Microserviços na otimização da ingestão de anúncios em massa na OLX

Rafael
DEV

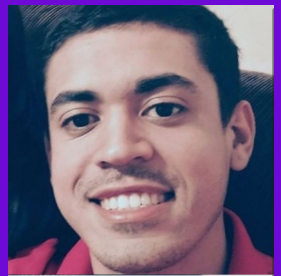
Drino
DEV

OLX

Edu
GE

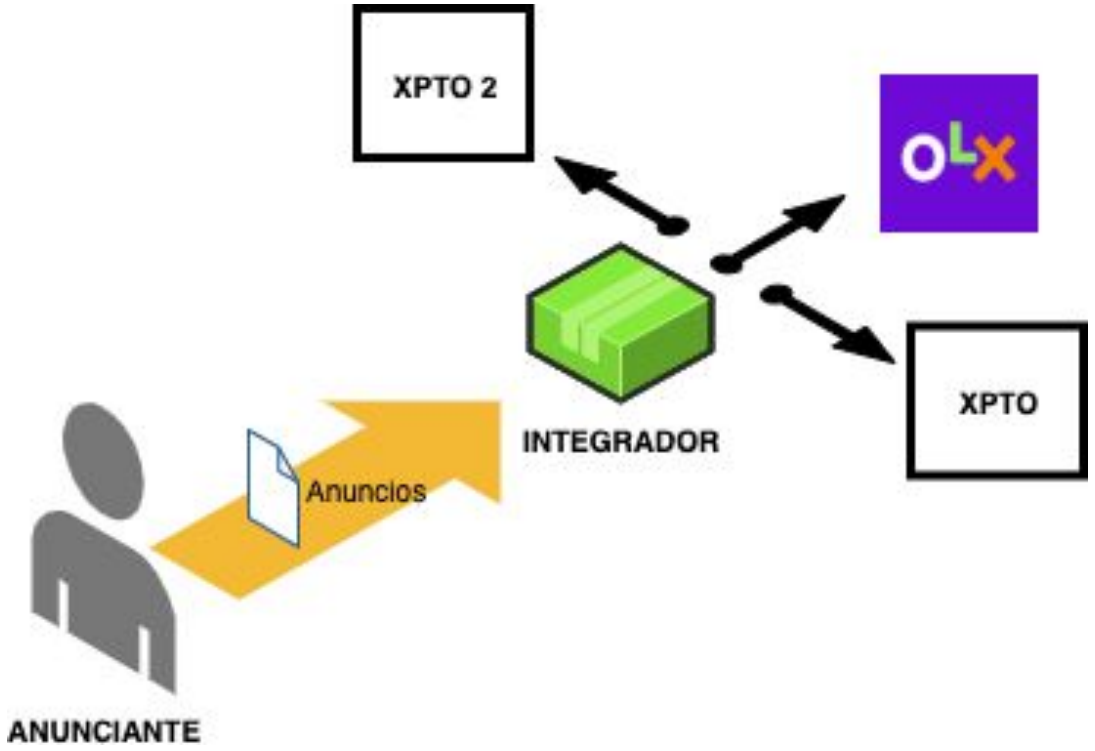
Renato
PM

Danilo
DEV



Edgard
DEV

Processo de integração de anúncios em massa



Nossos números

~4.4K

Contas importadas por dia

~3.6M

Anúncios processados por dia. Isso representa ~25% de todos os anúncios ativos da OLX

~141k

Requests por dia

Monolito: Toda a lógica de importação estava acoplada ao monolito da OLX

Cerberus: Sistema independente para realizar a importação

Microserviços: Dividir o sistema em domínios diferentes para simplificar e escalar a importação

1. *Aprendizado com o passado*

Positivos:

- Deploy independente
- Desacoplamento de infra

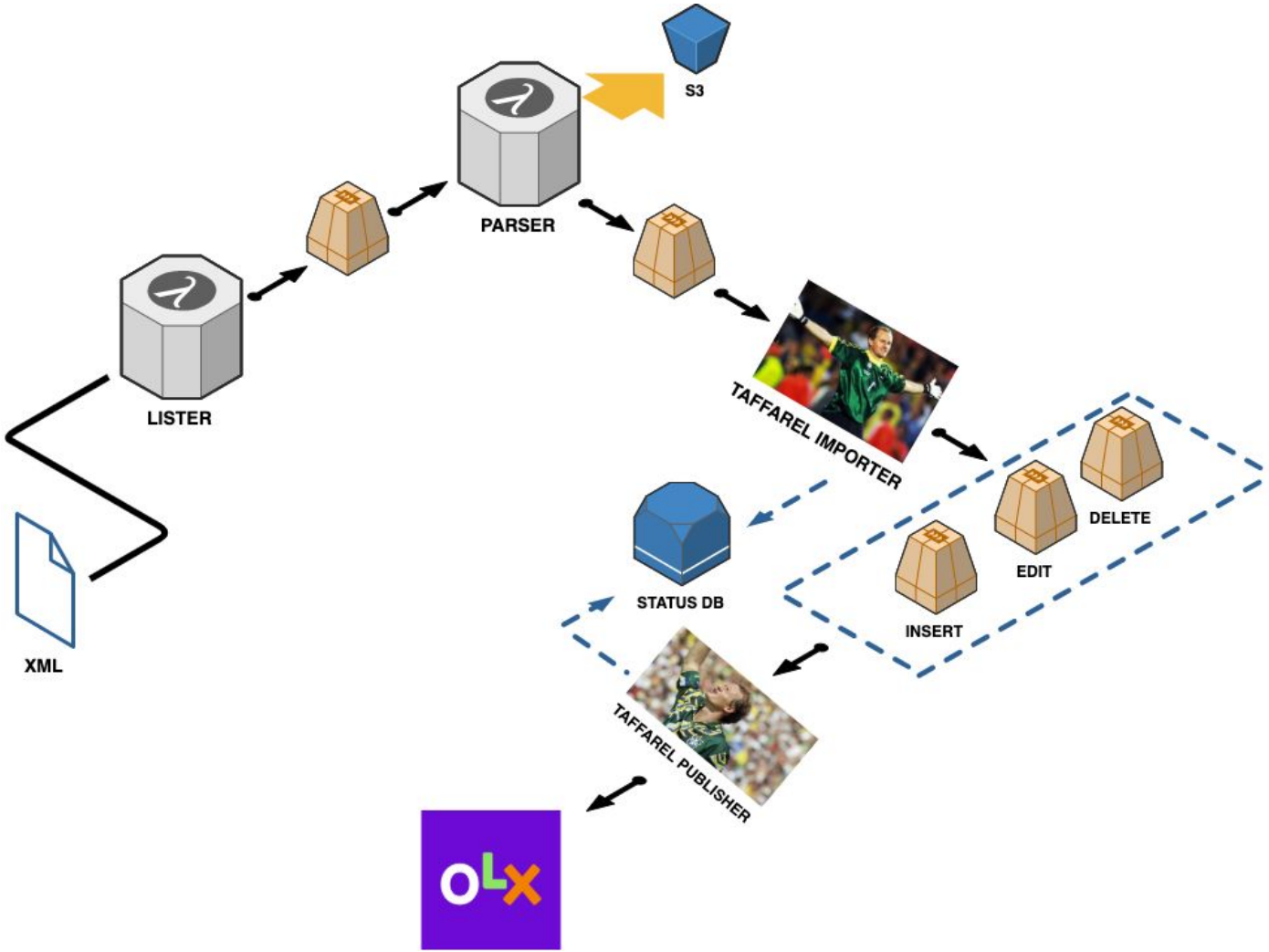
Negativos:

- Custo alto para escalar
- Acoplamento em código
- Troubleshooting

2. *Como começamos.*

Definição dos Serviços

- **Arquivo** Responsável por fazer toda manipulação e conversão dos arquivos dos clientes
- **Conta** Enriquecer os dados dos anúncios com informações relativas a conta (Ex.: telefone) e validações (Ex.: Limite do plano)
- **Anúncio** Operações a nível de anuncio (Ex.: Download de imagem)


S3
PARSER
LISTER
XML
TAFFAREL IMPORTER
STATUS DB
DELETE
EDIT
INSERT
TAFFAREL PUBLISHER
olx

22h

Demoramos quase 1 dia para fazer toda a importação e esse tempo não **não diminuía** com mais máquinas

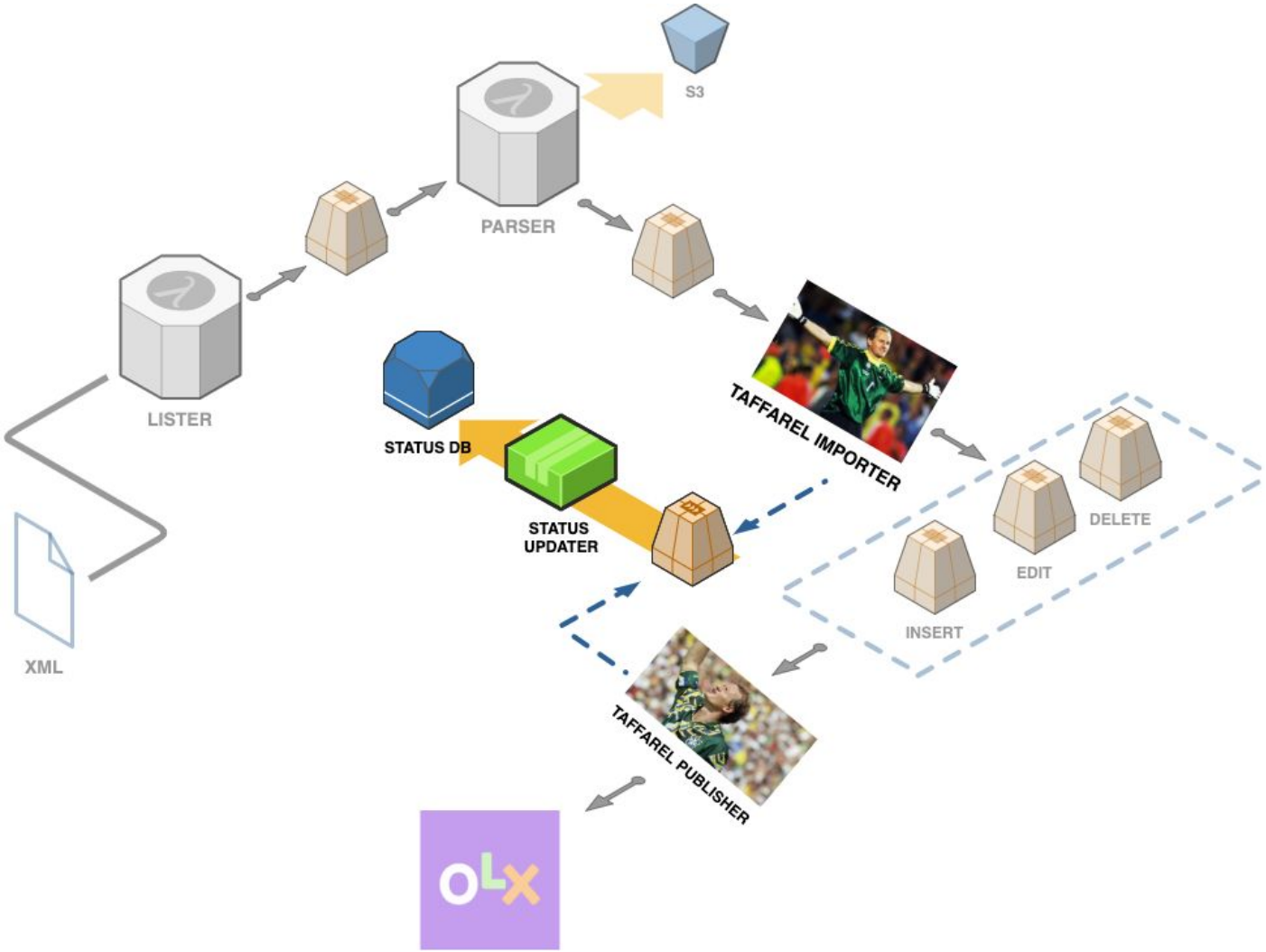
Destaques:

- Separação das operações
- Utilização do SQS
- Stateless

Problemas:

- Base de dados compartilhada
- Serviços externos
- Ausência de cache

Atualização de status assíncrono.


S3
PARSER
LISTER
XML
STATUS DB
STATUS
UPDATER
TAFFAREL IMPORTER
DELETE
EDIT
INSERT
TAFFAREL PUBLISHER
olx

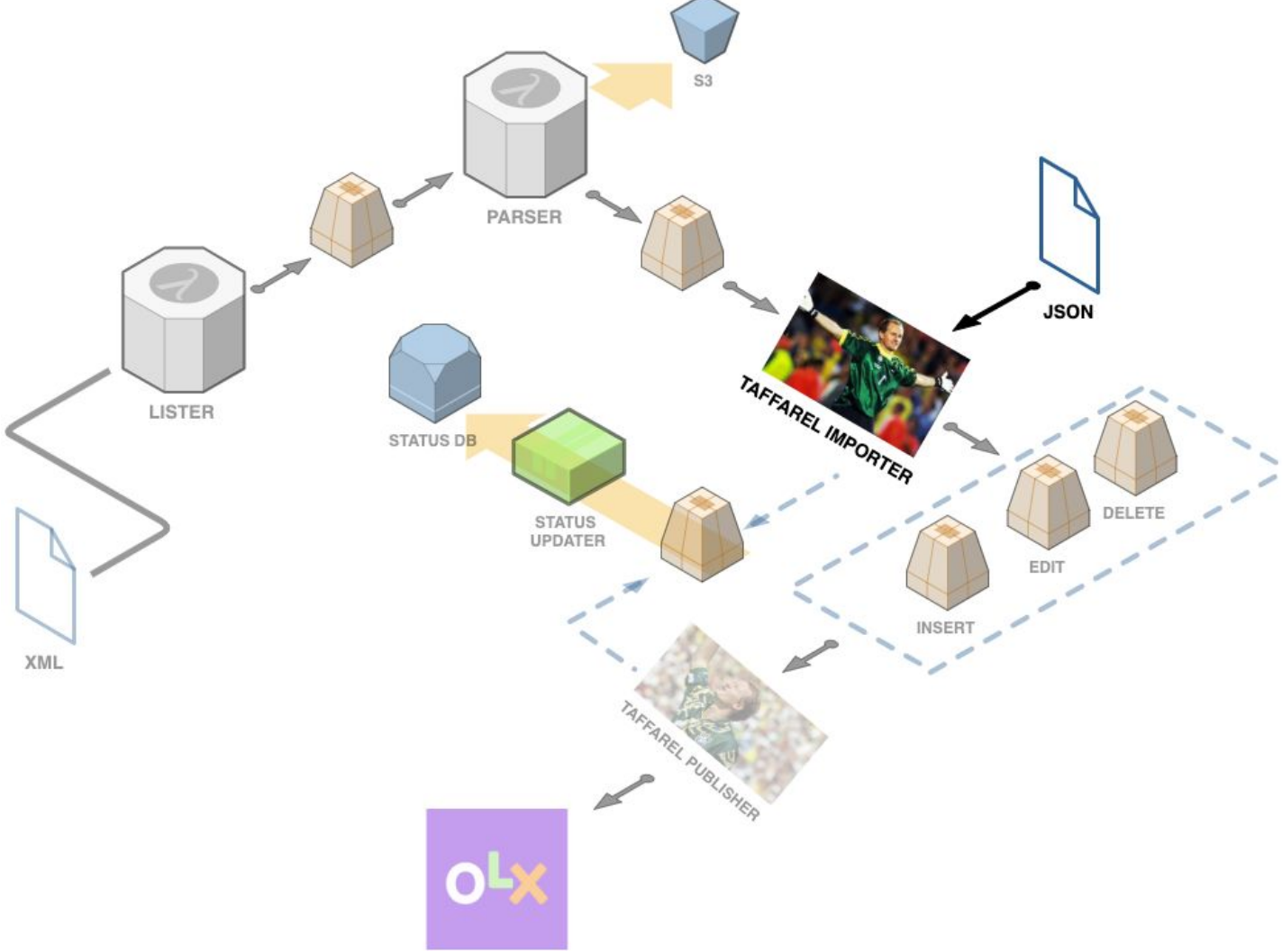
4h

Tempo total caiu para 4h e passou a ser **linear** com a quantidade de workers

Utilizamos um **cache** para evitar edição desnecessária de anúncio e um cache na consulta de CEP

Importação de arquivo finalizada com um tempo total de *projeto de 9 meses*.

Migração da API.


S3
PARSER
JSON
LISTER
STATUS DB
TAFFAREL IMPORTER
STATUS
UPDATER
DELETE
EDIT
INSERT
XML
TAFFAREL PUBLISHER
olx

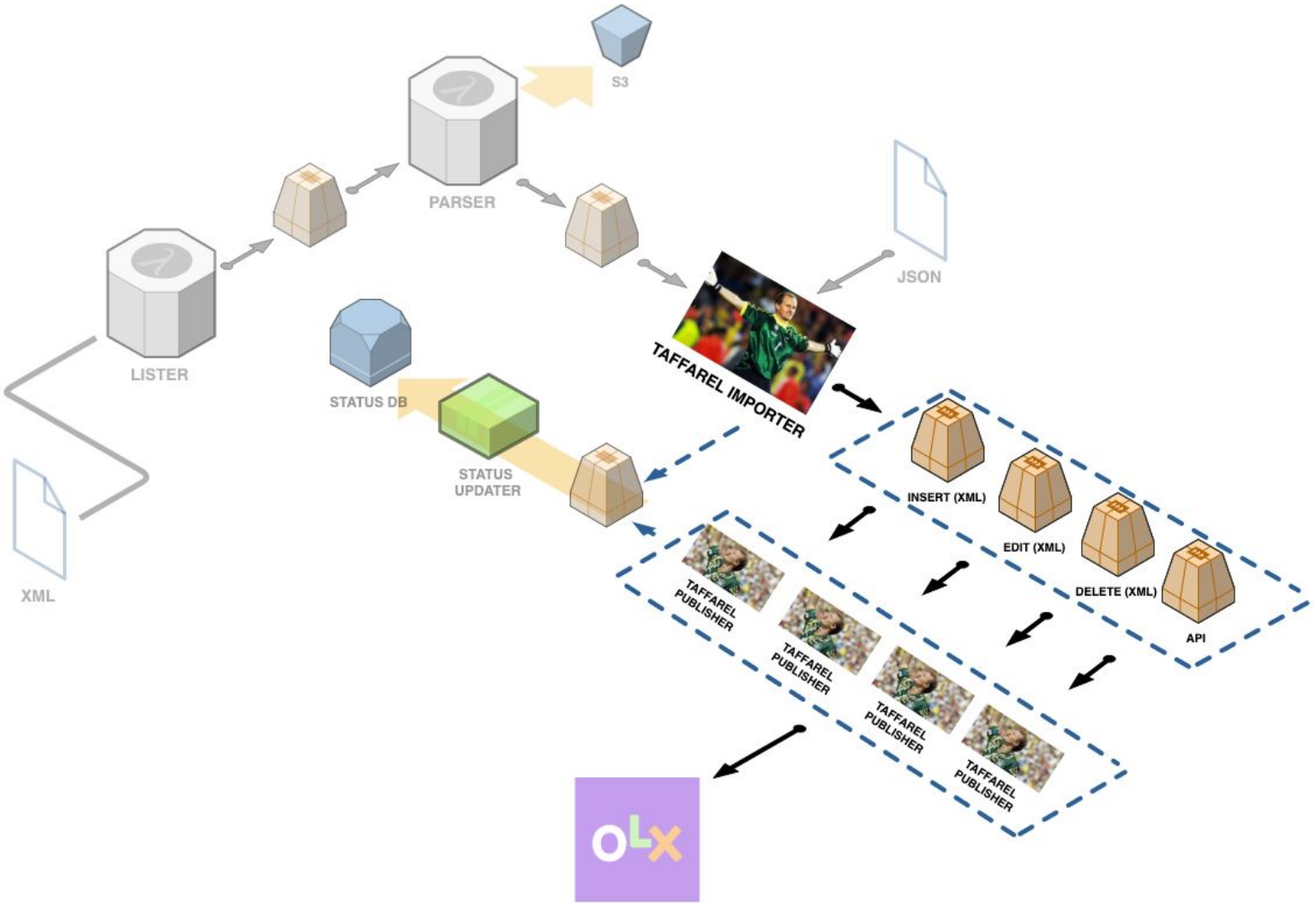
Destaques:

- Sem exceção para atender o fluxo de API

Problemas:

- Fluxo durante todo o dia
- Concorrência entra API e XML

Separação do taffarel publisher.


S3
PARSER
JSON
LISTER
STATUS DB
TAFFAREL IMPORTER
STATUS
UPDATER
INSERT (XML)
EDIT (XML)
DELETE (XML)
API
XML
TAFFAREL
PUBLISHER
TAFFAREL
PUBLISHER
TAFFAREL
PUBLISHER
TAFFAREL
PUBLISHER
olx

5. *Custos.*

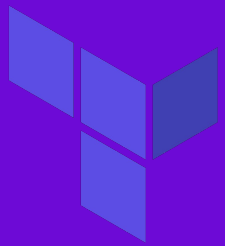
Chegamos em um máximo de *40 EC2 t2.micro* (1 vCpu e 1 GiB ram) rodando.

- $240 de EC2
- $65 de SQS

E se fosse um único serviço?

m4.larger	2 vCpu 8 GiB ram	$ 73.20
m4.xlarger	4 vCpu 16 GiB ram	$ 146.40
m4.2xlarger	8 vCpu 32 GiB ram	$ 292.80

5. *Operação.*



HashiCorp
Terraform

```
module "status_updater_workers" {
  source = "./modules/app_server/"

  team_name   = "${var.team_name}"
  environment = "${var.environment}"
  subnets     = "${var.subnets}"
  vpc         = "${var.vpc}"

  app_name  = "taffarel-status-updater"
  is_worker = "true"

  instance_type       = "${var.status_updater_workers["instance_type"]}"
  min_instances       = "${var.status_updater_workers["min_instances"]}"
  max_instances       = "${var.status_updater_workers["max_instances"]}"
  ec2_keypair         = "${var.status_updater_workers["ec2_keypair"]}"
  iam_profile         = "${var.status_updater_workers["iam_profile"]}"
  security_groups     = "${var.status_updater_workers["security_groups"]}"
  ssl_enable          = "${var.status_updater_workers["ssl_enable"]}"
  ssl_arn             = "${var.status_updater_workers["ssl_arn"]}"
  elb_subnets         = "${var.subnets}"
  elb_scheme          = "${var.status_updater_workers["elb_scheme"]}"
  elb_security_groups = "${var.status_updater_workers["elb_security_groups"]}"

  olx_tags = "${local.base_olx_tags}"
}
```

Gitlab CI

passed **Pipeline #133226** triggered 1 week ago by Danilo Oliveira

Merge branch 'MADS-1220-PT2' into 'master'

```
removing rosetta

See merge request !101
```

12 jobs for `master` in 18 minutes and 31 seconds (queued for 3 seconds)

`7eab73d6`

Pipeline Jobs 12

Build	Test	Publish-image	Deploy-to-development	Integration-test
Build	Test	Publish Image	Deploy API to D...	Integration T
			Deploy Worker...	

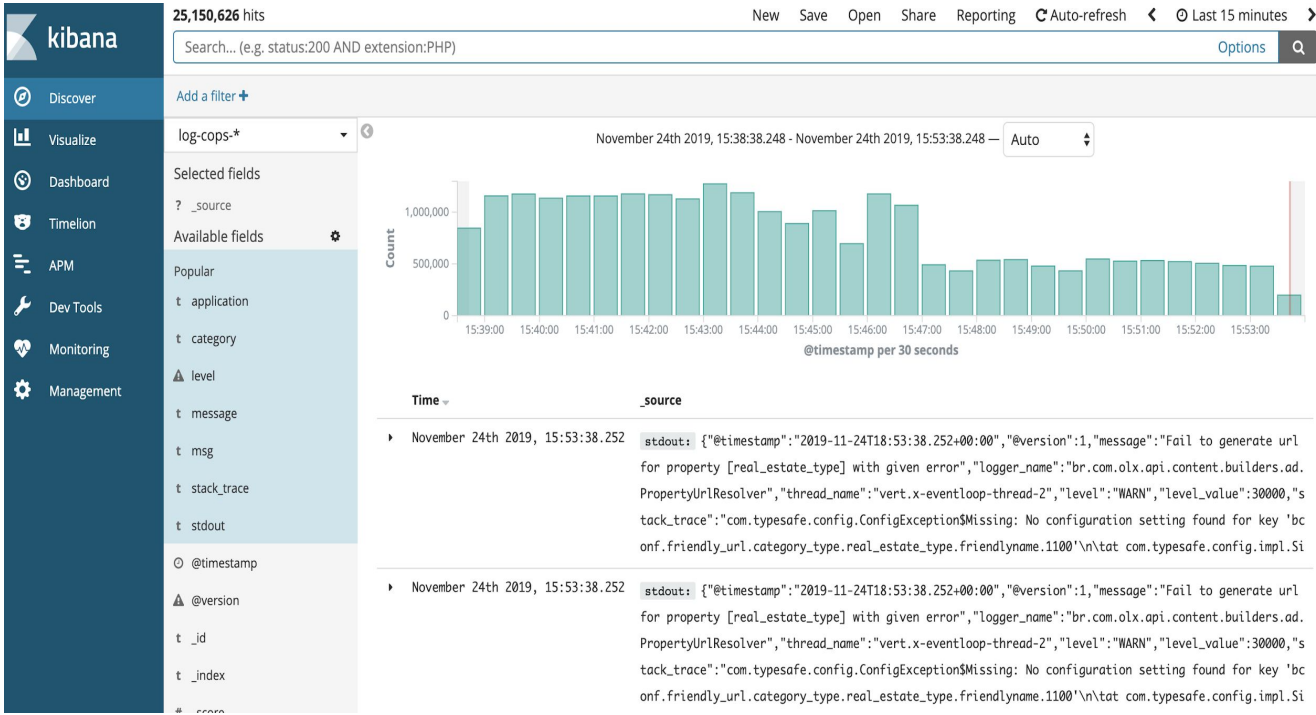
Prometheus

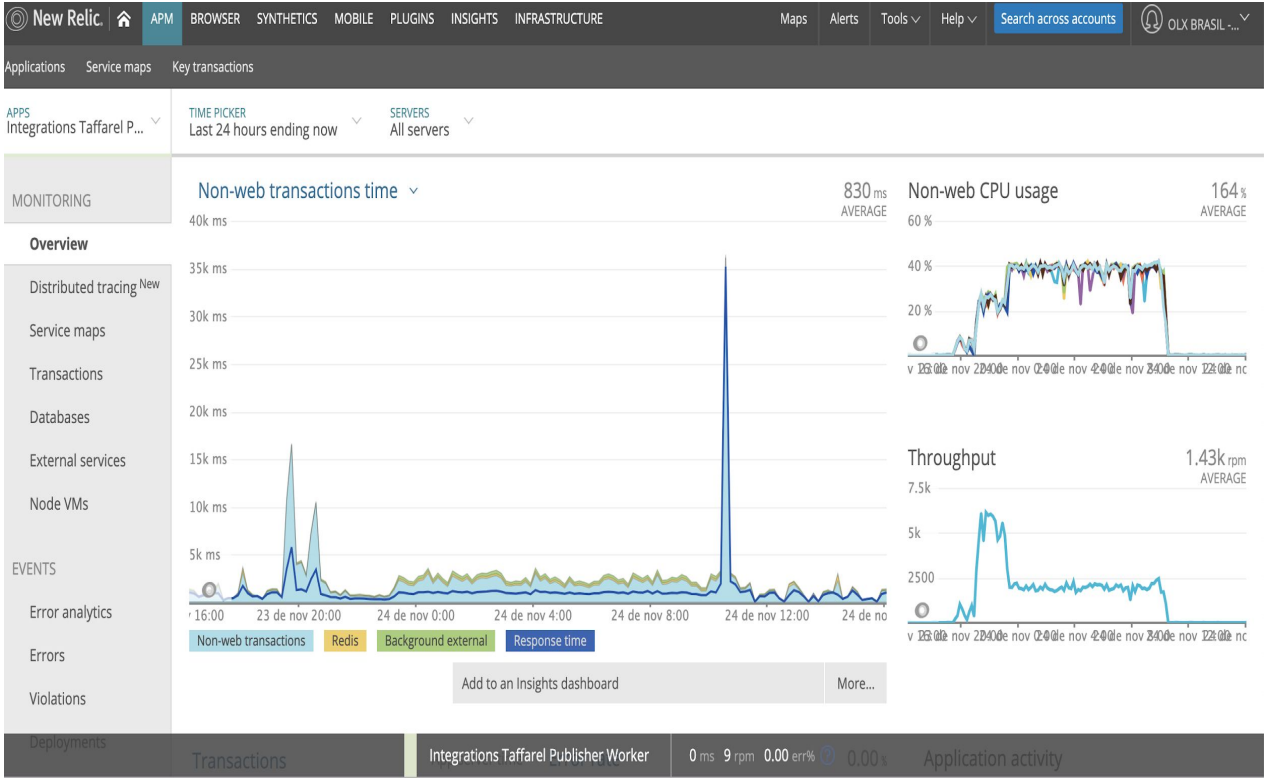
Grafana

Elasticsearch

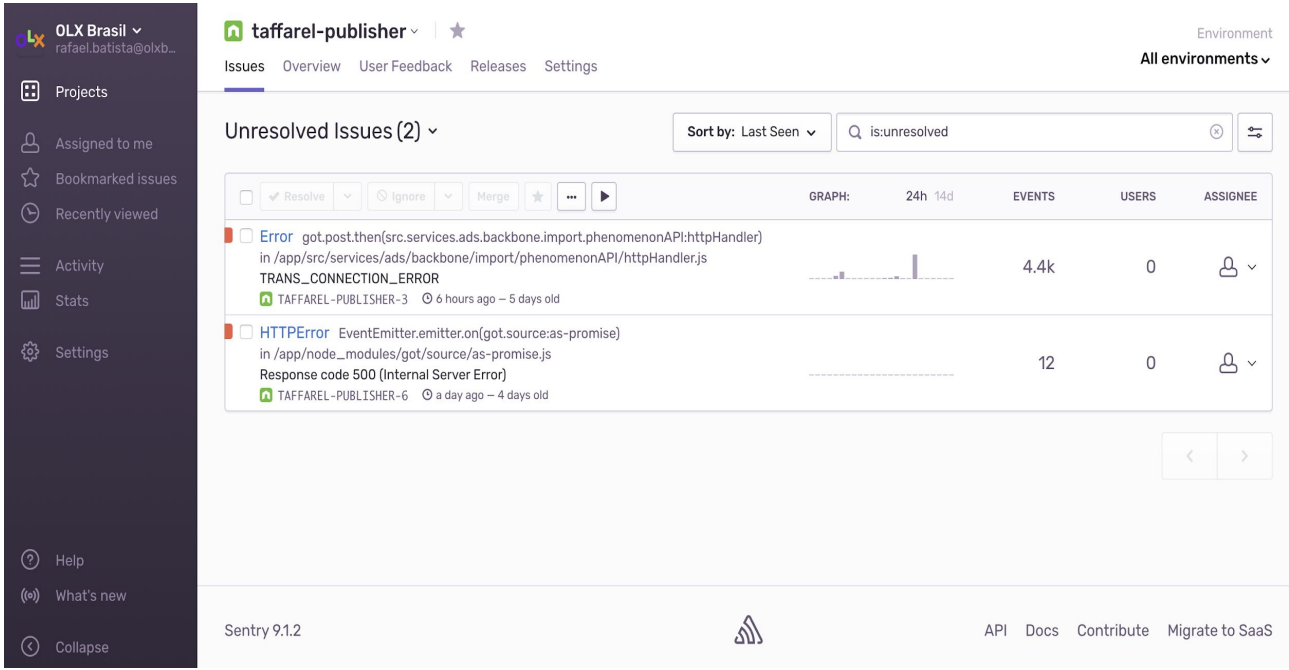
Logstash

Kibana




New Relic.
APM
BROWSER
SYNTHETICS
MOBILE
PLUGINS
INSIGHTS
INFRASTRUCTURE
Maps
Alerts
Tools
Help
Search across accounts
OLX BRASIL -...
Applications
Service maps
Key transactions
APPS
Integrations Taffarel P...
TIME PICKER
Last 24 hours ending now
SERVERS
All servers
MONITORING
Overview
Distributed tracing New
Service maps
Transactions
Databases
External services
Node VMs
EVENTS
Error analytics
Errors
Violations
Deployments
Non-web transactions time
830 ms
AVERAGE
40k ms
35k ms
30k ms
25k ms
20k ms
15k ms
10k ms
5k ms
16:00
23 de nov 20:00
24 de nov 0:00
24 de nov 4:00
24 de nov 8:00
24 de nov 12:00
Non-web transactions
Redis
Background external
Response time
Add to an Insights dashboard
More...
Non-web CPU usage
164 %
AVERAGE
60 %
40 %
20 %
Throughput
1.43k rpm
AVERAGE
7.5k
5k
2500
Transactions
Integrations Taffarel Publisher Worker
0 ms 9 rpm 0.00 err%
0.00 %
Application activity

Sentry


OLX Brasil
rafael.batista@olxb...
Projects
Assigned to me
Bookmarked issues
Recently viewed
Activity
Stats
Settings
Help
What's new
Collapse
taffarel-publisher
Issues
Overview
User Feedback
Releases
Settings
Environment
All environments
Unresolved Issues (2)
Sort by: Last Seen
is:unresolved
Resolve
Ignore
Merge
GRAPH:
24h 14d
EVENTS
USERS
ASSIGNEE
Error got.post.then(src.services.ads.backbone.import.phenomenonAPI:httpHandler)
in /app/src/services/ads/backbone/import/phenomenonAPI/httpHandler.js
TRANS_CONNECTION_ERROR
TAFFAREL-PUBLISHER-3
6 hours ago – 5 days old
4.4k
0
HTTPError EventEmitter.emitter.on(got.source:as-promise)
in /app/node_modules/got/source/as-promise.js
Response code 500 (Internal Server Error)
TAFFAREL-PUBLISHER-6
a day ago – 4 days old
12
0
Sentry 9.1.2
API
Docs
Contribute
Migrate to SaaS

OLX

Obrigado!

Dúvidas?