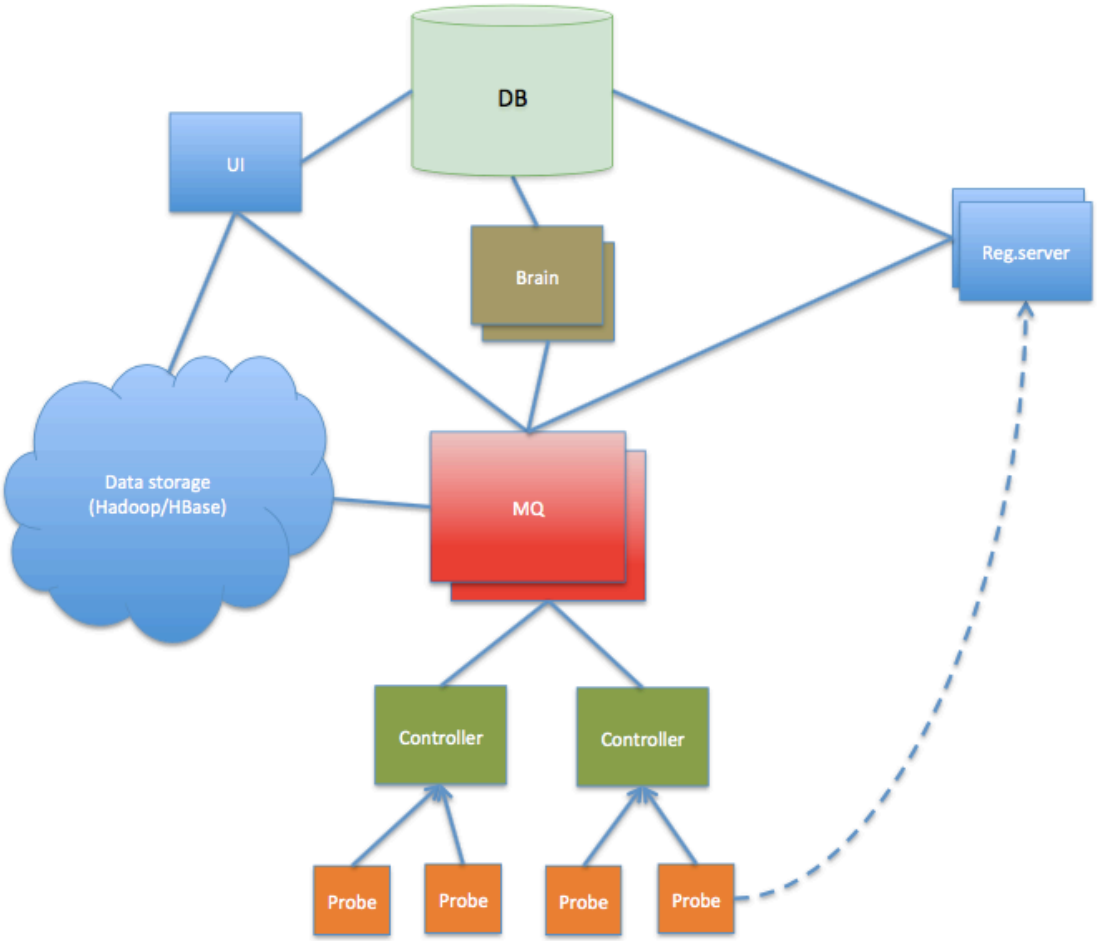
RIPE NCC

RIPE NETWORK COORDINATION CENTRE

RIPE Atlas & Related Infrastructures

Robert Kisteleki | CAIDA AIMS17

RIPE Atlas - Architecture & Co



RIPE Atlas - Architecture & Co

- Built-in redundancy for almost all components
- Probes are semi-autonomous, not just packet reflectors
  - Reporting results generally every 60-90 seconds
  - One-offs are immediately reported
  - They can work offline (crontab-like mechanism)
  - Resource constrained, running just any msm code is not really an option
  - Cannot use just any hardware, in order to avoid "reuse"
- VM proposals and v4 hardware are in the making

RIPE Atlas - Data Backend

- Main Hadoop cluster
  - 3.8PB total raw storage (1.3PB redundant)
  - 0.9PB raw storage in use (various compressions)
  - Total: ~8.5TB RAM
    - ~5TB RAM for Spark/YARN (job processing)
    - ~1.5TB RAM for Hadoop/HDFS/HBase
    - ~1.5TB RAM for OS
  - 1080 CPU cores (+HT)
- Backup cluster: 400TB storage, ~400 CPUs
- Storing ~160TB atlas results, 30TB RIS (BGP)

RIPE Atlas - Data Processing

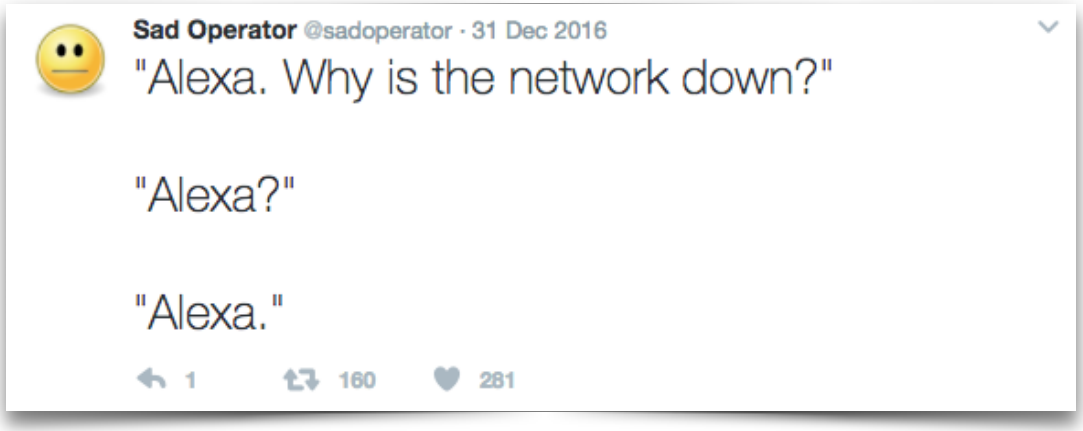
- Some really quickly changing technologies here
  - Map-reduce was sooo yesterday
  - Starting to use (py)Spark now
    - Also gives us the potential to combine RIS (control plane) and Atlas (data plane) data
- (PANDA reference)
  - Giving access to all this data is hard
  - Code-to-data is hard too…

RIPE Atlas - Data Inflow

- Baseline: ~350M results / day (~4K+ / sec)
  - All indexed and can be retrieved immediately
- Recently added:
  - Additional topology discovery measurements
    - ~2M results per day (could be more if needed)
  - Additional DNS root measurements:
    - ~3M results per day
- Technically possible now: measure over Wifi
  - Starting with EduRoam

RIPE Atlas - In The Pipeline

- Infrastructure scalability (for the next 5+ years)
- APIs, data access, visualisations, particular use cases
  - In particular: traceroute viz
- Measuring the cloud (maybe)
- OpenIPMap



OpenIPMap - Infrastructure Geoloc

- Continuing after Emile's prototype
- Goal is a probabilistic answer, based on:
  - Reverse DNS
  - triangulation
  - proximity
  - crowdsourcing (bulk input & interactively on a map),
  - external sources
- Simple combinator engine (to start with)
- Open data (and open code)

RIS (BGP) News

- The RIS infrastructure is being renewed
  - Ultimate goal: modernisation and building capacity to expand to more peers
- Expected new features:
  - Real-time data flow (using Kafka)
  - Streaming interface
  - Centralised looking glass
  - Likely a new riswhois service
  - … and more

Dealing With Unexpected Use Cases

Questions

robert@ripe.net
@kistel