HPC.NRW

APPLYING THE HPC SKILL TREE

Early experiences with course classification and design

December 14, 2021 | Marc-André Hermanns

DISCLAIMER

HPC.NRW

This presentation contains work in progress and should be regarded as a basis for discussion.

- For users:
  - Orientation in knowledge levels of different HPC topics
  - Self-assessment of personal knowledge level (after course participation)
  - Independently-assessed of personal knowledge level (after exam)

MOTIVATION FOR THE SKILL TREE

HPC.NRW

- For users:
  - Orientation in knowledge levels of different HPC topics
  - Self-assessment of personal knowledge level (after course participation)
  - Independently-assessed of personal knowledge level (after exam)
- For HPC centers:
  - Better assessment of users' knowledge ($\rightarrow$ "HPC Führerschein")
  - Strategically improve users' knowledge on HPC topics
  - Towards "standardization" of HPC curricula

APPLYING THE SKILL TREE IN COURSE RE-DESIGN

HPC.NRW

- Extending an existing MPI course
  - 2-day MPI introduction as part of PPCES @ RWTH Aachen University
  - Basis for HPC.NRW online MPI tutorial (in development)
    - Extend to broad coverage of MPI

APPLYING THE SKILL TREE IN COURSE RE-DESIGN

HPC.NRW

- Extending an existing MPI course
  - 2-day MPI introduction as part of PPCES @ RWTH Aachen University
  - Basis for HPC.NRW online MPI tutorial (in development)
    - Extend to broad coverage of MPI
- Initial assessment:
  - Classify the skill level of course content
    - Which skills does the current course build up?

APPLYING THE SKILL TREE IN COURSE RE-DESIGN

HPC.NRW

- Extending an existing MPI course
  - 2-day MPI introduction as part of PPCES @ RWTH Aachen University
  - Basis for HPC.NRW online MPI tutorial (in development)
    - Extend to broad coverage of MPI
- Initial assessment:
  - Classify the skill level of course content
    - Which skills does the current course build up?
  - Classify the skill level of requirements
    - Which skills does the participant need to follow the course?

APPLYING THE SKILL TREE IN COURSE RE-DESIGN

HPC.NRW

- Extending an existing MPI course
  - 2-day MPI introduction as part of PPCES @ RWTH Aachen University
  - Basis for HPC.NRW online MPI tutorial (in development)
    - Extend to broad coverage of MPI
- Initial assessment:
  - Classify the skill level of course content
    - Which skills does the current course build up?
  - Classify the skill level of requirements
    - Which skills does the participant need to follow the course?
  - Group content according to skil level
    - Which content is needed for “basic”, “intermediate”, and “expert” level?

- SD – Software Development
  - SD 1 – Programming Concepts for HPC
    - SD 1.2 – Parallel Programming
      - SD 1.2.3 – Programming Message Passing Systems

- K – HPC Knowledge
  - K 3 – Program Parallelization
    - K 3.1 – Level of Parallelization
    - K 3.2 – Parallelization Overheads
    - K 3.3 – Domain Decomposition

- SD – Software Development
  - SD 1 – Programming Concepts for HPC
    - SD 1.2 – Parallel Programming
      - SD 1.2.3 – Programming Message Passing Systems
- K – HPC Knowledge
  - K 3 – Program Parallelization
    - K 3.1 – Level of Parallelization
    - K 3.2 – Parallelization Overheads
    - K 3.3 – Domain Decomposition

- SD – Software Development
  - SD 1 – Programming Concepts for HPC
    - SD 1.2 – Parallel Programming
      - SD 1.2.3 – Programming Message Passing Systems ←

- K – HPC Knowledge
  - K 3 – Program Parallelization
    - K 3.1 – Level of Parallelization
    - K 3.2 – Parallelization Overheads
    - K 3.3 – Domain Decomposition

– Blocking vs. Nonblocking

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective ✓
- Detect Race Conditions & Deadlocks?

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective ✓
- Detect Understand Race Conditions & Deadlocks?

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective ✓
- Detect Understand Race Conditions & Deadlocks?
- Assess the impact of communication and synchronization on program performance

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective ✓
- ~~Detect~~ Understand Race Conditions & Deadlocks?
- ~~Assess~~ Understand the impact of communication and synchronization on program performance

- Blocking vs. Nonblocking ✓
- Point-to-Point vs. Collective ✓
- ~~Detect~~ Understand Race Conditions & Deadlocks?
- ~~Assess~~ Understand the impact of communication and synchronization on program performance

→ Detection and assessment are topics of “performance engineering” and “correctness checking”

– ~~Assess~~ Understand the impact of communication and synchronization on program performance

- ~~Assess~~ Understand the impact of communication and synchronization on program performance
- What else?
  - (see classification proposal)

SD 1.2.3 PROGRAMMING MESSAGE PASSING SYSTEMS

Expert Knowledge

HPC.NRW

- ~~Assess~~ Understand the impact of communication and synchronization on program performance
- Apply the concept of overlay networks

- ~~Assess~~ Understand the impact of communication and synchronization on program performance
- Apply the concept of overlay networks
  - What is meant here? Virtual Topologies? Why not earlier?

Classification of a large API

Marc-André Hermanns

CLASSIFICATION OF MPI TOPICS

A first proposal

- Semantic Terms
  - Local vs. nonlocal
  - Blocking vs. nonblocking
  - Collective vs. noncollective

CLASSIFICATION OF MPI TOPICS

A first proposal

- Semantic Terms [Basic]
  - Local vs. nonlocal
  - Blocking vs. nonblocking
  - Collective vs. noncollective

CLASSIFICATION OF MPI TOPICS

A first proposal

HPC.NRW

- Semantic Terms [Basic]
- Point-to-Point Communication
  - Blocking
  - Nonblocking
  - Persistent
  - Partitioned

CLASSIFICATION OF MPI TOPICS

A first proposal

HPC.NRW

- Semantic Terms [Basic]
- Point-to-Point Communication
  - Blocking [Basic]
    - Basic understanding of routine
  - Nonblocking [Basic]
    - Initial understanding of nonblocking semantics
  - Persistent [Intermediate]
    - Persistence as intermediate concept ($\rightarrow$ tuning opportunities)
  - Partitioned [Expert]
    - Partitioned communication requires understanding of persistence

CLASSIFICATION OF MPI TOPICS

A first proposal

HPC.NRW

- Semantic Terms [Basic]
- Point-to-Point Communication
  - Blocking [Basic]
  - Nonblocking [Basic]
  - Persistent [Intermediate]
  - Partitioned [Expert]
- Datatypes

CLASSIFICATION OF MPI TOPICS

A first proposal

HPC.NRW

- Semantic Terms [Basic]
- Point-to-Point Communication
  - Blocking [Basic]
  - Nonblocking [Basic]
  - Persistent [Intermediate]
  - Partitioned [Expert]
- Datatypes [Basic]
  - → Understanding of type map & type signature

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking
  - Nonblocking
  - Persistent
  - Neighborhood

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
    - Basic understanding of collective concept
  - Nonblocking [Basic]
    - Extend the nonblocking concept to collective concept
  - Persistent
  - Neighborhood

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
    - Basic understanding of collective concept
  - Nonblocking [Basic Intermediate]
    - Extend the nonblocking concept to collective concept
    - Time constraints may prohibit this topic to be in "Basic" level
  - Persistent
  - Neighborhood

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic Intermediate]
  - Persistent [Expert]
    - Persistence as intermediate concept ($\rightarrow$ tuning opportunities)
  - Neighborhood [Expert]
    - Relying on virtual topologies as intermediate concept

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic / Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators
  - Inter-communicators

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic / Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators [Basic]
  - Inter-communicators [Expert]

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators [Basic]
  - Inter-communicators [Expert]
- Process Topologies

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic / Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators [Basic]
  - Inter-communicators [Expert]
- Process Topologies [Intermediate]
  - Relying on virtual topologies as intermediate concept

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic / Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators [Basic]
  - Inter-communicators [Expert]
- Process Topologies [Intermediate]
- Info Object

CLASSIFICATION OF MPI TOPICS (CONT'D)

A first proposal

HPC.NRW

- Collective Communication
  - Blocking [Basic]
  - Nonblocking [Basic / Intermediate]
  - Persistent [Expert]
  - Neighborhood [Expert]
- Groups & Communicators
  - Intra-communicators [Basic]
  - Inter-communicators [Expert]
- Process Topologies [Intermediate]
- Info Object [Intermediate]
  - User-Library-Interaction as intermediate concept

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model
  - Sessions Model

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
    - Single-threaded initialization [Basic]
    - Hybrid initialization [Basic/Intermediate?]
  - Sessions Model [Intermediate]
    - Tune process startup
    - Isolate libraries

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
  - Relies on understanding nonblocking communication
  - Complex semantics

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests [Expert?]
  - Chicken-egg problem? (mostly unknown $\leftrightarrow$ mostly unused)

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests [Expert?]
- I/O

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests [Expert?]
- I/O [Intermediate]
  - Important scalability concept

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests [Expert?]
- I/O [Intermediate]
- Tool Support

CLASSIFICATION OF MPI TOPICS (CONT'D II)

A first proposal

HPC.NRW

- Process Initialization
  - World Model [Basic]
  - Sessions Model [Intermediate]
- One-sided Communication [Intermediate]
- Generalized Requests [Expert?]
- I/O [Intermediate]
- Tool Support [Expert]
  - Tool experts should have an expert understanding of MPI

A proposed curriculum

Marc-André Hermanns

PROPOSED BASIC LEVEL: SD 1.2.3-B

HPC.NRW

- Semantiv terms
- Process Initialization (World Model)
- Blocking & nonblocking point-to-point communication
- Datatypes
- Blocking collective communication
- Nonblocking collective communication
- Groups & Communicators (Intra-communicators)

PROPOSED INTERMEDIATE LEVEL: SD 1.2.3-I

HPC.NRW

- Process Initialization (Sessions Model)
- Persistent point-to-point communication
- Nonblocking collective communication
- One-sided communication
- I/O
- Process Topologies

PROPOSED INTERMEDIATE LEVEL: SD 1.2.3-I

HPC.NRW

- Process Initialization (Sessions Model)
- Persistent point-to-point communication
- Nonblocking collective communication
- One-sided communication
- I/O
- Process Topologies

- Increased understanding of performance implications
  - Require basic cost awareness (PE 1-B)
  - Require basic understanding of performance profiling (PE 2.2-B)
- Best practices on communication design

PROPOSED EXPERT LEVEL: SD 1.2.3-E

HPC.NRW

- Partitioned point-to-point communication
- Persistent collective communication
- Neighborhood collective communication
- Groups & Communicators (Intra-communicators)
- Generalized requests
- Tool Support

PROPOSED EXPERT LEVEL: SD 1.2.3-E

HPC.NRW

- Partitioned point-to-point communication
- Persistent collective communication
- Neighborhood collective communication
- Groups & Communicators (Intra-communicators)
- Generalized requests
- Tool Support

- Further detailed understanding of performance implications
  - Require intermediate knowledge level of related skills?
    - Which skills specifically?

HOW TO HANDLE RELATED SKILLS

A first proposal

HPC.NRW

- Domain Decomposition (K 3.3)
- Load Balancing (SD 1.2.4)
- Performance Engineering (PE)
- I/O Programming (SD 1.2.5)

HOW TO HANDLE RELATED SKILLS

A first proposal

HPC.NRW

- Domain Decomposition (K 3.3)
- Load Balancing (SD 1.2.4)
- Performance Engineering (PE)
- I/O Programming (SD 1.2.5)

- Not (direct) part of SD 1.2.3 course
  - Use in exercises?

HOW TO HANDLE RELATED SKILLS

A first proposal

HPC.NRW

- Domain Decomposition (K 3.3)
- Load Balancing (SD 1.2.4)
- Performance Engineering (PE)
- I/O Programming (SD 1.2.5)

- Not (direct) part of SD 1.2.3 course
  - Use in exercises?

→ Requirements for specific levels of SD 1.2.3
  - K 3.3-B should be a requirement for SD 1.2.3-I/E
    - K 3.3-B is a good motivation for virtual topologies
  - ...

HOW TO HANDLE RELATED SKILLS II

A first proposal

HPC.NRW

- → Requirements for specific levels of SD 1.2.3
  - ...
  - SD 1.2.3-B or SD 1.2.2-B requirement for K 3.3-I
    - At least one parallelization paradigm should be known

HOW TO HANDLE RELATED SKILLS II

A first proposal

→ Requirements for specific levels of SD 1.2.3

- ...
- SD 1.2.3-B or SD 1.2.2-B requirement for K 3.3-I
  - At least one parallelization paradigm should be known
- SD 1.2.5-B should be a requirement for SD 1.2.3-I
  - Information to fundamentals of filesystems are needed to motivate MPI I/O

HOW TO HANDLE RELATED SKILLS II

A first proposal

HPC.NRW

- → Requirements for specific levels of SD 1.2.3
  - ...
  - SD 1.2.3-B or SD 1.2.2-B requirement for K 3.3-I
    - At least one parallelization paradigm should be known
  - SD 1.2.5-B should be a requirement for SD 1.2.3-I
    - Information to fundamentals of filesystems are needed to motivate MPI I/O
- Focus on functionality first and performance later?
  - Enable participants to gain experience with the interface
  - Participants may get into bad practice without proper guidance?

SUMMARY

HPC.NRW

- Skill tree is still a work in progress (e.g., SD 1.2.3)
- Skill definition needs to be precise enough to full coverage in a block
  - A “course” with a skill label should cover the full skill

  $\rightarrow$ Participants can assume full coverage for the exam

SUMMARY

- Skill tree is still a work in progress (e.g., SD 1.2.3)
- Skill definition needs to be precise enough to full coverage in a block
  - A “course” with a skill label should cover the full skill
  - → Participants can assume full coverage for the exam
- Large interfaces not trivial to break down (due to cross-cutting concepts)
  - If non-blocking point-to-point and blocking collective communication is covered, why not also cover nonblocking collective communication?
  - → Time constraint

SUMMARY

- Skill tree is still a work in progress (e.g., SD 1.2.3)
- Skill definition needs to be precise enough to full coverage in a block
  - A “course” with a skill label should cover the full skill
  - → Participants can assume full coverage for the exam
- Large interfaces not trivial to break down (due to cross-cutting concepts)
  - If non-blocking point-to-point and blocking collective communication is covered, why not also cover nonblocking collective communication?
  - → Time constraint
  - → Unwanted spliting of similar concepts

SUMMARY

- Skill tree is still a work in progress (e.g., SD 1.2.3)
- Skill definition needs to be precise enough to full coverage in a block
  - A “course” with a skill label should cover the full skill
  - → Participants can assume full coverage for the exam
- Large interfaces not trivial to break down (due to cross-cutting concepts)
  - If non-blocking point-to-point and blocking collective communication is covered, why not also cover nonblocking collective communication?
  - → Time constraint
  - → Unwanted spliting of similar concepts
  - Do we need further subdivisioning: SD 1.2.3.X?
    - Where to place MPI I/O?

- Complex relationships among skill tree nodes
  - How to ensure requirements for a course are known and met?
  - How to encode optional knowledge?
  - How to build paths from basic to expert level?

- Complex relationships among skill tree nodes
  - How to ensure requirements for a course are known and met?
  - How to encode optional knowledge?
  - How to build paths from basic to expert level?
- Do we need working groups to generate specific progress in different parts of the skill tree?

Thank you.