Distributed latency monitoring

Anurag Bhatia, Hurricane Electric

Starts with idea of looking for smokeping alternative...

Smokeping

- Monitors latency, packet loss etc based on ICMP
- Supports ICMP, HTTP, DNS and many other “probes”
- Easy quick config
- Can send email if high latency, packet loss etc is detected

Challenges with Smokeping

1. Hard to scale up

2. Different locations need different setups / no easy federated view

3. No easy to club graphs based on source or destination
   E.g 5 locations, 50 endpoints = 250 graphs!

4. Limited alerting support

Why look for alternative?

- It’s important to watch out for latency between various endpoints
- With many sources to many destinations, number of graphs can be very high & hence an aggregate view is important
- It’s easier to have single tool to monitor latency, monitor servers, network devices, application and API endpoints
- Requirement to run setup in high availability design

Prometheus

Prometheus

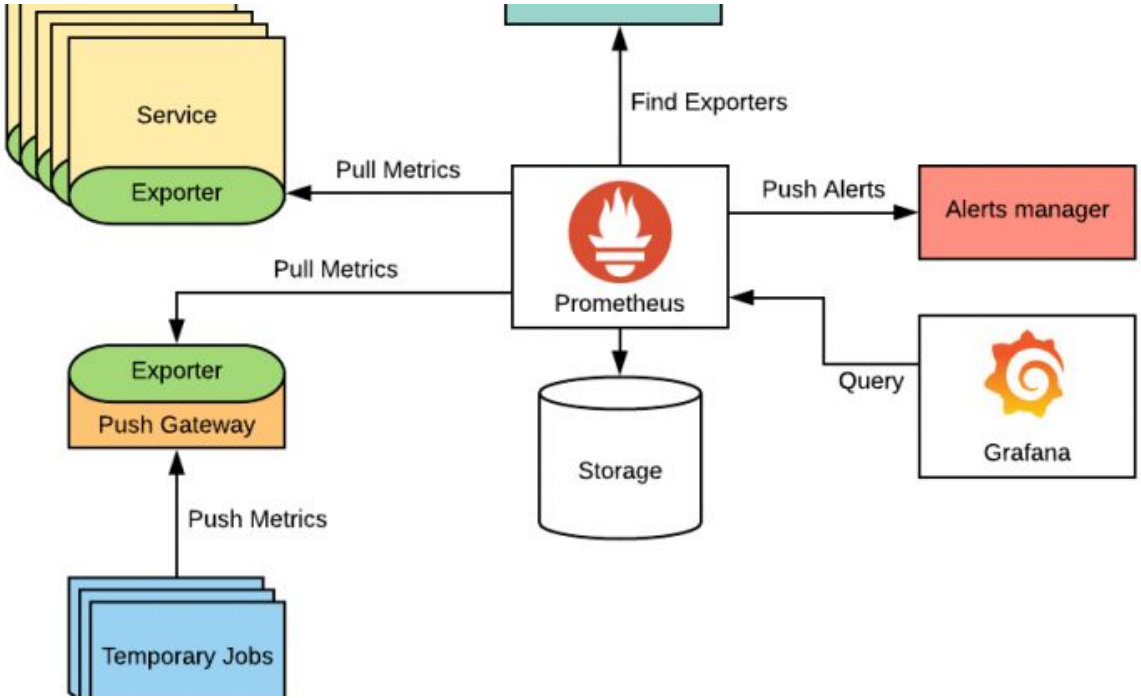
1. Tool which in itself includes a tool to retrieve various metrics, store them a Time Series Database (TSDB), make them available over HTTP endpoint

2. Works on a “pull model” by default where metrics can be pulled over from endpoints which run “agent”

3. Can store any metrics, with any set of labels like CPU, memory utilization, storage utilisation, network interface traffic and even the latency!

4. Prometheus server speaks to agent via HTTP(s) to pull these metrics at predefined intervals

Prometheus design



Everything is “metrics”...

Everything is “metrics”...

- Metrics can be interpret / graphed in way needed
- Possible to look at average (e.g 1 min average, 5 min average etc)
- One can attach various labels with metric (e.g dst_country: HK, dst_type: cloud etc)
- Support for alerting (via Alertmanager) based on predefined rule against a metric

Example of metrics

```
anurag@desktop -> curl lo.server7.anuragbhatia.com:9100/metrics
# HELP go_gc_duration_seconds A summary of the pause duration of garbage collection cycles.
# TYPE go_gc_duration_seconds summary
go_gc_duration_seconds{quantile="0"} 3.3299e-05
go_gc_duration_seconds{quantile="0.25"} 5.8645e-05
go_gc_duration_seconds{quantile="0.5"} 7.2725e-05
go_gc_duration_seconds{quantile="0.75"} 0.000100836
go_gc_duration_seconds{quantile="1"} 0.000839921
go_gc_duration_seconds_sum 45.901053136
go_gc_duration_seconds_count 352397
# HELP go_goroutines Number of goroutines that currently exist.
# TYPE go_goroutines gauge
go_goroutines 8
# HELP go_info Information about the Go environment.
# TYPE go_info gauge
go_info{version="go1.20.6"} 1
# HELP go_memstats_alloc_bytes Number of bytes allocated and still in use.
# TYPE go_memstats_alloc_bytes gauge
go_memstats_alloc_bytes 2.451792e+06
# HELP go_memstats_alloc_bytes_total Total number of bytes allocated, even if freed.
# TYPE go_memstats_alloc_bytes_total counter
go_memstats_alloc_bytes_total 7.08515897304e+11
# HELP go_memstats_buck_hash_sys_bytes Number of bytes used by the profiling bucket hash table.
# TYPE go_memstats_buck_hash_sys_bytes gauge
go_memstats_buck_hash_sys_bytes 2.104808e+06
# HELP go_memstats_frees_total Total number of frees.
# TYPE go_memstats_frees_total counter
go_memstats_frees_total 1.0299813362e+10
# HELP go_memstats_gc_sys_bytes Number of bytes used for garbage collection system metadata.
# TYPE go_memstats_gc_sys_bytes gauge
go_memstats_gc_sys_bytes 8.54596e+06
# HELP go_memstats_heap_alloc_bytes Number of heap bytes allocated and still in use.
# TYPE go_memstats_heap_alloc_bytes gauge
go_memstats_heap_alloc_bytes 2.451792e+06
# HELP go_memstats_heap_idle_bytes Number of heap bytes waiting to be used.
# TYPE go_memstats_heap_idle_bytes gauge
go_memstats_heap_idle_bytes 8.15104e+06
# HELP go_memstats_heap_inuse_bytes Number of heap bytes that are in use.
# TYPE go_memstats_heap_inuse_bytes gauge
go_memstats_heap_inuse_bytes 3.842048e+06
# HELP go_memstats_heap_objects Number of allocated objects.
# TYPE go_memstats_heap_objects gauge
go_memstats_heap_objects 36420
# HELP go_memstats_heap_released_bytes Number of heap bytes released to OS.
# TYPE go_memstats_heap_released_bytes gauge
go_memstats_heap_released_bytes 6.815744e+06
# HELP go_memstats_heap_sys_bytes Number of heap bytes obtained from system.
# TYPE go_memstats_heap_sys_bytes gauge
go_memstats_heap_sys_bytes 1.1993088e+07
# HELP go_memstats_last_gc_time_seconds Number of seconds since 1970 of last garbage collection.
# TYPE go_memstats_last_gc_time_seconds gauge
```

Introducing Blackbox exporter

Blackbox exporter

- Open source probing endpoint which can trigger measurement whenever probes
- Probed over HTTP(s) endpoint with requirement arguments of host to measure
- Supports HTTP, HTTPS, DNS, TCP, ICMP and gRPC
- Written in go, can be downloaded & executed as binary on server or as docker container

ICMP probe for “hknog.net” via Blackbox exporter

```
anurag@desktop ~> curl "http://lo.server7.anuragbhatia.com:9115/probe?module=icmp4&target=hknog.net"
# HELP probe_dns_lookup_time_seconds Returns the time taken for probe dns lookup in seconds
# TYPE probe_dns_lookup_time_seconds gauge
probe_dns_lookup_time_seconds 0.163528277
# HELP probe_duration_seconds Returns how long the probe took to complete in seconds
# TYPE probe_duration_seconds gauge
probe_duration_seconds 0.466092885
# HELP probe_icmp_duration_seconds Duration of icmp request by phase
# TYPE probe_icmp_duration_seconds gauge
probe_icmp_duration_seconds{phase="resolve"} 0.163528277
probe_icmp_duration_seconds{phase="rtt"} 0.302224495
probe_icmp_duration_seconds{phase="setup"} 0.000107769
# HELP probe_icmp_reply_hop_limit Replied packet hop limit (TTL for ipv4)
# TYPE probe_icmp_reply_hop_limit gauge
probe_icmp_reply_hop_limit 54
# HELP probe_ip_addr_hash Specifies the hash of IP address. It's useful to detect if the IP address changes.
# TYPE probe_ip_addr_hash gauge
probe_ip_addr_hash 1.634000219e+09
# HELP probe_ip_protocol Specifies whether probe ip protocol is IP4 or IP6
# TYPE probe_ip_protocol gauge
probe_ip_protocol 4
# HELP probe_success Displays whether or not the probe was a success
# TYPE probe_success gauge
probe_success 1
anurag@desktop ~>
```

http_2xx probe for “hknog.net” via Blackbox exporter

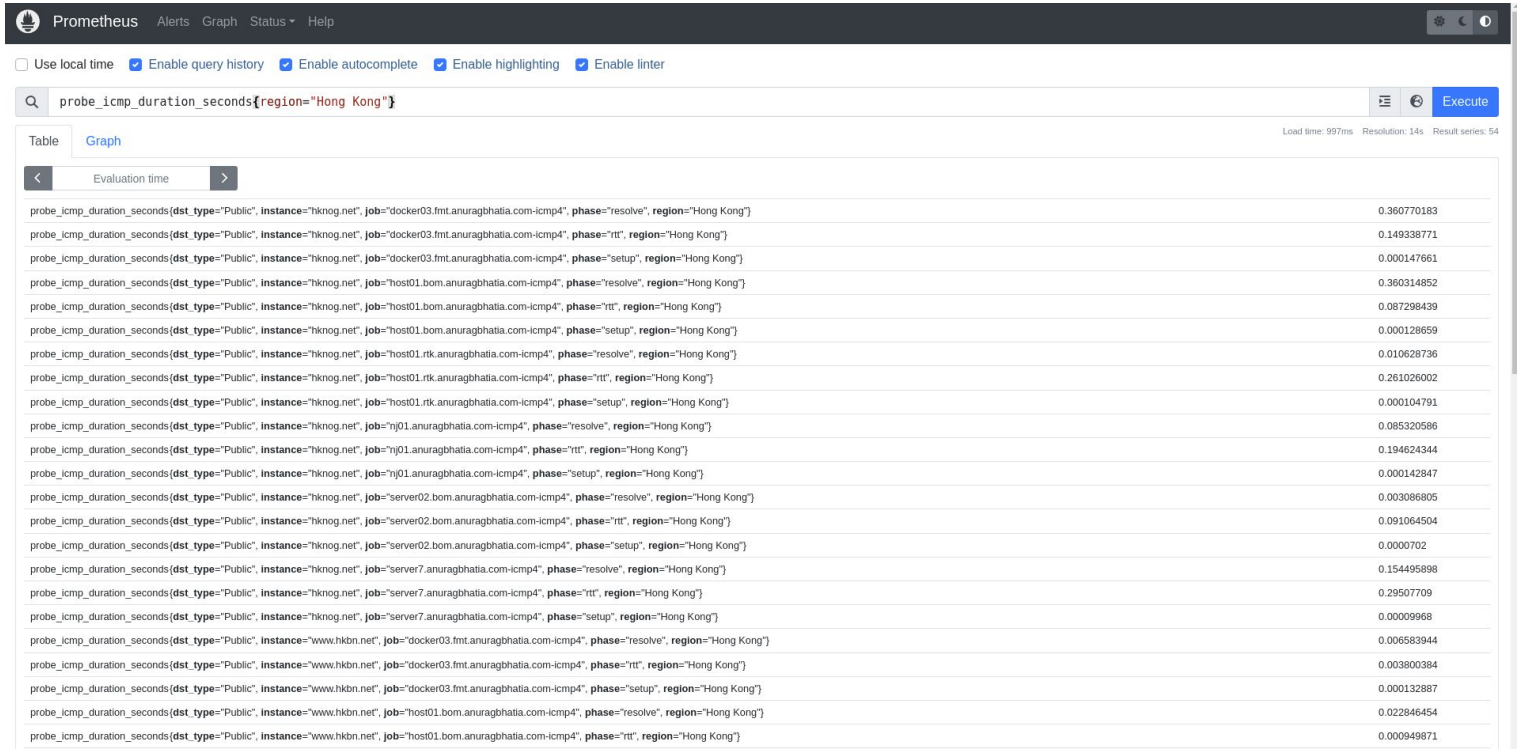
```
anurag@desktop -> curl "http://lo.server7.anuragbhatia.com:9115/probe?module=http_2xx&target=https://hknog.net"
# HELP probe_dns_lookup_time_seconds Returns the time taken for probe dns lookup in seconds
# TYPE probe_dns_lookup_time_seconds gauge
probe_dns_lookup_time_seconds 0.011924482
# HELP probe_duration_seconds Returns how long the probe took to complete in seconds
# TYPE probe_duration_seconds gauge
probe_duration_seconds 1.378284781
# HELP probe_failed_due_to_regex Indicates if probe failed due to regex
# TYPE probe_failed_due_to_regex gauge
probe_failed_due_to_regex 0
# HELP probe_http_content_length Length of http content response
# TYPE probe_http_content_length gauge
probe_http_content_length 19071
# HELP probe_http_duration_seconds Duration of http request by phase, summed over all redirects
# TYPE probe_http_duration_seconds gauge
probe_http_duration_seconds{phase="connect"} 0.292706958
probe_http_duration_seconds{phase="processing"} 0.484309093
probe_http_duration_seconds{phase="resolve"} 0.011924482
probe_http_duration_seconds{phase="tls"} 0.295746634
probe_http_duration_seconds{phase="transfer"} 0.292787411
# HELP probe_http_redirects The number of redirects
# TYPE probe_http_redirects gauge
probe_http_redirects 0
# HELP probe_http_ssl Indicates if SSL was used for the final redirect
# TYPE probe_http_ssl gauge
probe_http_ssl 1
# HELP probe_http_status_code Response HTTP status code
# TYPE probe_http_status_code gauge
probe_http_status_code 200
# HELP probe_http_uncompressed_body_length Length of uncompressed response body
# TYPE probe_http_uncompressed_body_length gauge
probe_http_uncompressed_body_length 19071
# HELP probe_http_version Returns the version of HTTP of the probe response
# TYPE probe_http_version gauge
probe_http_version 2
# HELP probe_ip_addr_hash Specifies the hash of IP address. It's useful to detect if the IP address changes.
# TYPE probe_ip_addr_hash gauge
probe_ip_addr_hash 1.717888314e+09
# HELP probe_ip_protocol Specifies whether probe ip protocol is IP4 or IP6
# TYPE probe_ip_protocol gauge
probe_ip_protocol 4
# HELP probe_ssl_earliest_cert_expiry Returns last SSL chain expiry in unixtime
# TYPE probe_ssl_earliest_cert_expiry gauge
probe_ssl_earliest_cert_expiry 1.706543999e+09
# HELP probe_ssl_last_chain_expiry_timestamp_seconds Returns last SSL chain expiry in timestamp
# TYPE probe_ssl_last_chain_expiry_timestamp_seconds gauge
probe_ssl_last_chain_expiry_timestamp_seconds 1.706543999e+09
# HELP probe_ssl_last_chain_info Contains SSL leaf certificate information
# TYPE probe_ssl_last_chain_info gauge
probe_ssl_last_chain_info{fingerprint_sha256="6039c601670757d626fd227f8015abe870b99bb248df6569838d6673d3282408",issuer="CN=AlphaSSL CA - SHA256 - G4,O=Globa
lSign nv-sa,C=BE",subject="CN=www.hknog.net",subjectalternative="www.hknog.net,hknog.net"} 1
# HELP probe_success Displays whether or not the probe was a success
# TYPE probe_success gauge
probe_success 1
# HELP probe_tls_version_info Returns the TLS version used or NaN when unknown
# TYPE probe_tls_version_info gauge
probe_tls_version_info{version="TLS 1.3"} 1
anurag@desktop ->
```

Prometheus Configuration examples...

```
- targets: ['hknog.net.']
  labels:
    dst_type: 'Public'
    name: 'HKNOG Website'
    region: 'Hong Kong'
```

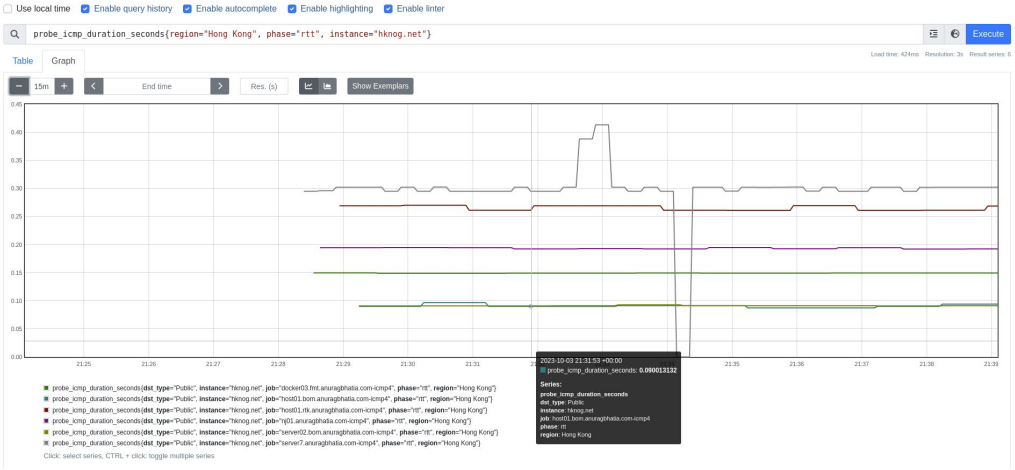
```
- targets:
  - hknog.net #HKNOG Website
  - www.hkbn.net
  - www.hkt.com
  labels:
    dst_type: 'Public'
    region: 'Hong Kong'
```

Prometheus query example



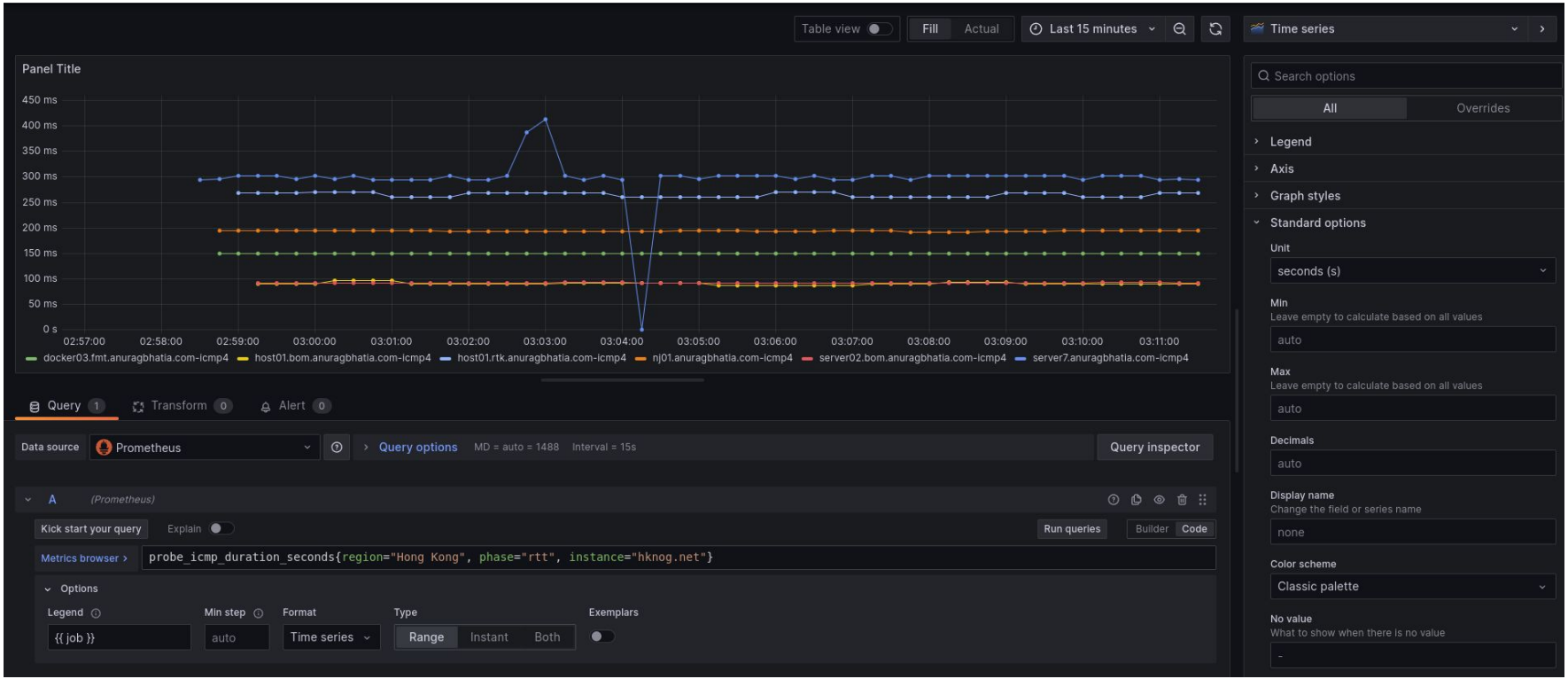
Prometheus query example

Prometheus query example

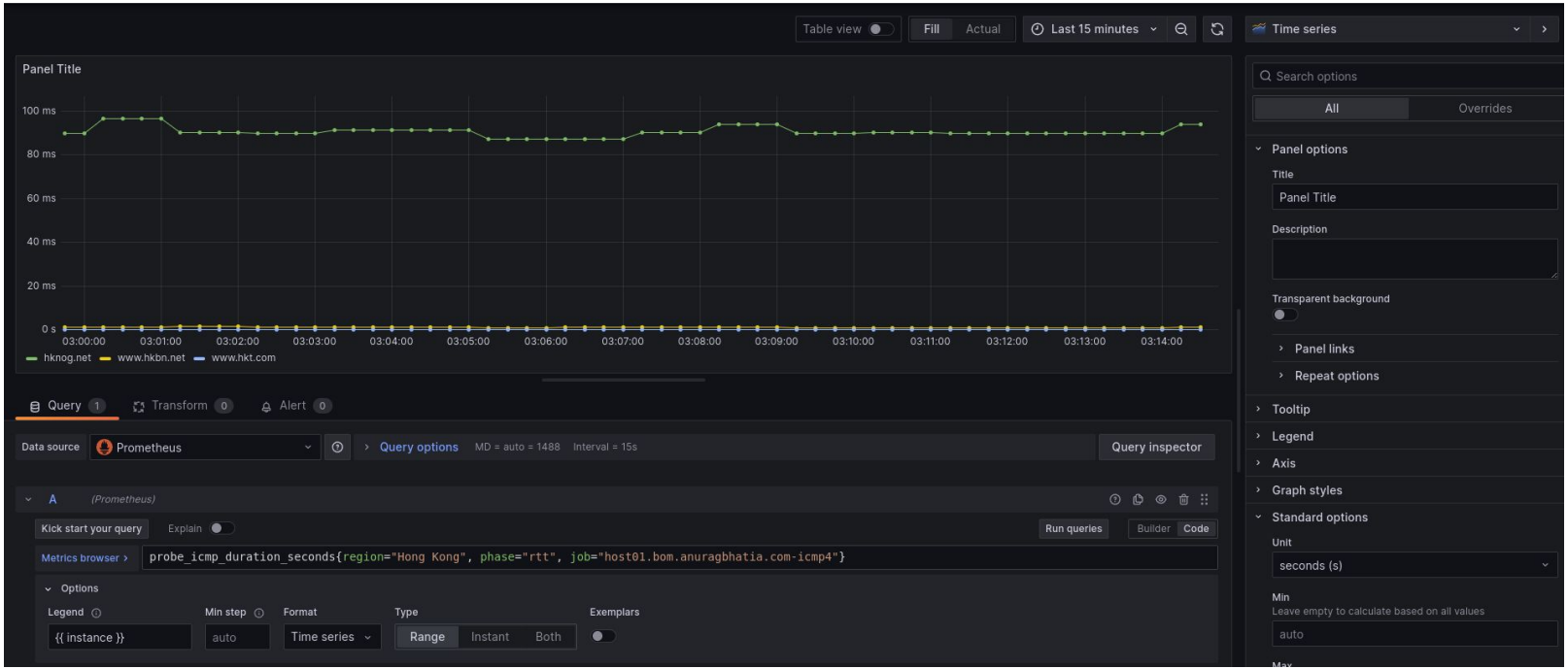


Whatever can be queried, can be plotted...

Grafana query



Grafana query



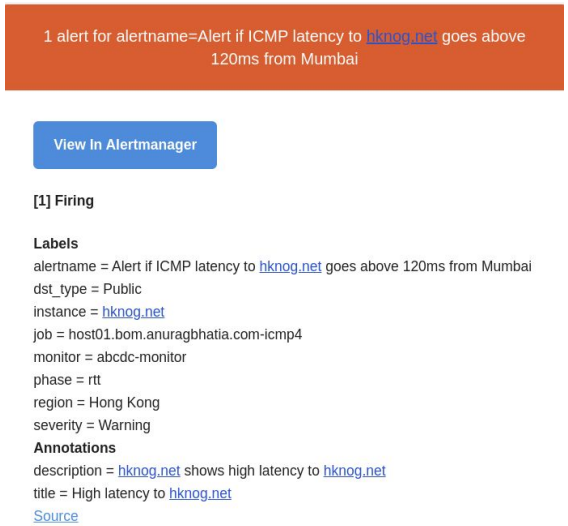
Whatever can be queried can be set to give alerts...

Alerts setup

```
- alert: Alert if ICMP latency to hknog.net goes above 120ms from Mumbai
  expr: probe_icmp_duration_seconds{region="Hong Kong", phase="rtt", instance="hknog.net", job="host01.bom.anuragbhatia.com-icmp4"} > 0.12
  for: 5m
  annotations:
    title: 'High latency to {{ $labels.instance }}'
    description: '{{ $labels.instance }} shows high latency to hknog.net'
  labels:
    severity: 'Warning'
```

```
- alert: Alert if ICMP latency to any destination in Hong Kong goes over 300ms from Germany
  expr: probe_icmp_duration_seconds{region="Hong Kong", phase="rtt", job="server7.anuragbhatia.com-icmp4"} > 0.3
  for: 5m
  annotations:
    title: 'High latency to {{ $labels.instance }}'
    description: '{{ $labels.instance }} shows high latency to Hong Kong'
  labels:
    severity: 'Warning'
```

Alerts setup



Scaling up

- Distribute monitoring endpoints to various probes with logical labels (country, region, type etc)
- Multiple prometheus servers for in hierarchical manner (support for federation)
- Long term retention on S3 endpoints
- Single alert manager running in HA to de-duplicate
- Support via Thanos, Cortex, Grafana mimir etc

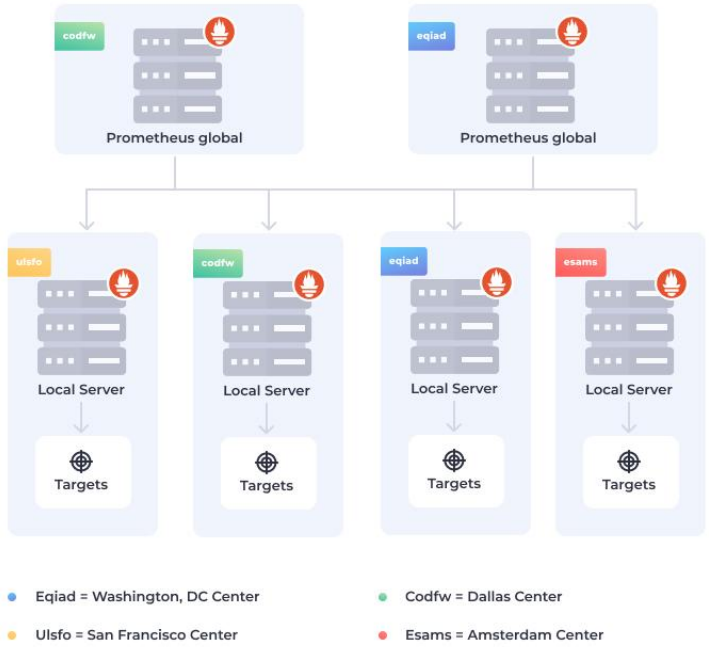
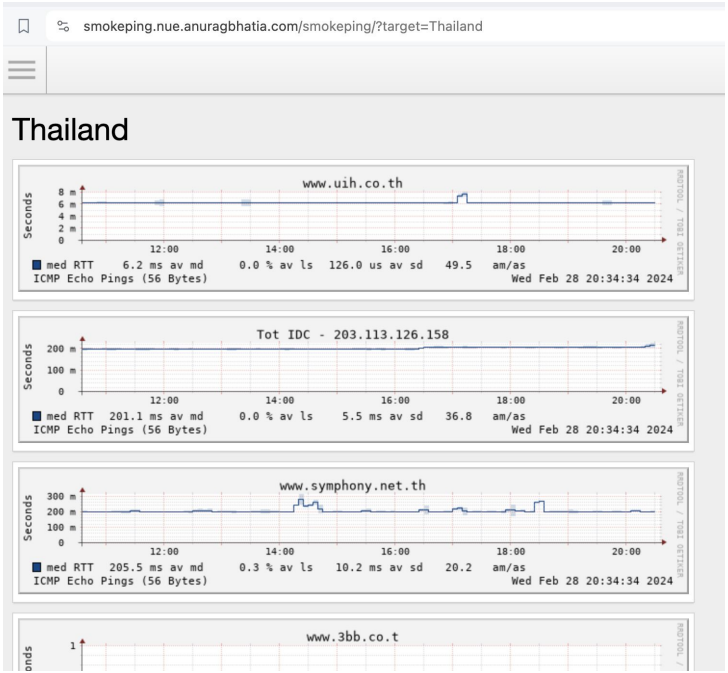
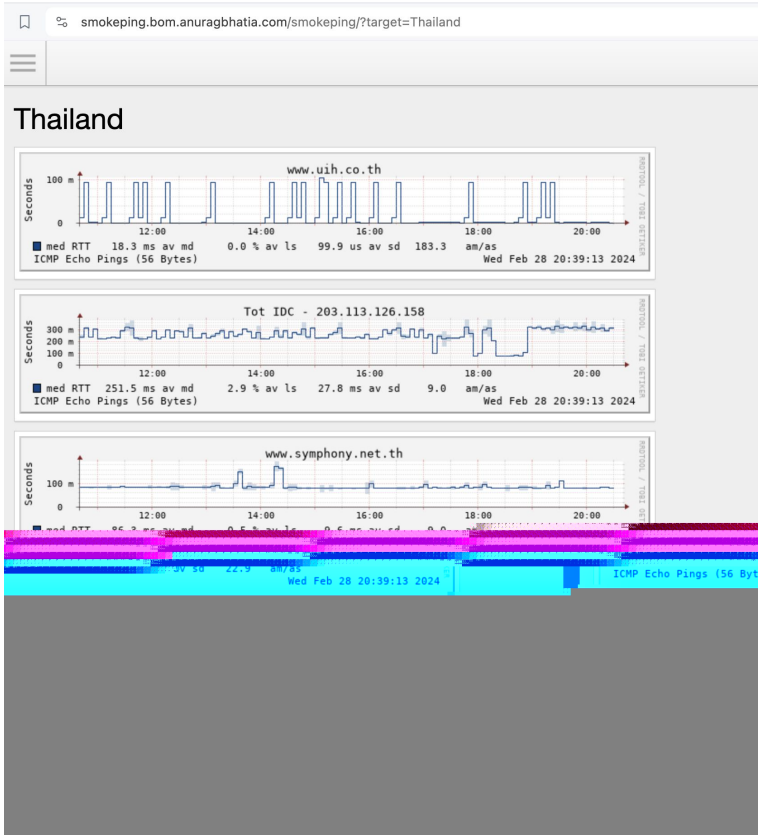


Image source here

How this all fits in?


smokeping.nue.anuragbhatia.com/smokeping/?target=Thailand
Thailand
www.uih.co.th
Seconds
8 m
6 m
4 m
2 m
0
12:00
14:00
16:00
18:00
20:00
med RTT 6.2 ms av md 0.0 % av ls 126.0 us av sd 49.5 am/as
ICMP Echo Pings (56 Bytes) Wed Feb 28 20:34:34 2024
Tot IDC - 203.113.126.158
200 m
100 m
0
med RTT 201.1 ms av md 0.0 % av ls 5.5 ms av sd 36.8 am/as
ICMP Echo Pings (56 Bytes) Wed Feb 28 20:34:34 2024
www.symphony.net.th
300 m
200 m
100 m
0
med RTT 205.5 ms av md 0.3 % av ls 10.2 ms av sd 20.2 am/as
ICMP Echo Pings (56 Bytes) Wed Feb 28 20:34:34 2024
www.3bb.co.t
1


smokeping.bom.anuragbhatia.com/smokeping/?target=Thailand
Thailand
www.uih.co.th
Seconds
100 m
0
12:00
14:00
16:00
18:00
20:00
med RTT 18.3 ms av md 0.0 % av ls 99.9 us av sd 183.3 am/as
ICMP Echo Pings (56 Bytes) Wed Feb 28 20:39:13 2024
Tot IDC - 203.113.126.158
300 m
200 m
100 m
0
med RTT 251.5 ms av md 2.9 % av ls 27.8 ms av sd 9.0 am/as
ICMP Echo Pings (56 Bytes) Wed Feb 28 20:39:13 2024
www.symphony.net.th
100 m
0
Wed Feb 28 20:39:13 2024
ICMP Echo Pings (56 Byt

Search or jump to...
cmd+k
Home › Dashboards › Dev › APRICOT2024
Add
Last 2 days
Nuremberg
300 ms
250 ms
200 ms
150 ms
100 ms
50 ms
0 s
02/27 00:00
02/27 08:00
02/27 16:00
02/28 00:00
02/28 08:00
02/28 16:00
203.113.126.158
www.3bb.co.th
www.dtac.co.th
www.jastel.co.th
www.symphony.net.th
www.uih.co.th
Mumbai
300 ms
250 ms
200 ms
150 ms
100 ms
50 ms
0 s
02/27 00:00
02/27 06:00
02/27 12:00
02/27 18:00
02/28 00:00
02/28 06:00
02/28 12:00
02/28 18:00
203.113.126.158
www.3bb.co.th
www.dtac.co.th
www.jastel.co.th
www.symphony.net.th
www.uih.co.th

What about from distributed?

RIPE Atlas...

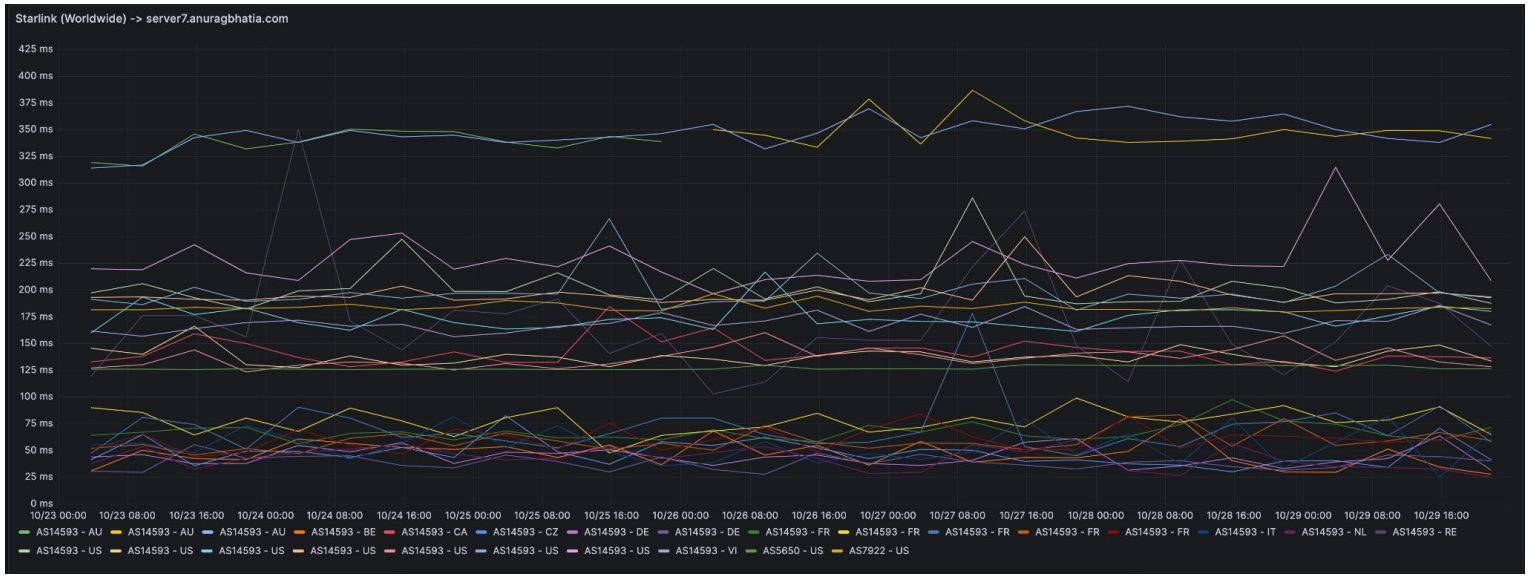
RIPE Atlas Exporter demo

```
anurag@Anurags-MacBook-Pro ~ [SIGINT]> curl -s "http://lo.server7.anuragbhatia.com:9400/metrics?measurement_id=61354401"
# HELP atlas_ping_avg_latency Average latency
# TYPE atlas_ping_avg_latency gauge
atlas_ping_avg_latency{asn="14593",country_code="AU",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="-34.9315",long="138.6015",measurement="61354401",probe="60892"} 344.193365
atlas_ping_avg_latency{asn="14593",country_code="AU",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="-38.1815",long="146.2495",measurement="61354401",probe="24742"} 348.14659175
atlas_ping_avg_latency{asn="14593",country_code="BE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.9605",long="4.9295",measurement="61354401",probe="1001356"} 40.37865875
atlas_ping_avg_latency{asn="14593",country_code="CA",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="45.4575",long="-76.2025",measurement="61354401",probe="60510"} 137.333319
atlas_ping_avg_latency{asn="14593",country_code="CZ",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="50.0885",long="14.4085",measurement="61354401",probe="1005623"} 45.92677475
atlas_ping_avg_latency{asn="14593",country_code="DE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.1185",long="9.1515",measurement="61354401",probe="1006382"} 45.236784
atlas_ping_avg_latency{asn="14593",country_code="DE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.7315",long="7.7715",measurement="61354401",probe="1006388"} 52.0565655
atlas_ping_avg_latency{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="43.8485",long="1.3995",measurement="61354401",probe="62843"} 70.4949145
atlas_ping_avg_latency{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="44.4075",long="6.4495",measurement="61354401",probe="13040"} 101.4165485
atlas_ping_avg_latency{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="48.6405",long="2.2315",measurement="61354401",probe="32686"} 51.4994225
atlas_ping_avg_latency{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="48.7475",long="2.4895",measurement="61354401",probe="61241"} 63.706624
atlas_ping_avg_latency{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="48.9475",long="2.5005",measurement="61354401",probe="16971"} 58.88381425
atlas_ping_avg_latency{asn="14593",country_code="IT",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="45.2505",long="8.8605",measurement="61354401",probe="1004876"} 44.22689575
atlas_ping_avg_latency{asn="14593",country_code="RE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="-20.8795",long="55.4515",measurement="61354401",probe="60797"} 117.216932
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="37.0415",long="-121.9915",measurement="61354401",probe="60929"} 204.364665
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="40.7105",long="-74.0115",measurement="61354401",probe="61537"} 138.8093585
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="40.8585",long="-102.8625",measurement="61354401",probe="62613"} 191.362956
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="43.2505",long="-124.3915",measurement="61354401",probe="23127"} 197.64231475
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="43.9305",long="-73.2925",measurement="61354401",probe="63017"} 135.19330125
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="46.5015",long="-122.9685",measurement="61354401",probe="62498"} 192.23980175
atlas_ping_avg_latency{asn="14593",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="61.5685",long="-149.0125",measurement="61354401",probe="61113"} 232.5192975
atlas_ping_avg_latency{asn="14593",country_code="VI",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="18.3375",long="-64.9325",measurement="61354401",probe="62911"} 173.17183475
atlas_ping_avg_latency{asn="5650",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="37.7985",long="-89.0225",measurement="61354401",probe="1005302"} 125.86889525
atlas_ping_avg_latency{asn="7922",country_code="US",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="47.6495",long="-122.5425",measurement="61354401",probe="61105"} 182.04190775
# HELP atlas_ping_dup Number of duplicate icmp repsponses
# TYPE atlas_ping_dup gauge
atlas_ping_dup{asn="14593",country_code="AU",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="-34.9315",long="138.6015",measurement="61354401",probe="60892"} 0
atlas_ping_dup{asn="14593",country_code="AU",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="-38.1815",long="146.2495",measurement="61354401",probe="24742"} 0
atlas_ping_dup{asn="14593",country_code="BE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.9605",long="4.9295",measurement="61354401",probe="1001356"} 0
atlas_ping_dup{asn="14593",country_code="CA",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="45.4575",long="-76.2025",measurement="61354401",probe="60510"} 0
atlas_ping_dup{asn="14593",country_code="CZ",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="50.0885",long="14.4085",measurement="61354401",probe="1005623"} 0
atlas_ping_dup{asn="14593",country_code="DE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.1185",long="9.1515",measurement="61354401",probe="1006382"} 0
atlas_ping_dup{asn="14593",country_code="DE",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="49.7315",long="7.7715",measurement="61354401",probe="1006388"} 0
atlas_ping_dup{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="43.8485",long="1.3995",measurement="61354401",probe="62843"} 0
atlas_ping_dup{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="44.4075",long="6.4495",measurement="61354401",probe="13040"} 0
atlas_ping_dup{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="48.6405",long="2.2315",measurement="61354401",probe="32686"} 0
atlas_ping_dup{asn="14593",country_code="FR",dst_addr="144.91.67.7",dst_name="144.91.67.7",ip_version="4",lat="48.7475",long="2.4895",measurement="61354401",probe="61241"} 0
```

RIPE Atlas Exporter config sample...

```yaml
- job_name: 'atlas_exporter-server7-Starlink'
  scrape_interval: 6h
  static_configs:
    - targets:
      - 61353502 # Starlink to IAXN Rohtak
      - 61354232 # Starlink to host01.fmt.anuragbhatia.com
      - 61354401 # Starlink to server7.anuragbhatia.com
  relabel_configs:
    - source_labels: [__address__]
      regex: (.*)(:80)?
      target_label: __param_measurement_id
      replacement: ${1}
    - source_labels: [__param_measurement_id]
      regex: (.*)
      target_label: instance
      replacement: ${1}
    - source_labels: []
      regex: .*
      target_label: __address__
      replacement: lo.server7.anuragbhatia.com:9400
```

Startlink -> My server in Nuremberg, Germany



Summary

- Metrics world is becoming standard for tooling
- Systems like Prometheus scale well. While initial learning curve is more than single integrated tools like Smokeping but these tools scale well
- There are many components which may or may not be used depending on the need like federation, object storage offload etc.
- Labels play an important role for both graphing as well as alerting.

References

1. Prometheus (here)
2. Node exporter (here)
3. Blackbox exporter (here)
4. Alert Manager (here)
5. Blog post Monitoring my home network by Karan Sharma (here)
6. Replacing Smokeping with Prometheus (here)
7. Smokeping_prober (here)
8. Scaling up - Thaos, Cortex and Grafana mimir
9. RIPE Atlas Exporter (here)