
Digitale Lehrbriefe mit Jupyter Notebooks

Prof. Dr. Heiner Giefers¹ und Prof. Dr. Christian Plessl²

¹Fachhochschule Südwestfalen

²Universität Paderborn

Forschendes Lernen erfordert Zugang zu Informationen und Inhalten, aber auch zu Technologien, auf deren Grundlage mit Erfahrungswissen Innovationen entstehen können. Ziel dieses Projekts war es, eine Plattform zu schaffen, die Studierende dabei unterstützt, niederschwellig und praxisorientiert Kompetenzen im Programmieren und in den datengetriebenen Wissenschaften zu entwickeln. Dazu erstellen wir interaktive Lernmaterialien auf Basis von Jupyter Notebooks, welche Lehrinhalte, Programmcode und daraus erzeugte Resultate in einem Dokument integrieren. Durch eine automatische Auslagerung der Berechnungen auf Hochleistungsrechner sind auch sehr komplexe, anwendungsrelevante Fallstudien möglich. Die erstellten Materialien und Werkzeuge wurden in unseren Lehrveranstaltungen eingesetzt und evaluiert. Die notwendige IT-Infrastruktur wurde im Projekt als Referenzplattform entwickelt, erprobt und veröffentlicht [1], um perspektivisch weit über den Kontext des Projektes Anwendung zu finden.

1 Motivation

Der technische Kern dieses Projekts ist eine für die Studierenden transparente Integration von digitalen Arbeitsblättern (sog. Notebooks), in denen am eigenen Laptop Lerninhalte und Programme dargestellt, bearbeitet und ausgeführt werden können, mit aktuellen Cloud- und High-Performance Computing (HPC)-Technologien. Dadurch wird die Voraussetzung für innovative, digitale Lehre geschaffen, die zum eigenständigen, interaktiven Erarbeiten von z. B. Data Science Anwendungen ermuntert, ohne den Studierenden die Komplexität der Nutzung eines Hochleistungsrechners aufzubürden. Stattdessen werden besonders rechen- und datenintensive Berechnungen automatisch auf einen Hochleistungsrechner ausgelagert, wodurch auch äußerst anspruchsvolle Analysen durchgeführt werden können, die auf den Laptops der Studierenden nicht machbar wären.

2 Beschreibung der Lehrinnovation

Im Rahmen dieses Projektes haben wir zwei Lehrinnovationen weiterentwickelt, um die Effizienz des Selbststudiums für die Studierenden zu verbessern. Der für die Studierenden sichtbare Teil sind **digitale, interaktive Lehrbriefe**, in denen Modulinhalte multimedial aufbereitet und anhand von Beispielen und Fallstudien demonstriert werden. Die Integration der Beispiele und Fallstudien als ausführbare und veränderbare Programme ermöglicht das individuelle Erfahren des Stoffes im Sinne eines konstruktivistischen Lernverständnisses. Ein zweiter wichtiger Teil ist eine Web-gestützte **Ausführungsplattform**, die den Studierenden über die Lehrbriefe den Zugriff auf hochperformante Computersysteme auf einfache Art und Weise ermöglicht. Wir schätzen diesen hürdenlosen Zugriff auf moderne Technologie als essenziell für das Erlangen der relevanten Fachkompetenzen ein.

Die Umsetzung dieses Ansatzes baut zu einem Großteil auf bestehende Werkzeuge und Ressourcen auf. Jupyter Notebooks [6, 10] sind eine interaktive Entwicklungsumgebung, die Programmcode und formatierte Inhalte wie Text, Formeln, Grafiken und Animationen auf einzigartige Weise integrieren. Die Dokumente bestehen aus einer Folge von Zellen, die allesamt editierbar sind. Der Leser wird so zum Co-Autor des Dokuments, der an jeder Stelle Änderungen, Ergänzungen, Experimente, etc. vornehmen kann. Die Programmcode-Elemente können Ausgaben erzeugen, die formatiert und interaktiv bedienbar im Notebook dargestellt werden. Diese Eigenschaften haben Notebooks zum produktivsten Werkzeug in den Disziplinen Data Science und KI gemacht. Die Art und Weise, wie Notebooks Dokumentation und Experiment in einem Artefakt miteinander verbinden, macht sie zudem zu einem idealen Medium für die Lehre.

Eine Multi-User-Umgebung, auf der Notebooks serverseitig ausgeführt werden können ist JupyterHub¹. Der Nutzer benötigt lediglich einen Internet-fähigen Computer und kann damit die Notebooks online editieren und ausführen oder zur offline-Bearbeitung herunterladen. Eine JupyterHub-Umgebung kann mit überschaubarem Investitionsausgaben auch an kleineren Bildungseinrichtungen realisiert werden. Allerdings ist die zugrundeliegende Hardware dann nicht für rechenintensive Algorithmen geeignet, wie sie heutzutage in vielen modernen Anwendungen aus Wissenschaft und Technik – wissenschaftliche Simulationen und KI sind hier nur zwei Beispiele – zum Einsatz kommen. Um diese Lücke zu schließen und den Studierenden die Möglichkeit zu geben, relevante Anwendungen zu erproben und weiterzuentwickeln, haben wir die JupyterHub-Umgebung an bestehende HPC-Rechenressourcen in Paderborn gekoppelt. Dazu musste eine benutzerfreundliche Schnittstelle entwickelt werden, über welche die Nutzer die Ausführung eines Jupyter Notebooks auf den HPC-Rechner auslagern können. Grundlegende Technologien für die Auslagerung auf Cloud-Rechenzentren wurden bereits vorgestellt [8, 12], können für HPC-Systeme aber nicht direkt angewendet werden, weil diese andere Technologien zum Software-Management und Deployment einsetzen. Im Rahmen dieses Projektes wurde eine passende Lösung für die HPC-Systeme in Paderborn entwickelt und als Open-Source-Referenzprojekt veröffentlicht.

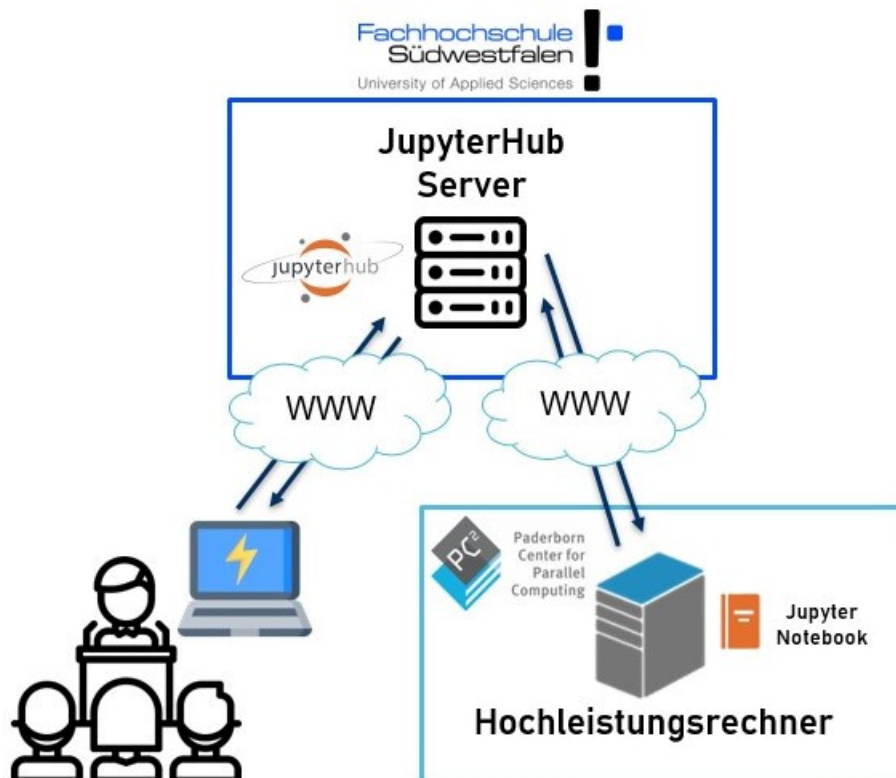


Abbildung 1: Systemaufbau der Lernumgebung

Abbildung 1 zeigt den Systemaufbau der Lernumgebung. Dozenten stellen über den JupyterHub die Lehrmaterialien in Form von Notebook-Dokumenten bereit. Die Studierenden loggen sich über einen Internet-

¹<https://jupyter.org/hub> (abgerufen am 31.05.2021)

Browser an der Plattform an und wählen einzelne Notebooks zur Bearbeitung aus. Durch die Verwendung von ausschließlich quelloffener und allgemein verfügbarer Software ist gewährleistet, dass die Studierenden die Notebooks auch auf ihren privaten Computern offline verwenden können. Über die `nbgrader`-Erweiterung sind Aufgaben zur Selbstkontrolle in die Lehrmaterialien integriert. Die Studierenden lösen die Aufgaben und können die Korrektheit ihrer Ergebnisse durch eingebaute Tests verifizieren. Abschließend übermitteln Sie die Lösungen zurück ins System, wo sie für den Dozenten zur Endkorrektur angezeigt und zur Leistungsstanderhebung genutzt werden können.

3 Umsetzung und Ergebnisse der Lehrinnovation

Aus dem Projekt hervorgegangen sind Innovationen im Bereich der Systemarchitektur sowie neu erstellte Lehrmaterialien für insgesamt 3 Module an der Fachhochschule Südwestfalen. Alle Materialien sowie die erstellte Software sind unter Permissive-Lizenzen auf [Github.com](https://github.com) veröffentlicht.

Systemarchitektur

An der Fachhochschule Südwestfalen wurden Server für die JupyterHub-Installation bereitgestellt. Im HPC-Cluster in Paderborn wurde ein Rechenknoten für die Ankopplung an den JupyterHub-Server zur Verfügung gestellt. Die Benutzeranmeldung erfolgt am JupyterHub auf Seiten der Fachhochschule. Damit beide Seiten (JupyterHub und HPC-System) verschlüsselt miteinander kommunizieren können, wurde ein *SSH-Tunnel* eingerichtet, der nach dem Benutzerlogin automatisch aufgebaut und während der gesamten Session aufrechterhalten wird. Um Notebooks remote auf dem Rechenknoten ausführen zu können, mussten auf dem JupyterHub-Server spezielle *Spawner* eingerichtet werden. Mithilfe des *Wrapspawners* werden userspezifische Optionen, wie zum Beispiel GPU-Typ oder CPU-Kerne, im Webbrowser abgefragt und an den *Batchspawner* übergeben. Der *Batchspawner* kommuniziert mit dem Workload-Manager des entfernten Systems und startet die Notebooks.

Das HPC-System stellt für die Berechnungen verschiedene Hardware (CPUs und GPUs) bereit. Nach dem Benutzerlogin auf dem FH-Server kann die gewünschte Hardware sowie die Größe des zu verwendenden Arbeitsspeichers ausgewählt werden. Auf dem HPC-System kommt *Slurm* als Workload-Manager zum Einsatz. Dieser ist für das Starten von Jobs verantwortlich und musste für unsere Zwecke entsprechend konfiguriert werden. Die Notebooks werden in vordefinierten *Singularity-Containern* gestartet. Um die Einrichtung des JupyterHub Dienstes zu vereinfachen, wurde ein *Ansible-Playbooks* entwickelt, das alle nötigen Installationsschritte automatisiert durchführt. Für die erforderlichen Installationen und Konfigurationen auf dem HPC-System wurde ein Installations-Wizard entwickelt, welcher durch die einzelnen Installationsschritte führt.

Nach der Benutzeranmeldung wird ein entsprechender Container geladen, in welchem die Ausführung der Notebooks abgekapselt von allen anderen Nutzern des Systems stattfinden. Dies hat zum einen den Vorteil, dass die Notebooks in einer gesicherten Umgebung ausgeführt werden, wodurch die meisten Angriffsszenarien von vornherein ausgeschlossen sind. Zum anderen wird es dem Benutzer auf diese Weise ermöglicht, eigene Pakete und Tools zu installieren, die er für seine Berechnungen einsetzen möchte. Beim Starten eines Containers werden die Benutzereigenen Informationen mit eingehängt, sodass alle Dateien, Notebooks sowie installierte Tools oder Pakete selbst nach erneuter Anmeldung erhalten bleiben. Zusätzlich haben die Studierenden die Möglichkeit, ihr Sciebo [9] Laufwerk einzubinden, um so direkten Zugriff auf ihre Daten, auch ausserhalb der JupyterHub Umgebung, zu haben.

Zur Authentifizierung mit dem Jupyterhub wurden zwei alternative Methoden evaluiert. Die Moodle e-Learning Plattform unterstützt den *Learning Tools Interoperability* (LTI) Standard [7]. Mit einem entsprechenden Plugin auf JupyterHub-Seite, können Studierende direkt über die Kursseite eines Moduls zu einem Jupyter Notebook geleitet werden. Der Inhalt wird als Moodle-Aufgabe angelgt, was es erlaubt, Ergebnisse an das Moodle System zurückzumelden. Um die Plattform auch Kurs- und Hochschul-übergreifend zugänglich zu machen, wurde im Rahmen einer Bachelorarbeit untersucht, wie man den Dienst über das im Hochschul- und Forschungsumfeld weit verbreitete und von DFN-Verein zentral gemanagte Shibboleth Protokoll abwickeln kann. Hierüber können Studierende der FH Südwestfalen und optional auch weiterer

Bildungseinrichtungen, mit ihren persönlichen Accounts Zugriff auf die Plattform erhalten, ohne das auf Seiten des Dienstes Benutzeraccounts gespeichert und verwaltet werden müssen.

Schon parallel zur Entwicklung wurde das System von einer studentischen Hilfskraft ausgiebig getestet, um frühzeitig Fehler beheben zu können. Verbesserungsvorschläge sind kontinuierlich mit in das Projekt eingeflossen, um Fehlentwicklungen zu vermeiden. Im Anschluss wurde die entwickelte Plattform einem *Benchmarktest* unterzogen. Insbesondere wurde die Performance von CPUs und GPUs untersucht, indem gängige Algorithmen und Trainingsmethoden aus dem Bereich der künstlichen Intelligenz ausgeführt wurden. Im Vergleich mit frei verfügbaren Machinelearning-Plattformen wie *Colab* und *Kaggle* schnitt unser System gut ab. Die Ergebnisse sind auf der GitHub-Projektseite einsehbar. Am Ende der Projektlaufzeit steht ein voll einsatzbereites und getestetes System bereit.

Inhalte

Im Rahmen des Projekts wurden Lehrmaterialien erarbeitet, die verschiedenen Modulen an der FH Südwestfalen eingesetzt werden. Seit dem Sommersemester 2021 werden die Jupyter Notebooks in den Lehrveranstaltungen *Skriptsprachen* (Verbundstudiengang Angewandte Informatik Bachelor), *Machine Learning* (Verbundstudiengang Angewandte Informatik Master) und *Einführung Machine Learning* (Informatik Bachelor) eingesetzt. Damit sich die Unterlagen vor allem für das Selbststudium eignen, wurden zahlreiche Übungsaufgaben integriert, die mittels eingebauter Funktionstests weitgehend automatisch evaluiert werden können. Die Notebooks werden über [Github.com](https://github.com) veröffentlicht [2–4] und vom den Jupyter Notebook Servern automatisch über `nbgitpuller` [11] ausgechecked, wenn die Studierenden auf eine entsprechende Aufgabe in Moodle zugreifen. Die Notebooks mit den Test- und Validierungscodes sind nicht-öffentlich und werden auf JupyterHub-Seite verwendet, um Lösungseinreichungen der Studierenden automatisiert zu überprüfen und zu bewerten. Hierzu kommt das *NBGrader*-System zum Einsatz [5]. Die Ergebnisse der Bewertung werden über die LTI-Schnittstelle zurück an den Moodle-Kurs gemeldet und sind so für die Studierenden über die Moodle-Seite zum Kurs einsehbar. Über diesen Aufbau lassen sich eine Vielzahl von Programmierübungen mit einem relativ geringen Betreuungsaufwand auch für eine große Gruppe von Studierenden realisieren. Die Studierenden begrüßen die Bereitstellung der Aufgaben und heben die Plattform bei der semesterweisen Veranstaltungsevaluation lobend hervor. Vereinzelt Kritik geht in Richtung der automatischen Funktionstests. Hier ist nicht immer ganz ersichtlich, welche Eigenschaften der Lösung genau überprüft werden und Tests für scheinbar richtige Lösungen können in Ausnahmefällen scheitern. Wir haben das Feedback des Studierenden im ersten Durchlauf genutzt, um die Materialien zu überarbeiten.

4 Verstetigung und Übertragbarkeit der Lehrinnovation

Die Lehrmaterialien werden in kommenden Durchläufen der o.g. Modulen erneut eingesetzt und erweitert. Darüber hinaus sollen weitere Module die Plattform nutzen, auch für diese Veranstaltungen werden neue Lehrmaterialien erstellt. Beispiele sind Module zu den Themengebieten Programmierung und Maschinelles Lernen des neuen Master-Verbundstudiengangs *Angewandte Künstliche Intelligenz*, der im Wintersemester 2021/22 erstmalig an der FH Südwestfalen angeboten wird. Die Veröffentlichung der Lehrmaterialien über eine Permissive-Lizenz erlaubt Dritten die Verwendung und Weiterentwicklung der Notebooks.

5 Fazit

Wir verstehen das entwickelte Projekt als prototypische Lehrplattform. Mit dem Projekt wurde eine Referenzplattform geschaffen, die zeigt, wie die IT-Infrastrukturen der Hochschulen so an landesweite Rechenzentren gekoppelt werden können, dass auch Nutzer ohne Erfahrungen im Hochleistungsrechnen die HPC-Ressourcen niedrigschwellig nutzen können. Eine Verstetigung und Weiterentwicklung der Plattform durch andere Fördermittel über die ursprüngliche Projektlaufzeit hinaus ist ein wesentliches Ziel. Nach der erfolgreichen Testphase und der Etablierung erster Module auf der Plattform ist eine Nutzung durch weitere Dozenten aus unterschiedlichen Fachbereichen an beiden Hochschulen vorgesehen. In den Informatik-Curricula

der FH Südwestfalen gibt es verschiedene Lehrveranstaltungen, in denen Notebooks sinnvoll eingesetzt werden können. Die Veröffentlichung der Lösung als Open Source erlaubt es auch, weitere Hochschulen an die Plattform anzubinden.

Referenzen

- [1] Universität Paderborn und Fachhochschule Südwestfalen. *JupyterHub HPC Interface*. <https://github.com/pc2/JHub-HPC-Interface>. (18.05.2021).
- [2] Heiner Giefers. *Materialien zum Modul Einführung Machine Learning*. <https://github.com/fh-swf-hgi/intro-ml>. 2021.
- [3] Heiner Giefers. *Materialien zum Modul Machine Learning*. <https://github.com/fh-swf-hgi/ml>. 2020.
- [4] Heiner Giefers. *Materialien zum Modul Skriptsprachen*. <https://github.com/fh-swf-hgi/skriptsprachen-ex>. 2020.
- [5] Project Jupyter u. a. „nbgrader: A Tool for Creating and Grading Assignments in the Jupyter Notebook“. In: *Journal of Open Source Education* 2.16 (2019).
- [6] Thomas Kluyver u. a. „Jupyter Notebooks – a publishing format for reproducible computational workflows“. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Hrsg. von F. Loizides und B. Schmidt. IOS Press. 2016, S. 87–90.
- [7] *Learning Tools Interoperability Standard*. <http://www.imsglobal.org/activity/learning-tools-interoperability>. (12.07.2021).
- [8] Michael B. Milligan. „Jupyter As Common Technology Platform for Interactive HPC Services“. In: *Proceedings of the Practice and Experience on Advanced Research Computing*. PEARC '18. ACM, 2018, 17:1–17:6.
- [9] *Sciebo hochschulcloud.nrw*. <https://hochschulcloud.nrw>. 2021.
- [10] Helen Shen. „Interactive notebooks: Sharing the code“. In: *Nature* 515 (2014), S. 151–152.
- [11] nbgitpuller Team. *nbgitpuller*. Stand 12.07.2021. URL: <https://jupyterhub.github.io/nbgitpuller>.
- [12] Christopher J. Woods. *Running HPC Workloads via Jupyter using Kubernetes and Fn serverless functions*. CNCF Meetup. Apr. 2019.