

Toward a Globally Acknowledged and Free HPC Certification



Julian Kunkel (+ HPC Certification Forum)

<https://hpc-certification.org>

Archer Seminar

2020-07-22

Outline



1 The Forum

2 Skills

3 Certification Process

4 Conclusions

Challenges for HPC 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

The HPC Certification Forum



Goals

- Fine-grained standardizing HPC knowledge representation
 - ▶ 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

- An independent international body
- Organized into
 - ▶ Steering board (elected)
 - ▶ Full members (with voting rights)
 - Contributors to the project (e.g., 1-2 hours per month)
 - ▶ Associate members (anyone and any institution)
 - ▶ Collaboration with e.g., SIGHPC Education Chapter

Responsibilities

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

Governance



We have governance rules splitting responsibility across roles

Steering Board

- General chair: Julian Kunkel (University of Reading)
- Skill-tree curator: Kai Himstedt (University of Hamburg)
- Topic curators:
 - ▶ HPC Knowledge: Lev Lafayette (University of Melbourne)
 - ▶ Performance Engineering: Anja Gerbes (University of Frankfurt)
 - ▶ Software Development: Roberto Villegas-Diaz (South Dakota State University)
 - ▶ Administration: Sudeep Banerjee (Indian Institute of Technology Gandhinagar)
- Other topics are jointly managed by the board
- Examination curator: Christian Meesters (University of Mainz)
- Publicity chair: Weronika Filingner

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
- Annual general assembly (form of a BoF at ISC or workshop)

Data handling

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

Outline



1 The Forum

2 Skills

3 Certification Process

4 Conclusions

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

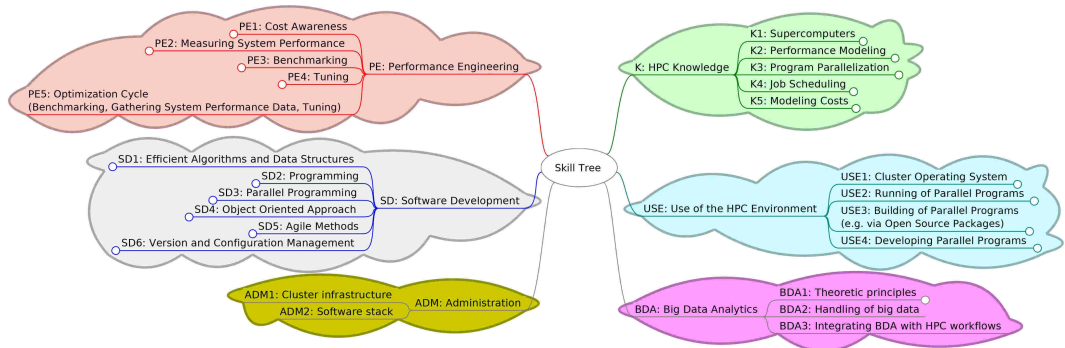


Figure: Top-levels of the skill tree (Initial ADM and BDA branches)

Example High-Level Skill (Excerpt)



- Name: SLURM Workload manager
- Id: USE4.2.2-B
- Background: SLURM is a widely used open-source workload manager providing various advanced features.
- Aim:
 - ▶ comprehend and describe the basic architecture of SLURM and its tools
 - ▶ use relevant tools to run and monitor (parallel) applications

Learning outcomes (these must be examinable)

- run interactive jobs with salloc, a batch job with sbatch
- explain the architecture of SLURM, i.e., the role of slurmd, srun
- explain the function of the tools: sacct, sbatch, salloc, ...
- explain time limits and the benefit of a backfill scheduler
- see <https://www.hpc-certification.org/wiki/>

Classification of HPC Competences



■ Granularity of skill descriptions

- ▶ Too fine \Rightarrow content of a skill is predefined at leaf level
- ▶ Too coarse \Rightarrow no help for structuring the material
- ▶ Guiding principle: leaf node should be coverable in 1-4 hour lecture/workshop

■ Organization of HPC skills

- ▶ Skills are typically depending on sub-skills \Rightarrow 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

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?

Status / Previous Activities



- The development version of the Competence Standard is online
- Released technical representations of the HPC skills
 - ▶ Markdown (embedded in a Wiki)
- Released JavaScript for visualization of skill tree ([demo](#))
 - ▶ Enables views: adjustable/embedable in your webpage
- Developed prototype for exam process and framework
- Developed various processes
- Designed seal of endorsement
- Engaged with various stakeholders (e.g., SIGHPC Edu)
- Conducted survey to verify the skill tree (more to come!)

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



This training covers (partially)
- K1.1 System architectures
- K1.2 Hardware architectures
See <https://hpc-certification.org/c/1.0>

Wiki for Skills



K-B HPC Knowledge [HPC x] +

hpc-certification.org/wiki/skill-tree/k/b

SkillTree

- ▶ ADM-B Administration
- ▶ BDA-B Big Data Analytics
- ▼ K-B HPC Knowledge
 - ▶ K1-B Supercomputers
 - ▶ K2-B Performance Modeling
 - ▶ K3-B Program Parallelization
 - ▶ K4-B Job Scheduling
 - ▶ K5-B Modeling Costs
- ▶ PE-B Performance Engineering
- ▶ SD-B Software Development
- ▶ USE-B Use of the HPC Environment

K-B HPC Knowledge

Background

The theoretical knowledge of HPC provides the background to understand how supercomputers and HPC environments operate. This enables practitioners to effectively use such environments.

Aims

- To provide background knowledge that is relevant for all other branches.
- To provide theoretical background to judge the behavior and efficiency of systems.
- To provide technical understanding of HPC systems

Outcomes

- Explain the hardware, software, and operation of HPC systems
- Construct and judge simple performance models for systems and applications
- Understand that there are performance frontiers
- Explain why it is a special challenge to achieve good speedups and good efficiencies if the number of processing elements is steadily increased
- Compare different paradigms for the parallelization of applications
- Construct and execute an HPC workflow on an HPC system
- Comprehend job scheduling principles
- Apply a cost model to compute the costs for running a workflow on an HPC system

Subskills

Table of Contents

- ▶ K-B HPC Knowledge
- ▶ Background
- ▶ Aims
- ▶ Outcomes
- ▶ Subskills

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](#)
- Details in our videos on YouTube: https://www.youtube.com/playlist?list=PL4b682pSp7MQbdhhvwisrPo7PjYya26_g

Outline



1 The Forum

2 Skills

3 Certification Process

4 Conclusions

Certification: Assessment Prototype



1. User takes multiple-choice test online (any time!)
 - ▶ A combination of JavaScript and a web service
 - ▶ System selects number of questions randomly from a pool
 - The questions are managed with rigorous license agreement
 - ▶ System draws 4-5 responses from 10 possible responses (some randomized)
2. Choices are submitted to the web server
3. *Approval* of the result
4. 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)

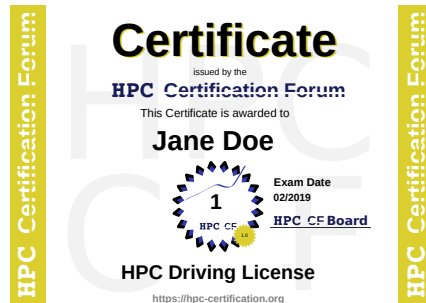
Certification: Certificate



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



Outline



1 The Forum

2 Skills

3 Certification Process

4 Conclusions

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!
- Visit us and join our Slack/ mailing lists: <https://hpc-certification.org>