Build Unique Search Experiences

Velocity

Own your reliability

Adam Surak
SRE & Security Engineer
adam.surak@algolia.com
@AdamSurak
#id

→ Algolia since 2014 (team of 8)
→ SRE & Security Engineer
→ Responsible for infrastructure
→ I like to sleep and break things
Algolia Today

2100+ customers in 100+ countries
40B+ Write operations per month
17B+ User-generated queries per month

15 regions, 50+ datacenters
Who owns your availability?

YOU
Basic principles

\[ T = A^2 \]

\[ T = 1-(1-A)^2 \]
Reality

- 99.95%
- 99.3%
- 99.6%
- 98%
- 99.95%
- 99.8%

**external service**

no documentation

does what?
Ideal state
What is SLA?
What is SLA?

“A service level agreement (SLA) is a contract between a service provider (either internal or external) and the end user that defines the level of service expected from the service provider.” by Palo Alto Networks

- Mostly uptime
- In advanced environments - response time, error rate
How much costs you a minute of downtime?
## Common SLA levels

<table>
<thead>
<tr>
<th>SLA</th>
<th>Downtime per month</th>
<th>Cost ($95/min, $50M/year)</th>
<th>Cost per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 %</td>
<td>7 hours and 18 minutes</td>
<td>$41 610</td>
<td>$500 000</td>
</tr>
<tr>
<td>99,9 %</td>
<td>43 minutes</td>
<td>$4 085</td>
<td>$49 000</td>
</tr>
<tr>
<td>99,95 %</td>
<td>21 minutes</td>
<td>$1 995</td>
<td>$24 000</td>
</tr>
<tr>
<td>99,99 %</td>
<td>4 minutes</td>
<td>$380</td>
<td>$4 560</td>
</tr>
<tr>
<td>99,999 %</td>
<td>26s</td>
<td>$41</td>
<td>$492</td>
</tr>
<tr>
<td>99,9999 %</td>
<td>2,6 seconds</td>
<td>$4</td>
<td>$48</td>
</tr>
<tr>
<td>100 %</td>
<td></td>
<td>Marketing level</td>
<td></td>
</tr>
</tbody>
</table>
SLA tricks

→ “100% uptime, 5% refund after each 0.05%”
  ➡ 99.95%

→ “99.9% SLA, downtime counts if backend responds with error during 2 consecutive 90s intervals”
  ➡ 99.8%

→ “…downtime counts from the moment of customer’s report”
  ➡ :facepalm:
Where is the SLA monitored from?
Independent monitoring

- Pingdom
- ServerDensity
- ThousandEyes
- CatchPoint
- TurboBytes Pulse
- Kentik
- Custom

DNS - 1.6k, API Status - 3.3k, Latency - 73k
Who runs a server?

Assuming the rest is #serverless with abstracted issues
Can you restart any server anytime?
Can you restart any rack anytime?
Can you restart any datacenter anytime?
Underestimated dependencies

→ Power/network segments
  • are two adjacent racks really independent?
  • how protected is the network? Rogue DHCP? IP hijack?
  • can you choose rack with your provider?
  • what happens if you order 3 servers at once?

→ A/C
  • influences a set of racks

→ Network cables and interfaces are always broken
  • is your 1Gbit interface really in 1Gbit mode?
Network maintenance issues

→ unplanned

→ planned

• but they forget to tell you

→ failing

• planned, they told you, but you have downtime
“But I run on cloud!”

– Cloud user
Clouds are not error-proof

- AWS has outages
  - route leak, broken DynamoDB
- GCP has outages
  - sequence of lightning strikes, global network issue
- Azure has outages
  - global multi-hour outage
- Salesforce has outages
  - EU2 read-only, NA14 outage
- ...you name it
- you can deploy multi-cloud! => APIs!
Network related issues

→ AWS EU-West-1 broke Direct Connect with OVH

→ ISP received 0.0.0.0/0 from a new peer => 75% traffic lost

→ Malaysia Telecom announced AWS prefixes => US-East-1 unavailable

→ ISP of CloudFlare misconfigured router and started to receive all CloudFlare’s worldwide traffic in Doha, Qatar

→ TCP proxy becomes your best friend
→ Servers in San Jose

→ Customer in Oregon
  • AWS US-West 2

→ 21 ms average latency
→ from 21 ms to 150-300ms

→ ?
→ new route via Denver

→ 20% packet loss on Seattle-Denver

→ issue out of AWS network

<table>
<thead>
<tr>
<th>HOST</th>
<th>Loss%</th>
<th>Snt</th>
<th>Last</th>
<th>Avg</th>
<th>Best</th>
<th>Wrst</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. ec2-54-186-253-146</td>
<td>0.0%</td>
<td>10</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
<td>1.3</td>
<td>0.3</td>
</tr>
<tr>
<td>5. 100.64.16.131</td>
<td>0.0%</td>
<td>10</td>
<td>1.3</td>
<td>7.7</td>
<td>0.8</td>
<td>43.9</td>
<td>14.1</td>
</tr>
<tr>
<td>8. 205.251.225.199</td>
<td>0.0%</td>
<td>10</td>
<td>8.5</td>
<td>8.1</td>
<td>7.1</td>
<td>9.6</td>
<td>0.9</td>
</tr>
<tr>
<td>9. sea-b1-link.telia.net</td>
<td>0.0%</td>
<td>10</td>
<td>7.9</td>
<td>8.1</td>
<td>7.7</td>
<td>9.9</td>
<td>0.6</td>
</tr>
<tr>
<td>10. den-b1-link.telia.net</td>
<td>20.0%</td>
<td>10</td>
<td>109.1</td>
<td>109.4</td>
<td>108.6</td>
<td>110.0</td>
<td>0.5</td>
</tr>
<tr>
<td>11. sjo-b21-link.telia.net</td>
<td>10.0%</td>
<td>10</td>
<td>137.8</td>
<td>137.5</td>
<td>136.7</td>
<td>138.0</td>
<td>0.4</td>
</tr>
<tr>
<td>12. leaseweb-ic-302376-sjo-b21.c</td>
<td>10.0%</td>
<td>10</td>
<td>136.4</td>
<td>137.0</td>
<td>136.4</td>
<td>137.8</td>
<td>0.6</td>
</tr>
<tr>
<td>13. 209.58.135.206</td>
<td>10.0%</td>
<td>10</td>
<td>47.3</td>
<td>47.5</td>
<td>47.2</td>
<td>47.9</td>
<td>0.3</td>
</tr>
<tr>
<td>14. 209.58.135.201</td>
<td>44.4%</td>
<td>9</td>
<td>47.3</td>
<td>47.1</td>
<td>46.8</td>
<td>47.3</td>
<td>0.2</td>
</tr>
<tr>
<td>15. 209.58.131.95</td>
<td>22.2%</td>
<td>9</td>
<td>47.0</td>
<td>47.0</td>
<td>46.8</td>
<td>47.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
“What happens when you put www.algolia.com to your browser and hit enter?”

–Networking 101
DNS

- No DNS -> no new connections
- Packet loss prone
- Latency of DNS is very tricky
  - language and OS dependent timeouts
- DNS providers are popular DDoS targets
- Having two DNS providers is perfectly doable -> APIs!
  - amazon.com doesn’t use Route53 and has two other providers
Software design

- TCP checksum is not 100% safe
- DNS resolving is not 100% working
- HTTP calls don't always succeed or return 200
  - what is the default timeout of your HTTP client?
Software operations

- Package repositories can get broken or out-of-sync
- Can you deploy when GitHub is down?
- Does your CI works when Docker signature is broken?
- Invest in introducing mistakes!

iptables -A INPUT -p udp --dport 33434:33523 -j REJECT
People

→ Who holds the knowledge about the system?

→ Do people know what to do?

→ How do you escalate?

→ Send people on vacation!
“Everything that can break will work and then we will make wrong assumptions about the reliability.”

–Sidney Dekker
We are hiring in Paris and SF

THANK YOU!

adam@algolia.com
@AdamSurak