

Dataflow with Apache NiFi/MiNiFi

Andy LoPresto - [@yolopey](#)

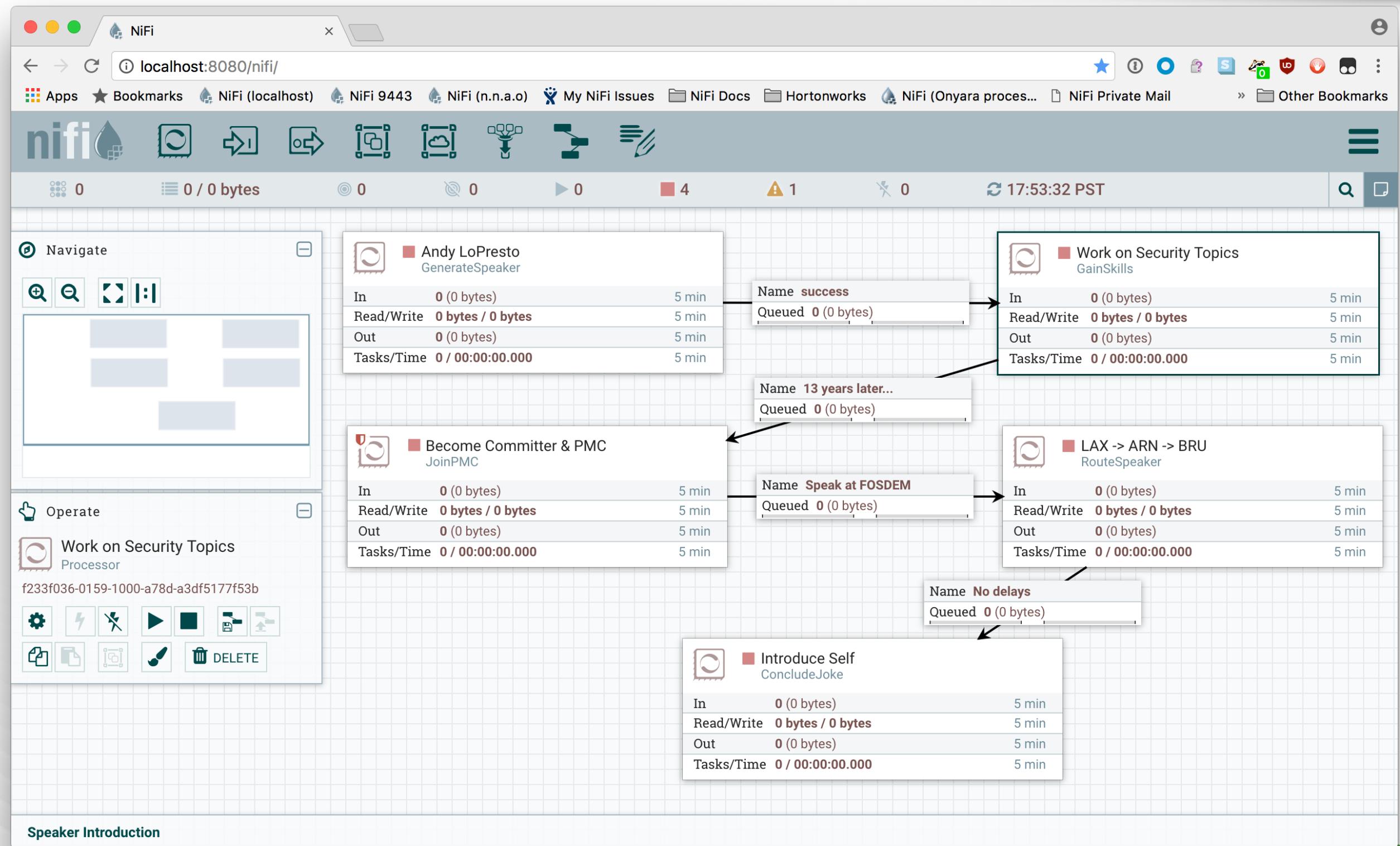
Intelligently Collecting Data at the Edge

Apache NiFi PMC

FOSDEM '17 - Brussels

04 Feb 2017





Agenda

What is dataflow and what are the challenges?
Apache NiFi
Apache MiNiFi
Architecture
Community

Agenda



What is dataflow and what are the challenges?

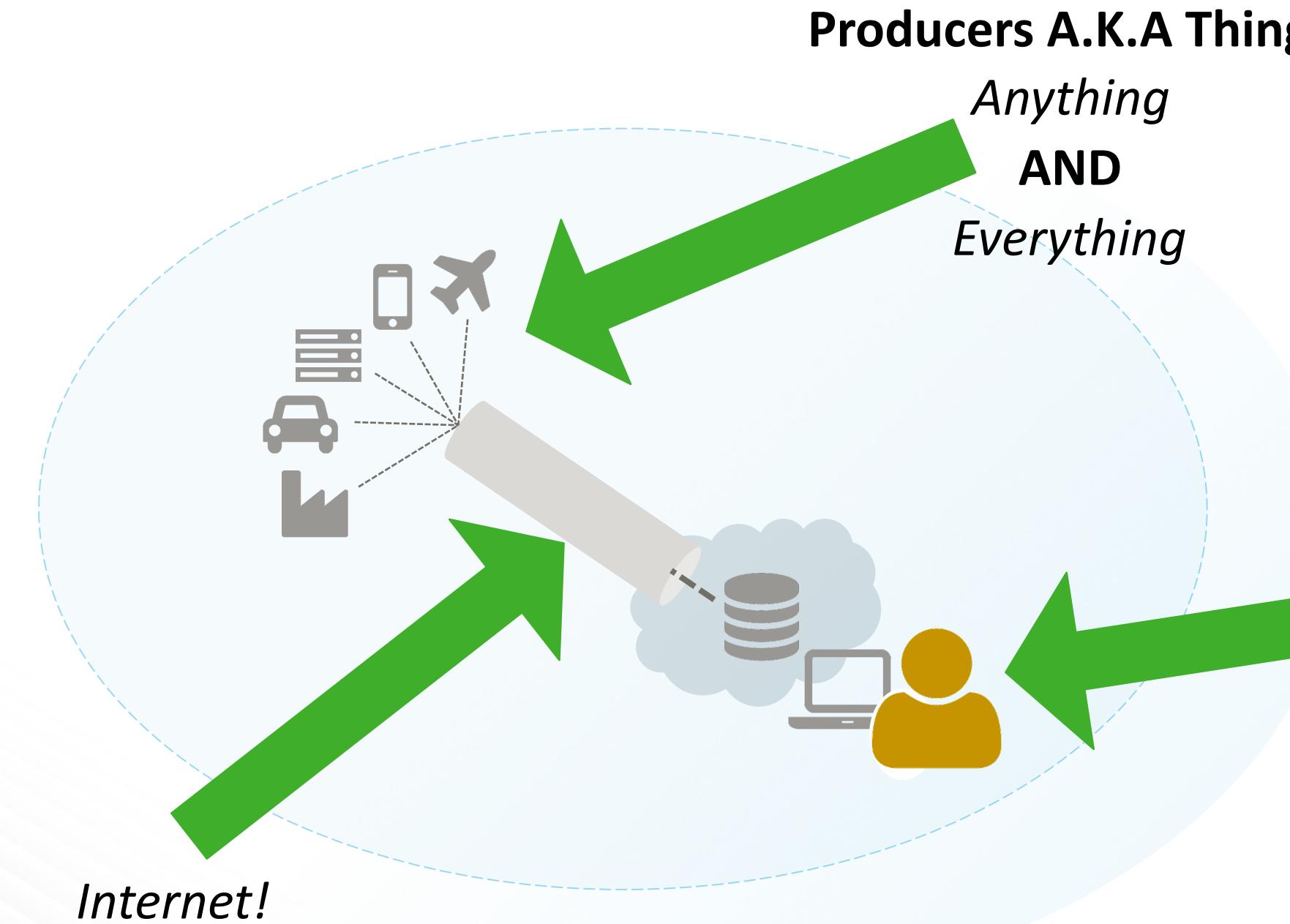
Apache NiFi

Apache MiNiFi

Architecture

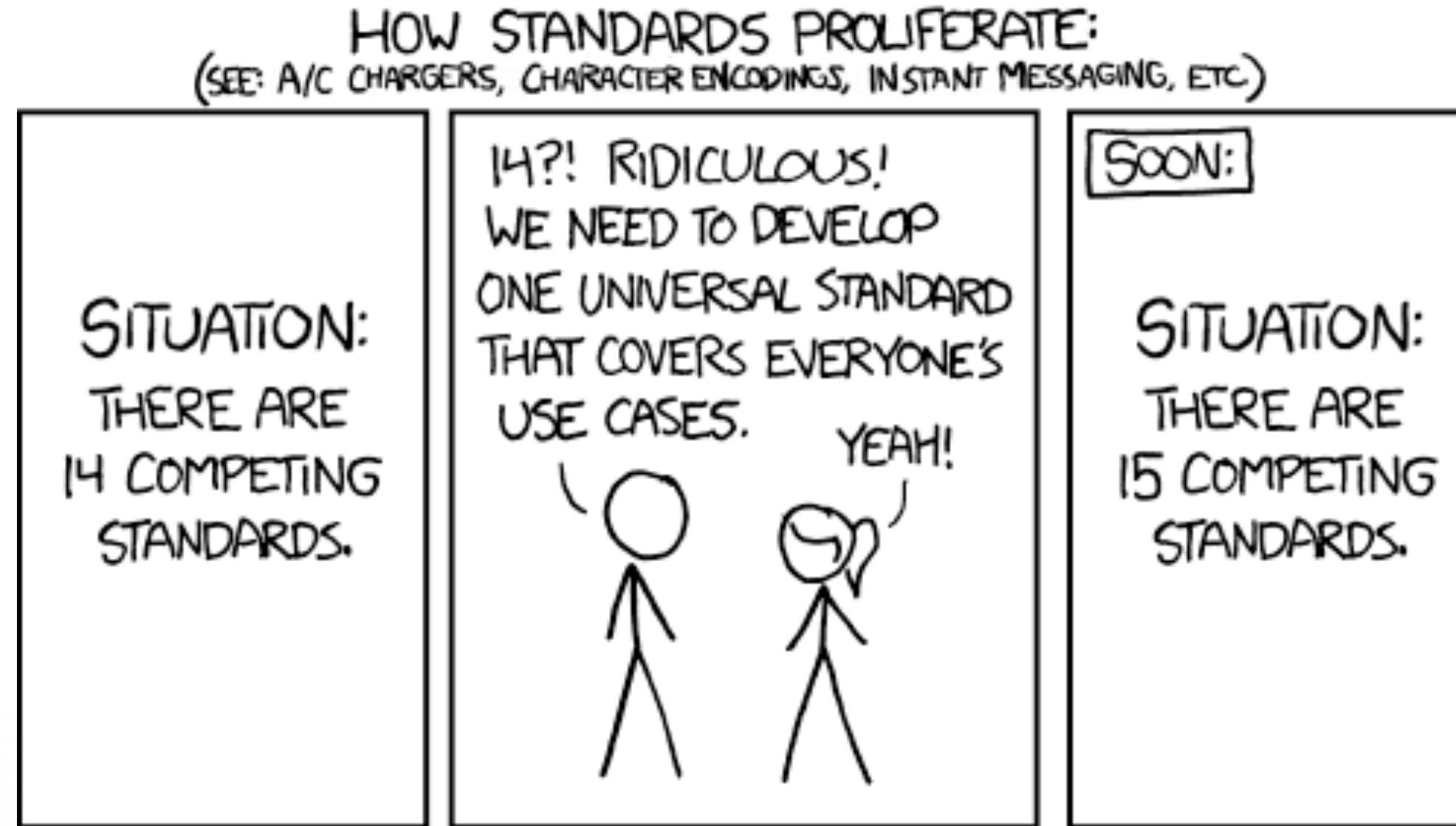
Community

Let's Connect A to B



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

Moving data *effectively* is hard



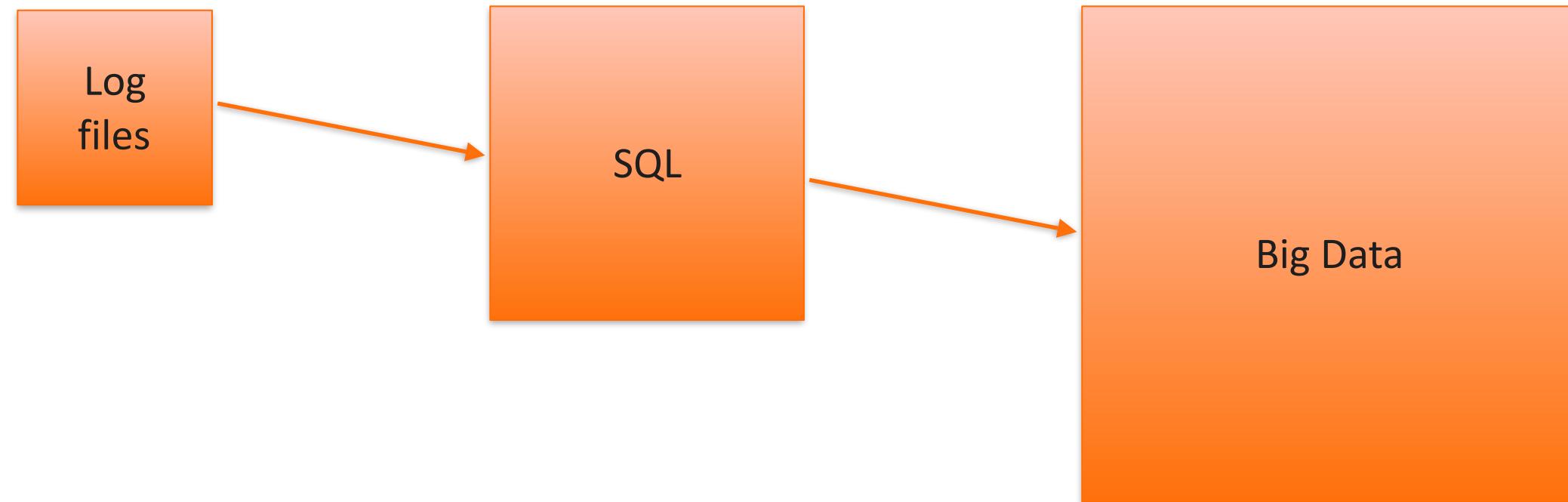
Standards: <http://xkcd.com/927/>

Why is moving data *effectively* hard?

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

Connecting A to B to C

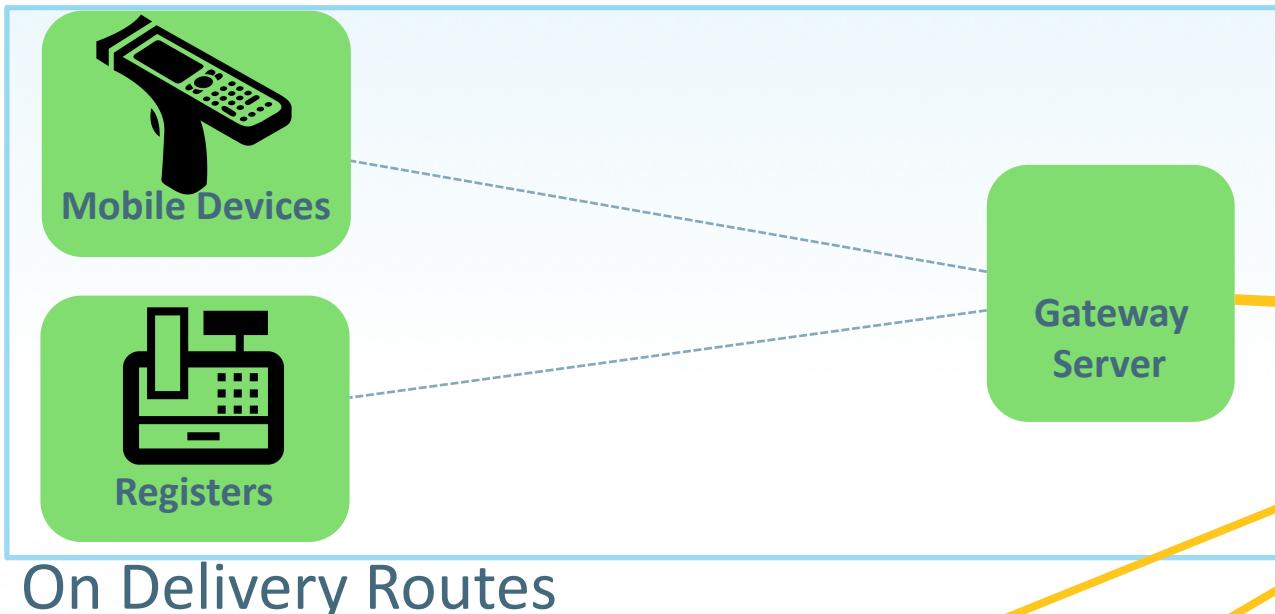
Easy enough with Bash scripts, Ruby/Python/Groovy, etc.



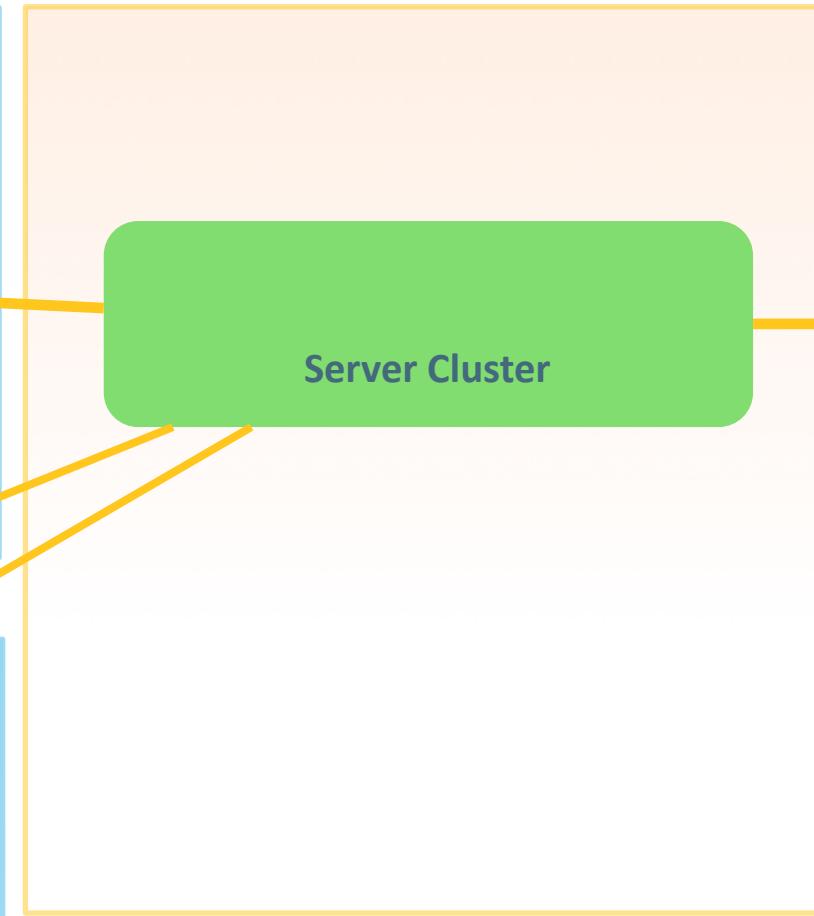
Let's Connect Lots of As to Bs to As to Cs to Bs to Δs to Cs to ϕs

Let's consider the needs of a courier service

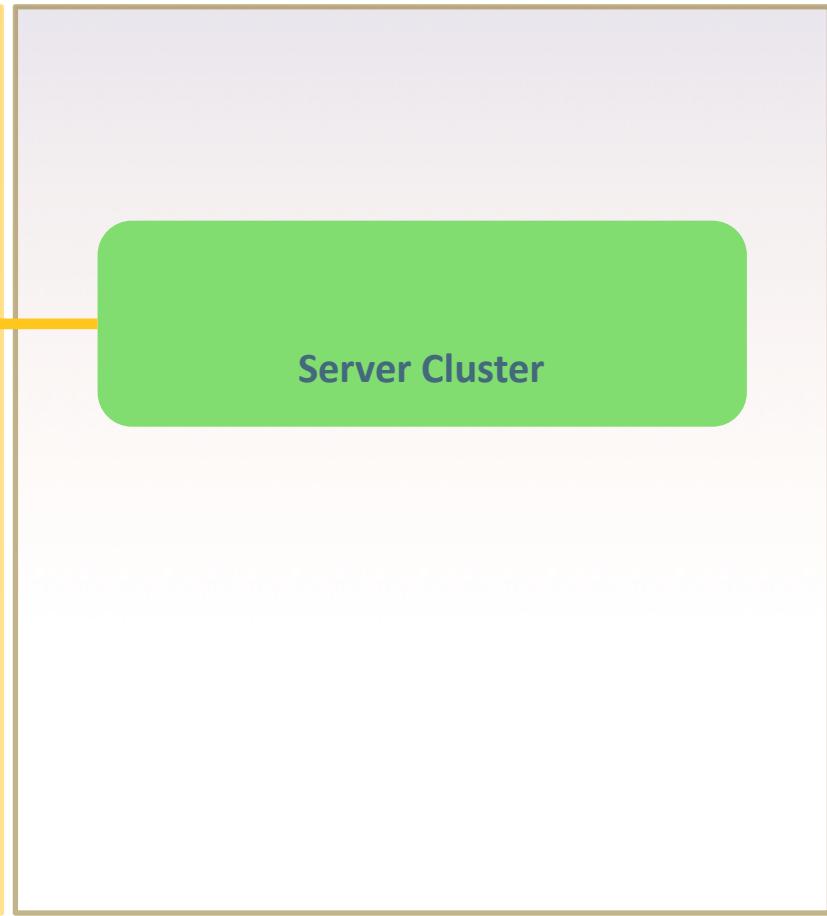
Physical Store



Distribution Center



Core Data Center at HQ



On Delivery Routes



Delivery Truck: Creative Stall, <https://thenounproject.com/creativestall/>

Deliverer: Rigo Peter, <https://thenounproject.com/rigo/>

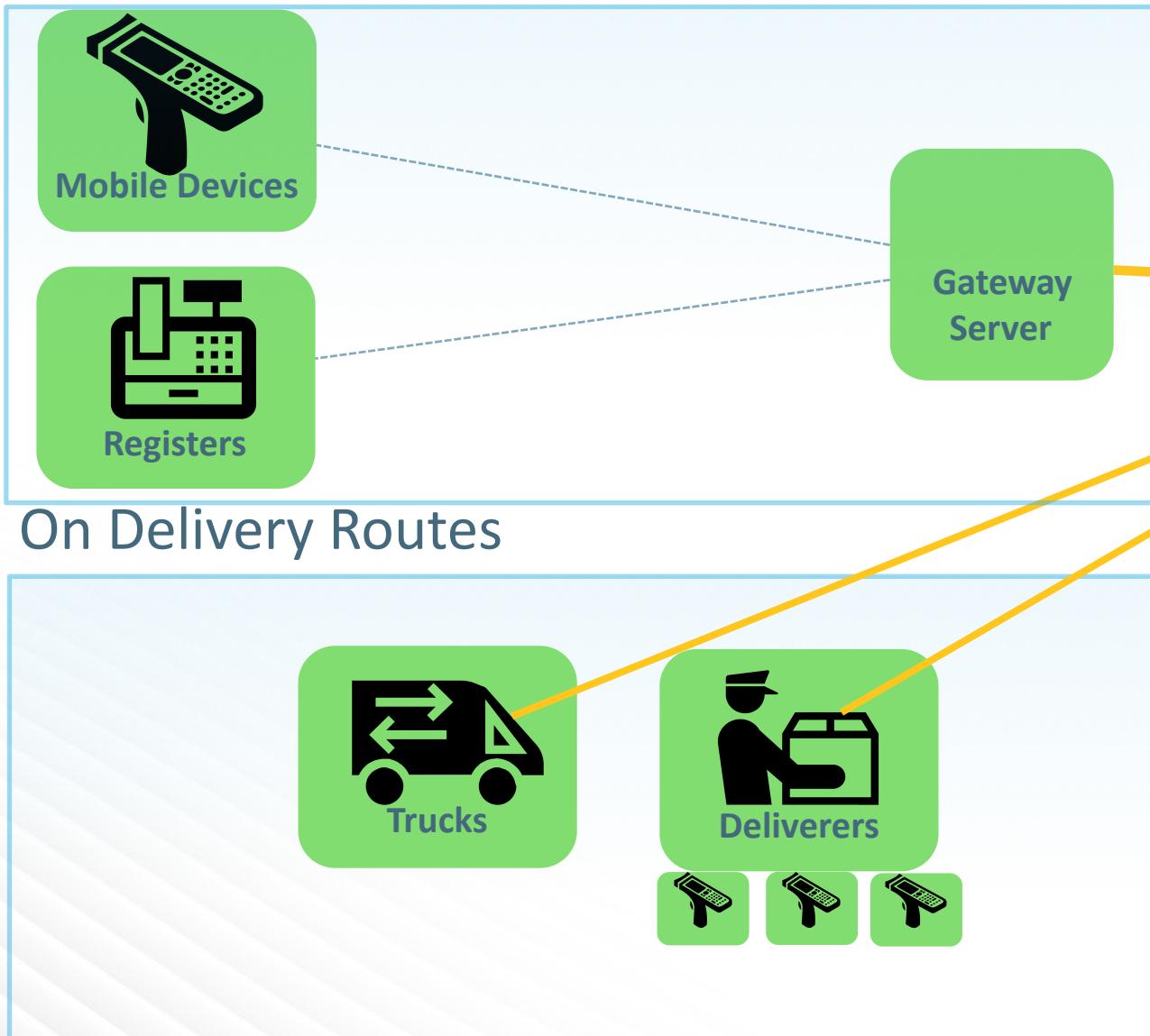
Cash Register: Sergey Patutin, <https://thenounproject.com/bdesign.by/>

Hand Scanner: Eric Pearson, <https://thenounproject.com/epearson001/>

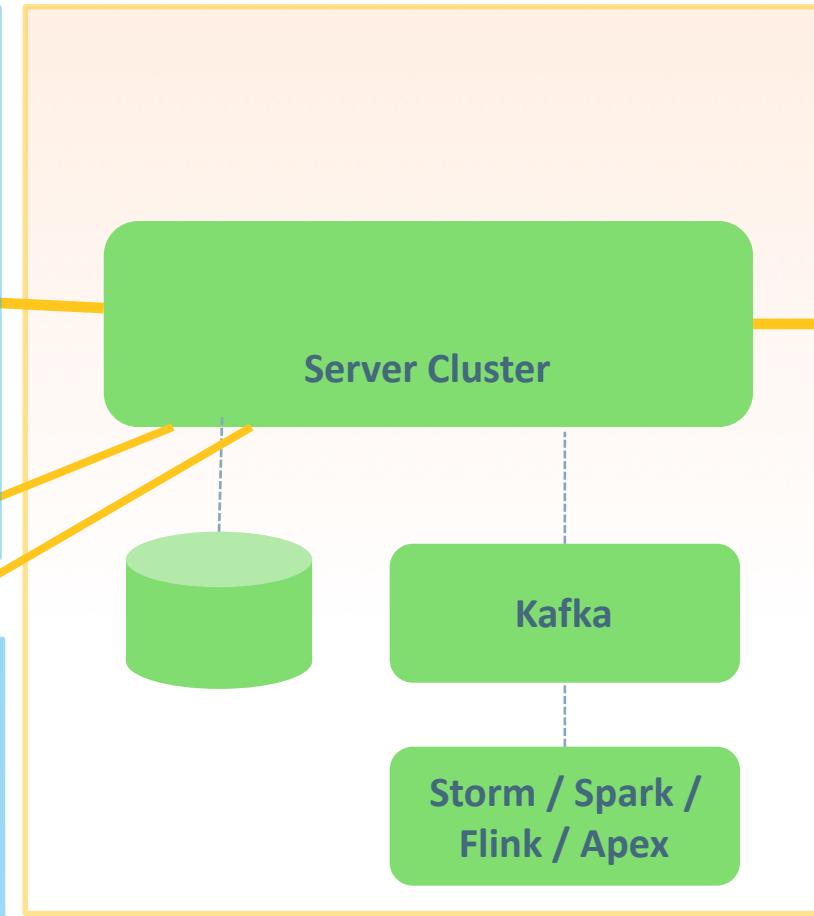
Great! I am collecting all this data! Let's use it!

Finding our needles in the haystack

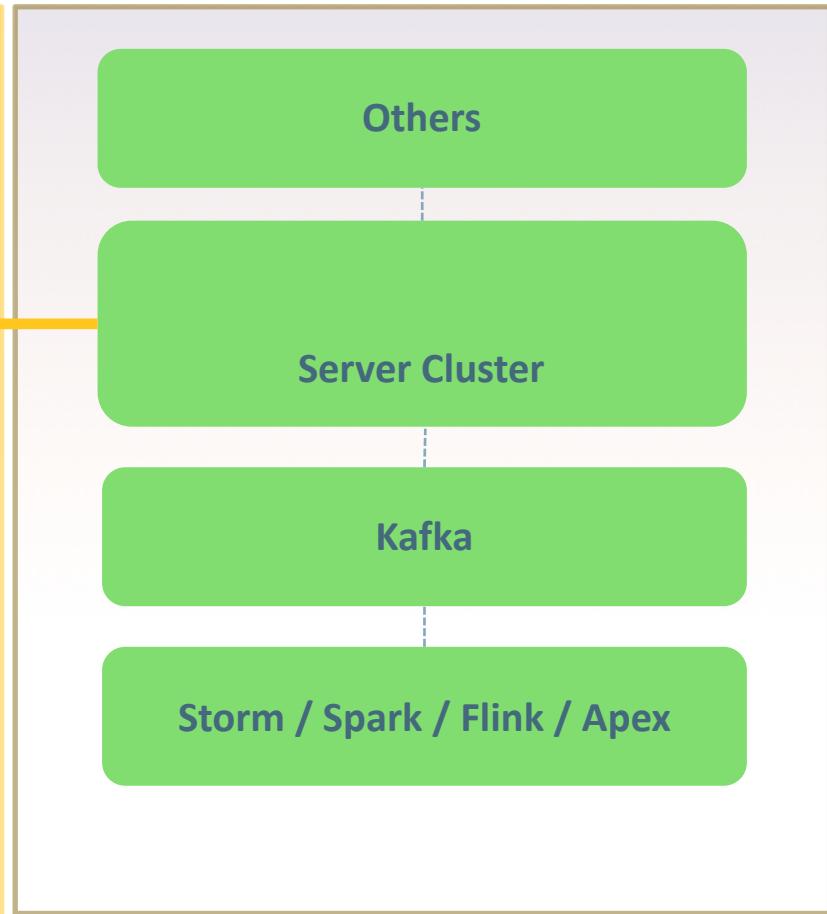
Physical Store



Distribution Center

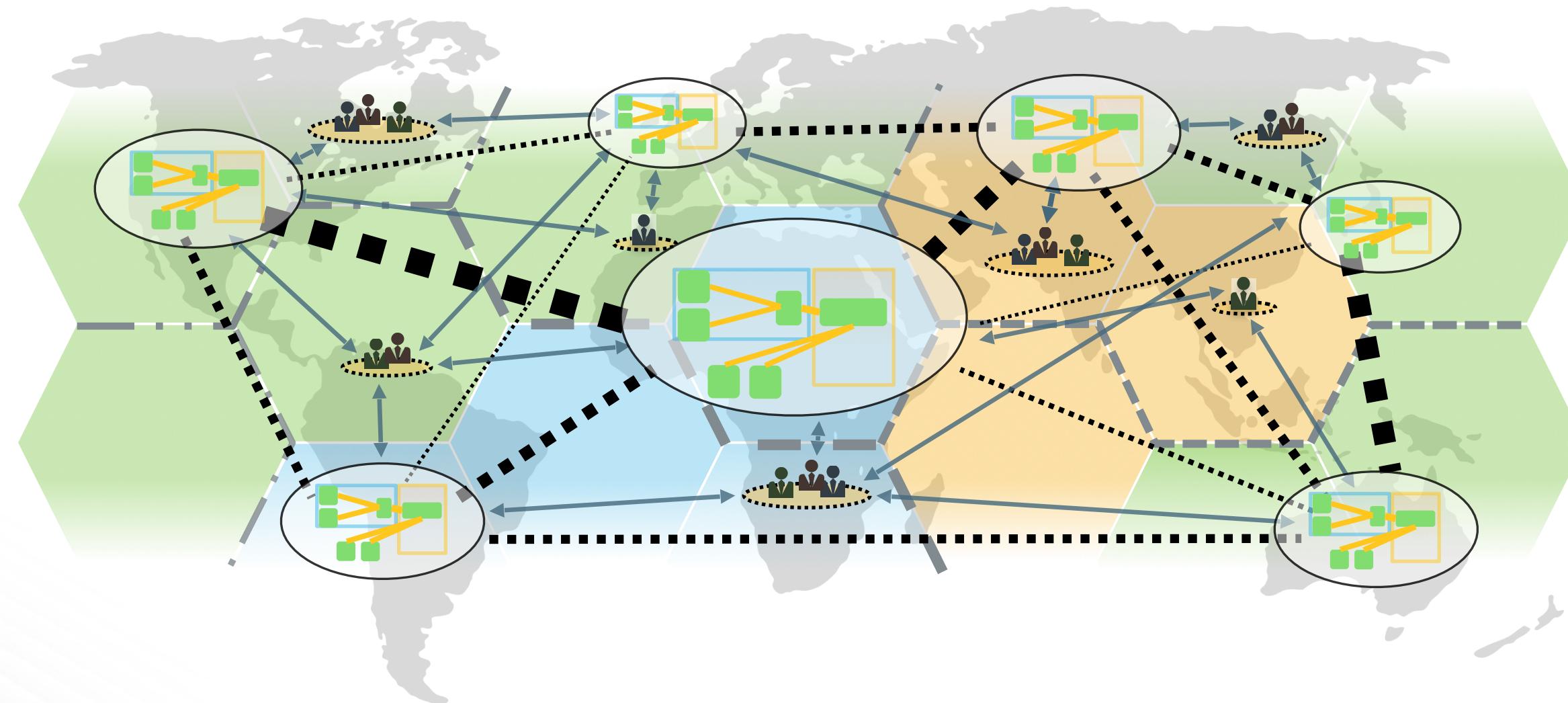


Core Data Center at HQ



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

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



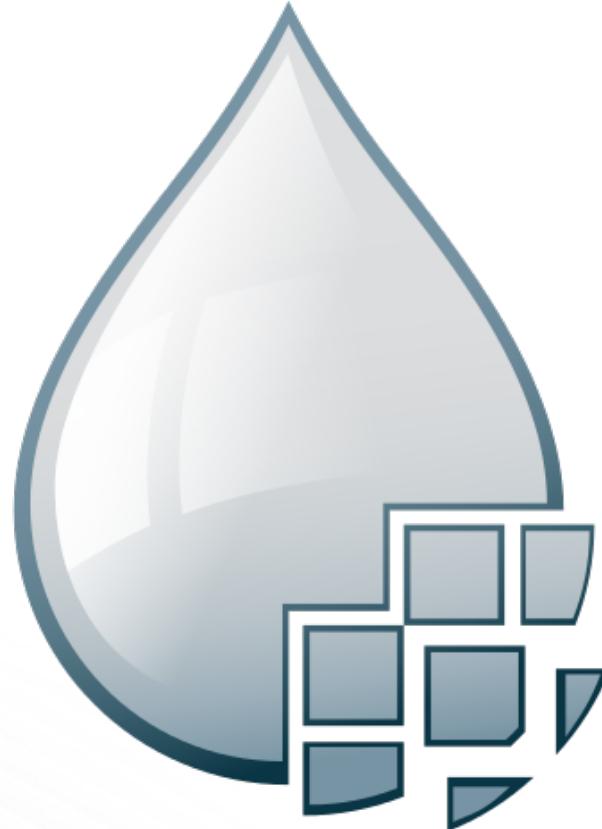
Agenda



- What is dataflow and what are the challenges?
- Apache NiFi
- Apache MiNiFi
- Architecture
- Community

Apache NiFi

Key Features



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

NiFi is based on Flow Based Programming (FBP)

FBP Term	NiFi Term	Description
Information Packet	FlowFile	Each object moving through the system.
Black Box	FlowFile Processor	Performs the work, doing some combination of data routing, transformation, or mediation between systems.
Bounded Buffer	Connection	The linkage between processors, acting as queues and allowing various processes to interact at differing rates.
Scheduler	Flow Controller	Maintains the knowledge of how processes are connected, and manages the threads and allocations thereof which all processes use.
Subnet	Process Group	A set of processes and their connections, which can receive and send data via ports. A process group allows creation of entirely new component simply by composition of its components.

FlowFiles & Data Agnosticism

- ◆ NiFi is data agnostic!
- ◆ But, NiFi was designed understanding that users can care about specifics and provides tooling to interact with specific formats, protocols, etc.

Robustness principle

“ Be conservative in what you do,
be liberal in what you accept from others

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS **THE** CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13

20130227 2013.02.27 27.02.13 27-02-13

27.2.13 2013. II. 27. 27½-13 2013.158904109

MMXIII-II-XXVII MMXIII ^{LVII}/_{CCCLXV} 1330300800

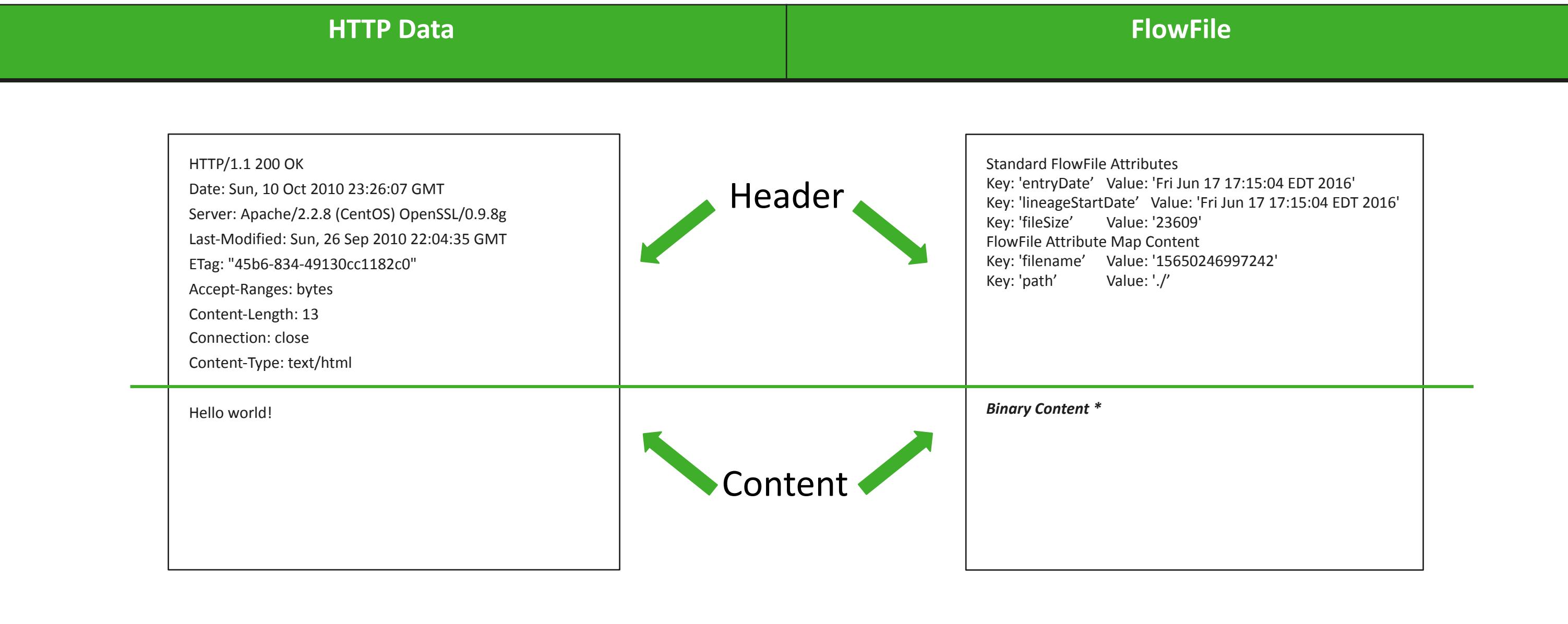
((3+3)×(111+1)-1)×3/3-1/3³ 2013 ²⁰¹³ HISSSS

10/11011/1101 02/27/20/13 ^{2 3 1 4} 0 1 2 3 7 ^{2 2 3 5} 8



ISO 8601 - <http://xkcd.com/1179/>

FlowFiles are like HTTP data



User Interface

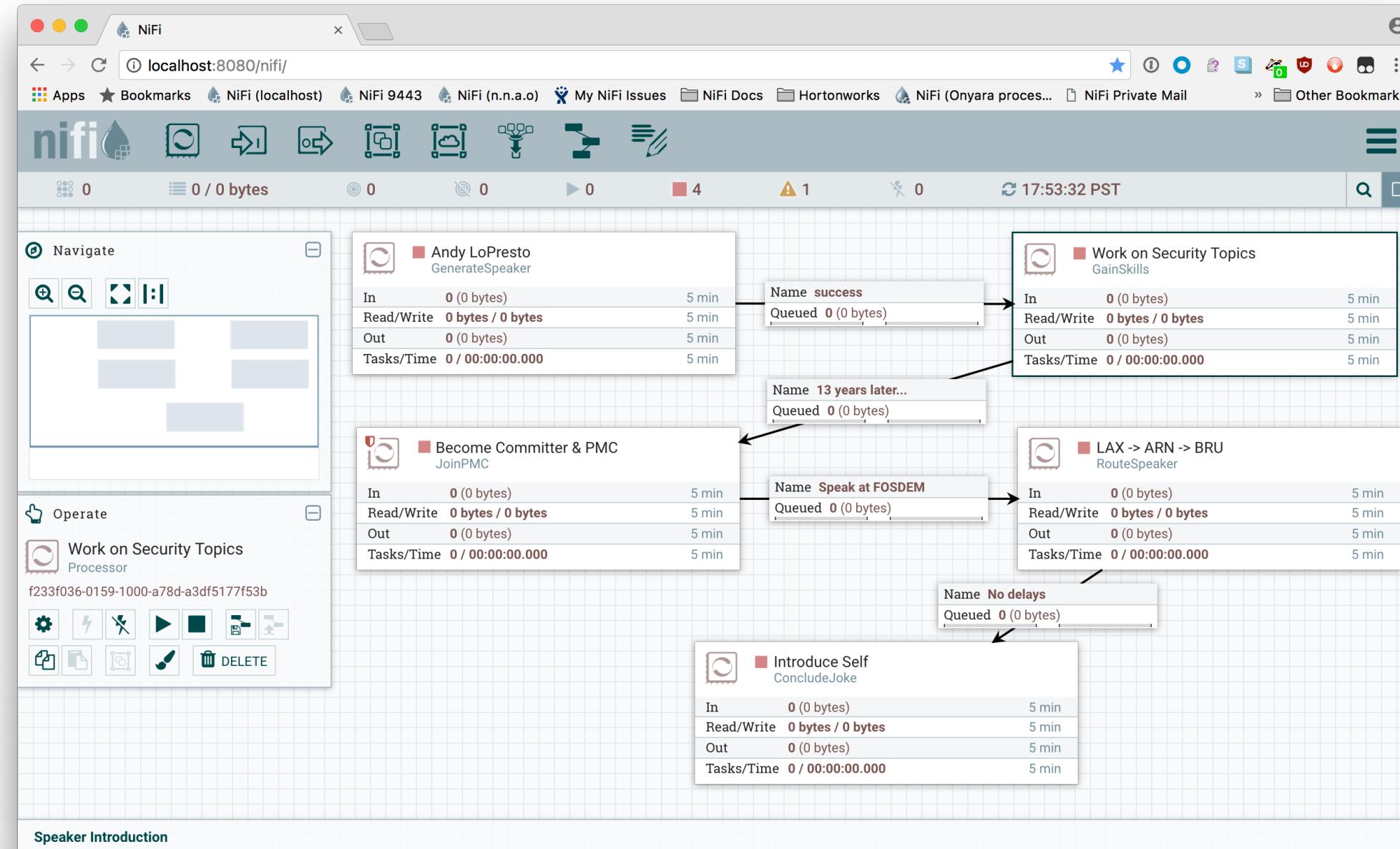
Less of this...

```
scratch/release_verification (master) alopresto hw12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.1.0-bin/nifi-1.1.0/conf/users.xml nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.1.0-bin/nifi-1.1.0/conf/total 144 drwxr-xr-x 16 alopresto staff 544B Nov 24 19:20 ./ drwxr-xr-x 17 alopresto staff 578B Nov 24 14:25 .. drwxr-xr-x 38 alopresto staff 1.3K Nov 24 14:33 archive/ drwxr-xr-x 1 alopresto staff 2.5K Nov 24 19:20 authorizations.xml drwxr-xr-x 1 alopresto staff 3.7K Nov 24 19:19 authorizers.xml drwxr-xr-x 1 alopresto staff 2.1K Nov 23 23:36 bootstrap-notification-services.xml drwxr-xr-x 1 alopresto staff 3.1K Nov 24 19:29 bootstrap.conf drwxr-xr-x 1 alopresto staff 2.5K Nov 24 14:33 flow.xml.gz drwxr----- 1 alopresto staff 3.0K Nov 24 19:16 keystore.jks drwxr-xr-x 1 alopresto staff 8.0K Nov 23 23:36 logback.xml drwxr-xr-x 1 alopresto staff 6.2K Nov 23 23:36 login-identity-providers.xml drwxr-xr-x 1 alopresto staff 9.0K Nov 24 19:17 nifi.properties drwxr-xr-x 1 alopresto staff 3.6K Nov 23 23:36 state-management.xml drwxr-xr-x 1 alopresto staff 911B Nov 24 19:16 truststore.jks drwxr-xr-x 1 alopresto staff 226B Nov 24 19:20 users.xml drwxr-xr-x 1 alopresto staff 1.4K Nov 23 23:36 zookeeper.properties h12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto nifi-1.1.0-RC1-failed/nifi-1.1.0/nifi-assembly/target/nifi-1.1.0-bin/nifi-1.1.0/conf/authorizations.xml <xml version="1.0" encoding="UTF-8" standalone="yes"?> <authorizations> <policies> <policy identifier="b2c0320b-0384-38d2-8ff7-58a26dd63897" resource="/flow" action="R"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="7e67308c-a837-3ba1-8b37-4a40ff8d43fe" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="b762618f-3636-3b45-943e-6d38605e276e" resource="/data/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="W"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="61df3e5b-837f-37f4-b636-b2e7bc4989c3" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="R"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="ff9f93ce-1fd4-3b08-9309-088ba6abc5a9" resource="/process-groups/9871f1da-0158-1000-c0b2-42aaca57d800" action="W"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="bc5512df-2a78-33ba-959e-3bc1ba5f781d" resource="/restricted-components" action="W"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="70197f52-afda-3938-b634-4986bf26032fa" 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-fca4-3813-baa8-f87949853e7e" resource="/policies" action="R"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="3906378a-8fb9-3a71-aa0b-44325bf1723cf" resource="/policies" action="W"> <user identifier="9860132a-283e-370b-ba63-b023393be562"/> </policy> <policy identifier="26c66358-8548-31cb-a38c-e0ebb2c3ddb4" 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> h12203:/Users/alopresto/Workspace/scratch/release_verification (master) alopresto & 9s @ 17:26:53 $ 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 h12203:...space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto & 157s @ 17:40:39 $ nfkl 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 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. h12203:...space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto & 9s @ 17:40:49 $ sub conf/nifi.properties h12203:...space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto & 5s @ 17:40:55 $ fgno 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: /Users/alopresto/.jenv/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 /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/log4j-over-slf4j-1.7.12.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/logback-classic-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/logback-core-1.1.3.jar:/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/lib/nifi-api-1.2.0-SNAPSHOT.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-utils-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/sl4j-api-1.7.12.jar:-Dorg.apache.jasper.compiler.disablejsr199=true -Xms512m -Xms512m -Dsun.net.http.allowRestrictedHeaders=true -Djava.net.preferIPv4Stack=true -Djava.awt.headless=true -Xmx1G -Djava.protocol.handler.pkgs=sun.net.www.protocol -Dnifi.properties.file.path=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT:/conf/nifi.properties -Dnifi.bootstrap.listen.port=59772 -Dapp=Nifi -Dorg.apache.nifi.bootstrap.config.log.dir=/Users/alopresto/Workspace/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT/logs org.apache.nifi.NiFi h12203:...space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto & 21s @ 17:41:17 $ nfkl 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. h12203:...space/nifi/nifi-assembly/target/nifi-1.2.0-SNAPSHOT-bin/nifi-1.2.0-SNAPSHOT (NIFI-3313) alopresto & 11595s @ 17:54:33 $
```

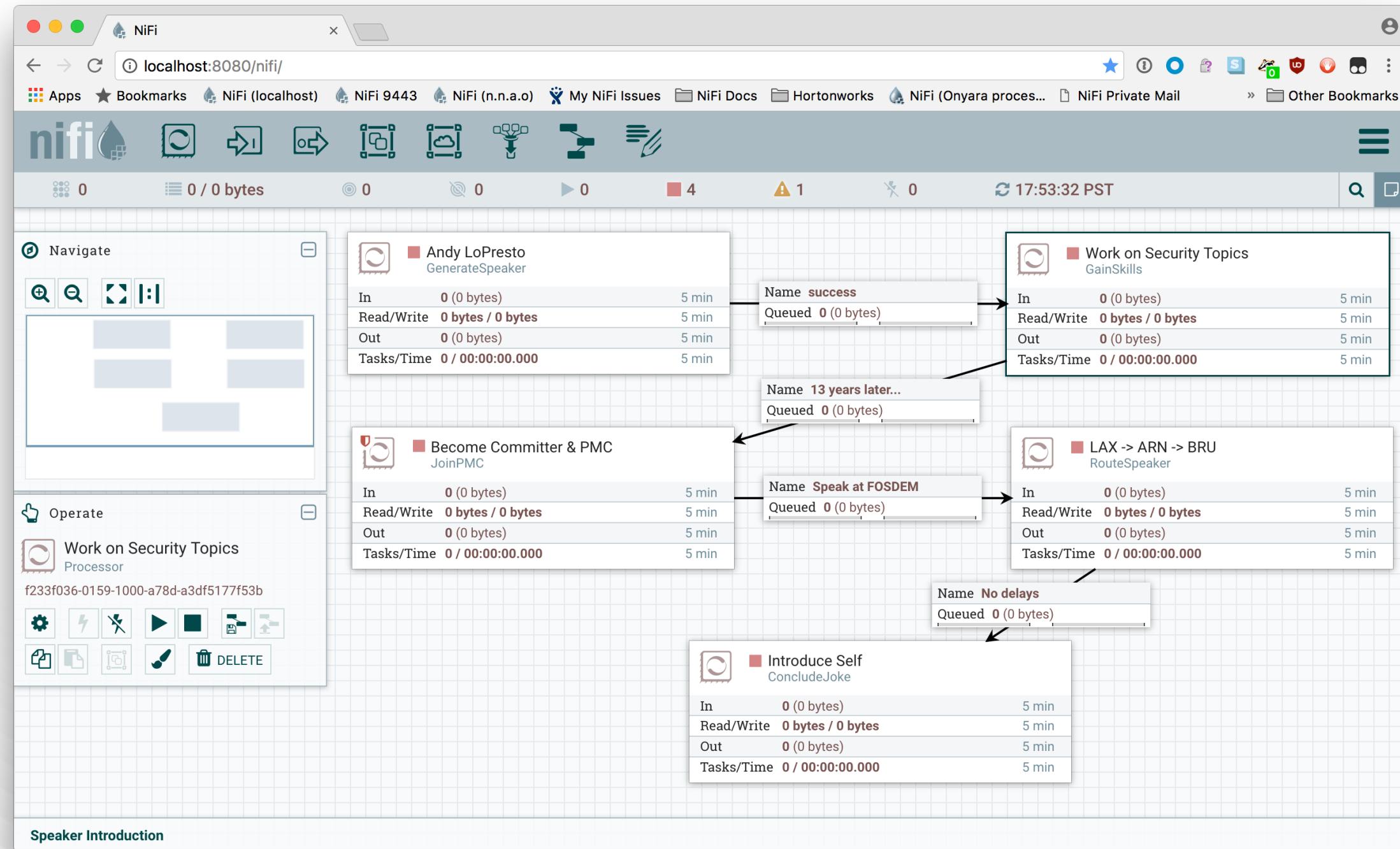


User Interface

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



User Interface



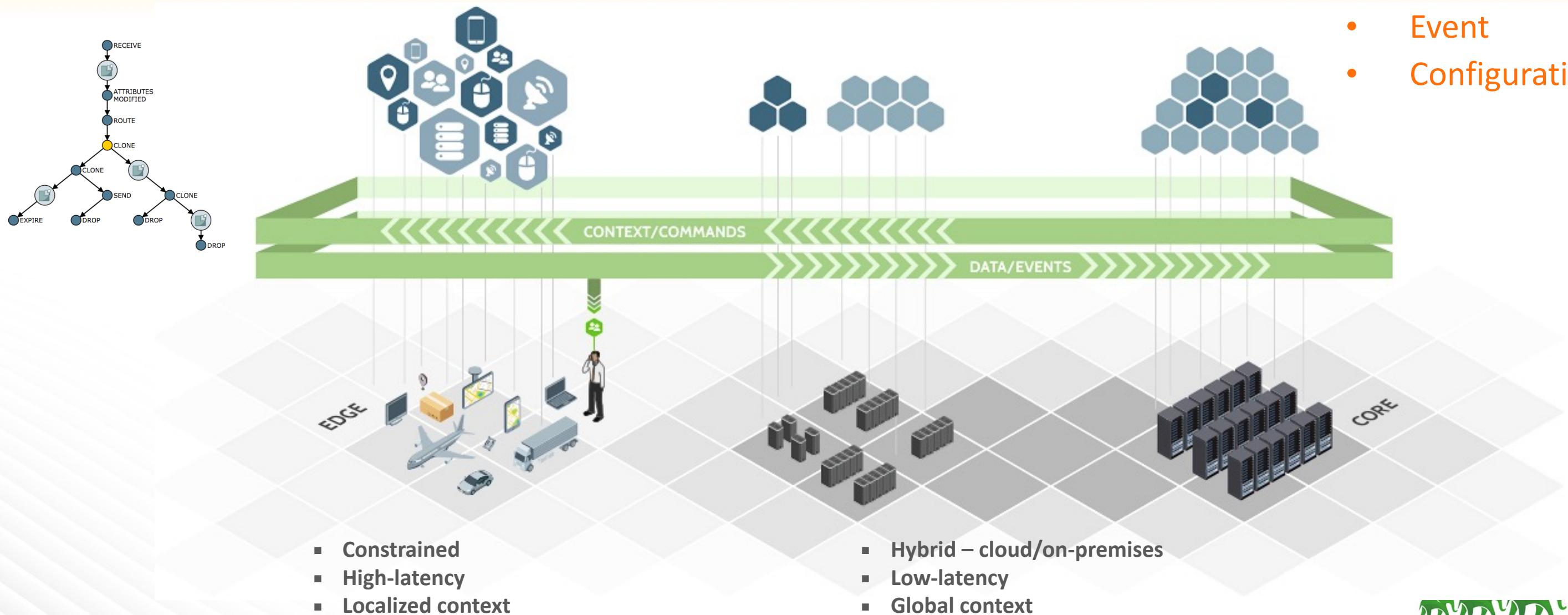
Data Provenance

Origin – attribution
Replay – recovery

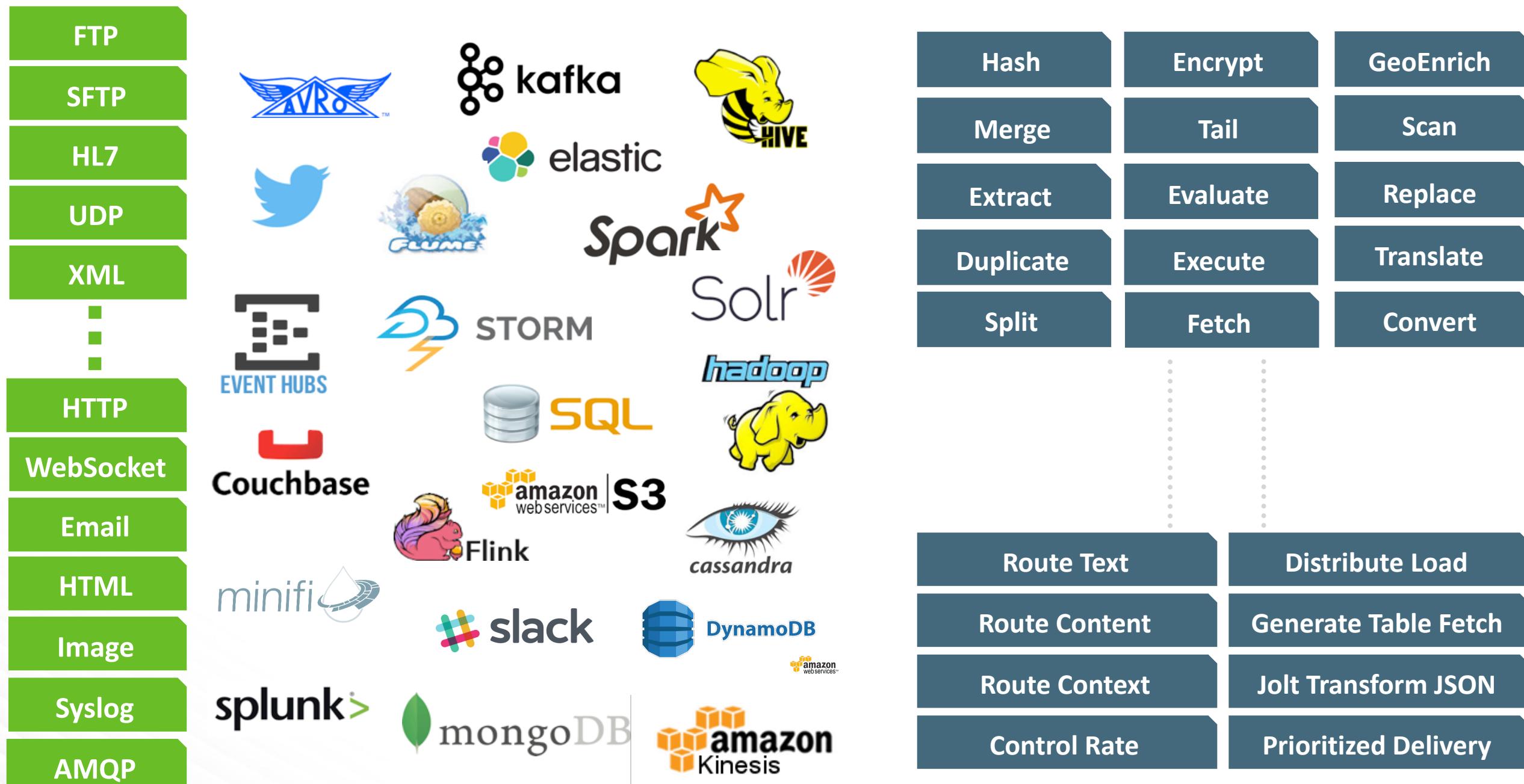
Evolution of topologies
Long retention

Types of Lineage

- Event
- Configuration



Deeper Ecosystem Integration: 180+ Processors



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

Agenda



- What is dataflow and what are the challenges?
- Apache NiFi
- Apache MiNiFi
- Architecture
- Community

Apache NiFi Subproject: MiNiFi

- ◆ Let me 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
 - ◆ S2S client libs



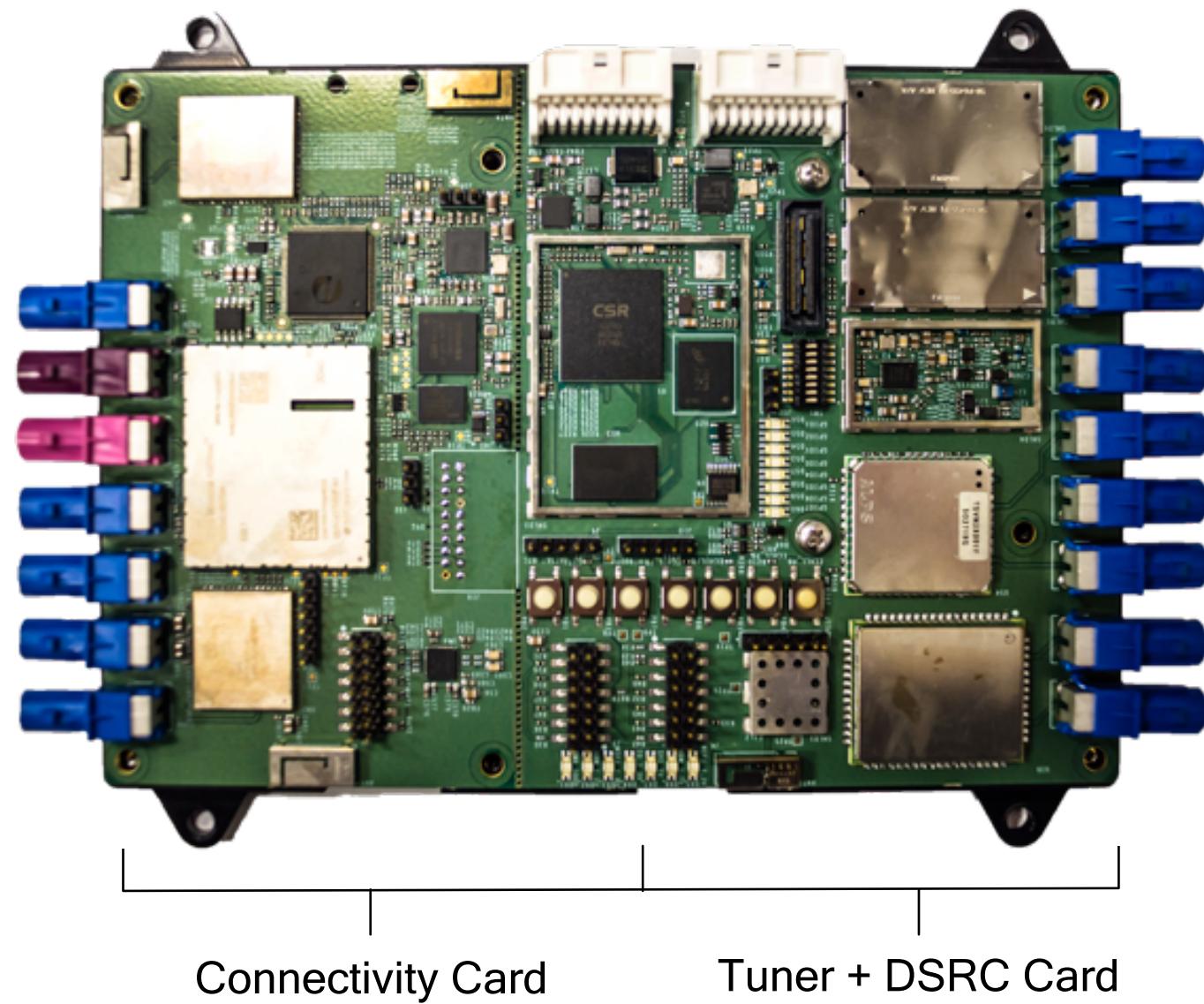
Why build MiNiFi?

- ◆ NiFi is big
 - ◆ 1.1.0 release is 726 MB compressed
 - ◆ Can be modified to run in restricted environments, but requires manual surgery
 - ◆ Provides UI, provenance query, etc.
 - ◆ Runs on dedicated machines/clusters — “owns the box”
- ◆ MiNiFi lives at the edge
 - ◆ No UI
 - ◆ 0.1.0 Java binary is 45 MB, C++ binary is 746 KB
 - ◆ “Good guest”

What does MiNiFi provide?

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

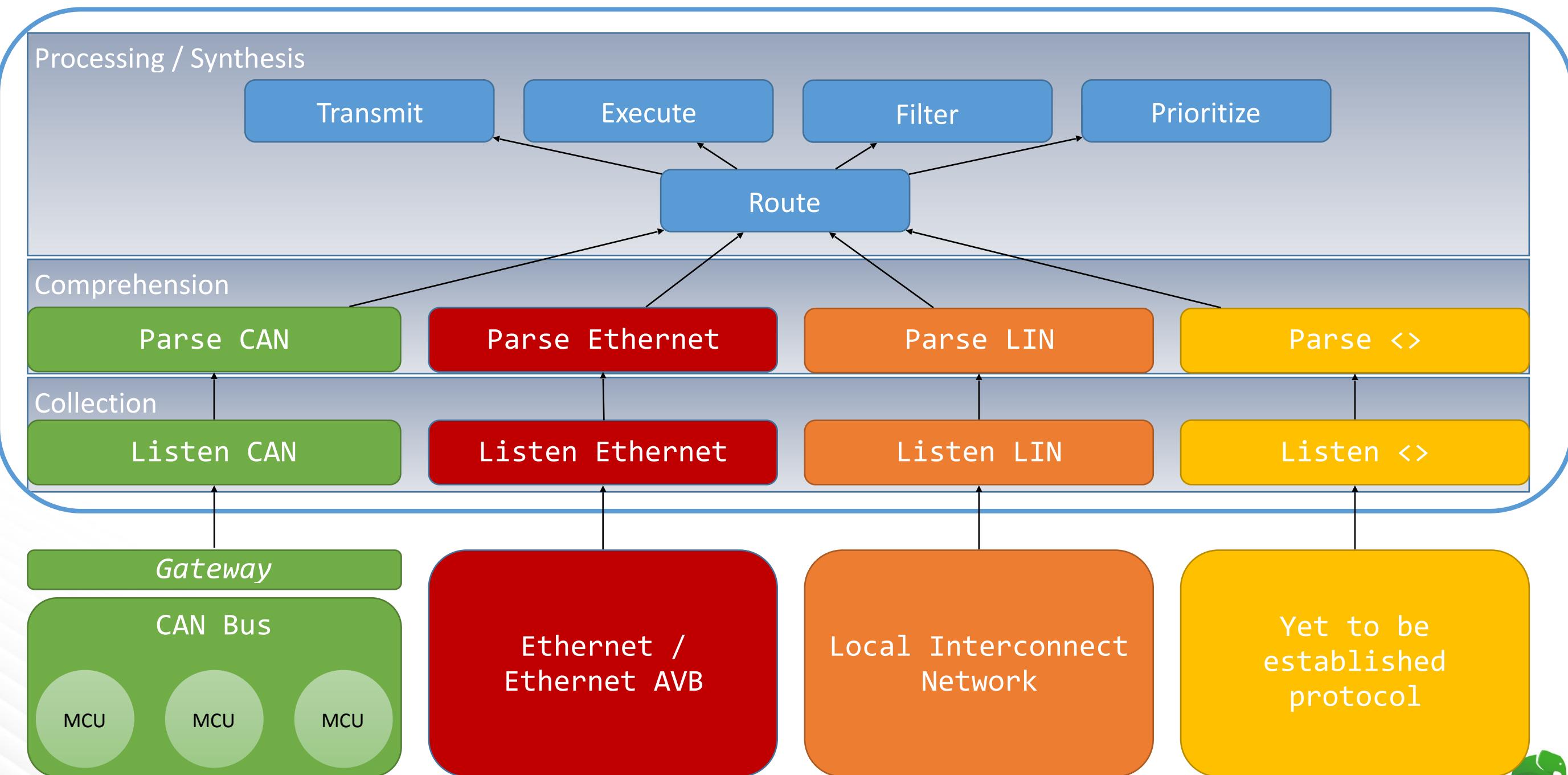
Connected Car Reference Platform Box



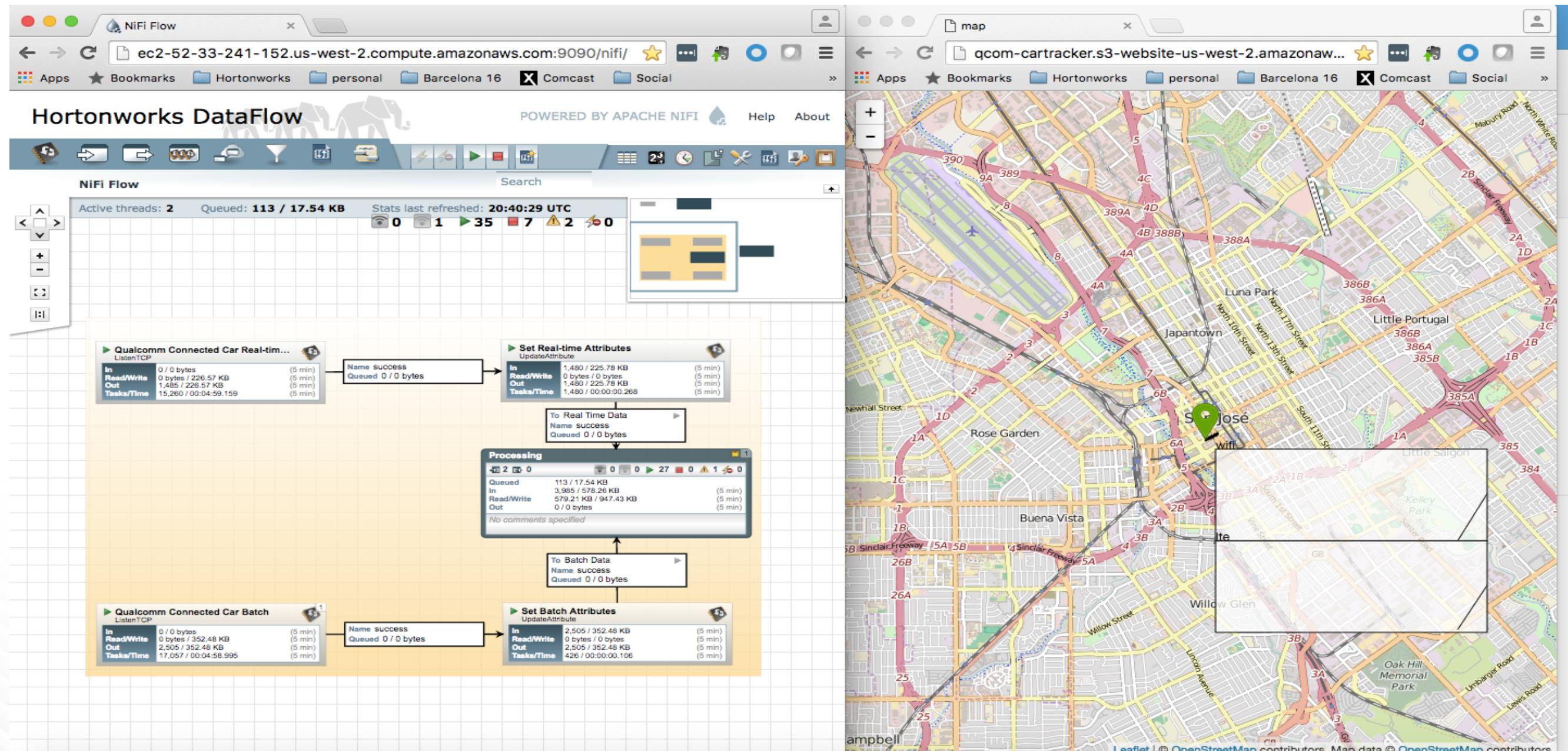
Connectivity Card

Tuner + DSRC Card

MiNiFi on a Connected Car



MiNiFi on a Connected Car



MiNiFi Feature Proposals

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

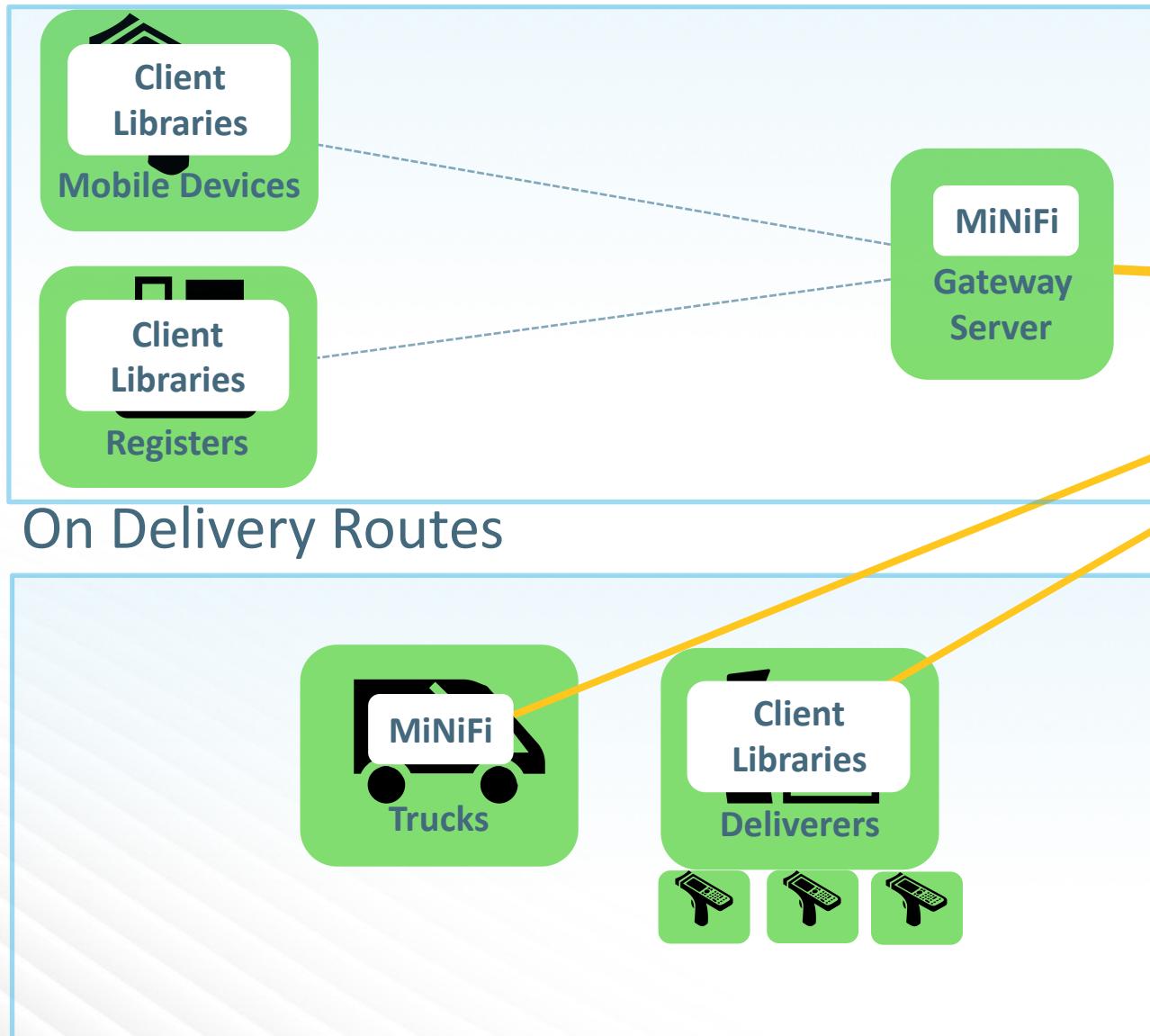
Agenda



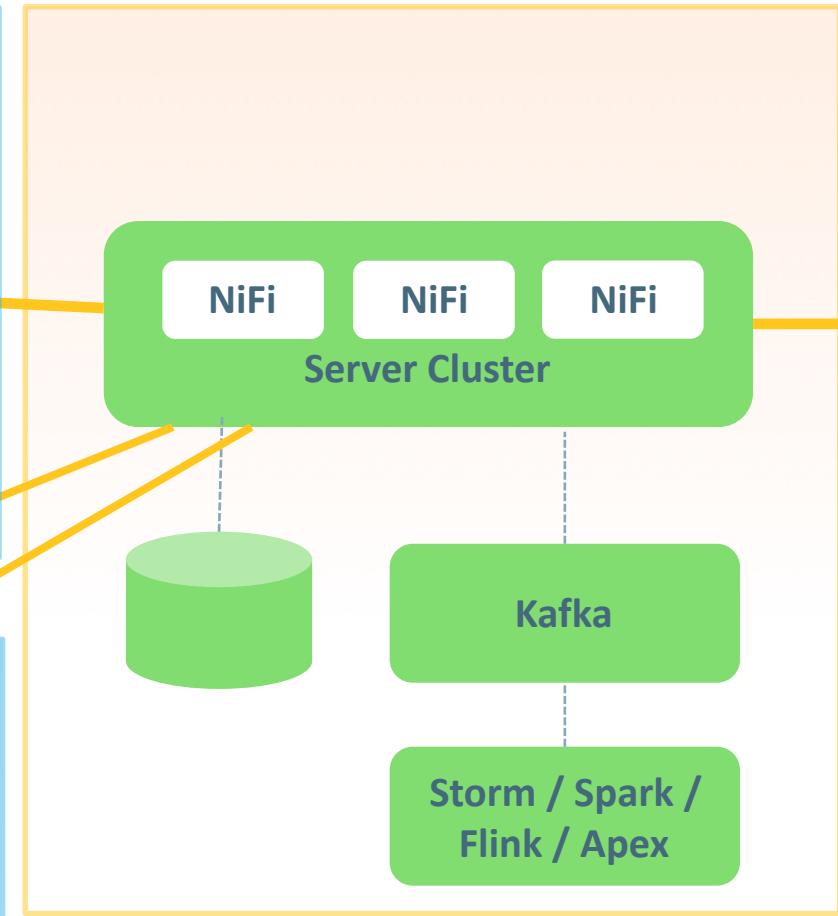
What is dataflow and what are the challenges?
Apache NiFi
Apache MiNiFi
Architecture
Community

Let's revisit our courier service from the perspective of NiFi

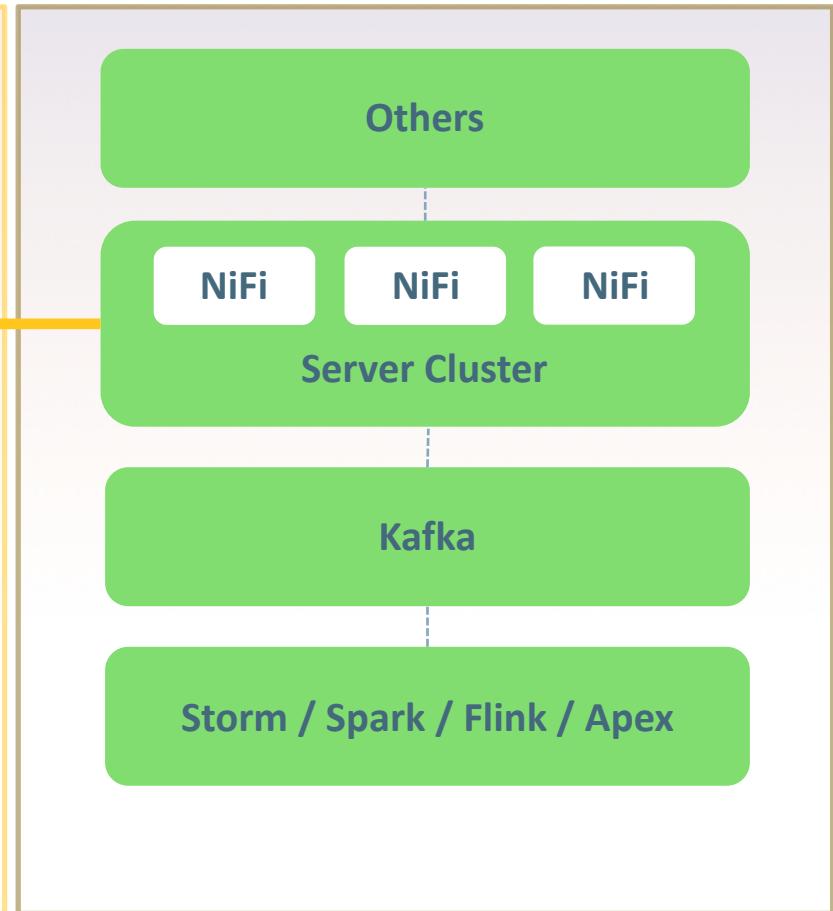
Physical Store



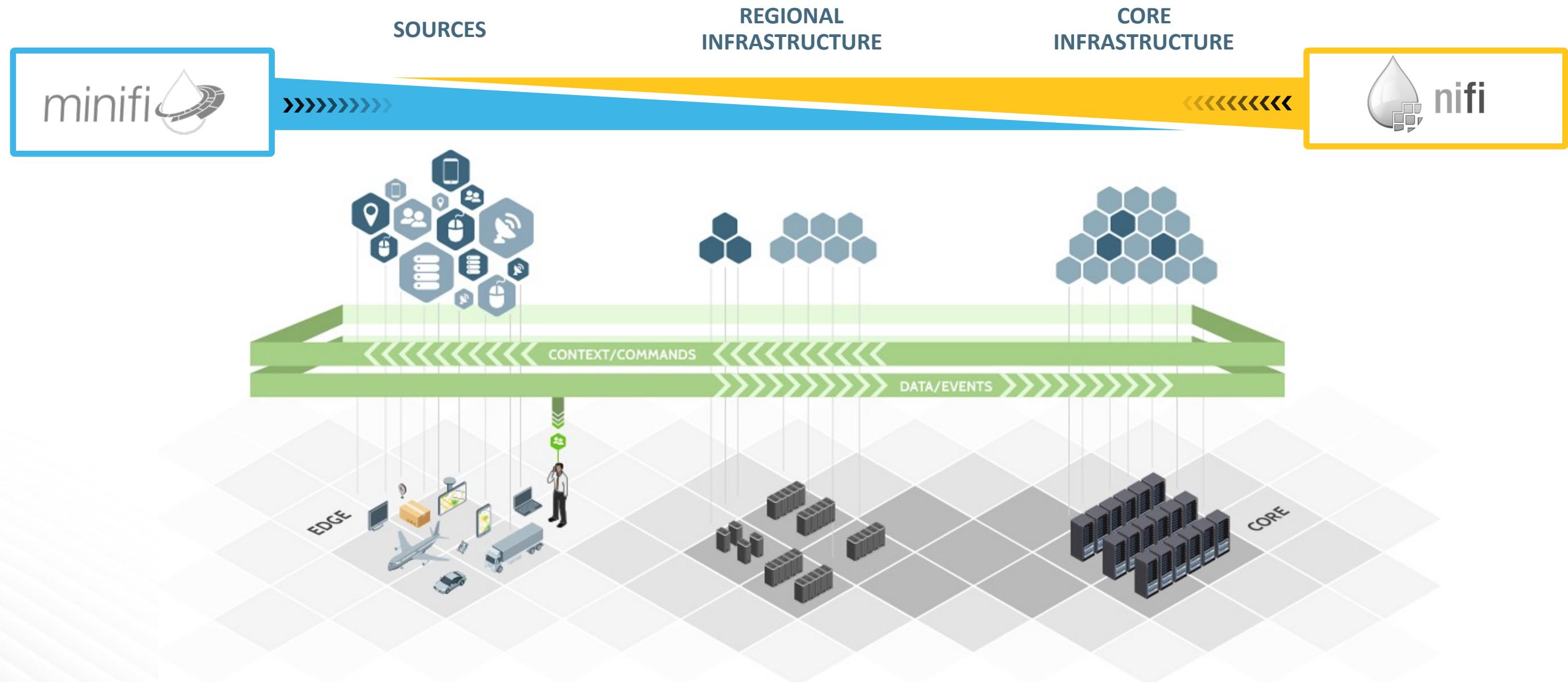
Distribution Center



Core Data Center at HQ



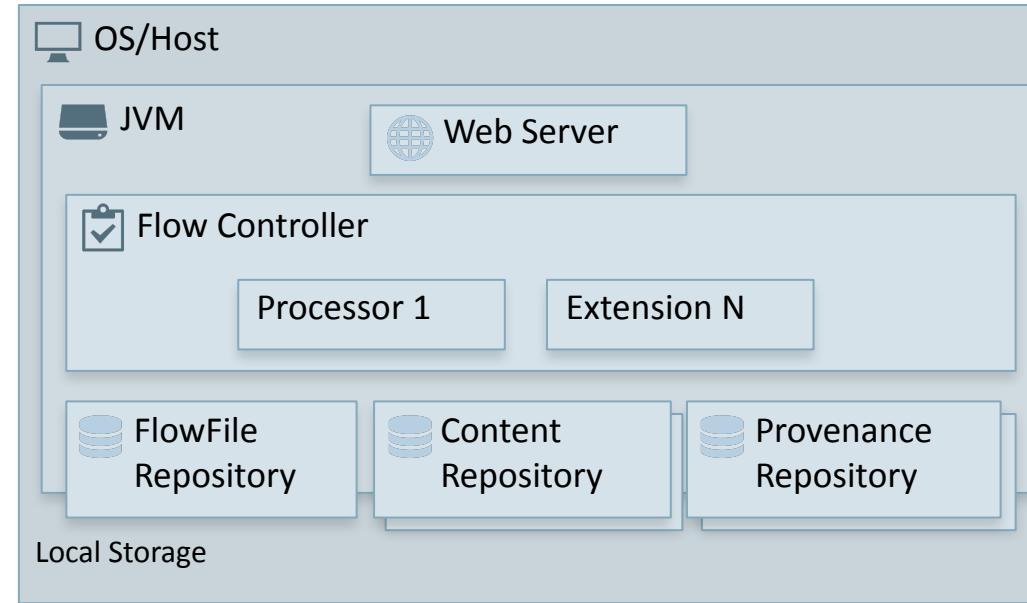
Apache NiFi Managed Dataflow



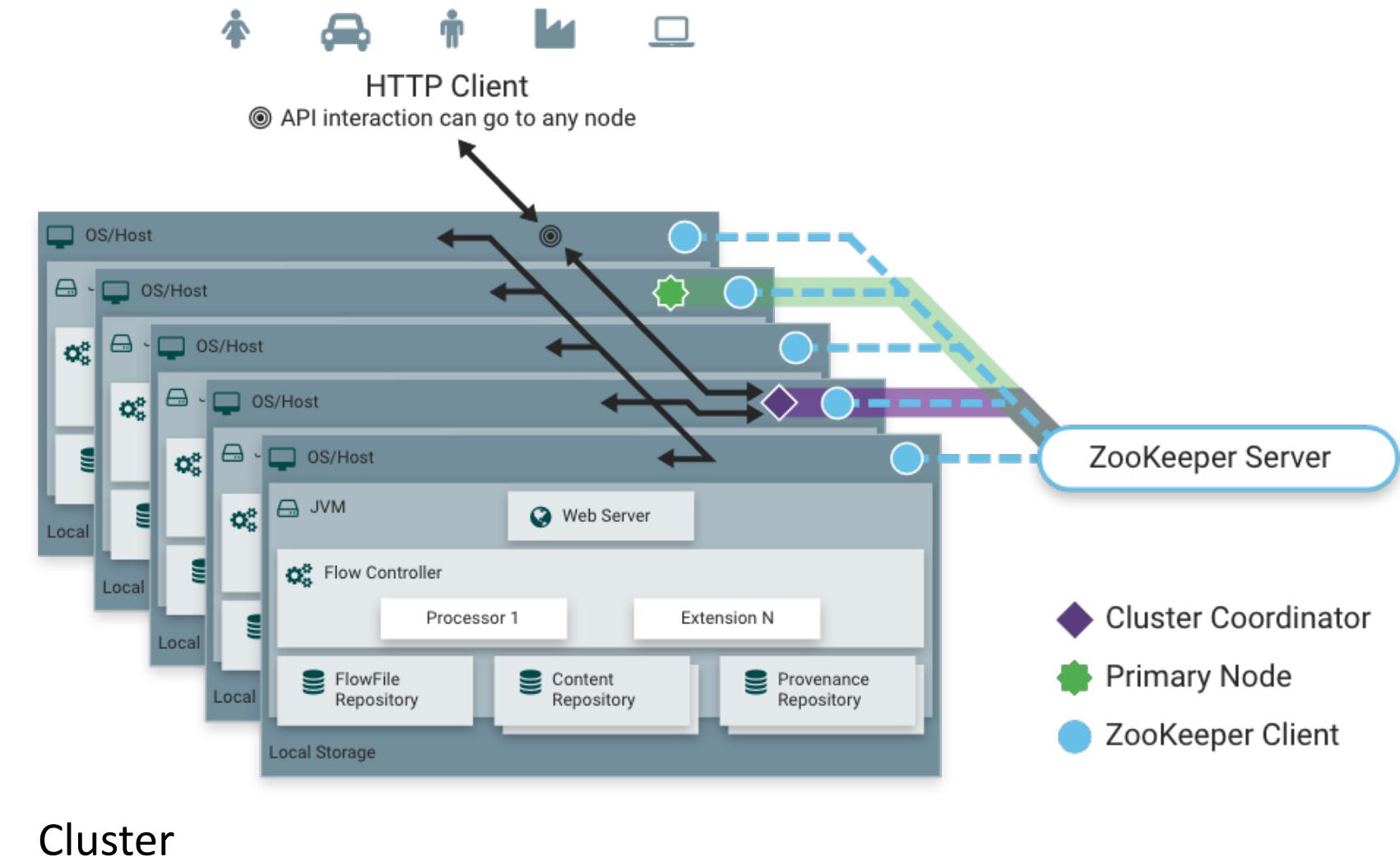
Extension / Integration Points

NiFi Term	Description
Flow File Processor	Push/Pull behavior. Custom UI
Reporting Task	Used to push data from NiFi to some external service (metrics, provenance, etc.)
Controller Service	Used to enable reusable components / shared services throughout the flow
REST API	Allows clients to connect to pull information, change behavior, etc.

Architecture



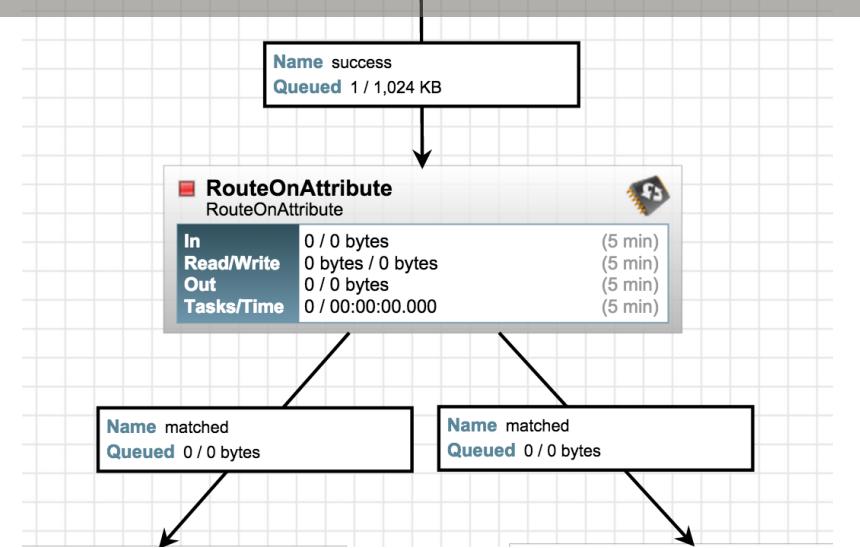
Standalone



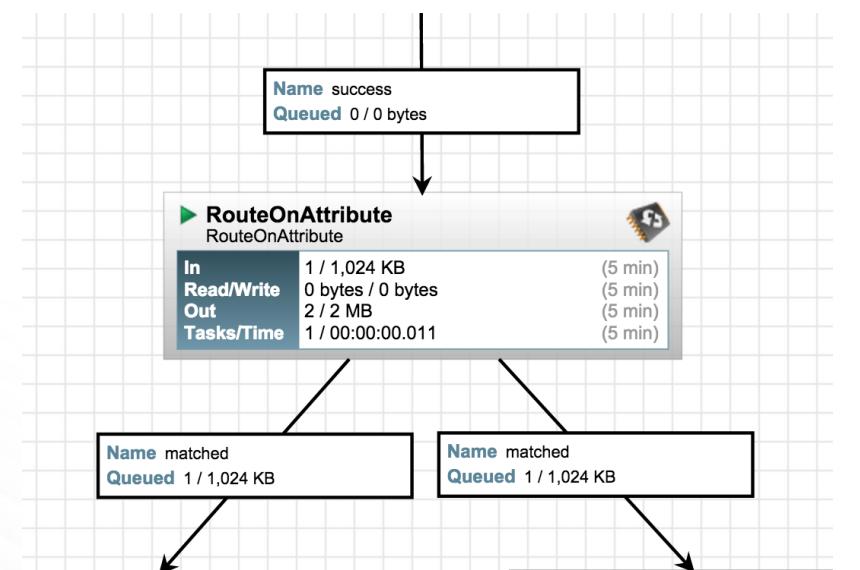
NiFi Architecture – Repositories - Pass by reference

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

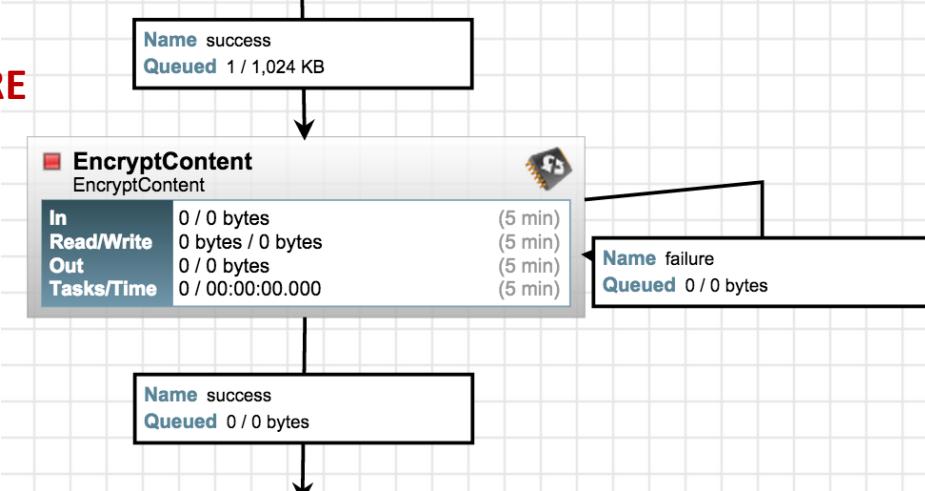
Content

Provenance

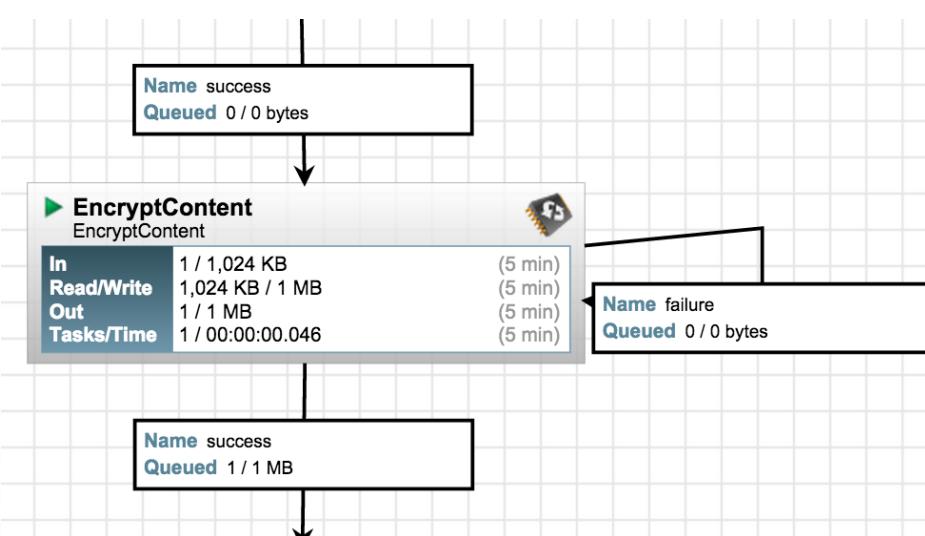
NiFi Architecture – Repositories – Copy on Write

Excerpt of demo flow...

BEFORE



AFTER



What's happening inside the repositories...

$F_1 \rightarrow C_1$

C_1

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

$F_1 \rightarrow C_1$
 $F_{1.1} \rightarrow C_2$

C_1 (plaintext)
 C_2 (encrypted)

$P_1 \rightarrow F_1 - \text{CREATE}$
 $P_2 \rightarrow F_{1.1} - \text{MODIFY}$



Agenda



What is dataflow and what are the challenges?
Apache NiFi
Apache MiNiFi
Architecture
Community

Why NiFi?

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

Healthy Community

apache / nifi
mirrored from <git://git.apache.org/nifi.git>

Unwatch 101 Unstar 446 Fork 416

Code Pull requests 96 Projects 0 Pulse Graphs

Mirror of Apache NiFi



Unresolved: By Priority

Priority	Issues	Percentage
Blocker	6	1%
Critical	19	2%
Major	495	45%
Minor	516	47%
Trivial	71	6%
View Issues		

Status Summary

Status	Issues	Percentage
Open	1008	30%
In Progress	20	1%
Reopened	10	1%
Resolved	2234	66%
Closed	63	2%
Patch Available	69	2%

[View Issues](#)

Learn more and join us

Apache NiFi site

<https://nifi.apache.org>

Subproject MiNiFi site

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

Subscribe to and collaborate at

dev@nifi.apache.org

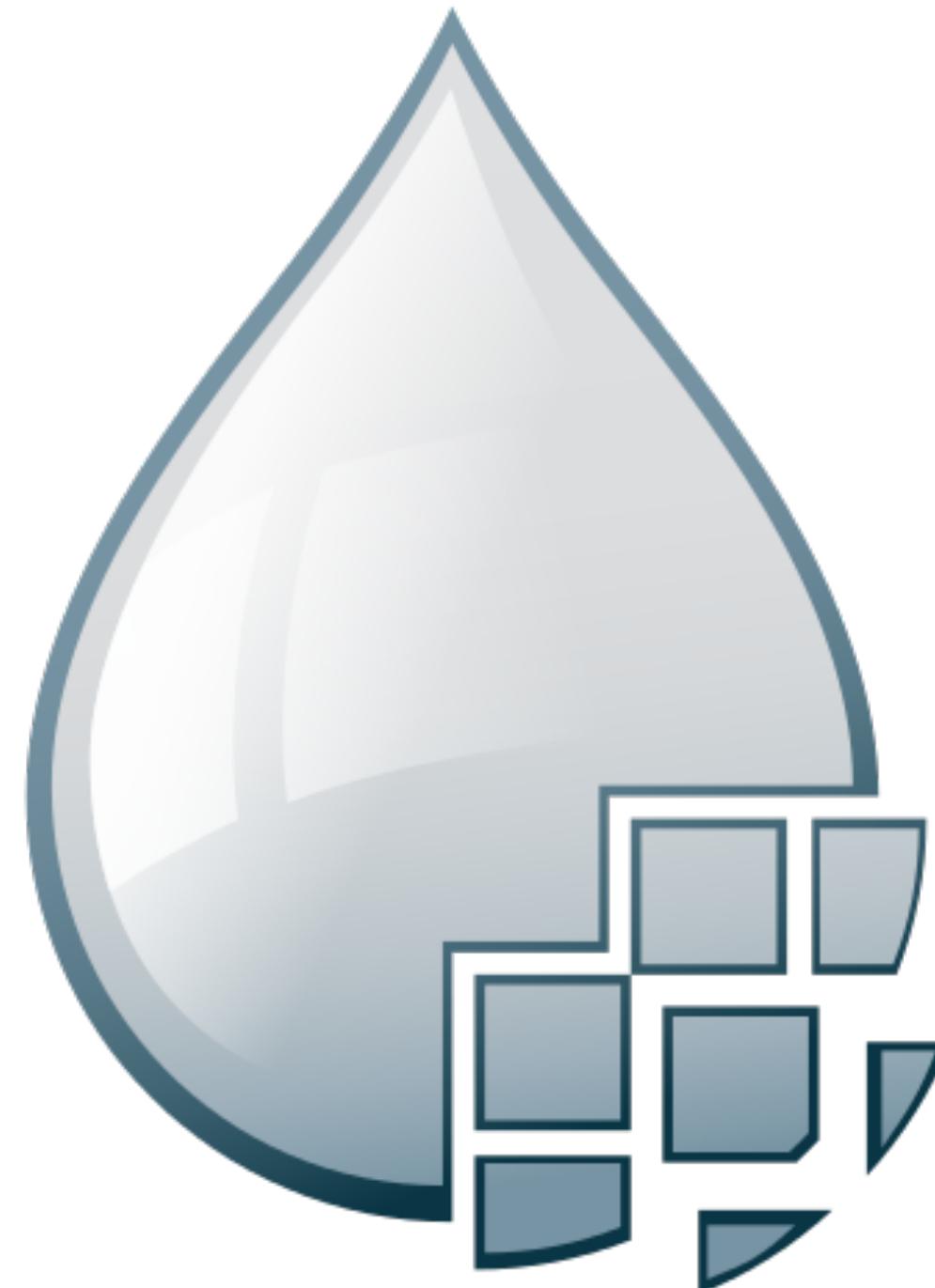
users@nifi.apache.org

Submit Ideas or Issues

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

Follow us on Twitter

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



Thank You

I'm sticking around for discussions/questions



@yolopey / @apachennifi

alopresto@apache.org

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