

# JupyterHub

Interaktive webbasierte Datenanalyse-, Visualisierungs- und Entwicklungsplattform

Dr. Markus Blank-Burian



# Forschungsdaten

- Forschungsdaten können in der Regel einfach erzeugt werden
  - Hardware (Messgeräte, Mikroskope, Kameras, ...)
  - HPC (Simulationen)
  - Händisch (z.B. Tabellenkalkulation)
- Verschiedene Formate
  - Binär (HDF5, diverse Bildformate, herstellerspezifische Formate, selbsterfundene Formate, ...)
  - Text (z.B. CSV)
  - Datenbank

# Speicherorte für Forschungsdaten

- Lokaler Rechner
  - Kleine bis mittelgroße Datenmengen
- Sciebo
  - Kleine bis mittelgroße Datenmengen
- Palma
  - Beliebig große Datenmengen
- WWU Cloud
  - Beliebig große Datenmengen

Daten können je nach Größe auch redundant auf mehreren Systemen liegen

# Auswertung der Daten

- Nicht-Interaktiv
  - Analyseprogramme und Scripts
  - Lokaler Rechner für kleine Datenmengen
  - HPC für große Datenmengen
- Interaktiv
  - Spezialsoftware
  - Gnuplot, Excel
  - Jupyter

Interaktive Aufbereitung meist auch nach Nicht-Interaktiven Analyse nötig!

# Jupyter Notebooks

- Ursprünglich entwickelt zur interaktiven Datenanalyse in **Julia**, **Python** und **R**
- Notebook ist Mischung aus Text, Code und Daten-Visualisierung
- Wird traditionell auf lokalem Rechner gestartet
- Mittlerweile viele Kernel für verschiedene Programmiersprachen verfügbar
- Zahlreiche Extensions zur Visualisierung verfügbar
  
- Problem: Wie komme ich an meine Forschungsdaten?
  - Kopieren auf lokalen Rechner wegen Datengröße nicht immer möglich
  - Teilweise umständlicher Workflow
  - Remote-Zugriff auf Notebook möglich

# JupyterHub

- Webanwendung zum Starten von Jupyter-Sessions
- Sessions laufen in Kubernetes-Cluster "zivkube" auf WWU Cloud als "Pods"
- Anzahl der CPUs und RAM auswählbar
  - Limits auf Basis von Benutzergruppen (Studierende, Mitarbeiter)
  - Gruppen von Studierenden werden auf Anfrage zugelassen (aktuell noch keine)
- Besonderheit: Benutzer starten Session mit korrekter UID, GID
  - Zugriff auf Palma (/home und /scratch) und Benutzerfreigaben in WWU Cloud möglich
  - Zugriff auf JupyterHub Home-Verzeichnis: `\\wwu.de\ddfs\Cloud\zivkube\jupyterhub_homes`

# Jupyter Images

- Jupyter Sessions werden durch Docker-Images erzeugt
- Vom ZIV bereitgestellte Jupyter-Images:
  - eScience: Fokus auf wissenschaftliche Datenanalyse und Visualisierung
  - Development: Fokus auf Softwareentwicklung in verschiedenen Programmiersprachen
  - Installation von Bibliotheken und Anwendungen auf Anfrage möglich
- Mitarbeiter dürfen auch beliebige Images starten, aber ohne Datenzugriff auf Palma/WWU Cloud

# JupyterLab

- Weiterentwicklung "JupyterLab" als interaktive Entwicklungsumgebung für Notebooks
  - Tabs für Launcher, Notebooks und Terminals
  - Sidebar für Filetree, Git und Sessions
- Geschrieben in TypeScript
- Erweiterbar durch JupyterLab-Extensions
  - Datenvisualisierung für Notebooks
  - Webanwendungen
  - X11-Anwendungen (mit noVNC als Webanwendung)
  - HPC Jobübersicht



# Web-Anwendungen

- Visual Studio Code
  - Web-Version der IDE zur Softwareentwicklung
  - Integrierter Python, Java-Debugger und C++-Debugger
- R Studio
  - IDE für Softwareentwicklung in R
- Shiny Server
  - Anzeige und Testen von in R geschriebenen Shiny Apps

# X11-Anwendungen

- Integration von X11-Anwendungen über noVNC
- Beliebige X11-Anwendungen im Browser verfügbar
  
- Blender
- Mathematica, Matlab
- Voreen, Paraview, VMD
- Grace
- Starten von eigenen X11-Anwendungen möglich!

**Serverseitige GPU-Unterstützung zur Datenvisualisierung in Planung**

# Notebook Kernel und Programmiersprachen

- Python, R, Julia
- Gnuplot
- SageMath, Octave, Scilab, Mathematica
- C/C++ mit GCC, Clang und Xeus-Kernel
- Go
- Java
- Scheme
- (Haskell)
- Clojure, Groovy, Kotlin, Scala
- SQL

# Anwendungen und Bibliotheken

- Machine-Learning: Scikit-Learn, Keras, Torch, Heat
- Diverse Python Mathematik-Pakete (numpy, scipy, sympy ...)
- Root
- Gromacs
  
- Dask
  - Starten von Worker-Tasks auf Palma möglich
  - Transparente Integration für scikit-learn
  - Langlebige Tensorflow-Tasks auf Palma

## Zielgruppe

- Mitarbeiter (insbesondere Wissenschaftliche Mitarbeiter, Doktoranden)
  - Mehr CPUs/RAM
  - GPU-Unterstützung (in Arbeit, zunächst nur Hardware-Rendering für X11-Anwendungen)
  - Custom Images
- Studierende
  - Limitierte Auswahl an CPU/RAM
  - Keine GPU-Unterstützung
  - Keine Custom Images
  - Vorerst nur Freigabe von Gruppen auf Anfrage der Fachbereiche

# Support und Knowledgebase

Mattermost (Chat): Team WWU, Kanal JupyterHub

Confluence-Integration:

- Upload von Notebooks nach Confluence möglich
- Aufbau einer Knowledgebase
- Alle Mitarbeiter können Schreibzugriff bekommen
- <https://zivconfluence.uni-muenster.de/display/jupyterhub/>

Support seitens ZIV nur für technische Probleme und Softwareinstallation

**Keine Programmierberatung außerhalb vom KB-Artikeln möglich**

# Vielen Dank für die Aufmerksamkeit!

<https://jupyterhub.wwu.de>