

PROCEEDINGS OF THE 5TH BWHPC SYMPOSIUM

Michael Janczyk
Dirk von Suchodoletz
Bernd Wiebelt
(Hrsg.)

PROCEEDINGS OF THE 5TH BWHPC SYMPOSIUM

**HPC Activities in Baden-Württemberg
Freiburg – September 2018**



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.



Der Text dieses Werks ist unter der Creative-Commons-Lizenz CC BY-SA 4.0 International (Namensnennung Share Alike – 4.0 International) veröffentlicht. Den Vertragstext finden Sie unter: <https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>

Die Online-Version dieser Publikation ist auf den Verlagswebseiten von Tübingen Library Publishing frei verfügbar (Open Access).

<http://hdl.handle.net/10900/87676>

<http://nbn-resolving.de/urn:nbn:de:bsz:21-dspace-876765>

<http://dx.doi.org/10.15496/publikation-29062>

1. Auflage 2019 Tübingen Library Publishing

Universitätsbibliothek Tübingen

Wilhelmstraße 32

72074 Tübingen

druckdienste@ub.uni-tuebingen.de

<https://tlp.uni-tuebingen.de>

ISBN (Softcover): 978-3-946552-26-0

ISBN (PDF): 978-3-946552-25-3

Umschlaggestaltung: Sandra Binder, Universität Tübingen

Coverfoto: pixabay

Satz: Michael Janczyk

Druck und Bindung: Pro BUSINESS digital printing Deutschland GmbH

Printed in Germany

Vorwort

Die nun fünfte Ausgabe des bwHPC-Symposiums zeigt, dass sich die Veranstaltung fest etabliert hat. Sie ist Plattform für den Austausch von Wissenschaftlerinnen und Wissenschaftlern mit den Betreibern der großen, föderierten Forschungsinfrastrukturen. Das Symposium versucht, eine Brücke zwischen den verschiedenen Fachdisziplinen und den technisch-organisatorischen Aspekten zu schlagen. Es ergänzt die etablierte Governance aus Landesnutzerausschuss und den lokalen Gremien. Dem diesjährigen Symposium fiel die zusätzliche Rolle zu, die erste Projektphase von bwHPC-S5 einzuleiten, nachdem das fünfjährige Vorgängerprojekt bwHPC-C5 abgeschlossen wurde.¹

Erfolgreiche Wissenschaft benötigt leistungsfähige Infrastrukturen: Forschung bildet eine zentrale Säule im Selbstverständnis moderner Gesellschaften. Digitalisierte Arbeitsprozesse prägen alle Wissenschaftsdisziplinen. Die Stärke eines Wissenschaftsstandorts leitet sich wesentlich von der Verfügbarkeit attraktiver, integrierter und skalierender Forschungsinfrastrukturen ab. Die Computational Science und damit HPC-Systeme als deren technisches Fundament gewinnen unablässig an Bedeutung. Deshalb besteht eines der Ziele der im Juli 2018 gestarteten bwHPC-S5 Begleitaktivitäten darin, die Forscher*innen in ihrer Arbeit mit den leistungsfähigen Infrastrukturen umfassend zu unterstützen, ohne sie jedoch aus der individuellen Verantwortung für ihre Daten und wissenschaftlichen Workflows zu entlassen.

Der Erkenntnis folgend, dass heutige Anforderungen nicht mehr sinnvoll von einzelnen Universitäten oder Forschungsinstitutionen bedient werden können, koordinieren die wissenschaftlichen Rechenzentren des Landes Baden-Württemberg ihre Aktivitäten im Bereich HPC und der Speicherung großer Datenmengen unter

¹Vgl. hierzu <http://www.bwhpc.de/projektaufgaben.php>

dem gemeinsamen Dach des »Umsetzungskonzept(es) der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS²DM)«. Damit baut das Land ein wesentliches Alleinstellungsmerkmal bei der Unterstützung der Wissenschaften aus und strebt eine Vorreiterrolle bei der Etablierung von Nationalen Forschungsdateninfrastrukturen (NFDI) und Science Data Centern an.

Der föderierte Betrieb großer Forschungsinfrastrukturen verändert das Umfeld und die Anforderungen der beteiligten Universitätsrechenzentren. Sie müssen sich verstärkt abstimmen und gemeinsame Lösungen finden. Nutzer der Infrastrukturen kommen nun nicht mehr allein aus der eigenen Einrichtung, sondern aus dem ganzen Land. Umgekehrt müssen Forscher*innen der eigenen Universität dabei unterstützt werden, Ressourcen an anderen Standorten zu verwenden. Leistungserbringung, Kosten und Zugehörigkeit zu einer Forschungseinrichtung fallen nicht mehr zwangsläufig zusammen, was neue Herausforderungen in der Leistungsverrechnung und Finanzierung mit sich bringt.

Mit der bereits vor mehr als zehn Jahren mit dem bwGRID gestarteten HPC-Initiative des Landes wurde eine neue Kultur der Bereitstellung und Nutzung großer Forschungsinfrastrukturen eingeleitet. Durch die Spezialisierung der HPC-Systeme und die Bündelung gemeinsamer Anforderungen der verschiedenen Forschungsgruppen wurden Anstrengungen und Anschaffungen zusammengeführt. Diese nutzt nicht nur Economies-of-Scale, die Erkenntnis, dass der Aufwand für das Design, die Beschaffung und den Betrieb eines Clusters mit zunehmender Größe fast nicht mehr ansteigt. Es erlaubt ebenfalls eine deutlich effizientere Auslastung der Ressourcen, da üblicherweise in Einzelsystemen anfallende Leerlaufzeiten durch andere Nutzer gefüllt werden. Damit entstehen klare Mehrwerte, die es erlauben, einige deutlich größere Systeme mit hoher Leistung zu gleichen Kosten wie viele kleine mit deutlich geringerer Gesamtleistung zu beschaffen. Dabei erlauben es Mechanismen wie »Fairshare« Nutzungsanteile am Gesamtsystem in beliebiger Stückelung herauszugeben. Dieses senkt die Einstiegsschwelle insbesondere für kleinere Gruppen und Nachwuchsforscher*innen. Für große Gruppen und traditionelle Großantragsteller besteht der Mehrwert in einem zügigen Start ihrer Forschung, einer potenziell noch größeren Ressource als in einer individuellen Beschaffung ohne die Nachteile der Aufwendungen für eine eigene Betriebsmannschaft. Durch eine gestaffelte Beschaffung und regelmäßige Erneuerungszyklen bleibt die Attraktivität der Ressourcen

langfristig erhalten, und die Beteiligung neuer Nutzergruppen kann sichergestellt werden. Das verändert die Landschaft großer Forschungsinfrastrukturen nachhaltig.

Die im »High Performance Computing« und »High Throughput Computing« zunehmend genutzten großen Datenmengen in stärker föderierten Forschungsinfrastrukturen erfordern geeignete Strategien für ihre Haltung und ihren Zugriff, da selbst bei hochperformanten Netzwerken das Kopieren von Datensätzen immer weniger eine effiziente Lösung darstellt. Im Zuge der Entstehung, Verarbeitung und nachhaltiger Langzeitlagerung und Nachnutzung ändern sich die Anforderungen an das Datenmanagement bezüglich Zugriffsrechten, Leistungserwartungen und Absicherungen im Zeitablauf. Hierzu zählt die zunehmend komplexe Datenhaltung, die von der Generierung der Daten in Messinstrumenten, Simulationen und vielfältigen weiteren Verfahren über das Pre- und Postprocessing bis zur Verarbeitung und Visualisierung in Spezialsystemen und der langfristigen Haltung und Publikation reicht. Im Lebenszyklus von Forschungsdaten entstehen bei den Forschenden verschiedenartige Anforderungen, die von der schnellen Speicherung bei der Datenerhebung, über die Verarbeitung in HPC- und Cloudsystemen bis hin zur Visualisierung reichen. Hinzu kommt die notwendige Aufarbeitung der Daten im Sinne »guter wissenschaftlicher Praxis«, deren Langzeitspeicherung und -verfügbarkeit für Publikation und Nachnutzung. Das Forschungsdatenmanagement (FDM), die nachhaltige und zukunftsorientierte Speicherung von Forschungsdaten, ihre Verfügbarmachung und optimalerweise Publikation, rückt damit in den Fokus moderner Forschungsprozesse. Nachnutzung und Reproduzierbarkeit von Forschungsergebnissen benötigen sowohl Verständnis und Bereitschaft der Forscher*innen, als auch neuartige Dienstleistungen seitens der zentralen Infrastrukturbieter.

Langfristig müssen geeignete Maßnahmen und Workflows umgesetzt werden, um den Wissenschaftler*innen nicht nur einen Großteil der Datenmanagementaufgaben abzunehmen, sondern auch die Effektivität der Ressourcennutzung zu steigern. Dies ist insbesondere bei der Nutzung von HPC-Ressourcen wünschenswert, indem der dort stattfindende Batch-Betrieb durch die automatische Versorgung mit Daten optimiert wird. Neben weiteren »lokalen« Softwarekomponenten, wie beispielsweise Versionsmanagement, ist auch der Anschluss an, beziehungsweise die einfache Nutzbarkeit bestehender externer Forschungsdateninfrastrukturen und der zukünftigen Nationalen Forschungsdateninfrastruktur notwendig, um Wissenschaft-

lern die Nutzung der fachspezifischen Repositoryn zu ermöglichen. Die Aufgabe der Betreiber und des bwHPC-S5-Projekts besteht demnach zusätzlich darin, beständig neue Lösungen zu suchen und die föderativ angelegten Forschungsinfrastrukturen im Sinne der Forschenden beständig weiterzuentwickeln. Die Anwendung von Virtualisierungs- und Containerisierungsstrategien ist nur ein Beispiel hierfür.

Der Tagungsband besteht aus drei Teilen, wobei im ersten Teil die konzeptionellen und projektbezogenen Darstellungen im Vordergrund stehen. Der zweite Teil beinhaltet die durch ein Peer-Review begutachteten Beiträge der Forscher*innen, die eine Verbindung von Forschungsfragestellungen mit konkreten Berechnungen auf den jeweiligen bwHPC-Systemen herstellen. Im dritten Teil finden sich die durch die Projektpartner begutachteten Texte zu administrativen und betrieblichen Fragestellungen und Überlegungen zur technischen Weiterentwicklung der Betriebsmodelle.

Die in diesem Tagungsband zusammengefassten Konzepte und Aktivitäten beziehen zu großen Teilen ihre Kraft aus der langfristig angelegten Unterstützung des Ministeriums für Wissenschaft, Forschung und Kultur in Baden-Württemberg. Hierzu zählen zuvorderst die finanziellen Beiträge für die Beschaffung und regelmäßige Erneuerung der HPC-Systeme als auch die Förderung der bwHPC-C5- und bwHPC-S5-Begleitprojekte. Weiterhin profitieren die beteiligten Forschungseinrichtungen durch die eScience-Initiative des Landes, durch die beispielsweise das ViCE-Projekt seine Unterstützung erfuhr.

Die Herausgeber
Freiburg, April 2019

Contents

Vorwort	V
Grußwort der Ministerin für Wissenschaft, Forschung und Kunst Baden-Württemberg	XI
Grußwort des Prorektors für Forschung der Universität Freiburg	XV
I Konzepte und Projekte/Concepts and Projects	1
I.1 Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das HPC, DIC und LS ² DM (gekürzte Fassung) (Gerhard Schneider et al.)	3
I.2 bwHPC-S5: Scientific Simulation and Storage Support Services (Robert Barthel und Jürgen Salk)	17
I.3 bwForCluster NEMO – Forschungscluster für die Wissenschaft (Michael Janczyk et al.)	29
I.4 Data Analysis for Improving High Performance Computing Operations and Research (Florina M. Ciorba et al.)	51
II Scientific Contributions	59
II.1 Performance of the bwHPC cluster in the production of → embedded events used for the prediction of background for H → analyses (Janek Bechtel et al.)	61
II.2 Simulating tactoids of chiral rod-like particle (Anja Kuhnhold)	75
II.3 Neutron Star Oscillations – Linking gravitational waves to micro-physics (Andreas Boden et al.)	87

II.4	Testing Einstein’s theory of gravity with simulations of tidal disruption events (Gela Hämerling et al.)	101
II.5	HPC with Python: An MPI-parallel implementation of the Lattice Boltzmann Method (Lars Pastewka and Andreas Greiner)	119
II.6	Numerical Investigation of Strongly Interacting Bosons at Zero Temperature (Laurent de Forges de Parny et al.)	135
III	Administrative and Technical Contributions	159
III.1	Dynamic Resource Extension for Data Intensive Computing with Specialized Software Environments on HPC systems (Christoph Heidecker et al.)	161
III.2	Unified Container Environments for Scientific Cluster Scenarios (Benjamin Schanzel et al.)	173
III.3	Integration of NEMO into an existing particle physics environment through virtualization (Felix Bührer et al.)	187
III.4	de.NBI Cloud Storage Tübingen – A federated and georedundant solution for large scientific data (Benjamin Gläßle et al.)	201
III.5	A Sorting Hat for Clusters (Jonathan Bauer et al.)	217
III.6	Feeding the Masses: DNBD3 (Simon Rettberg et al.)	231
III.7	Game of Templates (Jonathan Bauer et al.)	245
III.8	Integrated Storage Infrastructures (Dirk von Suchodoletz et al.)	263