

# 仮想計算機イメージ格納領域としての分散ファイルシステム/ ストレージシステムの性能比較

島 慶一\*

DANG Nam†

## 1 はじめに

仮想化技術の貢献により、これまでは非現実的と考えられていた数の計算機資源をインターネットサービスに導入運用できるようになってきた。計算機の仮想化の目的は、計算機機能の重畳による物理計算機資源の有効活用である。そのため、仮想計算機がデータセンターの負荷や拡張状況に応じて、高い自由度で再配置できることが重要な課題となる。今後、異なる場所に配置されたデータセンター間での仮想計算機移動なども考慮すると、広域運用可能な仮想計算機基盤の構築が必要になる。本論文では、仮想計算機のディスクイメージを広域分散ファイルシステム/ストレージシステムに格納することを想定し、既存の複数システムでの性能を比較した結果を報告する。

## 2 対象分散システム

計測対象とした分散ファイルシステム/ストレージシステム (以後、分散ストアと呼称) は Ceph<sup>1</sup>、Sheepdog<sup>2</sup>、GlusterFS<sup>3</sup>、XtreemFS<sup>4</sup> の4つである。仮想計算機を提供する環境は KVM<sup>5</sup>、分散ストアおよび KVM を運用するホスト OS は Ubuntu 12.04.1 LTS を用いた。

Ceph と Sheepdog は KVM によるネイティブサポートが提供されており、KVM の仮想ディスクとしてそれぞれ rbd 形式、sheepdog 形式として指定できる。GlusterFS と XtreemFS はネイティブサポー

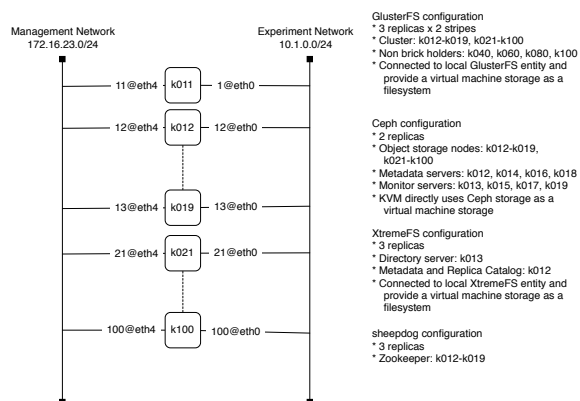


図 1: 分散ストアノードの構成

トが存在しないため、一旦ホスト OS でファイルシステムとして認識させた後、その領域に仮想計算機のディスクイメージを格納する。

Ceph、Sheepdog、XtreemFS は仮想計算機のイメージを分割し、それらを分散配置されたストアノードに格納する手法を採用している。GlusterFS の場合はファイル単位で分散ストアに配置する方式が基本となる。本実験の場合は仮想計算機のディスクイメージをファイルとして作成しているため、ディスクイメージ単位ということになる。GlusterFS では性能向上のためストライピングの設定が可能となっており、本実験でもストライピング機能を利用している。

## 3 ネットワーク構成

図 1 に分散ストアのネットワーク構成を示す。k012 から k019、k021 から k100 までの 88 台を分散ストアノード群として構成した。それぞれの分散ストアでの複製個数は 3 (Ceph のみ 2) としている。GlusterFS ではストライピング個数を 2 としている。各分散ス

\*keiichi@iijlab.net, IJ イノベーションインスティテュート

†namd@de.cs.titech.ac.jp, 東京工業大学

<sup>1</sup><http://ceph.com/>

<sup>2</sup><http://www.osrg.net/sheepdog/>

<sup>3</sup><http://www.gluster.org/>

<sup>4</sup><http://www.xtreemfs.org/>

<sup>5</sup><http://www.linux-kvm.org/>

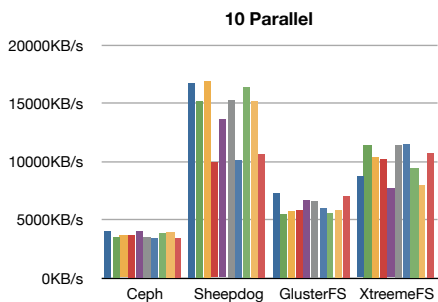


図 2: 10 台の仮想計算機による同時書き出し性能

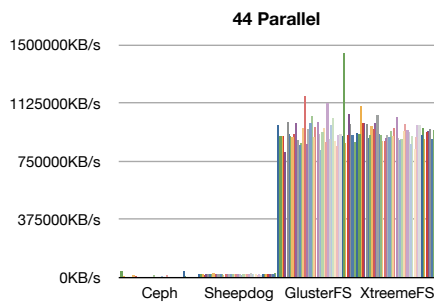


図 5: 44 台の仮想計算機による同時読み出し性能

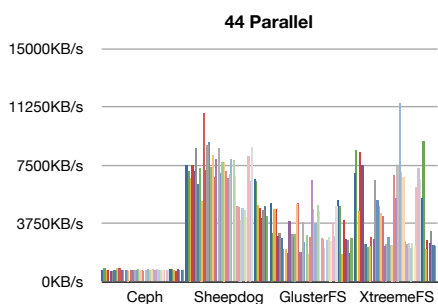


図 3: 44 台の仮想計算機による同時書き出し性能

トアノードは KVM ハイパーバイザーとしても機能し、それぞれ 1 台の仮想計算機を保持している。

## 4 ディスク性能計測

ディスク性能の計測は Bonnie++<sup>6</sup> を用いた。図 2 および図 3 は、それぞれ 10 台、44 台の仮想計算機による同時書き込み時の性能計測の結果である。図 4 および図 5 は、同条件での読み込み性能計測の結果である。

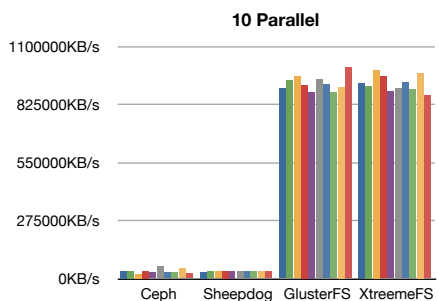


図 4: 10 台の仮想計算機による同時読み出し性能

## 5 考察

ディスクへの書き出しに関して Sheepdog が他の分散ストアよりもよい性能を出している。一方、ディスクからの読み出しは GlusterFS と XtremFS の性能が高い。今回、GlusterFS と XtremFS はファイルシステムとしてハイパーバイザーにマウントした状態で利用しているので、ファイルシステムレベルでの読み込みキャッシュが貢献しているのではないかと考えられる。Ceph は全体として性能の揺らぎの少ない安定した動作を実現しているが、相対的な性能は他の分散ストアよりも低いことが確認された。

## 謝辞

情報通信研究機構 北陸 StarBED 技術センターの宮地博士および中井氏には実験実施に際して多くの助言と支援をいただきました。ここにお礼を申し上げます。

<sup>6</sup><http://www.coker.com.au/bonnie++/>