Infrastructure

2018 RECAP
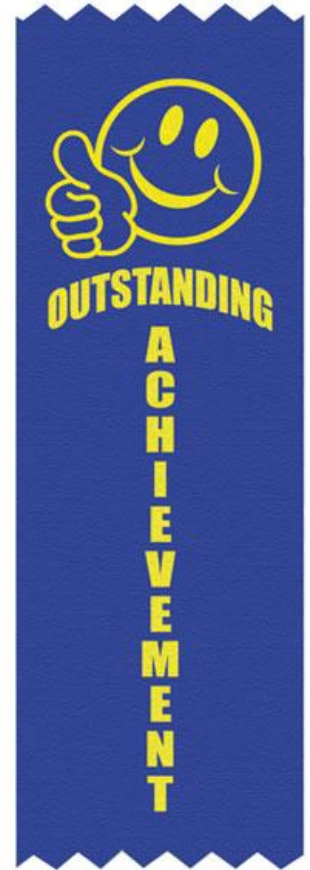
2018 Big Things

Reliability

Chris!

Ubuntu 16

Feature Highlights

Feature Highlights

Mux

Auth DNS

Zone Transfer

- Standardized Reason Codes
- Externalized Providers
- Benefit Report
- Removed Hughes from geo rollup

YYYYMMDD	2018-10-19 ▾		▾	⊞ openmix_customer_hour

⊟ YYYYMMDD (395)

s0. queryType	STRING	
s0. rttScores	RECORD	REPEATED
s0.rttScores. reasonCode	STRING	NULLABLE
s0.rttScores. decisionCount	INTEGER	NULLABLE
s0.rttScores. scores	RECORD	REPEATED
s0.rttScores.scores. pid	INTEGER	NULLABLE
s0.rttScores.scores. value	INTEGER	NULLABLE
s0. tputScores	RECORD	REPEATED
s0.tputScores. reasonCode	STRING	NULLABLE

1 Year History

Every Decision

Detail Score Values

Benefit report requires detailed scores for each decision.

Changes in Apps

Flow App & Static Failover

Network-only Scores

- Flow app can return multiple responses
- Ignored provider API for benefit report
- Can move providers from JS code to Portal

```json
{
  "app_type": "V1_FLOW",
  "app_version": 1,
  "ttl": 20,
  "default_cname": "no.available.candidates.g
  "app_flow_data": [
    {
      "condition": {
        "ip": [
          "123.45.26.23/24",
          "2.3.4.5"
        ]
      },
      "routing": [
        {
          "method": "FAILOVER",
          "platforms": {
            "platform1": {
              "priority": 1,
              "radar_availability": 80,
              "response": "www.mydomain1.com",
              "sonar": true
            },
```

```json
    {
      "condition": {
        "asn": [
          30512,
          30555
        ]
      },
```

```json
    {
      "condition": {
        "country": [
          70,
          223
        ],
        "market": [
          3
        ]
      },
```

Improvements to Fusion Platform

Streaming

onFusion() callback

- Two second fusion updates to appserver
- App can cache parsed fusion data
- No more fusion in benson 🙋
- RMQ still SPOF

Antoine 🏡 11:25 AM

I did some tests and I found that on average it's taking around 1s.

Features in Progress

# In Progress

GSLB Rates

Long Queries

DNSSEC

- DNSSEC is huge spec, lots of moving parts here. Q1 is tight 👷
- GSLB is stalled on Citrix/IT or design
- Query performance still in research 👶

joel 🐙 1:13 PM

As the RFC puts it:

> An alternating sequence of DNS public key (DNSKEY) RRsets and Delegation Signer (DS) RRsets forms a chain of signed data, with each link in the chain vouching for the next. A DNSKEY RR is used to verify the signature covering a DS RR and allows the DS RR to be authenticated. The DS RR contains a hash of another DNSKEY RR and this new DNSKEY RR is authenticated by matching the hash in the DS RR. This new DNSKEY RR in turn authenticates another DNSKEY RRset and, in turn, some DNSKEY RR in this set may be used to authenticate another DS RR, and so forth until the chain finally ends with a DNSKEY RR whose corresponding private key signs the desired DNS data.
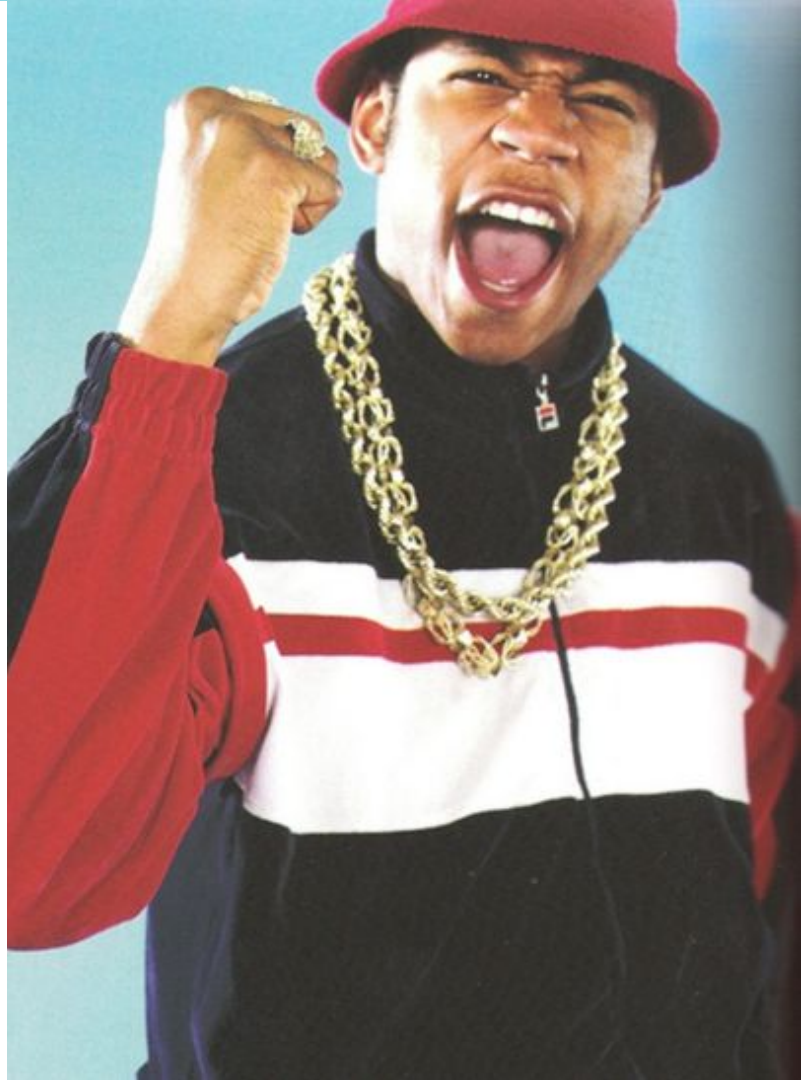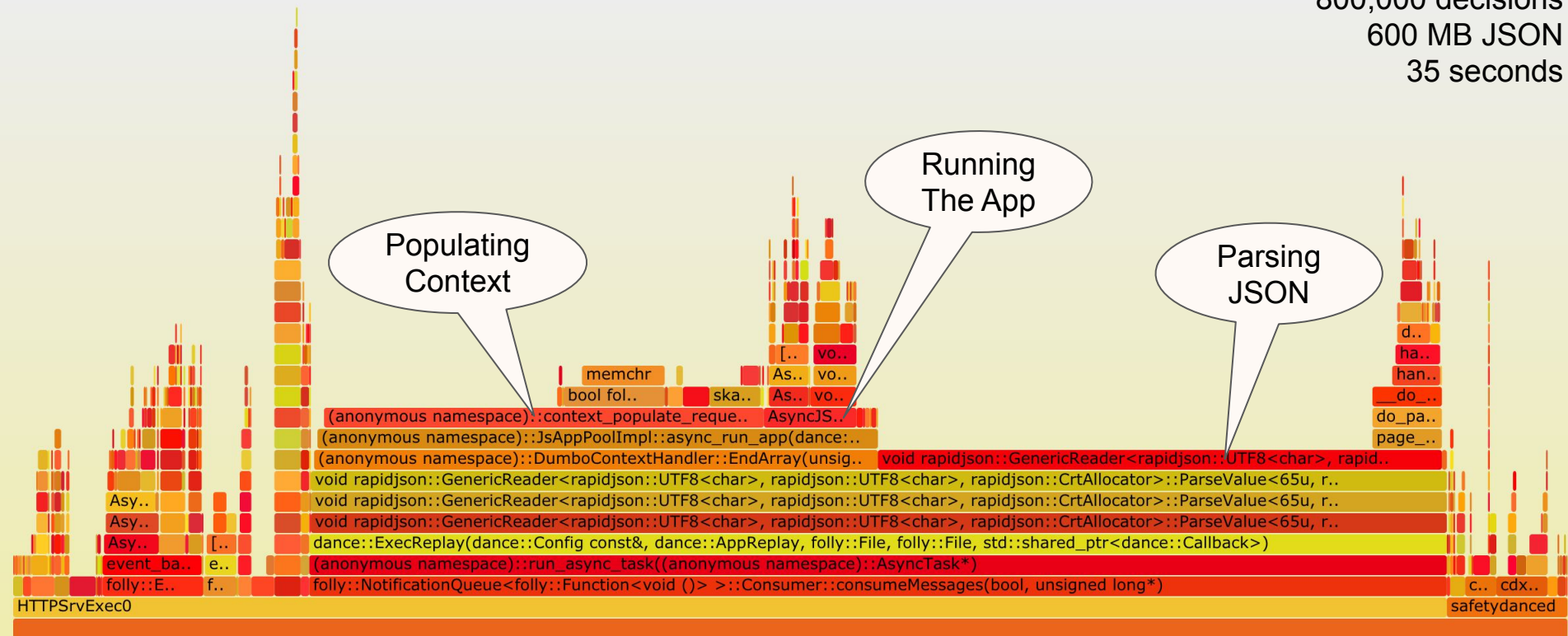
Backend Ready

Canary Apps

TCP DNS

App Replay

- All these are ready and pending time from UI or Ops to get it up and running.
- Sure there are bugs to be uncovered!

Flame Graph

800,000 decisions
600 MB JSON
35 seconds

This is a flame graph of replay, which shows running a JS App takes approximately the same amount of time as just parsing raw JSON from Dumbo

Platform Highlights
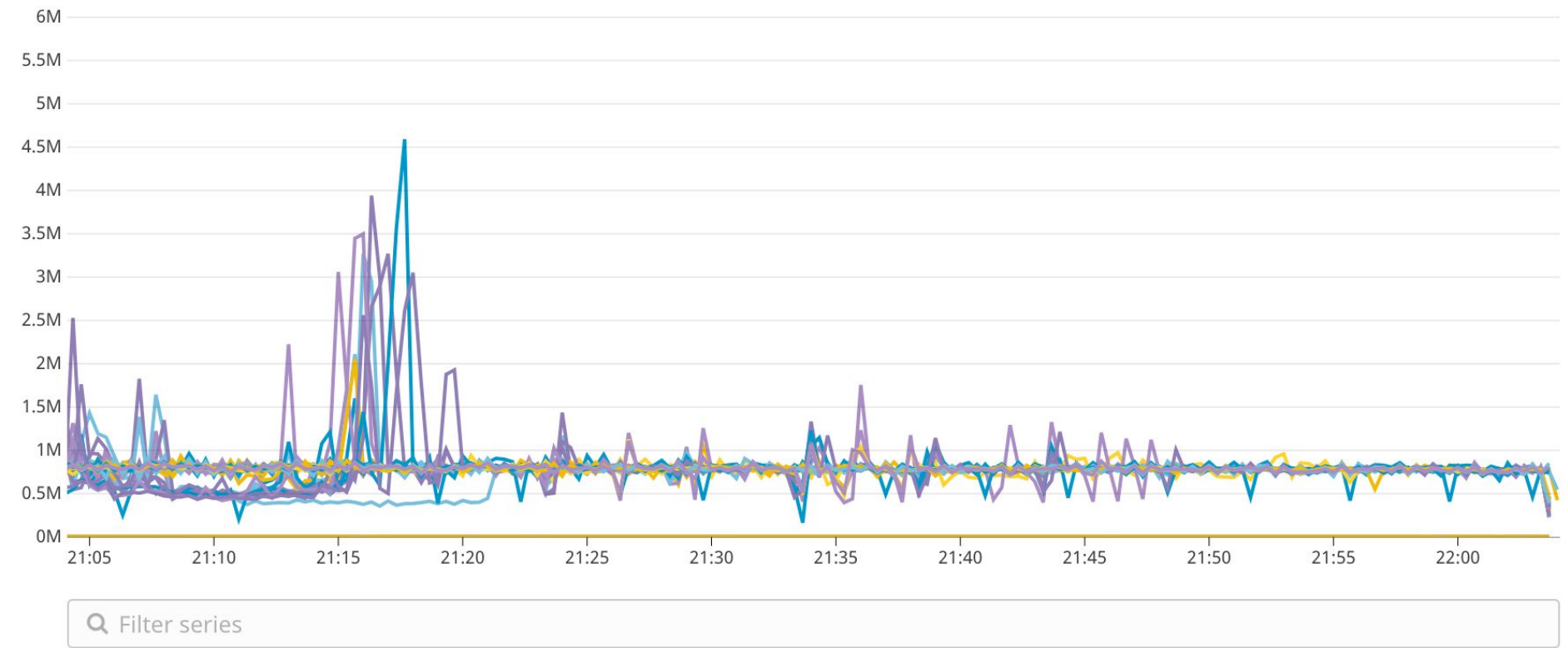
Better Connections

Autocon

Rate monitors

Reconnects

- Moved VPN responsibility to ops
- New library code for stream management
- Per-sender stats at receiver

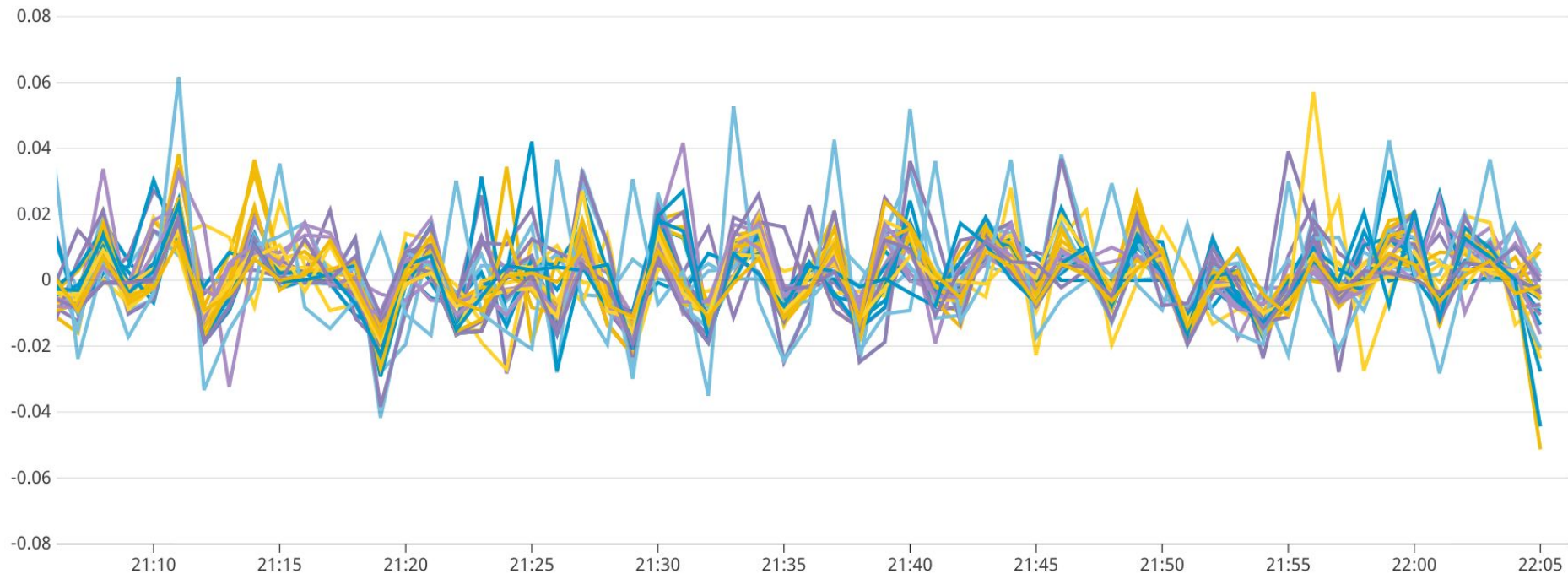# Avg of appserver.score_messages_accepted over prod by ...

How many scores received, as reported by each appserver

nz1.dca.sl.prod % scores dropped per ...

Zinter to appserver: difference of sender & receiver counts

Reduced Administration

Zero Config

No-Downtime Restarts

We run Auth DNS

- Our own auth dns code powers cedexis.net
- All upgrades are hand-over hand and downtime-free

```
"ns1.cedexis.net": { "A": { "response" : "204.48.34.10", "ttl": 86400 } },
"ns2.cedexis.net": { "A": { "response" : "209.107.211.14", "ttl": 86400 } },
"ns3.cedexis.net": { "A": { "response" : "69.28.180.4", "ttl": 86400 } },
"ns4.cedexis.net": { "A": { "response" : "172.85.88.4", "ttl": 86400 } },

"cedexis.net" : {
  "NS" : {
    "response" : [
      "flipm.cdxcn.net", "flipd.cdxcn.net", "flipg.cdxcn.net",
      "flipp.cdxcn.net"
    ],
    "ttl" : 28800
  },
  "SOA" : {
    "response": {
      "mname": "flipm.cdxcn.net",
      "rname": "admin.cedexis.com",
      "refresh": 14400,
      "retry": 7200,
      "expire": 604800,
      "rxttl": 20
    },
    "ttl" : 240
```

Same format emitted by portal for customer-configured domains
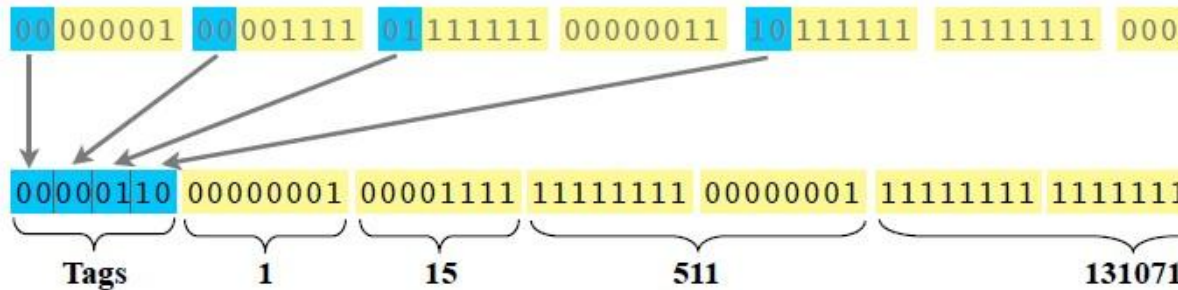
Bug Fixes & Miscellany

Ready for EDNS Flag Day

Throughput scores > 64Mbps

- JS App exceptions are now available in BQ
- Real custom query support is now in dumbo-api
- dumbo-prod-fusion is a thing that exists 💁

Group Varint Encoding

- Idea: encode groups of 4 values in 5-17 bytes
  - Pull out 4 2-bit binary lengths into single byte prefix

| 00 000001 | 00 001111 | 01 111111 | 00000011 | 10 111111 | 11111111 | 000 |

| 00 00 01 10 | 00000001 | 00001111 | 11111111 | 00000001 | 11111111 | 1111111 |

Tags   1   15   511   131071

- Decode: Load prefix byte and use value to lookup in 256-entry
  . . .

| 00 00 01 10 | → Offsets: +1,+2,+3,+5; Masks: ff, ff, ffff
  . . .

- Much faster than alternatives:
  - 7-bit-per-byte varint: decode ~180M numbers/second
  - 30-bit Varint w/ 2-bit length: decode ~240M numbers/second
  - Group varint: decode ~400M numbers/second

Row	byte_size	pct
1	1 byte	2.15
2	2 byte	78.18
3	3 byte	19.67

- Was uint16_t which was max at 64Mbps
- Most values still take 2 bytes

Appserver Performance Improvements

Immutable Apps

JSProc IPC Serialization

- Zero-allocation trivially copyable structs
- Removed runtime checks in apps; they now reload when config changes

Prep Time and Exec Time improvements
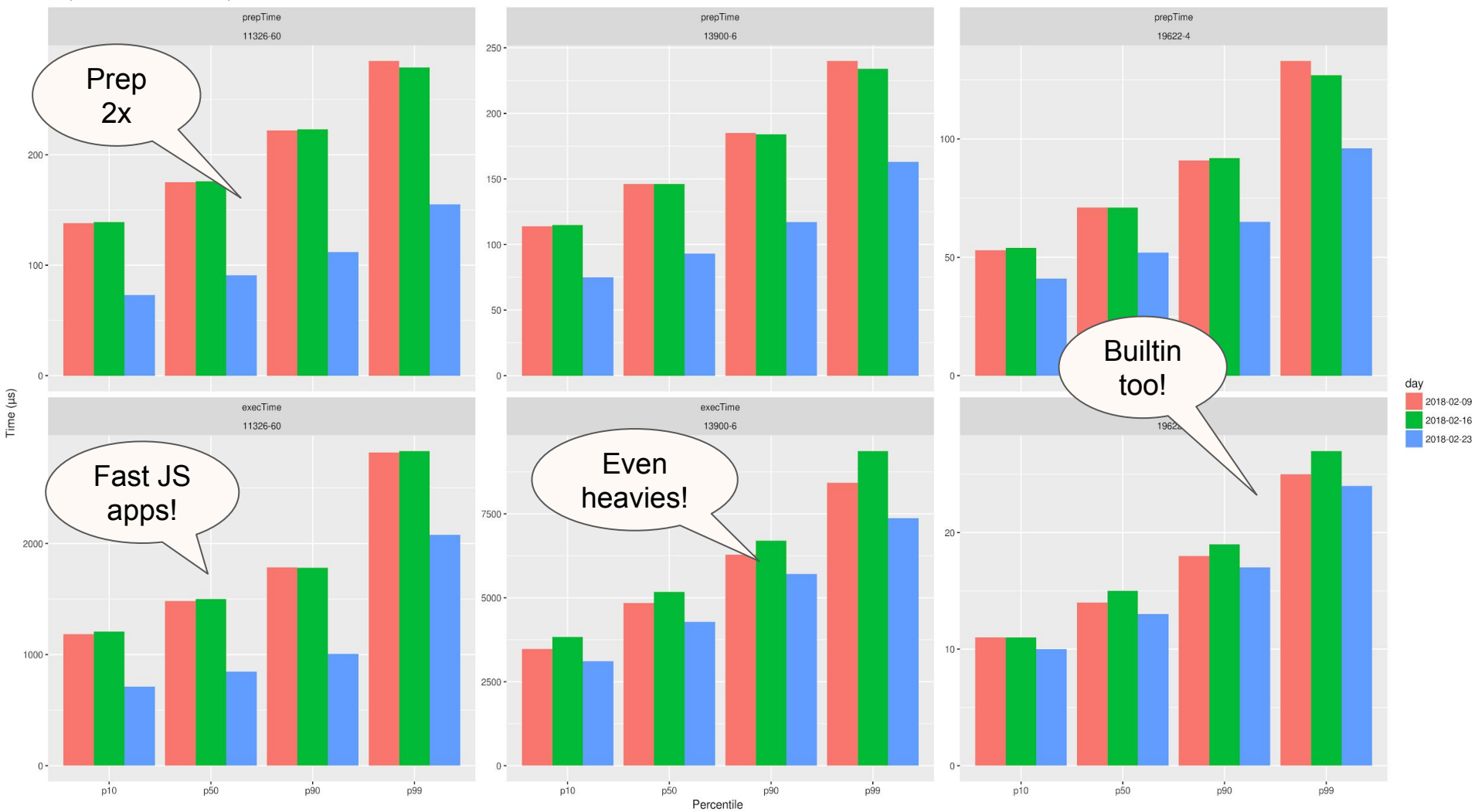
Other Performance Improvements

150k/req/sec DNS

3x URL parsing for reports

- Netmap parse shaved 20sec from boot
- New stats lib with no cross-thread sharing
- New rate limit implementations
- Faster logging too!

```cpp
// N2
template <>
void parse_path<n1_report>(RadarLog& log, folly::StringPiece path) {
  using NavTiming = openmix::messages::RadarNavtimingValues;

  log.set_report_type(ReportType::N1Report);

  PathParse p = {path};
  start_parse(p, log, "/n1")              //
      / &RadarLog::set_response_code      //
      ;
  parse_into(p, *log.mutable_navtiming())
      / &NavTiming::set_navigation_start
      / &NavTiming::set_unload_event_start
      / &NavTiming::set_unload_event_end
      / &NavTiming::set_redirect_start
      / &NavTiming::set_redirect_end
      / &NavTiming::set_fetch_start
      / &NavTiming::set_domain_lookup_start
      ;
  parse_into(p, log)
      / required(&RadarLog::mutable_signature)
      / optional(&RadarLog::mutable_partner_tags)
      / optional(&RadarLog::set_start_render)
      ;

  validate_parse(log, p.ok);
}
```

① (annotation marker)

```cpp
template <typename T, typename FieldT>
ParseContext<T>& operator/(ParseContext<T>& ctx, FieldT&& fld) {
  parse_segment(ctx, fld, ctx.parse.path.split_step('/'));
  return ctx;
}
```

② (annotation marker)

③ (annotation marker)

```cpp
#define CALL_MEMBER_FN(object, ptrToMember) ((object).*(ptrToMember))
template <typename T, typename V>
void parse_segment(ParseContext<T>& ctx, IntegralField<T, V> fld,
                   folly::StringPiece segment) {
  if (ctx.parse.ok) {
    if (auto v = folly::tryTo<V>(segment)) {
      CALL_MEMBER_FN(ctx.target, fld.setter)(ctx.basis + *v);
    } else if (fld.required) {
      ctx.parse.ok = false;
    }
  }
}
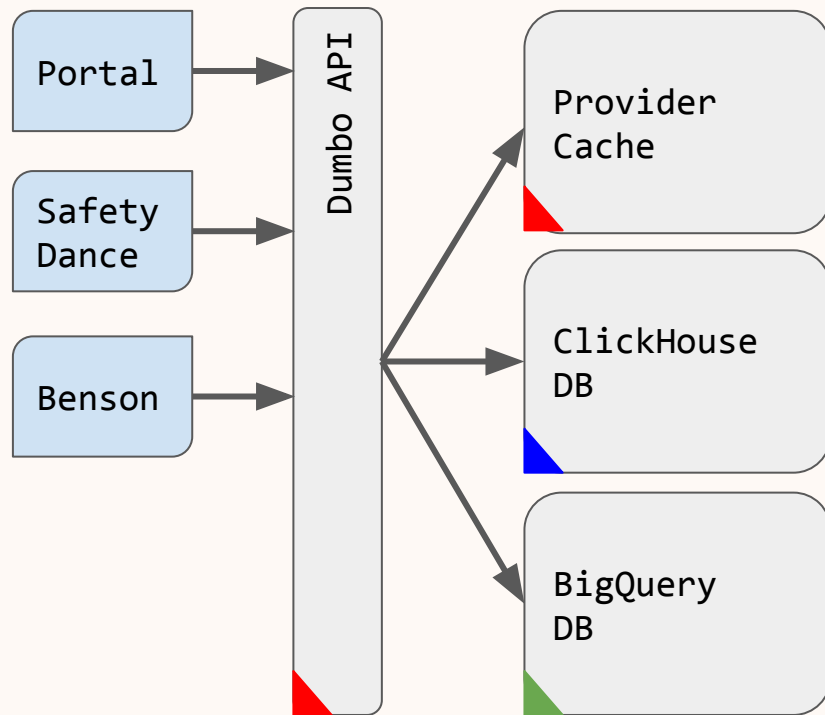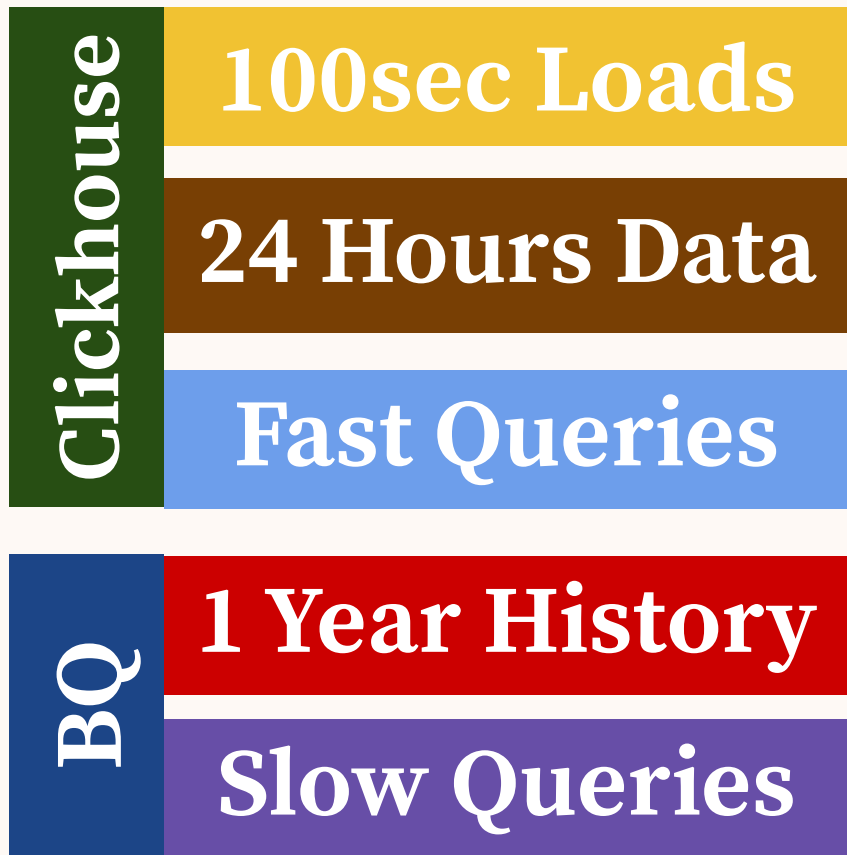```

# Dumbo Performance Improvements

**Gzip responses**

**Last-hour API queries are sub-second**

**30 day BQ works!**

- New CH hardware, better SQL
- Better BQ SQL
- Still super reliable
- Ingest performance is ~100s

# Dumbo API Architecture Review

# Team Highlights

# Disciplined Release Schedule

**Frequent Releases**

**184 Releases**
**163 Hosts**

appserver: 26
autocon: 3
belvedere: 3
benson: 10
button: 14
deliverator: 1
dumbo-api: 24
dumbo-ingest: 9
dumbo-notify: 4
kennylogger: 21

newzinter: 12
safety-dance: 16
spinderella: 5
tangled: 7
tps-report: 11
tribble: 2

**darren** 🥒 11:28 AM
wow

**alok** 10:27 AM
wow

**alok** 10:07 AM
YES! HUGE release!

**darren** 🥒 10:34 AM
wow, ok!

**alok** 12:39 PM
wow, huge release. 👍

# Code Highlights

**DNS Server Rewrite**

**New Unit Test Library**

- Stats lib
- Builtins lib
- Send/Recv lib
- CSV➡JSON config
- Direct apt deps

# Changes in Openmix2.0 in 2018



This code churn chart is the best chart in the entire world!

# Healthy Churn

Safe releases
Safe changes
Safe culture

Everybody is at the Party



Changes in Bigdata in 2018
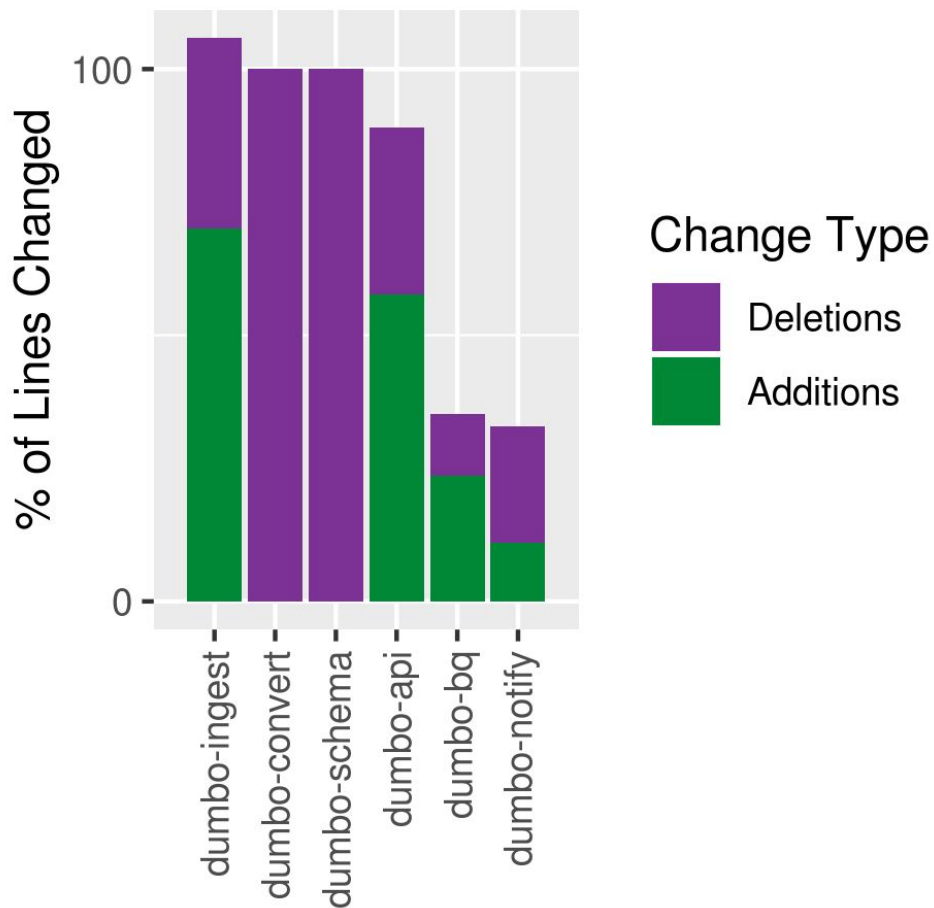
# Team Safety

# How We Do It

**No Bug Backlog**

**No JIRA**

- Big Active Project
- Individuals keep lists of things which bug them
- Trusted to fix those things
- 出る釘は打たれる

**dnsserver: Add code to generate RRSIGs**
#2196 by cincodenada was merged on Dec 12, 2018 • Approved
💬 86

**dnsserver: Rework ECS/EDNS handling**
#2134 by cincodenada was merged on Nov 9, 2018 • Approved
💬 63

**DNSServer: listen TCP**
#2057 by suspend0 was merged on Oct 11, 2018 • Approved
💬 94

**Safety: app replay feature [WIP]**
#2019 by suspend0 was merged on Sep 25, 2018 • Approved
💬 131

**Providers: single thread for all consumers**
#1795 by suspend0 was merged on Jun 12, 2018 • Approved
💬 52

**JSProc: Async implementation**
#1687 by suspend0 was merged on May 11, 2018 • Approved
💬 64

**Appserver: new multi-stage app**
#1653 by suspend0 was merged on Apr 21, 2018 • Approved
💬 107

Building consensus about the change: answer questions & debate implementation
Some pull requests don't get enough support and are abandoned

# Kaizen (改善) and Sekinin (責任)

**Team Responsibility**

**Small Improvements**

- Common to have 20/50/100 comments on a code review
- Safe to ask a question
- Safe to ask why
- Safe to try something
- Safe to be wrong

"If you are going to do kaizen continuously…you've got to assume that things are a mess.  Too many people just assume that things are alright the way they are.

Are you guys convinced that the way you're doing things is the right way?  That's no way to get anything done.

Kaizen is about changing the way things are. If you assume that things are alright the way they are, you can't do kaizen. So change something!"

–Taiichi Ohno

# Practicing Continuous Improvement

**Have a Standard**

**Change the Standard**

- Monorepo
- Shared code
- Copy & tweak
- Tension between change & consistency

# Team Safety

**Messy is OK**

**Temporary is OK**

**Fail is OK**

- Get something in, get it shipped
- Make better later
- But, actually do that

fin