



Service und Wartung

BeeGFS on NetApp with E-Series Storage

NetApp
August 23, 2024

Inhalt

- Service und Wartung 1
 - Failover- und Failback-Services 1
 - Versetzen Sie das Cluster in den Wartungsmodus 3
 - Beenden Sie den Cluster und starten Sie den Cluster 4
 - Datei-Nodes ersetzen 5
 - Erweitern oder verkleinern Sie den Cluster 7

Service und Wartung

Failover- und Failback-Services

BeeGFS-Services zwischen Cluster-Nodes verschieben

Überblick

BeeGFS-Services können ein Failover zwischen den Nodes im Cluster durchführen, um sicherzustellen, dass die Clients weiterhin auf das Filesystem zugreifen können, wenn ein Node einen Fehler aufweist, oder Sie müssen eine geplante Wartung durchführen. In diesem Abschnitt werden verschiedene Möglichkeiten beschrieben, wie Administratoren das Cluster nach der Wiederherstellung nach einem Ausfall reparieren oder Services manuell zwischen Nodes verschieben können.

Schritte

Failover und Failback

Failover (Geplant)

Wenn Sie einen einzelnen Datei-Node zur Wartung offline schalten müssen, möchten Sie in der Regel alle BeeGFS-Dienste von diesem Node verschieben (oder ablassen). Dies kann erreicht werden, indem zunächst der Knoten in den Standby-Modus versetzt wird:

```
pcs node standby <HOSTNAME>
```

Nach der Überprüfung mit `pcs status` Alle Ressourcen wurden auf dem alternativen Datei-Node neu gestartet. Sie können je nach Bedarf weitere Änderungen am Node vornehmen.

Failback (nach einem geplanten Failover)

Wenn Sie bereit sind, die BeeGFS-Dienste zuerst auf den bevorzugten Knoten wiederherzustellen `pcs status` Und überprüfen Sie in der „Knotenliste“, ob der Status Standby lautet. Wenn der Node neu gebootet wurde, wird er offline angezeigt, bis Sie die Cluster-Services in den Online-Modus versetzen:

```
pcs cluster start <HOSTNAME>
```

Sobald der Node online ist, bringen Sie ihn aus dem Standby-Modus mit:

```
pcs cluster node unstandby <HOSTNAME>
```

Schließlich verlagern alle BeeGFS-Dienste wieder auf ihre bevorzugten Knoten mit:

```
pcs resource relocate run
```

Failback (nach einem ungeplanten Failover)

Wenn auf einem Node ein Hardware- oder ein anderer Fehler auftritt, sollte der HA-Cluster automatisch reagieren und seine Services auf einen gesunden Node verschieben. So bleibt den Administratoren Zeit für Korrekturmaßnahmen. Bevor Sie fortfahren, lesen "[Fehlerbehebung](#)" Sie den Abschnitt, um die Ursache des Failovers zu ermitteln und alle offenen Probleme zu beheben. Sobald der Knoten wieder eingeschaltet ist und sich in einem ordnungsgemäßen Zustand befindet, können Sie mit dem Failback fortfahren.

Wenn ein Node nach einem ungeplanten (oder geplanten) Neubooten gebootet wird, werden Cluster-Services nicht automatisch gestartet. Sie müssen daher den Node zuerst in den Online-Modus versetzen:

```
pcs cluster start <HOSTNAME>
```

Bei der nächsten Bereinigung werden alle Ressourcenfehler behoben, und der Fechtverlauf des Node wird zurückgesetzt:

```
pcs resource cleanup node=<HOSTNAME>
pcs stonith history cleanup <HOSTNAME>
```

Verifizieren in `pcs status` Der Knoten ist online und in einem ordnungsgemäßen Zustand. Standardmäßig werden BeeGFS-Dienste nicht automatisch Failback durchführen, um zu vermeiden, dass Ressourcen versehentlich auf einen ungesunden Knoten zurückverschoben werden. Wenn Sie bereit sind, alle Ressourcen im Cluster wieder an die bevorzugten Nodes zurückzugeben, mit den folgenden Funktionen:

```
pcs resource relocate run
```

Einzelne BeeGFS-Services werden auf alternative Datei-Nodes verschoben

Verschieben Sie einen BeeGFS-Service dauerhaft auf einen neuen Datei-Node

Wenn Sie den bevorzugten Datei-Node für einen einzelnen BeeGFS-Service dauerhaft ändern möchten, passen Sie den Ansible-Bestand an, sodass der bevorzugte Node zuerst aufgelistet wird, und führen Sie das Ansible-Playbook erneut aus.

In dieser Beispieldatei ist `beegfs_01` beispielsweise `inventory.yml` der bevorzugte Datei-Node zum Ausführen des BeeGFS-Managementservice:

```
mgmt:
  hosts:
    beegfs_01:
    beegfs_02:
```

Durch eine Umkehrung des Auftrags würden die Managementservices am `beegfs_02` bevorzugt werden:

```
mgmt:
  hosts:
    beegfs_02:
    beegfs_01:
```

Verschieben Sie einen BeeGFS-Service vorübergehend auf einen alternativen Datei-Node

Im Allgemeinen, wenn ein Knoten gerade gewartet wird, möchten Sie die Schritte [Failover und Failback](#Failover-and-Failback) verwenden, um alle Dienste von diesem Knoten weg zu verschieben.

Wenn Sie aus irgendeinem Grund einen einzelnen Service auf einen anderen Dateiknoten verschieben müssen, führen Sie:

```
pcs resource move <SERVICE>-monitor <HOSTNAME>
```



Geben Sie keine einzelnen Ressourcen oder die Ressourcengruppe an. Geben Sie immer den Namen des Monitors für den BeeGFS-Dienst an, den Sie verschieben möchten. Um zum Beispiel den BeeGFS-Managementdienst auf beegfs_02 zu verschieben, führen Sie: Aus `pcs resource move mgmt-monitor beegfs_02`. Dieser Prozess kann wiederholt werden, um einen oder mehrere Services von den bevorzugten Nodes weg zu verschieben. Überprüfen Sie, ob `pcs status` die Services auf dem neuen Node verlegt/gestartet wurden.

Wenn Sie einen BeeGFS-Service wieder auf den bevorzugten Node verschieben möchten, löschen Sie zuerst die temporären Ressourcenbeschränkungen (diesen Schritt wird bei mehreren Services wiederholt):

```
pcs resource clear <SERVICE>-monitor
```

Wenn Sie bereit sind, den Service(s) dann wieder zurück zu den bevorzugten Knoten zu verschieben, werden die folgenden Aktionen ausgeführt:

```
pcs resource relocate run
```

Hinweis: Mit diesem Befehl werden Services verschoben, bei denen keine temporären Ressourcenbeschränkungen mehr vorhanden sind, die sich nicht auf den bevorzugten Nodes befinden.

Versetzen Sie das Cluster in den Wartungsmodus

Verhindern Sie, dass das HA-Cluster versehentlich auf geplante Änderungen in der Umgebung reagiert.

Überblick

Wenn Sie das Cluster in den Wartungsmodus versetzen, werden die gesamte Ressourcenüberwachung deaktiviert und Pacemaker kann nicht mehr Ressourcen im Cluster verschieben oder anderweitig verwalten. Alle Ressourcen werden auf den ursprünglichen Nodes weiterhin ausgeführt, unabhängig davon, ob es eine

temporäre Ausfallbedingung gibt, die den Zugriff auf sie verhindern würde. Dies wird empfohlen/ist u. a.:

- Netzwerkwartung, die vorübergehend Verbindungen zwischen Datei-Nodes und BeeGFS-Diensten unterbrechen kann.
- Block-Node-Upgrades:
- Dateiknoten-Betriebssystem, Kernel oder andere Paketaktualisierungen.

Im Allgemeinen ist der einzige Grund, das Cluster manuell in den Wartungsmodus zu versetzen, um zu verhindern, dass es auf externe Änderungen in der Umgebung reagiert. Wenn für einen einzelnen Node im Cluster die physische Reparatur erforderlich ist, verwenden Sie keinen Wartungsmodus und platzieren Sie den Node einfach gemäß dem oben beschriebenen Verfahren in Standby. Beachten Sie, dass bei der Umleitung von Ansible der Cluster automatisch der Wartungsmodus für die meisten Softwarewartungsarbeiten einschließlich Upgrades und Konfigurationsänderungen durchgeführt wird.

Schritte

So überprüfen Sie, ob das Cluster sich im Wartungsmodus befindet:

```
pcs property show maintenance-mode
```

Dies gibt FALSE zurück, wenn das Cluster ordnungsgemäß ausgeführt wird. Um den Wartungsmodus zu aktivieren, führen Sie folgende Schritte aus:

```
pcs property set maintenance-mode=true
```

Sie können überprüfen, indem Sie den PC-Status ausführen und sicherstellen, dass alle Ressourcen „(nicht verwaltet)“ anzeigen. Um das Cluster aus dem Wartungsmodus zu nehmen, führen Sie folgende Schritte aus:

```
pcs property set maintenance-mode=false
```

Beenden Sie den Cluster und starten Sie den Cluster

Graziös wird das HA-Cluster angehalten und gestartet.

Überblick

In diesem Abschnitt wird beschrieben, wie das BeeGFS-Cluster ordnungsgemäß heruntergefahren und neu gestartet wird. Beispielszenarien, bei denen dies möglicherweise erforderlich ist, sind beispielsweise die elektrische Wartung oder die Migration zwischen Rechenzentren oder Racks.

Schritte

Wenn Sie aus irgendeinem Grund das gesamte BeeGFS-Cluster beenden und alle Dienste herunterfahren müssen, laufen:

```
pcs cluster stop --all
```

Es ist auch möglich, das Cluster auf einzelnen Nodes anzuhalten (wodurch automatisch ein Failover von Services auf einen anderen Node erfolgt). Es wird jedoch empfohlen, den Node zunächst in den Standby-Modus zu versetzen (siehe "[Failover](#)" Abschnitt):

```
pcs cluster stop <HOSTNAME>
```

So starten Sie Cluster Services und Ressourcen auf allen Nodes:

```
pcs cluster start --all
```

Oder starten Sie Services auf einem bestimmten Knoten mit:

```
pcs cluster start <HOSTNAME>
```

An dieser Stelle Lauf `pcs status` Überprüfen Sie, ob die Cluster- und BeeGFS-Services auf allen Nodes gestartet werden und die Services auf den erwarteten Nodes ausgeführt werden.



Abhängig von der Größe des Clusters kann es irgendwann (Sekunden oder Minuten) dauern, bis der gesamte Cluster angehalten ist, oder es wird gestartet in angezeigt `pcs status`. Wenn `pcs cluster <COMMAND>` Hängt länger als fünf Minuten, bevor Sie „Strg+C“ ausführen, um den Befehl abubrechen, melden Sie sich bei jedem Node des Clusters an und verwenden Sie `pcs status` Um zu sehen, ob Cluster-Services (Corosync/Pacemaker) auf diesem Knoten noch ausgeführt werden. Von jedem Node, der das Cluster noch aktiv ist, können Sie überprüfen, welche Ressourcen das Cluster blockieren. Lösen Sie das Problem manuell, und der Befehl sollte entweder abgeschlossen werden oder kann erneut ausgeführt werden, um alle verbleibenden Services zu beenden.

Datei-Nodes ersetzen

Ersetzen eines Dateiknotens, wenn der ursprüngliche Server fehlerhaft ist.

Überblick

Dies bietet einen Überblick über die Schritte, die zum Austausch eines Datei-Nodes im Cluster erforderlich sind. Diese Schritte setzen voraus, dass der Datei-Node aufgrund eines Hardwareproblems ausgefallen ist und dass er durch einen neuen identischen File-Node ersetzt wurde.

Schritte

1. Ersetzen Sie den Datei-Node physisch und stellen Sie alle Kabel auf den Block-Node und das Storage-Netzwerk wieder her.
2. Installieren Sie das Betriebssystem auf dem Dateiknoten neu, einschließlich Hinzufügen von Red hat Subskriptionen.

3. Konfiguration von Management und BMC Networking auf dem Datei-Node
4. Aktualisieren Sie die Ansible-Bestandsaufnahme, wenn sich der Hostname, die IP, die Zuordnung der PCIe-zu-logischen Schnittstelle oder eine weitere Änderung bezüglich des neuen Datei-Nodes ergeben. Im Allgemeinen ist dies nicht erforderlich, wenn der Node durch identische Serverhardware ersetzt wurde und Sie die ursprüngliche Netzwerkkonfiguration verwenden.
 - a. Wenn sich beispielsweise der Hostname geändert hat, erstellen Sie die Bestandsdatei des Node (oder benennen Sie sie um) (`host_vars/<NEW_NODE>.yaml`) Und dann in der Ansible-Bestandsdatei (`inventory.yml`), ersetzen Sie den Namen des alten Knotens durch den neuen Knotennamen:

```
all:
  ...
  children:
    ha_cluster:
      children:
        mgmt:
          hosts:
            node_h1_new:  # Replaced "node_h1" with "node_h1_new"
            node_h2:
```

5. Entfernen Sie den alten Node von einem der anderen Nodes im Cluster: `pcs cluster node remove <HOSTNAME>`.



FAHREN SIE VOR AUSFÜHRUNG DIESES SCHRITTS NICHT FORT.

6. Auf dem Ansible-Steuerungsknoten:
 - a. Entfernen Sie den alten SSH-Schlüssel mit:

```
`ssh-keygen -R <HOSTNAME_OR_IP>`
```

- b. Konfigurieren Sie passwortloses SSH auf den Knoten Ersetzen mit:

```
ssh-copy-id <USER>@<HOSTNAME_OR_IP>
```

7. Führen Sie das Ansible-Playbook erneut aus, um den Node zu konfigurieren und dem Cluster hinzuzufügen:

```
ansible-playbook -i <inventory>.yaml <playbook>.yaml
```

8. An dieser Stelle, Lauf `pcs status` Und überprüfen Sie, ob der ersetzte Node jetzt aufgeführt ist und Services ausführt.

Erweitern oder verkleinern Sie den Cluster

Fügen Sie dem Cluster Bausteine hinzu oder entfernen Sie diese.

Überblick

In diesem Abschnitt werden verschiedene Überlegungen und Optionen dokumentiert, um die Größe Ihres BeeGFS HA-Clusters anzupassen. Normalerweise wird die Cluster-Größe durch Hinzufügen oder Entfernen von Bausteinen angepasst. Bei diesen handelt es sich in der Regel um zwei Datei-Nodes, die als HA-Paar eingerichtet wurden. Bei Bedarf können auch einzelne Datei-Nodes (oder andere Cluster-Nodes) hinzugefügt oder entfernt werden.

Hinzufügen eines Bausteins zum Cluster

Überlegungen

Das erweitern des Clusters durch Hinzufügen weiterer Bausteine ist ein unkomplizierter Prozess. Beachten Sie zunächst die Einschränkungen der minimalen und maximalen Anzahl von Cluster-Nodes in jedem einzelnen HA-Cluster und bestimmen Sie, ob Sie Nodes zum vorhandenen HA-Cluster hinzufügen oder ein neues HA-Cluster erstellen sollten. Normalerweise besteht jeder Baustein aus zwei Datei-Nodes, aber drei Nodes sind die Mindestanzahl an Nodes pro Cluster (um ein Quorum zu schaffen). Zehn davon ist das empfohlene Maximum (getestete). Für erweiterte Szenarien ist es möglich, einen einzelnen „Tiebreaker“ Node hinzuzufügen, auf dem keine BeeGFS-Services ausgeführt werden, wenn ein Cluster mit zwei Nodes implementiert wird. Bitte wenden Sie sich an den NetApp Support, wenn Sie eine solche Implementierung in Betracht ziehen.

Beachten Sie diese Einschränkungen und das erwartete zukünftige Cluster-Wachstum bei Ihrer Entscheidung über das erweitern des Clusters. Wenn Sie beispielsweise einen sechs-Node-Cluster haben und vier weitere Nodes hinzufügen müssen, empfiehlt es sich, nur einen neuen HA-Cluster zu starten.



Denken Sie daran, dass ein einziges BeeGFS-Dateisystem aus mehreren unabhängigen HA-Clustern bestehen kann. Dadurch können Filesysteme weit über die empfohlenen/harten Grenzen der zugrunde liegenden HA-Cluster-Komponenten hinaus skaliert werden.

Schritte

Wenn Sie dem Cluster einen Baustein hinzufügen, müssen Sie die `host_vars` Dateien für jeden der neuen Datei-Nodes und Block-Nodes (E-Series-Arrays) erstellen. Die Namen dieser Hosts müssen dem Bestand hinzugefügt werden, zusammen mit den neuen Ressourcen, die erstellt werden sollen. Die entsprechenden `group_vars` Dateien müssen für jede neue Ressource erstellt werden. "[Nutzung benutzerdefinierter Architekturen](#)" Weitere Informationen finden Sie im Abschnitt.

Nach dem Erstellen der richtigen Dateien müssen alle erforderlichen Dateien die Automatisierung mit dem Befehl erneut ausführen:

```
ansible-playbook -i <inventory>.yaml <playbook>.yaml
```

Entfernen eines Bausteins aus dem Cluster

Beachten Sie bei der Außerbetriebnahme eines Baublocks verschiedene Aspekte, z. B.:

- Welche BeeGFS-Services laufen in diesem Baustein?
- Werden nur die File-Nodes ausgemustert und die Block-Nodes mit neuen Datei-Nodes verbunden?
- Wenn der gesamte Baustein außer Betrieb genommen wird, sollten die Daten in einen neuen Baustein verschoben, in vorhandene Nodes im Cluster verteilt oder auf ein neues BeeGFS Filesystem oder ein anderes Storage-System verschoben werden?
- Kann dies bei einem Ausfall oder ohne Unterbrechung geschehen?
- Ist der Baustein aktiv genutzt oder enthält er in erster Linie Daten, die nicht mehr aktiv sind?

Aufgrund der vielfältigen möglichen Ausgangspunkte und gewünschten Endzustände wenden Sie sich bitte an den NetApp Support, damit wir die optimale Strategie basierend auf Ihrer Umgebung und Ihren Anforderungen identifizieren und implementieren können.

Copyright-Informationen

Copyright © 2024 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFT SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGEND EINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.