kruize / autotune Public

Code

Issues 84

Pull requests 22

Discussions

Actions

Projects

Wiki

Security

Insights

master

dinogun Merge pull request #428 from dinogun/prepare4hpoaas
...
12215c1 May 2, 2022
459 commits

.github/workflows
Update PR checks to use the appropriate operator and optuna images.
Nov 17, 2021

design
Update API doc to incorporate experiment_name and stack changes.
Dec 1, 2021

docs
Update autotune_install.md to reference kruize-demos
Apr 5, 2022

examples
Change "sla" to the more appropriate nomenclature "slo" globally.
Oct 20, 2021

extensions
inital directory structure of the autotune project
Nov 4, 2020

hyperparameter_tuning
Create ApplicationServiceStack class based on Container Image and not…
Nov 29, 2021

manifests
Merge pull request #407 from chandrams/operator_crd
Feb 17, 2022

scripts
Create separate container for Java and Python
Nov 15, 2021

src
Merge pull request #391 from dinogun/layer_presence_query_array_m
Jan 31, 2022

tests
Updated query_url based on changes in PR 362
Jan 27, 2022

CONTRIBUTING.md
Fixed a typo - changed PGP to GPG
Mar 22, 2022

Dockerfile.autotune
Create separate container for Java and Python
Nov 15, 2021

Dockerfile.optuna
Add missing jsonschema package in pip install.
Nov 17, 2021

LICENSE
added autotune minikube deployment scripts
Dec 8, 2020

README.md
Doc changes: autotune-demo repo is being renamed to kruize-demos
Apr 5, 2022

build.sh
Create separate container for Java and Python
Nov 15, 2021

deploy.sh
Updated usage information in deploy script and docs
Dec 6, 2021

pom.xml
Update pom.xml version to 0.0.7
Jan 28, 2022

README.md

Kruize Autotune - Autonomous Performance Tuning for Kubernetes !

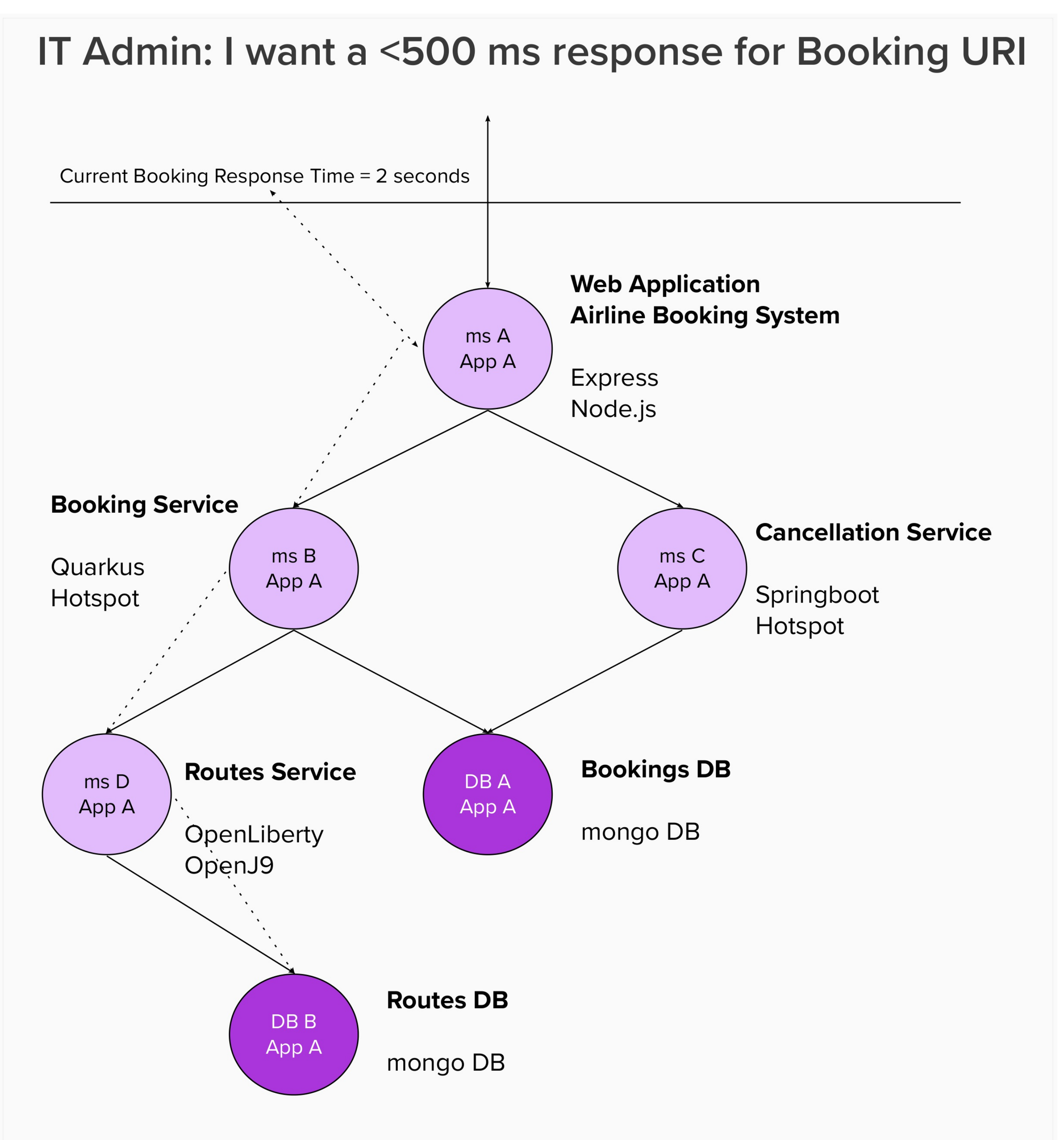
What is Kruize Autotune ?

Kruize Autotune is an Autonomous Performance Tuning Tool for Kubernetes. Autotune accepts a user provided "slo" goal to optimize application performance. It uses Prometheus to identify "layers" of an application that it is monitoring and matches tunables from those layers to the user provided slo. It then runs experiments with the help of a hyperparameter optimization framework to arrive at the most optimal values for the identified set of tunables to get a better result for the user provided slo.

Autotune can take an arbitrarily large set of tunables and run experiments to continually optimize the user provided slo in incremental steps. For this reason, it does not necessarily have a "best" value for a set of tunables, only a "better" one than what is currently deployed.

Motivation

Docker and Kubernetes have become more than buzzwords and are now the defacto building block for any cloud. We are now seeing a major transformation in the industry as every product/solution/offering is being containerized as well as being made kubernetes ready (Hello YAML!). This is throwing up a new set of challenges that are unique to this changing environment.



Consider an Flight Booking Application as shown in the figure. It consists of a number of microservices which are polyglot in nature and are running in a Kubernetes cluster. Consider a scenario where the user doing a booking is getting a very slow response time. The IT Admin is now tasked to reduce the overall response time for the booking URI.

Tracking down performance issues in a dynamic microservices environment can be challenging and more often than not, stack or runtime specific optimizations are written off as too complex. This is because runtime optimization is a very involved effort and requires deep expertise. Common fixes are mostly limited to increasing pod resources, fixing application logic to make it more optimal or increasing horizontal pod auto-scaling.

How do I start ?

Autotune helps to capture your performance tuning needs in a comprehensive way and runs experiments to provide recommendations that help achieve your slo goals. So how does it do it ? We recommend you check out the kruize-demos repo for a quick start !

Installation

See the Autotune Installation for more details on the installation.

REST API

See the API README for more details on the Autotune REST API.

Autotune Architecture

See the Autotune Architecture for more details on the architecture.

Contributing

We welcome your contributions! See CONTRIBUTING.md for more details.

License

Apache License 2.0, see LICENSE.

About

Autonomous Performance Tuning for Kubernetes !

#kubernetes #performance #hyperparameter-optimization #performance-tuning #sla #autotune #tunables

Readme
Apache-2.0 license
52 stars
14 watching
27 forks

Releases

4 tags

Packages

No packages published

Contributors 11

Languages

Shell 49.4% **Java** 46.6% **Python** 4.0%