

Incident Oct. 22, 2022: Degraded performance in our API Router

Availability: **~11% of sales degraded**

% of clients affected: **100%**

Duration of incident: **3 hours and 1 minute**

Symptoms

Between 12h00 and 15h01 UTC, customers that tried to use VTEX stores would experience degraded performance, especially increased latency to load uncached results. From 12h30 to 15h01 UTC, product changes were not being indexed.

Summary

Between 12h00 and 15h01 PM UTC, we experienced degraded performance in our main API Router service; it increased the response time of our platform services which partially impacted the sales flow.

This issue was caused by incorrect behavior where new Router instances added to handle the increasing load demand would not start correctly and hence would not receive any traffic. As new instances could not handle any incoming traffic, only previously existing servers would receive the load, leading to increased latency on the platform services.

We later identified that the new servers couldn't receive traffic due to consistent errors on startup where they could not retrieve essential configurations to run correctly.

We later identified that the service responsible for returning the configurations was not capable of doing it accordingly. A wrong process caused the insertion of duplicated configurations on the service, making it impossible for the system to reconcile the complete configuration set and return the correct configuration to the Router service.

During the incident, we turned off the indexing process to avoid increasing the blast radius and overloading some systems.

Timeline

[2022-10-22 12:06 UTC] Our monitoring systems warned of increased latency

[2022-10-22 12:17 UTC] We noticed that new instances from our router service were not becoming healthy

[2022-10-22 13:08 UTC] We detected errors in retrieving router configurations preventing that new instances became healthy

[2022-10-22 13:30 UTC] To reduce the load on the platform, we stopped the product indexing

[2022-10-22 14:35 UTC] We detected the root cause of the error in retrieving router configurations

[2022-10-22 14:36 UTC] We started the mitigation of the root cause

[2022-10-22 14:55 UTC] Services latency was gradually normalizing

[2022-10-22 15:01 UTC] We restarted the product indexing

[2022-10-22 15:01 UTC] Normal operations were fully restored.

Mitigation strategy

The mitigation work prioritized restoring the 11% of impacted sales as quickly as possible, preserving the other 89% of sales that were not affected, removing duplicated entries from the router configuration, and gradually restarting router service to load fixed settings.

Follow-up actions: preventing future failures

As a follow-up to this incident, we have already created new alarms to detect this behavior in the router service proactively.

We also made the Router Configuration API ignore duplicate entries preventing the scenario that degraded our Router Service.

We'll work on fixing the process that duplicates entries in the configurations.