

# Measurement Lab Tutorial

Georgia Bullen, [georgia@measurementlab.net](mailto:georgia@measurementlab.net)

Stephen Soltesz, [soltesz@google.com](mailto:soltesz@google.com)



# Question

Can the M-Lab dataset be useful to you?

# Goals

- Introduce two M-Lab datasets: NDT (speed test) & Paris-Traceroute
- Get Permission to Access and Run BigQuery queries
- Visualize results in Data Studio
- Optional: advanced queries & visualization in Jupyter Notebooks



Today — **500+** Servers in **130+** locations

# Running an M-Lab Speed Test

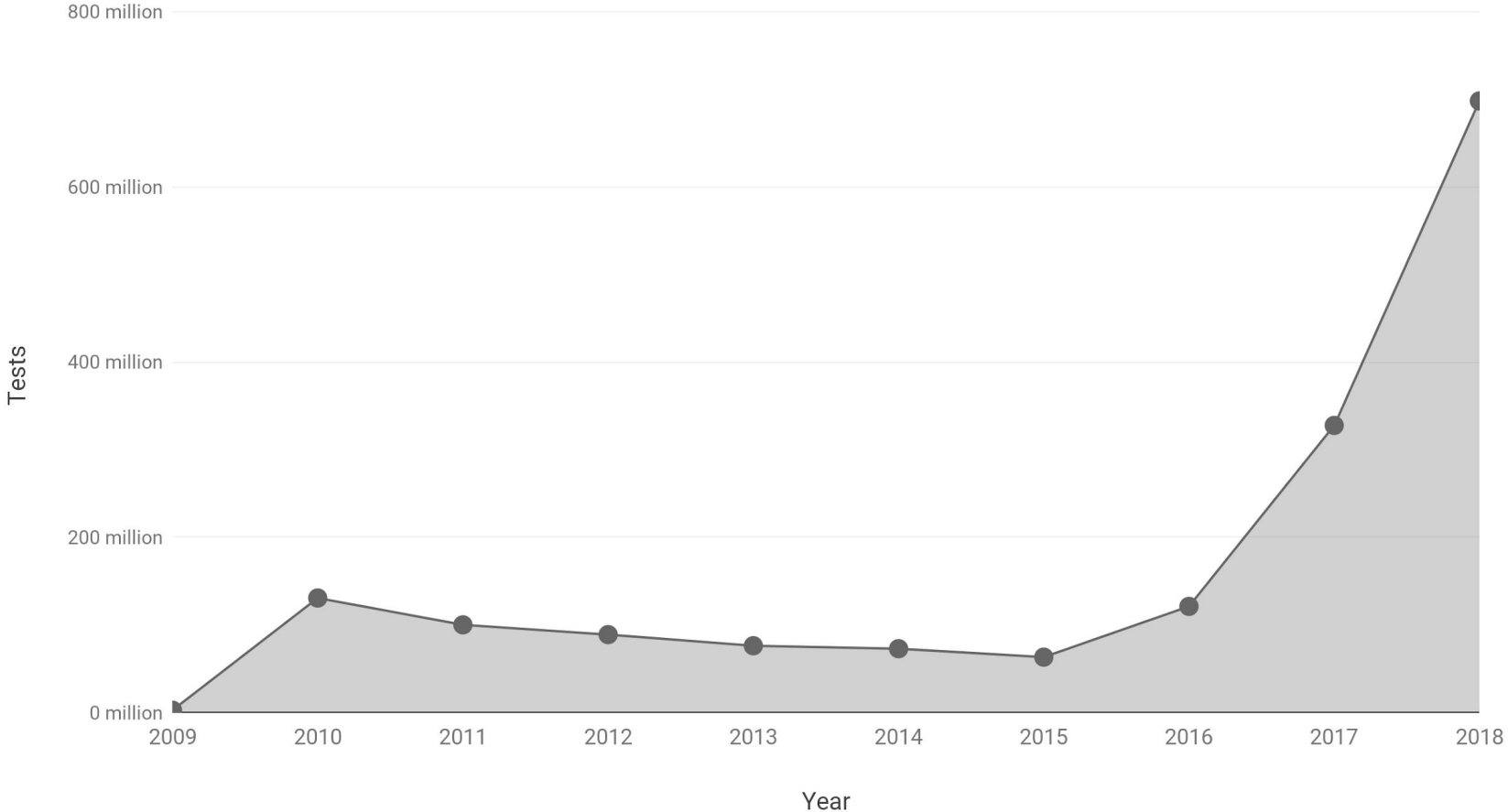
On demand:

- Browser - [speed.measurementlab.net](https://speed.measurementlab.net)
- uTorrent - [www.utorrent.com](http://www.utorrent.com)
- Android - OONI Probe App in Play Store

Scheduled:

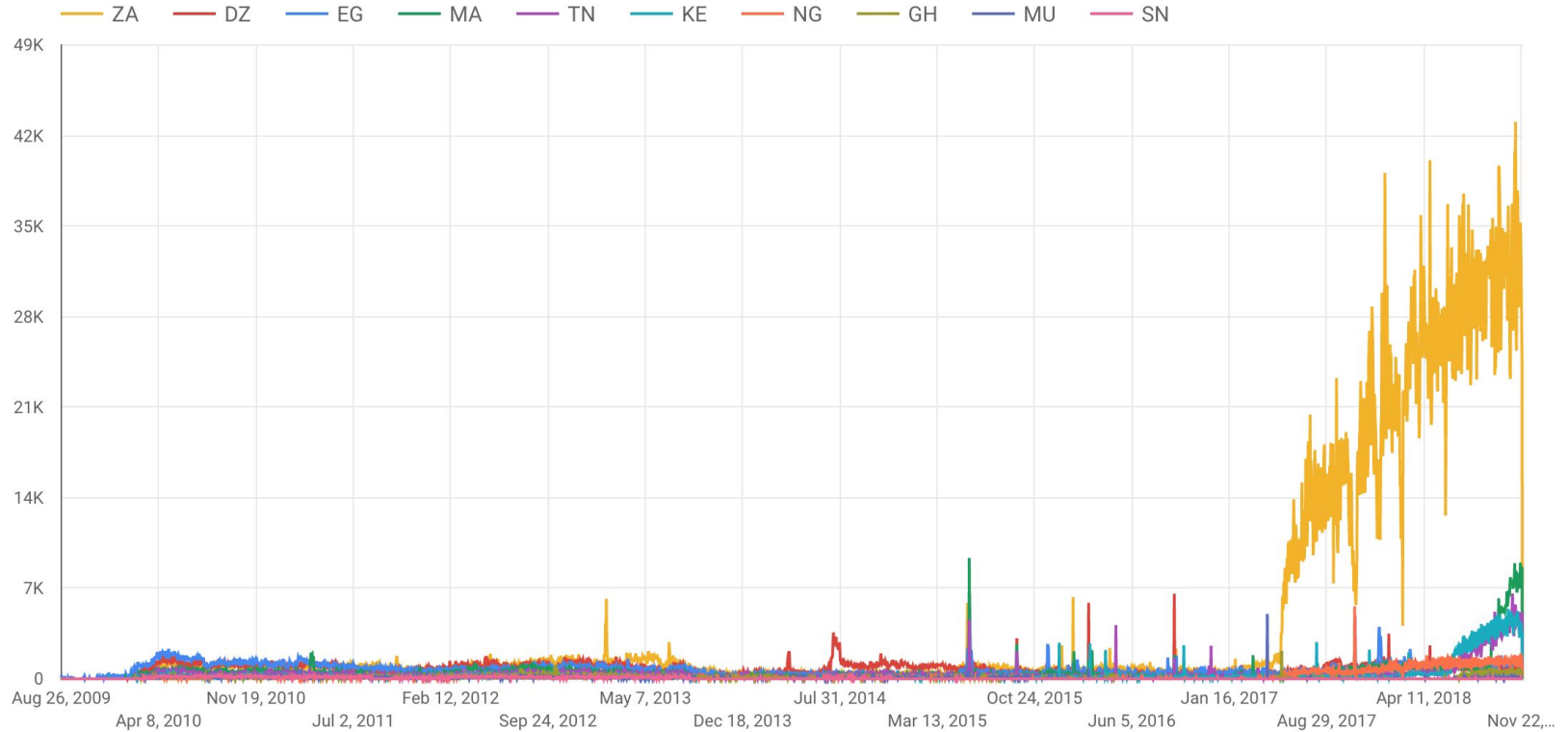
- Docker - Raspberry Pi - [github.com/m-lab/murakami](https://github.com/m-lab/murakami)
- Chrome Extension - “M-Lab Measure”

NDT Tests per Year



2018 numbers projected: current total / (7/12)

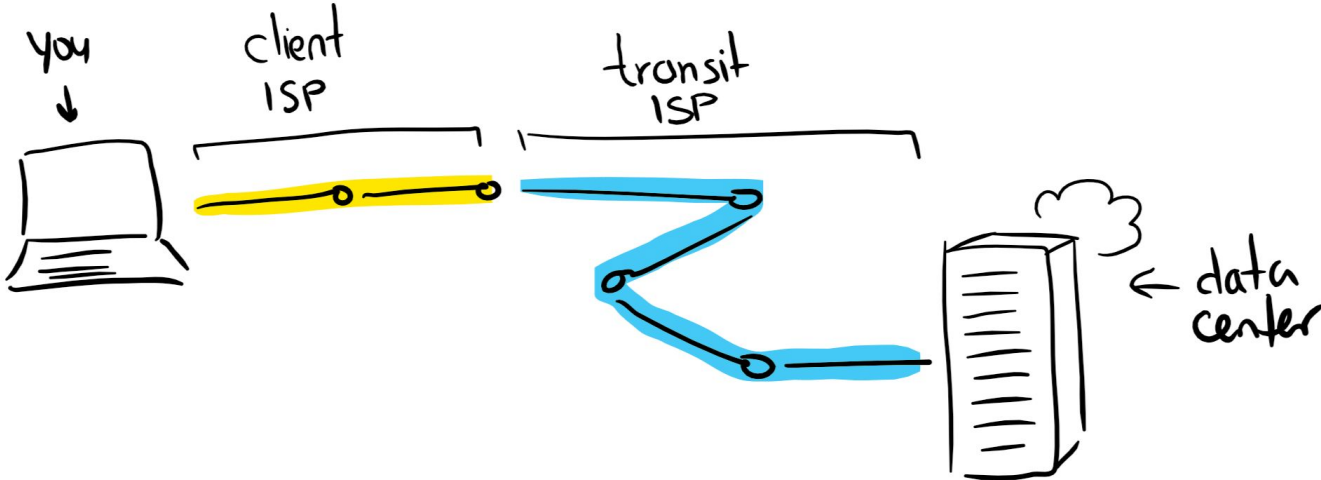
# NDT Tests in Africa



# What we measure

---

M-lab measures user experience of the **full route** from user to content

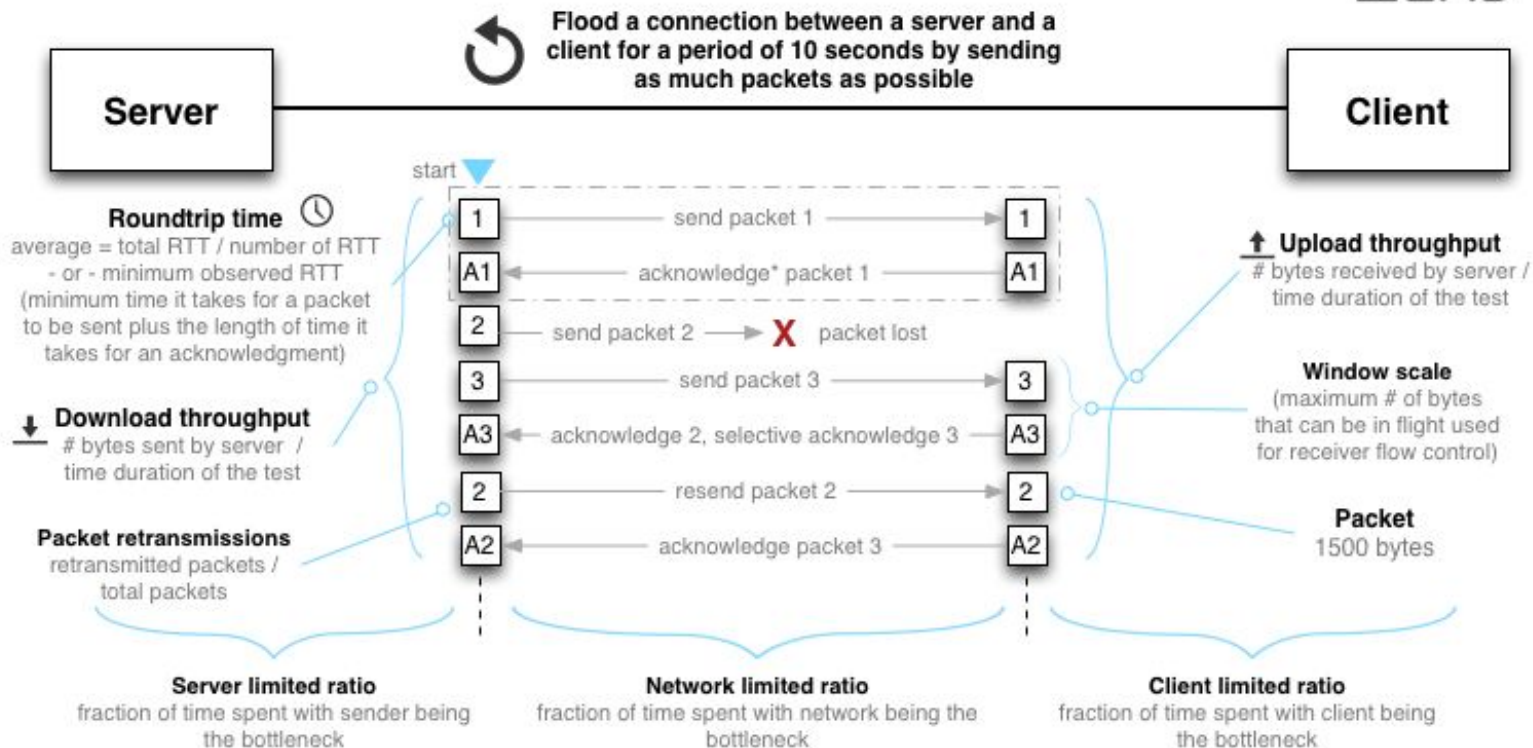




# NDT

## Network Diagnostics Tool

MLAB



\* acknowledge means cumulative acknowledgement

# NDT BigQuery Schema

`test_id & log_time & parse_time` - metadata for every row

## [NDT Archive Data](#)

`connection_spec.*` - client metadata

`connection_spec.client_geolocation.*` - lat/lon, country, region, etc

## [NDT BQ Schema](#)

`web100_log_entry.connection_spec.*` - server and client IP & ports

`web100_log_entry.snap.*` - Web100 metrics

## [Web100 Docs](#)

# NDT BigQuery Schema (cont)

`web100_log_entry.snap.*` - Web100 metrics

[NDT Archive Data](#)

`connection_spec.data_direction` = 1 / download - 0 / upload

[NDT BQ Schema](#)

`web100_log_entry.snap.HCThruOctetsAcked` - download byte count

`web100_log_entry.snap.HCThruOctetsReceived` - upload byte count

[Web100 Docs](#)

`web100_log_entry.snap.SndLimTimeRwin` - Receiver Limited Time

`web100_log_entry.snap.SndLimTimeCwnd` - Congestion Limited Time

`web100_log_entry.snap.SndLimTimeSnd` - Sender Limited Time

`web100_log_entry.snap.CongSignals` - Total congestion events

# Access to M-Lab Data

Quick Start - [www.measurementlab.net/quickstart/](http://www.measurementlab.net/quickstart/)

# Query: Counting NDT Tests in African Countries

```
SELECT
    connection_spec.client_geolocation.country_code AS country_code,
    COUNT(*) AS count

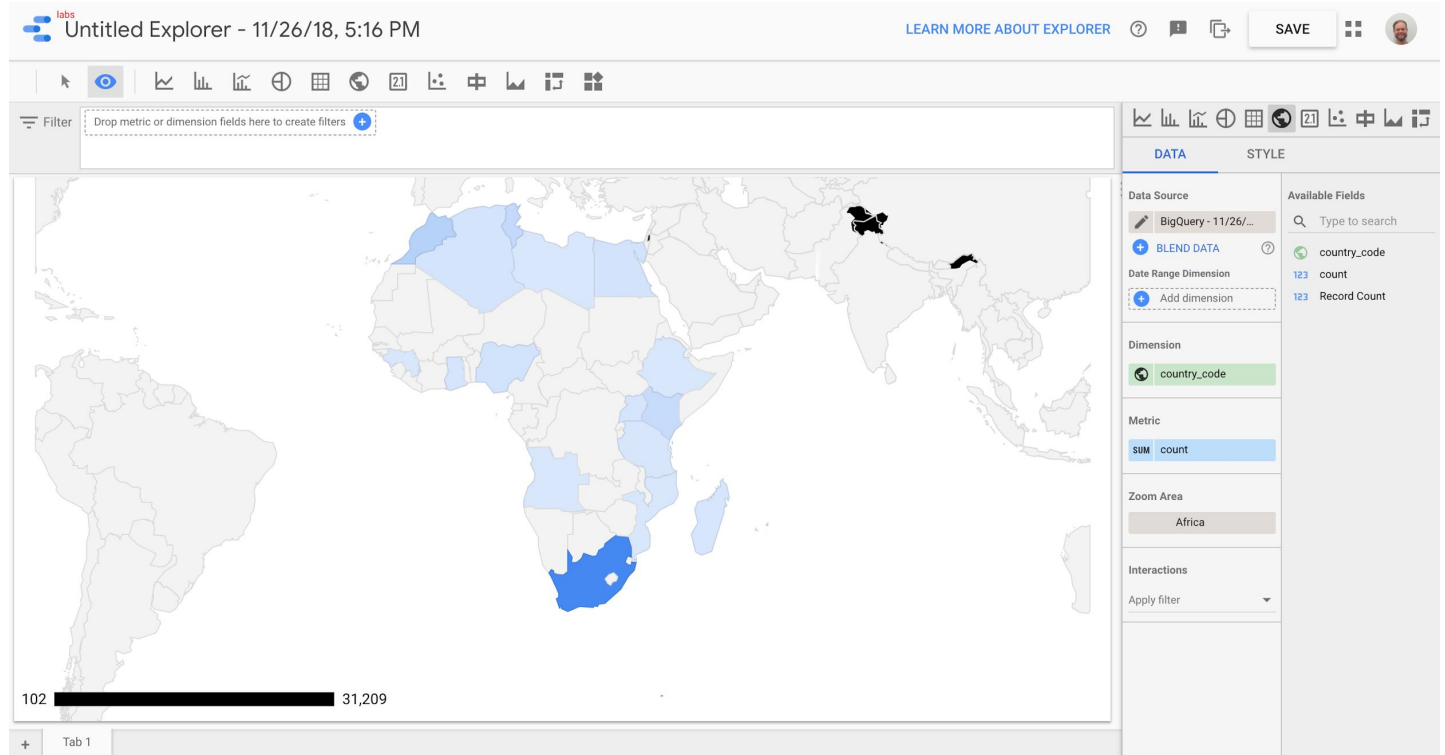
FROM
    `measurement-lab.ndt.base`

WHERE
    TIMESTAMP_TRUNC(log_time, DAY) >= TIMESTAMP("2018-10-01")
    AND connection_spec.client_geolocation.continent_code = "AF"

GROUP BY
    country_code
```

# BigQuery & Data Studio (NDT)

Query



# Query: Counting NDT Tests Over Time

```
SELECT
    TIMESTAMP_TRUNC(log_time, DAY) AS day,
    connection_spec.client_geolocation.country_code AS country_code,
    COUNT(*) AS count

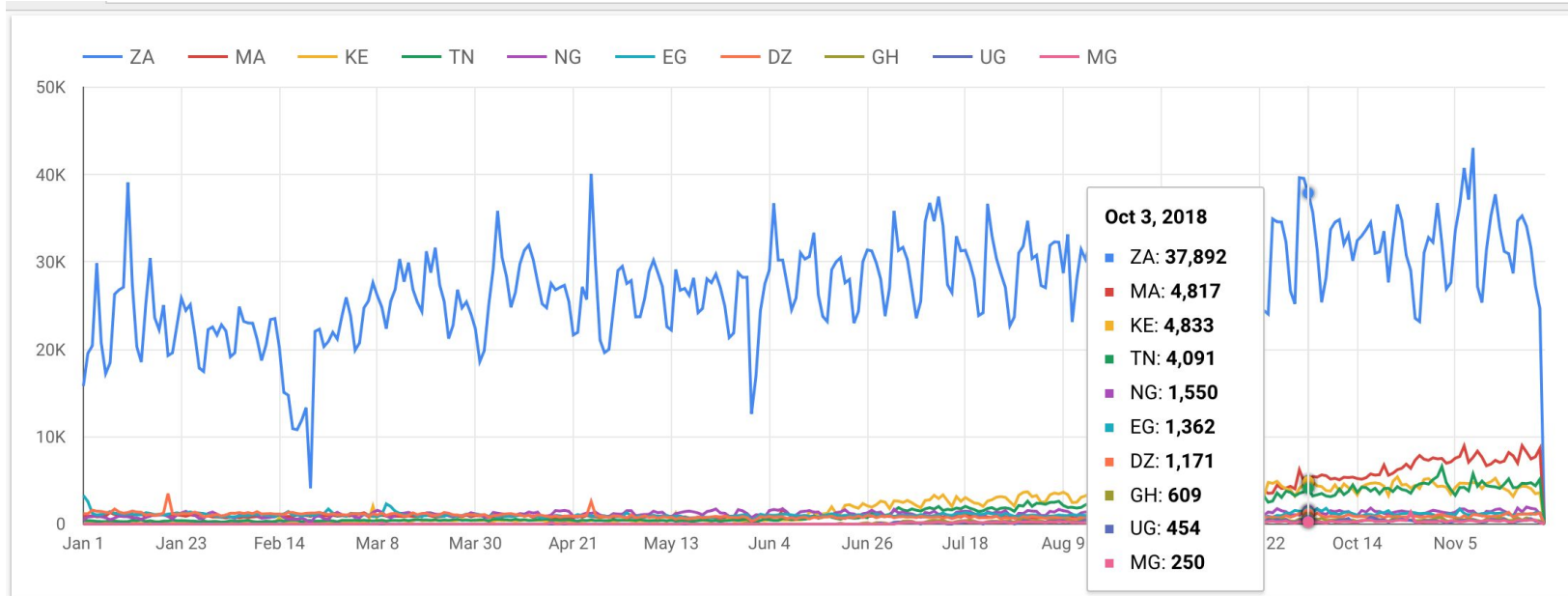
FROM
    `measurement-lab.ndt.base`

WHERE
    TIMESTAMP_TRUNC(log_time, DAY) >= TIMESTAMP("2018-01-01")
    AND connection_spec.client_geolocation.continent_code = "AF"

GROUP BY
    day, country_code
```

# BigQuery & Data Studio (NDT)

## Query





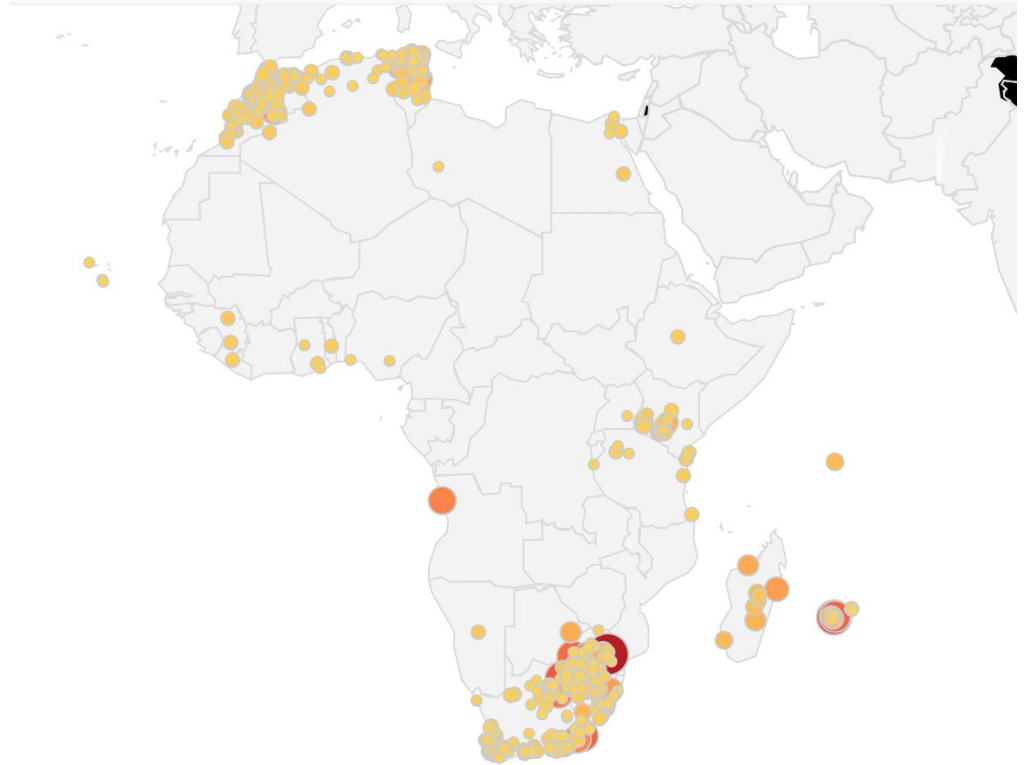
# Query: Average Download Rates in Africa

```
SELECT
  connection_spec.client_geolocation.country_code as country_code,
  AVG(8 * (web100_log_entry.snap.HCThruOctetsAcked /
    (web100_log_entry.snap.SndLimTimeRwin +
    web100_log_entry.snap.SndLimTimeCwnd +
    web100_log_entry.snap.SndLimTimeSnd))) AS download_Mbps
FROM
  `measurement-lab.ndt.base`
WHERE
  TIMESTAMP_TRUNC(log_time, DAY) = TIMESTAMP("2018-11-16")
  AND connection_spec.client_geolocation.continent_code = "AF"
  AND connection_spec.data_direction = 1
  AND web100_log_entry.snap.HCThruOctetsAcked >= 8192
  AND (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd
    + web100_log_entry.snap.SndLimTimeSnd) >= 9000000
  AND (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd
    + web100_log_entry.snap.SndLimTimeSnd) < 600000000
  AND web100_log_entry.snap.CongSignals > 0

GROUP BY
  country_code
```

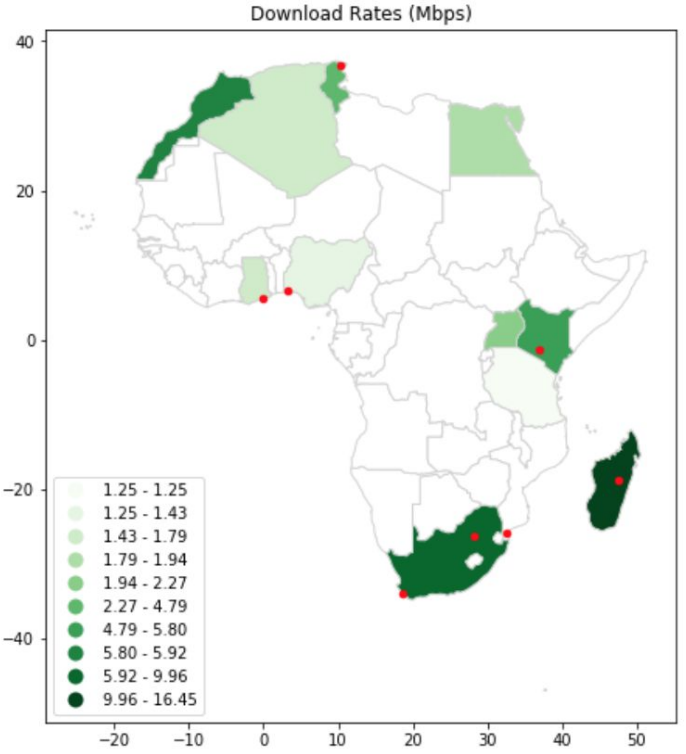
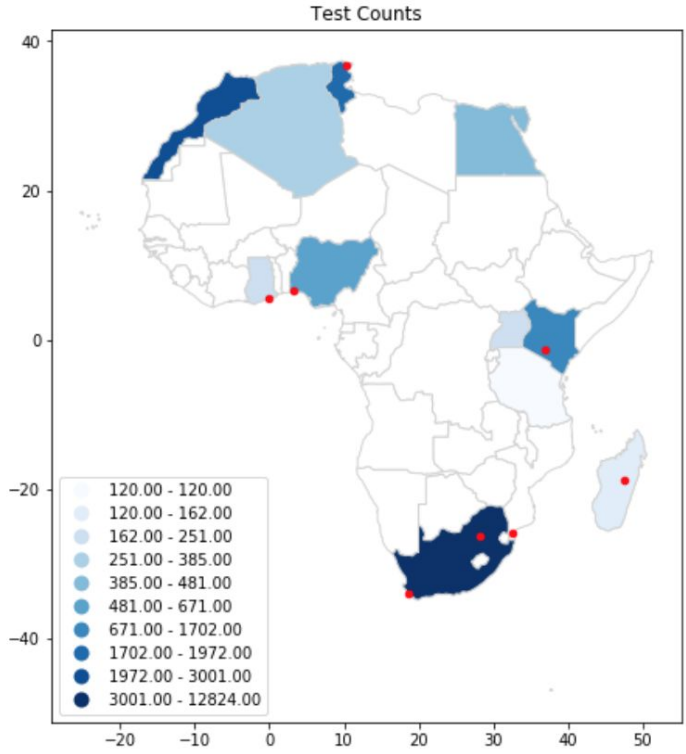
# BigQuery & Data Studio (NDT Downloads)

[Query](#)



# Exploring African Internet (with Jupyter)

[Notebook](#)

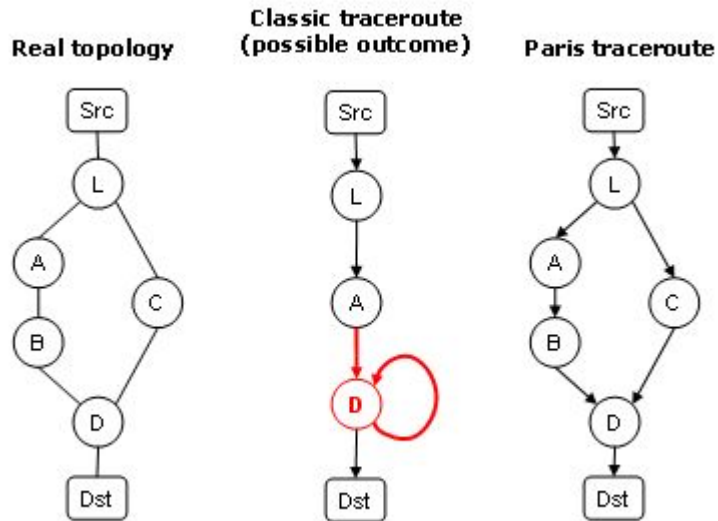


# Paris-Traceroute

---

Paris Traceroute detects load balancing, noting split paths in output.

M-Lab servers collect paris-traceroute information for every measurement.



# Paris-Traceroute BigQuery Schema

## [Traceroute Archive Data](#)

`test_id, log_time, parse_time` - same as NDT

## [Traceroute BQ Schema](#)

`connection_spec.*` - original client / server IPs

`connection_spec.client_geolocation.*` - lat/lon, country, region

## [paris-traceroute.net](#)

`paris_traceroute_hop.*` - each src-to-dest hop of traceroute

`paris_traceroute_hop.src_geolocation.*` - lat/lon country, etc.

`paris_traceroute_hop.dest_geolocation.*` - lat/lon country, etc.

`paris_traceroute_hop.rtt` - RTT of hop

# Query: Counting Traceroute Hops in a Country

```
SELECT
    paris_traceroute_hop.dest_geolocation.country_code AS country_code,
    COUNT(*) AS hops

FROM
    `measurement-lab.traceroute.base`

WHERE
    TIMESTAMP_TRUNC(log_time, DAY) = TIMESTAMP("2018-11-16")
    AND REGEXP_CONTAINS(test_id, r"mlab[1-4].los\d\d")
    AND connection_spec.client_geolocation.country_code = "NG"
    AND (
        REGEXP_CONTAINS(test_id, ".*3010.paris.gz")
        OR REGEXP_CONTAINS(test_id, ".*3001.paris.gz"))

GROUP BY
    country_code
```

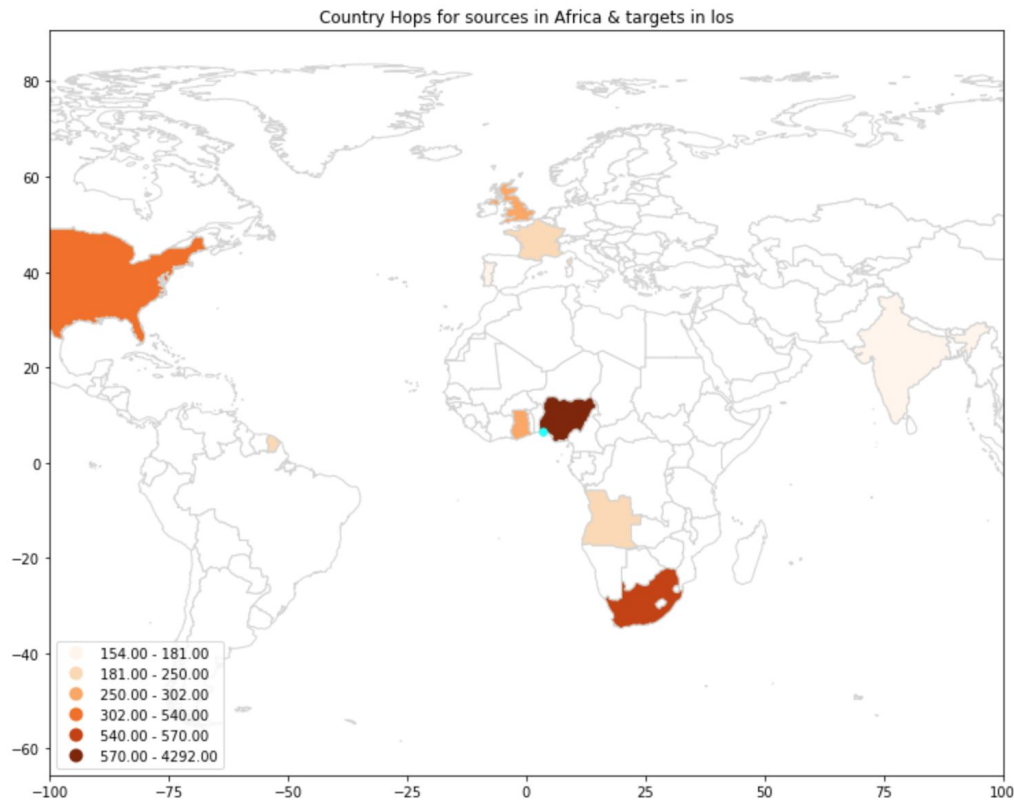
# BigQuery & Data Studio (traceroute)

[Query](#)



# Exploring African Internet (with Jupyter)

[Notebook](#)





# Question

Can the M-Lab dataset be useful to you?

# Resources

<https://www.submarinecablemap.com/>