BCNET

CONNECT

HIGHER ED & RESEARCH TECH SUMMIT

“Zeeking” at 100 Gigabits

Who am I?

Ryan M^{c}Ronald, PMP, CISSP

- I am a Senior Information Security and Data Protection Analyst with the University of Victoria
- I work on the team which supports ‘Arbutus’ which is Canada’s largest nationally funded cloud infrastructure for computational research
- I manage the Zeek implementation monitoring our research network links

What should you expect from this talk?

Sharing Uvic's experiences on what it takes to capture traffic at 100 Gigabits with Zeek. And, once we have it what can we do with OpenSearch?

- Why monitor network traffic?
- Why is monitoring difficult at 100 Gbs?
- What is Zeek?
- What does a Zeek cluster look like?
- What does Zeek capture?
- What do we do with all that data?
- What is OpenSearch?
- What is next?

Why monitor network traffic?

Monitoring network traffic is important for many reasons including:

- Understanding your environment and what “normal” traffic looks like
- Protecting systems and data against threats
- Detecting issues as they are occurring
- Supporting investigations, incident response, etc.
- Complying with standards, regulations, insurance or contractual requirements, etc.

Bottom Line: Monitoring network traffic is important for protecting, detecting, and responding to cyber security issues

Why is monitoring difficult at 100 Gigabits?

Monitoring at 100 Gigabits per second (Gbit/s) is difficult because:

- Research networks continually push the limits of networking
- 100 Gbit/s links may carry substantial amounts of information, up to:
  - 750 Gigabytes per minute (GB/m)
  - 45,000 Gigabytes per hour (GB/h)
  - 20,000 simultaneous Netflix HD streams
- Extensive resources are needed to log that amount of information
- Commodity monitoring components typically operate up to 10 Gbit/s
- Commercial 100 Gbit/s monitoring is very expensive and well beyond the budget of most higher education/research institutions

Bottom Line: Monitoring at 100 Gbit/s is difficult due to the amount of data which must be examined in order to be effective

What is Zeek?

Zeek is free and open-source network monitoring software which:

- Originated in the mid-90’s at Lawrence Berkeley National Lab
- Operates at the service layer necessitating network layer re-assembly
- Is scaleable, configurable, scriptable, and modular
- Is supported by a number of commercial vendors
- Has community contributed packages and support forums

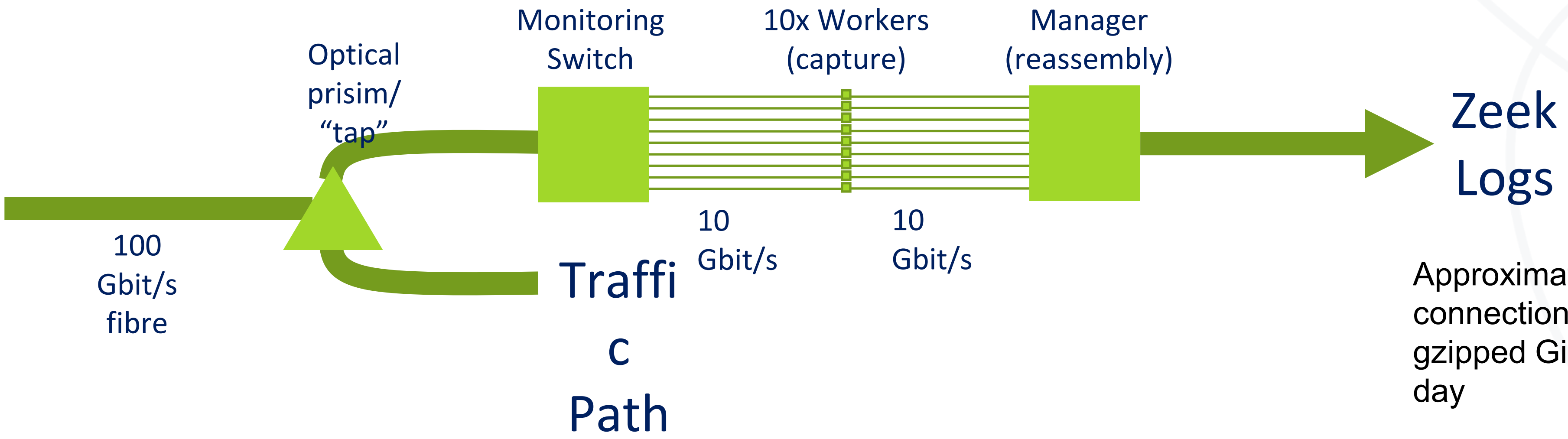
BCNET CONNECT

Bottom Line: Zeek provides passive network monitoring capable of capturing network traffic at 100 Gbit/s

What does a Zeek cluster look like?

A Zeek cluster uses multiple nodes to monitor network traffic:

- As many Zeek worker nodes as needed to perform packet capture
- One Zeek manager node to perform packet reassembly, protocol analysis and logging



Approximately 10 million lines connection lines per hour and 10 gzipped Gigabytes of logs per day

Bottom Line: Ten clustered workers capture data at 10 Gbits/s which is analyzed and reported into log files by the Zeek Manager

What does Zeek capture?

Here is a simplified example of how Zeek reports in the “conn.log”:

ts	uid	orig_h	orig_p	resp_h	resp_p	proto	service	duration	bytes	conn_state
1681901	ChP66d4j	1.2.3.4	3790	4.3.2.1	80	tcp	http	959	16340	SF

→ Every connection has a unique ID

→ Service details are logged separately

Here is a simplified example of the same connection in Zeek’s “http.log”:

ts	uid	method	host	uri	bytes	user_agent	status_code	mime_type
1681901	ChP66d4j	POST	apiserver.com	/api/status	16340	Mozilla/5.0	200	text/json

BCNET CONNECT

Bottom Line: Each network connection is logged in the “conn.log” and entries logged in service-specific files may be correlated using the “uid” in each logged entry

What does Zeek capture?

Example FIles	Contents	Example Fields (Timestamp, UID, and …)
conn.log	All network connections	Source and destination IP and ports, flags, duration, bytes
http.log	HTTP protocol request	HTTP directive, URL, query string, status code, content type
ssl.log	SSL protocol details	SSL/TLS version, cipher, server name, status
X509.log	X509 certificate details	Fingerprint, serial, subject, issuer, validity dates, signature type
ssh.log	SSH protocol details	Version, client string, server string, cipher, host_key fingerprint
files.log	Transferred files details	MIME type, filename, bytes, SHA hash
dns.log	DNS protocol details	Query string, Rcode, answers
ntp.log	NTP protocol details	Version, stratum, precision, various times
weird.log	Protocol anomalies	Unknown methods, flag anomalies, split routing
known_services.log	Last hour services	IP, port, service
known_hosts.log	Last hour hosts	IP addresses
known_certificates.log	Last hour certificates	IP, port, subject, serial

Bottom Line: There are a wide variety of service-specific files which contain useful information about how those services are being used

What do we do with all that data?

Using the Zeek logs we can:

- Understand what connections occurred:
  - “normal” traffic
  - “top talkers”
  - port scans
  - “botnet” activity
- Unencrypted service-level analysis:
  - Detailed specifics tailored to those protocols
- Encrypted service-level analysis:
  - SSH protocol details such as version, client string, server string, cipher
  - SSL/TLS details such as SSL/TLS version, cipher

Bottom Line: Zeek generates large amounts of log data which needs to be ingested into a repository and accessed by data analysis and visualization tools

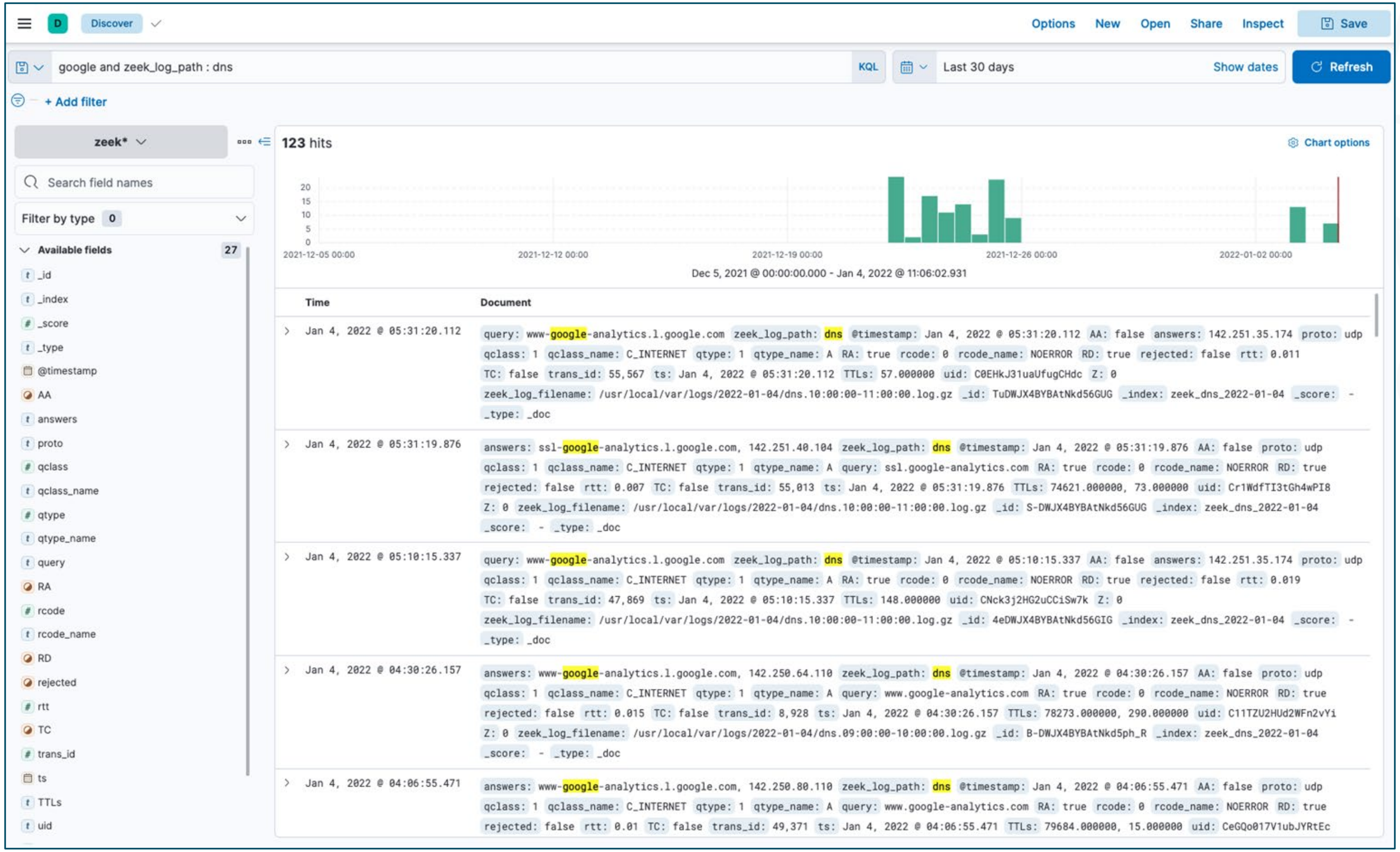
What is OpenSearch?

OpenSearch is free and open-source software which:

- Originated when Amazon forked open-source components of:
  - ElasticSearch forked into OpenSearch
  - Kibana forked into OpenSearch Dashboards
- OpenSearch is a scaleable storage and search software
- OpenSearch Dashboards is data visualization and analysis software
- Active development funded by Amazon
- Community support model

Bottom Line: OpenSearch tools provide search, analysis, and visualization capabilities which are essential to network monitoring and investigations

What is OpenSearch?



google and zeek_log_path : dns

Last 30 days

123 hits

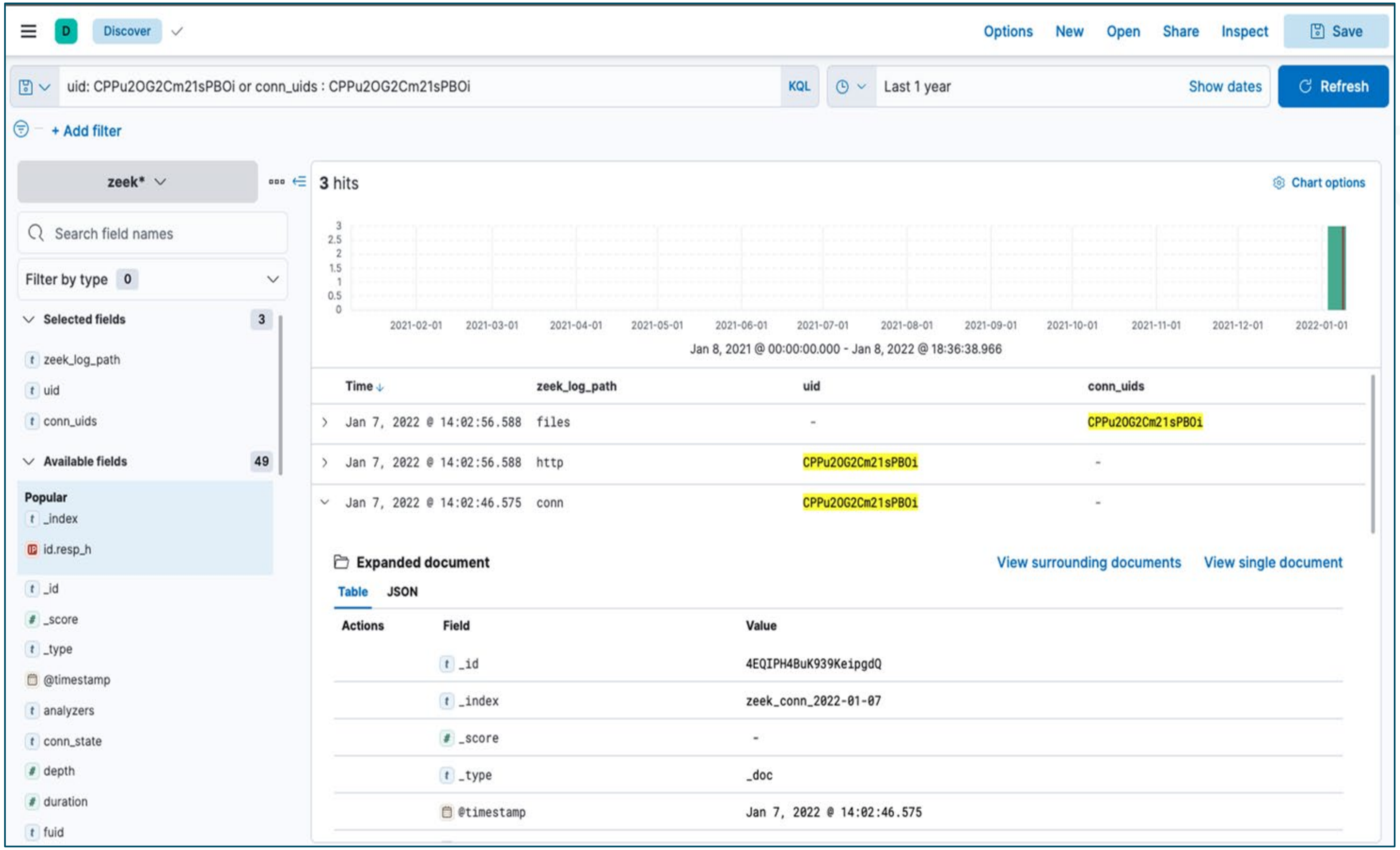
Dec 5, 2021 @ 00:00:00.000 - Jan 4, 2022 @ 11:06:02.931

```
Time                        Document
> Jan 4, 2022 @ 05:31:20.112   query: www-google-analytics.l.google.com zeek_log_path: dns @timestamp: Jan 4, 2022 @ 05:31:20.112 AA: false answers: 142.251.35.174 proto: udp
                               qclass: 1 qclass_name: C_INTERNET qtype: 1 qtype_name: A RA: true rcode: 0 rcode_name: NOERROR RD: true rejected: false rtt: 0.011
                               TC: false trans_id: 55,567 ts: Jan 4, 2022 @ 05:31:20.112 TTLs: 57.000000 uid: C0EHkJ31uaUfugCHdc Z: 0
                               zeek_log_filename: /usr/local/var/logs/2022-01-04/dns.10:00:00-11:00:00.log.gz _id: TuDWJX4BYBAtNkd56GUG _index: zeek_dns_2022-01-04 _score: -
                               _type: _doc
> Jan 4, 2022 @ 05:31:19.876   answers: ssl-google-analytics.l.google.com, 142.251.40.104 zeek_log_path: dns @timestamp: Jan 4, 2022 @ 05:31:19.876 AA: false proto: udp
                               qclass: 1 qclass_name: C_INTERNET qtype: 1 qtype_name: A query: ssl.google-analytics.com RA: true rcode: 0 rcode_name: NOERROR RD: true
                               rejected: false rtt: 0.007 TC: false trans_id: 55,013 ts: Jan 4, 2022 @ 05:31:19.876 TTLs: 74621.000000, 73.000000 uid: Cr1WdfTI3tGh4wPI8
                               Z: 0 zeek_log_filename: /usr/local/var/logs/2022-01-04/dns.10:00:00-11:00:00.log.gz _id: S-DWJX4BYBAtNkd56GUG _index: zeek_dns_2022-01-04
                               _score: - _type: _doc
> Jan 4, 2022 @ 05:10:15.337   query: www-google-analytics.l.google.com zeek_log_path: dns @timestamp: Jan 4, 2022 @ 05:10:15.337 AA: false answers: 142.251.35.174 proto: udp
                               qclass: 1 qclass_name: C_INTERNET qtype: 1 qtype_name: A RA: true rcode: 0 rcode_name: NOERROR RD: true rejected: false rtt: 0.019
                               TC: false trans_id: 47,869 ts: Jan 4, 2022 @ 05:10:15.337 TTLs: 148.000000 uid: CNck3j2HG2uCCiSw7k Z: 0
                               zeek_log_filename: /usr/local/var/logs/2022-01-04/dns.10:00:00-11:00:00.log.gz _id: 4eDWJX4BYBAtNkd56GIG _index: zeek_dns_2022-01-04 _score: -
                               _type: _doc
> Jan 4, 2022 @ 04:30:26.157   answers: www-google-analytics.l.google.com, 142.250.64.110 zeek_log_path: dns @timestamp: Jan 4, 2022 @ 04:30:26.157 AA: false proto: udp
                               qclass: 1 qclass_name: C_INTERNET qtype: 1 qtype_name: A query: www.google-analytics.com RA: true rcode: 0 rcode_name: NOERROR RD: true
                               rejected: false rtt: 0.015 TC: false trans_id: 8,928 ts: Jan 4, 2022 @ 04:30:26.157 TTLs: 78273.000000, 290.000000 uid: C11TZU2HUd2WFn2vYi
                               Z: 0 zeek_log_filename: /usr/local/var/logs/2022-01-04/dns.09:00:00-10:00:00.log.gz _id: B-DWJX4BYBAtNkd5ph_R _index: zeek_dns_2022-01-04
                               _score: - _type: _doc
> Jan 4, 2022 @ 04:06:55.471   answers: www-google-analytics.l.google.com, 142.250.80.110 zeek_log_path: dns @timestamp: Jan 4, 2022 @ 04:06:55.471 AA: false proto: udp
                               qclass: 1 qclass_name: C_INTERNET qtype: 1 qtype_name: A query: www.google-analytics.com RA: true rcode: 0 rcode_name: NOERROR RD: true
                               rejected: false rtt: 0.01 TC: false trans_id: 49,371 ts: Jan 4, 2022 @ 04:06:55.471 TTLs: 79684.000000, 15.000000 uid: CeGQo017V1ubJYRtEc
```

BCNET CONNECT

Bottom Line: In OpenSearch, Zeek log entries appear as ‘documents’ which can be queried and viewed through the OpenSearch Dashboards visualization interface shown here

What is OpenSearch?



uid: CPPu2OG2Cm21sPBOi or conn_uids : CPPu2OG2Cm21sPBOi

3 hits

Jan 8, 2021 @ 00:00:00.000 - Jan 8, 2022 @ 18:36:38.966

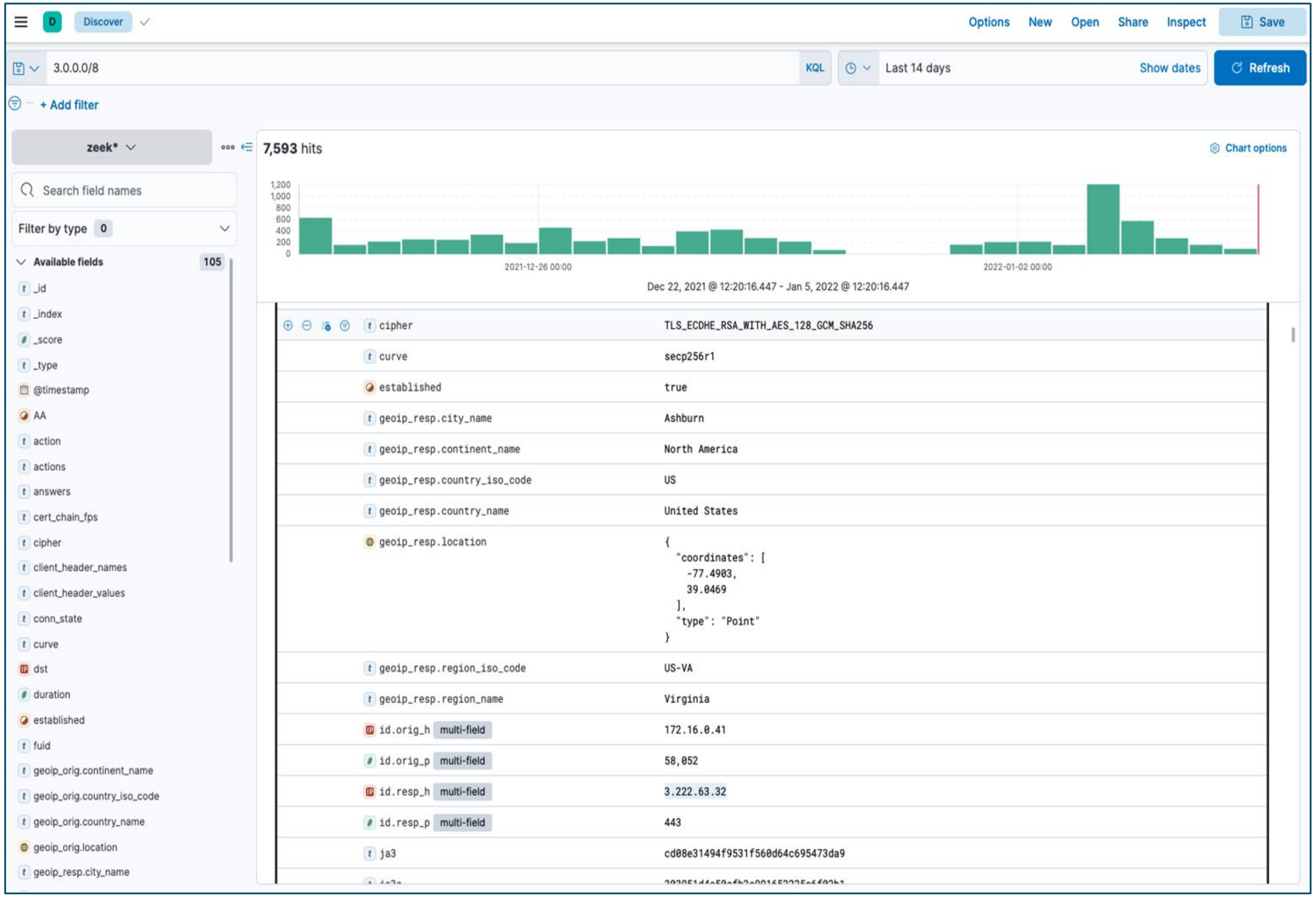
Time ↓	zeek_log_path	uid	conn_uids
Jan 7, 2022 @ 14:02:56.588	files	-	CPPu2OG2Cm21sPBOi
Jan 7, 2022 @ 14:02:56.588	http	CPPu2OG2Cm21sPBOi	-
Jan 7, 2022 @ 14:02:46.575	conn	CPPu2OG2Cm21sPBOi	-

Expanded document

Actions	Field	Value
	_id	4EQIPH4BuK939KeipgdQ
	_index	zeek_conn_2022-01-07
	_score	-
	_type	_doc
	@timestamp	Jan 7, 2022 @ 14:02:46.575

Bottom Line: A connection appears in multiple ‘documents’ identified by the UID and can be viewed in the visualization interface

What is OpenSearch?



3.0.0.0/8

Last 14 days

7,593 hits

Dec 22, 2021 @ 12:20:16.447 - Jan 5, 2022 @ 12:20:16.447

Available fields 105

- _id
- _index
- _score
- _type
- @timestamp
- AA
- action
- actions
- answers
- cert_chain_fps
- cipher
- client_header_names
- client_header_values
- conn_state
- curve
- dst
- duration
- established
- fuid
- geoip_orig.continent_name
- geoip_orig.country_iso_code
- geoip_orig.country_name
- geoip_orig.location
- geoip_resp.city_name

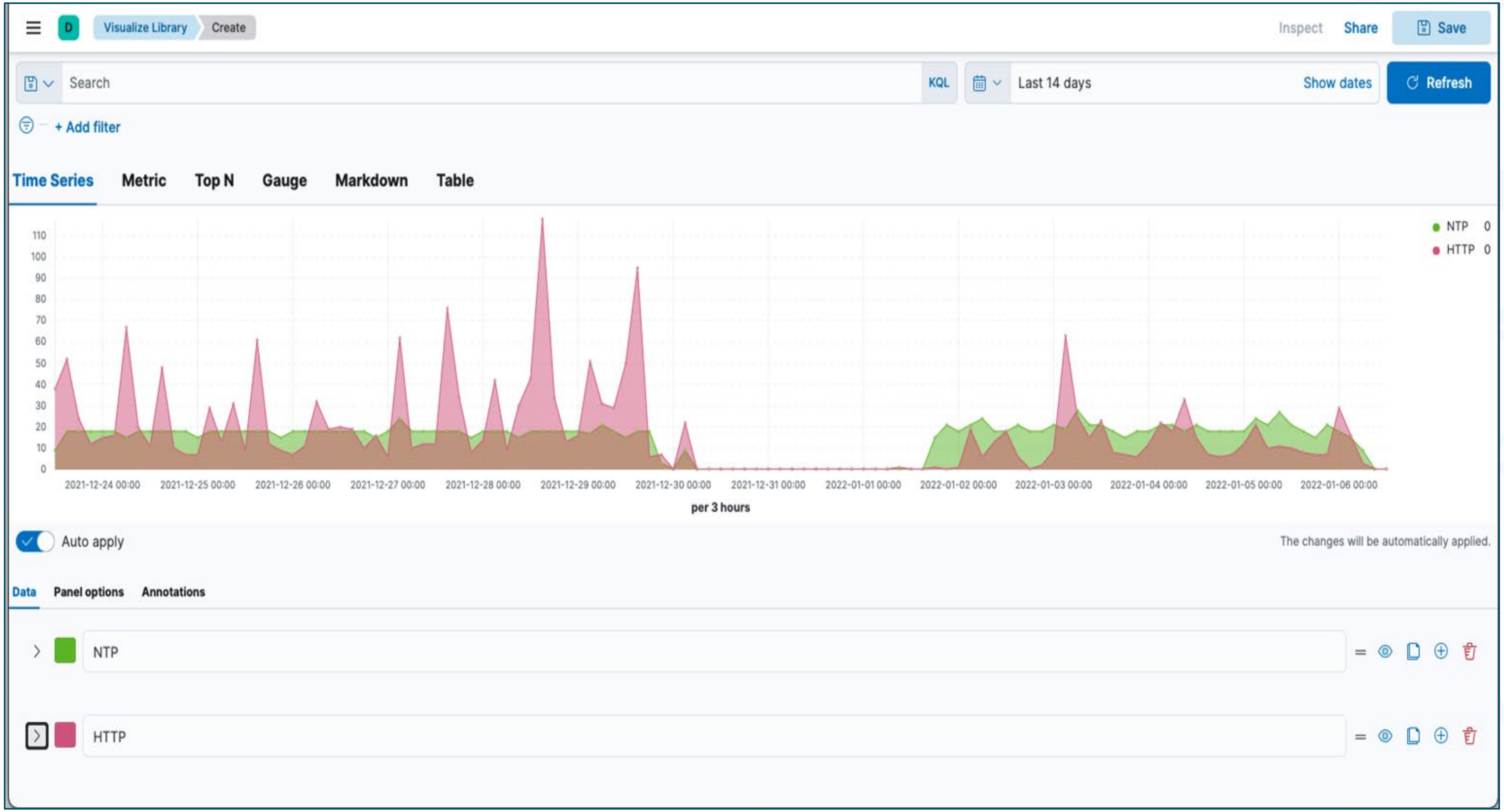
cipher	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256
curve	secp256r1
established	true
geoip_resp.city_name	Ashburn
geoip_resp.continent_name	North America
geoip_resp.country_iso_code	US
geoip_resp.country_name	United States

geoip_resp.location	{
	"coordinates": [
	-77.4903,
	39.0469
	],
	"type": "Point"
	}

geoip_resp.region_iso_code	US-VA
geoip_resp.region_name	Virginia
id.orig_h multi-field	172.16.0.41
id.orig_p multi-field	58,052
id.resp_h multi-field	3.222.63.32
id.resp_p multi-field	443
ja3	cd08e31494f9531f560d64c695473da9

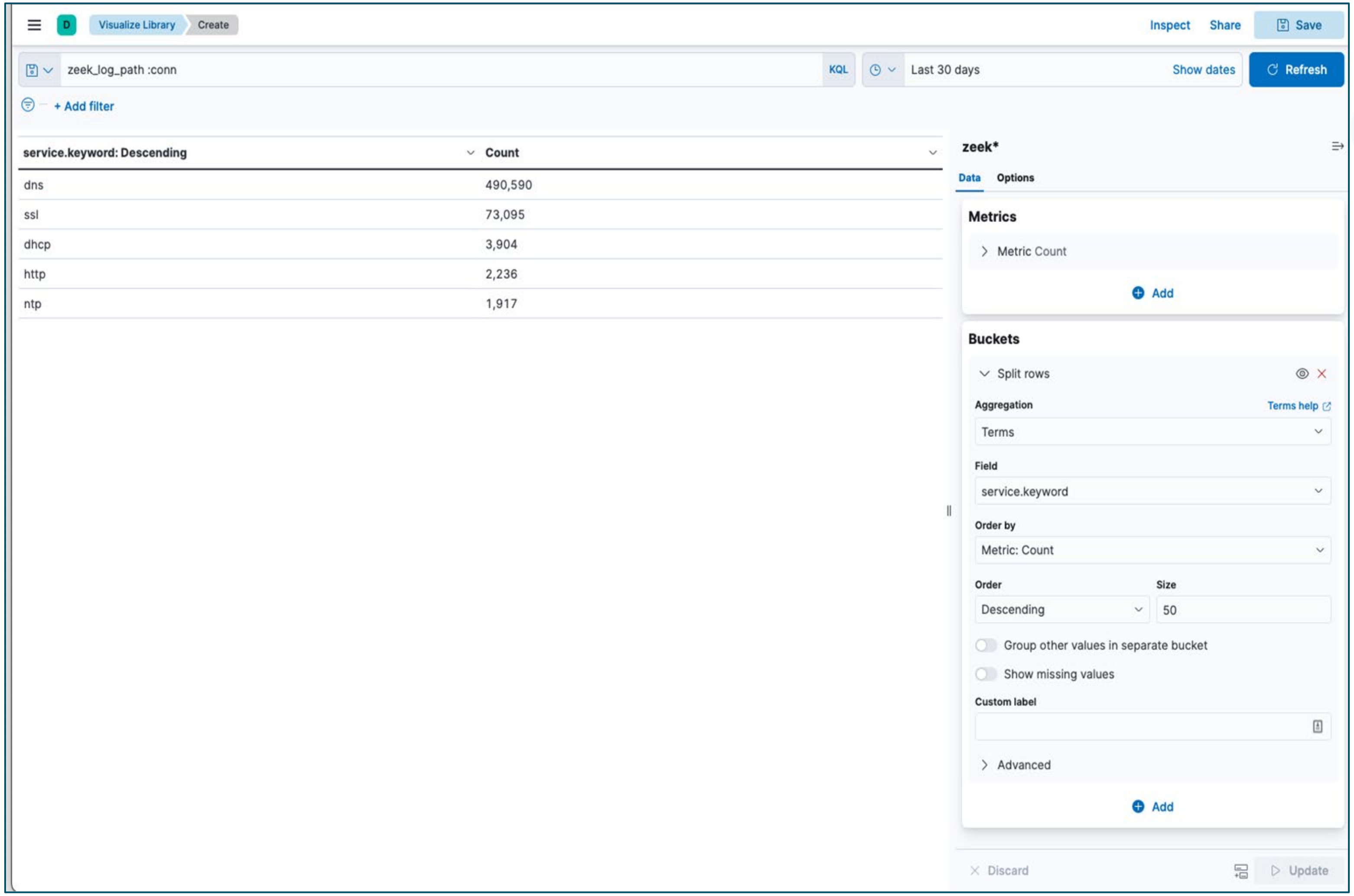
Bottom Line: Using the visualization interface, searching by subnet using CIDR notation can be accomplished using the 'addr' type

What is OpenSearch?



Bottom Line: Using the visualization interface, various time series visualizations can be created to view how services are used

What is OpenSearch?



zeek_log_path :conn

Last 30 days

service.keyword: Descending	Count
dns	490,590
ssl	73,095
dhcp	3,904
http	2,236
ntp	1,917

zeek*

Data Options

Metrics

Metric Count

Add

Buckets

Split rows

Aggregation: Terms

Terms help

Field: service.keyword

Order by: Metric: Count

Order: Descending

Size: 50

○ Group other values in separate bucket

○ Show missing values

Custom label

Advanced

Add

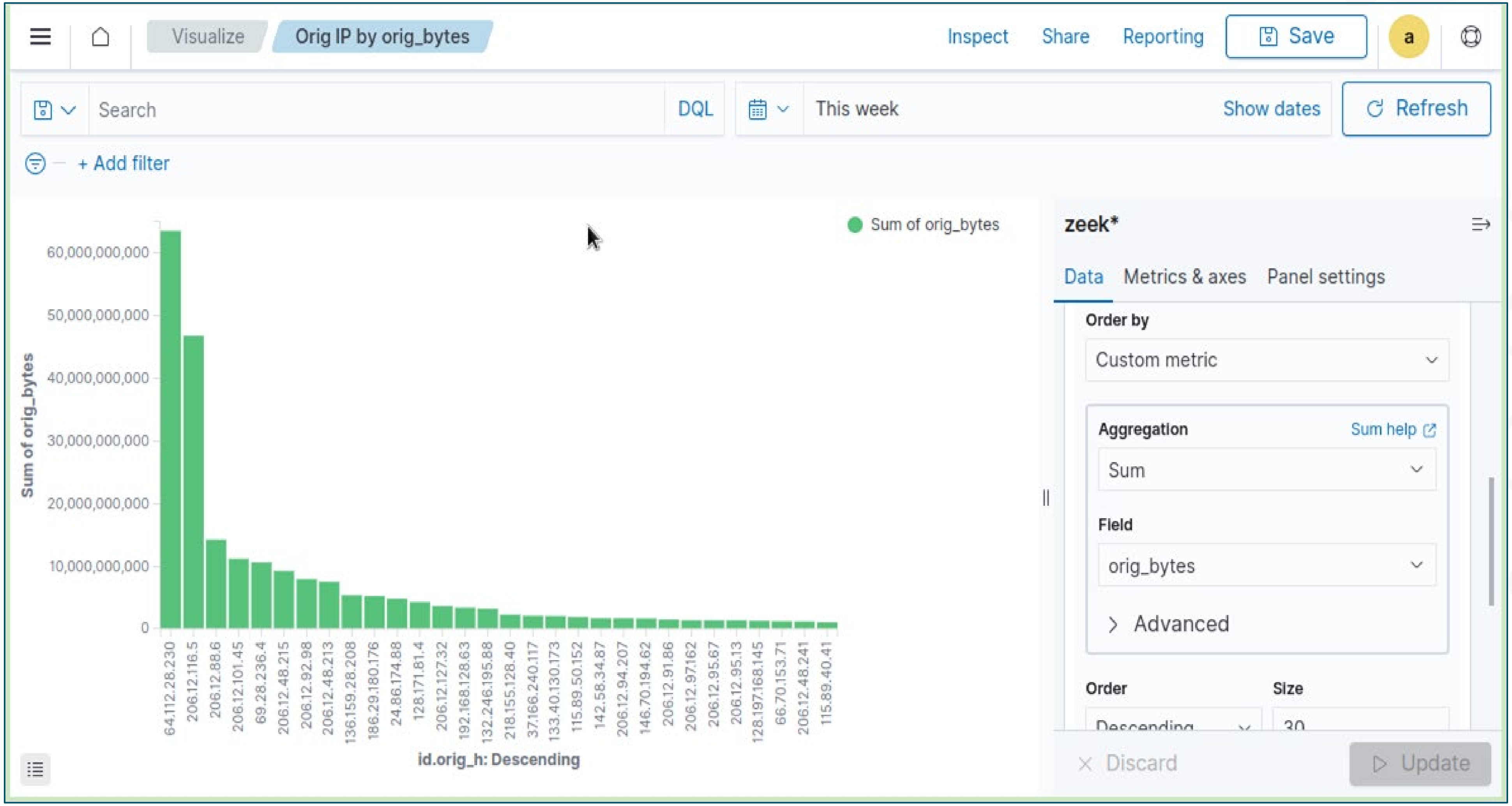
Discard

Update

BCNET CONNECT

Bottom Line: Using the visualization interface, ‘aggregations’ can be used to show service utilization in a summary form

What is OpenSearch?



Order by: Custom metric

Aggregation: Sum

Field: orig_bytes

Order: Descending

Size: 30

Bottom Line: Using the visualization interface, complex ‘aggregations’ go beyond counting the number of matching documents

What is next?

Future considerations for network monitoring include:

Future features	Description
Additional visualizations and dashboards	Design and develop further visualizations and dashboards for analyzing common scenarios
Consolidation with other logging and alerting functions	Ability to send alerts to other logging destinations (e.g. syslog) and export to Security Incident and Event Monitoring (SIEM)
Proactive notifications when issues occur	Ability to notify operational staff when potential issues occur so that they can be actioned on a timely basis
400 Gbit/s monitoring	The routers to upgrade to 400 Gbit/s have arrived and will be one of the first campuses in Canada to be connected at 400 Gbit/s.

Bottom Line: There is work to be done to enhance our network monitoring capabilities and to prepare for 400 Gbit/s networking

Questions?