Cloud-Scale BGP and NetFlow Analysis

Jim Frey, VP Product, Kentik Technologies

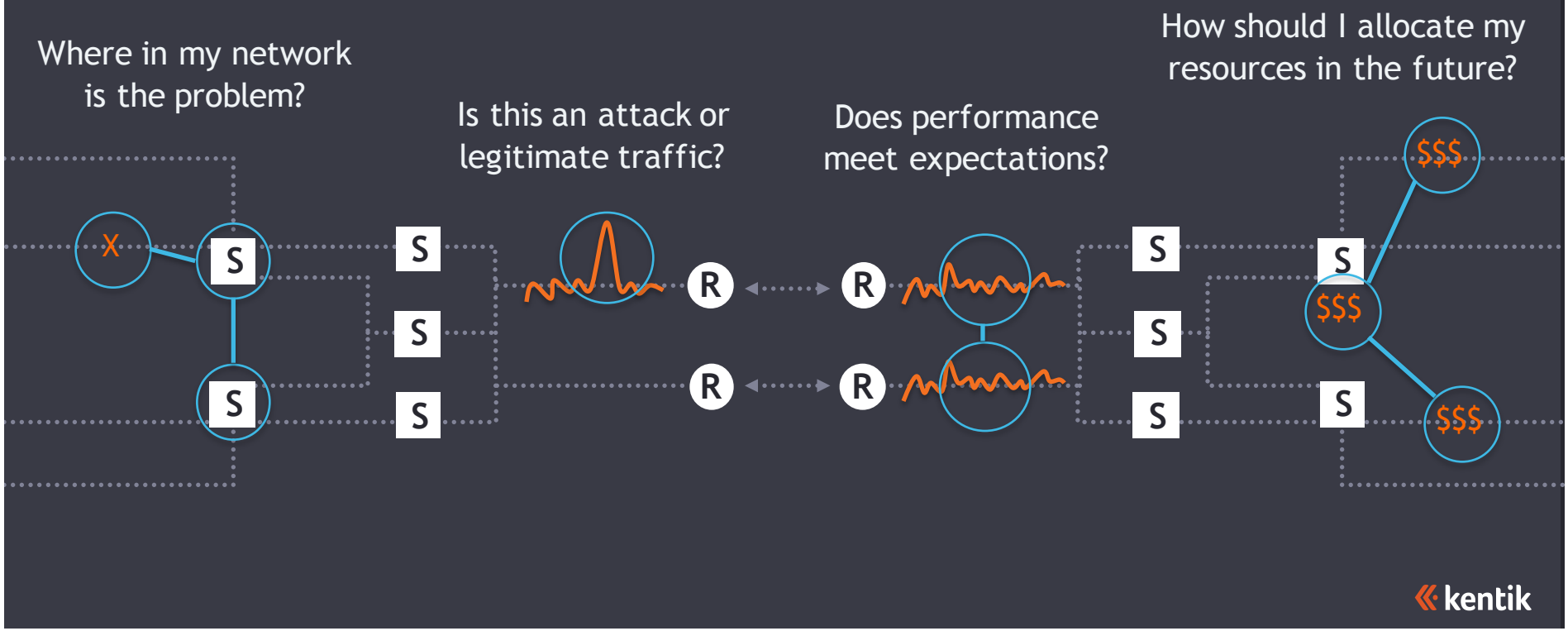
December 15, 2015

Agenda

- Common NetOps Stress points
- Helpful Data Sets – NetFlow, BGP
- Handling NetFlow and BGP at Cloud Scale
- Kentik's Approach
- Wrap-Up / Q&A

NetOps Stress Points: Needing *Instant* Answers

Things You Need Answers to About/From Your Network



kentik

What We Hear….

To Address These Questions, NetOps Needs:

- Accurate Visibility, Without Delay
- Relevant Alerts: No False Positives or Negatives
- Complete Data: Breadth + Depth
- Fast/Flexible Data Exploration
- Tools that don't suck (time or $$)

What Data Sets Can Help?

And which ones can do the job cost effectively?

Primary Network Monitoring Data Choices

Polled Stats

Examples
- SNMP, WMI

Advantages
- Ubiquitous
- Good for monitoring device health/status/activity

Disadvantages
- No traffic detail
- Typically no frequent than every 5 minutes truly anti-real-time

Flow Records

Examples
- NetFlow, sFlow, IPFIX

Advantages
- Details on traffic src/dest/content, etc.
- Very cost effective

Disadvantages
- NRT (near real-time) at best
- Incomplete app-layer detail
- Limited performance metrics
- Data volumes can be massive

Packet Inspection

Examples
- Packets -> xFlow
- Long term stream-to-disk

Advantages
- Most complete app layer detail
- True real-time (millisecond lvl)
- Complete vendor independent

Disadvantages
- Expensive to deploy at scale
- Requires network tap or SPAN
- Packet captures can be massive

Secondary Network Monitoring Data Choices

Log Records

Examples
- Syslog

Advantages
- Continuous/streaming
- Unique, device-specific info
- True real-time

Disadvantages
- No standards – must have very flexible search/mapping tools
- Data volumes can be massive

Routing/Path Data

Examples
- OSPF, IGRP, BGP

Advantages
- Details on traffic paths and provider volumes
- Insights into Internet factors

Disadvantages
- Address data only – no awareness of traffic
- Must peer with routers to get updates

Synthetic Agents

Examples
- IP SLA, Independent test sw

Advantages
- Assess functions/services 24x7
- Provides both availability and performance measures

Disadvantages
- Deploying/maintaining enough agents to achieve full coverage
- Only an approximation of real user experience (at best)

Key Assertion:

Use Multiple Data Types for Best Results

- You never know which data set will present the specific insights you need
- The challenge (real magic) comes from correlating multiple datasets, i.e.:
  - Behavioral observations with configuration changes
  - Trends with underlying traffic details
  - Routing data with traffic data

Why Correlate Routing Data with Traffic Data?

For Providers

- Recognizing new service opportunities based on subscriber (and peer) behavior
- Optimizing peering relationships for cost control

For Web Services / Commerce

- Recognizing where your customers are and how they reach you
- Managing peering relationships for best customer experience

For Enterprise

- Assessing how your connectivity providers perform/compare
- Building Internet IQ – how you connect/relate to the outside world

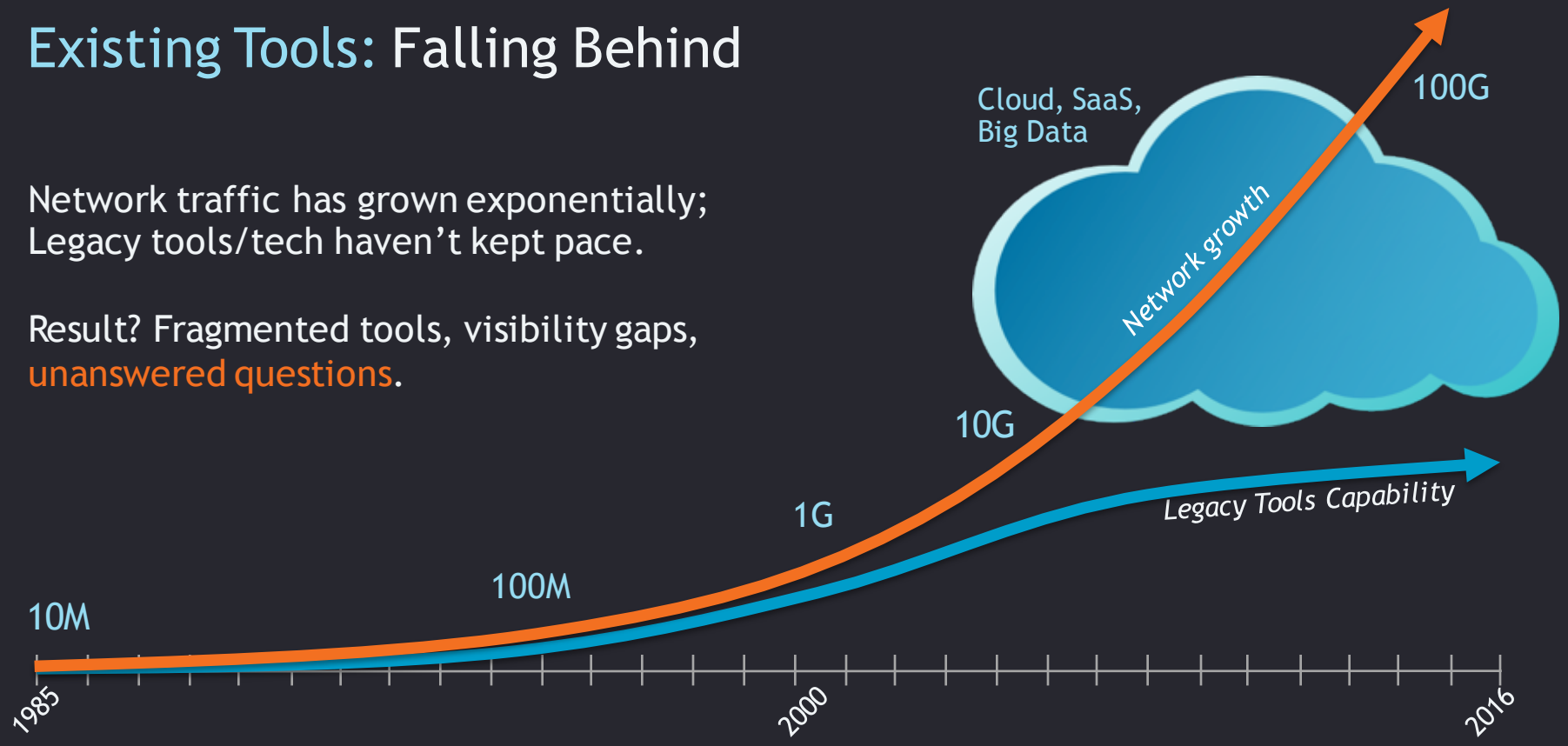
Cloud Scale for NetFlow and BGP: The Big Data Challenge

Why can't we just use our existing tools?

Existing Tools: Falling Behind

Network traffic has grown exponentially;
Legacy tools/tech haven't kept pace.

Result? Fragmented tools, visibility gaps,
unanswered questions.



kentik

Why Big Data?

- Network Monitoring Data *IS* Big Data
  - Meets Volume/Variety/Velocity Test
  - Billions of records/day (millions/second)
- Big Data architectures are considered best practices today for open/flexible correlation, analytics

Why Big Data?

- Network Monitoring Data *IS* Big Data
  - Meets Volume/Variety/Velocity Test
  - Billions of records/day (millions/second)
- Big Data architectures are considered best practices today for open/flexible correlation, analytics

Specific Challenges For NetFlow + BGP

Existing solutions shortfalls:

- Flexibility for moving between viewpoints and into full details
- Data Completeness due to reliance on summarized/aggregated flow data
- Speed: Generating new analysis in a timely manner

How to Get/Use Big Data Approach?

How to Get/Use Big Data Approach?

1. BYO – Build Your Own

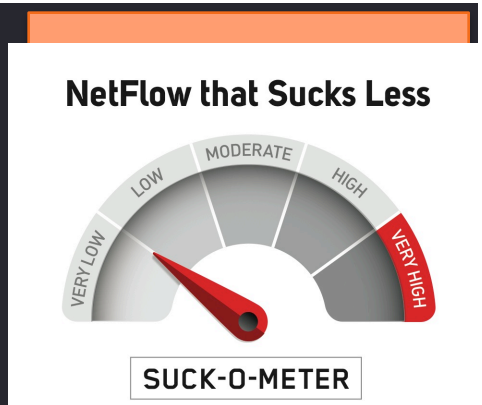
- Pick back end & reporting/analysis tools (open source = free?)
- Procure operating platforms (hard, virtual, or cloud servers = $$)
- Integrate, add data sources, and get it up and running (dev = $$)
- Keep it up and running (ops/admin = $$)

How to Get/Use Big Data Approach?

1. BYO – Build Your Own
   - Pick back end & reporting/analysis tools (open source = free?)
   - Procure operating platforms (hard, virtual, or cloud servers = $$)
   - Integrate, add data sources, and get it up and running (dev = $$)
   - Keep it up and running (ops/admin = $$)
2. Let SOMEONE ELSE build/optimize/operate
   - Subscribe to SaaS (ops $$)
   - Just Send Your Data and enjoy the ride!

Kentik's Answer

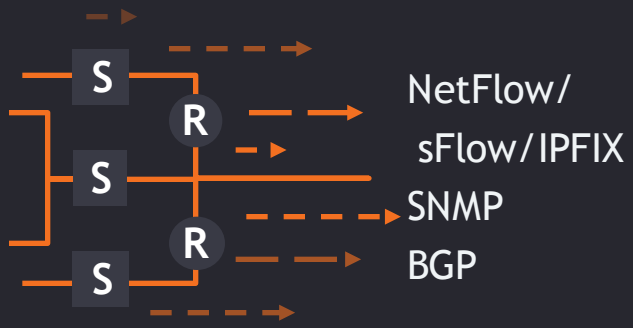
How we address the Big Data challenge to meet the needs of Network Operators now



kentik

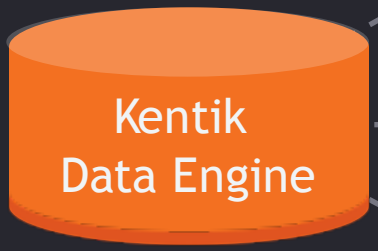
Kentik Detect: the first and only SaaS Solution

For Network Ops Management & Visibility at Terabit Scale

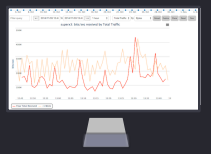


NetFlow/
sFlow/IPFIX
SNMP
BGP

The Network is
the Sensor



Big Data Network
Telemetry Platform



Web Portal
Real-time & historical
queries

Alerts
E-mail / Syslog / JSON

Open API
SQL / RESTful

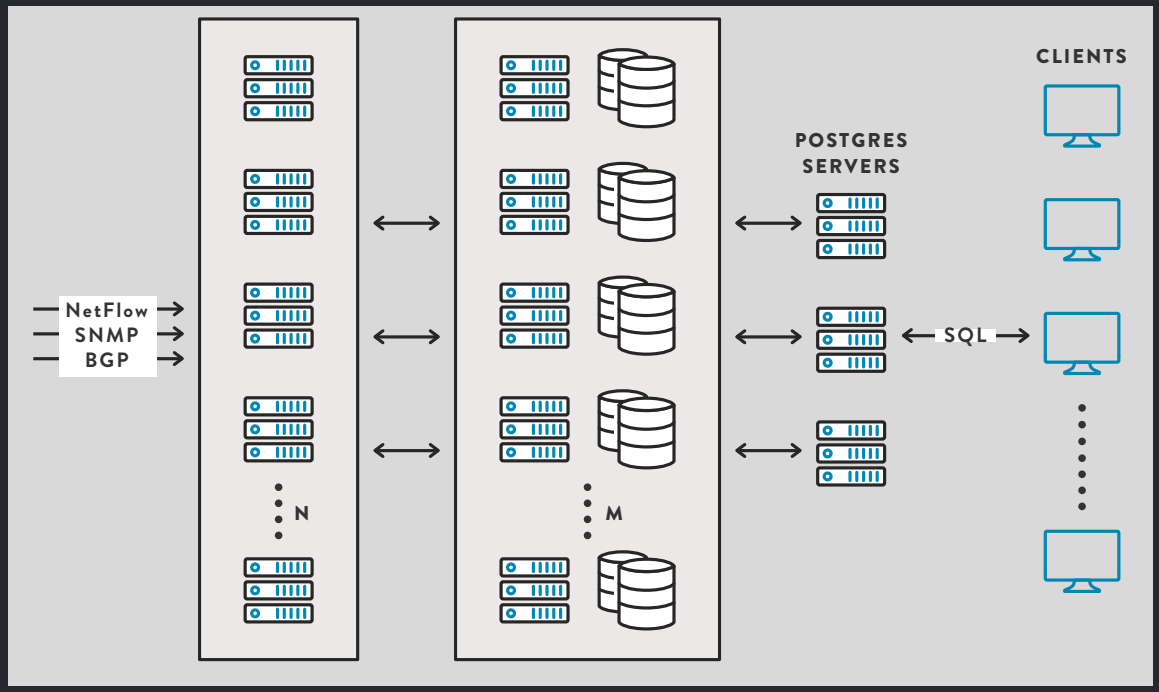
Analyze & Take Action

CLOUD-BASED | REAL-TIME | MULTI-TENANT | OPEN | GLOBAL

kentik

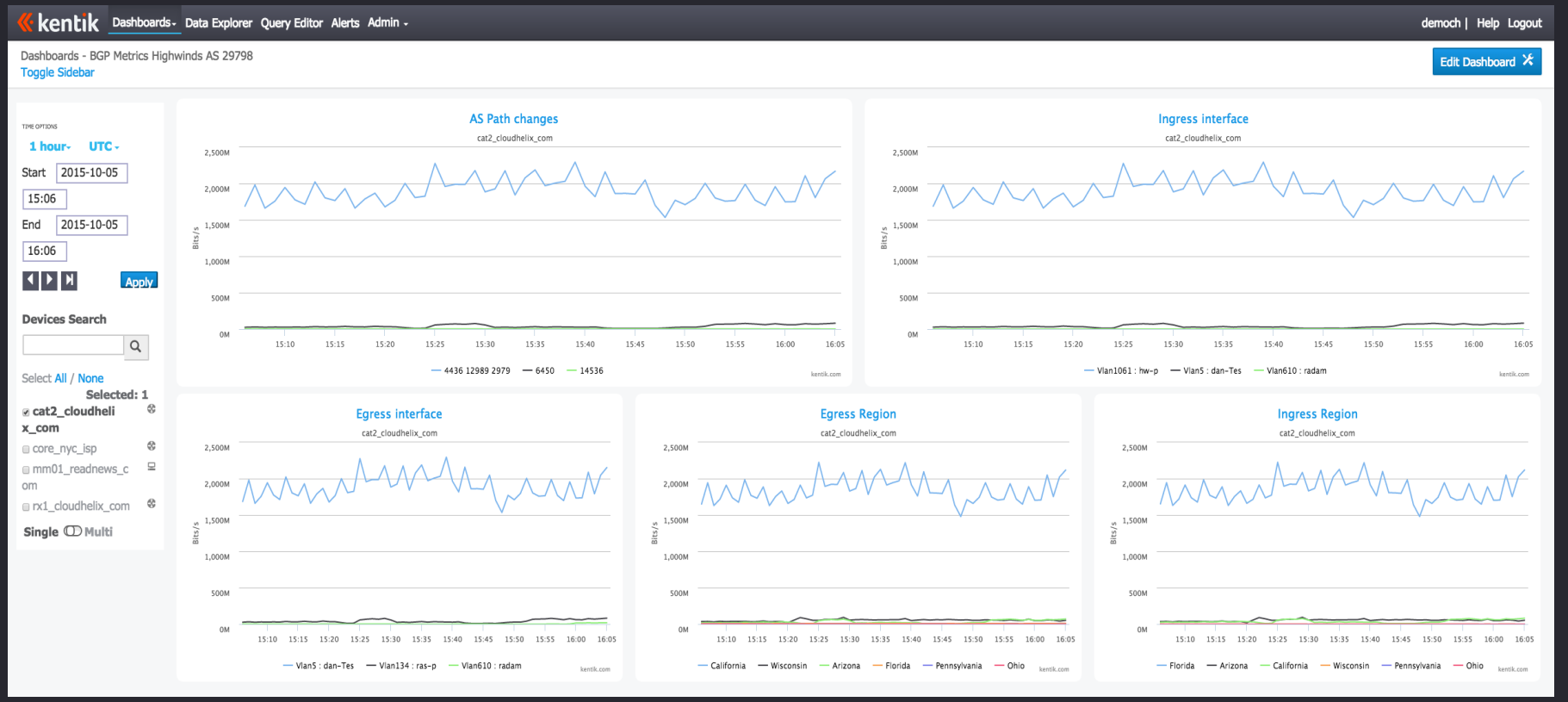
What's Behind the Kentik Data Engine

Multi-tiered/Clustered for Scale / Load Balancing / HA, Hosted by Kentik



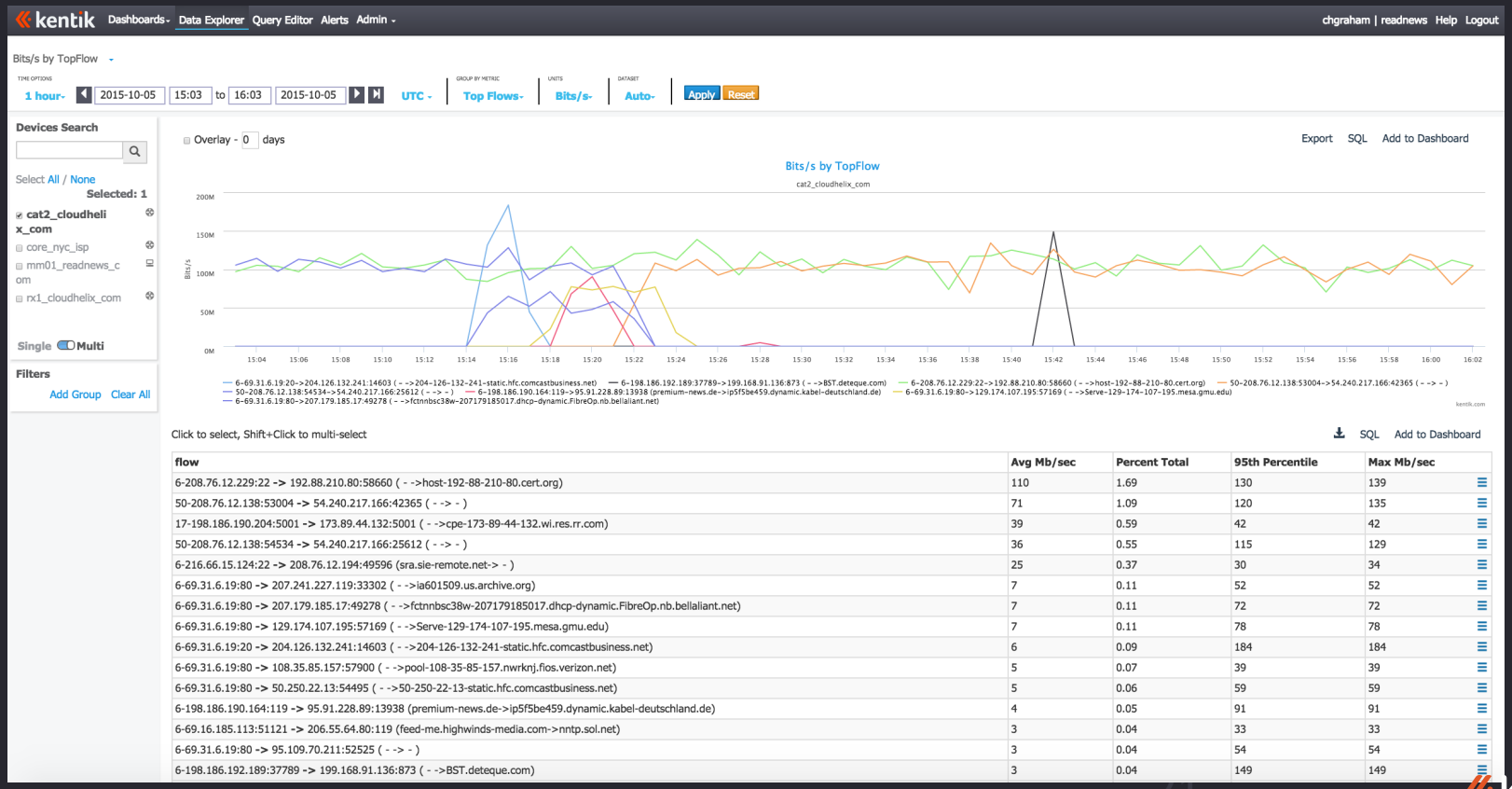
Optimized for Massive Data Ingest & Rapid Query Response

Kentik Portal Dashboard



kentik

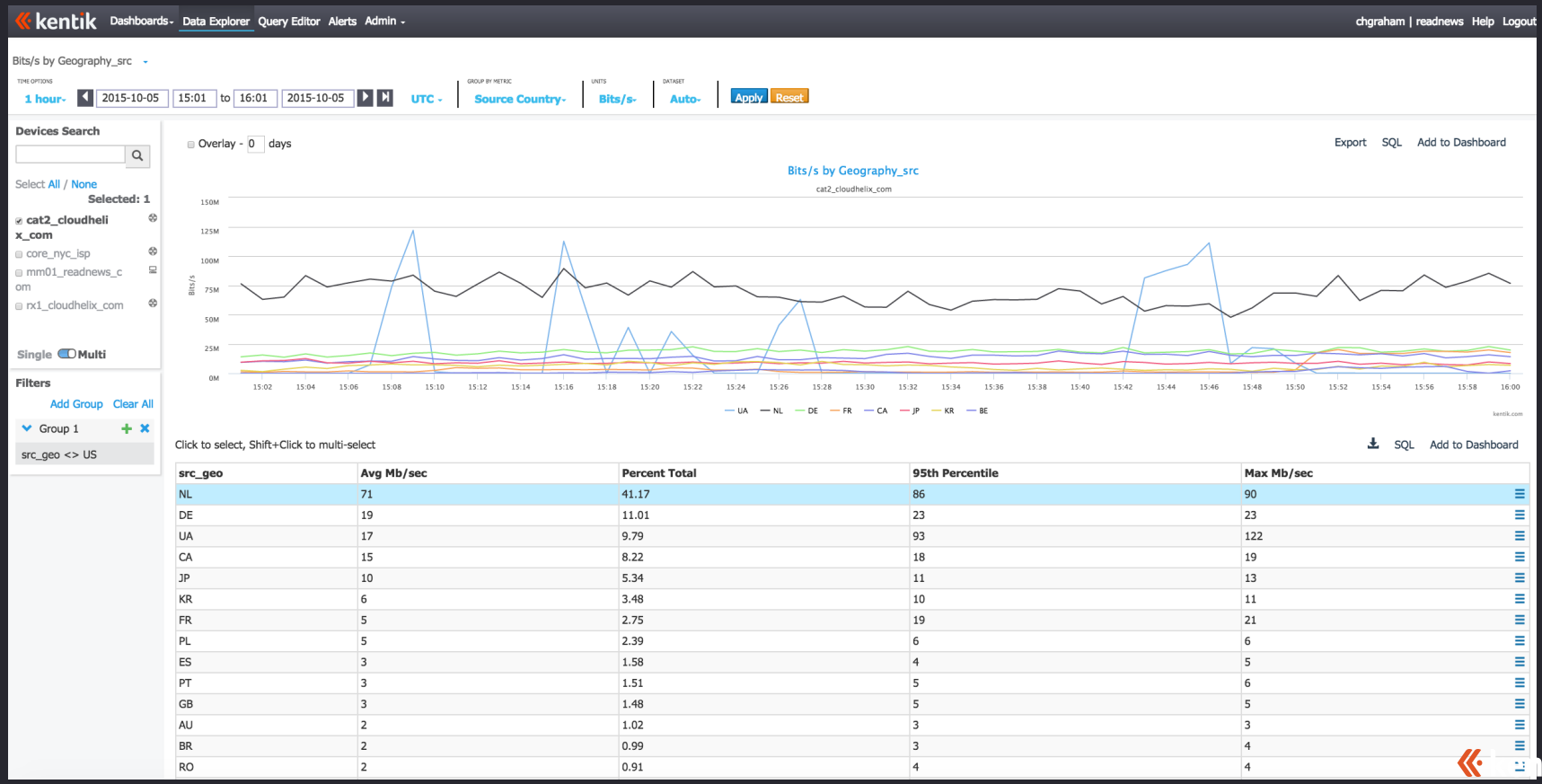
Top Traffic Flows



flow	Avg Mb/sec	Percent Total	95th Percentile	Max Mb/sec
6-208.76.12.229:22 -> 192.88.210.80:58660 (- ->host-192-88-210-80.cert.org)	110	1.69	130	139
50-208.76.12.138:53004 -> 54.240.217.166:42365 (- -> -)	71	1.09	120	135
17-198.186.190.204:5001 -> 173.89.44.132:5001 (- ->cpe-173-89-44-132.wi.res.rr.com)	39	0.59	42	42
50-208.76.12.138:54534 -> 54.240.217.166:25612 (- -> -)	36	0.55	115	129
6-216.66.15.124:22 -> 208.76.12.194:49596 (sra.sie-remote.net-> -)	25	0.37	30	34
6-69.31.6.19:80 -> 207.241.227.119:33302 (- ->ia601509.us.archive.org)	7	0.11	52	52
6-69.31.6.19:80 -> 207.179.185.17:49278 (- ->fctnnbsc38w-207179185017.dhcp-dynamic.FibreOp.nb.bellaliant.net)	7	0.11	72	72
6-69.31.6.19:80 -> 129.174.107.195:57169 (- ->Serve-129-174-107-195.mesa.gmu.edu)	7	0.11	78	78
6-69.31.6.19:20 -> 204.126.132.241:14603 (- ->204-126-132-241-static.hfc.comcastbusiness.net)	6	0.09	184	184
6-69.31.6.19:80 -> 108.35.85.157:57900 (- ->pool-108-35-85-157.nwrknj.fios.verizon.net)	5	0.07	39	39
6-69.31.6.19:80 -> 50.250.22.13:54495 (- ->50-250-22-13-static.hfc.comcastbusiness.net)	5	0.06	59	59
6-198.186.190.164:119 -> 95.91.228.89:13938 (premium-news.de->ip5f5be459.dynamic.kabel-deutschland.de)	4	0.05	91	91
6-69.16.185.113:51121 -> 206.55.64.80:119 (feed-me.highwinds-media.com->nntp.sol.net)	3	0.04	33	33
6-69.31.6.19:80 -> 95.109.70.211:52525 (- -> -)	3	0.04	54	54
6-198.186.192.189:37789 -> 199.168.91.136:873 (- ->BST.deteque.com)	3	0.04	149	149

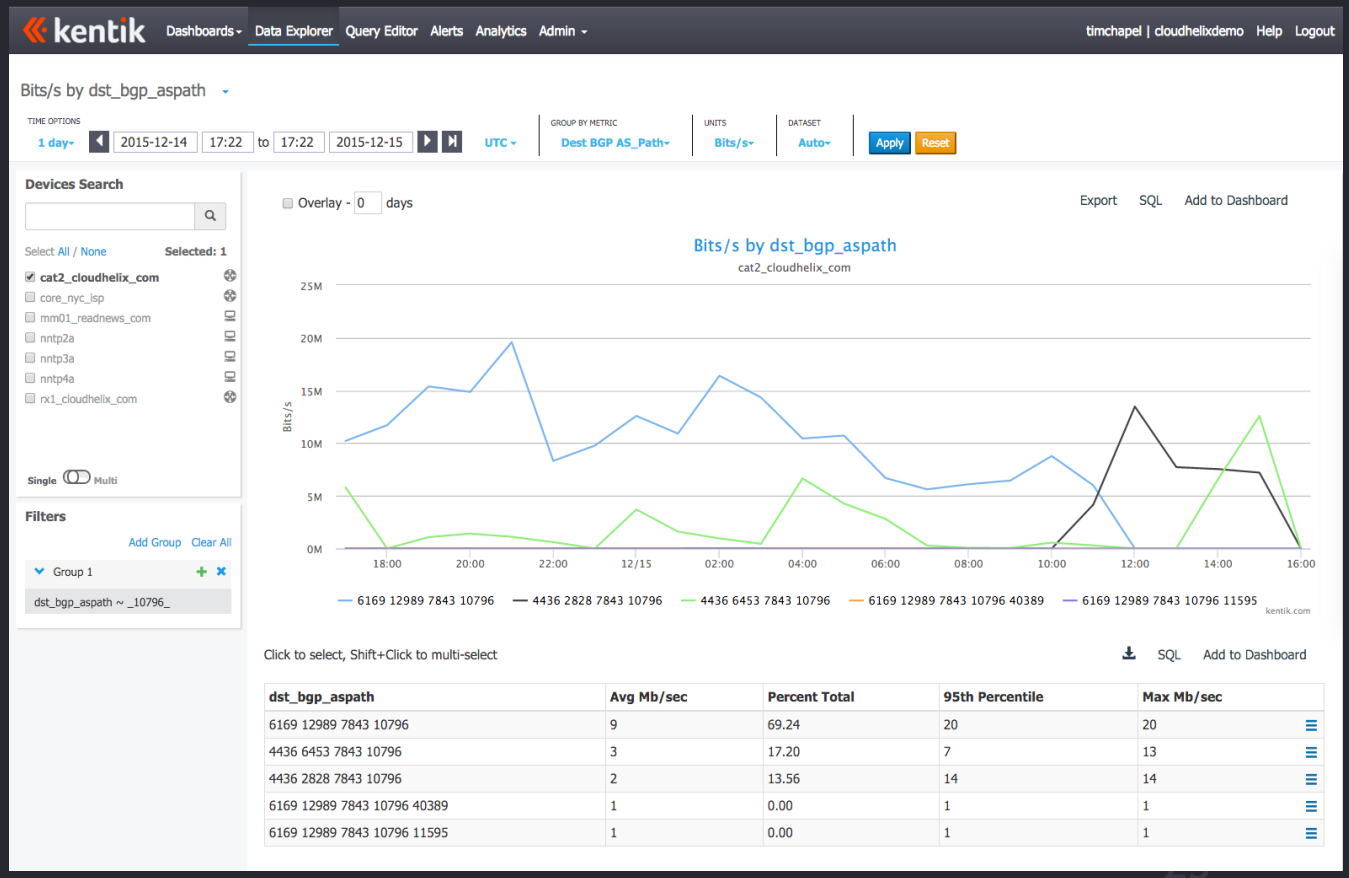
kentik

Traffic by Source Geography



src_geo	Avg Mb/sec	Percent Total	95th Percentile	Max Mb/sec
NL	71	41.17	86	90
DE	19	11.01	23	23
UA	17	9.79	93	122
CA	15	8.22	18	19
JP	10	5.34	11	13
KR	6	3.48	10	11
FR	5	2.75	19	21
PL	5	2.39	6	6
ES	3	1.58	4	5
PT	3	1.51	5	6
GB	3	1.48	5	5
AU	2	1.02	3	3
BR	2	0.99	3	4
RO	2	0.91	4	4

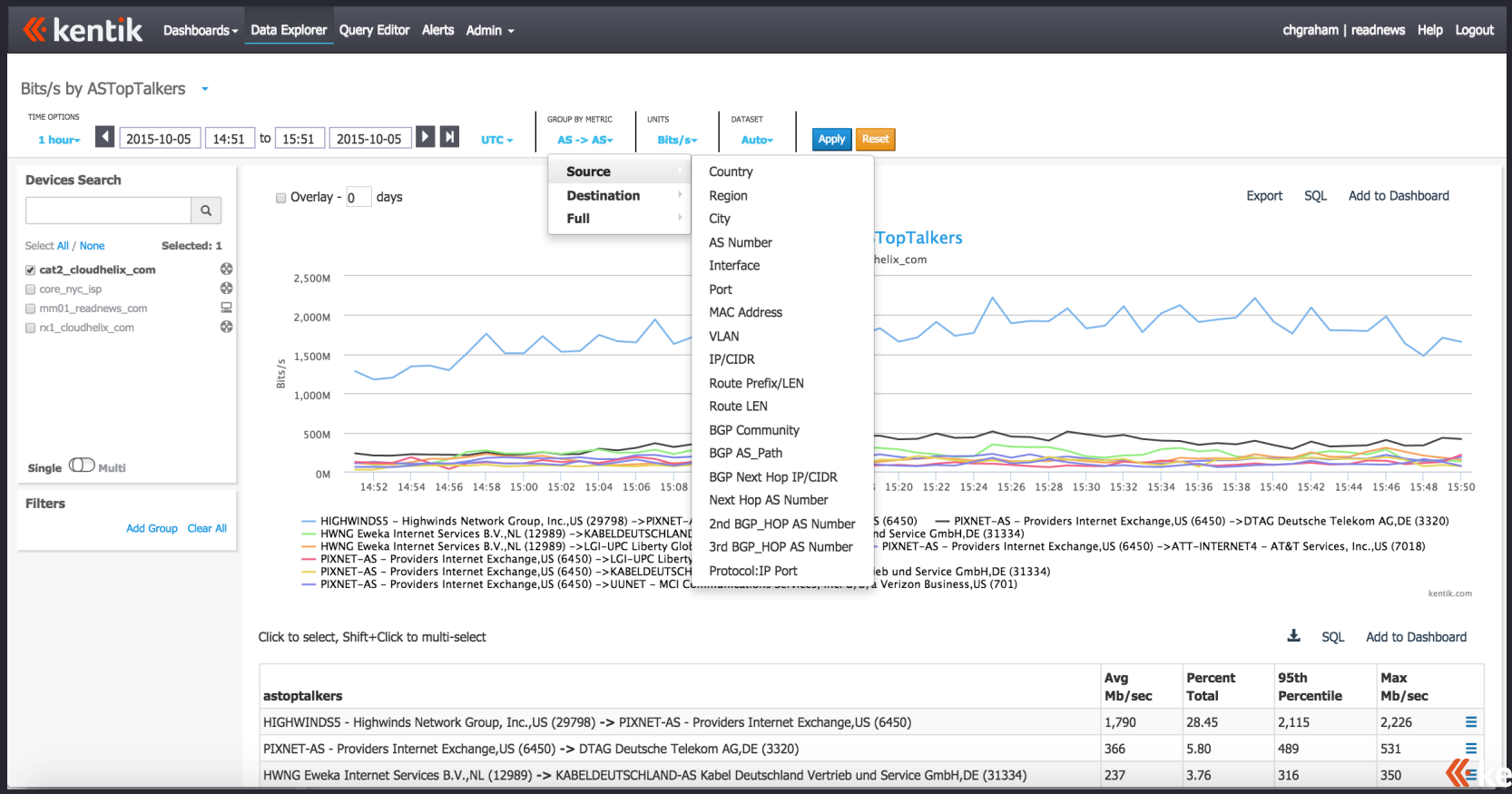
AS Path Changes



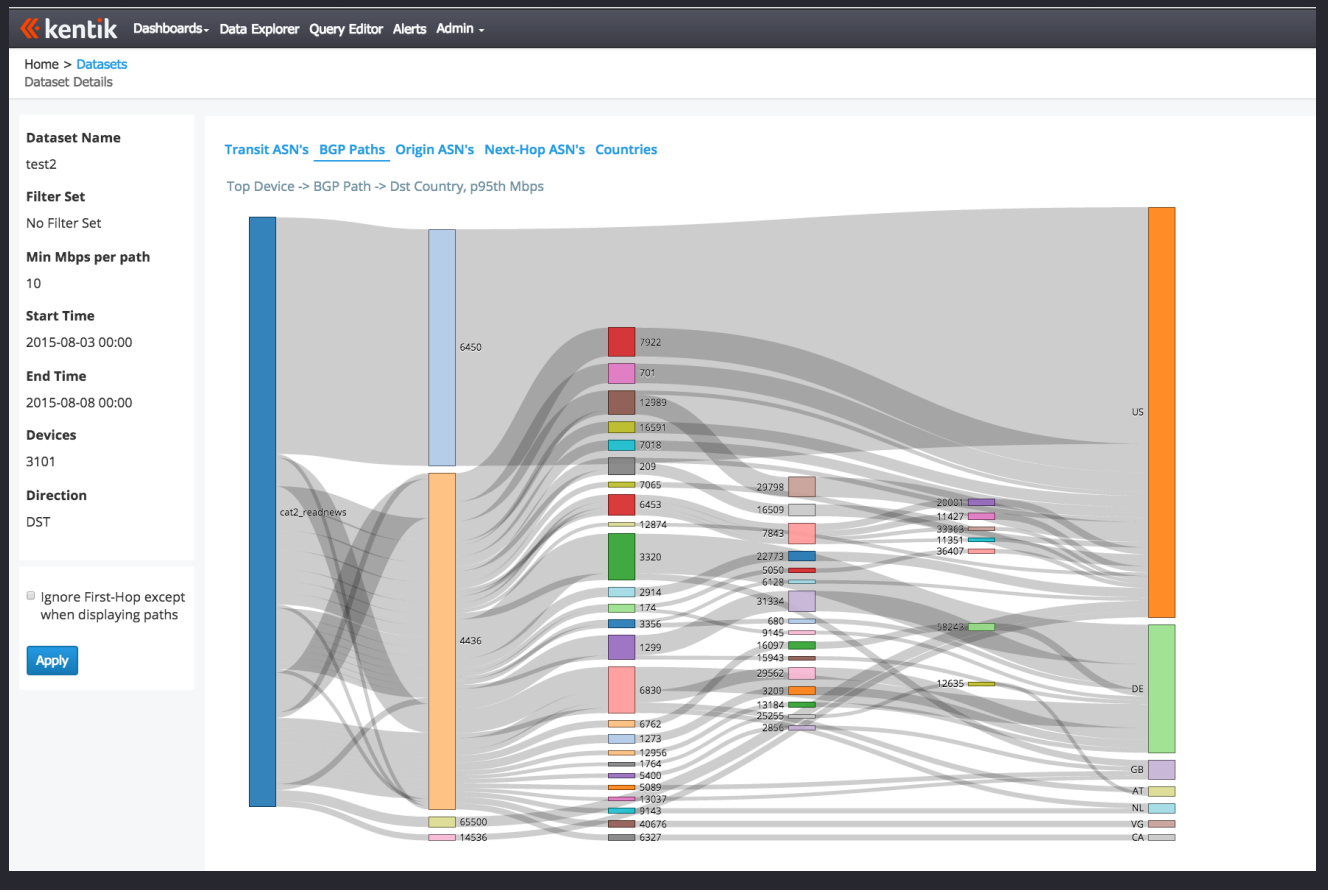
dst_bgp_aspath	Avg Mb/sec	Percent Total	95th Percentile	Max Mb/sec
6169 12989 7843 10796	9	69.24	20	20
4436 6453 7843 10796	3	17.20	7	13
4436 2828 7843 10796	2	13.56	14	14
6169 12989 7843 10796 40389	1	0.00	1	1
6169 12989 7843 10796 11595	1	0.00	1	1

kentik

AS Top Talkers and Drill Down Options

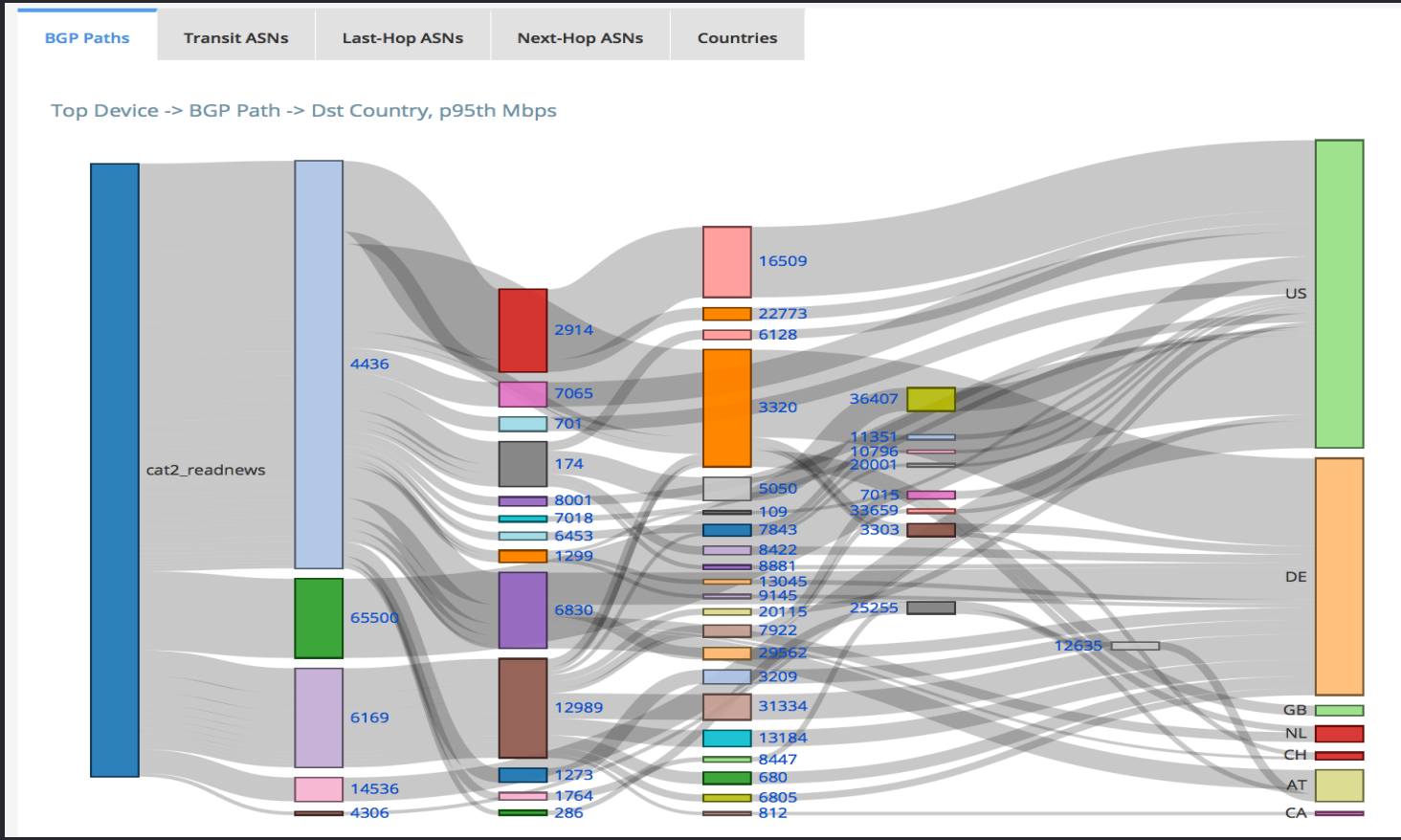


Peering Analytics: ASN by Dest Country Paths



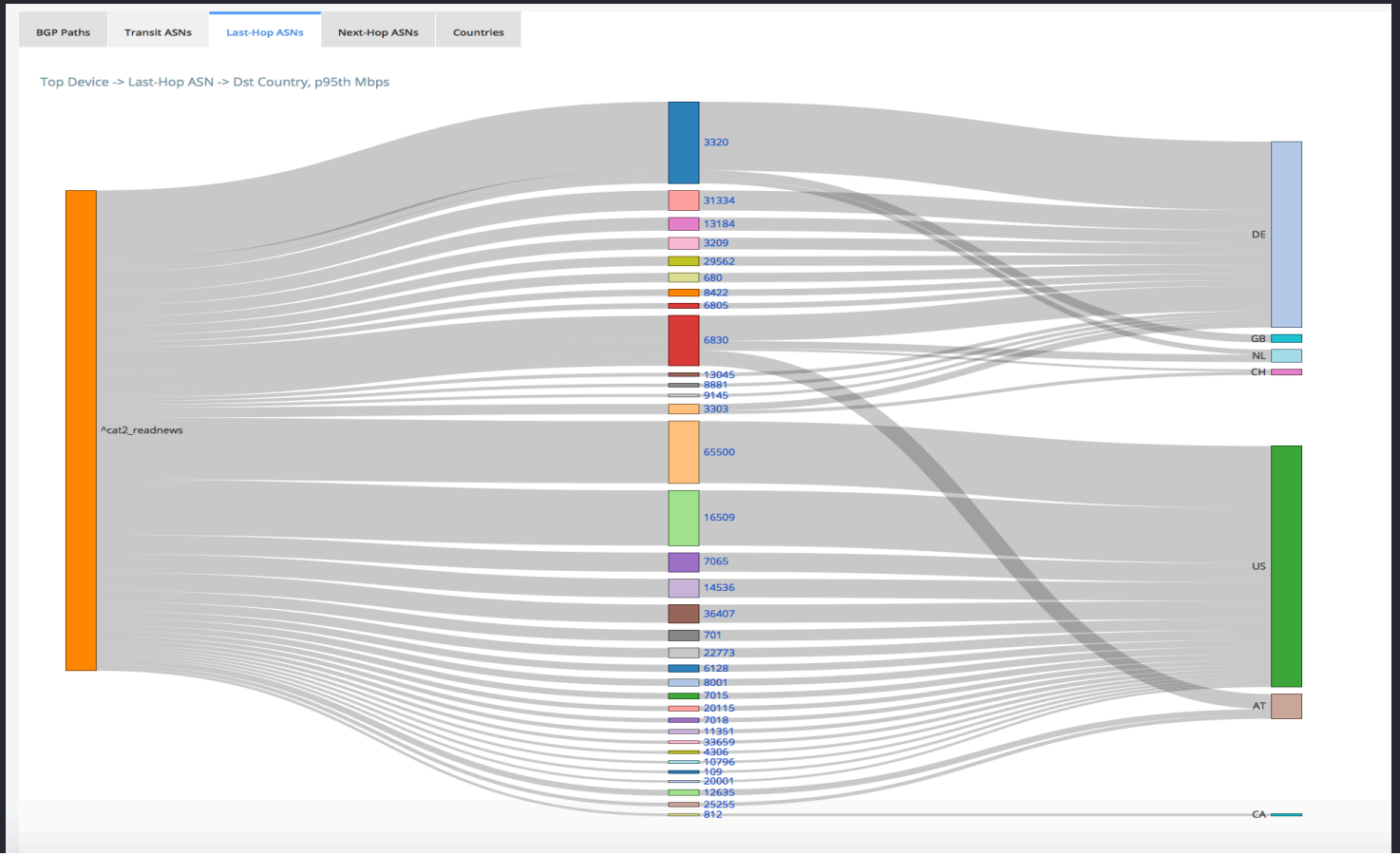
kentik

Peering Analytics: Traffic by BGP Paths



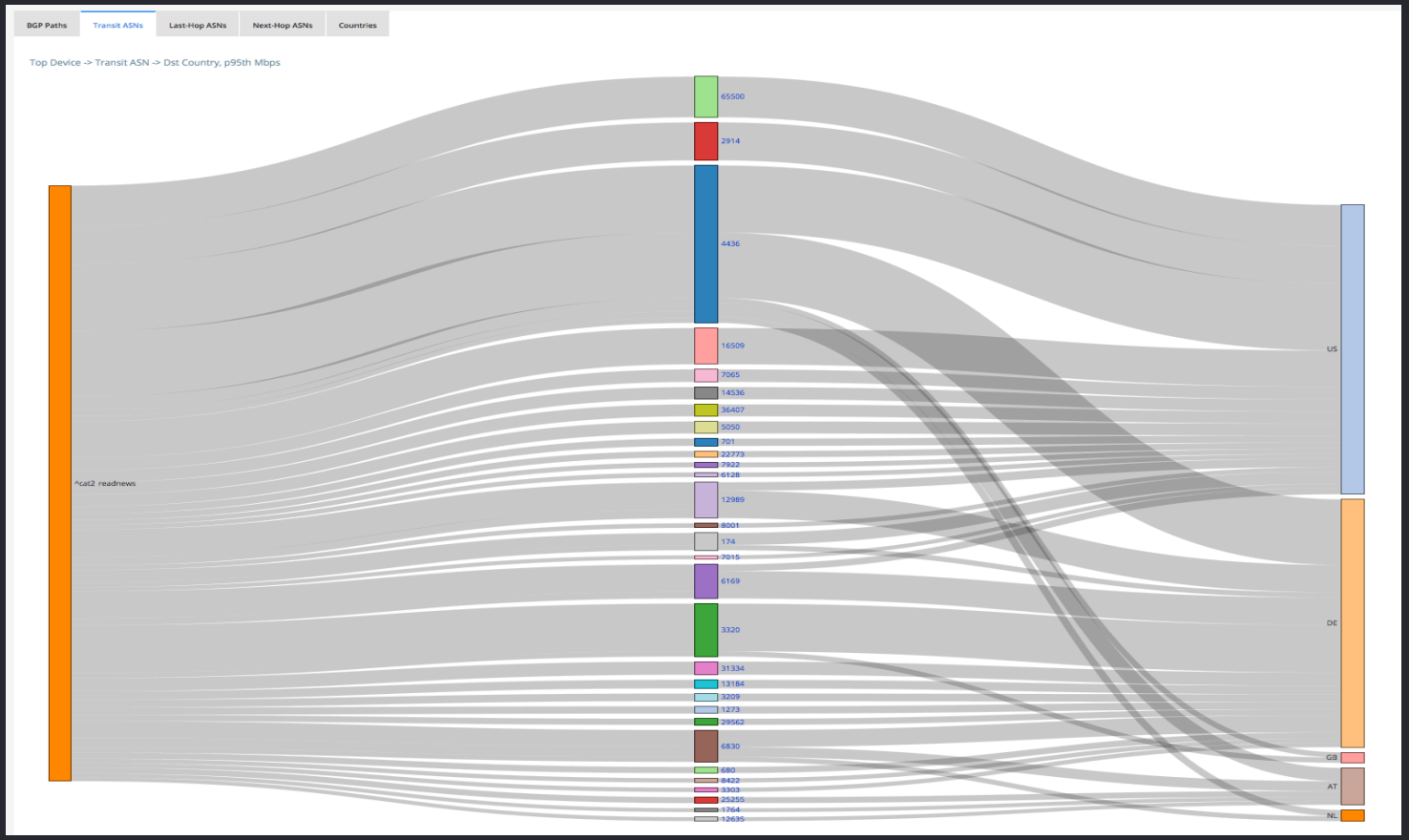
kentik

Peering Analytics: Traffic by Origin AS (“Last Hop”)



kentik

Peering Analytics: Traffic by Transit AS



kentik

Key Takeaways: Cloud Scale NetFlow + BGP

Why You Need It

- Clear Insight into external/Internet network traffic behaviors
- Improved customer/subscriber engagement
- Reduced network operating costs

Technical Path to Success

- This is a big data problem, requiring high capacity/speed for data management, correlation, exploration, and analytics
- SaaS solutions are a fully viable option

kentik

Network Intelligence at Terabit Scale

Thank You!

Jim Frey
VP Product
Kentik Technologies
jfrey@kentik.com
@jfrey80