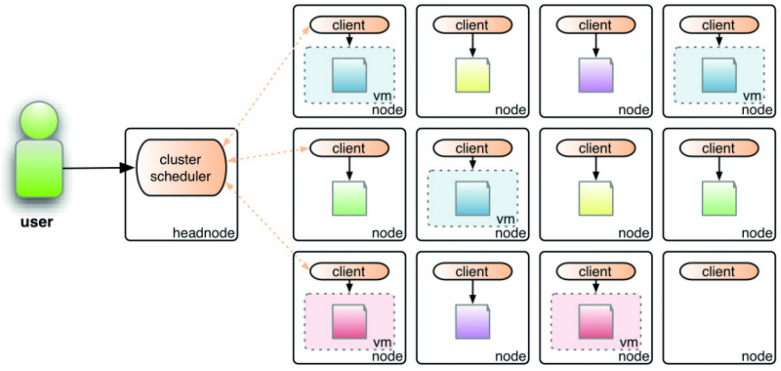
Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Management virtueller Maschinen in Linux High Performance Clustern



Researchers
Dr. Roland Schwarzkopf, Dr. Matthias Schmidt and Dr. Thomas Gebhardt

Principal Investigator
Prof. Dr. Bernd Freisleben

Project Term
2011 - 2012

Project Areas
Computer Science, Mathematics

Clusters
MaRC2 Cluster Marburg

Institute
Hochschulrechenzentrum Marburg

University
University of Marburg

Introduction

Traditionelle Linux High Performance Cluster (HPC) wie der Marburger Rechen-Cluster (MaRC2) basieren auf einem Verbund von physischen Rechenknoten, die zentral mit Hilfe eines Batch-Systems (in MaRC2: Sun Grid Engine) administriert werden. Neue Programme stellen den Administrator vor die Herausforderung, eine geeignete Umgebung bereit zu stellen. Diese haben oft neue Sprachen und eine Vielzahl von Bibliotheken, so dass eine unkomplizierte Installation einer Laufzeitumgebung unter Umständen erschwert oder gar nicht zur leisten ist. Unterschiedliche Laufzeitumgebungen beschränken zudem die Portierung von Software auf andere HPC-Cluster. Ziel des Projekts ist der Einsatz von Virtualisierungstechnologie im HPC-Umfeld, d.h. die Entwicklung einer Infrastruktur, die einen Mischbetrieb zwischen virtualisierten und nicht virtualisierten Jobs ermöglicht. Virtualisierter Job bedeutet, dass der Job innerhalb einer virtuellen Maschine (VM) ausgeführt wird. Diese bietet den Vorteil, dass der Anwender selbstständig eine Laufzeitumgebung für seine Programme einrichten kann, was den Administrator entlastet. Eine Alternative sind vorgefertigte VMs. Damit werden Auswirkungen auf das Gesamtsystem durch vom Anwender installierte Software vermieden. Zudem ist eine feingranulare Beschränkung der Ressourcen für jede VM möglich. Schließlich lösen VMs das Problem der Portierung von Software.

Methods

Zwei Konzepte zum Management von VMs wurden entwickelt: eine direkte Kopplung an den Cluster-Scheduler und die Bereitstellung eines „virtuellen Clusters“. Bei einer direkten Kopplung werden für jeden Job dediziert VMs erzeugt und zur Ausführung verwendet. Das zweite Konzept partitioniert das

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

System in einen virtuellen und einen nicht virtuellen Teil und verwendet virtuelle Rechenknoten, die sich ganz regulär als Rechenknoten am Cluster-Scheduler anmelden. In beiden Fällen ist die Virtualisierung für den Scheduler transparent, es sind jedoch beim zweiten Konzept Veränderungen der Partitionierung sichtbar. Für das erste Konzept spricht die höhere Flexibilität ohne vorgegebene Partitionierung und die Möglichkeit des Einsatzes Job-spezifischer VMs. Die dynamische Anpassung der Konfiguration einer VM bezüglich des Netzwerks und des verwendeten Scheduler-Clients basiert auf einem System zur Komposition von VMs aus verschiedenen Schichten. Eine Folge benutzerspezifischer VMs ist, dass eine große Zahl von VMs vorgehalten werden muss. Klassische Disk Images sind als Speicherformat für VMs nicht geeignet. Zur Lösung dieses Problems wurde der Marvin Image Store entwickelt, der neben der effizienten Speicherung von VMs mittels Deduplizierung die Historie der einzelnen VMs verwaltet.

Results

Zwei Lösungen zur Sicherheit von VMs wurden entwickelt: Der Update Checker prüft, ob für eine VM Sicherheits-Updates verfügbar sind, sogar wenn sie zum Zeitpunkt der Überprüfung nicht ausgeführt wird. Die Online Penetration Suite[4] prüft die Sicherheit der Konfiguration mittels Vulnerability Scanner. Dessen Ergebnisse können verwendet werden, um den Start einer veralteten bzw. unsicheren VM zu verhindern. Abschließend wurde eine Evaluation verschiedener Hypervisors (Xen, KVM, Vmware vSphere/Workstation, Oracle VM Server/VM Virtual Box) zum geplanten Einsatzzweck durchgeführt: KVM ist die passendste Lösung. Durchgeführte Messungen zeigten die grundsätzliche Eignung von VMs im HPC-Umfeld, wobei der durch die Virtualisierung erzeugte Overhead in bestimmten Konfigurationen noch optimiert werden kann.

Last Update: 2020-09-28 18:20