HPC CF

https://hpc-certification.org

Julian Kunkel

The HPC Certification Forum

Challenges for HPC (and Open Source) Training

- Not all users possess the right level of training
  - Inefficient usage of systems, frustration, lost potential
  - Good training saves compute time and costs!
- Diverse user background and goals
  - Science is the goal, HPC is the vehicle
  - Need to run an application to complete the PhD
- Learning is not easy
  - Users need to understand beneficial knowledge for tasks
  - There exist various different training material
  - Teaching of different data centers is hard to compare
- Data center have difficulties to verify the skills of users
- Confusion of trainers and HPC practitioners regarding skills
  - What should one person know/train regarding "MPI"?

Outline

The HPC Certification Forum

Goals

- Fine-grained HPC knowledge representation $\Rightarrow$ Competence Standard
  - What competences exist, how are they defined?
  - Puzzle of competences for everyone (practitioners, students, admins)
  - Supporting navigation and role-specific knowledge maps
- Establishing international certificates attesting knowledge
- Supporting an ecosystem around the HPC competences

Scope of the forum

- Central authority for competence representation, certification, and support
- Purposeful limitations of the forum:
  - We do not compete with content providers
  - We do not create a curriculum (university/centers responsibility)

The HPC Certification Forum

Organization Details

- Started in 2018 as a spin-off from a project
- An independent international body
- Organized into
  - Steering board (elected)
  - Full members (with voting rights) - Contributors
  - Associate members (anyone and any institution)
  - Collaboration with e.g., SIGHPC Education Chapter

Activities

- Curating and maintaining the **Competence Standard**
- Providing tools and ecosystem around the competences

Governance

Various processes are documented here.

Steering Board

- General chair: Julian Kunkel (University of Göttingen / GWDG)
- Skill-tree curator: Kai Himstedt (University of Hamburg)
- Topic curators:
  - HPC Knowledge: Lev Lafayette (University of Melbourne)
  - Performance Engineering: Anja Gerbes (University of Dresden)
  - Sofware Development: Marc-Andre Hermanns (RWTH Aachen)
  - Administration: Sudeep Narayan Banerjee (Indian Institute of Technology Gandhinagar)
- Publicity chair: Weronika Filinger
- Other topics are jointly managed by the board

Organization

Organization of the members

- Webpage is the central hub (`https://www.hpc-certification.org`)
- Mailinglists (news, members, board)
- Monthly public meetings on our Slack channel
- Monthly meeting of the board
- Annual general assembly (BoF at ISC or independent workshop)

Data handling

- Everything* is developed/available in the open GitHub (`https://github.com/HPC-certification-forum`)
- Exception are examination questions

Outline

Classification of Competences == Skills

- A **skill** defines background, objectives, learning outcomes
- The **skill tree** organizes the competences as hierarchical skills
- Certificates bundle several skills into attestable unit

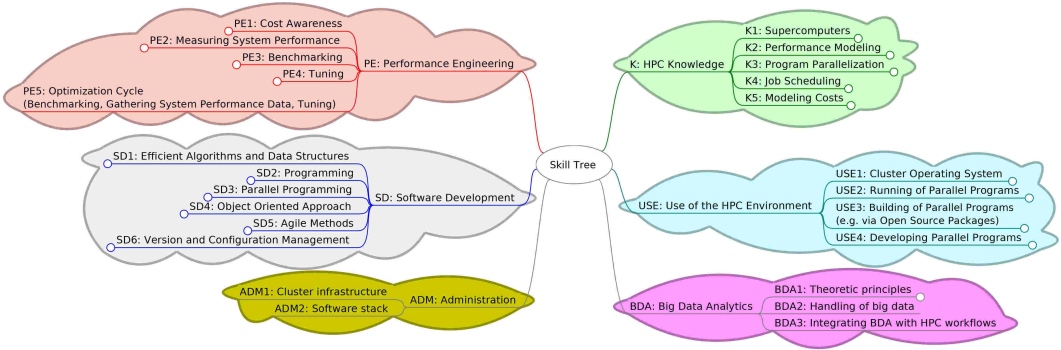


Figure: Top-levels of the skill tree (Initial ADM and BDA branches)

Example High-Level Skill (Excerpt)

- Name: Command Line Interface
- Id: USE1.1-B
- Background: HPC systems are usually accessed via a Linux-based Command Line Interface (CLI) that is provided by a shell. At its core, a shell is ...
- Aim:
  - describe the key principles of a shell
  - execute basic programs to query system information and manipulate...

Learning outcomes (these must be examinable)

- Utilize the bash shell to execute individual programs with arguments
- Describe the meaning of the exit code of a program
- Run multiple programs after another depending on the exit code ;, &&, ||
- List the set of basic programs and their tasks:
  - pwd

Classification of HPC Competences

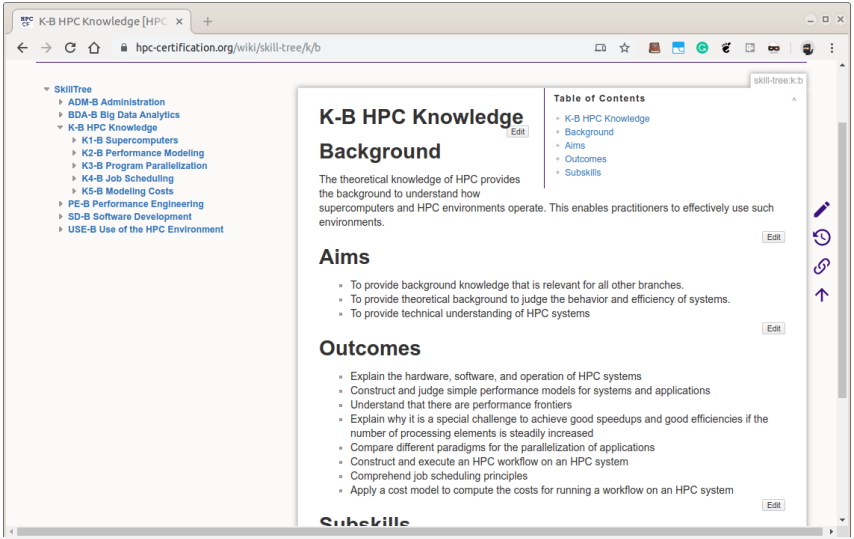
- Granularity of skill descriptions
  - Too fine ⇒ content of a skill is predefined at leaf level
  - Too coarse ⇒ no help for structuring the material
  - Guiding principle: leaf node should be covered in 1-4 hour lecture/workshop
- Organization of HPC skills
  - Skills are typically depending on sub-skills ⇒ tree structure
  - References to skills are possible; still skills are building blocks for various tasks
  - One skill can have multiple instances for different skill levels (basic, ..., expert)
- Verification of skill tree and certification approach
  - Feedback by the HPC community/practitioners justify the approaches

Status / Previous Activities

- Development version of the Competence Standard is online
  - ca. 250 skills identified - Git managed Markdown files
  - Easy access via Wiki and REST API
- Developed various processes
  - Supporting experts to adopt skills
    - Enable experts to curate skills that are in their field of expertise
    - Similar to code maintainer - various contributions this year
  - Working on sponsoring (but chicken/egg problem)
  - Developed examination prototype with first (but limited exam)
  - Adoption

All our developments are under open licenses (except the exam questions)

Wiki for Skills



Contribution to the Skill-Tree High-Level Editing

How can members contribute?

- Webpage with Markdown version controlled in Git
  - https://www.hpc-certification.org/wiki/skill-tree/b
  - GitHub: https://github.com/HPC-certification-forum/skill-tree
    - Pull requests, reviews, comments, ...
- Editing a MindMap, the structure of Skills
  - Synchronized with the skill tree in Git
  - Uses the OpenSource tool Freemind
- Discussion on our Slack
- Documented in our processes section
- See our videos on YouTube
- Visit us and join our Slack/mailing lists: https://hpc-certification.org

Certification: Assessment

1. User registers to test, receives email
2. User takes test online (any time!), consists of
   - Scenario – SSH login in a browser to a Docker environment
   - Multiple choice exam
     - System selects number of questions (and responses) randomly from a pool
3. Results are submitted to the web server
4. Automatic approval of response
5. Automatic creation of certificate and returned by email
   - Permanent computer-verifiable proof that skill is created
     - Return a text version with GPG signature
     - Return a link that can be verified on hpc-certification.org

- Privacy: minimize information stored on servers, keep some for statistics
- Includes some measure to prevent cheating and brute forcing (e.g., delay)

1. Registration with disclaimer

Disclaimer

Privacy statement

The certificate

The certificate will show your name and affiliation, the month the exam was conducted and a description about the particular certificate. It will also include a unique URL that can be given to a third-party allowing it to verify that you have obtained the certificate. It will be produced as a PDF and transferred via email to you.

Data recorded

To generate a meaningful certificate, we need the following personal information: **name**, **affiliation** and **email address**. On the server side, we will also attach the **date** the test was submitted to this metadata. When you interact with the webserver, the typical server logs will be recorded as stated in the privacy statement. These informations, will not be correlated to your personal information or the tests submitted here.

Data processing

We will use your personal data and the score to generate a certificate. This process may involve manual marking and approval of a staff member and the automatic generation of the certificate.

Your name and email is used solely to contact you and send you the test results and certificate (in case of your failure, we will send you some information that allows the discussion of your test results). We will not disclose your personal information, the results, or the generated certificate with any third party. In case of a successful examiniation, this information will be deleted from the server immediately after the results are processed and the email has been sent to you. In case of an unsuccesful examination, to prevent re-examination within a cooldown period, we will store a hash of your name together with the date of the examiniation for a duration that is the resit period of the individual certificate. After that duration, we will delete the information such that you can be re-examined.

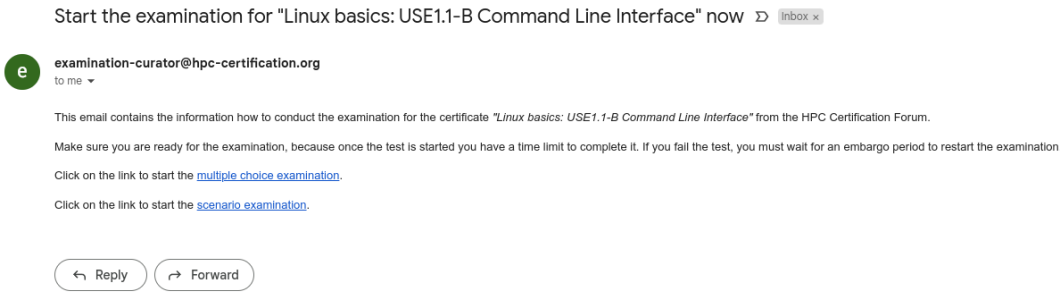
We will record the information of (date, affiliation), i.e., how many users with a certain affiliation have obtained a certain certificate. We will use this database of affiliations and date to promote the certification program (e.g., by mentioning that someone from an affiliation have obtained a certain certificate). These informations will **not** contain your name, email address, or the score.

We will record the achieved score separetely from any personal information with the answered questions for the purpose to optimize the examination.

Terms of this service

We will take industry-typical precaution to prevent any cybercrime including theft of your personal data while your data is stored on any IT system involved in the data processing. However, in case of any data loss, theft of the data provided, we do not take liability for data loss or thefts caused by cybercrime. Note that inclusion cases in which minor negligence actions of staff enabled

1. Received email with links



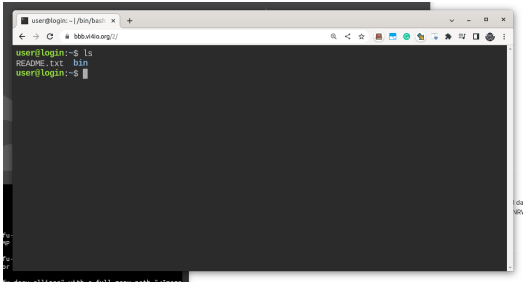
2. Scenario: Browser and PDF

Your scenario is ready, please follow the instructions to access it

The tasks are described in the following PDF

To access the sandbox, follow this link.
Then enter the credentials
username: *user*
password: *7f62ae2df4aebdd159e400efee5e79319436e00d*

Time is now: 2022-10-18 15:37:01
The deadline for the submission is: *2022-10-18 15:52:31*
The scenario will automatically be submitted by the deadline!



Scenarios

This document contains your individual scenario for the exam. Note that you can login multiple times by opening the provided URL in an individual browser tab. You may disconnect and reconnect at a later time as long as the time for the exam is not exceeded. This might indeed be necessary for some tasks!

It doesn't matter in which order you perform the individual tasks.

The execution of a program is indicated with the dollar sign, i.e., `$ myprog` means you shall execute on the shell literally the command called *myprog*. The exact arguments may be provided or you may have to specify them.

After your time is up, the container will be automatically be destroyed, disconnecting you. Your result will be assessed within a week.

1 Terminating a rogue program (1 mark)

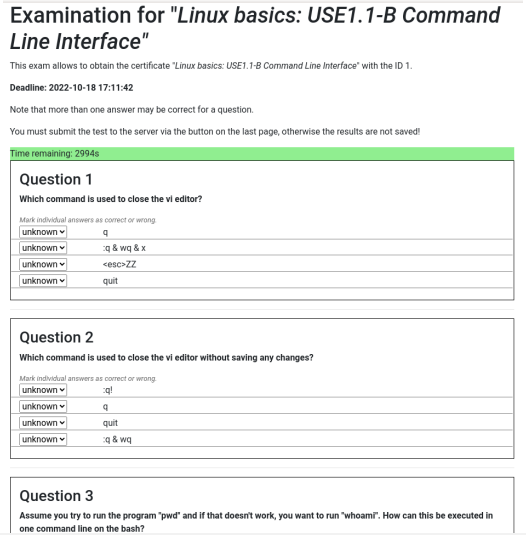
Assume you have accidentally started the program "sleep", now you need to terminate its execution. We already have started this program in your scenario. Hint: using the command `$ ps` providing the argument `-x`, you can find out the process ID (PID), that you may use to stop its execution.

2 Exiting editors (4 marks)

Please execute the following editors and close them successfully.

- nano
- vim
- vi
- emacs

2. Multiple Choice Exam



Examination for "*Linux basics: USE1.1-B Command Line Interface*"

This exam allows to obtain the certificate "*Linux basics: USE1.1-B Command Line Interface*" with the ID 1.

Deadline: 2022-10-18 17:11:42

Note that more than one answer may be correct for a question.

You must submit the test to the server via the button on the last page, otherwise the results are not saved!

Time remaining: 2994s

Question 1

Which command is used to close the vi editor?

Mark individual answers as correct or wrong.

unknown	q
unknown	:q & wq & x
unknown	<esc>ZZ
unknown	quit

Question 2

Which command is used to close the vi editor without saving any changes?

Mark individual answers as correct or wrong.

unknown	:q!
unknown	q
unknown	quit
unknown	:q & wq

Question 3

Assume you try to run the program "pwd" and if that doesn't work, you want to run "whoami". How can this be executed in one command line on the bash?

- Random questions from a pool
- Question(s) per learning objectives
- Question: Correct, incorrect

Certification: Certificate

Text representation

```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA512
HPC Certification Forum Certificate
This text confirms that "Jane Doe" has
successfully obtained the certificate
"HPC driving license" (id: 1) at 02/2019.
Verification URL: https://hpc-certification.org/[...]
-----BEGIN PGP SIGNATURE-----
[...]
-----END PGP SIGNATURE-----
```

Certificate

Adoption

- Positive feedback at regularly organized BoFs at SC and ISC
- Various Tier 2 Centers using references: ZIH Dresden, GWDG, Mainz, ...
- Informal agreement in NHR Training to use it
- Discussion with various Tier 1, 3 centers
- International, e.g., SIG-HPC EDU

Examples

- `https://tu-dresden.de/zih/hochleistungsrechnen/nhr-training/performance-engineering/pika`
- `https://hpc-en.uni-mainz.de/kurse-und-workshops/`

Outlook

- More NHR centers to refer to Competence Standard (skills)
- More "expert adopted skills"
- Having HPC "driving license" basic skills in V1.0 with assessment
- Certification of actual users for the "driving license"...
- Moving to another server
- More events, workshops planned
- Hopefully some funding for the activity

Outline

Needs of the HPC CF

- At the moment - all our activities on voluntary basis - this is slow
  - because we believe in it
- Need funding of at least 1 advisor for creating and curating skills (essential)
  - Would also organize workshops to train the trainers and videos in YouTube
- A technician to support development of the ecosystem (desirable)
  - from prototype to production
- Official endorsment by data centers and programmes (e.g. EuropeHPC JU)
- Collaboration and contribution by programmes to skills (e.g. EuropeHPC JU)
- Support in organizing BoFs and workshops (desirable)
- Need to expand our server infrastructure to another home (doable)

Outline

Outlook and Expected Benefits

HPC practitioners

- Increase motivation to participate as the certificates are recognized in a CV
- Validate knowledge via tests
- Browse relevant competences
- Identify recommended and required skills related to certain tasks
- Understand and compare teaching offers across sites

Data centers

- Increase sharing of teaching materials
- Simplifies documentation of taught skills
- Identify missing teaching activities
- Tailor skill-representation specifically to users
- Correlate lack of skills with efficient use

Summary

HPC Certification Program

- Effort to standardize representation/certification of relevant HPC skills
  - Hierarchical definition of skills for practitioners
  - Building blocks that can be cherry-picked for different tasks
  - It's goal is **NOT** to provide content or a linear curriculum
- Perspective for data centers
  - Use statistics and machine learning to direct users to right skills
  - Make certain skills a mandatory requirement?
- Customizable representation and navigation for data centers/domains
  - Interactive viewer to browse skills and related content
  - We will use the viewer to link good content to the skills, too!

Appendix

Further Considerations

- Certificate definition
  - Bundles a set of useful skills together
  - A users' HPC qualification is certified by successful exams
  - Testing a single (fine-grained) skill may be too easy with a cheat sheet
- Separation of **skill**, **certificates** and **content provider**
  - Similar to the concept of a high school graduation exam
  - Learning material can be provided by different institutions
  - Teachers can put badges on material: this "trains skills X, Y, Z"
- External information can be linked to the skills providing different **views**
  - Suitability for a user role (Tester, Builder, Developer)
  - Suitability for a scientific domain (Chemistry, Physics, ...)
  - View: purpose-specific representation / coloring / content
    - Groups/institutions can derive a new skill tree with their own emphasis
    - What should people know to effectively work in your environment?