株式会社サードウェア
Thirdware Inc.

リアルタイムレプリケーションツール

DRBD
アドバンスド・チュートリアル

株式会社サードウェア
久保 元治
(LINBIT認定コンサルタント)

株式会社サードウェア
Thirdware Inc.

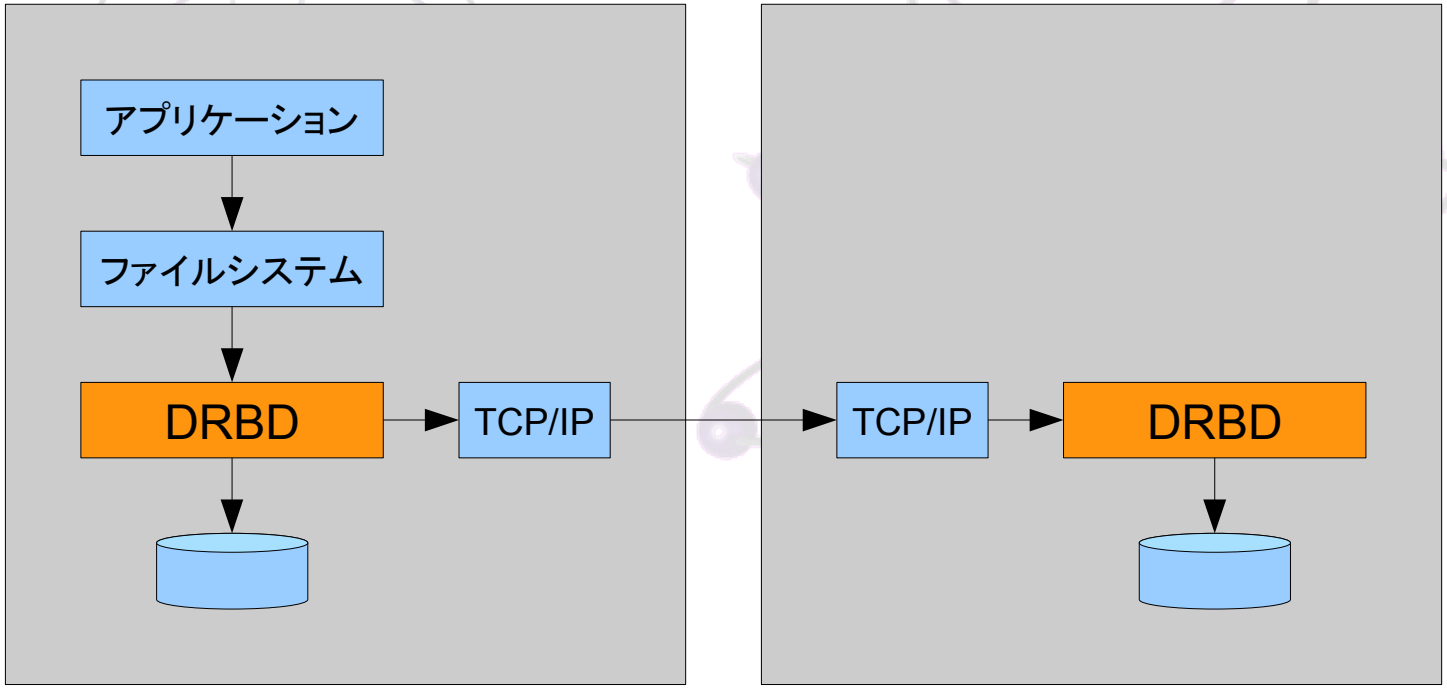
アジェンダ

- DRBDのおさらい
- スプリットブレインの防止・検出・復旧
- データ整合性の検証
- 3ノード、4ノードへの拡張

株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

- ブロックデバイスをレプリケート
  - ブロック単位
  - コピー元(primary)とコピー先(secondary)



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

- 各種障害に対するデータ保護
  - 片肺運転と再同期
  - ストレージ交換後のフル同期
  - データ整合性の検証
  - スプリットブレインの防止・検出・復旧

株式会社サードウェア
Thirdware Inc.

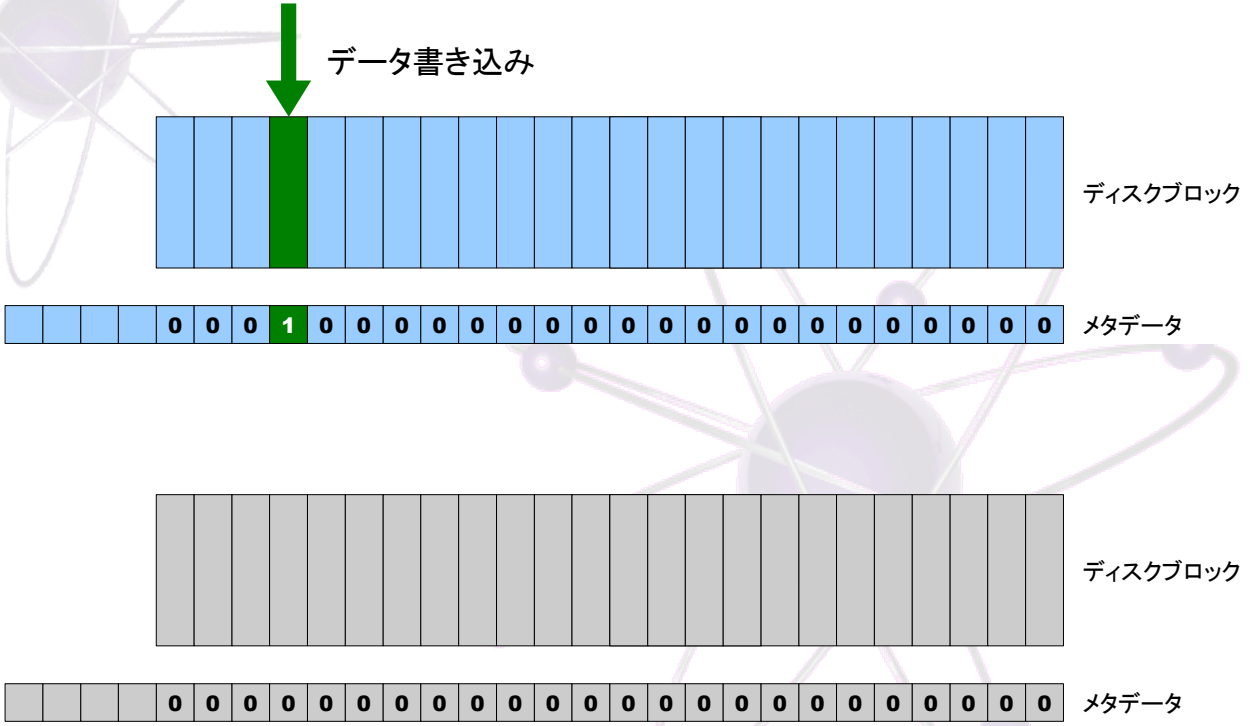
DRBDのおさらい

- HAクラスタ
  - Pacemaker ＋ Heartbeat (Corosync)
- バックアップ
  - LVMスナップショットと併用など
- 災害対策
  - 遠隔地へのリアルタイムバックアップ
  - アクセラレータとしてDRBD Proxy

株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

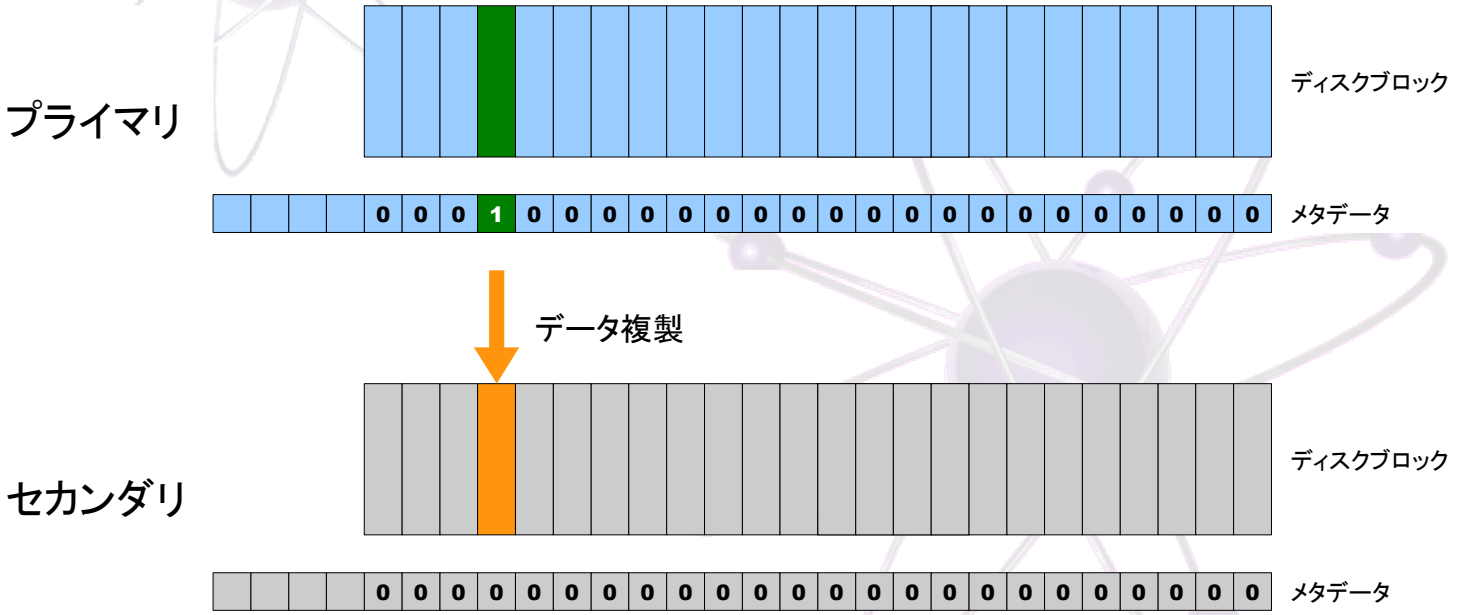
正常動作時



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

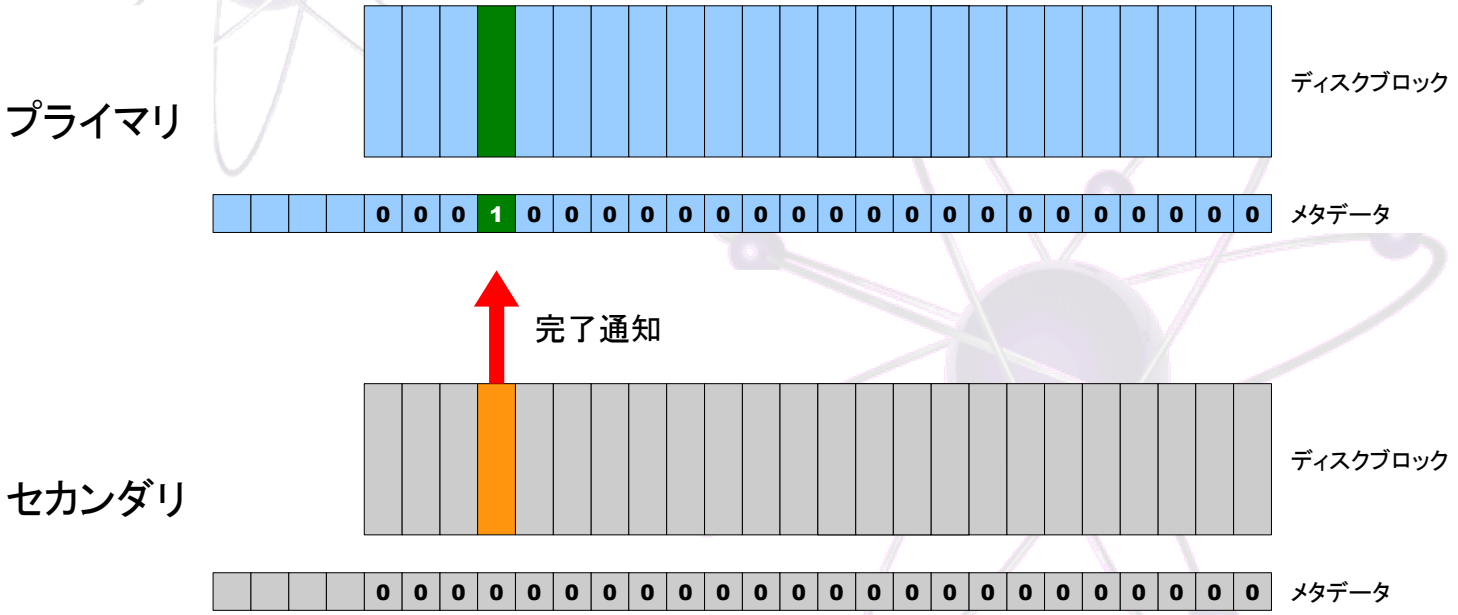
正常動作時



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

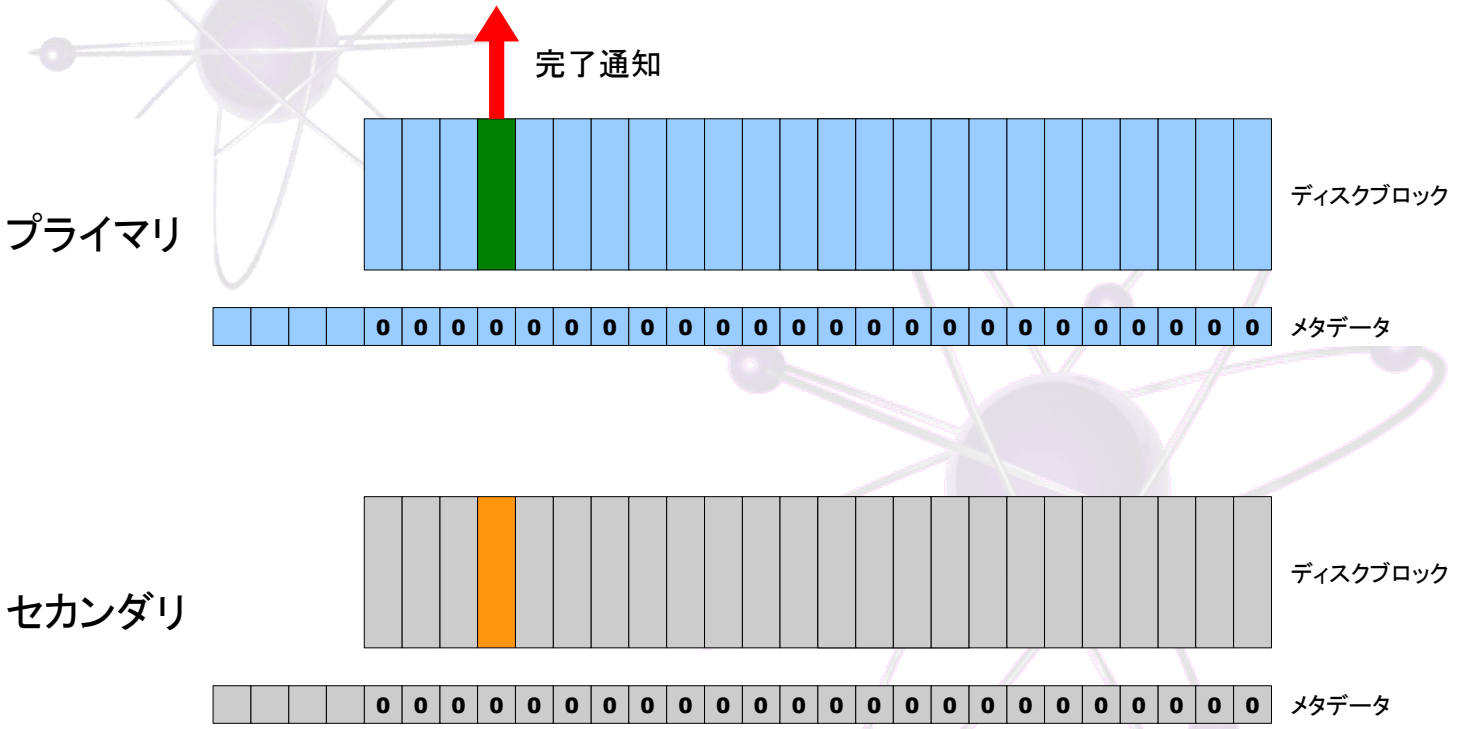
正常動作時



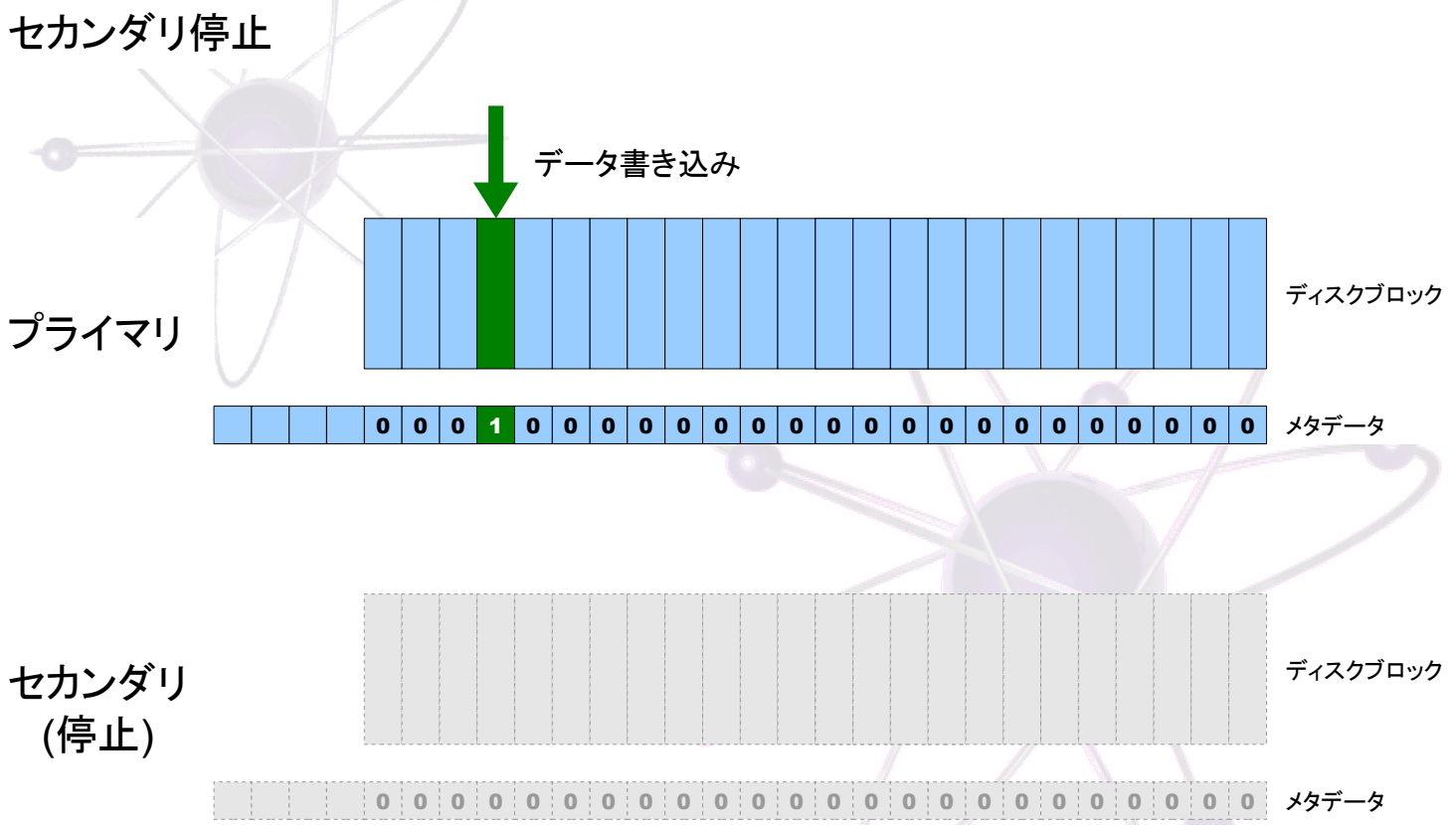
株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

正常動作時

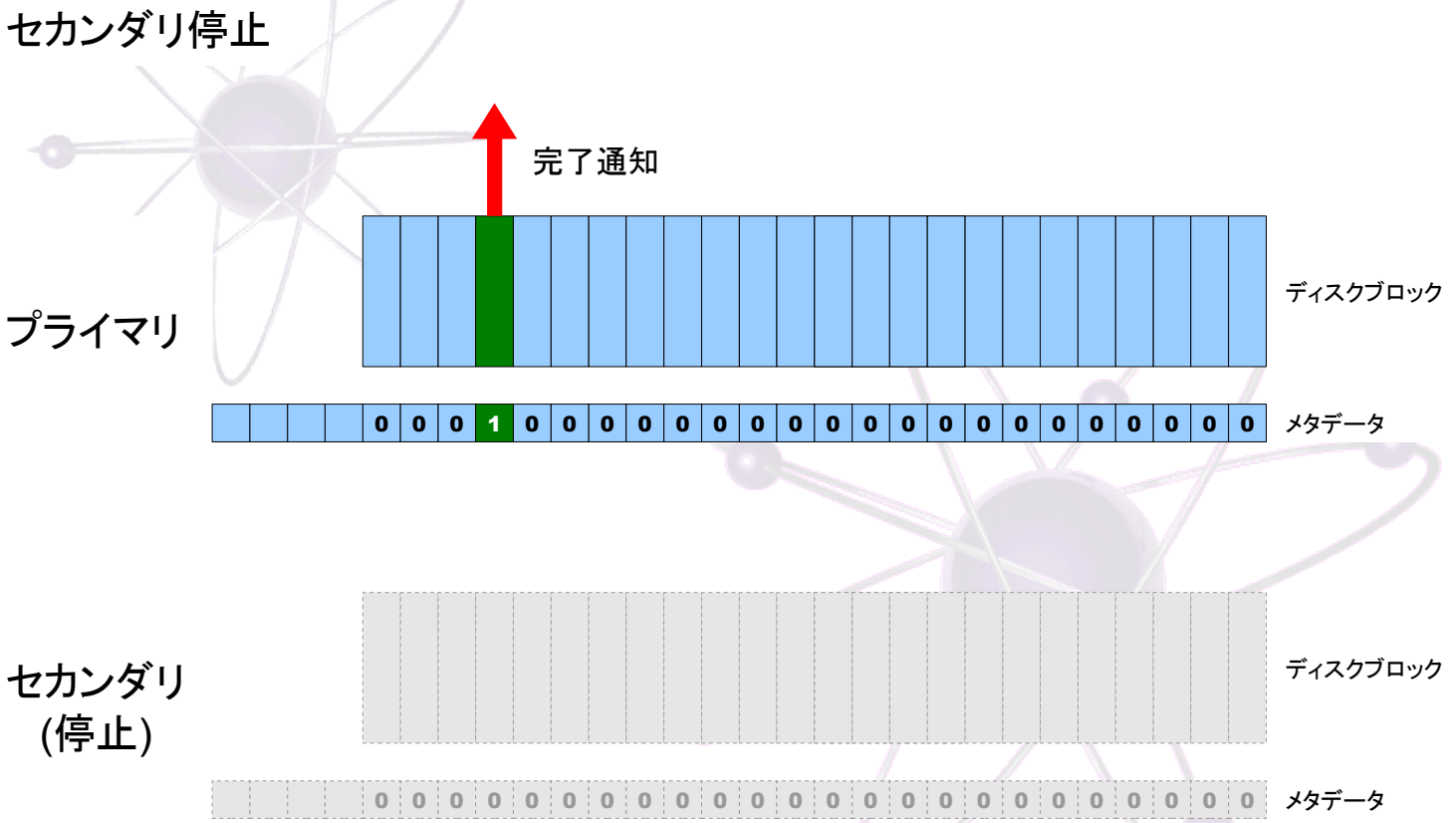


DRBDのおさらい



株式会社サードウェア
Thirdware Inc.

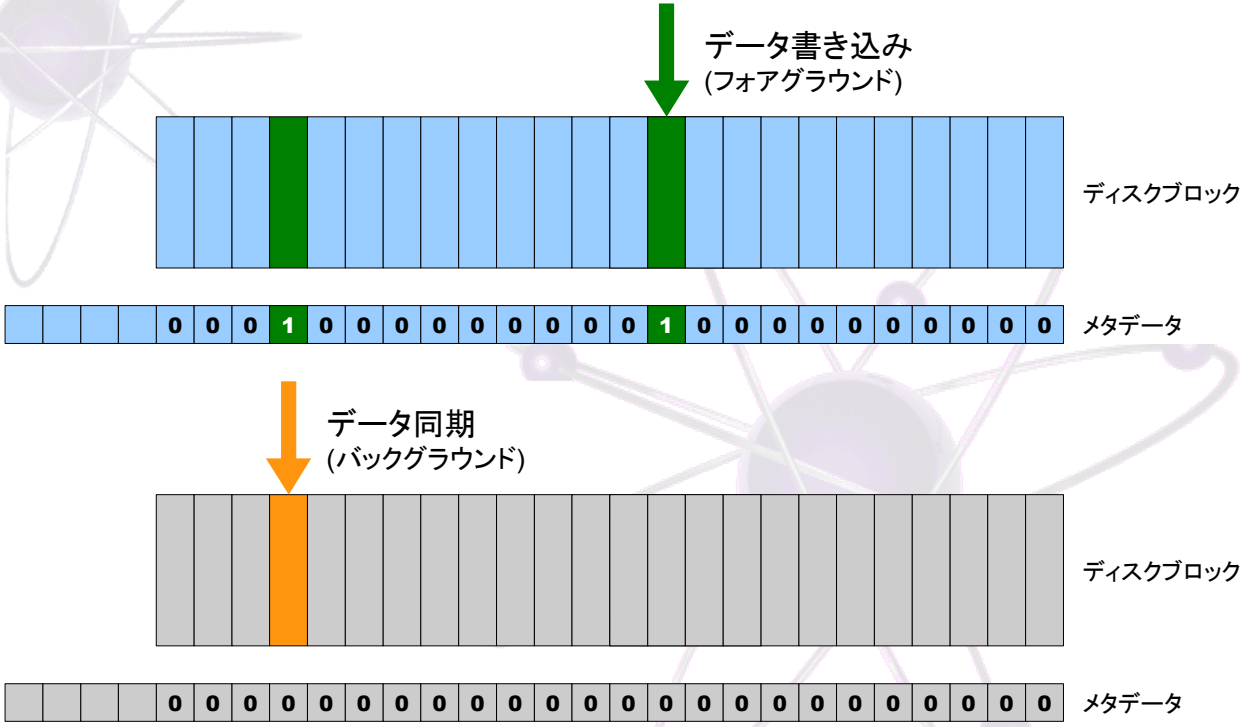
DRBDのおさらい



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

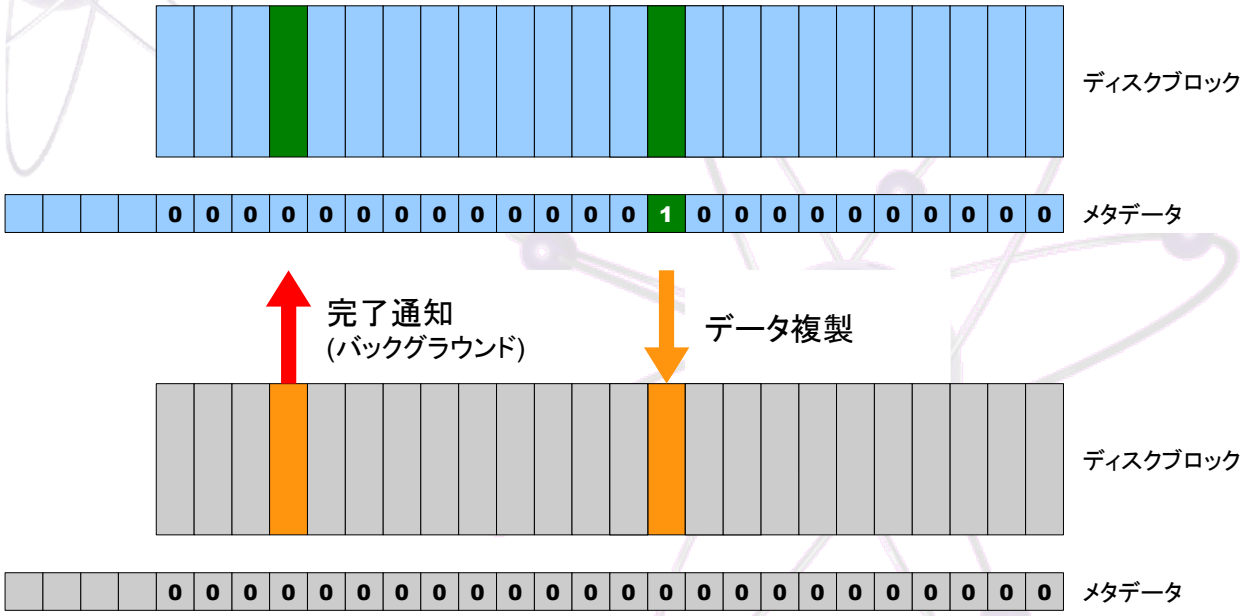
セカンダリ回復



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

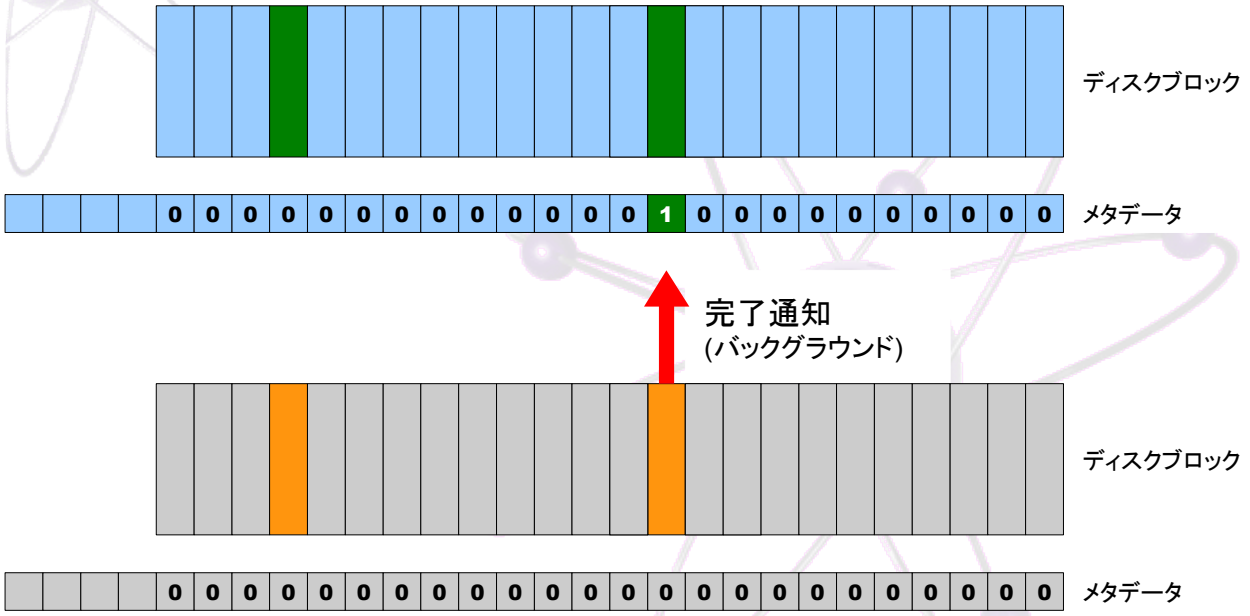
セカンダリ回復



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

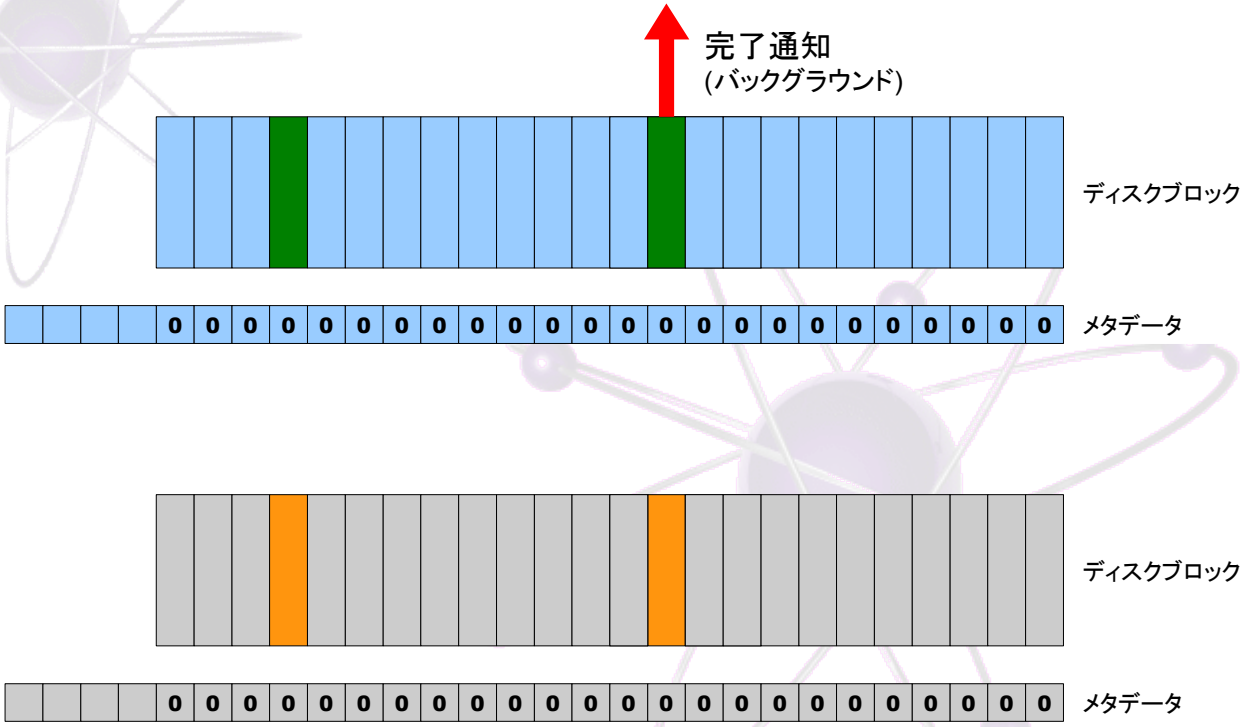
セカンダリ回復



株式会社サードウェア
Thirdware Inc.

DRBDのおさらい

セカンダリ回復



株式会社サードウェア
Thirdware Inc.

スプリットブレイン

- DRBDは両方がプライマリになることを許さない
  - デュアルプライマリモードは除く
  - 双方が正常に通信できている場合のみ
- 通信が途絶えると
  - セカンダリはプライマリに昇格できる
- スプリットブレイン発生！
  - 各ノードに別々のデータを書き込める

株式会社サードウェア
Thirdware Inc.

dopd

- スプリットブレイン発生を抑止
- Heartbeatと組み合わせる
  - drbd-peer-outdaterハンドラ
  - fencing resource-only

株式会社サードウェア
Thirdware Inc.

dopd

- /etc/ha.d/ha.cf

```
respawn hacluster /usr/lib/heartbeat/dopd
apiauth dopd gid=haclient uid=hacluster
```

64ビットOSでは/usr/lib64ディレクトリを指定

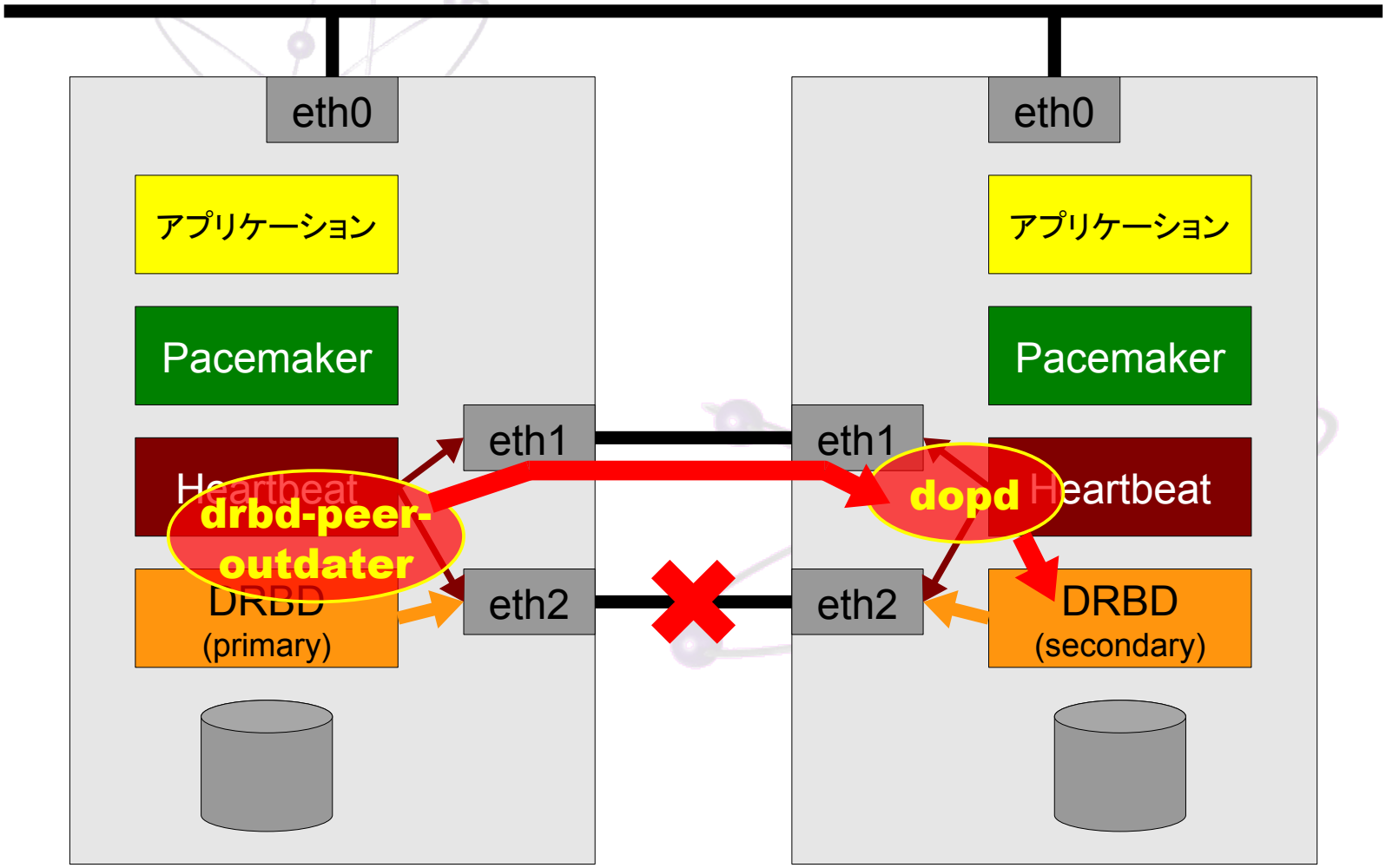
- drbd.conf

```
resource resource {
    handlers {
        outdate-peer “/usr/lib/heartbeat/drbd-peer-outdater -t 5”;
        .....
    }
    disk {
        fencing resource-only;
        ....
    }
    ....
}
```

64ビットOSでは/usr/lib64ディレクトリを指定

株式会社サードウェア
Thirdware Inc.

dopd



株式会社サードウェア
Thirdware Inc.

dopd

- Heartbeat通信経路の冗長化が必須
  (シリアル回線もOK、これを推奨)
- DRBD通信経路が途絶えると
  - drbd-peer-outdaterがdoptdにシグナル送信
  - セカンダリ側がds: Outdatedに変わる
  - Outdatedだとプライマリになれない
- DRBD通信経路が回復すると
  - 自動的にOutdatedは解除される

株式会社サードウェア
Thirdware Inc.

スプリットブレインの検出

- メタデータのGI (世代識別子)
  - ノード間の接続/切断履歴を記録
  - 過去にスプリットブレインが生じたことを確実に検出できる

株式会社サードウェア
Thirdware Inc.

スプリットブレインになったら

- DRBDは相手との接続を拒否
  - cs:StandAloneまたはcs:WFConnection
- スプリットブレイン状態の解除が必要
  - 2ノードの状態変化履歴や
  - 2ノードのデータ状態にもとづいて
  - どちらかのノードのデータを修復
  - 他方のノードのデータを破棄

株式会社サードウェア
Thirdware Inc.

discard-my-data

- データを破棄するノードで

```
drbdadm -- --discard-my-data connect <リソース>
```

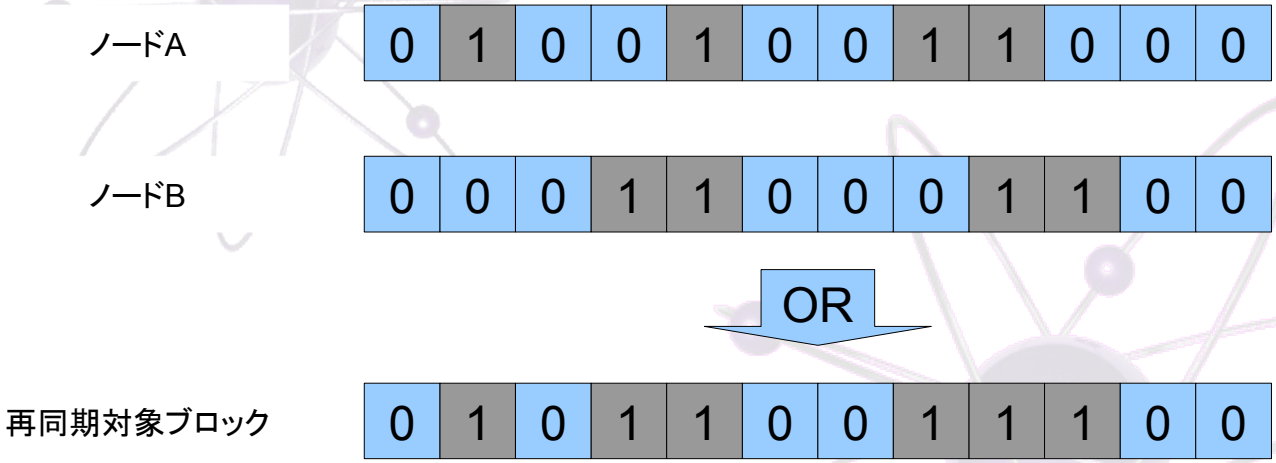
- 他ノードで

```
drbdadm disconnect <リソース>
drbdadm connect <リソース>
```

- DRBD間の接続が回復し、データ同期が始まる

株式会社サードウェア
Thirdware Inc.

discard-my-data



データ整合性

- データ整合性に疑いが生じた場合
  - ネットワークトラブルがあった
  - 間違って下位ブロックデバイスをいじった
  - データ整合性の保証が必要なアプリケーション

株式会社サードウェア
Thirdware Inc.

drbdadm verify

- DRBD 8.2.5から
- バックグラウンドで全ブロックを照合
  - verify-algが必要
  - カーネルのダイジェストアルゴリズムを指定 (md5、sha1、crc32cなど)
- 照合結果はsyslogに表示される

株式会社サードウェア
Thirdware Inc.

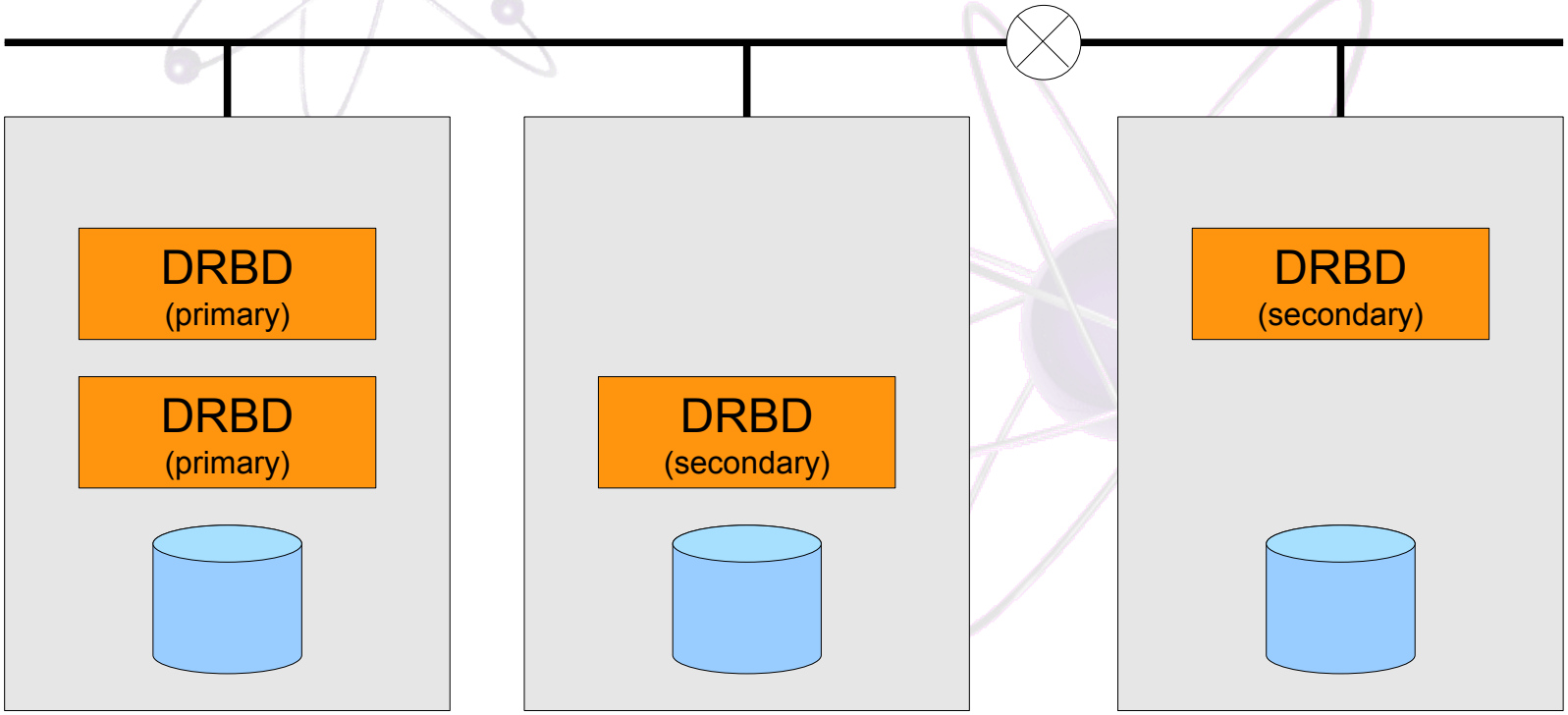
drbdadm verify

- データ不整合が見つかったら
  - drbdadm disconnect、drbdadm connect
  - 再接続時にバックグラウンドで同期しなおす

株式会社サードウェア
Thirdware Inc.

3ノードレプリケーション

- DRBD 8.3.0から
- DRBDをスタックする



株式会社サードウェア
Thirdware Inc.

3ノードレプリケーション

- 下位DRBD
  - HAクラスタの両ノードで動作
- 上位DRBD
  - 下位プライマリDRBDの上で動作
  - HAクラスタのリソースとして登録、実行

株式会社サードウェア
Thirdware Inc.

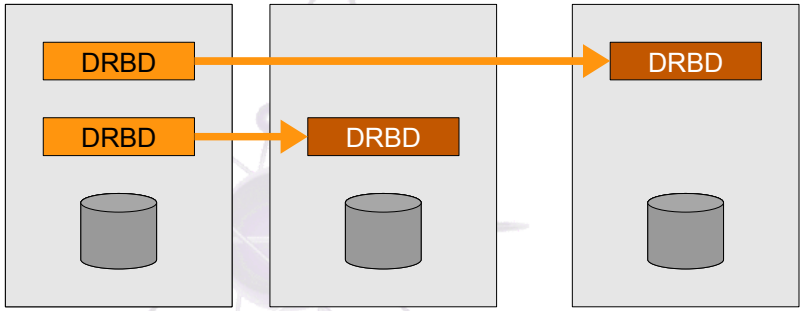
drbd.conf

```
resource r0 {
    protocol C;
    ....
    on node1 {
        device /dev/drbd0;
        disk /dev/sda5;
        address 10.0.0.1:7788;
        meta-disk internal;
    }
    on node2 {
        device /dev/drbd0;
        disk /dev/sda5;
        address 10.0.0.2:7788;
        meta-disk internal;
    }
}
```

```
resource r0U {
    protocol A;
    ....
    stacked-on-top-of r0 {
        device /dev/drbd10;
        address 222.22.2.2:7778;
    }
    on node3 {
        device /dev/drbd10;
        disk /dev/vg00/lv00;
        meta-disk internal;
    }
}
```

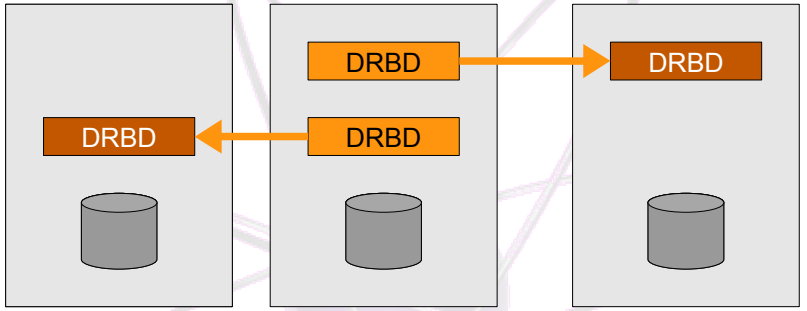
株式会社サードウェア
Thirdware Inc.

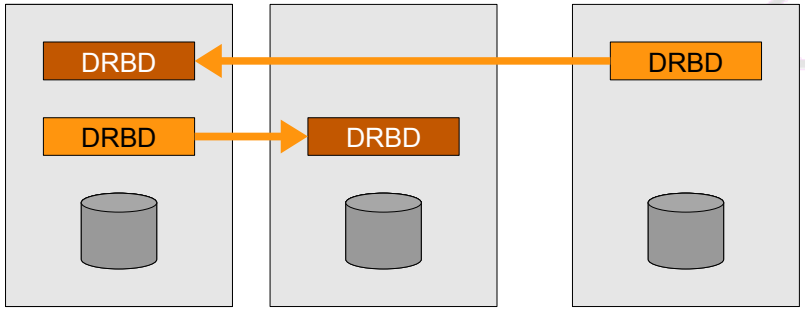
3ノードレプリケーション



クラスタとバックアップ

クラスタとバックアップ
(フェールオーバ)





バックアップサーバから
データをリストア

株式会社サードウェア
Thirdware inc.

おまけ

DRBDを使って
バックアップするのに

おまけ

LVMスナップショットとの
併用がお勧め
だと

株式会社サードウェア
Thirdware inc.

おまけ

言ってきました。

株式会社サードウェア
Thirdware inc.

おまけ

朗報です！！

おまけ

もっといい方法がありました

株式会社サードウェア
Thirdware Inc.

おまけ

- --discard-my-dataをうまく活用
- セカンダリノードをdisconnect
- セカンダリノードをprimaryに変更
  (意図的にスプリットブレイン状態にする)
- バックアップを実行
- セカンダリノードをsecondaryに戻す
- --discard-my-data付きで再接続

株式会社サードウェア
Thirdware inc.

おまけ

ただし

株式会社サードウェア
Thirdware Inc.

おまけ

- /dev/drbd*n*経由でアクセスすることは必須です。
- dopdは使えません。
- バックアップ中、プライマリサーバのデータが単一障害点になります。
- 3ノード構成の3台目で使うのに向いたソリューションです。