CLOUDERA APACHECON

Logging modernization and cybersecurity at scale with (Mi)NiFi, Kafka and Flink

Pierre Villard

APACHE nifi

Pierre Villard

Director of Product Management, Data in Motion @ Cloudera

Committer and PMC member of Apache NiFi (involved since 2015)

Twitter/Github - @pvillard31

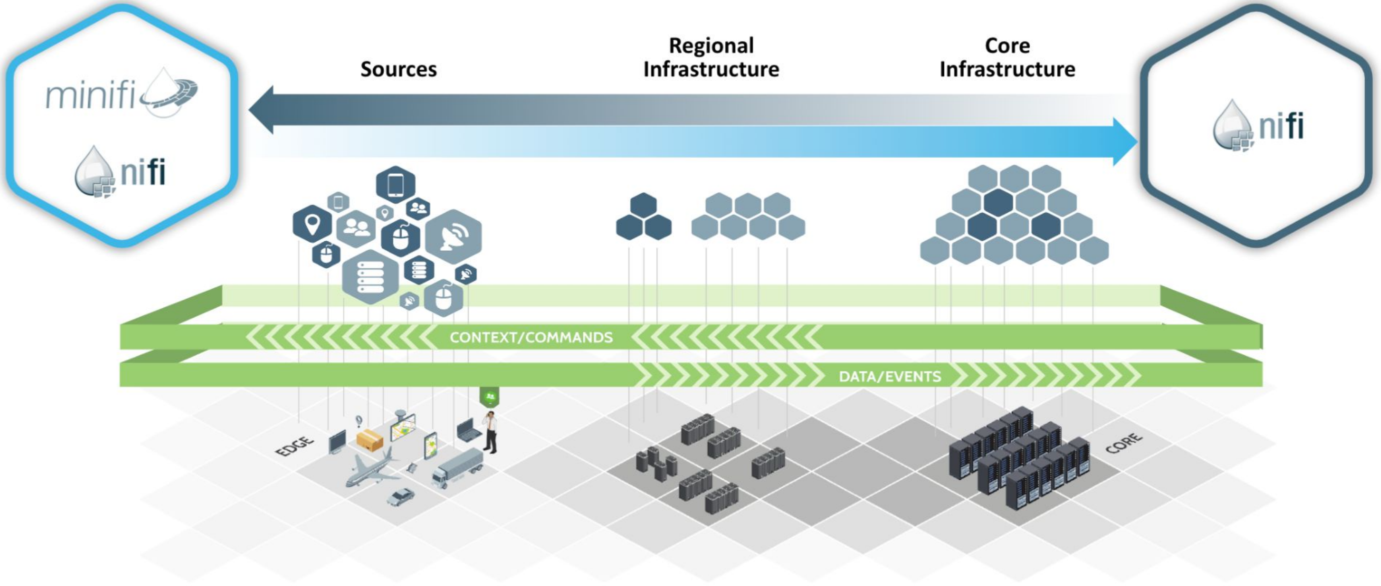
Apache NiFi - a software developed 16 years ago by the NSA

2006
NiagaraFiles (NiFi) was first incepted at the National Security Agency (NSA)

November 2014
NiFi is donated to the Apache Software Foundation (ASF) through NSA's Technology Transfer Program and enters ASF's incubator.

July 2015
NiFi reaches ASF top-level project status

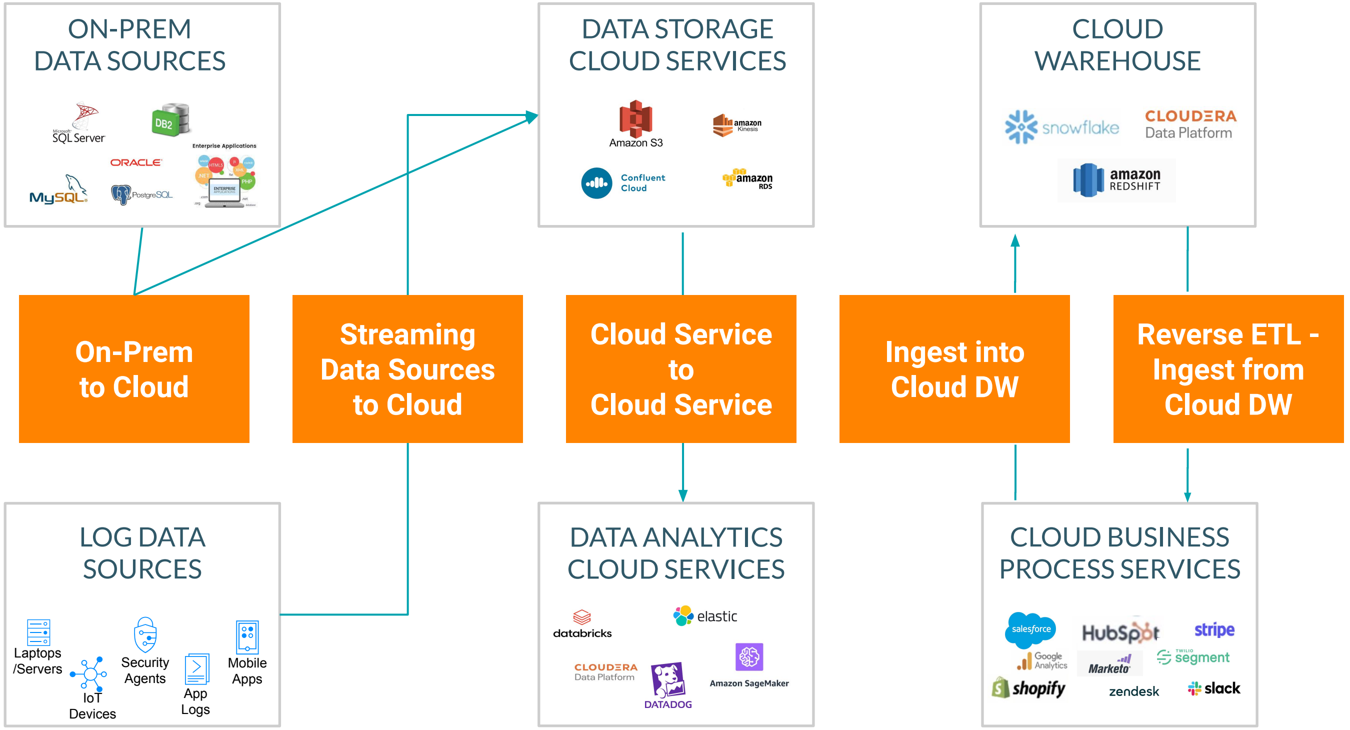
What is NiFi used for?



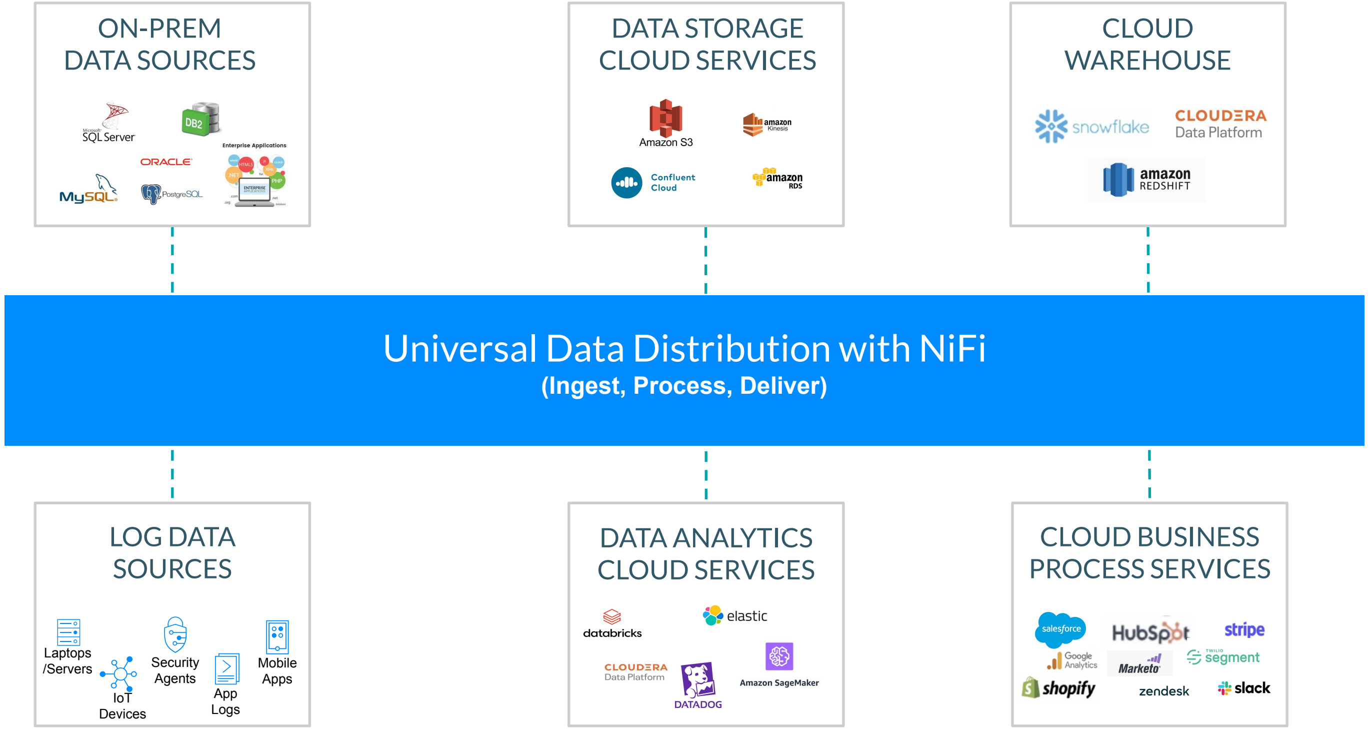
Common challenges for the modern data stack

DIFFERENT SYSTEMS

SIMILAR PROBLEMS



NiFi, the gateway for Universal Data Distribution



500+ components for Universal Data Distribution

- FTP
- SFTP
- HL7
- UDP
- XML
- • •
- • •
- HTTP
- WebSocket
- Email
- HTML
- Image
- Syslog
- AMQP

AVRO
kafka
HIVE
elastic
FLUME
Spark
Solr
EVENT HUBS
STORM
hadoop
SQL
Couchbase
amazon web services S3
Flink
cassandra
minifi
slack
DynamoDB amazon web services
splunk>
mongoDB
amazon Kinesis

Hash	Encrypt	GeoEnrich
Merge	Tail	Scan
Extract	Evaluate	Replace
Duplicate	Execute	Translate
Split	Fetch	Convert

• •
• •
• •
• •
• •

Route Text	Distribute Load
Route Content	Generate Table Fetch
Route Context	Jolt Transform JSON
Control Rate	Prioritized Delivery

CLOUDERA APACHE nifi

The NiFi ecosystem

- **NiFi** - Powerful and scalable directed graphs of data routing, transformation, and system mediation logic.
- **MiNiFi (Java version)** - Complementary data collection approach that supplements the core tenets of NiFi in dataflow management, focusing on the collection of data at the source of its creation.
- **MiNiFi (C++ version)** - The C++ implementation is an additional implementation to the one in Java with the aim of an even smaller resource footprint. Perspectives of the role of MiNiFi should be from the perspective of the agent acting immediately at, or directly adjacent to, source sensors, systems, or servers.
- **NiFi Registry** - Complementary application that provides a central location for storage and management of shared resources across one or more instances of NiFi and/or MiNiFi.
- **NiFi C2 Server** - Command and control server to manage many disparate agents running on all sorts of devices, to coordinate their work and to push out revised flows/configurations.
- **NiFi Fluid Design System** - Atomic reusable platform providing consistent set of UI/UX components.

Apache NiFi in a few numbers

A very active project with a dynamic community & comparison with ACEU 2019

2400+ members on the Slack channel (535+ - 3 years ago)
450+ contributors on Github across the repositories (260+ - 3 years ago)
60 committers in the Apache NiFi community (45 - 3 years ago)
Apache NiFi 1.18.0 to be released soon (RC vote in progress!) (NiFi 1.10 - 3 years ago)
12M+ docker pulls of the Apache NiFi image (1M+ - 3 years ago)

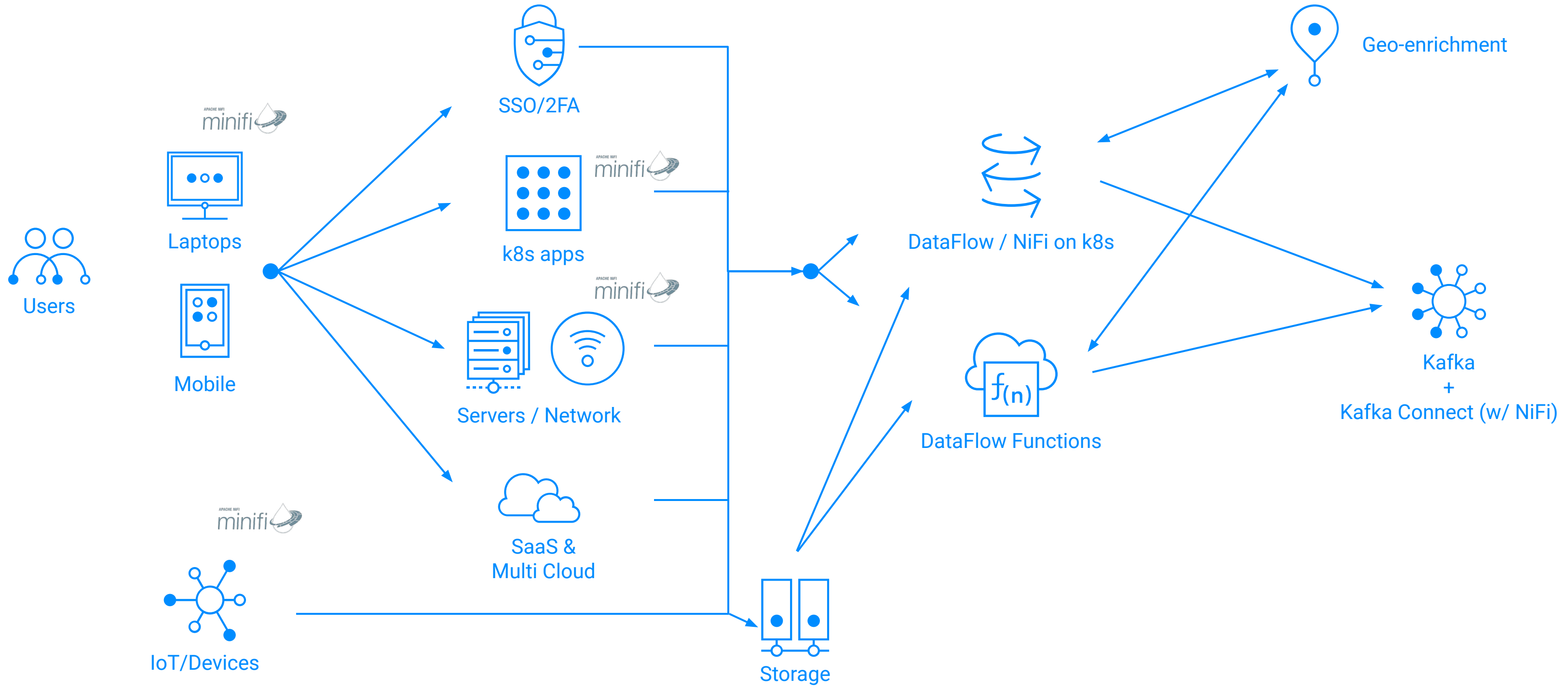
Log Modernization & Cybersecurity

Challenges and problem statement

- A lot of systems / softwares across the company
- Different AuthN/AuthZ mechanisms
- Various formats/schemas
- Need for normalization of the data and enrichment
- Need to take actions in real time

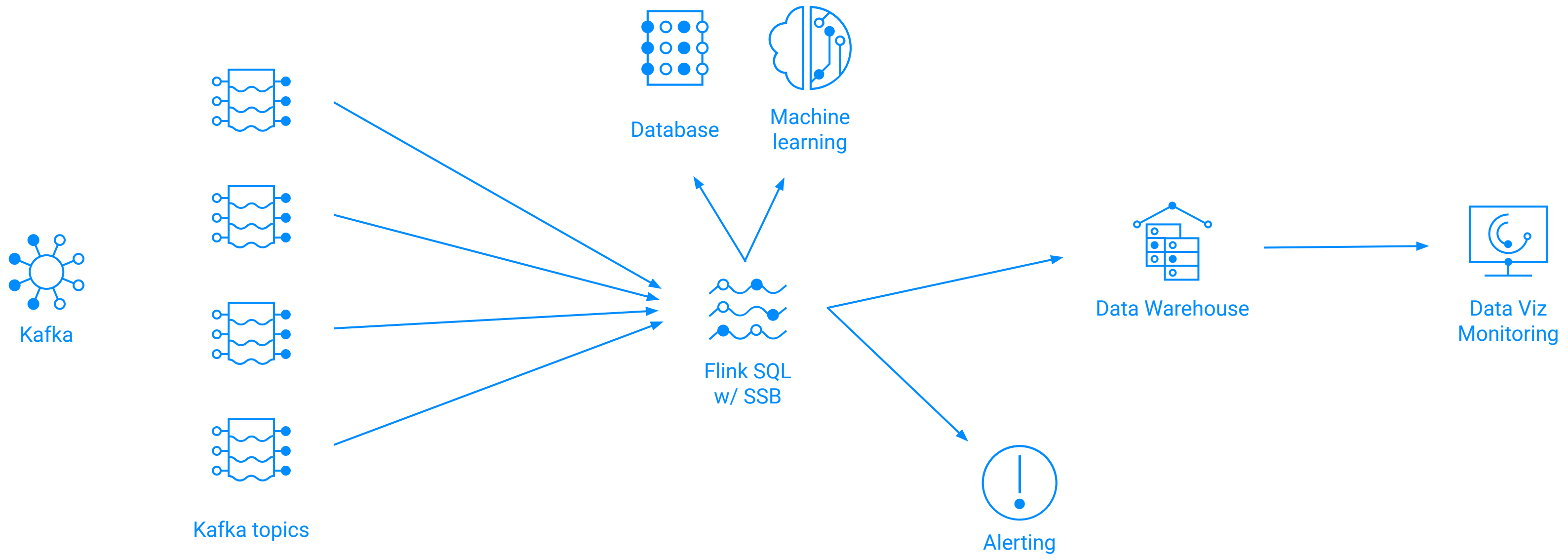
Architecture in the context of cybersecurity use cases (1/2)

Apache NiFi for Universal Data Collection & Distribution



Architecture in the context of cybersecurity use cases (2/2)

Kafka & Flink (Flink SQL with Stream SQL Builder) for real time analytics



Logs formats & normalization

NiFi Record Readers

Some Record Readers are available in NiFi to help you with logs processing:

- CiscoEmblemSyslogMessageReader
- CEFReader
- GrokReader
- IPFIXReader
- SyslogReader
- WindowsEventLogReader
- Netflow
- etc

+ scripted ones + extensibility

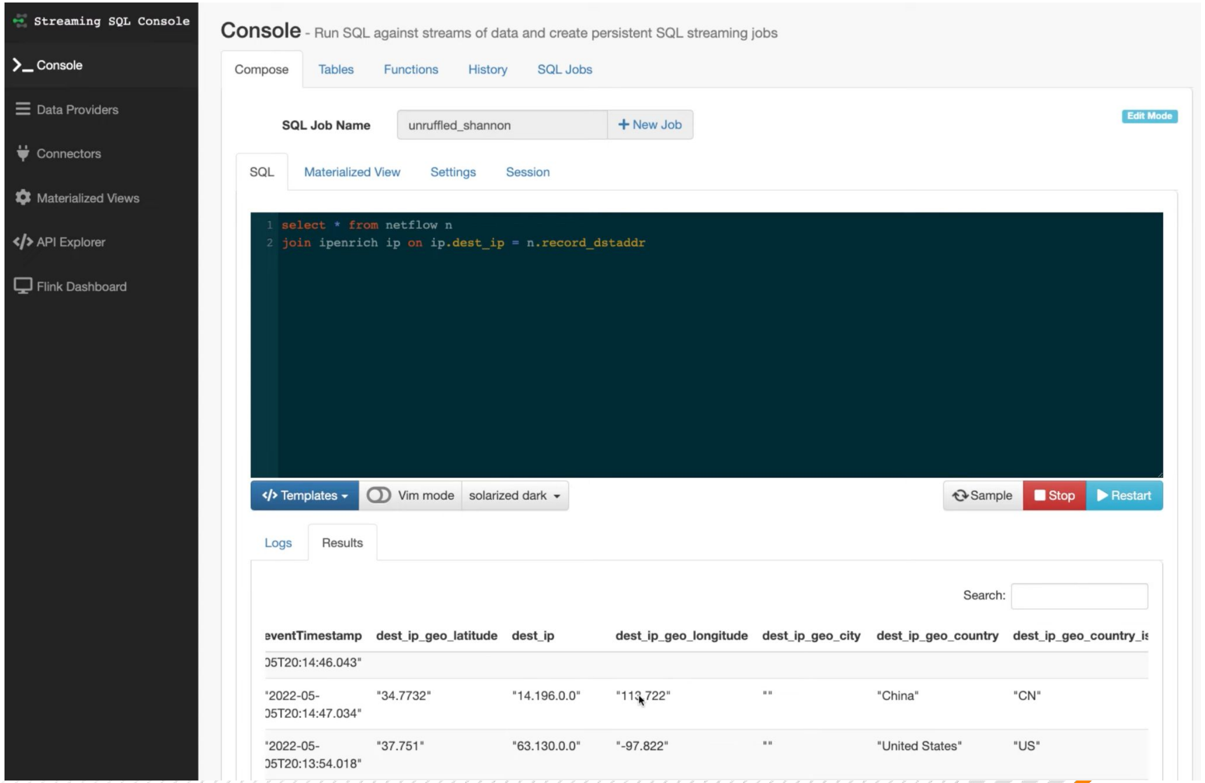
Example with CiscoEmblemSyslogMessageReader

https://medium.com/cloudera-inc/processing-cisco-asa-logs-with-cloudera-flow-management-f09cdf7382c3

```
2022-01-01T12:00:00Z
APPLIANCE-1 %ASA-6-725001:
Starting SSL handshake
with client
OUTSIDE:10.0.0.1/1024 to
192.168.1.100/443 for TLS
session
```

```
{
  "format": "PARSED",
  "log": "2022-01-01T12:00:00Z APPLIANCE-1 %ASA-6-725001: Starting SSL
handshake with client OUTSIDE:10.0.0.1/1024 to 192.168.1.100/443 for
TLS session",
  "timestamp": 1641038400000,
  "hostname": "APPLIANCE-1",
  "message": "Starting SSL handshake with client OUTSIDE:10.0.0.1/1024
to 192.168.1.100/443 for TLS session",
  "facility": "ASA",
  "level": 6,
  "messageNumber": 725001,
  "peerType": "client",
  "sourceInterface": "OUTSIDE",
  "sourceAddress": "10.0.0.1",
  "sourcePort": "1024",
  "destinationAddress": "192.168.1.100",
  "destinationPort": "443",
  "handshakeProtocol": "TLS"
}
```

FlinkSQL & SSB

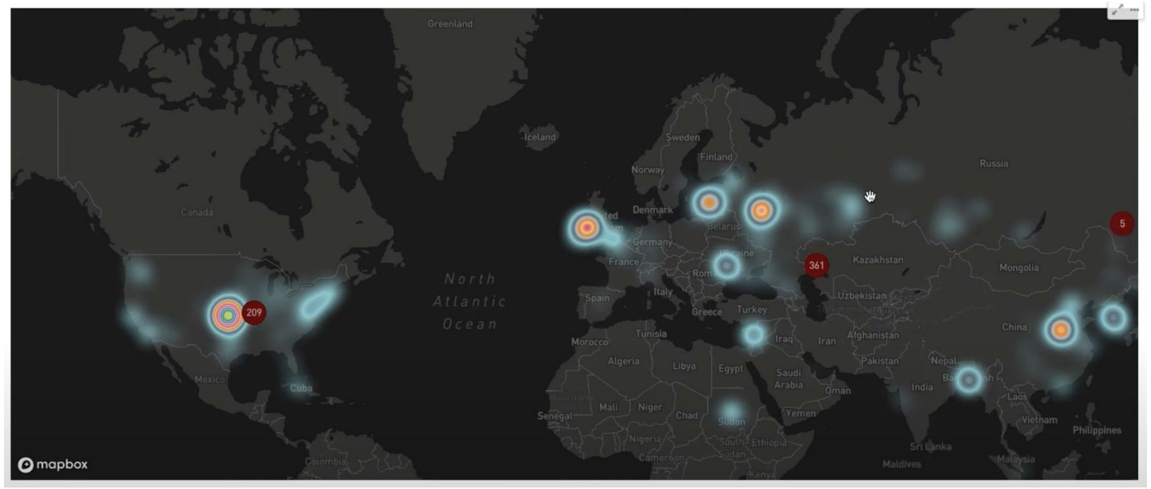


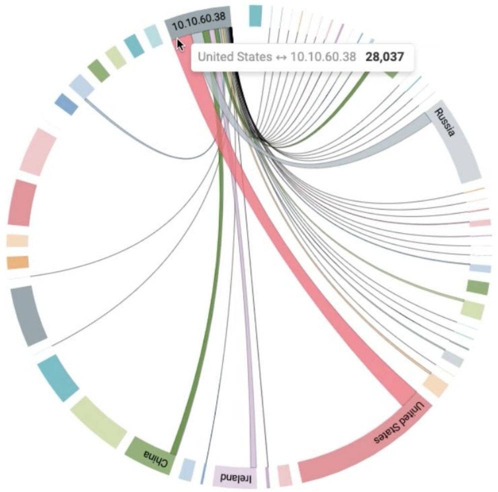
SQL Job Name: unruffled_shannon + New Job

```
select * from netflow n
join ipenrich ip on ip.dest_ip = n.record_dstaddr
```

eventTimestamp	dest_ip_geo_latitude	dest_ip	dest_ip_geo_longitude	dest_ip_geo_city	dest_ip_geo_country	dest_ip_geo_country_is
)5T20:14:46.043"						
"2022-05-)5T20:14:47.034"	"34.7732"	"14.196.0.0"	"113.722"	""	"China"	"CN"
"2022-05-)5T20:13:54.018"	"37.751"	"63.130.0.0"	"-97.822"	""	"United States"	"US"

Dashboarding & DataViz





THANK YOU

CLOUDERA