# Compute Component

## 3rd PUNCH 2.0 Workshop
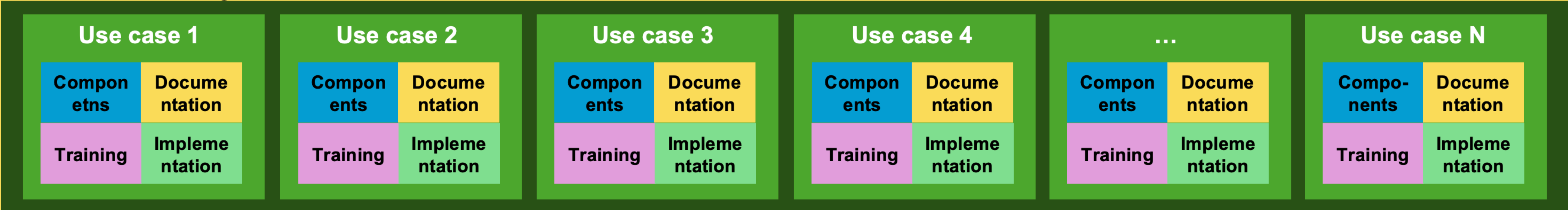
Manuel Giffels, 14.04.2025

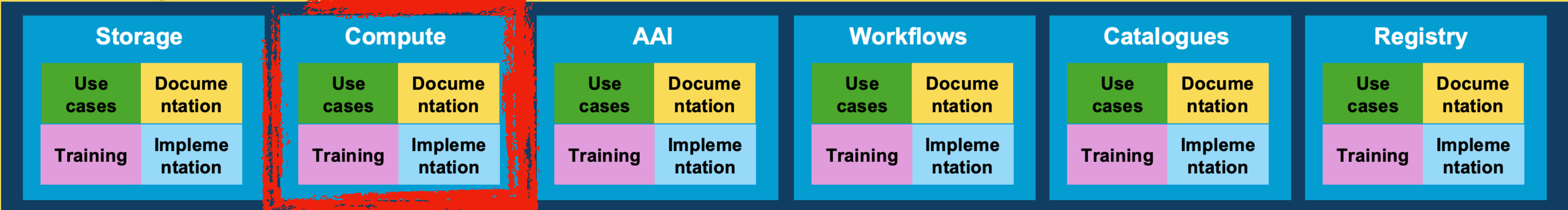# TA 4: Management

| Project Management | Finances | Communi-cation | Training |
| --- | --- | --- | --- |
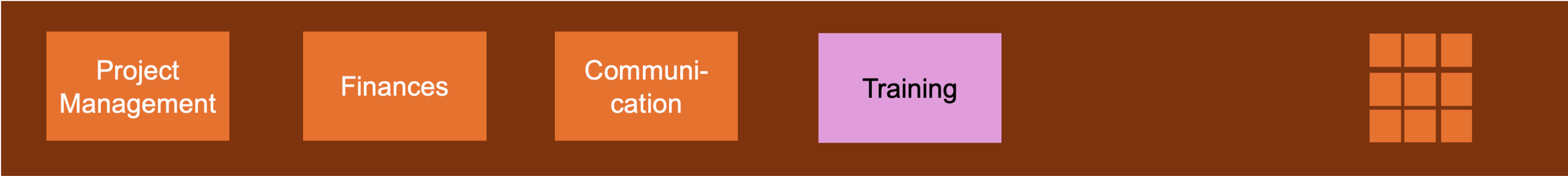
# TA 3: Services and Results

# TA 2: Assembly

| Use case 1 | Use case 2 | Use case 3 | Use case 4 | ... | Use case N |
| --- | --- | --- | --- | --- | --- |
| Componetns / Documentation / Training / Implementation | Components / Documentation / Training / Implementation | Components / Documentation / Training / Implementation | Components / Documentation / Training / Implementation | Components / Documentation / Training / Implementation | Compo-nents / Documentation / Training / Implementation |

# TA 1: Components

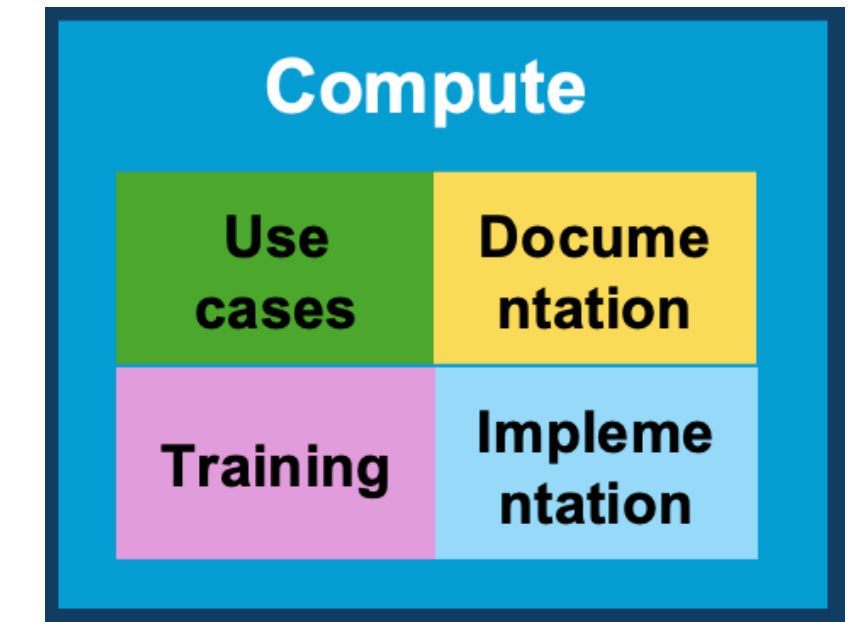| Storage | Compute | AAI | Workflows | Catalogues | Registry |
| --- | --- | --- | --- | --- | --- |
| Use cases / Documentation / Training / Implementation | Use cases / Documentation / Training / Implementation | Use cases / Documentation / Training / Implementation | Use cases / Documentation / Training / Implementation | Use cases / Documentation / Training / Implementation | Use cases / Documentation / Training / Implementation |

# Compute Component
## What is in this box?



- Federated Distributed Heterogeneous Compute Cloud

  - Covering mainly HTC and single node parallelism
    plus distributed processing frameworks like dask

  - Take into account site local data (parallel FS, S3, etc.), requires local account mapping

- Community specific entry points/interfaces
  (Login node, JupyterHub, REANA, Grid CE etc.)

- Generalised access to large scale computing resources (HPC) via AAI

  - Large scale multi node parallelism

- Container Registry including automized CI/CD builds and deployment

- Scalable software and container distribution (CVMFS4PUNCH)

# Compute Component

**What are the 3 main deliverables / milestones?**
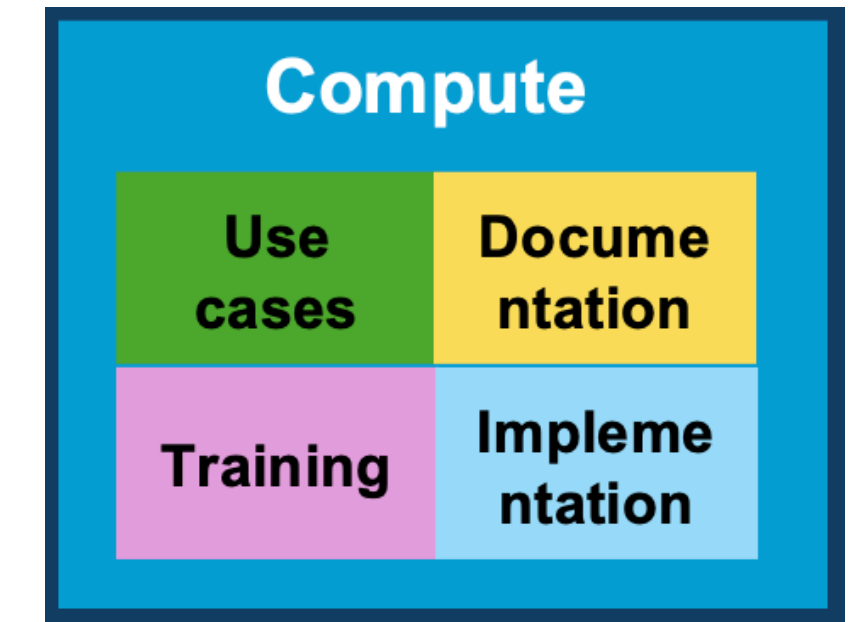


Hard to say without specifying requirements from generic use-cases

- Could define goals on which fraction of use-cases needs to be supported by when

I guess we will end up with something like:

- Set-up **enhanced** distributed computing & parallel processing infrastructure

- Establish user-centric access & interfaces lowering the threshold to use Compute4PUNCH

- Support for containerized workflows & scalable software distribution
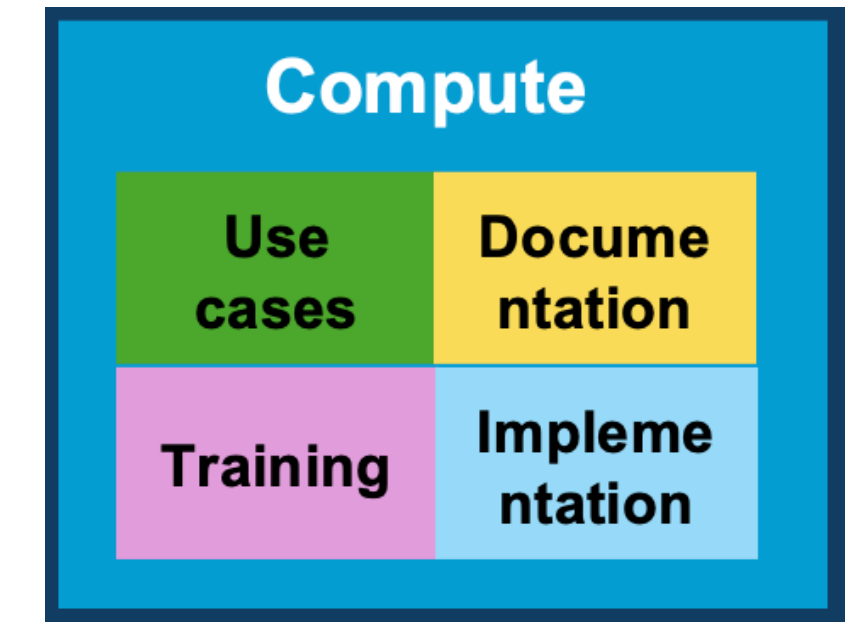
# Compute Component
## involved / interested / necessary institutions?



- AIP, Bonn, DESY, KIT, Göttingen?
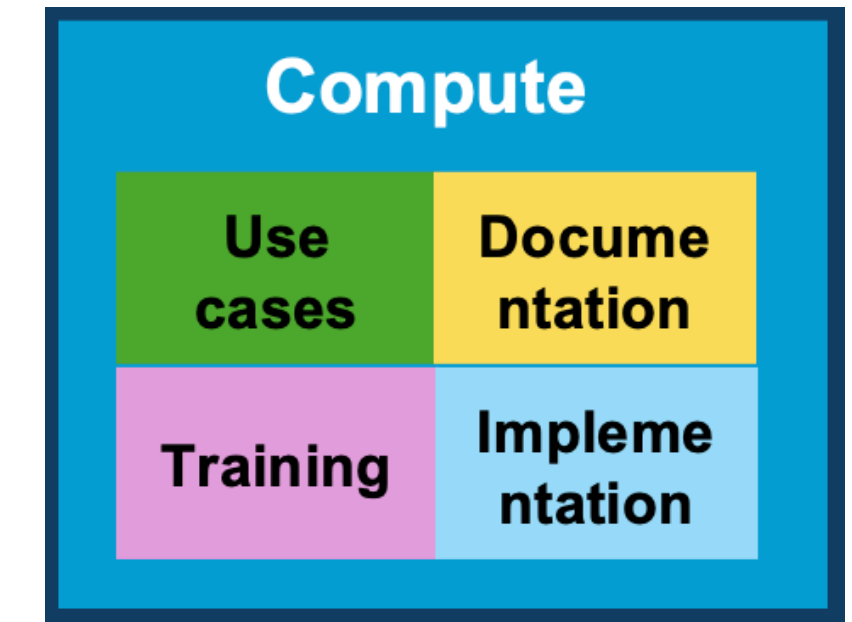
- NHR (external)?
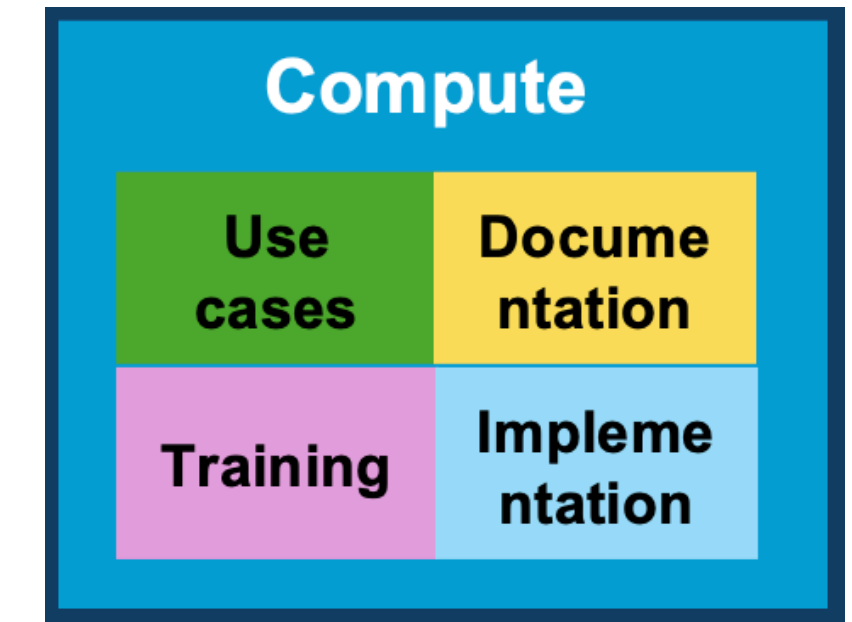
# Compute Component

## Approx. resource requirements?



- 5 well-qualified FTE fully committed to PUNCH Compute Component + in-kind person power
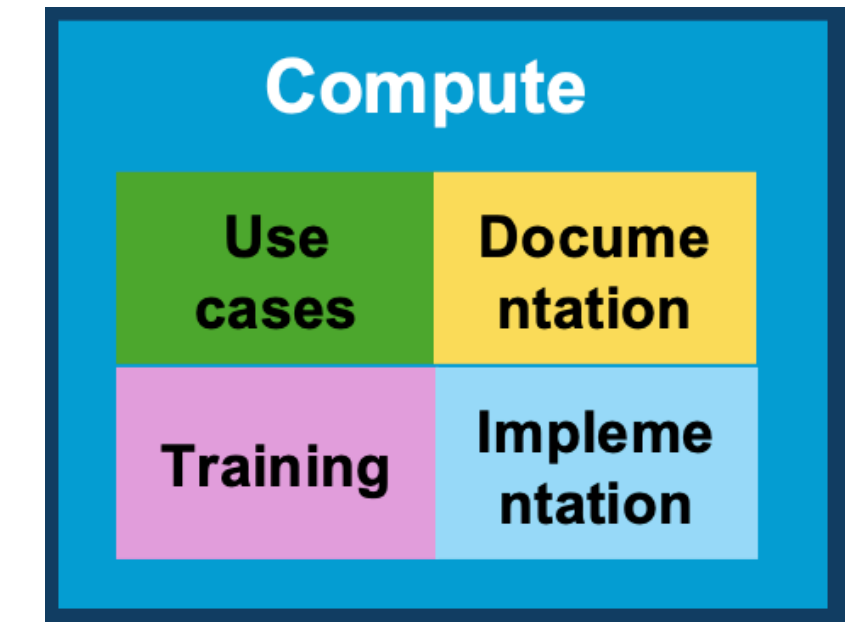
# Compute Component
## Community coverage - who is potentially left behind?

- Compute4PUNCH covers HTC, parallel processing (dask) and single node parallelism

- Variety of access points (login node, JupyterHub, Grid-CE, REANA)

- Use-cases requiring large scale multi node parallelism, will directly go to their favourite HPC cluster

  - However, as possible access should be simplified by utilizing login via PUNCH AAI

- IMHO that would cover pretty much of the community needs

# Compute Component
**Gap analysis - are these the right boxes?**



- IMHO software management and distribution is completely missing so far, I have added it for the time-being to the compute component. However, should we think about a relocation to a dedicated component?