

DSASのここんとこ

～ ネットワークブート編 ～

第3回 KLab 勉強会

<http://dsas.blog.klab.org/>



KLab

2007年11月2日

KLab 株式会社
Kラボラトリー
安井 真伸

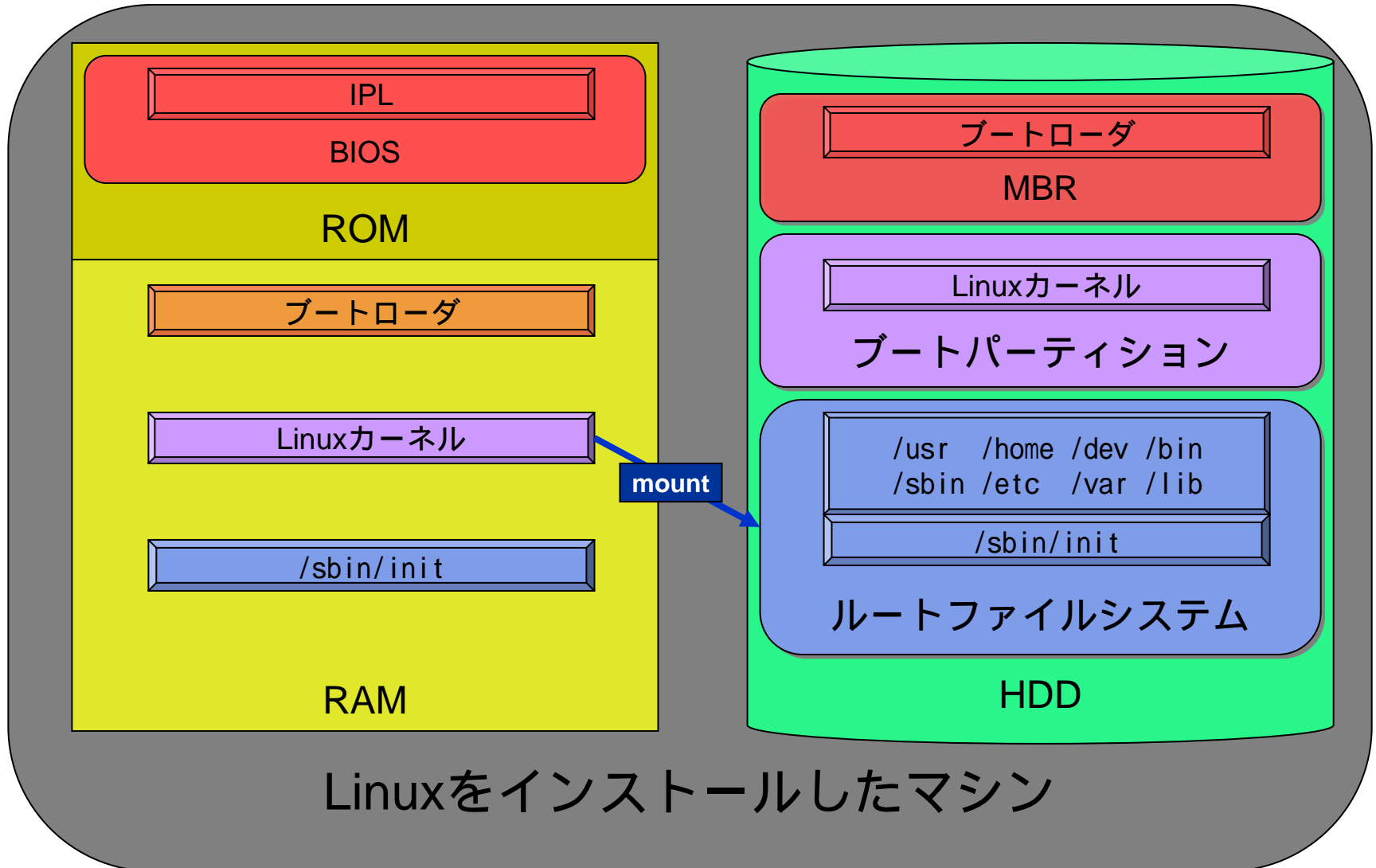


アジェンダ

- ネットワークブートってどんなもの？
 - ブートシーケンス
 - 構築に必要なもの
- 具体的な構成例
 - NFSを利用した構成
 - HDDを利用した構成
 - RamDiskを利用した構成
- DSASはどうやってるの？
 - 機能による構成の違い
 - ネットワークブートの効能
- 運用テクニックの紹介
 - 運用上の懸念点
 - DSASの運用方法

ネットワークブートの話の前に、

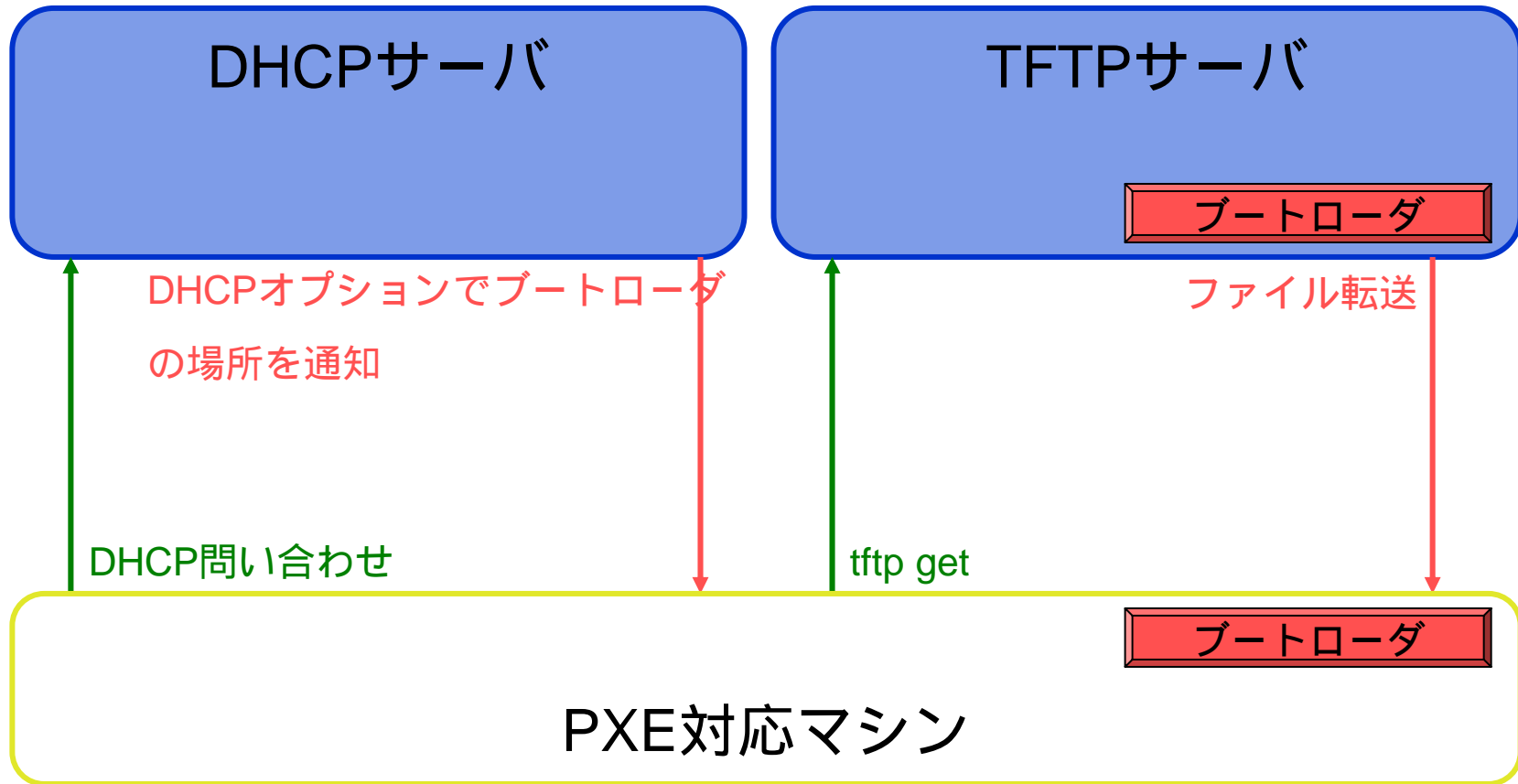
通常のブートシーケンス



ネットワークブートってどんなもの？

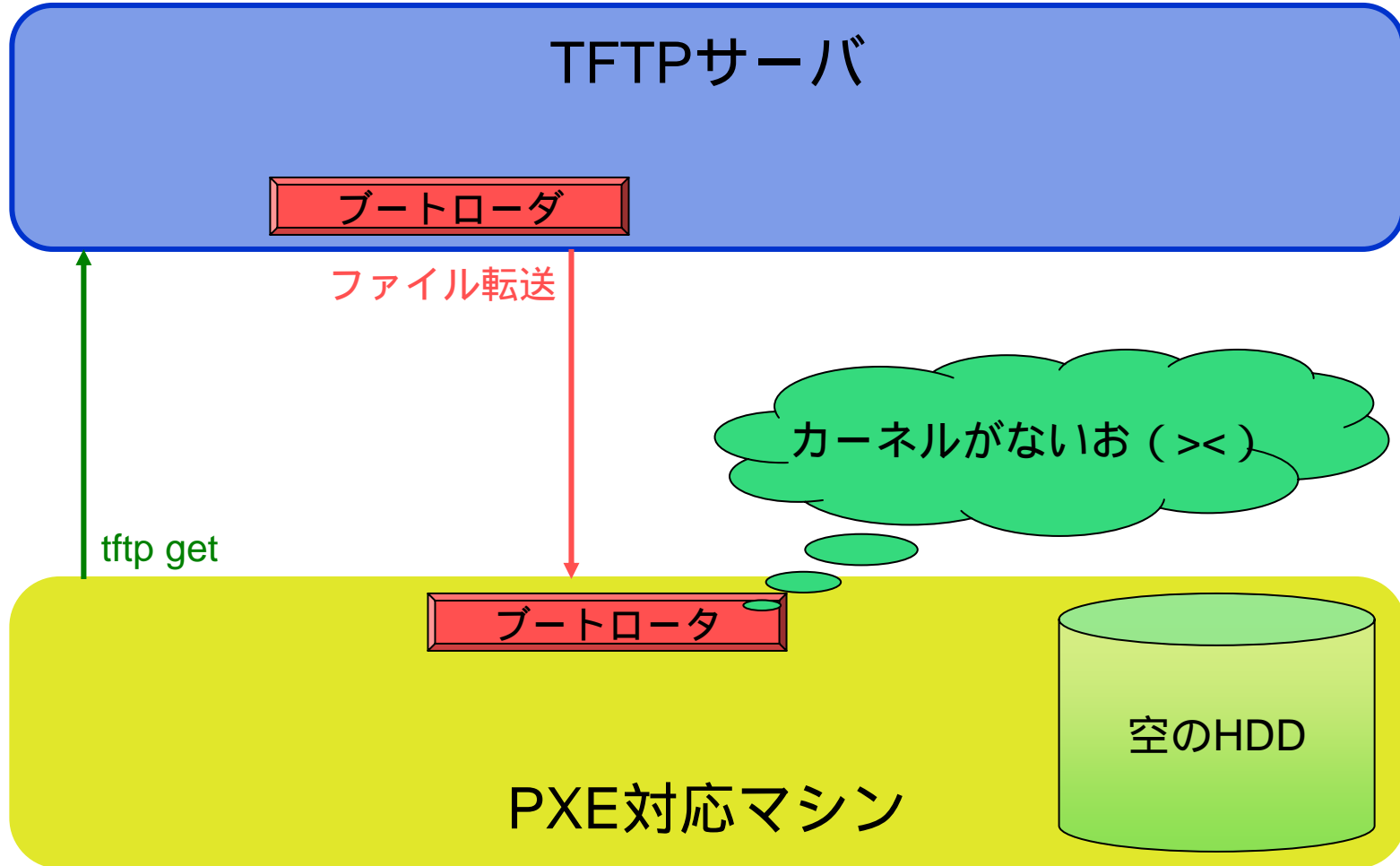
ネットワークブートとは

必要なファイルをネットワーク経由で取得して起動する仕組みです

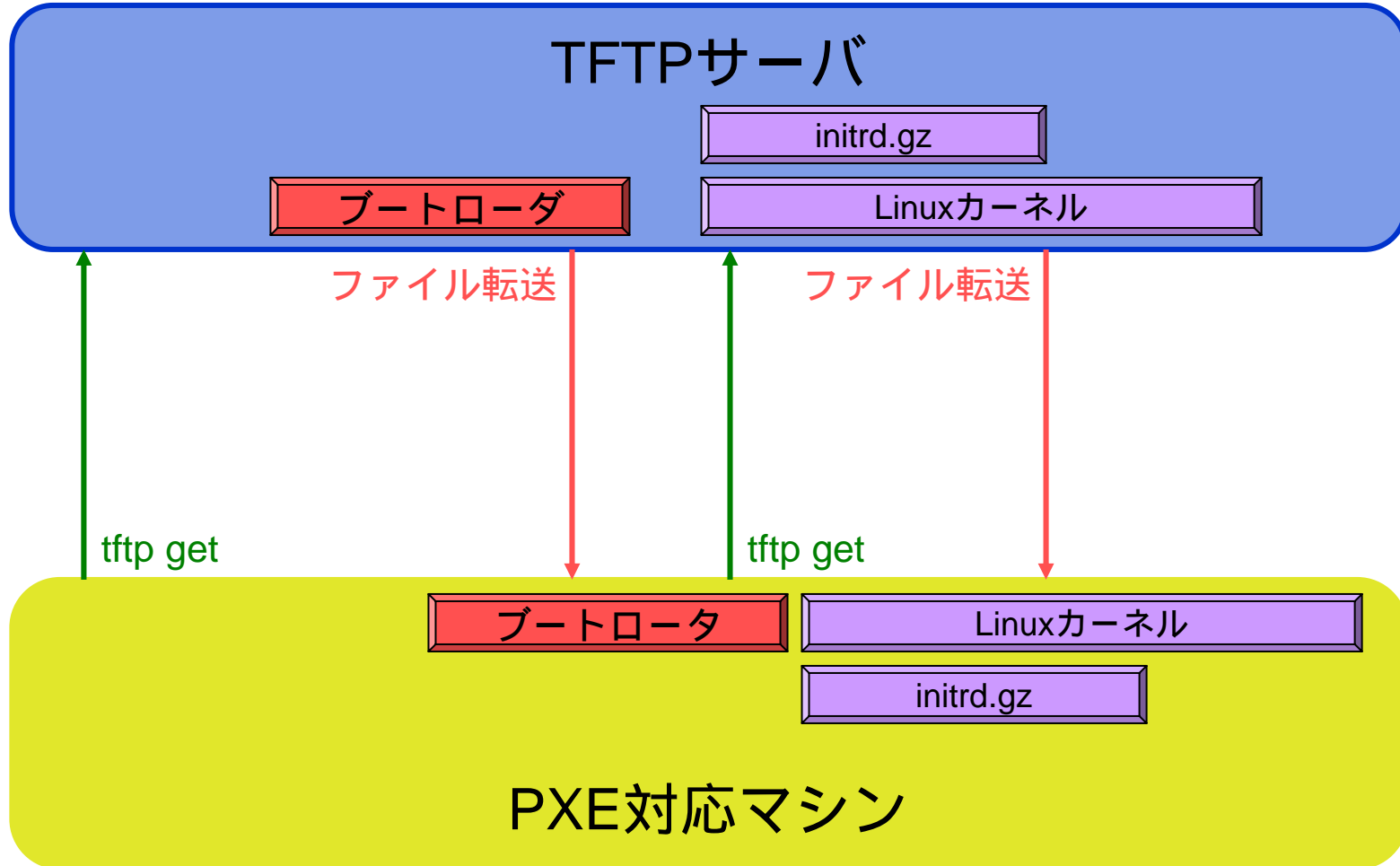


ここまでの処理はハードウェア (NIC) がやってくれます

普通のブートローダだと



PXE対応のブートローダなら





initrd.gzってなに？

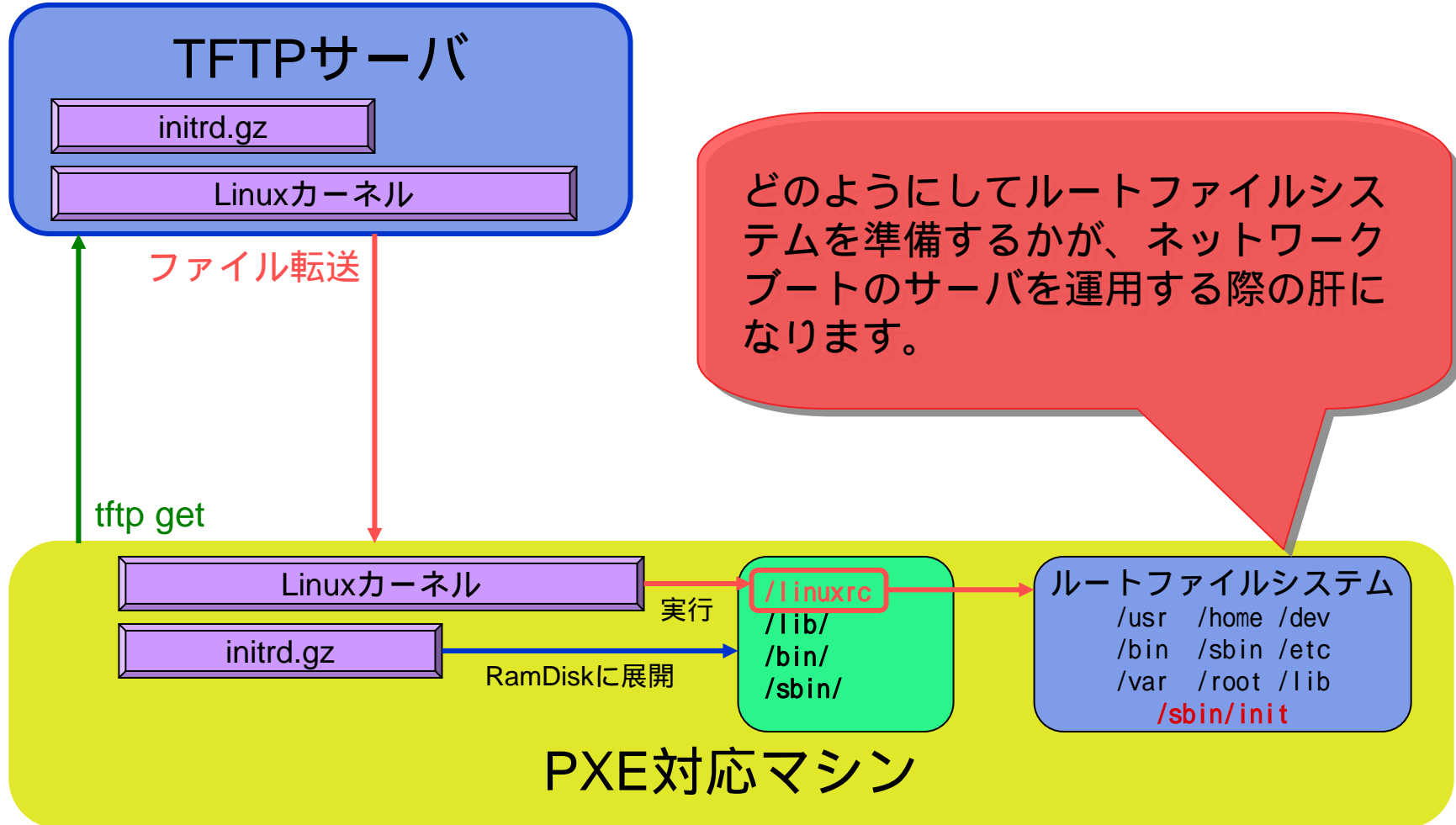
「ファイルシステムイメージを圧縮したもの」です

```
# gzip -d initrd.gz
# mount -o loop initrd /initrd/
# ls -l /initrd/
total 12
drwxr-xr-x 2 root root 2048 2004-10-01 20:39 bin
drwxr-xr-x 2 root root 1024 2005-04-15 00:04 dev
drwxr-xr-x 2 root root 1024 2005-03-10 03:45 etc
-rwx----- 1 root root 2094 2007-09-25 07:21 linuxrc
drwxr-xr-x 2 root root 1024 2004-10-08 15:11 proc
drwxr-xr-x 2 root root 1024 2004-10-01 20:39 sbin
```

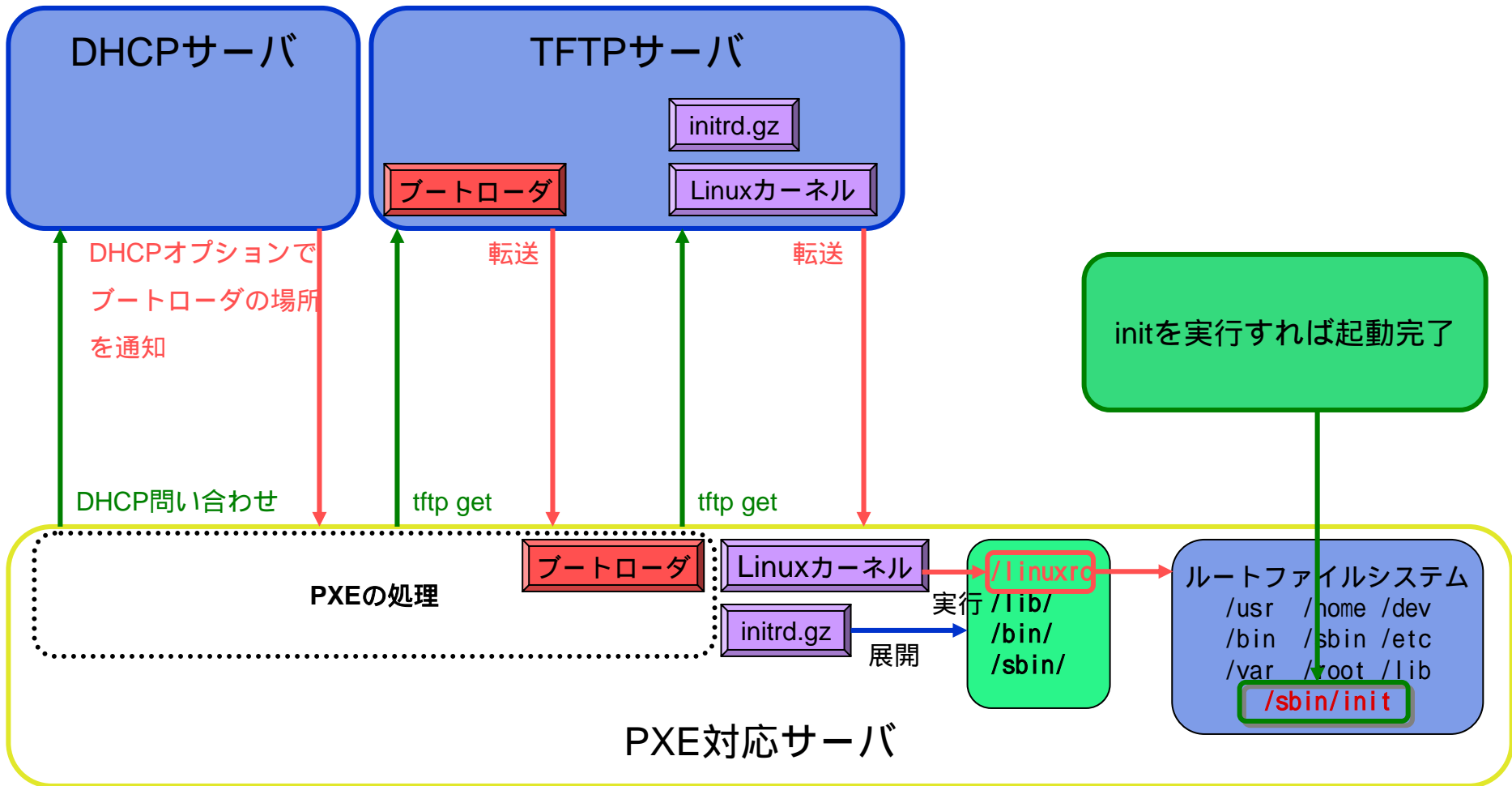
起動時にinitrdをロードすると、これらの内容がRamDiskに展開されて、linuxrcというスクリプトが実行されます。

linuxrcには、起動に必要な環境を整備する処理を記述しておきます。
(例えば、ドライバのロードとかNICの初期化とかあれとかこれとか……)

initrdの動作



ネットワークブートのまとめ



ネットワークブートに必要なもの

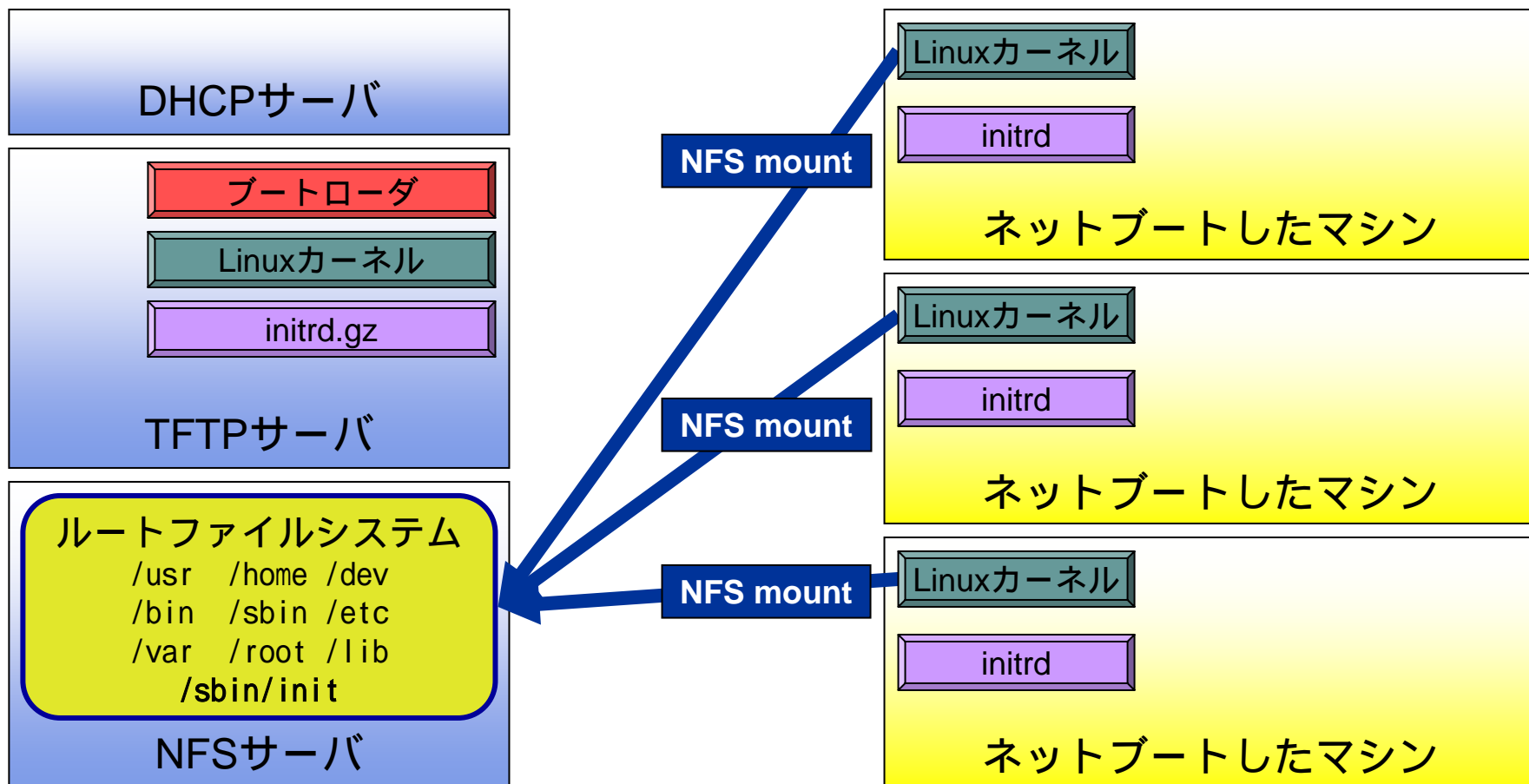
- PXE対応のマシン
- DHCPサーバ
- TFTPサーバ
- PXE対応のブートローダ
- Linuxカーネル
- initrd.gz
- ルートファイルシステム

具体的な構成例



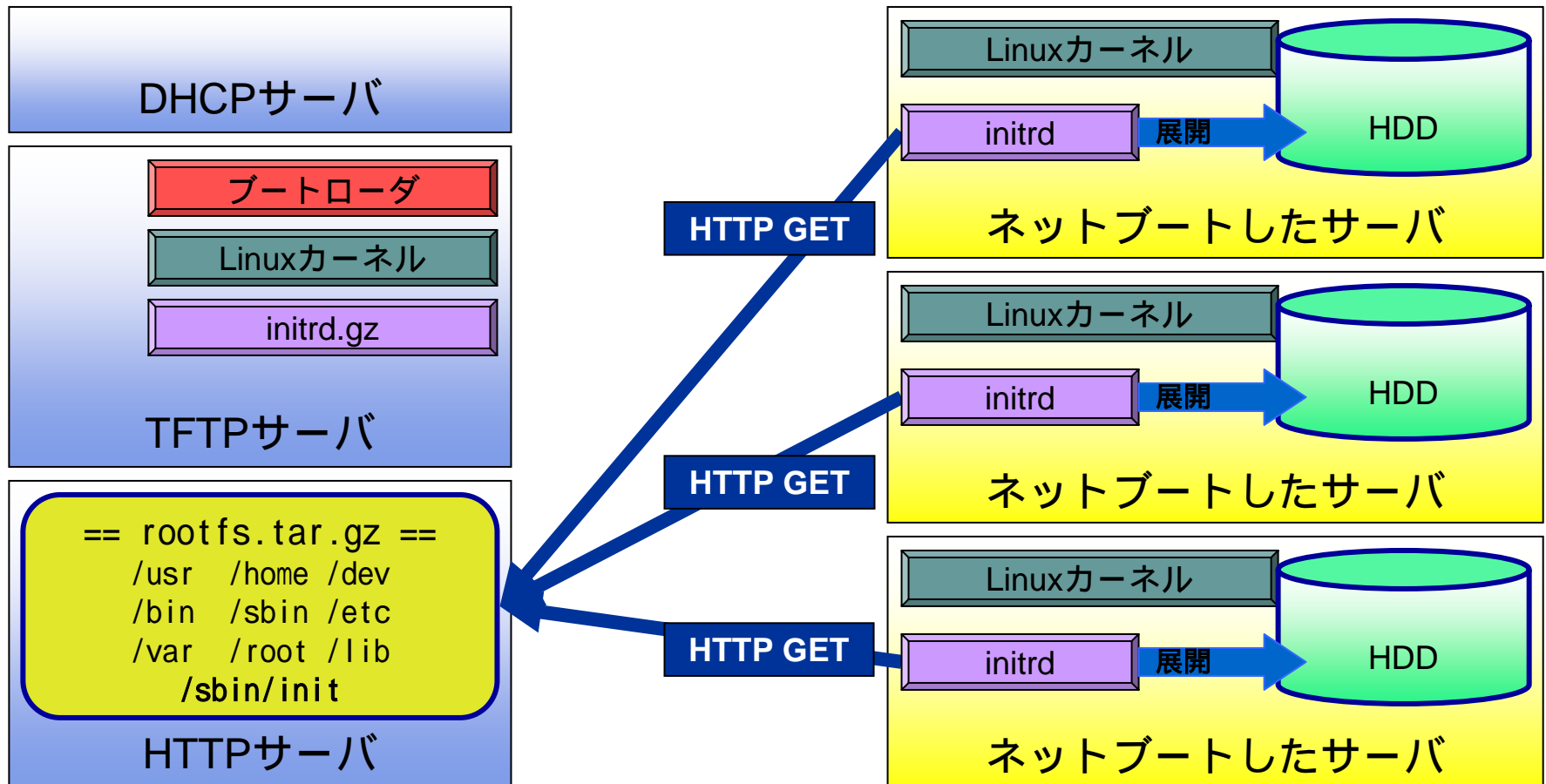
構成例(1)

- NFSを利用した構成 (nfs-root)



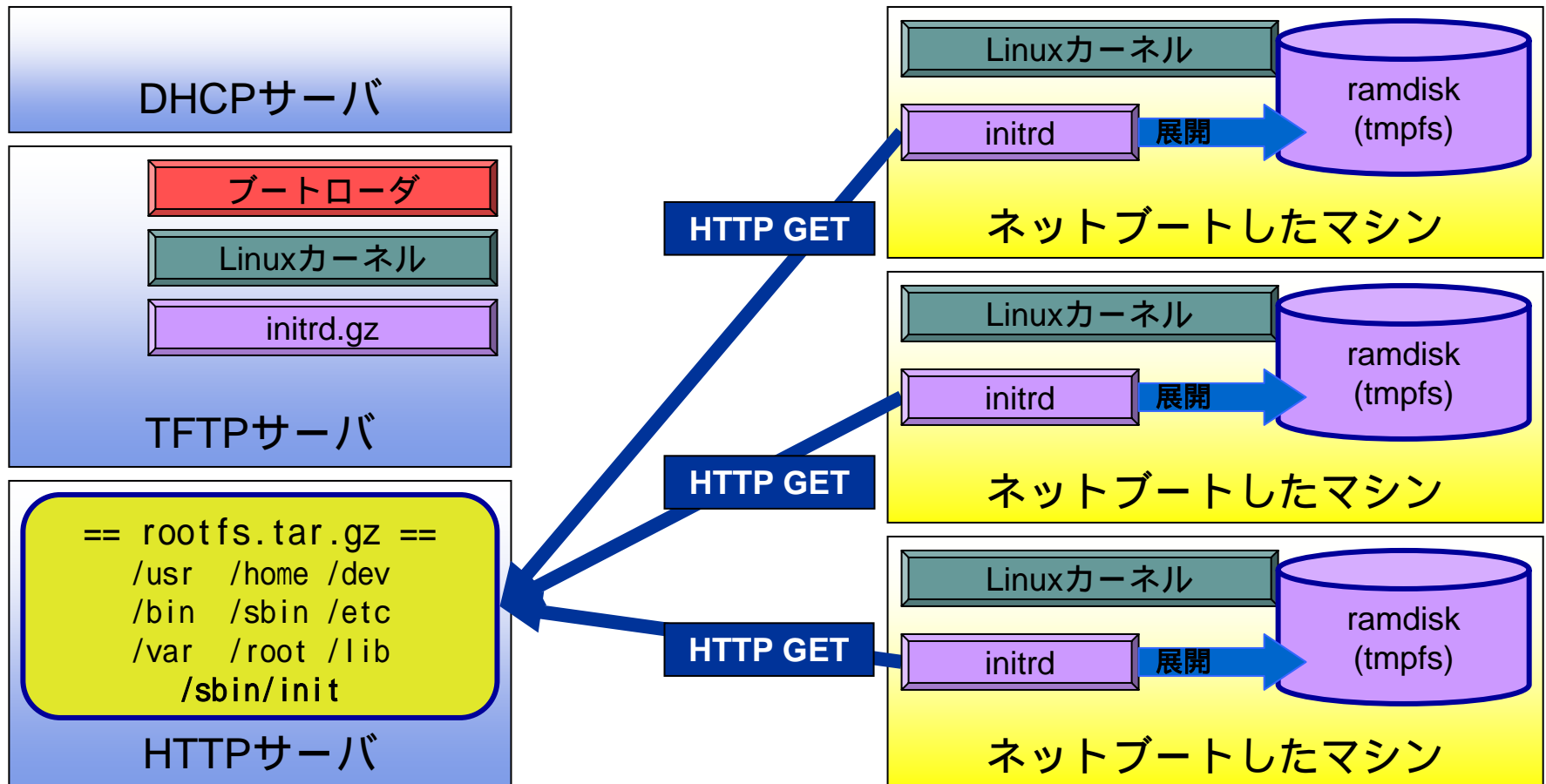
構成例(2)

- HDDを利用した構成



構成例(3)

- RamDisk(tmpfs)を利用した構成



DSASは怎么样了の？



DSASの全体像

- フロントエンドサービス向けサーバ群
- バックエンドサービス向けサーバ群

LVS (Active)	LVS (Backup)
-----------------	-----------------

マスタサーバ (Primary)	マスタサーバ (Secondary)	WEBサーバ	WEBサーバ	WEBサーバ
WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ

DBサーバ (Master)	TS (Active)	LLS (Active)	PS/WS (Active)	ログサーバ (Active)
DBサーバ (Slave)	TS (Backup)	LLS (Backup)	PS/WS (Backup)	ログサーバ (Backup)
DBサーバ (Slave)			BK	



DSASの全体像

- フロントエンドサービス向けサーバ群
- バックエンドサービス向けサーバ群

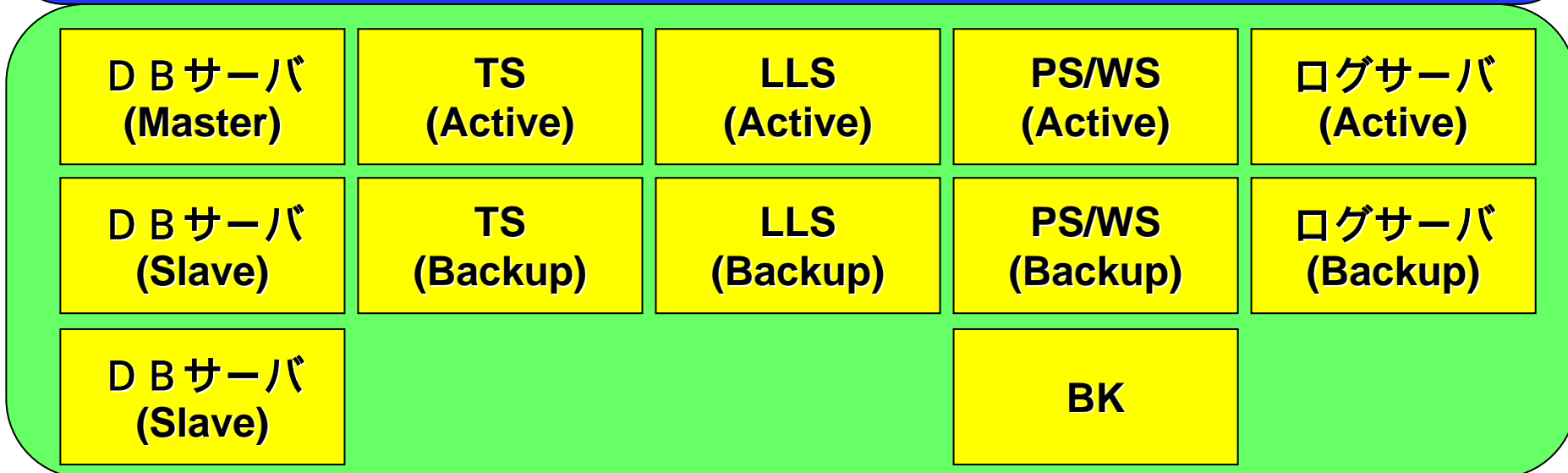
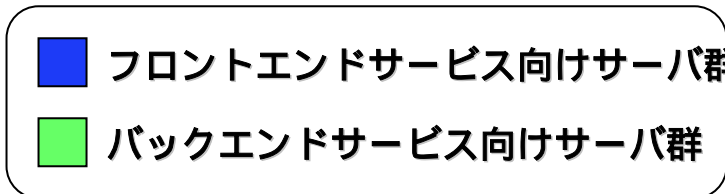
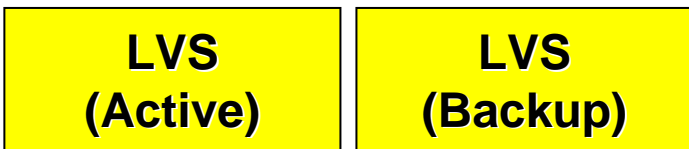
LVS (Active)	LVS (Backup)
-----------------	-----------------

マスタサーバ (Primary)	マスタサーバ (Secondary)	WEBサーバ	WEBサーバ	WEBサーバ
WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ

DBサーバ (Master)	TS (Active)	LLS (Active)	PS/WS (Active)	ログサーバ (Active)
DBサーバ (Slave)	TS (Backup)	LLS (Backup)	PS/WS (Backup)	ログサーバ (Backup)
DBサーバ (Slave)			BK	

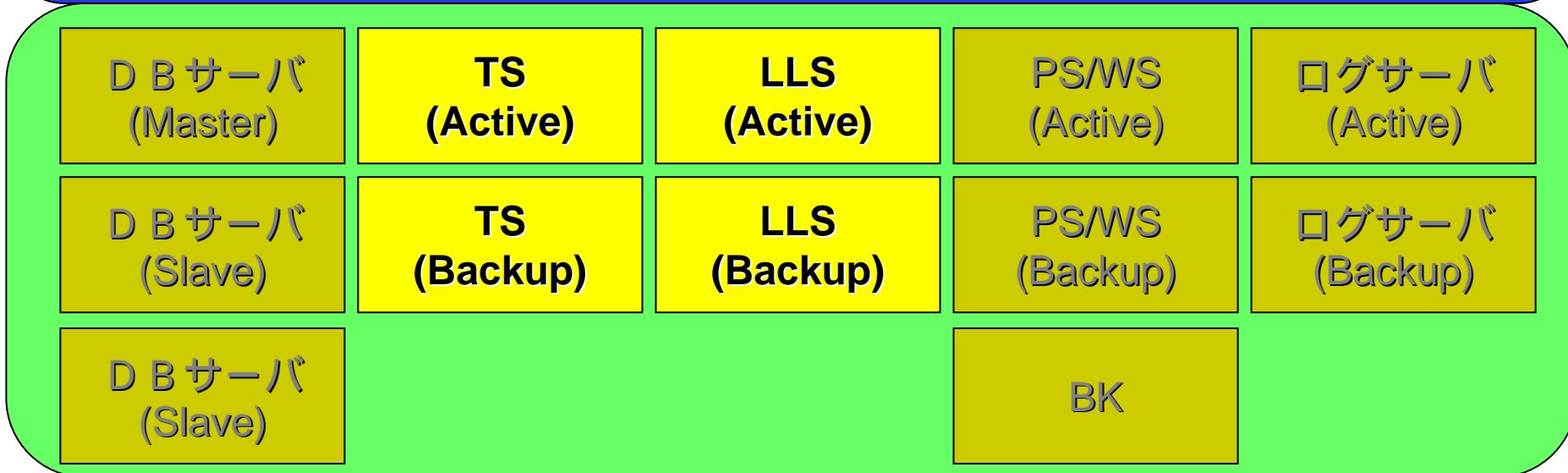
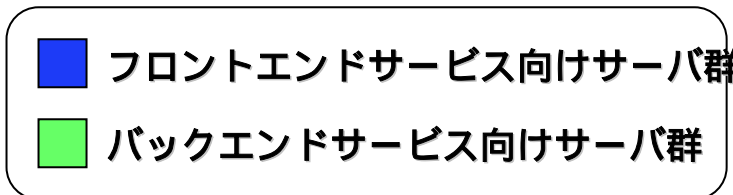
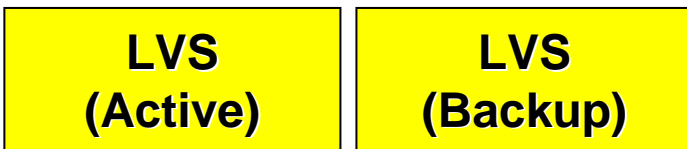


DSASの全体像



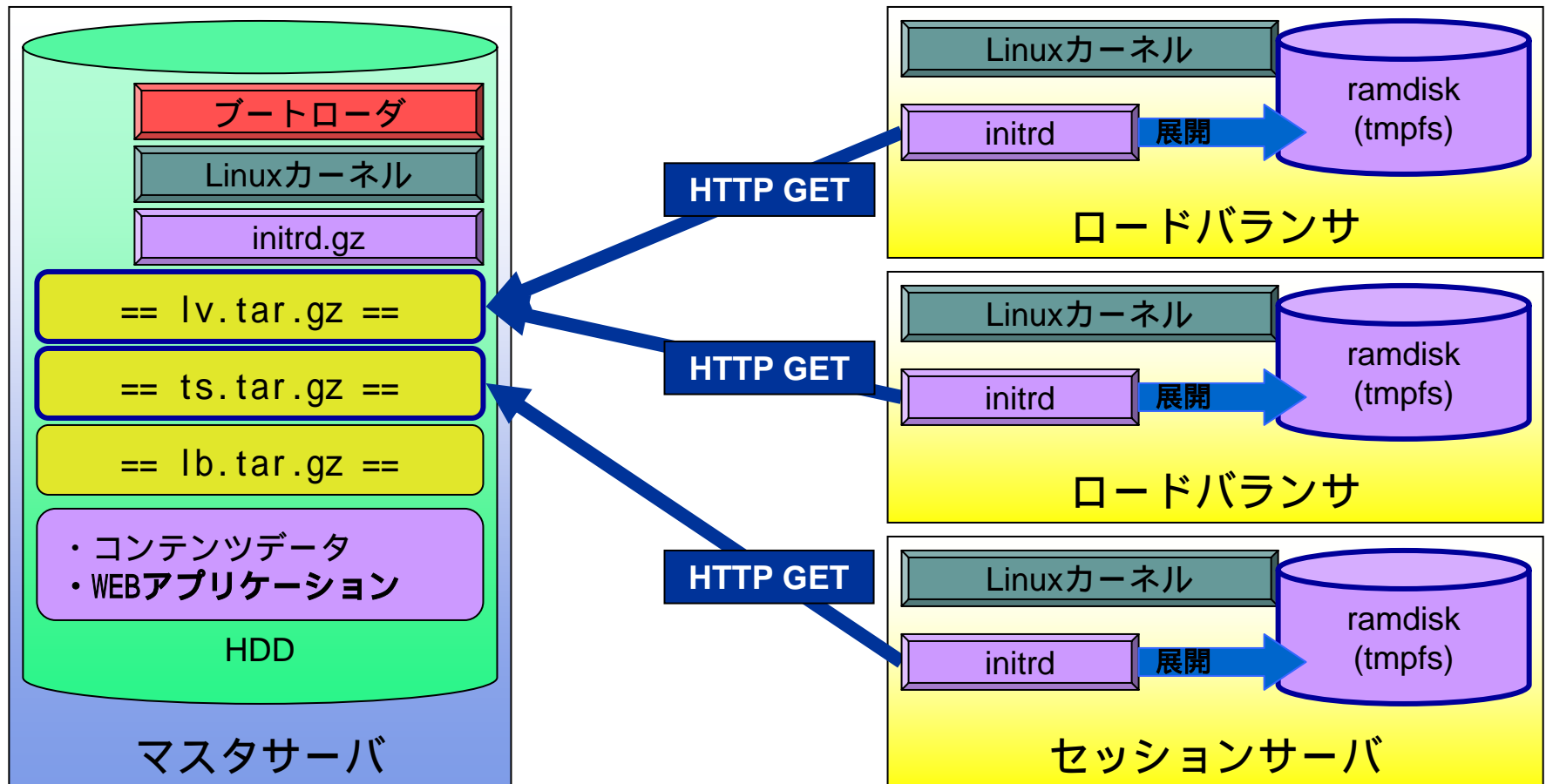


DSASの全体像



DSASの構成

- ロードバランサなどはディスクレス





DSASの全体像

- フロントエンドサービス向けサーバ群
- バックエンドサービス向けサーバ群

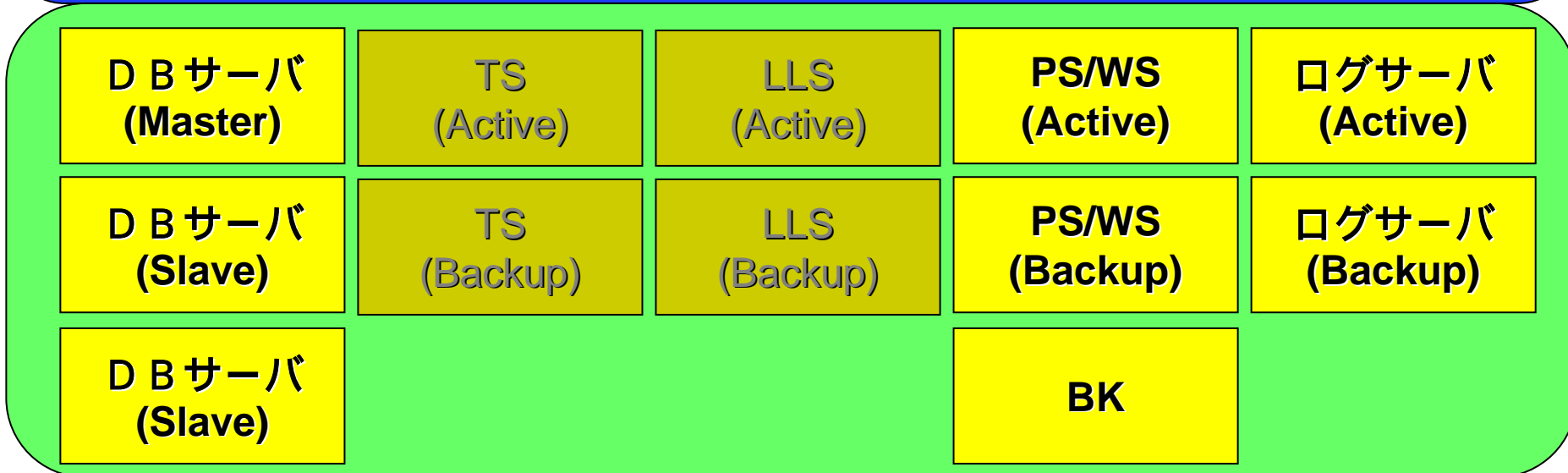
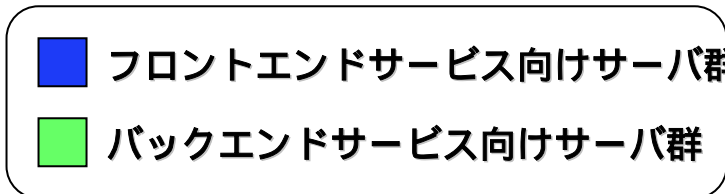
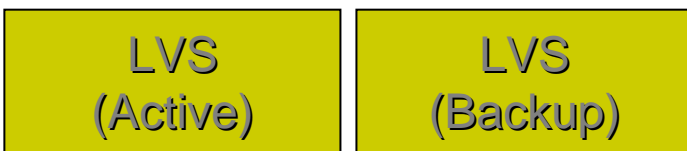
LVS (Active)	LVS (Backup)
-------------------------	-------------------------

マスタサーバ (Primary)	マスタサーバ (Secondary)	WEBサーバ	WEBサーバ	WEBサーバ
WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ	WEBサーバ

DBサーバ (Master)	TS (Active)	LLS (Active)	PS/WS (Active)	ログサーバ (Active)
DBサーバ (Slave)	TS (Backup)	LLS (Backup)	PS/WS (Backup)	ログサーバ (Backup)
DBサーバ (Slave)			BK	

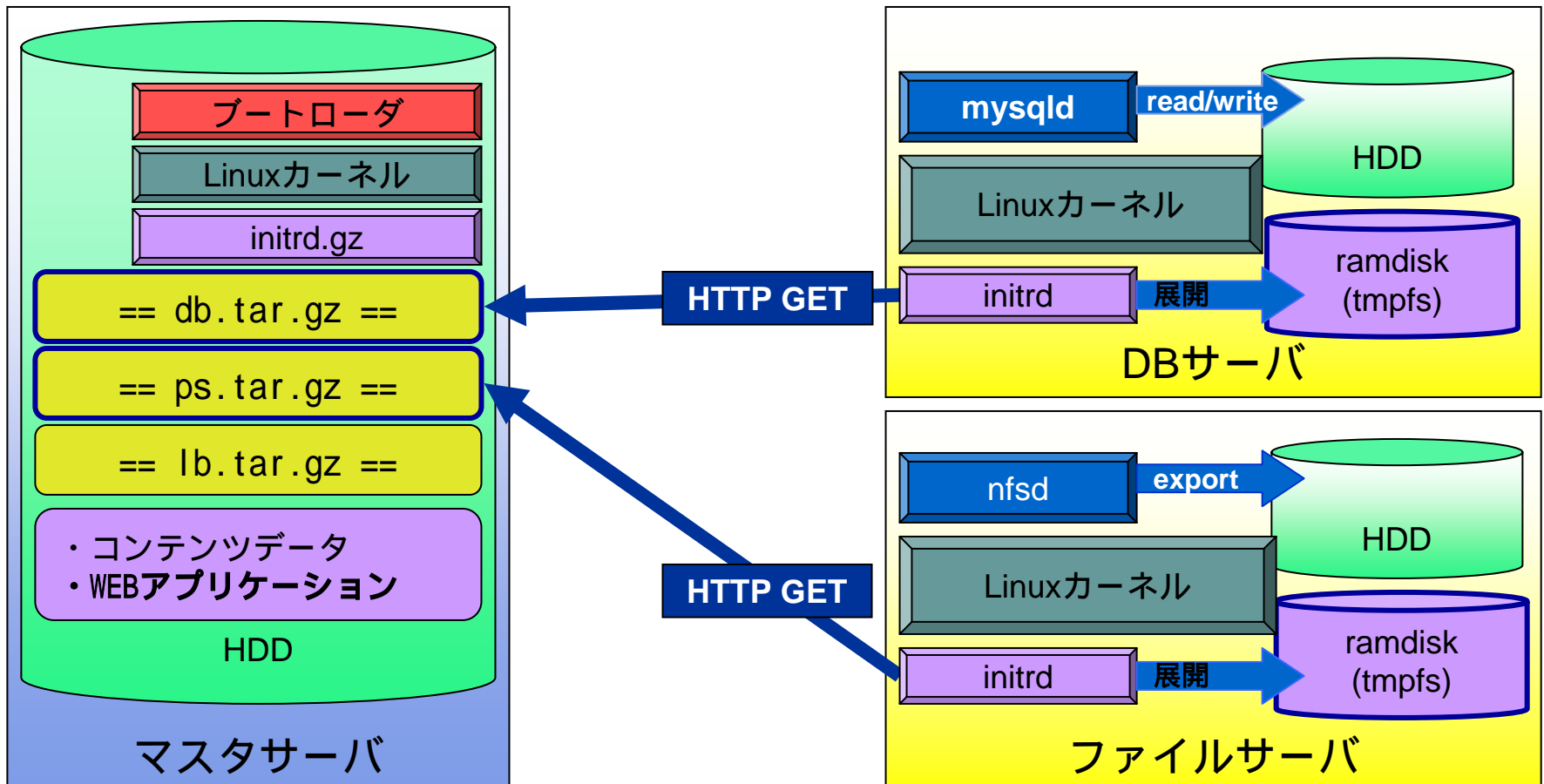


DSASの全体像



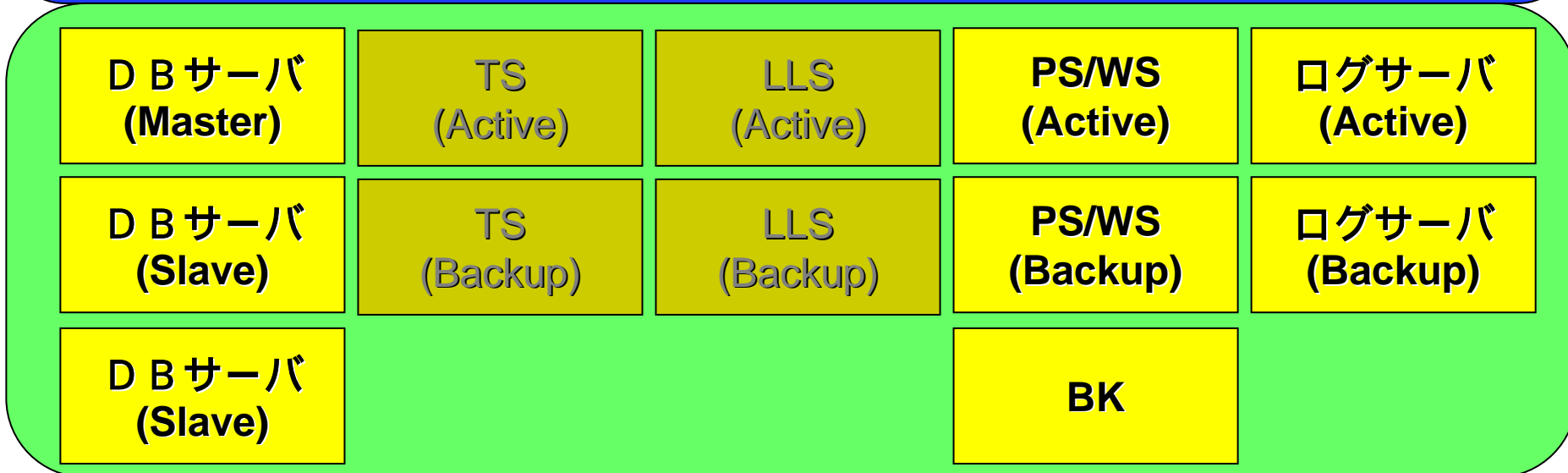
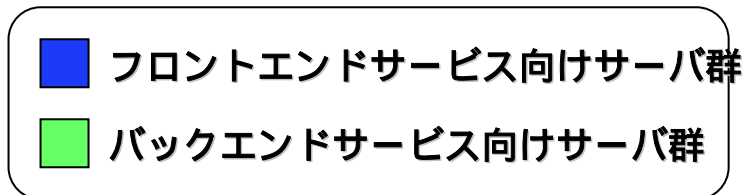
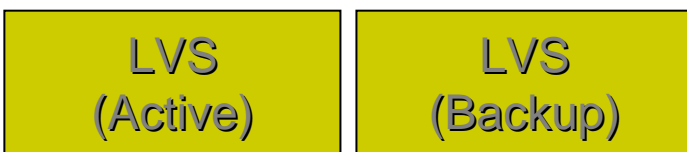
DSASの構成

- DBサーバやファイルサーバはHDDを併用



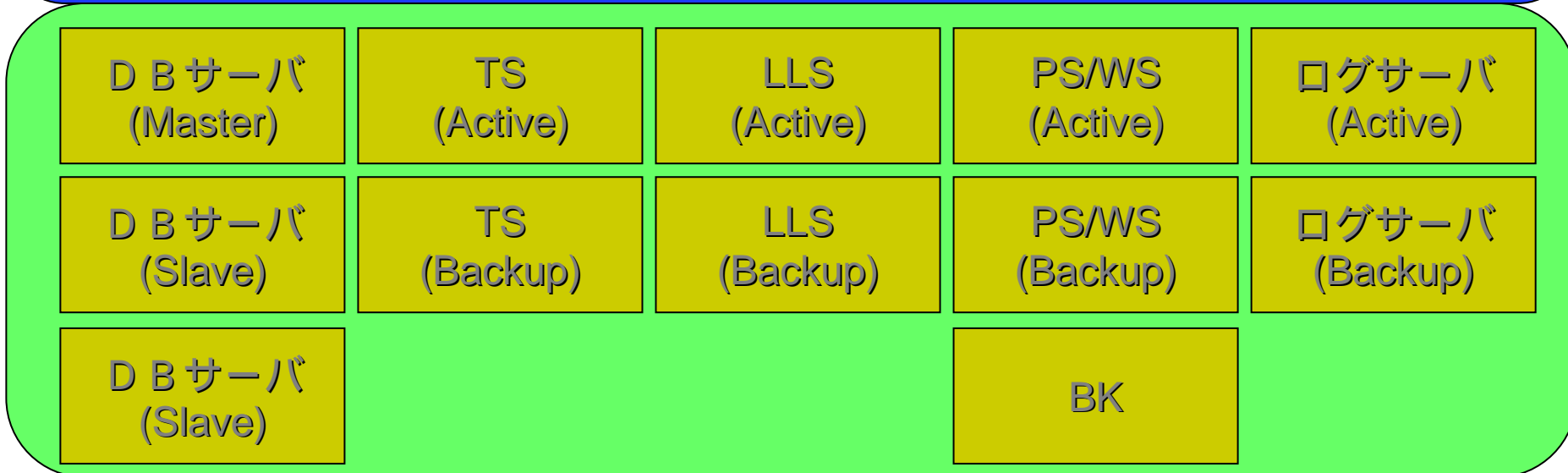
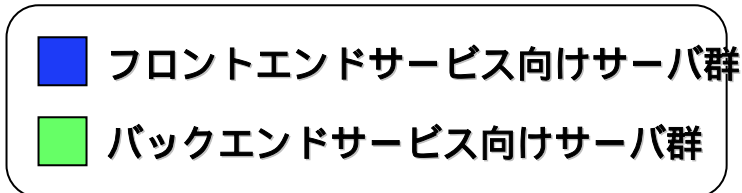
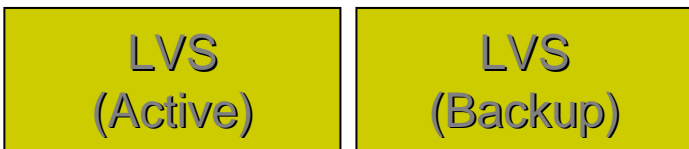


DSASの全体像



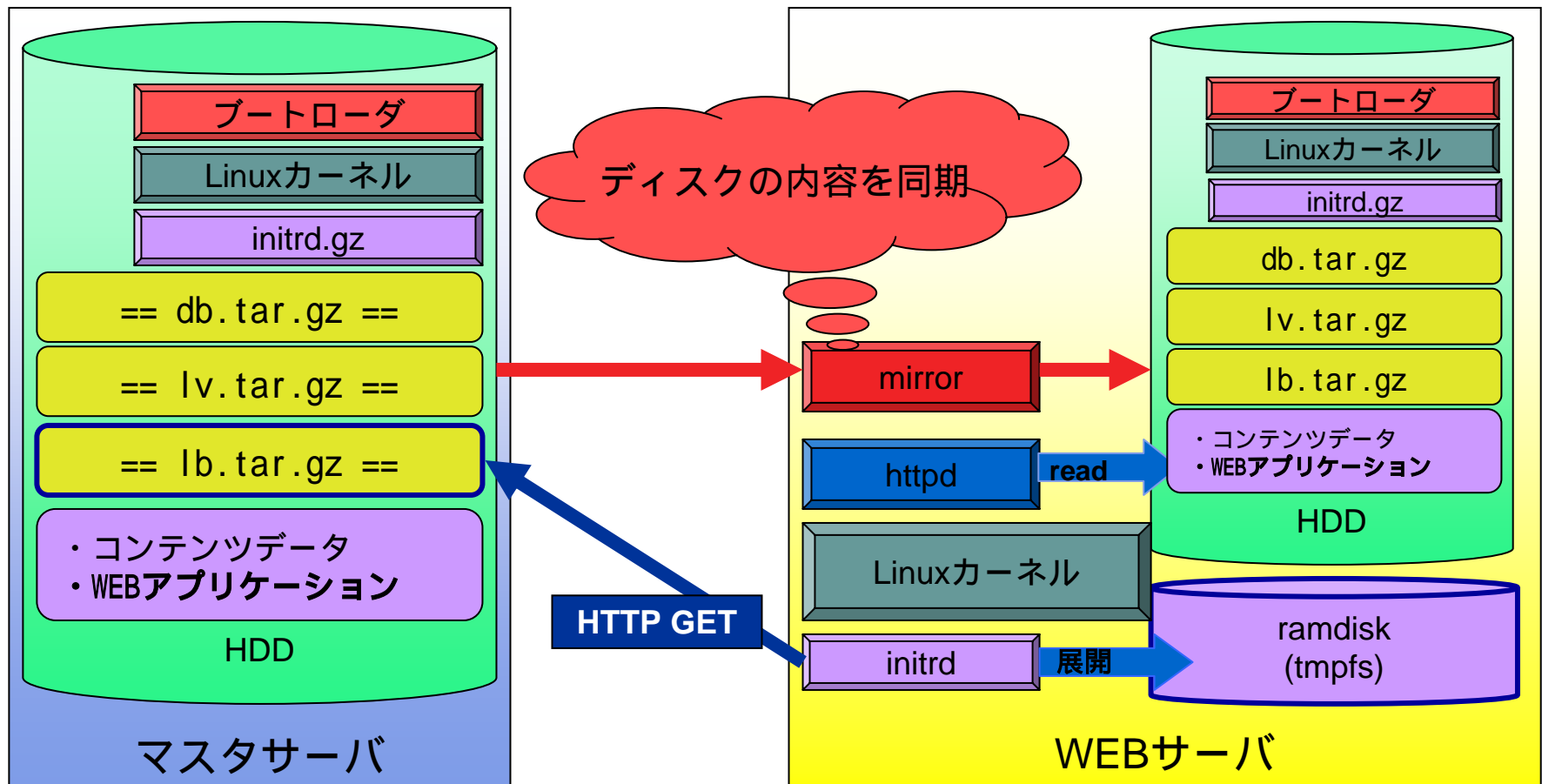


DSASの全体像



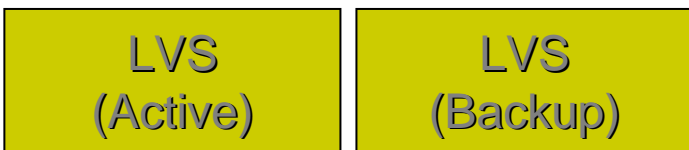
DSASの構成

- WEBサーバもHDDを併用



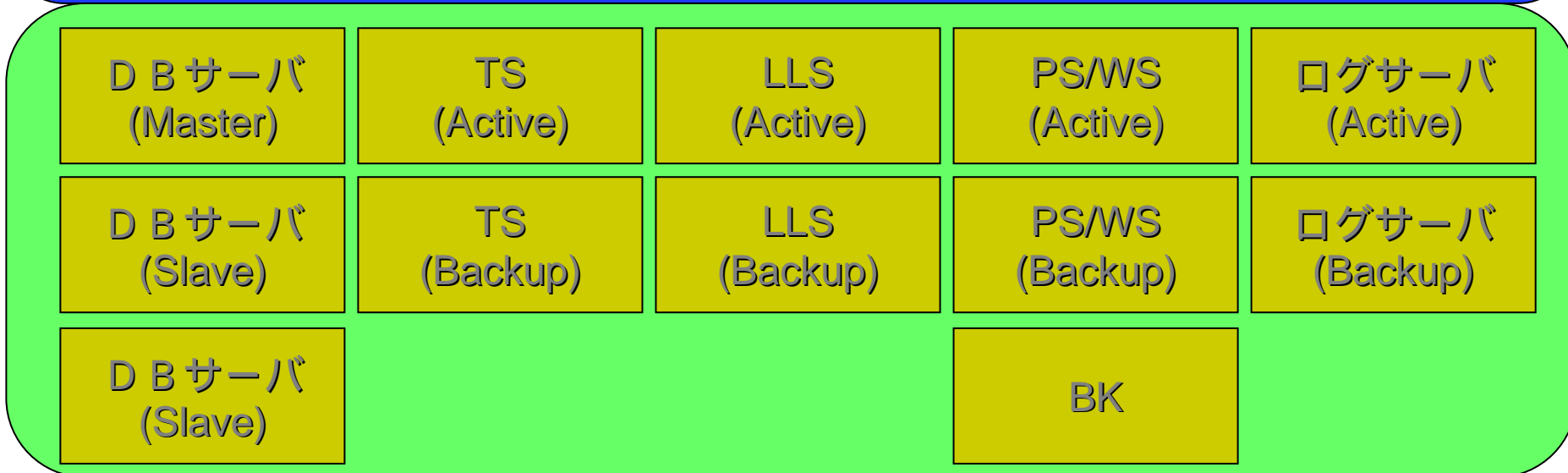


DSASの全体像



■ フロントエンドサービス向けサーバ群
■ バックエンドサービス向けサーバ群

HDDで再起動!



ネットワークブートの効果

ネットワークブートの効果

- インストール作業が不要です
 - なにはともあれ超らくちん！
 - 新しいサーバをすぐにサービスに投入できます
 - 新しいサーバが納品される前に環境構築できちゃいます
- 再起動すればもとどおり
 - ファイルを消してしまっても、設定を壊してしまっても大丈夫
 - 再起動するだけでももとどおりになります
 - バージョンアップなどの作業手順を事前に検証することができます
 - 新しい機能をちょっと試してみたいときにも重宝します
- レスキューディスクや診断ツールもネットワークブート
 - メンテナンス用のCDやFDを探す必要がなくなりました
 - memtest86なんかもネットワークブートでいけちゃいます



ネットワークブートの効果

- アプリケーションがディスクI/Oを占有できます
 - DBサーバのHDDはMySQLが独り占め
 - ストレージサーバのHDDはNFSが独り占め
 - ほとんどのプログラムはオンメモリで実行されます
- ディスク故障でOSが止まる事はありません
 - ディスクが壊れると・・・
 - Apacheがエラーを吐くかもしれませんが
 - MySQLが止まるかもしれませんが
 - バッチ処理などが失敗するかもしれませんが
 - ヘルスチェック機能やメール送信機能は影響を受けません
 - ディスク故障は速やかに正確に管理者に通知されます

運用テクニックの紹介

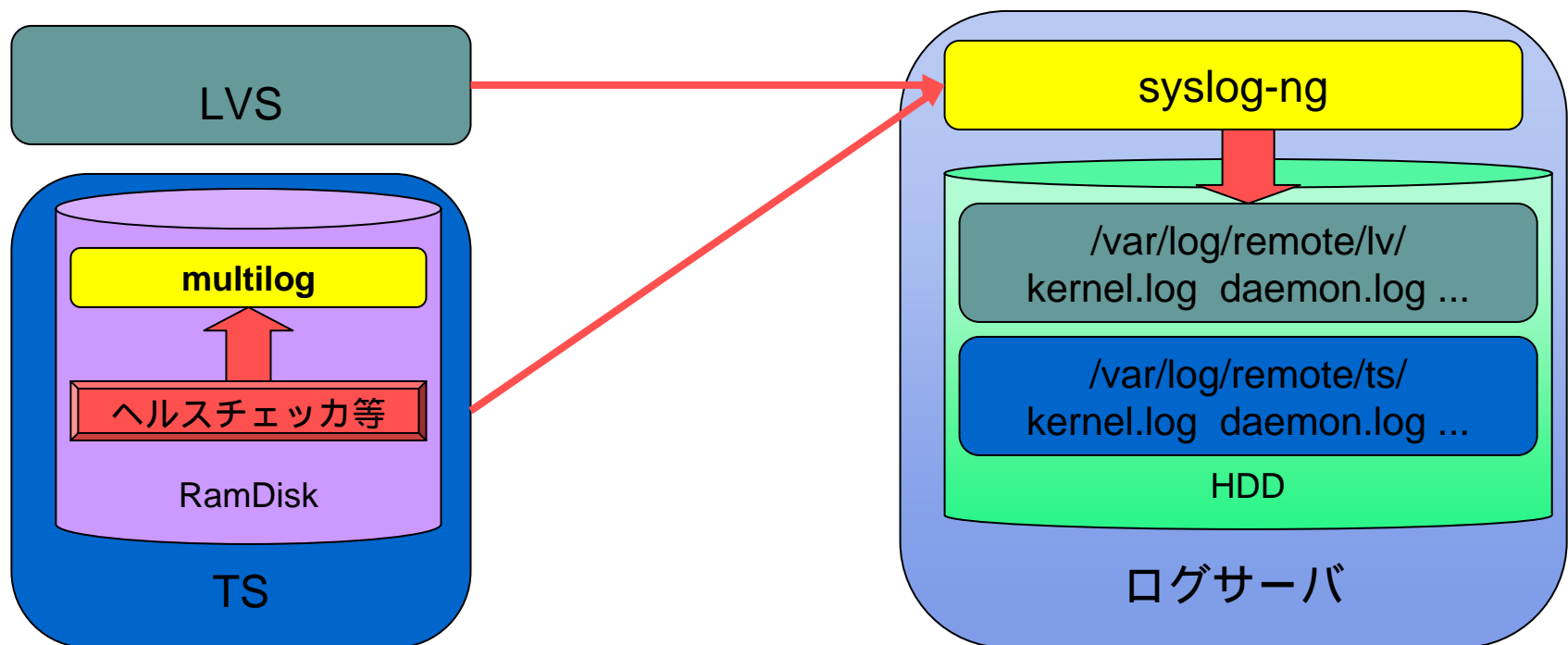


運用上の懸念点

- ログ管理
 - ディスクレス構成の場合、ログはどこにだせばいいの？
 - RamDiskに書いてたらメモリがいくらあっても足りなくね？
 - かといって、ログのためだけにディスク使うのもなんだかねえ…
- ルートファイルシステムのメンテナンス
 - バージョンアップはどうやるの？
 - 設定変更したいときはどうするの？
 - 再起動したら元に戻っちゃうよね？
- サーバ固有情報の処理
 - ホスト名やIPアドレスはどうやって指定するの？

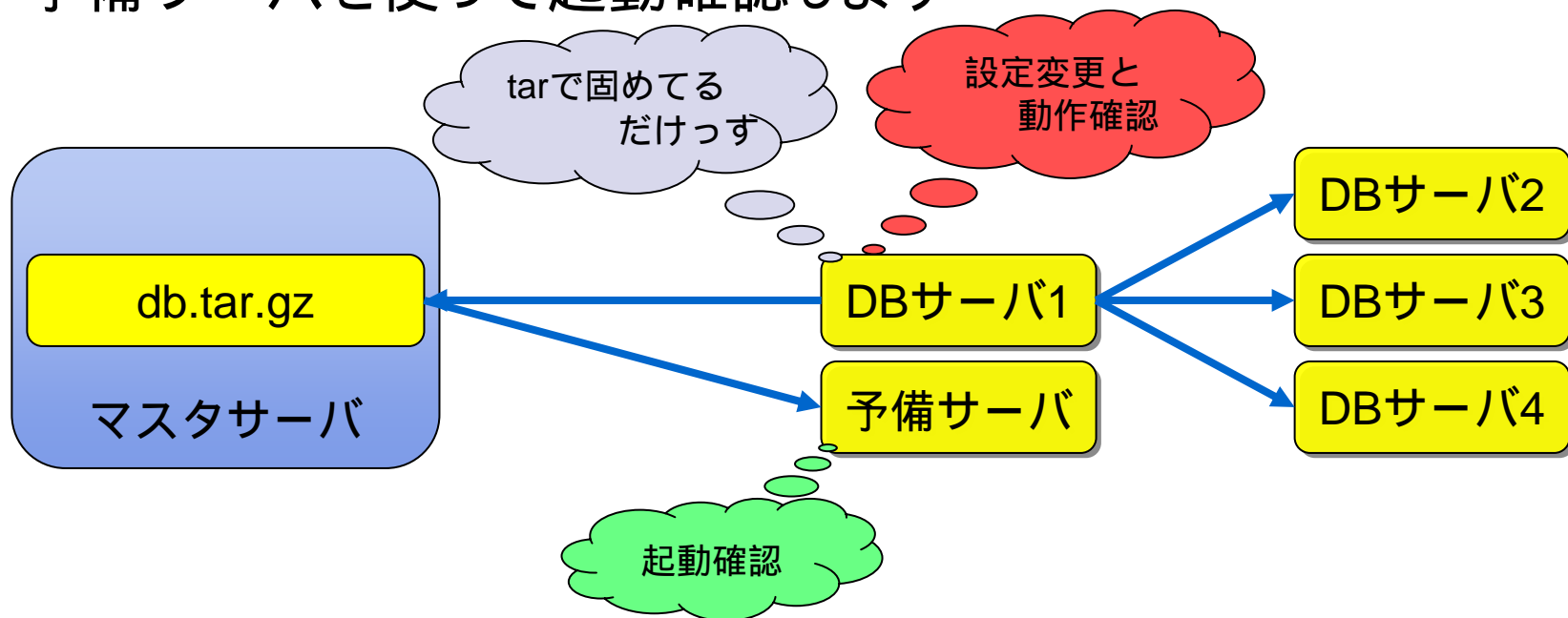
ログ管理

- ディスクレスマシンのログはネットワーク経由でログサーバに流します
- syslog-ngを使えばサーバごとにディレクトリを分けることができます
- 転送するまでもないログはdaemontoolsに付属のmultilogが便利です



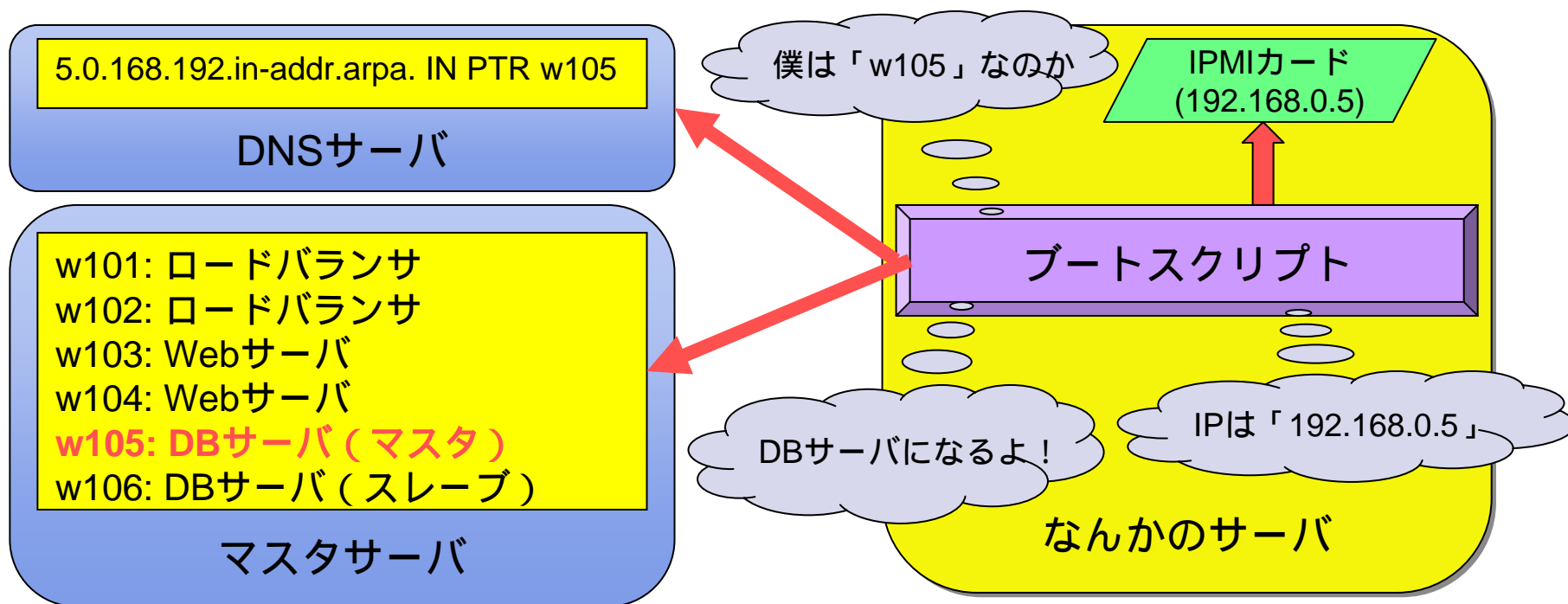
ルートファイルシステムのメンテナンス

- 設定変更やバージョンアップをしたい場合など
- 実機上で作業をして動作確認をします
- 変更内容は同じ種類の全サーバに適用します
- ルートファイルシステムイメージを再生成します
- 予備サーバを使って起動確認をします



サーバ固有情報の処理

- IPアドレスはIPMIカード(ハードウェア)に設定しています
 - ブートスクリプトはIPMIカードを参照して、IPアドレスを取得します
 - 次にDNSを参照して、IPアドレスからホスト名を取得します
 - 最後にマスタサーバを参照して、ホスト名から役割情報を取得します
- IPMIカードを使えないサーバは、起動時のカーネルパラメータで指定します



ってな感じになればいいなあ (^ ^ ;

と考えていま

す

最後になりましたが

ネットブートの導入を検討している方へ

最初から完璧な構成を目指すと、途中でくじけそうになります。
「ネットワークブート = ディスクレスシステム」とは限りません。
単にネットワークからブートするだけであれば、

- `dhcpd` (DHCPサーバ)
- `atftpd` (`tsize`オプション対応のTFTPサーバ)
- `pxelinux` (PXE対応のブートローダ)
- HDDにLinuxをインストールしたマシン (PXE対応のもの)

これだけあればできちゃいます。

PXEでブートローダをロードして、HDDをマウントして起動したシステムも、立派な「ネットワークブートシステム」ですよ。

「ネットブート使いたいけど、なんとなくめんどくさそう」という方は、まず、ここから手を付けて感触をつかんでみることをオススメします。

ご清聴ありがとうございました