



Driving Connected Vehicle Innovation with Apache NiFi and MiNiFi

Andy LoPresto | @yolopey

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

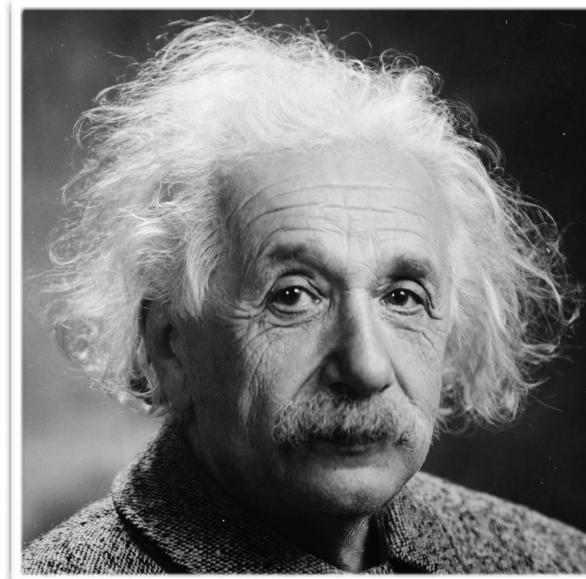
18 June 2018 Silicon Valley Automotive Open Source Meetup San Jose

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

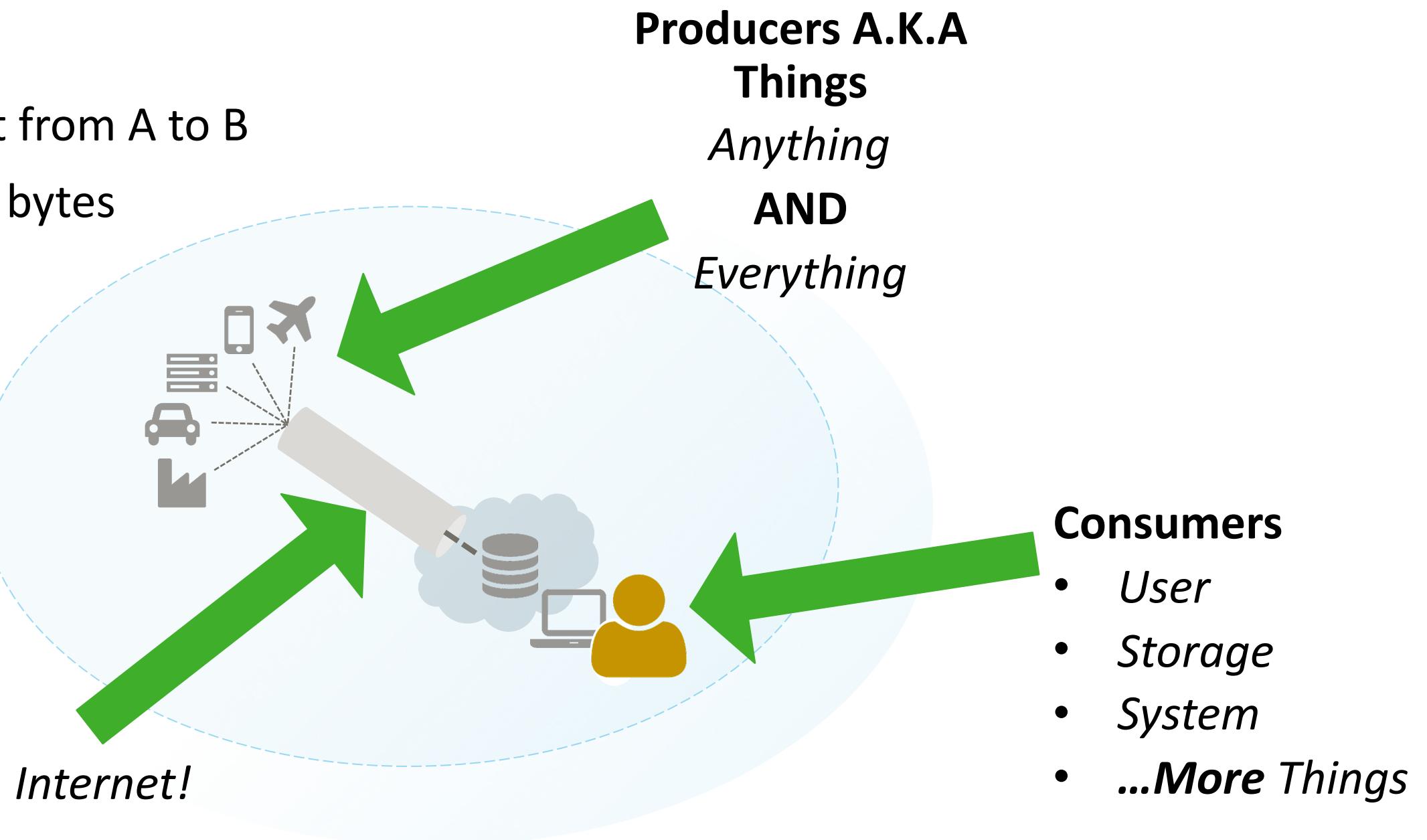
- Introduction
 - What is dataflow?
 - What are NiFi & MiNiFi?
 - How can NiFi/MiNiFi enhance connected vehicles?
 - What's next?
-
- All slides provided online, so no need to transcribe



What is dataflow?

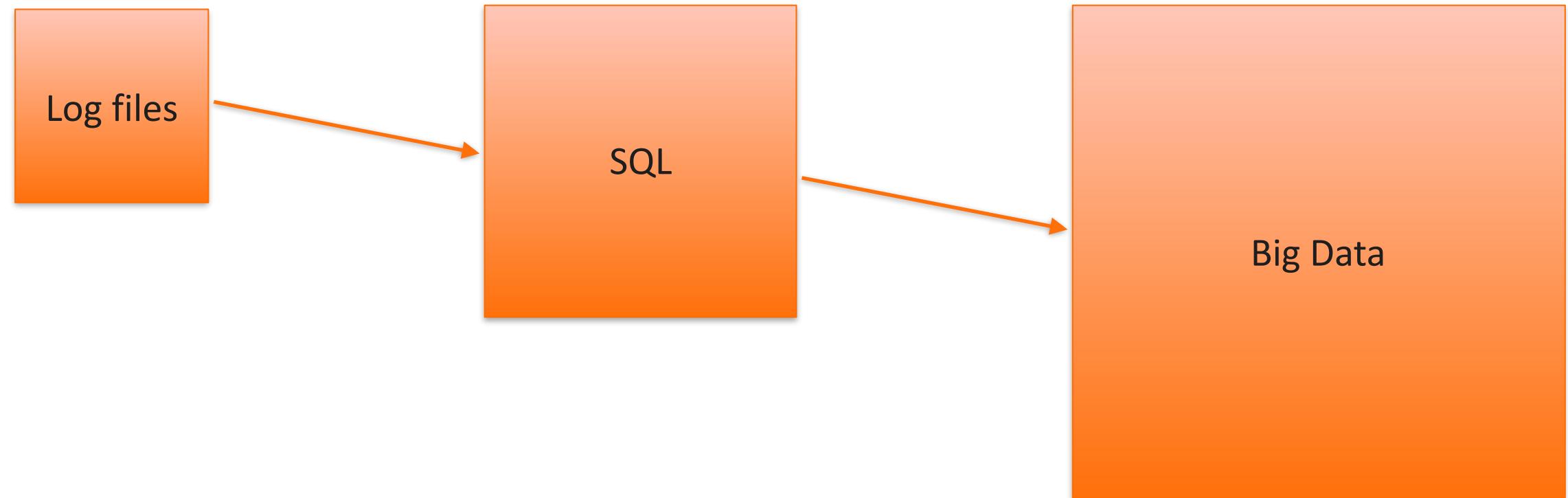
What is dataflow?

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



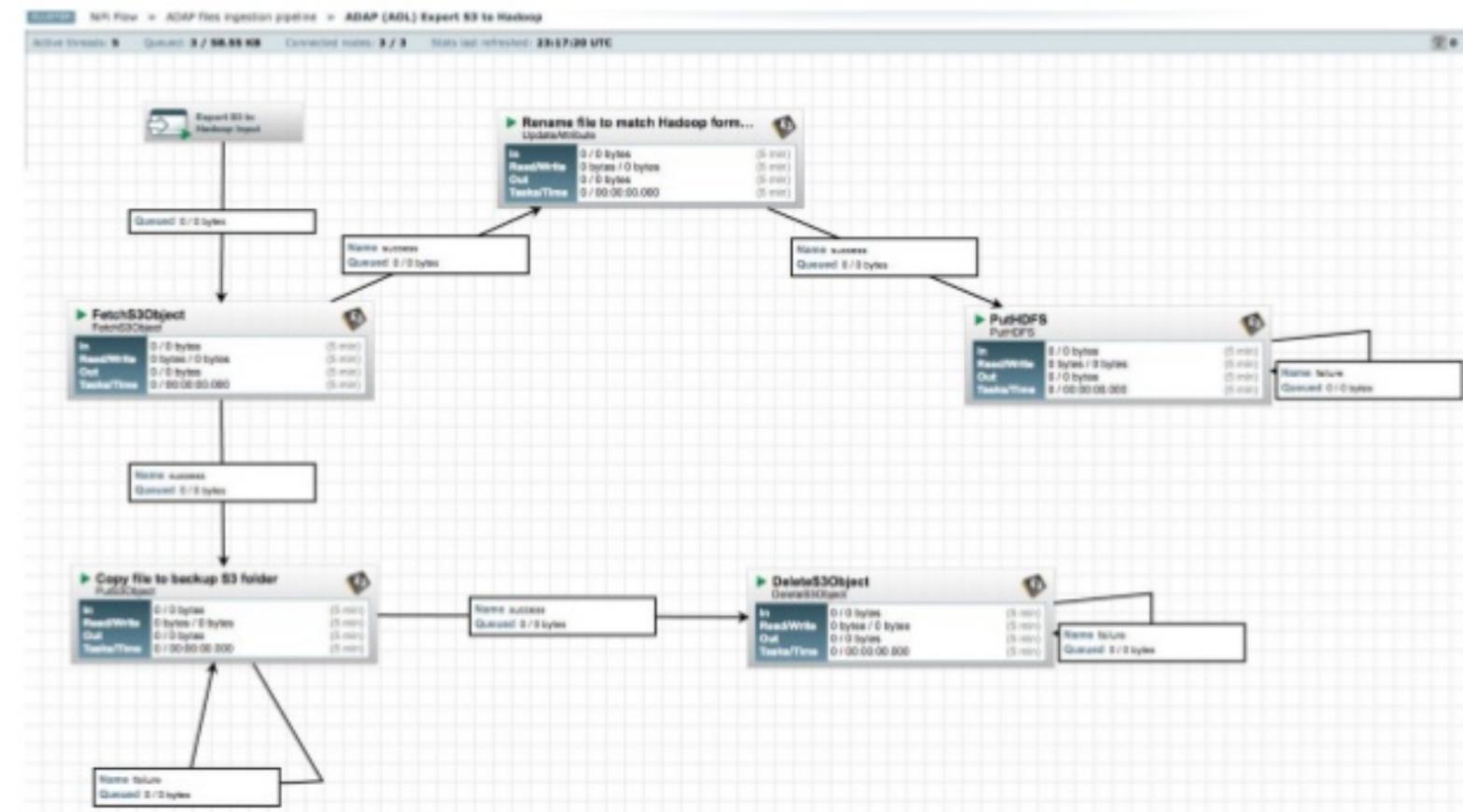
Connecting Data Points Is Easy

- Simple enough to write a process
 - Bash/Ruby/Python
 - SQL proc
 - etc.



Big Data Is About Scale...

- ...and this doesn't scale
- Example use case:
 - AOL Data Processing
 - AWS -> HDFS
 - 20 TB ingested/day
 - Lev Brailovskiy, “Data Ingestion and Distribution with Apache NiFi”, Slide 27, 02/2017
 - <https://www.slideshare.net/LevBrailovskiy/data-ingestion-and-distribution-with-apache-nifi>



Dataflow Challenges In 3 Categories

Data

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

Infrastructure

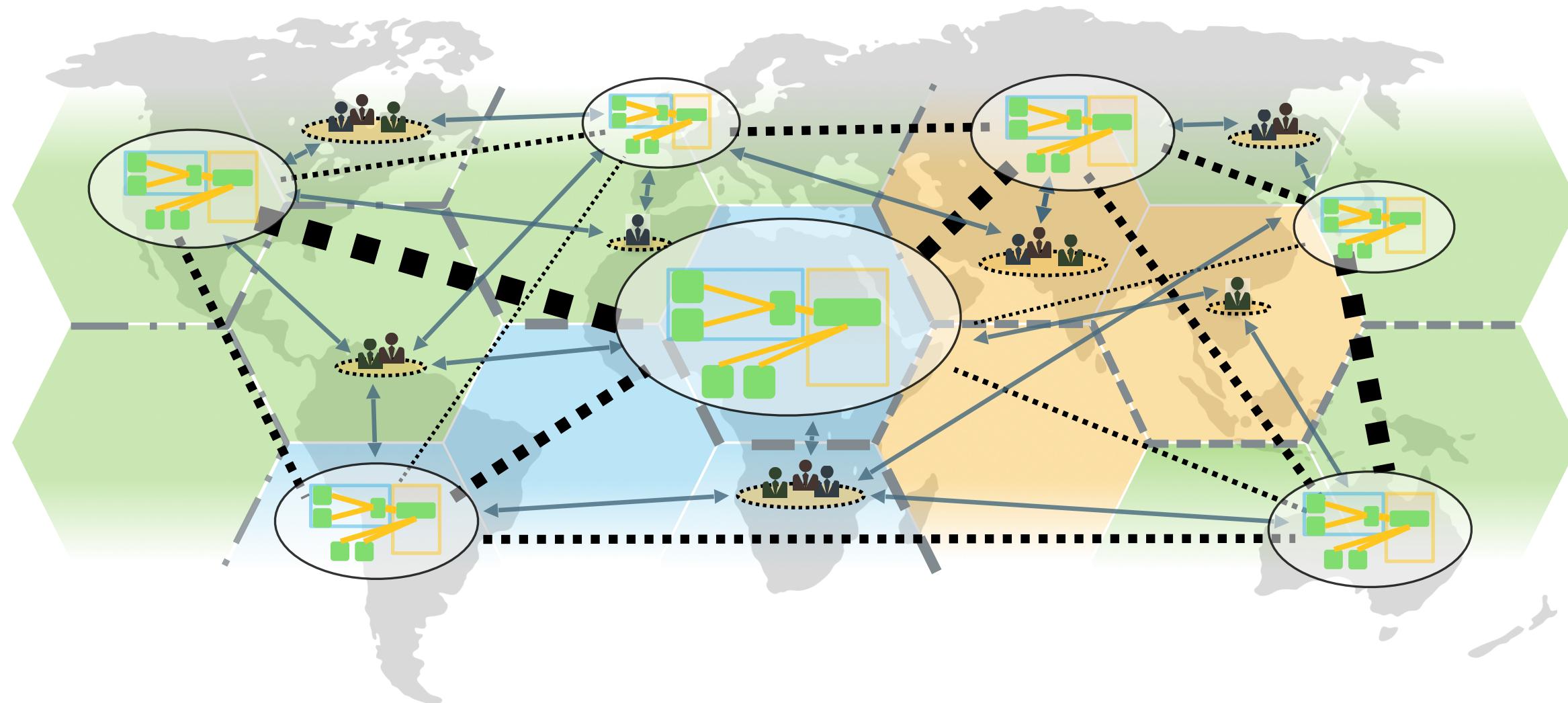
- “Exactly Once”
Delivery
- Ensuring
Security
- **Overcoming**
Security
- Credential
Management
- Network

People

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

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

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



What are Apache NiFi and MiNiFi?

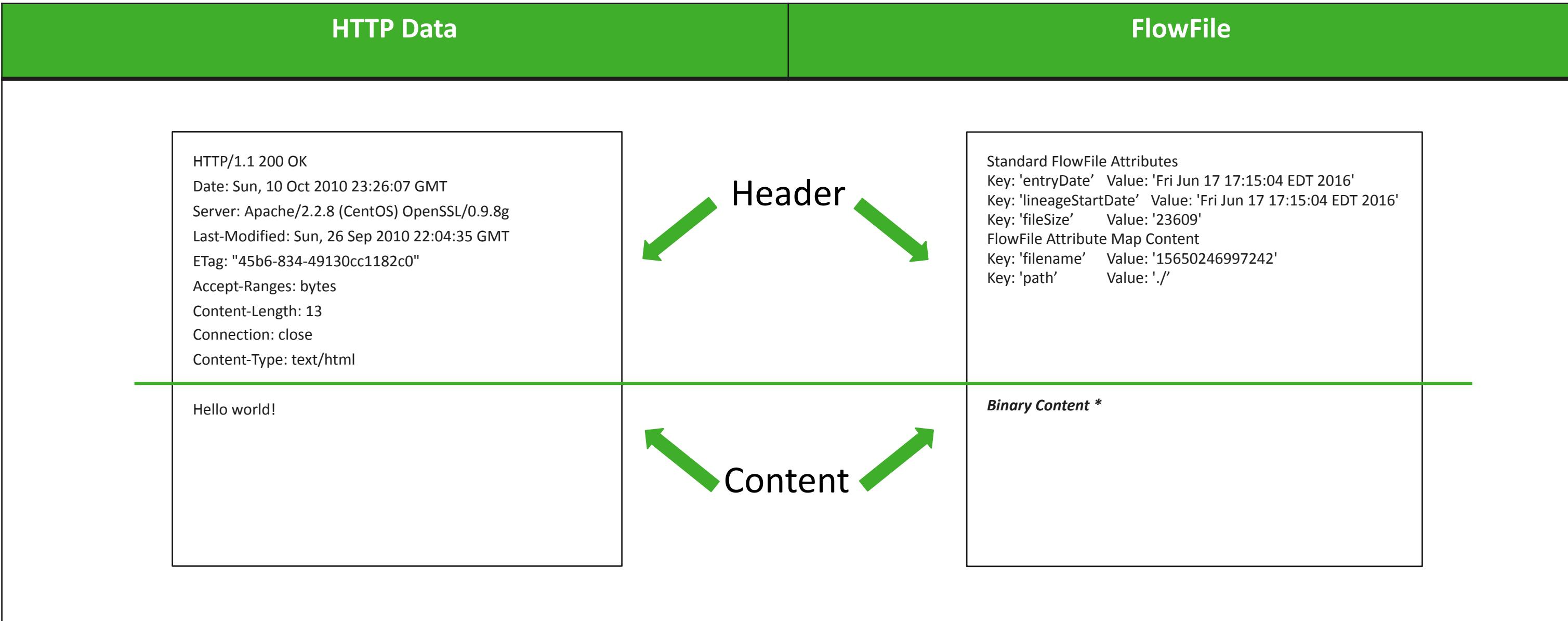
Apache NiFi

Key Features



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

Flowfiles Are Like HTTP Data



User Interface

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

```
1. scratch/release_verification (bash)

nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (bash)
nifi.flowfile.repository.always.sync=false
nifi.swap.manager.implementation=org.apache.nifi.controller.FileSystemSwapManager
nifi.queue.swap.threshold=20000
nifi.swap.in.period=5 sec
nifi.swap.in.threads=1
nifi.swap.out.period=5 sec
nifi.swap.out.threads=4

# Content Repository
nifi.content.repository.implementation=org.apache.nifi.controller.repository.FileSystemRepository
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
nifi.157s @ 17:40:39 $ nkfl

Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT

Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf

2017-01-30 17:40:43,812 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now Name.success
2017-01-30 17:40:43,854 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
2017-01-30 17:40:45,867 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.

hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
nifi.21s @ 17:40:49 $ subl conf/nifi.properties
hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
nifi.5s @ 17:40:55 $ nkfl

Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT

Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf

2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Starting Apache NiFi...
2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Working Directory: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT
2017-01-30 17:41:13,874 INFO [main] org.apache.nifi.bootstrap.Command Command: /Users/alopresto/.jenv/versions/1.8/bin/java -classpath /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/conf:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/jcl-over-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/jul-to-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/log4j-over-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/logback-classic-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-logback-core-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-api-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-documentation-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-framework-api-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-nar-util-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-properties-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-runtime-1.2.0-SNAPSHOT.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-slf4j-api-1.7.12.jar:/Dorg.apache.jasper.compiler-dislejir19=true-Xmx512m-Xms512m-Dsun.net.http.allowRestrictedHeaders=true-Djava.net.preferIPv4Stack=true-Djava.awt.headless=true-XX:+UseGCC-0java.protocolHandler.pcks=sun.net.www.protocol-Dnifi.properties.path=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT:/conf/nifi.properties-Dnifi.bootstrap.listen.port=59722-Dapp=Nifi-Dorg.apache.nifi.bootstrap.config.log.dir=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT/logs org.apache.nifi.NiFi

hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
nifi.21s @ 17:41:17 $ nkfl

Java home: /Users/alopresto/.jenv/versions/1.8
NiFi home: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT

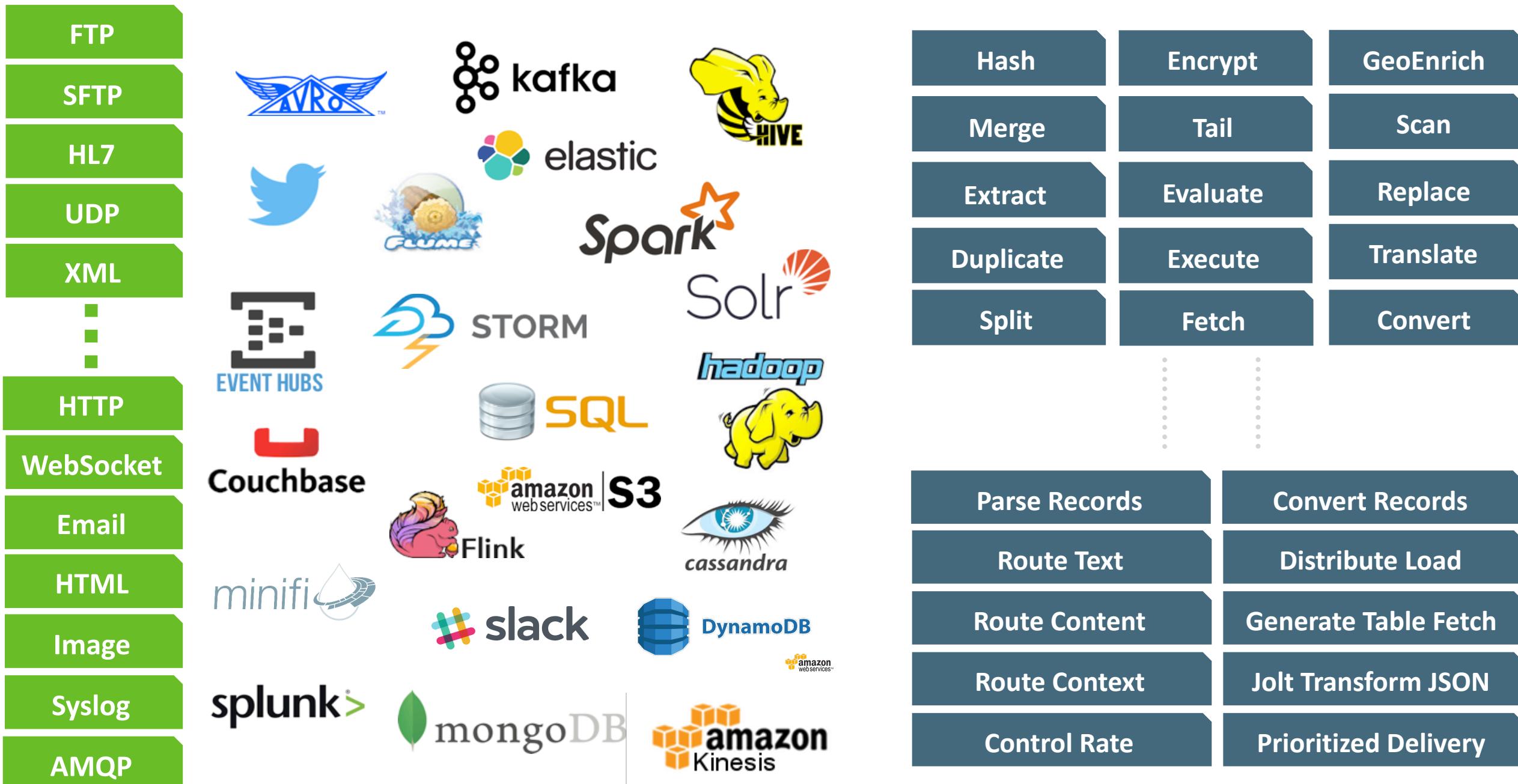
Bootstrap Config File: /Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/conf/bootstrap.conf

2017-01-30 20:54:26,971 INFO [main] org.apache.nifi.bootstrap.Command Apache NiFi has accepted the Shutdown Command and is shutting down now
2017-01-30 20:54:27,075 INFO [main] org.apache.nifi.bootstrap.Command Waiting for Apache NiFi to finish shutting down...
2017-01-30 20:54:29,085 INFO [main] org.apache.nifi.bootstrap.Command NiFi has finished shutting down.

hw12203:.../space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto
nifi.11595s @ 20:54:33 $ 
```



Deeper Ecosystem Integration: 260+ Processors, 48 Controller Services



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

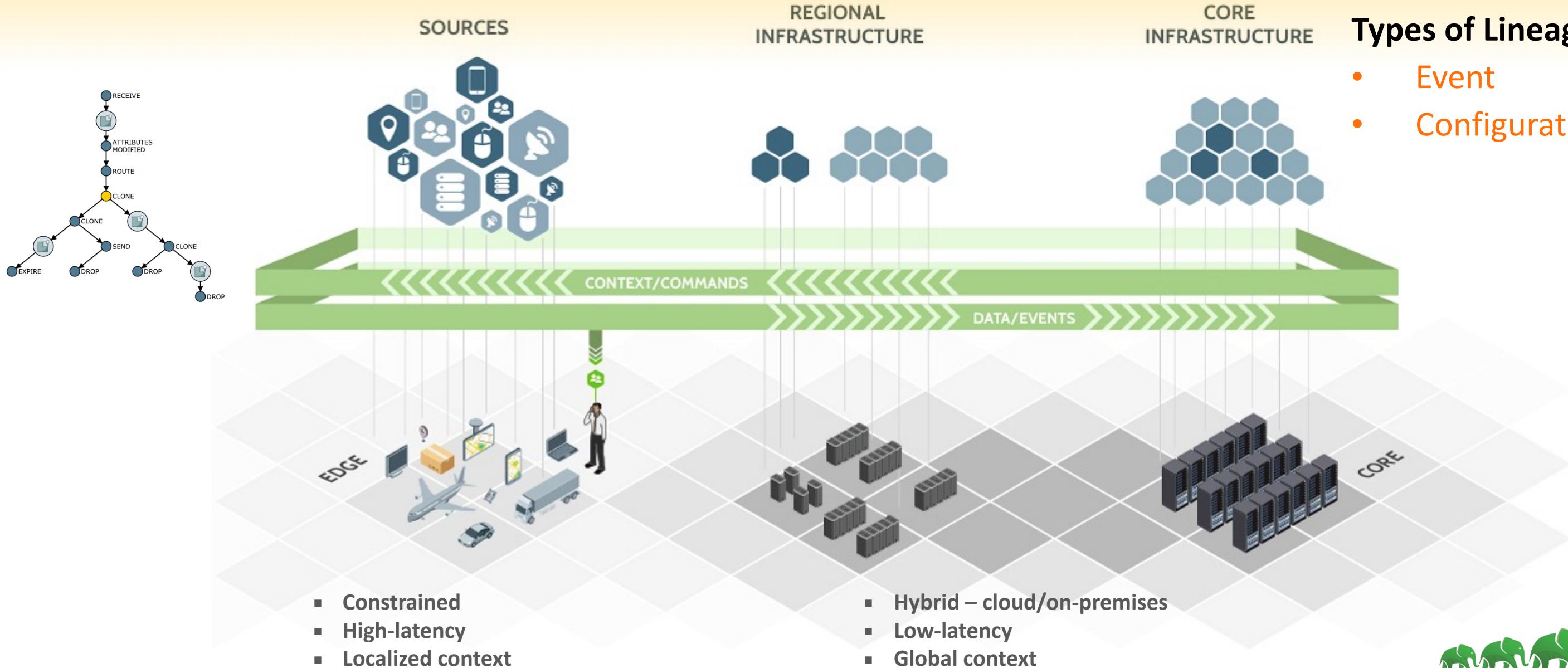
Data Provenance

Origin – attribution
Replay – recovery

Evolution of topologies
Long retention

Types of Lineage

- Event
- Configuration



IoT Challenges

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

Recent Examples

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



IoTPOT: Analysing the Rise of IoT Compromises

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

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

Google search results for "mirai botnet". The search bar shows "mirai botnet". The results page shows the following information:

- About 478,000 results (0.36 seconds)
- Mirai (Japanese for "the future", 未来) is malware that turns computer systems running Linux into remotely controlled "bots", that can be used as part of a botnet in large-scale network attacks. It primarily targets online consumer devices such as remote cameras and home routers.**
- Mirai (malware) - Wikipedia**
[https://en.wikipedia.org/wiki/Mirai_\(malware\)](https://en.wikipedia.org/wiki/Mirai_(malware))
Block en.wikipedia.org

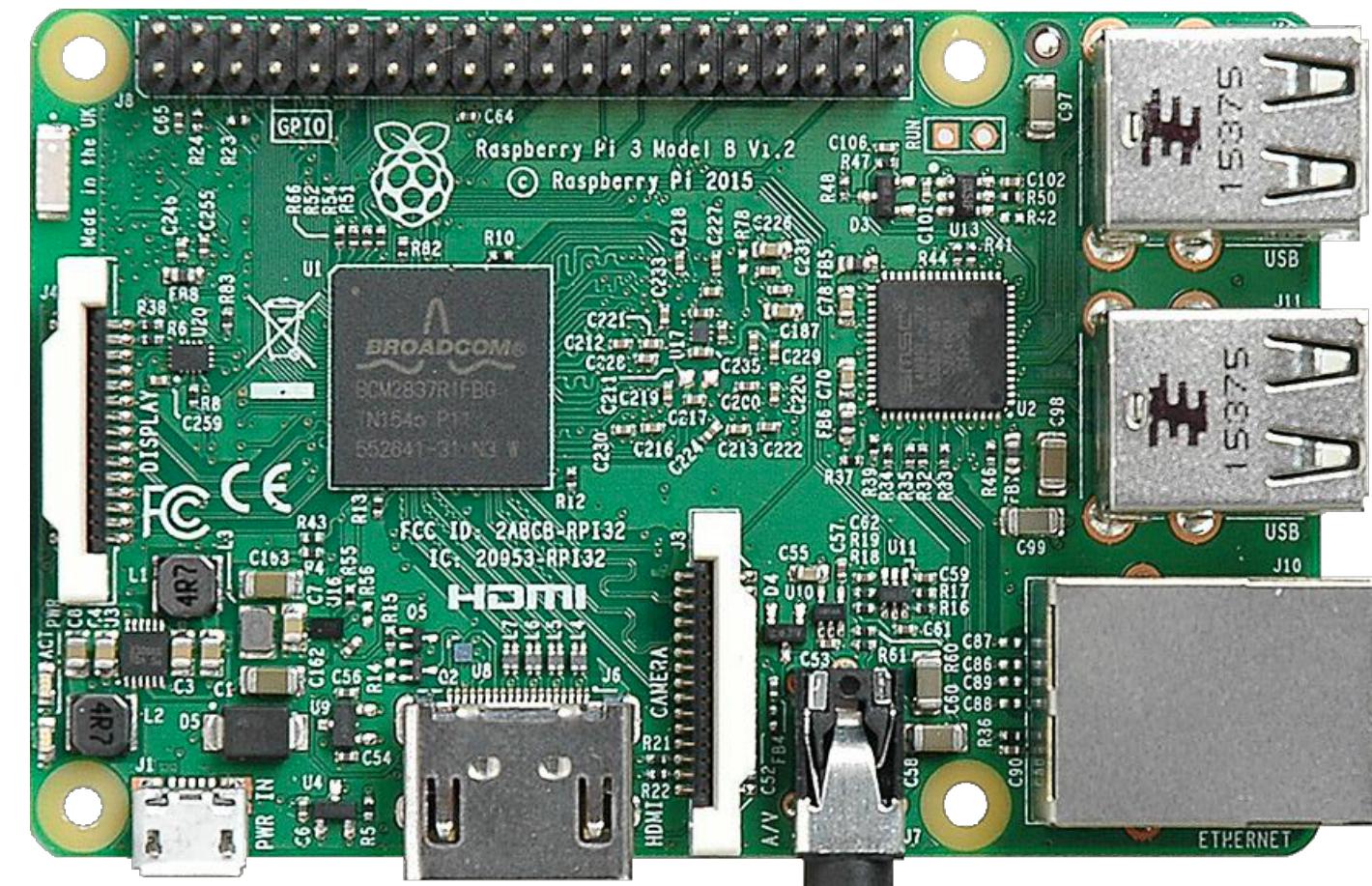
NiFi Solves Everything*

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

NiFi for IoT

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

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



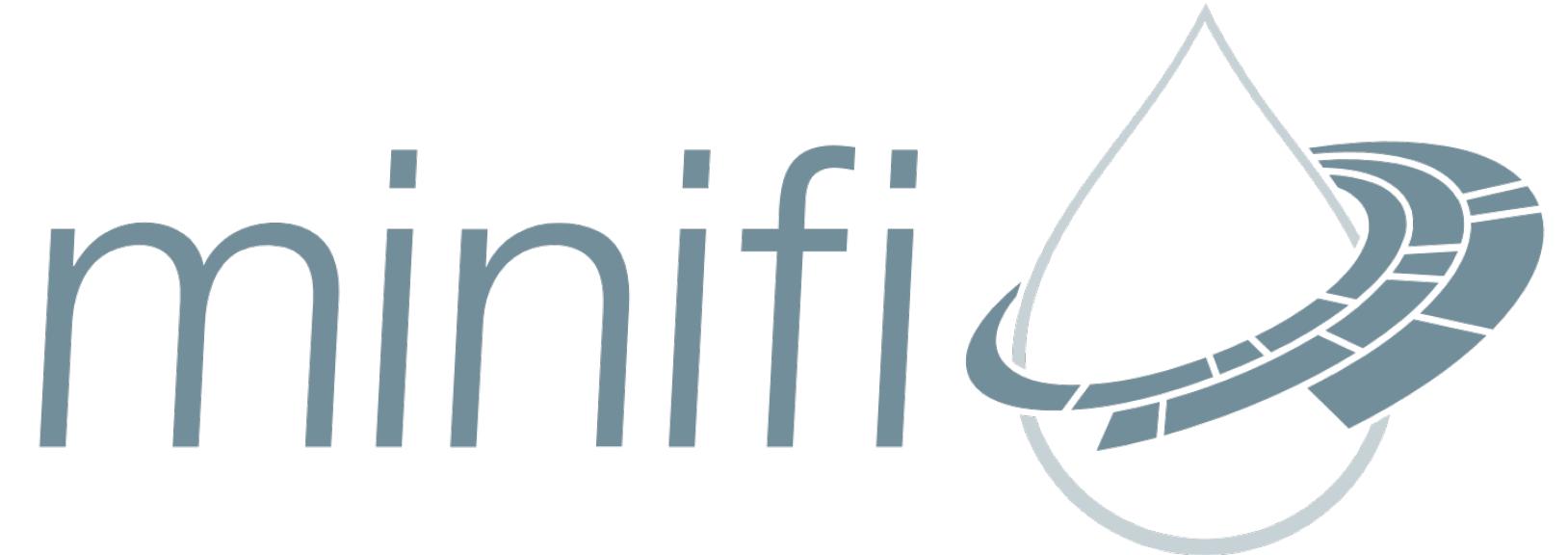
So Why Do We Need A Different Solution?

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

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

Apache NiFi Subproject: MiNiFi

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

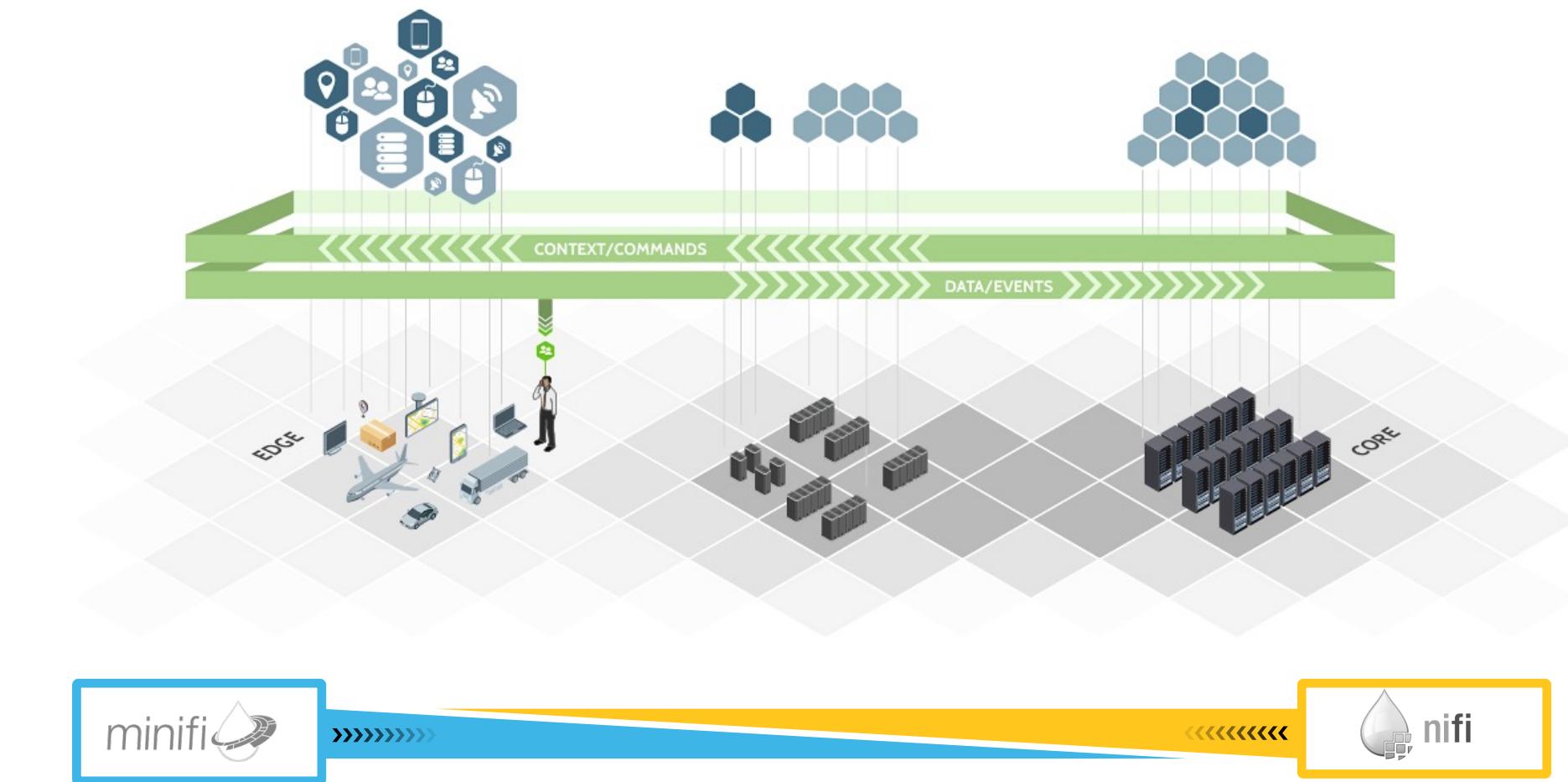


Why build MiNiFi?

- NiFi is big
 - 1.6.0 release is 1.2 GB compressed
 - Can be modified to run in restricted environments, but requires manual surgery
 - Provides UI, provenance query, etc.
 - Runs on dedicated machines/clusters — “owns the box”
- MiNiFi lives at the edge
 - No UI
 - 0.4.0 Java binary is 65 MB, 0.5.0 C++ binary is 5.2 MB (**0.2.0 fits on a floppy disk**)
 - “Good guest”

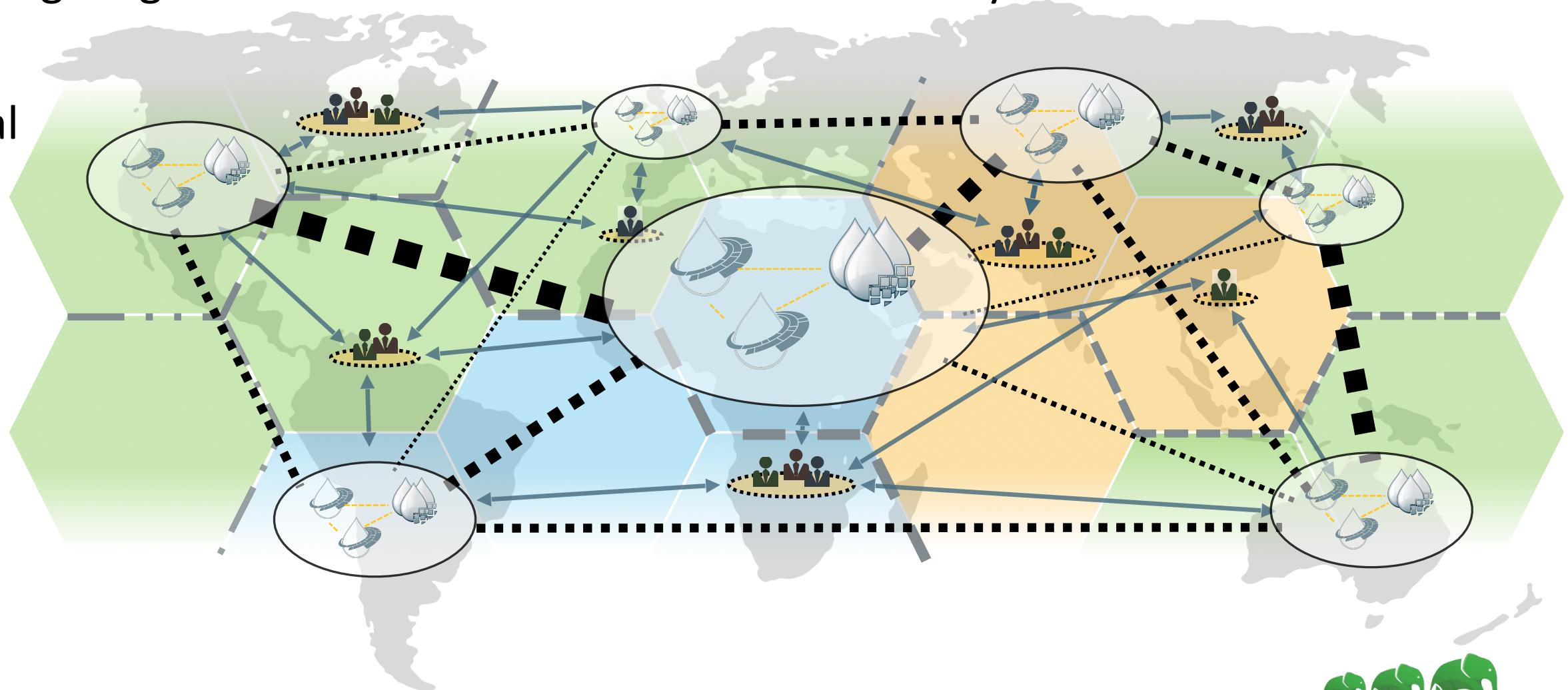
How Does MiNiFi Interact With NiFi?

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



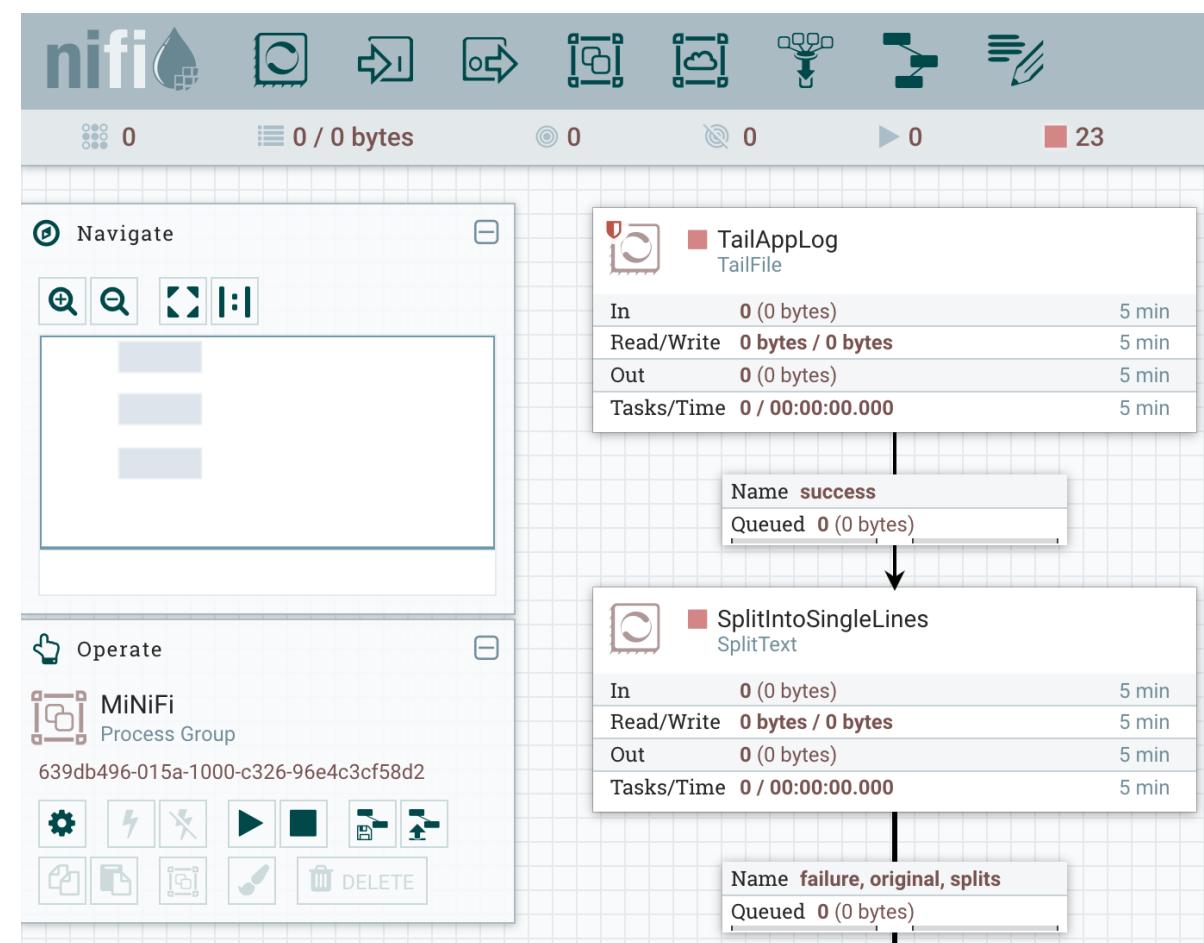
Let's Add Dimensionality

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

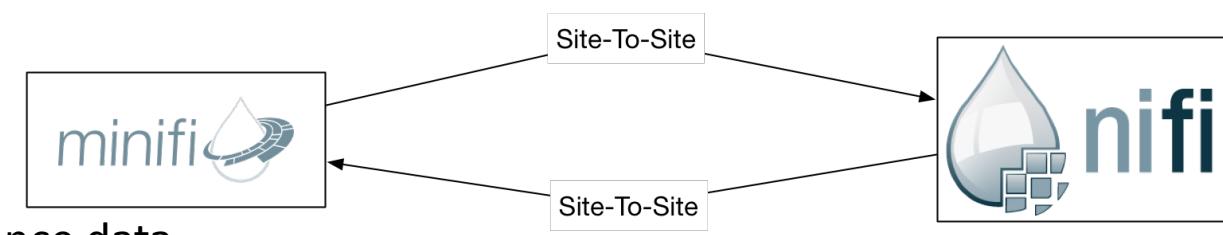


Flavors of MiNiFi

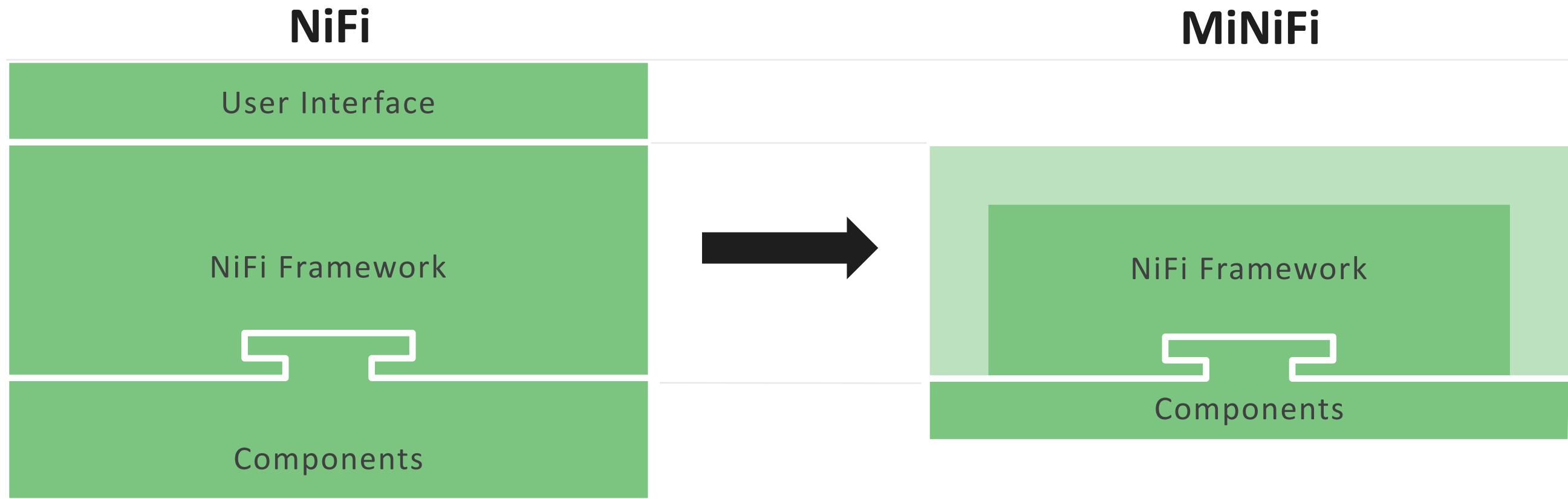
- MiNiFi Java (v0.4.0)
 - Modified version of NiFi
 - No UI
 - YAML configuration
 - Reduced processor count
 - 110+ by default, more available with additional NARs
- MiNiFi C++ (v0.5.0)
 - Written from scratch
 - 28 processors by default
 - Bi-directional site-to-site & provenance data



```
Security Properties:  
keystore: /tmp/ssl/localhost-ks.jks  
keystore type: JKS  
keystore password: localtest  
key password: localtest  
truststore: /tmp/ssl/localhost-ts.jks  
truststore type: JKS  
truststore password: localtest  
ssl protocol: TLS  
Sensitive Props:  
key:  
algorithm: PBWEWITHMD5AND256BITAES-CBC-OPENSSL  
provider: BC  
  
Processors:  
- name: TailAppLog  
  class: org.apache.nifi.processors.standard.TailFile  
  max concurrent tasks: 1  
  scheduling strategy: TIMER_DRIVEN  
  scheduling period: 10 sec  
  penalization period: 30 sec  
  yield period: 1 sec  
  run duration nanos: 0  
  auto-terminated relationships list:  
    Properties:  
      File to Tail: logs/minifi-app.log  
      Rolling Filename Pattern: minifi-app*  
      Initial Start Position: Beginning of File  
- name: SplitIntoSingleLines  
  class: org.apache.nifi.processors.standard.SplitText  
  max concurrent tasks: 1  
  scheduling strategy: TIMER_DRIVEN  
  scheduling period: 0 sec  
  penalization period: 30 sec  
  yield period: 1 sec  
  run duration nanos: 0  
  auto-terminated relationships list:  
    - failure  
    - original  
  Properties:  
    Line Split Count: 1  
    Header Line Count: 0  
    Remove Trailing Newlines: true
```



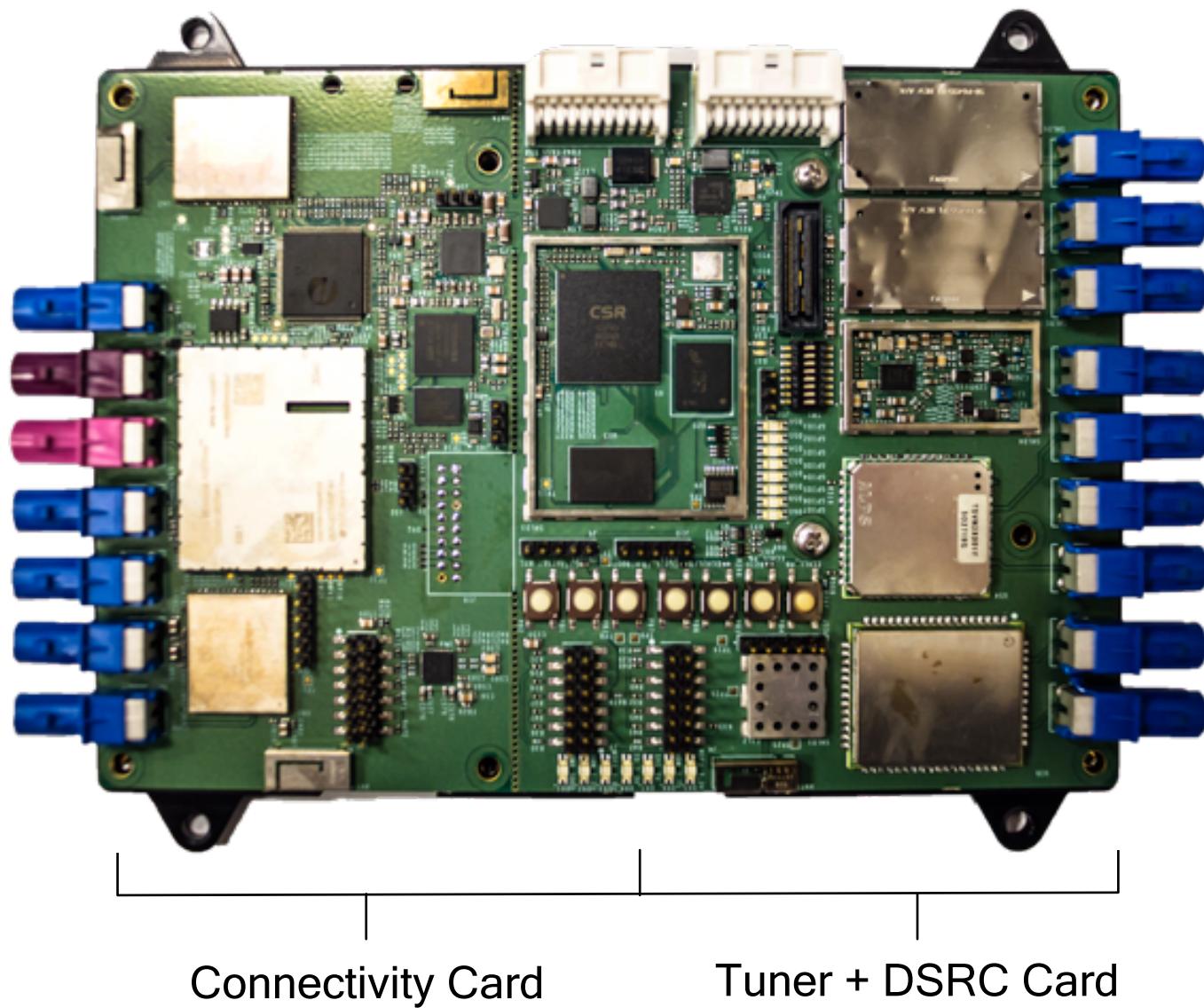
NiFi vs MiNiFi Java Processes



What does MiNiFi provide?

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

Connected Car Reference Platform Box



MiNiFi Exfil

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

Use Cases

Exploratory

- Introduce data collection to platform
- Scenario:
 - Capture vehicle velocity once per second
 - Capture GPS location once per 5 seconds
 - Send to data center via LTE



Intermediate

- Start to perform simple data processing on edge
- Scenario:
 - Capture data as described
 - If WiFi hotspot available, send all data
 - If no WiFi in 10 seconds, send all queued vehicle velocity readings and last GPS location to data center via LTE
 - Store all unsent data in queue for WiFi connection

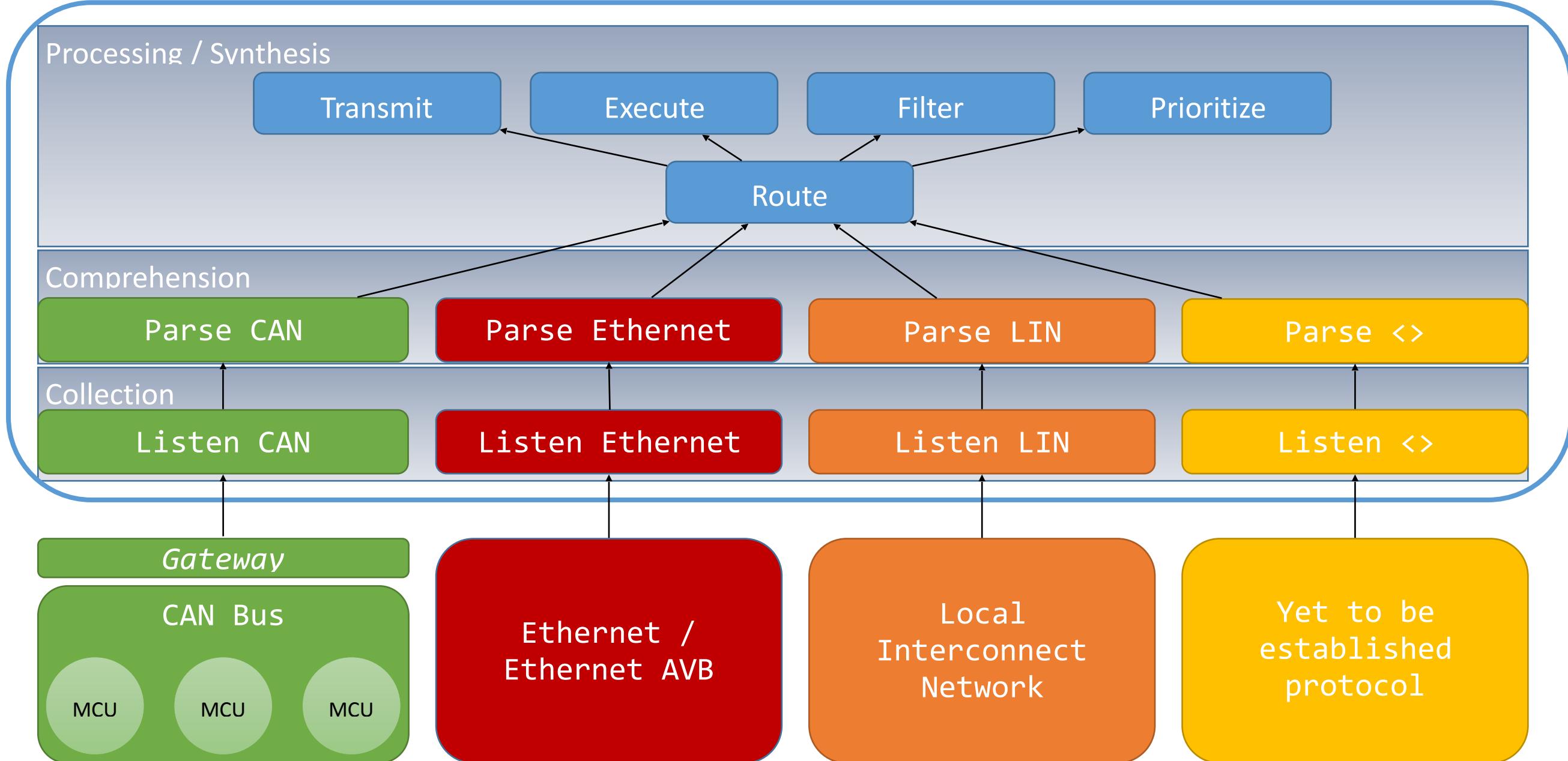


Advanced

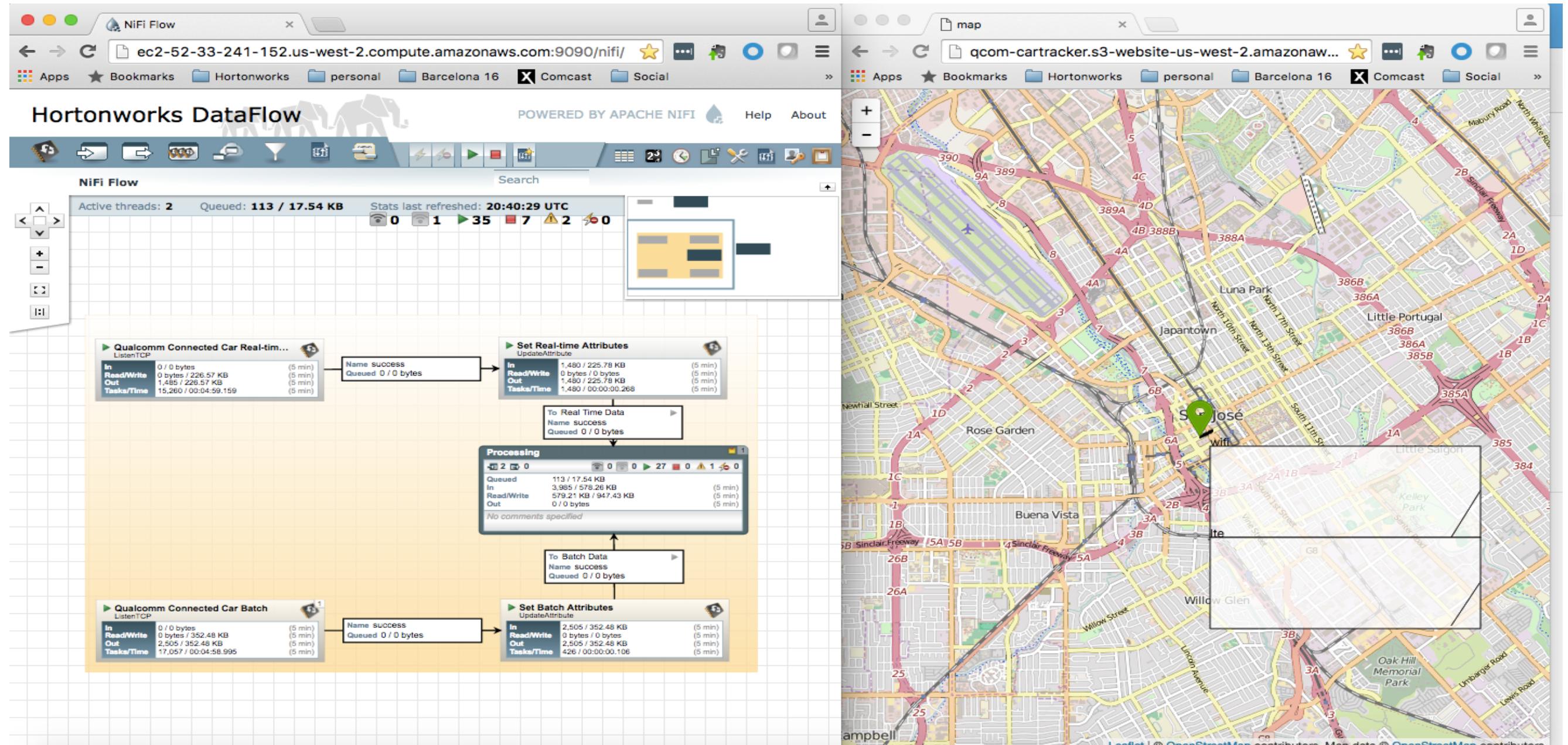
- Combine edge processing and core processing
- Scenario:
 - Capture vehicle data as described
 - Capture infotainment data once per 3 seconds
 - Capture braking data 100/s
 - Encrypt all data with per-car key
 - Drop GPS coordinates if in China
 - Send data according to existing plan, but if braking delta > 100psi OR > 30°C in the last second, prioritize and send immediately
 - Perform SEP on infotainment data to build profiles
 - Driver A prefers news on 1500AM during morning commute, but BT audio from iPhone during lunch (recent podcasts) and drive home (Genius playlist from “Working for the Weekend”)



MiNiFi on a Connected Car



MiNiFi on a Connected Car



A wide-angle photograph of a modern airport terminal. The ceiling is a massive glass and steel structure, allowing natural light to flood the space. Several escalators lead up to different levels. In the background, there are directional signs in Chinese and English. One sign points to "3号航站楼 4F" (Terminal 3, 4F) and "机场快轨" (Airport Express Train). Another sign indicates "行李提取厅" (Luggage Collection Hall) and "行李分拣厅" (Baggage Claim Hall). A few people are visible on the escalators.

What's Next?

New Announcements

Introducing Apache NiFi Registry

- NiFi 1.7.0 — ... June 2018 (200+ Jiras)
 - Hive, HBase, DB improvements
 - Asynchronous UI
 - Record processor improvements
- MiNiFi C++ 0.5.0 — 6 June 2018
- MiNiFi Java 0.5.0 — ... June 2018
- NiFi Registry 0.2.0 — ... June 2018



NiFi Registry for Dataflows

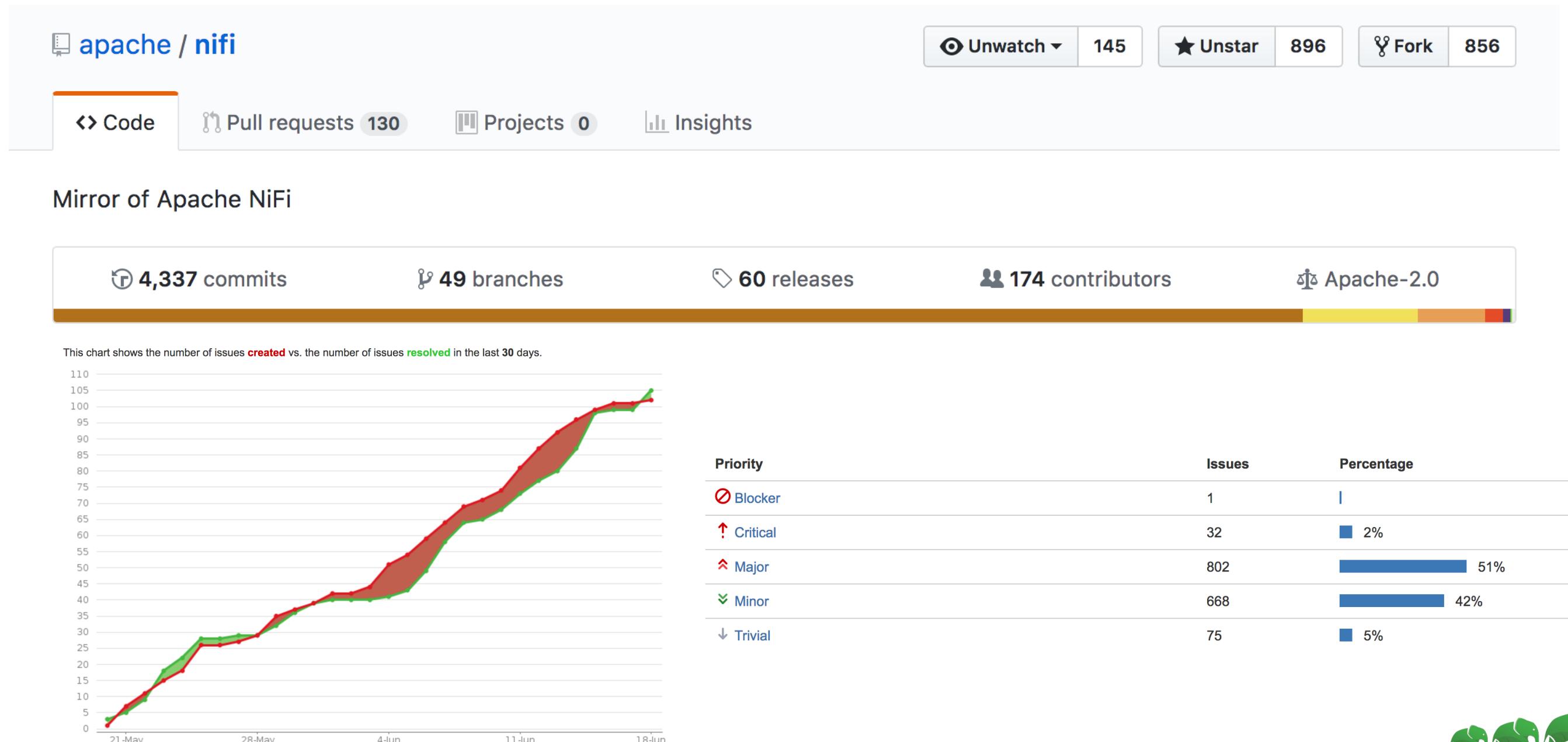
Introducing Apache NiFi Registry 0.2.0

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

The screenshot shows the Apache NiFi Registry interface. At the top, there's a navigation bar with the title "NiFi Registry / All" and user information "registry_user LOGOUT". Below the navigation bar, there's a search bar and sorting options "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", the "DESCRIPTION" column shows "Description 2". The "CHANGE LOG" section for Version 2 shows a blue circular icon with "Version 2 - 40 minutes ago" and "by registry_user". It also lists an action "Add processors" from "Dec-26-2017 at 11:23 PM". Another blue circular icon with "Version 1 - 41 minutes ago" and "by registry_user" is also present.

Learn more at [SDLC with Apache NiFi](#) by [Kevin Doran](#)
Thursday 21 June @ 1130, Executive Ballroom 210B/F

Community Health



Learn more and join us

Apache NiFi site

<https://nifi.apache.org>

Subproject MiNiFi site

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

Subscribe to and collaborate at

dev@nifi.apache.org

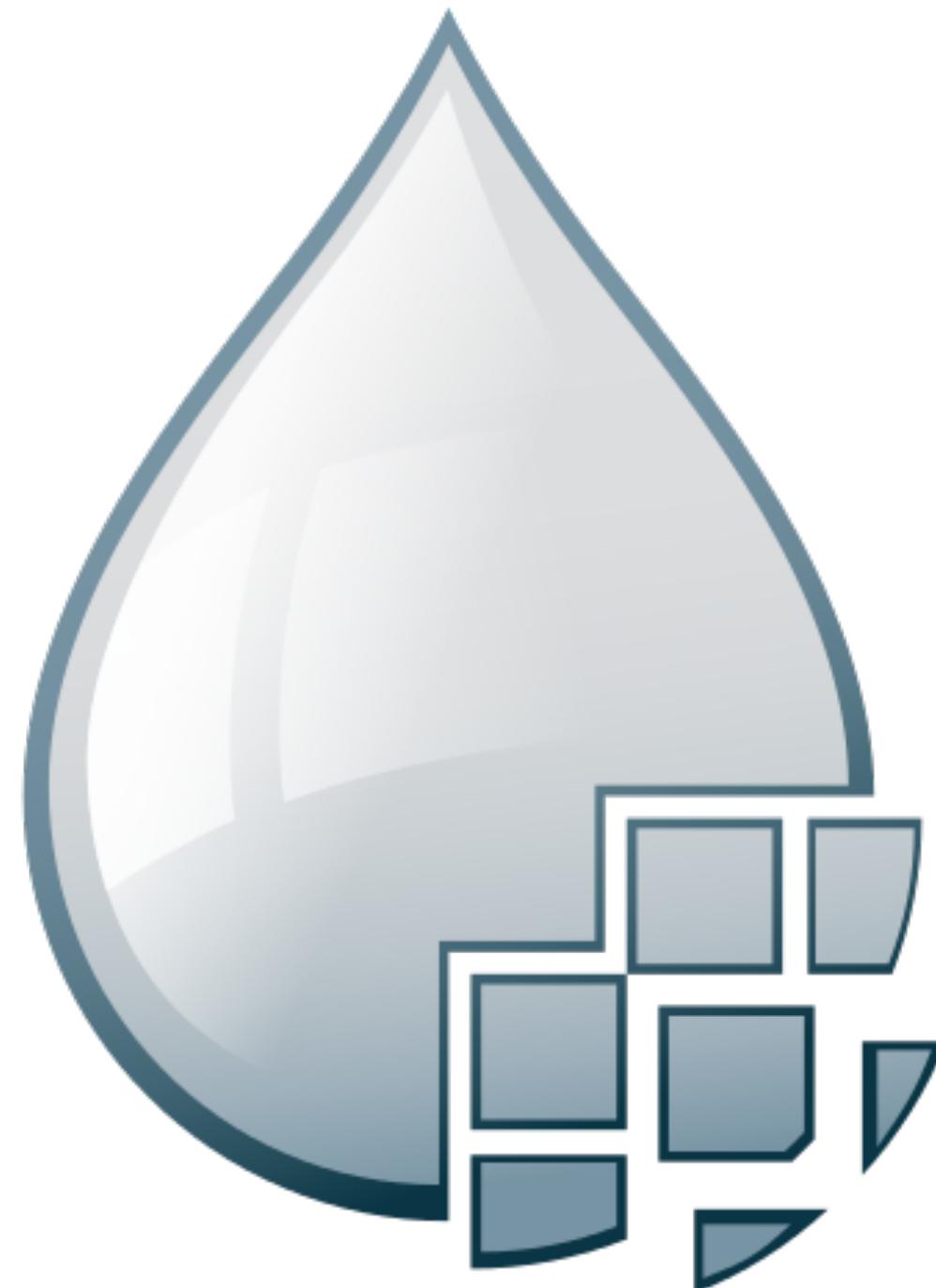
users@nifi.apache.org

Submit Ideas or Issues

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

Follow us on Twitter

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



More NiFi This Week...

Tuesday

Title	Time	Speaker(s)
Using NiFi to simplify data flow & streaming use cases @ Mastercard	1450 - 1530	Vishal Anand
Open source computer vision with TensorFlow, Apache MiNiFi, Apache NiFi, OpenCV, Apache Tika, and Python	1600 - 1640	Tim Spann
Experience Trimble's Live Blockchain-Based System powered by Apache Kafka and Apache NiFi	1740 - 1820	Timothy Leonard
Unlock Value from Big Data with Apache NiFi and Streaming CDC	1940 - 2020	Jordan Martz
Apache NiFi Crash Course	1100 - 1330	Andy LoPresto
Using Spark Streaming and NiFi for the next generation of ETL in the enterprise	1400 - 1440	Darryl Dutton
Acquisition of Seismic, Hydroacoustic, and Infrasonic Data with Apache NiFi and Apache Accumulo	1400 - 1440	Charles Houchin
Supercharge your NiFi environment for data lake management with open source Kylo	1450 - 1530	Matt Laudato
The Power of Intelligent Flows: Realtime IoT Botnet Classification with Apache NiFi	1400 - 1440	Andy LoPresto
Real-time Freight Visibility: How TMW Systems uses NiFi and SAM to create sub-second transportation visibility	1640 - 1720	Krishna Potluri
IoT, Streaming, and Dataflow Birds of a Feather	1740 - 1855	Aldrin Piri, Davor Bonaci, Venkatesh Ramanathan, Jeremy Dyer
SDLC with Apache NiFi	1130 - 1210	Kevin Doran

Wednesday

Thursday

Questions?

(Just kidding, hold them until after Saumitra)



Thank you

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