

И первый паттерн, с которого мы начнём – довольно простой для понимания. Во всяком случае, для меня, мне достаточно просто его понять. Это Health probe – проба жизни. Нравится, название, очень фэнтезийное звучит. Суть проста, при запуске вашего pod-а вы определяете некоторые правила, которые определяют: жив pod или его надо перезапустить. И Kubernetes просто, время от времени, этот pod проверяет на какие-то формальные условия и его перезапускает, если pod не отвечает.

Зачем нужна Health probe? Соответственно, мы знаем, что docker-контейнеры при остановке главного процесса, то есть если процесс рухнул с какой-либо ошибкой – перезапускаются, чтобы понять этот процесс заново. Это такой встроенный в docker механизм и в Kubernetes-е он тоже есть. При остановке главного процессора ваш контейнер будет перезапущен. И этого достаточно, чтобы подстраховаться от самых базовых ошибок, когда вы там какую-то ошибку в приложении сделали, или приложение слишком много памяти выжирает, у него больше памяти нет. Оно сколлапсировалось в чёрную дыру, поглотило вселенную и перезагрузилось.

Но этого недостаточно, потому что у вас может быть, например, веб-сервер, который сам по себе работает, но, например, не отвечает ни на одну команду по той причине, что, например, у него нет связи с базой данных. Ошибка? – Ошибка. Критичная? – Нет. Процесс будет продолжен, и веб-сервер, в принципе, довольно-таки живучая штука сама по себе, то есть эти платформы Java, Golang, неважно, что, – они довольно-таки живучие, они не падают при первых проблемах, если вы только их не уроните сознательно.

И чтобы вас немножко подстраховаться, чтобы подстраховаться разработчиков этих сервисов, мы можем добавить проверку, например, на то, что мы обращаемся к какому-то endpoint-у этого веб-сервера, и endpoint веб-сервера нам возвращает

что-то осмысленное. Если он не возвращает что-то осмысленное, значит, контейнер мёртв, контейнер не здоров, контейнер надо перезапустить – это позволяет нам сделать Kubernetes.

И соответственно, это используется как механизм Auto Recovery – во времена, когда я был наивен и юн, сервера перезапускали руками, если что-то прилегло. Сейчас естественно, с современными нагрузками, особенно если вы host-итесь в облаках – это значит, что у вас уже какие-то там запросы идут в секунду и остановка даже на 5 минут..., даже на 3 минуты... Время человеческой реакции – оно довольно большое, будет значить, что вы какое-то количество клиентов потеряли, поэтому Kubernetes, естественно, умеет перезапускать эти контейнеры сам, когда Health probe-а у нас отказывает.

Следующий паттерн, который я объединил в 1 раздел с Health probe, потому что они очень похожи по смыслу и по сути – это Readiness probe. По факту это то же самое, что и Health probe. То есть Kubernetes периодически по какому-то условия проверяет платформу, но здесь всё немножко сложнее. То есть если Health probe у нас проваливается – контейнер считается мертвым и подлежит немедленному рестарту для того, чтобы приложение снова могло работать.

В случае с Readiness probe, при провале этой пробы, контейнер..., pod просто исключается из общей ротации контейнеров, то есть он считается как неготовый обрабатывать запросы. Как так? Например, у вас веб-сервер, который быстро поднимается довольно, быстро соединяется с базой данных и готов к работе. Здоров ли он? – Он здоров, он может что-то принять и что-то отдать. Но, например, ему надо раскатить миграции, а до этого момента мы не можем туда клиентов пускать, потому что он фигню им отдаст всякую. Он не отдаст им нужные данные. И в этом момент пока наш сервис накатывает миграции – он считается

здоровым, но не готовым. То есть он не включен в общую ротацию наших серверов. И запросы туда, соответственно, приходить не будут.

Так что, в принципе, ответ на вопрос: «зачем Readiness probe?» – Он уже есть. Для Kubernetes-а внутренний смысл в том, что Kubernetes не должен полагаться, что запущенный процесс может выполнять свои функции. Kubernetes не должен просто ждать, когда контейнер поднимется с процессом и сразу туда лить трафик. Возможно, вам нужно раскатить какие-то миграции, может у вас контейнер должен создать служебные файлы. Может у вас контейнер должен соединиться с сервисом авторизации, получить сначала от него какие ключи. Короче, между живым здоровым контейнером и контейнером, который готов обслуживать – может пройти какое-то время, иногда довольно-таки большое.

Соответственно, запущенный и работающий с основной процесс – это тоже совершенно ничего не значит по той же причине. И в production-е форвард трафика на неготовые к работе приложения может привести к тому, что клиент увидит пустую страничку, увидит внутренку ваших скриптов. Увидит что-то совсем неприличное, расскажет об этом друзьям. Друзья будут над вами долго смеяться, в общем это не здорово, поэтому Health probe и Readiness probe надо очень чётко определять, очень четко им задавать разницу.

То есть Health probe может быть несколько более грубым, никакой логики в себе не содержать, если какой-то запрос на страничку вашего сервиса, отдающий просто статические какие-то данные, готов, то есть выдает 200, значит приложение здорово – его не надо перезагружать. Если, соответственно, вы там попытаетесь получить под определенными пользователями список определенных каких-то компонентов с авторизацией, которая у вас идет в сервис авторизации. И сервис может не отдавать чего-то или отдавать пустую страницу, вот значит, он ещё не готов. Значит, пусть он соединится с вашим сервисом авторизации,

поднимет свою миграции, и потом уже вы сможете этот список получить, например, список и объявить его готовым.

Итак, причем тут вообще GO, зачем нам на этих Health/Readiness probe-ах нужен Golang? Начнём с того, что [07:04] HPЗБ и Readiness probe нам Kubernetes предоставляет варианты либо обращаться по http какому-то endpoint-у, либо выполнять отдельную команду. Отдельная команда поднимает маленький контейнер, который выполняет эту команду и завершается. Соответственно, он ест ресурсы, пробы запускаются там. Вы можете их настроить, но они, как правило, должны запускаться довольно часто, чтобы у вас там мертвые контейнеры, на них не падал трафик раз в секунду, раз в 2-3-5 секунд. Поэтому чем меньше ресурсов вы используете, тем лучше. И также бывает, что нужно пробу положить внутрь контейнера. То есть для каких-нибудь баз данных, где у них секундно, нужно как-то обращаться через socket, чтобы получить какие-то метрики, снаружи их получить нельзя.

И если вы это делаете, то опять же, мы помним, что размер контейнера – это довольно критичная величина. Чем больше контейнер, тем больше памяти он занимает, тем он неповоротливый, тем сложнее его заскейджелить, то есть положить на какую-то node-у, найти ресурс, и тем больше будут зазоры между контейнерами, тем менее оптимально вы будете использовать ваши же ресурсы. Поэтому чем меньше само приложение мониторинга, тем лучше. Поэтому во многих случаях для многих баз данных приложение проб пишется как раз на Golang-е.

Давайте рассмотрим пример Health probe-ы и Readiness probe-ы и начнём с файлов, которые мы можем как раз..., которыми мы можем описывать состояние Kubernetes-овского кластера. Соответственно, эти файлы..., вот тот файл, который вы видите сейчас передо мной, который вы видите сейчас на экране – это файл

описания состояния, которого Kubernetes должен достичь при помощи команды Apply. То есть вы передаете команду Apply, вы передаете её туда в файл. Файл сконвертится в Kubernetes.api, и Kubernetes сам решает, каким образом он этот контейнер, этот файл поднимет. Соответственно, давайте посмотрим, что здесь есть в этом файле.

Итак, начинаем с того, что этот файл описывает версию Kubernetes-а, которую мы используем. Есть v1, v1 beta1, v1 beta2 и так далее. Тип ресурса, который мы используем, мы уже видели в команде `kubectl get pods`, то есть команда `pod`-ы получает. И соответственно, ресурс у нас тоже `pod`. И здесь у нас есть `metadata`, то есть это такие специальные данные, которые нам нужны для уже определения непосредственно каких-то специальных ролей или просто маркировки того, что вы задеплоили. То есть вы можете указывать лейблы, можете указывать имя – это нужно вам для отличия контейнеров друг от друга.

И соответственно, дальше у нас идет спецификация..., отличие ресурсов, вернее, друг от друга. Дальше у нас идет спецификация. Спецификация это уже что мы непосредственно будем делать? В данном случае в спецификации мы указываем, что мы хотим загрузить контейнеры, и дальше у нас идет перечисление. В описании формата `yaml` я уже указывал, что перечисление начинается с такого значка минус.

Мы загружаем контейнер с именем `liveness`, который у нас будет лежать в `pod`-е с именем `liveness-exec`. И естественно, указываем ему `image`. В нашем случае мы указываем `image: busybox` – это такой стандартный `image` Kubernetes-а маленький, в котором можно запускать какие-то программы. И мы указываем аргумент. Аргумент – это тоже самое, что и `entry point` в `docker`-контейнере, то есть это какая-то..., какой-то основной процесс, который будет у нас запускаться в

контейнере, и в данном случае в shell-е мы создаем файл tmp/healthy, спим 30 секунд и удаляем этот файл и спим 5 минут, чтобы показать, как работает проба.

Соответственно, проба у нас с помощью команды `exec` выполняет какое-то действие. И ожидая от этой команды, что она выйдет с кодом 0. Если команда выходит с каким-нибудь другим кодом, проба считает, что pod не жив, pod, скорее, мёртв. Его надо перезагрузить, перерескейджелить, то есть поднять заново. И начинаем мы проверять после 5 секунд, после того как pod заскейджелился. То есть цикл у нас какой? – У нас Kubernetes получает команду, ищет node-у на которой есть свободные ресурсы, там ей говорит: «Node-a, подними вот это вот». И node-a говорит: «Окей». И значит, после того как node-a сказала «окей», приняла к исполнению, через 5 секунд начнёт запускаться вот эта probe-a. То есть за 5 секунд ваш контейнер должен..., ваш процесс должен подняться и соответственно, проба будет у вас запускаться с периодом каждые 5 секунд и проверять этот процесс. Можно запускать чаще, можно позже, лучше запускать чаще. Вернее, чаще реже..., лучше запускать чаще, чем реже, потому что таким образом у вас будет более быстрая реакция, то есть если у вас что-то упадет, а у вас там проба 60 секунд стоит, у вас 60 секунд всё это будет лежать. Лучше, конечно, запускать почаще. Давайте посмотрим, как это работает.

Соответственно, можно было уже увидеть, что в namespace default мой pod уже создан, поэтому давайте создадим новый namespace, посмотрим на команду `create`. Create namespace, назовём его testcluster. Соответственно вот, у нас теперь namespace создан, и мы можем на него посмотреть. Вот наш testcluster, соответственно в нём нет довольно-таки ничего, поэтому давайте туда как раз deploy наш healthprobe-exec. И используем для этого команду `kubectl apply`. Соответственно, в kubectl-е есть такое правило, что если вы указываете команду без namespace-a, то она по умолчанию будет ссылаться на default.

Я хочу в данный момент загрузить мой liveness probe в test cluster. Для этого я буду также, как и в случае с командой get pods указывать мой namespace. Testcluster apply, дальше я указываю, что мне нужно здесь сделать из файла. И я забыл healthprobe-exec. Я соответственно, да, попутал имя файла. И так, жмём apply. Да, опять я пишу тут кучу интересных команд, которые не нужны.

Соответственно, вот, наш pod создан. Мы можем посмотреть в нашем testcluster-e get pods наш pod. Видим, что он запущен и можем сделать..., делать describe pod, то есть описать, что у нас происходит в pod-е. Здесь мы что видим? Здесь мы видим имя, видим куда он загружен, видим его приоритет, node-у с ip-адресом, когда он стартанул. Вот его лейблы, нет аннотации, потому что не указаны, его внутренний IP. И ID контейнера, тут у нас docker-контейнер, тут у нас docker под капотом, поэтому docker. Соответственно какой image, Image ID. Port и host, но у нас указаны, аргументы для запуска. То есть всё, что мы указали, в принципе мы видим, плюс служебная информация от Kubernetes-а, когда что случилось. И соответственно, здесь мы видим ещё, что с ним происходило конкретно.

Соответственно, мы создали... Нет, вернее не так. Мы заскейджелили, мы соответственно, на единственную доступную нам node-у, поскольку у нас docker работает только на нашем компьютере, положили, значит, наш контейнер, наш scheduler это сделал, то есть тот самый механизм Kubernetes-а, который указывает, куда что будет лежать. Туда мы стянули изображения busybox. Стянули его за 690 миллисекунд. Соответственно, создали контейнер, и он у нас успешно запустился. Запустился у нас с пробой, видя, что проба работает, через 30 секунд она..., у нас файл будет удален, и ещё через 5 секунд у нас проба должна упасть. Она уже упала 1 разок, пока я тут разглагольствовал и соответственно, мы теперь можем посмотреть на её... Давайте, ещё раз опишем её. И вот, мы видим, что она Unhealthy. Не получилось открыть у нас файл и соответственно, проба у нас restart-анула. То же самое у меня уже происходит некоторое время в моём

default контейнере – уже 22 раза это произошло. И опять же, мы можем описать наш ресурс. И вот мы видим, сколько раз там у нас уже всё случилось. У нас уже там warning-ы, backup-ы, то есть эта история – она чистится, и поэтому всё вы там не увидите, но видим, что у нас всё перезагрузилось. Вот так примерно работает Health probe-а, если она что-то не может найти в какой-то промежуток, она перезагружает контейнер. Вернее, перезагружает pod вместе со всеми его контейнерами.

И перед тем, как мы перейдем к следующей части, где я покажу другой тип Liveness probe-ы, мы удалим из нашего default-а нашу liveness-ехес, которая уже дофига раз перезагрузилась. С помощью команды delete мы удаляем pod. Опять же, ваш клиент дождется, пока pod сменится статус на удаленный, то есть если pod активен, ему отправится команда завершить процесс. Он какое-то время будет завершаться, потом scheduler его должен с node-а удалить. Node-а должна отпортовать, что всё прошло хорошо, поэтому это занимает некоторое время.

Видим, что наши pod-ы удалены. Больше у нас никаких ресурсов нет и теперь давайте посмотрим на второй наш..., нашу probe-у на вторую, нашу http-probe-у. Соответственно, что здесь происходит? Здесь у меня немножко отличаются уже всё с самого начала. Это не связано с probe-ой, это просто другой вид deployment-а... Другой вид описания, который я вам хотел показать. Это другой вид ресурса, соответственно мы видим, что здесь отличается kind, то есть здесь deployment, там был pod. И мы видим, что у нас отличаются версии api, то есть здесь api-версия – это aPps/v1, здесь был просто v1. В чем прикол?

Если мы откроем kubectl api-versions, мы увидим кучу различных версий api. А если мы откроем команду kubectl api-resources, то у нас будет ещё и описание, какие версии api чему соответствуют. То есть здесь ресурсы, которые вы можете использовать в Kubernetes-е, они все перечислены. У них даже есть сокращения,

то есть `get ns` и `get namespaces` – это одно и то же. То же самое с `node`-ами, то же самое с..., например, `cronjob`-ами, которые мы будем разбирать несколько попозже. Соответственно, здесь перечислены виды ресурсов, и мы видим, что наш `deployment` находится в `api-version apps/v1`, потому что Kubernetes разделил ресурсы по всяким различным версиям `api` для разграничения логических..., логически разграничивая их. И теперь вам нужно указывать ресурсы в..., и теперь можете указываться версия `api` в непосредственно файле описания иначе у вас просто не заработает то, что версия `api v1` не существует `deployment`-а. Значит, что, когда вы попробуете запустить, он вам скажет, что не знает, что такое `deployment` для `v1`. Не надо паниковать, открываем `kubectl api-resources`. Смотрим, что у вас на кластере доступно и соответственно, ставим соответствующую версию `api`.

То есть мы смотрим на тип `deployment`. Почему я его использую? Это просто другой способ задеплоить ваш ресурс и скорее всего, тот которым вы будете пользоваться, потому что по умолчанию что происходит? Я удалил `pod`, как вы видели, и в результате удаления этого `pod`-а он удалился. Удивительно, правда? В смысле, в том плане, что если у вас, например, `pod` принадлежит `node-e`. То есть если у вас `node`-а по какой-то причине выпала из сети: провалился (HP3 21:05), перезагрузился Kubernetes, перезапустилась сама `node-e`, `pod` не восстановится, потому что `pod` принадлежит `node-e`, `pod` не считается живучим объектом. То есть то, что может случиться с `pod`-ом, это если приложение упадет – он `restart`-анет, но, если упадет `node-e` или что-то в таком духе – `pod` заново никуда не заскеджуалится, потому что считается, что вот он существует на `node`, он умрет вместе с `node`-ой. Это не рабочий подход для распределенных приложений, естественно.

Рабочий подход – это, во-первых, держать больше, чем один `instance` приложений на тот случай, если что-то вдруг откажет, упадет и так далее. И держать его так,

чтобы в случае выхода из строя какой-то `node`-ы включилась другая. На ней поднялись все эти приложения и продолжило всё работать. Для этого используют..., раньше использовали тип ресурса `ReplicaSet`, теперь используют тип ресурса `deployment`, который считается приемником `ReplicaSet`-а. И соответственно, я не буду углубляться в чем их отличие – у нас, в конце концов лекция не про `Kubernetes` как таковой, а про применение..., про то, как вам базово менеджить `Kubernetes` с помощью `Golang`-а.

Соответственно, у `deployment`-а точно также есть `metadata` и лейблами и именем. Существует спецификация, и вот спецификация здесь отличается. Как минимум, мы видим число реплик, то есть вы можете указать здесь `1-2-3-15`, если вам хочется, и ваше приложение поднимет столько `pod`-ов..., ваш `deployment` поднимает столько `pod`-ов приложения, сколько надо для удовлетворения этого атрибута. До сих пор, пока на `node`-ах пока хватает места. Потом есть `selector`, на что применяется наш `deployment`, и есть `template`. `Template` – это уже описание непосредственно того, что здесь будет происходить. То есть `pod`-у была достаточна спецификация, `deployment`-у нужен ещё `template`.

У каждого `template` внутри `deployment`-а будет тоже `metadata`, будут какие-то лейблы. То есть у нас `label app`, и у нас здесь есть пометка `liveness`, то есть `selector` сработает, посмотрит, что `app`-ы совпадают. И соответственно, в этом `template` выполнит спецификацию. Спецификация у нас точно также включает контейнеры. В данном случае, мы берем тестовый контейнер, который из себя представляет такой маленький веб-сервер. У него есть аргумент `/сервер`, который запустит непосредственно приложение, отвечающее на какие-то запросы. И есть `endpoint`, который называется `healthz`, который как раз нам отдаст состояние этого контейнера. И здесь мы уже обращаемся по `pod`-у. И соответственно, передаем какие-то `http-header`-ы и начинаем по 3 секунды, с периодом в 3 секунды. Давайте задеплоим, да и посмотрим, что у нас получается.

Итак, применяем наш файл. Естественно, не пишем на латинице..., на кириллице то есть. и healthprobe у нас будет http. Итак, видим, что у нас здесь отличается. Это ж не просто pod создан, у нас deployment.apps created. Если мы посмотрим на pod-а, мы увидим, что у pod-а есть приписка – это такой label от deployment-а для того, чтобы отличать этот pod. И если мы как раз сделаем describe наш pod, то мы увидим, что соответственно, у нас здесь есть label app-liveness, есть pod-template-hash, то есть то, что у нас уже создано с ReplicaSet-ом. Вот у нас есть контроль ReplicaSet-а, то есть это уже от нашего deployment-а.

И соответственно, мы видим здесь то же самое с Liveness probe-ой. Liveness probe-а нам говорит, что код должен быть 200, если код какой-то другой, то у нас значит fail. И тут, в принципе, больше особо не о чем рассказывать. Здесь всё понятно. Мы можем также определить, кстати, в каких случаях probe-а считается, что достигла успеха. Сколько раз ей нужно сработать, чтобы успешно считать pod здоровым и сколько раз ей нужно провалиться, чтобы считать pod нездоровым. Можно увидеть, что default – это 1 и 3. То есть 3 провала, replica говорит: «Нет, что-то не так. Перезагружаемся». 3 успеха, replica говорит, что: «Всё так. Стартуем». Соответственно, да, мы видим, что replica у нас проваливается, у нас тут опять какие-то restart-ы, но нас интересует следующая вещь также в pod-ах.

То есть мы видим, что pod у нас вот такой: liveness-http. И теперь самое главное: что будет, если я удалю этот pod? Я такой удалил это pod, у меня опять Kubernetes ждёт удаления этого pod-а. И, казалось бы, если я сейчас сделаю get pods, то у меня ничего там не будет, на самом деле будет – создастся который pod, можно видеть, что у него hash отличается. Hash controller на отличается, hash pod-а отличается. Это происходит, потому что этот ресурс этого pod-а принадлежит deployment-у. То есть мы можем как раз посмотреть kubectl get deployments. Можем также точно его описать. И здесь мы увидим, да, какому Namespace-у он

принадлежит, какие у него label-ы, по какому selector-у он будет у нас, значит, поднимать приложение. Что будет происходить, когда у нас будет происходить update этого приложения. То есть если мы выкатываем новую версию – каким образом pod-ы должны restart-овать? RollingUpdates, значит, что pod-ы у нас будут выключаться 1 по одному, если у вас в реплике 3 pod-а – сначала перезагрузится первый с новой версией, потом второй с новой версией и так далее. И собственно, event-ы, которые у нас происходили с этой репликой.

То есть здесь мы видим, что это просто другой способ... Вернее, это не другой, это на мой взгляд правильный способ создавать ваше приложение. То есть если вы deploy-ите напрямую pod-ы – они у вас не вечные. То есть у вас что-то перезагрузилось в кластере, ваши pod-ы слетели, приложение не работает, клиенты недовольны. Если вы сделаете deployment, deployment у вас хранится в master node-ах. То есть deployment принадлежит, конечно, namespace-у, но конфигурация его хранится в master node-ах, и master node-ы сами с помощью своих контроллеров, вот они, контролируют ваши pod-ы. И, если у вас что-то случится с node-ой, ваши pod-ы будут перезагружены куда-то ещё.

Для того чтобы удалить все pod-ы, нам соответственно, необходимо в данном случае удалить deployment. То есть мы делаем delete deployment..., что мы там удаляем? Значит, liveness-http у нас deployment удалился, в данном случае быстро, потому что никакие pod-ы нам гасить не надо. И вот мы видим pod в статусе Terminating, то есть всё у нас погасилось вместе с deployment-ом, потому что deployment в менеджере. То есть controller видит, что deployment-а нет, а pod, который принадлежит deployment-у – есть. И говорит: «Выключись».

Значит, последнее, что мы рассмотрим в рамках наших probe. Получился у нас довольно большой отрезок, потому что это мимо probe ещё рассказал пару концептов важных нам достаточно. И здесь мы видим, значит, наш контейнер

горгоху, у которого есть какие-то port-ы, и у нас есть здесь readinessProbe, то есть та probe-a, которая нам показывает ready контейнер или нет, и LivenessProbe. Соответственно, они используют метод, механизм tcpSocket, который просто пытается открыть tcp-соединение низкоуровневое по определенному порту. Если ему это удастся – проба считается здоровой. Если ему это не удастся..., вернее, проба считается здоровой к работе. Если ему не удастся, то он выключается. Соответственно, можно увидеть, что ReadinessProbe у нас начинает работать раньше, чтобы мы могли посмотреть, что такое (HP3 29:42). LivenessProbe у нас начинает работать несколько позже. И давайте теперь запустим и посмотрим, как это работает.

Я применил команду apply, опять же в нашем default-ном namespace. Вот мы видим, у нас, значит, одна проба – она healthy, но не ready. То есть один контейнер. Кстати, можно ещё с помощью describe pod горгоху посмотреть на непосредственно список контейнеров. То есть если у вас их несколько, здесь они будут отражены. здесь их будет несколько. И соответственно, мы видим, что у нас всё ещё not ready. Теперь у нас заработал Readiness probe, она увидела, что всё готового и она у нас ready. То есть трафик можно лить на этот pod.

Соответственно, пара слов о ехес против, значит, http. Tcp-connection-ехес у нас требует открыть (HP3 30:45) контейнера, заходы внутрь контейнера. Для этого раньше во всяком случае из того, что я помню, Kubernetes создавал свой небольшой sрасе для того, чтобы получить свой контейнер маленький, для того чтобы получить результаты выполнения ехес.

Плюс, была проблема с тем, что ехес создавал зомби-контейнеры. И поэтому рекомендуется, если у вас есть какой-то процесс, который внутри машины болтается и не имеет никаких способов ответить по http или tcp, создать маленький сервис вместе с этим процессом, нарушая принцип ответственности. То

есть контейнер у вас сделает 2 вещи, но добавив этот процесс, потому что вы тем самым сэкономите ресурсы вашего контейнера. Это процесс может быть, например, написан на Golang-е. Вы можете получать список процессов. Смотрите, если ваш процесс запущен вместо того, чтобы каждый раз создавать пользователя, создавать (HP3 31:41) раз в 5 секунд и что-то туда добавлять. Разработчики Kubernetes-а работают над тем, чтобы exec был быстрее. Я думаю, что в какой-то момент эта цель будет неактуальна, но на данный момент мы всё ещё этим пользуемся.