

Verteilte Berechnung biomedizinischer Anwendungen mit Docker Swarm und OpenStack

Christoph Jansen¹, Michael Witt¹, Dagmar Krefting¹
¹Hochschule für Technik und Wirtschaft Berlin

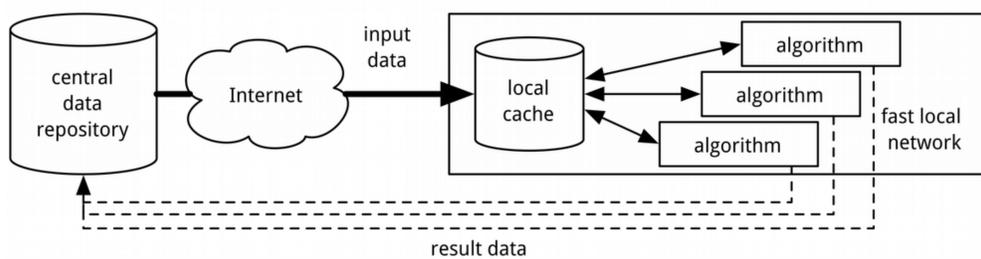
Biomedizinische Anwendungen

Algorithmen und Anwendungen der Biomedizin stellen besondere Anforderungen an eine sichere und zuverlässige Berechnung in einer verteilten Infrastruktur:

- Rechenintensive Anwendungen
→ Benötigt verteilte Berechnung in einem Compute-Cluster
- Ausführung der Anwendungen benötigt Abhängigkeiten
→ Betriebssystem mit Bibliotheken (z.B. Debian, Ubuntu, Fedora)
→ Laufzeitumgebungen (z.B. Matlab, Python, Java, Mono)
→ Skripte und Binaries
→ Konfiguration
- Große Ein- und Ausgabedaten (z.B. Bilder / 3D-Scans, Biosignale)
- Verschlüsselter Zugriff auf externe Datenquellen (XNAT, PACS)
→ Nur für autorisierte Nutzer
- Zugriff auf Daten nur für die Zeit der Berechnung
- Protokollierung von Datenzugriffen, Berechnungen und Fehlern

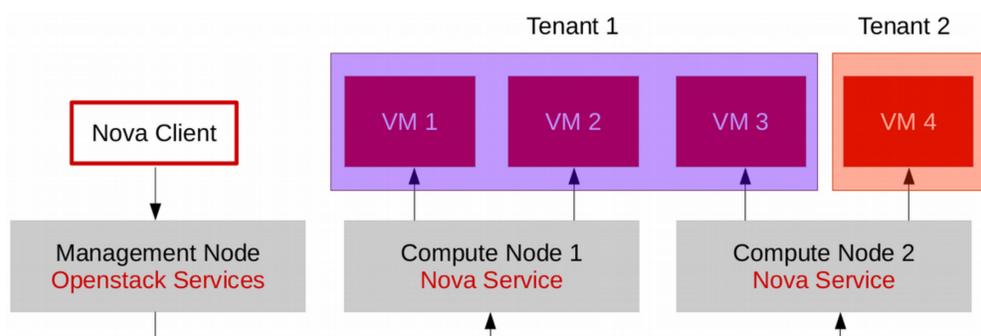
Konzept der Infrastruktur

Die Daten werden aus der Datenquelle in das Cluster geladen und stehen im Cache zur Verfügung, solange sie von autorisierten Algorithmen benötigt werden.



Komponente: OpenStack

Mit OpenStack können virtuelle Rechencluster verwaltet werden, die eine Abstraktion der verwendeten Hardware darstellen. Über Nova können nach Bedarf virtuelle Maschinen gestartet werden, die Ressourcen wie CPU, RAM, Festplattenspeicher und Netzwerke verfügbar machen.



Vorteile:

- Das Tenant-System schirmt die Ressourcen einzelner Nutzer und Projekte voneinander ab und bietet Sicherheit.
- Virtuelle Maschinen können kurzlebig sein und nach der Ausführung einer Anwendung gelöscht werden.

Nachteile:

- Das Starten von virtuellen Maschinen ist träge und erhöht die Ausführungszeit

Komponente: Docker Swarm

Docker ermöglicht die Ausführung von Anwendungen in einem Linux Container und bietet eine Abstraktion des Dateisystems, sowie der CPU und RAM-Ressourcen. Mit Docker Swarm können diese Container in einem Cluster parallel ausgeführt werden.

Vorteile:

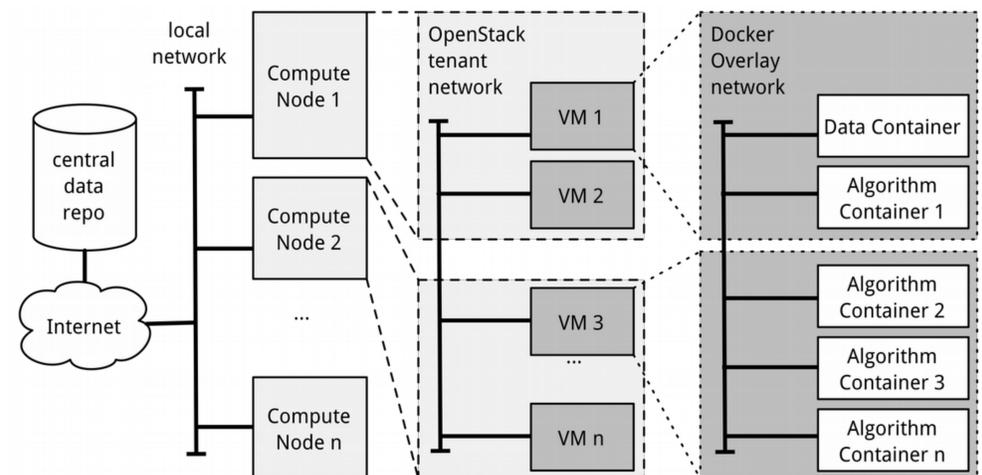
- Startet schneller als virtuelle Maschinen, da kein Betriebssystem gestartet werden muss
- Über Docker Images können Anwendungen mit allen Abhängigkeiten im Voraus verpackt werden

Nachteile:

- Kein Tenant-System zur Abschirmung der Ressourcen
- Container sind potenziell unsicherer als Virtuelle Maschinen

Docker Swarm auf OpenStack

Die Ausführung von kurzlebigen Containern wird mit Hilfe von Docker Swarm in einem statischen OpenStack-Cluster realisiert. Somit können die Vorteile beider Komponenten vereint werden.



Komponente: Curious Containers

Die **Eigenentwicklung** Curious Containers ist eine **quelloffene** Software zur Verwaltung der Anwendungen in einem Docker Swarm Cluster. Die Software besteht aus zwei Hauptkomponenten, dem CC-Server und dem CC-Container-Worker, deren Funktionsweise im Folgenden beschrieben ist.

- Die Ausführung einer Anwendung, welche zuvor in einem Docker Image verpackt wurde, wird als Task über eine Web-Schnittstelle des CC-Server angemeldet
- Sobald Ressourcen zur Verfügung stehen, werden die Container der angemeldeten Tasks ausgeführt
- Bei Bedarf werden Daten-Container gestartet, die als Cache für Eingabedaten dienen
- Im Container verwaltet CC-Container-Worker den Download von Eingabedaten, die Ausführung der Anwendung und den Upload der Ergebnisse
- Der CC-Container-Worker sendet den Status der Anwendung regelmäßig an CC-Server
- Das Protokoll der Ausführung wird in einer Datenbank hinterlegt

Quellcode und Doku unter: www.curious-containers.cc