Incident 24 Sep. 2020: Incident with Internal DNS resolutions

Availability: **Sales down, admin down**

% of clients hit: **100% of the clients**

Duration of incident: **44 minutes**

Symptom

From 14h36 to 15h20 UTC, a regular customer that would try to shop on any VTEX store would receive internal server (HTTP 500) errors. In addition, a client trying to access the admin panel of their store would not be able to login or do any major operations.

Summary

At 14h36 UTC, we identified a massive increase in internal errors on our platform. We discovered that our internal DNS resolutions stopped working, and therefore our services had difficulties communicating to each other. Once we fixed our DNS resolutions, our services gradually stabilized and stores and admins were working again.

To be the trusted partner to your success, our team is working on follow-up actions to make sure that this incident does not happen again, and that we identify and recover from future incidents faster. We are committed to improving our systems to guarantee a reliable and trusted experience.

Timeline

[24-09-2020 14:36 UTC] Our monitoring systems warned of anomalous drop behavior in orders and sessions, and increased error rates in internal modules.

[24-09-2020 14:41 UTC] We identified failures in our service communication layer.

[24-09-2020 14:44 UTC] We confirmed failures in our internal DNS resolution and began mitigation actions.

[24-09-2020 14:58 UTC] All actions required to mitigate the crisis were performed.

[24-09-2020 15:02 UTC] We confirmed that the actions were effective and our services gradually stabilized.

[24-09-2020 15:18 UTC] Normal operations were fully reestablished, and all enqueued operations were completed.

Mitigation Strategy

We restarted our internal DNS resolution service. Gradually, our services started communicating with each other again. For some services, manual intervention was necessary to accelerate their recovery.

Follow up Actions: preventing future failures

As follow-ups of this incident, we will work on strengthening our internal DNS by increasing monitoring, so that we can detect service communication failures faster. We're also creating a DNS failover so that our internal communications are more resilient.