

RIPE Atlas in Google BigQuery

Stephen Strowes <<u>sds@ripe.net</u>> | 2020-10-26 | Academic and NREN Session

RIPE Atlas

- Continuously measures the Internet
 - 145 million traceroutes per day,
 - 190 million DNS queries per day,
 - 500 million pings per day
- This becomes unwieldy!



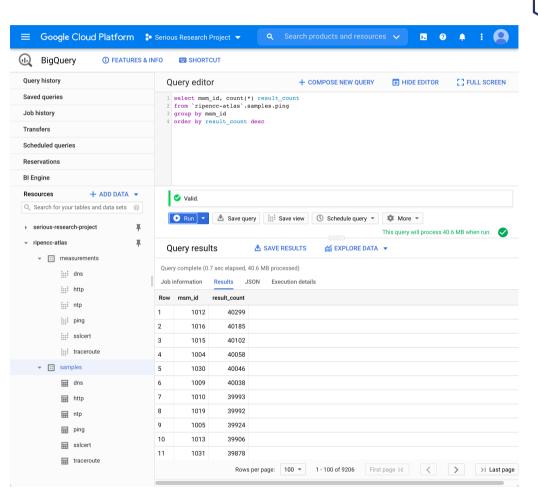




Announcing...

From today,

We're opening public access to our data on BigQuery



BigQuery: project & datasets



- https://console.cloud.google.com/bigquery?project=ripencc-atlas
 - measurements dataset:
 - public measurement results for DNS, traceroute, ping, HTTP, SSL, and NTP
 - dataset is current: streaming results
 - starting from 1 January 2020; we'll increase the date range as we backfill
 - samples dataset:
 - 1% of results from 1 week in October 2020, for each of the above
 - play with it!
- You need a Google account
 - we're covering storage, queries are on you

Why BigQuery?

- Can manage data like this without flinching
- Few hardware limitations
- Queries are often fast, but not real-time
 - think: O(seconds minutes), not O(milliseconds)
- Allows iterative exploration of the data
 - ask for results related to ANY measurement
 - combine with other datasets you have access to



More soon



- Currently we have measurement data
- We are planning to add other common resources soon
 - probe meta data
 - BGP data

This is an active beta project

- We've been building experience but in reality:
 - we need to know what you use this for
 - what data works, and what does not?
 - what common datasets could we build and store?

More Information



- Information:
 - <u>https://labs.ripe.net/tools/ripe-atlas-on-bigquery/ripe-atlas-on-bigquery</u>
 - <u>https://labs.ripe.net/Members/stephen_strowes/announcing-ripe-atlas-data-on-google-bigquery</u>
- Docs, HOWTOs, schemas, and so forth:
 - <u>https://github.com/RIPE-NCC/ripe-atlas-bigquery/</u>



Questions

atlas-bq@ripe.net