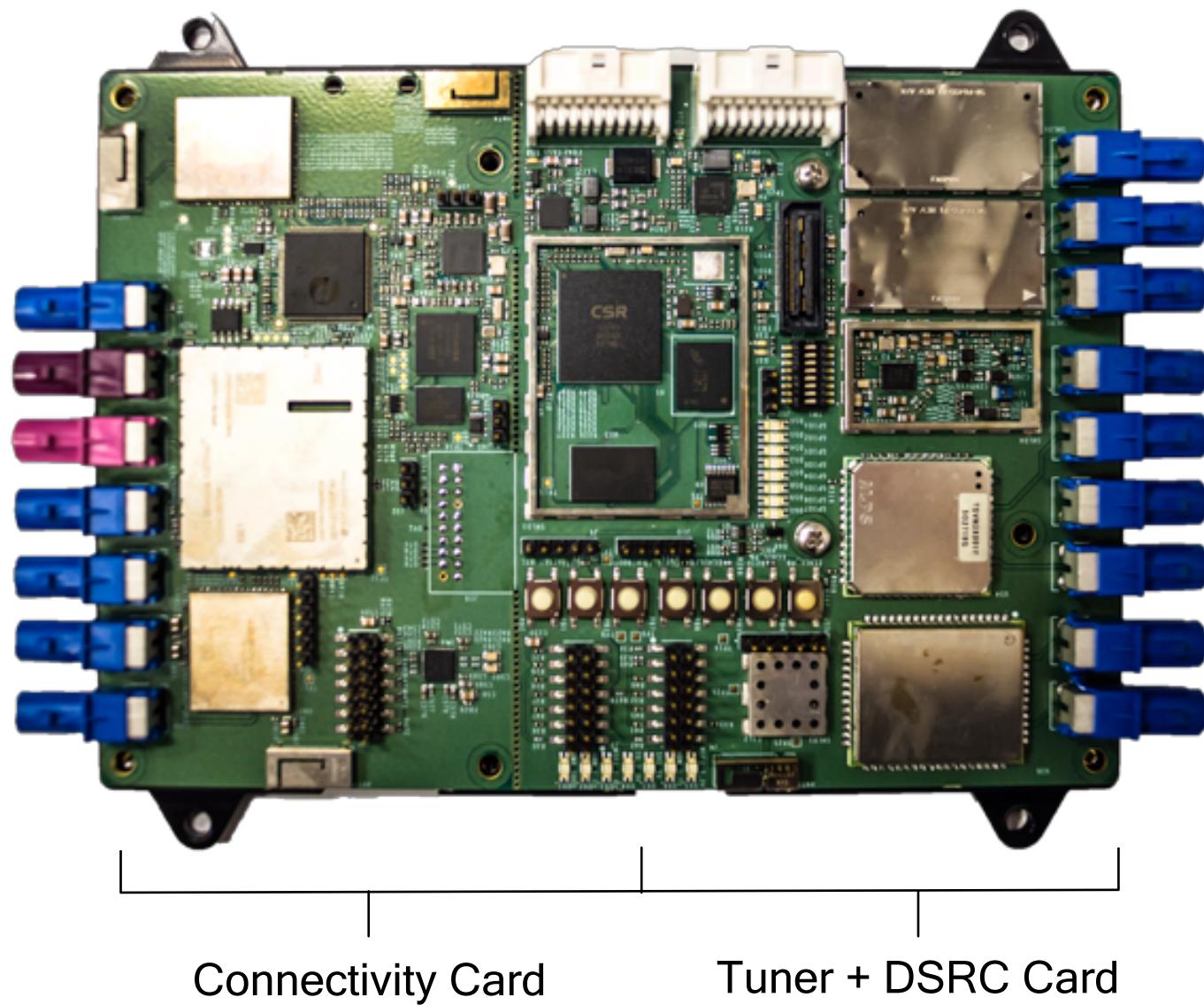
- NiFi
 - Design flows
 - Aggregate data from many sources
 - Perform routing/analysis/SEP
- MiNiFi
 - Receive flows
 - Collect data
 - Send for processing



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



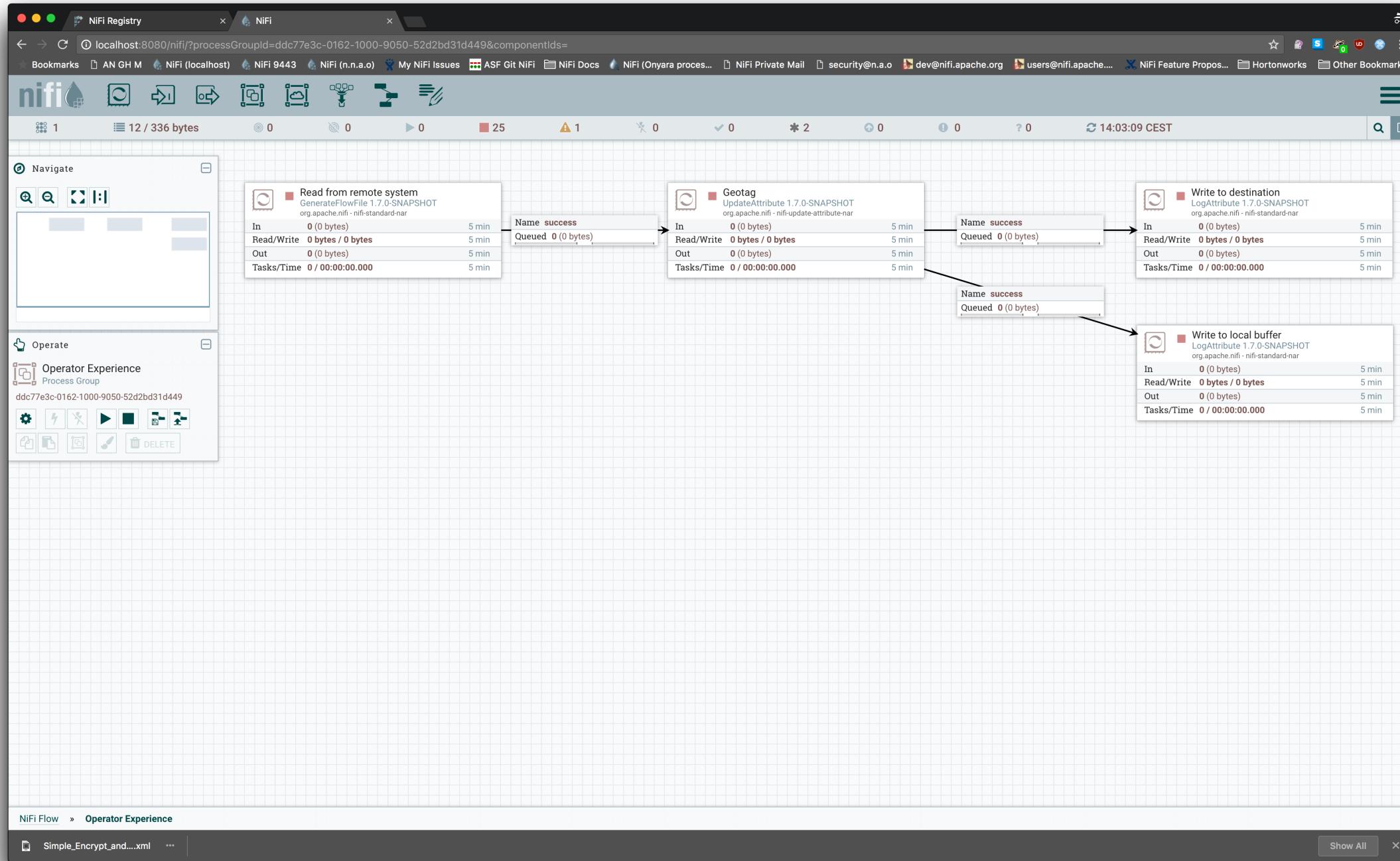
Apache NiFi Registry

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_soq.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

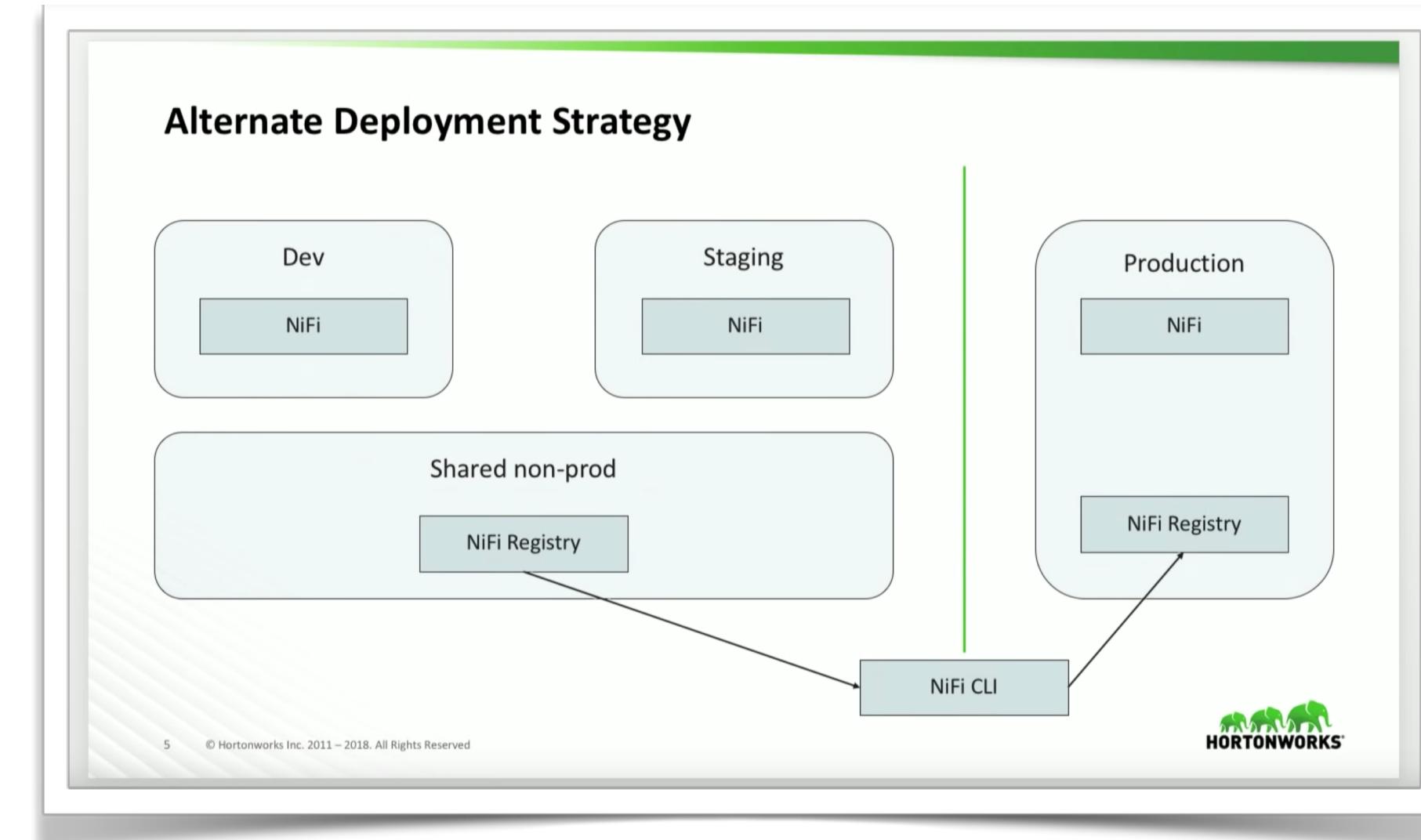
Introducing Apache NiFi Registry 0.3.0

- 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

The screenshot shows the Apache NiFi Registry interface. At the top, there's a header with the NiFi Registry logo, a dropdown menu labeled 'NiFi Registry / All ▾', and user information like 'registry_user' and 'LOGOUT'. Below the header, there's a search bar and a sorting dropdown set to 'Sort by: Name (a - z)'. The main area displays two flows: 'Flow 1 - Bucket 1' (1 version) and 'Flow 2 - Bucket 2' (2 versions). For 'Flow 2 - Bucket 2', a detailed view is shown with a 'DESCRIPTION' field containing 'Description 2'. To the right, a 'CHANGE LOG' section shows two entries: 'Version 2 - 40 minutes ago by registry_user' and 'Add processors Dec-26-2017 at 11:23 PM'. Another entry, 'Version 1 - 41 minutes ago by registry_user', is also listed.

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

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"  
  }  
}
```

Security

Security

- Secure the instance
 - Sensitive properties
 - HTTPS
 - Authentication & Authorization
- Secure the configuration
 - Encrypt the configs
- Secure the data
 - EncryptContent
 - Encrypted repositories

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

Sensitive Component Properties

- NiFi encrypts all sensitive component properties (*database password, FTP password, etc.*) with a configurable algorithm
 - Default is PBEWITHMD5AND256BITAES-CBC-OPENSSL
- If no key material is provided by the admin, a default is used
 - TBC in 2.0.0
 - **Populate this with a random, unique value when deploying NiFi** `nifi.sensitive.props.key=thisIsBadButWayBetterThanNothing`

```
148 # security properties #
149 nifi.sensitive.props.key=
150 nifi.sensitive.props.key.protected=
151 nifi.sensitive.props.algorithm=PBEWITHMD5AND256BITAES-CBC-OPENSSL
152 nifi.sensitive.props.provider=BC
153 nifi.sensitive.props.additional.keys=
```

```
/***
 * This constructor creates an encryptor using <em>Password-Based Encryption</em> (PBE). The <em>key</em> value is
 * the direct value provided in <code>nifi.sensitive.props.key</code> in
 * <code>nifi.properties</code>, which is a <em>PASSWORD</em> rather than a <em>KEY</em>, but is named such for
 * backward/legacy logical compatibility throughout the rest of the codebase.
 * <p>
 * For actual raw key provision, see {@link #StringEncryptor(String, String, byte[])}.
 *
 * @param algorithm the PBE cipher algorithm ({@link EncryptionMethod#algorithm})
 * @param provider the JCA Security provider ({@link EncryptionMethod#provider})
 * @param key      the UTF-8 characters from nifi.properties -- nifi.sensitive.props.key
 */
public StringEncryptor(final String algorithm, final String provider, final String key) {
    this.algorithm = algorithm;
    this.provider = provider;
    this.key = null;
    this.password = new PBEKeySpec(key == null
        ? DEFAULT_SENSITIVE_PROPS_KEY.toCharArray()
        : key.toCharArray());
    initialize();
}
```

First Step After Uncompressing NiFi

- Change the default key material
- Input to KDF to generate the symmetric key used for sensitive processor property value encryption

```
139 # security properties #
140 nifi.sensitive.props.key=This is not the default NiFi key material.
141 nifi.sensitive.props.key.protected=
142 nifi.sensitive.props.algorithm=PBEWITHMD5AND256BITAES-CBC-OPENSSL
143 nifi.sensitive.props.provider=BC
144 nifi.sensitive.props.additional.keys=
145
```

```
125     <name>Encryption Algorithm</name>
126     <value>AES_GCM</value>
127   </property>
128   <property>
129     <name>allow-weak-crypto</name>
130     <value>not-allowed</value>
131   </property>
132   <property>
133     <name>Password</name>
134   </property>
135   <property>
136     <name>raw-key-hex</name>
137     <value>enc{9D6F3EE98D4D04CD7875ED77E9EDBC30526820CDD2C764C45DCBBC0CBBC839FB2D80B
B0477779BFB49DD8FC262E05C9F2EB6BDED343BA8D116395BC4C34B4DA41C3978DEC4CA939D08552
A8DE9E089481959180E3DEAE537BE1BA278FA8F90C9}</value>
138   </property>
```

Configure HTTPS

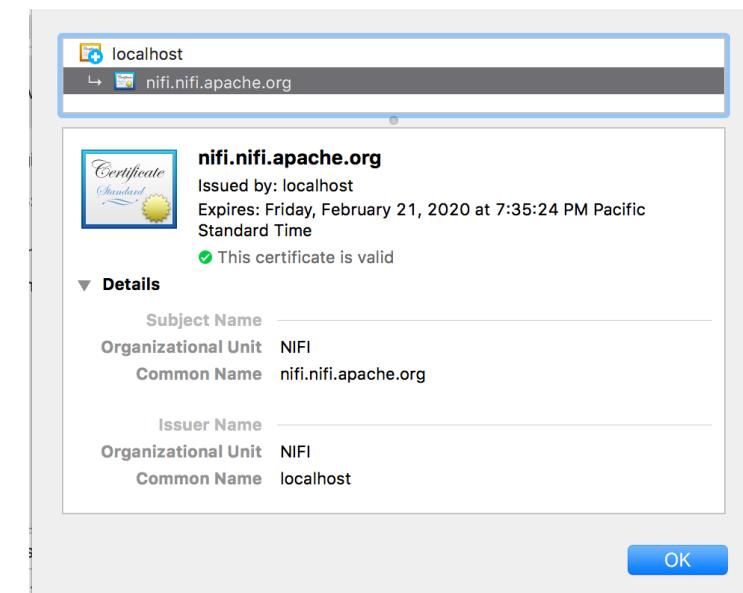
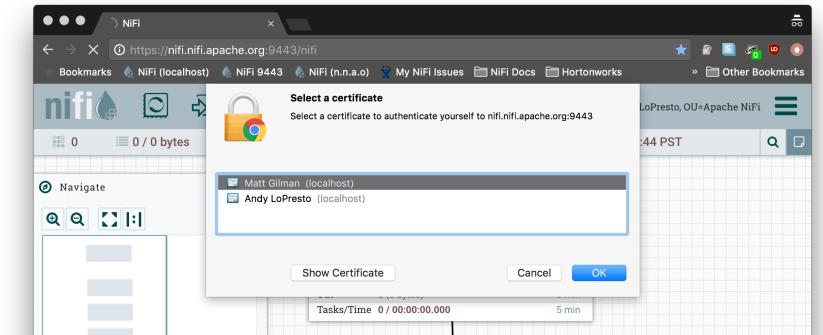
- Deploying a keystore & truststore secures the network connections and enables authentication
 - Keystore = “who am I?”
 - Truststore = “who should I listen to?”

```
nifi.properties (nifi) | nifi.properties (scratch)
A >> > nifi-1.4.0-SNAPSHOT > nifi-1.4.0-SNAPSHOT > conf > nifi.properties <-> B >> > Workspace > scratch > secure_nifi > nifi.properties <-
A nifi.properties ...Workspace/scratch/secure_nifi A nifi.properties ...bin/nifi-1.4.0-SNAPSHOT/conf

120 # Site to Site properties
121 nifi.remote.input.host=-
122 nifi.remote.input.secure=false
123 nifi.remote.input.socket.port=-
124 nifi.remote.input.http.enabled=true
125 nifi.remote.input.http.transaction.ttl=30 sec
126
127 # web properties #
128 nifi.web.war.directory=./lib
129 nifi.web.http.host=
130 nifi.web.http.port=8080
131 nifi.web.http.network.interface.default=-
132 nifi.web.https.host=-
133 nifi.web.https.port=-
134 nifi.web.https.network.interface.default=-
135 nifi.web.jetty.working.directory=./work/jetty
136 nifi.web.jetty.threads=200
137
138 # security properties #
139 nifi.sensitive.props.key=
140 nifi.sensitive.props.key.protected=
141 nifi.sensitive.props.algorithm=PBEWITHMD5AND256BITAES-CBC-OPENSSL
142 nifi.sensitive.props.provider=BC
143 nifi.sensitive.props.additional.keys=
144
145 nifi.security.keystore=-
146 nifi.security.keystoreType=jks
147 nifi.security.keystorePasswd=password
148 nifi.security.keyPasswd=password
149 nifi.security.truststore=-
150 nifi.security.truststoreType=jks
151 nifi.security.truststorePasswd=password
152 nifi.security.truststorePasswd=password
153 nifi.security.needClientAuth=-
154
155 # Site to Site properties
156 nifi.remote.input.host=nifi.nifi.apache.org
157 nifi.remote.input.secure=true
158 nifi.remote.input.socket.port=10443
159 nifi.remote.input.http.enabled=true
160 nifi.remote.input.http.transaction.ttl=30 sec
161
162 # web properties #
163 nifi.web.war.directory=./lib
164 nifi.web.http.host=
165 nifi.web.http.port=-
166 nifi.web.https.host=nifi.nifi.apache.org
167 nifi.web.https.port=9443
168
169 nifi.web.jetty.working.directory=./work/jetty
170 nifi.web.jetty.threads=200
171
172 # security properties #
173 nifi.sensitive.props.key=
174 nifi.sensitive.props.key.protected=
175 nifi.sensitive.props.algorithm=PBEWITHMD5AND256BITAES-CBC-OPENSSL
176 nifi.sensitive.props.provider=BC
177 nifi.sensitive.props.additional.keys=
178
179 nifi.security.keystore=/conf/nifi.nifi.apache.org/keystore.jks
180 nifi.security.keystoreType=jks
181 nifi.security.keystorePasswd=password
182 nifi.security.keyPasswd=password
183 nifi.security.truststore=/conf/nifi.nifi.apache.org/truststore.jks
184 nifi.security.truststoreType=jks
185 nifi.security.truststorePasswd=password
186 nifi.security.needClientAuth=
```

Enable TLS (SSL)

- Three ways to do this
 - Self-signed certificate generation
(manual / hard)
 - Existing enterprise certificates
(manual / hard)
 - NiFi TLS Toolkit
(automatic / easy)



■ 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).

NiFi TLS Settings

- HDF 3.0 (NiFi 1.2.0+) API/UI supports only TLSv1.2+
 - Strong cipher suites
 - No SHA-1*
 - Prevents downgrade attacks
- Can still support legacy protocols for external service ingest

NiFi TLS Toolkit

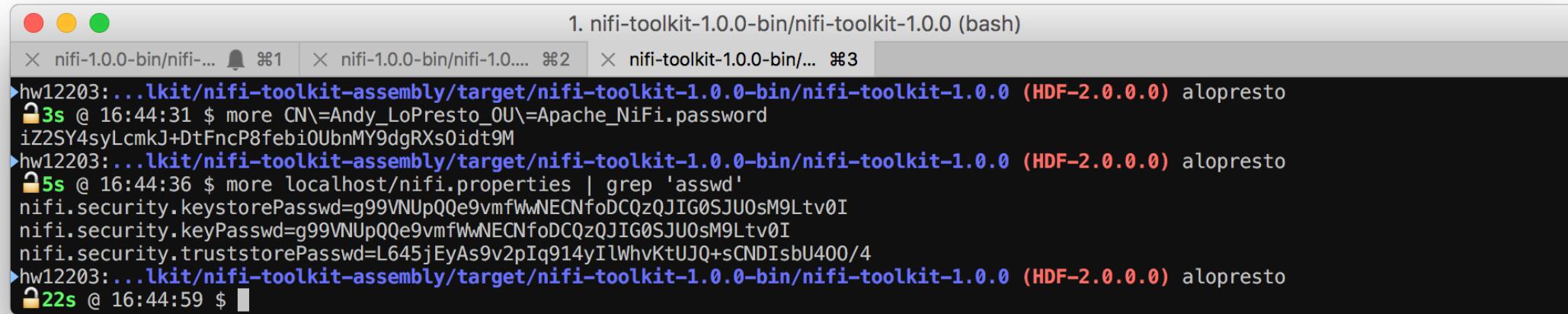
- Can run in *standalone* or *client/server* mode
- Automatically generates NiFi Certificate Authority (CA)
- Generates certificates for n nodes
- Signed by CA via CSR with shared-secret token
- Builds keystore & truststore
- Populates passwords in `nifi.properties`
- Generates client certificate for authentication
- Strong random passwords generated if not provided

```
Run toolkit in standalone mode ./bin/tls-toolkit.sh standalone
Generate a certificate for this hostname -n 'nifi.nifi.apache.org'
Generate a client certificate for this user -C 'CN=alopresto, OU=Apache NiFi'
Set the truststore password -P someLongTruststorePassword
Set the keystore password -S someLongKeystorePassword
Set the client cert password -B someLongClientCertPassword
Provide the input/output nifi.properties -f ../../../../../../nifi-assembly/target/
nifi-1.3.0-bin/nifi-1.3.0/conf/nifi.properties
Provide the output directory for the keys -o ../../../../../../nifi-assembly/target/
nifi-1.3.0-bin/nifi-1.3.0/conf/
```

```
hw12203:nifi-toolkit-1.3.0 (master) alopresto
└─ 15s @ 16:51:23 $ ./bin/tls-toolkit.sh standalone -n 'nifi.nifi.apache.org'
-C 'CN=alopresto, OU=Apache NiFi' -P someLongTruststorePassword -S
someLongKeystorePassword -B someLongClientCertPassword -f .../conf/
nifi.properties -o .../conf/
o.a.n.t: CommandLine: Using .../conf/nifi.properties as template.
o.a.n.t: Running standalone certificate generation with output directory ...
conf
o.a.n.t: Using existing CA certificate .../conf/nifi-cert.pem and key .../conf/
nifi-key.key
o.a.n.t: Writing new ssl configuration to .../conf/nifi.nifi.apache.org
o.a.n.t: Successfully generated TLS configuration for nifi.nifi.apache.org 1 in
.../conf/nifi.nifi.apache.org
o.a.n.t: Generating new client certificate .../conf/
CN=alopresto_OU=Apache_NiFi.p12
o.a.n.t: Successfully generated client certificate .../conf/
CN=alopresto_OU=Apache_NiFi.p12
o.a.n.t: tls-toolkit standalone completed successfully
hw12203:nifi-toolkit-1.3.0 (master) alopresto
└─ 6s @ 16:51:29 $
```

Security Mechanisms

- Private key is not communicated from client to CA; CSR is issued and signed
- Token and HMAC/SHA-256 is used to prevent MitM
- Default passwords generated for keystores, certificates are 32 bytes of secure random, Base64-encoded (~43 chars)



The screenshot shows a macOS terminal window titled "1. nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (bash)". It has three tabs open: "nifi-1.0.0-bin/nifi...", "nifi-1.0.0-bin/nifi-1....", and "nifi-toolkit-1.0.0-bin/...". The terminal window displays the following command-line session:

```
▶ hw12203:...lkit/nifi-toolkit-assembly/target/nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (HDF-2.0.0.0) alopresto
  3s @ 16:44:31 $ more CN\=Andy_LoPresto_0U\=Apache_NiFi.password
  iZ2SY4syLcmkJ+DtFncP8feb0UbnMY9dgRXs0idt9M
▶ hw12203:...lkit/nifi-toolkit-assembly/target/nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (HDF-2.0.0.0) alopresto
  5s @ 16:44:36 $ more localhost/nifi.properties | grep 'asswd'
  nifi.security.keystorePasswd=g99VNUpQqe9vmfWwNECNfoDCQzQJIG0SJu0sM9Ltv0I
  nifi.security.keyPasswd=g99VNUpQqe9vmfWwNECNfoDCQzQJIG0SJu0sM9Ltv0I
  nifi.security.truststorePasswd=L645jEyAs9v2pIq914yIlWhvKtUJQ+sCNDIsbU400/4
▶ hw12203:...lkit/nifi-toolkit-assembly/target/nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (HDF-2.0.0.0) alopresto
  22s @ 16:44:59 $
```

Authentication

- Supported authentication mechanisms
 - Client certificate
 - LDAP
 - Kerberos
 - OpenID Connect
 - KnoxSSO
 - OAuth 2.0 *under review

Restricted Components

- Controller services, reporting tasks, and processors which could maliciously affect the host system are now **RESTRICTED**
- Special annotation and access policy required

Add Processor

Source	Type	Version	Tags
all groups	Type ▾		
amazon	DeleteHDFS	1.4.0-SNAPSHOT	restricted, HDFS, hadoop, delete...
attributes	ExecuteFlumeSink	1.4.0-SNAPSHOT	sink, restricted, flume, hadoop, p...
avro	ExecuteFlumeSource	1.4.0-SNAPSHOT	restricted, flume, get, hadoop, s...
aws	ExecuteProcess	1.4.0-SNAPSHOT	process, external, restricted, inv...
consume	ExecuteScript	1.4.0-SNAPSHOT	luaj, python, jython, clojure, js, ex...
csv	ExecuteStreamCommand	1.4.0-SNAPSHOT	command execution, stream, re...
database	FetchFile	1.4.0-SNAPSHOT	ingress, input, restricted, get, file...
fetch	FetchHDFS	1.4.0-SNAPSHOT	restricted, get, fetch, hdfs, hado...
files	FetchParquet	1.4.0-SNAPSHOT	restricted, get, fetch, HDFS, had...
get	GetFile	1.4.0-SNAPSHOT	ingress, input, restricted, get, file...
hadoop	GetHDFS	1.4.0-SNAPSHOT	restricted, get, fetch, HDFS, had...
ingest	InvokeScriptedProcessor	1.4.0-SNAPSHOT	luaj, python, jython, ivthon, restr...
insert			
json			
listen			
logs			
message			
put			
remote			
restricted			
source			
split			
sql			
text			
update			

DeleteHDFS 1.4.0-SNAPSHOT org.apache.nifi - nifi-hadoop-nar

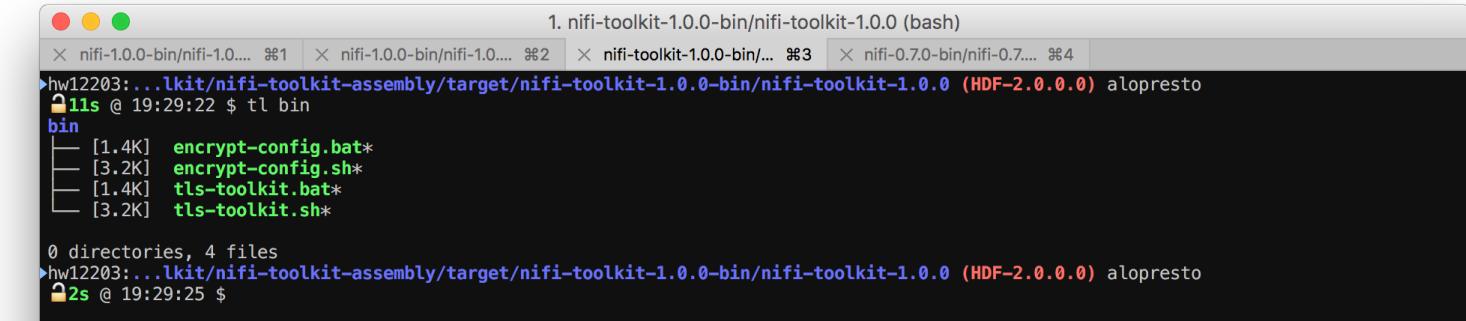
Deletes one or more files or directories from HDFS. The path can be provided as an attribute from an incoming FlowFile, or a statically set path that is periodically removed. If this processor has an incoming connection, it will ignore running on a periodic basis and instead rely on incoming FlowFiles to trigger a delete. Note that you may use a wildcard character to match multiple files o...

CANCEL ADD



Secure Configuration

- Previously, sensitive configuration values were exposed in **plaintext**
 - Recommendation for POSIX file permissions to prevent unauthorized access
- Pluggable architecture which allows **NiFiProperties** to be secured until application load
- Passwords protected by default; arbitrary values can be protected via **nifi.sensitive.props.additional.keys**



```
1. nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (bash)
hw12203:...lkit/nifi-toolkit-assembly/target/nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (HDF-2.0.0.0) alopresto
1ls @ 19:29:22 $ tl bin
bin
[1.4K] encrypt-config.bat*
[3.2K] encrypt-config.sh*
[1.4K] tls-toolkit.bat*
[3.2K] tls-toolkit.sh*
0 directories, 4 files
hw12203:...lkit/nifi-toolkit-assembly/target/nifi-toolkit-1.0.0-bin/nifi-toolkit-1.0.0 (HDF-2.0.0.0) alopresto
2s @ 19:29:25 $
```



```
73 # security properties #
74 nifi.sensitive.props.key=n2z+tTbHuZ4V4V2 | jwWhdasxvXD4ZG2lMAes/vqh6u4vaz4xgL4aEbF4Y/dXevgk3uLRc0wf1vc4RDQ
75 nifi.sensitive.props.key.protected=aes/gcm/256
76 nifi.sensitive.props.algorithm=PBEWITHMD5AND256BITAES-CBC-OPENSSL
77 nifi.sensitive.props.provider=BC
78 nifi.sensitive.props.additional.keys=
79
80 nifi.security.keystore=/path/to/keystore.jks
81 nifi.security.keystoreType=JKS
82 nifi.security.keystorePasswd=oBjT92hIGRElIG0h | JMZ6uYuWNBBr0A6usq/Jt3DaD2e4otNirZDytac/w/KFe0H0krJR03ycho
83 nifi.security.keystorePasswd.protected=aes/gcm/256
84 nifi.security.keyPasswd=ac/BaE35SL/esLiJ | +ULRvRLYdIDA2VqpE0eQXDEMjaLBMG2kbK0d0wBk/hGebDKlVg
85 nifi.security.keyPasswd.protected=aes/gcm/256
86 nifi.security.truststore=
87 nifi.security.truststoreType=
88 nifi.security.truststorePasswd=X/RSINr2QCJ1Kwe | dENJevX5P61ix+97airrtoB0oyasMFS6DG6fHbX+SZtw2VAMlISSnDeT97Q
89 nifi.security.truststorePasswd.protected=aes/gcm/256
90 nifi.security.needClientAuth=
91 nifi.security.user.authorizer=
```

Config Encryption Tool

- First implementation
 - Future options: HSM integration, Vault, KeyWhiz, etc.
- Command-line tool which accepts plaintext `nifi.properties` file and encrypts values using AES/GCM
 - 128 or 256-bit key depending on system JCE unlimited strength cryptographic jurisdiction policies
- Persists master key in `bootstrap.conf`
- Can also handle `login-identity-providers.xml`, `authorizers.xml`, and `flow.xml.gz`

Encrypted Provenance Repository

- Every provenance event record is encrypted with AES G/CM before being persisted to disk
 - Decrypted on deserialization for retrieval/query
 - Random access via offset seek
 - Handles key migration & rotation

NiFi Data Provenance							
Displaying 3 of 3				Showing the events that match the specified query. Clear search			
Filter		by component name					
Date/Time	Type	FlowFile UUID	Size	Component Name	Component Type		
06/05/2017 20:17:4...	CONTENT_MODIFIED	d602bdfd-9d14-4c2e...	77 bytes	ConvertRecord	ConvertRecord	 →	 →
06/05/2017 20:17:4...	ROUTE	d602bdfd-9d14-4c2e...	46 bytes	LookupRecord	LookupRecord	 →	 →
06/05/2017 20:17:4...	FORK	f540f7cf-1e41-4cb7...	40 bytes	LookupRecord	LookupRecord	 →	 →

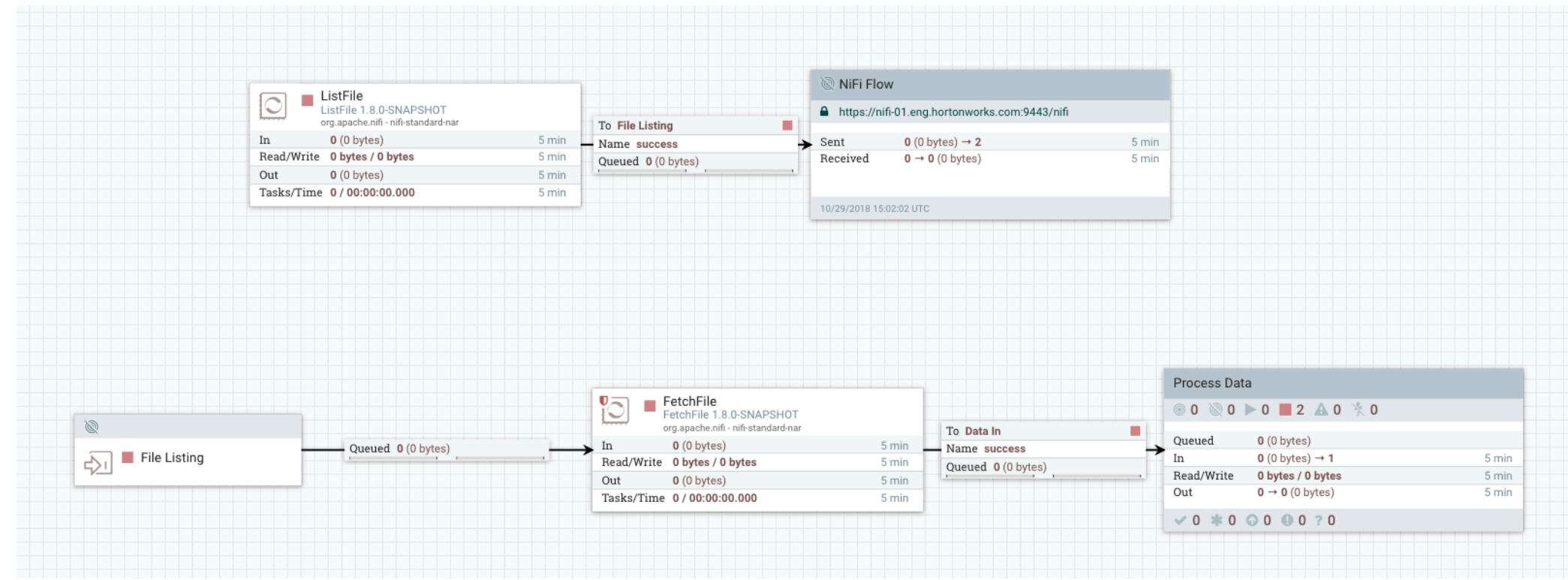
New Features

New Features in 1.8.0

- Cluster Data Management
 - Load-Balanced Connections
 - Node Decommissioning
- SQL results to record format
- Elasticsearch lookup service
- Docker improvements
- TLS Toolkit signing w/ external CA (standalone only)

Previously, on *NiFi Clusters*...

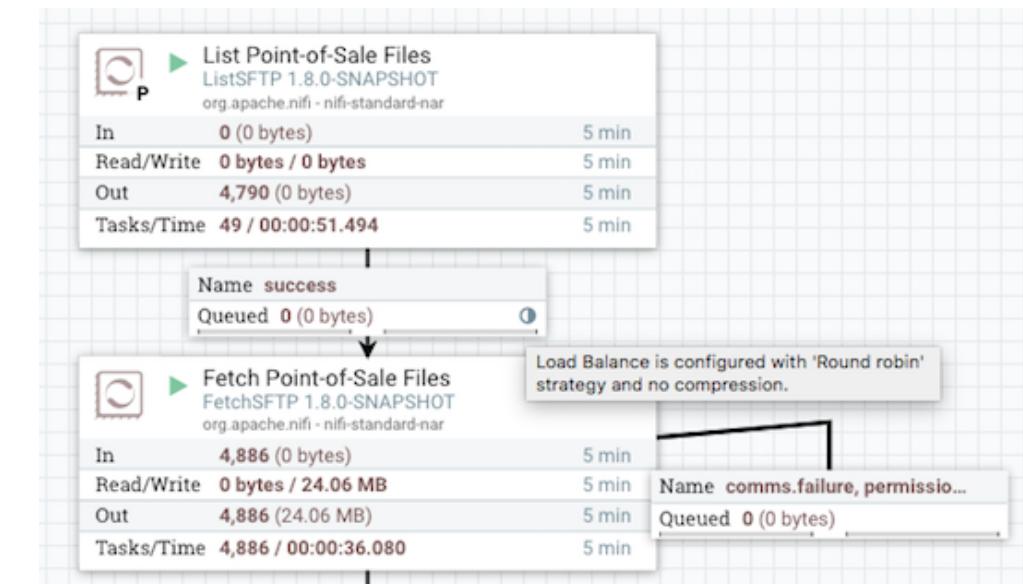
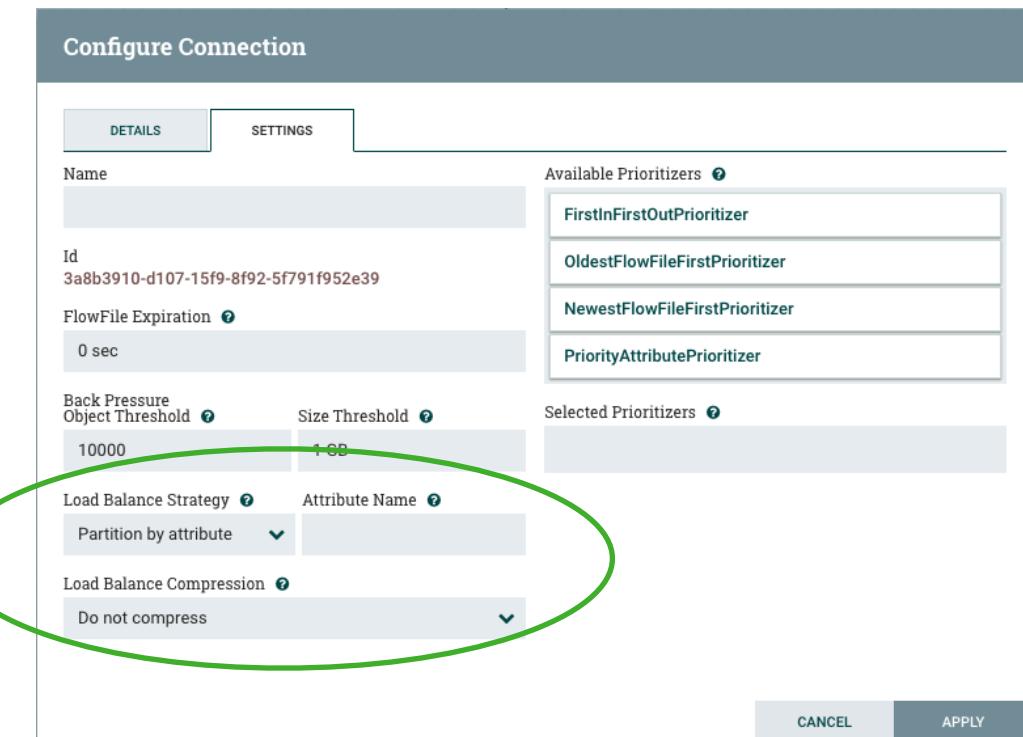
- Ingest activity (FTP, file read, etc.) performed on single node
- Routed via RPG to Input Port in cluster
- Flowfiles distributed across all nodes



[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

- **Encrypted content and flowfile repositories**
- **Encryption component enhancements**
- Cryptographic proof on sequences
- **Sensitive properties & variables**
- Sensitive Attributes
- **TLS Improvements**
- Encrypted Logs

Encrypted Content & Flowfile Repositories

- Similar to encrypted provenance repository
- Data stored on disk encrypted with referenceable keys & metadata
- Still allow random access (offset seek vs. linear read)
- Use framework-level **KeyProvider** interface
- HDF/NiFi ready for deployment on untrusted hardware w/o writing sensitive data* to disk
 - TLS certs and encryption keys still persisted

Encryption Component Improvements

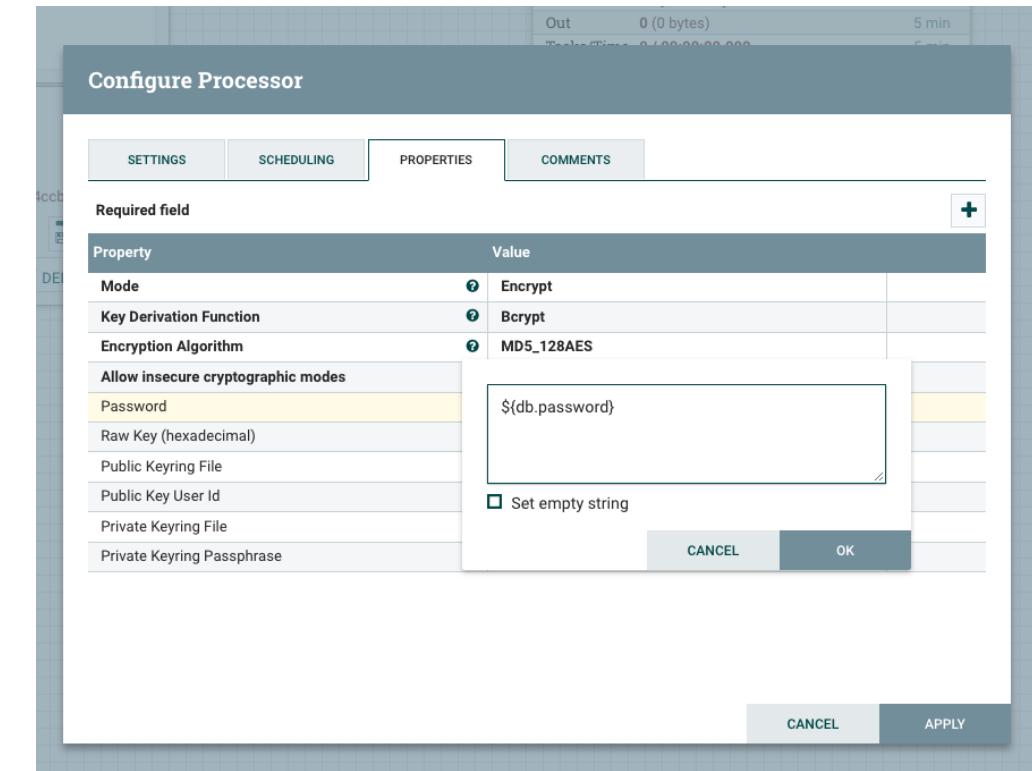
- **EncryptContent**
 - Improvements to key management
 - Enhanced metadata in attributes
- **EncryptAttribute**
- **EncryptRecord**
- Use framework-level **KeyProvider** controller service

Sensitive Properties & Variables

- Users want to promote flows between environments
- NiFi Registry enables this, but templates and versioned flows do not export sensitive properties
- Not all properties (and no passwords) support Expression Language
 - If `def = 123`, is `abc${def}` a 9 char password, or “`abc`” + `value(def)` = `abc123` ?
- Should enable:
 - Secure storage of sensitive variables on disk (i.e. encryption)
 - Secure export of sensitive variables to shared storage (i.e. encryption via Registry)
 - Key management (i.e. localized, derived, unique keys & encryption translation)
 - Access control for sensitive variables
 - Enable on-demand sensitive protection for dynamic properties

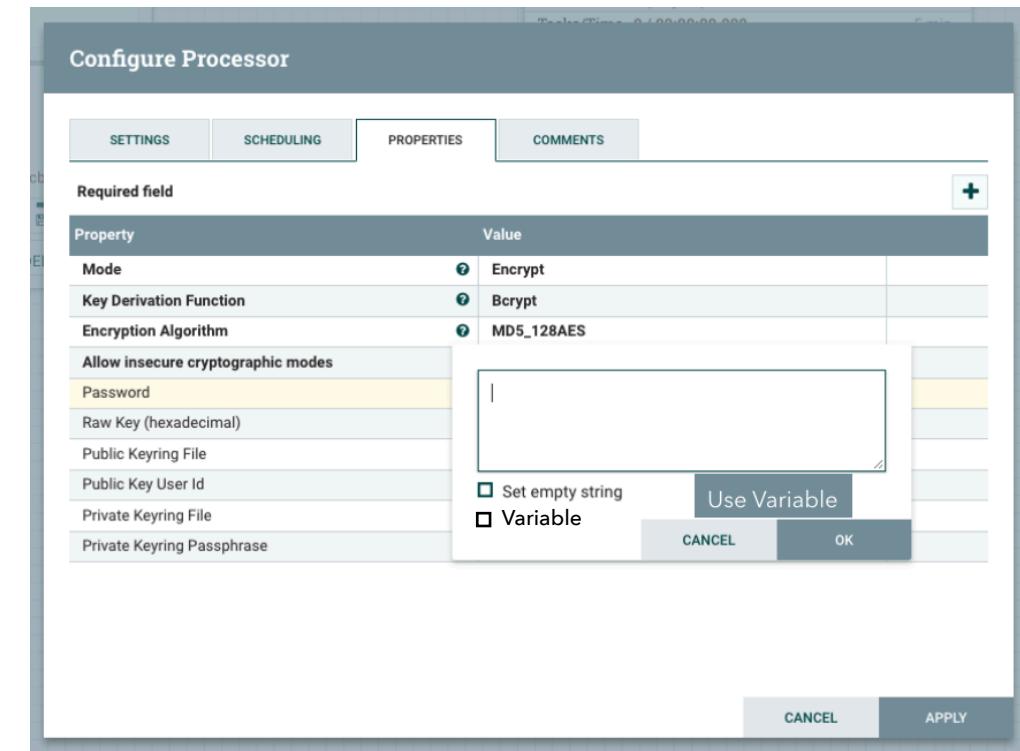
Current Situation

- No way to determine EL evaluation
- No way to protect dynamic properties



Proposed Changes

- Add “Variable” checkbox
 - Checked by default if property descriptor supports EL evaluation
 - If user un-checks, PD does not interpret EL (used for literal values containing `$\{...\}`)
- Add “Use Variable” button (auto-complete)



Use Variable

- Opens text field
- Allows free typing

Enter Variable Name...

Use Variable

- Typing triggers AJAX request to return variables
- New API endpoint
 - `GET /process-groups/abcd.../variables?q=d`
 - Text query filter
 - Permissions applied on server
 - Scope applied on server

d

data.endpoint

db.password

db.url

db.username

Variable Export

- Options
 - Never allow
 - Allow based on Registry admin setting (per-bucket?)
- Key management policies
 - Export in plaintext (**not secure**)
 - Export encrypted with original key (can't be deployed to other NiFi instances/rolled-over instances)
 - Export encrypted with per-Registry key (maintains independence of NiFi instances; preferred)
 - Import operation would then re-encrypt with destination NiFi's **nifi.sensitive.props.key**
 - Also possible to export with derived key to isolate buckets

TLS Improvements

- Handle keystores with multiple certificates to allow for varied identification by role
- Automatic Protocol/Cipher Suite Upgrades
 - Custom list of cipher suites and protocol versions loaded into Jetty at startup (work done; PR to be opened)
 - Integration with Mozilla TLS Labs tools
 - Plug n' Play Simplicity - “Set it and forget it”
 - Automatically sets and continuously analyzes supported cipher suites to ensure compatibility & risk minimization
- Extend TLS Toolkit
 - Better handle pre-existing certificates/certificate authorities
 - Help enterprise users securely deploy NiFi with less friction
 - Provide “sanity check” feature to test connectivity with provided keystore/truststore combos & identify potential issues
 - Refactor underway

Community

New Announcements

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



Community Health

apache / nifi

Unwatch ▾ 154 Unstar 1,103 Fork 1,041

Code Pull requests 145 Projects 0 Insights

Mirror of Apache NiFi

4,720 commits

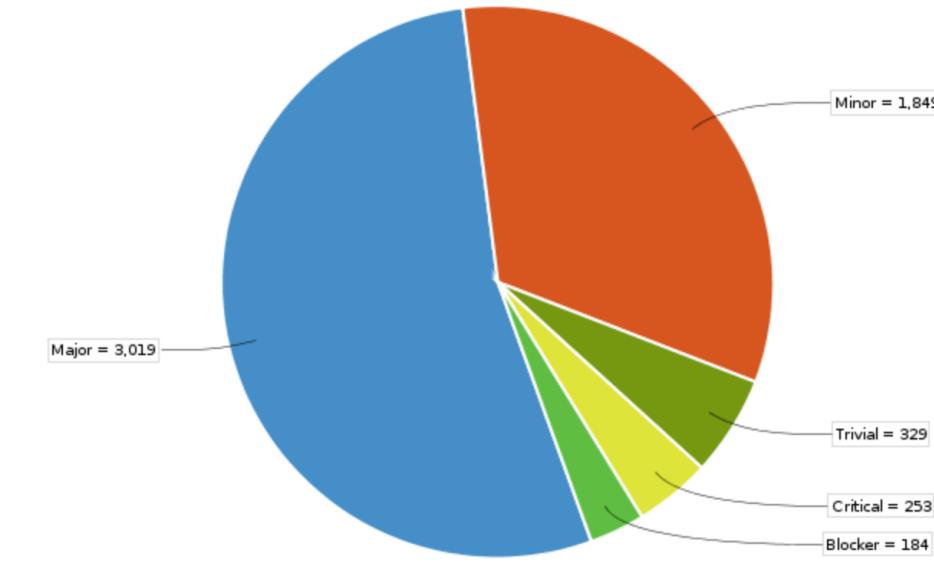
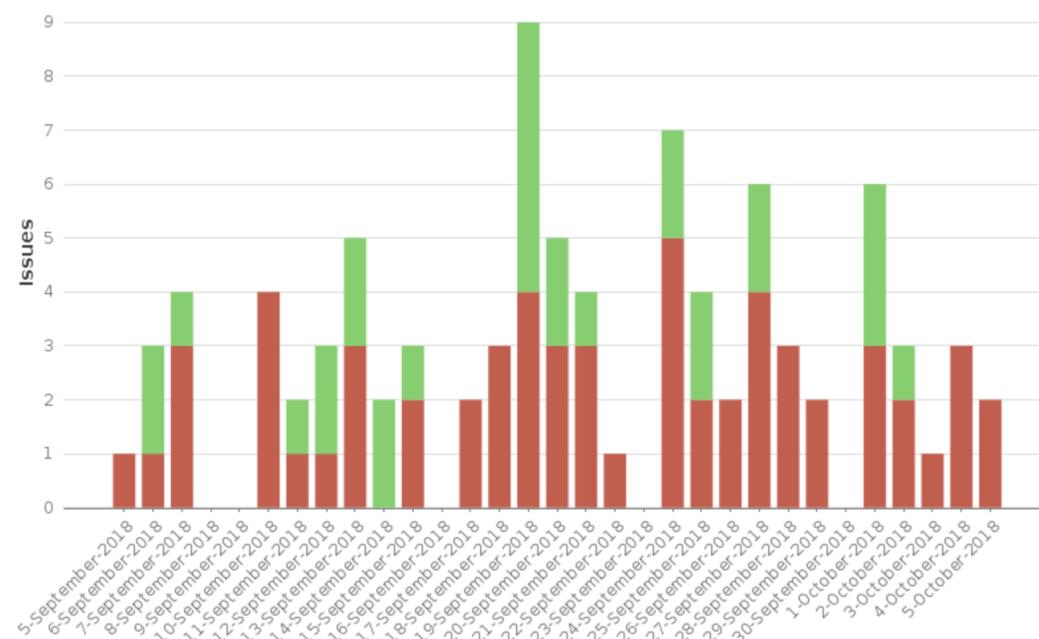
52 branches

63 releases

200 contributors

Apache-2.0

This chart shows the issues created in the last 30 days



Learn more and join us

Apache NiFi site

<https://nifi.apache.org>

Subproject MiNiFi site

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

Subscribe to and collaborate at

dev@nifi.apache.org

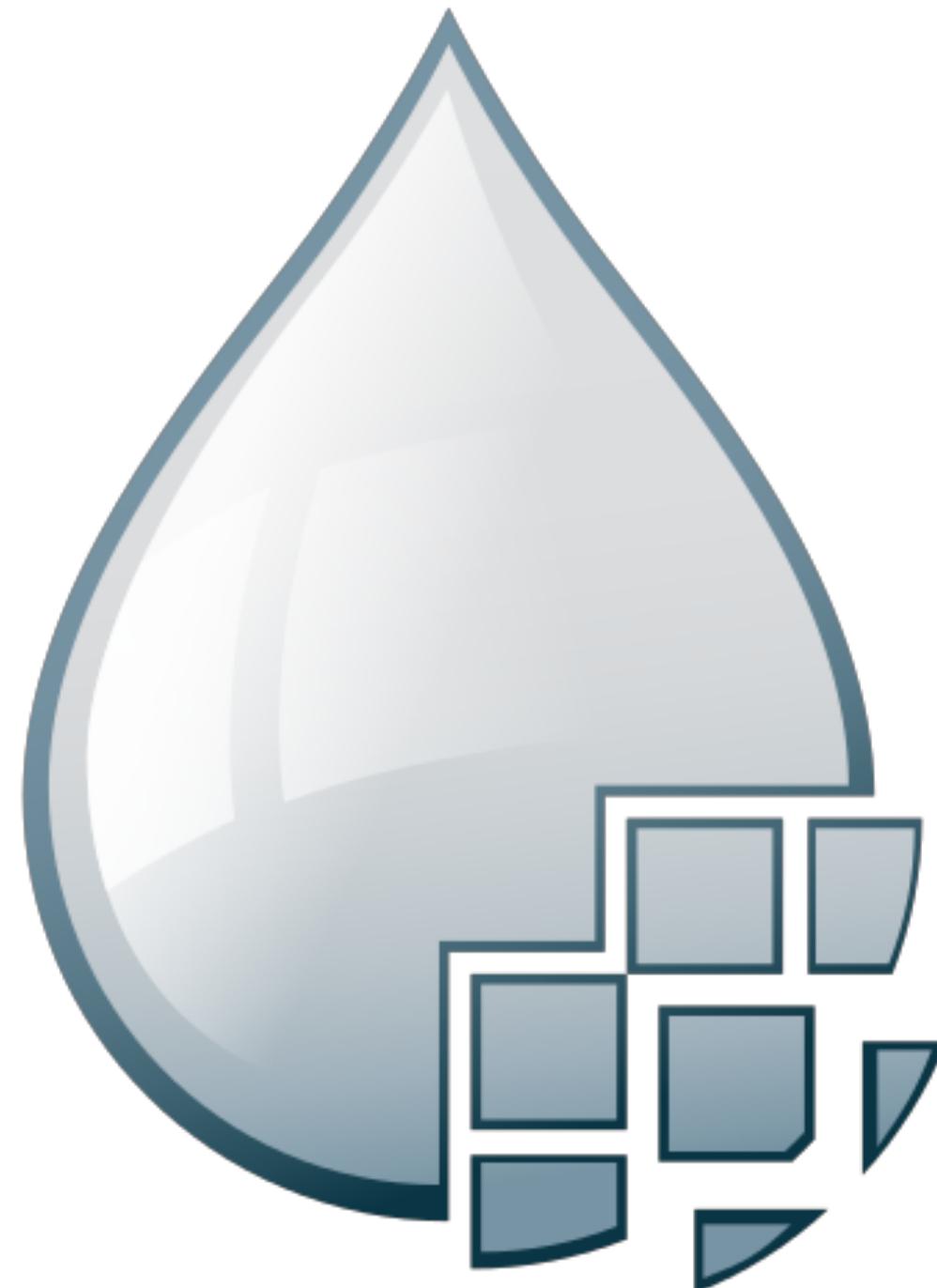
users@nifi.apache.org

Submit Ideas or Issues

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

Follow us on Twitter

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





Thank you

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