Sachbericht zum Verwendungsnachweis – Teil I (Kurzbericht)

ZE: Bayerische Akademie der Wissenschaften	Förderkennzeichen: 16HPC040

Vorhabenbezeichnung:
REGALE – Offene Architektur für die nächste Generation von Exascale-fähigen HPC-Anwendungen im zukünftigen Supercomputing

Laufzeit des Vorhabens:
01.04.2021 – 31.03.2024

Ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Der Stromverbrauch von Hochleistungsrechnern steigt kontinuierlich an. Bedingt durch das Ende des Dennard-Scaling und der Verlangsamung von Moore's Law werden Leistungssteigerungen bei Supercomputern primär durch immer größere Systeme mit stetig wachsendem Strombedarf erreicht. Das stellt die Betreiber solcher Systeme sowohl vor technische, finanzielle, als auch ökologische Probleme.

Ziel des REGALE-Projekts war es, eine offene Architektur für Hoch- und Höchstleistungsrechner zu entwickeln, die autonom die Ressourcennutzung überwacht, das Verhalten des Systems lernt, Ineffizienzen erkennt und die Ressourcenzuteilung an Anwendungen optimiert. Ein besonderer Schwerpunkt soll dabei auf der Steigerung der Energieeffizienz beim Betrieb solcher Systeme liegen. Zur Umsetzung der Ziele sind Änderungen an so gut wie allen Komponenten des HPC-Softwarestacks erforderlich.

Die Arbeiten in REGALE fokussieren sich einerseits auf klassische HPC-Anwendungen und Möglichkeiten um deren Ausführung energieeffizienter zu gestalten und kooperiert dabei mit der HPC PowerStack-Initiative, der mehrere Partner des Projekt-Konsortiums angehören. Andererseits liegt ein weiterer Schwerpunkt auf neuartigen, workflow-basierten Anwendung wie beispielsweise High Performance Data Analytics (HPDA) oder AI-Pipelines und deren effizienter Ausführung auf HPC-Systemen.

Für die Projektziele werden etablierte Softwarekomponenten des HPC Software-Stacks wie EAR, SLURM, oder DCDB weiterentwickelt und um die nötigen Features ergänzt. Relevante Anwendungen werden für die Zusammenarbeit mit dem REGALE-Stack optimiert.

Ablauf des Vorhabens

Das REGALE Projekt war in 6 Arbeitspakete organisiert:

- **WP1: Requirements, architecture and evaluation** (Leitung TU München)
  Anforderungsanalyse und Definition der REGALE-Architekur, Erstellung von Evaluierungskriterien.
- **WP2: Sophisticated resource allocation and management** (Leitung Université Grenoble Alpes)
  Verbesserung des Ressourcen-Managements und Applikationsunterstützung.
- **WP3: REGALE prototyping** (Leitung TU München)
  Entwicklung des REGALE Software-Stacks.
- **WP4: Exascaling the REGALE pilots** (Leitung National Technical University of Athens)
  Anpassung von relevanten Anwendungen an die REGALE-Architektur.

- **WP5: Innovation management** (Leitung LRZ)
  Außendarstellung des Projekts auf Konferenzen und Messen, Webpräsenz und Social Media, Zusammenarbeit mit Standardisierungsgremien.
- **WP6: Project management** (Leitung Institute of Communication and Computer Systems)
  Koordinierung des Projekts, Fortschrittsüberwachung, Qualitätssicherung und Erstellen der Finanzreporte.

Das Leibniz-Rechenzentrum (LRZ) leitete das Arbeitspaket 5 „Innovation management“ und beteiligte sich darin außerdem an den Tasks 5.1 „Communication Planning & Stakeholders’ Engagement“, 5.2 „Dissemination activities“ und 5.6 „Standardization“. Im Arbeitspaket 1 „Requirements, architecture and evaluation“ arbeitete das LRZ in allen Tasks mit, im Arbeitspaket 3 „REGALE prototyping“ hat sich das LRZ in Task 3.4 „Monitoring and Data Management“ eingebracht und seine Monitoring-Lösung DataCenter DataBase (DCDB) erweitert.

Wesentliche Ergebnisse sowie die Zusammenarbeit mit anderen Forschungseinrichtungen

Als ein von EuroHPC kofinanziertes Forschungsprojekt war REGALE von Anfang an auf intensive Zusammenarbeit mit anderen europäischen Forschungseinrichtungen ausgelegt. Insgesamt 16 Partner aus 6 europäischen Ländern trugen zum Gelingen des Projekts bei. Darüber hinaus gab es Kooperationen mit anderen internationalen Forschungsinstituten aus Asien und den USA im Rahmen der HPC PowerStack Initiative.

Das DEEP-SEA Projekt hat wichtige Systemsoftwarekomponenten und Anwendungen für zukünftige europäische Exascale-Systeme entwickelt und optimiert. Dabei wurden bestehende Softwarekomponenten um neue Funktionalitäten erweitert und für den Einsatz auf Supercomputern optimiert. Der Einsatz des im Projekt entwickelten Software-Stacks wurde exemplarisch anhand von 6 Co-Pilot Anwendungen demonstriert. Es wurde eine API für die Kommunikation der Software-Stack Komponenten untereinander definiert und prototypisch in einer Softwarebibliothek umgesetzt.

Das Leibniz-Rechenzentren hat sich im Bereich der technischen Forschungs- und Entwicklungsarbeiten hauptsächlich im Arbeitspaket 1 „Requirements, architecture and evaluation“ eingebracht“ und an der Anforderungsanalyse, der Definition der REGALE-Architektur und der Entwicklung der Use-Cases mitgewirkt. Im Arbeitspaket 3 „REGALE prototyping“ hat das LRZ sein Monitoring Framework „DataCenter DataBase“ (DCDB) um neue Features erweitert und die Integration mit anderen Komponenten des REGALE-Stacks umgesetzt. Darüber hinaus hat das LRZ das Arbeitspaket zur Öffentlichkeitsarbeit geleitet und alle Aktivitäten zur Außendarstellung (Webauftritt, Social Media, Messestände) koordiniert und größtenteils auch umgesetzt.

Sachbericht zum Verwendungsnachweis – Teil II (Eingehende Darstellung)

ZE: Bayerische Akademie der Wissenschaften	Förderkennzeichen: 16HPC040
Vorhabenbezeichnung: **REGALE – Offene Architektur für die nächste Generation von Exascale-fähigen HPC-Anwendungen im zukünftigen Supercomputing**	
Laufzeit des Vorhabens: 01.04.2021 – 31.03.2024	

Durchgeführte Arbeiten

Nach dem Ausscheiden eines Mitarbeiters Ende 2021 konnte die Stelle für die technischen Arbeiten in REGALE bis zum Projektende nicht neu besetzt werden. Die für den Projektfortschritt dringend erforderlichen Arbeiten wurden zwischenzeitlich von anderen Kollegen erbracht, die jedoch ihre Arbeitszeit aus buchhaltungstechnischen Gründen nicht im Projekt abrechnen konnten.

Arbeitspaket 2 „Requirements, architecture and evaluation"

- Mitwirkung an der Definition und Entwicklung der REGALE Architektur, Anforderungen und Use-Cases.
- Mitarbeit an Deliverable D1.1 „REGALE requirements, initial architecture, and evaluation plan"
- Mitarbeit an Deliverable D1.2 „REGALE Intermediate Architecture".
- Mitarbeit an Deliverable D1.3 "REGALE Final Architecture".

Arbeitspaket 3 „REGALE prototyping"

- Erweiterung der Query-API von DCDB um Aggregatsfunktionen (min, max, avg, etc.) und anderer mathematischer Funktionen (Integral, Ableitung).
- Definition von Schnittstellen und Entwicklung eines DCDB-Plugins zum Datenaustausch zwischen DCDB und der im Projekt von BSC entwickelten Energy Aware Runtime (EAR).
- Entwicklung eine Python-Moduls zur Abfrage von Monitoring-Daten aus DCDB als Pandas Dataframes.
- Mitwirkung an der Definition und Entwicklung der Integrationsszenarien und APIs für die in REGALE verwendeten Tools.
- Mitwirkung an der API-Spezifikation der REGALE Library.
- Mitarbeit am Deliverable D3.2 „REGALE prototype v2.0".

Arbeitspaket 6 „Innovation management"

- Erstellung eine Kommunikationsplans und Dissemination Toolkits (mit Logos, Farbpalette & Powerpoint-Template).
- Pflege und Generierung von Inhalten für die Projekt Website (https://regale-project.eu) und Social Media Kanäle (X & LinkedIn)
- Erstellung von Info- und Werbematerial, inkl. Grafiken, Flyer & Poster.
- Klassische PR & Medienarbeit, beispielsweise Medien-Pitch zur ISC23, Begleitung von Interviews mit Journalisten
- Projektvideos:

    - Konzeption, Dreh und Projektmanagement Projektvideo (https://youtu.be/wkzpgqxURI0?feature=shared)
    - Produktion und Veröffentlichung eines Videos, das die Ergebnisse des Projekts zusammenfasst. (https://youtu.be/yL3gYJ0iMeI?feature=shared)
- Kommunikation rund um das Webinar "Toolchains for Energy Efficient Supercomputing in the Exascale Era", in dem REGALE-Ergebnisse der Öffentlichkeit präsentiert wurden.
- Deliverables:
    - Hauptverantwortliche Erstellung von Deliverable D5.1 „Project Presentation & Web Portal".
    - Hauptverantwortliche Erstellung von Deliverable D5.3 „Dissemination, Communication & Stakeholder Engagement Plan".
    - Hauptverantwortliche Erstellung des Deliverable D5.5 „Dissemination, Communication & Exploitation Activities Report v1.0".
    - Hauptverantwortliche Erstellung des Deliverables 5.11 „Final Outreach Activity".
- Präsentation des Projekts auf Messen und Konferenzen:
    - Organisation des Projektauftritts auf der ISC23-Konferenz in Hamburg am EuroHPC-Stand.
    - Messestand auf der HiPEAC 24 Konferenz in Garching.
- Abschließende Medienarbeit, einschließlich einer abschließenden Pressemitteilung, die zu Berichterstattung in dem Branchenmedium InsideHPC sowie dem deutschen IT-Medium Datacenter Insider führte:
    - https://insidehpc.com/2024/03/eurohpc-ju-energy-efficiency-project-regale-comes-to-end/
    - https://www.datacenter-insider.de/forscher-wollen-den-strombedarf-von-supercomputern-senken-a-71f900d79c0f667569fe189caef222f9/

Wichtigste Positionen des zahlenmäßigen Nachweises

Personal

Nach dem Ausscheiden eines Mitarbeiters Ende 2021 konnte die Stelle für die technischen Arbeiten in WP1 und WP3 nicht neu besetzt werden. Die Personalmittel wurden deshalb nicht ausgeschöpft.

Dienstreisen

Aufgrund der COVID-19 Pandemie waren Dienstreisen zum Projektbeginn stark eingeschränkt, wodurch ursprünglich geplante Teilnahmen an Konferenzen nicht durchgeführt werden konnten. Die veranschlagten Mittel wurden deshalb nicht vollständig abgerufen.

Öffentlichkeitsarbeit

Ebenfalls bedingt durch die COVID-19 Pandemie war die Öffentlichkeitsarbeit zu Projektbeginn stark eingeschränkt, da Messen und Konferenzen abgesagt oder auf Online-Präsenz umgestellt wurden. Dadurch wurde das Budget für Öffentlichkeitsarbeit nicht ausgeschöpft.

Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Alle Projektarbeiten waren notwendig und wurden wie im Arbeitsplan festgelegt durchgeführt. Es gab keine nennenswerten Abweichungen vom Plan, deren Angemessenheit überprüft werden müsste.

Voraussichtlicher Nutzen

Alle im REGALE Projekt entwickelten Softwarekomponenten wurden als gemeinsamer Software-Stack veröffentlicht, der von anderen Rechenzentren genutzt werden kann. Viele der in REGALE entwickelten Technologien helfen bei der effizienteren Nutzung moderner HPC-Ressourcen und tragen somit zu einer besseren Auslastung und einem energie-effizienterem Betrieb selbiger bei.

Alle Arbeiten, die das LRZ an DCDB durchgeführt hat sind Teil der Open-Source Distribution von DCDB und im Rahmen der GNU Public License frei verfügbar. Das LRZ setzt DCDB auf seinen eigenen Produktivsystemen ein und nutzt es vorrangig zur Optimierung der Energieeffizienz des HPC-Betriebs.

Die im Projekt gewonnenen Erkenntnisse und Entwicklungen werden in zukünftigen Forschungsprojekten weitergeführt werden, beispielsweise im Rahmen des genehmigten EuroHPC-Projekts SEANERGYS.

Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Der Trend zu steigendem Stromverbrauch bei HPC-Systemen setzt sich ungebrochen fort. Die in REGALE entwickelten Technologien sind daher für den effizienten Einsatz solcher Systeme weiterhin essenziell.

Berichtsblatt

1. ISBN oder ISSN	**2. Berichtsart (Schlussbericht oder Veröffentlichung)** Schlussbericht
3. Titel REGALE – Offene Architektur für die nächste Generation von Exascale-fähigen HPC-Anwendungen im zukünftigen Supercomputing (BAdW-LRZ Teilprojekt)	
4. Autor(en) [Name(n), Vorname(n)] Schuffert Anita, Ott Michael	**5. Abschlussdatum des Vorhabens** 31.03.2024
	6. Veröffentlichungsdatum 30.09.2024
	7. Form der Publikation Document Control Sheet
8. Durchführende Institution(en) (Name, Adresse) Bayerische Akademie der Wissenschaften - Leibniz-Rechenzentrum (LRZ)	**9. Ber.-Nr. Durchführende Institution**
	10. Förderkennzeichen 16HPC040
	11. Seitenzahl 5
12. Fördernde Institution (Name, Adresse) BMBF	**13. Literaturangaben** 0
	14. Tabellen 0
	15. Abbildungen 0
16. DOI (Digital Object Identifier)	
17. Vorgelegt bei (Titel, Ort, Datum)	

18. Kurzfassung

Der Stromverbrauch von Hochleistungsrechnern steigt kontinuierlich an. Ziel des REGALE Projekts war es daher eine offene Softwarearchitektur für Hoch- und Höchstleistungsrechner zu entwickeln, die autonom die Ressourcennutzung überwacht, das Verhalten des Systems lernt, Ineffizienzen erkennt und die Ressourcenzuteilung an Anwendungen optimiert und dadurch die Energieeffizienz solcher Systeme erhöht.

Das REGALE Projekt hat wichtige Systemsoftwarekomponenten und Anwendungen für zukünftige europäische Exascale-Systeme entwickelt und optimiert. Dabei wurden bestehende Softwarekomponenten um neue Funktionalitäten erweitert und für den Einsatz auf Supercomputern optimiert. Der Einsatz des im Projekt entwickelten Software-Stacks wurde exemplarisch anhand von 6 Co-Pilot Anwendungen demonstriert. Es wurde eine API für die Kommunikation der Software-Stack Komponenten untereinander definiert und prototypisch in einer Softwarebibliothek umgesetzt.

Das Leibniz-Rechenzentren (LRZ) hat im Bereich der technischen Forschungs- und Entwicklungsarbeiten hauptsächlich bei der Anforderungsanalyse, der Definition der REGALE-Architektur und der Entwicklung der Use-Cases mitgewirkt. Außerdem hat das LRZ sein Monitoring Framework „DataCenter DataBase“ (DCDB) um neue Features erweitert und die Integration mit anderen Komponenten des REGALE-Stacks umgesetzt. Darüber hinaus hat das LRZ das Arbeitspaket zur Öffentlichkeitsarbeit geleitet und alle Aktivitäten zur Außendarstellung (Webauftritt, Social Media, Messestände) koordiniert und größtenteils auch umgesetzt.

19. Schlagwörter REGALE, HPC, Supercomputing, Energieeffizienz	
20. Verlag	**21. Preis**

Nicht änderbare Endfassung mit der Kennung 2462074-6

Document control sheet

1. ISBN or ISSN	**2. type of document (e.g. report, publication)** Veröffentlichung (Publikation)
3. title REGALE – Offene Architektur für die nächste Generation von Exascale-fähigen HPC-Anwendungen im zukünftigen Supercomputing (BAdW-LRZ Teilprojekt)	
4. author(s) (family name, first name(s)) Schuffert Anita, Ott Michael	**5. end of project** 31.03.2024
	6. publication date 30.09.2024
	7. form of publication Document Control Sheet
8. performing organization(s) name, address Bayerische Akademie der Wissenschaften - Leibniz-Rechenzentrum (LRZ)	**9. orginators report no.**
	10. reference no. 16HPC040
	11. no. of pages 5
12. sponsering agency (name, address) BMBF	**13. no. of references** 0
	14. no. of tables 0
	15. no. of figures 0
16. DOI (Digital Object Identifier)	
17. presented at (title, place, date)	
18. abstract The power consumption of high-performance computers (HPC) is continuously increasing. The aim of the REGALE project was therefore to develop an open software architecture for HPC systems that autonomously monitors resource usage, learns the behaviour of the system, detects inefficiencies and optimises the allocation of resources to applications, thereby increasing the energy efficiency of such systems. The REGALE project has developed and optimised important system software components and applications for future European exascale systems. New functionalities were added to existing software components and optimised for use on supercomputers. The use of the software stack developed in the project was demonstrated using 6 co-pilot applications as examples. An API for communication between the software stack components was defined and prototypically implemented in a software library. In terms of research and development work, the Leibniz Supercomputing Centre (LRZ) was mainly involved in the requirements analysis, the definition of the REGALE architecture and the development of the use cases. The LRZ also added new features to its monitoring framework 'DataCenter DataBase' (DCDB) and implemented integration with other components of the REGALE stack. In addition, the LRZ managed the dissemination work package and organised all public relations activties (web presence, social media, etc.). Translated with DeepL.com (free version)	
19. keywords REGALE, HPC, Supercomputing, energy efficiency	
20. publisher	**21. price**

Nicht änderbare Endfassung mit der Kennung 2409887-4