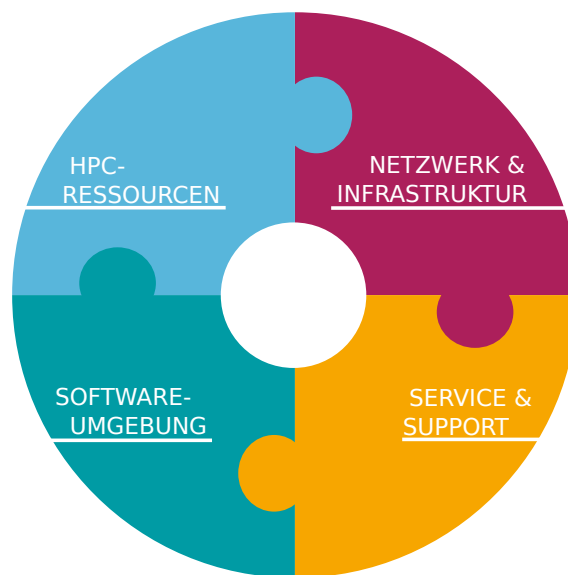


Auswertung der HPC.NRW Nutzendenbefragung

HPC.NRW*

2022



*Kontakt: support@hpc.nrw

Inhaltsverzeichnis

1	Einleitung	2
2	Überblick über die teilnehmenden Rechenzentren	4
3	Allgemeine Auswertung	6
4	Durchlässigkeit der Ebenen	9
5	Programmierung	11
6	HPC-Rechnungen	15
7	Performance-Analyse und Debugging	22
8	HPC-Basissoftware	26
9	Wissenschaftliche Software	38
10	HPC Ressourcen	40
11	Support	42
12	Schulungen	46
13	Erkenntnisse und Maßnahmen einzelner Standorte	51
14	Zusammenfassung	61

1 Einleitung

High-Performance Computing (HPC) wird als elementarer Bestandteil für Forschung und Entwicklung gesehen. Darum ist es wichtig, dass für die betreffenden Nutzendengruppen ausreichend Ressourcen bereitstehen und die Nutzung dieser Ressourcen optimal von technischer Seite betreut wird.

Im Rahmen des Projektes HPC.NRW wurde daher vom 07.07.2022 bis zum 30.09.2022 eine HPC-Nutzendenumfrage durchgeführt. Ziel der Umfrage war es herauszufinden, wo und im welchem Umfang Bedarf für HPC-Support, HPC-Software oder HPC-Anwenderschulungen in NRW besteht. Auf Basis der Ergebnisse sollen zukünftig Support, Software, Tools und Schulungen an die Bedürfnisse der Nutzenden angepasst und die Ergebnisdarstellung optimiert werden. Langfristig sollen dadurch Barrieren bezüglich der Nutzung von HPC-Ressourcen abgebaut und die Arbeit der Anwender effizienter gestaltet werden.

Initiator der Umfrage war das Kompetenznetzwerk HPC.NRW, ein Verbund aller Universitäten in Nordrhein-Westfalen. Hierzu zählen im Detail die drei Tier-2 Zentren Aachen, Köln und Paderborn und die neun Tier-3 Zentren Bielefeld, Bochum, Bonn, Dortmund, Duisburg-Essen, Düsseldorf, Münster, Siegen und Wuppertal. Zielgruppe der Umfrage waren alle HPC-Nutzenden in den Rechenzentren dieser Universitäten.

Der Fragenkatalog wurde in gemeinsamer Arbeit aller Standorte entwickelt. Jedes Rechenzentrum war für die Versendung der Umfrage an seine Nutzenden über interne Verteiler selbst verantwortlich. Durchgeführt wurde die Umfrage mit LimeSurvey, einer freien Online-Umfrage-Applikation, die es ermöglicht, anwenderfreundlich Online-Umfragen zu entwickeln, zu veröffentlichen und deren Ergebnisse datenbankseitig zu erfassen. Die LimeSurvey-Installation der Universität Paderborn wurde dafür genutzt.

Die Rohdaten der Umfrage stehen allen Partnern im Kompetenznetzwerk zur Verfügung, sodass fachbereichs-, standort- oder softwarespezifische Auswertungen und Korrelationsanalysen, die über die in diesem Dokument beschriebene Auswertung hinausgehen, durchgeführt werden können.

In der Auswertung werden auch standortspezifische Statistiken gezeigt, da hohe Teilnahmequoten aus einigen Standorten die Gesamtangaben sonst stark verzerren würden. Außerdem sollte beachtet werden, dass durch die teilweise sehr geringe Anzahl an teilnehmenden Nutzenden, insbesondere für die Standorte Bochum und Duisburg-Essen, deren Ergebnisse in Bezug auf die Repräsentativität hinterfragt werden müssen.

Um standortspezifische Auswertungen zu erzeugen, sind zwei verschiedene Wege denkbar, Teilnehmende der Umfrage einzelnen Standorten zuzuordnen: Eine Zuordnung basierend auf der Angabe zum physikalischen Standort oder eine Zuordnung basierend auf den Angaben zu genutzten HPC-Systemen. In dieser Auswertung wird der physikalische Standort gewählt, da es dabei sehr viel weniger Mehrfachnennungen gab als bei der Nutzung von HPC-Systemen.

Eine typische standortspezifische Auswertung als zweidimensionales Gitter ist in Abbildung 1 am Beispiel der Systemnutzung gezeigt. Dabei ist auf der horizontalen Achse der Standort des Umfrageteilnehmenden aufgetragen. Die Zahl in Klammern hinter dem Standort gibt die Anzahl an Teilnehmenden von diesem Standort an. Auf der vertikalen Achse sind die Antworten, hier die genutzten Typen von HPC-Systemen, aufgetragen. Jedes Feld enthält zum einen die prozentuale Anzahl wie oft diese Antwortmöglichkeit bezogen auf die Anzahl an Teilnehmenden von diesem Standort ausgewählt worden ist. Zum zweiten enthält ein Feld auch die absolute Anzahl an Antworten in Klammern. Da Mehrfachnennungen in Antworten sehr oft möglich waren, können die relativen Angaben in einer Spalte zu mehr als 100% aufsummieren. Die Spalte "Alle" repräsentiert die Summe über alle Standorte. In manchen Fällen ist die Summe abgegebener Stimmen geringer als bei anderen Fragen, was darauf zurückzuführen ist, dass die Frage nur jenen Nutzenden gestellt wurde, die zuvor auch eine bestimmte Antwort gegeben haben. So wurden die Nutzenden beispielsweise nicht nach der Qualität des Supports gefragt, wenn sie zuvor angegeben haben, dass dieser nie in Anspruch genommen wurde. Aus sprachlichen Gründen werden in dieser Auswertung an einigen Stellen Nutzende der HPC-Systeme und Teilnehmende der Nutzendenbefragung gleichgesetzt, was

jedoch keine direkte Repräsentativität der Nutzendenbefragung implizieren soll.

Die Daten der Auswertung wurden bereits von allen Standorten genutzt, um eigene Auswertungen durchzuführen und zukünftige Maßnahmen daraus abzuleiten. Diese wurden von den Standorten zusammengefasst und sind in Kapitel 13 dargestellt.

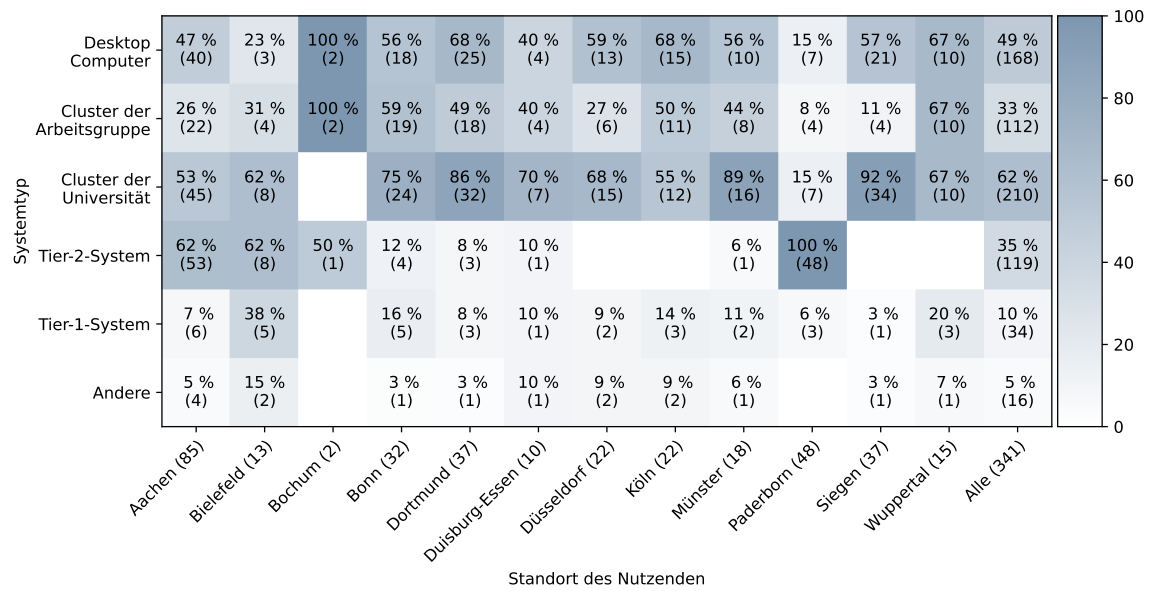


Abbildung 1: Beispiel für eine typische Ergebnisdarstellung im Dokument anhand der Nutzung von Cluster-Typen nach Standorten. Die Darstellung ist in Sektion 1 genauer beschrieben.

2 Überblick über die teilnehmenden Rechenzentren

Die folgende Tabelle liefert einen Überblick über die Rechenzentren (beginnend mit Tier-2), deren zugehörige Cluster mit den entsprechenden technischen Ausprägungen sowie derzeit aktive Nutzende auf den Systemen. Bochum verfügt derzeit noch nicht über ein lokales zentrales Cluster, weshalb hier noch keine Angaben gemacht werden können.

Standort	Systemname	Systemgröße	Spezielle Architekturen	Anzahl Nutzende
Tier-2				
Aachen	CLAIX	Claix-2016-MPI: ~600 Knoten Claix-2016-SMP: 8 Knoten Claix-2018-MPI: ~1250 Knoten	Claix-2016-GPU: 9 Knoten (je 2 Pascal) Claix-2018-GPU: 54 Knoten (je 2 Volta) SX-Aurora: 1 Knoten (8 VEs)	~8000
Köln	CHEOPS	~800 Knoten: 24x(32 Kerne/500 GB), 592x(12 Kerne/24-48 GB), 180x(8 Kerne/24-48 GB)	keine	200
Paderborn	Noctua 2	1121 Knoten (143488 CPU Kerne)	GPU: 32 Knoten (je 4 A100) FPGA: 16 Knoten (je 3 AMD Xilinx U280) FPGA: 16 Knoten (je 2 Intel Stratix 10 GX 2800)	630
Paderborn	Noctua 1	274 Knoten (10960 CPU Kerne)	GPU: 18 Knoten (je 1 NVIDIA RTX 2080 Ti)	749
Tier-3				
Bielefeld	-	28 Knoten (560 Kerne) mit 384 GB	8x NVIDIA Tesla V100 je Knoten	~30
Bochum	N/A			
Bonn	Bonna	70 Knoten (2240 Kerne)	-	~50
Dortmund	LiDO3	364 Knoten (14880 Kerne): 316x (40 Kerne/64 GB) 28x (48 Kerne/256 GB) 2x (48 Kerne/1024 GB) 18x (40 Kerne/64 GB, 2x K40)	18 Knoten mit je 2 K40 28 mit 48 Kernen/256 GB 2 mit 48 Kernen/1024 GB	~100
Duisburg-Essen	magnitUDE	492 (64GB)	72 FatNodes (128 GB) 60 SuperFatNodes (256 GB)	~75
Düsseldorf	Hilbert	~ 220 Knoten (24/32 Kerne, 192/256 GB)	5 FatNodes (1.5/3 TB) 7 A100 Knoten (8 GPUs, 64/128 Kerne, 1/2 TB) 20 GPU Knoten (10 GPUs, 20/48 Kerne, 256 GB)	329
Münster	PALMA	514 Knoten	8 TitanXP, 4 V100, 48 2080 Ti, 4 TitanRTX, 16 RTX 3090, 20 A100-40G, 16 A100-80G	153

Siegen	OMNI	434 Knoten	10 GPU-Knoten, 5 Kubernetes-Knoten, 2 SMP-Knoten, 8 Openstack-Knoten	~300
Wuppertal	PLEIADES	268 Knoten (17152 Kerne)	5 Knoten mit je 8xA100, 128 Kernen und 2TB RAM	~180

3 Allgemeine Auswertung

An der Umfrage haben insgesamt 341 Personen teilgenommen. Diese ordneten sich folgenden Standorten zu: Aachen (85), Bielefeld (13), Bochum (2), Bonn (32), Dortmund (37), Duisburg-Essen (10), Düsseldorf (22), Köln (22), Münster (18), Paderborn (48), Siegen (37), Wuppertal (15). Damit hat sich nach Einschätzung der Rechenzentren nur ein geringer bis mittlerer Anteil der aktiven Nutzenden an der Umfrage beteiligt (<20%). Insgesamt sinkt die Beteiligung gegenüber der Nutzendenbefragung 2020 um knapp 22%. Der Grund für den Rückgang lässt sich aufgrund der anonymisierten Erfassung nicht benennen, es sind aber mehrere Faktoren denkbar:

- Nutzende aus der Umfrage 2020 wollten nicht wieder eine Umfrage ausfüllen
- Informationen zur Umfrage haben weniger Nutzende erreicht
- Geringere Anzahl an Nutzenden, die genügend Umgang mit den HPC-Systemen haben, um gewillt zu sein, die Umfrage auszufüllen
- Insgesamt weniger Nutzende
- Umfragezeitraum lag 2022 zum Großteil in den Semesterferien

Die Befragten arbeiten, wie in Abbildung 2 zu erkennen, meist auf dem lokal verfügbaren HPC-System und sind nur sehr vereinzelt auch auf Systemen anderer Standorte aktiv. Knapp die Hälfte nutzt für ihre Arbeit die Workstation, 33% ein eigenes Cluster innerhalb der Arbeitsgruppe, 62% das lokale Universitätscluster und 35% ein Tier-2, bzw. 10% ein Tier-1 System. Bochum verfügt derzeit noch nicht über ein lokales zentrales Cluster. Gegenüber der vorherigen Befragung ist in Bonn eine Ausdehnung der wissenschaftlichen Rechnungen von überwiegend auf dezentralen Systemen hin zum Cluster der Universität (52 % zu 75%) festzustellen. Weiter fällt auf, dass in Paderborn 100% der Befragten angaben, ein Tier-2-System zu nutzen und dass die Nutzung dezentraler Systeme gegenüber 2020 stark gefallen ist. Dies könnte sowohl auf die Inbetriebnahme des neuen Clusters zurückzuführen sein, als auch auf die HPC-Lehrmaßnahmen von HPC.NRW, da es sich bei dem System in Paderborn um ein Tier-2-System handelt.

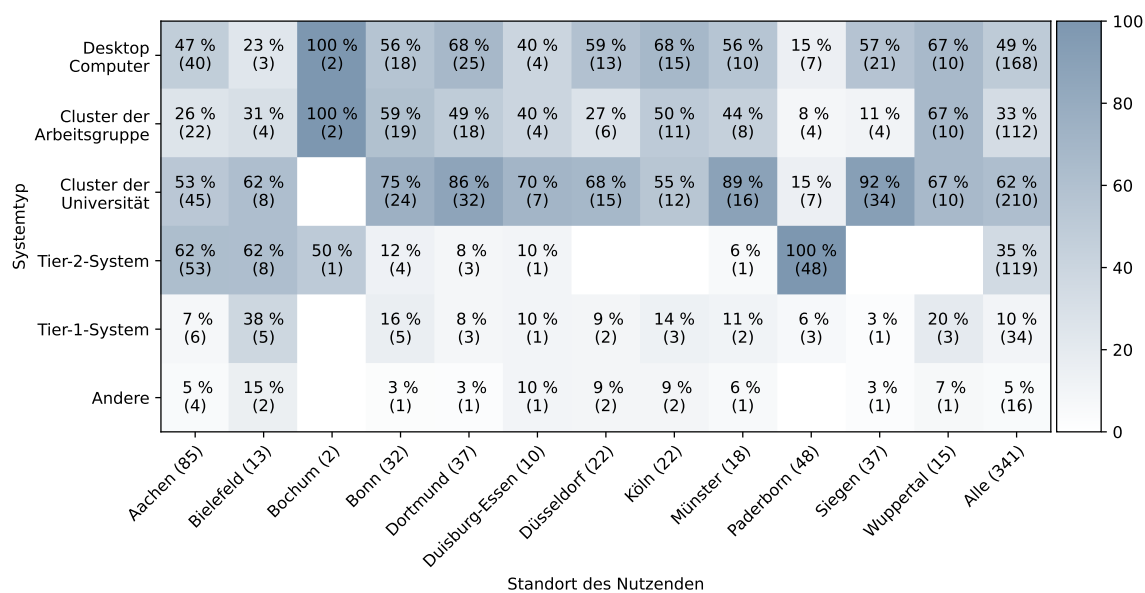


Abbildung 2: Nutzung von Cluster-Typen nach Standorten. Die Darstellung ist in Sektion 1 genauer beschrieben.

Der akademische Grad der Befragten war, wie in Abbildung 3 gezeigt, zur Hälfte Master/Diplom, ein Viertel Dr./PhD. und jeweils knapp 10% Bachelor und Prof.. Überdurchschnittlich hoch, mit teilweise über 50%, ist der Anteil von Befragten mit Dr./PhD in Bielefeld und Köln. Hingegen rechnen in Düsseldorf und Paderborn überdurchschnittlich viele Studierende mit Bachelorabschluss. Ausgedehnt auf das gesamte HPC.NRW-Netzwerk kann man jedoch sagen, dass sich in den letzten zwei Jahren die Landschaft der Nutzenden im Bezug auf akademischen Titel kaum verändert hat.

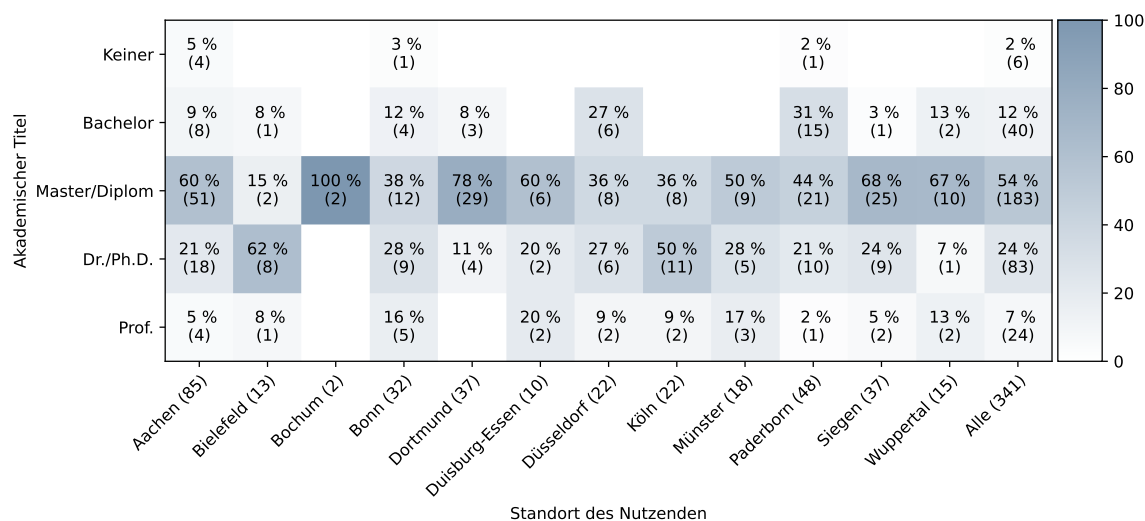


Abbildung 3: Verteilung der akademischen Grade unter den Befragten nach Standorten. Die Darstellung ist in Sektion 1 genauer beschrieben.

In den primären Forschungsgebieten waren die Fachbereiche Physik (23%) und Informatik (23%), System- und Elektrotechnik (23%) führend, gefolgt von Chemie (12%) und Biologie (10%). In geringerem Umfang waren folgende Fachbereiche vertreten: Maschinenbau und Wirtschaftsingenieurwesen (7%), Mathematik (6%), Materialwissenschaft und Werkstofftechnik (5%), Geowissenschaften (4%), Medizin (3%), Sozial- und Verhaltenswissenschaften (3%), Bauingenieurwesen und Architektur (2%), Wärme- und Verfahrenstechnik (2%), Geisteswissenschaften (1%), Landwirtschaft, Forstwirtschaft und Veterinärmedizin (<1%). Die Verteilung der Fachbereiche ist in Abbildung 4 gezeigt.

Beeindruckend ist die Relevanz von HPC in der Forschung. Rund 79% aller Nutzenden gaben an, dass HPC von größter Wichtigkeit für ihre Arbeit ist und weitere 20% erwähnten, dass HPC nicht erforderlich ist, aber die Arbeit in hohem Maße unterstützt. Lediglich 2% aller Befragten haben den Eindruck, dass HPC nicht wirklich relevant ist, allerdings ein Trend dorthin zu verzeichnen sei. Keiner der Befragten gab an, dass HPC keine Rolle in der Forschung spielt.

In der Lehre ist das Thema allerdings noch nicht sehr verbreitet. Jeweils 26% aller Befragten gaben an, dass das Thema HPC keine Rolle spielt, da es ein sehr spezielles Thema ist oder dass es noch keine Rolle spielt, aber ein Trend dorthin zu verzeichnen sei. Etwas mehr, mit 28%, geben an, dass HPC eine untergeordnete Rolle spielt und bislang nur in speziellen Veranstaltungen vermittelt wird. Der Anteil der Befragten, die HPC als fest im Lehrplan integriert angaben, stieg gegenüber 2020 um 8 Prozentpunkte auf 20 % an.

Die Diskrepanz zwischen HPC-Bedarf in der Forschung und HPC-Integration in der Lehre zeigt auf, dass ein großes Potential für das Lehrangebot vorliegt, welches ausgeschöpft werden sollte, um den Ansprüchen der Forschung zu genügen.

Knapp die Hälfte der Befragten (48%), haben innerhalb der letzten 18 Monate bereits einen Beitrag in einem Journal, einer Konferenz oder einem Workshop veröffentlicht, der auf Ergebnissen

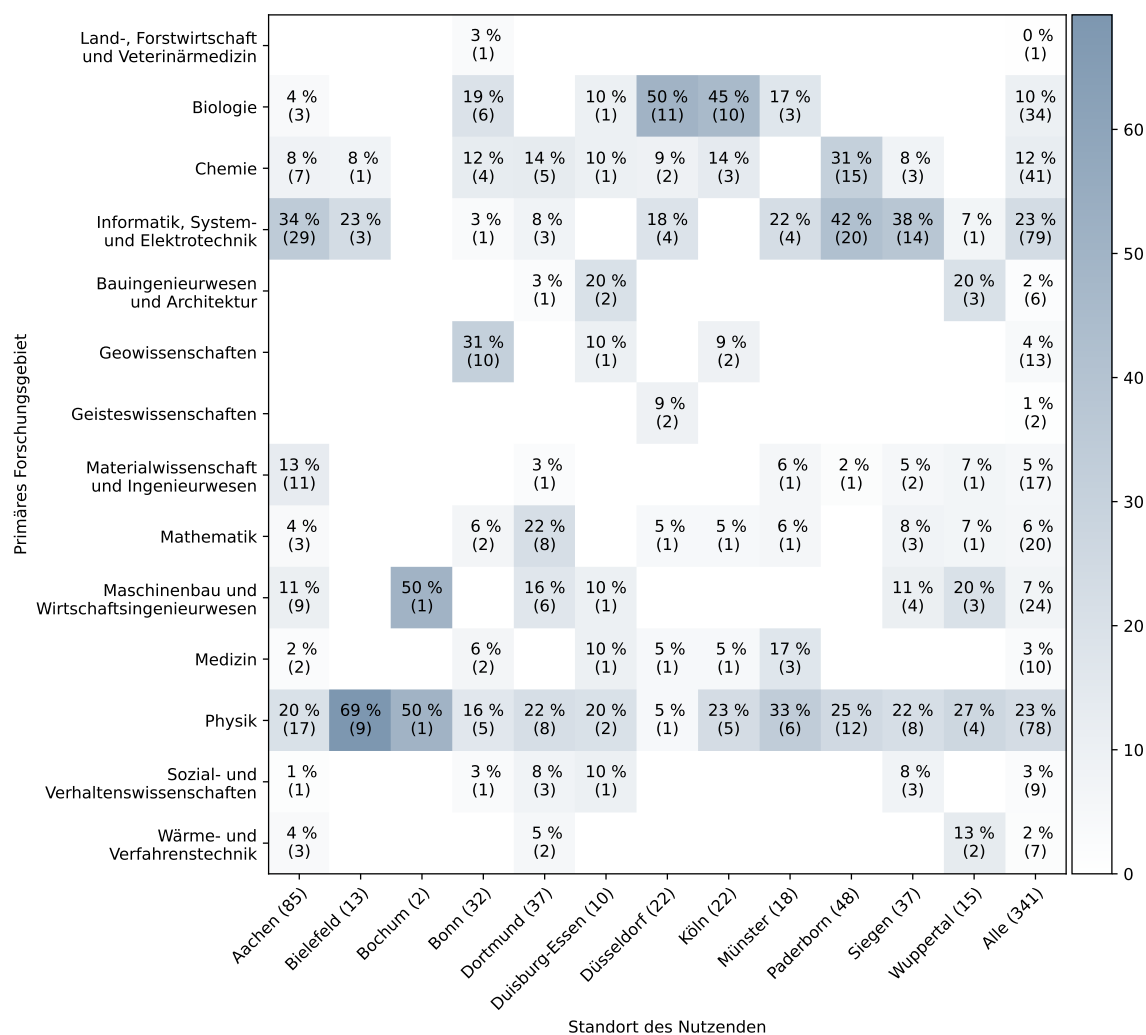


Abbildung 4: Verteilung der Fachbereiche pro Standort. Die Darstellung ist in Sektion 1 genauer beschrieben.

basiert, welche mithilfe eines Clusters erzielt wurden. 68% der Befragten gaben an, dass eine solche Veröffentlichung derzeit in Vorbereitung ist oder sich bereits in einem Review-Prozess befindet.

4 Durchlässigkeit der Ebenen

Eine weitere wichtige Fragestellung von HPC.NRW beschäftigt sich mit der Durchlässigkeit zwischen den Ebenen. Ziel hierbei ist es, Nutzenden mit erhöhtem Rechenbedarf Hilfestellungen bei Projektanträgen auf höheren Tier-Ebenen zu geben. Abbildung 5 zeigt die Nutzenden-Mobilität zwischen den einzelnen Zentren.

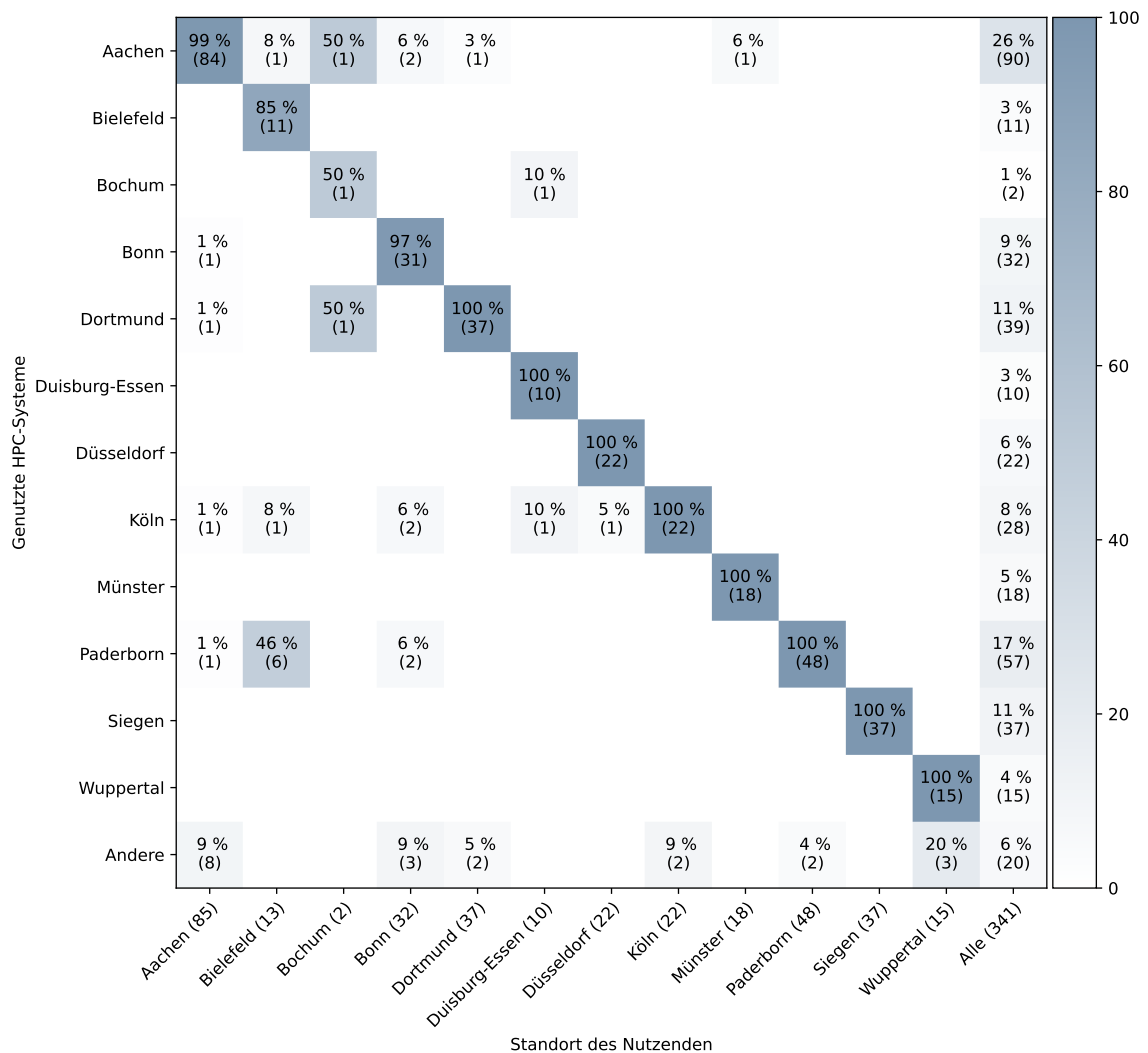


Abbildung 5: User-Mobilität zwischen den HPC-Zentren innerhalb und außerhalb von HPC.NRW. Die Darstellung ist in Sektion 1 genauer beschrieben.

Tier-3 zu Tier-2 Am Beispiel des Tier-3-Standorts Bielefeld sieht man, dass 85% aller Nutzenden (11 von 13) zwar die lokalen Ressourcen verwenden, allerdings werden von 46% auch das System des Tier-2-Standorts Paderborn und von je 8% die Systeme in Köln und Aachen verwendet.

Ähnlich verhält es sich in Bonn, wo 97% der Nutzenden zwar die lokalen Ressourcen nutzen, aber jeweils 6% die Tier-2 Standorte Aachen, Köln und Paderborn verwenden. Außerdem nutzen in Bonn 9% Angebote außerhalb des HPC.NRW-Konsortiums, was leicht oberhalb des netzwerkweiten Wertes von 6% liegt.

Obwohl einzelne Nutzende auch auf die Ressourcen der Tier-3-Zentren in Dortmund, Bonn und Bochum ausweichen, ist klar erkennbar, dass bevorzugt Zentren der Ebene 2 als Alternative

verwendet werden, so wie es auch in der HPC-Landesstrategie vorgesehen ist. Gegenüber der Umfrage von 2020 ist ein deutlicher Rückgang des Standortes “Andere” zu verzeichnen, welcher auch Standorte der Ebene 1 enthält. Insgesamt ist die Mobilität zwischen den Ebenen 2 und 3 relativ gering. Ob dies an den ausreichend vorhandenen Tier-3 Ressourcen oder der höheren Einstiegshürde für die Tier-2-Ebene liegt, geht aus der Umfrage nicht hervor. Im Rahmen von HPC.NRW wird versucht, letzteres durch gezielte Maßnahmen wie Intensivberatungen, vereinfachte Projektkategorien, einheitliche Beantragungs-/Begutachtungssysteme, Tutorials und verbesserte Dokumentation zu adressieren.

Tier-2 zu Tier-1 Insgesamt wurden in der Umfrage von Nutzenden, die HPC.NRW-Standorten zuzuordnen sind, 10 mal die Nutzung von Zentren auch außerhalb HPC.NRW genannt (Freitextangaben in Kategorie “Andere”). Bei der Analyse zeigt sich hier ein besonderer Fokus auf die Tier-1-Zentren. Hierbei werden die Tier-1-Systeme des Forschungszentrums Jülich besonders häufig genannt (64% von “Andere”). Ein Grund hierfür ist sicher die gemeinsame Antragsschiene der RWTH Aachen University mit dem Forschungszentrum Jülich und die geographische Nähe innerhalb NRW. Es zeigt sich aber auch, dass Forschenden der Standorte Bonn, Dortmund, Köln, Paderborn und Wuppertal ebenfalls Rechenzeit in Jülich bewilligt worden ist.

Diese Auswertung zeigt, dass eine Durchlässigkeit der Ebenen grundsätzlich gegeben ist, jedoch noch verbessert werden kann.

5 Programmierung

Auf die Frage, ob die Befragten ihre Anwendungen selbst bauen/kompilieren, antworteten 60% mit ja und 40% mit nein wie in Abbildung 6 gezeigt.

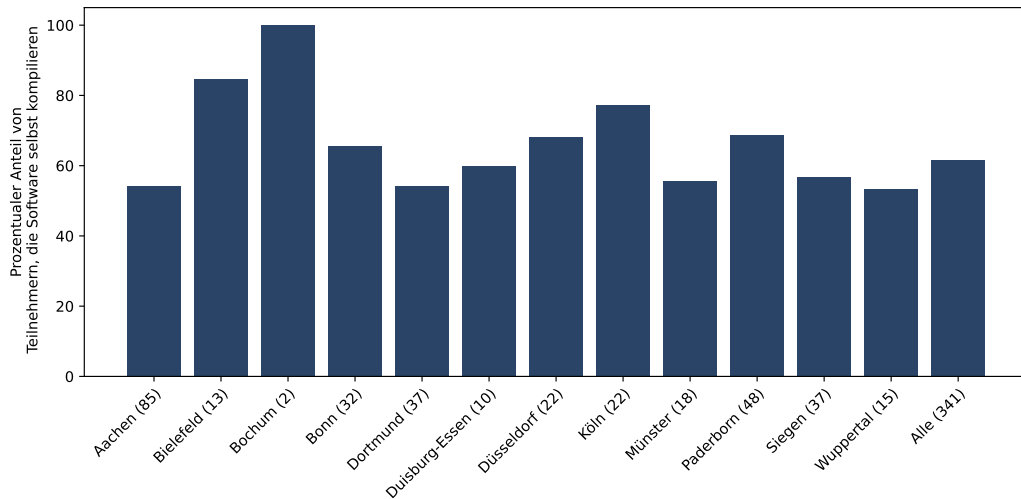


Abbildung 6: Prozentualer Anteil der Nutzenden, die Programme zur Nutzung auf HPC-Clustern selbst kompilieren.

In die Entwicklung von Programmen sind 66% aller Befragten direkt involviert und die Verteilung nach Standorten ist in Abbildung 7 zu sehen. Weitere 11% würden gerne oder werden demnächst selbst Entwicklungstätigkeiten übernehmen. Nur 14% gaben an, ausschließlich bestehende Programme zu verwenden, bei 17% findet die Entwicklung der Software in derselben Arbeitsgruppe statt. Ein besonders hoher Anteil an Eigenentwicklung liegt in Bielefeld (10/13 der Befragten), Bonn (24/32), Siegen (29/37) und Wuppertal (12/15) vor.

Zusätzlich zu typischen Entwicklungsaufgaben, wie die Erweiterung der Programmfunktionalität (55%), Erstellen neuer Programme (53%), Fehlerbehebung/Bug fixing (50%) und Performanceoptimierung (42%), siehe Abb. 7, führen die Befragten, die selbst Programme entwickeln, auch vereinzelt folgende Aufgaben aus: Parallelisierung, Software-Paketierung, Erstellen von Skripten und Wrappern zur Abbildung von Arbeitsprozessen und Portierung.

Das Gesamtbild zur Entwicklung von Programmen ist seit 2020 größtenteils unverändert geblieben. Leichte Rückgänge sind für Bonn, Duisburg-Essen, Köln und Paderborn zu verzeichnen, wobei wie anfangs angemerkt unklar ist, ob diesmal die gleichen Nutzenden die Umfrage ausgefüllt haben.

Programmiersprachenkenntnisse sind in Abbildung 8 ausgewertet und am weitesten verbreitet sind Python (72%), C/C++ (63%), Matlab (47%) und Java (40%). Bis auf Python überwiegen bei diesen die grundlegenden Kenntnisse. Als weitere vorgegebene Antworten wurden Fortran mit 27%, R (31%), Mathematica (22%) und Julia (16%) angegeben. In diesen Sprachen sind bei den Befragten überwiegend Basiskenntnisse vorhanden. In den Freitextantworten stach die Verwendung von Bash-Skripten am stärksten hervor, hielt sich aber mit knapp 4% in Grenzen. Die Befragten konnten sich auch dazu äußern, welche Grundlagenkurse (Abbildung 9, oben) und Fortgeschrittenenkurse (Abbildung 9, unten) sie zu programmierrelevanten Themen interessieren würden. Bei den Grundlagen wurden am häufigsten MPI (38%), CUDA (35%) und C/C++ (30%) genannt. Bei dem Wunsch nach fortgeschrittenen Kursen liegen Python mit 51% und C/C++ mit 39% vorne, gefolgt von MPI und CUDA mit jeweils 31%.

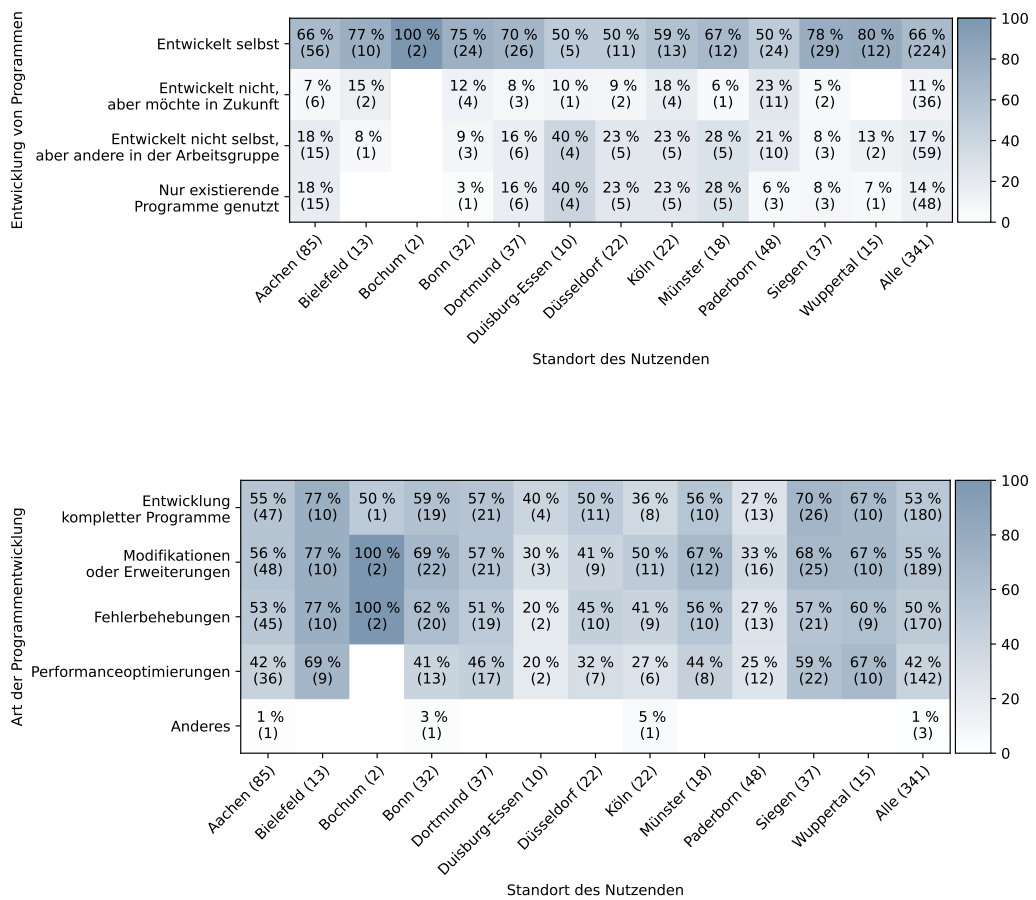


Abbildung 7: Verteilung der Antworten zur Entwicklung von Programmen (oben) und Art der Entwicklung (unten) nach Standorten. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

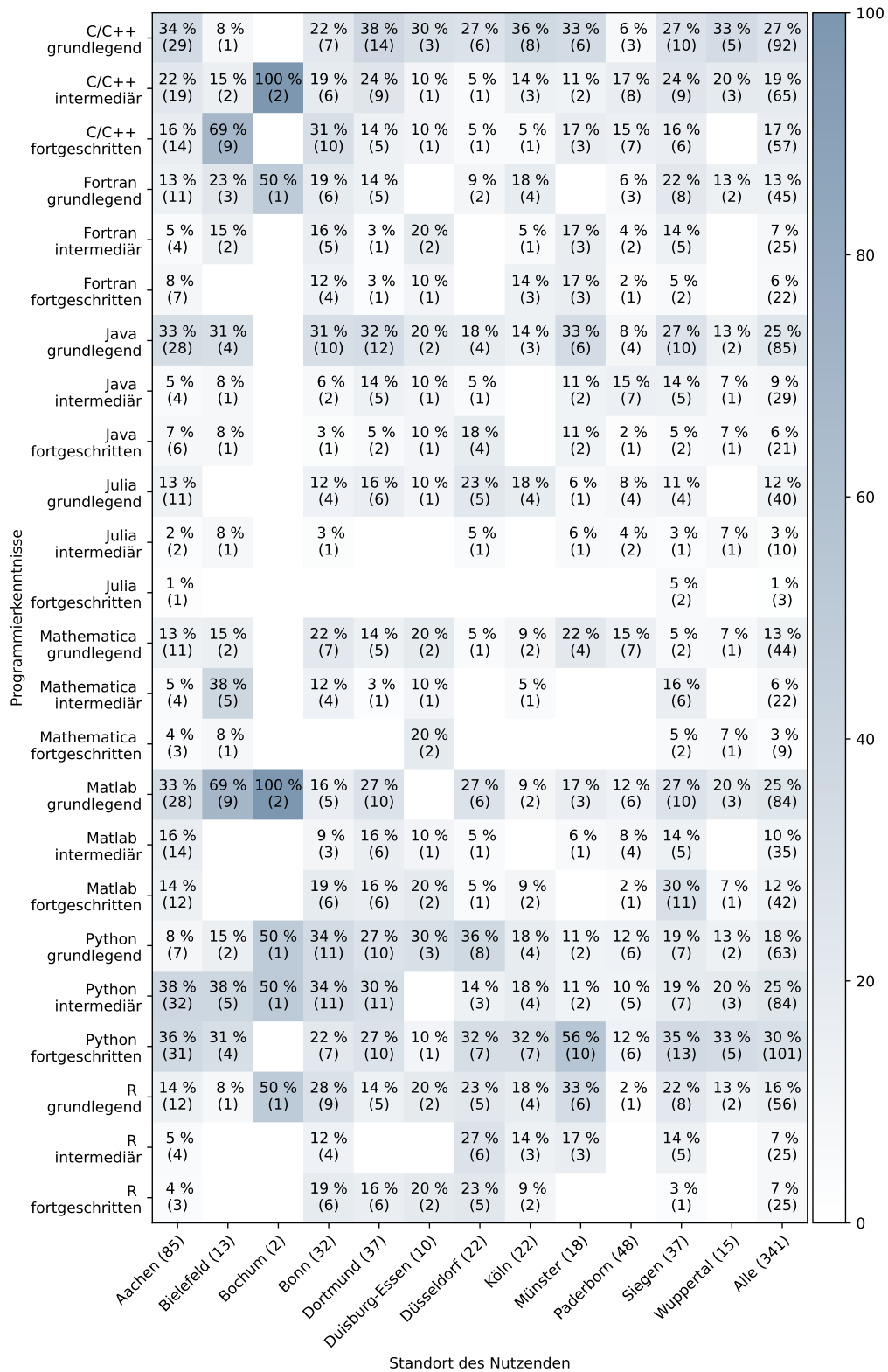


Abbildung 8: Verteilung der Programmiersprachenkenntnisse nach Standorten. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

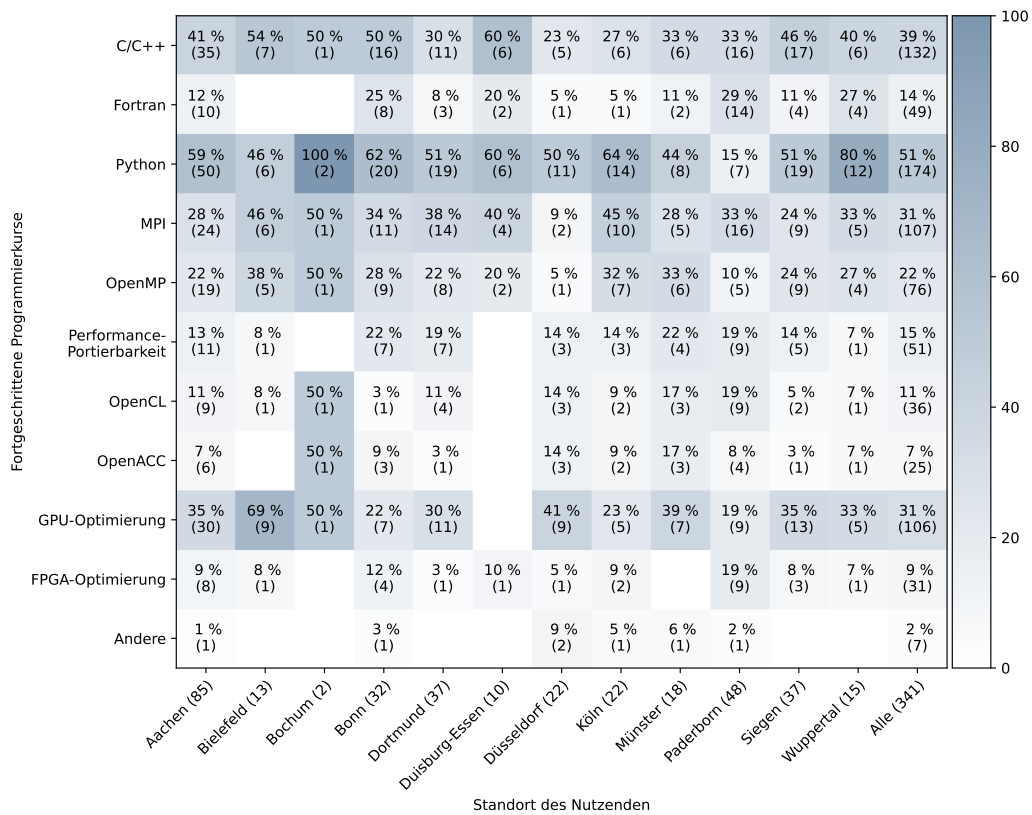
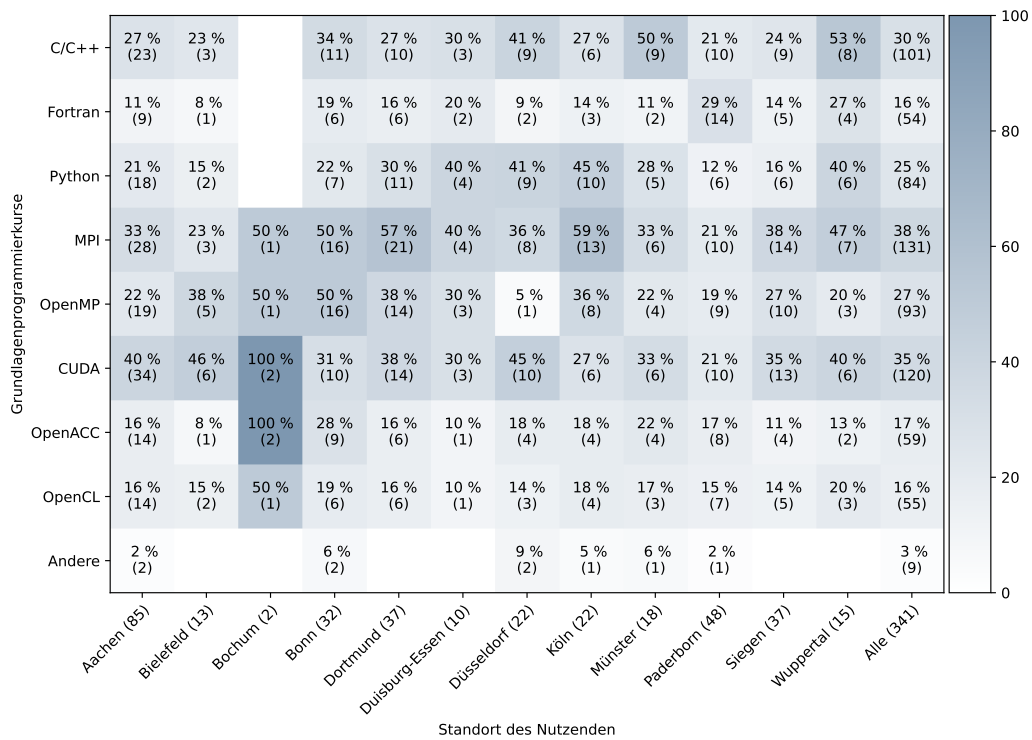


Abbildung 9: Verteilung der gewünschten Grundlagenkurse (oben) und Fortgeschrittenenurse (unten) mit Programmierbezug nach Standorten. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

6 HPC-Rechnungen

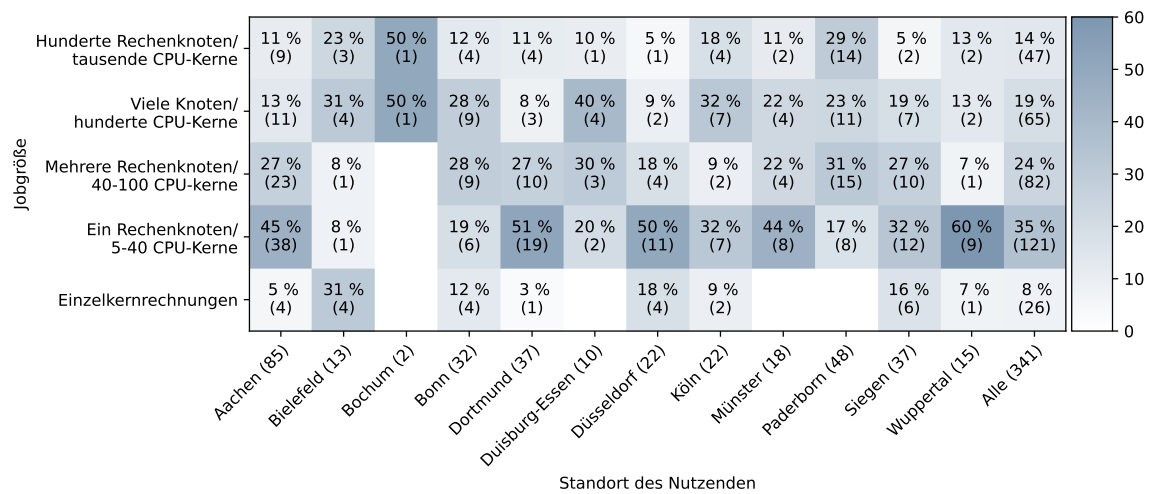


Abbildung 10: Jobgrößen der Befragten nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

Ein geringer Anteil an Befragten (8%) führt weiterhin hauptsächlich serielle Berechnungen aus wie in [Abbildung 10](#) zu erkennen ist. Gegenüber 2020 stellt dies einen Rückgang um 5 Prozentpunkte dar. Parallele Rechnungen im kleinen Stil (5-40 CPU-Kerne oder alle Kerne eines Knotens) führen 35% der Befragten regelmäßig durch. Im unteren mittleren Bereich (mehrere Knoten oder 40-100 Kerne) bewegen sich 24% und im oberen mittleren Bereich (<viele Knoten oder hunderte CPU-Kerne) 19%. Überwiegend auf größeren Systemen (hunderte Knoten oder tausende Kerne) bewegen sich 14%. Hervorzuheben sind Aachen, Dortmund, Düsseldorf und Wuppertal, die bei den Rechnungen mit nur einem Rechenknoten oder 5-40 Kernen deutlich oberhalb des netzwerkweiten Wertes liegen. In Bielefeld fällt auf, dass entweder nur mit einzelnen Kernen oder aber mit sehr vielen gerechnet wird. Außerdem liegen die beiden Tier-2 Standorte Köln und Paderborn in den beiden größten Jobkategorien oberhalb des Durchschnitts.

Zu Zwecken der Parallelisierung wird von 48% der Befragten MPI, 38% OpenMP, 24% CUDA, 6% PThreads, 4% OpenCL, 2% OpenACC und 1% Intel TBB eingesetzt. Keine Parallelisierung zu nutzen gaben 7% an, während 17% keine Angaben dazu machen konnten. Das Fehlen einer Parallelisierung ist hier in guter Übereinstimmung mit den Nutzenden, die nur auf einem Kern rechnen. Eine standortspezifische Auswertung ist in [Abbildung 11](#) gezeigt. Besonders häufig wurde in Düsseldorf, Duisburg-Essen und Wuppertal angegeben, dass man nicht wisse, welche Parallelisierungstechnik zum Einsatz kommt. Das Gesamtbild betrachtend lässt sich jedoch kein signifikanter Unterschied zu den Umfrageergebnissen 2020 feststellen.

Bei der Anzahl der maximal verwendeten CPU-Kerne pro Job ist wie in [Abbildung 12](#) zu sehen keine klare Verteilung ersichtlich. Sie liegt bei 27% der Befragten im Bereich zwischen 25-48, gefolgt von 25% mit mittleren und kleineren Jobgrößen von 49-512 bzw. 2-24 Kernen, 9% großen Allokationen mit >1024 Kernen und 6% im Bereich zwischen 512-1024 Kernen. Serielle Jobs werden ausschließlich von 5% der Nutzenden ausgeführt. Weitere 3% konnten hierzu keine Angaben machen. Bis auf die üblichen Schwankungen ist auch hier kein neuer Trend gegenüber 2020 zu erkennen.

Bei den Angaben zum Hauptspeicher pro CPU-Kern in [Abbildung 13](#) ergibt sich dagegen eine sehr klarer Trend zu mittleren bis größeren Speicherkapazitäten. So benötigen 30% der Befragten 4-16 GB, 21% zwischen 2-4 GB, 18% schon 16-40 GB. Werte darunter als auch darüber werden von

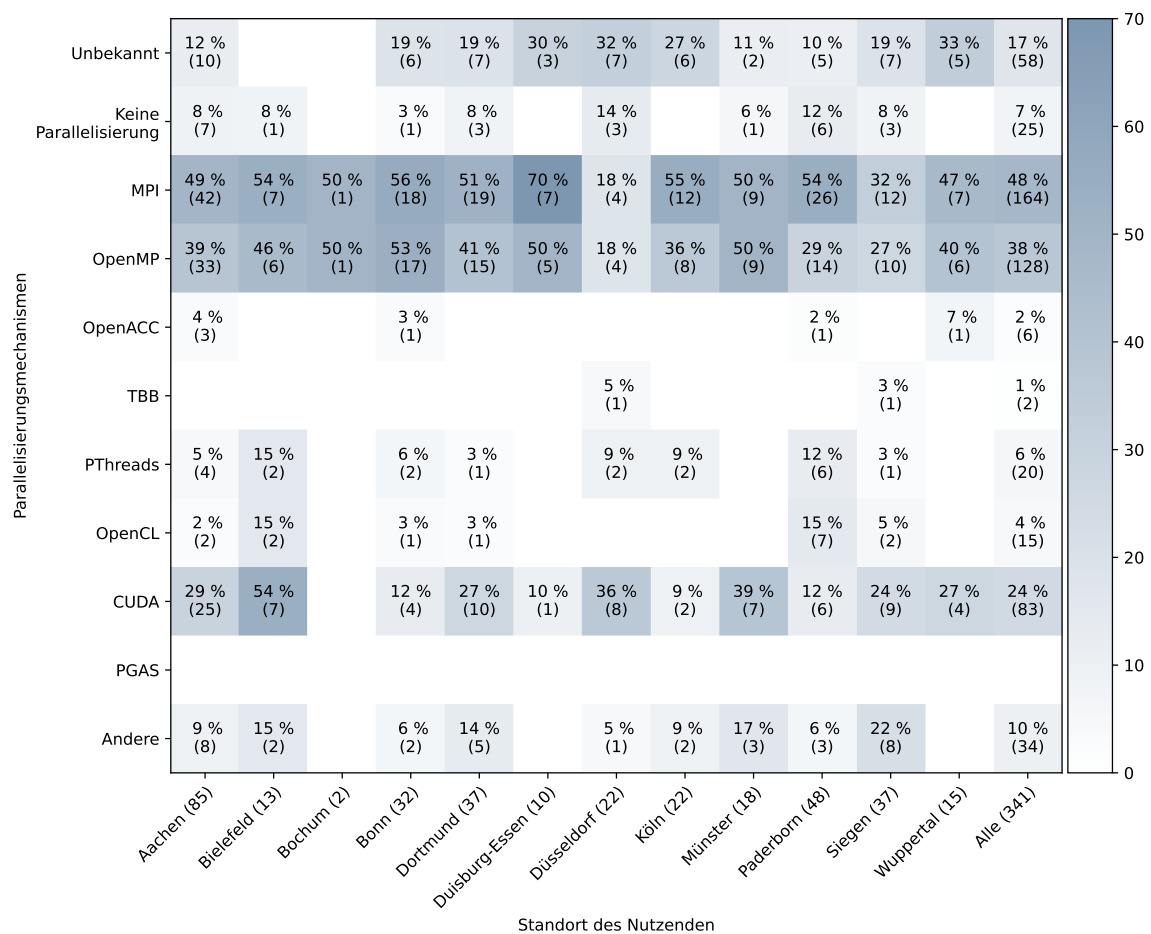


Abbildung 11: Nutzung von Parallelisierungsmechanismen der Befragten nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

jeweils knapp 10% der Nutzenden angegeben. Keine Angaben dazu kamen von 9% der Befragten.

Der Bedarf an Festplattenspeicher pro Job ([Abb. 14](#)) liegt über alle Standorte betrachtet hauptsächlich in der mittleren Region mit 5-100 GB (49%), gefolgt von einem sehr geringen Bedarf unterhalb von 5 GB (27%). Der Anteil an Jobs oberhalb von 100 GB beträgt weniger als 1/4 aller Rechnungen. Dieses Bild ergibt sich auch größtenteils für die einzelnen Standorte, jedoch haben Düsseldorf (32%), Münster (28%), Paderborn (25%) und Bonn (22%) einen überdurchschnittlich hohen Anteil an Jobs mit höherem Festplattenspeicherbedarf.

Nicht alle Systeme im HPC.NRW-Verbund verfügen über ein systematisches Job-Monitoring. Um einen Einblick zu bekommen, welche Anforderungen die Anwendungen an das System stellen, wurden auch Fragen zu typischen Anwendungscharakteristiken gestellt. Diese sind in [Abbildung 15](#) ausgewertet.

Konkret gaben die Befragten als Bottleneck am häufigsten die Größe des Hauptspeichers (43%) an. Dies ist ein deutlicher Anstieg, 12 Prozentpunkte, gegenüber 2020. Dass dieses Bottleneck so häufig genannt wird, ist nicht verwunderlich, da das die einzige Systemgröße ist, die von den Nutzenden durch Programmabbrüche direkt erkennbar ist. Andere Systemgrößen sind lediglich aus Performance-Sicht relevant, behindern aber nicht die Ausführung der Anwendungen. Von daher stellt die Größe des Hauptspeichers ein essenzielles Problem dar.

Danach folgen Hauptspeicherbandbreite (21%), Dateizugriffe (17%) und Gleitkommazahlendurchsatz (14%). Die Cache-/Hauptspeicherlatenz sehen nur 9% als ein Problem, Netzwerklatenz

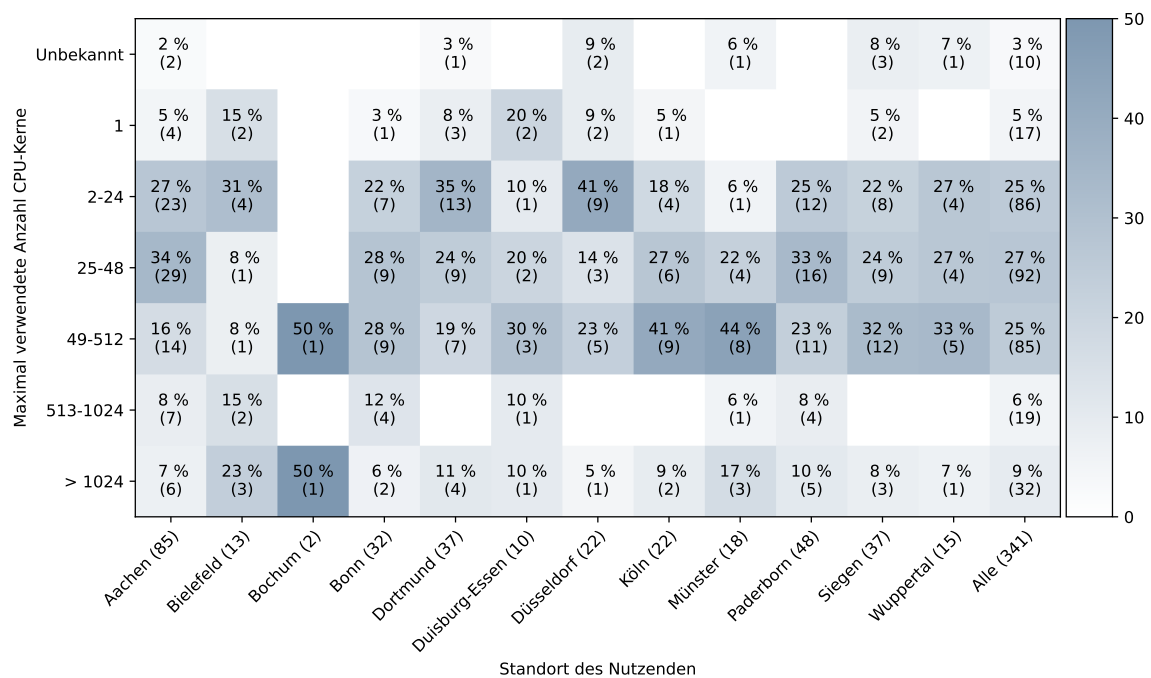


Abbildung 12: Maximal genutzte Anzahl CPU-Kerne pro Job der Befragten nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

und -bandbreite wirken sich für 11% bzw. 8% der Befragten auf das Verhalten ihrer Anwendungen aus.

Bemerkenswert hoch ist die Unkenntnis der Nutzenden über das parallele Dateisystem. Fast 51% gaben an, nicht zu wissen, welche parallelen Dateisysteme auf den genutzten HPC Systemen im Einsatz sind. Diese Wissenslücke hat in der Praxis aber kaum Auswirkungen, da über generische Bibliotheken darauf zugegriffen wird und ein Dateisystem-spezifisches Tuning erst ab Tier2/Tier1 Ebene durchgeführt wird. Zudem sind gerade an den kleineren Zentren oft mehrere Systeme im Einsatz und die Dokumentationen darüber mangelhaft.

Die Verbreitung von Beschleunigern (GPUs, FPGAs usw.) hat in den letzten Jahren stark zugenommen. Von den Befragten setzen indes 26% Nvidia Data Center GPUs und 25% Nvidia GTX/RTX GPUs routinemäßig ein wie in [Abbildung 16](#) zu sehen ist. Gegenüber den Umfrageergebnissen von 2020 stellt dies in beiden Fällen einen Anstieg von über 10 Prozentpunkten dar. In Bielefeld und Paderborn ist der Anteil an GPU-Nutzenden unter den Befragten mit >50% besonders hoch. Gemeinsam mit der Tatsache, dass es mittlerweile keinen Standort mehr gibt, an dem keine Beschleuniger verwendet werden (wie es 2020 noch mehrfach der Fall war), ist ein klarer Trend zu erkennen.

Auch bei den Multibeschleunigern (z.B. mehrere GPUs mit P2P, Nvidia NVLINK, FPGAs mit direktem Interconnect) ist das Interesse an der Technologie ausgeprägt ([Abb. 17](#)). Eingesetzt werden Multibeschleuniger bereits bei 9% der Befragten. Vorreiter sind hier Bielefeld und Paderborn mit 23% und 25%. NRW-weit arbeiten zudem 5% der Befragten an der Entwicklung von Multibeschleuniger-fähigen Programmen und weitere 17% würden diese Technologie gerne einsetzen. Natürlich sind nicht alle Anwendungen hierfür geeignet und so antworteten 24%, dass der Einsatz von Multibeschleunigern bei ihnen keinen Sinn machen würde. Etwa die Hälfte der Befragten ist mit dieser Technologie noch nicht vertraut.

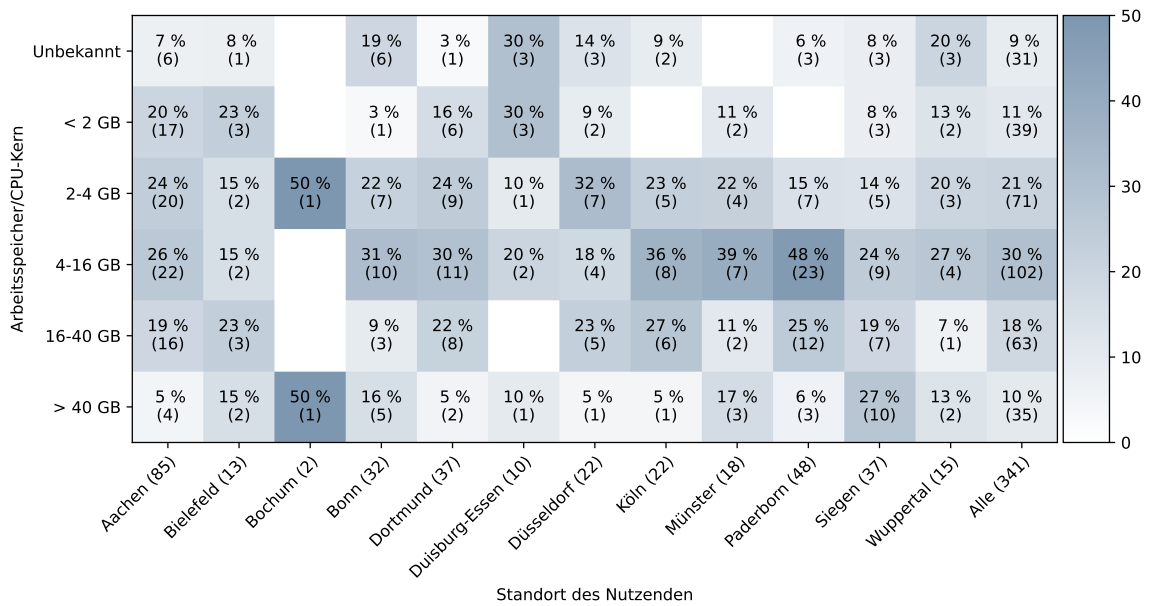


Abbildung 13: Benötigter Arbeitsspeicher je CPU-Kern nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

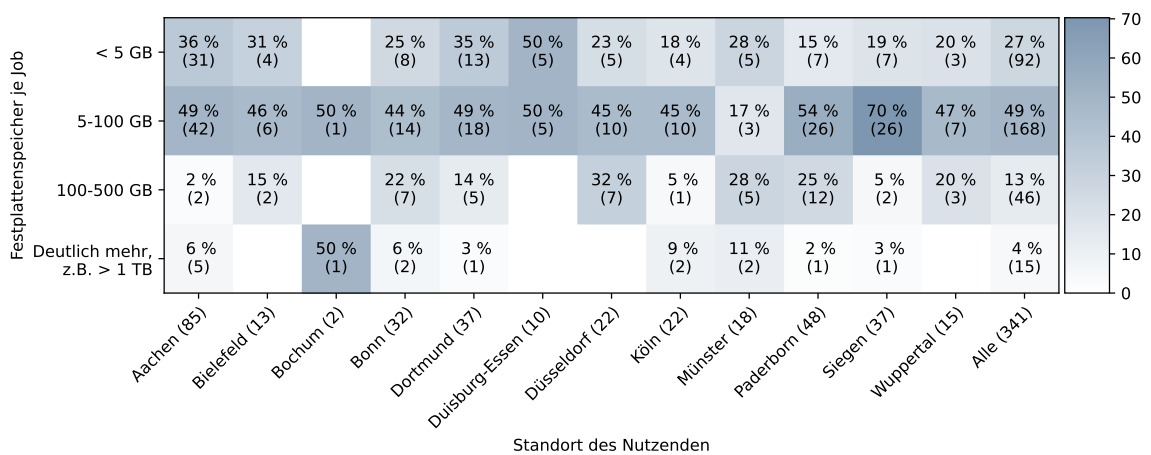


Abbildung 14: Benötigter Festplattenspeicher je Job nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

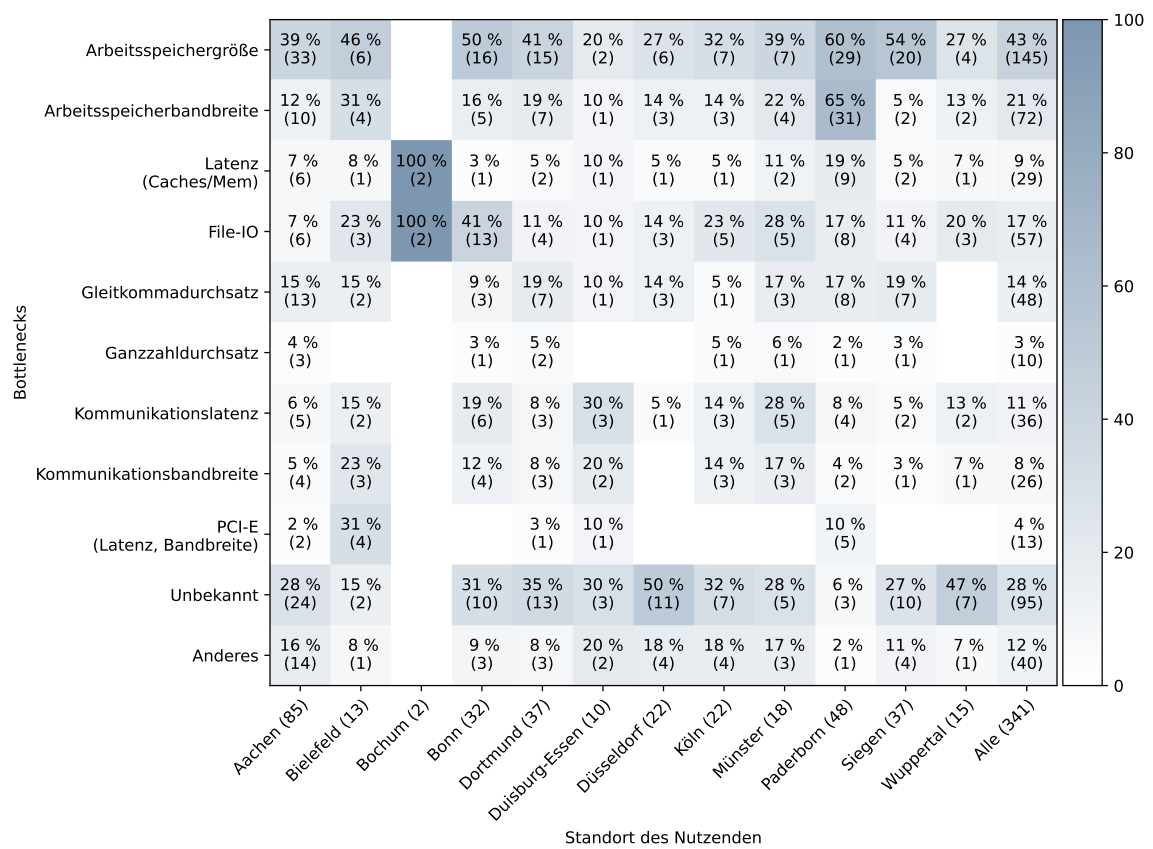


Abbildung 15: Bottlenecks der Befragten nach Standort. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

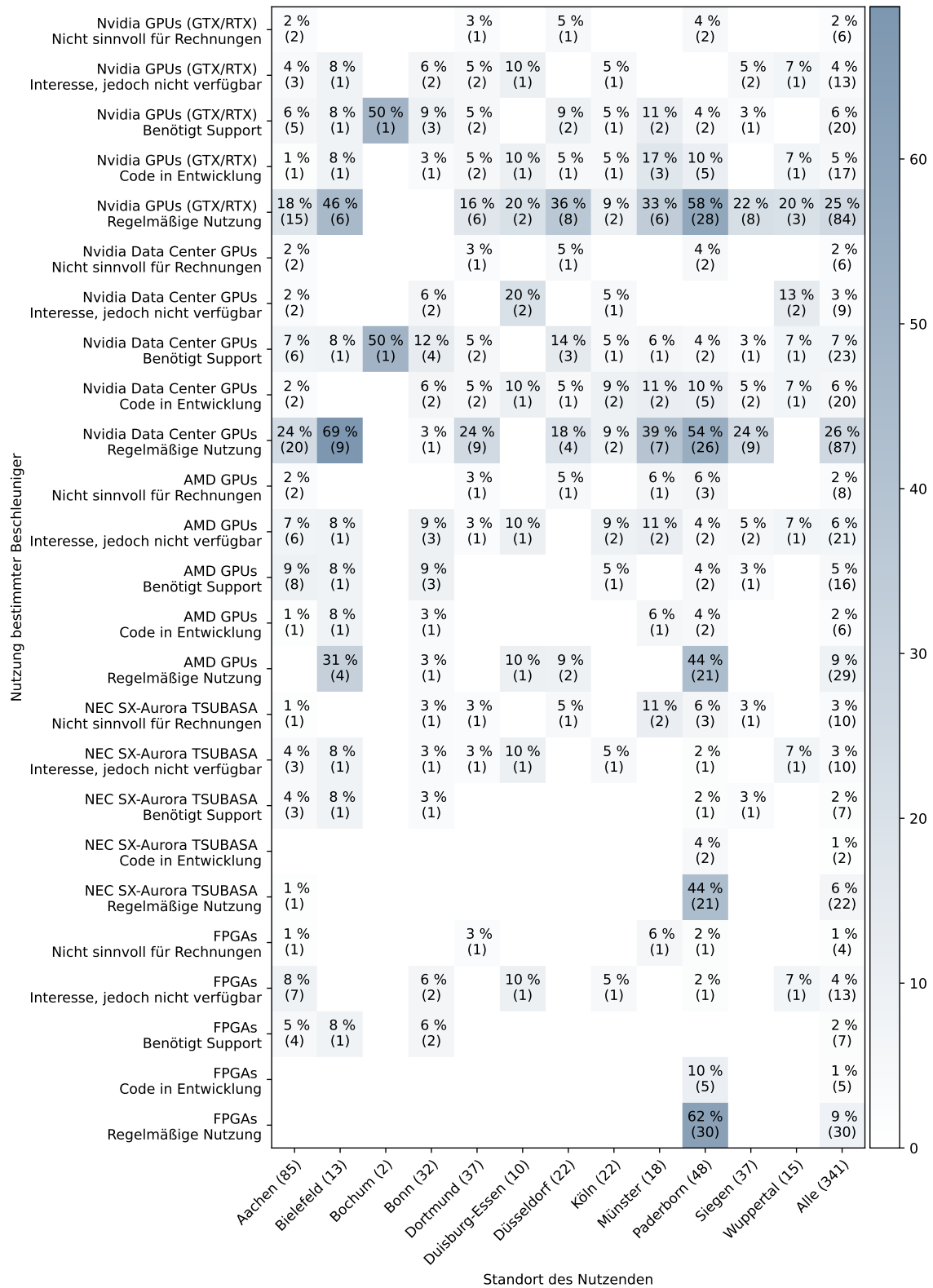


Abbildung 16: Verwendung von Beschleunigern der Befragten nach Standort. Die Darstellung ist in Sektion 1 genauer beschrieben.

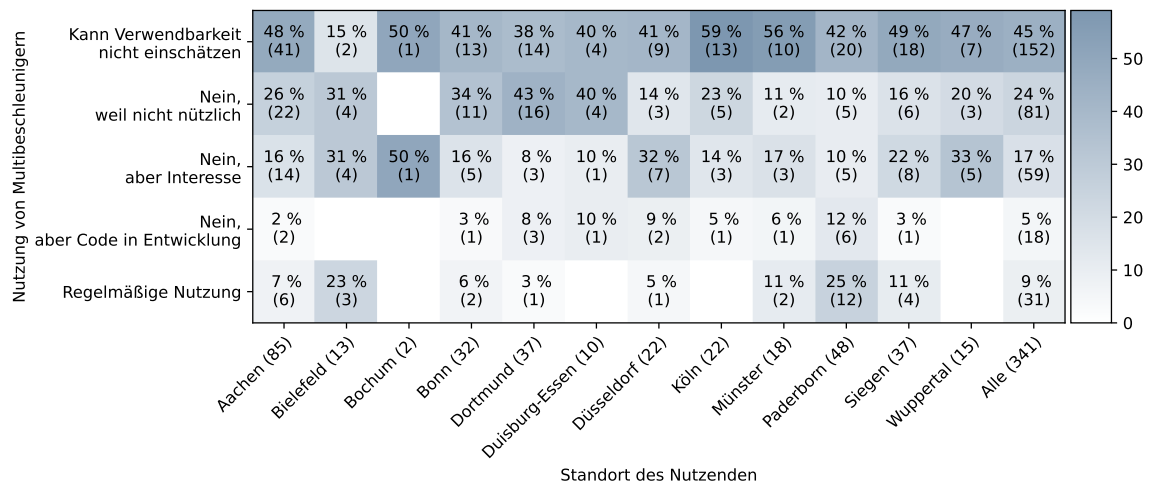


Abbildung 17: Verwendung oder Interesse an Nutzung von mehreren Beschleunigern nach Standort. Die Darstellung ist in Sektion 1 genauer beschrieben.

7 Performance-Analyse und Debugging

Mehr als ein Drittel der Befragten (39%) führt keine Performance-Analysen oder Profiling durch. Davon geben 60% als Grund an, dass sie nicht mit den entsprechenden Tools und Techniken vertraut sind, weitere 27% dass sie es nicht für notwendig erachten und 7% wissen, dass sich jemand anderes darum kümmert. Dagegen werden elementare Analysen von 48% der Befragten durchgeführt und 13% setzen gezielt Tools für die Performance-Analyse ein. Damit hat sich das Bild im Vergleich zu 2020 nicht wesentlich verändert.

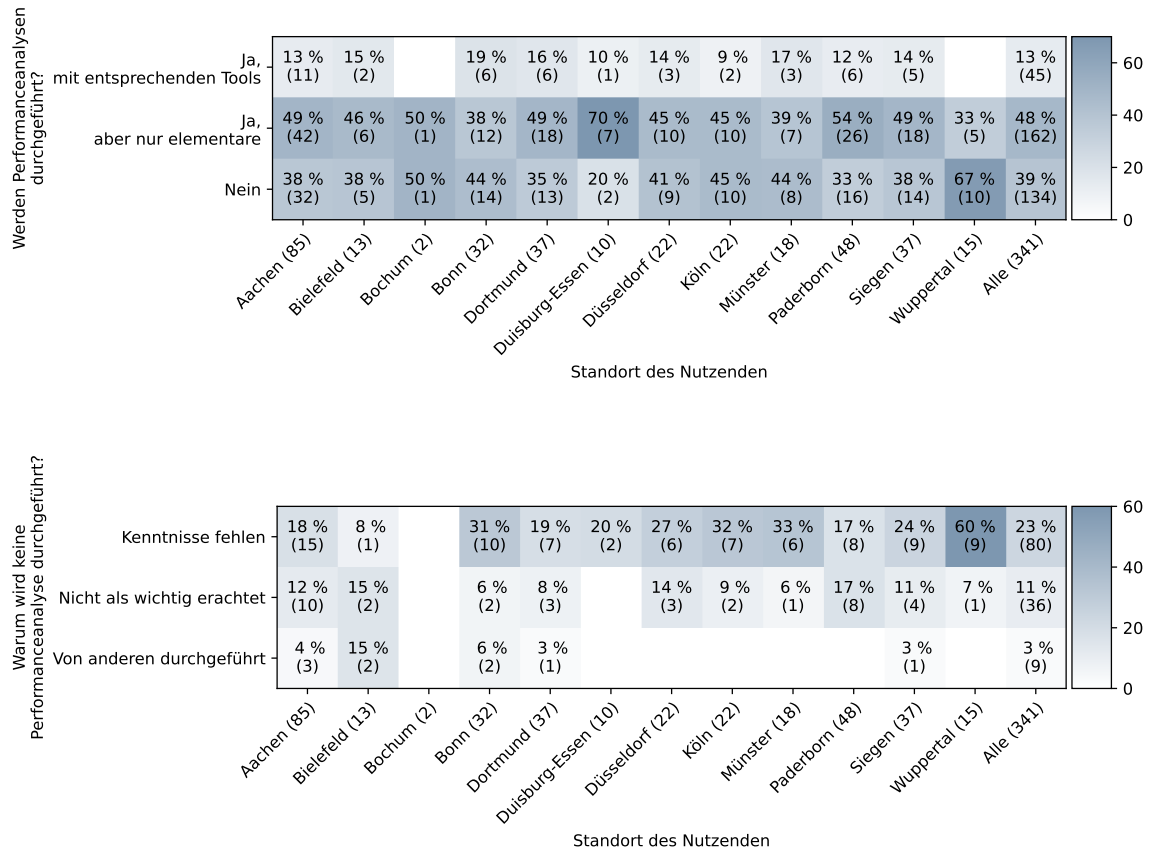


Abbildung 18: Durchführung von Performance-Analysen (oben) und Gründe (unten). Die Darstellung ist in Sektion 1 genauer beschrieben.

Auf die Frage, wie in parallelen Programmen die Fehlersuche und -behebung durchgeführt wird (Abb. 19), gaben 34% der Antwortenden an, eine Fehlersuche (Debugging) nicht zu benötigen. Aufgrund fehlender Kenntnisse zu Tools und Techniken führen 38% der Befragten keine Fehlersuche in parallelen Programmen durch. Somit sind weniger als ein Drittel der Befragten mit dem Debugging vertraut. Davon verwenden 2/3 (18% aller Befragten) einfache textuelle Ausgaben (print-Statements) und 1/3 (9%) setzen auch entsprechende Tools ein.

Wie in Abbildung 20 zu sehen, werden als häufigste HPC-Tools Valgrind (9%) und der GNU-Debugger (8%) verwendet. Weitere Tools wie Intel Performance Tools (6%), Gprof (4%) und Totalview (3%) wurden seltener genannt.

Zusätzlich zu den vorgegebenen Antwortmöglichkeiten wurden im zugehörigen Freitextfeld einzeln weitere Tools zum Profiling wie cProfile, Intel and Xilinx Profiler, Julia Profiling Library, NVIDIA Nsight und Profiler, perf und pstats genannt.

Etwa 47% der Befragten wollen in Zukunft keine weiteren Tools nutzen, da sie nach eigener Einschätzung keine weiteren Werkzeuge benötigen (Abb. 21). Etwa 13% der Befragten möchten in

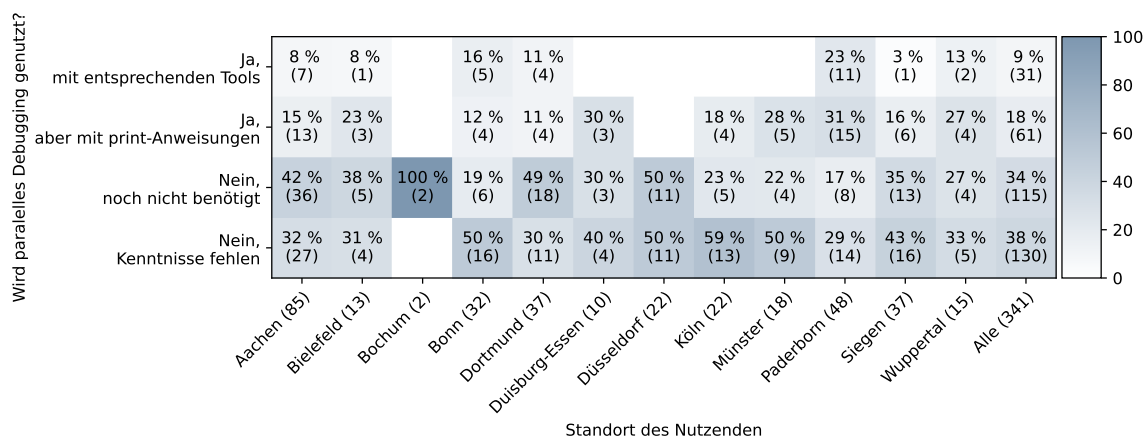


Abbildung 19: Nutzung von (parallelem) Debugging. Die Darstellung ist in Sektion 1 genauer beschrieben.

Zukunft zusätzlich Intel-Tools zur Performanceanalyse nutzen sowie 11% den GNU-Debugger und 7% Valgrind.

Insgesamt haben aber 12% zusätzlich im Freitext angegeben, dass sie keine Kenntnisse über die aufgeführten Tools haben oder sich unsicher sind, welche Tools für ihre Anwendungen geeignet sind, und somit keine Auswahl treffen konnten. Es empfehlen sich also Überblicksveranstaltungen, in denen viele Tools kurz besprochen werden, um bei den Nutzenden Bewusstsein für die Existenz und die Möglichkeiten zu schaffen.

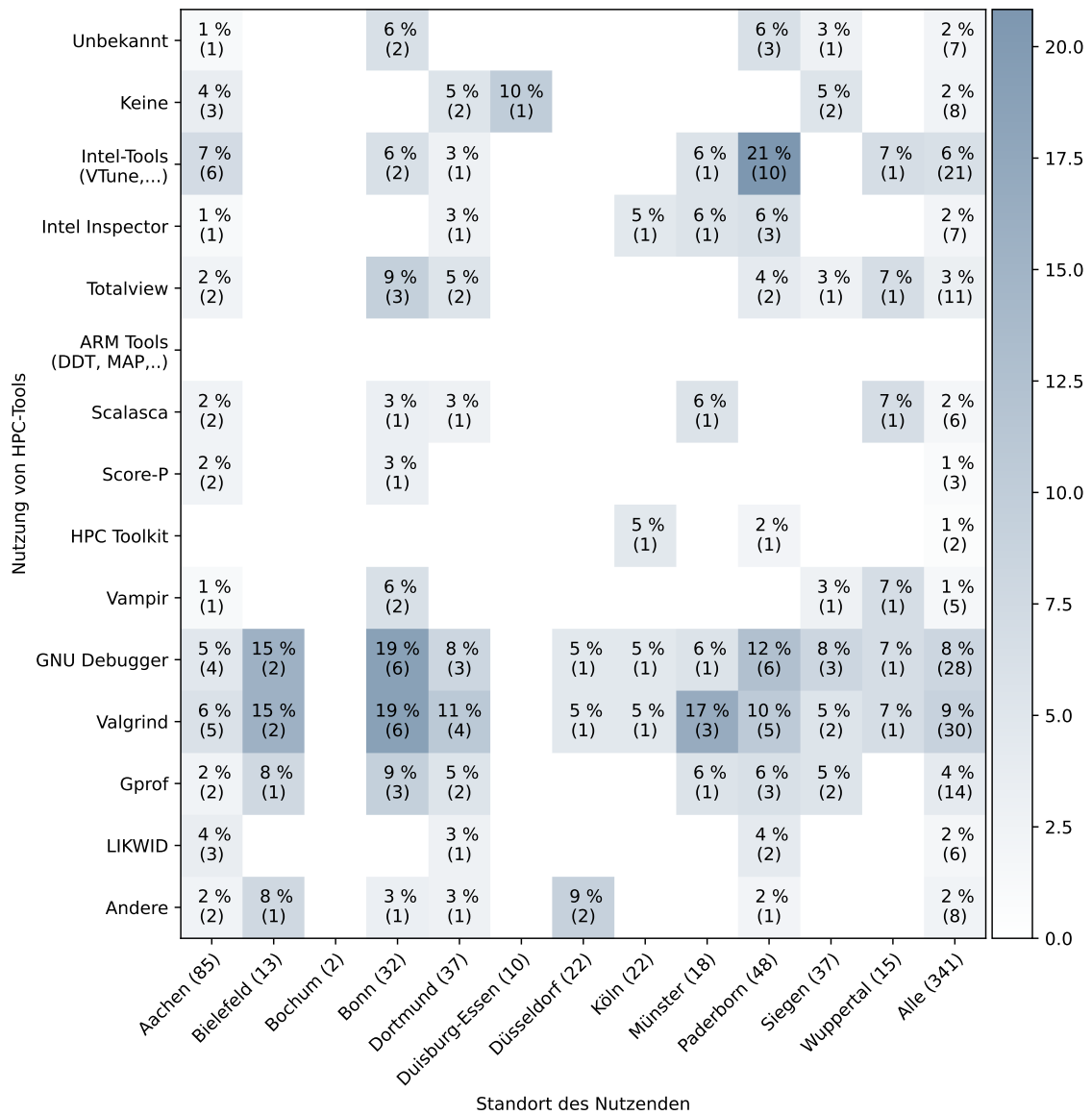


Abbildung 20: Nutzung von HPC-Tools durch die Befragten. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

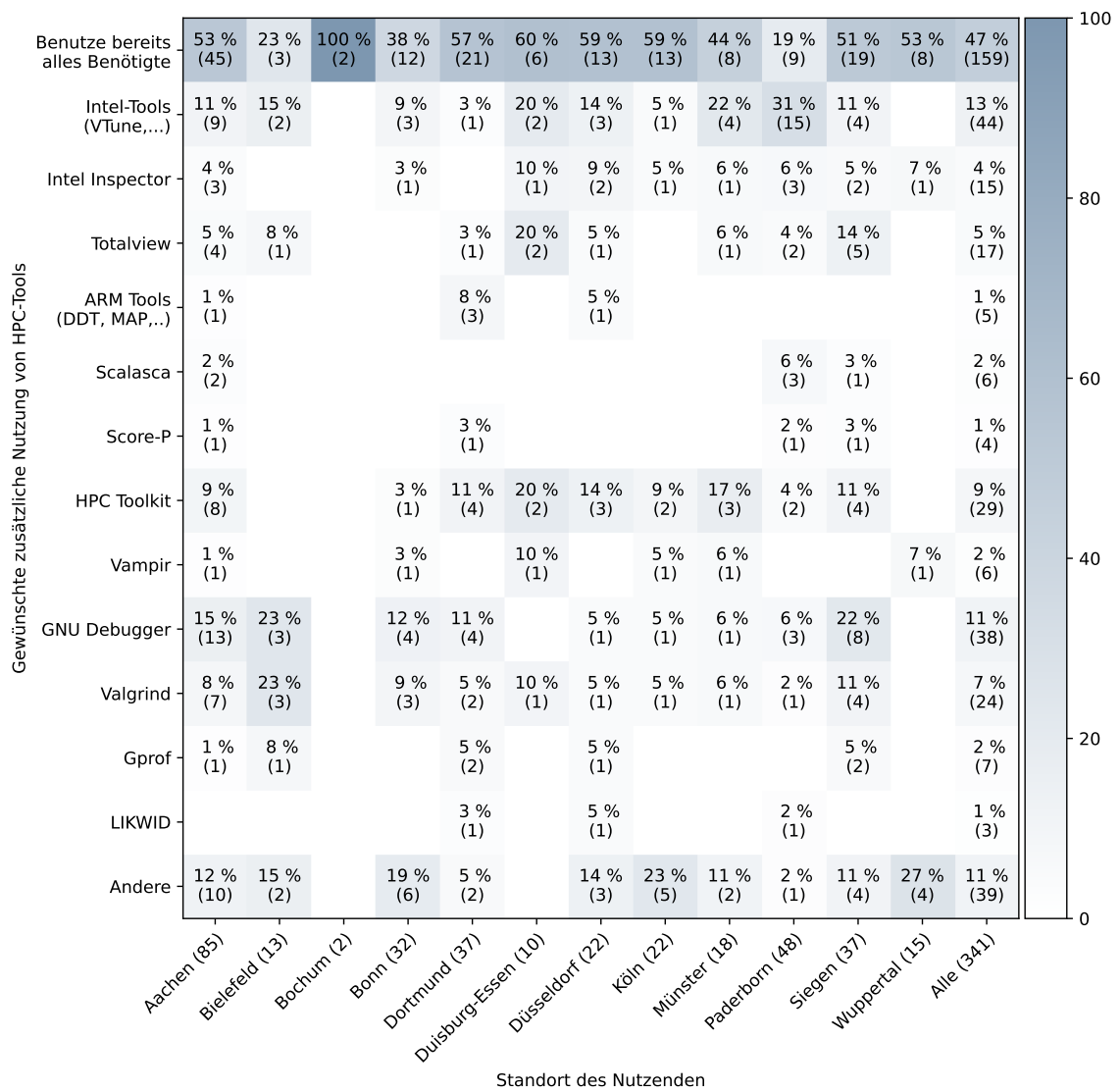


Abbildung 21: Gewünschte Nutzung von HPC-Tools durch die Befragten. Die Darstellung ist in Sektion 1 genauer beschrieben.

8 HPC-Basissoftware

Compiler und MPI

Zusätzlich zur Nutzung von HPC-Tools, die bereits in Kapitel 7 beschrieben worden ist, wurde die Nutzung von HPC-Basissoftware wie Compiler, Bibliotheken und Werkzeuge in Multiple-Choice und optionalen zusätzlichen Freitextfeldern in der Befragung erfasst. In diesem Bereich ist die erste relevante Frage zur Nutzung von Software, ob Nutzende selbst Software zur Nutzung auf den HPC-Systemen kompilieren. Die Statistik der Antworten wurde bereits in Abbildung 6 gezeigt. Im Durchschnitt kompilieren 60% der Befragten Software zur Nutzung auf den HPC-Systemen (auch) selbst. Dies zeigt, wie wichtig die Bereitstellung einer Bandbreite von verschiedenen Compilern, Bibliotheken und Entwicklungswerkzeugen auf HPC-Clustern und eine entsprechende Aufklärung ist. Besonders Nutzende des Bielefelder HPC-Systems fallen in dieser Auswertung durch eine hohe Rate auf. Ein wahrscheinlicher Grund dafür ist die spezielle Architektur dieses Clusters als reines GPU-Cluster (<https://www2.physik.uni-bielefeld.de/gpu-cluster.html>) und die starke anwendungsspezifische Ausrichtung des Systems.

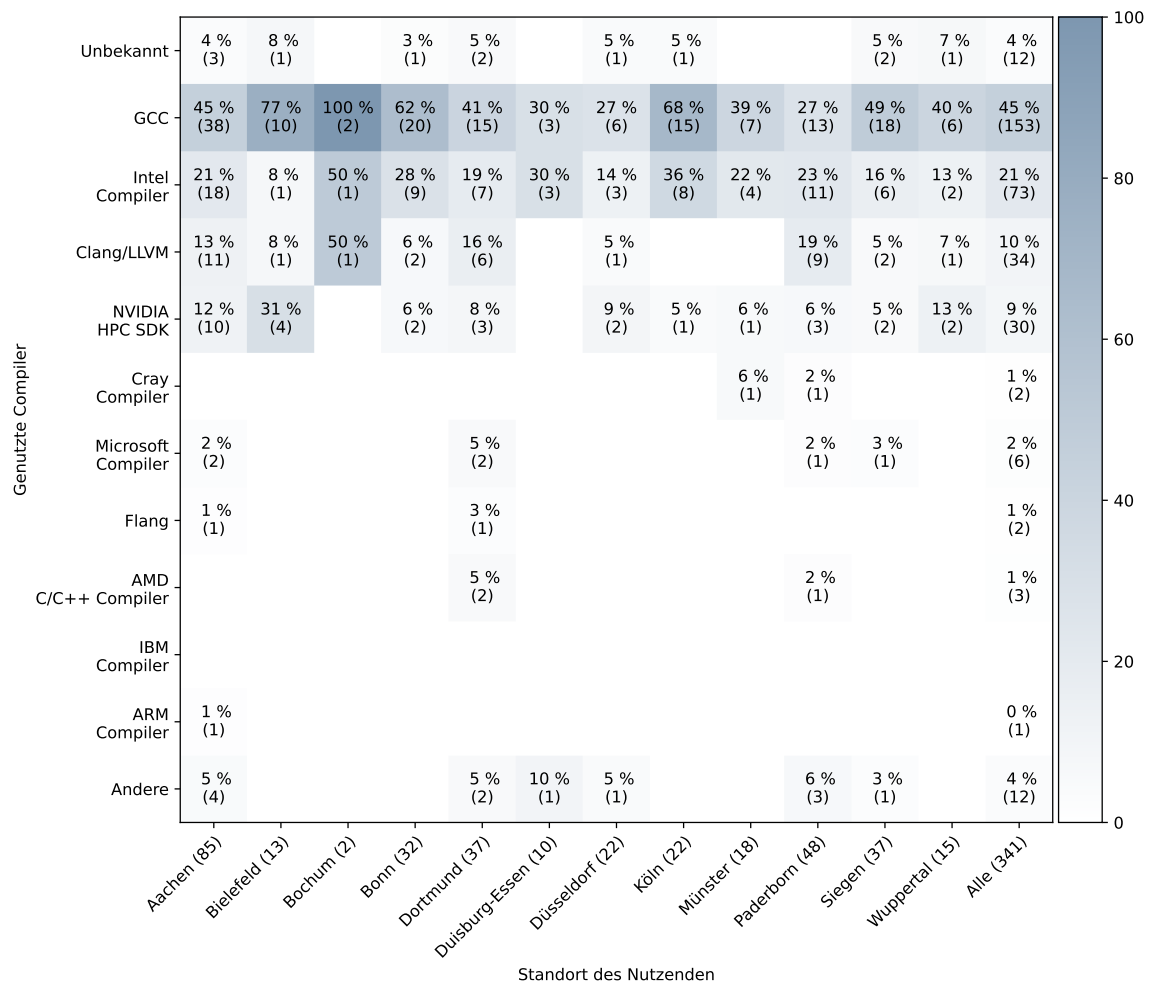


Abbildung 22: Von Befragten angegebene Nutzung von Compilern. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

In [Abbildung 22](#) sind die Angaben zur Nutzung von Compilern zusammengefasst. Dabei sind die Compiler der GNU Compiler Collection (GNU, gcc, g++, gfortran) mit 45% klar dominant in der

Nutzung, gefolgt von Intel Compilern (21%) und Clang als LLVM-basierten Compiler (10%), der im HPC-Umfeld noch relativ neu ist. Die in der Umfrage 2020 nicht genannten Flang Compiler werden diesmal von 2 Nutzenden (1%) verwendet, wohingegen IBM Compiler diesmal von niemandem verwendet werden.

Insbesondere durch die wieder wachsende Bedeutung von AMD-CPU's und ARM-basierten Systemen im HPC-Umfeld, wird die Bedeutung des GCC-Compilers sowie von Clang als auch anderen auf LLVM-basierenden Compilern, wie dem AMD Optimizing C/C++ Compiler (aocc) und Flang voraussichtlich weiter wachsen. Außerdem erzeugt die wachsende Verwendung von nicht-Intel-CPU's wichtige performancekritische Herausforderungen an die korrekte Verwendung von Intel-Compilern, sodass diese auch auf anderen Architekturen effizienten Code generieren. Für die Herausforderungen muss zuerst ein allgemeines Bewusstsein geschaffen und diese dann vor allem im HPC-Support adressiert werden.

In Bezug auf die Verwendung von MPI-Implementationen ergibt sich ein ähnliches Muster: eine freie Implementation, hier OpenMPI mit insgesamt 33%, führt die Statistik an. Dem schließt sich eine Implementation von Intel (Intel MPI) mit insgesamt 18% an, gefolgt von MPICH mit 8%. Je etwas unter 10% verwenden keine MPI-Implementation oder wissen nicht, was sie verwenden.

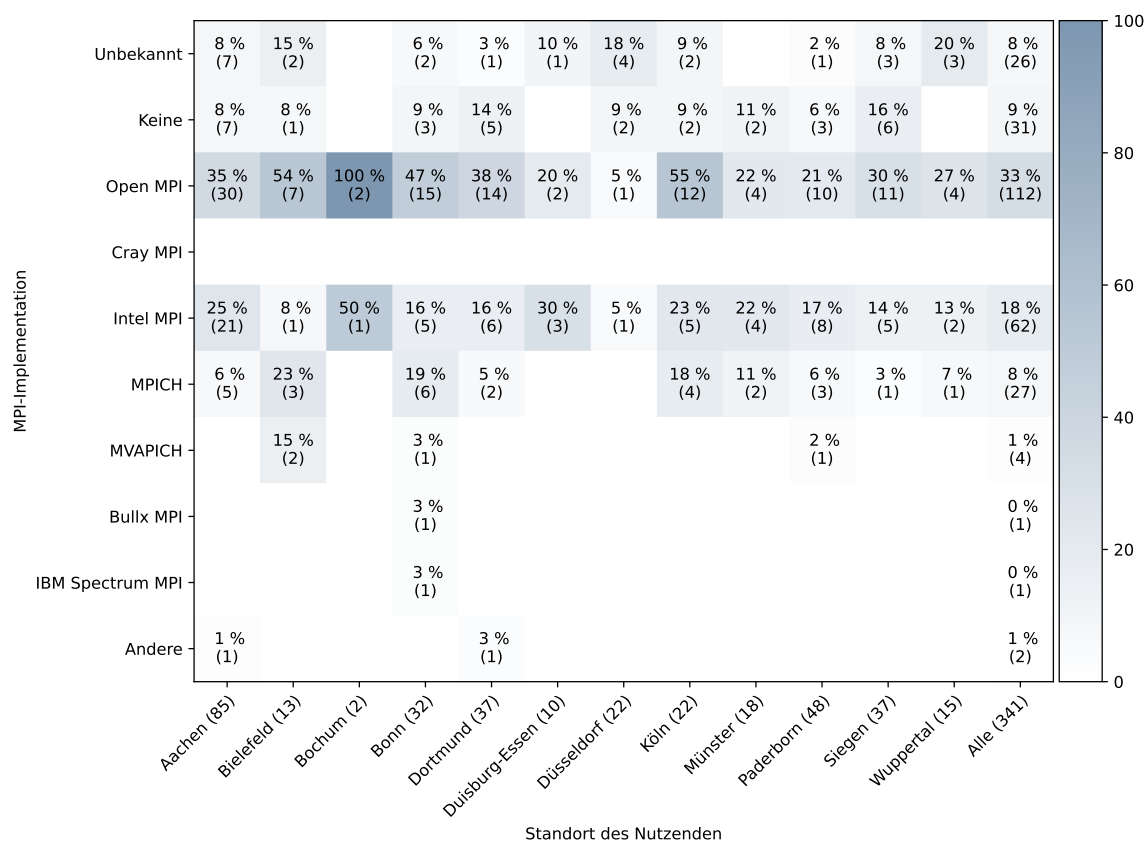


Abbildung 23: Von Befragten angegebene Nutzung von MPI-Implementationen. Die Darstellung ist in Sektion 1 genauer beschrieben.

Bibliotheken

Neben Fragen zur Nutzung von Compilern und MPI-Implementationen, enthielt die Nutzendenbefragung auch Fragen zur Verwendung von Bibliotheken. Dabei waren die in Abbildung 24 gelisteten Antworten vorgegeben und es wurden hier auch weitere im zugehörigen Freitextfeld angegeben. Für CPU-Systeme werden vor allem die MKL (16%), OpenBLAS (15%), FFTW (9%) und Eigen (7%) benutzt. Ähnlich viele Befragte, die angaben, die Intel MKL zu nutzen, führten auf, die numerischen Bibliotheken von Nvidia (cuBLAS,...) (15%) zu nutzen. Anders als 2020 kommen diese Angaben nicht mehr vorwiegend von Clustern, die GPUs ausschließlich, wie das Bielefelder Cluster, oder in großem Umfang, wie Aachen oder Paderborn, zur Verfügung haben. Es kann jedoch nicht ausgeschlossen werden, dass Nutzende von Tier-3 Standorten hierfür auf die Ressourcen von Tier-2 Standorten zurückgreifen. Der Trend des maschinellen Lernens ist auch hier ersichtlich, da knapp ein Fünftel der Nutzenden bereits eine entsprechende Bibliothek verwendet (11% pyTorch und 10% Tensorflow). Zu Bibliotheken wurden nur wenige Freitextantworten gegeben. In diesen kamen vereinzelt SLATE, OpenMP, oneAPI, NetCDF, NAG, mpi4py, LSDyna, HIP, HDF5, booster und Armadillo vor.

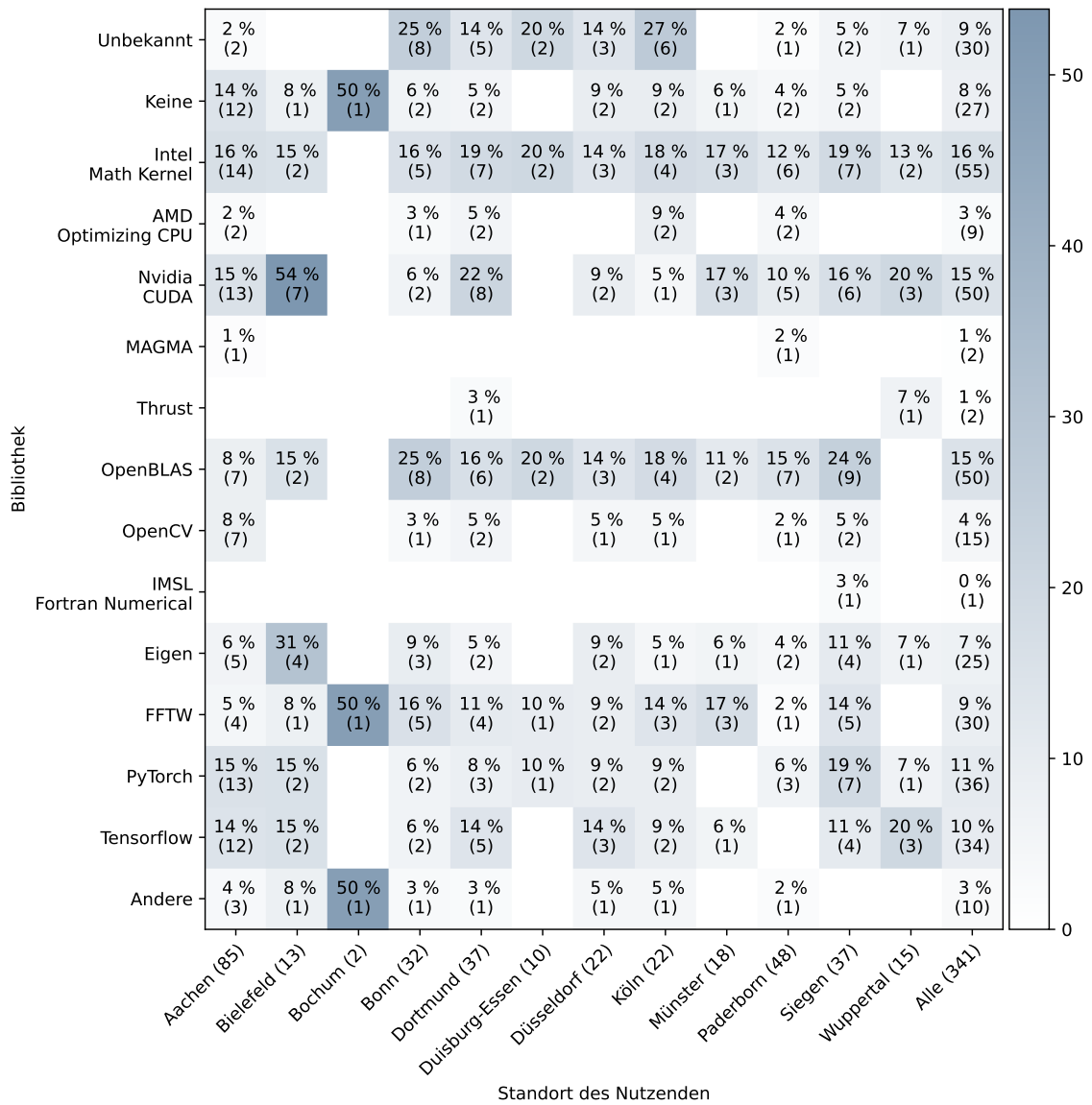


Abbildung 24: Von Befragten angegebene Nutzung von Bibliotheken. Die Darstellung ist in Sektion 1 genauer beschrieben.

Nutzung von Zugangssoftware

Die Erstellung und Submittierung von Compute-Jobs bilden die Kerninteraktion von Nutzenden mit einem HPC-System. Abbildung 25 zeigt die von den Teilnehmenden dafür verwendeten Programme/Wege. Wie erwartet, bilden kommandozeilenbasierte und grafische SSH-Clients den Hauptzugangsweg. Insbesondere Nutzende des Paderborner Systems verwenden jedoch auch in signifikantem Maße Remote-Desktop-Lösungen, Grafische SSH Clients oder Submittierung über Anwendungen. Das ist ein deutlicher Unterschied zur Umfrage 2020, in der die Job-Submittierung über Anwendungen und per Remote-Desktop kaum Anklang fand. Dieser Trend liegt unter anderem darin begründet, dass in Paderborn Remote-Desktop-Lösungen entwickelt worden sind, die auch Verwendung in zahlreichen Kursen, Workshops und anderen Veranstaltungen fanden. Dadurch konnten die Nutzenden bei mehreren Gelegenheiten die benötigte Erfahrung sammeln. Diese beiden Punkte sind an den meisten anderen Standorten weiterhin in der Unterzahl. Bei den in Zukunft gewünschten Zugangswegen sind diese Mechanismen jedoch zusammen mit Remote-Desktop-Lösungen stark vertreten, wie in Abbildung 26 deutlich wird. Das zeigt, dass Betriebskonzepte für HPC-Systeme solche Wege weiter in den Fokus rücken müssen.

Bezüglich des Zugangs zum HPC-System gaben die Befragten im Freitextfeld an, dass die Wartezeiten bis zum Start des eigenen Jobs manchmal zu lang ausfallen. Außerdem gibt es gelegentliche Probleme mit der Übersetzung von auf Workstations funktionierender Software zum Cluster-System. Der Wunsch nach einem Internetzugang vom Cluster aus wird auch mehrfach genannt.

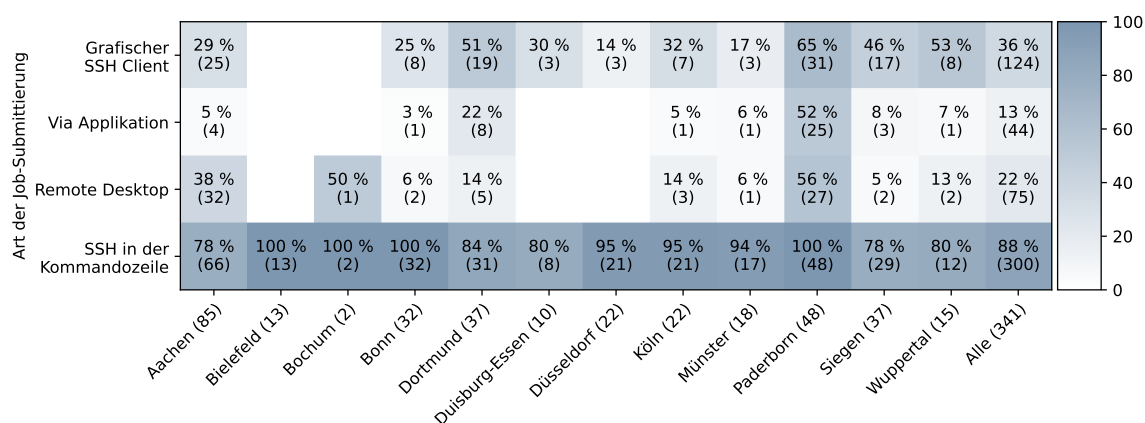


Abbildung 25: Für die Job-Submittierung genutzte Software. Die Darstellung ist in Sektion 1 genauer beschrieben.

Für den Zugang zu Daten auf dem HPC-System werden laut Abbildung 27 vorwiegend SCP oder die direkte Bearbeitung der Daten auf dem HPC-System verwendet. Seit 2020 ist ein immenser Anstieg bei der Nutzung von CFS/SMB (von 21% auf 60%) und NFS Exports (von 9% auf 60%) in Paderborn zu verzeichnen. Mäßigen Anklang finden diese auch in Bonn, Dortmund, Düsseldorf und Münster. Wie bereits bei der Job-Submittierung, wird der Anstieg in der Art und Weise wie Nutzende in Paderborn auf ihre Clusterdaten zugreifen darauf zurückzuführen sein, dass wiederholt in Kursen auf den Gebrauch von CIFS und NFS eingegangen wurde. Weiterhin ist nicht klar, ob der mangelnde Gebrauch an anderen Standorten an technischen Herausforderungen für die Nutzenden liegt oder die Bekanntheit solcher Wege verbessert werden muss. In der gewünschten zukünftigen Nutzung liegen die Exports (Abb. 28) im oberen Bereich des ansonsten recht homogen verteilten Spektrums von Antworten. Die Befragten gaben als weitere Software, die sie für den Datentransfer nutzen, noch sftp, rsync und XRootD/WebDAV an. Einige beschrieben nicht den Zugriff auf die Daten als Problem, sondern vielmehr das Austauschen der Daten mit Personen

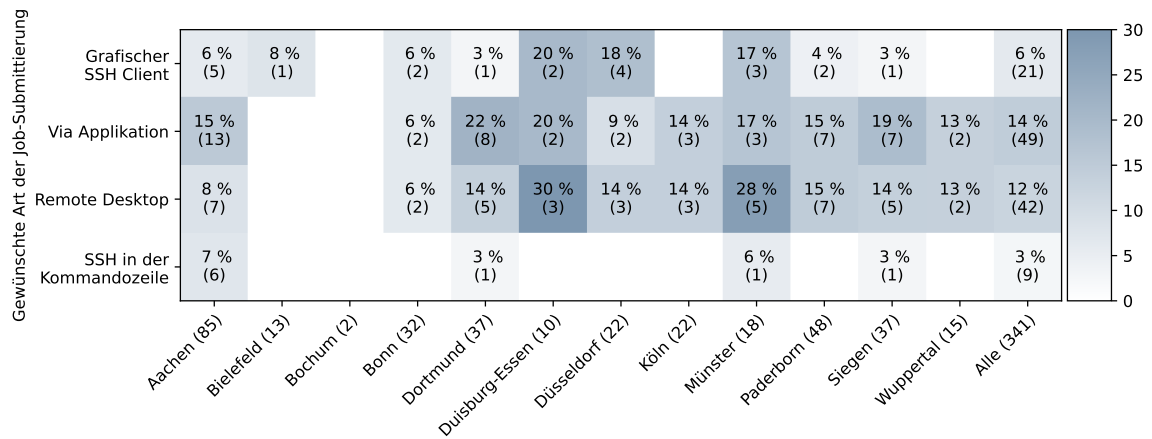


Abbildung 26: Für die Job-Submittierung gewünschte Software. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

anderer Institutionen. Hier werden Lösungen wie GridFTP oder Globus gewünscht. Seitdem ein Zugang zu vielen HPC-Systemen nur noch über VPN möglich ist, erfahren die Befragten nach eigenen Angaben längere Transferzeiten.

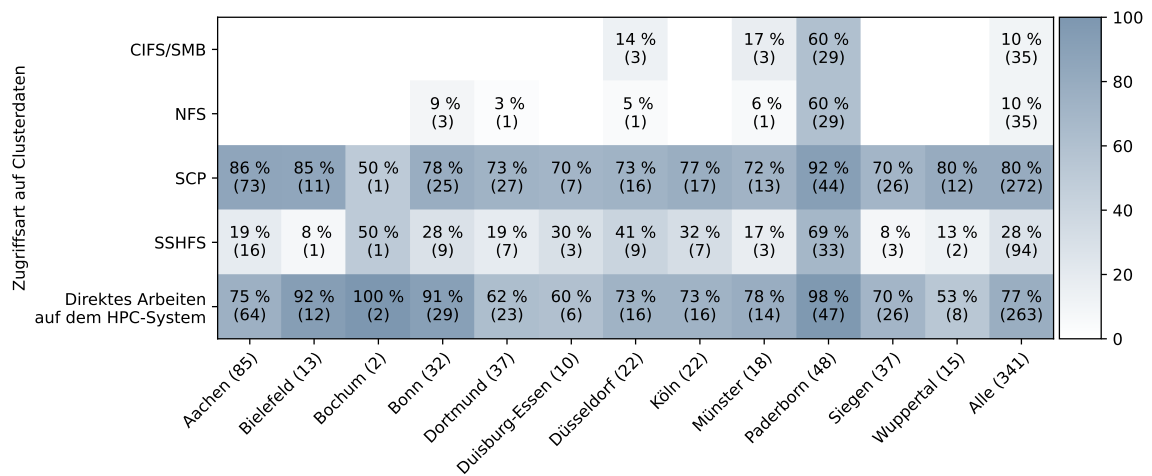


Abbildung 27: Für den Zugang zu Daten genutzte Software. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

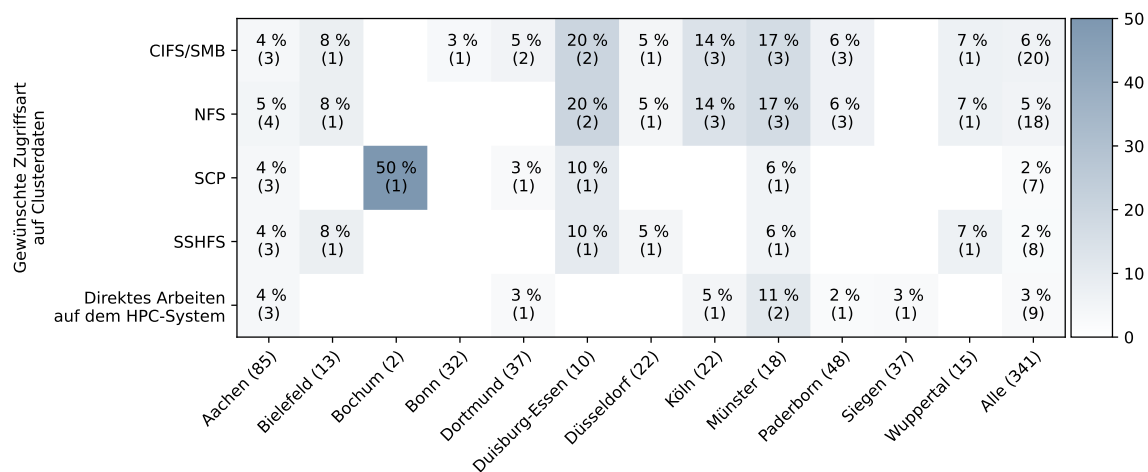


Abbildung 28: Für den Zugang zu Daten gewünschte Software. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

Nutzung von Jupyter-Notebooks

Jupyter Notebooks stellen einen modernen und flexiblen Weg dar, HPC-Systeme mit Sprachen wie Python, Julia oder anderen zu benutzen. Wie Abbildung 29 zeigt, ist die Etablierung solcher Dienste und die effiziente Integration in HPC-Systeme noch am Anfang, da an 7 Standorten unter 10% der Befragten Jupyter Notebooks nutzen (davon 4 Standorte gar nicht). Gleichzeitig ist anzumerken, dass gegenüber der Nutzung im Jahr 2020 (Abb. 29, oben) größtenteils ein Anstieg zu verzeichnen ist (Abb. 29, unten). So geben an 6 von den 8 Standorten, an denen Jupyter Notebooks genutzt wird, mindestens 10% der Befragten an, Jupyter Notebooks zu verwenden. In Bielefeld und Wuppertal liegt die Nutzung sogar bei über 20%. Aus allen Standorten sticht Paderborn hier klar hervor, da knapp 70% der Befragten von Jupyter Notebooks Gebrauch machen. Ein wesentlich Grund hierfür wird sein, dass seit 2020 in Paderborn eine integrierte Lösung für den Einsatz von Jupyter Notebooks im HPC-System entwickelt und den Nutzenden bereitgestellt wurde. Zusätzlich zum reinen Angebot von Jupyter Notebooks wurden diese auch in mehreren Veranstaltungen genutzt, so dass es wiederholte Hands-On Erfahrungen gab. Die jetzige Nutzung steht jedoch weiterhin im starken Kontrast zur von den Teilnehmenden gewünschten zukünftigen Nutzung in Abb. 30. Zwischen etwa 20% und 50% der Befragten (variiert nach Standort) wünschen sich solche Dienste oder den Ausbau von existierenden Diensten. Der geringe Wunsch Jupyter Notebooks künftig in Paderborn weiter zu verwenden, ist darauf zurückzuführen, dass Nutzende des Dienstes (knapp 70% in Paderborn) diese Frage nicht mehr gestellt bekommen haben.

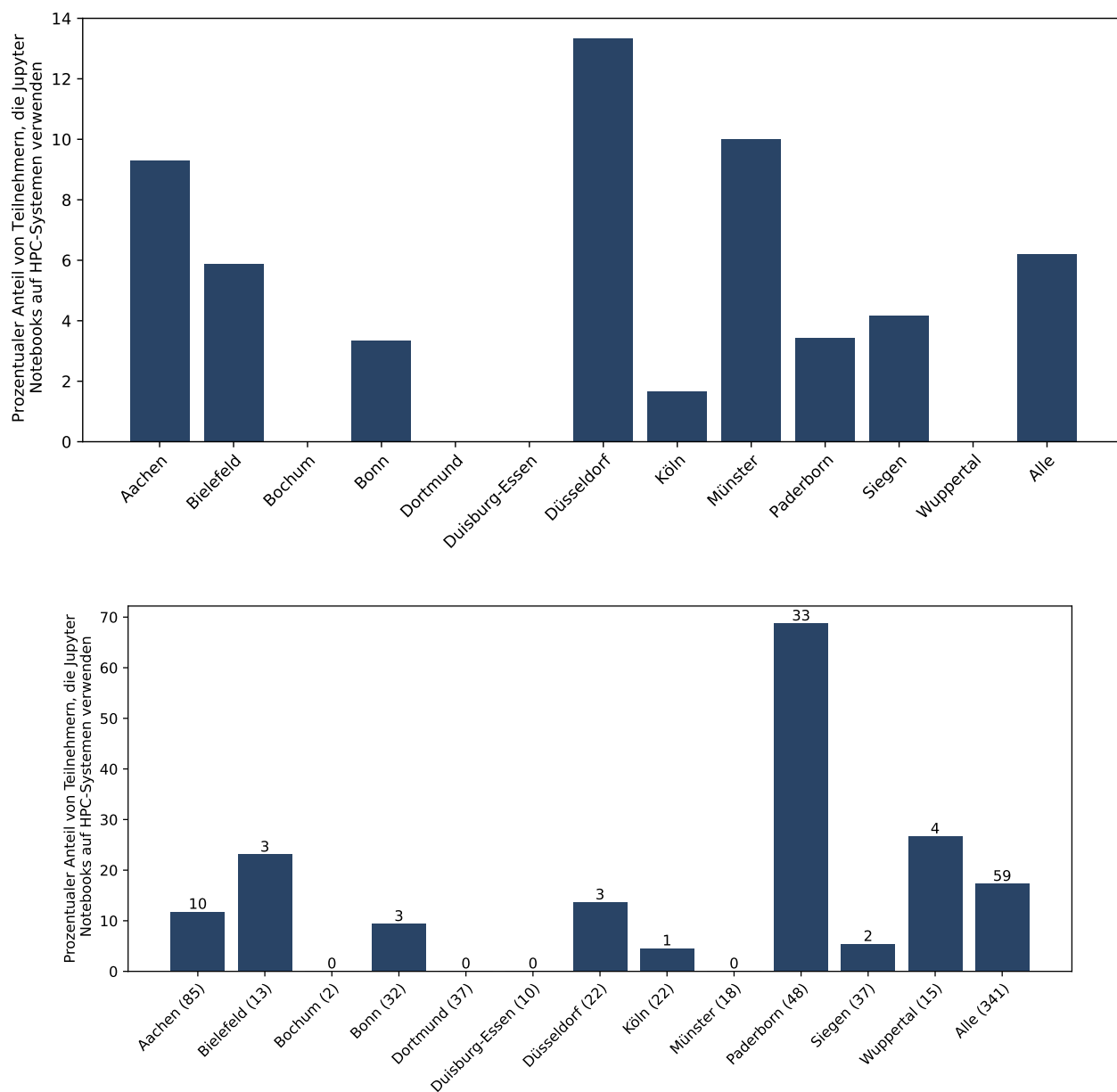


Abbildung 29: Nutzung von Jupyter-Notebooks auf HPC-Systemen im Jahr 2020 (oben) und 2022 (unten). Die Werte an den Balken zeigen die absolute Anzahl an entsprechenden Antworten.

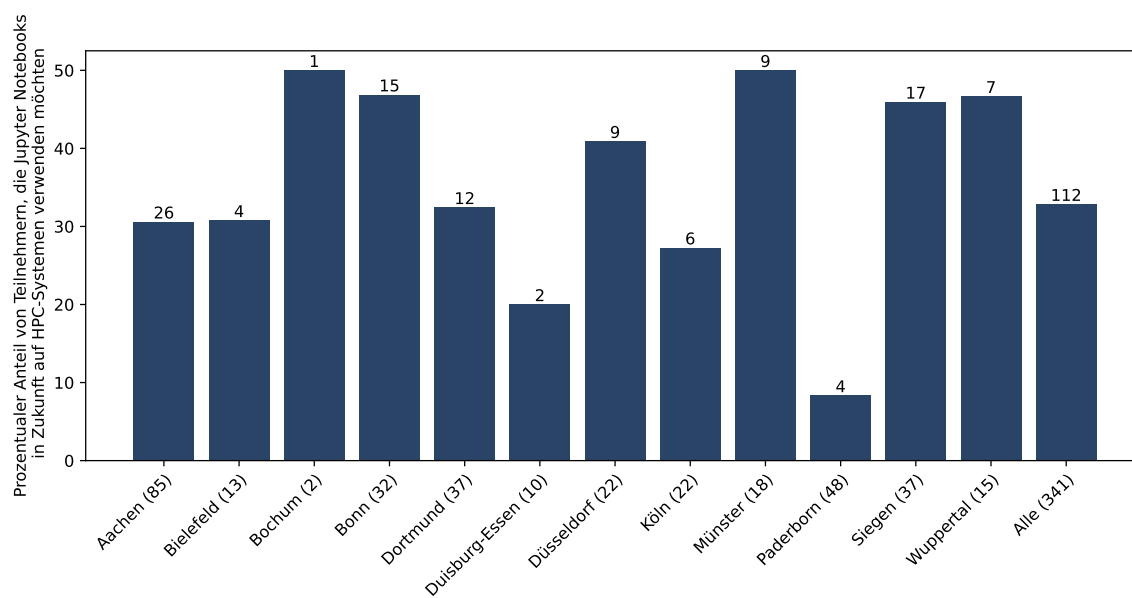


Abbildung 30: Gewünschte Nutzbarkeit von Jupyter-Notebooks auf HPC-Systemen. Die Werte an den Balken zeigen die absolute Anzahl an entsprechenden Antworten.

Nutzung von Softwarecontainern

Eine ähnlich disruptive Technologie wie Jupyter Notebooks sind Softwarecontainer. Diese können das Deployment von Software und insbesondere die Reproduzierbarkeit von wissenschaftlichen Simulationen stark verbessern, bringen jedoch auch interessante neue Herausforderungen mit sich. Die Verteilung der Nutzung auf die bekannten Softwarecontainerlösungen ist in Abbildung 31 gezeigt. Dabei dominieren Singularity und Docker die derzeitige Nutzung klar. Eine Ausnahme stellt Paderborn dar, wo je 48% der Nutzenden auch Charliecloud, Kubernetes, Podman oder Shifter kennen. Gefragt nach der zukünftigen gewünschten Nutzung, die in Abbildung 32 dargestellt ist, legen die Teilnehmenden eine Priorität auf Docker-basierte Container, die in der Auswertung von Singularity und Kubernetes gefolgt wird. Da konventionelle Docker-Container für das HPC-Umfeld ungeeignet sind, sollten die Standorte mit hoher Nachfrage nach Docker-Containern entweder vollständig unprivilegierte Docker-Container auf den HPC-Systemen ermöglichen oder die Nutzenden über die Konvertierung von Docker- zu Singularity-Containern schulen.

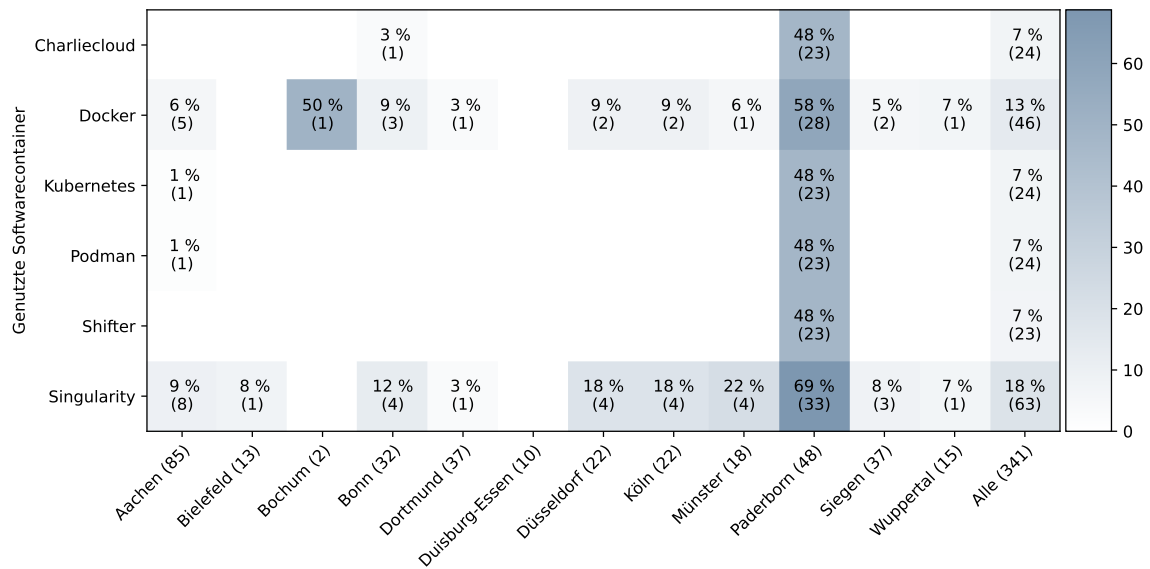


Abbildung 31: Nutzung von Softwarecontainern auf HPC-Systemen.

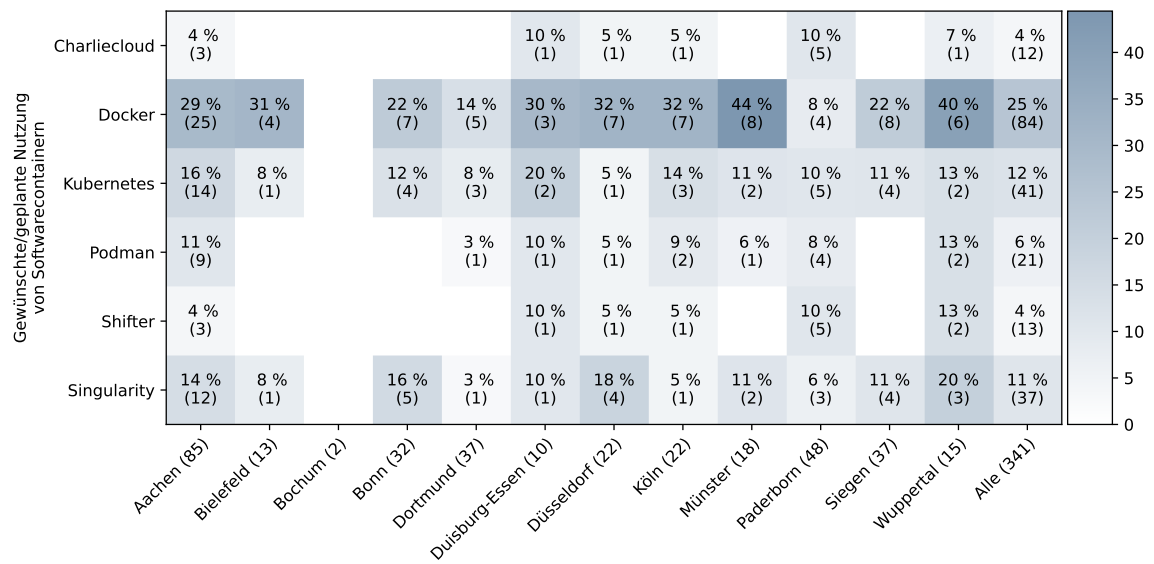


Abbildung 32: Gewünschte Nutzung von Softwarecontainern auf HPC-Systemen.

9 Wissenschaftliche Software

Softwaretypen

Gefragt nach der Nutzung von hauseigener, Open Source oder kommerzieller Software gaben die meisten Befragten an, Open Source Software (73%) zu nutzen, gefolgt von Custom Software (57%) und kommerzieller Software (27%). Hierbei sind die Fluktuationen nach Standorten gering. Auch die relativ geringe Abhängigkeit von kommerzieller Software ist erwähnenswert. Die Verwendung von viel selbst entwickelter oder hauseigener Software stellt hohe Anforderungen an die Programmier- und Optimierungskennntnisse der Nutzenden, was damit auch sehr hohe Ansprüche an die HPC-Weiterbildungsprogramme der Standorte stellt.

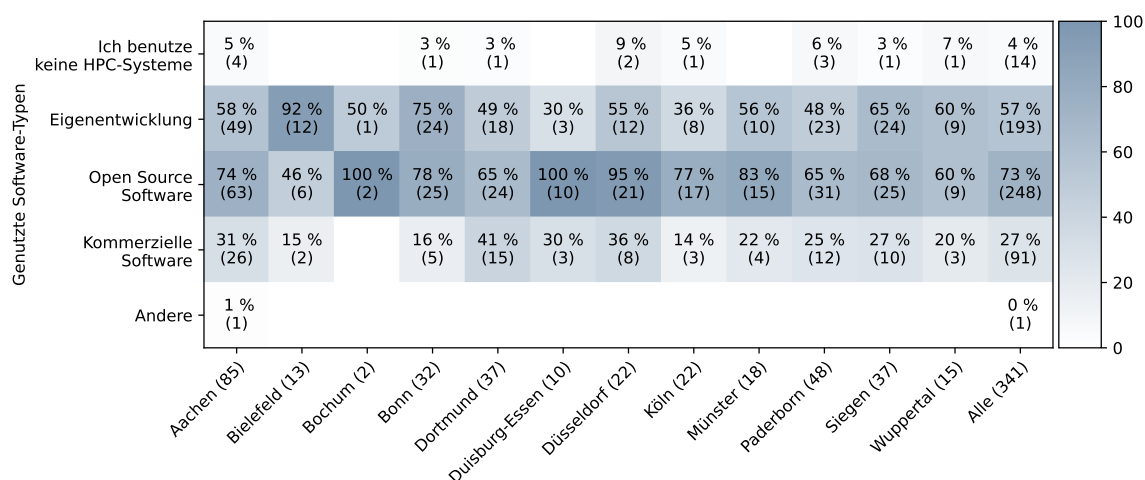


Abbildung 33: Nutzung von verschiedenen Software-Typen. Die Darstellung ist in Sektion 1 genauer beschrieben.

Erhebung

Die Erfassung der Nutzung von wissenschaftlicher Software über die Nutzendenbefragung ist wegen der Vielzahl an denkbarer Software eine große Herausforderung. Hier hatten wir uns für ein großes Freitextfeld entschieden, was den Teilnehmenden der Umfrage das Ausfüllen leicht machte, da nicht lange Listen oder Untermenüs von Software durchsucht werden mussten, jedoch den Aufwand auf die Auswertung verschob.

Nutzung

In der Analyse der über 600 Nennungen wurden mehrere hundert verschiedene Softwares gefunden. Im Durchschnitt verwendet jeder Nutzende zwei verschiedene Softwares. Die über alle Standorte hinweg am häufigsten angegebene Software ist in Abbildung 34 dargestellt. Außerdem sollte angemerkt werden, dass knapp 4% der Nutzenden angegeben haben, eigens oder am Lehrstuhl entwickelte Software zu verwenden.

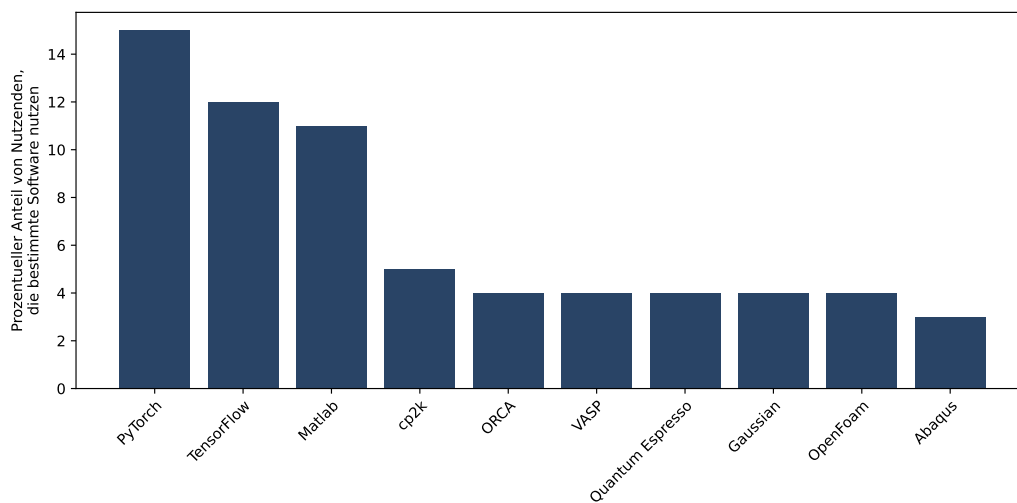


Abbildung 34: Software die am häufigsten von den Befragten angegeben wurde. Es ist nur Software gelistet, die über alle Standorte hinweg mindestens 10 mal genannt wurde.

10 HPC Ressourcen

Die verfügbaren **Infrastrukturen und Dienste** bewerten 72% der Befragten als ausreichend (Abb. 35). Sie können effizient arbeiten. Weitere 18% kommen über Umwege zurecht.

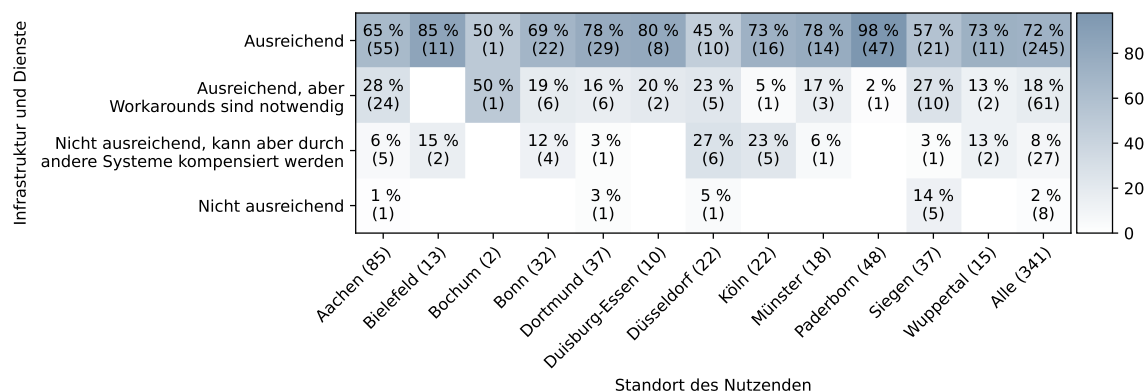


Abbildung 35: Einschätzung der vorhandenen Infrastruktur und Dienste. Die Darstellung ist in Sektion 1 genauer beschrieben.

Im Folgenden werden Freitext-Antworten mit Bezug zur Infrastruktur zusammengefasst. Natürlich ist nicht jeder Punkt für jeden Standort zutreffend. Weiter ist hervorzuheben, dass es sich hierbei nur um vereinzelte Antworten handelt:

- Aktualität und Verfügbarkeit von Software
 - Veraltete Versionen von GNU Compiler, R, Python
 - Nicht genügend verschiedene Versionen von Software (inklusive veralteter)
 - Software nicht verfügbar (Ansys, Jupyter Notebook, bestimmte Compiler, VASP)
 - Beschränkter Internetzugang hinderlich
- Verfügbarkeit von Ressourcen (Knoten, Dateisystem)
 - Häufige Wünschäußerung nach mehr GPUs
- Qualität der Ressourcen
 - Zugriff auf das Dateisystem zu langsam
 - Datentransfer nach außerhalb des Clusters zu langsam (auch wegen VPN) und zu umständlich, schlechte Verbindung (z.B. zum NCBI Server), zu geringe Bandbreite von den Compute-Knoten nach außen
 - Compute-Knoten zu langsam, Knoten haben zu wenig Speicher
 - Ressourcenmanagement auf den Knoten bei mehreren Nutzenden
 - Vereinfachter Zugang über SSH
- Lange Wartezeiten/Scheduling:
 - Zu lange Wartezeiten für Jobs
 - MPI-Jobs werden durch Single-Node Anwendungen anderer User geblockt
 - Zu lange Wartezeiten für viele sehr kurze Jobs (Wartezeit \geq Rechenzeit)
 - Maximale Anzahl von Jobs zu niedrig
- Besserer Wissenstransfer:
 - Nachfrage nach Basiswissen/Einführungskursen
 - Workflow/Arbeitsschritte sind oft unstrukturiert und unorganisiert

- Wunsch nach Rezepten zur Job-Submittierung
 - anwendungsspezifischer Support (z.B. ORCA)
 - Einarbeitungszeit für Studierende/Mitarbeiter zu lang
 - Trainingsmöglichkeiten auf dem Cluster
 - mehr Schulungen und Integration von HPC in die Lehrpläne nötig
- Möglichkeit, Rechenzeit ohne großen Aufwand (Antrag) zu bekommen

Die verfügbare **Hardware** (HPC-Ressourcen, Abb. 36) ist für 69% der Befragten ausreichend. Bei weiteren 10% mussten zeitaufwendige Optimierungen vorgenommen werden, um mit den verfügbaren HPC-Ressourcen auszukommen. Ein anderes Bild zeigt sich in Bielefeld, wo für 54% der Befragten die HPC-Ressourcen nicht genügen. Auch in Köln und Münster ist bei knapp 30% ein Mehrbedarf vorhanden. Im wesentlichen scheint die Wahrnehmung der Hardware sich gegenüber den Ergebnissen von 2020 kaum verändert zu haben. Dies ist auch schlüssig, da entweder keine neuen Beschaffungen getätigt wurden oder rechtzeitig neu beschafft wurde. Eine Ausnahme stellt Paderborn dar, wo seit der letzten Umfrage Noctua 2 in Betrieb genommen wurde. Dies hat zur Folge, dass dort nun 96% der Nutzenden die Hardware für ausreichend erachten (statt 67% in 2020). Ähnlich verhält es sich in Wuppertal, wo 2021 das neue System in Betrieb genommen wurde. Hier erhöht sich der Anteil jener Nutzenden, die zufrieden mit dem Hardwareangebot sind, von 50% (in 2020) auf 80%. In den Freitextantworten ist die Nachfrage nach GPUs besonders stark. Falls GPUs schon

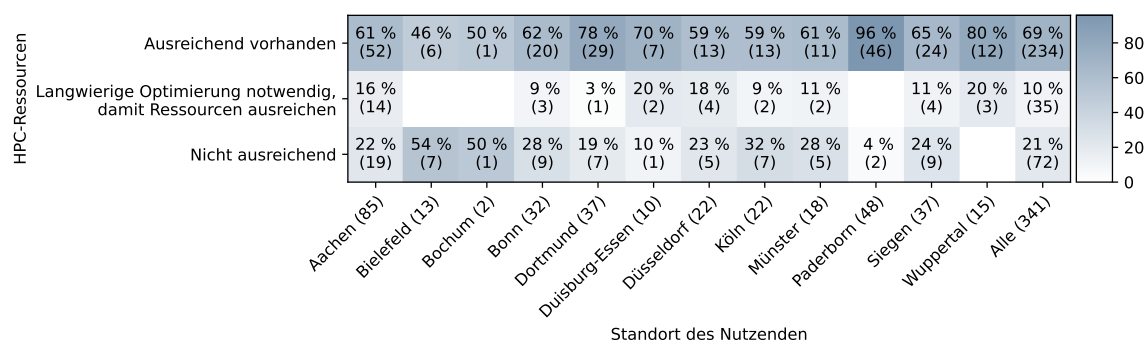


Abbildung 36: Einschätzung der vorhandenen Hardware-HPC-Ressourcen. Die Darstellung ist in Sektion 1 genauer beschrieben.

vorhanden sind, wünschen sich die Befragten mehr dieser Ressourcen, sowie neuere Modelle und Ausführungen mit mehr GPU-Speicher. Ebenfalls häufig werden insgesamt mehr Rechenknoten und Knoten mit mehr Arbeitsspeicher gewünscht, bei letzteren sind den Nutzenden die Wartezeiten oft zu lang. Ebenfalls erwähnt wurden mehr Rechenzeit, mehr permanenter Speicher im Dateisystem oder ein Dateisystem zum Archivieren sowie Containerlösungen und mehr Support.

11 Support

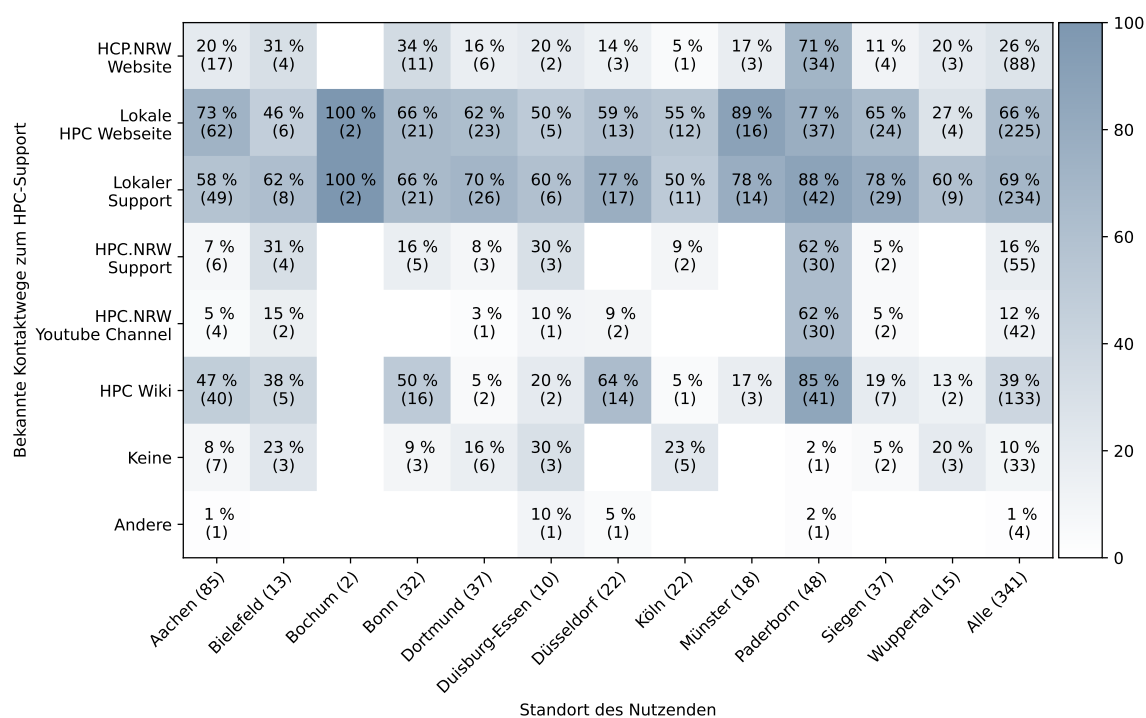


Abbildung 37: Anteile der Befragten, denen verschiedene Kontaktwege zum HPC-Support bekannt sind. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

Die Bekanntheit des lokalen HPC-Supportes, Nutzung des HPC-Supportes und weitere Fragestellungen wurden in der Nutzendenbefragung ebenfalls abgedeckt. [Abbildung 37](#) zeigt die Bekanntheit des lokalen HPC-Supports und der verschiedenen Support-Angebote von HPC.NRW, die sich nicht nur auf den direkten Support beschränken, sondern auch das Angebot der Wiki-Seite und des Youtube Kanals beinhalten. Hier zeigt sich, dass maßgeblich der lokale Support (69%) und die lokale HPC Webseite (66%) bei den Nutzenden bekannt sind. Das bekannteste Element des HPC.NRW Netzwerkes stellt die HPC Wiki dar. Ein Grund hierfür kann sein, dass in HPC-Veranstaltungen und auch durch den lokalen Support häufig auf die Wiki-Seite verwiesen wird. Auf Basis der durchweg erhöhten Wahrnehmung von HPC.NRW über alle nennbaren Kanäle in Paderborn wird diese Vermutung nochmals bekräftigt. Durch das breite und regelmäßige Angebot an HPC-Veranstaltungen (User-Day, Kurse, HPC Sprechstunden) in Paderborn können die verschiedenen Kanäle wiederholt kommuniziert werden, so dass sich diese bei den Nutzenden einprägen.

Wie in [Abbildung 38](#) zu sehen, wissen nicht alle Befragten wie sie den lokalen HPC-Support kontaktieren können. Vor allem in Köln und Wuppertal ist mehr als den durchschnittlichen 8% nicht bekannt, dass es einen HPC-Support gibt und wie dieser zu erreichen ist. Im Allgemeinen lässt sich aber festhalten, dass das Wissen über die Kontaktwege zum HPC-Support mit 87% konstant hoch geblieben ist. Vom Support Gebrauch gemacht ([Abb. 39](#)) haben bisher 69% der Befragten. Weitere 23% wissen, wie sie den Support erreichen können, haben ihn aber bisher nicht benötigt.

Die häufigsten Gründe, den Support zu kontaktieren sind, wie in [Abbildung 40](#) gezeigt, anwendungsspezifische Probleme (33%), Fragen zum Cluster-Zugang (28%) und Fragen zur Job Submittierung (24%), gefolgt von gewünschter Unterstützung bei Rechenzeitanträgen (18%) und Hilfe bei der Programmkompilierung (18%)¹. In Düsseldorf (59%), Duisburg-Essen (50%) und einigen

¹Mehrfachnennungen möglich.

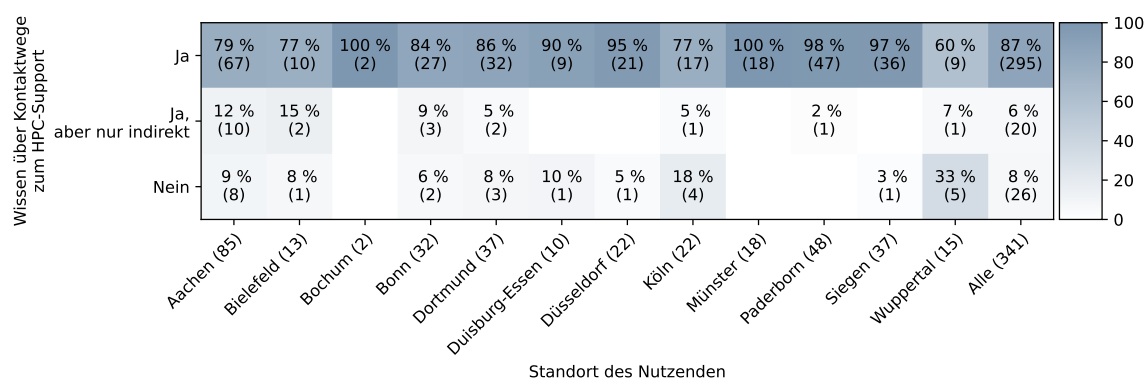


Abbildung 38: Anteile der Befragten, die wissen wie sie den lokalen HPC-Support kontaktieren können. Die Darstellung ist in Sektion 1 genauer beschrieben.

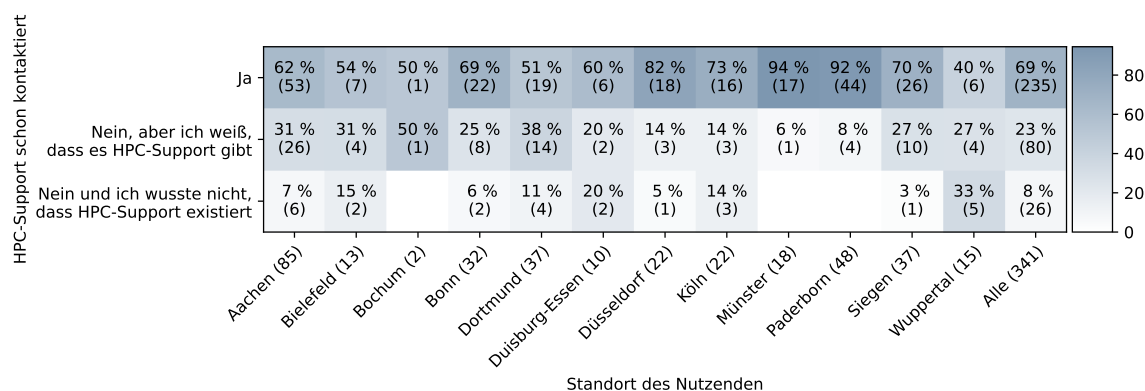


Abbildung 39: Anteile der Befragten, die den lokalen HPC-Support schon genutzt haben. Die Darstellung ist in Sektion 1 genauer beschrieben.

weiteren Standorten gibt es überdurchschnittlich viele Anfragen zu anwendungsspezifischen Problemen. In Paderborn hingegen hebt sich der Bedarf an Unterstützung für Rechenzeitanträge im Verhältnis zur Anzahl der Antworten hervor (38%), was letztlich darauf zurückgeführt werden kann, dass es sich um einen Tier-2 Standort handelt, der durch die Inbetriebnahme eines zweiten Großrechners seine Ressourcen massiv aufwerten konnte und somit neue Forschende anziehen konnte. Außerdem haben Bonn, Münster und Paderborn vermehrt Anfragen zum Cluster-Zugang (44%, 44% und 40%). Überdurchschnittlich viel Hilfe bei der Submittierung von Jobs wird in Münster (56%), Siegen (49%) und Köln (41%) benötigt. Das gesamte HPC.NRW Netzwerk betrachtend hat sich die Verteilung der Gründe für Kontakt zum HPC-Support gegenüber 2020 kaum verändert.

Derzeit haben 65% (Abb. 41) der Befragten keinen akuten Support-Bedarf. Eine Ausnahme stellt hier Paderborn dar, wo nur 23% der Nutzenden momentan keinen Bedarf für Support angeben. Der Vergleich mit dem Ergebnis von 2020 (70% hatten keinen Bedarf) lässt vermuten, dass auch hier das neue HPC-System zu einer größeren Auslastung des Supports führt.

46% der Befragten geben an, dass das bisher von HPC.NRW zur Verfügung gestellte Schulungsangebot vollkommen ausreicht (Abb. 42). Einen Ausbau des Angebots wünschen sich 54%. Hierbei wollen 28% allgemeine HPC Schulungen und 24% wünschen sich diese auf einem fortgeschrittenen Niveau. Anwendungsspezifische Hilfestellung, z.B. bei Code- und Performanceoptimierung, sollten laut 24% der Befragten ausgebaut werden und 21% sehen zusätzlichen Bedarf beim individuellen

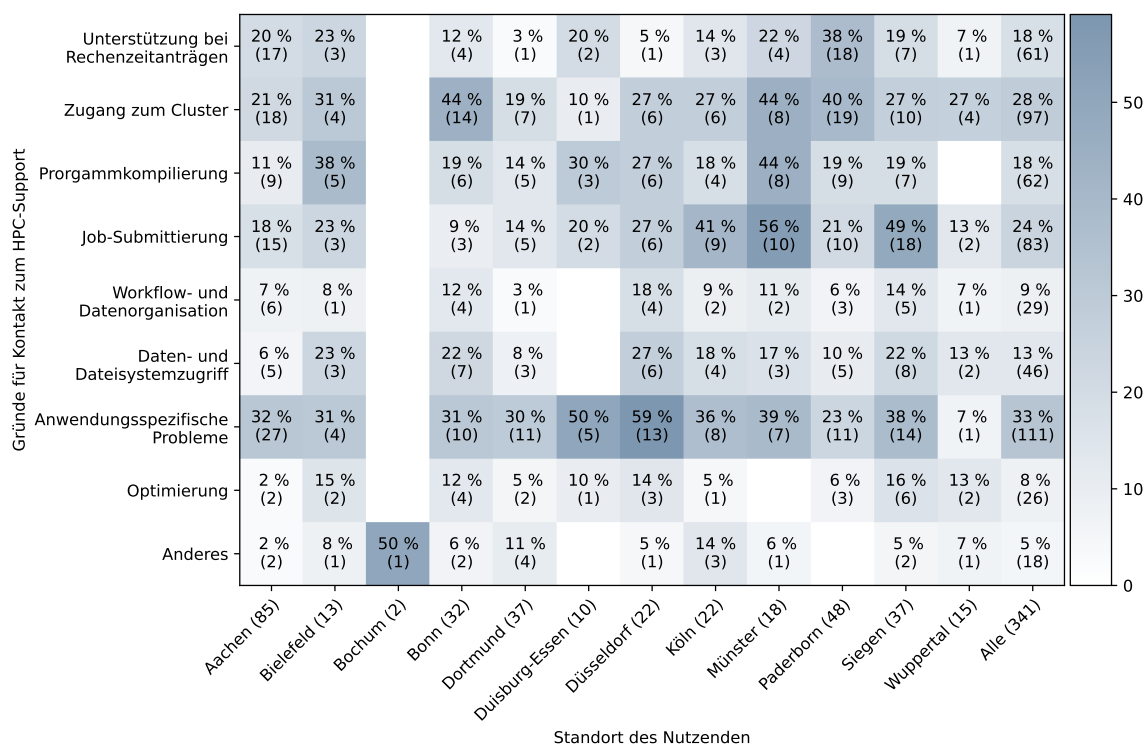


Abbildung 40: Gründe warum die Befragten den lokalen HPC-Support kontaktiert haben. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

Support bei HPC-Problemen. Auffällig hoch ist der Bedarf an fortgeschrittenem HPC-Training in Bielefeld (46%) und Wuppertal (47%). Insbesondere bei letzterem Bestand 2020 noch gar kein entsprechender Bedarf. Dies könnte auf Wuppertals erhöhte Teilnehmerzahl 2022 (von 4 hoch auf 15) zurückzuführen sein oder auf das von HPC.NRW zur Verfügung gestellte HPC-Kursangebot.

Die Auswertung der Freitext-Antworten zum Thema Support liefert verständlicherweise ein ähnliches Bild wie die Auswertung der Multiple-Choice Antworten. Es ist jedoch davon auszugehen, dass in Freitext ausgeführte Antworten ein besonderes Anliegen der Nutzenden wiedergeben. Von den gegebenen Freitext-Antworten wurden 85 der Kategorie Support zugeordnet. Das größte Anliegen ist mit 31% ein Ausbau der Tutorials und Workshops im Bereich der HPC-Grundlagen. Auch wenn Lob für die Arbeit von HPC.NRW geäußert wurde, besteht weiterhin der Bedarf nach (regelmäßigen) Veranstaltungen, welche die Grundlagen des Umgangs mit HPC-Systemen vermitteln. Die Erhebung zur verwendeten wissenschaftlichen Software ([Abschnitt 9](#)) hat zuvor gezeigt, dass die Nutzenden bereits mit mehr als nur einer Software arbeiten. Daher verwundert es nicht, dass sich einige der Nutzenden Rezepte für einzelne Anwendungsfälle wünschen, um den Mehraufwand für weitere Softwarekenntnisse zu reduzieren (z.B. für batch-Skripte). Als nächstes folgt mit 24% der Wunsch nach mehr Hilfe bei der Installation von Software. Oftmals läuft diese auf der eigenen Workstation, aber nicht auf dem HPC-System. Hier ist davon auszugehen, dass eine Korrelation zum Bedarf für Grundlagenveranstaltungen vorliegt. Grundsätzlich ist den Aussagen eine Dankbarkeit für den verfügbaren Support zu entnehmen, jedoch gibt es wiederholt Anfragen nach schnelleren Rückmeldungen, einem Ausbau des Angebotes und einer generellen Aufstockung des Personals.

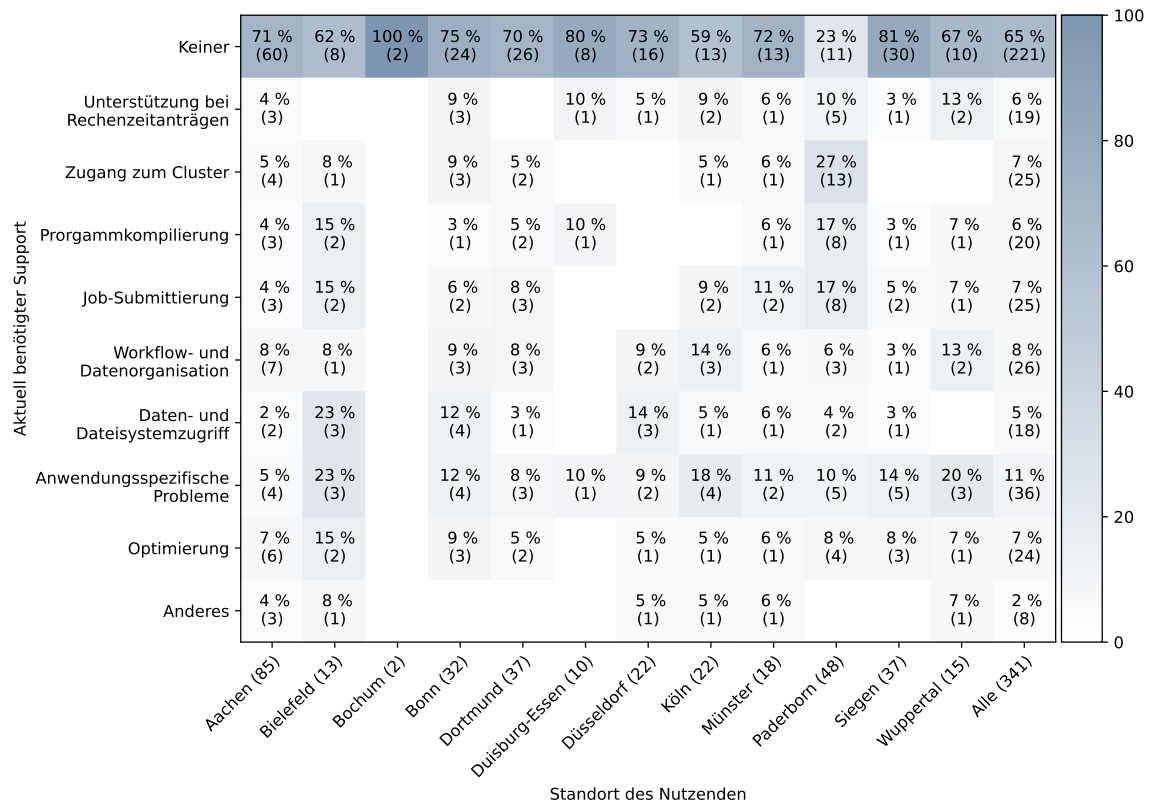


Abbildung 41: Zur Zeit benötigte Hilfestellungen der Befragten. Die Darstellung ist in Sektion 1 genauer beschrieben.

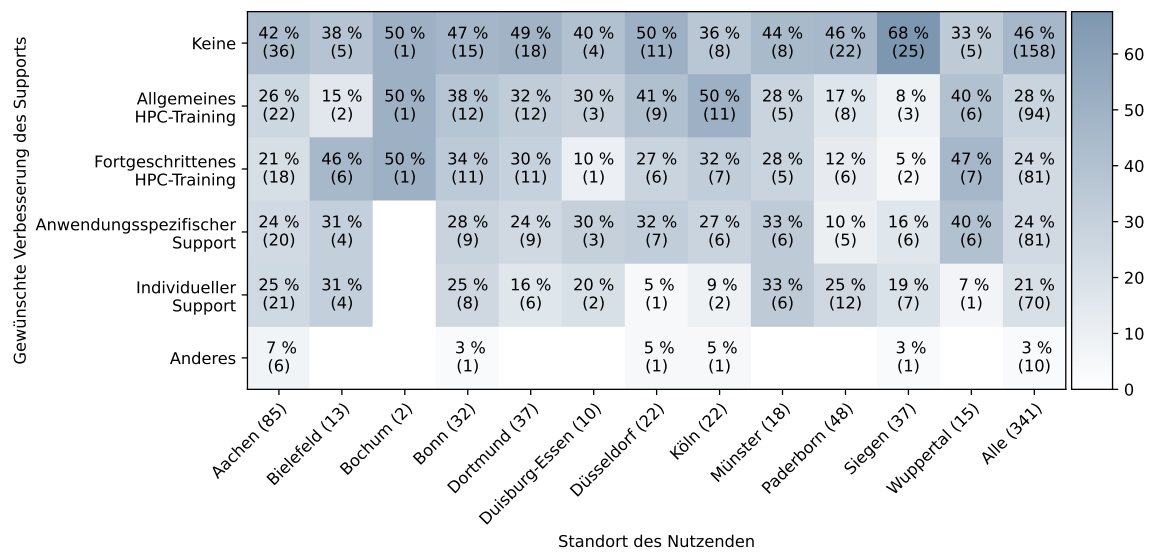


Abbildung 42: Von Befragten gewünschte Verbesserungen/Erweiterungen im lokalen HPC-Support. Die Darstellung ist in Sektion 1 genauer beschrieben.

12 Schulungen

Aus der Umfrage (Abb. 43) wird ersichtlich, dass 31% der Befragten bisher an einem HPC-Kurs oder einem Tutorial teilgenommen haben. Mangelnde Informationen über angebotene Kurse sind der Hauptgrund, warum von den Befragten noch keine Schulung besucht wurde. Als weitere Gründe werden mangelndes lokales Kursangebot, COVID-19 Pandemie, der Besuch von Kursen an anderen Institutionen, ein unzureichendes Kursangebot (zu schnell ausgebucht, nur auf deutsch, Kurs findet nur einmal im Jahr statt) und Unwissen über stattfindende Kurse aufgeführt.

Von den 122 Nutzenden, die Freitextantworten gegeben haben, wurde am häufigsten angegeben, bisher an Einführungskursen (Cluster oder Linux), MPI und OpenMP Kursen und an Veranstaltungen zur Performance Analyse teilgenommen zu haben. Ebenfalls häufiger genannt wurden (in dieser Reihenfolge) Kurse zur folgenden Themen: Parallele Programmierung, CUDA/GPU und verschiedenen Programmiersprachen (C++, Python, Julia).

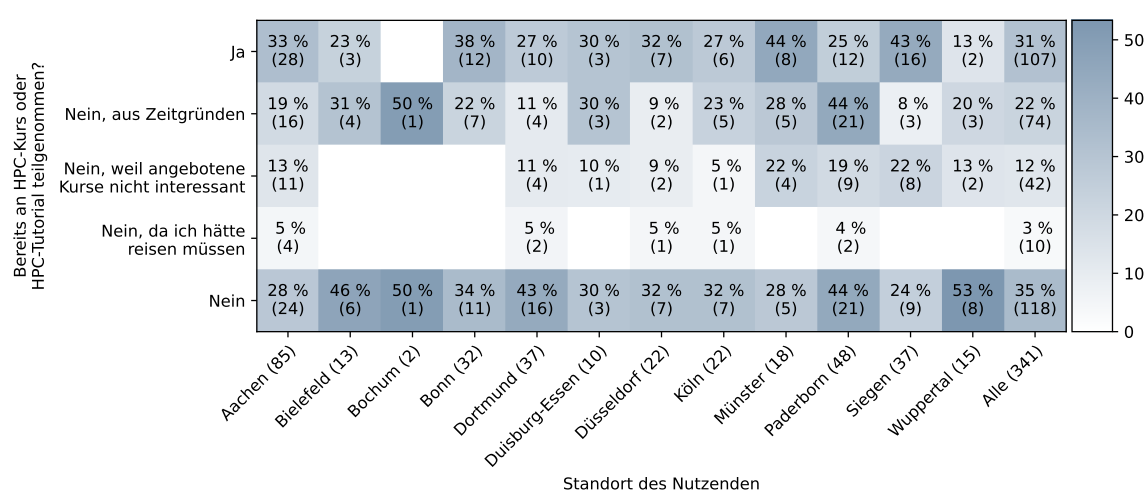


Abbildung 43: Anteile der Befragten, die bereits an einem HPC-Kurs/Tutorial teilgenommen haben inkl. Gründe falls noch nicht. Die Darstellung ist in Sektion 1 genauer beschrieben.

Die Präferenz der Befragten für das Format einer Weiterbildung ist in Abbildung 44 ausgewertet. Dominant sind dabei kurze Onlinetutorials mit kurzen Videos, Texten und interaktiven Elementen wie sie auch bereits im Arbeitspaket 2 entwickelt wurden und werden.

Die Reisebereitschaft der Befragten für HPC-Kurse ist leider weiterhin recht gering (Abb. 45). Bereit zu Reisen, um einen Kurs an einem anderen Standort innerhalb NRWs zu besuchen, sind 19% der Befragten. Für weitere 47% kommt das nur für besonders interessante Kurse in Betracht. Dagegen würden 26% nur lokale Kurse besuchen und 6% gaben keine Antwort. Daraus ergibt sich die Notwendigkeit viele Kurse lokal anzubieten bzw. die virtuellen Angebote weiter auszubauen.

Die Nachfrage nach Schulungen ist groß. Durchschnittlich 87% der Befragten interessieren sich für Grundlagen- und/oder Fortgeschrittenenkurse.

Als beliebte Basis-Themen wurden Einführungen ins Profiling/HPC-Code-Optimierung (48%), Softwarecontainer (41%), effizienten Datenspeicherung (36%), Debugging (36%) und Linux-Skripting (35%) angegeben. Weitere Antworten sind in Abbildung 46 dargestellt.

Bei den fortgeschrittenen Kursen gaben die Befragten an, dass Techniken zur Code-Optimierung (47%), zum parallelen Debugging (43%), zum parallelen Profiling (40%) sowie zu Kompilern und Bibliotheken (38%) vorrangig interessant sind (siehe Abbildung 47). Kein Interesse an fortgeschrittenen Kursen wird lediglich von 26% bekundet. Der Bedarf ist jedoch nicht gänzlich homogen über die Standorte verteilt. So gibt es beispielsweise in Bielefeld einen überdurchschnittlich hohen Bedarf

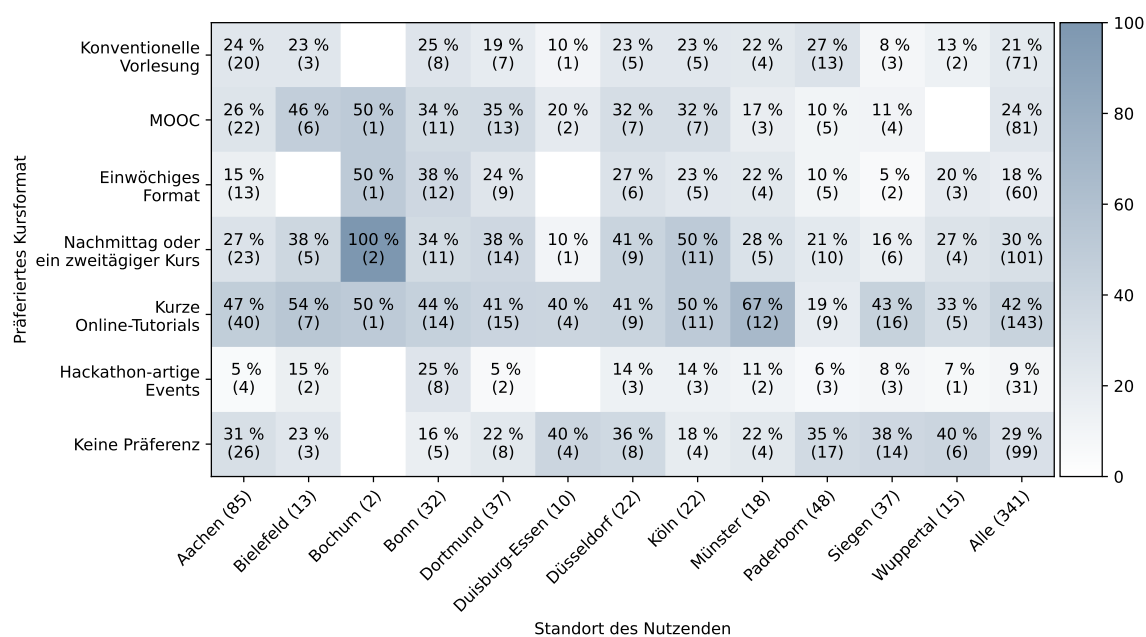


Abbildung 44: Präferenz der Befragten für verschiedene Kursformate. Die Darstellung ist in Sektion 1 genauer beschrieben.

nach parallelem Debugging und parallelem Profiling (je 69%) und in Köln wird häufig (68%) der Wunsch nach entsprechenden Kursen zu Compilern und Bibliotheken geäußert.

Das Interesse an Programmiergrundlagenkursen ist für alle vorgegebenen Sprachen/Konzepte mit >15% durchweg hoch. Die Antworten der Befragten verteilen sich wie folgt auf die Themen: MPI (38%), GPU-Programmierung mit CUDA (35%), C/C++ (30%), OpenMP (27%) und Python (25%). Kein Interesse an Grundlagenkursen haben 27% der Befragten (siehe Abbildung 48). Vergleichend mit den Umfragewerten von 2020 kann hervorgehoben werden, dass nun an 5 Standorten mindestens jede zweite Person Grundlagenkurse zum Thema MPI wünscht (zuvor war dies nur in Duisburg-Essen der Fall). Ansonsten ist keine signifikante Änderung gegenüber den Ergebnissen von 2020 festzustellen.

Etwas deutlicher heben sich bei den fortgeschrittenen Programmierkursen die Themen Python (51%) und C/C++ (39%) ab. Aber auch GPU-Optimierung (31%), MPI (31%) und OpenMP (22%) würden mehrere Kurse füllen.

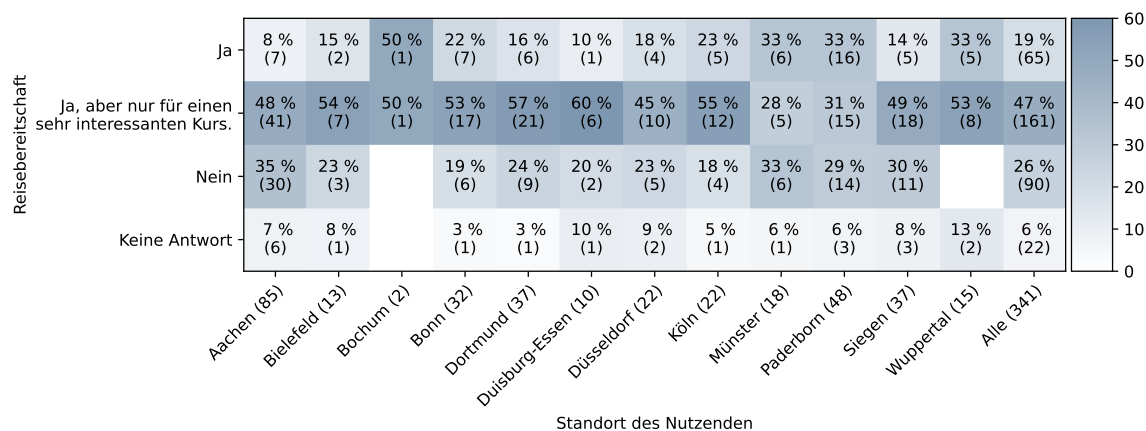


Abbildung 45: Reisebereitschaft der Befragten. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

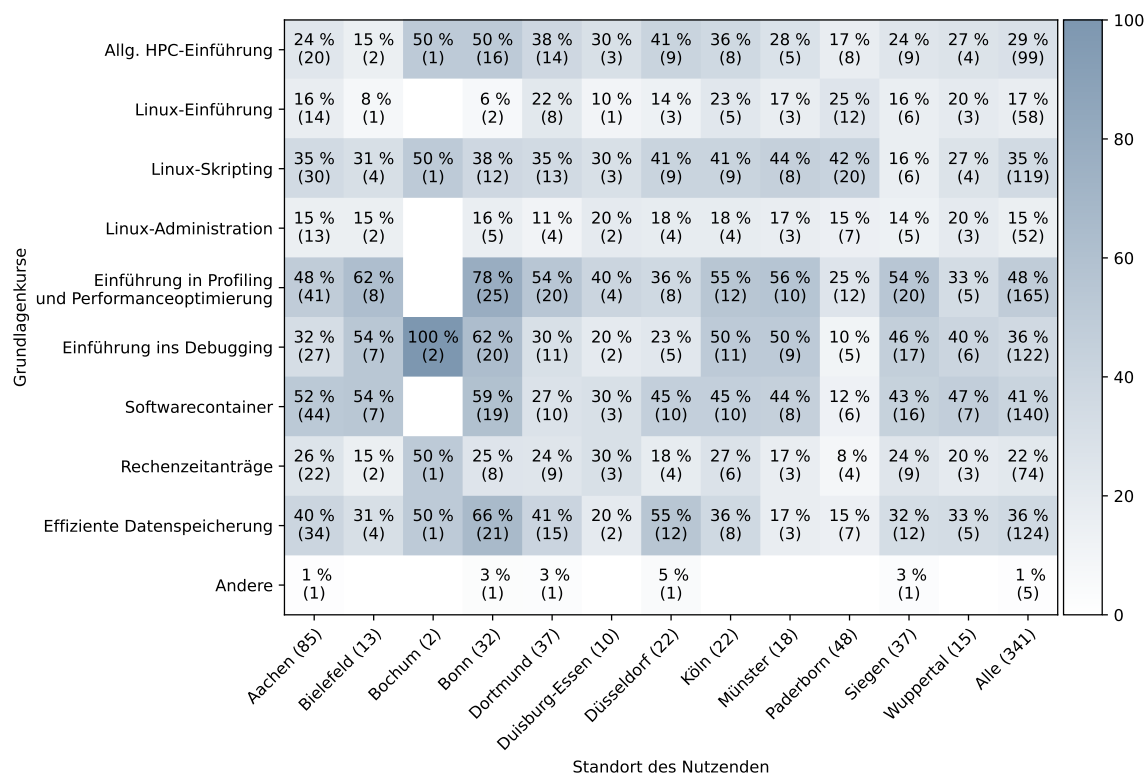


Abbildung 46: Interesse der Befragten an Grundlagenkursen. Die Darstellung ist in [Sektion 1](#) genauer beschrieben.

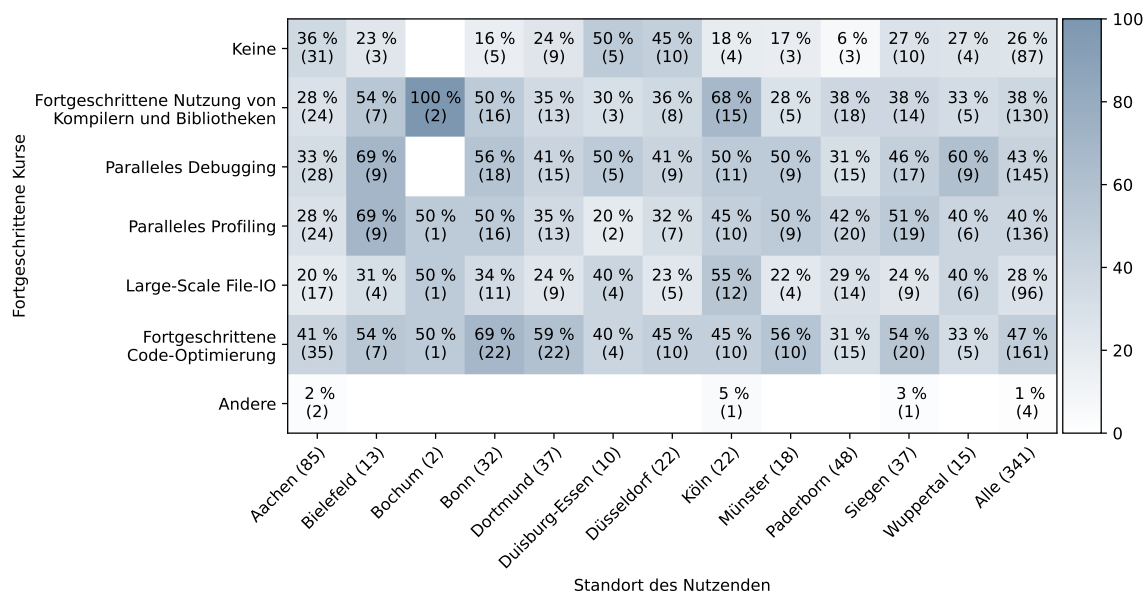


Abbildung 47: Interesse der Befragten an fortgeschrittenen Kursen. Die Darstellung ist in Sektion 1 genauer beschrieben.

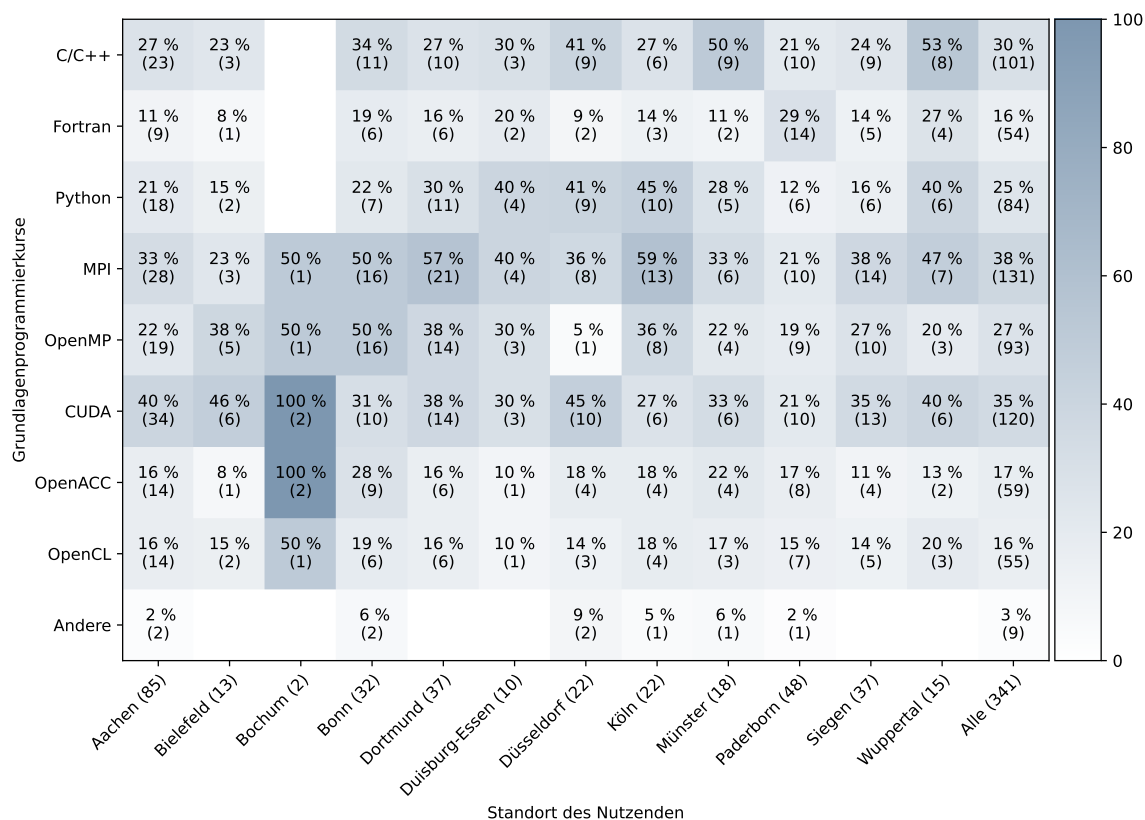


Abbildung 48: Interesse der Befragten an Grundlagen-Programmierkursen. Die Darstellung ist in Sektion 1 genauer beschrieben.

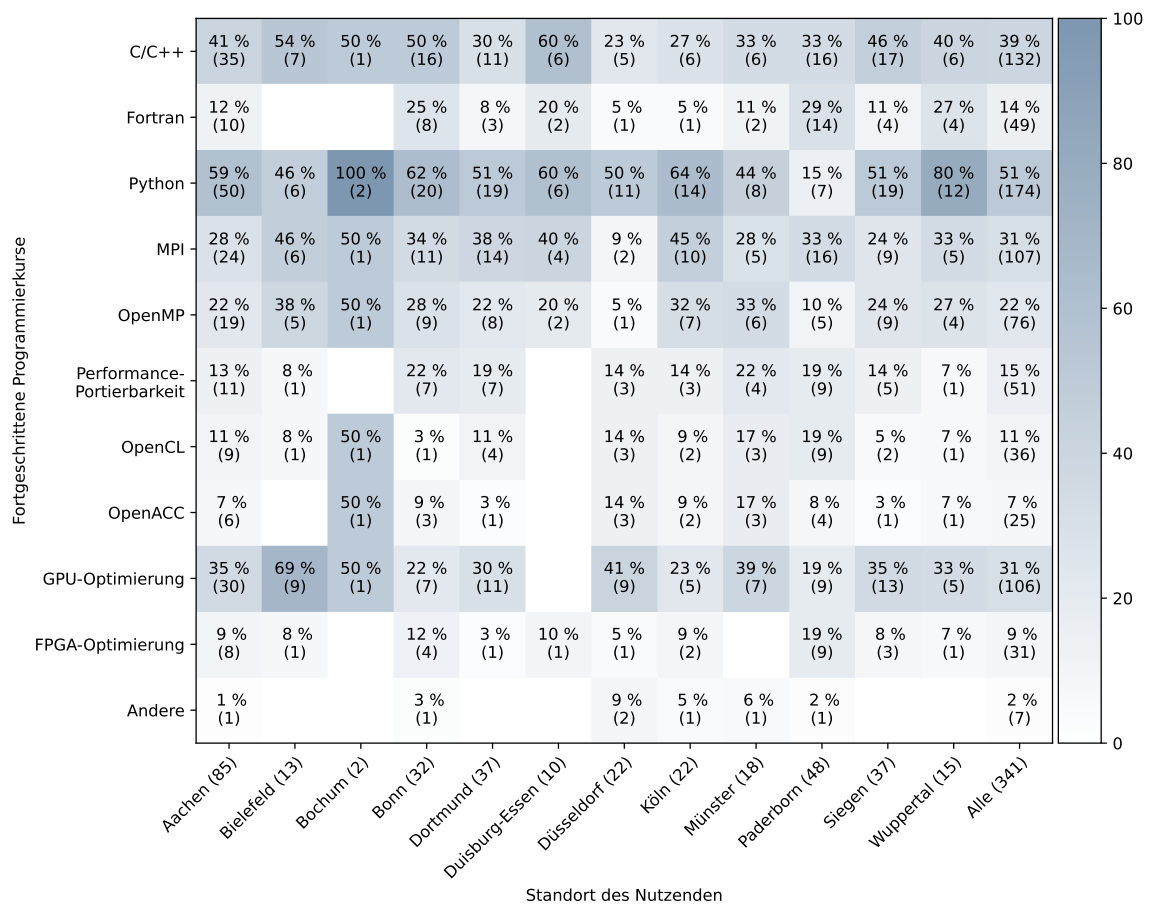


Abbildung 49: Interesse der Befragten an fortgeschrittenen Programmierkursen. Die Darstellung ist in Sektion 1 genauer beschrieben.

13 Erkenntnisse und Maßnahmen einzelner Standorte

Die Daten der Auswertung wurden bereits von allen Standorten genutzt, um eigene Auswertungen durchzuführen und zukünftige Maßnahmen daraus abzuleiten. Diese wurden von den Standorten zusammengefasst und werden in den folgenden Abschnitten unbearbeitet wiedergegeben.

13.1 Standortsspezifische Auswertung - Aachen

Von den 85 teilnehmenden Nutzenden vom Standort Aachen gaben 45 an den CLuster der Universität (Tier-3) zu nutzen und 53 gaben an das Tier-2 System zu nutzen. Beide Rechenressourcen werden durch verschiedene Teile des Aachener Clusters (CLAIX) abgebildet. Signifikante Änderungen gegenüber den Ergebnissen der Umfrage von 2020 sind hier nicht zu erkennen. Die Befragten kommen wie schon 2020 zum Größten Teil in den Bereichen “Informatik, System- und Elektrotechnik” (34 %), “Physik” (20 %), “Materialwissenschaften und Ingenieurwesen” (13 %) und “Maschinenbau und Wirtschaftsingenieurwesen” (11 %). Im Vergleich zur Umfrage in 2020 gaben in 2022 weniger Teilnehmer “Chemie” (8 % [12 % in 2020]) als primäres Forschungsgebiet an.

Von den teilnehmenden Nutzern geben 51 einen Master/Diplom Abschluss und 18 einen Dr./Ph.D. als höchsten akademischen Grad an.

Weiterhin entwickeln viele der Befragten Programme selbst. So geben zwei Drittel der Befragten an, Programme selbst zu entwickeln und weitere 7 Prozent dies zu planen. Dies weist den zur Verfügung stehenden Compiler- und Entwicklungsumgebungen auch weiterhin eine erhebliche Wichtigkeit zu. Bei der Programmierung geben meist mehr als die Hälfte der Befragten an, sowohl komplette Programme zu entwickeln, aber auch bestehende Programme zu modifizieren und zu erweitern. Auch Fehlerbehebung und Performanceoptimierung stehen oft im Fokus und unterstreichen hier die Wichtigkeit von Softwareentwicklungswerkzeugen wie Debuggern und Performance Analyse Werkzeugen. Programmierkenntnisse sind weiterhin hoch. So geben mehr als die Hälfte der Befragten an Programmierkenntnisse in C/C++, Matlab, oder Python zu haben. Aber auch für Fortran, Julia, Mathematica, Java und R gibt es Kenntnisse bei einem signifikanten Anteil der Befragten. Bei den Wünschen der Befragten nach Weiterbildungsmöglichkeiten in der Programmierung—sowohl für Anfänger als auch Fortgeschrittene—ist GPU-basierte Programmierung sehr gefragt. Aber auch klassische Themen wie MPI und OpenMP, werden von bis zu einem Drittel der Befragten gewünscht. Bei den Fortgeschrittenen Themen stehen außerdem C/C++ und Python Programmierung hervor, beides Themen die öfter in moderneren Simulationscodes Verwendung finden. Die Nachfrage insbesondere nach Kursen für Fortgeschrittene Themen zeigt den Wunsch der Befragten die eigenen Programmierfähigkeiten über ein Grundmaß hinaus zu entwickeln.

Bei den getätigten Jobs der Befragten dominiert die Gruppe der Nutzer, die einzelne Rechenknoten mit mehreren Kernen nutzen, was einen aktuellen auf Thread-parallele Programmiermodelle nahelegt. Dennoch lassen mehr als die Hälfte der Befragten auch regelmäßig Jobs mit mehreren Rechenknoten laufen, was wiederum die Wichtigkeit der Programmiermodelle mit verteiltem Speicher unterstreicht. Ein besonderer Augenmerk könnte hier darauf gelegt werden hybride Programmierung zu unterstützen, um die Nutzenden dabei zu unterstützen, die Schwelle zu mehreren Rechenknoten erfolgreich zu überschreiten. Dies passt auch zu den von den Befragten angegebenen genutzten Programmiermodellen, wo fast die Hälfte der Befragten MPI einsetzen und 39% der Befragten OpenMP. Desweiteren ist CUDA als drittes Programmiermodell unter 29% der Befragten sehr beliebt. Und insgesamt 42% der Befragten nutzen NVIDIA GPUs, was den hohen Prozentsatz an direkter CUDA Nutzung erklärt. Immerhin 48% der Befragten geben an, dass sie nicht einschätzen können, ob die Nutzung von mehreren Beschleunigern ihrer Applikation nutzen würde. Hier könnte dedizierte Beratung mehr Klarheit bei den Nutzern schaffen. MPI und OpenMP werden in Aachen stark durch das Kursprogramm unterstützt. Bei GPU Programmierung wurde in der Vergangenheit eher auf Portabilität mittels OpenACC gesetzt, eine direktere Unterstützung von CUDA könnte insgesamt hier die Nutzer ebenfalls sinnvoll unterstützen.

Bei Fragen des Speicherverbrauchs pro Kern lässt sich erkennen, dass knapp die Hälfte der Befragten mit dem durchschnittlichen Speicher von 4 GB pro Kern auskommen. Die restlichen 50% der

Befragten nutzen mehr als 4 GB Speicher pro Kern. Somit bieten die 4 GB pro Kern des CLAIX System einen guten Kompromiss beide Gruppen unter den Befragten zu unterstützen. Dennoch geben fast 40% der Befragten an, dass die Größe des Arbeitsspeichers das Bottleneck ihrer Applikation darstellt. Nutzer mit solchen Anforderungen besonders bei der Wahl der Rechenressourcen unterstützt werden.

Immerhin untersuchen mehr als die Hälfte der die Leistung ihrer Applikationen, davon aber die meisten nur nach elementaren Metriken. Hier zeigt sich somit Potential, an der Leistung ihrer Applikationen interessierten Nutzern den Gebrauch von entsprechenden Tools nahe zu bringen. Hier bietet sich z.B. der etablierte und jährlich veranstaltete aiXcelerate Workshop an.

In Bezug auf externe Bibliotheken sind in dieser Umfrage zum ersten Mal Machine-Learning Frameworks abgefragt worden, und knapp 30% der Nutzer geben an diese zu nutzen. Dies unterstreicht die wachsende Wichtigkeit dieses Bereiches für den Bereich Schulungen und Support. Insbesondere ML Software wird in Aachen über Container bereit gestellt und mit insgesamt fast zwei Dritteln der Teilnehmenden besteht ein großes Interesse an Containern. Der Anteil der Teilnehmenden, die Jupyter-Notebooks auf HPC Systemen ist im vergangen Jahr minimal angestiegen und liegt stabil bei 10%. Dies erweitert sich auf 26% der Teilnehmenden wenn man die zukünftige Nutzung mit einschließt. Beide Technologien (Container und Jupyter-Notebooks) werden an der RWTH Aachen University in Zukunft weiter ausgebaut.

Im Bereich Nutzendenunterstützung geben immerhin 93% der Befragten aus Aachen an zu wissen, dass es entsprechenden Nutzendensupport gibt und wie er erreichbar ist. Dennoch wissen fast 10% der Befragten nicht, wie der HPC Support zu erreichen ist. Dies könnte durch stärkere Bewerbung der Kontaktwege verbessert werden. Die Hauptinformationsquellen zum Fragen rund um HPC sind, neben dem lokalen Support, die lokale HPC Webseiten und das HPC Wiki. Dies unterstreicht, dass die Wichtigkeit des Ausbaus und der Vernetzung dieser Quellen. Fast die Hälfte der Befragten aus Aachen wünscht sich anwendungsspezifischen oder individuellen Support. An dieser Stelle könnte das Angebot der Nutzendenunterstützung ausgeweitet werden. Fast die Hälfte der Befragten aus Aachen wünscht sich kurze Online-Tutorials, was die Wichtigkeit dieses Formates unterstreicht.

13.2 Standortspezifische Auswertung - Bielefeld

In Bielefeld haben 13 Nutzer an der Befragung teilgenommen (19 im Jahr 2020). Die Zusammensetzung der Nutzer ist gegenüber der Umfrage von 2020 relativ unverändert, 68% der Nutzer sind Physiker, der Rest verteilt sich auf die Fachgebiete Informatik (23%) und Chemie (8%). Leider gibt es einen Rückgang der studentischen Nutzer von 9 auf 3 zu verzeichnen (möglicherweise haben mehrere studentische Nutzer dieses Jahr nicht an der Umfrage teilgenommen).

Generell handelt es sich in Bielefeld zum Großteil um erfahrende Nutzer, da die Hälfte (46%) der Nutzer auch den Tier-2 Standort Paderborn verwendet und viele darüber hinaus auch Tier-1 Cluster verwenden. Dies äußert sich auch in den Programmierfähigkeiten der Nutzer: über 80% geben an, ihren Code selbst zu kompilieren, und 77% entwickeln ihn auch selbst. Weiterhin geben 69% der Nutzer an, dass sie fortgeschrittene Kenntnisse in C/C++ besitzen. Dies ist mit Abstand die höchste Zahl unter allen Standorten.

Die spezielle Architektur des Bielefelder Clusters (Fokus auf GPUs mit 8 NVIDIA V100 Karten pro Knoten) bedingt auch, dass 46% der Nutzer in CUDA entwickeln. Dementsprechend ist der meistgewünschte HPC-Kurs ein GPU-Optimierungskurs. Weiterhin sind die meisten HPC-Rechnungen (54%) mit CUDA parallelisiert, jedoch viele auch mit MPI (54%) und openMP (46%).

Alles in allem handelt es sich in Bielefeld um sehr spezielle Hardware, die stark auf GPUs fokussiert ist. Dies spiegelt sich in den Kenntnissen und Wünschen der Nutzer wider. Ermöglicht aber auch in zunehmendem Maße Anwendungen, die auf maschinellem Lernen basieren und die einschlägigen Bibliotheken PyTorch und TensorFlow nutzen. Nutzer, die keine GPUs verwenden können, werden von unserem Support an das Tier-2 Zentrum Paderborn verwiesen.

13.3 Standortspezifische Auswertung - Bochum

In Bochum befindet sich der Aufbau zentraler Tier-3 Infrastruktur derzeit in der Ausschreibungsphase. Nur zwei Nutzende der vielfältigen dezentralen HPC Ressourcen in Bochum haben an der Umfrage teilgenommen, diesmal aus den in der Umfrage 2020 nicht vertretenen Disziplinen Physik und Maschinenbau.

Ähnlich wie in den in der vorherigen Umfrage vertretenen Feldern Medizin, Mathematik, Geisteswissenschaften, Ingenieurwesen und Informatik wird auch in Physik und Maschinenbau in Bochum die HPC Software weitgehend selbst programmiert und gepflegt. Entsprechend werden auch weiterhin Informationen und Schulungen in den Bereichen MPI, OpenMP und GPU Programmierung mit verschiedenen Schnittstellen nachgefragt, wobei derzeit MPI und OpenMP zum Einsatz kommen und der Einsatz von GPUs geplant wird. Es kommen derzeit viele CPU Knoten zum Einsatz, weshalb teils auch bereits Tier-2 Ressourcen in Aachen genutzt werden.

Es wird erwartet, dass auch die in der diesjährigen Umfrage beschriebenen Ressourcenbedarfe (CPU Knoten, Arbeitsspeicher) vom zukünftigen Tier-3 Rechencluster mit heterogener Architektur (inklusive Fat-Knoten mit besonders viel Arbeitsspeicher sowie GPU-Knoten) abgedeckt werden können, wobei insbesondere bei den in der aktuellen Umfrage genannten Bottlenecks Latenz und File-IO Verbesserungen zu erwarten sind.

Zur Performance-Analyse und zum Debugging kommen bisher nur einfache Tools zum Einsatz; ein ähnliches Bild ergab die Umfrage 2020. Informationsangebote in diesem Themenfeld erscheinen daher weiterhin sinnvoll und werden von den Nutzenden auch gewünscht.

Im Unterschied zur Umfrage 2020 und im Einklang mit der Entwicklung an anderen Standorten wächst in Bochum das Interesse an Alternativen zum Clusterzugang via SSH, insbesondere via Jupyter Notebooks.

Der lokale HPC Support ist den Nutzenden bekannt, allerdings muss die Sichtbarkeit von HPC.NRW Ressourcen in Bochum noch erhöht werden. Dazu wurde bereits eine zentrale Webseite aufgebaut.

Zusammenfassend steht in Bochum im kommenden Jahr neben Beschaffung, Aufbau und Inbetriebnahme des Tier-3 Clusters vor allem der Ausbau des Schulungs- und Informationsangebots im Zentrum.

13.4 Standortspezifische Auswertung - Bonn

Am Standort Bonn haben sich insgesamt 32 Nutzende an der Umfrage beteiligt, was einer sehr ähnlichen Beteiligung im Vergleich zur letzten Iteration der Nutzendenumfrage von 2020 mit 31 Antworten entspricht. Auffällig ist jedoch eine stark veränderte Zusammensetzung der vertretenen Fachbereiche von Antworten, die primär aus der Physik stammen (50%), hin zu einer diverseren Landschaft, wobei die Geowissenschaften am stärksten vertreten (31%) sind, gefolgt von Biologie (19%), Physik (16%) und Chemie (12%). Ebenso wurden etwas mehr Antworten von Nutzenden mit Bachelor- und Masterabschlüssen abgegeben, mit einem gleichzeitigen leichten Rückgang an Antworten von Promovierten etc.

Der lokale Cluster, Bonna, wird mittlerweile von 3 von 4 Nutzenden verwendet, wobei eine deutliche Steigerung im Vergleich zu 2020, wo es nur jeder zweite Nutzende war, zu beobachten ist. Mögliche Gründe dafür könnten in der steigenden Bekanntheit der zentralen Ressourcen am Standort Bonn sowie eines Relaunches von Bonna Anfang 2022 liegen. Scheinbar wird der Cluster dabei als zusätzliche Ressource verwendet, da die Nutzung von Arbeitsgruppenrechnern sowie von Tier-1- und Tier-2-Rechnern weiterhin in einer ähnlichen Größenordnung angegeben wird. Der beobachtete Anstieg bei den Bedarfen in Bezug auf Hauptspeicher pro Kern und des Festplattenspeichers wird sich mit der Anschaffung eines neuen zentralen Tier-3-Rechners aller Voraussicht nach gut abdecken lassen. Des Weiteren ist die Verwendung von Beschleunigern immer noch eine Seltenheit, was eventuell auch an (noch) mangelnden zentralen Kapazitäten in diesem Bereich liegen könnte.

Im Bezug auf das Durchführen von Performanceanalysen und Debugging ist bei den Nutzenden ein leichter absteigender Trend zu beobachten, welcher allerdings auch entweder ein rein statistischer

Effekt sein könnte oder an der Zusammensetzung der Antworten aus den unterschiedlichen Fachbereichen liegen könnte. Trotz diesen Trends ist weiterhin ein ähnliches Interesse an Schulungen zu diesen Themenbereichen zu beobachten bzw. sogar eine deutliche Steigerung an Einführungskursen zum Thema Performance-Profiling (ca. jede/r zweite/r in 2020, jetzt über 3 von 4), was positiv zu bewerten ist. Bei den identifizierten Flaschenhälsen in Nutzendenanwendungen haben der Hauptspeicher pro Kern (30% → 50%) und der Datendurchsatz (30% → 41%) an Bedeutung zugenommen.

Im Supportbereich konnten die Möglichkeiten zur direkten Kontaktaufnahme etwas besser vermittelt werden (74% → 84%). Je ungefähr ein Drittel wünscht sich weiterhin ein wachsendes Angebot an Schulungen und Kursen, welches in Zukunft durch ein wachsendes lokales Team umgesetzt werden wird. Als Themenbereich für Einstiegskurse konnte unter anderem die Komplexe Performance-Profiling (48% → 78%) und MPI (16% → 50%) an Interesse gewinnen. Es ist schwierig zu beurteilen, ob diese Entwicklung einem generell gestiegenen Interesse an HPC-Themen zu Grunde liegt oder eventuell an einer unterschiedlichen Zusammensetzung der Antworten aus den verschiedenen Fachbereichen.

Als lokale Maßnahmen am Standort Bonn ergeben sich aus der Nutzendenumfrage Impulse für neue Schulungen, (Online-)Kurse und Tutorials. Ein Hauptaugenmerk könnte dabei auf die Vermittlung von Techniken und der Bedeutung von Performanceanalysen und parallelem Debugging fallen. Außerdem werden bestehende Bestrebungen, die Bekanntheit lokaler Supportinfrastrukturen und zentraler Ressourcen zu steigern, weiter verfolgt und intensiviert. Bestrebungen, die Kontaktaufnahme möglichst niederschwellig zu gestalten, werden untersucht und soweit möglich umgesetzt. Geäußerte Bedarfe in der Befragung werden sich voraussichtlich mit dem neuen zentralen Tier-3-System erfüllen lassen.

13.5 Standortsspezifische Auswertung - Dortmund

37 Nutzende des LiDO3-Clusters der TU Dortmund haben vollständig an der Umfrage teilgenommen. Dies entspricht etwa 35% der aktiven Nutzenden. Das Feedback der Befragten war überwiegend positiv.

Zirka die Hälfte der Nutzenden führt Rechnungen auf nur einem Knoten pro Job aus. Dieses Ergebnis lässt sich qualitativ auch durch Accountingdaten bestätigen.

Etwa 40% der Nutzer wünschen sich mehr Arbeitsspeicher pro Rechenknoten. Der neue LiDO4-Cluster wird dieser Anforderung Rechnung tragen, den höheren Bedarf nach GPU-Beschleunigern berücksichtigen und Unterstützung für Jupyter und Container mitbringen.

Des Weiteren wurde vielfach der Wunsch nach Schulungen in den Bereichen Grundlagen, MPI, Multithreading und Performanceanalyse geäußert, da ein Großteil der Nutzenden keine oder nur sehr rudimentäre Erfahrungen damit haben aber sehr viele ihre eigene Software für den Cluster programmieren.

Abschließend wird das Thema Forschungsdatenmanagement auch im Bereich HPC immer wichtiger und für die Nutzenden praktisch relevant.

13.6 Standortsspezifische Auswertung - Düsseldorf

Obwohl auf dem Düsseldorfer HPC-Cluster "Hilbert" pro Jahr drei- bis vierhundert Forschende aktiv sind, haben lediglich 22 von ihnen die Nutzendenumfrage vollständig ausgefüllt. Die Hälfte dieser Befragten kommt aus dem Gebiet der Biologie, gefolgt von Informatik, Physik und nicht zuletzt auch aus den Bereichen Chemie, Medizin und Sozialwissenschaften. Im Gegensatz zur letzten HPC Nutzendenumfrage, wo die meisten Befragten einen Master- oder Dokortitel hatten, sieht man nun mit 27 % Bachelor- und 36 % Masterstudenten eine Verschiebung seitens der Nutzenden, bei dem der Einstieg ins HPC bereits viel früher in deren wissenschaftlicher Karrierelaufbahn erfolgt. Dies zeigt unseren Erfolg, dass wir den Zugang zum HPC noch niederschwelliger gestalten konnten, was vom Projekt HPC.NRW auch explizit gewünscht wurde.

Bis auf wenige Ausnahmen, z. B. Professuren nach dem Jülicher Modell - die dadurch sowohl in Düsseldorf als auch in Jülich die HPC Systeme nutzen, rechnen fast alle Befragten ausschließlich

auf dem lokalen HPC System in Düsseldorf und nutzen darüber hinaus kein weiteres HPC System (z. B. Tier-2). Zusammen mit Aachen nutzen die HPC-Nutzenden in Düsseldorf am wenigsten Arbeitsgruppencluster (also sogenannte Tier-4 Systeme). Dies ist sehr positiv zu bewerten, da dieser Trend “weg von Tier-4” explizit vom neuen HPC-Landeskonzept gewollt ist. Damit gehören wir hier zu den Vorreitern und stehen damit für die Zukunft schon sehr gut da.

Für 79 % aller HPC Nutzenden ist HPC von größter Wichtigkeit für ihre Arbeit und bei 20 % der weiteren Befragten unterstützt HPC deren Arbeit zumindest in hohem Maße. Dies unterstreicht noch einmal ganz deutlich die Wichtigkeit von HPC in Forschung und Lehre.

Mittlerweile kompilieren und entwickeln mehr als die Hälfte der Befragten ihre Anwendungen (auch) selber. Dies ist eine deutliche Steigerung gegenüber der letzten Umfrage, wo davon nur etwa ein Drittel der Nutzenden Gebrauch machten. Ein Grund für diese positive Entwicklung ist sicher auch, dass dafür extra spezielle Kurse und Schulungen (z. B. der zweiteilige Fortgeschrittenenkurs “Compiling and installing tools for HPC”) in Düsseldorf angeboten wurden.

Der Trend geht eindeutig weg von sequentiellen Codes. Das heißt, die meisten Jobs in Düsseldorf sind bereits parallelisiert und nutzen zwischen 5 und 40 CPUs pro Job. Auch hier hat die Beratung und die Optimierung von Codes sicher entscheidend zu dieser lobenswerten Entwicklung beigetragen. So wurden in Intensivberatungen bspw. selbst Lösungen zur Parallelisierung von Anwendungen in eher HPC-untypischen Sprachen wie “R” entwickelt. Diejenigen, die hingegen nicht selbst entwickeln, sondern einfach “fertige” Anwendungen nutzen, wissen oft nicht, welche Parallelisierungstechnik sich dahinter verbirgt. Ein Großteil unserer Nutzer hat auch Interesse an der gleichzeitigen Nutzung von mehreren GPUs. Dies wird durch unsere GPU-Knoten mit je 8 oder 10 GPUs bereits ermöglicht, aber oft können die Nutzenden deren Verwendbarkeit nicht richtig einschätzen. Leider machen die meisten Nutzenden keine Performanceanalysen ihrer Anwendung(en) und begründen dies mit fehlendem Wissen. Genau an diesen Stellen könnten zukünftige Schulungen und Kurse anknüpfen und damit Wissenslücken in diesen Bereichen für die Nutzenden schließen. Darüber hinaus wünschen sich viele HPC Nutzende auch generelle (i. S. v. nicht direkt HPC-relevante) Kurse zu Python, aber auch zur effizienten Datenspeicherung.

Als größtes Bottleneck für die HPC Anwendungen wird am häufigsten die Größe des verfügbaren Hauptspeichers angegeben. Dieser Punkt wurde bereits vorhergesehen und in der gerade erst erfolgten Beschaffung gut adressiert. Dabei wurden die alten “Ivybridge”-Knoten mit je nur 128 GB RAM durch neue “Icelake”-Knoten mit jeweils 256 GB RAM ersetzt. Beim benötigten Festplattenspeicher pro Job gehört Düsseldorf - bedingt durch die Haupt-Fachrichtungen - mit zu den Spitzenreitern in NRW. Dies wird aktuell in einem Folgeantrag für einen Storage-Nachfolger mit berücksichtigt.

KI-Initiativen wie “KI für alle”, gefördert durch das Bund-Länder-Programm “Künstliche Intelligenz in der Hochschulbildung”, sowie der deutliche Ausbau von GPU-basierten HPC-Knoten tragen sichtbar Früchte: So gaben schon ganze 23 % der Befragten an, hauptsächlich maschinelles Lernen zu betreiben. Dabei hilft sicher auch die Bereitstellung eines JupyterHubs in Düsseldorf, mit dem sehr einfach Jupyter-Notebooks verwendet werden können. Dies wird auch von etwa 13 % der Befragten aktiv genutzt.

Einen regelrechte Verdopplung sieht man bei der Nutzung von sogenannten “freien” Anwendungen: So verwenden 95 % der Befragten mittlerweile Open-Source-Software, 55 % Eigenentwicklungen und 36 % kommerzielle Anwendungen. Ein gestiegenes Interesse findet sich bei der Nutzung von Containerlösungen.

Der bereits eingeschlagene Richtungswechsel des HPC-Systems hin zu einem stabilen Basisdienst wird auch als richtig und wichtig erachtet. Bemängelt wurden die Hardware-bedingten Ausfälle in der Vergangenheit, die hauptsächlich auf Kühlungs- und Stromversorgungsprobleme oder Probleme mit dem Storage-System zurückzuführen waren. Viele dieser Problemstellen wurden bspw. durch die bereits erfolgte Erneuerung und den redundanten Ausbau des Kühlungssystems, der demnächst anstehende Austausch der Stromschienen, und die aktuelle Beantragung eines neuen, redundanteren Storage-Systems angegangen.

Zusammen mit dem Standort Paderborn ist Düsseldorf in der Nutzendenbefragung Spitzenrei-

ter, was den Support betrifft. Dabei haben wir die meisten Anfragen zu konkreten anwendungsspezifischen Problemen. Es wurde flächendeckend sehr viel Lob und Anerkennung gegenüber dem bestehenden Support geäußert. Darüber hinaus wurde auch positiv anerkannt, dass Problemlösungen schnell und sogar außerhalb der üblichen Arbeitszeiten erfolgten. Die Befragten wünschen sich explizit noch mehr Ressourcen für den Support, um eben auch ausführlichere Beratungen und Codeanalysen durchführen zu können.

13.7 Standortspezifische Auswertung - Duisburg-Essen

Im Vergleich zur vorherigen Umfrage konnte die Teilnehmendenzahl aus der Universität Duisburg-Essen nicht gesteigert werden, sondern sank um eine Person auf 10. Liegt damit aber unter dem Schnitt des Teilnehmerrückgangs über alle Standorte. Jedoch wird angestrebt, die Bekanntmachungen für zukünftige Nutzendenbefragungen weiter zu intensivieren. Aufgrund des niedrigen Anteils von 13,3 % der Befragten an der Anzahl aktiver Nutzenden (75) sind die Umfrageergebnisse nur bedingt auswertbar. Trotzdem werden einige Punkte aufgegriffen und resultierende oder geplante Maßnahmen benannt.

Die Nutzenden wünschen sich sowohl im Bereich der Grundlagenkurse als auch Fortgeschrittenkurse mehr Angebote besonders hinsichtlich Parallelsierungsmethoden (OpenMP und MPI). Dies wird zum Anlaß genommen, die entsprechenden Kurse an den NHR und Tier-1 Zentren stärker zu bewerben. Gleichzeitig wird angestrebt, ausgewählte Kurse am UDE-Standort durch das HPC-Support-Team durchzuführen.

Bei den Jobgrößen konzentriert sich der Hauptanteil (70 %) auf die Nutzung von mehreren Rechenknoten (40-100 Kerne) und viele Knoten (hunderte Kerne). Damit werden im Vergleich zu den anderen Tier-3 Systemen ein größerer Anteil großer Jobs durchgeführt. Dieses Ergebnis bestätigt auch die Auslegung der Jobpriorisierung im Scheduler, die u.a. auf der Basis der allokierten Knotenanzahl durchgeführt wird.

Die Anforderung an die Hardwareausstattung (Arbeits- und Festplattenspeicher) verglichen mit dem Rahmenparameter des aktuellen Systems zeigen, dass die Nutzenden ausreichende Ressourcen vorfinden. Bei den Wünschen nach möglichen Beschleunigern zeigt es sich, dass die Entscheidung, im zukünftigen HPC-System (voraussichtlich ab Mitte 2023) eine GPU-Knoten Partition vorzusehen, dem Wunsch der Nutzer entspricht.

Die Befragungsergebnisse hinsichtlich Performanceanalyse und Debugging zeigt, dass dort ein gewisser Nachholbedarf besteht. Aus diesem Grund wurde bereits aus den ISV-Mitteln von HPC.NRW die Software Vampir beschafft. Demnächst werden in entsprechenden Schulungsangeboten den Nutzenden das notwendige Wissen zur Nutzung dieser Analysesoftware vermittelt.

Den Wünschen hinsichtlich des Datenzugriffs mittels CIFS/SMB bzw. NFS wurde bereits bei der Auslegung des nächsten HPC-Systems entsprochen. So wird dabei ein separater Gateway-Server eingesetzt, der den Export des parallele Filesystem über verschiedene Protokolle ermöglicht.

13.8 Standortspezifische Auswertung - Köln

Von den etwa 300-400 jährlich aktiven Nutzenden des CHEOPS Clusters haben dieses Jahr 22 Personen an der Umfrage teilgenommen, was im Vergleich zur Umfrage von 2020 mit 54 Antworten einer um 60% geringeren Beteiligung entspricht. Die fachliche Zusammensetzung ist weitgehend gleich geblieben - es sind vorrangig Lebenswissenschaften (65%) vertreten, womit auch der Schwerpunkt des Kölner Standorts widerspiegelt wird. Die akademische Ausbildung der Nutzenden ist auch ähnlich geblieben, mit leichtem Anstieg an Antworten von Promovierten (9% Prof., 50% Dr./Ph.D, 36% Master).

Über die Hälfte (55%) der Befragten veröffentlichten Peer Reviewed Artikel in den letzten 18 Monaten und fast alle (91%) planen bald zu publizieren. Weiterhin liegt der Schwerpunkt für HPC Nutzung klar in der Forschung, allerdings hat sich der Anteil derer, die es für ihr Curriculum als sehr wichtig erachten, auf 32% verdoppelt.

Die genutzten CPU-Compiler und ihre Häufigkeiten sind in etwa gleich geblieben (GCC 68%, Intel 36%). Zusätzlich wird der GPU-Compiler aus dem NVIDIA HPC SDK von 5% der Umfrageteilnehmer eingesetzt. Aus den genannten MPI-Implementationen stechen OpenMPI (55%), Intel MPI (23%) und MPICH (18%) hervor. Die drei häufigst genutzten HPC-Bibliotheken sind OpenBLAS (18%), Intel MKL (18%) und FFTW (14%). Als am meisten am Standort genutzte Parallelisierungstechniken werden MPI (58%) und OpenMP (36%) genannt.

Die meisten User verwenden Open Source Programme (77%), etwas über ein Drittel (36%) schreibt eigene Programme und eine kleine Menge (14%) nutzen kommerzielle Software. Die wichtigste Zugriffsart auf die HPC-Infrastruktur ist direkter Zugriff über ssh, welches die meisten User sowohl zur Ausführung von Software (91%) als auch zum Datenzugriff (68%) nutzen.

Einige Nutzende wünschen sich aktuelle Container Softwaretechnologien wie Docker (32%) oder Singularity (5%), wobei letzteres schon von 18% der Befragten eingesetzt wird. Dies zeigt zum einen, dass das im Oktober durchgeführte Upgrade des CHEOPS Clusters auf CentOS7 mit einhergehender Zurverfügungstellung von Singularity noch nicht breit genug kommuniziert worden ist. Zum anderen fehlen noch die Kenntnisse über die Konvertierung von Docker- zu Singularity-Workflows. Ähnlich sieht es bei Jupyter-Notebooks aus - 27% der Befragten melden ein Interesse an der Technologie während sie bei nur 5% bereits im Einsatz ist.

Das CHEOPS Cluster unterliegt aufgrund der benutzten Hardware bzw. Ressourcenverfügbarkeit einigen Beschränkungen. Trotzdem äußern sich 73% der Befragten positiv zur Frage, ob Infrastruktur und Dienste auf dem Cluster für ihre Arbeit ausreichend ist. Bei Hardware sieht es geringfügig schlechter aus (59% ausreichend, 9% ausreichend mit langwieriger Optimierung). Die Beschaffung von CHEOPS2 sollte den bemängelten Bottlenecks in Workflows wie Arbeitsspeichergröße (32%), File-IO (23%), fehlende GPUs oder AVX (14%) sowie Kommunikationslatenz/-bandbreite (14%) entgegenwirken.

Die Kontaktwege zum Support durch die HPC Mitarbeiter des RRZK sind den meisten Nutzenden bekannt (82%) und dieser wird für alle Arten von Themen (z.B. Job Submission, Workfloworganisation, anwendungsspezifische Probleme) in Anspruch genommen. Alle Befragten, die den Support kontaktiert haben, bewerten ihre Erfahrungen mit „sehr zufrieden“. In den Jahren 2020 - 2022 sind sowohl die allgemeinen HPC-Kurse als auch die Parallelisierungskurse der HPC Gruppe Corona-bedingt ausgefallen. Dementsprechend werden in den Verbesserungswünschen des Supports nun vorrangig HPC Trainings genannt - das Interesse an Allgemeinem HPC Training ist von 19% auf 50%, am Fortgeschrittenen HPC Training von 17% auf 32% gestiegen.

Im Allgemeinen werden sowohl mehr online als auch Präsenz HPC Kurse angefragt. Die gewünschten Themen sind dabei Debugging und Profiling, Large-Scale File-IO sowie Code Optimierung und Compiler-/Bibliotheksnutzung. Bei den Programmierkursen sind vorrangig Python, C/C++ sowie CUDA und bei den Parallelisierungsmethoden hauptsächlich MPI und OpenMP gefragt.

Neue Kursformen, darunter auch Ringvorlesungen, Q&A Sessionen mit Lightning Talks sowie Kurse mit externen Experten werden vom RRZK entwickelt. Ferner plant das RRZK in Zusammenarbeit mit den Herstellern des CHEOPS2 Systems weitere hochwertige Kurse und Workshops anzubieten, die die nachgefragten Themen umfassen. Weiterhin wird die Darstellung der Support Informationen für das CHEOPS Cluster überarbeitet und ausgebaut.

13.9 Standortspezifische Auswertung - Münster

Die Anzahl der Teilnehmer der Umfrage in Münster ist stark von 40 auf nur 18 zurückgegangen. Die größte Gruppe kam dabei jedoch wie zuvor aus der Physik (6) gefolgt von Informatik (4) sowie Medizin und Biologie mit je 3 Teilnehmern.

Der Anteil der Befragten die zufrieden mit dem angebotenen Support sind, ist mit 44% in etwa gleich hoch wie bei der vorherigen Umfrage (40%). Ein leichter Anstieg ist im Bedarf bei anwendungsspezifischem und individuellem Support zu verzeichnen. Hier hat die WWU IT bereits reagiert und eine wöchentliche, persönliche Sprechstunde für alle Nutzerinnen und Nutzer eingerichtet.

Stark zurückgegangen ist, entgegen des allgemeinen Trends, die Nutzung von Jupyter-Notebooks

auf dem PALMA HPC Cluster. Die Nutzung dieser ist bereits seit 2020 möglich und der Rückgang vermutlich teilweise auf die andere Zusammensetzung der Befragten zurückzuführen. Allerdings wünschen sich ca. die Hälfte der Teilnehmer diese Funktion zu nutzen, was auf mangelnde Informationen hinweist. Wir haben uns daher vorgenommen die Nutzung der Jupyter-Notebooks ebenfalls als Inhalt mit in die lokalen HPC Schulungen einzubauen und die Sichtbarkeit der Funktion in unserem lokalen Wiki zu erhöhen.

Ein wichtiger Punkt bei der Nutzung des Clusters ist zudem die fehlende Verfügbarkeit von Nodes mit genug RAM und Festplattenspeicher. Da dies bereits bei der ersten Umfrage deutlich wurde, konnte die WWU in 2022 erfolgreich Mittel bei der DFG für eine Erweiterung von Nodes mit entsprechend größerem Hauptspeicher und Festplatten beantragen. Die Ausschreibung hierfür wird Anfang 2023 stattfinden, mit einer Installation ist vrs. im Q2 2023 zu rechnen.

13.10 Standortspezifische Auswertung - Paderborn

Von den insgesamt etwa 341 Personen, die an der Nutzendenbefragung teilgenommen haben, nutzen 57 eines der Paderborner HPC-Systeme (Noctua 1 oder Noctua 2). In den Zugehörigkeiten zu Fachbereichen spiegeln sich die Schwerpunkte des PC², sprich atomistische Simulationen, Optoelektronik und Quantenphotonik, sowie intelligente Systeme wieder.

Die Befragten aus Paderborn nutzen vor allem das Tier-2-System, das das PC² anbietet. Nach der Installation des neuen Hochleistungsrechners Noctua 2 wurden die Ausschreibungen für Rechenzeitanträge für Forschenden aus Deutschland veröffentlicht. Daher benutzen die Meisten das Tier-2-System. Im Vergleich zur letzten HPC.NRW Nutzendenumfrage, wo 28% der Befragten die Arbeitsgruppencluster (sogenannte Tier-4-Systeme) benutzten, sieht man nun, dass der Anteil auf nur 8% gesunken ist. Dies ist bereits einen Fortschritt in der Richtung des neuen Trends des HPC-Landeskonzepts.

In der Nutzendenumfrage haben etwa 98% der Befragten aus Paderborn angegeben, einen Weg zu kennen, mit dem HPC-Support bei Problemen oder Fragen in Kontakt treten zu können. Etwa 92% haben angegeben, den Support schon einmal genutzt zu haben. Insgesamt entfielen die meisten Gründe für Kontakt zum HPC-Support auf den Zugang zum Cluster (40%) und die Unterstützung bei Rechenzeitanträgen (38%). Dies könnte sowohl auf die Inbetriebnahme des Noctua 2 zurückzuführen sein, als auch auf unsere Aktivitäten in den Aufrufen zur Antragstellung auf Rechenzeit. Bezüglich des Zugangs zu den HPC-Systemen werden ausführliche Informationen in der öffentlichen Nutzerdokumentation mitgeteilt² und werden auch bei regelmäßigen HPC-Kursen dargestellt³. Der größte Aufwand für die Antragstellende entsteht in der Vorbereitung des Rechenzeitantrags. Um diesen Bedarf abzudecken, wurden bei jährlichen PC² User Day 2021⁴ und 2022⁵ die Informationen zur Beantragung von Rechenzeit im PC² vorgestellt, wie z.B. die Voraussetzungen, das Beantragungsverfahren, insbesondere der Einsatz des PC²-JARDS⁶, sowie die Vereinfachungen für mehrjähriges Projekt, wissenschaftliche Begutachtung und Verwendung bekannter Anwendungen.

Im Gegensatz zur letzten HPC.NRW Nutzendenumfrage, wo unter 5% der Befragten das Jupyter Notebooks auf den HPC-Systemen verwendeten, benutzen knapp 70% der Befragten nun das Jupyter Notebooks auf den HPC-Systemen im PC². In den vergangenen zwei Jahren wurde eine integrierte Lösung für den Einsatz von Jupyter Notebooks auf den HPC-Systemen entwickelt und als einer unserer HPC-Services für Nutzenden bereitgestellt. Außerdem wurde in der Nutzerdokumentation eine Schritt-für-Schritt-Anleitung zur Nutzung des Jupyter Notebooks auf den HPC-Systemen erstellt⁷. Darüber hinaus wurde Jupyter Notebooks in den letzten Jahren in HPC-

²Siehe <https://upb-pc2.atlassian.net/wiki/spaces/PC2D0K/pages/1903514/Access+and+Login>.

³Siehe <https://events.uni-paderborn.de/category/1/>.

⁴Siehe <https://events.uni-paderborn.de/e/userday2021>.

⁵Siehe <https://events.uni-paderborn.de/e/userday2022>.

⁶Siehe <https://proposals.pc2.uni-paderborn.de/>.

⁷Siehe <https://upb-pc2.atlassian.net/wiki/spaces/PC2D0K/pages/1903131/JupyterHub>.

Kursen eingesetzt, z.B. „Introduction to PC² JupyterHub“⁸ und „oneAPI for CPUs, GPUs and FPGAs“⁹.

Der Zugang zu Daten auf dem HPC-System wurde in der letzten Nutzendumfrage von kommandozeilenbasierte `scp` und ähnlichen Programmen dominiert. Es gibt nun eine deutliche Steigerung bei der Nutzung von speziellen NFS/CIFS/SMB Exports (zu 60%), die für alle Betriebssysteme der Nutzenden verfügbar sind, z.B. Windows, macOS und Linux. Um die Bekanntheit solcher Wege für die Nutzenden zu verbessern, wurden die Anleitungen dazu in der Nutzerdokumentation des PC² überarbeitet¹⁰. Außerdem wurden neue zusätzliche Folien dafür erstellt und bei regelmäßigen HPC-Kursen dargestellt.

Um die Sichtbarkeit des HPC.NRW zu erhöhen, stellen wir es auf PC² Webseiten vor¹¹ und verlinken auf die entsprechenden Webseiten, z.B. <https://hpc.dh.nrw/> und https://hpc-wiki.info/hpc/HPC_Wiki. Unsere Veranstaltungen sind weitere wichtige Kanäle, über die wir mit den Nutzenden besser kommunizieren können, z.B. auf der Seite von jährlichen PC² User Day 2021 und 2022 wurden NHR und HPC.NRW hervorgehoben¹².

13.11 Standortsspezifische Auswertung - Siegen

In Siegen haben rund 40% der aktiv Nutzenden an der Umfrage teilgenommen. Die mit Abstand größte Gruppe gehört zum Bereich „Informatik, Systeme und Elektrotechnik“. Seit der letzten Umfrage wurde der HoRUS Cluster durch den OMNI Cluster ersetzt, der Kreis der Nutzenden hat sich dadurch etwas erweitert. Die Rechenressourcen wurden deutlich erhöht, dennoch fällt auf, dass mit rund einem Viertel der Nutzer die Nachfrage nach mehr Ressourcen immer noch relativ hoch ist. Dies ist insbesondere im Bereich der GPUs der Fall.

Im Vergleich zum Vorjahr sind die Publikationen, die mit Unterstützung des Clusters entstanden sind, deutlich gestiegen, 54% der Befragten haben bereits etwas veröffentlicht, bei 70% ist sogar eine Publikation in Planung.

Insgesamt fällt auf, dass die Nutzenden relativ fortgeschritten sind. Rund 80% der Befragten entwickeln ihre Programme selbst und ca. 85% nutzen parallelisierte Jobs. Nach wie vor sind die parallelen Dateisysteme nicht hinreichend bekannt, hier besteht durchaus noch Handlungsbedarf.

Die Umfrage zeigt, dass insbesondere der lokale Support bekannt ist, gerne genutzt wird und auch sehr positiv bewertet wurde, HPC.NRW und alles, was damit in Verbindung steht, wie das HPC-Wiki, der Youtube-Channel etc. haben allerdings noch keine hohe Visibilität erreicht. Da die Nutzenden sich, was die Schulungen betrifft, insbesondere vermehrt Tutorials mit kurzen, unabhängigen Videos wünschen, wurde dies zum Anlass genommen, proaktiv das HPC-Wiki mit seinen Online-Tutorials sowie den Youtube-Channel zu bewerben. Auch HPC.NRW wurde in der Mail erläutert, findet nun auf der lokalen Homepage der Universität Siegen Erwähnung und ist dort auch entsprechend verlinkt.

13.12 Standortsspezifische Auswertung - Wuppertal

In Wuppertal haben hauptsächlich Nutzende aus den Ingenieurwissenschaften und der Physik an der Umfrage teilgenommen. Zwar ist die Anzahl der Personen von 4 im Jahr 2020 auf diesmal 15 deutlich angestiegen, jedoch sind weiterhin alle Schlussfolgerungen unter dem Aspekt einer relativ niedrigen Teilnahme zu sehen. Seit der letzten Umfrage ist das Pleiades Cluster im Juni 2021 durch neue Hardware ersetzt worden, wobei gleichzeitig auch der Nutzendenkreis deutlich erweitert wurde. Eine weitere Ursache für die gestiegene Beteiligung könnte eine verstärkte Wahrnehmung von HPC.NRW als Folge von universitäts- und clusterweiten Veranstaltungen sein, die sowohl von

⁸Siehe <https://events.uni-paderborn.de/event/209/>.

⁹Siehe <https://events.uni-paderborn.de/event/176/>.

¹⁰Siehe <https://upb-pc2.atlassian.net/wiki/spaces/PC2D0K/pages/1903056/Data+Transfer+File+Staging>.

¹¹Siehe <https://pc2.uni-paderborn.de/> und <https://pc2.uni-paderborn.de/go/hpc-nrw/>.

¹²Siehe <https://events.uni-paderborn.de/e/userday2021> und <https://events.uni-paderborn.de/e/userday2022>.

lokal Beschäftigten als auch anderen Mitgliedern des Kompetenznetzwerkes gehalten wurden. Ein Blick auf Abb. 37 verdeutlicht dennoch, dass die Bekanntheit der HPC.NRW-spezifischen Kanäle noch deutlich verbessert werden könnte. Hinsichtlich des Hintergrunds der Nutzenden sollte auch beachtet werden, dass $\frac{2}{3}$ ein abgeschlossenes Master- oder Diplom-Studium besitzen und damit wahrscheinlich an einer Promotion arbeiten. Dadurch sind die Ansprüche jener forschungsorientierten Nutzenden entsprechend höher anzusetzen als bei Nutzenden ohne abgeschlossenes Studium. Gleichzeitig kann davon ausgegangen werden, dass aus Eigeninitiative heraus mehr Weiterbildung im Bereich HPC-relevanter Themen stattfand als alleine durch das Kursangebot der Universität oder HPC.NRW möglich wäre. Zuletzt sei angemerkt, dass alle Befragten in Wuppertal ihre Rechnungen ausführen (20% auch am JSC) und sich somit das Feedback hauptsächlich auf den lokalen Cluster Pleiades bezieht. Alle Personen geben an, dass sich Hochleistungsrechnen positiv auf ihre Arbeit auswirkt und ganze 87% bezeichnen es als unerlässlich. Trotz des hohen Wertes geben 60% an, dass HPC keine Rolle in ihrem Curriculum spielt. Nur 7% (1 Person) geben an, dass Vorlesungen mit HPC-Inhalten fest in ihrem Studium eingebettet sind. Die verbleibenden 33% geben an, dass es nur als Randthema und in spezialisierten Vorlesungen thematisiert wird.

Setzt man dies in Relation zum Anspruch, den ein Promotionsstudium stellt, fällt auf, dass hier Nachholbedarf besteht, an welchen HPC.NRW ansetzen kann. Durch Einführungsveranstaltungen zum Kompetenznetzwerk, regelmäßigen Austausch mit den Forschungskräften an den Lehrstühlen und Entwicklung von HPC-Veranstaltungen (zunächst an Ingenieure gerichtet) soll allmählich eine bessere Grundlage für angehende Forschende zur Verfügung gestellt werden. Dies wird umso wichtiger, wenn man bedenkt, dass 80% ihre Programme selber entwickeln, aber 67% kein Profiling durchführen und 33% nicht wissen, welche Form der Parallelisierung genutzt wird.

Die von den Clusterbetreibenden zur Verfügung gestellten HPC Werkzeuge scheinen weitestens zufriedenstellend zu sein, da 53% angeben, keine weiteren Werkzeuge zu benötigen. Die verbleibenden Antworten verteilen sich gleichmäßig und es ist keine klare Tendenz zu einem bestimmten HPC Tool zu erkennen.

Direkt auf Containisierungs-Software angesprochen zeichnet sich ein leicht anderes Bild ab, in welchem 40% Docker nutzen würden und 20% Singularity. Besonders in Anbetracht von nachhaltiger Forschung und Forschungsdatenmanagement sind Containerisierungstechnologien positiv zu sehen, weshalb die Dokumentation über die Verwendung von Containern auf dem Cluster ausgebaut werden sollte. Ähnlich verhält es sich bei der direkten Frage nach einer gewünschten Nutzung von Jupyter Notebooks, die von 47% bejaht wird (weitere 27% nutzen es bereits). Einer der Gründe hierfür ist sicherlich die Popularität von Python wie standortübergreifend in dieser Umfrage festgestellt werden konnte.

14 Zusammenfassung

In der aktuellen Nutzendenbefragung wurde eine ähnliche Herangehensweise wie bei der letzten Iteration verfolgt. Dementsprechend präsentiert dieses Dokument eine allgemeine und zusammenfassende Auswertung sowie einige standortspezifische Erkenntnisse der Partner im Kompetenznetzwerk HPC.NRW. Die Rohdaten stehen auch dieses Mal allen Standorten zur weiteren Analyse zur Verfügung. Obwohl die Gesamtbeteiligung an der Nutzendenbefragung im Vergleich zur letzten Iteration 2020 etwas nachgelassen hat, ist der Anteil relativ zu den aktiven Nutzenden ähnlich geblieben.

Ein besonderes Augenmerk bei der Auswertung lag in einer vergleichenden Betrachtung der Ergebnisse, wo dies angemessen war. Die allgemeine Auswertung spiegelt dabei hauptsächlich die reine Entwicklung der Daten wider. In den standortspezifischen Abschnitten wird zum Teil auf mögliche kausale Zusammenhänge in diesen Entwicklungen eingegangen. Da sich die Einführung und Art von Maßnahmen sowie die Reaktion auf mögliche Maßnahmen an den Standorten stark unterscheiden können, wurde die Auswertung auf diese Weise aufgeteilt.

Zusammenfassend zeigen die Antworten, dass der HPC-Support in den einzelnen HPC-Zentren in NRW weiterhin gut aufgestellt und an einigen Punkten ihr Angebot noch ausbauen konnten. Dies wird zum Beispiel an der insgesamt stärkeren Nutzung von Jupyter-Notebooks deutlich.

Die Antworten verdeutlichen die hohe Relevanz von HPC in der Forschung, bei gleichzeitig eher geringer Abdeckung entsprechender Themen in regulären Studiengängen. Eine entsprechende Lücke in der Ausbildung kann durch regelmäßige Workshops, lokale Einführungsveranstaltungen und online Tutorials gefüllt werden.