

# DNSql

## Processing Massive DNS Collections

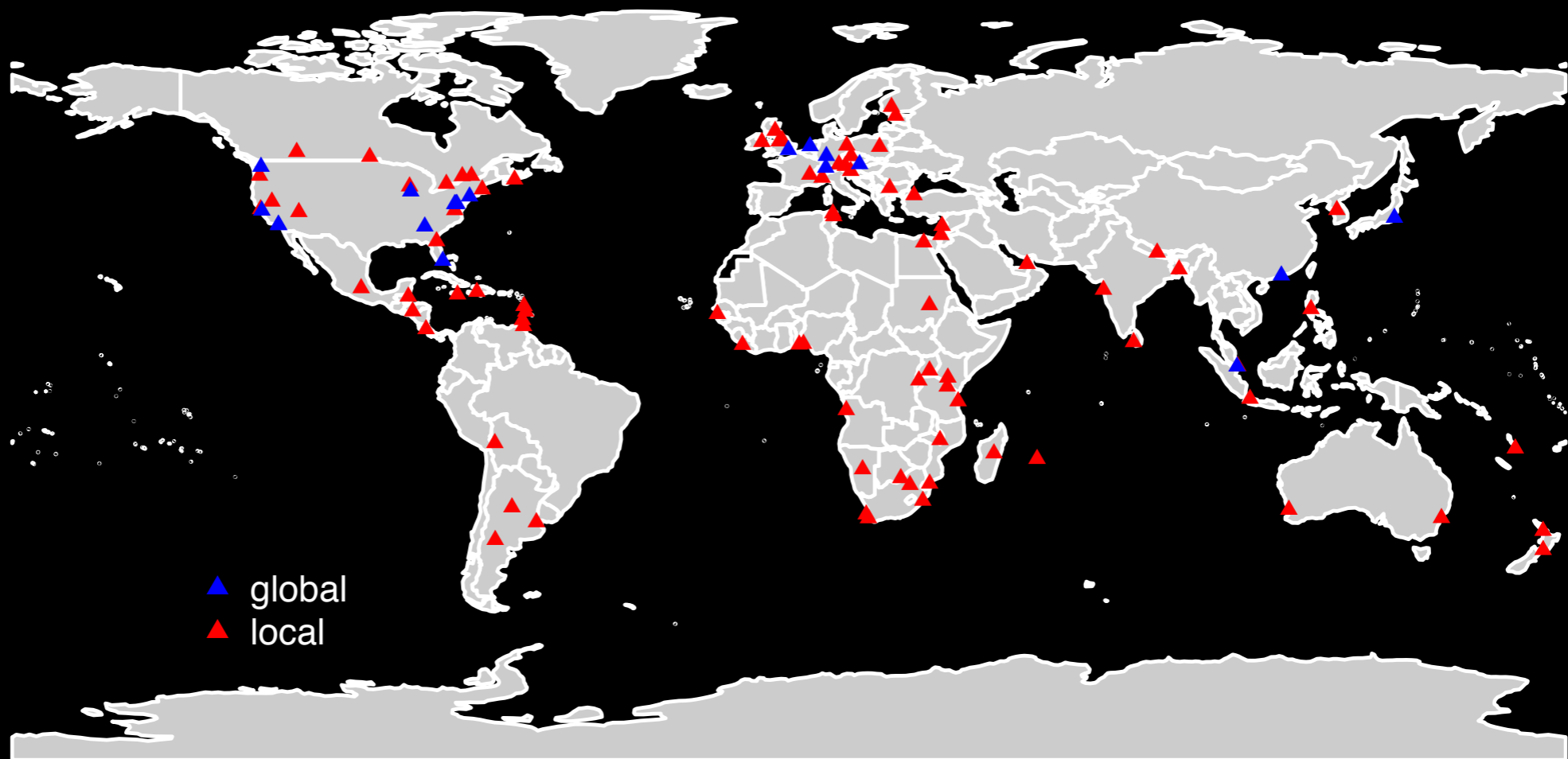
Stephen Herwig, Dave Levin, Bobby Bhattacharjee, Neil Spring  
*University of Maryland, College Park*

# D-root

Operated by UMD

Anycast with 109 replicas

Hourly sampled collection by replica



# Problem

Lots of data

~140 GiB / day

Serial processing is slow

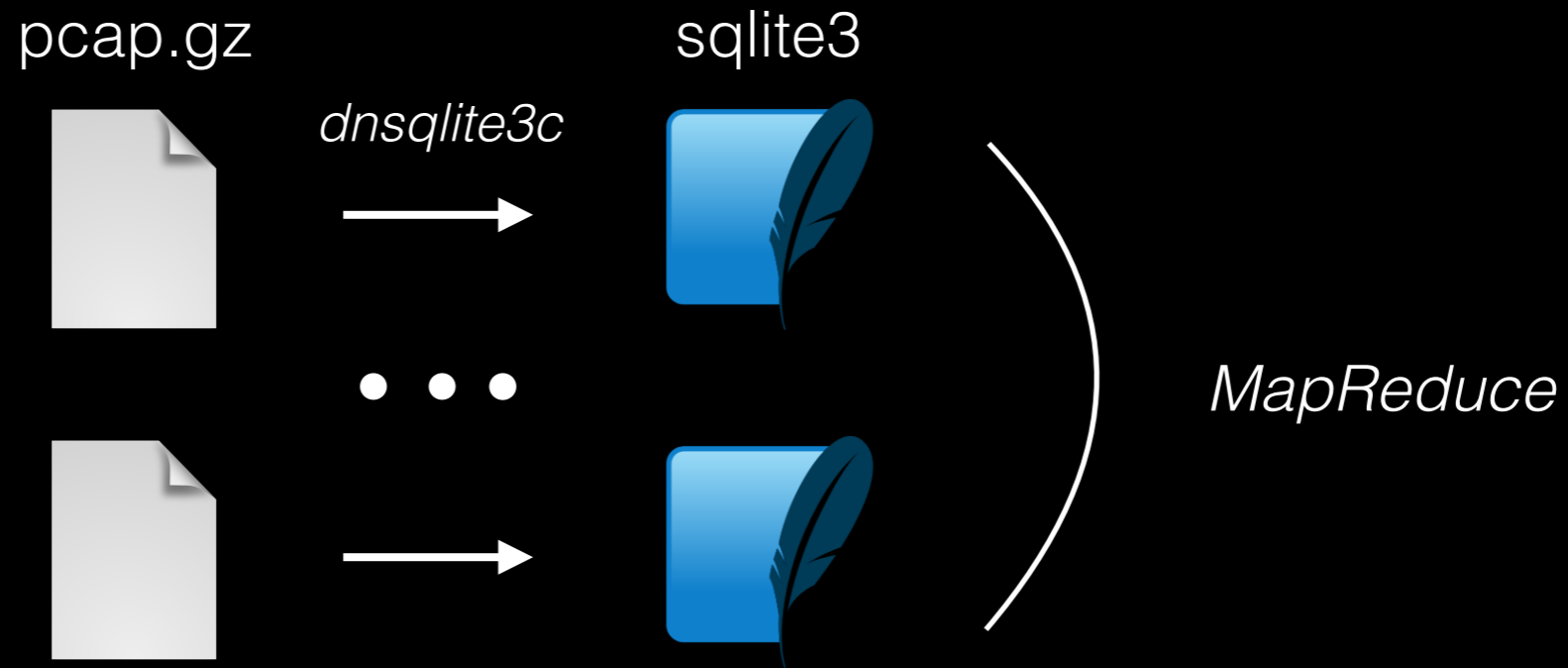
~8h to read a month's worth of collection for CPMD replica

Diverse analyses

Short-term, Long-Term

Aggregation by source, replica, geography, topology

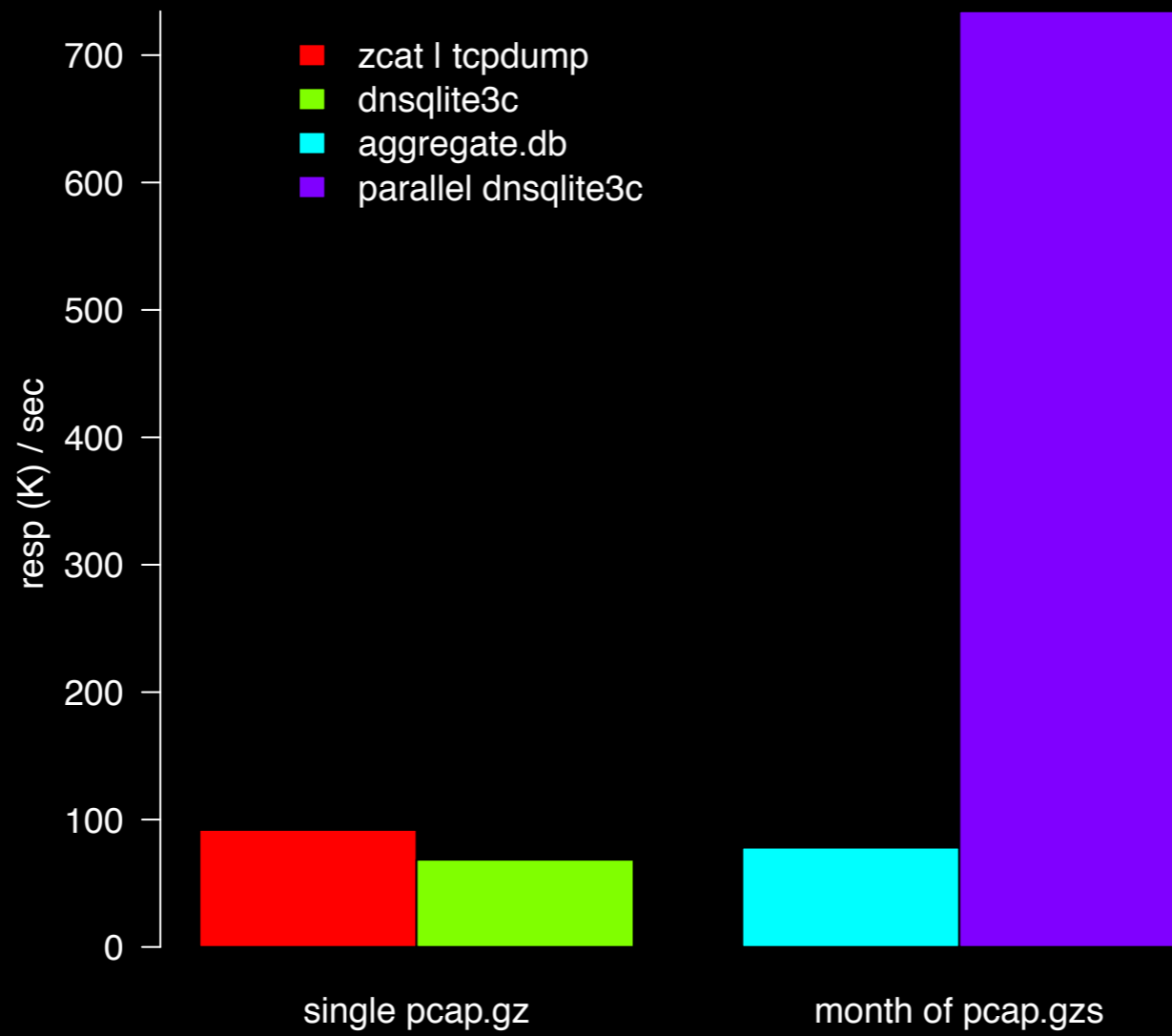
# Approach



```
CREATE TABLE queryresp (  
  id      INTEGER PRIMARY KEY,  
  sec     INTEGER,  
  usec   INTEGER,  
  src     BLOB,  
  sport  INTEGER,  
  opcode INTEGER,  
  qclass  INTEGER,  
  qtype   INTEGER,  
  rcode   INTEGER,  
  qname   TEXT  
);  
CREATE INDEX qname_index ON queryresp(qname);  
CREATE INDEX src_index ON queryresp(src);  
CREATE TABLE qps (sec INTEGER, n INTEGER);
```

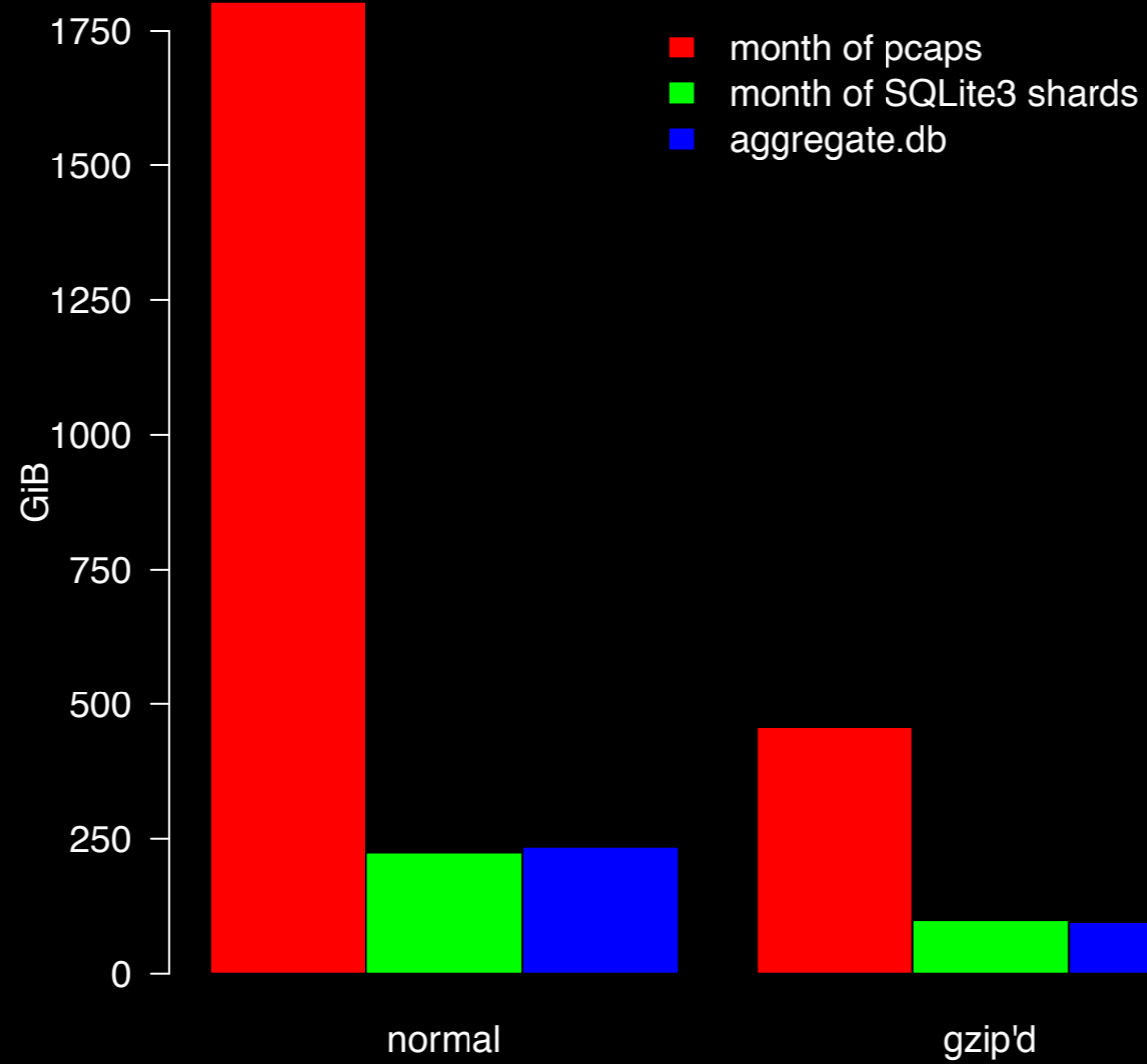
# Processing Speed

CPMD March 2015

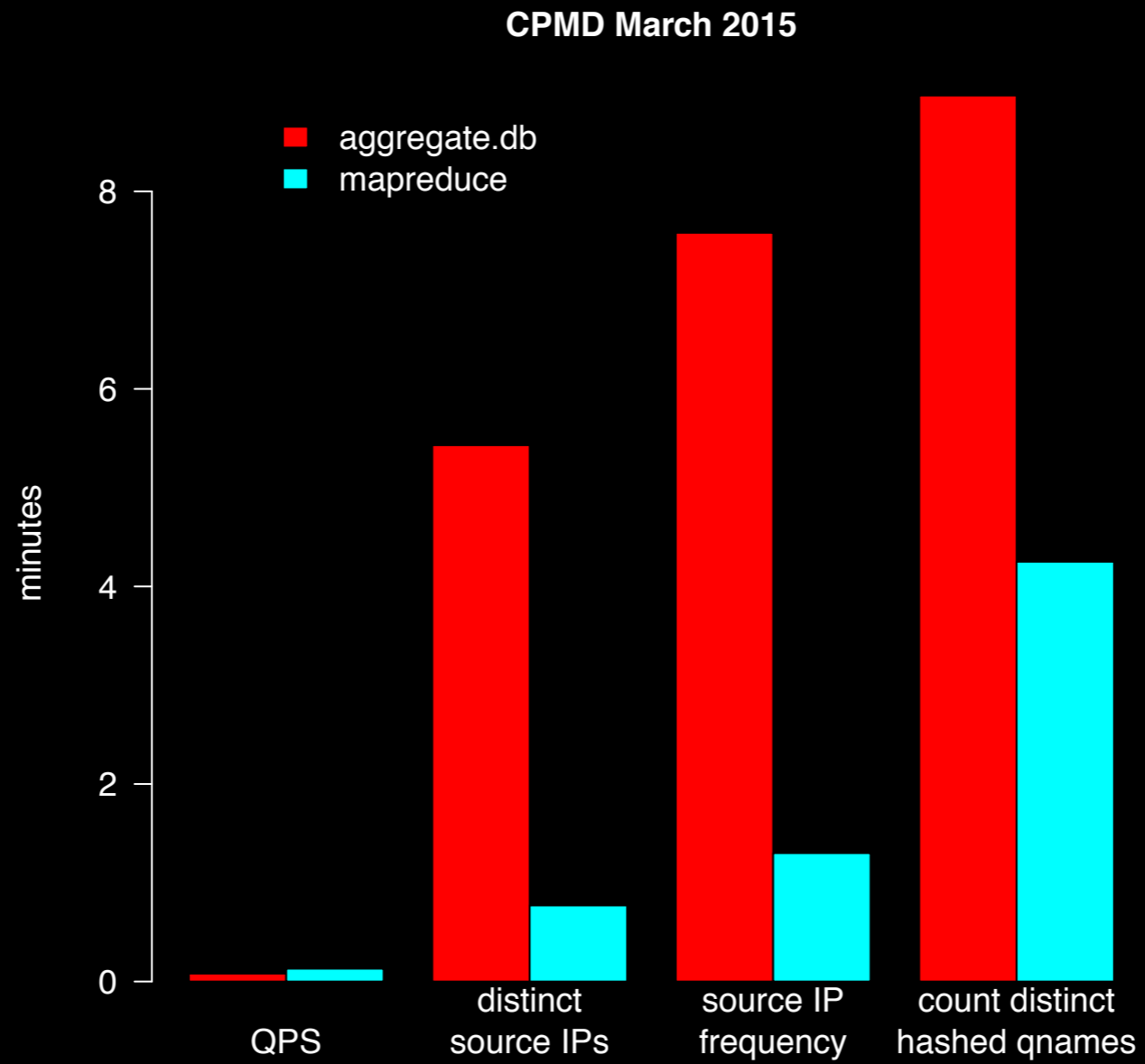


# Database Size

CPMD March 2015

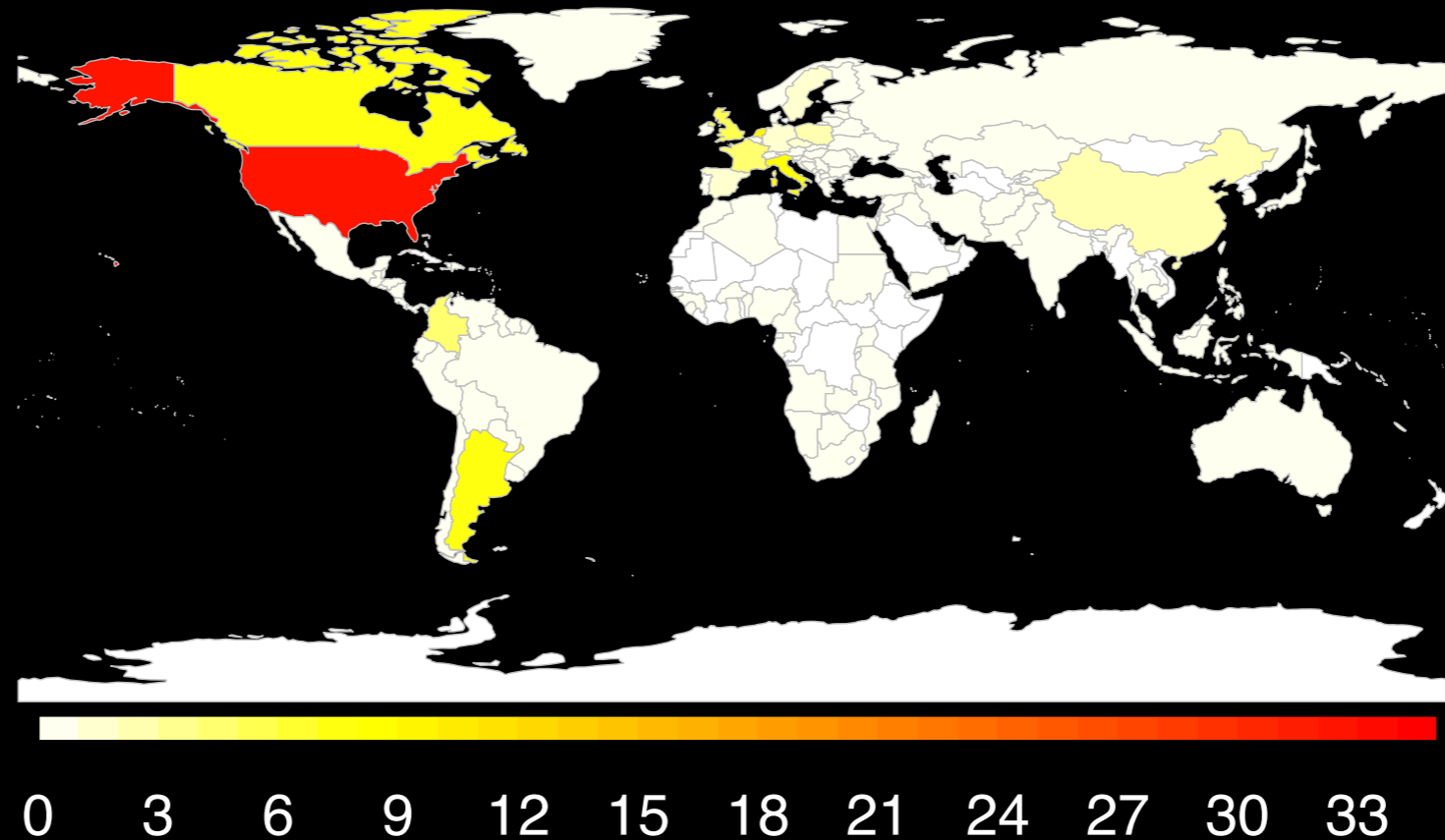


# Query Speed



# Additional Data Sources

## Percent of Queries to CPMD By Country (March 2015)

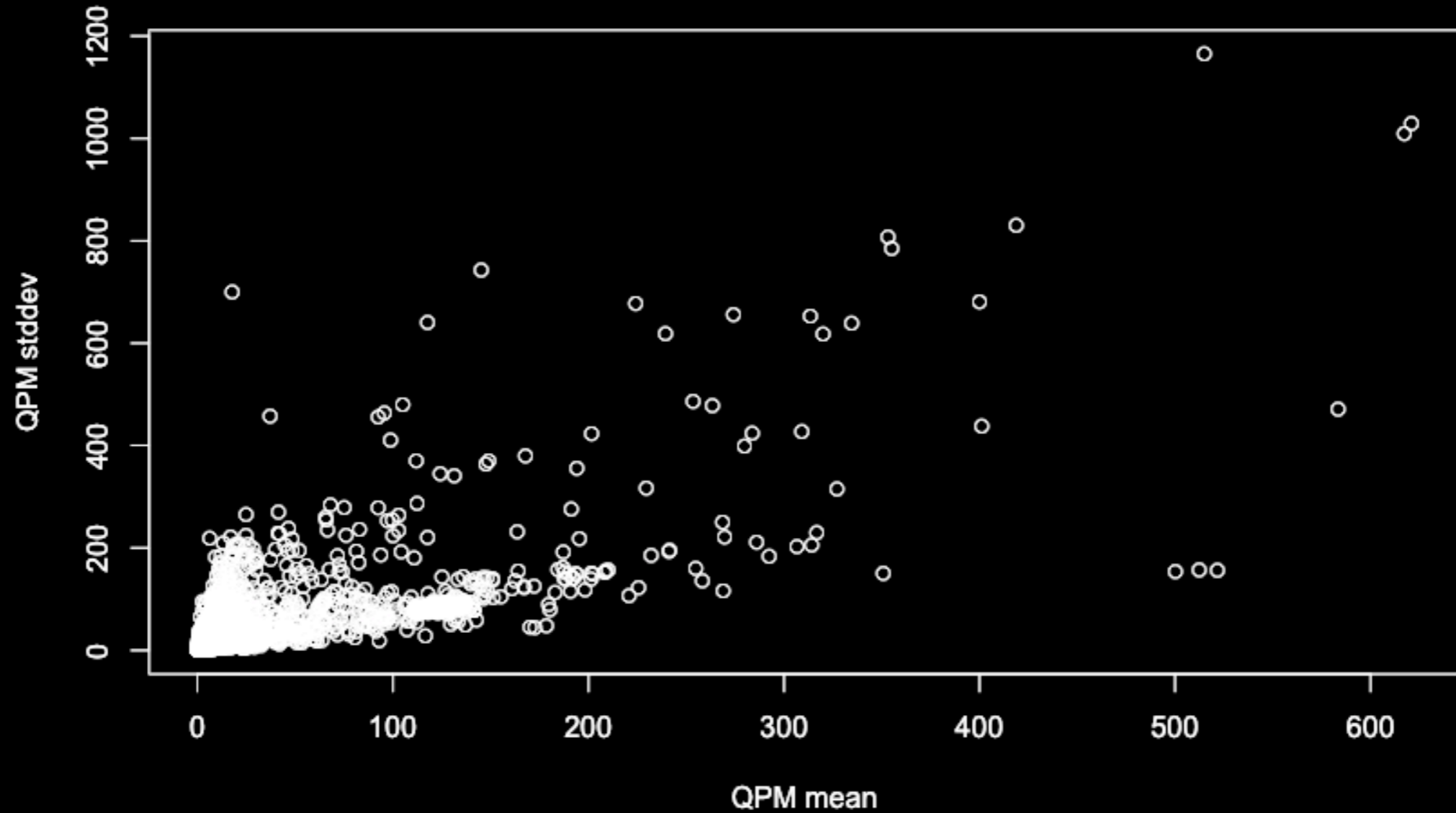


MaxMind GeoLite database  
7m query time



# Per-Source Metrics

Per-Source Query Per Minute (QPM) mean vs stddev for CPMD March 2015



466,021 unique sources  
1h 10m query time

# Discussion

Additional queries?

Optimizations?

Extension to non-root servers?