



VIVER

[váivər]

次世代Linuxディスクレスブート
テクノロジー VIVER

6/22 2007 © FURUHASHI Sadayuki

1. VIVERとは？
2. VIVERの使い方
3. VIVERの仕組み
4. V-FIELD

自己紹介



氏名：古橋貞之（ふるはしただゆき）

年齢：19

キーワード：冗長化、負荷分散、動的、
C++、Ruby、筑波大学、未踏ユース

自己紹介

2006年度未踏ユースで
VIVER 0.3を開発

氏名 スーパークリエーター(ふるはしただゆき)

年齢：19 認定

キーワード：冗長化、負荷分散、動的、

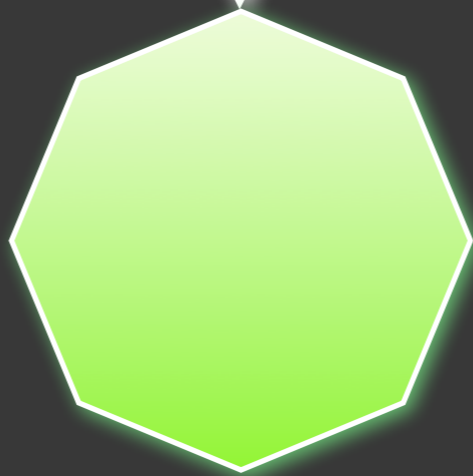
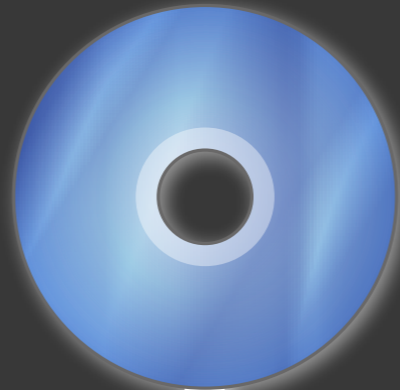
C++、Ruby、筑波大学、未踏ユース



I

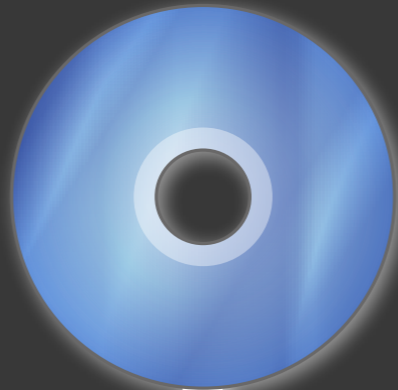
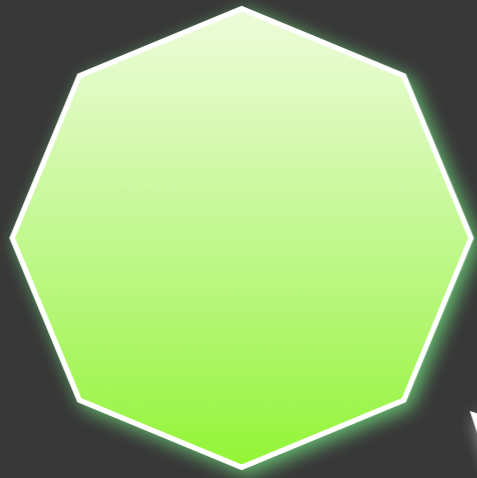
WINTER

とは？

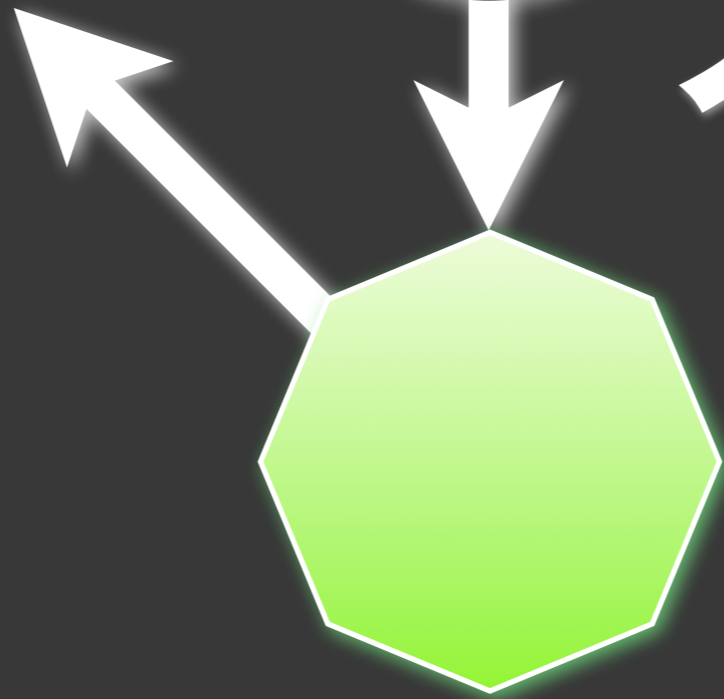


ブート
メディア

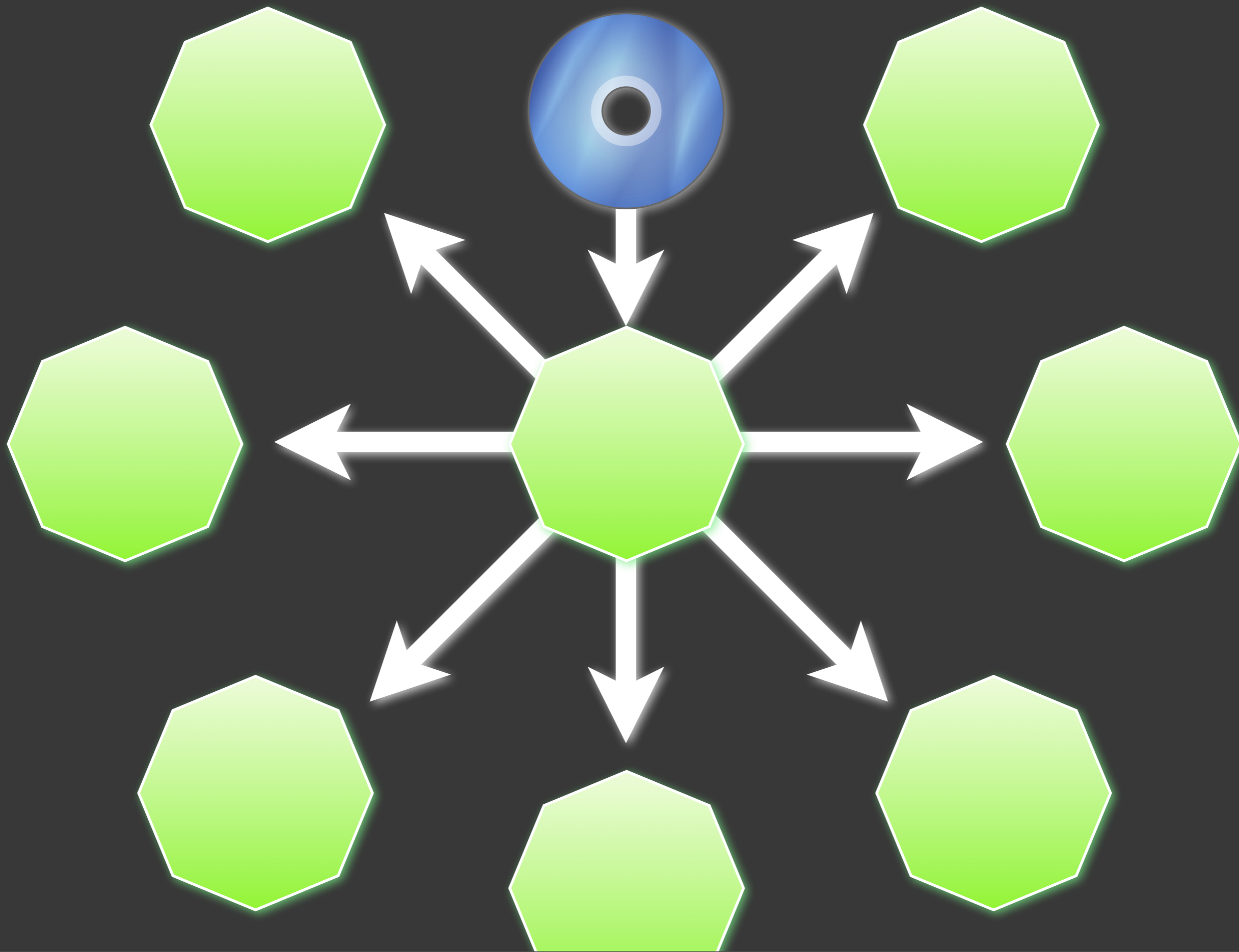
起動



ブート
メディア



ネットワーク
ブート (PXE)



Live Linux

(KNOPPIXなど)

とは違う？

全然違ふ

似ている

とこる：

**最初の一は
CDから起動**

→USBメモリ
でも可

→ HDD

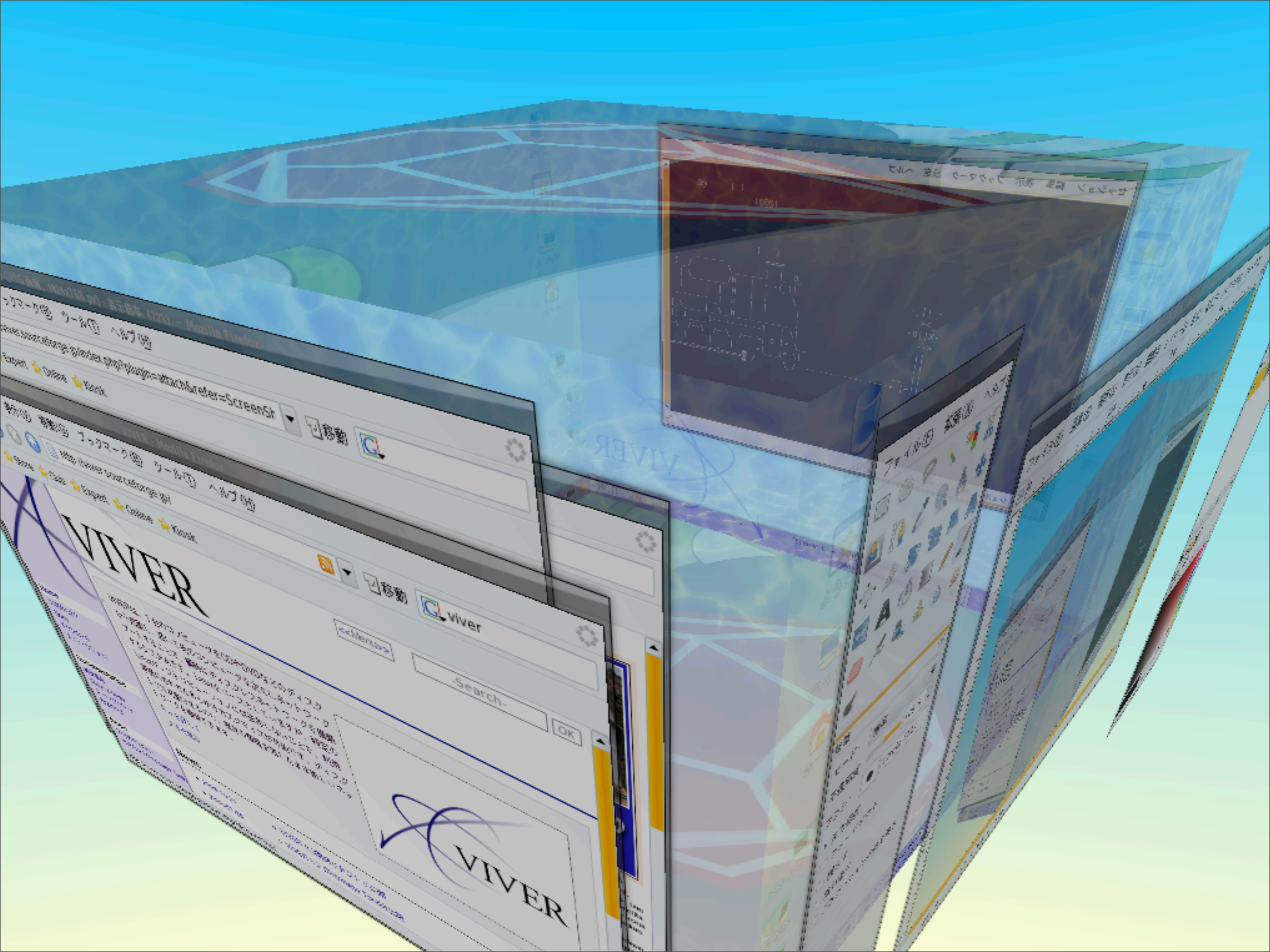
でも可

→SDカード
でも可

ハードウェア

自動検出

→ 設定不要で
すぐ使える



NFS-Root

(NFSを/にして
ネットワークブート)

とは違う？

すこし

似ている

とくらひ違ふ。：

**最初の一は
CDから起動**

→サーバーの

設定不要

ハードウェア

自動検出

→ 異種混合環境

でもOK

**NFSを
使わない**

高速分散多重化

共有ブロック

デバイス

「V-FIELD」

A decorative graphic consisting of several overlapping, flowing lines in shades of blue and purple, starting from the left edge and curving upwards and then downwards, framing the word 'Demo'.

Demo

VIVER

ネットワークブート

II

VIEWER

の使い方

Step Ø

HDDに

Linuxを

インストール

Debianでも

CentOSでも

Momongaでも

Mandrivaでも

Step 1

**VIVER化キット
をインストール**

<http://viver.sourceforge.jp/>

or

VIVER

検索

カチッ♪

Step 2

ブートディスク を作る

squashfs

+

mkisofs

Step 3

起動！

起動！

起動！

Step 4

コーヒーを飲む

Step 5

Hackする

III

WIVVER

の仕組み

通常Linuxのブートプロセス

BIOS



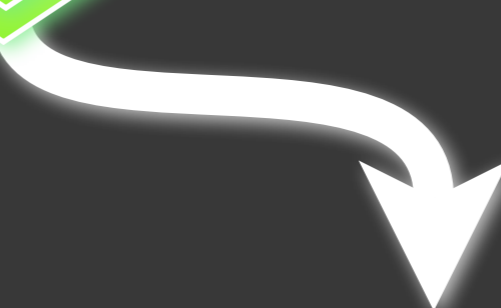
GRUB

RAMにロード



initrd

kernel



/

BIOS



GRUB

RAMにロード



initrd

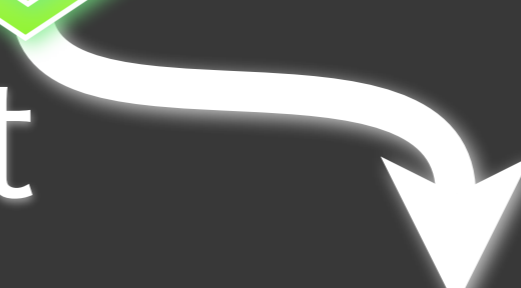
kernel

/init → shell script

SATAドライバ

ext3ドライバ

...



BIOS



RAMにロード

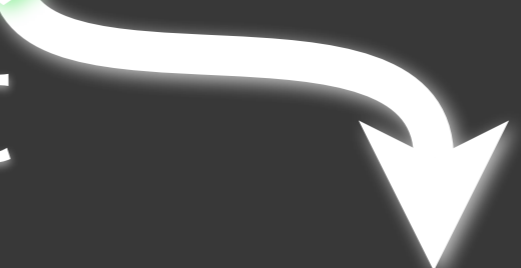
GRUB



initrd

kernel

→ shell script



HDD



mount



VIVERのブートプロセス

BIOS



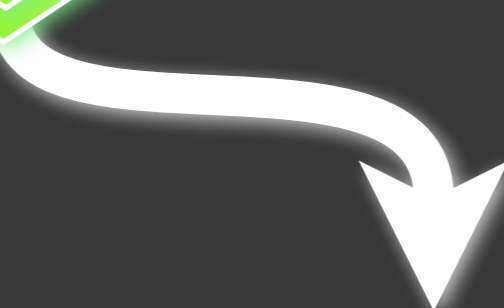
GRUB

RAMにロード



initrd

kernel



/



BIOS

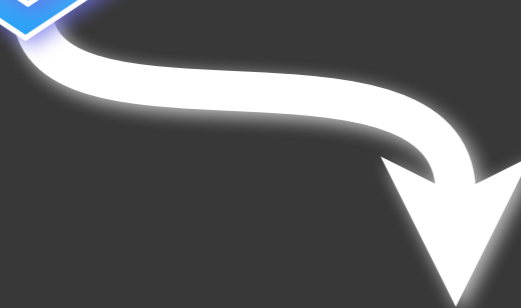


RAMにロード

GRUB

VIVER

kernel



/

BIOS



GRUB

RAMにロード

VIVER

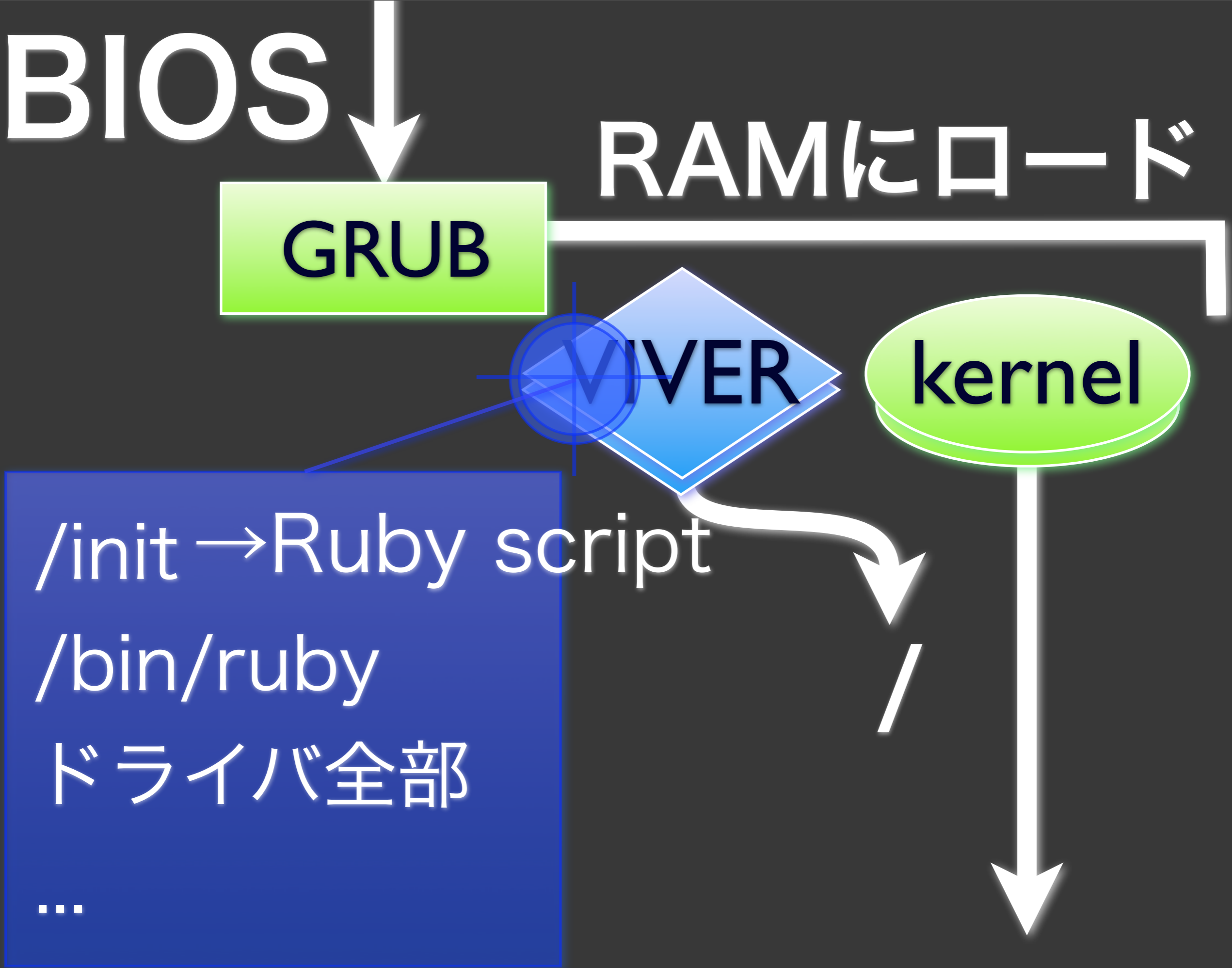
kernel

/init → Ruby script

/bin/ruby

ドライバ全部

...



BIOS



GRUB

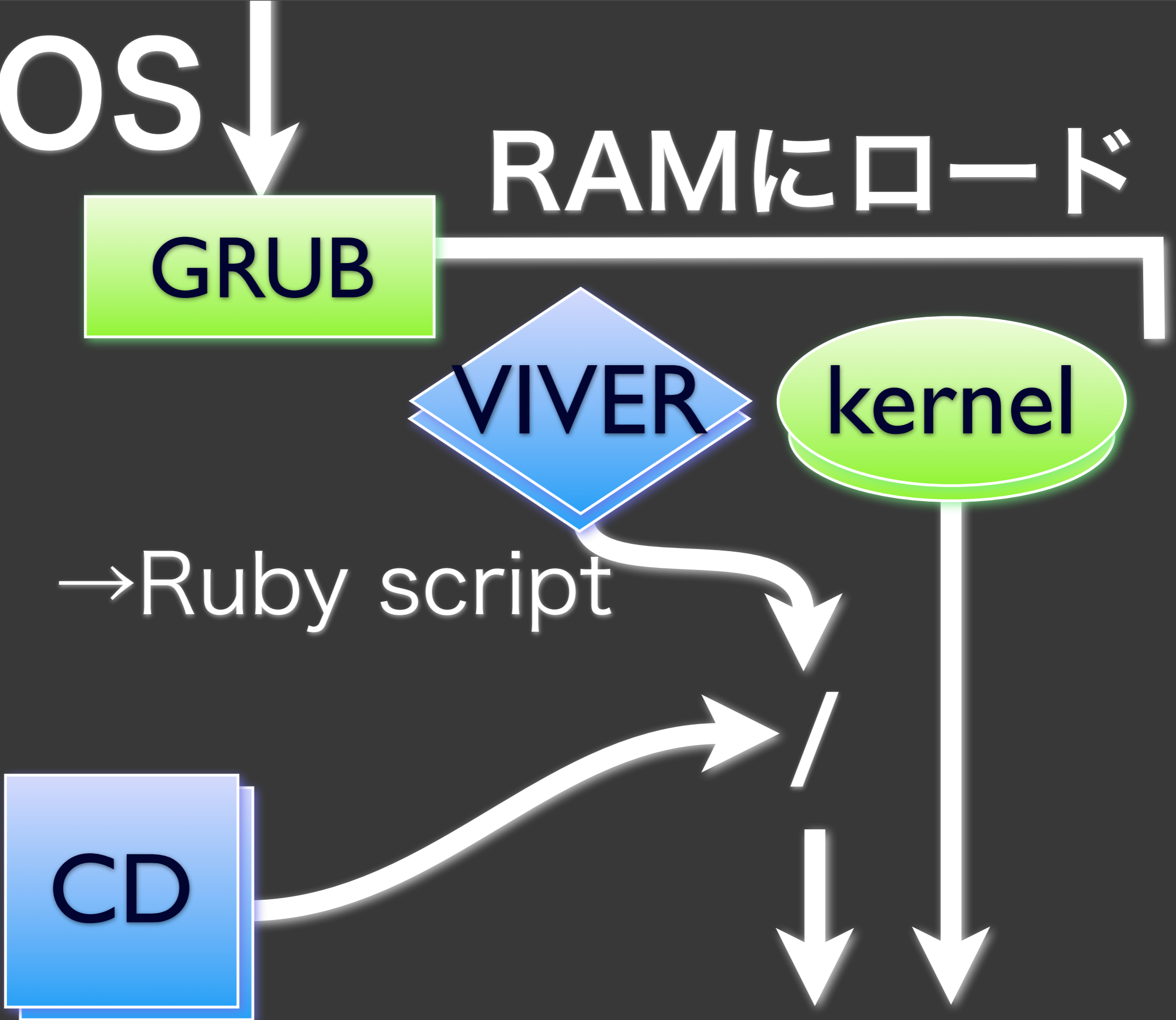
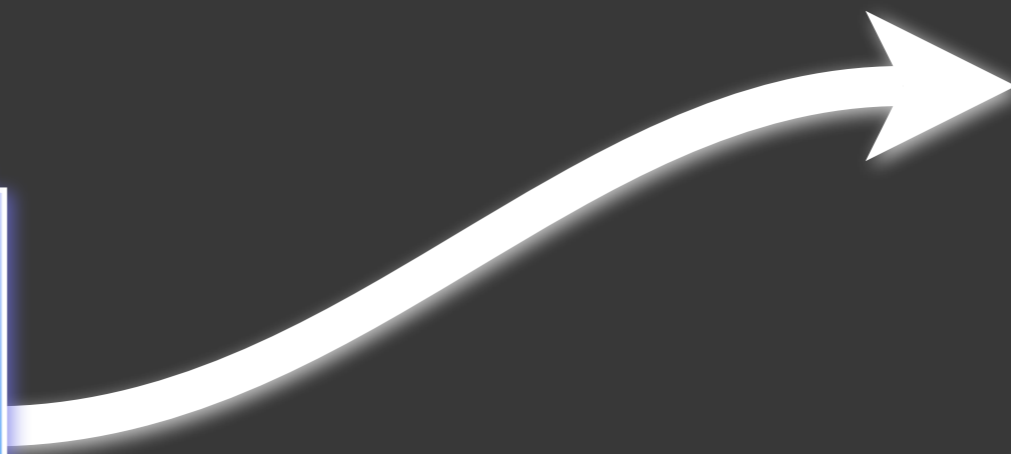
RAMにロード

VIVER

kernel

→ Ruby script

CD



BIOS



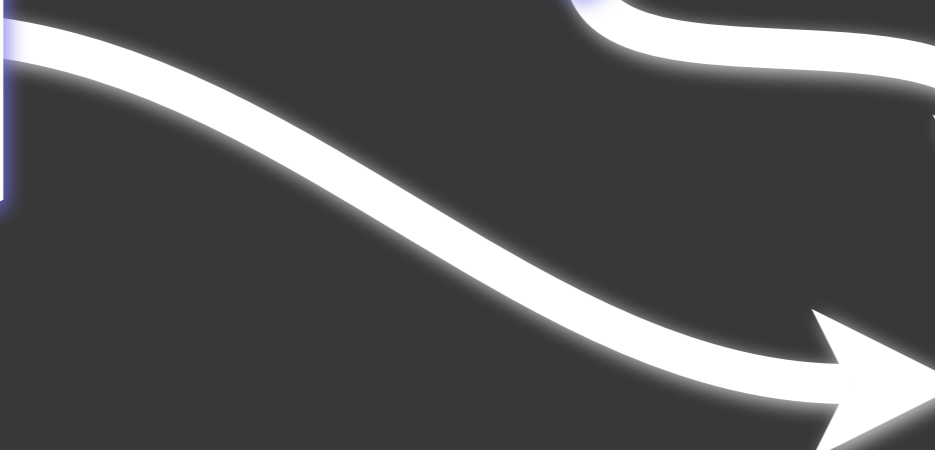
RAMにロード

GRUB

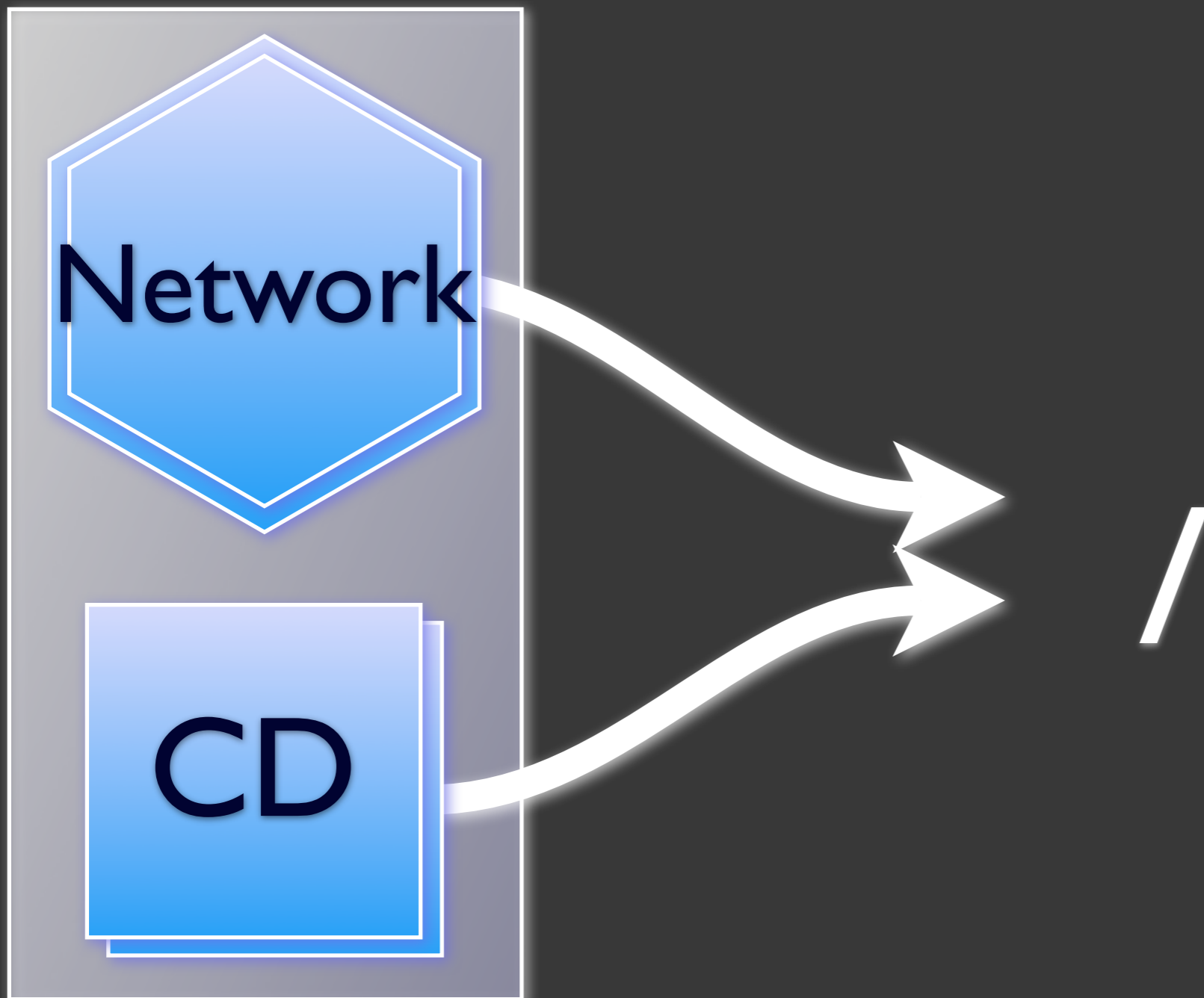
VIVER

kernel

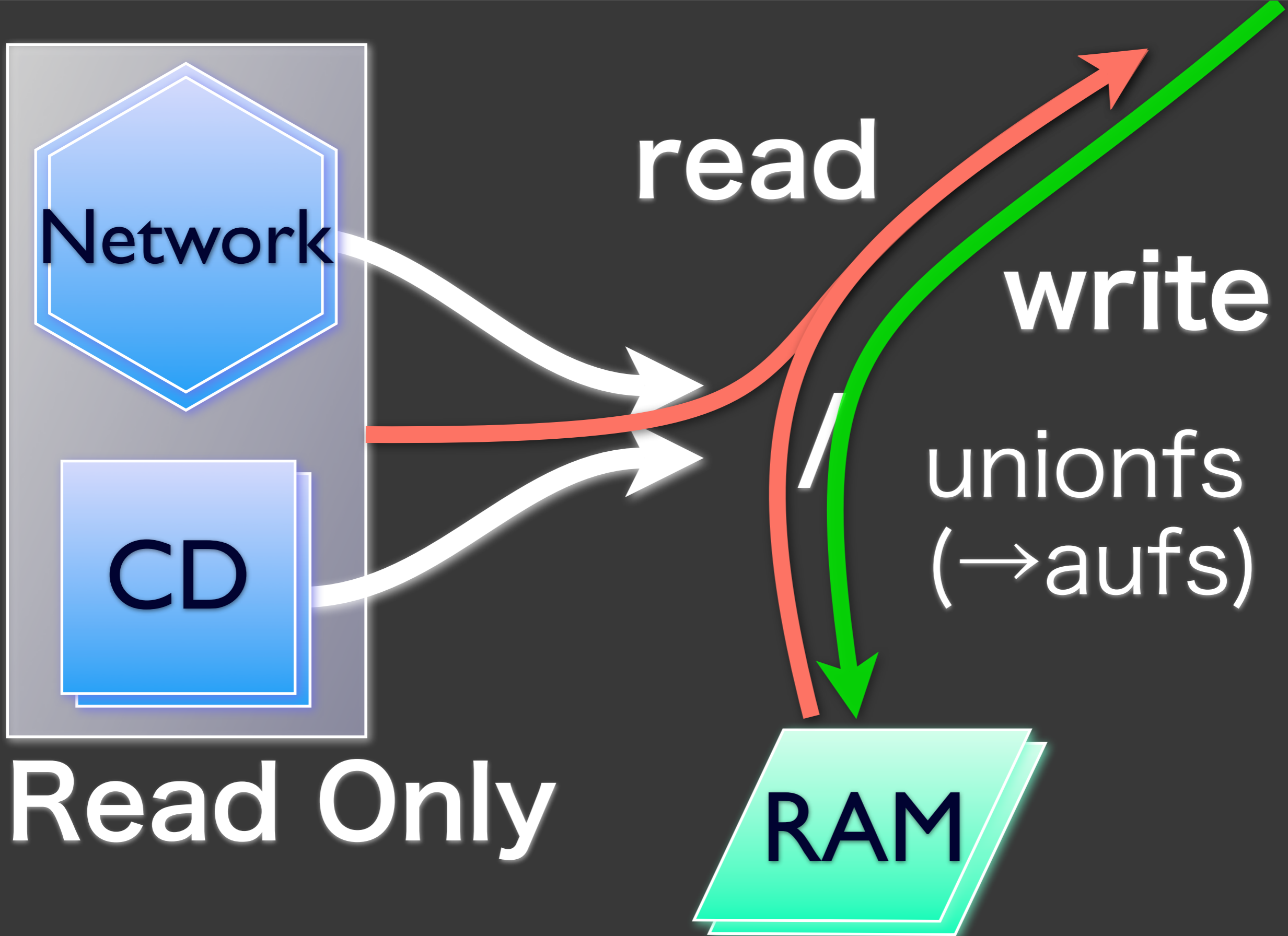
Network



Copy-on-Write



Read Only



Network

CD

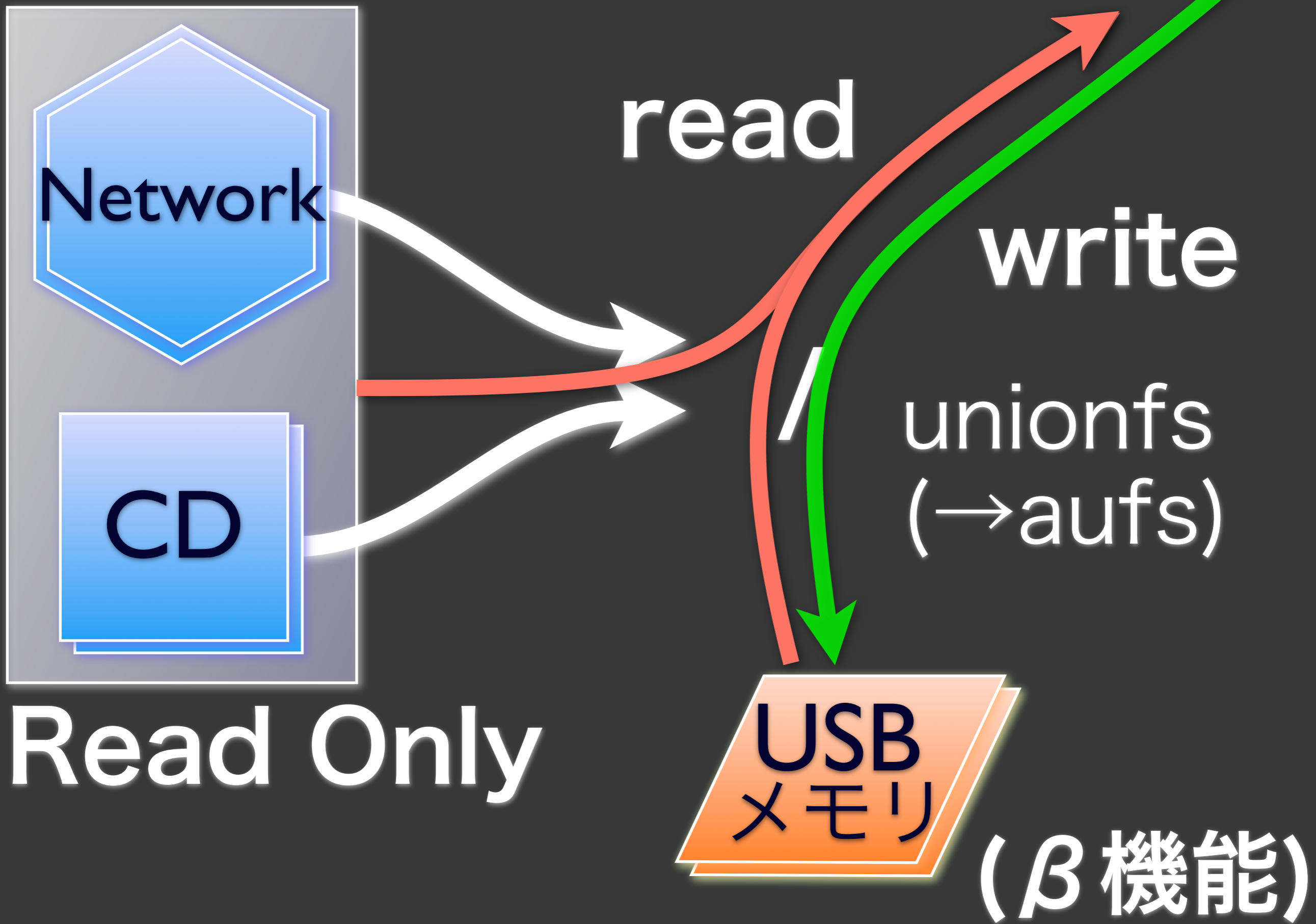
Read Only

read

write

unionfs
(\rightarrow aufs)

RAM



Network

CD

Read Only

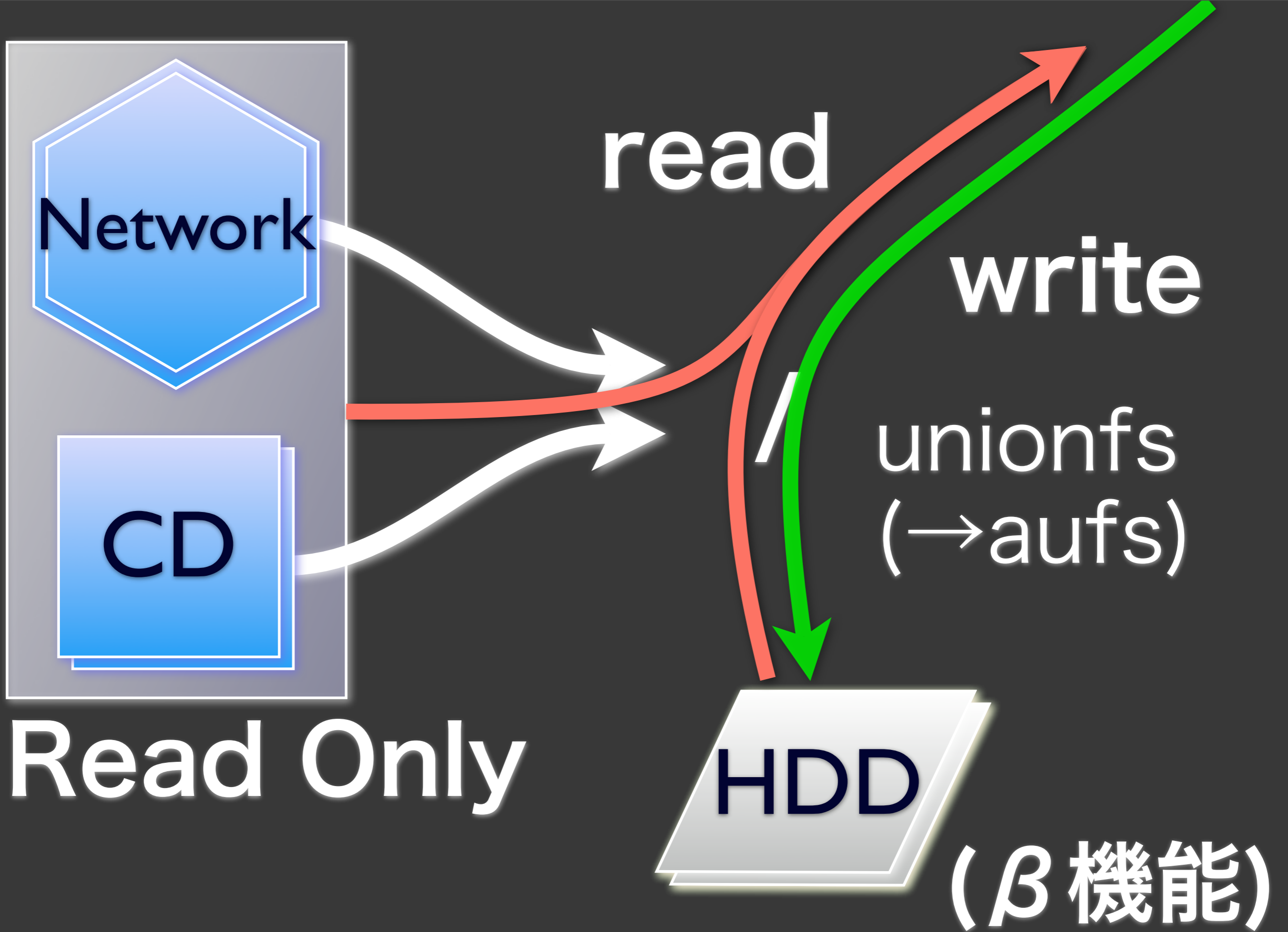
read

write

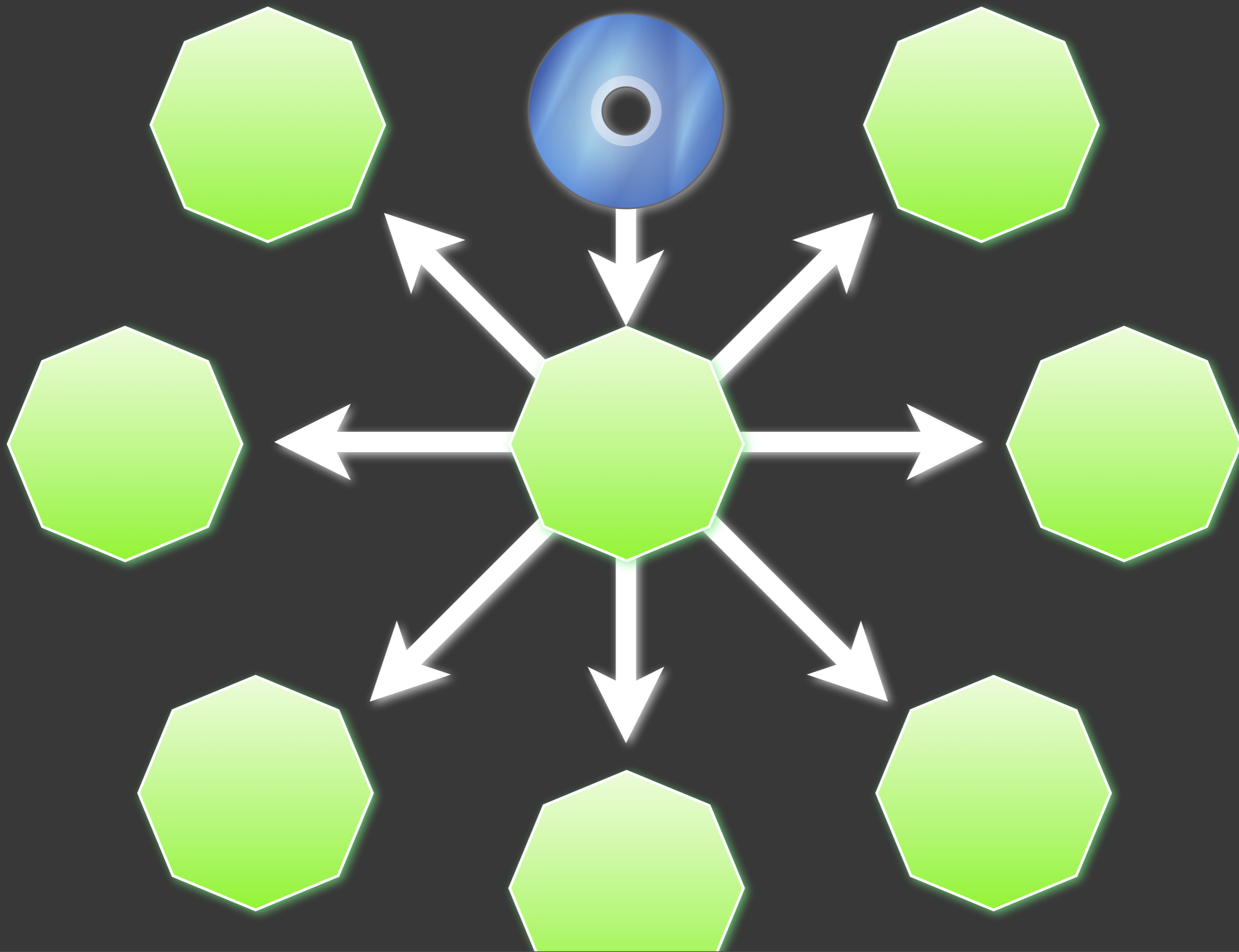
unionfs
(→aufs)

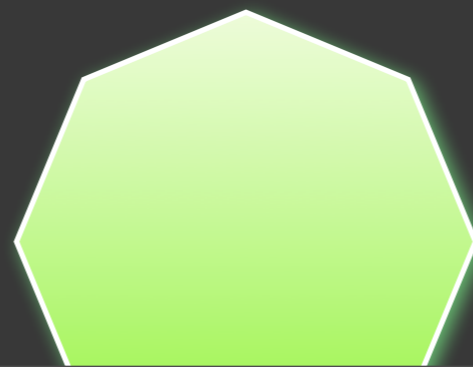
USB
メモリ

(β機能)



とこるで...



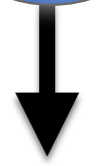
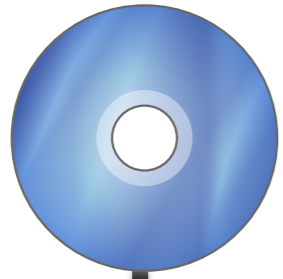


というのは
ダメですよね

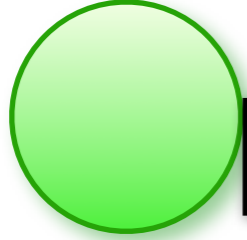
で じ ぞ

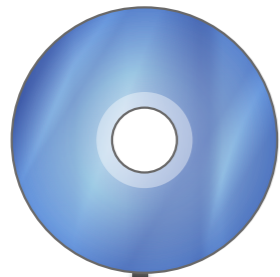
IV

V-FIELD

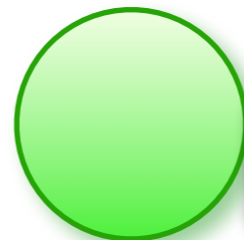
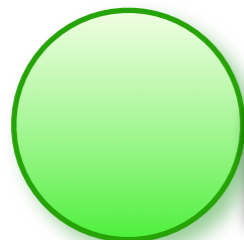
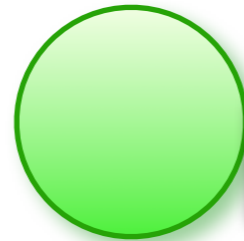
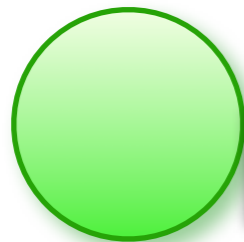
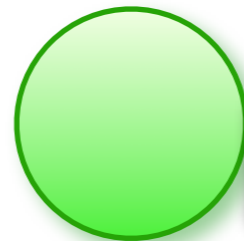
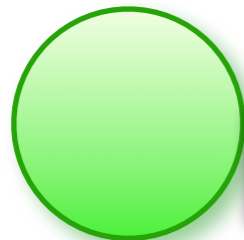
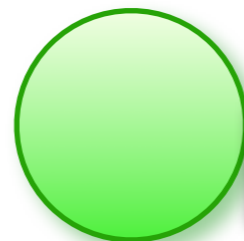
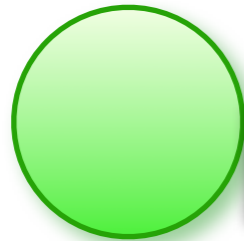
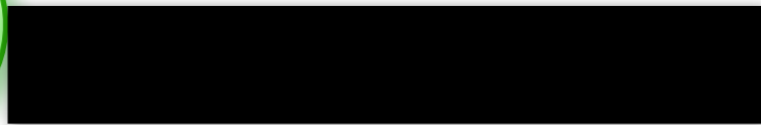
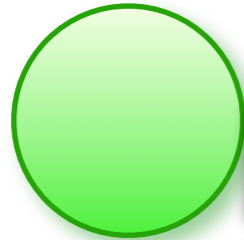


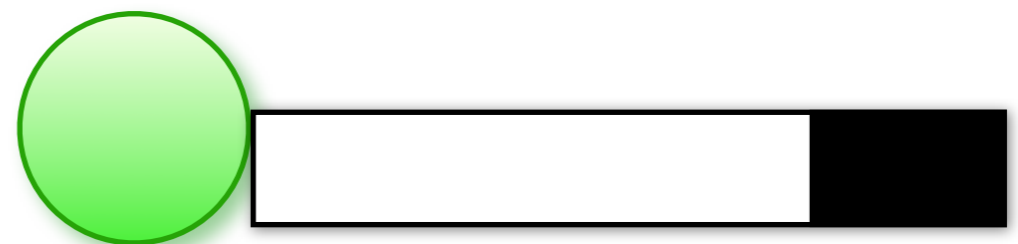
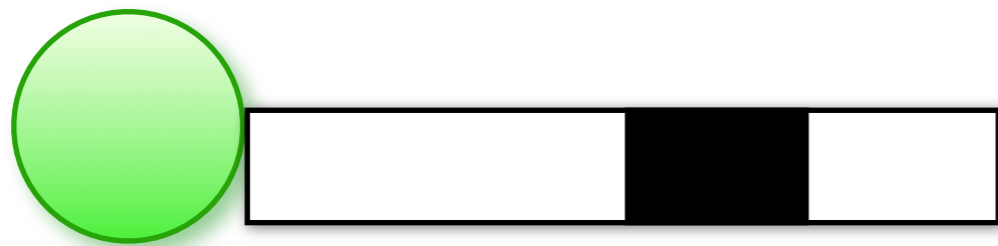
↓ディスクイメージ





↓ディスクイメージ





元 長 化

負荷

分散

動的な

サーバー

の追加

A decorative graphic consisting of several overlapping, flowing lines in shades of blue and purple, starting from the left edge and curving upwards and then downwards, framing the word 'Demo'.

Demo

**高速分散多重化
共有ブロックデバイス
「V-FIELD」**

転送性能

V-FIELD

P2P型

NBD

(Network Block Device)

クライアント/サーバー型

300MBのディスクイメージを
共有し、全ノードで一斉に
シーケンシャルリード

IBM Blade Center HS20

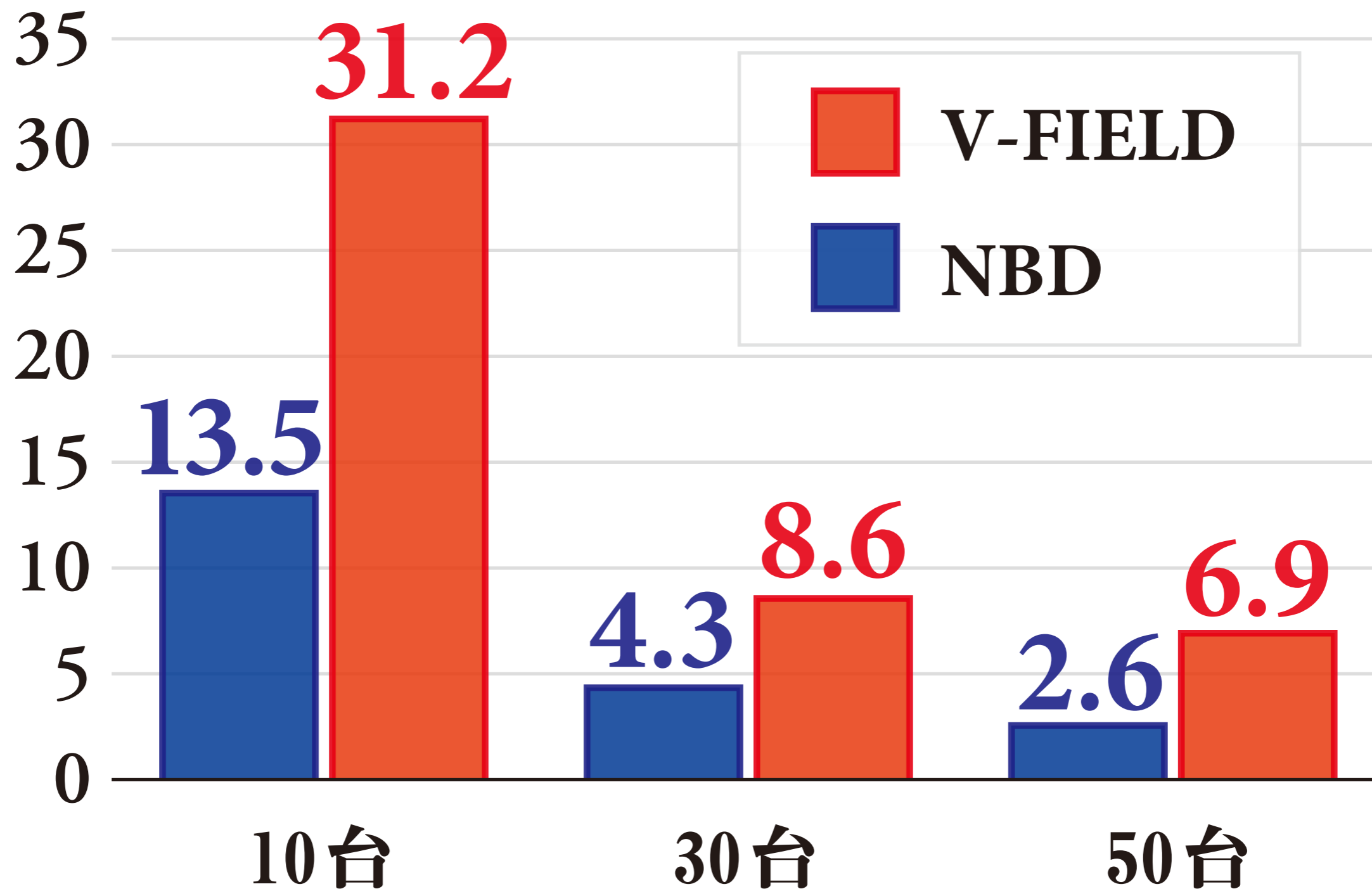
Xeon 3.6GHz ×2

Memory 2GB

NIC BCM5604s (GbE)

NBD: 1对多, V-FIELD: 多对多

(MB/s)



問題点

**V-FIELDは
Read Only**

→元のディスクイメージ
も書き換え不可

正常に終了

できない

→ **電源ボタン長押し**

→ **Copy-on-Write**で
RAMを使っているなら
問題なし

（どちらにしても）
（データは消える）

**まだまだ開発を
続けています**

分散ネットワークを自己組織化

MailServer

log (msg)

LogServer

mail (msg)

Server C

HealthCheck

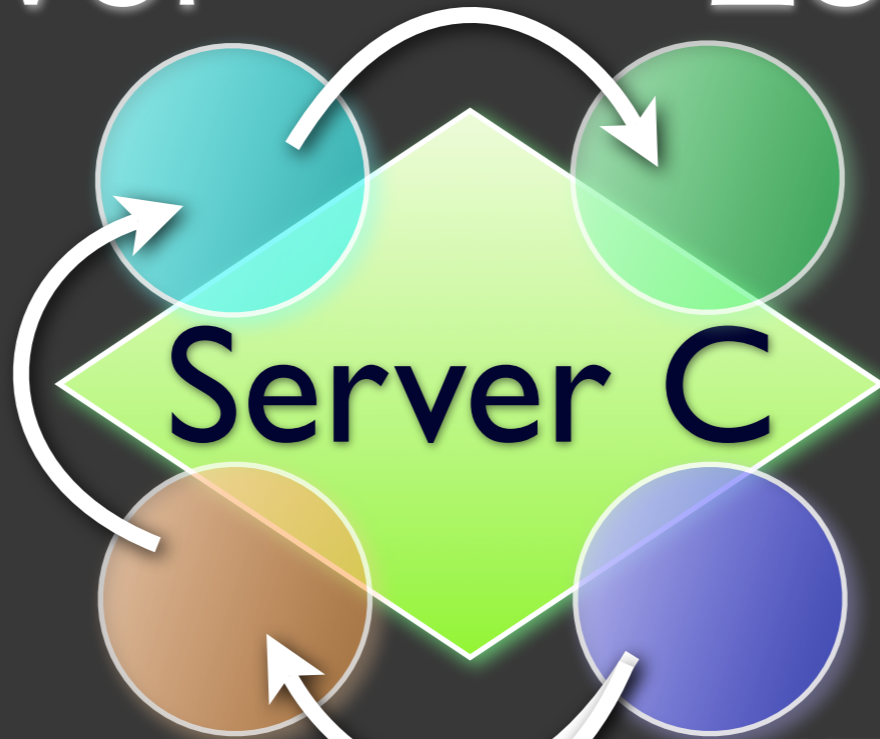
HTTPServer

regist (:http)



MailServer

LogServer



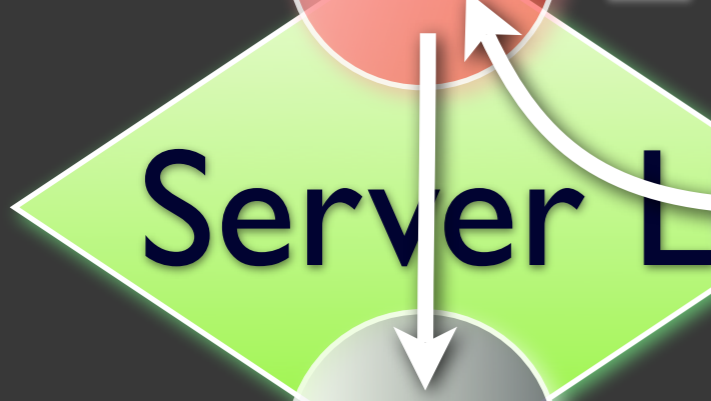
Server C

HTTPServer

regist(myip)



LoadBalance



Server L

HTTPServer



VRRP



MailServer

LogServer

Server C

HTTPServer

LoadBalance

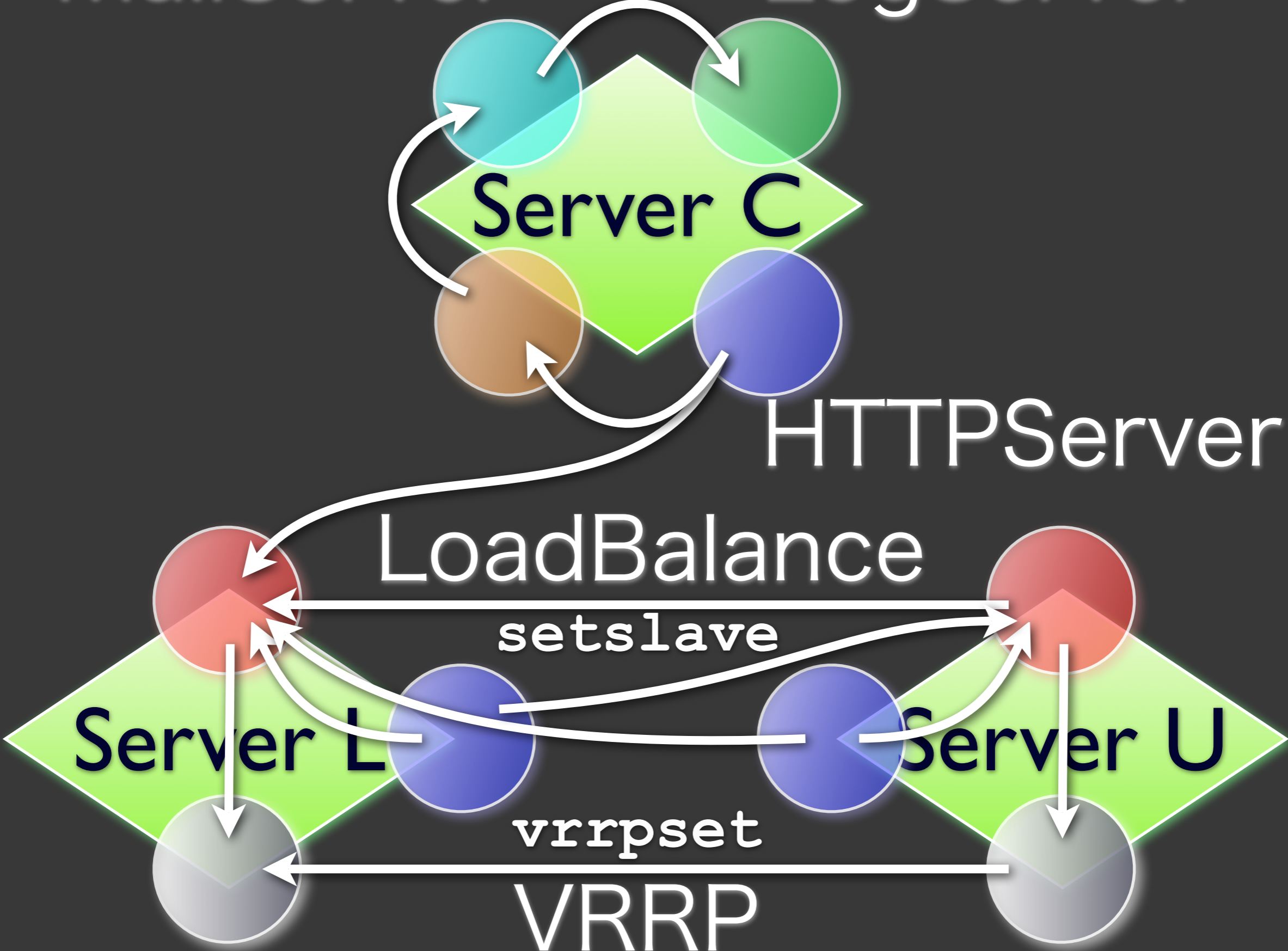
setslave

Server L

Server U

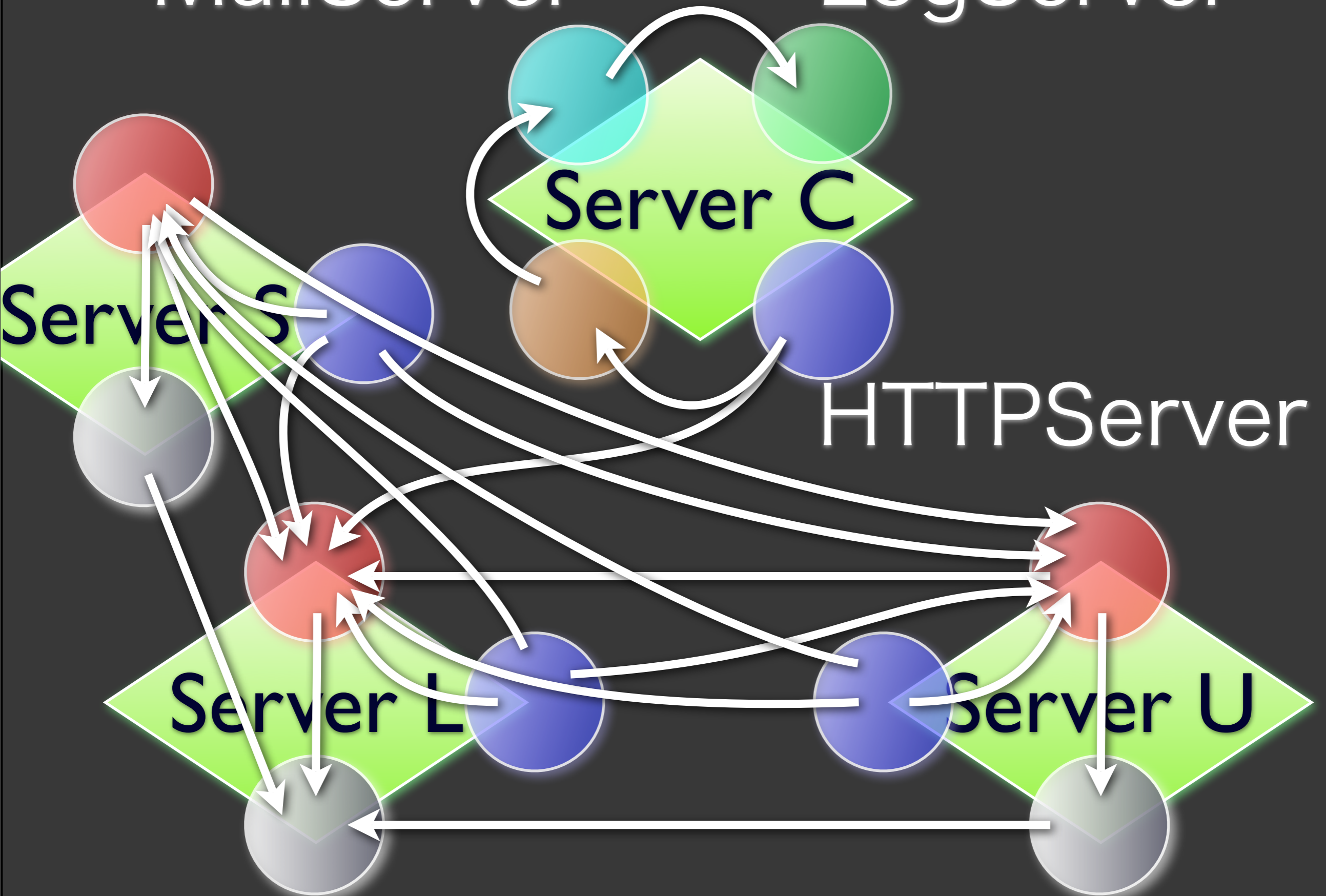
vrrpset

VRRP



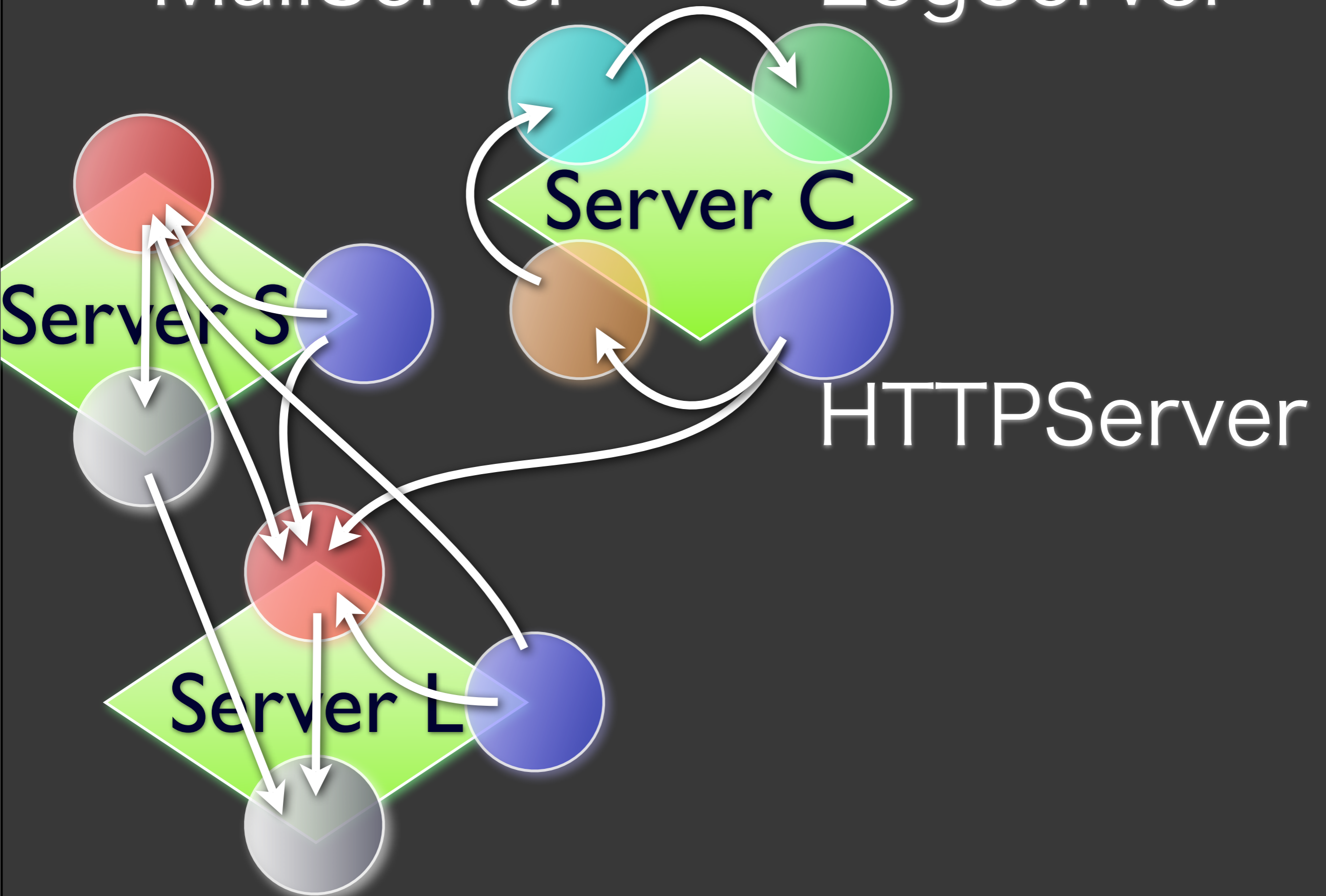
MailServer

LogServer



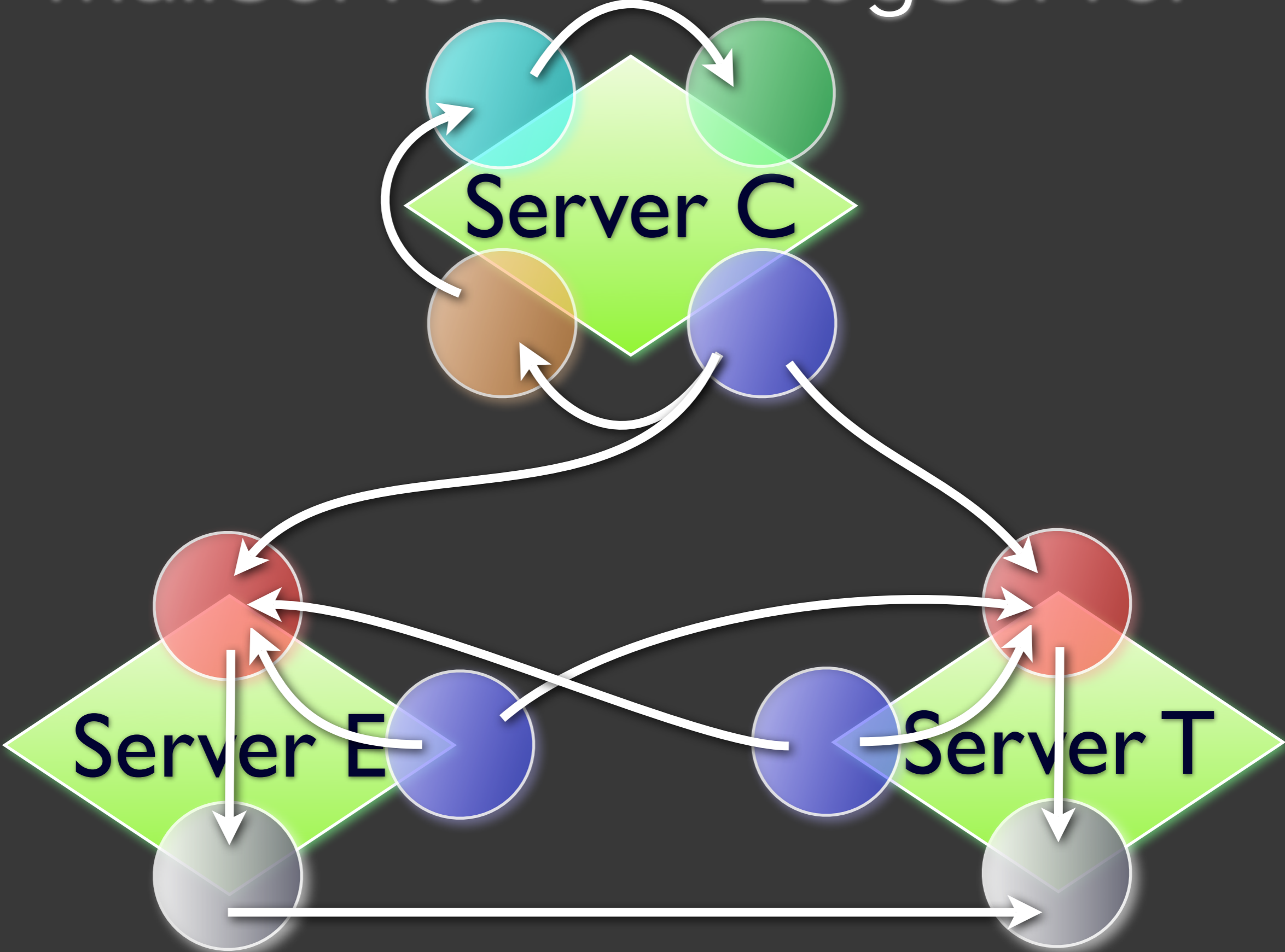
MailServer

LogServer



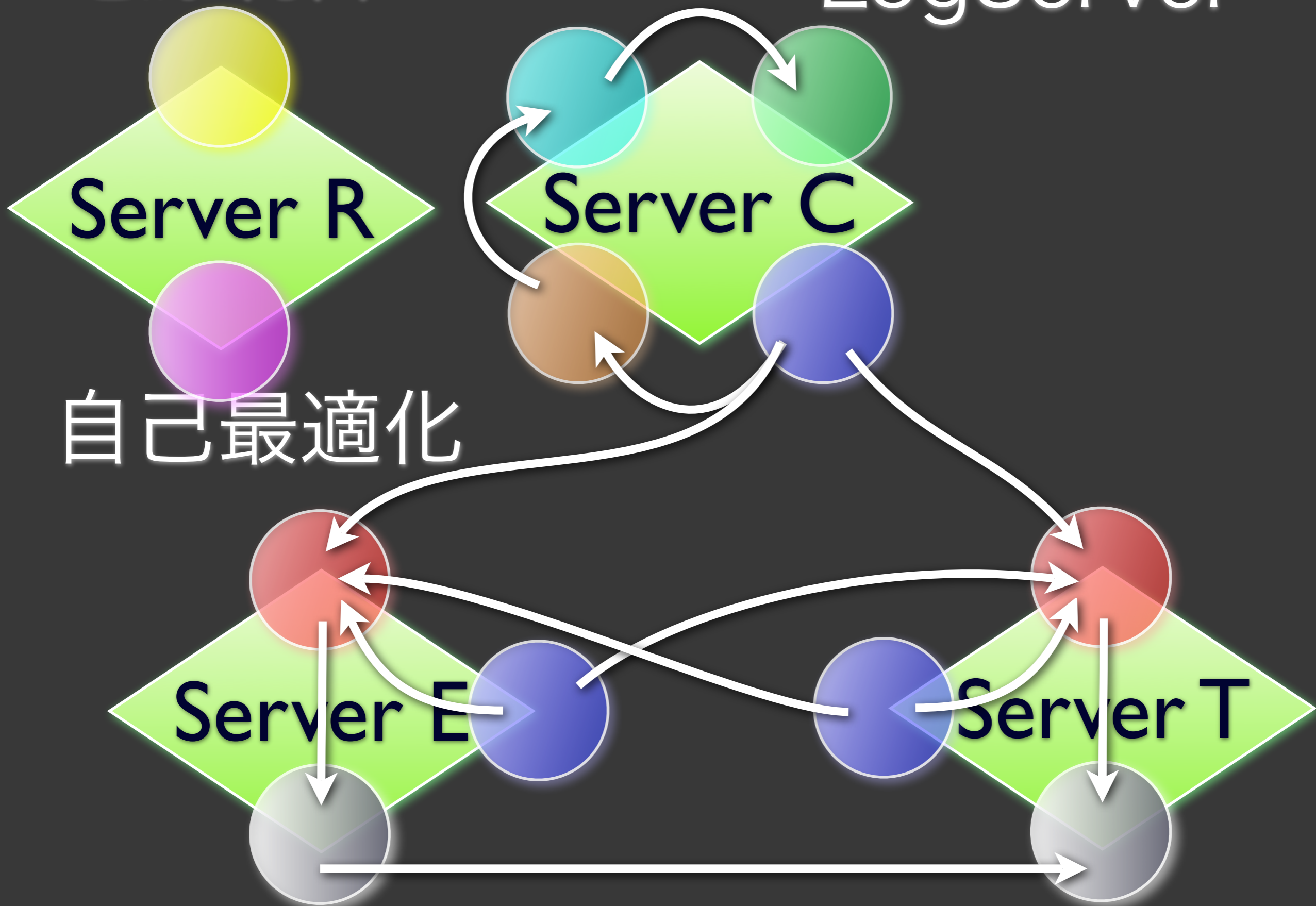
MailServer

LogServer



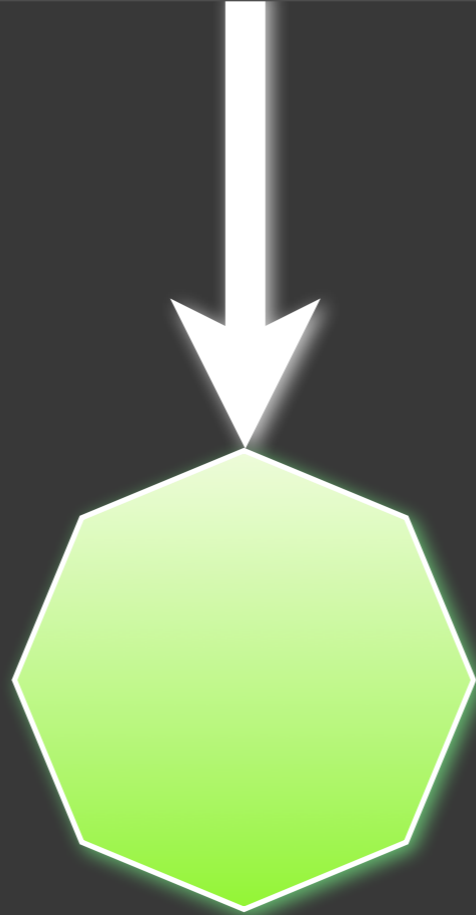
電源制御

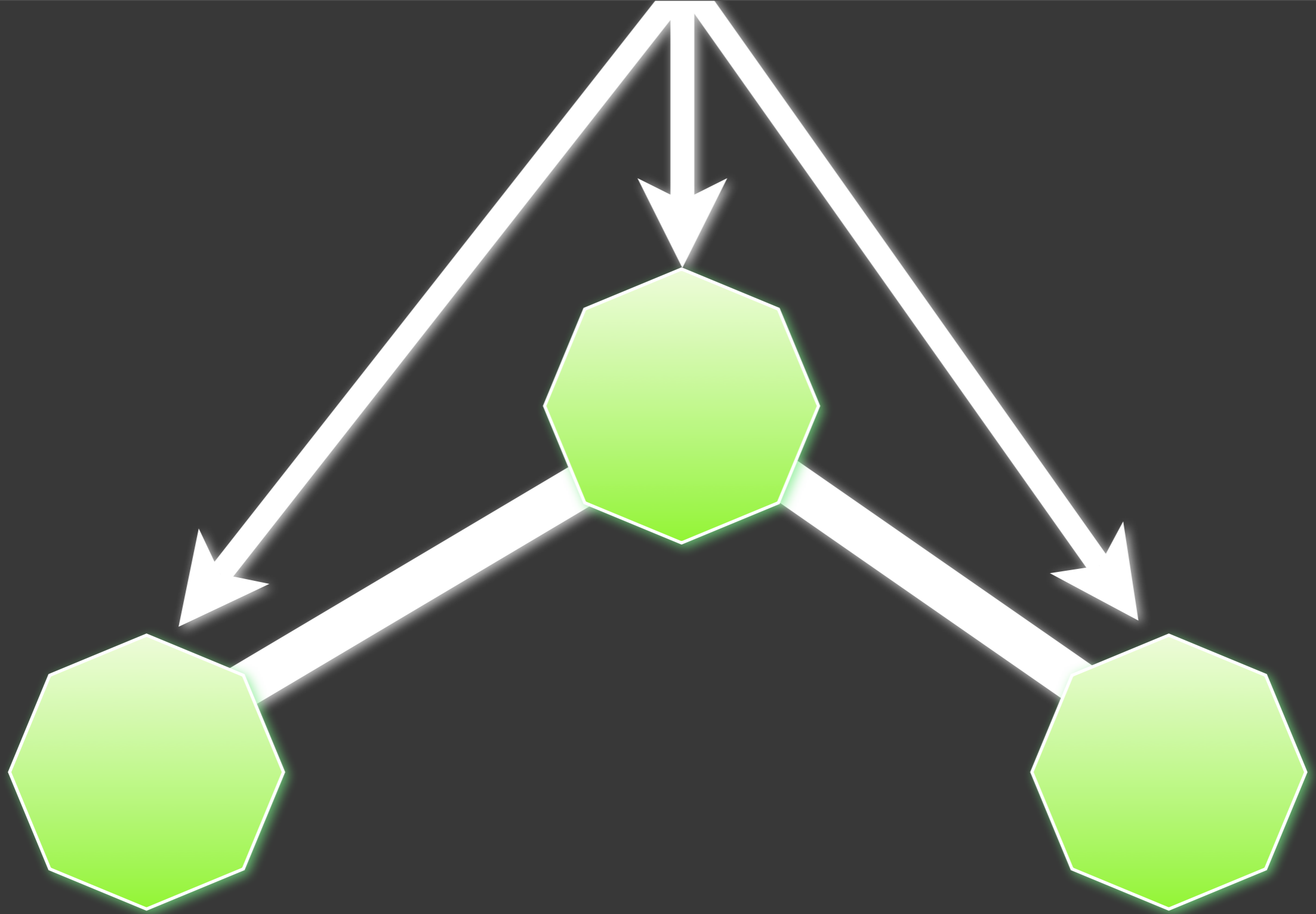
LogServer

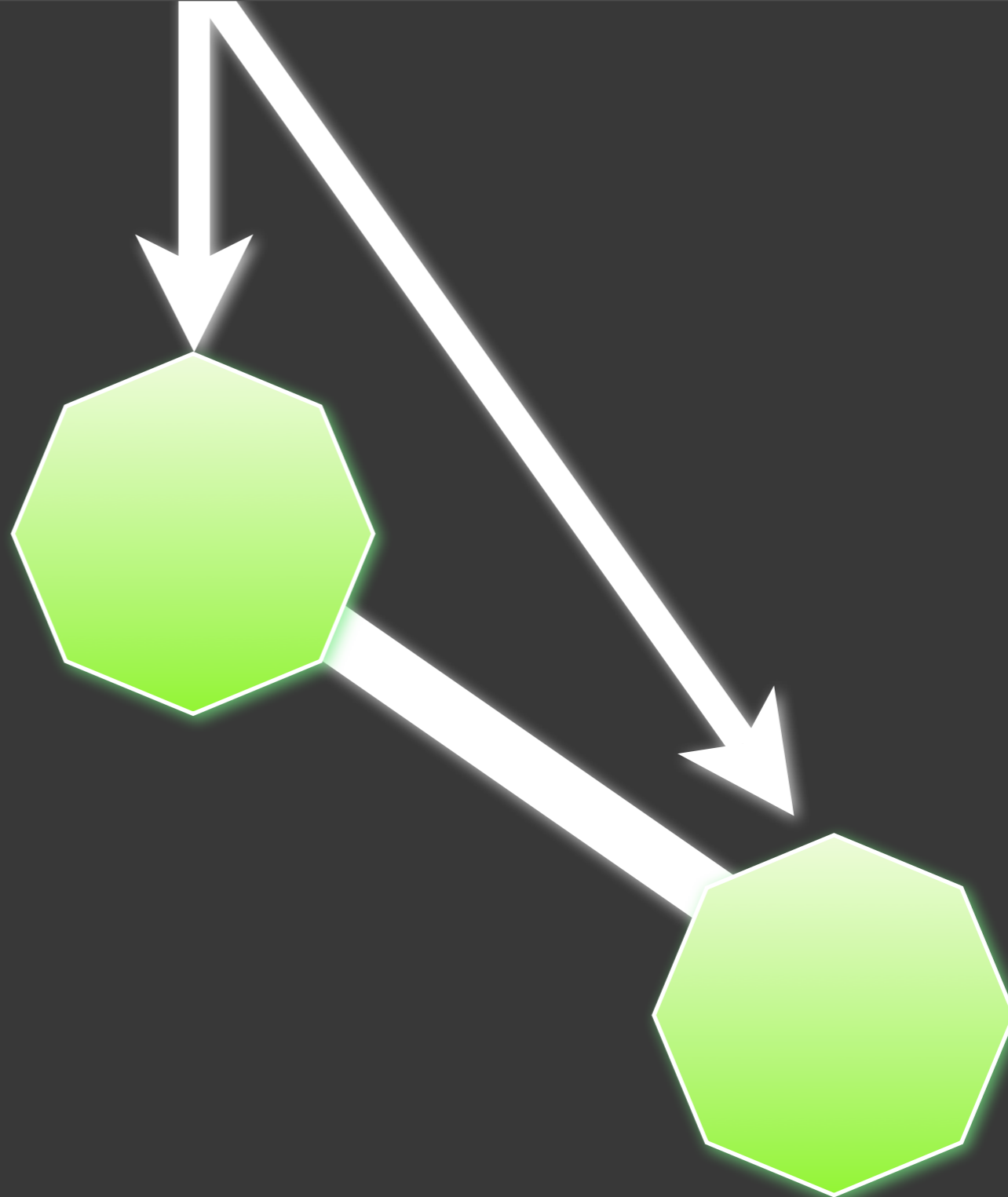


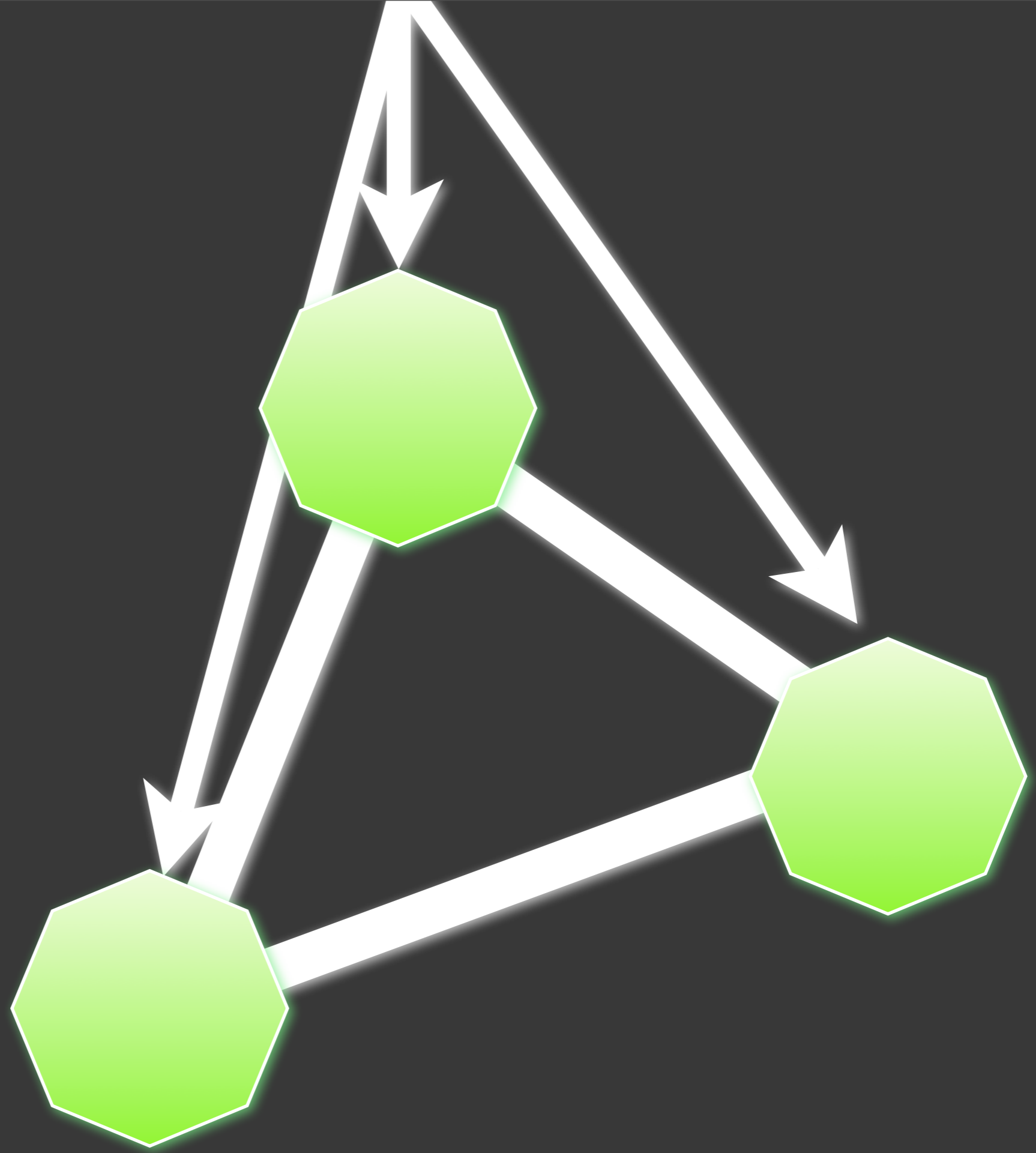
自己最適化

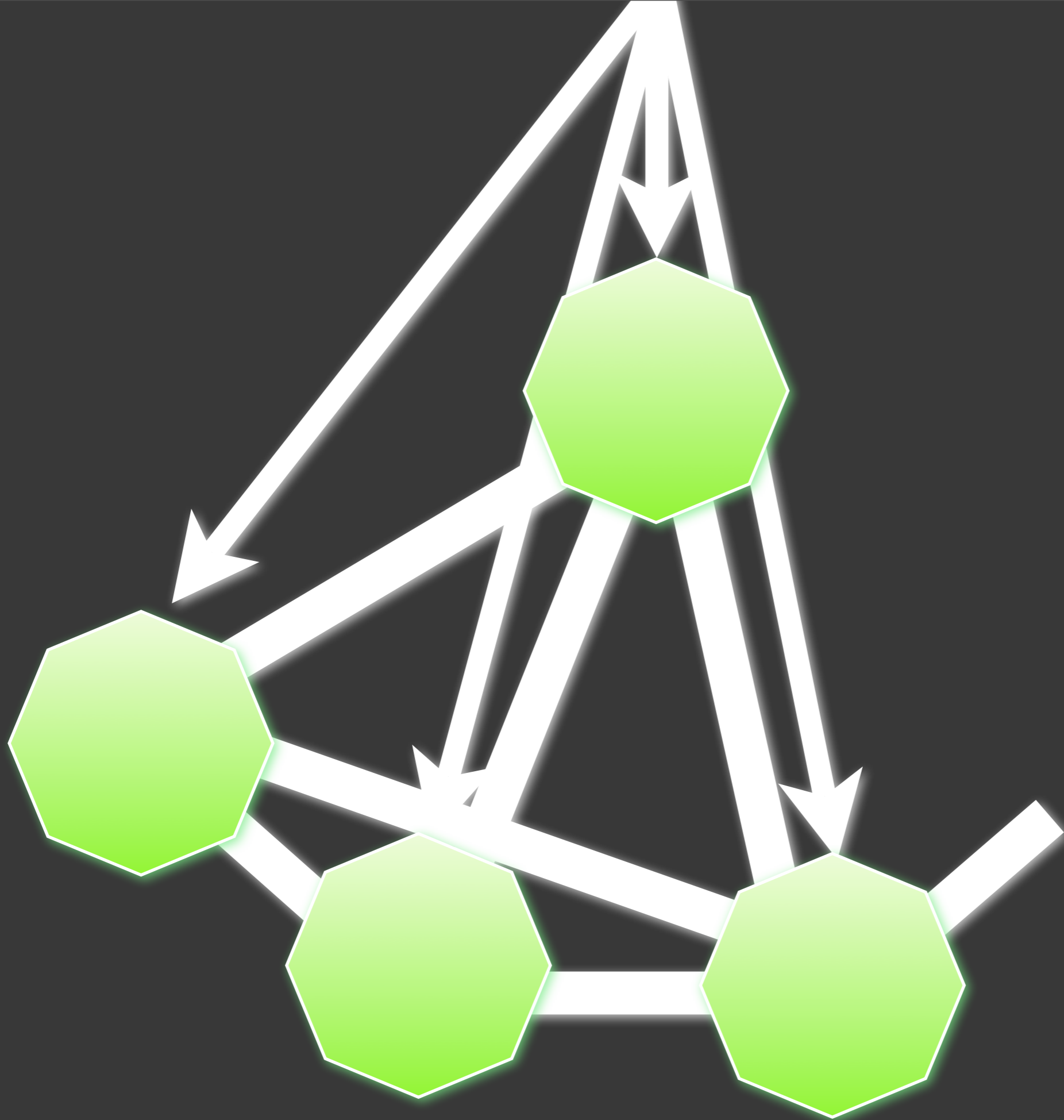
分散ファイルシステム



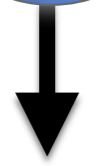
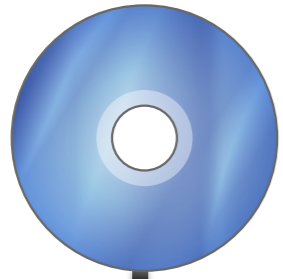




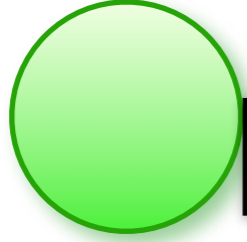


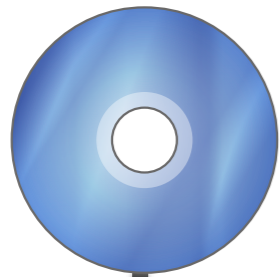


ご静聴ありがとうございました。

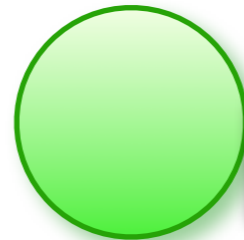
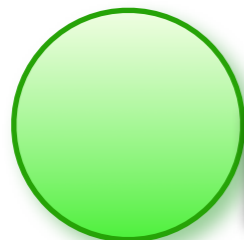
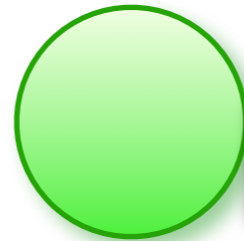
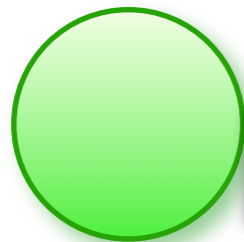
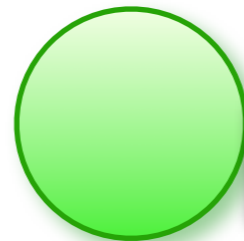
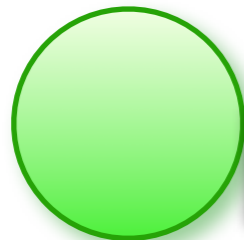
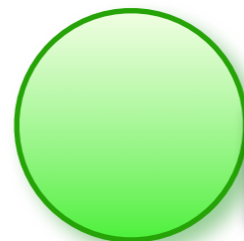
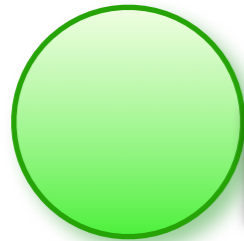
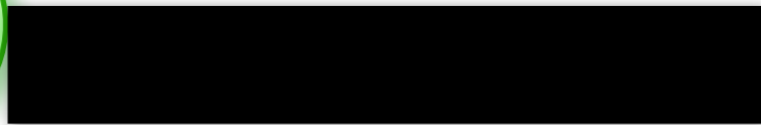
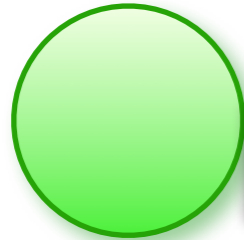


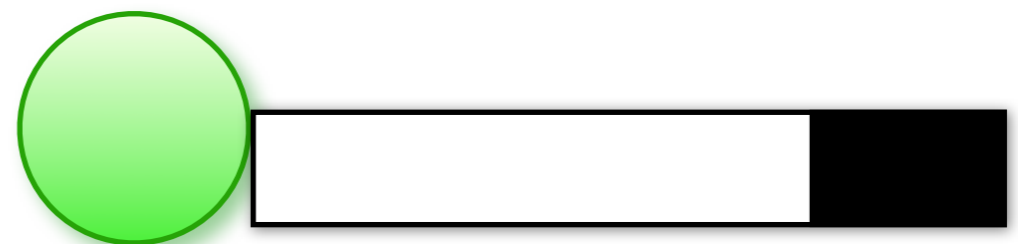
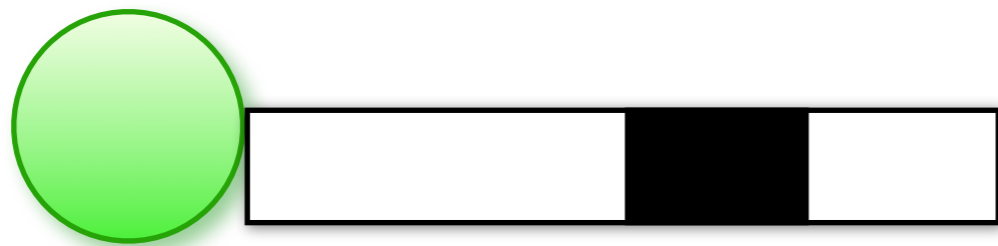
↓ディスクイメージ





↓ディスクイメージ





分散ディスク共有

プロトコル

TCP

Join

ネットワークに参加する

Get Data

データをダウンロードする

UDP

Notify Up

持っているデータ範囲を知らせる

Notify Down

今からダウンすることを伝える

Find

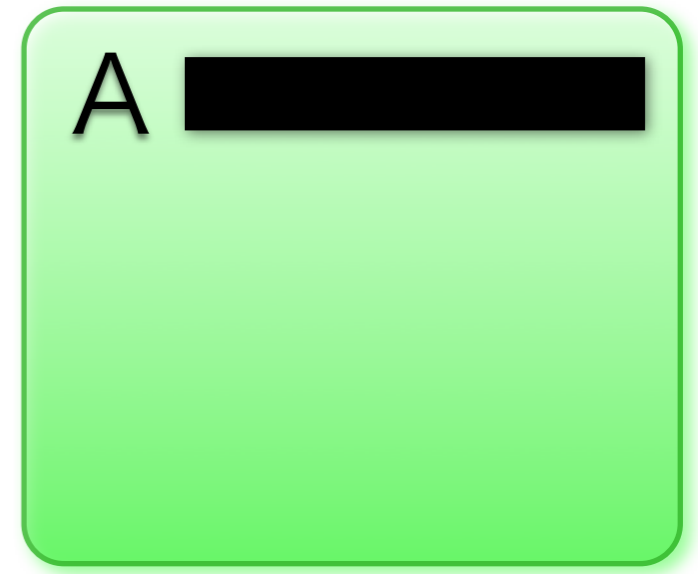
データを持っていないか聞く

分散ディスク共有

ノードの起動



データ



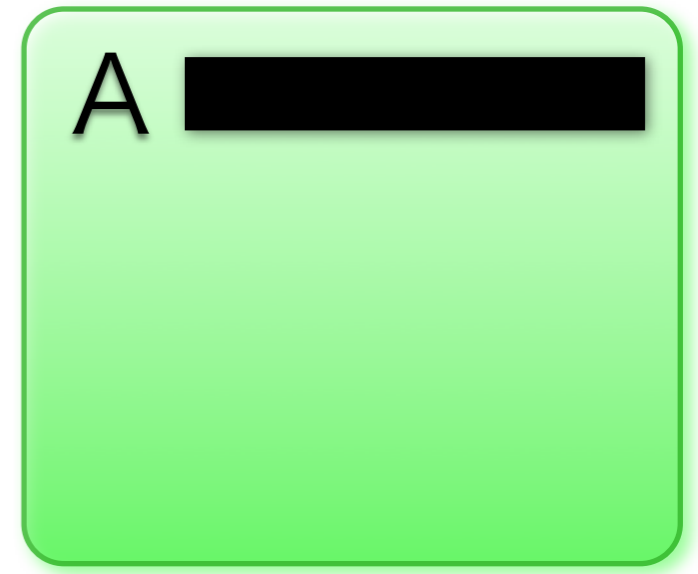
V-Table

分散ディスク共有

ノードの起動



データ



V-Table

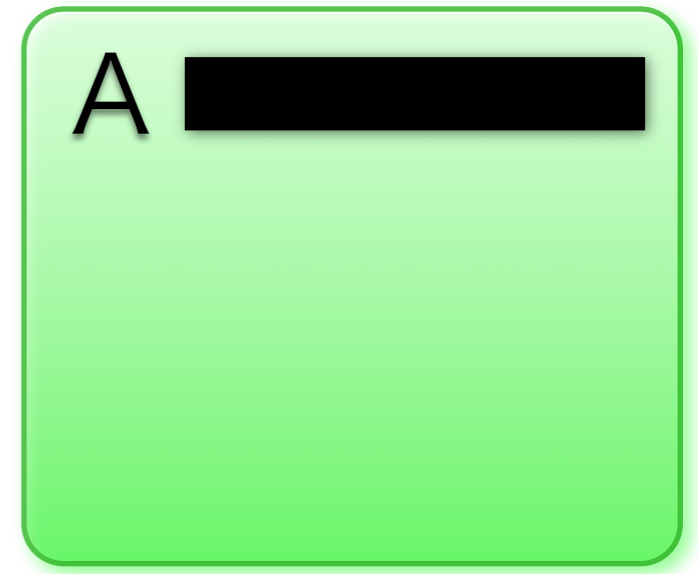


分散ディスク共有

ノードの起動

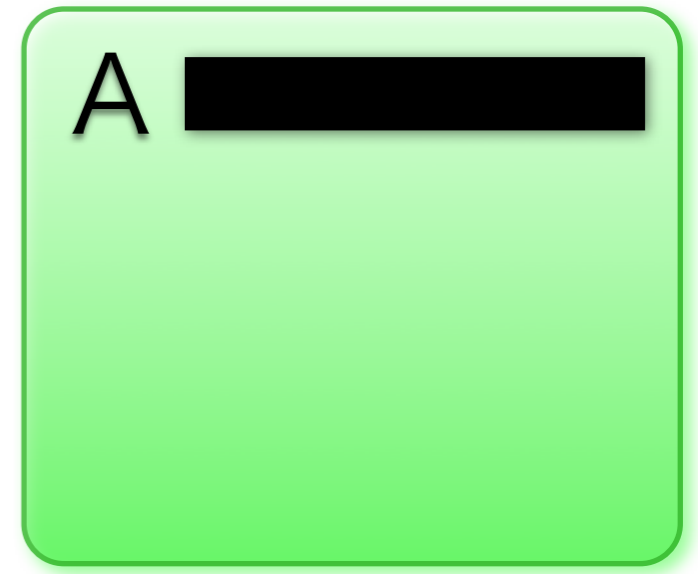
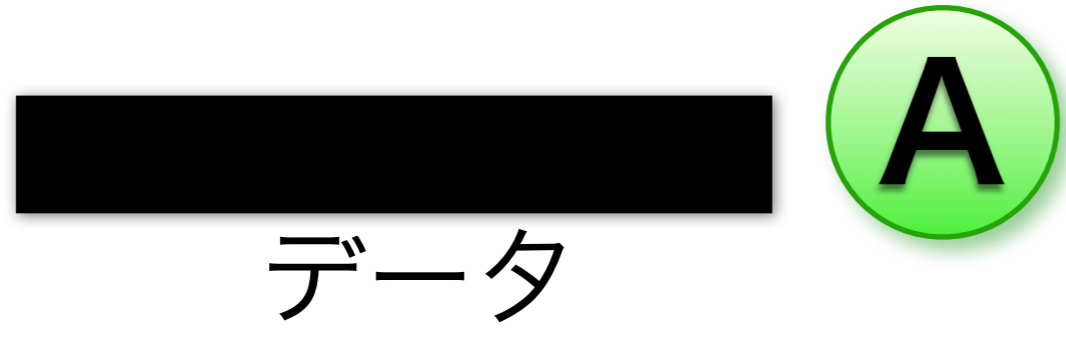


Join



分散ディスク共有

ノードの起動

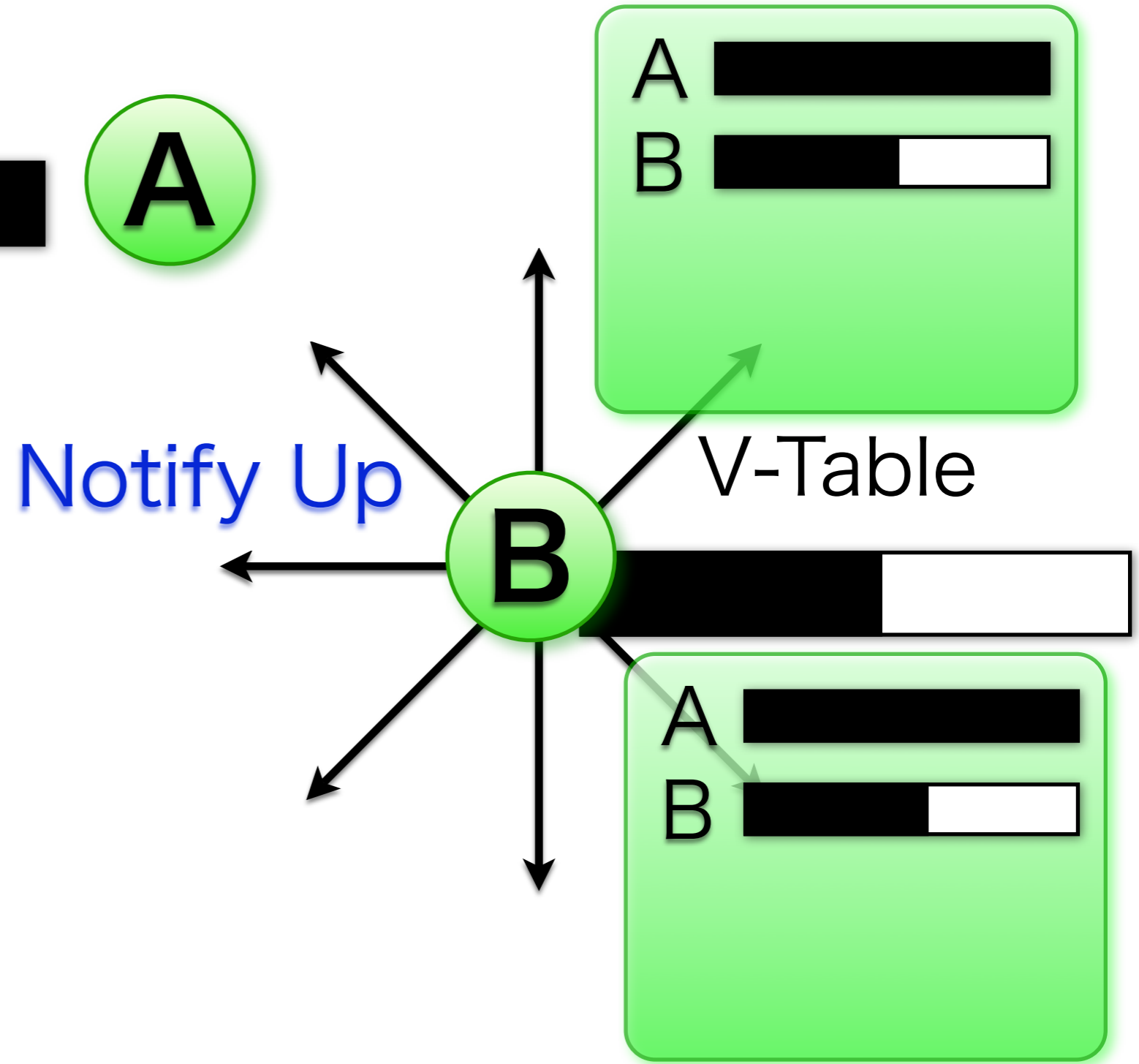
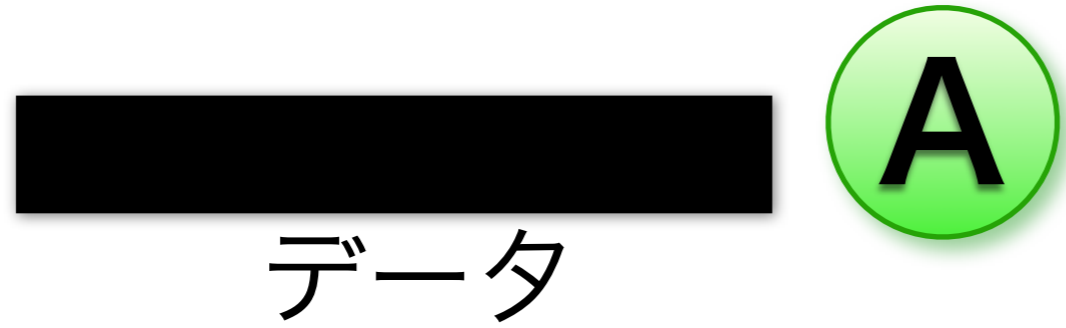


V-Table



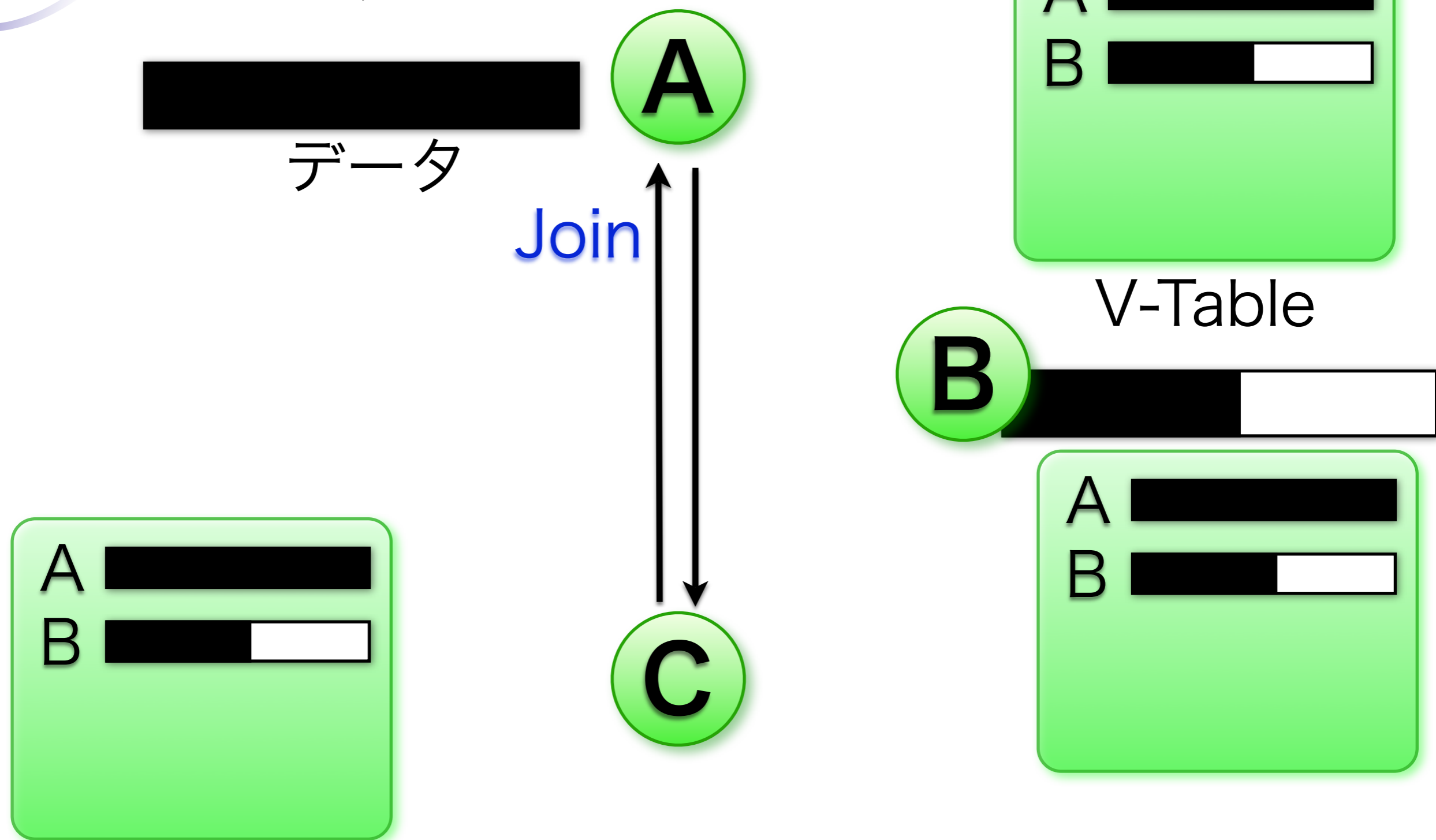
分散ディスク共有

ノードの起動



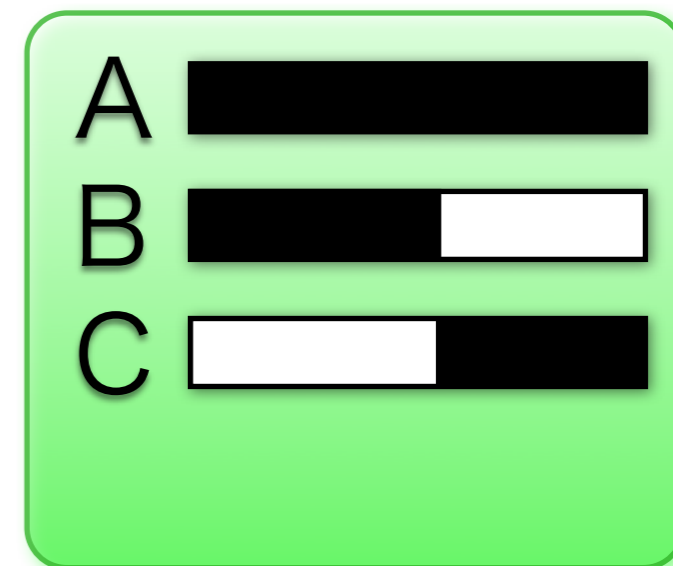
分散ディスク共有

ノードの起動

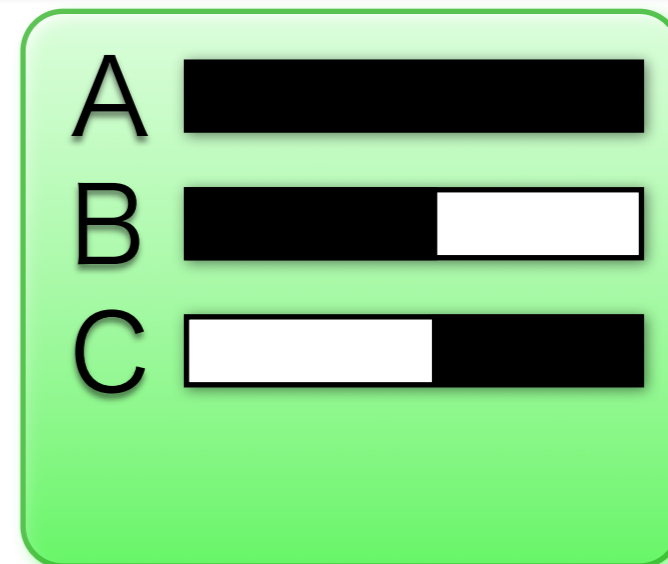


分散ディスク共有

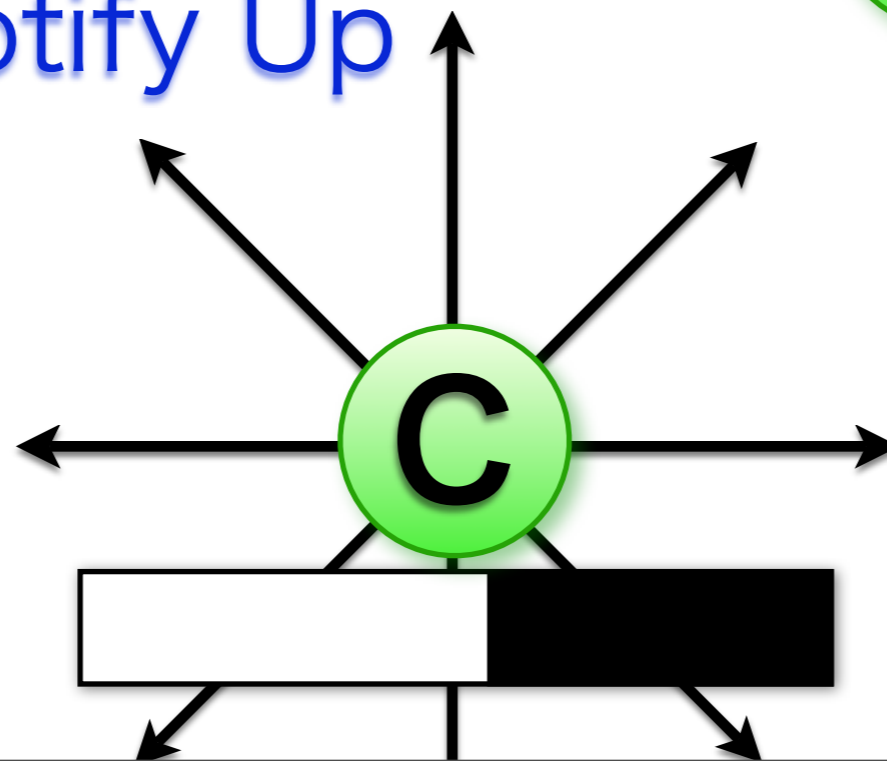
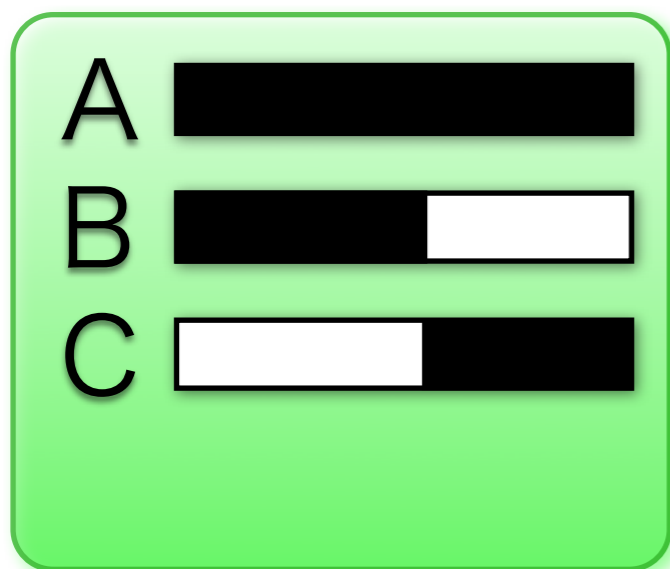
ノードの起動



V-Table

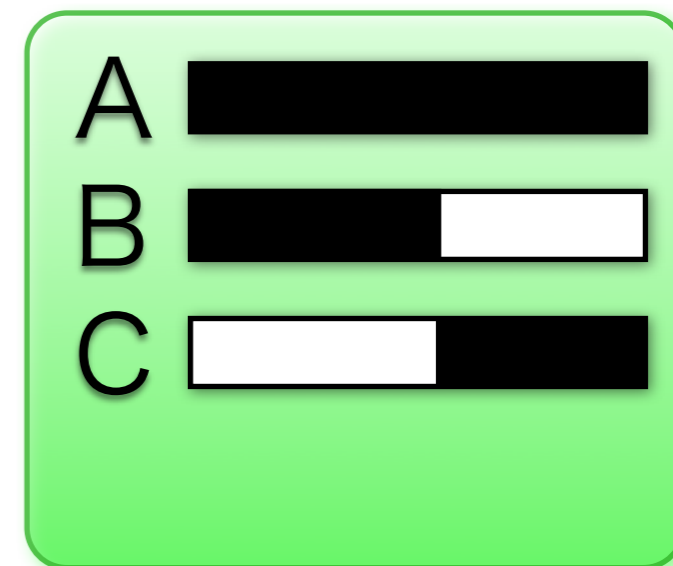


Notify Up

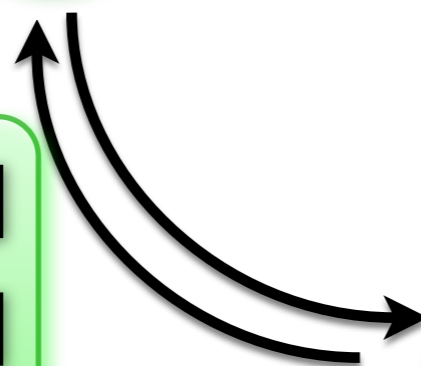
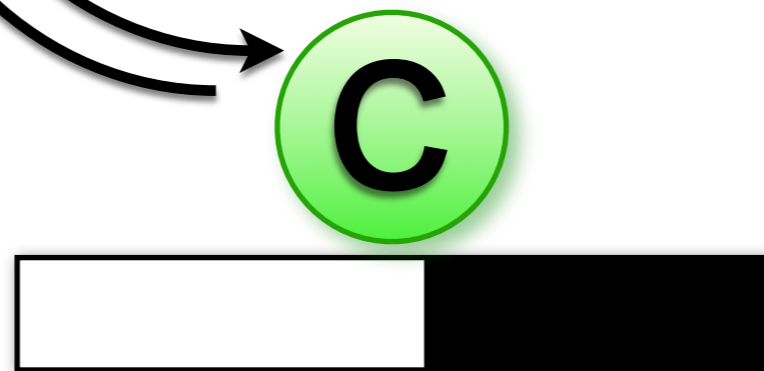
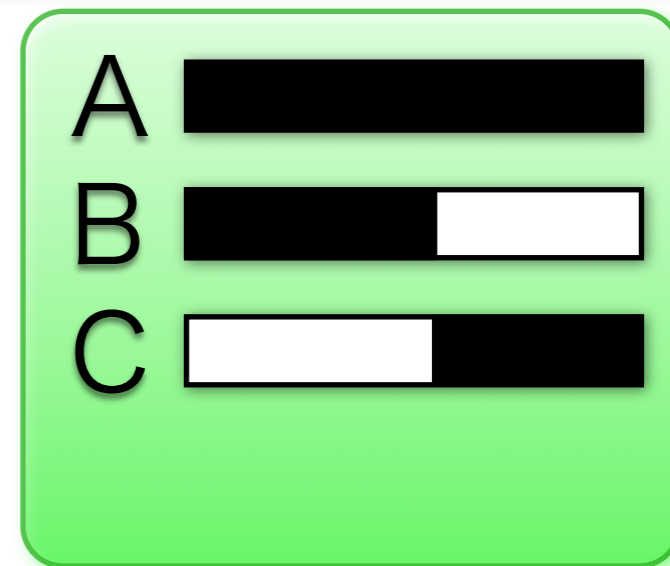
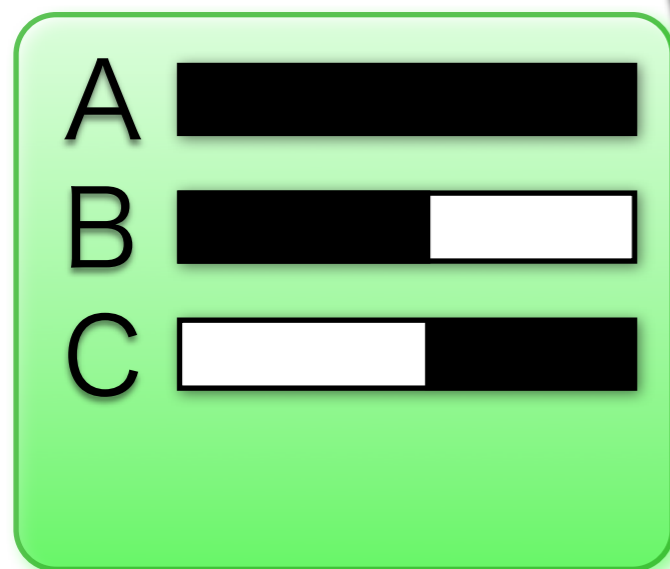
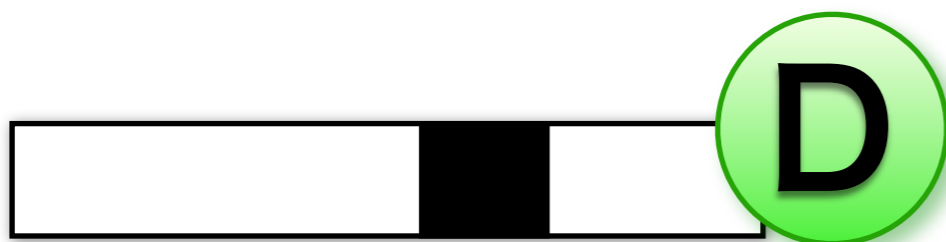


分散ディスク共有

ノードの起動

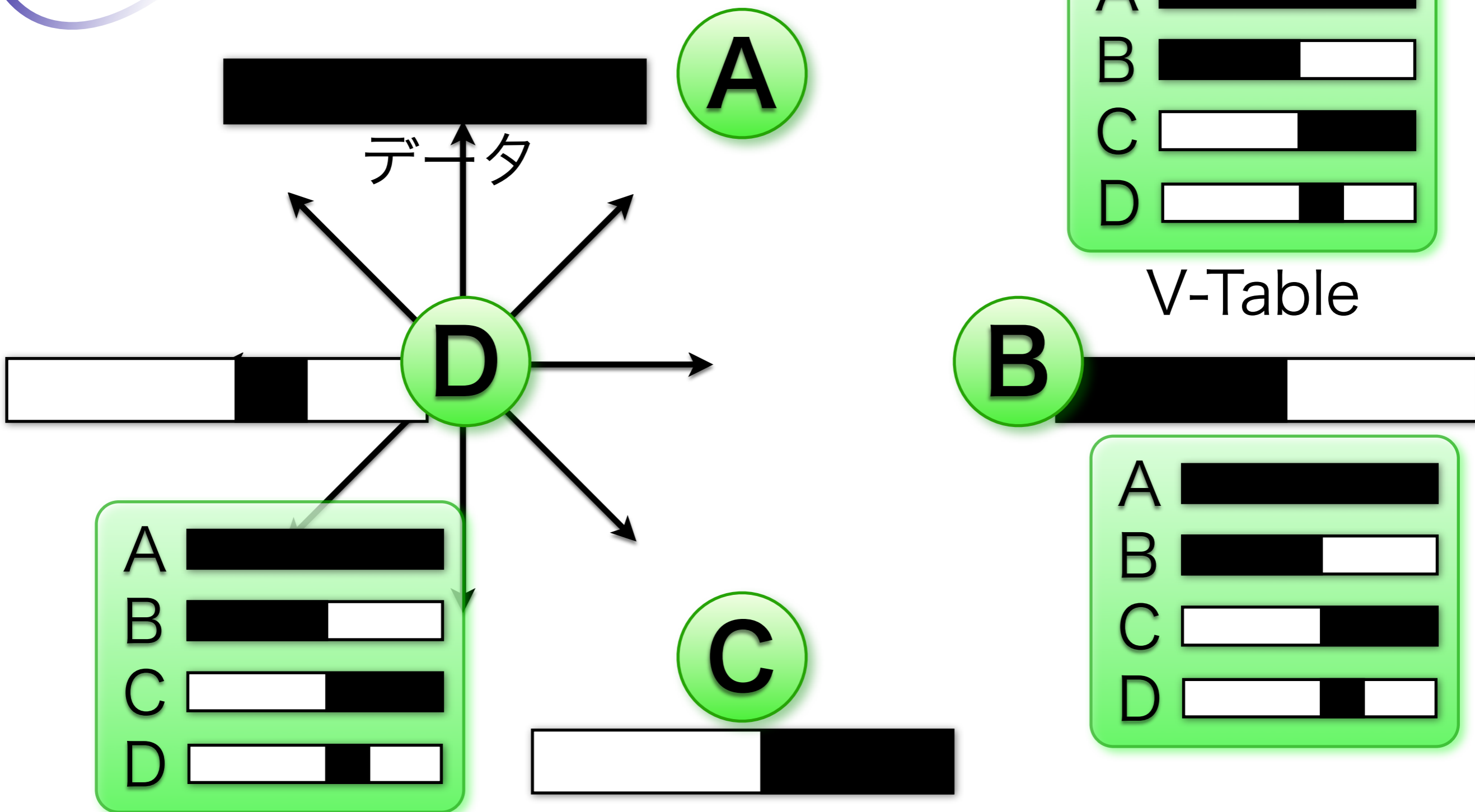


V-Table



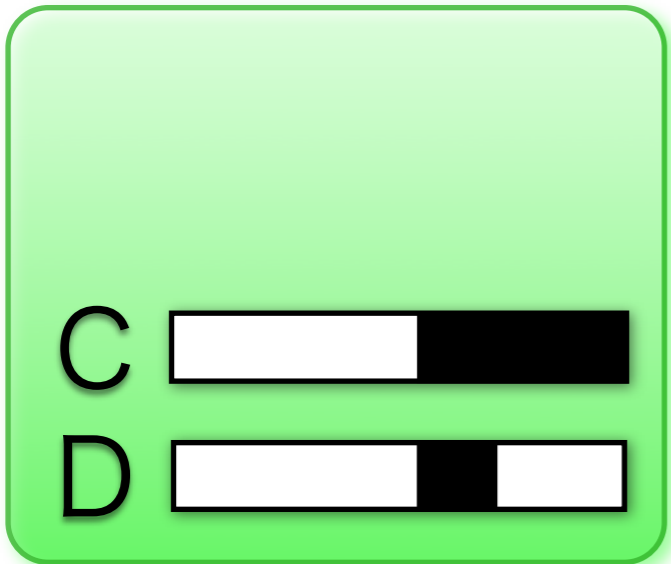
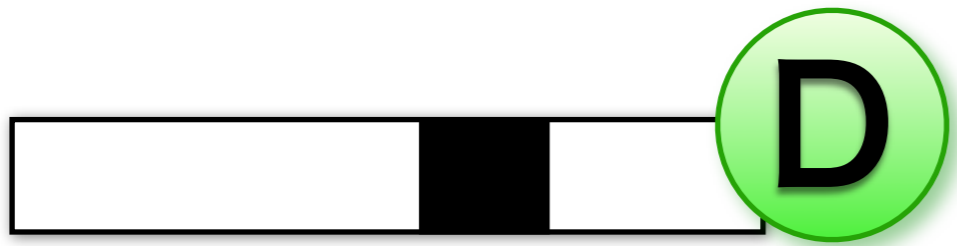
分散ディスク共有

ノードの起動



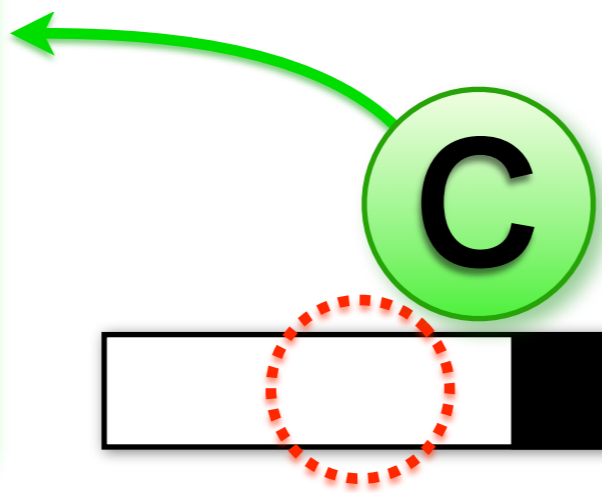
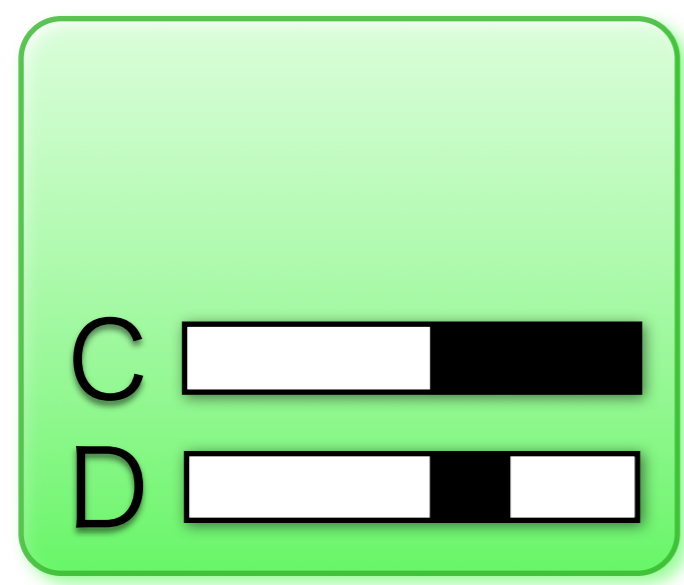
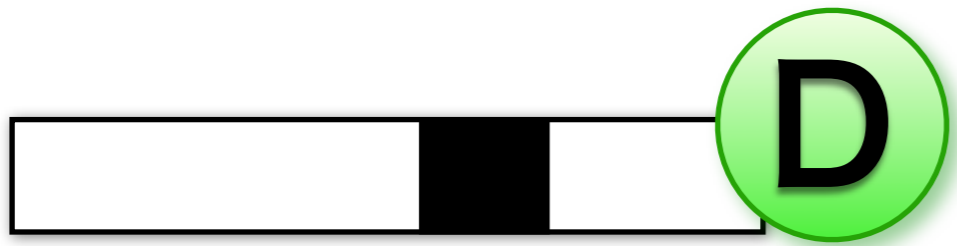
分散ディスク共有

V-Tableの補完



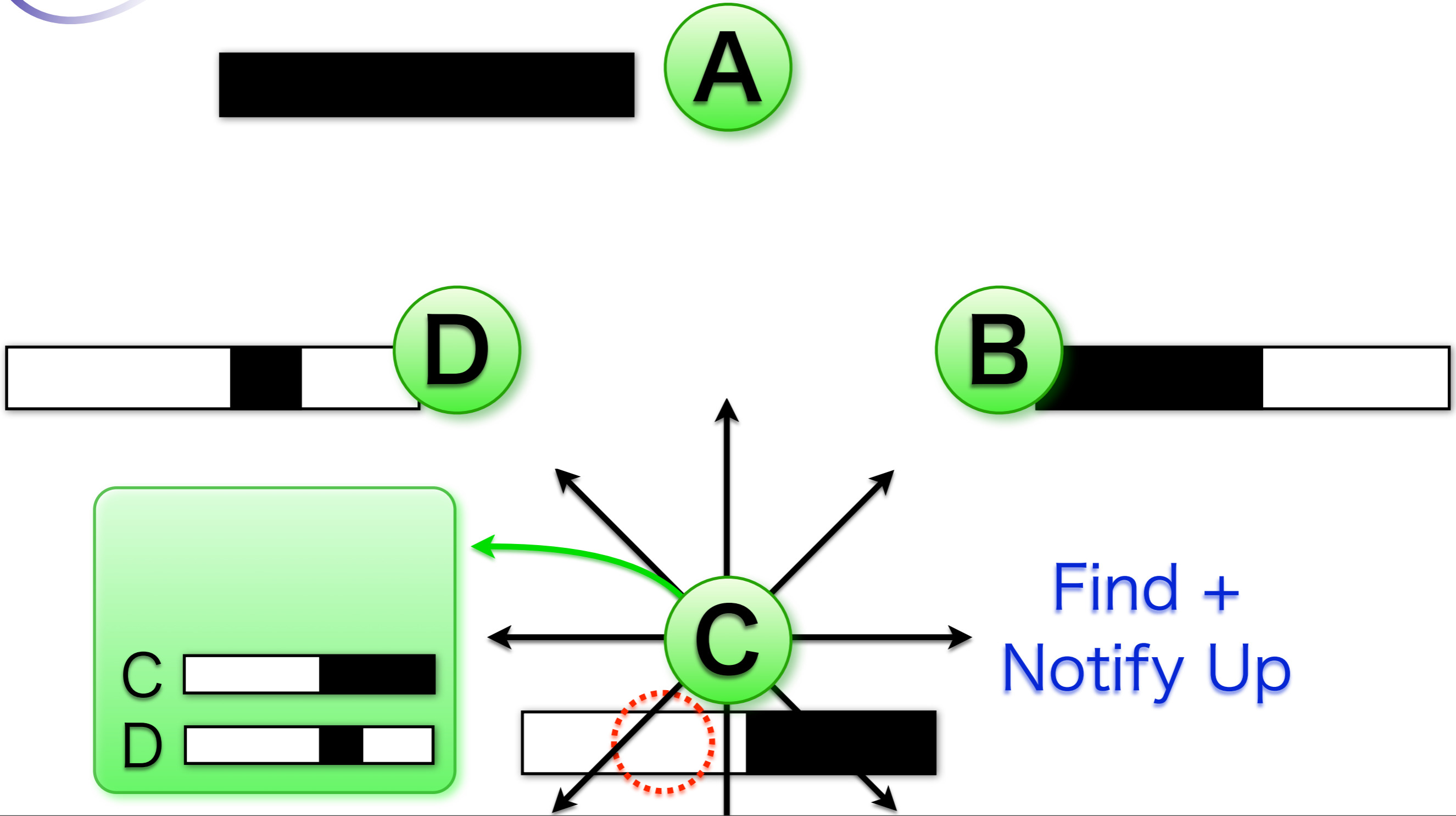
分散ディスク共有

V-Tableの補完



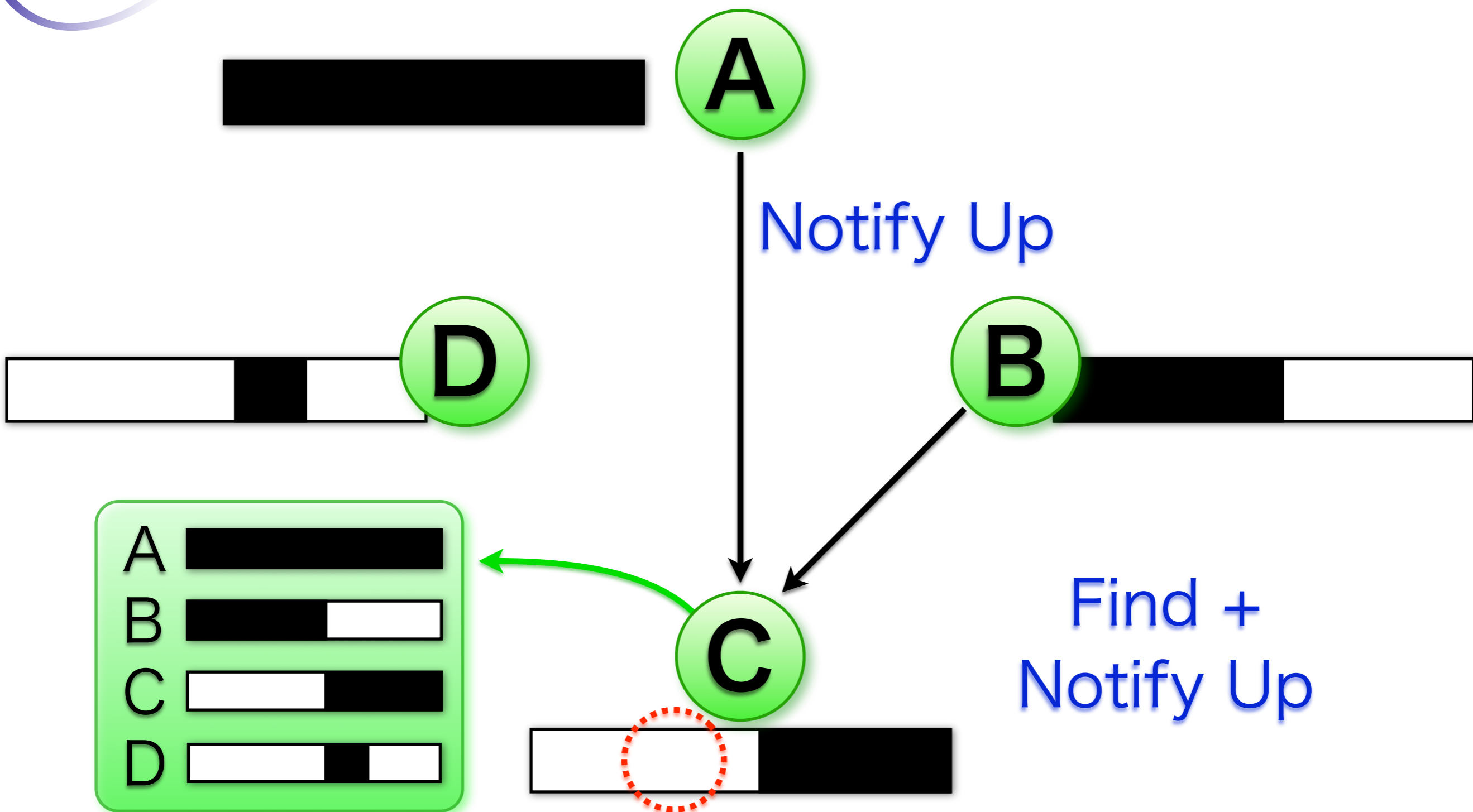
分散ディスク共有

V-Tableの補完



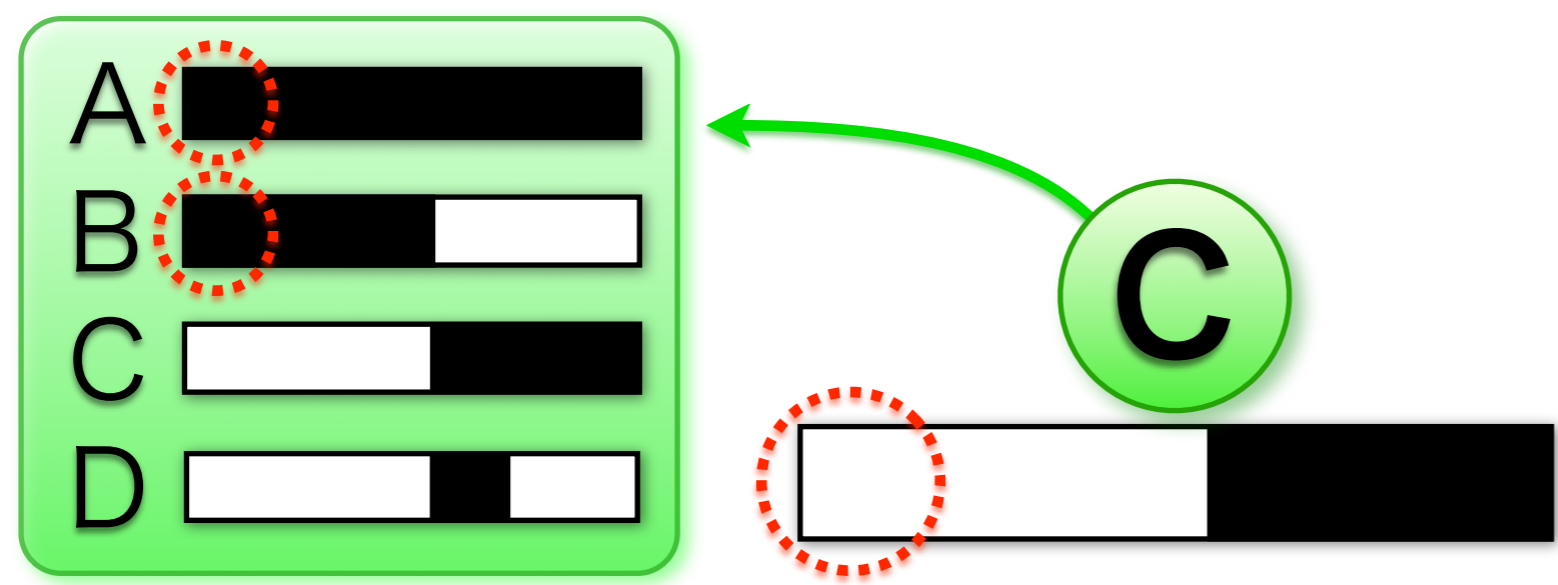
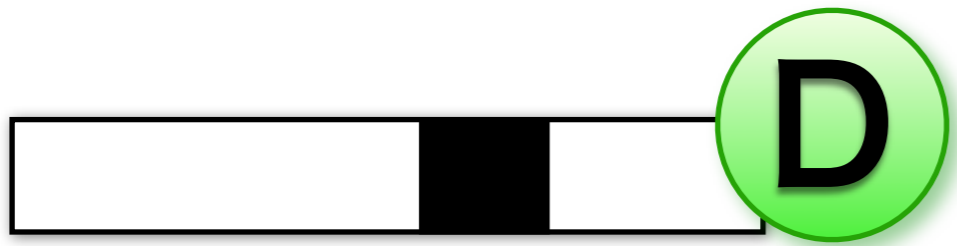
分散ディスク共有

V-Tableの補完



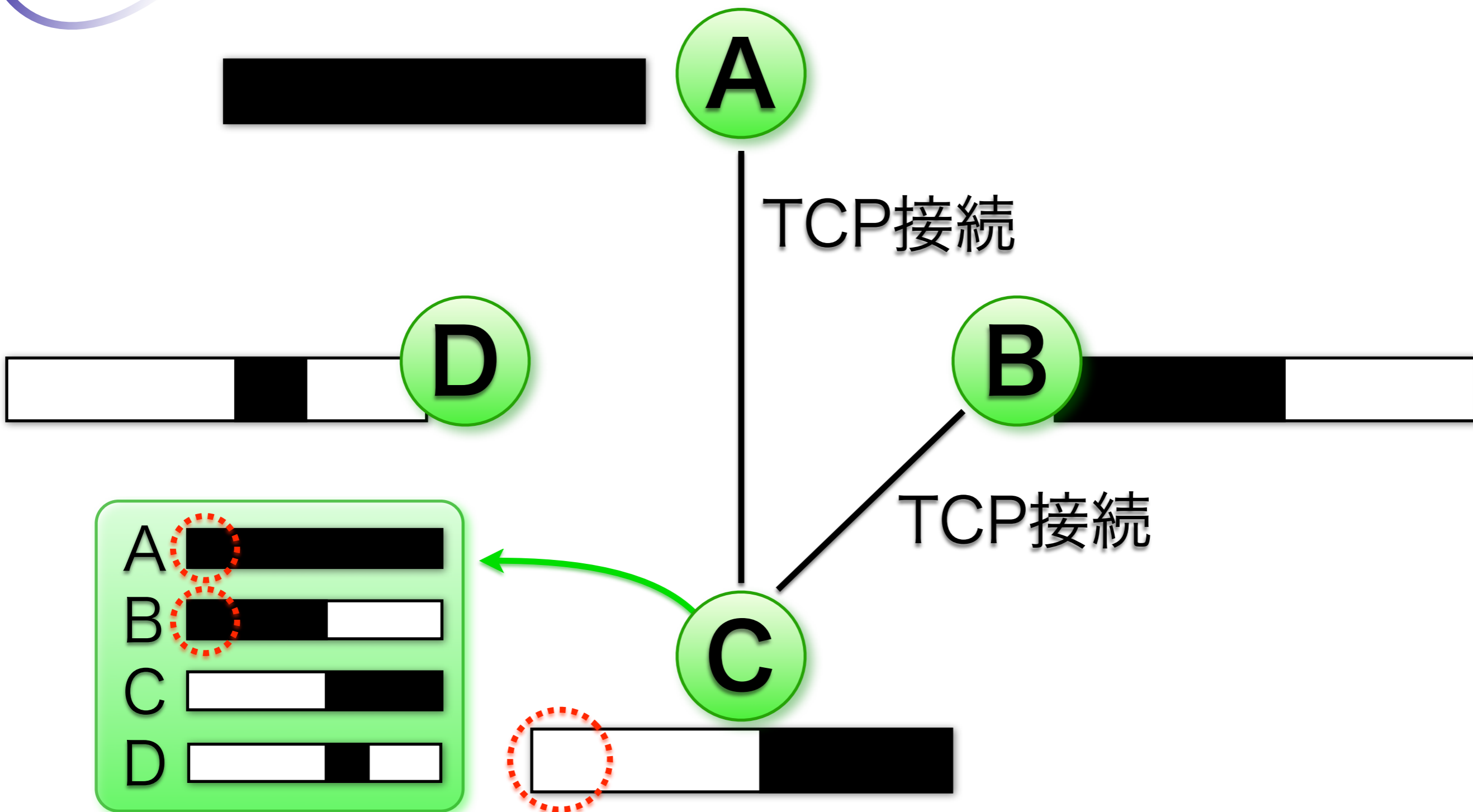
分散ディスク共有

データの取得



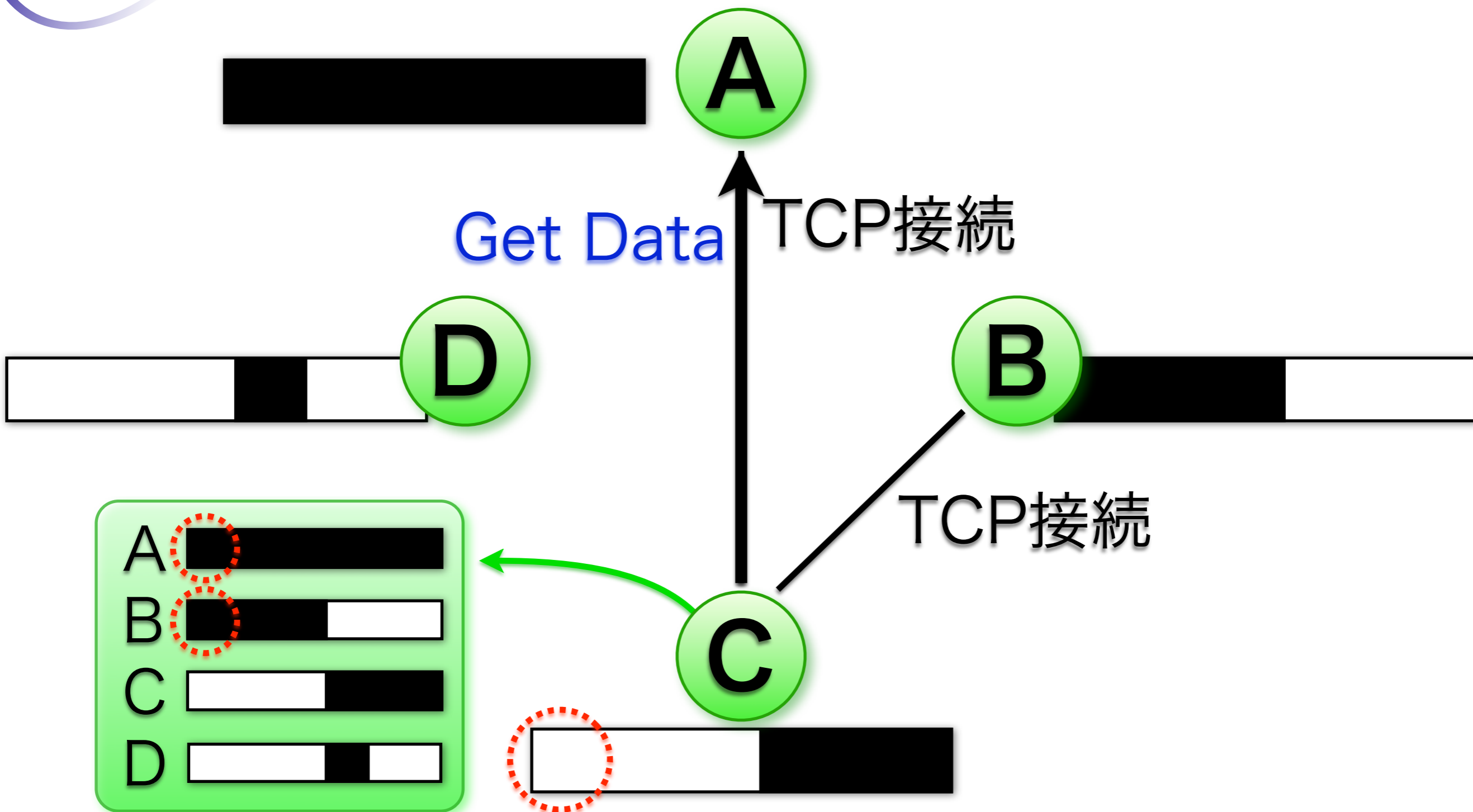
分散ディスク共有

データの取得



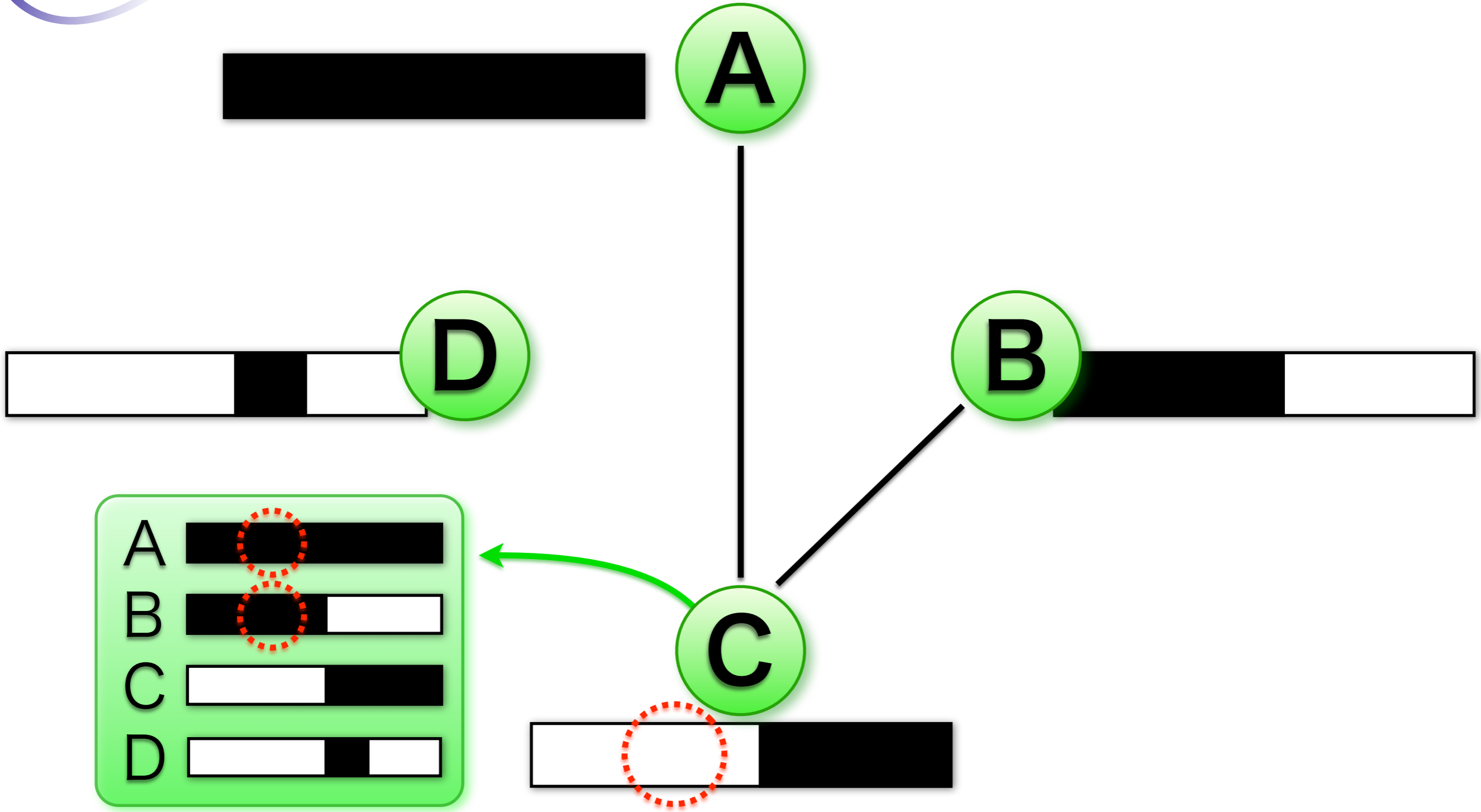
分散ディスク共有

データの取得



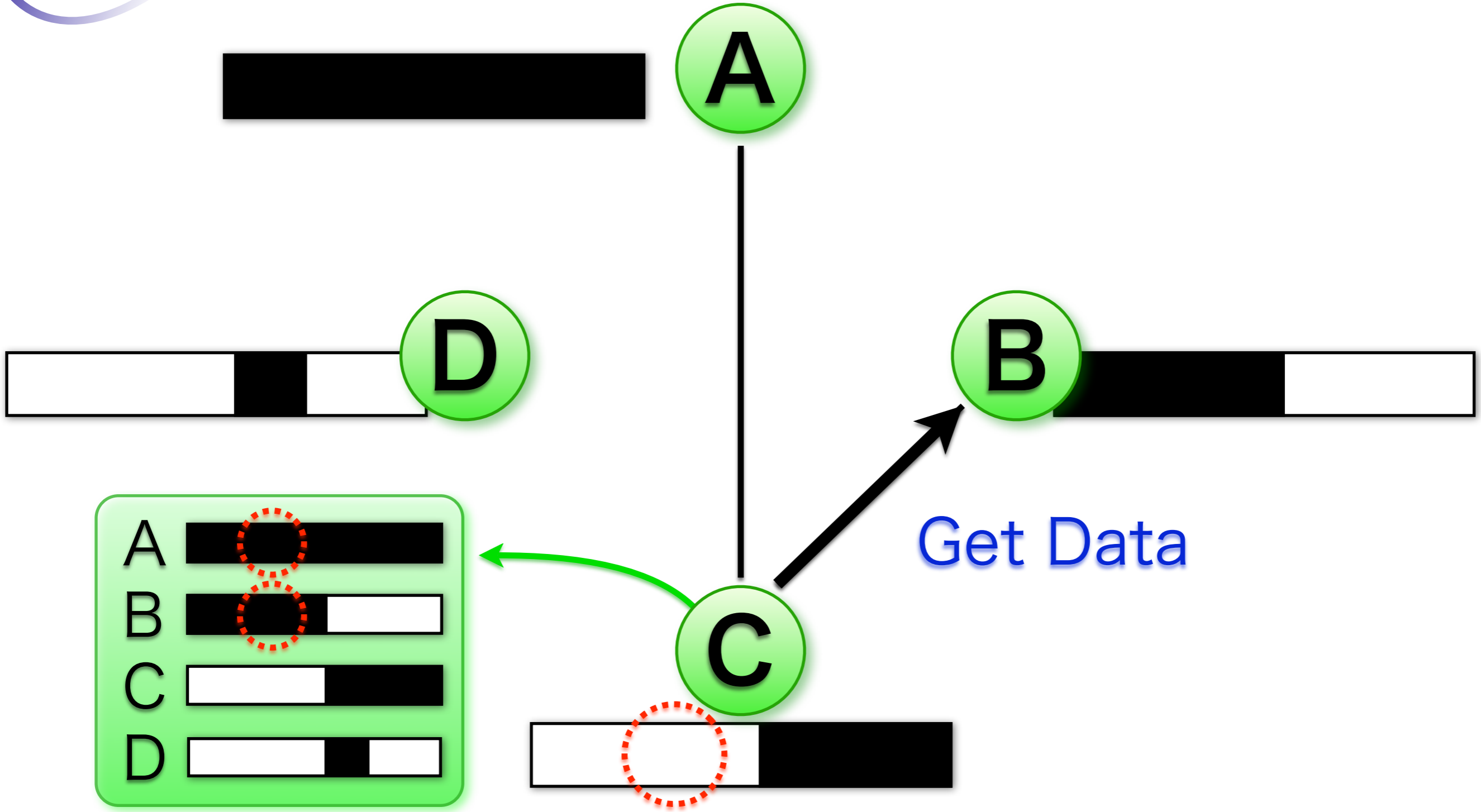
分散ディスク共有

TCP接続の維持



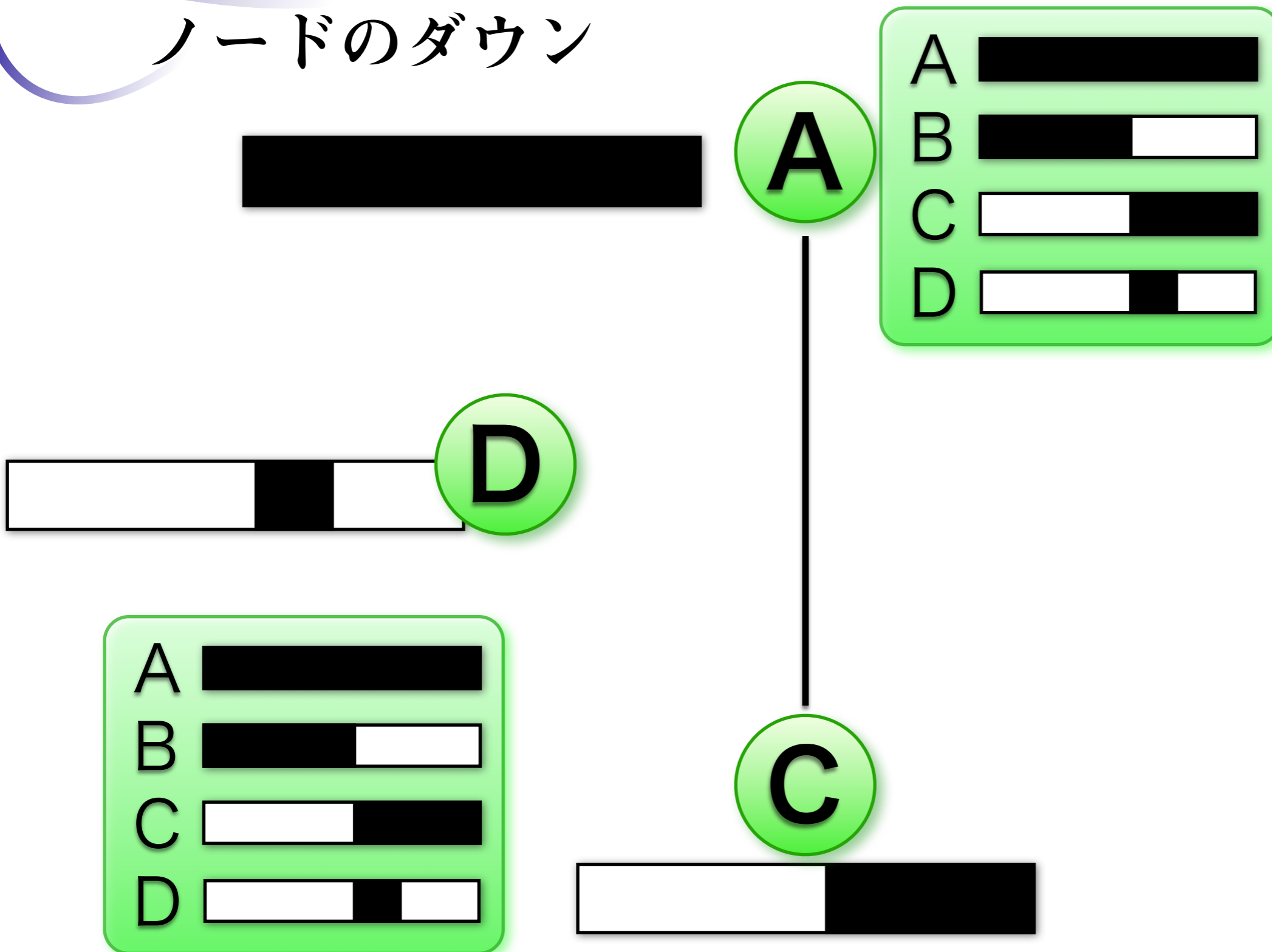
分散ディスク共有

TCP接続の維持



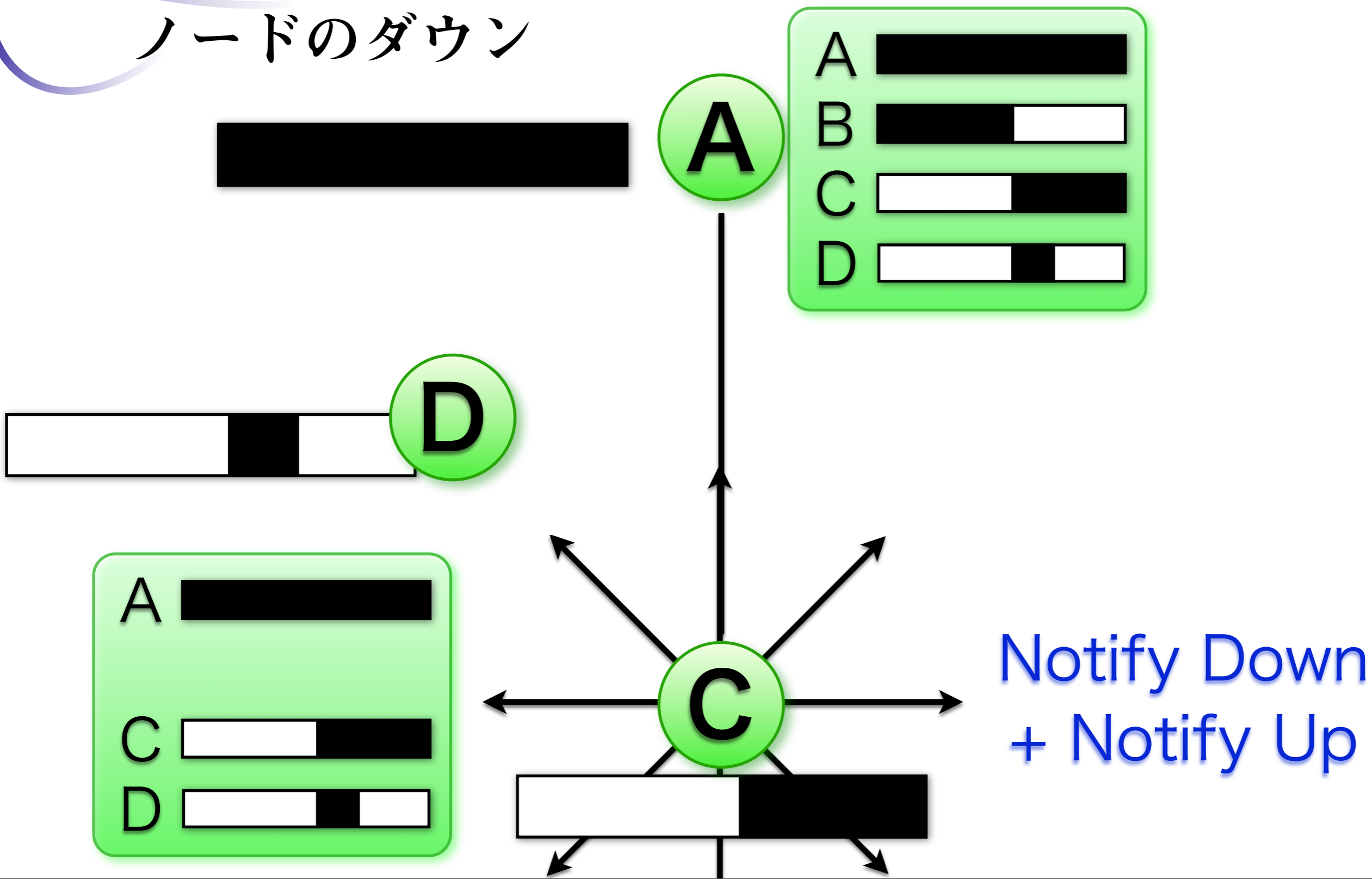
分散ディスク共有

ノードのダウン



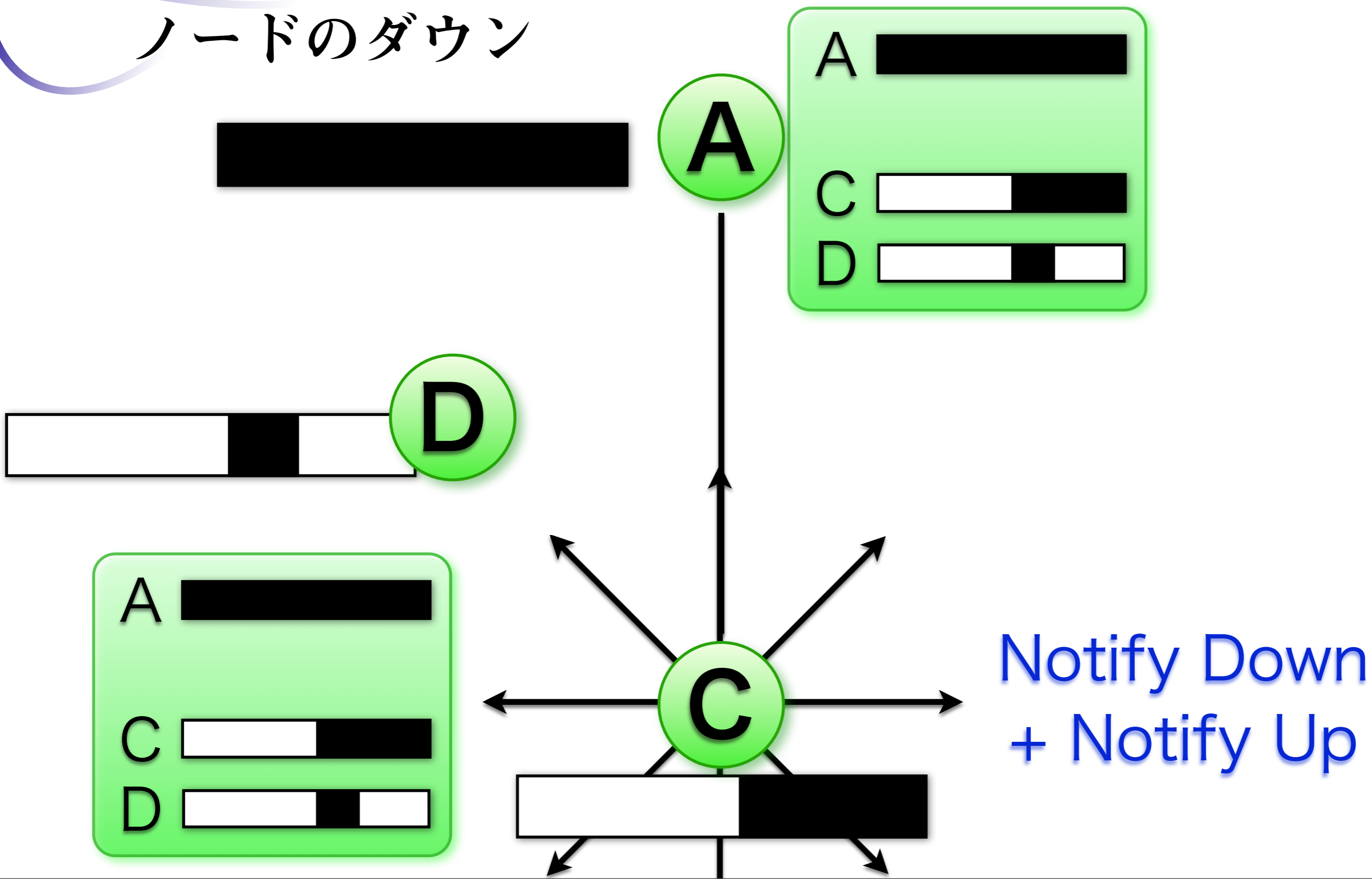
分散ディスク共有

ノードのダウン



分散ディスク共有

ノードのダウン



分散ディスク共有

- TCP接続を維持する

3-way handshakeを減らす
ノードのダウンを検出

- **Find**と**Notify Down**にも**Notify Up**を載せる

V-Tableの精度が上がる

