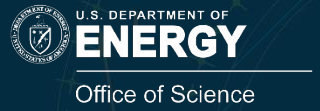
Vera C. Rubin Observatory

APDB Update

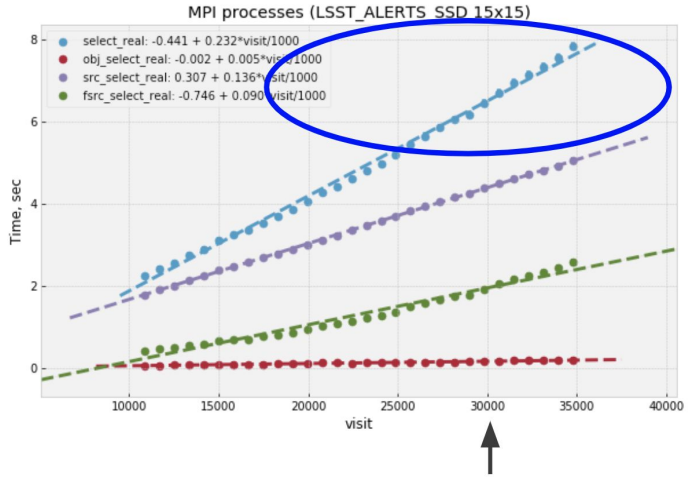
Fritz Mueller / Andy Salnikov
SLAC National Accelerator Laboratory
Vera C. Rubin Observatory Data Management



Rubin Observatory

Previously…

DiaSource SELECT time dominates



One Month

- Using SSD, parallel “image processing” nodes
- ~6 seconds after 1 month -> 72 seconds for 12 month history
- Query time is proportional to both data size on disk in the DB and returned result size, don’t currently have data to disambiguate

From DMLT F2F, October 2019

Summary:

- Relational DB testing write-up: DMTN-113
- Off required perf. by factor ~few
- No clear benefit to further studies down this path
- Evaluate Cassandra NoSQL next
- Use new Qserv czar hardware, available Jan 2020

Cassandra NoSQL Evaluation

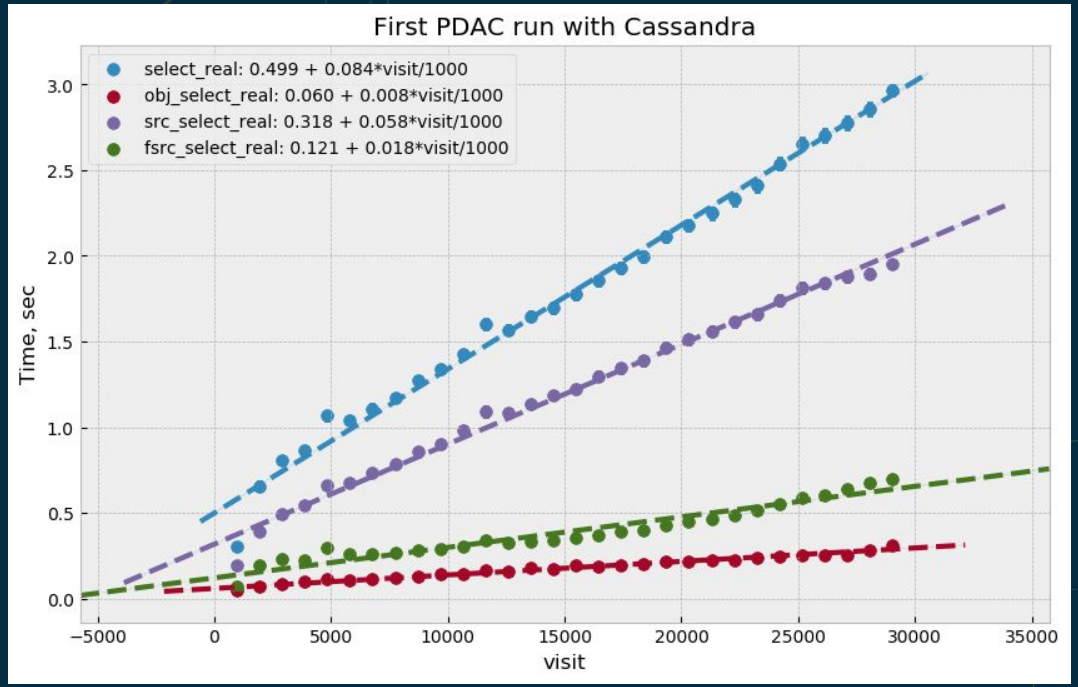
- Track progress at DM-20580 and containing epic

- Hardware procured, racked, and available on schedule (thank you, NCSA!) Configured as 2.5 effective nodes:
  - 2x new Qserv czar nodes
    - 32 cores @ 2.3GHz, 256G RAM, 20 TB NVMe
  - 1x old Qserv czar “half” node
    - 28 cores @ 2.2GHz, 256G RAM, 5 TB NVMe

- Cassandra installed and configured on nodes

- ap_proto being run on verification cluster
  - Reduced scale (9x9 tiles instead of 15x15 tiles) due to contention with PDR2 runs

Some Early Results



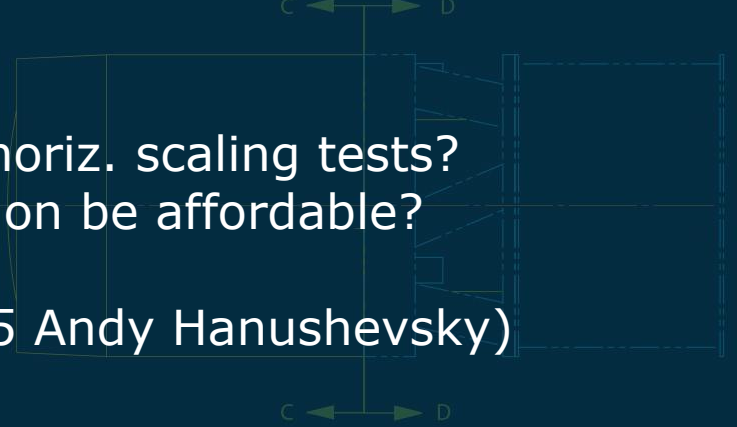
Quite early, but we can so far see:

- Cassandra doesn't completely not work
- Query times are so far scaling linearly sensibly
- Relative query costs per type roughly consistent with SQL solutions
- Good news: we now have horiz. scaling knob to turn
- Bad news: could take a lot of nodes to get there; don't have enough info yet to know how many

Rubin Observatory

Ongoing Work

- Understand better how the current deployment is functioning
  - Why are writes slow; isn't Cassandra supposed to be fast there?
  - Many, MANY, tuning options yet to be tried and understood...
  - Hooking up monitoring to see under the hood
- Will query perf. remain linear out to 12 mos. of simulated visits?
- What *is* the horizontal scaling factor?
  - Can we make cloud work for our horiz. scaling tests?
  - Will enough hardware for production be affordable?
- ~1 FTE available (.5 Andy Salnikov, .5 Andy Hanushevsky)



Beyond Cassandra (Backup Plans)

(Per Oct. 2020 DMLT)

- Experiment with custom solutions
  - Can we put together a system from smaller stock parts, write some of our own code?
  - E.g. use an object store for static “blobs” of records from past nights + combine with DB results for tonight’s latest updates.
  - Goal would be to better exploit the structure of the problem

- Push back on requirements
  - Most significant is probably alert time-series as currently conceived. Perhaps less history, or simplify “sliding window” design?
  - What could be gained by relaxing 60-second alert constraint?