Betriebskonzept für das Wissenschaftliche Rechnen (WR) an der TU Dortmund (zweite Fassung)

(Version: DoWiR-20140602)

I Einleitung

Die Bedeutung des Wissenschaftlichen Rechnens (WR) hat in den letzten Jahren kontinuierlich zugenommen und ist aus Forschung und Lehre nicht mehr wegzudenken. In vielen Disziplinen – insbesondere den MINT-Fächern - sind numerische Simulation und Optimierung auf leistungsfähigen Rechnersystemen eine Schlüsseltechnologie. Laut Wissenschaftsrat „erfordert Wissenschaftliches Rechnen eine Synthese aus fachspezifischer Kompetenz, mathematischer Methodik und informationstechnischem Instrumentarium auf höchstem Niveau.“[1]

An der TU Dortmund haben sowohl die methodisch-fachlichen Kompetenzen als auch der Betrieb des informationstechnischen Instrumentariums, das insbesondere hochperformante Rechnerressourcen umfasst, eine lange Tradition. Für die Universitätsallianz Ruhr (UAR), insbesondere aber für den Eigenbedarf der Dortmunder Forscher, wurde und wird Infrastruktur in Form von High-Performance-Computing-(HPC-)Ressourcen (z.B. LIDOng und DGRZR) für WR bereitgestellt.

Wie aus dem Zwischenbericht für LiDOng[2] ersichtlich wird, wurde die Infrastruktur für zahlreiche Promotionen, Master-, Diplom- und Bachelorarbeiten eingesetzt. Für verschiedene u.a. von der DFG geförderte kollaborative Drittmittelprojekte (z.B. SFB 823, SFB 708, SFB 876 und SFB TRR 30, GRK 1032) wurden und werden die HPC-Ressourcen durch die Forscherinnen und Forscher in wesentlichen Teilen für ihre Arbeit genutzt.

Für die Weiterentwicklung des WR am Standort Dortmund in den kommenden Jahren ist es erforderlich, den aktuellen Bedarf an HPC- und anderen Rechenressourcen zu erfassen, um auf dieser Basis ein Betriebskonzept für WR an der TU Dortmund zu entwickeln, das die unterschiedlichen Anforderungen adäquat abbildet. Hierzu wurden vom Dortmunder Zentrum für Wissenschaftliches Rechnen (DoWiR) Ende 2012 und 2013 zwei Umfragen durchgeführt, die erste unter den Arbeitsgruppen der MINT-Bereiche der TU Dortmund, die zweite unter allen Arbeitsgruppen. Die Ergebnisse werden im Folgenden zusammengefasst und aus dem resultierenden Bedarf (bewusst beschränkt auf die TU Dortmund) wird ein zukunftsweisendes Konzept entwickelt, wie die unmittelbare forschungs- und auch lehrnahe (IT-)Grundversorgung für das WR, die in das IT-Serviceangebot der TU eingebettet ist und zu der z.B. Serverhosting- sowie Web- und E-Mail-Dienste gehören, kosteneffizient und bedarfsorientiert gestaltet werden kann. Neben dem Betrieb der zugehörigen HPC-IT-Infrastruktur spiegelt sich im nachfolgenden drei-gegliederten Betriebskonzept auch die Bereitstellung fachspezifischer Kompetenz im WR wider, ohne die eine effiziente Nutzung der vorhandenen und weiter auszubauenden Infrastruktur nicht möglich ist. Zudem wird auf den Kontext der Gauß-Allianz und ihre Rolle für die Entwicklung des HPC-Standorts Westfalen hingewiesen.

[1] Wissenschaftsrat, Drs. 8619-08, Berlin, 2008

[2] Bericht für die DFG zur Nutzung des Großforschungsgerätes INST212/204-1 (LIDOng), Dortmund, 2012

II Ergebnisse der Umfragen

An der ersten Umfrage Ende 2012 haben 27 Arbeitsgruppen mit insgesamt 307 Mitarbeitern (davon ca. 50% drittmittelfinanziert) teilgenommen, die ein Drittmittelvolumen von durchschnittlich 17 Mio. Euro p.a. in den letzten 5 Jahren vertreten und die sehr aktiv bzgl. Lehre und Ausbildung im WR sind. Bei der zweiten Umfrage Ende 2013 gab es Rückmeldungen von 51 Arbeitsgruppen mit 507 Mitarbeitern. Im nachfolgenden sind ausschließlich die Ergebnisse der zweiten Umfrage dargestellt.

Für die Gesamtheit der Teilnehmer an der Umfrage ist die Nutzung von WR-Ressourcen für ihre **Forschung und Lehre** unverzichtbar bzw. sehr wichtig. Bislang wird WR an der TU Dortmund zentral in Gestalt des Hochleistungsrechners LiDOng, der Grid-Ressource DGRZR und dezentral auf mehreren lokalen Cluster-Installationen an den Lehrstühlen betrieben.

Von den zentralen Ressoucen LiDOng und DGRZR wurden im Zeitraum 11/2012 bis 10/2013 9 Mio. Core-Stunden von LiDOng (Dortmunder Anteil im Rahmen der UAR bzw. Dortmund allein) und 6 Mio. Core-Stunden von DGRZR (gesamt) verbraucht. Die Nutzung von LiDOng hat sich in Q1/2013 verändert. Der Wegfall der Nutzung aus Bochum und Duisburg-Essen hat nicht zu einer geringeren Auslastung des Clusters geführt, sondern zu einer verstärkten Nutzung durch Dortmunder Wissenschaftler. Die Kapazität der lokalen Cluster wurde in der Umfrage mit 423 Servern und 3406 Kernen angegeben. Geschätzt ergeben sich damit für ein Jahr bei einer angenommenen Auslastung von 50% insgesamt 0,5 x 3406 x 365 x 24 = 15 Mio. Core-Stunden. Zusammen mit den verbrauchten Ressourcen von LiDOng und DGRZR ergab die Umfrage somit einen jährlichen Bedarf von ca. 30 Mio. Core-Stunden, die von der neuen zentralen Ressource zur Verfügung gestellt werden müssen.

Das Umfrageergebnis zeigt eine starke Heterogenität des Bedarfes: Die Anforderungen der Anwender unterscheiden sich sehr stark in Hinblick auf Anzahl Cores pro Knoten, Hauptspeichergröße, Zusatzhardware (GPU etc.), Netzverbindung und Gesamtkapazität des Systems. 70% der Teilnehmer, die sich zur Frage Hauptspeichernutzung für serielle Jobs geäußert haben, benötigen für ihre Jobs bis zu 32 GB Hauptspeicher, 16% benötigen bis zu 64 GB Hauptspeicher, 13% sogar bis zu einem Terabyte. Ca. 50% der Benutzer rechnen parallel (MPI oder OpenMP). 9% der Benutzer benötigen für ihre parallelen MPI-Jobs mehr als 512 GB Gesamt-Hauptspeicher; die neue Ressource sollte daher in der Lage sein, eine hinreichend große Anzahl dieser Jobs unterstützen zu können. 20% der Umfrageteilnehmer benötigen für OpenMP-Rechnungen bis zu 256 GB Hauptspeicher (gleichbedeutend mit dem Gesamtspeicherausbau eines Knotens). 70% der Benutzer können ihre Jobs mit einer Ausstattung von 8 Cores oder weniger laufen lassen, 3% geben 64 Cores als Maximalanforderung an, für die anderen sind bis zu 32 Cores pro Knoten nötig. Für 20% ist die Verfügbarkeit eines Hochleistungsnetzwerkes wichtig, für weitere 40% der mögliche Einsatz von GPUs.

Neben den wissenschaftlichen Aspekten hat für 88% der Umfrage-Teilnehmer das WR auch in der **Lehre** einen sehr wichtigen oder unverzichtbaren Stellenwert, d.h. dass die WR-Ressource auch in der Lehre genutzt wird und weiterhin werden soll (Ressourcen zur Nutzung in Lehrveranstaltungen, entsprechende Kurse). Dies ist insbesondere vor dem Hintergrund wichtig, spätere wissenschaftliche Mitarbeiter schon frühzeitig in der Nutzung von WR-Ressourcen auszubilden und an diese Werkzeuge heranzuführen.

Ein Großteil der Anwender (86%, bzw. 88%) erweitert vorhandene Software mit eigenen Elementen oder entwickelt selbst Software, so dass hier eine umfangreiche Entwicklungsumgebung (Compiler, Debugger, etc.) vorzusehen ist.

54% der Gruppen nutzen bereits LiDOng, 19% zudem die Höchstleistungsrechner im deutschen und europäischen Rahmen, wodurch der große Bedarf an HPC-Leistung dokumentiert wird.

Seitens der Umfrageteilnehmer wurde ein großer Bedarf an Schulungen, Kursen und Unterstützung für lokale Administratoren zum Ausdruck gebracht. Bei den Grundlagenkursen wurden häufig die Themen Sprachkurse C++, Fortran, Crashkurs Benutzung von Cluster-Systemen mit Batch-Betrieb, parallele Programmiertechniken MPI/OpenMP und Crashkurs GPU/CUDA genannt.

III Eckpunkte des Betriebskonzepts

Die nachfolgenden Eckpunkte beschreiben die Klassifikation und Details der WR-Ressourcen, wie sie aus Sicht der Dortmunder Forscher (innerhalb von DoWiR) zum wissenschaftlichen Arbeiten an der TU Dortmund benötigt werden. Grundlagen sind die oben ausgeführten Ergebnisse der von DoWIR durchgeführten zweiten Umfrage, sowie die Verbrauchs-Istdaten der letzten 12 Monate von LiDOng (Dortmunder Anteil) und DGRZR für die Kapazitätsschätzung.

Die Heterogenität des aktuellen Bedarfs an WR-Ressourcen ist charakteristisch für die heterogenen Arbeitsfelder der Gruppen, die WR in Forschung und Lehre einsetzen. Das Betriebskonzept muss diesen Bedarf abbilden und unter dem Gesichtspunkt der ökonomischen Optimierung strukturierend steuern. Sinnvoll ist in diesem Sinne die Klassifizierung der benötigten Ressourcen in drei Betriebsmodelle,

(a) lokal aufgestellte Ressource, lokal verwaltet („lokal/lokal“),

(b) zentral aufgestellte Ressource, lokal verwaltet („zentral/lokal,“ „Housing“-Konzept),

(c) zentral aufgestellte Ressource, zentral verwaltet („zentral/zentral“),

wobei die im WR aktiven Gruppen aus Gründen intern heterogener Anforderungen teilweise auch mehr als nur eine Klasse nutzen. Vor dem Hintergrund der reinen Kosteneffizienz (gemessen am Gesamtaufwand für Betriebspersonal, Energie, Klimatisierung, Wiederbeschaffung pro verbrauchter Core-Stunde) sinkt der entsprechende finanzielle Aufwand von (a) nach (c). Gleichzeitig müssen aber Faktoren berücksichtigt werden, die bei ausschließlichem Angebot des Modells (c) die wissenschaftliche Arbeit erschweren bzw. unmöglich machen (Details s.u.). Diese Faktoren sind schwer mit eindeutigen Kennzahlen zur Entscheidungsgrundlage zu beziffern, sodass ein TU-weites Betriebskonzept nicht strikte Vorgaben machen kann, sondern *Anreize* bieten muss, um die Arbeitsgruppen graduell hin zu den ökonomisch effizienteren Modellen - d.h. zu (c) oder wenigstens zu (b) - zu führen (die auch nur dann optimal eingesetzt werden, wenn die Auslastung gleichbleibend hoch ist), ohne dass qualitative und quantitative Einbußen befürchtet werden.

Unter dieser Prämisse greift das Betriebskonzept gleichzeitig das Konzept des „Cloud Computing“ auf, das prinzipiell die erforderliche Flexibilität zumindest teilweise bereitstellen kann, das allerdings nicht in allen an der TU praktizierten Einsatzszenarios vollständig alle Randbedingungen erfüllt und zudem gemessen an der Leistungsfähigkeit eines realen im Vergleich mit einem virtuellen Core um bis zu eine Größenordnung penalisiert ist, wie durch Benchmarks belegbar ist (Evaluation der IBM-Testcloud mit Anwendungen von DoWiR-Mitgliedern).

(a) „lokal/lokal“

Grundsätzlich besteht Bedarf an solchen Ressourcen, da sie alleinig bestimmte Anforderungen erfüllen bzw. Nutzungsszenarios abbilden können. So ist bei der Verarbeitung großer Datenmengen oder bei Visualisierung in Echtzeit die Vernetzungslatenz bzw. das sehr hohe zu transferierende Datenvolumen hier nicht zu vernachlässigen. Das Modell „lokal/lokal“ bietet eine sehr hohe Flexibilität um „ad hoc“-Anpassungen vorzunehmen und über die Ressourcen auch außerhalb von Servicezeiten im direkten Zugriff zu verfügen. Zudem besteht häufig Bedarf für Spezialhardware, Jobdauern über die starren Queue-Vorgaben hinaus sowie für die instantane Verfügbarkeit zu Entwicklungs- und Testzwecken. Wenn das Messen von Systemparametern für das Testen verschiedener, häufig selbst entwickelter Ausführungsumgebungen das Erkenntnisinteresse ist,

brauchen die Nutzer selbst den direkten Zugriff. Darüber hinaus sind Software-Lizenzen häufig an lokale Hardware gebunden bzw. in zentralen Ressourcen oder in Cloud-Umgebungen nicht einsetzbar. Die Personalkosten zur lokalen Administration können weiterhin nicht als reine Vollkostenbelastung gesehen werden, da die betreffenden Mitarbeiter an der Lehrstühlen im Laufe ihrer Arbeit auch entsprechende Kompetenzen für ihre künftigen Arbeitsfelder außerhalb der Universität erwerben sollen und häufig auch wollen. Um jedoch insbesondere Betriebskosten einzusparen und für eine höhere Auslastung der Ressourcen zu sorgen, ist es zweckmäßig, eine Reduzierung der Anzahl lokaler Serverräume anzustreben. Ebenso sollen geeignete Maßnahmen ergriffen werden, um den Energieverbrauch zu senken, z.B. das gezielte Ausschalten im Moment nicht benötigter CPU-Cores.

(b) „zentral/lokal“

Mit der Verfügbarkeit der neuen Räume des ITMC wird die zentrale Aufstellung (und auch koordinierte Beschaffung) lokal beschaffter und administrierter Ressourcen erleichtert, um weitere Kosten einzusparen. Arbeitsgruppen müssen motiviert werden, bisherige Systeme nach Modell (a) soweit wie möglich in diesem Modell (wenn (c) nicht möglich ist) zu betreiben, das einen Kompromiss zwischen den lokalen Bedürfnissen (s.o.) und der Anforderung der Kosteneffizienz darstellt.

(c) „zentral/zentral“

Die zentrale Ressource muss eine Vielzahl an unterschiedlichen Aufgaben bewältigen können. Ein Teil dieser Ressource ist primär für die klassischen HPC-Anwendungen, insbesondere massiv parallele numerische Simulationen, andererseits aber auch für serielle Jobs, insbesondere für solche, die über die Grid-Schnittstelle eingestellt werden, vorgesehen. Die oben eingeführte zentrale Cloud-Ressource für wissenschaftliche Anwendungen soll ebenfalls in diesem Rahmen implementiert werden. Diese Teil-Ressource der zentralen Installation wird Anwendern angeboten, die für ihre wissenschaftliche Arbeit z.B. eigene Betriebssystemumgebungen benötigen oder die keine zeitkritischen bzw. ressourcenintensiven Jobs ausführen wollen. Von zentraler Bedeutung ist, dass Compute-Ressourcen flexibel und bedarfsorientiert zwischen diesen beiden WR-Instanzen verschoben werden können, um die Auslastung zu maximieren. Für die Datenspeicherung ist eine zentrale Storage-Ressource mit einem parallelen Filesystem vorgesehen, auf das beide zentralen WR-Instanzen Zugriff haben.

Im nachfolgenden Abschnitt wird der notwendige Ressourcenbedarf von 30 Mio. Corestunden (wobei man von anwachsendem Bedarf ausgehen muss) komplett in der zentralen Ressource abgebildet. Auch wenn der Bedarf und die Notwendigkeit an lokalen Ressourcen unbestritten ist, sollen doch auf längere Sicht die Ressourcen zum größten Teil zentral gebündelt und bei zukünftigen Beschaffungen der TU Dortmund berücksichtigt werden, entsprechend der Favorisierung des Modells „zentral/zentral.“

IV Realisierung

Die im Folgenden dargestellte Realisierung der Anforderungen stellt nur eine Momentaufnahme vor dem Hintergrund der aktuellen technischen Entwicklung dar.

Um die Anforderungen unter wirtschaftlichen Gesichtspunkten erfüllen zu können, ist eine Einteilung in folgende Knotentypen – dem oben erwähnten heterogenen Ansatz folgend - sinnvoll:

1. Standardknoten,
2. Knoten mit schneller Vernetzung und Beschleunigerhardware (z.B. GPU, Xeon Phi),
3. SMP-Knoten mit vielen Cores und großem Hauptspeicher.

Die folgende Kapazitätskalkulation geht von der minimal zu verarbeitenden Core-Stundenanzahl von 30 Mio. pro Jahr aus, wobei versucht wird, möglichst viele Anforderungen der Benutzer aus der Umfrage zu erfüllen. Das Scheduling der Jobs soll attributorientiert arbeiten; die Queues des Systems erfassen alle Ressourcen und dienen nur zur reinen Rechenzeitverteilung. Hierfür wird die benötigte Hardware durch Angabe entsprechender Attribute ausgesucht, um eine hohe Auslastung zu gewährleisten, z.B. wird ein Knoten mit schneller Vernetzung dann von einem seriellen Job belegt, wenn zu diesem Zeitpunkt kein Job vorhanden ist, um die schnelle Vernetzung auszunutzen. Stellt ein Job keine besonderen Anforderungen an die Ressource, kann er auf jedem Knoten zur Ausführung kommen.

Um 30 Mio. Core-Stunden pro Jahr erbringen zu können, werden minimal 3425 Cores benötigt; um jedoch dem steigenden Bedarf in der Zukunft, der in der Umfrage zum Ausdruck gebracht wurde, gerecht zu werden, sollte diese Zahl auf 3584 Cores erhöht werden. 50% der Benutzer arbeiten ausschließlich seriell (1792 Cores), bei einem Doppel-Quad-Core-System werden dafür 224 Knoten benötigt. Sechs- oder Mehrkern-Prozessoren sind für diese Zielgruppe nicht zu empfehlen, da bei einer höheren Core-pro-CPU-Zahl die Speicherbandbreite zu gering sein wird. Um die Majorität der Anforderungen bezüglich der Hauptspeichergröße zu erfüllen, sind 32 GB Hauptspeicher pro Knoten in serieller Nutzung vorzusehen. Um aber den Durchsatz bei IO-Operationen im Netzwerk zu erhöhen, ist es sinnvoll, darüberhinaus den Hauptspeicher auf 64 GB zu erhöhen, um die Daten im Hauptspeicher puffern zu können. Der Anteil der Benutzer, die Beschleunigerhardware nutzen, beträgt 40%, während 20% ein leistungsfähiges Netzwerk benötigen. Es ist sinnvoll, einen Teil der Knoten mit schnellem Netzwerk und Beschleunigerkarten (GPUs, Xeon Phi) auszustatten. Die Anzahl der schnell vernetzten beträgt somit 96, die Hälfte dieser Knoten soll mit je zwei Beschleunigerkarten ausgestattet werden. Die SMP-Knoten benötigen für die der Umfrage entnommenen Anforderung 32 CPU-Cores (4x Achtkernprozessor, 64 Cores mit Hyper-Threading) und einen Hauptspeicher von 256 GB. Ein Teil dieser Knoten soll mit 1024 GB Hauptspeicher ausgestattet werden, um Ressourcen für große Datenmengen anbieten zu können. Für diesen Knotentyp ist eine hohe Core-Zahl wichtig, sodass Einschränkungen der Speicherbandbreite hingenommen werden können. Da der SMP-Knotentyp auch für den Einsatz in der Cloud-Ressource vorgesehen ist, ist die Anzahl dementsprechend zu erhöhen, so dass hier 32 Knoten zum Einsatz kommen sollen.

Um die Anforderungen bezüglich des Plattenspeichers erfüllen zu können, wird ein Ausbau von (netto) 1,2 Petabyte (PB) benötigt. Um einen schnellen Zugriff von allen Ressourcen aus zu gewährleisten, ist der Einsatz eines parallelen Filesystems zwingend notwendig. Ferner ist jedoch eine ausreichende Ausstattung an lokalem Plattenplatz vorzusehen (mindestens 1 TB pro Knoten), weil der Zugriff auf ein paralleles Filesystem immer durch die Leistungsfähigkeit des Netzwerks limitiert ist und (bei ausschließlicher Verwendung) IO-intensive Anwendungen massiv ausbremsen würde.

Zusammenfassung

Knotentyp	Anzahl
Standard-Knoten, 8 Cores, 64 GB Hauptspeicher	224 Knoten mit 1792 Cores
Stark vernetzte Knoten mit/ohne Acceleratoren, 8 Cores, 64 GB Hauptspeicher, volle Infiniband-Bandbreite	96 Knoten mit 768 Cores, davon 32 mit je zwei Acceleratoren
SMP/Cloud-Knoten, 32 Cores, 256 GB/ 1 TB Hauptspeicher	32 Knoten mit 1024 Cores, davon 4 mit 1 TB Hauptspeicher
Gesamt:	352 Knoten mit 3584 Cores und 31744 GB Hauptspeicher

Das System soll von der Infrastruktur her auf eine spätere Erweiterung ausgelegt sein, sofern später mehr Ressourcen benötigt werden sollten. Ebenso soll das System auch für Drittmittelinvestitionen der beteiligten Arbeitsgruppen offenstehen.

Ausgehend von den Anforderungen an ein energieeffizientes System sollte der Energieverbrauch pro CPU 80 W nicht überschreiten. Aufgrund der Erfahrungen mit LiDOng kommen dazu noch ca. 60 W anteilige Verbrauchskosten pro CPU (Speicher, Festplatten, Netzwerk, Infrastruktur) hinzu, ferner verbrauchen GPUs und ähnliche Acceleratoren ca. 200 W pro Knoten, so dass das hier skizzierte System pro Jahr (Knoten x Anzahl CPUs x Stunden pro Jahr x Leistungsaufnahme pro CPU) + (Anzahl GPU-Knoten x Stunden pro Jahr x Leistungsaufnahme pro Grafikkarte) (224 x 2 +96 x 2 + 32 x 4) x (80 + 60) x (24 x 365) + (64 x 200 x 24 x 365) Wh = 1054003 kWh verbrauchen wird. Bei einem angenommenen Strompreis von 0,17 EUR/kWh und der Annahme, dass für jede verbrauchte kWh auch 0.73 kWh an Kühlleistung aufgewendet werden muss, ergeben sich somit jährliche Kosten für Strom und Kühlung von ca. 310.000 EUR.

Die Gesamtkosten für einen angenommenen Betrieb über drei Jahre betragen damit ca. 3,93 Mio. EUR (bei einer Investitionssumme von 3 Mio. EUR für das zentrale System (mit 300.000 EUR Eigenanteil der TU) und jährlichen Kosten für Strom und Kühlung von ca. 310.000 EUR).

Der Zugriff auf die HPC-Ressource erfolgt batchorientiert, zusätzlich ist eine Schnittstelle vorgesehen, die die Ressource als Grid-Service anbietet. Für die Cloud-Verwaltung ist eine geeignete Benutzerschnittstelle vorzusehen. Um flexibel genug auf die sich ändernden Anforderungen reagieren zu können, aber einen bezahlbaren und verlässlichen Betrieb nicht aus den Augen zu verlieren, muss ein solches hybrides Auslastungs- bzw. Betriebskonzept stetig weiterentwickelt werden.

Eine Alternative zur Beschaffung von Hardware und dem Betrieb der Hardware vor Ort ist die Anmietung von Ressourcen in Form einer Cloud-Lösung. Die Firma IBM bietet im Rahmen ihres „SmartCloud“-Angebotes derartige Ressourcen an. Die Abrechnung erfolgt in Stunden pro virtuellen Server (aktuell besteht solch ein Server aus 16 Cores und 32 GB Hauptspeicher), ferner sind die Kosten für den permanenten Speicherplatz und den Netzwerktransfer zu berücksichtigen. Für die 30 Mio. Corestunden wären 30 Mio./16 = 215 virtuelle Serverknoten notwendig, für die bei einem Preis von 0,647 EUR pro Server und Stunde jährliche Kosten von 1,22 Mio. EUR entstehen (Stand: Mai 2013). Für den Speicher von 1,2 PB und einen angenommenen externen Netzwerktransfer von 40 TB/Monat fallen jährliche Kosten in Höhe von 1,45 Mio. EUR an. Zusammengefasst würde eine Cloud-Realisierung über eine Laufzeit von 3 Jahren Kosten in Höhe von ca. 8 Mio. EUR verursachen. Jedoch erfüllt diese Realisierung nicht alle technischen Anforderungen (Hauptspeicher jenseits der 32GB, viele Cores, Beschleunigerkarten, latenzarmes Hochleistungsnetzwerk), zudem ergibt sich aufgrund der notwendigen Virtualisierung ein Leistungsverlust an Rechenleistung, der je nach Anwendung einen Faktor zwischen drei und sieben ausmachen kann.

Daher ergibt sich im Vergleich zu einer ausschließlich externen Cloud-Lösung sowohl aus wirtschaftlichen als auch anwendungstechnischen Gründen klar die Empfehlung, TU-lokale zentrale Hardware anzuschaffen und dort zu betreiben. Um jedoch Spitzenlasten flexibel abfedern zu können, empfiehlt sich die Anbindung der zentralen Ressource an externe Cloud-Ressourcen (Hybrid-Cloud). Da die Nutzung der Compute-Ressourcen in der Cloud insbesondere durch die interne Netzwerkleistung begrenzt sind und die Daten über die limitierte Außenanbindung der TU transferiert werden müssen, ist dieses Nutzungsszenario nur für serielle Jobs mit geringem Datenaufkommen geeignet. Einerseits kann ein Benutzer explizit Jobs in die Cloud schicken, anderseits kann das System automatisch bei Überlast Jobs verschieben, sofern der Benutzer sein Einverständnis gegeben hat. Dieses Einverständnis ist notwendig, weil einerseits die Datensicherheit bei Cloudnutzung im Gegensatz zur lokalen Nutzung nicht mehr gewährleistet ist und andererseits die Inanspruchnahme von Cloud-Ressourcen direkte zusätzliche Kosten verursacht.

V Fazit

Das WR an der TU Dortmund ist ein unverzichtbarer Bestandteil und Werkzeug der Forschungs- und Lehraktivitäten. Vor dem Hintergrund des Zeithorizonts der existierenden Installation, der Bedürfnisse der Dortmunder Nutzergruppen und im Kontext der HPC-Planungen der Nachbaruniversitäten ist die Weiterentwicklung des Standorts mit einer starken lokalen (= TU Dortmund-eigenen) Installation unabdingbar. Das beschriebene Betriebskonzept greift das schwierige Wechselspiel zwischen Kosteneffizienz und Bedarf auf und beschreibt insbesondere durch das dreistufige Modell mit Anreizbildung zum Übergang zur effizienteren Nutzung einen gangbaren Weg.

Auf längere Sicht sollen die Rechenressourcen zum größten Teil zentral organisiert werden (Modell „zentral/zentral“ bzw. zentral/lokal), die kontinuierlich durch die Fakultäten erweitert und durch Drittmittel ausgebaut werden sollen. „Lokal/lokal“-Ressourcen sollen nur in begründeten Ausnahmefällen vorgesehen werden; die Übernahme der Betriebs- und Investitionsfolgekosten werden in diesem Fall zwischen lokalem Betreiber und der Universitätsleitung ausgehandelt. DoWiR wird den WR-interessierten Arbeitsgruppen in der Planung und Organisation ihrer Ressourcen beratend vor dem Hintergrund dieses Betriebskonzept zur Seite stehen.

Um zentrale Ressourcen beschaffen zu können, ist die externe Förderung notwendig. Hierzu ist eine assoziierte Mitgliedschaft der TU Dortmund in der Gauß-Allianz – als Ergebnis eines im UAR Kontext schon länger verfolgten (Teil)-Ziels – anzustreben. Mit einer assoziierten Mitgliedschaft wird der Wille, aber auch die Expertise im WR zum Betrieb von HPC-Ressourcen dargelegt. Zur Realisierung werden eine Investitionssumme von ca. 3 Mio. EUR und jährliche Kosten für Strom und Kühlung von ca. 310.000 EUR erforderlich sein. Eine Realisierung als ausschließlich externe Cloud ist aus wirtschaftlichen und anwendungstechnischen Gründen nicht empfehlenswert.