TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Dr. Christoph Lehmann
Zentrum für Informationsdienste und Hochleistungsrechnen – TU Dresden

AI for Teaching

Cloud-Computing, Data Management and High Performance Computing

Dresden, 15 October 2020

Introduction – Agenda

01 Introduction

- Goal: teaching methods and approaches from areas as simulation, data analytics, machine learning etc.
- today's audience: teachers, not: developers
- students should be able to work on practical examples themselves
- Common main requirements:
  1. unified software environment
  2. appropriate hardware
  3. central access to course material, esp. prepared examples, data sets, source codes etc.
- different modes for permission and access rights:
  - anonymous use, no change of data
  - personalized use, TU Dresden login (aka ZIH login) necessary
    (i) changing data is possible, results are *not* kept permanently
    (ii) changing data is possible, results are kept permanently
  - HPC login needed: can be personalized, but also temporary logins possible
- no universal approach, typically individual combination of tools and services depending on application and number of students

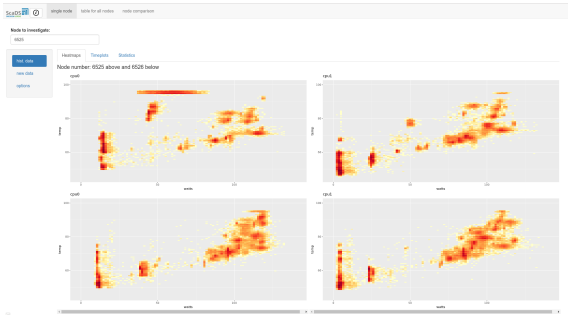
01 Introduction

- today's goal: provide some ideas about tools and services and how to combine them
- provide the overall picture, nevertheless there are some details behind the scenes
- starting point: what students will see and how to make this reality
- all use cases are real world examples (already tested and applied)
- typically, there is some effort needed by the teacher (installation, configuration etc.)

Use Cases – Agenda

02 Use Case I – Simple Demonstrator

- students get in first touch with a topic; only a few degrees of freedom for interactivity
- accessible within the TU Dresden data network, anonymous use
- typically some kind of web app using web frameworks as e. g. flask (for Python) or shiny (for R)
- level of interactivity strongly depends on the web app itself
- is running on a virtual machine (VM)
- example for shiny app based on R: http://172.24.249.12:3405/

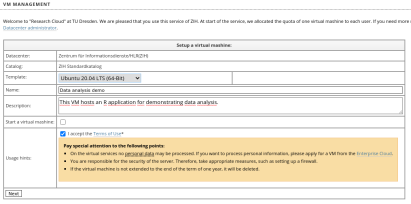


02 Use Case I – Simple Demonstrator

- VM configuration: RAM 4GB, 2 CPUs, HDD 40GB, Ubuntu 18.04.
- setting was applied successfully with approximately 20-30 students
- prerequisites and effort:
  1. create web app
  2. apply for a VM (research cloud or enterprise cloud)
  3. install and configure VM according to the app's requirements
  4. move web app to VM
  5. maintain/update VM and your web app if necessary
- helpful links:
  a) cloud services at ZIH: https://selfservice.zih.tu-dresden.de/...cloud_dienste/
  b) web framework shiny for R: https://mastering-shiny.org/, https://shiny.rstudio.com/tutorial/
  c) web frameworks for Python: https://palletsprojects.com/p/flask/, http://bottlepy.org/

02 Cloud Service at ZIH – Create Virtual Machine

- How to create a VM? (example research cloud)
  1. application form: https://selfservice.zih.tu-dresden/.../cloud_dienste/vm_create


VM MANAGEMENT

Welcome to "Research Cloud" at TU Dresden. We are pleased that you use this service of ZIH. At start of the service, we allocated the quota of one virtual machine to each user. If you need more resources, you can request them directly under "My Settings" in your Datacenter administrator.

Setup a virtual machine:

Datacenter:	Zentrum für Informationsdienste/HLR(ZIH)
Catalog:	ZIH Standardkatalog
Template:	Ubuntu 20.04 LTS (64-Bit)
Name:	Data analysis demo
Description:	This VM hosts an R application for demonstrating data analysis.
Start a virtual machine:	
Usage hints:	I accept the Terms of Use* Pay special attention to the following points: • On the virtual services no personal data may be processed. If you want to process personal information, please apply for a VM from the Enterprise Cloud. • You are responsible for the security of the server. Therefore, take appropriate measures, such as setting up a firewall. • If the virtual machine is not extended to the end of the term of one year, it will be deleted.

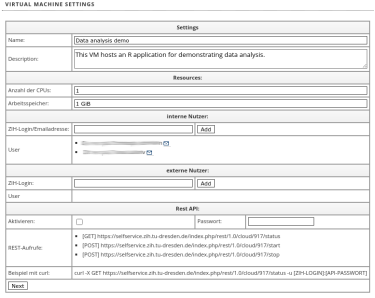
Next

  2. select an operating system that matches your needs (currently available: some Linux distros and Windows Server 2016/2019)
  3. newly created VM in research cloud is available within a couple of minutes
  4. check your VMs at: https://selfservice.zih.tu-dresden.de/.../cloud_dienste/vm

Name ∧	Host	Datacenter	State	Lease	Control	Action
Data analysis demo	172.	Zentrum für Informationsdienste/HLR(ZIH)	Stoped	2020-12-19		
Image Retrieval Demo	172.	Zentrum für Informationsdienste/HLR(ZIH)	Running	2021-08-05		

02 Cloud Service at ZIH – Create Virtual Machine

– choose "configure" (red circle, see slide above) to select final settings (number of CPUs and RAM) that fit your needs https://selfservice.zih.tu-dresden.de/.../cloud_dienste/vm


VIRTUAL MACHINE SETTINGS

Settings	
Name:	Data analysis demo
Description:	This VM hosts an R application for demonstrating data analysis.
Resources:	
Anzahl der CPUs:	1
Arbeitsspeicher:	1 GB
interne Nutzer:	
ZIH-Login/Emailadresse:	Add
User	[redacted]
externe Nutzer:	
ZIH-Login:	Add
User	
Rest API:	
Aktivieren:	Passwort:
REST-Aufrufe:	• [GET] https://selfservice.zih.tu-dresden.de/index.php/rest/1.0/cloud/917/status • [POST] https://selfservice.zih.tu-dresden.de/index.php/rest/1.0/cloud/917/start • [POST] https://selfservice.zih.tu-dresden.de/index.php/rest/1.0/cloud/917/stop
Beispiel mit curl:	curl -X GET https://selfservice.zih.tu-dresden.de/index.php/rest/1.0/cloud/917/status -u [ZIH-LOGIN]:[API-PASSWORT]

Next

– hint: default resources are limited to 2 CPUs, 8 GB RAM, 100 GB disk space (if necessary it is possible to get more)

02 Cloud Service at ZIH – Create Virtual Machine

- you will be notified by mail when the VM is available
- after starting the VM at the management page, it can be accessed by ssh (for Windows users: use an ssh-client or VMware)
- login credentials can be found at https://selfservice.zih.tu-dresden.de/.../cloud_Dienste/task
- general hints for the research cloud (details can be found at https://selfservice.zih.tu-dresden.de/.../cloud_dienste/AGB)
  - VMs are allowed only to be used for research, teaching and administration tasks for TU Dresden
  - VM-lifetime is limited to one year (can be extended)
  - processing of personal data is not allowed
  - VMs are no systems for high availability (this is different for the enterprise cloud)
  - VMs are not protected by ZIH-firewalls and the teacher/originator is responsible for updates and maintenance of the operating system
  - research cloud VM can be transformed into enterprise cloud VM: use research cloud for prototyping

02 Agenda

02 Use Case II – Advanced Interactivity

- students should be able to interact and try things out (including failures)
- working through prepared examples to get a deeper understanding or developing on their own
- a uniform software environment forms the basis, here: Python environment within a jupyterhub
- jupyterhub allows personalized access to jupyter notebooks and allows for user-specific workspace on shared resources
- remarks:
  1. login management comes into play as jupyterhub considers different users
  2. jupyterhub and the notebooks are not restricted to Python (more on that later)
- Thanks to Jens Brose, who provided this use case.
- live-demo example: jupyterhub for study of physics with login management via OPAL, https://bildungsportal.sachsen.de/opal/.../22770614272/

02 Use Case II – Advanced Interactivity

- VM (enterprise cloud) configuration: RAM 8GB, 4 CPUs
- sufficient resources for parallel access of ≈ 20 students, overall ≈ 400 students over the semester, 2 virtual experiments per student
- prerequisites and effort:
  1. create jupyter notebook(s) with your teaching content
  2. apply for a VM (research cloud or enterprise cloud)
  3. install and configure necessary software packages and jupyterhub on VM
  4. use connection to OPAL course for login management by using the LTI module (https://www.bps-system.de/help/display/LMS/LTI-Tool)
  5. maintain/update VM and your notebook(s) if necessary
- helpful links:
  a) cloud services at ZIH: https://selfservice.zih.tu-dresden.de/.../cloud_dienste/
  b) jupyterhub: https://jupyter.org/hub
  c) great flexibility for jupyter notebooks; available kernels: https://github.com/jupyter/jupyter/wiki/Jupyter-kernels

02 Agenda

02 Use Case III – Interactivity on HPC

- students should be able to interact and try things out (including failures) based on strong hardware, e.g. using multiple CPUs, GPUs and/or large memory
- working through prepared examples to get a deeper understanding or developing on their own
- basis here: uniform Python environment within a jupyterhub on HPC machine (taurus)
- login management for HPC machine necessary
- Thanks to Simon Stone, who provided this use case.
- live-demo example: jupyterhub for „Praktikum, Angewandte Intelligente Signalverarbeitung“
- this example is uses data from github repo https://github.com/nullpunktTUD/tud-ais
  1. request resources at https://taurus.hrsk.tu-dresden.de/jupyter/hub/home
  2. pull data from github repo into user home:
     https://taurus.hrsk.tu-dresden.de/jupyter/hub/user-redirect/
     git-pull?repo=https%3A%2F%2Fgithub.com
     %2FnullpunktTUD%2Ftud-ais&urlpath=lab%2Ftree%2Ftud-ais%2F

02 Use Case III – Interactivity on HPC

- resources were used by ≈ 20 students
- used a weekly reservation (8 weeks) for 90min, 120 CPUs
- prerequisites and effort (details on resources, login and reservation see slides 24, 26, 28):
  1. create jupyter notebook(s) with your teaching content
  2. apply for an HPC project according to your resource needs
  3. install and configure software packages for your project if necessary (e.g. create new kernels)
  4. apply for reservation
  5. provide link for students to join the HPC project
  6. maintain/update your own kernels and notebook(s) if necessary
- helpful links:
  a) general hints for HPC project in teaching: https://tu-dresden.de/zih/hochleistungsrechnen/zugang/hpclehre
  b) apply for HPC project: https://hpcprojekte.zih.tu-dresden.de/
  c) general hints for jupyterhub on taurus: https://doc.zih.tu-dresden.de/.../Compendium/JupyterHub
  d) details for jupyterhub in teaching: https://doc.zih.tu-dresden.de/.../JupyterHubForTeaching

02 Agenda

02 Hints for HPC – Available Software

– check available software from module system on Taurus at:
https://doc.zih.tu-dresden.de/hpc-wiki/.../SoftwareModulesList

Tcl	lang	2020-08-11	-	8.6.9 8.6.8 8.6.7 8.6.10
TensorFlow	lib	2020-03-20	-	2.1.0 2.0.0 1.8.0 1.15.0 1.10.0
Tk	vis	2020-08-11	-	8.6.9 8.6.8 8.6.10
Tkinter	lang	2020-08-11	-	3.8.2 3.7.4 3.7.2

Pillow	vis	2020-06-24	-	6.2.1
PowerAI	data	2019-12-10	-	1.7.0_a0 1.6.1
PyTorch	devel	2020-09-29	-	1.6.0 1.3.1 1.1.0
PyTorch-Geometric	devel	2020-09-29	-	1.6.1
PyYAML	lib	2020-02-18	-	5.1.2 3.13
Python	lang	2020-07-30	-	3.7.4 3.7.2 3.6.6 3.6.4 2.7.16 2.7.15 2.7.14
PythonAnaconda	lang	2019-12-10	-	3.7 3.6

– especially available python modules on jupyterhub: check out package list-link on jupyterhub spawner page (https://taurus.hrsk.tu-dresden.de/jupyter/hub/home)

– possibility to create own kernel/environment for jupyterhub: see docs at
https://doc.zih.tu-dresden.de/hpc-wiki/.../JupyterHub#Creating_and_using_your_own_environment

02 Hints for HPC – Urls for Jupyterhub

- url alchemy with jupyterhub: support students to start notebooks with the right configuration
- login at jupyterhub (https://taurus.hrsk.tu-dresden.de/jupyter/hub/home) on taurus before
- link examples
  a) use git pull and open single file + jupyterlab with python (tensorflow example):
  https://taurus.hrsk.tu-dresden.de/jupyter/hub/user-redirect/
  git-pull?repo=https%3A%2F%2Fgitlab.hrz.tu-chemnitz.de%2Fclehm-tu-dresden.de%2Fai-for-teaching-r.git
  &urlpath=lab %2Ftree%2Fai-for-teaching-r.git/tensorflow-python/testing_tensorflow.ipynb/%2F&branch=master

  b) use git pull + rstudio + resource request with GPU (spawn parameters) (keras example):
  https://taurus.hrsk.tu-dresden.de/jupyter/hub/user-redirect/
  git-pull?repo=https%3A%2F%2Fgitlab.hrz.tu-chemnitz.de%2Fclehm-tu-dresden.de
  %2Fai-for-teaching-r.git &urlpath=rstudio %2F&branch=master #/~(environment~'test~partition~'gpu2~gres~'gpu:1)
  remark: for RStudio choose environment 'test' as it is available only there

02 Hints for HPC – Urls for Jupyterhub

- these links work within TU Dresden data network
- see the docs at https://doc.zih.tu-dresden.de/hpc-wiki/.../JupyterHubForTeaching
- link generator for nbgittpuller: https://jupyterhub.github.io/nbgitpuller/link.html
  (i) https://taurus.hrsk.tu-dresden.de/jupyter/ in field JupyterHub URL
  (ii) especially for using gitlab from TU Chemnitz add .git to the repo address in the final link

Organisational Issues – Agenda

03 Organizational Issues

- preparation effort: prior knowledge, installation, configuration, maintenance
- HPC project and resources
- login management
- reservation of HPC resources

03 Agenda

03 HPC Project and Resources

- apply for HPC project (ZIH login needed) at: https://hpcprojekte.zih.tu-dresden.de/
- resources example:
  - 20 students, 2 hours per week (weekly event), 16 weeks (approx. one semester)
  - every student should have 5 cores for a job
  - you would apply for a total of: $20\ \text{students} \times 5\,\frac{\text{cores}}{\text{student}} \times 2\,\frac{\text{hours}}{\text{week}} \times 16\ \text{weeks} = 3200$ CPU hours
- up to 5000 CPU hours: no detailed project description is necessary (make clear that it is for teaching)
- think about your memory (RAM and storage) needs as well
- project home can be small: for storage of large data use workspaces (see the docs at https://doc.zih.tu-dresden.../WorkSpaces)

03 Agenda

03 Login and Project Management

- HPC access for teaching purposes: students join HPC project
- joining an HPC project requires personal ZIH login and project activation link
- use HPC project management (login only for project originator) at https://hpcprojekte.zih.tu-dresden.de/managers/
  a) get activation link to join project
  b) remove students from project
- further information for HPC project management: https://doc.zih.tu-dresden.de/hpc-wiki/.../ProjectManagement
- if not all students have ZIH login, please contact us at hpcsupport@zih.tu-dresden.de

03 Agenda

03 Reservation

- for immediate job processing during classroom sessions use reservations for your project (at latest 7-8 days beforehand via hpcsupport@zih.tu-dresden.de), repetitive reservations are possible
- more details on reservations: https://doc.zih.tu-dresden.de/.../Slurm#Reservations

Conclusions – Agenda

04 Conclusions

- overview on use cases

	use case I – simple	**use case II – advanced**	**use case III – HPC**
hardware	VM (research cloud)	VM (enterprise cloud)	HPC machine
data	static	dynamic (git)	dynamic (git)
save changes	no	temporary	permanent
login management	no	yes	yes
access	anonymous	personalized	personalized

- Think carefully about your goals and needs, referring software, data and hardware.
- Scalability of an approach (login management, hardware): What if the number of students/users doubled or tripled?
- trade-off between powerful hardware for certain time slots vs. availability and flexibility of VMs

04 Conclusions

- tools ecosystems are changing dynamically, esp. with HPC systems it is hard to react immediately
- everything shown should act as inspiration for own ideas
- Contact us with your own ideas and wishes! We will check what is possible.

Thanks for your attention

Christoph Lehmann
+49 351 - 463 42489
christoph.lehmann@tu-dresden.de

Dresden, Germany