



DNS TRAFFIC ANALYSIS WITH ENTRADA

*Measuring, Analyzing &
Evaluating DNS Performance*

VNIX-NOG 2026
Tri Nguyen, VNNIC



DNS Performance is not just “low latency”

For large-scale authoritative DNS, performance must be viewed across **latency, availability, correctness, capacity, and traffic distribution**.

Latency

Time from query → response; p50/p95/p99 percentiles

Reliability

No-response, timeout, SERVFAIL, retry/retransmission

Correctness

RCODE, flags, EDNS, DNSSEC, TCP fallback

Capacity

QPS, packet rate, response size, saturation

Distribution

Per anycast site / server / ASN / geography

A DNS change can simultaneously affect multiple layers:

BGP / Anycast catchment

Load / queue processing time

EDNS / DNSSEC response size

TTL / cache query mix

Incident / abuse rcode + volume

*Operational question: “**Is DNS fast?**” → must become “Where and for whom is it fast/stable, over which protocol, and under what conditions?”*

DNS traffic measurement: Active measurement and Passive telemetry complement each other

Active provides “the view from the outside”; passive shows “what actually passes through the system”.

ACTIVE MEASUREMENT

RIPE Atlas / probe-based

- DNS RTT to resolver / authoritative
- IPv4 vs IPv6
- UDP vs TCP
- Multiple vantage points
- Path / reachability context
- Advantage: Direct measurement from the probe's perspective

PASSIVE TELEMETRY

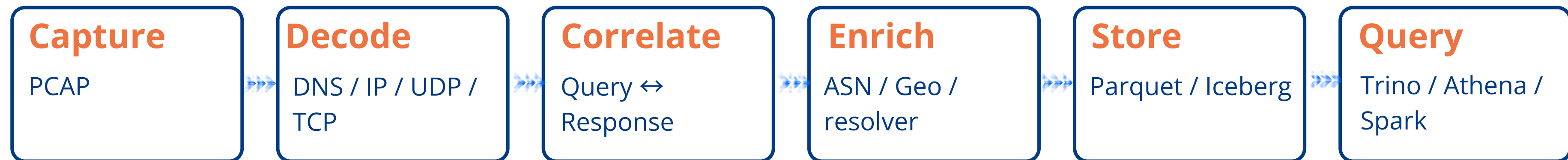
PCAP / DNS query-response stream

- QNAME / QTYPE / RCODE / flags
- Request ↔ response timing
- Source ASN / country / resolver
- Server / anycast site
- EDNS / DNSSEC / response size
- Advantage: Preserves long-term history for investigation and comparison

Important: passive traffic at the authoritative side does not see the full stub → recursive path and does not directly measure users' cache-hit latency.

From packets to a “DNS event”: turning hard-to-query data into an analytical unit

ENTRADA's goal is not only to store PCAP, but to create records that pair query + response and include enriched metadata.



```
time      cname      qtype      rcode      proc_time      ip_version      asn      site
2026-...  www.example.vn  AAAA      NOERROR    1.82 ms        6              AS...    HCM
```

Key design principle: correlate once → query many times. This reduces CPU/I/O cost and investigation complexity.

ENTRADA: from Big Data DNS Warehouse to ENTRADA2

The original ENTRADA (2016) used Hadoop/HDFS + Impala/Spark; ENTRADA2 modernizes it with object storage and an open table format.

ENTRADA • 2016

- PCAP → Parquet
- HDFS
- Impala / Spark
- 6-node on-prem cluster
- ~100 TB of PCAP ingested over >1.5 years
- **Historical benchmark:**
≈52 TB PCAP-equivalent → <3.5 minutes
on 4 data nodes

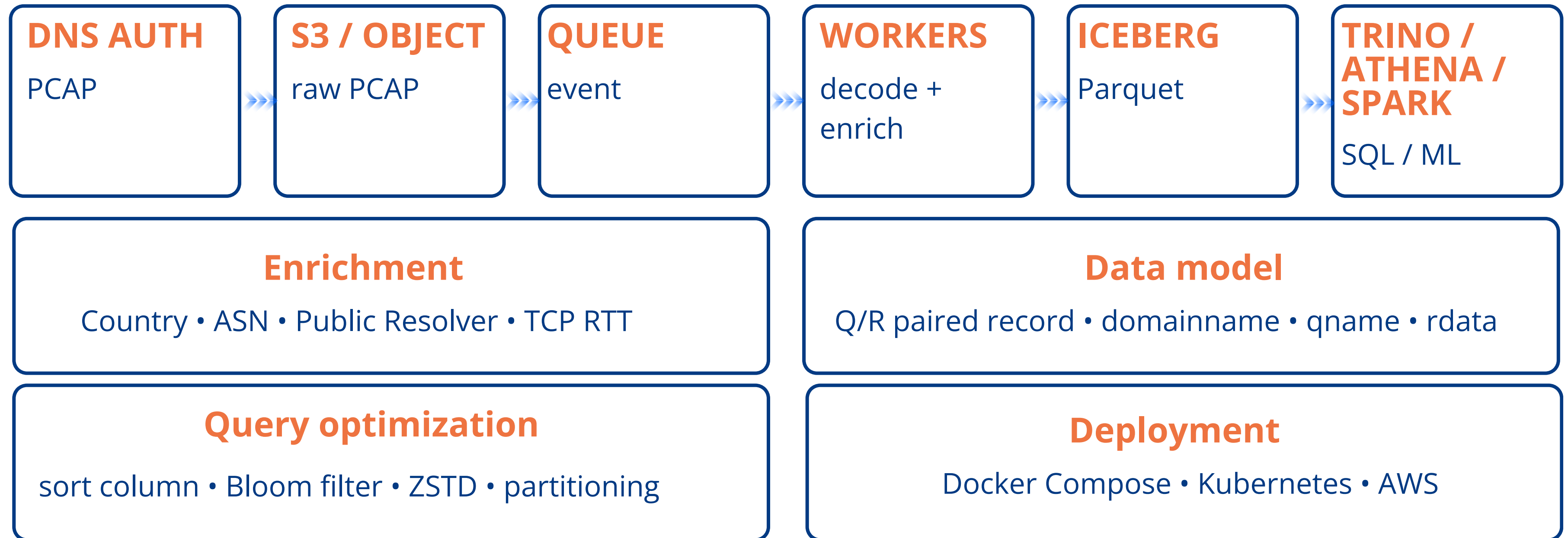
ENTRADA2 • PRESENT

- PCAP → Parquet → Iceberg
- S3-compatible object storage
- RabbitMQ / AWS SQS
- Trino / Athena / Spark
- Geo / ASN / public resolver / TCP RTT
- **Notable additions:**
Iceberg + sorting + Bloom filter + ZSTD
→ optimized queries by domainname

The 2016 benchmark should not be directly compared with ENTRADA2 today; the appropriate comparison is architectural philosophy: columnar + query engine + scalable storage.

ENTRADA2 architecture: decoupled storage-compute, open for analytics

The current design supports both streaming/batch workflows and ad-hoc investigation and dashboards.



DNS event model: low-level enough for tracing, high-level enough for analysis

DNS SEMANTICS

- dns_qname
- dns_domainname
- dns_qtype
- dns_rcode
- dns_tc
- dns_aa
- dns_do

TIMING / TRANSPORT

- time
- dns_proc_time
- tcp_rtt
- prot
- ip_version
- dns_req_len
- dns_res_len

ORIGIN / TOPOLOGY

- ip_src
- ip_asn
- ip_asn_org
- ip_geo_country
- server
- server_location

MODERN DNS CONTEXT

- edns_udp
- edns_version
- edns_options
- edns_ecs
- edns_ext_error
- dns_rdata
- dns_cname

From DNS event → Performance Insight

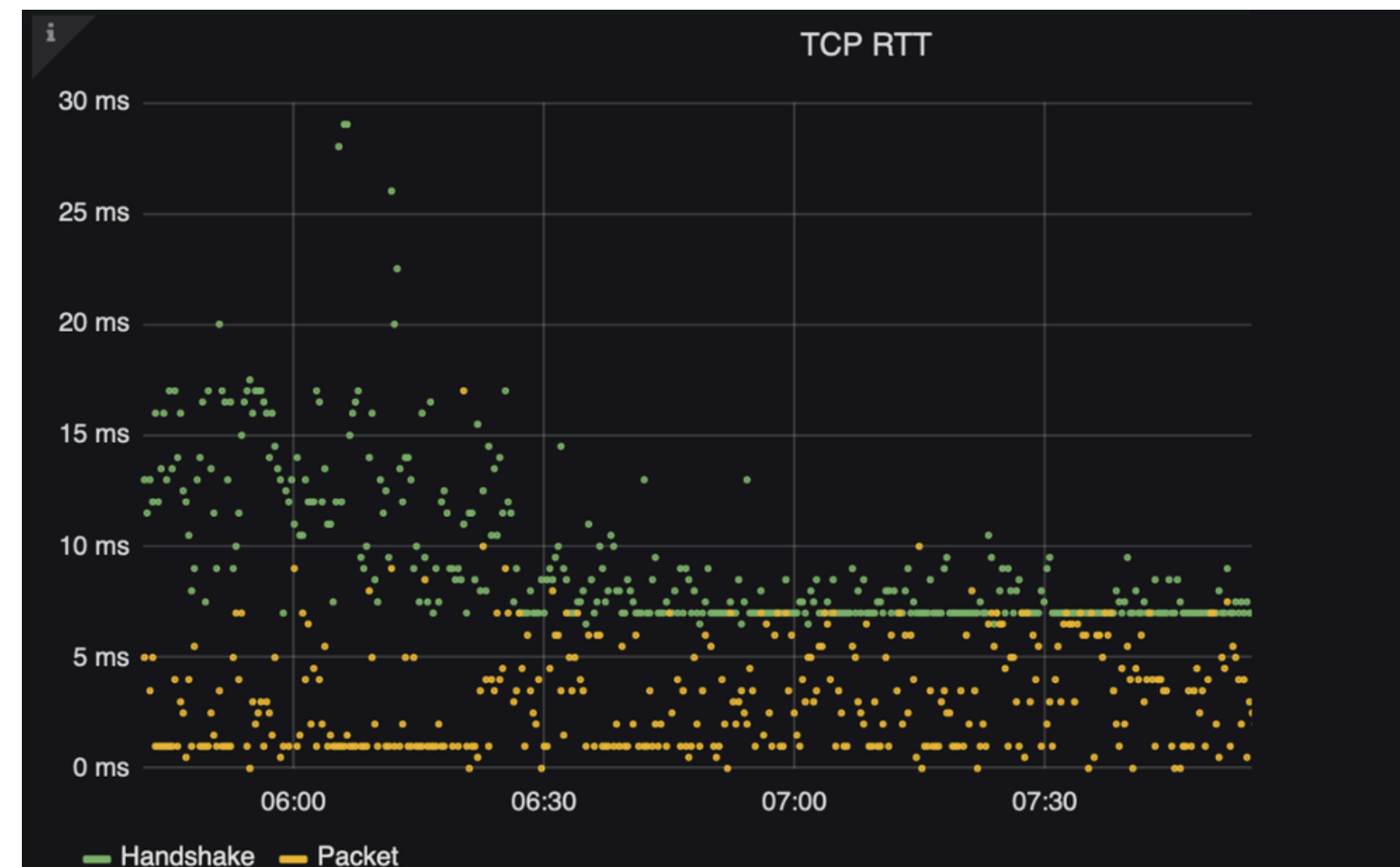
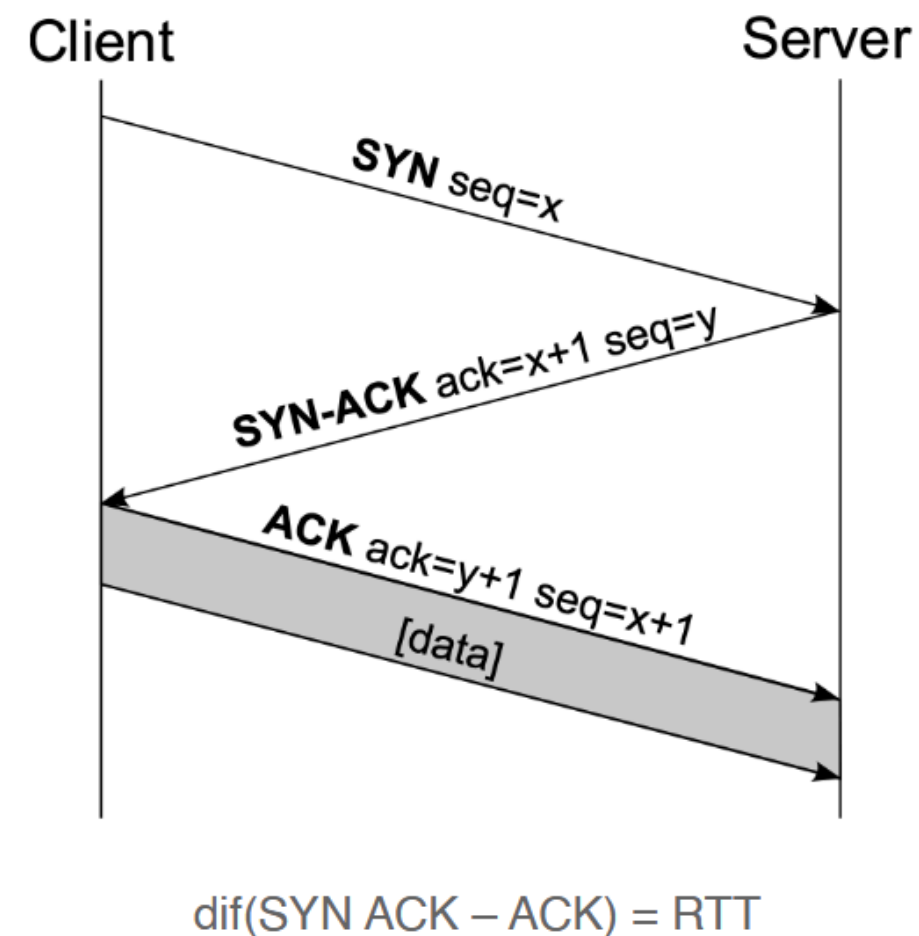
QPS
volume

p50/p95/p99
latency

**NOERROR/
NXDOMAIN/
SERVFAIL**
reliability

**UDP/TCP +
IPv4/IPv6**
transport

**site / ASN /
country**
distribution



Example: turning an operational question into a query

ENTRADA2 uses SQL through Trino/Athena/Spark; domain-centric filtering is supported by sorting + Bloom filter.

-- RCODE BY DAY

```
SELECT day(time) AS day,  
       dns_rcode, COUNT(*) AS n  
FROM dns  
WHERE time >= TIMESTAMP '2026-09-01'  
      AND time < TIMESTAMP '2026-09-08'  
GROUP BY 1, 2  
ORDER BY 1, 3 DESC;
```

*baseline response behavior → detect SERVFAIL /
NXDOMAIN shifts*

Top domains GROUP BY dns_domainname

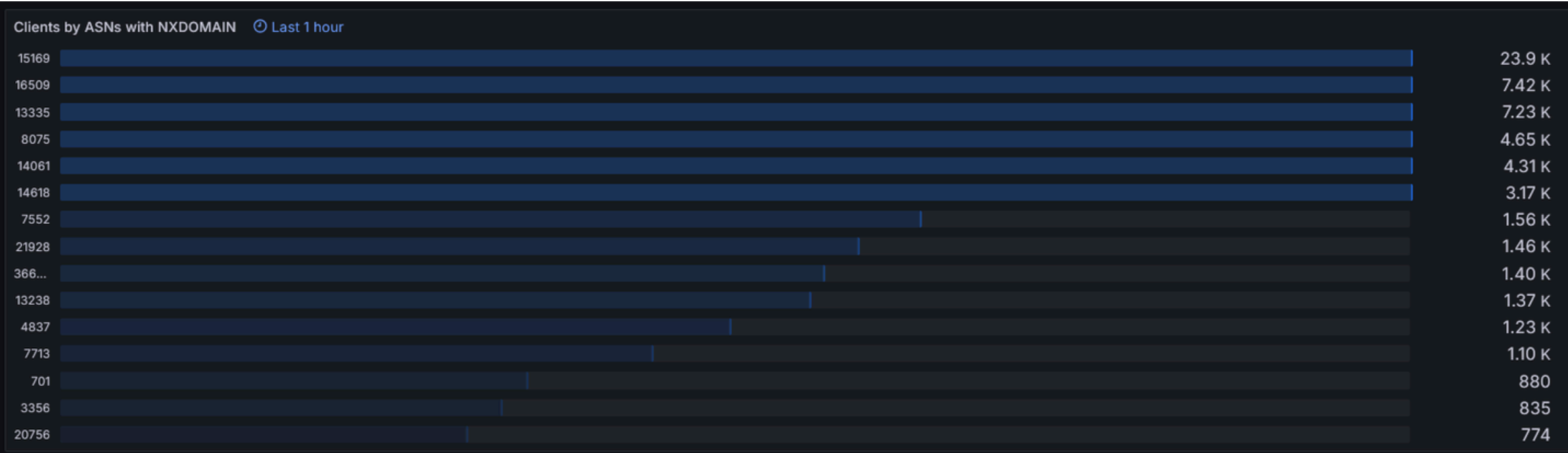
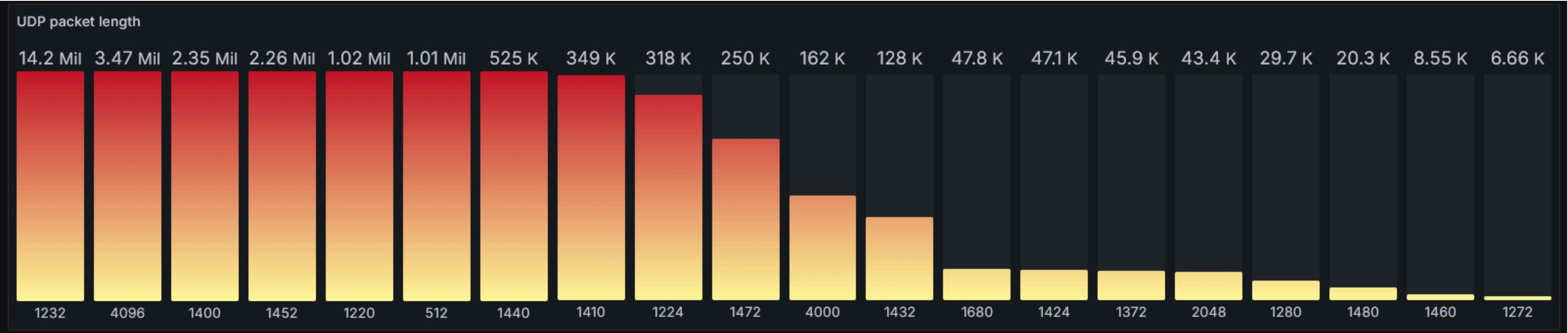
Top ASNs GROUP BY ip_asn

Anycast sites GROUP BY server_location

Large answers dns_res_len > X

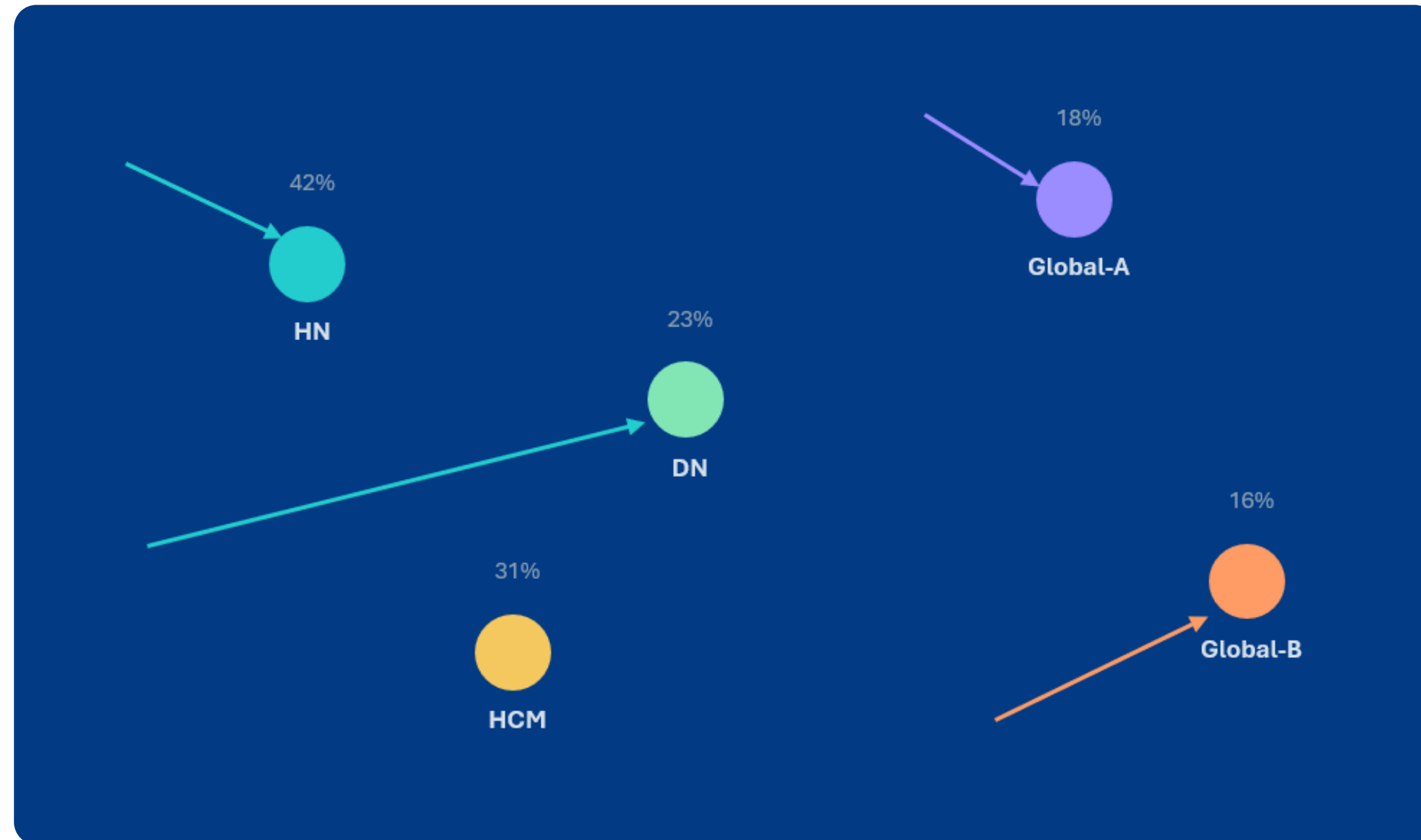
08 / SQL

Example: turning an operational question into a query



Anycast performance: do not look only at total QPS

Traffic distribution by anycast site and routing context determines actual latency and resilience.



QUESTIONS TO MEASURE

- Which site receives more than baseline?
- Does p95 latency differ across sites?
- Did an ASN suddenly change catchment?
- Does the BGP change coincide with a traffic shift?
- Is route engineering causing load skew?

Anomaly detection: baseline first, rules next, ML when needed

Good anomaly detection should create an investigable signal, not just more alerts.

Volume anomaly

QPS spike • source concentration • site skew

Behavior anomaly

RCODE shift • qtype shift • TCP/IPv6 shift

Content anomaly

long labels • entropy • TXT pattern • tunneling clues

Baseline → deviation → triage



rolling median / seasonality



delta / z-score / percentile



ASN / geo / site / qtype



known event? deploy? attack?



route / capacity / mitigation

DNS tunneling research uses ENTRADA to apply rules based on record type, continuous patterns, and query characteristics; “potential activity” does not mean definitive evidence.

DNS Traffic Analytics is an observability layer for both security and operations

Good anomaly detection should create an investigable signal, not just more alerts.

Phishing / malicious domains

query popularity + domain signals

DNS tunneling

TXT / long labels / sequence patterns

DDoS / overload

volume + site skew + RCODE shift

Protocol compliance

EDNS / TCP / DNSSEC behavior

Performance anomaly → root cause evidence → security context → controlled operations

DNS “.VN”: the scale is large enough to require “data-driven operations”

~346 billion

Total

26%

Queries using
IPv6

11%

Queries for
AAAA records

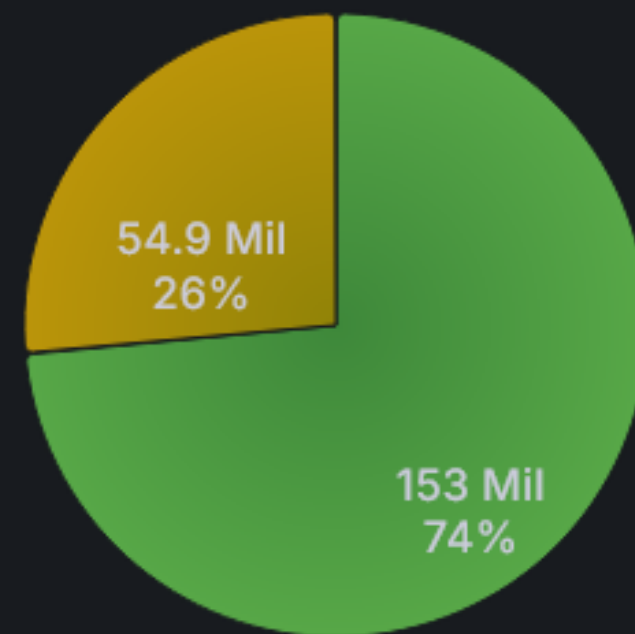
71%

NOERROR
responses

24%

NXDOMAIN
responses

Queries by IP version ⓘ



— IPv4 — IPv6

Queries by transport version ⓘ



— UDP — TCP

What .VN traffic revealed — and what we decided

Three findings that total QPS and server dashboards could not show

1 EDNS & DNSSEC

54% of EDNS queries advertise a 1232 B buffer

SEEN

≈14% advertise >1500 B; ≈4% only 512 B; DS = 15% of queries

MEANS

Large DNSSEC answers risk UDP fragmentation and silent loss

DECISION

Enforce a 1232 B UDP cap and size TCP capacity per site

VERIFY

UDP answers >1232 B = 0; TC rate and TCP share (~2%) per site

2 NXDOMAIN sources

~83% of NXDOMAIN from 6 cloud & public-resolver ASN

SEEN

NXDOMAIN = 24% of answers; AS15169 ≈39%, VN ISPs ≈2.5%

MEANS

Driven by global resolvers and cloud clients, not VN users

DECISION

Per-ASN baselines + alerts, notify owners, tune RRL per source

VERIFY

NXDOMAIN share per ASN, weekly

3 Anycast & IPv6

26% of queries arrive over IPv6 vs ~65% of VN users

SEEN

Some VN ASNs land on overseas anycast sites, with higher TCP RTT

MEANS

Domestic queries detour abroad; resolvers lag users on IPv6

DECISION

BGP tuning + VNIX peering; IPv6 on ISP resolver egress

VERIFY

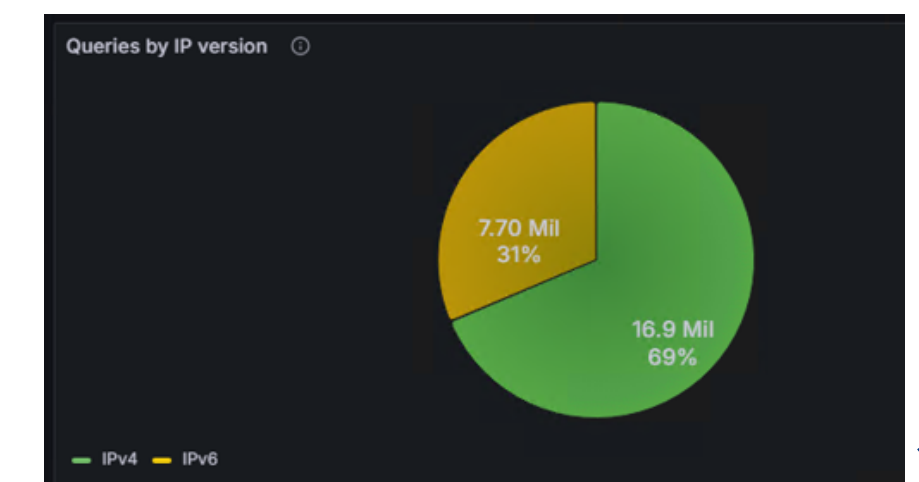
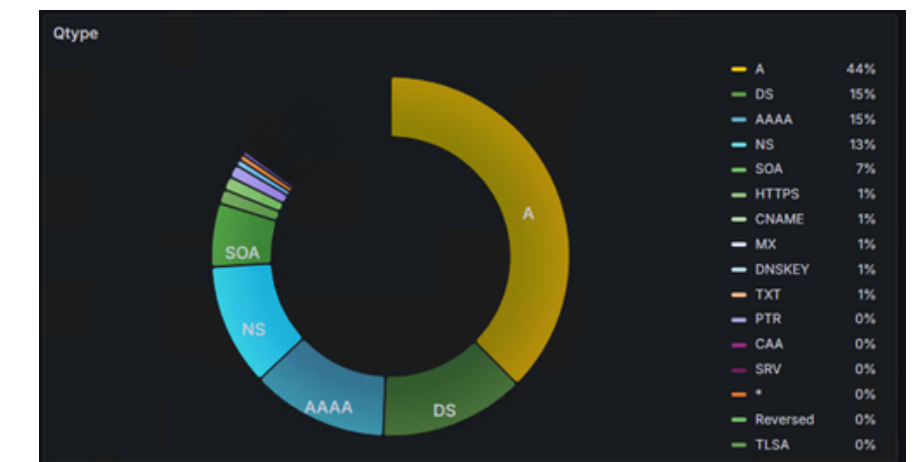
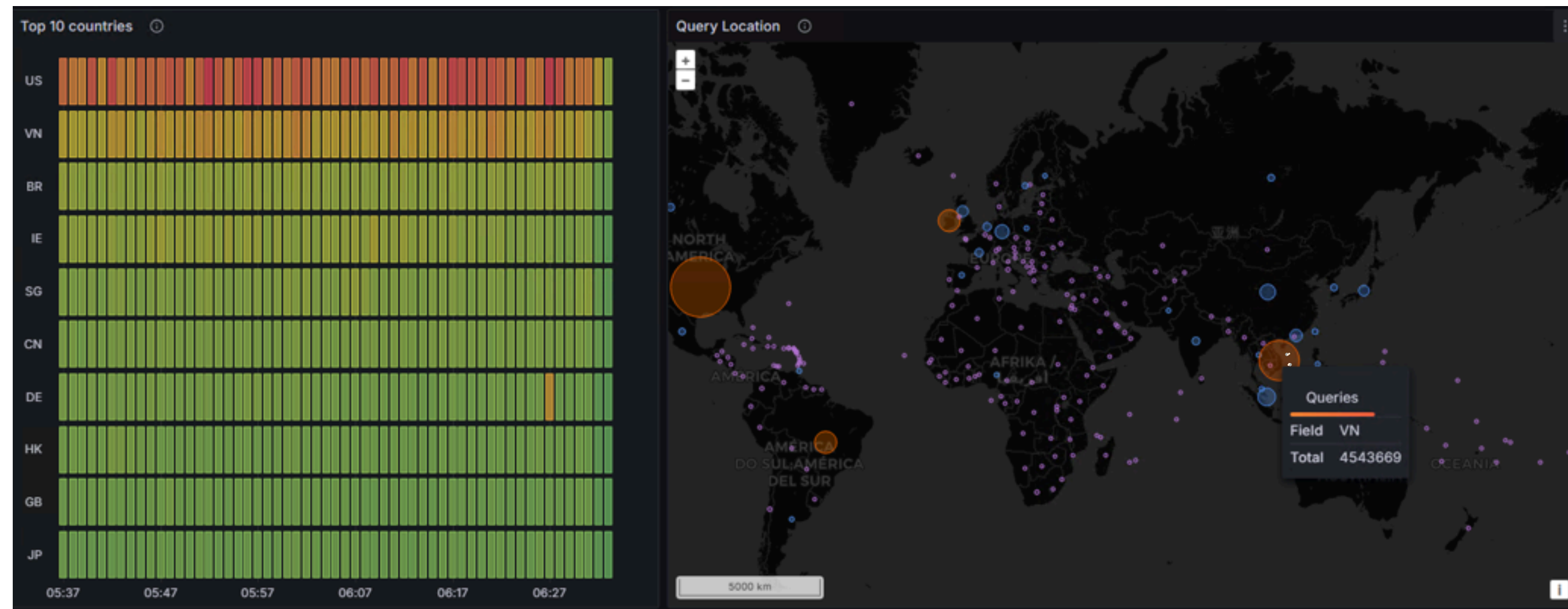
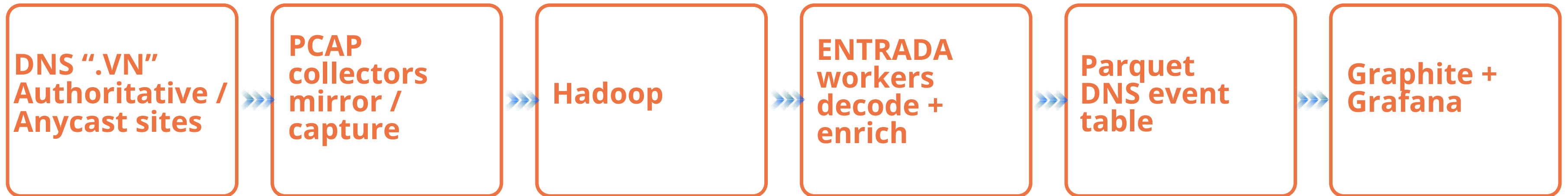
Per-ASN site share and TCP RTT before/after; RIPE Atlas

From public data → an operational question set for DNS “.vn”

-  **Capacity**
Peak QPS by site / server?
-  **Latency**
Where do p95/p99 change?
-  **Availability**
When do no-response / SERVFAIL increase?
-  **Traffic mix**
Do IPv6 / TCP / EDNS / DNSSEC change?
-  **Distribution**
Is there abnormal catchment skew?
-  **Root cause**
route, server, resolver, or domain?

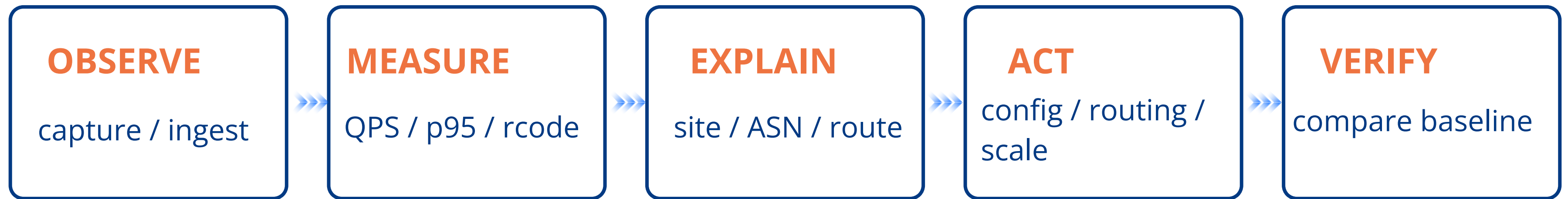
Public evidence already shows protocol / IPv6 / RCODE mix. ENTRADA-style telemetry can extend this to “where + when + what changed”.

ENTRADA integration model for DNS “.VN” — technical proposal



Data-driven DNS operations loop

The value of analytics appears when insights feed back into technical decisions.



CONCLUSION

1

WHY DNS TRAFFIC ANALYSIS MATTERS

It reveals traffic patterns, protocol behavior, failures, latency shifts, and anycast imbalance - turning raw packets into evidence for operations, capacity planning, and security.

2

A FEASIBLE IMPLEMENTATION PATH

ENTRADA provides a practical pipeline: capture PCAP, correlate and enrich DNS events, store them in Parquet/Iceberg, then analyze with SQL and dashboards. Start with one-site pilot, then scale.

3

TOWARD DNS QUALITY ASSESSMENT

Establish baselines and thresholds for latency, reliability, correctness, capacity, and distribution; segment results by site, ASN, protocol, and domain; link findings to action and verification.



THANK YOU