



# 3 ESXの操作と設定

---

ESX上での本装置固有のセットアップや操作について説明します。CPU/IOモジュールにはプロセッサ機能とIO機能の部分が存在しています。本章では各種ユーティリティでは前者をCPUモジュール、後者をPCIモジュールと記載します。

# ディスクの操作

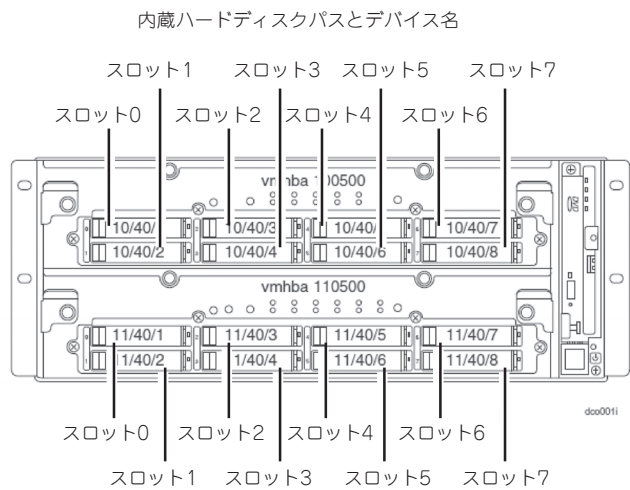
Express5800/ftサーバでは、Software-RAIDにより内蔵ディスクの二重化を行い、データ保全を図ります。



- ESXを含むディスクにはシステムパーティションのみ作成することを推奨します。
- ESXを含むディスクにVMFSデータストアを作成している場合は、ESXの再インストール時にはディスクの全領域がクリアされますので注意してください。

## 操作可能なディスク構成について

Express5800/ftサーバではすべての内蔵ディスクを二重化する必要があります。  
表1で対応している各スロットの内蔵ディスクは、Software-RAIDを用いて冗長化構成を構築します。



ミラーリング処理に対応するスロット

対応するスロット	
スロット0(10/40/1)	↔ スロット0(11/40/1)
スロット1(10/40/2)	↔ スロット1(11/40/2)
スロット2(10/40/3)	↔ スロット2(11/40/3)
スロット3(10/40/4)	↔ スロット3(11/40/4)
スロット4(10/40/5)	↔ スロット4(11/40/5)
スロット5(10/40/6)	↔ スロット5(11/40/6)
スロット6(10/40/7)	↔ スロット6(11/40/7)
スロット7(10/40/8)	↔ スロット7(11/40/8)

内蔵ディスクの操作には、カーネルデバイス名を使用します。カーネルデバイス名は、ブート時やディスクの挿入によりシステムがディスクを検出することによって決定されます。カーネルデバイス名は「vmhba $nn$ 0500:C0:T $x$ :L0」のように表されます。「vmhba $nn$ 0500」の $nn$ はPCIモジュール(10、11)を表し、「T $x$ 」の $x$ はターゲット番号を表し、検出順に0以上が割り当てられます。

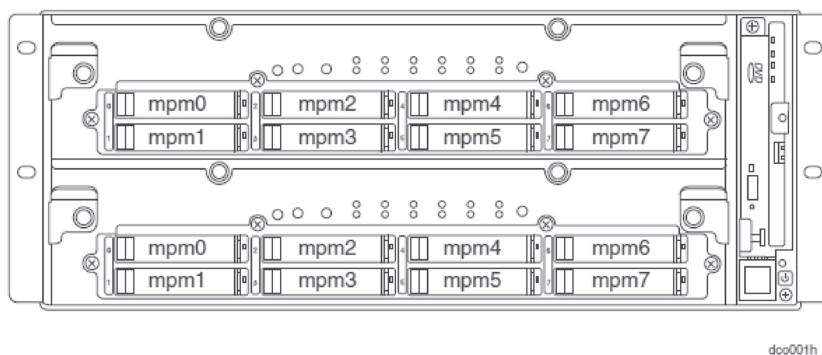
スロットに対応したカーネルデバイス名を確かめるためには/opt/ft/bin/ftsmaintコマンドを使用します。

PCIモジュール0(10)のスロット2に挿入されているディスクのカーネルデバイス名を知りたい場合、次のコマンドを実行してください。以下の例では、カーネルデバイス名はvmhba100500:C0:T1:L0となります。

```
# /opt/ft/bin/ftsmaint ls 10/40/2
H/W Path           : 10/40/2
Description        : Disk Drive
State              : ONLINE
Op State           : DUPLEX
Reason             : NONE
Modelx             : SEAGATE: ST973452SS
Firmware Rev       : 0005
Serial #           : 3TA02A9M00009929YFLA
Device Name        : disk_b
Udev Device Names   : -
Kernel Device Names : vmhba100500:C0:T1:L0
```

冗長化構成の構築は、mpmadmコマンドを使用しますが、その際RAIDデバイス名を下記の図で示すようにmpm $n$ ( $n$ は0～7)で表します。

内蔵ハードディスクのRAIDデバイス名





重要

- 各RAIDデバイスの状態が「resync」「recover」「check」または「repair」の状態の間は、ディスクの抜き差しやシステム停止、再起動をしないでください。RAIDデバイスの状態表示が消え、各ディスク状態が「in\_sync」になるまでしばらくお待ちください。RAIDデバイスの状態は後述のmpmadmコマンドで確認することができます。
- 弊社で指定しているハードディスクドライブのみを使用してください。サードパーティのハードディスクドライブなどを取り付けると、ハードディスクドライブだけでなく本装置が故障する恐れがあります。冗長化構成を構築するハードディスクドライブは、同じモデルを2台1組でお買い求めください。本装置に最適なハードディスクドライブについては、お買い求めの販売店にお問い合わせください。

ディスクの状態を確認するには/opt/ft/sbin/mpmadm -l コマンドを使用します。  
以下は/opt/ft/sbin/mpmadm -l コマンドを実行したときの表示例です。

```
# cd /opt/ft/sbin/  
# ./mpmadm -l  
mpm0 [2/2]  
L_ vmhba110500:C0:T0:L0          [ in_sync ]  
L_ vmhba100500:C0:T0:L0          [ in_sync ]
```

# ハードディスクドライブの交換について

ハードディスクドライブの故障による交換は以下の手順で行います。ハードディスクドライブの交換はCPU/IOモジュール0、1の電源がONの状態で行います。

## 障害ディスクの特定方法

障害が発生しているハードディスクドライブの特定方法を説明します。



この操作は、rootユーザで実行しなければなりません。

1. `mpmadm -l` を実行します。
2. 表示された情報から障害ディスクを確認します。

以下はPCIモジュール1のスロット0に挿入されている内蔵ディスクに障害が発生している例です。

```
# cd /opt/ft/sbin/  
# ./mpmadm -l  
mpm0 [2/2]  
  _ vmhba110500:C0:T0:L0      [ faulty ]  
  _ vmhba100500:C0:T0:L0      [ in_sync ]
```

この場合、PCIモジュール1のスロット0を`/opt/ft/bin/ftsmaint ls`で確認すると次のようになっています。

```
# cd /opt/ft/bin/  
# ./ftsmaint ls 11/40/1  
  
H/W Path           : 11/40/1  
Description         : Disk Drive  
State               : BROKEN  
Op State            : SHOT  
Reason              : NONE  
以下省略  
...  
...
```

## 冗長構成の復旧

問題が発生した内蔵ディスクを交換して、再度、二重化する手順について説明します。



- この操作は、rootユーザで実行しなければなりません。
- 交換したディスクが冗長化構成へ復旧している間、再構築された各RAIDデバイスが「recover」状態になっていた場合はシステムの停止や再起動を行わないでください。それらの状態表示が消え、各ディスク状態が「in\_sync」になるまでしばらくお待ちください(ディスク容量に比例して長くなります)。

1. `mpmadm -f`、`mpmadm -r` にRAIDデバイス名とディスクを表すカーネルデバイス名を指定し実行することで、冗長構成から切り離します。
2. システムからディスクを抜き取り、新しいディスクを挿入します。
3. システムが自動的に再同期化を開始するまで、5分程度待ってください。
4. 手順3にて自動的に再起動が行われない場合は、`mpmadm -a` にRAIDデバイス名とディスクを表すカーネルデバイス名を指定し実行することで、冗長構成への復旧を行うことができます。

以下はPCIモジュール1のスロット0に挿入されている内蔵ディスクの切り離しから復旧までの例です。

```
# cd /opt/ft/sbin/

# ./mpmadm -l
mpm0 [2/2]
l_ vmhba110500:C0:T0:L0      [ in_sync ]
l_ vmhba100500:C0:T0:L0      [ in_sync ]

<<<lO モジュール1 のスロット0 のカーネルデバイス名を指定してディスクを冗長構成から切り離します>>>
# ./mpmadm -f mpm0 vmhba110500:C0:T0:L0
Successfully set disk vmhba110500:C0:T0:L0 faulty in volume mpm0.

# ./mpmadm -r mpm0 vmhba110500:C0:T0:L0
Successfully removed disk vmhba110500:C0:T0:L0 from volume mpm0.
```

<<<指定したディスクがRAID から切り離されていることを確認します。>>>

```
# ./mpmadm -l  
mpm0 [1/2]  
  _ vmhba100500:C0:T0:L0          [ in_sync ]
```

Unused disks:

- vmhba110500:C0:T0:L0

<<<指定したディスクが正常に冗長構成から切り離されている(Unused disks になっている)ことを確認したら、ディスクの交換を行ってください。

新しいディスクを挿入後、5 分程度はシステムが自動的に再同期化を開始するのを待ってください。

自動的に再同期化しなかった場合は、ディスクの復旧を手動で行ってください。>>>

```
# ./mpmadm -l  
mpm0 [1/2]  
  _ vmhba100500:C0:T0:L0          [ in_sync ]
```

<<<新しいディスクがシステムに認識されていることを確認します。>>>

```
# ./mpmadm -l  
mpm0 [1/2]  
  _ vmhba100500:C0:T0:L0          [ in_sync ]
```

Unused disks:

- vmhba110500:C0:T1:L0

<<<システムが新しいディスクを自動的に再同期化しなかった場合に、本コマンドを実行することにより冗長構成へ復旧を手動で行います。>>>

```
# ./mpmadm -a mpm0 vmhba110500:C0:T1:L0  
Successfully added disk vmhba110500:C0:T1:L0 to volume mpm0.
```

<<<再同期化が開始されていることを確認します。>>>

```
# ./mpmadm -l
```

```
mpm0 [2/2] recover=0.6% (455808/71484736) finish=18.1min (65115K/s)
```

```
l_ vmhba110500:C0:T1:L0          [ syncing ]
```

```
l_ vmhba100500:C0:T0:L0          [ in_sync ]
```

<<<RAID デバイス(mpm)に冗長構成への復旧「recover」を行います。しばらくした後、再度下記コマンドにて状態を確認します。「syncing」は復旧中であることを示します。>>>

```
# ./mpmadm -l
```

```
mpm0 [2/2]
```

```
l_ vmhba110500:C0:T1:L0          [ in_sync ]
```

```
l_ vmhba100500:C0:T0:L0          [ in_sync ]
```

<<<RAID デバイス(mpm)に「recover」が表示されておらず、交換したディスクの状態が「in\_sync」になっていれば冗長構成は正常に復旧しており、ディスクの交換作業は完了です。>>>



# ネットワークの二重化

ここでは操作可能なネットワーク構成について説明します。

## 機能概要

LANの二重化は、同一仮想スイッチに複数のアダプタをバインドすること(NICチーミング)で実現しています。1つのアップリンクアダプタが故障した場合、直ちに別のアダプタに切り替え運用を継続させます。

## 操作可能なネットワーク構成について

Express5800/ftサーバでは、ネットワークインタフェース名は以下のとおりの命名規則となります。ネットワークの二重化はCPU/IOモジュール0、CPU/IOモジュール1の同じPCIスロットのネットワークインタフェースを対として構成されます(1つのスイッチに対して2つのネットワークインタフェースをバインドします)。ネットワークインタフェースの設定および確認はVMware vSphereから行ってください。

具体的な設定方法については、VMware vSphereドキュメント「ESX構成ガイド」を参照してください。

ただし、NICチーミングの構成でロードバランシングを「IPハッシュに基づいたルート」に設定した場合、故障などでモジュールのフェールオーバーが発生した時にネットワークが最大30秒間不通となる場合があります。クライアントやアプリケーションの設定などを十分に検討した上でご使用ください。

ロードバランシングを「発信元MACハッシュに基づいたルート」または「発信元仮想ポートIDに基づいたルート」に設定する場合、VMWareの仕様上フェールオーバーが発生した時にネットワークが一時的に不通となる場合があります。フェールオーバーの検出から「ピーコンの検知」を選択してご使用ください。

PCIスロットとネットワーク名

PCIスロット	ポート	CPU/IOモジュール0	CPU/IOモジュール1
On Board	#1	vmnic100600	vmnic110600
	#2	vmnic100601	vmnic110601
PCI-e slot 1	#1	vmnic100100	vmnic110100
	#2	vmnic100101	vmnic110101
PCI-e slot 2	#1	vmnic100200	vmnic110200
	#2	vmnic100201	vmnic110201
PCI-e slot 3	#1	vmnic100300	vmnic110300
	#2	vmnic100301	vmnic110301
PCI-e slot 4	#1	vmnic100400	vmnic110400
	#2	vmnic100401	vmnic110401

# モジュールの二重化動作確認方法

システム導入時や再インストール時などに、システムが正しく動作することを確認する方法について解説します。



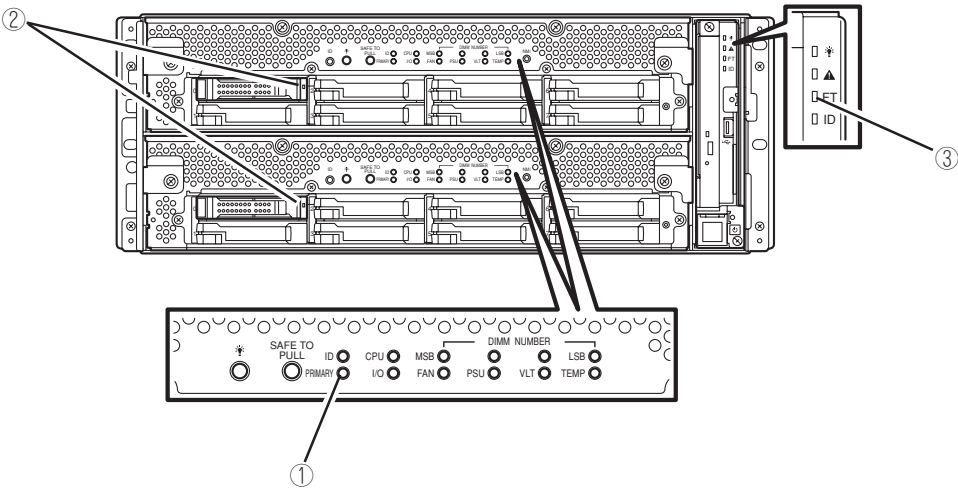
## ヒント

CPU/IOモジュールにはプロセッサ機能とIO機能部分が存在しそれぞれの部分について監視、管理しています。ここでは前者をCPUモジュール、後者をPCIモジュールと記載します。

## PCIモジュールの起動停止評価

プライマリのCPU/IOモジュールを停止させても、フェールオーバによりシステムが継続して稼動することを確認する方法について解説します。

- 1. プライマリ側のCPU/IOモジュールがどれであることを確認する。  
PRIMARY(プライマリ)ランプが点灯しているCPU/IOモジュールがプライマリとなります。
- 2. CPU/IOモジュールが二重化していることを確認する。  
CPU/IOモジュールが二重化できているかどうかは、システムFTランプで確認することができます。



[ PCIモジュールが二重化状態時のステータスランプ状態 ]

ランプ		プライマリ	セカンダリ
1	プライマリランプ	緑点灯	—
2	DISK ACCESSランプ	緑点灯	緑点灯
ランプ		システム	
3	システムFTランプ	緑点灯	

\* 表中の各番号は上図の番号に対応しています。  
②のDISK ACCESSランプは、ハードディスクドライブへアクセスがあった時に点灯します。

3. ftsmaintコマンドでプライマリ側のPCIモジュールの動作を停止する。

PCIモジュール0がプライマリの場合、次のコマンドを実行します。

```
# cd /opt/ft/bin  
# ./ftsmaint bringdown 10 (*)
```

※[(プライマリ側の)PCIモジュール]については、PCIモジュール0がプライマリである場合は[PCIモジュール(ID:10)]を、PCIモジュール1がプライマリである場合は[PCIモジュール(ID:11)]を選択します。プライマリ側のPCIモジュールを停止すると、フェールオーバーが発生し、それまでセカンダリ側であったPCIモジュールへプライマリが変更されます。

プライマリのPCIモジュールを停止し、フェールオーバーが発生すると、PCIモジュールのステータスランプが以下のように変化する。

【ステータスランプ状態】

ランプ		セカンダリ*	プライマリ*
1	プライマリランプ	—	緑点灯
2	DISK ACCESSランプ	—	アンバー点滅または緑点滅 (DISKアクセス時に緑点灯)
ランプ		システム	
3	システムFTランプ	—	

\* フェールオーバー後のプライマリ、セカンダリを示しています。

4. 手順3で停止したPCIモジュールを起動する。

PCIモジュール0が停止している場合、次のコマンドを実行することで、PCIモジュールを起動します。

```
# cd /opt/ft/bin  
# ./ftsmaint bringup 10
```

PCIモジュールが起動されると、PCIモジュールの診断、PCIモジュールの二重化が行われます。  
PCIモジュールのステータスランプは以下のように遷移します。

**[ステータスランプ状態]**

PCIモジュール起動直後から診断が完了するまで

ランプ		セカンダリ	プライマリ
1	プライマリランプ	—	緑点灯
2	DISK ACCESSランプ	—	アンバー点滅または緑点滅 (DISKアクセス時に緑点灯)
ランプ		システム	
3	システムFTランプ	—	



PCIモジュールの診断が完了し、ディスクの二重化が開始されたとき

ランプ		セカンダリ	プライマリ
1	プライマリランプ	—	緑点灯
2	DISK ACCESSランプ	アンバー点滅または緑点滅 (DISKアクセス時に緑点灯)	アンバー点滅または緑点滅 (DISKアクセス時に緑点灯)
ランプ		システム	
3	システムFTランプ	—	



ディスクの二重化が完了し、PCIモジュールが二重化されたとき

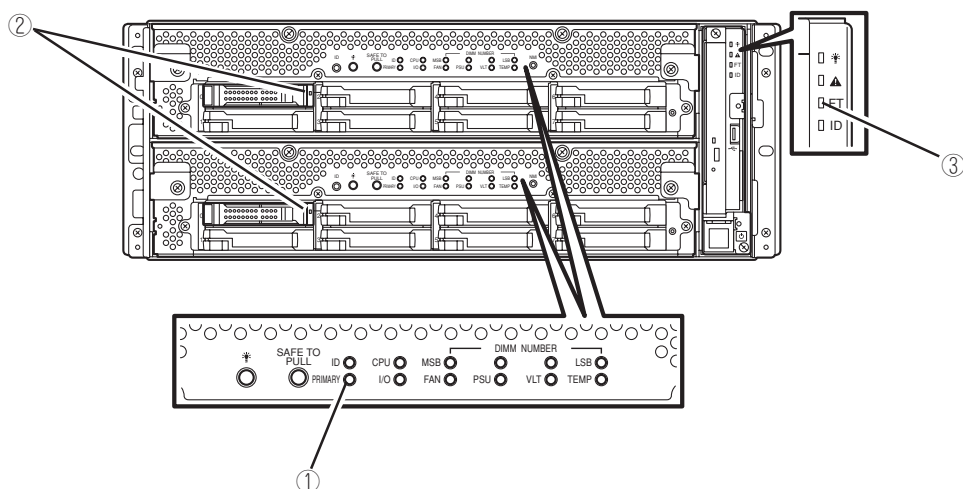
ランプ		セカンダリ	プライマリ
1	プライマリランプ	—	緑点灯
2	DISK ACCESSランプ	緑点灯	緑点灯
ランプ		システム	
3	システムFTランプ	緑点灯	

# CPUモジュールの起動停止評価

一方のCPUモジュールを停止させても、システムが継続して稼動することを確認する方法について解説します。

1. CPUモジュールが二重化していることを確認する。

CPUモジュールが二重化しているかどうかは、CPUモジュールのステータスランプで確認することができます。



【CPUモジュールが二重化状態時のステータスランプ状態】

ランプ		CPU/IOモジュール0 (稼働中)	CPU/IOモジュール1 (稼働中)
1	プライマリランプ	緑点灯	—
2	DISK ACCESSランプ	緑点灯	緑点灯
ランプ		システム	
3	システムFTランプ	緑点灯	

2. ftsmaintコマンドで、取り外す方のCPUモジュールの動作を停止する。

CPUモジュール0を停止させたい場合、次のコマンドを実行します。

```
# cd /opt/ft/bin
# ./ftsmaint bringdown 0 (*)
```

CPUモジュールを停止すると、以下のようにステータスランプが変化します。これはCPUモジュールが片系運転になったことを示しています。

**【ステータスランプ状態】**

ランプ		CPU/IOモジュール0 (停止) *	CPU/IOモジュール1 (稼働中)
1	プライマリランプ	緑点灯	—
2	DISK ACCESSランプ	緑点灯	緑点灯
ランプ		システム	
3	システムFTランプ	—	

\* ここでは例として、CPU/IOモジュール0側が停止した場合を示しています。

**3. 停止したCPUモジュールを起動する。**

手順2で停止したCPUモジュールは、次のコマンドを実行することで、起動します。

```
# cd /opt/ft/bin
# ./ftsmaint bringup 0
```

CPUモジュールが起動されると、[ハードウェアの診断]→[メモリの同期(メモリコピー)]→[二重化完了]へと動作が移行します。

メモリの同期中はメモリコピーを行うためにシステムが一時的に停止します。

**【二重化完了後のステータスランプの状態】**

ランプ		CPU/IOモジュール0 (稼働中)	CPU/IOモジュール1 (稼働中)
1	プライマリランプ	緑点灯	—
2	DISK ACCESSランプ	緑点灯	緑点灯
ランプ		システム	
3	システムFTランプ	緑点灯	

**重要**

二重化完了後、メモリ状態のチェックが行われます。

この処理が完了するまで、次のPCIおよびCPUモジュールの起動停止評価は実施しないでください。処理が完了すると以下のイベントログが出力されます。

```
kernel: EVLOG: INFORMATION - Memory consistency check has completed
memory scan.
```