OpenShift Dedicated 4

클러스터 관리

OpenShift Dedicated 클러스터 구성

Last Updated: 2024-09-04

OpenShift Dedicated 4 클러스터 관리

OpenShift Dedicated 클러스터 구성

법적 공지

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at
http://creativecommons.org/licenses/by-sa/3.0/
. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux ® is the registered trademark of Linus Torvalds in the United States and other countries.

Java ® is a registered trademark of Oracle and/or its affiliates.

XFS ® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL ® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js ® is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack ® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

초록

이 문서에서는 OpenShift Dedicated 클러스터 구성에 대한 정보를 제공합니다.

차례

1장. 클러스터 알림 **3**
1.1. 추가 리소스 3
1.2. 클러스터 알림에서 예상되는 사항 3
1.3. RED HAT HYBRID CLOUD CONSOLE을 사용하여 클러스터 알림 보기 5
1.4. 클러스터 알림 이메일 6
1.5. 문제 해결 7

2장. 프라이빗 연결 구성 **8**
2.1. AWS의 개인 연결 구성 8
2.2. 프라이빗 클러스터 설정 12

3장. 클러스터 자동 스케일링 **15**
3.1. 클러스터 자동 스케일러 정보 15
3.2. OPENSHIFT CLUSTER MANAGER를 사용하여 클러스터 생성 중 자동 스케일링 활성화 16
3.3. OPENSHIFT CLUSTER MANAGER를 사용하여 클러스터 생성 후 자동 스케일링 활성화 17
3.4. OPENSHIFT CLUSTER MANAGER를 사용한 클러스터 자동 스케일링 설정 17

4장. 노드 **21**
4.1. 머신 풀 정보 21
4.2. 컴퓨팅 노드 관리 22
4.3. 클러스터의 노드 자동 스케일링 정보 27

1장. 클러스터 알림

클러스터 알림은 클러스터의 상태, 상태 또는 성능에 대한 메시지입니다.

클러스터 알림은 Red Hat site Reliability Engineering(SRE)이 관리형 클러스터의 상태에 대해 귀하와 통신하는 기본 방법입니다. SRE는 클러스터 알림을 사용하여 클러스터 문제를 해결하거나 방지하기 위해 작업을 수행하도록 요청할 수도 있습니다.

클러스터 소유자 및 관리자는 클러스터가 정상 상태로 유지되고 지원되는지 확인하기 위해 클러스터 알림을 정기적으로 검토하고 조치를 취해야 합니다.

클러스터의 클러스터 **기록 탭에서 Red Hat Hybrid Cloud Console에서 클러스터**알림을 볼 수 있습니다. 기본적으로 클러스터 소유자만 이메일로 클러스터 알림을 수신합니다. 다른 사용자가 클러스터 알림 이메일을 수신해야 하는 경우 각 사용자를 클러스터에 대한 알림 연락처로 추가합니다.

1.1. 추가 리소스

- 고객 담당: 클러스터 알림 검토 및 작업
- 클러스터 알림 이메일
- 문제 해결: 클러스터 알림

1.2. 클러스터 알림에서 예상되는 사항

클러스터 관리자는 클러스터의 상태 및 관리 요구를 효과적으로 이해하기 위해 클러스터 알림과 유형 및 심각도 수준을 전송하는 시기와 이유를 알고 있어야 합니다.

1.2.1. 클러스터 알림 정책

클러스터 알림은 클러스터의 상태와 영향을 미치는 높은 영향을 미치는 이벤트에 대한 정보를 유지하도록 설계되었습니다.

대부분의 클러스터 알림은 자동으로 생성되고 전송되어 즉시 문제에 대한 정보 또는 클러스터 상태에 대한 중요한 변경 사항을 확인할 수 있습니다.

특정 상황에서 Red Hat 사이트 안정성 엔지니어링(SRE)은 클러스터 알림을 생성하고 전송하여 복잡한 문제에 대한 추가 컨텍스트 및 지침을 제공합니다.

영향을 받지 않는 이벤트, 위험이 낮은 보안 업데이트, 일상적인 운영 및 유지 관리 또는 SRE가 신속하게 해결하는 일시적인 문제에 대해서는 클러스터 알림이 전송되지 않습니다.

Red Hat 서비스는 다음과 같은 경우 자동으로 알림을 보냅니다.

- 원격 상태 모니터링 또는 환경 확인 검사에서는 작업자 노드에 디스크 공간이 부족한 경우와 같이 클러스터에서 문제를 감지합니다.
- 예를 들어 예정된 유지 관리 또는 업그레이드가 시작되는 경우 심각한 클러스터 라이프 사이클 이벤트가 발생하거나 클러스터 작업이 이벤트의 영향을 받지만 고객의 개입은 필요하지 않습니다.
- 예를 들어 클러스터 소유권 또는 관리 제어가 한 사용자에서 다른 사용자로 전송되는 경우와 같이 중요한 클러스터 관리 변경이 발생합니다.
- 예를 들어 Red Hat이 클러스터에서 서브스크립션 조건 또는 기능을 업데이트할 때 클러스터 서브스크립션이 변경 또는 업데이트됩니다.

SRE는 다음과 같은 경우 알림을 생성하고 보냅니다.

- 사고로 인해 클러스터의 가용성 또는 성능에 영향을 미치는 성능 저하 또는 중단이 발생합니다(예: 클라우드 공급자의 경우 지역 중단). SRE는 사고 해결 진행 상황을 알려주기 위해 후속 알림을 보냅니다.
- 클러스터에서 보안 취약점, 보안 위반 또는 비정상적인 활동이 감지됩니다.
- Red Hat은 변경 사항이 생성 중이거나 클러스터 불안정성을 초래할 수 있음을 감지합니다.
- Red Hat은 워크로드가 클러스터에서 성능 저하 또는 불안정성을 초래하고 있음을 감지합니다.

1.2.2. 클러스터 알림 심각도 수준

각 클러스터 알림에는 비즈니스에 가장 큰 영향을 미치는 알림을 식별하는 데 도움이 되는 관련 심각도 수준이 있습니다. 클러스터의 클러스터 **기록 탭에서 Red Hat Hybrid Cloud Console의 심각도 수준에 따라 클러스터** 알림을 필터링할 수 있습니다.

Red Hat은 가장 심각한 클러스터 알림에 다음과 같은 심각도 수준을 사용합니다.

심각

즉각적인 조치가 필요합니다. 서비스 또는 클러스터의 하나 이상의 주요 기능이 작동하지 않거나 곧 작동하지 않습니다. 중요한 경고는 직원에게 호출하고 일반 워크플로우를 중단할 수 있을 만큼 중요합니다.

major

즉각적인 조치가 강력히 권장됩니다. 클러스터의 하나 이상의 주요 기능이 곧 작동하지 않습니다. 주요 문제는 적시에 해결되지 않는 경우 중요한 문제로 이어질 수 있습니다.

경고

가능한 한 빨리 조치가 필요합니다. 클러스터의 하나 이상의 주요 기능이 최적으로 작동하지 않고 추가로 성능이 저하될 수 있지만 클러스터 작동에 즉각적인 위험이 발생하지 않습니다.

정보

작업이 필요하지 않습니다. 이 심각도는 처리해야 하는 문제를 설명하지 않으며 의미 있거나 중요한 라이프 사이클, 서비스 또는 클러스터 이벤트에 대한 중요한 정보만 설명합니다.

Debug

작업이 필요하지 않습니다. 디버그 알림은 예기치 않은 동작을 디버깅하는 데 도움이 되도록 덜 중요한 라이프사이클, 서비스 또는 클러스터 이벤트에 대한 낮은 수준의 정보를 제공합니다.

1.2.3. 클러스터 알림 유형

각 클러스터 알림에는 역할 및 역할과 관련된 알림을 식별하는 데 도움이 되는 관련 알림 유형이 있습니다. 클러스터의 클러스터 **기록 탭에서 Red Hat Hybrid Cloud Console에서 이러한 유형에 따라 클러스터**알림을 필터링할 수 있습니다.

Red Hat은 다음 알림 유형을 사용하여 알림을 표시합니다.

용량 관리

노드 풀, 시스템 풀, 컴퓨팅 복제본 또는 할당량(로드 밸런서, 스토리지 등) 업데이트, 생성 또는 삭제와 관련된 이벤트 알림.

클러스터 액세스

STS 인증 정보가 만료되었거나 AWS 역할에 구성 문제가 있는 경우 또는 ID 공급자를 추가하거나 제거할 때와 같은 그룹, 역할 또는 ID 공급자 추가 또는 ID 공급자와 관련된 이벤트에 대한 알림입니다.

클러스터 애드온

애드온의 애드온 관리 또는 업그레이드 유지 관리와 관련된 이벤트의 알림(예: 애드온을 설치, 업그레이드 또는 제거)하거나 미해결 요구 사항으로 인해 설치할 수 없습니다.

클러스터 구성

클러스터 튜닝 이벤트, 워크로드 모니터링 및 진행 중인 검사에 대한 알림입니다.

클러스터 라이프사이클

클러스터 또는 클러스터 리소스 생성, 삭제 및 등록에 대한 알림 또는 클러스터 또는 리소스 상태 변경(예: 준비 또는 누락).

클러스터 네트워킹

HTTP/S 프록시, 라우터 및 수신 상태를 포함한 클러스터 네트워킹과 관련된 알림.

클러스터 소유권

클러스터 소유권과 관련된 알림은 한 사용자의 다른 사용자로 이동합니다.

클러스터 스케일링

노드 풀, 머신 풀, 컴퓨팅 복제본 또는 할당량 업데이트, 생성 또는 삭제와 관련된 알림입니다.

클러스터 보안

예를 들어 클러스터 보안과 관련된 이벤트(예: 실패한 액세스 시도 횟수, 신뢰 번들에 대한 업데이트 또는 보안에 영향을 미치는 소프트웨어 업데이트 등)

클러스터 서브스크립션

클러스터 만료, 평가판 클러스터 알림 또는 무료에서 유료로 전환.

클러스터 업데이트

업그레이드 유지 관리 또는 활성화와 같은 업그레이드 관련 항목입니다.

고객 지원

지원 케이스 상태에 대한 업데이트

일반 알림

기본 알림 유형입니다. 이는 더 구체적인 카테고리가 없는 알림에만 사용됩니다.

1.3. RED HAT HYBRID CLOUD CONSOLE을 사용하여 클러스터 알림 보기

클러스터 알림은 클러스터 상태에 대한 중요한 정보를 제공합니다. Red Hat Hybrid Cloud Console의 클러스터 **기록 탭에서** 클러스터로 전송된 알림을 볼 수 있습니다.

사전 요구 사항

- 하이브리드 클라우드 콘솔에 로그인되어 있습니다.

절차

1. 하이브리드 클라우드 콘솔의 클러스터 페이지로 이동합니다.
2. 클러스터 이름을 클릭하여 클러스터 세부 정보 페이지로 이동합니다.
3. **클러스터 기록** 탭을 클릭합니다.
   클러스터 알림은 클러스터 기록 제목 아래에 표시됩니다.
4. 선택 사항: 관련 클러스터 알림에 대해 필터링
   필터 컨트롤을 사용하여 사용자의 전문 영역에 집중하거나 중요한 문제를 해결할 수 있도록 사용자와 관련이 없는 클러스터 알림을 숨깁니다. 알림 설명의 텍스트, 심각도 수준, 알림 유형, 알림이 수신될 때 및 알림을 트리거한 시스템 또는 사람을 기반으로 알림을 필터링할 수 있습니다.

1.4. 클러스터 알림 이메일

기본적으로 클러스터 알림이 클러스터로 전송되면 클러스터 소유자의 이메일도 전송됩니다. 적절한 모든 사용자가 클러스터 상태에 대한 정보를 유지할 수 있도록 알림 이메일에 대한 추가 수신자를 구성할 수 있습니다.

1.4.1. 클러스터에 알림 연락처 추가

알림 연락처는 클러스터 알림이 클러스터로 전송될 때 이메일을 수신합니다. 기본적으로 클러스터 소유자만 클러스터 알림 이메일을 수신합니다. 클러스터 지원 설정에서 다른 클러스터 사용자를 추가 알림 연락처로 구성할 수 있습니다.

사전 요구 사항

- 클러스터가 Red Hat Hybrid Cloud Console에 배포 및 등록되었습니다.
- 하이브리드 클라우드 콘솔에 로그인되어 있습니다.
- 의도된 알림 수신자의 클러스터에 사용자 계정이 있습니다.

절차

1. 하이브리드 클라우드 콘솔의 클러스터 페이지로 이동합니다.
2. 클러스터 이름을 클릭하여 클러스터 세부 정보 페이지로 이동합니다.
3. **지원** 탭을 클릭합니다.
4. **지원** 탭에서 **알림 연락처** 섹션을 찾습니다.
5. **알림 연락처 추가를** 클릭합니다.
6. **Red Hat 사용자 이름 또는 이메일**필드에 이메일 주소 또는 새 수신자의 사용자 이름을 입력합니다.
7. **연락처 추가를** 클릭합니다.

검증 단계

- "알림 연락처가 성공적으로 추가됨" 메시지가 표시됩니다.

1.4.2. 클러스터에서 알림 연락처 제거

알림 연락처는 클러스터 알림이 클러스터로 전송될 때 이메일을 수신합니다.

클러스터 지원 설정에서 알림 연락처를 제거하여 알림 이메일을 수신하지 못하도록 할 수 있습니다.

사전 요구 사항

- 클러스터가 Red Hat Hybrid Cloud Console에 배포 및 등록되었습니다.
- 하이브리드 클라우드 콘솔에 로그인되어 있습니다.

절차

1. 하이브리드 클라우드 콘솔의 클러스터 페이지로 이동합니다.
2. 클러스터 이름을 클릭하여 클러스터 세부 정보 페이지로 이동합니다.
3. **지원** 탭을 클릭합니다.
4. **지원** 탭에서 **알림 연락처** 섹션을 찾습니다.
5. 제거하려는 수신자 옆에 있는 옵션 메뉴(t)를 클릭합니다.
6. **삭제를 클릭합니다.**

검증 단계

- **"알림 연락처가 성공적으로 삭제됨" 메시지가 표시됩니다.**

1.5. 문제 해결

클러스터 알림 이메일을 수신하지 못하는 경우

- **@redhat.com 주소에서 전송된 이메일이 이메일 수신함에서 필터링되지 않았는지 확인합니다.**
- **올바른 이메일 주소가 클러스터의 알림 연락처로 나열되어 있는지 확인합니다.**
- **클러스터 소유자 또는 관리자에게 알림 연락처로 추가하도록 요청합니다. 클러스터 알림 이메일.**

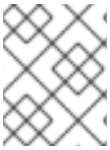
클러스터가 알림을 수신하지 않는 경우

- **클러스터가 api.openshift.com 에서 리소스에 액세스할 수 있는지 확인합니다.**
- **문서화된 사전 요구 사항에 따라 방화벽이 구성되었는지 확인합니다. AWS 방화벽 사전 요구 사항**

2장. 프라이빗 연결 구성

2.1. AWS의 개인 연결 구성

2.1.1. AWS 클라우드 인프라 액세스 이해



참고

AWS 클라우드 인프라 액세스는 CCS 클러스터가 계정에 배포되므로 클러스터를 생성할 때 선택한 CCS(Customer Cloud Subscription) 인프라 유형에는 적용되지 않습니다.

AWS(Amazon Web Services) 인프라 액세스를 통해 고객 포털 조직 관리자와 클러스터 소유자가 AWS IAM(Identity and Access Management) 사용자가 OpenShift Dedicated 클러스터에 대한 AWS 관리 콘솔에 액세스할 수 있습니다. 고객 AWS 사용자에 대해 AWS 액세스 권한을 부여할 수 있으며 OpenShift Dedicated 환경의 요구에 맞게 프라이빗 클러스터 액세스를 구현할 수 있습니다.

1. OpenShift Dedicated 클러스터에 대한 AWS 인프라 액세스 구성을 시작하십시오. AWS 사용자 및 계정을 생성하고 해당 사용자에게 OpenShift Dedicated AWS 계정에 대한 액세스 권한을 제공합니다.
2. OpenShift Dedicated AWS 계정에 액세스한 후 다음 방법 중 하나 이상을 사용하여 클러스터에 대한 개인 연결을 설정합니다.
   - AWS VPC 피어링 구성: VPC 피어링을 활성화하여 두 개의 개인 IP 주소 간에 네트워크 트래픽을 라우팅합니다.
   - AWS VPN 구성: 가상 사설 네트워크를 설정하여 프라이빗 네트워크를 Amazon Virtual Private Cloud에 안전하게 연결합니다.
   - AWS Direct Connect 구성: AWS Direct Connect를 구성하여 프라이빗 네트워크와 AWS Direct Connect 위치 간의 전용 네트워크 연결을 설정합니다.

클라우드 인프라 액세스를 구성한 후 프라이빗 클러스터 구성에 대해 자세히 알아보십시오.

2.1.2. AWS 인프라 액세스 구성

AWS(Amazon Web Services) 인프라 액세스를 통해 고객 포털 조직 관리자와 클러스터 소유자가 AWS IAM(Identity and Access Management) 사용자가 OpenShift Dedicated 클러스터에 대한 AWS 관리 콘솔에 액세스할 수 있도록 할 수 있습니다. 관리자는 네트워크 관리 또는 읽기 전용 액세스 옵션 중에서 선택할 수 있습니다.

사전 요구 사항

- IAM 권한이 있는 AWS 계정.

절차

1. AWS 계정에 로그인합니다. 필요한 경우 AWS 문서에 따라 새 AWS 계정을 생성할 수 있습니다.
2. AWS 계정 내에서 **STS:AllowAssumeRole** 권한이 있는 IAM 사용자를 생성합니다.
   a. AWS 관리 콘솔의 IAM 대시보드를 엽니다.
   b. Policies 섹션에서 정책 생성을 클릭합니다.

c. JSON 탭을 선택하고 기존 텍스트를 다음과 같이 변경합니다.

```
{
 "Version": "2012-10-17",
 "Statement": [
    {
       "Effect": "Allow",
       "Action": "sts:AssumeRole",
       "Resource": "*"
    }
 ]
}
```

d. Next:Tags 를 클릭합니다.

e. 선택 사항: 태그를 추가합니다. Next:Review를 클릭합니다.

f. 적절한 이름과 설명을 입력한 다음 정책 만들기 를 클릭합니다.

g. 사용자 섹션에서 사용자 추가 를 클릭합니다.

h. 적절한 사용자 이름을 제공합니다.

i. AWS 액세스 유형으로 AWS Management Console 액세스를 선택합니다.

j. 조직에 필요한 암호 요구 사항을 조정한 다음 Next:Permissions 를 클릭합니다.

k. 기존 정책 직접 연결 옵션을 클릭합니다. 이전 단계에서 생성한 정책을 검색하고 확인합니다.

참고

권한 경계를 설정하지 않는 것이 좋습니다.

l. Next: Tags 를 클릭한 다음Next: Review 를 클릭합니다. 구성이 올바른지 확인합니다.

m. 사용자 만들기를 클릭하면 성공 페이지가 표시됩니다.

n. IAM 사용자의 ARM(Amazon Resource Name)을 수집합니다. ARN에는 **arn:aws:iam::000111222333:user/username** 형식이 있습니다. 닫기를 클릭합니다.

3. 브라우저에서 OpenShift Cluster Manager 를 열고 AWS 인프라 액세스를 허용할 클러스터를 선택합니다.

4. 액세스 제어 탭을 선택하고AWS Infrastructure Access 섹션으로 스크롤합니다.

5. AWS IAM ARN을 붙여넣고네트워크 관리 또는 읽기 전용 권한을 선택한 다음역할 부여 를 클릭합니다.

6. AWS OSD 콘솔 URL을 클립보드에 복사합니다.

7. 계정 ID 또는 별칭, IAM 사용자 이름 및 암호를 사용하여 AWS 계정에 로그인합니다.

8. 새 브라우저 탭에서 AWS Switch Role 페이지로 라우팅하는 데 사용할 AWS OSD 콘솔 URL을 붙여넣습니다.

9. 계정 번호와 역할이 이미 입력될 것입니다. 필요한 경우 표시 이름을 선택한 다음 역할 전환을 클릭합니다.

검증

- 이제 최근 방문한 서비스 아래 VPC 가 표시됩니다.

2.1.3. AWS VPC 피어링 구성

VPC(Virtual Private Cloud) 피어링 연결은 프라이빗 IPv4 주소 또는 IPv6 주소를 사용하여 트래픽을 라우팅할 수 있는 두 VPC 간 네트워킹 연결입니다. 다른 AWS VPC 네트워크와 피어링하도록 OpenShift Dedicated 클러스터가 포함된 AWS(Amazon Web Services) VPC를 구성할 수 있습니다.

주의

클러스터를 제거하기 전에 클러스터의 VPC 피어링 연결을 제거해야 합니다. 이렇게 하지 않으면 클러스터가 제거 프로세스를 완료하지 못할 수 있습니다.

AWS는 중국을 제외한 모든 상용 리전 간에 리전 간 VPC 피어링을지원합니다.

사전 요구 사항

- 피어 요청을 시작하는 데 필요한 고객 VPC에 대한 다음 정보를 수집합니다.
  - 고객 AWS 계정 번호
  - Customer VPC ID
  - 고객 VPC 리전
  - 고객 VPC CIDR
- OpenShift Dedicated Cluster VPC에서 사용하는 CIDR 블록을 확인합니다. 고객 VPC의 CIDR 블록과 겹치거나 일치하는 경우 이 두 VPC 간의 피어링을 수행할 수 없습니다. 자세한 내용은 Amazon VPC Unsupported VPC 피어링 구성 문서를 참조하십시오. CIDR 블록이 겹치지 않으면 절차를 진행할 수 있습니다.

절차

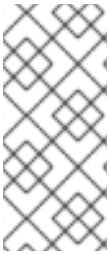
1. VPC 피어링 요청을 시작합니다.
2. VPC 피어링 요청을 수락합니다.
3. VPC 피어링 연결의 경로 테이블을 업데이트합니다.

추가 리소스

- 자세한 내용은 AWS VPC 가이드를 참조하십시오.

2.1.4. AWS VPN 구성

고객의 온사이트 하드웨어 VPN(Virtual Private Network) 장치를 사용하도록 AWS(Amazon Web Services) OpenShift Dedicated 클러스터를 구성할 수 있습니다. 기본적으로 AWS VPC(Virtual Private Cloud)로 시작하는 인스턴스는 자체(remote) 네트워크와 통신할 수 없습니다. AWS Site-to-Site VPN 연결을 생성하고 연결을 통해 트래픽을 전달하도록 라우팅을 구성하여 VPC에서 원격 네트워크에 대한 액세스를 활성화할 수 있습니다.



참고

AWS VPN은 현재 VPN 트래픽에 NAT를 적용하는 관리형 옵션을 제공하지 않습니다. 자세한 내용은 AWS Knowledge Center 에서 참조하십시오.

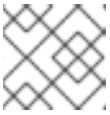
개인 연결을 통해 모든 트래픽(예: **0.0.0.0/0**)을 라우팅하는 것은 지원되지 않습니다. 이를 위해서는 SRE 관리 트래픽을 비활성화하는 인터넷 게이트웨이를 삭제해야 합니다.

사전 요구 사항

- 하드웨어 VPN 게이트웨이 장치 모델 및 소프트웨어 버전 (예: Cisco ASA 버전 8.3을 실행함) AWS 문서를 참조하여 AWS 에서 게이트웨이 장치를 지원하는지 확인합니다.
- VPN 게이트웨이 장치의 공용 고정 IP 주소입니다.
- BGP 또는 정적 라우팅: BGP인 경우 ASN이 필요합니다. 정적 라우팅인 경우 하나 이상의 정적 경로를 구성해야 합니다.
- 선택 사항: 연결할 수 있는 서비스의 IP 및 포트/프로토그래프로 VPN 연결을 테스트합니다.

절차

1. 고객 게이트웨이를 생성 하여 VPN 연결을 구성합니다.
2. 원하는 VPC에 가상 프라이빗 게이트웨이가 아직 연결되어 있지 않은 경우 가상 프라이빗 게이트웨이를 생성하고 연결합니다.
3. 라우팅을 구성하고 VPN 경로 전파를 활성화합니다.
4. 보안 그룹을 업데이트하십시오.
5. 사이트 간 VPN 연결을 설정합니다.



참고

구성에 추가해야 하는 VPC 서브넷 정보를 원격 네트워크로 기록해 둡니다.

추가 리소스

- 자세한 내용은 AWS VPN 가이드를 참조하십시오.

2.1.5. AWS Direct Connect 구성

AWS(Amazon Web Services) Direct Connect에는 동일한 계정의 원격 VPC(Virtual Private Cloud)에 액세스하기 위해 가상 게이트웨이(VWW) 또는 전환 게이트웨이에 차례로 연결된 다이렉트 연결 게이트웨이(VXGateway)에 연결된 호스트형 가상 인터페이스(VIF)가 필요합니다.

기존 DXGateway가 없는 경우 일반적인 프로세스는 AWS 계정에 DXGateway 및 VGW를 생성하여 호스팅된 VIF를 생성하는 것입니다.

하나 이상의 기존 VGW에 연결된 기존 DXGateway가 있는 경우 프로세스에는 DXGateway 소유자로 association Proposal을 전송하는 AWS 계정이 포함됩니다. DXGateway 소유자는 제안된 CIDR이 연결된 다른 VGW와 충돌하지 않는지 확인해야 합니다.

사전 요구 사항

- OpenShift Dedicated VPC의 CIDR 범위가 연결된 다른 VGW와 충돌하지 않는지 확인합니다.
- 다음 정보를 수집합니다.
  - Direct Connect Gateway ID입니다.
  - 가상 인터페이스와 연결된 AWS 계정 ID입니다.
  - DXGateway에 할당된 BGP ASN입니다. 선택사항: Amazon 기본 ASN도 사용할 수 있습니다.

절차

1. VIF를 생성하거나 기존 VIF를 보고 생성해야 하는 직접 연결 유형을 결정합니다.
2. 게이트웨이를 만듭니다.
   a. Direct Connect VIF 유형이 Private인 경우 가상 개인 게이트웨이를 만듭니다.
   b. Direct Connect VIF가 공용인 경우 Direct Connect 게이트웨이를 만듭니다.
3. 사용하려는 기존 게이트웨이가 있는 경우 연관 제안을 작성하고 승인을 위해 DXGateway 관리자에게 제안을 보냅니다.

주의

기존 DXGateway에 연결할 때 비용에 대한 책임이 있습니다.

추가 리소스

- 자세한 내용은 AWS Direct Connect 가이드를 참조하십시오.

2.2. 프라이빗 클러스터 설정

내부 애플리케이션을 회사 네트워크 내에서 호스팅할 수 있도록 OpenShift Dedicated 클러스터를 프라이빗으로 만들 수 있습니다. 또한 보안을 강화하기 위해 내부 API 엔드포인트만 갖도록 프라이빗 클러스터를 구성할 수 있습니다.

OpenShift Dedicated 관리자는 OpenShift Cluster Manager 내에서 퍼블릭 및 프라이빗 클러스터 구성 중에서 선택할 수 있습니다. 클러스터 생성 중 또는 클러스터가 설정된 후 개인 정보 설정을 구성할 수 있습니다.

2.2.1. 클러스터 생성 중 개인 클러스터 활성화

새 클러스터를 만들 때 개인 클러스터 설정을 활성화할 수 있습니다.

사전 요구 사항

- 개인 액세스를 허용하도록 다음 개인 연결을 구성해야 합니다.
  - VPC 피어링
  - Cloud VPN
  - DirectConnect (AWS만 해당)
  - TransitGateway (AWS 전용)
  - Cloud Interconnect (GCP 전용)

절차

1. OpenShift Cluster Manager 에 로그인합니다.
2. 클러스터 생성 → OpenShift Dedicated → 클러스터 생성을 클릭합니다.
3. 클러스터 세부 정보를 구성합니다.
4. 기본 네트워크 구성을 선택하는 경우 고급 을 선택합니다.
5. Private 을 선택합니다.

주의

Private 로 설정하면 사전 요구 사항에 설명된 대로 클라우드 공급자의 프라이빗 연결을 구성하지 않으면 클러스터에 액세스할 수 없습니다.

6. 클러스터 생성을 클릭합니다. 클러스터 생성 프로세스가 시작되고 완료하는 데 약 30~40분이 걸립니다.

검증

- Installing cluster heading는 Overview 탭에서 클러스터가 설치되고 이 제목에서 설치 로그를 볼 수 있음을 나타냅니다. Details (세부 정보) 제목 아래의 상태 표시기는 클러스터를 사용할준비가 된 시기를 나타냅니다.

2.2.2. 기존 클러스터를 프라이빗으로 활성화

클러스터가 생성되면 나중에 클러스터를 프라이빗으로 활성화할 수 있습니다.

사전 요구 사항

- 개인 액세스를 허용하도록 다음 개인 연결을 구성해야 합니다.
  - VPC 피어링
  - Cloud VPN

- DirectConnect (AWS만 해당)
- TransitGateway (AWS 전용)
- Cloud Interconnect (GCP 전용)

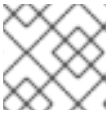
절차

1. OpenShift Cluster Manager 에 로그인합니다.
2. 비공개로 설정하려는 공용 클러스터를 선택합니다.
3. Networking (네트워킹) 탭의 Control Plane API 엔드포인트에서 Make API Private을 선택합니다.

주의

Private 로 설정하면 사전 요구 사항에 설명된 대로 클라우드 공급자의 프라이빗 연결을 구성하지 않으면 클러스터에 액세스할 수 없습니다.

4. 설정 변경을 클릭합니다.



참고

프라이빗과 퍼블릭 간에 클러스터를 전환하는 데 몇 분이 걸릴 수 있습니다.

2.2.3. 기존 프라이빗 클러스터가 공용으로 활성화

개인 클러스터가 생성되면 나중에 클러스터를 공용으로 활성화할 수 있습니다.

절차

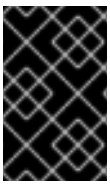
1. OpenShift Cluster Manager 에 로그인합니다.
2. 공개하려는 프라이빗 클러스터를 선택합니다.
3. Networking (네트워킹) 탭의 Control Plane API 엔드포인트에서 비공개 API를 선택 취소합니다.
4. 설정 변경을 클릭합니다.

참고

프라이빗과 퍼블릭 간에 클러스터를 전환하는 데 몇 분이 걸릴 수 있습니다.

3장. 클러스터 자동 스케일링

OpenShift Dedicated 클러스터에 자동 스케일링을 적용하려면 클러스터 자동 스케일러를 구성한 다음 클러스터에서 하나 이상의 머신 풀에 대한 머신 자동 스케일러를 구성해야 합니다.



중요

머신 API가 작동하는 클러스터에서만 클러스터 자동 스케일러를 구성할 수 있습니다.

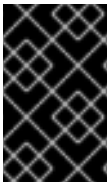
클러스터당 하나의 클러스터 자동 스케일러만 생성할 수 있습니다.

3.1. 클러스터 자동 스케일러 정보

클러스터 자동 스케일러는 현재 배포 요구 사항에 맞게 OpenShift Dedicated 클러스터의 크기를 조정합니다. 이는 Kubernetes 형식의 선언적 인수를 사용하여 특정 클라우드 공급자의 개체에 의존하지 않는 인프라 관리를 제공합니다. 클러스터 자동 스케일러에는 클러스터 범위가 있으며 특정 네임 스페이스와 연결되어 있지 않습니다.

리소스가 부족하여 현재 작업자 노드에서 pod를 예약할 수 없거나 배포 요구를 충족시키기 위해 다른 노드가 필요한 경우 클러스터 자동 스케일러는 클러스터 크기를 늘립니다. 클러스터 자동 스케일러는 사용자가 지정한 제한을 초과하여 클러스터 리소스를 늘리지 않습니다.

클러스터 자동 스케일러는 컨트롤 플레인 노드를 관리하지 않더라도 모든 노드에서 클러스터의 총 메모리와 CPU를 계산합니다. 이러한 값은 단일 시스템 지향이 아닙니다. 전체 클러스터에 있는 모든 리소스를 집계합니다. 예를 들어 최대 메모리 리소스 제한을 설정하면 클러스터 자동 스케일러에 현재 메모리 사용량을 계산할 때 클러스터의 모든 노드가 포함됩니다. 그런 다음 해당 계산은 클러스터 자동 스케일러에 작업자 리소스를 더 추가할 수 있는 용량이 있는지 결정하는 데 사용됩니다.



중요

작성한 **ClusterAutoscaler** 리솟스 정의의 **maxNodesTotal** 값이 클러스터에서 예상되는 총 머신 수를 대응하기에 충분한 크기의 값인지 확인합니다. 이 값에는 컨트롤 플레인 머신 수 및 확장 가능한 컴퓨팅 머신 수가 포함되어야 합니다.

10초마다 클러스터 자동 스케일러는 클러스터에서 불필요한 노드를 확인하고 해당 노드를 제거합니다. 다음 조건이 적용되는 경우 클러스터 자동 스케일러는 제거 노드를 고려합니다.

- 노드 사용률은 클러스터의 노드 *사용률 수준*임계값보다 적습니다. 노드 사용률 수준은 노드에 할당된 리소스로 구분된 요청된 리소스의 합계입니다. **ClusterAutoscaler** 사용자 정의 리소스에 값을 지정하지 않으면 클러스터 자동 스케일러는 기본값 **0.5** 를 사용하며 이 값은 50%에 해당합니다.
- 클러스터 자동 스케일러는 노드에서 실행 중인 모든 Pod를 다른 노드로 이동할 수 있습니다. Kubernetes 스케줄러는 노드에서 Pod를 예약하는 역할을 합니다.
- 클러스터 자동 스케일러에는 축소 비활성화된 주석이 없습니다.

노드에 다음 유형의 pod가 있는 경우 클러스터 자동 스케일러는 해당 노드를 제거하지 않습니다.

- 제한적인 PDB (Pod Disruption Budgets)가 있는 pod
- 기본적으로 노드에서 실행되지 않는 Kube 시스템 pod
- PDB가 없거나 제한적인 PDB가있는 Kube 시스템 pod
- deployment, replica set 또는 stateful set와 같은 컨트롤러 객체가 지원하지 않는 pod

- 로컬 스토리지가 있는 pod
- 리소스 부족, 호환되지 않는 노드 선택기 또는 어피니티(affinity), 안티-어피니티(anti-affinity) 일치 등으로 인해 다른 위치로 이동할 수 없는 pod
- **"cluster-autoscaler.kubernetes.io/safe-to-evict": "true"** 주석이없는 경우**"cluster-autoscaler.kubernetes.io/safe-to-evict": "false"** 주석이 있는 pod

예를 들어 최대 CPU 제한을 64개 코어로 설정하고 각각 8개의 코어가 있는 머신만 생성하도록 클러스터 자동 스케일러를 구성합니다. 클러스터가 30개 코어로 시작하는 경우 클러스터 자동 스케일러는 총 62개 코어에서 최대 4개의 노드를 추가할 수 있습니다.

클러스터 자동 스케일러를 구성하면 추가 사용 제한이 적용됩니다.

- 자동 스케일링된 노드 그룹에 있는 노드를 직접 변경하지 마십시오. 동일한 노드 그룹 내의 모든 노드는 동일한 용량 및 레이블을 가지며 동일한 시스템 pod를 실행합니다.
- pod 요청을 지정합니다.
- pod가 너무 빨리 삭제되지 않도록 해야 하는 경우 적절한 PDB를 구성합니다.
- 클라우드 제공자 할당량이 구성하는 최대 노드 풀을 지원할 수 있는 충분한 크기인지를 확인합니다.
- 추가 노드 그룹 Autoscaler, 특히 클라우드 제공자가 제공하는 Autoscaler를 실행하지 마십시오.

HPA (Horizond Pod Autoscaler) 및 클러스터 자동 스케일러는 다른 방식으로 클러스터 리소스를 변경합니다. HPA는 현재 CPU 로드를 기준으로 배포 또는 복제 세트의 복제 수를 변경합니다. 로드가 증가하면 HPA는 클러스터에 사용 가능한 리소스 양에 관계없이 새 복제본을 만듭니다. 리소스가 충분하지 않은 경우 클러스터 자동 스케일러는 리소스를 추가하고 HPA가 생성한 pod를 실행할 수 있도록 합니다. 로드가 감소하면 HPA는 일부 복제를 중지합니다. 이 동작으로 일부 노드가 충분히 활용되지 않거나 완전히 비어 있을 경우 클러스터 자동 스케일러가 불필요한 노드를 삭제합니다.

클러스터 자동 스케일러는 pod 우선 순위를 고려합니다. Pod 우선 순위 및 선점 기능을 사용하면 클러스터에 충분한 리소스가 없는 경우 우선 순위에 따라 pod를 예약할 수 있지만 클러스터 자동 스케일러는 클러스터에 모든 pod를 실행하는 데 필요한 리소스가 있는지 확인합니다. 두 기능을 충족하기 위해 클러스터 자동 스케일러에는 우선 순위 컷오프 기능이 포함되어 있습니다. 이 컷오프 기능을 사용하여 "best-effort" pod를 예약하면 클러스터 자동 스케일러가 리소스를 늘리지 않고 사용 가능한 예비 리소스가 있을 때만 실행됩니다.

컷오프 값보다 우선 순위가 낮은 pod는 클러스터가 확장되지 않거나 클러스터가 축소되지 않도록합니다. pod를 실행하기 위해 추가된 새 노드가 없으며 이러한 pod를 실행하는 노드는 리소스를 확보하기 위해 삭제될 수 있습니다.

머신 API를 사용할 수 있는 플랫폼에서 클러스터 자동 스케일링이 지원됩니다.

3.2. OPENSHIFT CLUSTER MANAGER를 사용하여 클러스터 생성 중 자동 스케일링 활성화

OpenShift Cluster Manager를 사용하여 클러스터 생성 중에 자동 스케일링할 수 있습니다.

절차

1. 클러스터 생성 중에 자동 스케일링 활성화 확인란을 선택합니다. Edit cluster autoscaling settings 버튼을 선택합니다.
    a. 자동 스케일링할 최소 또는 최대 노드 양을 선택할 수도 있습니다.

2. 클러스터 자동 스케일링 설정 편집을클릭합니다.

3. 원하는 설정을 편집한 다음 닫기 를 클릭합니다.

3.3. OPENSHIFT CLUSTER MANAGER를 사용하여 클러스터 생성 후 자동 스케일링 활성화

OpenShift Cluster Manager를 사용하여 클러스터 생성 후 자동 스케일링할 수 있습니다.

절차

1. OpenShift Cluster Manager에서 자동 스케일링할 클러스터의 이름을 클릭합니다. 클러스터의 개요 페이지에는 자동 확장 항목이 있으며 활성화되어 있는지 여부를 나타냅니다.

2. 머신 풀 탭을 클릭합니다.

3. Edit cluster autoscaling버튼을 클릭합니다. Edit cluster autoscalingsettings 창이 표시됩니다.

4. 창 상단에 있는 자동 스케일링 클러스터 토글을 클릭합니다. 이제 모든 설정을 편집할 수 있습니다.

5. 원하는 설정을 편집한 다음 저장을 클릭합니다.

6. 화면 오른쪽 상단에 있는 x 를 클릭하여 설정 창을 닫습니다.

모든 자동 스케일링 설정을 기본값으로 되돌리려면 Revert all to defaults 버튼을 클릭합니다.

3.4. OPENSHIFT CLUSTER MANAGER를 사용한 클러스터 자동 스케일링 설정

이 표에서는 OpenShift Cluster Manager에서 클러스터 자동 스케일링을 사용할 때 구성 가능한 모든 UI 설정을 설명합니다.

3.4.1. 일반 설정

표 3.1. OpenShift Cluster Manager를 사용할 때 클러스터 자동 스케일링을 위한 구성 가능한 일반 설정

설정	설명	유형 또는 범위	Default
log-verbosity	자동 스케일러 로그 수준을 설정합니다. 기본값은 1입니다. 수준 4는 디버깅에 권장됩니다. 레벨 6은 거의 모든 것을 가능하게 합니다.	**integer**	1
skip-nodes-with-local-storage	**true** 인 경우 클러스터 자동 스케일러는 로컬 스토리지가 있는 Pod(예: EmptyDir 또는 HostPath)가 있는 노드를 삭제하지 않습니다.	**boolean**	true

설정	설명	유형 또는 범위	Default
max-pod-grace-period	축소하기 전에 Pod의 정상 종료 시간(초)을 제공합니다.	**integer**	600
max-node-provision-time	클러스터 자동 스케일러가 노드를 프로비저닝할 때까지 대기하는 최대 시간입니다.	**string**	15m
pod-priority-threshold	사용자가 "best-effort" Pod를 예약할 수 있습니다. 이 Pod는 클러스터 자동 스케일러 작업을 트리거할 수 없습니다. 이러한 Pod는 예비 리소스를 사용할 수 있는 경우에만 실행됩니다.	**integer**	-10
ignore-daemonsets-utilization	축소할 리소스 사용률을 계산할 때 클러스터 자동 스케일러가 데몬 세트 Pod를 무시하는지 여부를 결정합니다.	**boolean**	false
balance-similar-node-groups	**true** 인 경우 이 설정은 동일한 인스턴스 유형 및 동일한 레이블 세트로 노드 그룹을 자동으로 식별하고 해당 노드 그룹의 각 크기를 균형 있게 유지하려고 합니다.	**boolean**	false
balancing-ignored-labels	이 옵션은 노드 그룹 유사성을 고려할 때 클러스터 자동 스케일러가 무시해야 하는 레이블을 지정합니다. 이 옵션에는 공백을 포함할 수 없습니다.	**배열(문자열)**	형식은 쉼표로 구분된 레이블 목록이어야 합니다.

3.4.2. 리소스 제한

표 3.2. OpenShift Cluster Manager를 사용할 때 클러스터 자동 스케일링에 대해 구성 가능한 리소스 제한 설정

설정	설명	유형 또는 범위	Default

설정	설명	유형 또는 범위	Default
cores-total-min	클러스터의 최소 코어 수입니다. 클러스터 자동 스케일러는 이 수보다 클러스터를 확장하지 않습니다.	**object**	0
cores-total-max	클러스터의 최대 코어 수입니다. 클러스터 자동 스케일러는 이 수보다 클러스터를 확장하지 않습니다.	**object**	180 * 64 (11520)
memory-total-min	클러스터의 최소 메모리 수입니다. 클러스터 자동 스케일러는 이 수보다 클러스터를 확장하지 않습니다.	**object**	0
memory-total-max	클러스터의 최대 메모리 수입니다. 클러스터 자동 스케일러는 이 수보다 클러스터를 확장하지 않습니다.	**object**	180 * 64 * 20 (230400)
max-nodes-total	모든 노드 그룹의 최대 노드 수입니다. 자동으로 확장되는 노드가 아닌 모든 노드를 포함합니다. 클러스터 자동 스케일러는 이 수보다 클러스터를 늘리지 않습니다.	**integer**	180
GPU	클러스터에 있는 다른 GPU의 최소 및 최대 수입니다. 클러스터 자동 스케일러는 이러한 숫자보다 작거나 큰 클러스터를 확장하지 않습니다.	**array**	형식은 쉼표로 구분된 "<min>:<max>"{p> 목록이어야 합니다.

3.4.3. 구성 축소

표 3.3. OpenShift Cluster Manager를 사용할 때 클러스터 자동 스케일링을 위한 구성 가능한 축소 설정

설정	설명	유형 또는 범위	Default
scale-down-enabled	클러스터 자동 스케일러가 클러스터를 축소해야 합니까.	**boolean**	true

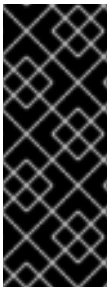
설정	설명	유형 또는 범위	Default
scale-down-utilization-threshold	요청된 리소스를 용량으로 나눈 총합으로 정의된 노드 사용률 수준보다 노드를 축소할 수 있습니다.	**플로트**	0.5
scale-down-unneeded-time	노드를 축소할 수 있기 전에 필요하지 않은 노드 수입니다.	**string**	10m
scale-down-delay-after-add	확장 후 축소 평가가 재개되는 시간입니다.	**string**	10m
scale-down-delay-after-delete	노드를 삭제한 후 축소 평가가 재개되는 기간입니다.	**string**	0s
scale-down-delay-after-failure	스케일 다운 평가가 다시 시작되는 실패 후의 시간입니다.	**string**	3m

4장. 노드

4.1. 머신 풀 정보

OpenShift Dedicated는 머신 풀을 클라우드 인프라 상단에 탄력적이고 동적인 프로비저닝 방법으로 사용합니다.

기본 리소스는 머신, 머신 세트 및 머신 풀입니다.



중요

OpenShift Dedicated 4.11부터 기본 Pod당 PID 제한은 **4096** 입니다. 이 PID 제한을 활성화하려면 OpenShift Dedicated 클러스터를 이 버전 이상으로 업그레이드해야 합니다. 4.11 이전 버전을 실행하는 OpenShift Dedicated 클러스터는 기본 PID 제한을 **1024** 로 제한합니다.

OpenShift Dedicated 클러스터에서는 Pod별 PID 제한을 구성할 수 없습니다.

4.1.1. Machine

머신은 작업자 노드의 호스트를 설명하는 기본 단위입니다.

4.1.2. 머신 세트

MachineSet 리소스는 컴퓨팅 머신 그룹입니다. 더 많은 시스템이 필요하거나 규모를 줄여야 하는 경우 컴퓨팅 시스템 세트가 속하는 시스템 풀의 복제본 수를 변경합니다.

4.1.3. 머신 풀

머신 풀은 컴퓨팅 머신 세트에 대한 더 높은 수준의 구성 요소입니다.

시스템 풀은 모두 가용성 영역에서 동일한 구성을 복제하는 컴퓨팅 머신 세트를 생성합니다. 머신 풀은 작업자 노드에서 모든 호스트 노드 프로비저닝 관리 작업을 수행합니다. 더 많은 머신이 필요하거나 규모를 줄여야하는 경우 컴퓨팅 요구 사항에 맞게 시스템 풀의 복제본 수를 변경합니다. 스케일링을 수동으로 구성하거나 자동 스케일링을 설정할 수 있습니다.

기본적으로 클러스터는 하나의 머신 풀로 생성됩니다. 기존 클러스터에 머신 풀을 추가하고, 기본 머신 풀을 수정하고, 머신 풀을 삭제할 수 있습니다.

단일 클러스터에 여러 머신 풀이 존재할 수 있으며 각각 다른 유형 또는 크기 노드를 사용할 수 있습니다.

4.1.4. 여러 영역 클러스터의 머신 풀

다중 가용성 영역(Multi-AZ) 클러스터에 머신 풀을 생성할 때 하나의 머신 풀에 3개의 영역이 있습니다. 시스템 풀은 차례로 클러스터의 각 영역에 하나씩 총 3개의 컴퓨팅 머신 세트를 생성합니다. 각 컴퓨팅 시스템 세트는 해당 가용성 영역에 있는 하나 이상의 시스템을 관리합니다.

새 Multi-AZ 클러스터를 생성하면 머신 풀이 해당 영역에 자동으로 복제됩니다. 기존 Multi-AZ에 머신 풀을 추가하면 해당 영역에 새 풀이 자동으로 생성됩니다. 마찬가지로 머신 풀을 삭제하면 모든 영역에서 삭제됩니다. 이러한 다기능 효과 때문에 Multi-AZ 클러스터에서 머신 풀을 사용하면 머신 풀을 생성할 때 특정 리전에 더 많은 프로젝트의 할당량을 사용할 수 있습니다.

4.1.5. 추가 리소스

- 자동 스케일링 정보

4.2. 컴퓨팅 노드 관리

이 문서에서는 OpenShift Dedicated를 사용하여 컴퓨팅 (작업자라고도 함) 노드를 관리하는 방법을 설명합니다.

컴퓨팅 노드의 대부분의 변경 사항은 시스템 풀에 구성됩니다. 머신 풀은 동일한 구성이 있어 쉽게 관리할 수 있는 클러스터의 컴퓨팅 노드 그룹입니다.

스케일링, 노드 레이블 추가, 테인트 추가와 같은 머신 풀 구성 옵션을 편집할 수 있습니다.

4.2.1. 머신 풀 생성

OpenShift Dedicated 클러스터를 설치할 때 머신 풀이 생성됩니다. 설치 후 OpenShift Cluster Manager를 사용하여 클러스터에 대한 추가 머신 풀을 생성할 수 있습니다.

중요

사용 가능한 컴퓨팅 (작업자) 노드 인스턴스 유형, 자동 스케일링 옵션 및 노드 수는 OpenShift Dedicated 서브스크립션, 리소스 할당량 및 배포 시나리오에 따라 다릅니다. 자세한 내용은 영업 담당자 또는 Red Hat 지원에 문의하십시오.

사전 요구 사항

- OpenShift Dedicated 클러스터를 생성하셨습니다.

절차

1. OpenShift Cluster Manager 로 이동하여 클러스터를 선택합니다.
2. 머신 풀 탭에서 머신 풀 추가 를 클릭합니다.
3. 머신 풀 이름을 추가합니다.
4. 드롭다운 메뉴에서 컴퓨팅 노드 인스턴스 유형을 선택합니다. 인스턴스 유형은 시스템 풀의 각 계산 노드에 대한 vCPU 및 메모리 할당을 정의합니다.

참고

풀을 생성한 후에는 머신 풀의 인스턴스 유형을 변경할 수 없습니다.

5. 선택 사항: 머신 풀의 자동 스케일링을 구성합니다.
   a. 배포 요구에 맞게 자동 스케일링 활성화를 선택하여 머신 풀의 머신 수를 자동으로 스케일링합니다.

참고

자동 스케일링 활성화 옵션은 **capability.cluster.autoscale_clusters** 서브스크립션이 있는 경우에만 OpenShift Dedicated에서 사용할 수 있습니다. 자세한 내용은 영업 담당자 또는 Red Hat 지원에 문의하십시오.

b. 자동 스케일링에 대한 최소 및 최대 노드 수 제한을 설정합니다. 클러스터 자동 스케일러는 지정한 제한을 초과하여 머신 풀 노드 수를 줄이거나 늘리지 않습니다.

- 단일 가용성 영역을 사용하여 클러스터를 배포한 경우 최소 및 최대 노드 수를 설정합니다. 이는 가용성 영역에 최소 및 최대 컴퓨팅 노드 제한을 정의합니다.
- 여러 가용성 영역을 사용하여 클러스터를 배포한 경우 영역당 최소 노드 및 영역당최대 노드를 설정합니다. 이는 영역당 최소 및 최대 컴퓨팅 노드 제한을 정의합니다.

참고

또는 머신 풀을 생성한 후 머신 풀에 대한 자동 스케일링 기본 설정을 설정할 수 있습니다.

6. 자동 스케일링을 활성화하지 않은 경우 컴퓨팅 노드 수를 선택합니다.
   - 단일 가용성 영역을 사용하여 클러스터를 배포한 경우 드롭다운 메뉴에서 컴퓨팅 노드 수를 선택합니다. 이 명령은 영역의 시스템 풀에 프로비저닝할 컴퓨팅 노드 수를 정의합니다.
   - 여러 가용성 영역을 사용하여 클러스터를 배포한 경우 드롭다운 메뉴에서 컴퓨팅 노드 수(영역당) 를 선택합니다. 이는 영역당 시스템 풀에 프로비저닝할 컴퓨팅 노드 수를 정의합니다.
7. 선택사항: 머신 풀의 노드 레이블 및 테인트를 추가합니다.
   a. Edit node labels and taints메뉴를 확장합니다.
   b. 노드 라벨 에서 노드 레이블에 대한Key 및 Value 항목을 추가합니다.
   c. 테인트에서 테인트의 키 및 값 항목을 추가합니다.

참고

테인트를 사용하여 머신 풀을 생성하는 것은 클러스터에 테인트 없이 하나 이상의 머신 풀이 이미 있는 경우에만 가능합니다.

   d. 각 테인트에 대해 드롭다운 메뉴에서 Effect 를 선택합니다. 사용 가능한 옵션에는 **NoSchedule**,**PreferNoSchedule** 및 **NoExecute** 가 포함됩니다.

참고

또는 머신 풀을 생성한 후 노드 레이블 및 테인트를 추가할 수 있습니다.

8. 선택 사항: 이 머신 풀의 노드에 사용할 추가 사용자 지정 보안 그룹을 선택합니다. 보안 그룹을 이미 생성하여 이 클러스터에 대해 선택한 VPC와 연결되어야 합니다. 머신 풀을 생성한 후에는 보안 그룹을 추가하거나 편집할 수 없습니다. 자세한 내용은 "추가 리소스" 섹션의 보안 그룹에 대한 요구 사항을 참조하십시오.
9. 선택 사항: CCO(Customer Cloud Subscription) 모델을 사용하여 AWS에 OpenShift Dedicated를 배포한 경우 머신 풀을 구성하여 머신을 보장되지 않는 AWS Spot 인스턴스로 배포하려는 경우 Amazon EC2 Spot 인스턴스를 사용합니다.
   a. Amazon EC2 Spot 인스턴스 사용을선택합니다.
   b. 온 디맨드 인스턴스 가격을사용하도록 선택한 온 디맨드 인스턴스 가격은 그대로 둡니다. 또는 최대 가격 설정을 선택하여 Spot 인스턴스의 최대 시간당 가격을 정의합니다.

Amazon EC2 Spot 인스턴스에 대한 자세한 내용은 AWS 설명서 를 참조하십시오.

중요

언제든지 Amazon EC2 Spot 인스턴스가 중단될 수 있습니다. 중단을 허용할 수 있는 워크로드에만 Amazon EC2 Spot 인스턴스를 사용합니다.

참고

머신 풀에 Amazon EC2 Spot 인스턴스 사용을 선택하는 경우 머신 풀을 생성한 후에는 옵션을 비활성화할 수 없습니다.

10. 머신 풀 추가를 클릭하여 머신 풀을 생성합니다.

검증

- 시스템 풀이 머신 풀 페이지에 표시되고 구성이 예상대로 표시되는지 확인합니다.

4.2.2. 머신 풀 삭제

워크로드 요구 사항이 변경되고 현재 머신 풀이 더 이상 요구 사항을 충족하지 않는 경우 머신 풀을 삭제할 수 있습니다.

OpenShift Cluster Manager를 사용하여 머신 풀을 삭제할 수 있습니다.

사전 요구 사항

- OpenShift Dedicated 클러스터를 생성했습니다.
- 클러스터가 ready 상태입니다.
- 테인트 없이 기존 머신 풀이 있고 Single-AZ 클러스터에는 두 개 이상의 복제본이 있거나 Multi-AZ 클러스터의 복제본 3개가 있습니다.

절차

1. OpenShift Cluster Manager 에서 Cluster List 페이지로 이동하여 삭제할 시스템 풀이 포함된 클러스터를 선택합니다.
2. 선택한 클러스터에서 머신 풀 탭을 선택합니다.
3. 머신 풀 탭에서 삭제하려는 머신 풀의 옵션 메뉴 를 클릭합니다.
4. 삭제를 클릭합니다.

선택한 머신 풀이 삭제됩니다.

4.2.3. 수동으로 컴퓨팅 노드 확장

머신 풀에 자동 스케일링을 활성화하지 않은 경우 배포 요구에 맞게 풀의 컴퓨팅 (작업자라고도 함) 노드 수를 수동으로 스케일링할 수 있습니다.

각 머신 풀을 별도로 스케일링해야 합니다.

사전 요구 사항

- OpenShift Dedicated 클러스터를 생성하셨습니다.
- 기존 머신 풀이 있습니다.

절차

1. OpenShift Cluster Manager 로 이동하여 클러스터를 선택합니다.
2. 머신 풀 탭에서 스케일링할 머신 풀의 옵션 메뉴 를 클릭합니다.
3. 스케일링 을 선택합니다.
4. 노드 수를 지정합니다.
   - 단일 가용성 영역을 사용하여 클러스터를 배포한 경우 드롭다운 메뉴에서 노드 수를 지정합니다.
   - 여러 가용성 영역을 사용하여 클러스터를 배포한 경우 드롭다운 메뉴에서 영역당 노드 수를 지정합니다.

   참고

   서브스크립션에 따라 선택할 수 있는 노드 수가 결정됩니다.
5. 적용을 클릭하여 머신 풀을 확장합니다.

검증

- 머신 풀 탭에서 시스템 풀의 노드 수가 예상대로 표시되는지 확인합니다.

4.2.4. 노드 라벨

레이블은 **Node** 오브젝트에 적용되는 키-값 쌍입니다. 라벨을 사용하여 오브젝트 세트를 구성하고 Pod 예약을 제어할 수 있습니다.

클러스터 생성 중 또는 이후에 라벨을 추가할 수 있습니다. 레이블은 언제든지 수정하거나 업데이트할 수 있습니다.

추가 리소스

- 라벨에 대한 자세한 내용은 Kubernetes 라벨 및 선택자 개요를 참조하십시오.
- 사용자 지정 추가 보안 그룹 요구 사항에 대한 자세한 내용은 추가 사용자 지정 보안 그룹을 참조하십시오.

4.2.4.1. 머신 풀에 노드 라벨 추가

언제든지 compute(작업자라고도 함) 노드의 라벨을 추가하거나 편집하여 사용자와 관련된 방식으로 노드를 관리합니다. 예를 들어 특정 노드에 워크로드 유형을 할당할 수 있습니다.

레이블은 키-값 쌍으로 할당됩니다. 각 키는 할당된 오브젝트에 고유해야 합니다.

사전 요구 사항

- OpenShift Dedicated 클러스터를 생성하셨습니다.
- 기존 머신 풀이 있습니다.

절차

1. OpenShift Cluster Manager 로 이동하여 클러스터를 선택합니다.
2. 머신 풀 탭에서 레이블을 추가할 머신 풀의 옵션 메뉴 를 클릭합니다.
3. 라벨 편집을 선택합니다.
4. 제거할 머신 풀에 기존 레이블이 있는 경우 레이블 옆에 있는 x 를 선택하여 삭제합니다.
5. **< key>=<value >** 형식을 사용하여 레이블을 추가하고 Enter 키를 누릅니다. 예를 들어**app=db** 를 추가한 다음 Enter를 누릅니다. 형식이 올바르면 키 값 쌍이 강조 표시됩니다.
6. 추가 레이블을 추가하려면 이전 단계를 반복합니다.
7. 저장 을 클릭하여 머신 풀에 레이블을 적용합니다.

검증

1. 머신 풀 탭에서 머신 풀옆에 있는 >를 선택하여 보기를 확장합니다.
2. 확장된 뷰의 라벨 아래에 라벨 이 나열되어 있는지 확인합니다.

4.2.5. 머신 풀에 테인트 추가

머신 풀에서 compute(작업자라고도 함) 노드의 테인트를 추가하여 예약된 Pod를 제어할 수 있습니다. 머신 풀에 테인트를 적용하면 Pod 사양에 테인트에 대한 허용 오차가 포함되지 않는 한 스케줄러는 풀의 노드에 Pod를 배치할 수 없습니다.

참고

클러스터에 테인트가 포함되지 않은 하나 이상의 머신 풀이 있어야 합니다.

사전 요구 사항

- OpenShift Dedicated 클러스터를 생성하셨습니다.
- 기존 머신 풀에는 테인트가 포함되지 않고 두 개 이상의 인스턴스가 포함됩니다.

절차

1. OpenShift Cluster Manager 로 이동하여 클러스터를 선택합니다.
2. 머신 풀 탭에서 테인트를 추가할 머신 풀의 옵션 메뉴 를 클릭합니다.

3. 테인트 편집을 선택합니다.
4. 테인트에 대한 키 및 값 항목을 추가합니다.
5. 드롭다운 메뉴에서 테인트에 대한 Effect 를 선택합니다. 사용 가능한 옵션에는 **NoSchedule**, **PreferNoSchedule** 및 **NoExecute** 가 포함됩니다.
6. 머신 풀에 테인트 를 추가하려면 테인트 추가를 선택합니다.
7. 저장 을 클릭하여 머신 풀에 테인트를 적용합니다.

검증

1. 머신 풀 탭에서 머신 풀 옆에 있는 >를 선택하여 보기를 확장합니다.
2. 확장된 뷰에서 테인트가 테인트에 나열되어 있는지 확인합니다.

4.2.6. 추가 리소스

- 머신 풀 정보
- 자동 스케일링 활성화
- 자동 스케일링 비활성화
- OpenShift Dedicated 서비스 정의

4.3. 클러스터의 노드 자동 스케일링 정보

중요

자동 스케일링은 Google Cloud Marketplace 및 Red Hat Marketplace를 통해 구매한 클러스터에서만 사용할 수 있습니다.

자동 스케일러 옵션은 클러스터의 머신 수를 자동으로 확장하도록 구성할 수 있습니다.

리소스가 부족하여 현재 노드에서 pod를 예약할 수 없거나 배포 요구를 충족시키기 위해 다른 노드가 필요한 경우 클러스터 자동 스케일러는 클러스터 크기를 늘립니다. 클러스터 자동 스케일러는 사용자가 지정한 제한을 초과하여 클러스터 리소스를 늘리지 않습니다.

또한 클러스터 자동 스케일러는 리소스 사용이 적고 중요한 pod가 모두 다른 노드에 적합한 경우와 같이 상당한 기간 동안 일부 노드가 지속적으로 필요하지 않은 경우 클러스터 크기를 줄입니다.

자동 스케일링을 활성화하는 경우 최소 및 최대 작업자 노드 수를 설정해야 합니다.

참고

클러스터 소유자 및 조직 관리자만 클러스터를 확장하거나 삭제할 수 있습니다.

4.3.1. 클러스터에서 노드 자동 스케일링 활성화

작업자 노드에서 자동 스케일링을 활성화하여 기존 클러스터에 대한 머신 풀 정의를 편집하여 사용 가능한 노드 수를 늘리거나 줄일 수 있습니다.

Red Hat OpenShift Cluster Manager를 사용하여 기존 클러스터에서 노드 자동 스케일링 활성화
OpenShift Cluster Manager 콘솔에서 머신 풀 정의에서 작업자 노드에 대한 자동 스케일링을 활성화합니다.

절차

1. OpenShift Cluster Manager 에서 클러스터 목록 페이지로 이동하여 자동 스케일링을 활성화할 클러스터를 선택합니다.
2. 선택한 클러스터에서 머신 풀 탭을 선택합니다.
3. 자동 스케일링을 활성화하려는 머신 풀 끝에 있는 옵션 메뉴 를 클릭하고 스케일링을 선택합니다.
4. 노드 수 편집 대화 상자에서 자동 스케일링 활성화 확인란을 선택합니다.
5. 적용을 선택하여 이러한 변경 사항을 저장하고 클러스터에 대한 자동 스케일링을 활성화합니다.

4.3.2. 클러스터에서 자동 스케일링 노드 비활성화

작업자 노드에서 자동 스케일링을 비활성화하여 기존 클러스터에 대한 머신 풀 정의를 편집하여 사용 가능한 노드 수를 늘리거나 줄일 수 있습니다.

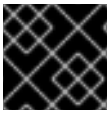
OpenShift Cluster Manager 콘솔을 사용하여 클러스터에서 자동 스케일링을 비활성화할 수 있습니다.

Red Hat OpenShift Cluster Manager를 사용하여 기존 클러스터에서 자동 스케일링 노드 비활성화
OpenShift Cluster Manager 콘솔에서 시스템 풀 정의에서 작업자 노드의 자동 스케일링을 비활성화합니다.

절차

1. OpenShift Cluster Manager 에서 Cluster List 페이지로 이동하여 비활성화해야 하는 자동 스케일링이 있는 클러스터를 선택합니다.
2. 선택한 클러스터에서 머신 풀 탭을 선택합니다.
3. 자동 스케일링을 사용하여 머신 풀 끝에 있는 옵션 메뉴 를 클릭하고 스케일링 을 선택합니다.
4. "노드 수 편집" 대화 상자에서 Enable autoscaling 확인란을 선택 취소합니다.
5. 적용을 선택하여 이러한 변경 사항을 저장하고 클러스터에서 자동 스케일링을 비활성화합니다.

OpenShift Dedicated 클러스터에 자동 스케일링을 적용하려면 클러스터 자동 스케일러를 배포한 다음 클러스터의 각 머신 유형에 대한 머신 자동 스케일러를 배포해야 합니다.



중요

Machine API가 작동하는 클러스터에서만 클러스터 자동 스케일러를 구성할 수 있습니다.

4.3.3. 클러스터 자동 스케일러 정보

클러스터 자동 스케일러는 현재 배포 요구 사항에 맞게 OpenShift Dedicated 클러스터의 크기를 조정합니

다. 이는 Kubernetes 형식의 선언적 인수를 사용하여 특정 클라우드 공급자의 개체에 의존하지 않는 인프라 관리를 제공합니다. 클러스터 자동 스케일러에는 클러스터 범위가 있으며 특정 네임 스페이스와 연결되어 있지 않습니다.

리소스가 부족하여 현재 작업자 노드에서 pod를 예약할 수 없거나 배포 요구를 충족시키기 위해 다른 노드가 필요한 경우 클러스터 자동 스케일러는 클러스터 크기를 늘립니다. 클러스터 자동 스케일러는 사용자가 지정한 제한을 초과하여 클러스터 리소스를 늘리지 않습니다.

클러스터 자동 스케일러는 컨트롤 플레인 노드를 관리하지 않더라도 모든 노드에서 클러스터의 총 메모리와 CPU를 계산합니다. 이러한 값은 단일 시스템 지향이 아닙니다. 전체 클러스터에 있는 모든 리소스를 집계합니다. 예를 들어 최대 메모리 리소스 제한을 설정하면 클러스터 자동 스케일러에 현재 메모리 사용량을 계산할 때 클러스터의 모든 노드가 포함됩니다. 그런 다음 해당 계산은 클러스터 자동 스케일러에 작업자 리소스를 더 추가할 수 있는 용량이 있는지 결정하는 데 사용됩니다.

중요

작성한 **ClusterAutoscaler** 리솟스 정의의 **maxNodesTotal** 값이 클러스터에서 예상되는 총 머신 수를 대응하기에 충분한 크기의 값인지 확인합니다. 이 값에는 컨트롤 플레인 머신 수 및 확장 가능한 컴퓨팅 머신 수가 포함되어야 합니다.

10초마다 클러스터 자동 스케일러는 클러스터에서 불필요한 노드를 확인하고 해당 노드를 제거합니다. 다음 조건이 적용되는 경우 클러스터 자동 스케일러는 제거 노드를 고려합니다.

- 노드 사용률은 클러스터의 노드 *사용률 수준*임계값보다 적습니다. 노드 사용률 수준은 노드에 할당된 리소스로 구분된 요청된 리소스의 합계입니다. **ClusterAutoscaler** 사용자 정의 리소스에 값을 지정하지 않으면 클러스터 자동 스케일러는 기본값 **0.5** 를 사용하며 이 값은 50%에 해당합니다.
- 클러스터 자동 스케일러는 노드에서 실행 중인 모든 Pod를 다른 노드로 이동할 수 있습니다. Kubernetes 스케줄러는 노드에서 Pod를 예약하는 역할을 합니다.
- 클러스터 자동 스케일러에는 축소 비활성화된 주석이 없습니다.

노드에 다음 유형의 pod가 있는 경우 클러스터 자동 스케일러는 해당 노드를 제거하지 않습니다.

- 제한적인 PDB (Pod Disruption Budgets)가 있는 pod
- 기본적으로 노드에서 실행되지 않는 Kube 시스템 pod
- PDB가 없거나 제한적인 PDB가있는 Kube 시스템 pod
- deployment, replica set 또는 stateful set와 같은 컨트롤러 객체가 지원하지 않는 pod
- 로컬 스토리지가 있는 pod
- 리소스 부족, 호환되지 않는 노드 선택기 또는 어피니티(affinity), 안티-어피니티(anti-affinity) 일치 등으로 인해 다른 위치로 이동할 수 없는 pod
- **"cluster-autoscaler.kubernetes.io/safe-to-evict": "true"** 주석이없는 경우**"cluster-autoscaler.kubernetes.io/safe-to-evict": "false"** 주석이 있는 pod

예를 들어 최대 CPU 제한을 64개 코어로 설정하고 각각 8개의 코어가 있는 머신만 생성하도록 클러스터 자동 스케일러를 구성합니다. 클러스터가 30개 코어로 시작하는 경우 클러스터 자동 스케일러는 총 62개 코어에서 최대 4개의 노드를 추가할 수 있습니다.

클러스터 자동 스케일러를 구성하면 추가 사용 제한이 적용됩니다.

- 자동 스케일링된 노드 그룹에 있는 노드를 직접 변경하지 마십시오. 동일한 노드 그룹 내의 모든 노드는 동일한 용량 및 레이블을 가지며 동일한 시스템 pod를 실행합니다.
- pod 요청을 지정합니다.
- pod가 너무 빨리 삭제되지 않도록 해야 하는 경우 적절한 PDB를 구성합니다.
- 클라우드 제공자 할당량이 구성하는 최대 노드 풀을 지원할 수 있는 충분한 크기인지를 확인합니다.
- 추가 노드 그룹 Autoscaler, 특히 클라우드 제공자가 제공하는 Autoscaler를 실행하지 마십시오.

HPA (Horizond Pod Autoscaler) 및 클러스터 자동 스케일러는 다른 방식으로 클러스터 리소스를 변경합니다. HPA는 현재 CPU 로드를 기준으로 배포 또는 복제 세트의 복제 수를 변경합니다. 로드가 증가하면 HPA는 클러스터에 사용 가능한 리소스 양에 관계없이 새 복제본을 만듭니다. 리소스가 충분하지 않은 경우 클러스터 자동 스케일러는 리소스를 추가하고 HPA가 생성한 pod를 실행할 수 있도록 합니다. 로드가 감소하면 HPA는 일부 복제를 중지합니다. 이 동작으로 일부 노드가 충분히 활용되지 않거나 완전히 비어 있을 경우 클러스터 자동 스케일러가 불필요한 노드를 삭제합니다.

클러스터 자동 스케일러는 pod 우선 순위를 고려합니다. Pod 우선 순위 및 선점 기능을 사용하면 클러스터에 충분한 리소스가 없는 경우 우선 순위에 따라 pod를 예약할 수 있지만 클러스터 자동 스케일러는 클러스터에 모든 pod를 실행하는 데 필요한 리소스가 있는지 확인합니다. 두 기능을 충족하기 위해 클러스터 자동 스케일러에는 우선 순위 컷오프 기능이 포함되어 있습니다. 이 컷오프 기능을 사용하여 "best-effort" pod를 예약하면 클러스터 자동 스케일러가 리소스를 늘리지 않고 사용 가능한 예비 리소스가 있을 때만 실행됩니다.

컷오프 값보다 우선 순위가 낮은 pod는 클러스터가 확장되지 않거나 클러스터가 축소되지 않도록합니다. pod를 실행하기 위해 추가된 새 노드가 없으며 이러한 pod를 실행하는 노드는 리소스를 확보하기 위해 삭제될 수 있습니다.

머신 API를 사용할 수 있는 플랫폼에서 클러스터 자동 스케일링이 지원됩니다.

4.3.4. 추가 리소스

- machinepools 정보