

Если вы видите что-то необычное, просто сообщите мне.

Cluster Autoscaler: как он работает и решение частых проблем

Что такое Cluster Autoscaler

Kubernetes представляет несколько механизмов для масштабирования нагрузки. Три главные механизмы это : VPA, HPA, CA.

CA автоматически подбирает количество нод в кластере под требования. Когда число подов, которые находятся в очереди назначения или при отсутствии возможности назначить, показывает что ресурсов не хватает в кластере, CA добавляет новые ноды в кластер. Он так же может уменьшить количество нод если они не до конца используются долгое время.

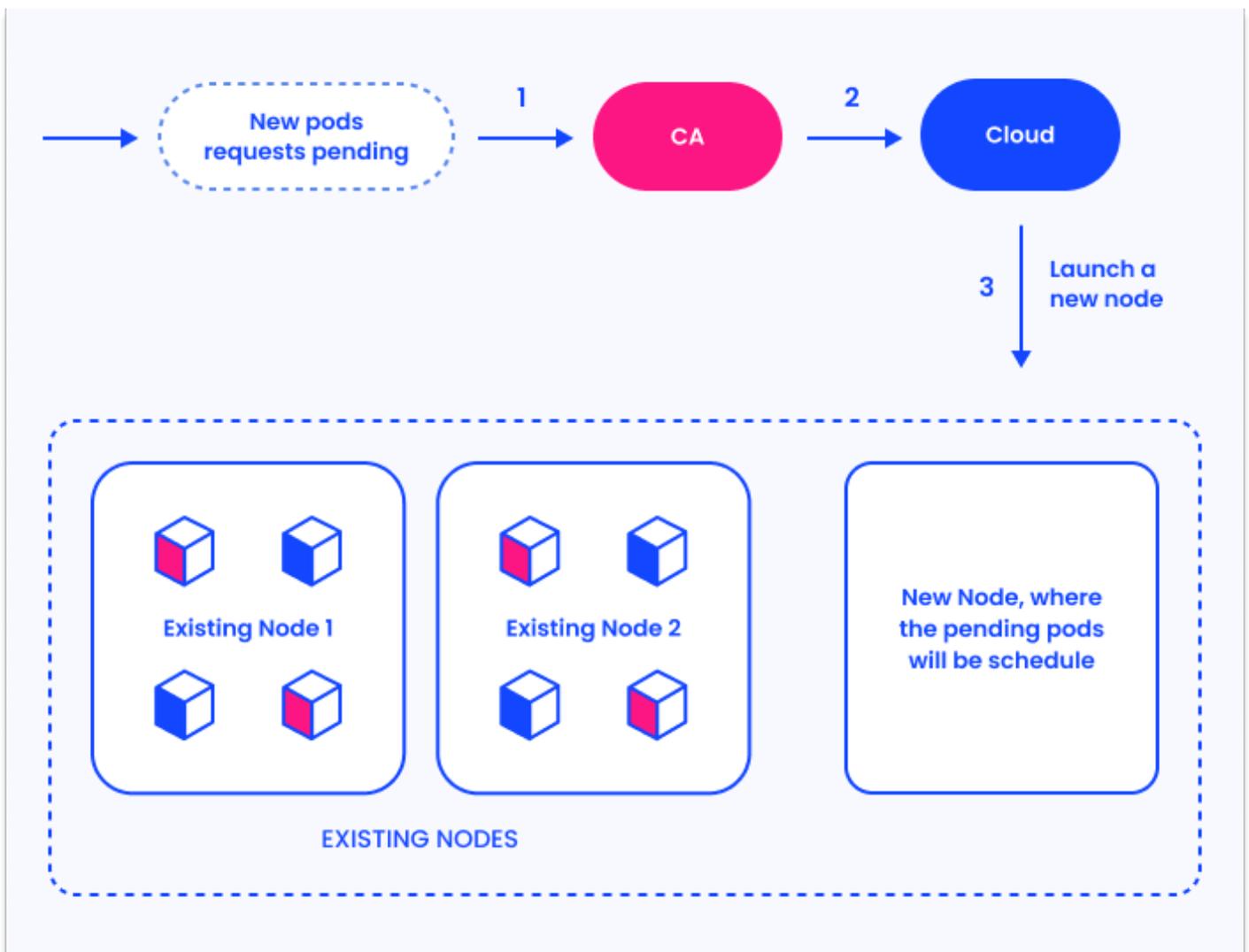
Обычно Cluster Autoscaler устанавливается как объект Deployment в кластере. Он работает только одной репликой и использует выборный механизм для того, чтобы быть уверенным, что он полностью доступен.

Как работает Cluster Autoscaler

Для простоты, мы объясним процесс Cluster Autoscaler в режиме масштабирования. Когда число назначенных подов в кластере увеличивается, указывая на недостаток ресурсов, CA автоматически запускает новые ноды.

Это проявляется в четырех шагах:

1. CA проверяет назначенные поды, время проверки 10 секунд(для настройки можно указать флаг `--scan-interval`)
2. Если есть назначенные поды, CA запускаем новые ноды для масштабирования кластера, в рамках конфигурации кластера. CA встраивается в облачную платформу, например AWS или Azure, используя их возможности масштабирования для того, чтобы можно было управлять vm.
3. K8s регистрирует новые vm в качестве нод, позволяя K8s запускать поды на свежих ресурсах.
4. K8s планировщик запускает назначенные поды на новые ноды. ew nodes.



Обнаружение проблем с Cluster Autoscaler

CA полезный механизм, но он может работать не так, как ожидает администратор. Вот первые шаги, чтобы найти проблему с CA.

Логирование на нодах

План управления K8s создает логи активности CA по следующему пути: `/var/log/cluster-autoscaler.log`

События

`kube-system/cluster-autoscaler-status` ConfigMap производят следующие события:

- `ScaledUpGroup` - это событие говорит, CA увеличивает размер группы нод(предоставляется прошлый и текущий размеры)
- `ScaleDownEmpty` - это событие означает, что CA убирает ноду, которая не имеет подов(системные поды при этом не рассматриваются)
- `ScaleDown` - это событие создается, когда CA убирает ноду, которая имеет запущенные поды. Событие содержит имена всех подов, которые будут переназначены на другие ноды в результате действия.

События нод

- `TriggeredScaleUp` - это событие говорит, что CA увеличивает кластер, так как появились поды в очереди.

- NotTriggerScaleUp - событие говорит, что CA не может увеличить количество нод в группе.
- ScaleDown - это событие значит, что CA пробует перенести поды с ноды, чтобы затем освободить ноду и удалить из кластера.

Cluster Autoscaler: работа с определенными ошибками

Предлагаем несколько определенных ситуаций, которые могут повяится при работе CA и возможные решения этих проблем.

Эта инструкция позволит выяснить простые ошибки работы CA, но для более сложных проблем, включающие множество двигающихся частей в кластере, возможно придется автоматизировать инструментарий решения проблем.

Ноды с недостаточной нагрузкой не удалятся из кластера.

Вот причины по которы CA не может уменьшить количество нод, и что можно с этим сделать.

Причин а пробле мы	Что мо жн о сде лат ь
---------------------------------------	--

В опи сан ии под а ест ь ука зан ие, что его нел ьзя пер ене сти на дру гую нод у.	Про вер ьте отс утсв ую щий Conf igM ар и соз дай те его, или исп оль зуй те дру гой.
Гру ппа нод уже име ет мин има льн ое зна чен ие.	Сок рат ите мин има льн ое зна чен ие в нас тро йка х СА.

Нод а име ет дир ект иву "sca le- dow n disa bled ".	Убе рит е дир ект иву с нод ы.
---	--

СА
ожи
дае
т
вре
мен
и
сог
лас
но
одн
ому
из
ука
зан
ных
сле
дую
щих
фла
гов:
--
scale
-
down
-
unne
eded
-
time
--
scale
-
down
-
delay
-
after-
add ,
--
scale
-
down
-
delay
-
after-
failur
e ,
--
scale
-
down
-
delay

Сок
рат
ите
вре
мя
ука
зан
ное
во
соо
тве
тсву
ющ
ем
фла
ге,
или
дож
дис
ь
ука
зан
ног
о
вре
мен
и.

Неу	Под
дач	ожд
на	ите
япо	5
пыт	мин
ка	ут и
уда	про
лен	вер
ия	ьте
нод	реш
ы(С	ила
А	сь
буд	ли
ет	про
жда	бле
ть 5	ма. .
мин	
ут	
пее	
ред	
пов	
тор	
ной	
поп	
ытк	
ой)	

Поды в состоянии pending, но НОВЫЕ НОДЫ НЕ СОЗДАЮТСЯ.

Ниже приведены причины почему CA может не увеличивать количество нод в кластере, и что с этим можно сделать.

Причина	Что можно сделать
Создаваемый подимет запросы превыщающие характеристики узлов.	Дать возможность иметь САдобавлять болышие ноды, или сократить требования ресурсов для пода.

Причина	Что можно сделать
Все подходы к группе имеют максимум разрешенное значение.	Увеличьте максимальное значение неограниченной группы.

Причина	Что можно сделать
Новый поднезначается, но новые ноды.	Изменить тип описания пода, чтобы бы предоставит возможность поду назначаться на определенной группенод.

NoVolumeZoneConflict error— показывает, что StatefulSet требует запуск в той же зоне что и PersistentVolume(PV), но эта зона уже имеет доступный лимит .| начиная с Kubernetes 1.13, вы можете разделить группу нод на зоны и использовать флаг --balance-similar-node-groups

Cluster Autoscaler прекратил работать

Если CA не работает, пройдитесь по следующим шагам, чтобы понять проблему.

1. Проверьте что CA запущен. Это можно проверить по последнему событию, которое генерируется в `kube-system/cluster-autoscaler-status` ConfigMap. Оно не должно превышать 3 минуты.
2. Проверьте если кластер и группы нод находятся в здоровом состоянии, это так же можно найти в configMap
3. Проверьте наличие неготовых нод - если какие-то ноды оказываются `unready` проверьте число `resourceUnready`. Если какие-то ноды помечены, проблема, скорее всего, в том, что не было установлено необходимое ПО.
4. Если состояние CA и кластера здоровое, проверьте:
 - Control plane CA logs - могут указать на проблему, которая может не давать масштабировать кластер.
 - CA события для pod объекта — может дать понимание почему CA не переназначает поды.
 - Cloud provider resources quota— если есть неудачные попытки добавить ноду, проблема может быть в квотах ресурсов у провайдера.
 - Networking issues— если провайдер пытается создать ноду, но она не подключается к кластеру, это может говорить о проблеме с сетью.

Revision #6

Created 23 December 2022 07:49:40 by gasick

Updated 16 April 2023 19:36:18 by gasick