GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

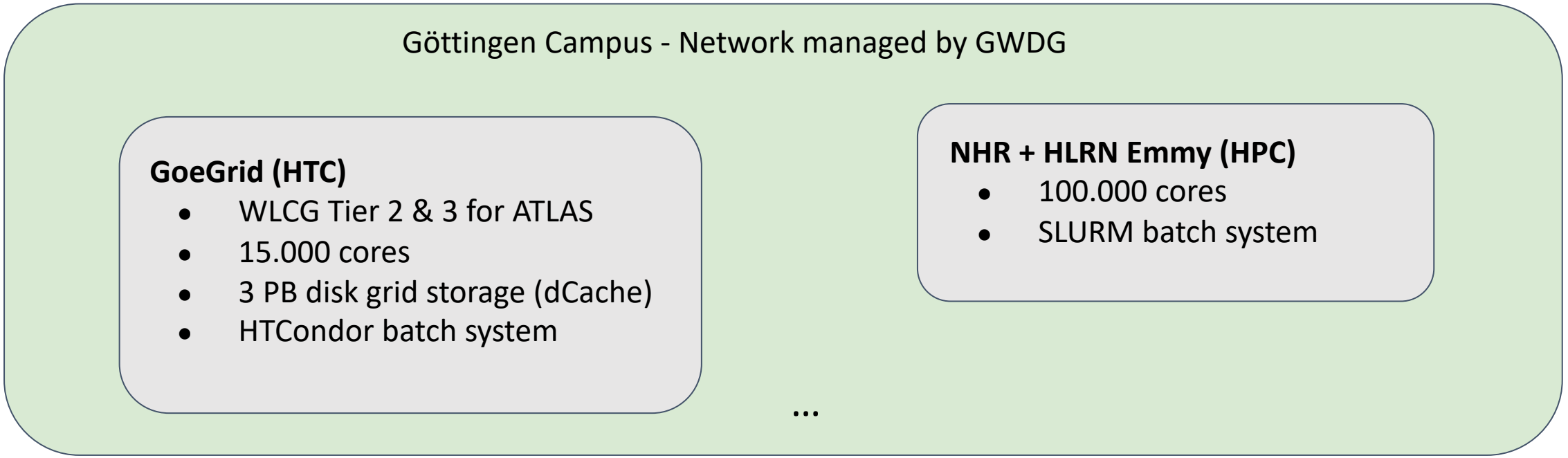
HPC Integration into WLCG Tier-2 GoeGrid

FIDIUM Collboration Meeting 2022

Status of Work Package 1

Arnulf Quadt, Sebastian Wozniewski, Saidev Polisetty, Daniel Schindler

HEP Computations in Göttingen



In future (LHC Run4) high computing resources for HEP computations are needed

Extension of GoeGrid with the local HPC system „Emmy“

- Opportunistic resource: usage on limited time frame for free
- In future: Non-opportunistic usage planned

Possible long-term perspective: A single merged cluster

FIDIUM Team @ Uni Göttingen

Distribution of work loads (FTE)

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

				
	Arnulf Quadt Professor Group Leader	**Sebastian Wozniewski** PostDoc GoeGrid	**Daniel Schindler** (since 08/2022) PostDoc HPC Integration	**Saidev Polisetty** (since 09/2022) PhD Student Performance Studies
Area 1			0.5	0
Area 2			0.5	0
Area 3			0	0.67

Typical Requirements of ATLAS (HTC)

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

- 300.000 cores and 300 PB storage needed per year (2022)
- Requirements of single-core jobs and single-node multi-core jobs
  - 8 - 64 cores
  - ~ 2 GB memory per core
  - ~ 4 GB local scratch space per core
  - ~ 1 Mbit/s per core network usage (remote storage access)
  - 12 hours walltime with large variance (due to pilot model)
- The load subdivides into different physical calculations
  - Simulation of physical processes (50% CPU)
  - Reprocessing of experiment data (30% CPU)
  - Analysis (15% CPU)
  - Statistical modelling and ML (5% CPU)

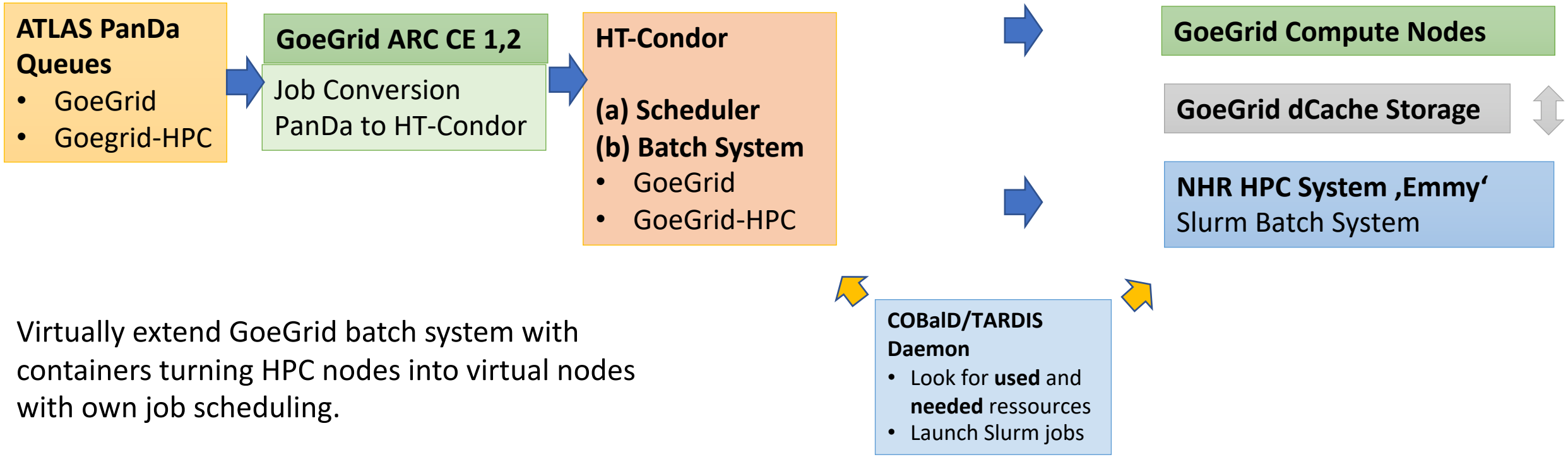
Conceptional Challenges

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

- Scheduling policy:
  - HPC Batch System only allows Whole-Node-Scheduling.
  - Targeted to run different job types (single-core or multi-core)
  - Need a solution to dynamically schedule jobs on booked nodes.
- Opportunistic and Non-Opportunistic usage of Emmy is targeted.
  - Opportunistic usage of Emmy would limit the consumable time for jobs to max. 2h. Usually jobs by ATLAS run with t>2 h.
  - Short lifetime of booked nodes limits flexibility and efficiency

Setup of GoeGrid and Emmy

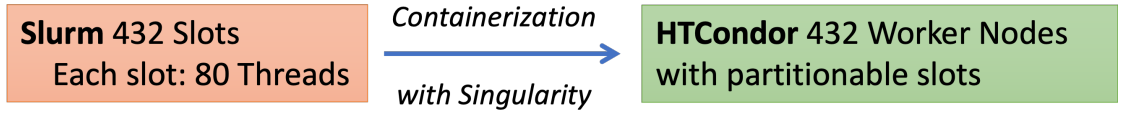
First goal and test solution



Virtually extend GoeGrid batch system with containers turning HPC nodes into virtual nodes with own job scheduling.

Usage of COBalD/TARDIS

- COBalD: Look what resources are used
- TARDIS: Interface to the resource and the overlay batch system



Current Issues

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Network

- **Outbound connections** from nodes not possible by default - using proxies problematic due to high traffic - allow connections to known IPs as a compromise?
- GoeGrid Network has to be connected to the net of Emmy to allow for **high-bandwidth connections** from HPC to the local **grid storage**.
- Allow for remaining outbound connections to outside grid services e.g. **cvmfs** or via GoeGrid squid proxies

Software permissions

- No unprivileged user namespaces - Prevents multiple layers of containers as needed for using COBalD/TARDIS with ATLAS jobs
- Possible solutions:
  - User-specific temporary permission based on setuid-script (under discussion)
  - Additionally leave out network namespaces which is not needed but main reason for security concerns?

Walltime preferences

- Drone lifetime should cover multiple jobs in sequence for efficient usage
- HPC limits long-term jobs (most resources allow only 12h); ~2h for opportunistic usage

Summary and Outlook

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

1. The LHC Run4 has higher computational demands which cannot be met by current resources.
2. Our approach: COBalD/TARDIS will be used for the integration of the NHR HPC system „Emmy“ to GoeGrid.
3. The usage of a virtual node has been tested on a GoeGrid server.
4. Fast network connection of GoeGrid storage elements to the Emmy cluster is currently being setup.
5. Work for Area 1 is ongoing, first steps have been successfully achieved.