# Virtualisierte wissenschaftliche Forschungsumgebungen und die zukünftige Rolle der Rechenzentren

Konrad Meier[1], Björn Grüning[2], Clemens Blank[3], Michael Janczyk[4], Dirk von Suchodoletz[5]

**Abstract:** Virtualisierungs- und Cloud-Technologien haben den Wandel der IT-Landschaft der letzten Jahre erheblich geprägt und werden von der Wissenschaft zunehmend für ihre eigenen Zwecke genutzt. Rechenzentren sollten auf den steigenden Bedarf in Forschung und Lehre mit der Bereitstellung geeigneter Infrastrukturen antworten. Es wird untersucht, wie Cloud-, Compute- und Lehrpool-Umgebungen so bereitgestellt werden können, dass eine bestmögliche Versorgung der Wissenschaft bei gleichzeitig effizienter Nutzung vorhandener Ressourcen erreicht werden kann. Durch die Erzeugung geeigneter virtualisierter Forschungs- und Lehrumgebungen können verschiedene Wissenschafts-Communities ihre Vorhaben durchführen, ohne hierfür weiterhin eigene Hardware-Infrastrukturen betreiben zu müssen. Es wird anhand der Umsetzung zweier prototypischer virtualisierter Forschungsumgebungen (VFU) im Bereich der Elementarteilchenphysik und der Bioinformatik diskutiert, wie zukünftige Schnittstellen zwischen Forschenden und Infrastrukturbetreibern aussehen sollten. Als Basis kommen die baden-württemberg-weit angebotenen kooperativen Forschungsinfrastrukturen des Hybrid-Clusters NEMO, der bwCloud und für den wissenschaftlichen Desktop bwLehrpool zum Einsatz. Die auf diesen Systemen eingesetzte Abstraktion durch Virtualisierung ermöglicht eine Skalierung der Ressourcen in den VFUs.

**Keywords:** VFU, virtualisierte Forschungs-, Lehr- und Lernumgebung, Cloud, Rechenzentrum

## 1 Einleitung

Digitale Datenerhebung, -verarbeitung, -verteilung und Archivierung in fast allen Bereichen der Forschung und Lehre an den Universitäten sind Kernbestandteile jeder wissenschaftlichen Tätigkeit. Diese Entwicklungen spiegeln sich in den veränderten Erwartungen an Rechenzentren wider. Mit der Verbreitung IT-gestützter Forschung und der abnehmenden Stellung von Großrechnern fand eine Verschiebung von IT-Personal in die Fakultäten statt. Sie sollten die Freiheit bei der Wahl der eingesetzten elektronischen Werkzeuge sicherstellen und Unterstützung für Forschungstätigkeiten leisten. Die Komplexität und Vielfalt der eingesetzten Systeme und Software ist in den vergangenen Jahren jedoch erheblich gestiegen und daher nicht mehr mit dem ursprünglichen Personaleinsatz zu bewältigen. Zusätzlich dazu entwickeln sich interdisziplinärer Austausch und fächerübergreifende Kollaboration zu zentralen Aspekten an heutigen Fakultäten. Entwicklungen wie virtualisierte Forschungsumgebungen ([Bu14], [Sc15]) können helfen, eine Aufgabenverteilung zwischen Forschenden und Infrastrukturanbietern zu erreichen.

[1] Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, konrad.meier@rz.uni-freiburg.de
[2] Technische Fakultät, Albert-Ludwigs-Universität Freiburg, gruening@informatik.uni-freiburg.de
[3] Zentrum für Biosystemanalyse, Albert-Ludwigs-Universität Freiburg, blankc@informatik.uni-freiburg.de
[4] Rechenzentrum, Albert-Ludwigs-Universität Freiburg, michael.janczyk@rz.uni-freiburg.de
[5] Rechenzentrum, Albert-Ludwigs-Universität Freiburg, dirk.von.suchodoletz@rz.uni-freiburg.de

## 2 Von der Virtualisierung zur Science-Cloud

Der physikalische Ort, an dem eine Software ausgeführt wird, wird durch schnelle Netze und Virtualisierung von Rechenleistung und Storage zunehmend unwichtiger. Ein Zugriff auf geographisch entfernte Standardanwendungen, die am besten in der Nähe zu den zu verarbeitenden Daten laufen, kann heutzutage sehr performant erfolgen. Die Voraussetzungen an die lokale Hardware des Anwenders sinken. Diese Entwicklung verbirgt sich hinter dem Begriff Cloud. Kommerzielle Rechenzentren setzen bei Cloud-Diensten massiv auf die „Economies of Scale"-Effekte und übersteigen die Kapazitäten – auch großer Universitäten – inzwischen um Größenordnungen. Nach dem Schritt von der Mainframe auf den ubiquitären PC stellt die Cloud die nächste Iteration der wesentlichen IT-Innovationen dar. Nachdem eine Hardware-Innovation den ersten Umbruch auslöste, wird die zweite Umwälzung nun durch eine Software- und Service-Innovation angestoßen und stellt die Alternativlosigkeit von Hochschulrechenzentren in Frage. Auf diese externen Vorgänge sollten Rechenzentren reagieren und ihr Service-Portfolio geeignet anpassen.
Virtualisierte Forschungsumgebungen (VFU) beginnen, sich Cloud-Technologien zunutze zu machen und in den verschiedenen Wissenschaftsdisziplinen zu etablieren. Sie schaffen (virtuelle) Arbeitsplattformen, die eine kooperative Forschungs- oder Lehrtätigkeit durch mehrere Wissenschaftlerinnen und Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen erlauben. Eine VFU unterstützt in verschiedenen Phasen des wissenschaftlichen Prozesses und kann je nach Disziplin von der Erhebung, Berechnung oder Verarbeitung von Eingangsdaten, der Simulation von Prozessen bis hin zur Generierung von Ergebnissen reichen. Vielfach sind Werkzeuge für Recherche und Bearbeitung von Texten und Referenzen zur Erstellung der finalen Publikation enthalten. VFUs konzentrieren sich auf die Softwareseite, technologisch basieren sie vielfach auf Virtualisierung oder Containerisierung. Sie können so die gesamte Breite von virtuellen Desktops über Cloud-Instanzen bis hin zu abgeschotteten Umgebungen im High Performance Computing abdecken. Sie vereinigen das disziplinspezifische Know-How in einem Paket, welches aus Sicht eines RZs als Black-Box auf geeigneten Forschungsinfrastrukturen ausgeführt werden kann. Einen ähnlichen Weg gehen Virtual Desktop Infrastructures für den Arbeitsplatz mit der Bereitstellung von Lehr- und Lernumgebungen [Ri16].
Durch die so erreichbare Trennung von Inhalt und Infrastruktur muss der wissenschaftliche Nachwuchs nicht mehr Zeit auf die Installation von Betriebssystem und Softwarepaketen ohne Garantie auf Erfolg verwenden. Die Aufgabenteilung ist hochgradig standardisierbar, vermeidet Fehler bei der Installation und ist eine Grundvoraussetzung für eine zuverlässig archivierbare und nachnutzbare VFU. Durch standardisierbare und wiederholbare Prozesse wird die Dokumentation der Arbeit der Forschenden vereinfacht. Zwar können unbemerkte Veränderungen durch Systemupgrades nicht komplett verhindert werden, sie sind jedoch zumindest nachvollziehbar dokumentiert und bei Bedarf rückholbar.
Wird eine wissenschaftliche Forschungsumgebung von Beginn an virtualisiert, wird lokale Hardware-Abhängigkeit eher vermieden und späteres Verschieben der Umgebung auf eine neue Virtualisierungsplattform deutlich erleichtert. So könnte ein Wissenschaftler eine Simulationssoftware auf dem lokalen Desktop bereits in einem Container oder virtuellen Maschine entwickeln und in diesem Schritt das Debugging erleichtern, da Algorithmen und Prozesse interaktiv mit direktem Feedback getestet werden können. Die erfolgreich

getestete Simulation kann auf eine Cloud oder in einem Cluster berechnet werden, da der Arbeitsrechner üblicherweise mit der Anzahl der Simulationsläufe überfordert ist. Die Ressourcen dafür können durch RZs in Form einer Compute-Cloud bereitgestellt werden. Virtualisierung von Beginn an stellt sicher, dass das Image der VM direkt in die Cloud-Umgebung kopiert und eine oder mehrere der Problemstellung angemessene virtuelle Instanzen gestartet werden können. Gleichzeitig kann das Image Kooperationspartnern zur direkten Verwendung oder Anpassung an die eigene Fragestellung wie auch Dritten zur Verifikation des Workflows zur Verfügung gestellt oder im Rahmen einer Lehrveranstaltung genutzt werden.
Die Vorteile solcher Ansätze sind erheblich, so dass die Abstraktion der darunterliegenden Infrastruktur (Hardware) sowie die Bereitstellung von standardisierten Zugriffsschnittstellen (APIs) von der Europäischen Kommission gefördert werden [Eu15]. Die interdisziplinäre Kooperation der Wissenschaftsbereiche durch die Bereitstellung von Diensten und Werkzeugen soll durch speziell bereitgestellte Ressourcen aller Infrastrukturbereiche (Netzwerk, Berechnungen, Daten, Software, Benutzer-Schnittstellen) vorangebracht werden. Die Kommission erwartet durch den Einsatz von virtualisierten Forschungsumgebungen eine effizientere Kooperation sowie eine gesteigerte Produktivität durch den einfacheren Zugang zu Daten.

### 2.1 Veränderte Rolle der Rechenzentren

Rechenzentren (und neuerdings dedizierte E-Science-Abteilungen) sollen als zentrale Einrichtungen Hilfestellung leisten und den Betrieb technischer Infrastruktur sicherstellen. Doch sind sie gleichfalls mit diesen neuen Aufgabenstellungen überfordert, da aufgrund der Breite der technischen Anforderungen, der zahllosen fachspezifischen Werkzeugen und deren Abhängigkeiten eine fallbezogene Hilfestellung nur noch in Ausnahmefällen geleistet werden kann. Beispielsweise sind die von RZs angebotenen Standarddienste wie Storage oder Server-Hosting oft zu wenig passgenau, um eine ausreichende Entlastung der einzelnen Fachdisziplinen zu gewährleisten.
VFUs müssen technisch dem Stand der Entwicklung entsprechen, bedürfen zur Entfaltung ihrer vollen Wirksamkeit jedoch der Akzeptanz in den jeweiligen Fach-Communities. Der direkte Kontakt der RZs zu den Forschenden ist notwendige Voraussetzung dafür, ihre Angebote besser zu planen und auf die Anforderungen der Forschenden bestmöglich einzugehen. Das Standardisieren und Zentralisieren von Infrastruktur am RZ entlastet Wissenschaftler von IT-administrativen Aufgaben, die beim Betreiben eigener Infrastrukturen zwangsläufig anfallen. Auf Basis der angestrebten RZ-Infrastruktur ist es möglich, VFUs zu betreiben, die speziell auf die unterschiedlichen Anforderungen der verschiedenen Arbeitsgruppen ausgelegt sind. Die Bereitstellung der Ressourcen muss dabei flexibel und zeitnah erfolgen können.
Um die notwendigen Abstimmungen und Entwicklungen gemeinsam durch Wissenschaft und Infrastrukturanbieter RZ voranzutreiben, wurde das Projekt „ViCE – Virtual Open Collaboration Environment“ ins Leben gerufen. Es entwickelt nachhaltige Geschäfts- und Steuerungsmodelle für die Kooperation von unterschiedlichsten Fach-Communities mit Rechenzentren auf Basis von VFUs, die dem dynamischen Charakter der Wissenschaften

und ihren wechselnden Anforderungen angepasst sind. Es schafft hierzu eine RZ-übergreifende Kollaborations- und Austauschplattform für virtualisierte Forschungsumgebungen, die versioniert, annotiert und geteilt werden können. Die Beschreibung der enthaltenen Tools und Workflows erlaubt die einfache Nachnutzung für neue Forschungsfragestellungen, eine schnelle Einbindung des wissenschaftlichen Nachwuchses und den Einsatz in der Lehre. Notwendige Basisinfrastrukturen der RZs werden so aufbereitet, dass sie abstrakt von verschiedenen Disziplinen einfach und ohne Startverzögerung eingebunden werden können. Hierzu wird Know-How aufgebaut, das die Wissenschaft in ihren Bedürfnissen unterstützt und eine einfache Ausdehnung auf weitere Communities erlaubt (Abb. 1).

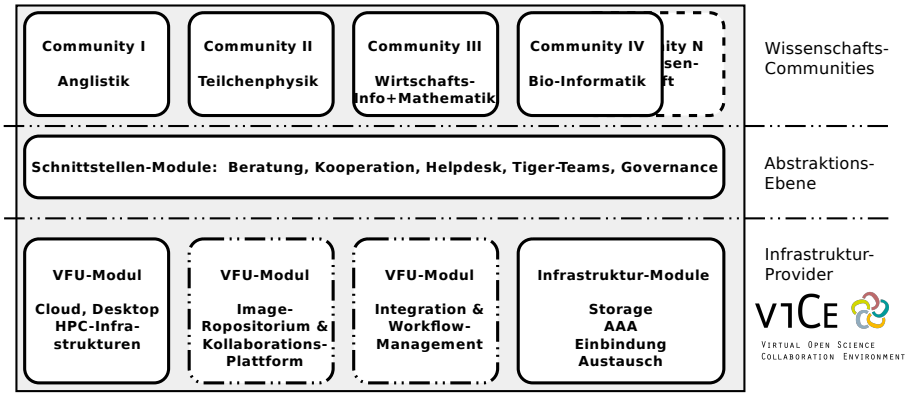


Abb. 1: ViCE sucht gemeinsam mit Infrastruktur-Providern und Wissenschafts-Communities nach einer guten Aufgabenverteilung in Nutzung und Betrieb aktueller Forschungsinfrastrukturen.

Schrittweise entstehen Best-Practices für den Betrieb von VFUs, die die Grundlagen für einen Austausch innerhalb und zwischen Standorten erlauben und perspektivisch in gemeinsamen Standards für die teilnehmenden IT-Zentren beziehungsweise deren Basisinfrastrukturen münden können. Diese orientieren sich an den Erfordernissen und Schnittstellen nationaler und internationaler Projekte und stellen so sicher, dass die Verschiebbarkeit auf breiter Ebene für weitere Fach-Communities nachnutzbar ist. Die Rekalibrierung der Aufgabenverteilung erlaubt es den Rechenzentren, zentrale und skalierende Infrastrukturen zu betreiben, und entlastet gleichzeitig die Wissenschaftler vom Betrieb eigener Infrastrukturen.

## 3 Virtualisierte Forschungs-, Lern- und Lehrumgebungen

Grundlegende Aufgabe beim Bereitstellen von VFUs ist eine flexible und schnelle Provisionierung von Infrastruktur-Ressourcen. Die bei der Bereitstellung entstehenden Herausforderungen gilt es zu lösen, um eine einheitliche Schnittstelle für die Ressourcenverwaltung zu etablieren. Dabei ist vor allem die Skalierung der Infrastruktur für bestehende Systeme eine Herausforderung, die durch neue Konzepte angegangen wird. Die sich schnell ändernden Anforderungen aus der Wissenschaft können sonst nicht adäquat beantwortet werden.

Abbildung 2 demonstriert, wie die Vereinheitlichung der Schnittstellen darauf aufbauende virtualisierte Forschungsumgebungen erlaubt. Das Beispiel zeigt eine HPC- und Cloud-Ressource eines Rechenzentrums. Auf beiden Umgebungen wird die OpenStack-Virtualisierungsplattform aufgesetzt, die eine OpenStack-API als Schnittstelle bereitstellt. Auch wenn die beiden Ressourcen im Wesentlichen unterschiedliche Anforderungen abbilden und unterschiedliche Betriebskonzepte aufweisen, sind sie aus Sicht der VFU homogen und ermöglichen einen einfachen Wechsel zwischen den Ressourcen ohne aufwändige Schnittstellen-Anpassungen. Ergänzt wird die HPC- und Cloud-Ressource um eine Desktop-Virtualisierung, um dem Wissenschaftler das Arbeiten in speziell angepassten, virtualisierten Desktop-Umgebungen am Arbeitsplatz zu ermöglichen. Wird eine wissenschaftliche Forschungsumgebung von Beginn an virtualisiert aufgesetzt, ist sie nicht abhängig von der darunterliegenden Hardware. Ein späteres Verschieben der Umgebung auf eine neue Virtualisierungsplattform wird somit erleichtert.

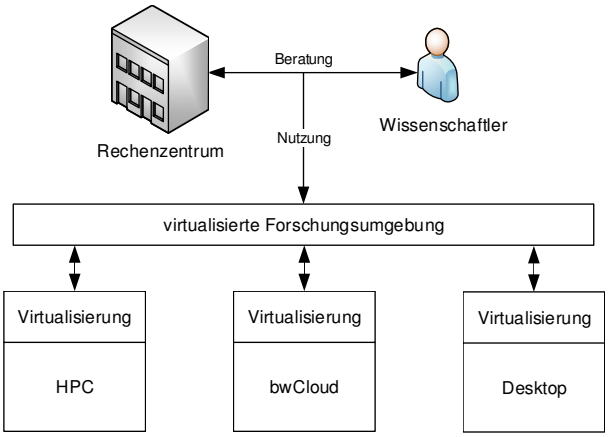


Abb. 2: Eine VFU entsteht auf dem Desktop, wird dort getestet und später zur besseren Skalierung je nach Ressourcenprofil auf das Hybrid-Cluster oder auf die bwCloud zur Berechnung verschoben.

**bwCloud:** Das Landesprojekt bwCloud[6] betreibt eine Infrastruktur-Cloud (IaaS) auf Basis von OpenStack für Wissenschaftler und Studierende. Das Betriebskonzept umfasst eine standortübergreifende Cloud-Plattform, die es den Benutzern erlaubt, virtuelle Maschinen selbst zu provisionieren. Dies ermöglicht es im Rahmen von Forschung und Lehre, die benötigten Ressourcen zeitnah zur Verfügung zu stellen. Die dafür notwendige Hardware wird von den Projektpartnern im Verbund an mehreren Rechenzentren betrieben [DSS15].

**bwHPC:** Für rechenintensive VFUs sind die Ressourcen einer Cloud-Umgebung nicht immer ausreichend. Rechencluster aus dem High Performance Computing (HPC) stellen die nächste Skalierungsstufe dar und sind für diese Fragestellungen ausgelegt. Den Wissenschaftlern in Baden-Württemberg stehen im Rahmen des bwHPC-Konzepts hochmoderne HPC-Systeme aller Leistungsklassen zur Verfügung [HWC13]. Das vom Rechenzentrum

[6] bwCloud Projektwebseite: http://www.bw-cloud.org [letzter Aufruf 11.3.2017]

der Universität Freiburg entwickelte neuartige Hybrid-Cluster-Konzept[7] gestattet es, in virtuellen Maschinen auf einem HPC-System zu rechnen. Das ermöglicht es VFUs auf einem HPC-Cluster auszuführen, ohne die darunterliegende Hardware des HPC-Systems zu partitionieren [Me16].

**bwLehrpool:** Die durch bwLehrpool[8] angebotene Desktop-Virtualisierung kombiniert auf Grundlage eines einzigen Basissystems ein flexibles Angebot verschiedenster Betriebssystemumgebungen mit einer einfachen Administrierbarkeit großer PC-Landschaften [Ri16]. Labor-, Lehr- und Lernumgebungen müssen auf diese Weise nicht mehr auf den Arbeitsplatzrechnern installiert sein, was den Wartungsaufwand der IT-Administratoren erheblich reduziert sowie Lehrenden und Forschenden vollkommen neue Gestaltungsmöglichkeiten der Lehre einräumt. Sie können Arbeitsumgebungen in einem weiten Spektrum selbst gestalten und aufgrund der Abstraktion des Systems sogar einrichtungsübergreifend austauschen. Das erlaubt Hochschulen, auf aktuelle Entwicklungen schnell zu reagieren, eine deutlich flexiblere Nutzung ihrer vorhandenen Ressourcen zu erreichen und die zentralen IT-Dienste von repetitiven Standardaufgaben zu entlasten.

### 3.1 VFU – Anwendungsfall Experimentelle Teilchenphysik

Aktuelle sowie zukünftige Experimente der Hochenergiephysik (HEP) generieren enorme Datenmengen. Um diese Daten in angemessener Zeit zu speichern, zu verarbeiten und zu analysieren, sind viele Rechenressourcen notwendig. Zusätzlich zu den aufgezeichneten Daten sind Monte-Carlo-Simulationen in gleichem Umfang erforderlich, um Beobachtungen aus dem Teilchenbeschleuniger mit theoretischen Modellen zu vergleichen.
In der Vergangenheit wurde die benötigte Rechenleistung von speziellen HEP-Clustern bereitgestellt. Die jeweils erforderlichen Betriebs- und Softwareumgebungen konnten auf den Clustern ohne Probleme installiert werden, da diese exklusiv zur Verfügung standen. Die Rechenaufgaben wurden aufgrund ihrer Unabhängigkeit in kleine Compute-Jobs aufgeteilt und gleichmäßig auf die HEP-Cluster verteilt. Eine gleichzeitige Ausführung oder eine Interprozesskommunikation (IPC) war nicht erforderlich. Durch die lose Koppelung der Compute-Jobs zählen sie zum High Throughput Computing (HTC). Die HEP-Cluster wurden dementsprechend mit geeigneten Schedulern und Infrastrukturen betrieben, um diese Compute-Jobs effizient zu verarbeiten.
Demgegenüber stehen Forschungs- und Universalcluster, die eine größere Nutzergruppe und damit andere Anforderungen bedienen müssen. Diese Cluster müssen Compute-Jobs ausführen, die auf mehreren Rechenknoten parallel laufen und damit einen Cluster-Interconnect mit geringer Latenz für die IPC benötigen. Diese Cluster zählen zur Klasse des High Performance Computings (HPC). Betriebssystem- und Softwarekonfigurationen unterscheiden sich teilweise deutlich voneinander. Sie sind darauf optimiert, die spezialisierte Hardware für IPC optimal zu unterstützen. So werden beispielsweise spezielle Compiler und Bibliotheken für das HPC-Cluster angeboten. Im Gegensatz dazu setzen die meisten

[7] Das Konzept des Hybrid-Clusters wurde erstmals als Poster auf dem 2. bwHPC Symposium 2015 in Ulm vorgestellt; http://www.uni-ulm.de/index.php?id=69869 [letzter Aufruf 3.4.2017]
[8] bwLehrpool Projektwebseite: https://www.bwlehrpool.de [letzter Aufruf 11.3.2017]

HEP-Experimente auf eine genau definierte Softwareumgebung inklusive Betriebssystem und Software aus einem zentralen Software-Repositorium. Nur so kann die Vergleichbarkeit der Ergebnisse garantiert werden, die auf unterschiedlichen Clustern berechnet werden. Darüber hinaus wird die HEP-Softwareumgebung in regelmäßigen Abständen aktualisiert und muss zeitnah auf den Clustern verfügbar sein. Das kann auf Clustern, die nicht dediziert von der HEP-Community administriert werden, nicht gewährleistet werden.
Diese Ausgangssituation macht es für HEP-Anwender schwierig, die Ressourcen von HPC-Clustern zu verwenden. Da diese Cluster jedoch einen signifikanten Anteil zu den benötigten Rechenressourcen beitragen können, wurde ein Betriebskonzept entwickelt, das es erlaubt, VFUs auf einem HPC-System zu starten. Die Ressourcen stehen für mehrere wissenschaftliche Fachbereiche und damit nicht exklusiv für einen Forschungsschwerpunkt bereit.
Der entwickelte Ablauf ist in Abbildung 3 dargestellt und zeigt, wie die VFU basierend auf dem Hybrid-Cluster-Konzept umgesetzt wurde. Dabei übermitteln Wissenschaftler ihre Compute-Jobs wie gewohnt an HTCondor. Daraufhin skaliert der Resource-Broker ROCED bedarfsabhängig und automatisiert die VFUs. HTCondor ist dabei ein Scheduling-System, das Priorisierung sowie Ressourcen-Monitoring und -Management erlaubt. ROCED (Rapid On-Demand Cloud-enabled Deployment) wird wiederum verwendet, um HTCondor zu überwachen, um je nach Ressourcenbedarf dynamisch virtuelle Maschinen zu starten oder zu stoppen.

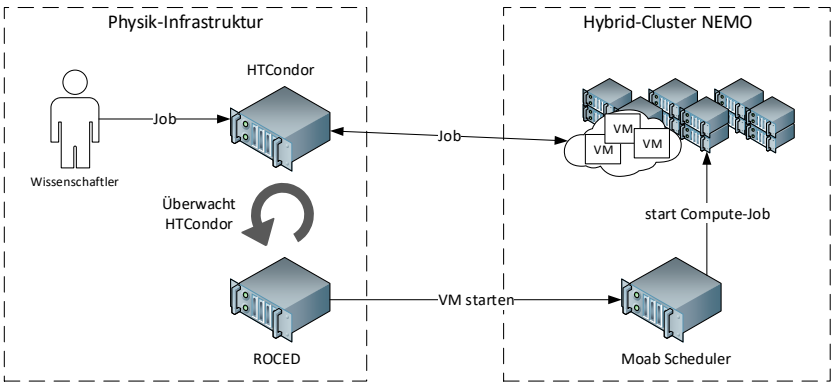


Abb. 3: Ein Resource-Broker skaliert bedarfsorientiert das Hybrid-Cluster nach Job-Übermittlung.

Das Setup zeigt eine Möglichkeit, wie spezielle Softwareumgebungen auf einem HPC-System ausführt werden können und wie die Aufgabenverteilung zwischen Wissenschaftlern und Rechenzentren in der Praxis aussehen kann. Das Bereitstellen der virtualisierten Infrastruktur (Hybrid-Cluster NEMO) wird vom Rechenzentrum übernommen. Die VFU wird vom Wissenschaftler erzeugt und nutzt die vom Rechenzentrum bereitgestellten Ressourcen. Im Fall der Teilchenphysik handelt es sich dabei um eine komplexe VFU, die unterschiedliche Ressourcen gleichzeitig verwendet. So laufen rechenintensive virtuelle Maschinen auf dem Hybrid-Cluster, während die für die Umgebung benötigten Dienste (HTCondor und ROCED) auf der bwCloud-Infrastruktur ausgeführt werden.

### 3.2 VFU – Anwendungsfall Bioinformatik

In der der Bioinformatik in Freiburg wird seit vier Jahren eine fortwährend wachsende Galaxy-Instanz betrieben, wodurch nicht nur die für die Wissenschaftler notwendige Hard- und Software verfügbar gemacht wird, sondern ebenso umfangreiche Metadaten über die technischen Anforderungen gesammelt werden können. Job-Logs mit detaillierter Aufzeichnung von Job-Metriken wie dem Speicherbedarf, der CPU-Leistung und der Laufzeit erlauben eine umfangreiche Analyse des Bedarfs sowie des Optimierungspotentials. Durch stetig wachsende Nutzerzahlen, Datenmengen und wachsender Komplexität der Software zeigen sich hierdurch Grenzen des bisherigen Setups, welches auf einem eigenen Cluster der Bioinformatik läuft und über die Sun Grid Engine (SGE) angesprochen wird.
Die Galaxy-Plattform ist ein auf Unix-Maschinen lauffähiges Software-Framework für das Management von wissenschaftlichen Workflows und Pipelines, für die Datenverarbeitung und für die Durchführung von Rechenjobs. Motiviert insbesondere durch die technischen Probleme, die mit dem Next Generation Sequencing (NGS) und der damit einhergehenden High-Throughput-Datenerzeugung aufkommen, ist es das Ziel, aus den großen Datenmengen möglichst viel wissenschaftliche Information zu gewinnen. Dies bedarf komplexer statistischer und algorithmischer Methoden, die nur aufgrund großer zur Verfügung stehender Rechenleistung realistisch geworden sind. Das führt zwangsweise zu Problemen in biologischen Fachbereichen, da hierdurch Wissenschaftler ohne einen informationstechnischen Hintergrund auf hochkomplexe computergestützte Analysen angewiesen sind. Das seit 2005 bestehende Galaxy-Projekt nutzt ein einfach zugängliches Webinterface, um unabhängig vom IT-Hintergrund Zugang zu umfangreichen Tools und den benötigten technischen Ressourcen zu ermöglichen und um komplexe Analysen durchzuführen.
Am Beispiel von Galaxy wird in der Softwarelandschaft der Bioinformatik eine Infrastruktur aufgebaut, die für den Nutzer das Handling auf Seiten der Software enorm verbessert. Darüber hinaus werden mittels Virtualisierung und Containerisierung die Hardware-Abhängigkeiten auf ein Minimum reduziert. Dadurch gelingt es, Forschenden abgestimmte, virtualisierte Forschungsumgebungen anzubieten, welche sowohl auf einem Desktop als auch in der Cloud und im HPC laufen [Af16].
Hierzu gibt es zusammen mit der Bioinformatik-Community Bestrebungen, verfügbare und zukünftige Software in einer Paketverwaltung (Conda) einzugliedern. Diese läuft im Userspace und erlaubt es, automatisiert, transparent und reproduzierbar vorkompilierte Pakete zu erstellen, welche sich in einem weiteren automatisierten Schritt zusätzlich in verschiedene Container-Formate übersetzen lassen. Dies macht sowohl eine Versionierung, als auch einen Austausch von Forschungsdaten inklusive deren Reproduzierung und die automatische Erfassung von Metadaten bezüglich eines verwendeten Workflows und dessen Tools einfach.
Aufgrund von wachsender Nachfrage und der angestrebten Nutzung, den bisher auf Freiburg und Umgebung beschränkten Dienst national und international zu öffnen, ist eine flexible Nutzung bereitgestellter Ressourcen notwendig. Dies erlaubt, auch auf zukünftige Veränderungen reagieren zu können und mit dem Dienst weiter zu skalieren.

In Abbildung 4 ist das entwickelte Setup der Freiburger Galaxy-Installation zu sehen. Zusätzlich zum bisherigen Cluster der Bioinformatik können nun über HTCondor Rechenjobs verwaltet werden, um sie unter anderem auf dem Hybrid-Cluster NEMO oder

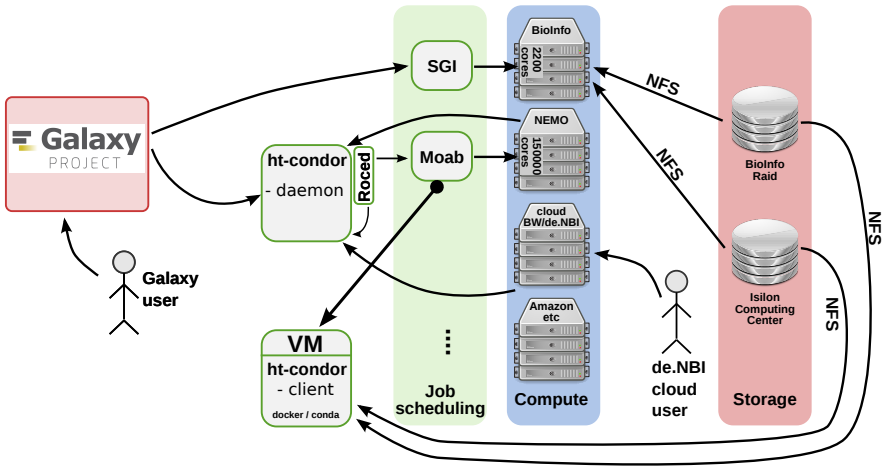


Abb. 4: Die Freiburger Galaxy-Instanz entscheidet je nach Workflow, ob eine Berechnung auf dem Bioinformatik-Cluster (SGI) oder in der Bioinformatik-VFU erfolgen soll.

auf Compute-Clouds wie der bwCloud oder de.NBI-Cloud zu starten. Dies erlaubt ein Setup, das weitgehend unabhängig von der zugrundeliegenden Hardware-Infrastruktur ist. Dadurch können grundsätzlich auch weitere und kommerzielle Anbieter genutzt werden (Eucalyptus, OpenNebula, OpenStack, Amazon EC2, etc.). Anhand der in Galaxy gesammelten Job-Metriken wurden Anwendungsprofile erstellt, die Berechnungen automatisch auf geeignete Hardware schicken. Für die Nutzer ändert sich dadurch nichts. Sie können weiterhin mit dem Webfrontend arbeiten, das sie auch von einer lokalen Galaxy-Instanz auf dem Desktop kennen. Die Verwaltung der Tools in einem Paketmanager und die Virtualisierung oder Containerisierung stellen sicher, dass das wissenschaftliche Arbeiten flexibel auf jeglicher Hardware transparent und mit reproduzierbaren Ergebnissen durchgeführt werden kann.

## 4 Fazit

Eine zukünftige IT-Strategie moderner Hochschulrechenzentren konzentriert sich neben der Grundversorgung auf notwendige Forschungsinfrastrukturen. Da diese einerseits von „Economies of Scale“-Effekten profitieren und damit andererseits eine attraktive Vielfalt an Diensten angeboten werden kann, bieten sich hier kooperative Ansätze an. Virtualisierte Forschungsumgebungen helfen dabei, die Ortsabhängigkeit aufzulösen und gemeinschaftlich zentrale Dienste anzubieten. Mit dieser Sichtweise erlauben Virtualisierung und Compute-Clouds eine neue Aufteilung der Aufgaben zwischen Wissenschaftlern, Instituts-IT und dem Rechenzentrum. Das Rechenzentrum stellt notwendige Basis- und Forschungsinfrastrukturen, angefangen vom Netz bis hin zu hochwertigen Storage-, Servervirtualisierungs- und HPC-Diensten zur Verfügung. Auf Basis dieser können Forschende ihre eigenen Umgebungen und Methoden frei realisieren. Sie werden dabei zusätzlich durch Angebote zu automatischem Backup und Beratung zum Forschungsdatenmanagement un-

terstützt.
Der Übergang vom klassischen wissenschaftlichen Rechenzentrum der 1990er zu einem Rechenzentrum für die Wissenschaft kann oft nicht aus den vorhandenen Ressourcen gestemmt werden. Um solche Angebote erfolgreich weiterentwickeln zu können, muss sich ein Rechenzentrum auch an Forschungs- und Infrastrukturbegleitprojekten beteiligen. Diese Neuausrichtung ist längst nicht abgeschlossen und wird anhalten, so lange es gesellschaftliche, politische, hochschulspezifische und technische Veränderungsprozesse gibt. Rechenzentren müssen sich diesen Herausforderungen stellen und Weiterentwicklungen meistern: „Ohne ständige Innovationen an einem IZ/RZ wird es obsolet“[9].

## Literaturverzeichnis

[Af16] Afgan, Enis; Baker, Dannon; van den Beek, Marius; Blankenberg, Daniel; Bouvier, Dave; Čech, Martin; Chilton, John; Clements, Dave; Coraor, Nate; Eberhard, Carl; Grüning, Björn; Guerler, Aysam; Hillman-Jackson, Jennifer; Von Kuster, Greg; Rasche, Eric; Soranzo, Nicola; Turaga, Nitesh; Taylor, James; Nekrutenko, Anton; Goecks, Jeremy: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Research, 44(W1):W3–W10, 2016.

[Bu14] Buddenbohm, Stefan; Enke, Harry; Hofmann, Matthias; Klar, Jochen; Neuroth, Heike; Schwiegelshohn, Uwe: A life cycle model für collaborative research environments. In (Müller, Paul; Neumair, Bernhard; Reiser, Helmut; Rodosek, Gabi Dreo, Hrsg.): 7. DFN-Forum - Kommunikationstechnologien, 16.-17. Juni 2014, Fulda, Germany. Jgg. 231 in LNI. GI, S. 1–10, 2014.

[DSS15] Dulov, Oleg; Scheibenberger, Klaus; Schulz, Janne Chr.: bwCloud – Standortübergreifende Servervirtualisierung. SCC-News, 01/2015, 2015.

[Eu15] European Commission: EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE). http://cordis.europa.eu/programme/rcn/664625_en.html, 2015.

[HWC13] Hartenstein, Hannes; Walter, Thomas; Castellaz, Peter: Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechnen und datenintensive Dienste. Praxis der Informationsverarbeitung und Kommunikation, 36(2):99–108, 2013.

[Me16] Meier, Konrad; Fleig, Georg; Hauth, Thomas; Janczyk, Michael; Quast, Gürnter; von Suchodoletz, Dirk; Wiebelt, Bernd: Dynamic provisioning of a HEP computing infrastructure on a shared hybrid HPC system. Journal of Physics: Conference Series, 762(1):012012, 2016.

[Ri16] Ritter, Steffen; Trahasch, Stephan; Slotosch, Sven; von Suchodoletz, Dirk; Münchenberg, Jan: bwLehrpool: Durchführung von elektronischen Prüfungen in virtualisierten Umgebungen. In (Lucke, Ulrike; Schwill, Andreas; Zender, Raphael, Hrsg.): DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik, 11.-14. September 2016, Potsdam. Jgg. P-262 in LNI. GI, S. 149–154, 2016.

[Sc15] Schwarzkopf, R.: Virtual machine lifecycle management in grid and cloud computing. Philipps-Universität, 2015.

[9] Stefan Wesner, Keynote auf der 17. ZKI-Herbsttagung, https://www.uni-ulm.de/fileadmin/website_uni_ulm/zkiherbst2016/praesentationen/dienstag/wesner_gestaltung_von_innovationsprozessen_an_rechen_und_informationszentren.pdf, S. 41