

FUJITSU Software PRIMECLUSTER

他社製 PC サーバ使用時の留意事項

Linux

SWPCL-T061-01(00)

第1.1版

2019年12月

変更履歴

版数	作成日	修正箇所	修正概要
1.0	2019/06		初版
1.1	2019/12	第1章 はじめに	対象バージョンを4.5A10から4.5A10以降に変更。 問合せが必須であることを明記。
		1.1 対象製品	バージョンに「以降」を追記。
		1.2 前提条件	IPMIのバージョンをRHEL7以前とRHEL8以降に 場合分けして記載。
		全体	参照マニュアルの名称からバージョンを削除。
		2.2.3 IPMIシャットダウンエージェントの設定	[RHEL7の場合]を[RHEL7以降の場合]に変更。
		2.3 クラスタ環境のチェック	メッセージが出力されるのは4.5A10の場合のみであることを明記。

目次

第1章	はじめに.....	3
1.1	対象製品.....	4
1.2	前提条件.....	4
1.3	PRIMERGY環境との差異.....	5
第2章	導入.....	6
2.1	IPMIシャットダウンエージェントに必要な設定.....	6
2.2	シャットダウン機構の設定.....	7
2.2.1	シャットダウンエージェントの情報確認.....	7
2.2.2	シャットダウンデーモンの設定.....	8
2.2.3	IPMIシャットダウンエージェントの設定.....	9
2.2.4	シャットダウン機構の起動.....	11
2.2.5	クラスタノード強制停止テスト.....	11
2.3	クラスタ環境のチェック.....	12
第3章	運用.....	13
3.1	クラッシュダンプの採取.....	13

第1章 はじめに

本書は、他社製PCサーバ上にPRIMECLUSTER 4.5A10以降を導入する場合の留意事項について説明します。
なお、PRIMECLUSTERを他社製PCサーバ上に導入する場合、必ず事前に下記の窓口までお問い合わせください。

[お問合せ窓口]

高信頼基盤ソフトウェア FUJITSU Software PRIMECLUSTER

<https://www.fujitsu.com/jp/products/software/infrastructure-software/infrastructure-software/primecluster/>

→「本製品のお問い合わせ」

略称

- Red Hat Enterprise Linux をRHELと略しています。

輸出管理規制について

本ドキュメントを輸出または第三者へ提供する場合は、お客様が居住する国および米国輸出管理関連法規等の規制をご確認のうえ、必要な手続きをおとりください。

商標について

Red Hat は米国およびそのほかの国において登録されたRed Hat, Inc. の商標です。

Linux は、Linus Torvalds 氏の登録商標です。

PRIMECLUSTERは、富士通株式会社の登録商標です。

その他各種製品名は、各社の製品名称、商標または登録商標です。

お願い

- 本書を無断で他に転載しないようお願いします。
- 本書は予告なしに変更されることがあります。

著作権表示

All Rights Reserved, Copyright (C) 富士通株式会社 2019

1.1 対象製品

他社製PCサーバ環境で使用できるPRIMECLUSTER製品を以下に示します。

- PRIMECLUSTER Enterprise Edition 4.5A10以降
- PRIMECLUSTER HA Server 4.5A10以降
- PRIMECLUSTER Clustering Base 4.5A10以降
- PRIMECLUSTER Lite Pack 4.5A10以降
- PRIMECLUSTER GL 4.5A10以降
- PRIMECLUSTER GD Snapshot 4.5A10以降
- PRIMECLUSTER Wizard for NAS 4.5A00以降
- PRIMECLUSTER Wizard for NetVault 4.5A10以降
- PRIMECLUSTER Wizard for NetWorker 4.5A10以降
- PRIMECLUSTER Wizard for Oracle 4.5A00以降

1.2 前提条件

PRIMECLUSTERを他社製PCサーバ環境で使用する際のサーバに関する前提条件を説明します。

- 物理環境(物理マシン、または、KVM環境の管理OS上にPRIMECLUSTERを導入)の場合
 - 2WAYサーバであること。
 - 2ノード以上のクラスタシステムを構築する場合、下記のIPMIに準拠したBMC (Baseboard Management Controller)が搭載されていること。
 - ◇ RHEL7以前のOSを使用する場合：IPMI1.5またはIPMI2.0
 - ◇ RHEL8以降のOSを使用する場合：IPMI2.0
- 仮想環境(仮想マシンのゲストOS上だけにPRIMECLUSTERを導入)の場合
前提条件はありません。

注意

-
-
- 2WAYサーバ以外の機種の場合、下記にお問合せください。
高信頼基盤ソフトウェア FUJITSU Software PRIMECLUSTER
<https://www.fujitsu.com/jp/products/software/infrastructure-software/infrastructure-software/primecluster/>
→「本製品のお問い合わせ」
 - 異なるPCIカード間(LANカード間やファイバーチャネルカード間)でI/Oを冗長化することを推奨します。このため、必要な数のPCIスロットが存在するサーバ機種を選定してください。
-
-

1.3 PRIMERGY環境との差異

PRIMECLUSTERを他社製PCサーバ環境で使用する場合、PRIMERGY環境で使用する場合と比較して、以下の差異があります。

- 物理環境(物理マシン、または、KVM環境の管理OS上にPRIMECLUSTERを導入)の場合
 - クラスタ高速切替機能は使用できません。
このため、物理環境のOSパニックやOSハング発生時に調査資料としてクラッシュダンプが採取されないことがあります。

設定できるシャットダウンエージェントは、IPMIシャットダウンエージェントのみです。クラスタシステムを構成するノードの1つに異常が発生しクラスタシステムのノード間のハートビートが失敗した場合、シャットダウン機構は異常が発生したノードの電源を強制的に切断します。ハートビートの失敗要因がパニックであった場合、クラッシュダンプ採取中に強制電源切断し、クラッシュダンプ採取が途中で終了します。
 - かんたん設計構築機能は使用できません。
- 仮想環境(仮想マシンのゲストOS上のみ)にPRIMECLUSTERを導入の場合
差異はありません。

注意

PRIMECLUSTER GLのGS連携方式を使用する場合、事前により下記にお問合せください。

高信頼基盤ソフトウェア FUJITSU Software PRIMECLUSTER

<https://www.fujitsu.com/jp/products/software/infrastructure-software/infrastructure-software/primecluster/>

→「本製品のお問い合わせ」

第2章 導入

PRIMECLUSTERのマニュアルを参照してインストールと構築を行います。

物理環境(物理マシン、または、KVM環境の管理OS上にPRIMECLUSTERを導入)の場合、PRIMERGY RX/TXシリーズの場合と同様に設定します。ただし、一部の設定方法に違いがあります。以下の表に従って設定してください。

PRIMERGY RX/TXシリーズの場合	他社製PCサーバの場合
インストールガイド 第3章 インストール	CLIインストーラ(cluster_install)実行時は、PRIMEQUESTの場合と同じオプションを指定する。
PRIMECLUSTER導入運用手引書 3.1.4 ServerViewによるハードウェア監視の設定	設定不可
PRIMECLUSTER導入運用手引書 3.1.6 クラスタ高速切替機能の設定	本書 2.1 IPMIシャットダウンエージェントに必要な設定
PRIMECLUSTER導入運用手引書 5.1.2 シャットダウン機構の設定	本書 2.2 シャットダウン機構の設定
PRIMECLUSTER導入運用手引書 6.9 クラスタ環境のチェック	本書 2.3 クラスタ環境のチェック

仮想環境(仮想マシンのゲストOS上にのみPRIMECLUSTERを導入)の場合、設定方法はPRIMERGYの仮想環境の場合と同じです。“PRIMECLUSTER導入運用手引書”に従って設定してください。

2.1 IPMIシャットダウンエージェントに必要な設定

BMCの設定を行ってください。

- IPアドレス
- IPMIシャットダウンエージェント用のユーザ作成(*1)

詳細は、各機種のマニュアルを参照してください。

*1) 作成するユーザには管理者権限を付与してください。また、作成するユーザのパスワードは以下の文字以外の7ビットASCII文字で設定してください。

> < " / ¥ = ! ? ; , &

2.2 シャットダウン機構の設定

IPMIシャットダウンエージェントを設定します。

ここでは、IPMIシャットダウンエージェントをシャットダウン機構に設定する方法について説明します。生存優先度については、“PRIMECLUSTER導入運用手引書”の“5.1.2.1 生存優先度”を参照してください。

シャットダウン機構の設定は、以下の流れで行います。

1. シャットダウンエージェントの情報確認
2. シャットダウンデーモンの設定
3. IPMIシャットダウンエージェントの設定
4. シャットダウン機構の起動
5. クラスタノード強制停止テスト

注意

- シャットダウンエージェント設定後は、正しいノードが強制停止できることを確認するため、クラスタノード強制停止テストを実施してください。クラスタノード強制停止テストの詳細については、“PRIMECLUSTER導入運用手引書”の“1.4 テスト”を参照してください。
- SA_ipmi.cfg、rcsd.cfgファイルの内容はすべてのノードで同一にしてください。同一でない場合誤動作します。

2.2.1 シャットダウンエージェントの情報確認

IPMIシャットダウンエージェントの設定に必要な、BMC(Baseboard Management Controller)の以下の設定を確認してください。

- BMCのIPアドレス
- BMCに定義したユーザ
- BMCに定義したユーザのパスワード

また、以下について確認してください。

- BMCに定義したユーザに管理者権限が設定されていること
- BMCに定義したユーザのパスワードが以下の文字以外の7ビット ASCII文字で設定されていること
> < " / ¥ = ! ? ; , &

2.2.2 シャットダウンデーモンの設定

全ノードで/etc/opt/SMAW/SMAWsf/rcsd.cfgを以下のような内容で作成します。
rcsd.cfgファイルはrootユーザで作成し、パーミッションを600にしてください。
また、全ノードで同一内容にしてください。

```
CFName1, weight=weight, admIP=myadmIP: agent=SA_ipmi, timeout=timeout  
CFName2, weight=weight, admIP=myadmIP: agent=SA_ipmi, timeout=timeout
```

CFNameX : クラスタホストのCFノード名を指定します。
weight : SFのノードの重みを指定します。
myadmIP : クラスタホストのシャットダウン機構で使用する管理LANのIPアドレスを指定します。
BMCのIPアドレスではありません。
指定可能なアドレス形式は、IPv4アドレスおよびIPv6アドレスです。
IPv6のリンクローカルアドレスは使用できません。
IPv6アドレスを指定する場合は、角括弧[]で囲んでください。
(例 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
SA_ipmi : IPMIシャットダウンエージェントです。
timeout : IPMIシャットダウンエージェントのタイムアウト時間(秒)を指定します。
25秒を指定します。
シャットダウン機構で使用する管理LANでSTP(スパンニングツリープロトコル)を使用する場合には、STPがツリーを作成する時間と余裕を考慮し、50秒加算してください。この設定により、フェイルオーバーに要する時間は長くなります。

例)

```
node1, weight=1, admIP=10.20.30.100: agent=SA_ipmi, timeout=25  
node2, weight=1, admIP=10.20.30.101: agent=SA_ipmi, timeout=25
```

参考

/etc/opt/SMAW/SMAWsf/rcsd.cfgファイルを作成する場合、下記のファイルを雛形として使用することができます。

ファイル名 : /etc/opt/SMAW/SMAWsf/rcsd.cfg.template

2.2.3 IPMIシャットダウンエージェントの設定

1. IPMIサービスの起動

[RHEL6の場合]

全ノードで以下のコマンドを実行し、IPMIサービスの起動状態を確認します。

```
# /sbin/service ipmi status
ipmi_msghandler module in kernel.
ipmi_si module in kernel.
ipmi_devintf module not loaded.
/dev/ipmi0 does not exist.
```

"/dev/ipmi0 does not exist."の場合は、以下のコマンドを実行します。

"/dev/ipmi0 exists."の場合は、以下のコマンド実行は不要です。

```
# /sbin/service ipmi start
Starting ipmi drivers: [ OK ]
```

[RHEL7以降の場合]

全ノードで以下のコマンドを実行し、IPMIサービスの起動状態を確認します。

```
# /usr/bin/systemctl status ipmi.service
ipmi.service - IPMI Driver
   Loaded: loaded (/usr/lib/systemd/system/ipmi.service; disabled)
   Active: inactive (dead)
```

"Active:"の項目が "inactive"の場合は、以下のコマンドを実行します。

"Active:"の項目が "active"の場合は、以下のコマンド実行は不要です。

```
# /usr/bin/systemctl start ipmi.service
```

2. IPMIサービスの起動設定

[RHEL6の場合]

起動時にIPMIサービスを読み込むように、全ノードで以下のコマンドを実行します。

```
# /sbin/chkconfig --level 2345 ipmi on
```

[RHEL7以降の場合]

全ノードで、現在のIPMIサービスが有効化されているかを確認します。

```
# /usr/bin/systemctl list-unit-files --type=service | grep ipmi.service
ipmi.service disabled
```

上記で "disabled" と表示された場合は、以下のコマンドを実行します。

"enabled" と表示された場合は、以下のコマンド実行は不要です。

```
# /usr/bin/systemctl enable ipmi.service
```

3. パスワードの暗号化

sfcipherコマンドを実行し、シャットダウン機構用のユーザのパスワードを暗号化します。

例) BMC設定時に定義したパスワードが"bmcpwd1\$"の場合

```
# sfcipher -c
Enter User's Password: ←bmcpwd1$ を入力
Re-enter User's Password: ←bmcpwd1$ を入力
oGYN60R2t31qrdL51IPKBg==
```

注)パスワードとして定義した記号の前に'¥'の記載は不要です。

sfcipherコマンドの使用法については、sfcipherのマニュアルページを参照してください。

注意

BMC設定時に定義したパスワードには、7ビット ASCIIコードの文字が使用可能です。
その中で以下の記号は問題が発生する可能性があるため、使用しないようにしてください。

> < " / ¥ = ! ? ; , &

4. シャットダウンエージェントの設定

全ノードで/etc/opt/SMAW/SMAWsf/SA_ipmi.cfgを以下のような内容で作成します。

SA_ipmi.cfgファイルはrootユーザで作成し、パーミッションを600にしてください。

また、全ノードで同一内容にしてください。

○ IPv4アドレスの場合

```
CFName1 ip-address:user:passwd {cycle | leave-off} [IPMI2.0]
CFName2 ip-address:user:passwd {cycle | leave-off} [IPMI2.0]
```

○ IPv6アドレスの場合

```
CFName1 [ip-address]:user:passwd {cycle | leave-off} [IPMI2.0]
CFName2 [ip-address]:user:passwd {cycle | leave-off} [IPMI2.0]
```

CFNameX : クラスタホストのCFノード名を指定します。
ip-address : CFノード名が*CFNameX*であるサーバのBMCのIPアドレスを指定します。
指定可能なアドレス形式は、IPv4アドレスおよびIPv6アドレスです。
IPv6のリンクローカルアドレスは使用できません。
IPv6アドレスを指定する場合は角括弧[]で囲んでください。
(例 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
user : BMC設定時に定義したユーザを指定します。
passwd : BMC設定時に定義したパスワードです。
3.で暗号化したパスワードを指定します。
cycle : ノード強制停止後、再起動します。
leave-off : ノード強制停止後、電源切断します。
IPMI2.0 : BMCがIPMI2.0をサポートしている場合は指定します。

例1)

node1のBMCのIPアドレスが10.20.30.50、node2のBMCのIPアドレスが10.20.30.51、BMCがIPMI1.5のみサポートしている場合

```
node1 10.20.30.50:root:oGYN60R2t31qrdL51IPKBg== cycle
node2 10.20.30.51:root:oGYN60R2t31qrdL51IPKBg== cycle
```

例2)

node1のBMCのIPアドレスが10.20.30.50、node2のBMCのIPアドレスが10.20.30.51、BMCがIPMI2.0をサポートしている場合

```
node1 10.20.30.50:root:oGYN60R2t31qrdL51IPKBg== cycle IPMI2.0
node2 10.20.30.51:root:oGYN60R2t31qrdL51IPKBg== cycle IPMI2.0
```

例3)

node1のBMCのIPアドレスが1080:2090:30a0:40b0:50c0:60d0:70e0:80f0、node2のBMCのIPアドレスが1080:2090:30a0:40b0:50c0:60d0:70e0:80f1、BMCがIPMI1.5のみサポートしている場合

```
node1 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0]:root:oGYN60R2t31qrdL51PKBg== cycle
node2 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f1]:root:oGYN60R2t31qrdL51PKBg== cycle
```

例4)

node1のBMCのIPアドレスが1080:2090:30a0:40b0:50c0:60d0:70e0:80f0、node2のBMCのIPアドレスが1080:2090:30a0:40b0:50c0:60d0:70e0:80f1、BMCがIPMI2.0をサポートしている場合

```
node1 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0]:root:oGYN60R2t31qrdL51PKBg== leave-off IPMI2.0
node2 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f1]:root:oGYN60R2t31qrdL51PKBg== leave-off IPMI2.0
```

参考

/etc/opt/SMAW/SMAWsf/SA_ipmi.cfgファイルを作成する場合、下記のファイルを雛形として使用することができます。

ファイル名 : /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg.template

2.2.4 シャットダウン機構の起動

全ノードでシャットダウン機構を起動または再起動します。

詳細については、“PRIMECLUSTER導入運用手引書”の“5.1.2.3.6 シャットダウン機構の起動”を参照してください。

2.2.5 クラスタノード強制停止テスト

シャットダウン機構の設定後に、正しいノードが強制停止できることを確認するため、クラスタ強制停止テストを実施します。

クラスタノード強制停止の詳細については、“PRIMECLUSTER導入運用手引書”の“1.4 テスト”を参照してください。

2.3 クラスタ環境のチェック

PRIMECLUSTER環境の設定後、PRIMECLUSTER環境チェックツールを用いて設定ミスがないことを確認します。詳細については、“PRIMECLUSTER導入運用手引書”の“6.9 クラスタ環境のチェック”を参照してください。

PRIMECLUSTER 4.5A10の場合、`clchkcluster`コマンド実行時に以下のメッセージが出力されることがありますが、問題はありません。無視してください。

```
[警告]
  シャットダウンエージェントが冗長化されていません。
[対処]
  クラスタ内のサーバに異常が発生した際に、確実にサーバをシャットダウンできるよう、シャットダウンエージェントを冗長化してください。“PRIMECLUSTER導入運用手引書”の“シャットダウン機構の設定”を参照し、当ハードウェア機種で利用可能なエージェントを登録してください。

SA configuration: Passed
```

また、PRIMECLUSTER 4.5A10の場合、以下のメッセージが出力されることがあります。以下のメッセージが出力された場合、`sdtool -s`コマンドを実行してください。AgentにSA_ipmiのみが表示される場合、問題はありません。AgentにSA_ipmi以外が表示される場合、“2.2 シャットダウン機構の設定”を参照してもう一度設定してください。

```
[問題]
  /etc/opt/SMAW/SMAWsf/rcsd.cfgファイルが正しく設定されていません。
[対処]
  “PRIMECLUSTER導入運用手引書”の“シャットダウン機構の設定”を参照してもう一度設定してください。

LKCD SA      : Failed
```

第3章 運用

他社製PCサーバ環境に構築されたPRIMECLUSTERシステムの運用時における注意点について説明します。

3.1 クラッシュダンプの採取

物理環境(物理マシン、または、KVM環境の管理OS上にPRIMECLUSTERを導入)のクラスタシステムにおいて、手動でクラッシュダンプを採取する場合、以下の手順で採取してください。この手順に従っていない場合、クラッシュダンプ採取中にサーバの電源が切断され、クラッシュダンプの採取が途中で終了します。

1. シャットダウン機構の停止

全ノードで以下のコマンドを実行し、シャットダウン機構を停止します。

```
# sdttool -e
```

2. クラッシュダンプの採取

手動でクラッシュダンプを採取します。

3. LEFTCLUSTER状態の確認

任意の1ノードで以下のコマンドを実行し、クラッシュダンプを採取しているノードがLEFTCLUSTER状態になっていることを確認します。LEFTCLUSTER状態になっていない場合、10秒ほど待ってから再度確認します。

```
# cftool -n
```

注意

CFハートビートのタイムアウト検出時間を変更している場合(/etc/default/cluster.configファイルにCLUSTER_TIMEOUTを設定している場合)、ハートビートタイムアウト時間だけ待ってから確認してください。

4. LEFTCLUSTER状態からの回復

“PRIMECLUSTER Cluster Foundation導入運用手引書”の“5.2 LEFTCLUSTERからの回復”を参照し、LEFTCLUSTER状態から回復します。

5. シャットダウン機構の起動

クラッシュダンプを採取しているノードを除いたすべてのノードで以下のコマンドを実行し、シャットダウン機構を起動します。

```
# sdttool -b
```