

ERLANGEN REGIONAL COMPUTING CENTER



Components for practical
performance engineering in a
computing center environment:
The ProPE project

Jan Eitzinger

7. HPC-Status-Konferenz der Gauß-Allianz

Overview



Call:

Performance Engineering für wissenschaftliche Software

Partners:



Duration:

03/2017 – 02/2020

Coordination: Prof. G. Wellein (RRZE)

Current state

- HPC competence in German HPC centers distributed across country
- Gauss-Allianz is an initiative to integrate and organize TIER 2/3 HPC landscape in Germany
- Multiple local efforts and island projects:
bwHPC, KONWIHR, HKHLR, HLRN ...



Our contribution

- Similar targets as sketched in **GA Strategiepapier**, but focus on Performance-Engineering sub-topic

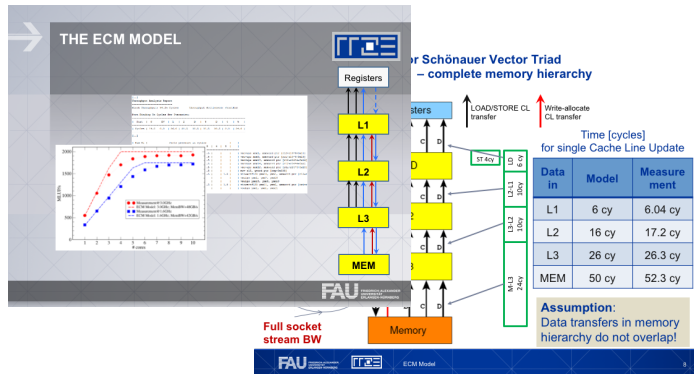
Integrate with and **built on** already existing efforts and further foster collaboration among German Tier 2/3 centers with respect to PE.

Major Building Blocks

- **Dissemination** – Increase publicity of project and raise general awareness for performance issues

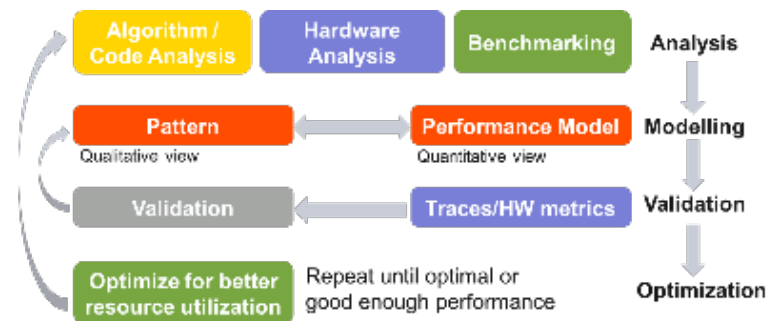


We want to talk with you about your PE problem!



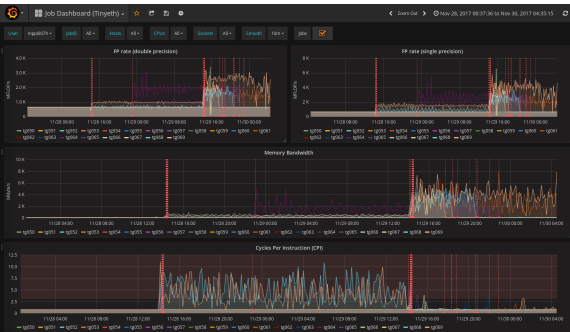
- **Documentation** – Build a central web offering, create content and provide resources to maintain it

- **Structured PE-Process** – Systematic bottleneck centric performance analysis and optimization process



Major Building Blocks cont.

- **PE Support Infrastructure** – Process blueprint for nation-wide aligned support effort



- **Application Monitoring and Analysis** – Automatic profiling and bottleneck analysis for all applications running on a HPC-System

- **HPC Curriculum** – Coordinated nation-wide Workshop and Tutorial program



WP1 PE Process: Initial Experiences

- Application Performance Monitoring in place:
 - Resource allocation issues (# processes, memory, load balance and affinity)
 - File IO issues on parallel file system
 - Other user issues (batch script, MPI)
- Multiple meetings with users, but no PE code optimization project so far from the production users
- **BUT:** Many requests from outside for collaboration on PE topics in a research context

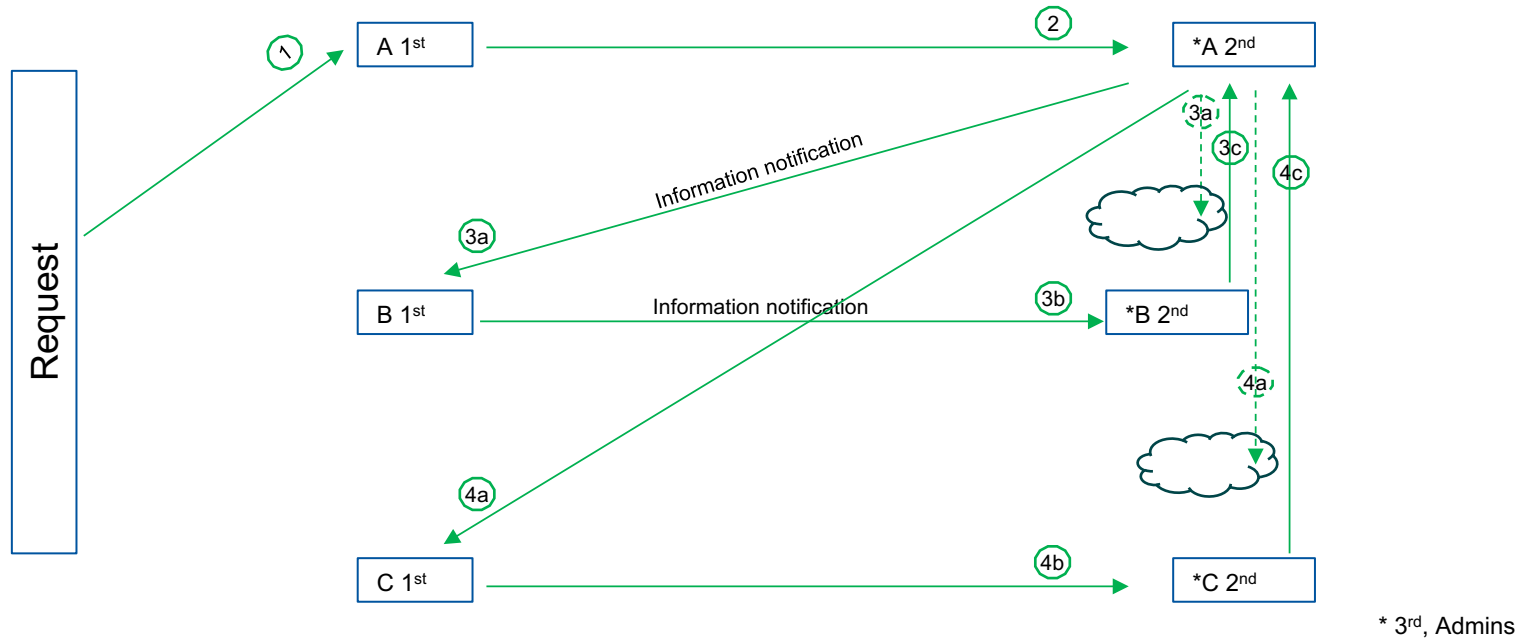
WP1 PE Process: Current activities

- Author whitepapers on **ECM Performance Model** and **Pattern-based PE Process**
- Apply PE process on Proxy Apps:
 - HPCG (almost done)
 - Mini-MD (Mantevo)
 - SPEC OMP 2012 botsalgn and smithwa
- Integrate common community codes into RRZE HPC-Bench for automated benchmarking and performance analysis

WP2 - Process management

Creation of a multi-tier distributed support structure

- Establish a service structure for a performance engineering process
- Formulate terms and conditions of a distributed support structure
- Describe a support infrastructure allowing to transfer requests between sites based on a defined sequence of actions for supporters at three support levels.



WP3 – Performance Monitoring and Analysis

● Objectives

- Establish an automatic rating of the performance footprint of applications running on a production system

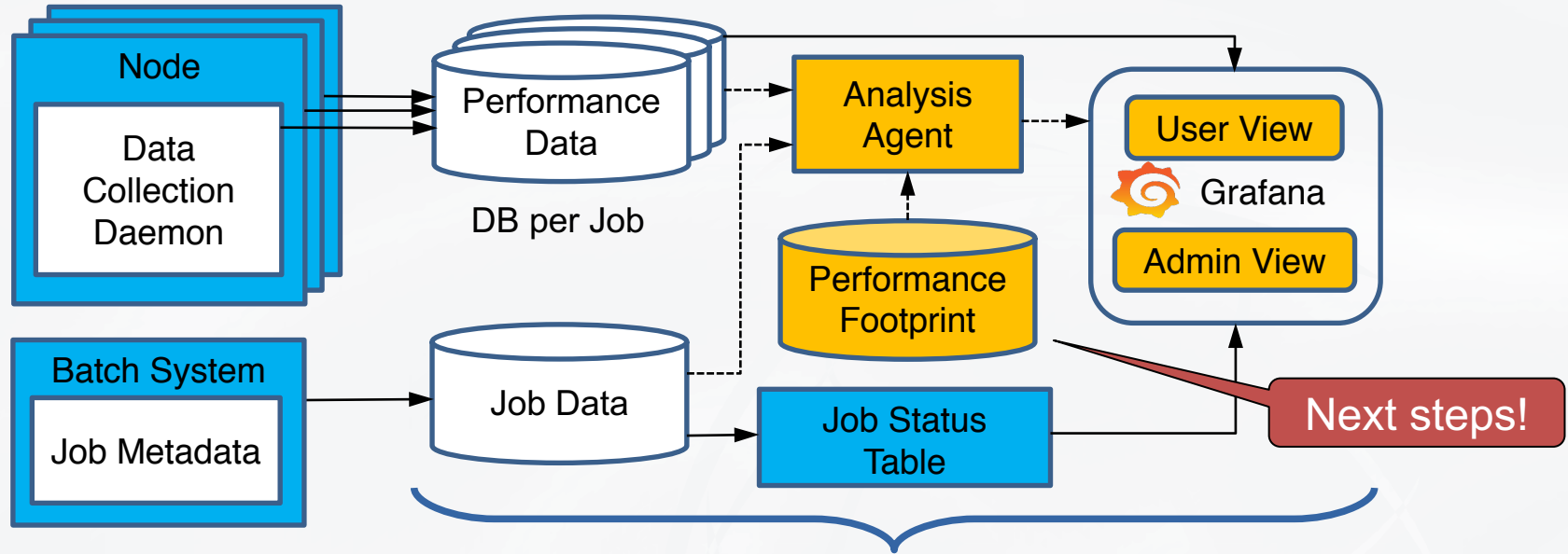
● Current State

- Identification of suitable performance metrics
 - CPI, FLOP/s, Main Memory Bandwidth, I/O, Network
- Collection of job-specific performance metrics without instrumenting individual applications
 - Diamond
- Live (and post-mortem) visualization of performance data
 - Grafana

● Next Steps

- Analysis of performance footprints to automatically detect performance issues

WP3 – Performance Monitoring and Analysis



Service VM

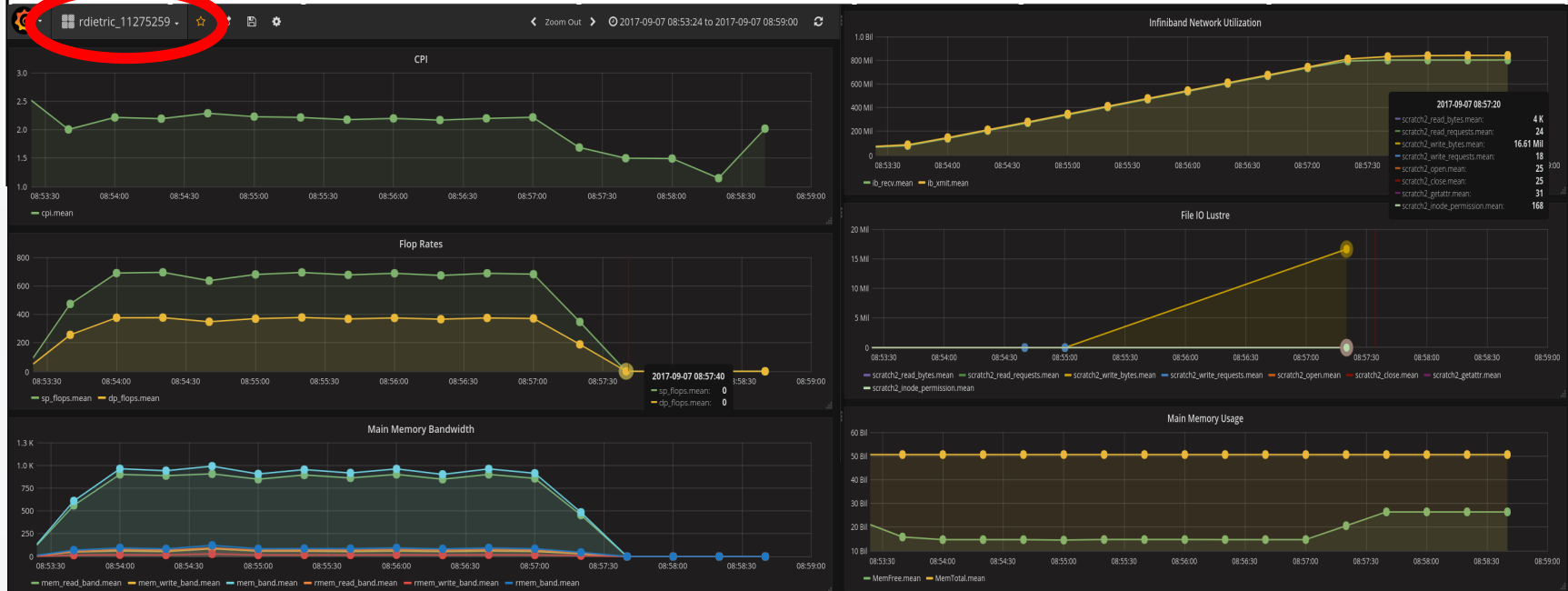
Collectors	Performance Data
Likwid	<ul style="list-style-type: none"> - FLOP/s, CPI - Memory bandwidth
Lustre	<ul style="list-style-type: none"> - Read/write bytes/requests - Read/write calls for individual block sizes - Metadata
Infiniband	<ul style="list-style-type: none"> - Xmit, Recv
Main Memory	<ul style="list-style-type: none"> - total/free/available/usage

Batch System	Job Data
SLURM	<ul style="list-style-type: none"> - Start and end time - All SLURM environment variables that are available in prolog and epilog e.g. SLURM_JOB_ID SLURM_JOB_USER

WP3 – Performance Monitoring and Analysis

Job Status Table View → Grafana

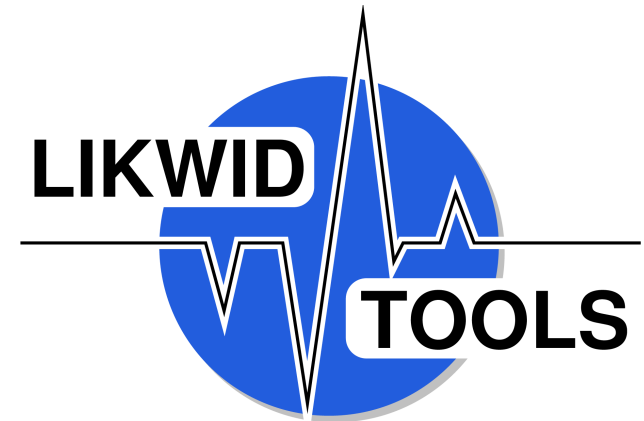
JOB ID	USER	START	END	STATUS	NUM_NODES	NODELIST
11275259	dietric	07/09/2017 08:53	07/09/2017 08:58	completed	2	taurusi[3179-3180]
11275061	rotscher	07/09/2017 06:55	07/09/2017 06:56	completed	1	tauruskn1
11275055	rotscher	07/09/2017 06:48	07/09/2017 06:48	completed	1	tauruskn1
11275054	rotscher	07/09/2017 06:47	07/09/2017 06:47	completed	1	tauruskn1



WP3 Performance Monitoring RRZE

- LIKWID 4.3.0 will be released soon:
 - Intel Skylake support
 - AMD ZEN/EPYC support
 - perf_event backend

- Likwid Monitoring Stack released
<https://github.com/RRZE-HPC/LMS>



T. Röhl, J. Eitzinger, G. Hager, and G. Wellein: *LIKWID Monitoring Stack: A flexible framework enabling job specific performance monitoring for the masses*. [HPCMASPA 2017](#), held in conjunction with IEEE Cluster 2017, Honolulu, HI, September 5, 2017

Develop a coherent, nationwide HPC curriculum

- Examine and structure online course material in Germany (GA, GCS), EU (PRACE) and USA
- Target groups: user, developer, admin, HelpDesk, staff member, HPC experts, domain experts

HPC Knowledge Base

- Setup a central web platform for coherent site independent documentation of HPC related material.
- A MediaWiki has been set up to enable moderated user contributions and discussions
- Examine existing online material in Germany (GA, GCS), EU (PRACE) and USA

Conclusion and Outlook

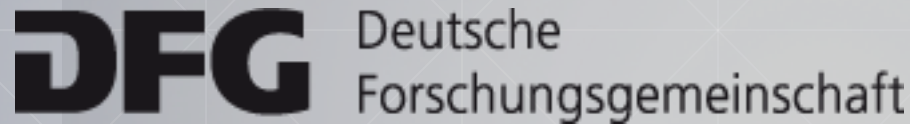
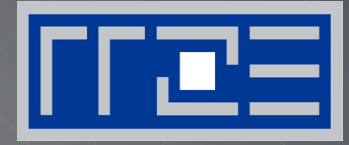
- Work packages are **on track**
- Further establish **closer contact** to other projects of call

Outlook:

- We need to review the working plan with regard to user software optimization projects
- Eventually initiate own PE analysis effort on community codes with user provided input
- **Focus** on effective activities due to limited work force

Do you have a candidate for a PE project? Contact us!

ERLANGEN REGIONAL COMPUTING CENTER



TECHNISCHE
UNIVERSITÄT
DRESDEN

Thank you for your attention!