クラスタ構築システムRocksを用いた仮想クラスタの構築

中田 秀基[1] 横井 威[1] 江原 忠士[1,2] 谷村 勇輔[1] 小川 宏高[1] 関口 智嗣[1]

1. 産業技術総合研究所, 2. 数理技研

1 概要

データセンタにおける計算機運用の運用率を向上させる方法として，資源の一部を適当なサービスレベルアグリーメントの元に，予約ベースで貸し出すことによる方法が考えられる．この方法は，実際の計算機，ネットワーク，ストレージを用いて実現することもできるが，配線の変更，OSのインストール，アプリケーションのデプロイなど，膨大な作業が必要となる．これを低コストで実現する方法として，仮想化を用いる方法が考えられる．われわれは，クラスタの3つの側面，すなわち，計算機，ネットワーク，ストレージをそれぞれ仮想化することで，仮想クラスタを実現した [1]．計算機の仮想化にはVMWareを用い，ネットワークの仮想化にはVLANを，ストレージの仮想化にはiSCSIを用いた．また，仮想クラスタの構築にはクラスタデプロイシステムであるNPACI Rocks[2] を用いる．Rocksを用いることによって，単なる仮想ノードの集合ではなく,「仮想クラスタ」を構築することが提案システムの特徴である．さらに，提案システムそのものもRocksによって簡便に配備することを可能とした．

2 提案システムの概要

本システムには，クラスタプロバイダ，サービスプロバイダ，ユーザの三者が関与する．クラスタプロバイダは本システムを使用して所有するクラスタを管理し，サービスプロバイダに提供する主体である．サービスプロバイダはクラスタを利用して，ユーザにサービスを提供する．

仮想クラスタインストールの様子を図1に示す．本システムでは，クラスタマネージャ，ゲイトウェイ，ノードの三種類の計算機を仮定する．クラスタマネージャは物理クラスタ仮想クラスタすべてを管理するノードで，管理者はクラスタマネージャに対してWebインターフェイスでアクセスし，クラスタの予約を行う．ゲイトウェイは外部ネットワークと内部ネットワークの双方へのアクセスできる．

仮想クラスタを利用するサービスプロバイダは，クラスタプロバイダに対して仮想クラスタの予約を行う．この際に開始･終了時刻，ノード数，メモリ，ディスクサイズ，インストールするべきソフトウェア，共有の外部ストレージ容量を指定する．システムは，開始時刻が来るとゲイトウェイ上に仮想フロントエンドをインストールする (図1中段)．さらに仮想ノードを起動し，仮想フロントエンドからインストールを行う (図1下段)．この際，仮想フロントエンドと仮想ノードの間のネットワークとしては，内部ネットワーク上に設けたVLANを用いる．さらに，privateネットワーク上に存在するiSCSIターゲットマシン上の論理ボリュームを払い出し，アクセス可能に設定する．仮想フロントエンドがinitiatorとなりこの論理ボリュームをマウントし，NFSで仮想クラスタ全体に公開する．

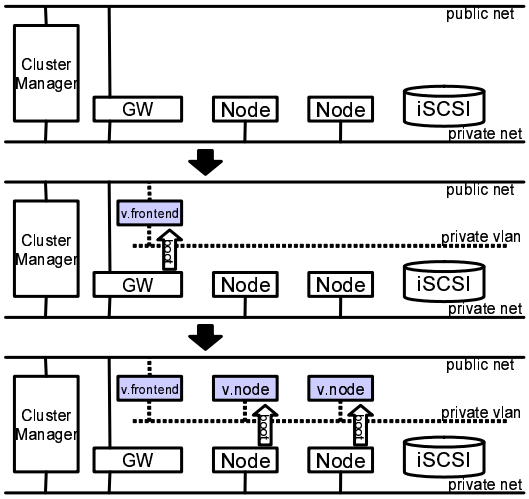

図 1: 仮想クラスタのインストール

3 今後の課題

- ストレージ管理の高度化：現在仮想ノードのディスクイメージは各実ノード上のディスク上に配備されている．これをiSCSIターゲットに集中させることで，より柔軟な運用が可能になる．また，共有ディスクに関しても，GFSやPVFSなどのクラスタファイルシステムをiSCSIと組み合わせることで，仮想フロントエンドがボトルネックになることを避ける．
- 複数クラスタの統合的運用：単一のクラスタでは，提供できる資源には限界がある．複数のクラスタ上に仮想的な単一クラスタを形成することを検討する．

参考文献

[1] 中田秀基, 横井威, 関口智嗣: Rocksを用いた仮想クラスタ構築システム, 情報処理学会 HPC研究会 2006-HPC-106 (2006).

[2] : Rocks. http://rocks.npaci.edu/.