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

Cluster Autoscaler: как он работает и решение частых проблем

Что такое Cluster Autosc

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 остутствии 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.

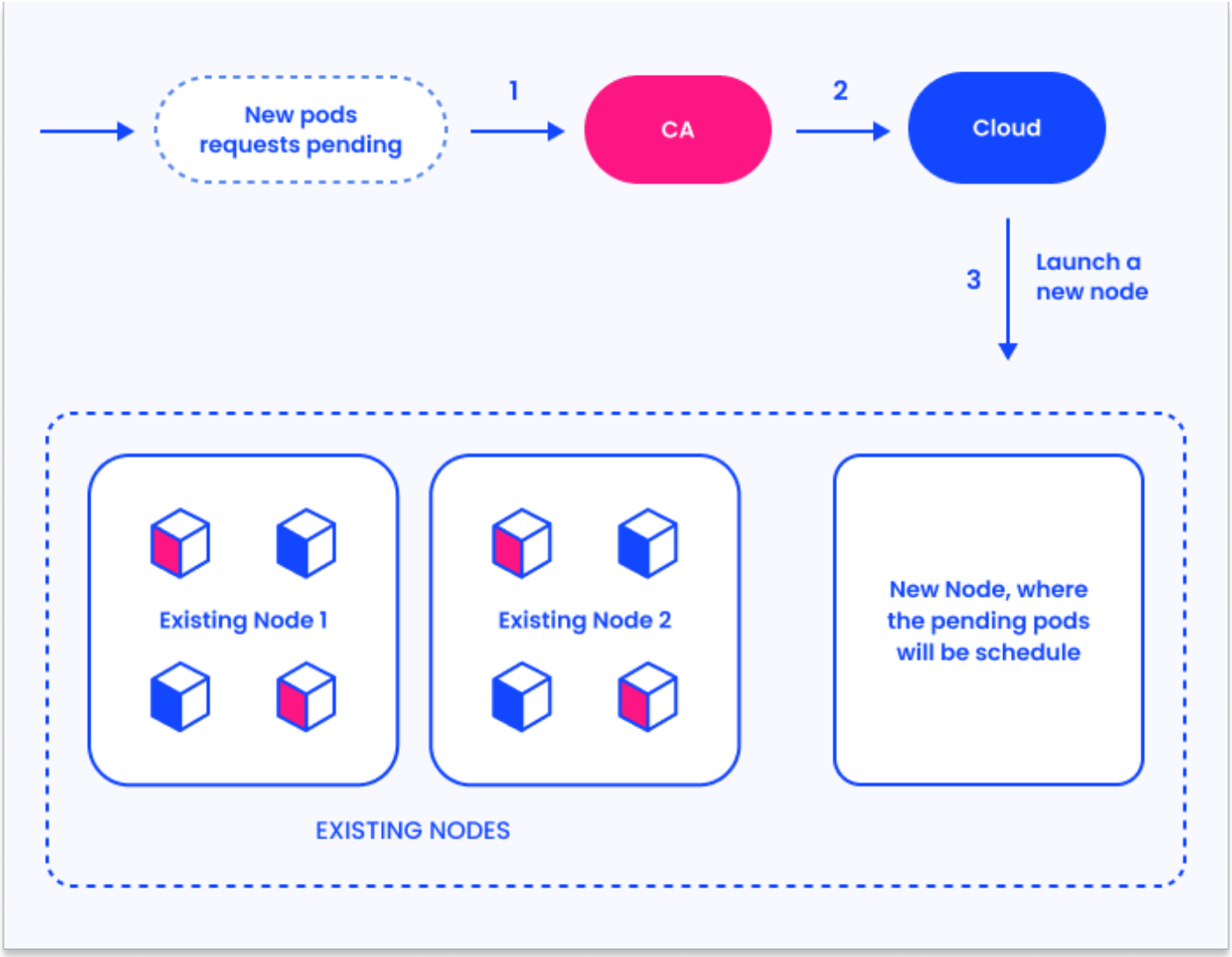
Обычно Cluster Autoscaler устанавливается как объект Deployment в кластере. Он работает только одной репликой и использует выборный механизм для того. чтобы быть уверенным, что он полностью доступен.

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 первшые 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 означение, что CA убирает ноду, которая не имеет подов(системные поды при этом не рассматриваюца)
- ScaleDown - это событие создается, когда CA убирает ноду, которая имеет запущенные поды. Событие содержит имена всех подов, которые будет перезазначены 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 нагрузой не удалются из кластера.

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

Причина проблемы	Что можно сделать
В описании пода есть указание, что его нельзя перенести на другую ноду.	Проверьте отсутсвующий ConfigMap и создайте его, или используйте другой.
Группа нод уже имеет минимальное значение.	Сократите минимальное значение в настройках CA.
Нода имеет директиву “scale-down disabled”.	Уберите директиву с ноды.

Причина проблемы	Что можно сделать
CA ожидает времени согласно одному из указанных следующих флагов: `--scale-down-unneeded-time` `--scale-down-delay-after-add`, `--scale-down-delay-after-failure`, `--scale-down-delay-after-delete`, `--scan-interval`	Сократите время указанное во соответсвующем флаге, или дождись указанного времени.
Неудачна япопытка удаления ноды(CA будет ждать 5 минут пееред повторной попыткой)	Подождите 5 минут и проверьте решилась ли проблема. .

Поды в состоянии penind, 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 под имеет запросы превыщающие характеристики ноды.	Дать возможность CA добавлять большие ноды, или сократить требования ресурсов для пода.
Все подходящие группы нод имеют максимально разрешенное значение.	Увеличьте максимальное значение необходимой группы.
Новый под не назначается но новые ноды.	Изменити описание пода, чтобы предоставить возможность поду назначаться на определенной группе нод.

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 для балансировки.|

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 resoureceUnready. Если какие-то ноды помечены, проблема, скорей всего, в том, что не было установленно 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 2022-12-23 07:49:40 UTC by gasick
Updated 2023-04-16 19:36:18 UTC by gasick