pnda.io

PNDA.io: when BGP meets Big-Data

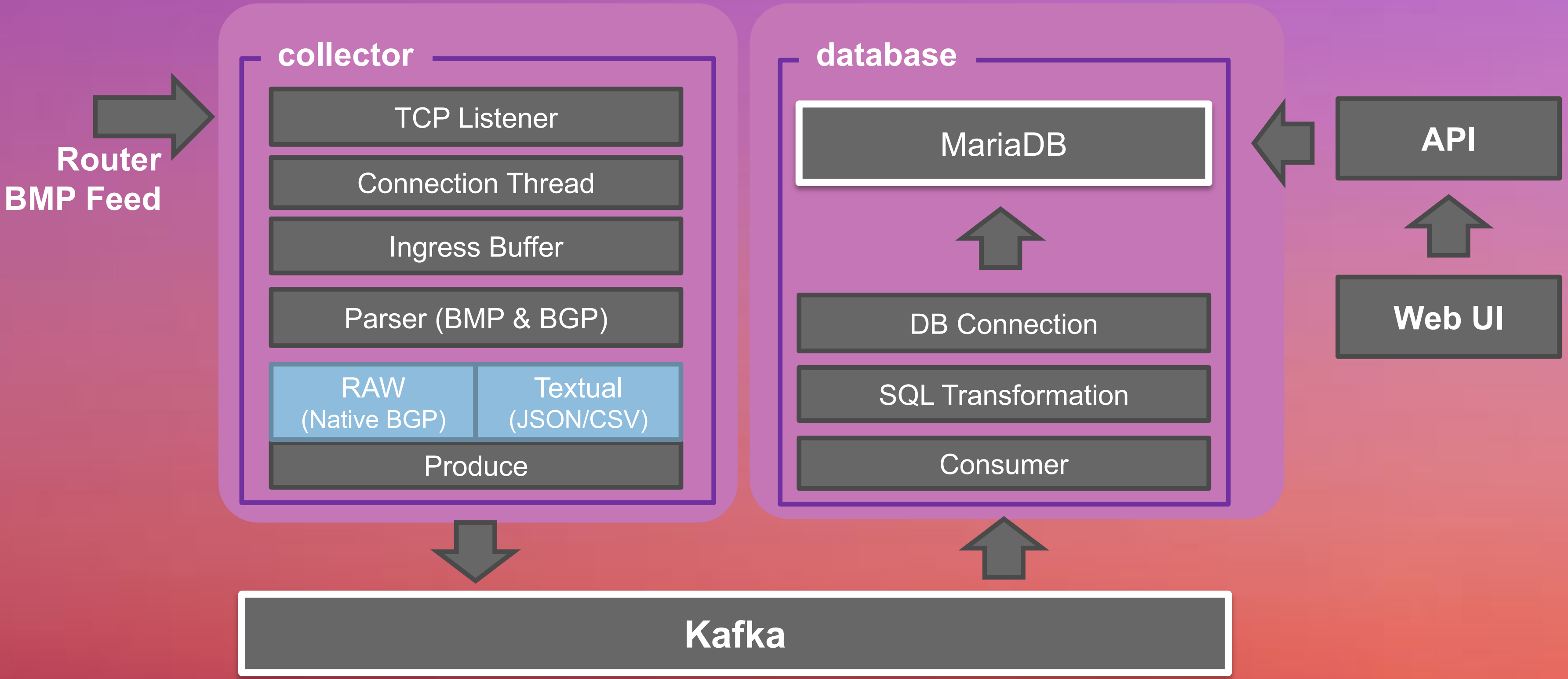
26th April 2017

Let’s go back in time…

The Internet is very much ‘alive’

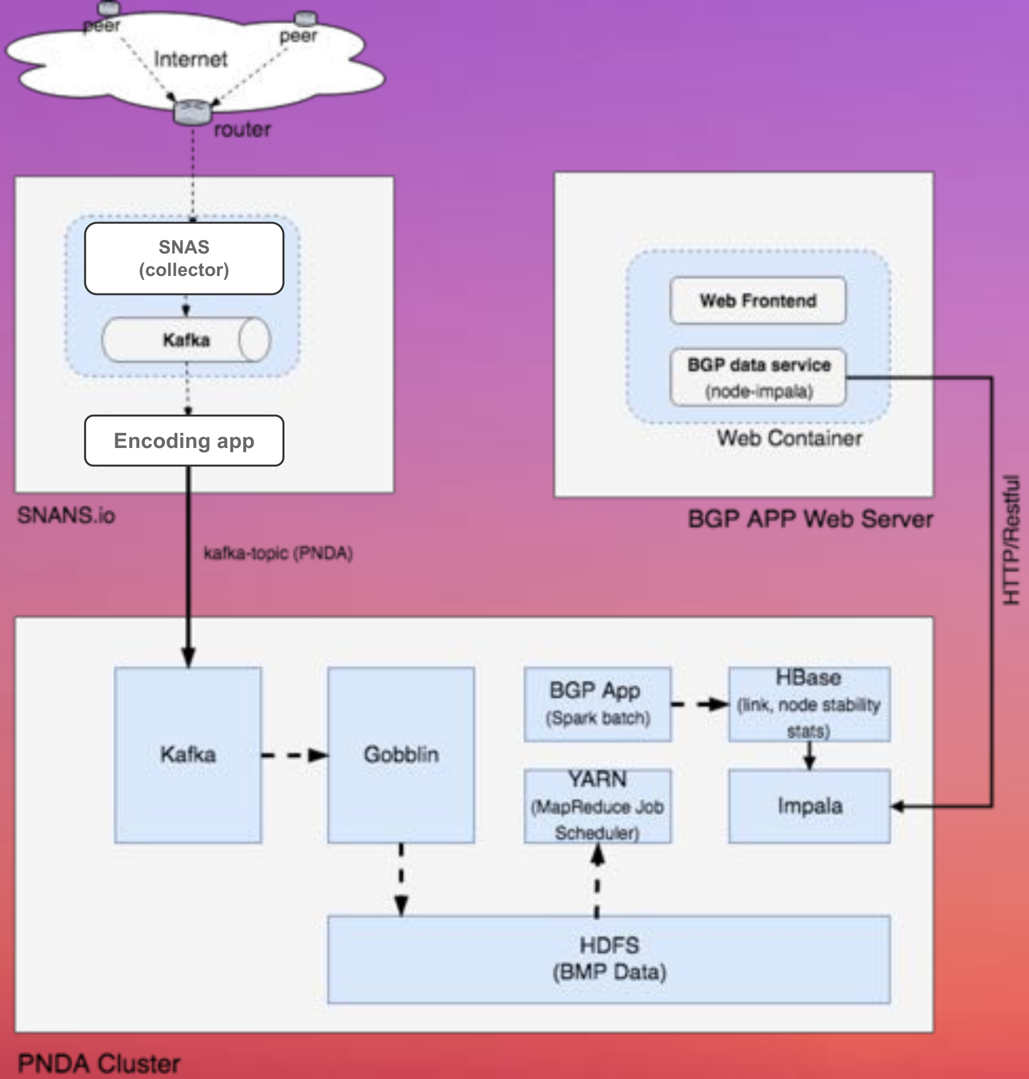
- Millions of BGP events occurring every day
  - 15 Routers Monitored
  - 410 active peers (both IPv4 and IPv6)
  - ~120,000,000 Prefixes Advertised
  - ~950,000 events per day from a single transit peer
  - ~202,000,000 changes per day
  - ~6,000,000,000 changes per month
- How do we extract ‘signal’ from ‘noise’?
- Can we apply techniques from other domains in this pursuit?

SNAS Architecture



E2E architecture

- Encoding app required to perform ‘avro’ encoding of BMP data
- BGP App runs as a Spark batch job, running periodically
- Can be converted to a Spark ‘streaming’ application for near-real-time processing



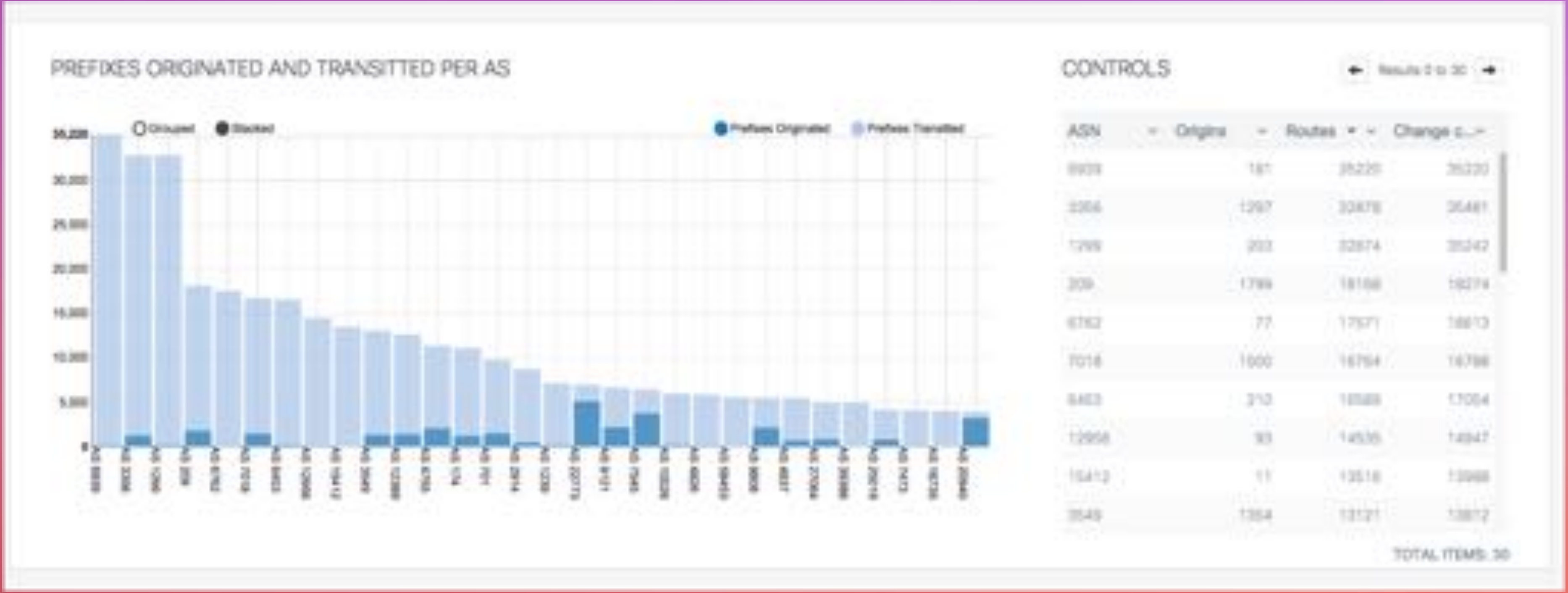
What does this give us?

SNAS.io gives us the ability to record the dynamics of the Internet

PNDA platform enables -

- ‘Raw’ event recording capability, with horizontal scaling (HDFS)
- Run analysis over very large data-sets with parallelism
- Ask questions of the aggregate data about the Internet
- Ask specific question
  - Per-prefix
  - Per-AS
  - Per AS-Path

Top-N analysis



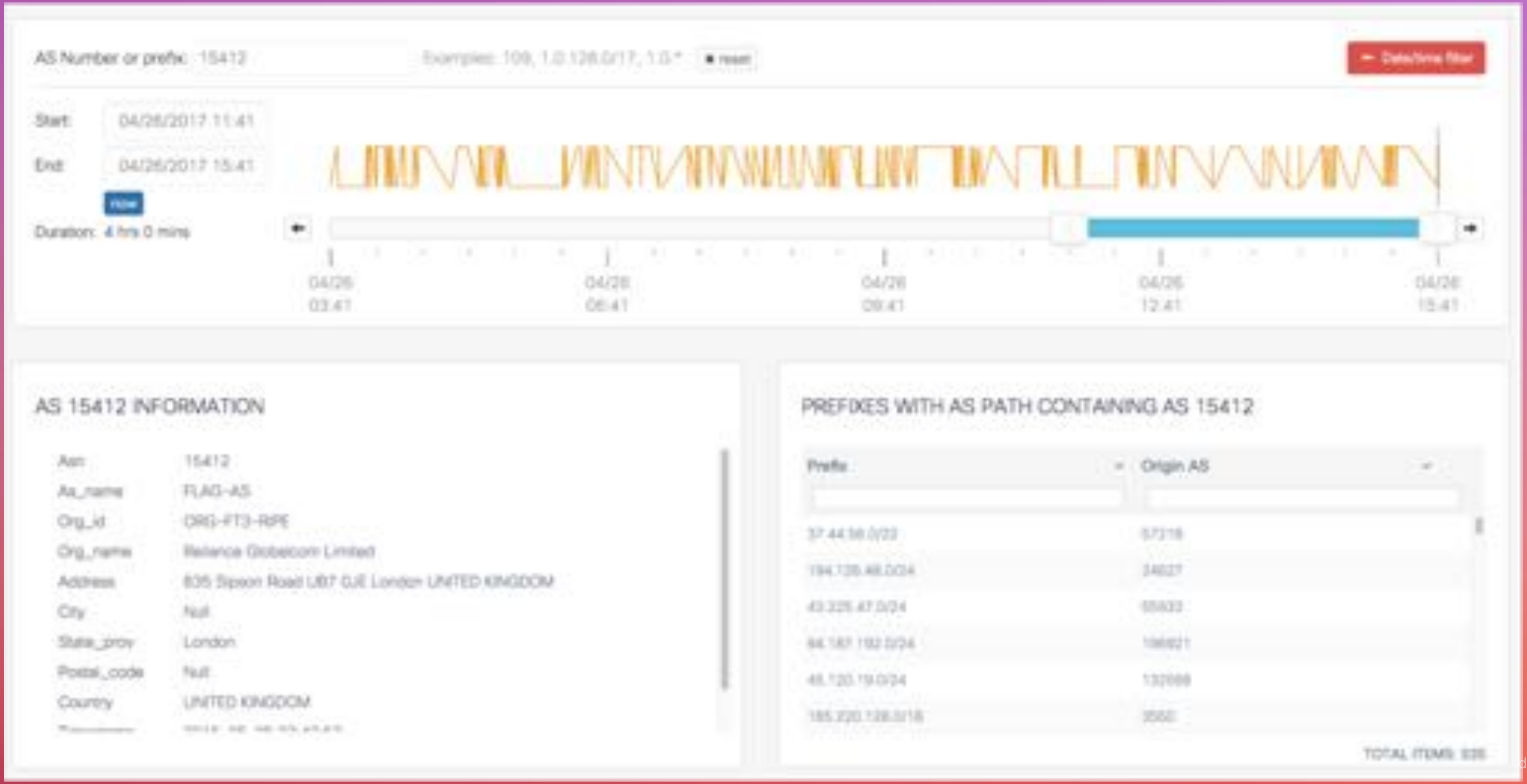
PREFIXES ORIGINATED AND TRANSITTED PER AS

CONTROLS

ASN	Origins	Routes	Change c...
6939	181	35220	35220
3356	1297	33478	35481
1299	203	32874	35242
209	1799	18168	18274
6762	77	17571	18613
7018	1000	16764	16798
6453	212	16589	17004
12956	93	14535	14947
15412	11	13516	13988
3549	1354	13121	13872

TOTAL ITEMS: 30

Path stability



AS Number or prefix: 15412 Examples: 109, 1.0.128.0/17, 1.0.*

Start: 04/26/2017 11:41

End: 04/26/2017 15:41

Duration: 4 hrs 0 mins

AS 15412 INFORMATION

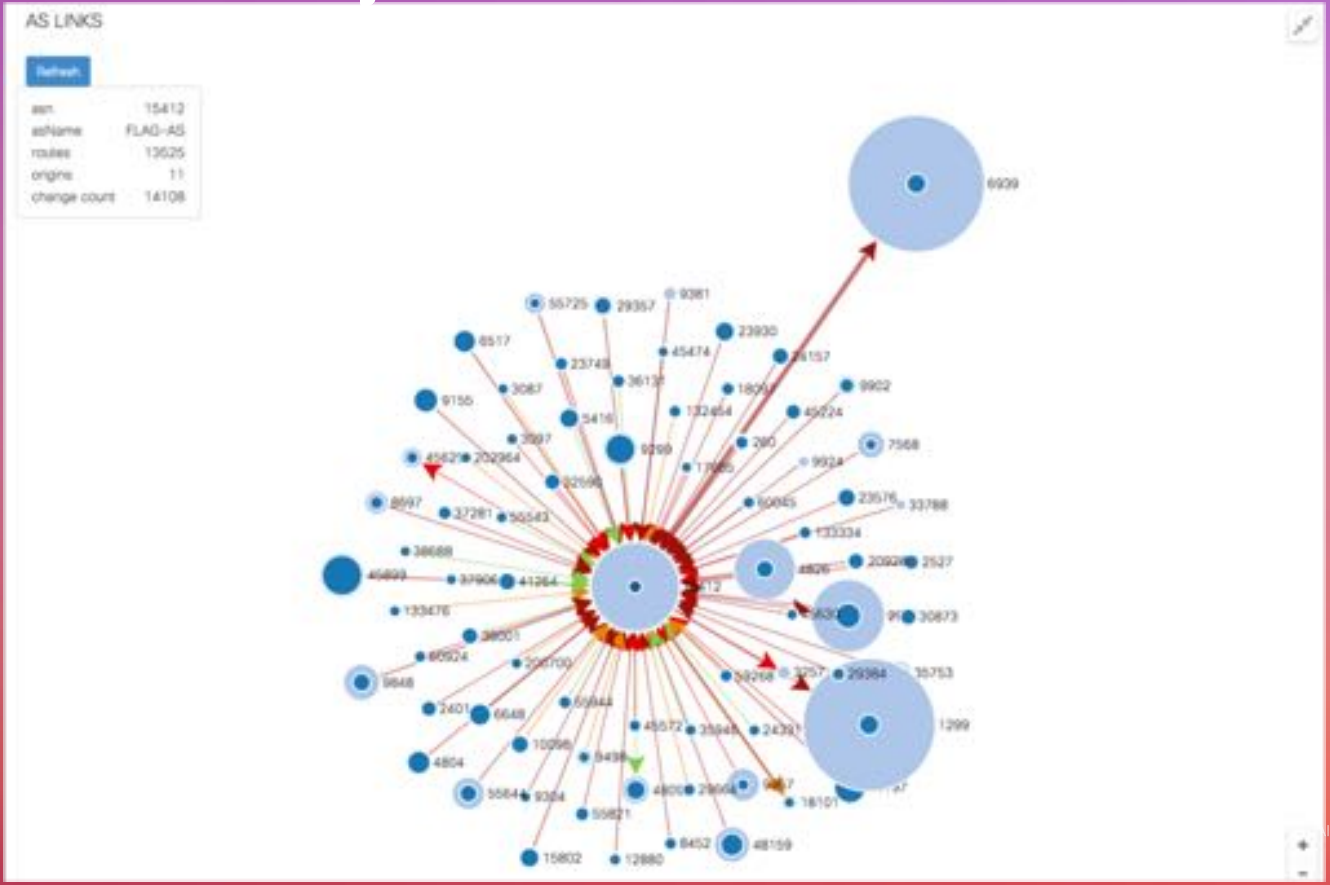
Asn	15412
As_name	FLAG-AS
Org_id	ORG-FT3-RIPE
Org_name	Reliance Globalcom Limited
Address	835 Sipson Road UB7 0JE London UNITED KINGDOM
City	Null
State_prov	London
Postal_code	Null
Country	UNITED KINGDOM
[illegible]	[illegible]

PREFIXES WITH AS PATH CONTAINING AS 15412

Prefix	Origin AS
37.44.56.0/23	57218
194.126.48.0/24	24627
43.225.47.0/24	55933
84.187.192.0/24	136821
45.120.19.0/24	132668
185.220.128.0/18	3560

TOTAL ITEMS: 326

AS Connectivity - FLAG

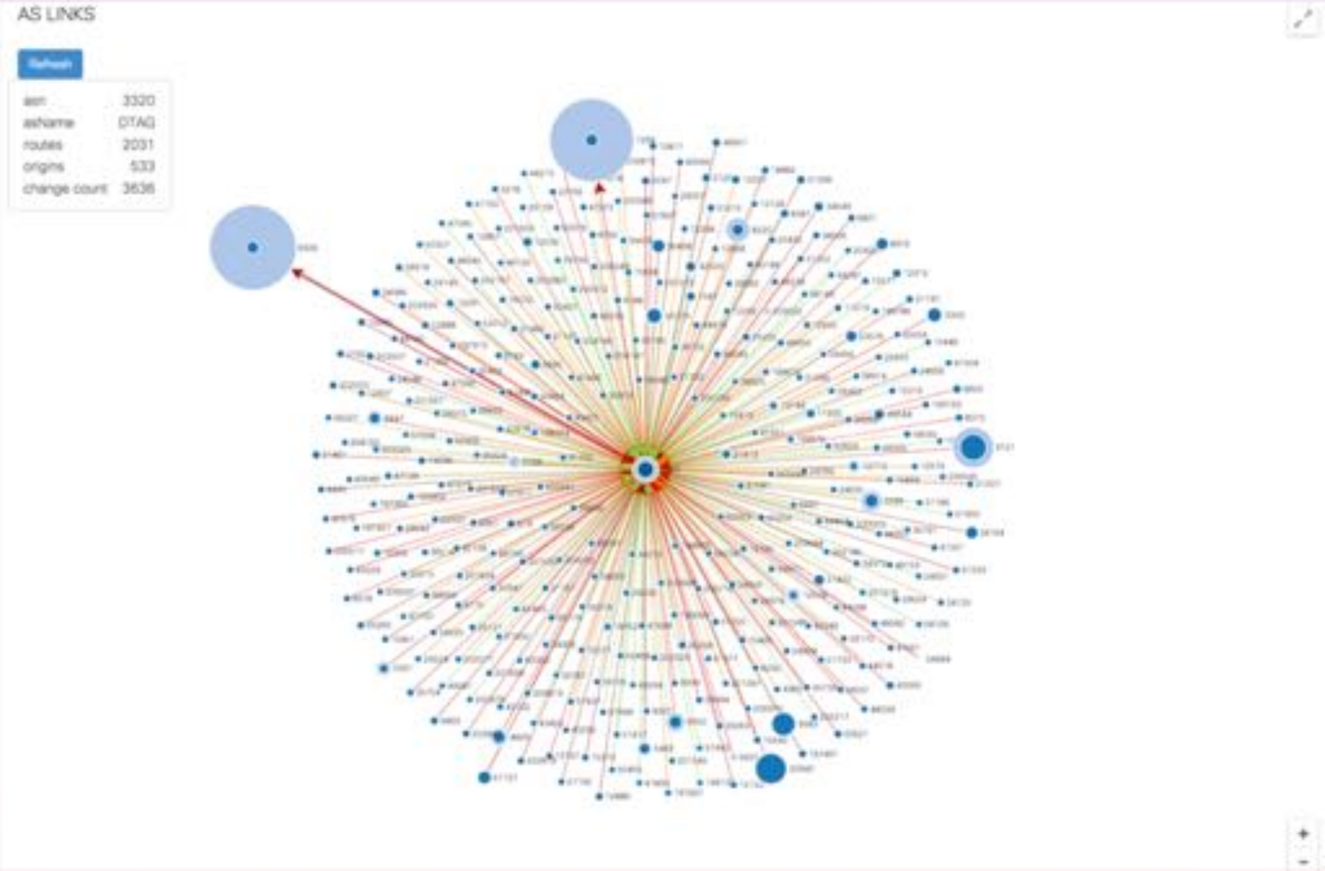


AS LINKS

Refresh

asn	15412
asName	FLAG-AS
routes	13525
origins	11
change count	14108

AS Connectivity – Deutsche Telekom



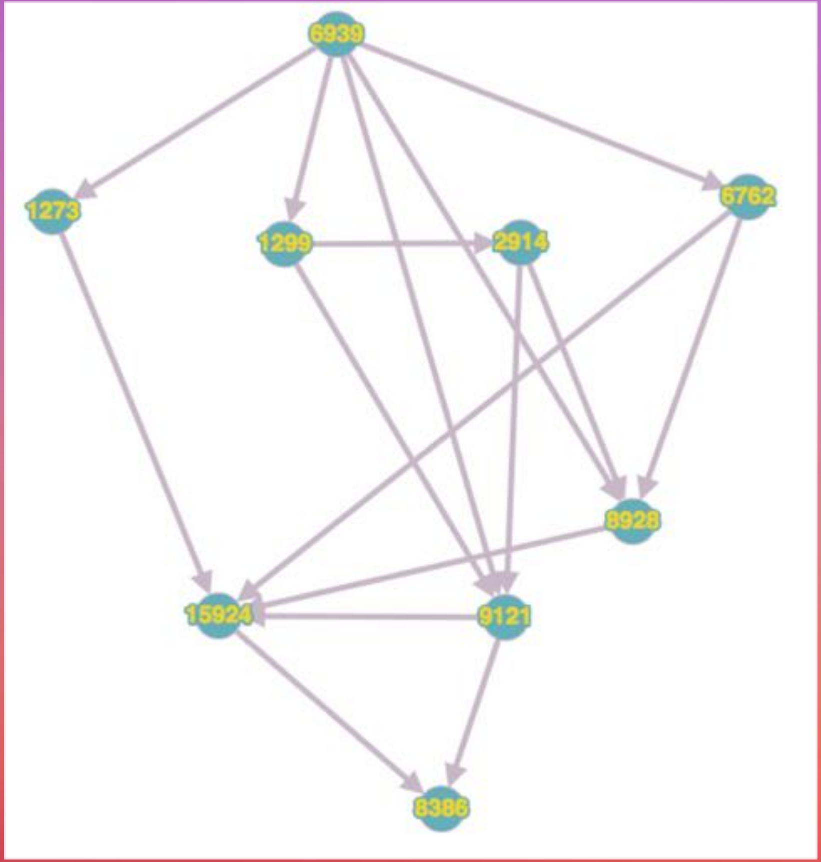
AS LINKS

asn	3320
asname	DTAG
routes	2031
origins	533
change count	3636

AS Path variance – 6939 to 8386

Shortest path – 4 hops
Longest path – 29 hops
Longest unique AS path – 6
Unique paths - 9
Largest prepend count – 17
Prepend variation – [7-17]
Path with most updates – via AS1273

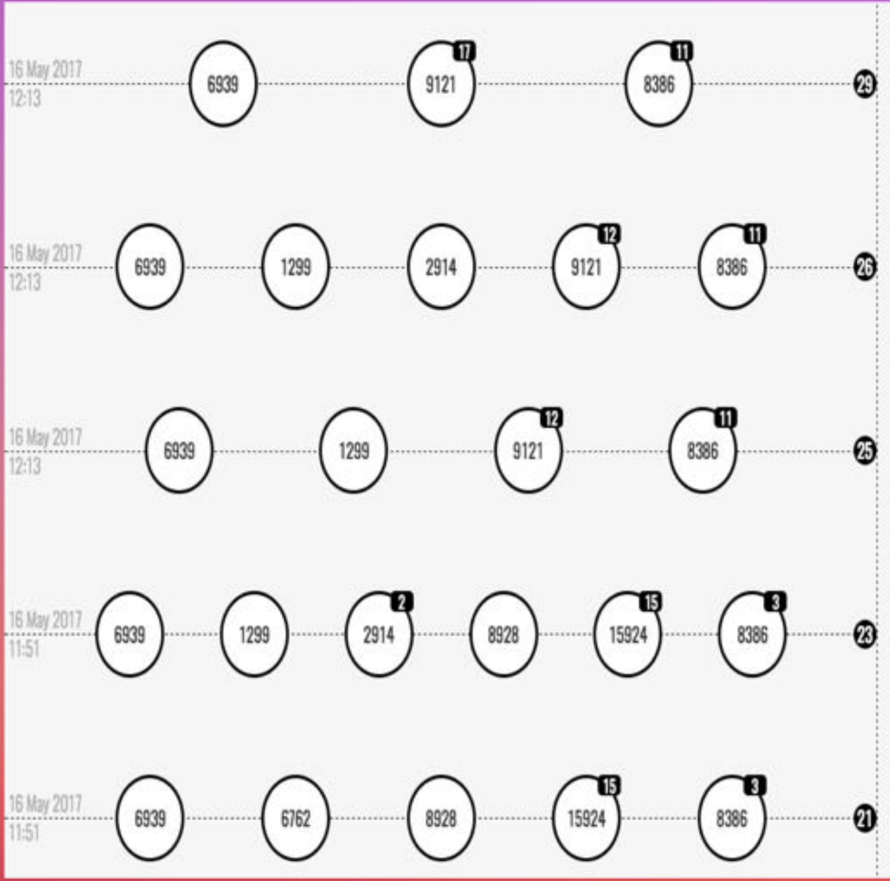
Data recorded in a 24hr period



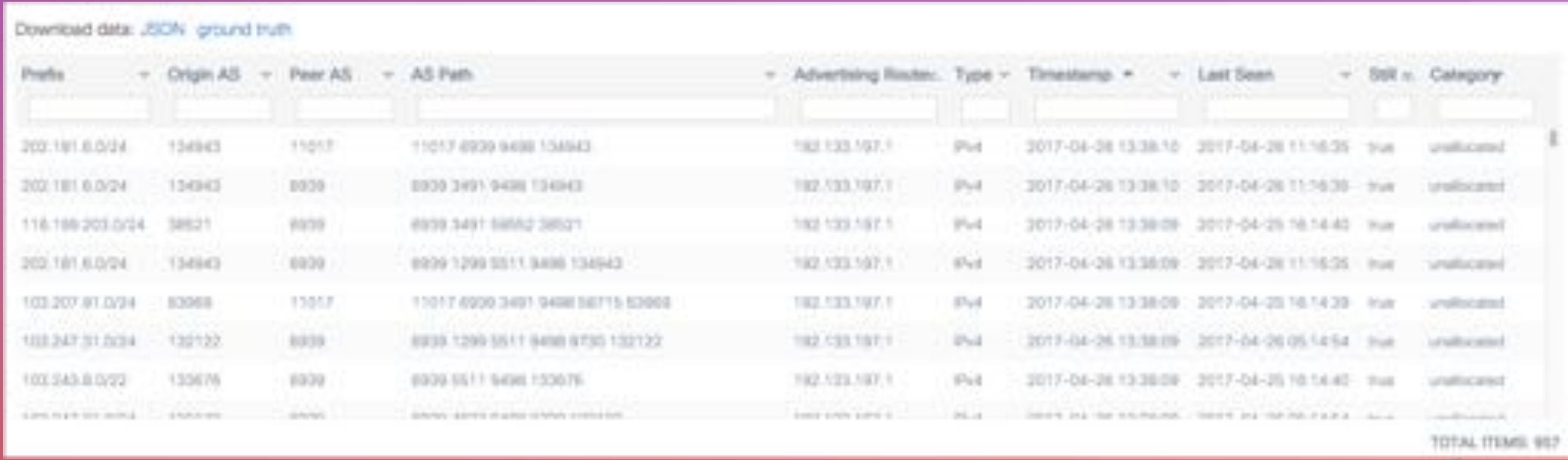
AS Path variance – 6939 to 8386

Shortest path – 4 hops
Longest path – 29 hops
Longest unique AS path – 6
Unique paths - 9
Largest prepend count – 17
Prepend variation – [7-17]
Path with most updates – via AS1273

Data recorded in a 24hr period



Security – Unallocated prefixes



Download data: JSON ground truth

Prefix	Origin AS	Peer AS	AS Path	Advertising Router	Type	Timestamp	Last Seen	Still ...	Category
202.181.6.0/24	134943	11017	11017 6939 9498 134943	192.133.197.1	IPv4	2017-04-26 13:38:10	2017-04-26 11:16:35	true	unallocated
202.181.6.0/24	134943	6939	6939 3491 9498 134943	192.133.197.1	IPv4	2017-04-26 13:38:10	2017-04-26 11:16:39	true	unallocated
116.199.203.0/24	38521	6939	6939 3491 58552 38521	192.133.197.1	IPv4	2017-04-26 13:38:09	2017-04-25 16:14:40	true	unallocated
202.181.6.0/24	134943	6939	6939 1299 5511 9498 134943	192.133.197.1	IPv4	2017-04-26 13:38:09	2017-04-26 11:16:35	true	unallocated
103.207.91.0/24	63968	11017	11017 6939 3491 9498 58715 63968	192.133.197.1	IPv4	2017-04-26 13:38:09	2017-04-25 16:14:39	true	unallocated
103.247.31.0/24	132122	6939	6939 1299 5511 9498 9730 132122	192.133.197.1	IPv4	2017-04-26 13:38:09	2017-04-26 05:14:54	true	unallocated
103.243.8.0/22	133676	6939	6939 5511 9498 133676	192.133.197.1	IPv4	2017-04-26 13:38:09	2017-04-25 16:14:40	true	unallocated
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

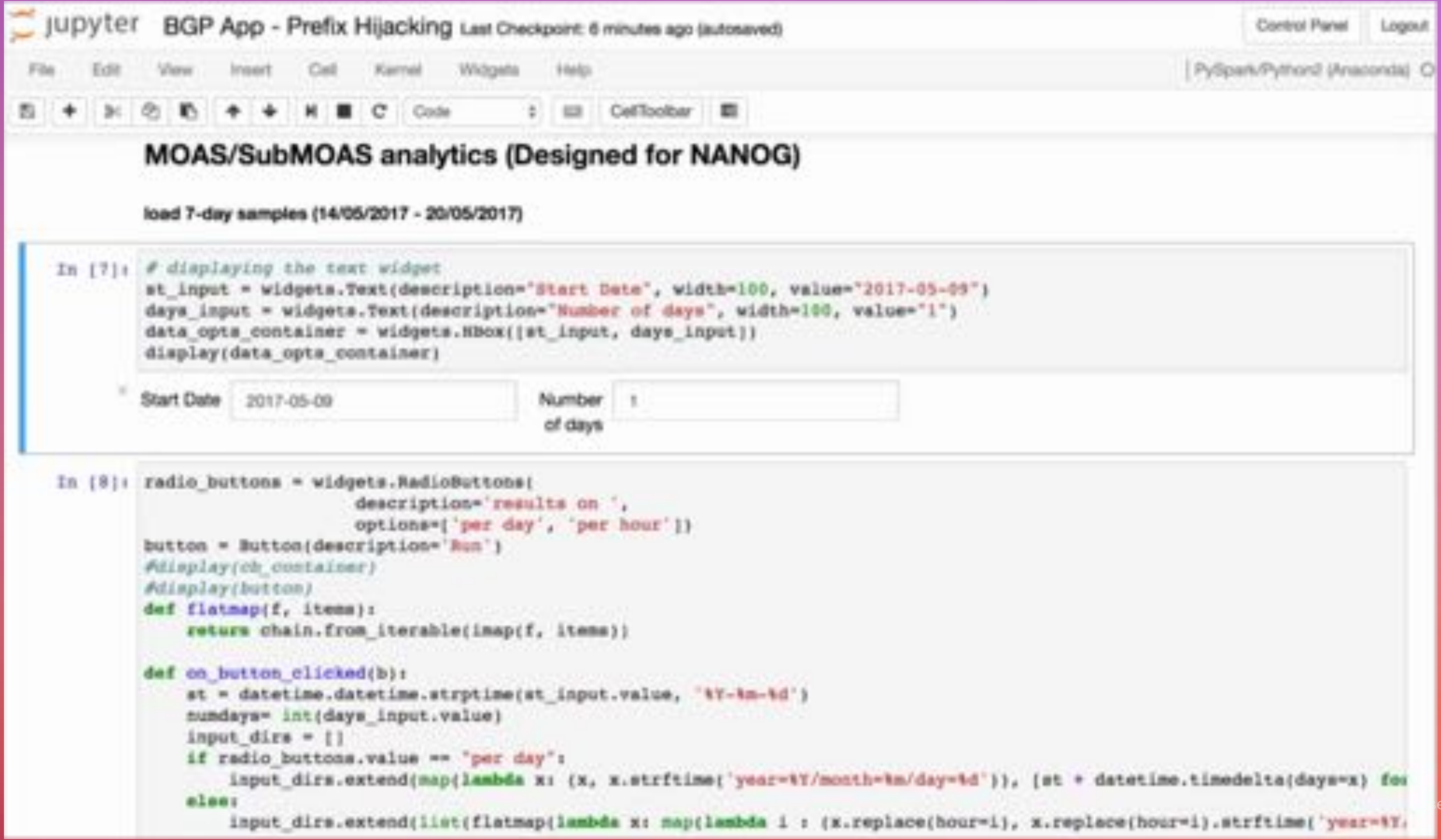
TOTAL ITEMS: 957

Observed over a 12 hour period

More specific prefix detection

- AS 12345 originates 100.100.0.0/18
- Hijacker originates 100.100.63.0/24
- Basically a needle in a large haystack, does anyone notice?

Looking for the needle using Jupyter Notebook



BGP App - Prefix Hijacking Last Checkpoint: 6 minutes ago (autosaved)

MOAS/SubMOAS analytics (Designed for NANOG)

load 7-day samples (14/05/2017 - 20/05/2017)

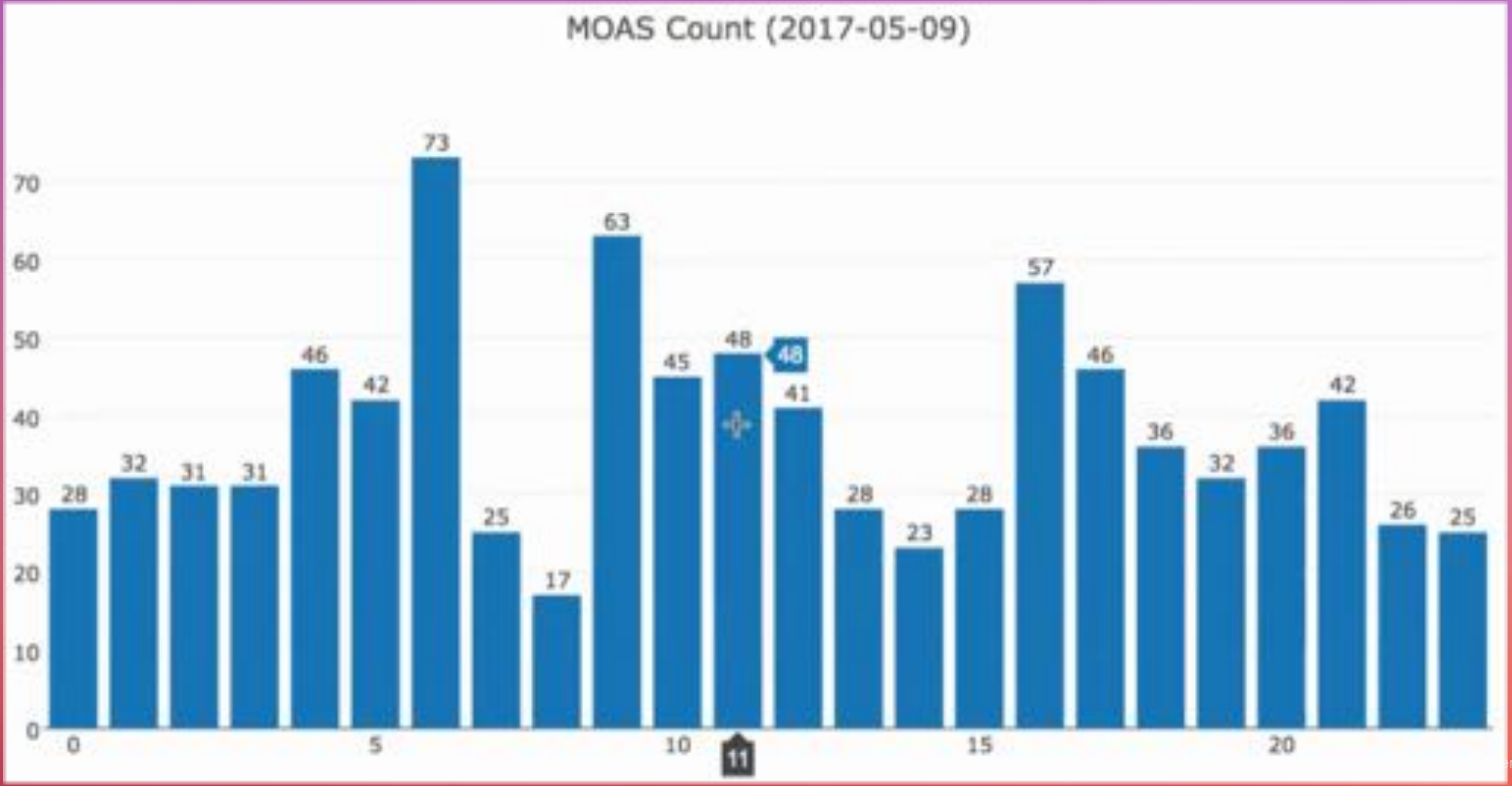
```
In [7]: # displaying the text widget
        st_input = widgets.Text(description="Start Date", width=100, value="2017-05-09")
        days_input = widgets.Text(description="Number of days", width=100, value="1")
        data_opts_container = widgets.HBox([st_input, days_input])
        display(data_opts_container)
```

Start Date: 2017-05-09   Number of days: 1

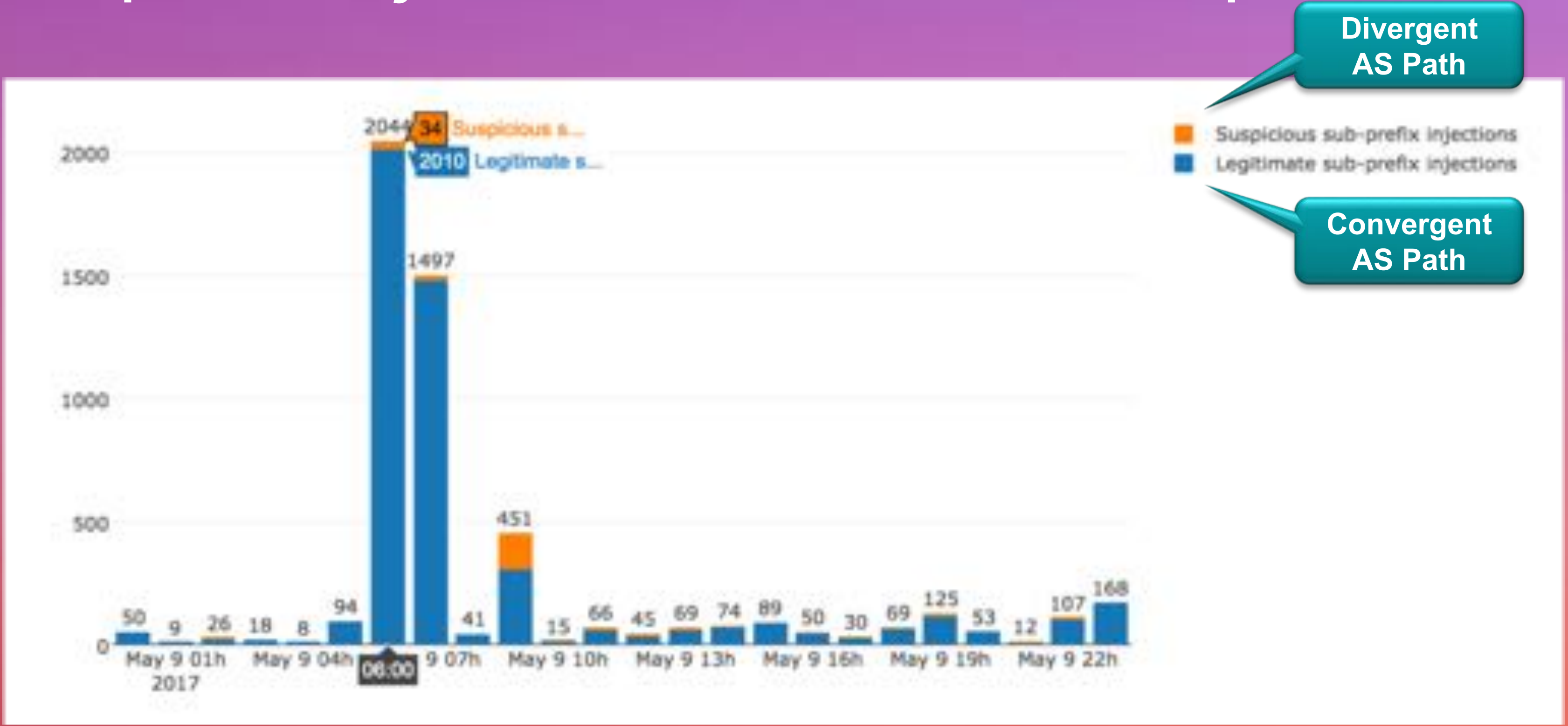
```
In [8]: radio_buttons = widgets.RadioButtons(
                             description='results on ',
                             options=['per day', 'per hour'])
        button = Button(description='Run')
        #display(cb_container)
        #display(button)
        def flatmap(f, items):
            return chain.from_iterable(imap(f, items))

        def on_button_clicked(b):
            st = datetime.datetime.strptime(st_input.value, '%Y-%m-%d')
            numdays= int(days_input.value)
            input_dirs = []
            if radio_buttons.value == "per day":
                input_dirs.extend(map(lambda x: (x, x.strftime('year=%Y/month=%m/day=%d')), [st + datetime.timedelta(days=x) fo
            else:
                input_dirs.extend(list(flatmap(lambda x: map(lambda i : (x.replace(hour=i), x.replace(hour=i).strftime('year=%Y,
```

Multi-Origin AS prefixes ‘add’ detected – 24 hour period



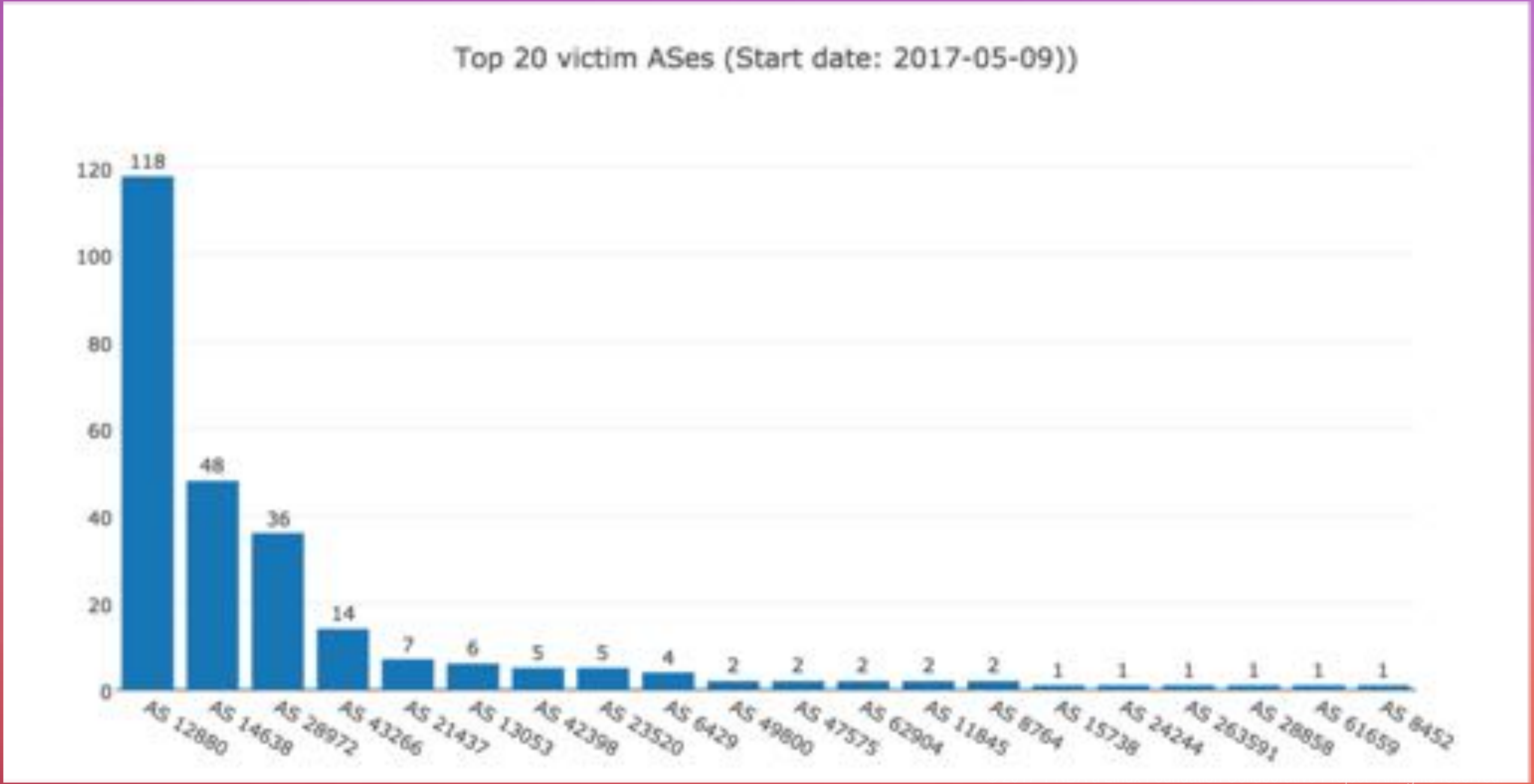
Sub-prefix injections over a 24 hour period



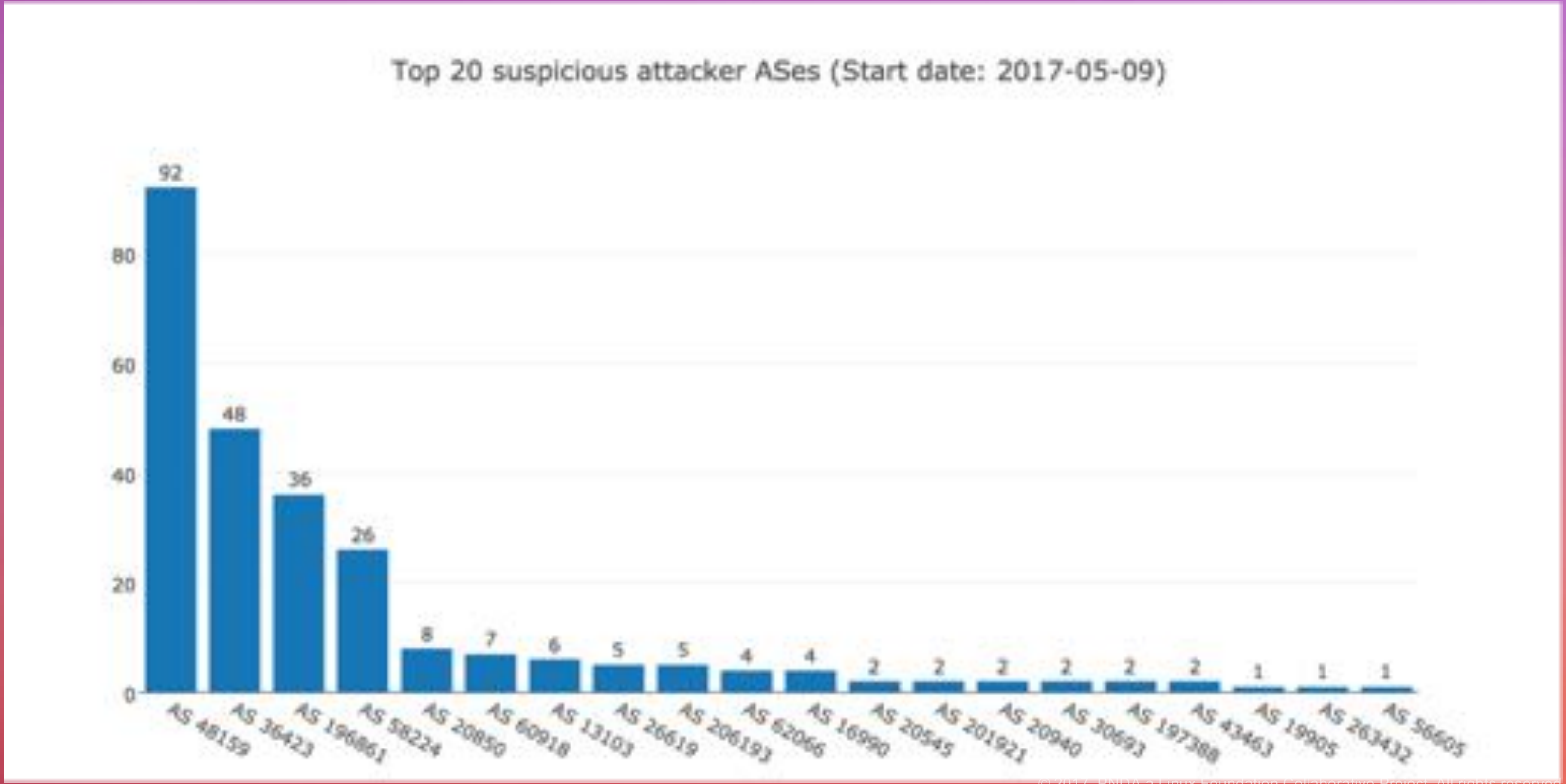
Divergent AS Path

Convergent AS Path

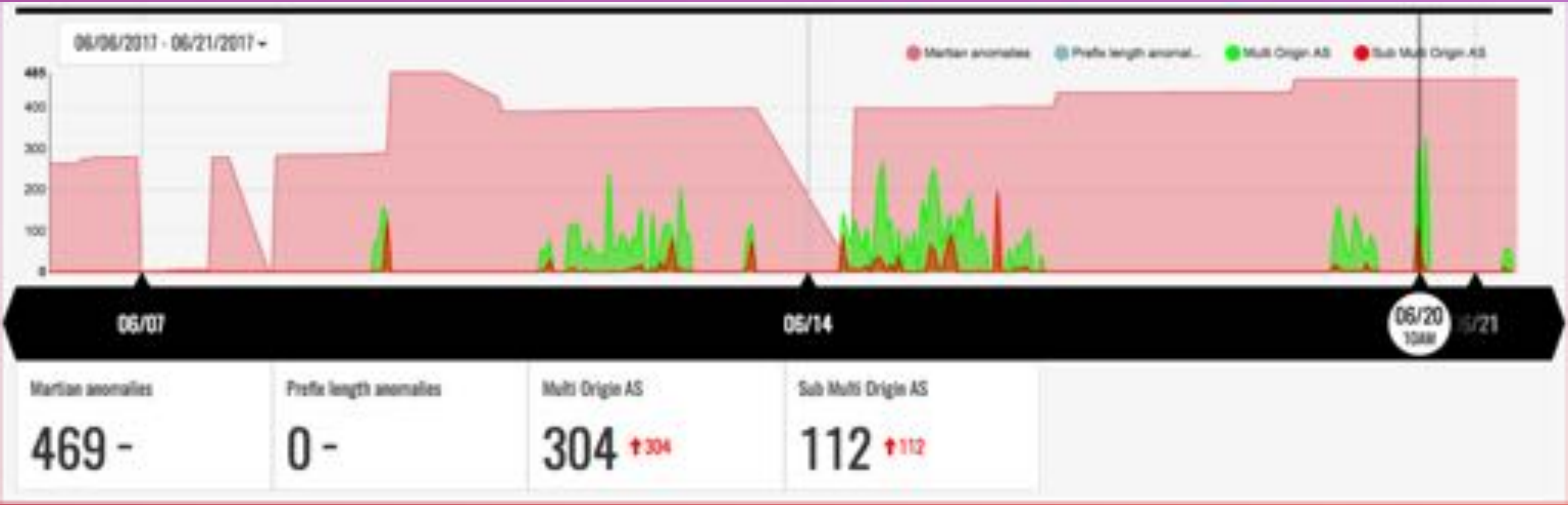
Sub-prefix injection victims – 24 hour period



Sub-prefix injection attackers – 24 hour period



Move to real-time analytics



Martian anomalies

469 -

Prefix length anomalies

0 -

Multi Origin AS

304 ↑304

Sub Multi Origin AS

112 ↑112

PNDA.io – the platform

What is PNDA?

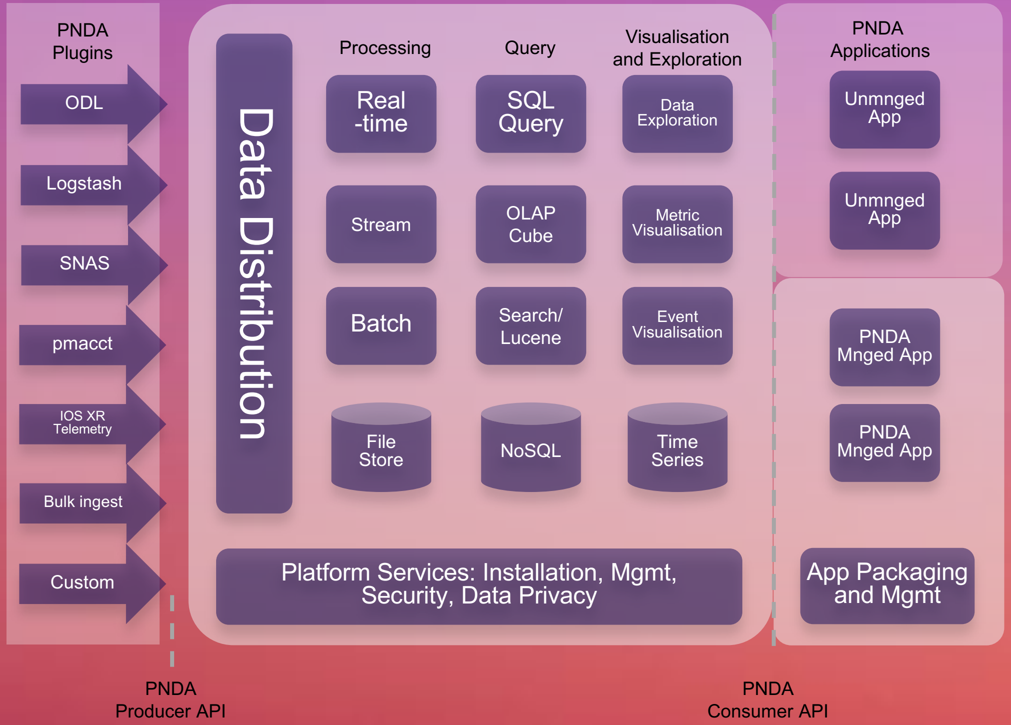
PNDA brings together a number of open source technologies to provide a simple, scalable open big data analytics Platform for Network Data Analytics

Linux Foundation Collaborative Project based on the Apache ecosystem

Where is PNDA today?

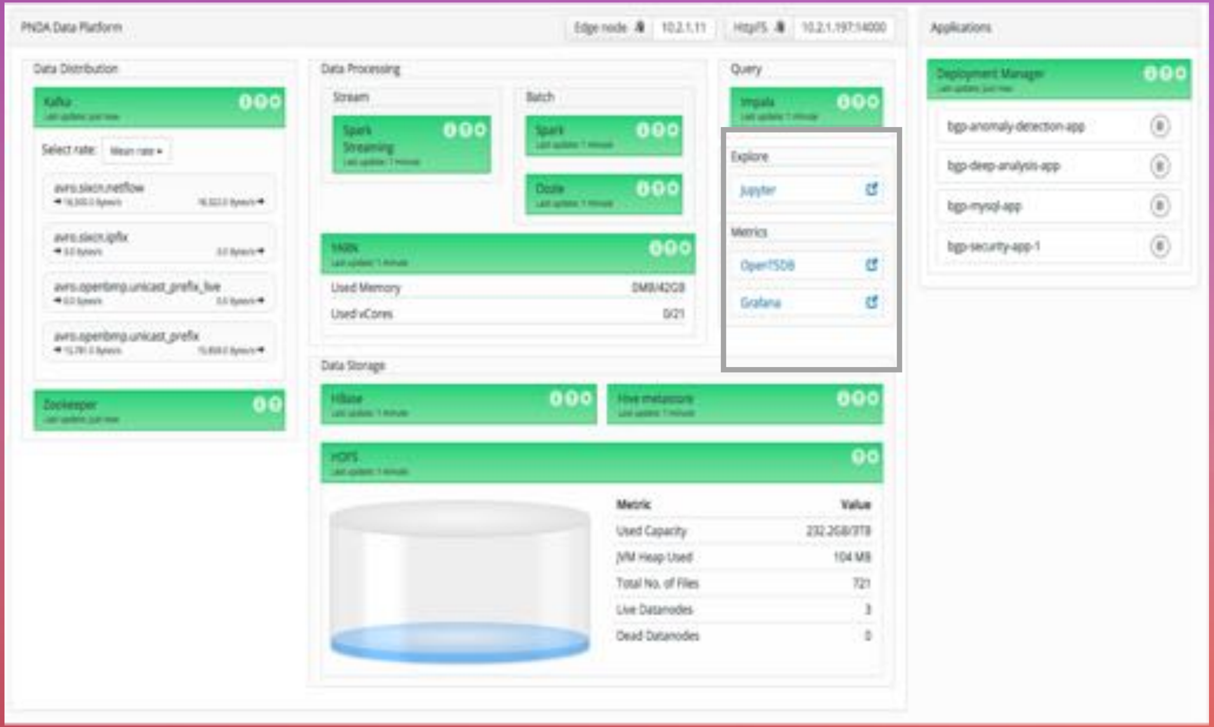
- In service trials with two Service Providers
- One platform supporting a range of use-cases including
  - Network security – Apache Spot
  - 6CN
  - Virtualization infrastructure monitoring and analysis
  - Smart Cities
  - Smart Transportation use-cases

PNDA



- Horizontally scalable platform for analytics and data processing applications
- Support for near-real-time stream processing and in-depth batch analysis on massive datasets
- Decouples data collection and aggregation from data analysis
- Consuming applications can be either platform apps developed for PNDA or client apps integrated with PNDA
- Client apps can use one of several structured query interfaces or consume streams directly.
- Leverages best current practise in big data analytics

PNDA



PNDA Data Platform

Edge node 10.2.1.11 | HttpFS 10.2.1.197:14000

Data Distribution

Kafka
Last update: just now

Select rate: Mean rate ▾

awrs.slacn.netflow
◄ 16,300.0 bytes/s 16,323.0 bytes/s ►

awrs.slacn.ipfix
◄ 0.0 bytes/s 0.0 bytes/s ►

awrs.openbmp.unicast_prefix_live
◄ 0.0 bytes/s 0.0 bytes/s ►

awrs.openbmp.unicast_prefix
◄ 15,791.0 bytes/s 15,808.0 bytes/s ►

Zookeeper
Last update: just now

Data Processing

Stream

Spark Streaming
Last update: 1 minute

Batch

Spark
Last update: 1 minute

Oozie
Last update: 1 minute

YARN
Last update: 1 minute

Used Memory	0MB/42GB
Used vCores	0/21

Query

Impala
Last update: 1 minute

Explore

Jupyter

Metrics

OpenTSDB

Grafana

Data Storage

HBase
Last update: 1 minute

Hive metastore
Last update: 1 minute

HDFS
Last update: 1 minute

Metric	Value
Used Capacity	232.2GB/3TB
JVM Heap Used	104 MB
Total No. of Files	721
Live Datanodes	3
Dead Datanodes	0

Applications

Deployment Manager
Last update: just now

bgp-anomaly-detection-app

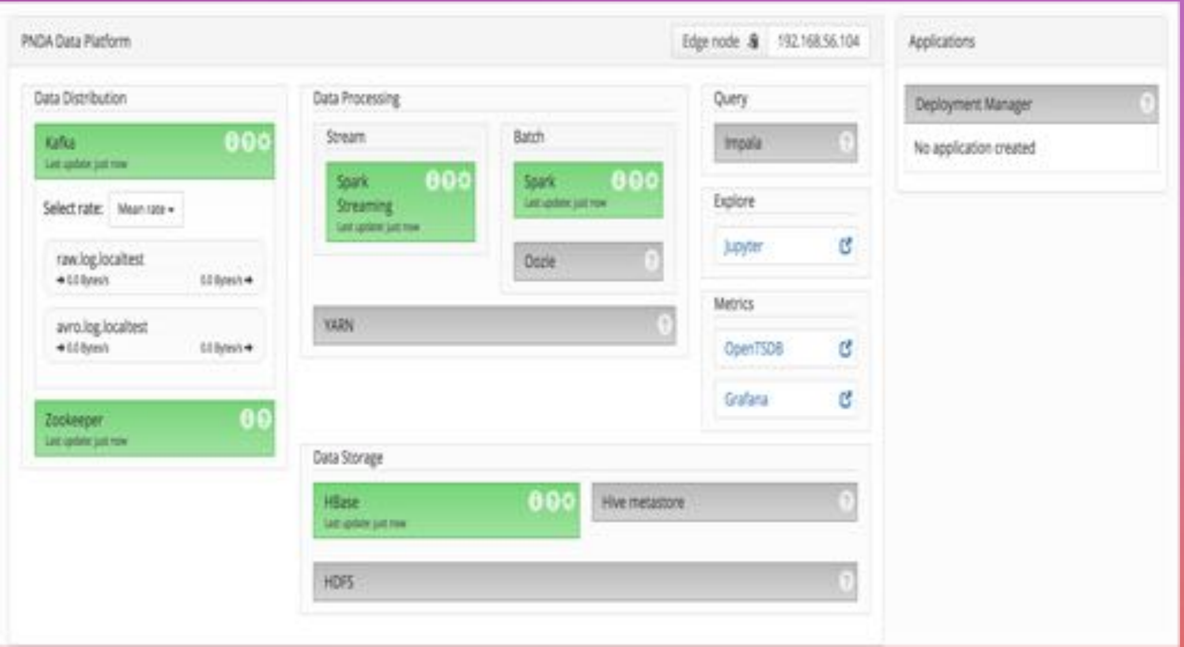
bgp-deep-analysis-app

bgp-mysql-app

bgp-security-app-1

- Simple, scalable open data platform
- Provides a common set of services for developing analytics applications
- Accelerates the process of developing big data analytics applications whilst significantly reducing the TCO
- PNDA provides a platform for convergence of network data analytics

Red-PNDA



Edge node 192.168.56.104

No application created

Select rate: Mean rate

raw.log.localtest
0.0 Bytes/s 0.0 Bytes/s

avro.log.localtest
0.0 Bytes/s 0.0 Bytes/s

Last update: just now

- A reduced set of components providing a PNDA-like environment for education and basic prototyping
- Miniature PNDA – fits your laptop
  - Lightweight simplified “Big Data” platform

Potential

What can we do with large-scale collection of historical event information?

- Event impact analysis –
  - Stability
  - Security
  - Misconfiguration
  - Forensics
- Application of ML/DL to data-set
- Pattern-detection and network ‘weather forecasting’

Where can I learn more?

- www.pnda.io
- https://github.com/pndaproject
- https://github.com/pndaproject/red-pnda
- www.snas.io

pnda.io