Methodology for Development Effort & Productivity Estimation in HPC –

Getting HPC Developers Involved

Sandra Wienke

IT Center, RWTH Aachen University, Germany

GPU Hackathon, Lugano, 07/04/2016

IT Center | RWTH AACHEN UNIVERSITY

Whoami?

- M.Sc. in Technomathematics
- HPC Group, IT Center, RWTH Aachen University
  - Organizations: SPEC HPG
  - Teaching assistant (HPC software labs & lectures)
  - Tutorials, Workshops
  - GPU support for the RWTH Compute Cluster
- Research focus
  - Analysis of Total Cost of Ownership (TCO) & Productivity of HPC Centers

- Why I came as mentor to the Hackathon?
  - Help researcher to accelerate their code
  - Own research on real development effort
    - Impact factors (surveys)
    - Productivity (electronic developer diaries)

  → Accepted paper at SC16 „Development Effort Estimation in HPC“


Source: Wikipedia

Productivity of HPC Centers

- In pursuit of exaflop computing: expenses of HPC centers increase
  - Acquisition, energy, staff/ labor costs, programming
- Informed decision on how to invest available budgets
  - Quantifiable metric for benefit cost ratio needed[3]

$$productivity = \frac{value}{cost} = \frac{\#app.runs}{total\ cost\ of\ ownership\ [\$]}$$

$$\$ = HW + energy + development\ cost + \ldots$$

$$development\ effort\ [\text{day}] * salary\ [\$/\text{day}]$$

- **How to estimate development efforts in HPC?**

[3] S. Wienke, H. Iliev, D. an Mey, M.S. Müller: Modeling the Productivity of HPC Systems on a Computing Center Scale. LNCS, vol. 9137, pp. 358–375. Springer International Publishing (2015)

Development Effort Estimation in HPC

- **Aim**: model for development effort in HPC
  - Assume: $\text{effort} = S \cdot f(\text{performance})^R$
    with S, R combination of parameters
  → Relationship of development effort & performance

- **Challenge**: numerous impact factors on effort

- **Needed**: data base to rely statistical methods on!

[5]E. Nuhfer, D. Knipp: The knowledge survey: a tool for all reasons. To Improve the Academy, vol 21, pp. 59-78 (2003).

Getting HPC Developers Involved

Thank you for your participation!

- Tools
  - **Knowledge Survey**[5] (https://www.soscisurvey.de/ksHack16/?q=CUDA, https://www.soscisurvey.de/ksHack16/?q=OpenACC)
    - Do not really answer questions, instead rate level of confidence to answer question
    - Impact of PRE-knowledge on development productivity (4.7.2016)
    - Questions on application, parallel programming, GPUs (~30 min)
  - **Effort Log Tool** (https://github.com/julianmi/effort-log)
    - Electronic developer diary (Lin, Win, Mac)
    - Interval-based: What did you do? Tuning knobs?
    - Milestones for performance-effort relationship!
  - **Impact factors on development effort**
    - Rank which factors have most impact (e.g. tools)

ELA, CSCS
(use binary which statically inlcudies all needed libraries)

```
cp /project/hck/effortlog/
effort-log_v0.7.3_linux64 .

./effort-log_v0.7.3_linux64
```

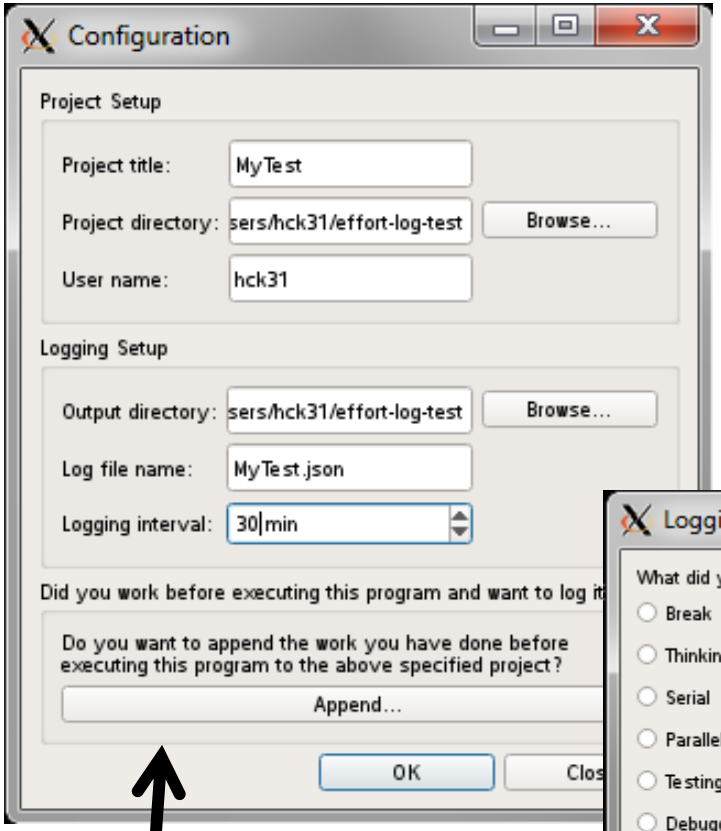
- **Mapping** of surveys, effort logs, performance makes our research more valuable
  - Please use the same (arbitrary) nick name (data will be anonymously treated)

Selected Publications

1) S. Wienke, J. Miller, M. Schulz, M.S. Müller: **Development Effort Estimation in HPC**. Paper accepted at International ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC16), November 2016, Salt Lake City, UT, USA.
2) S. Wienke, T. Cramer, M.S. Müller, M. Schulz: **Quantifying Productivity - Towards Development Effort Estimation in HPC**. Scientific poster at the International ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC15), November 2015, Austin, TX, USA.
3) S. Wienke, H. Iliev, D. an Mey, M.S. Müller: **Modeling the Productivity of HPC Systems on a Computing Center Scale**. In: Kunkel, J.M., Ludwig, T. (eds.) High Performance Computing, Lecture Notes in Computer Science, vol. 9137, pp. 358–375, Springer International Publishing, 2015.
4) S. Wienke, D. an Mey, D., M.S. Müller: **Accelerators for Technical Computing: Is It Worth the Pain? A TCO Perspective**. In: Kunkel, J., Ludwig, T., Meuer, H. (eds.) Supercomputing, Lecture Notes in Computer Science, vol. 7905, pp. 330–342, Springer Berlin Heidelberg, 2013.

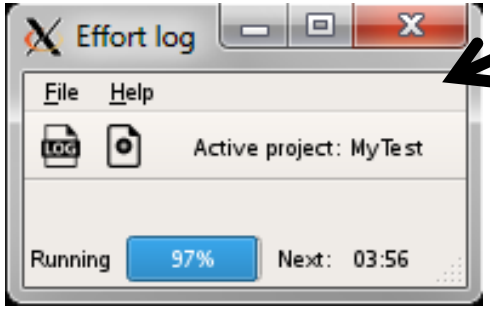
Further Information on *EffortLog*

EffortLog - Interface



Setup frame

- Specify interval (e.g. 30min)
- Specify local file name

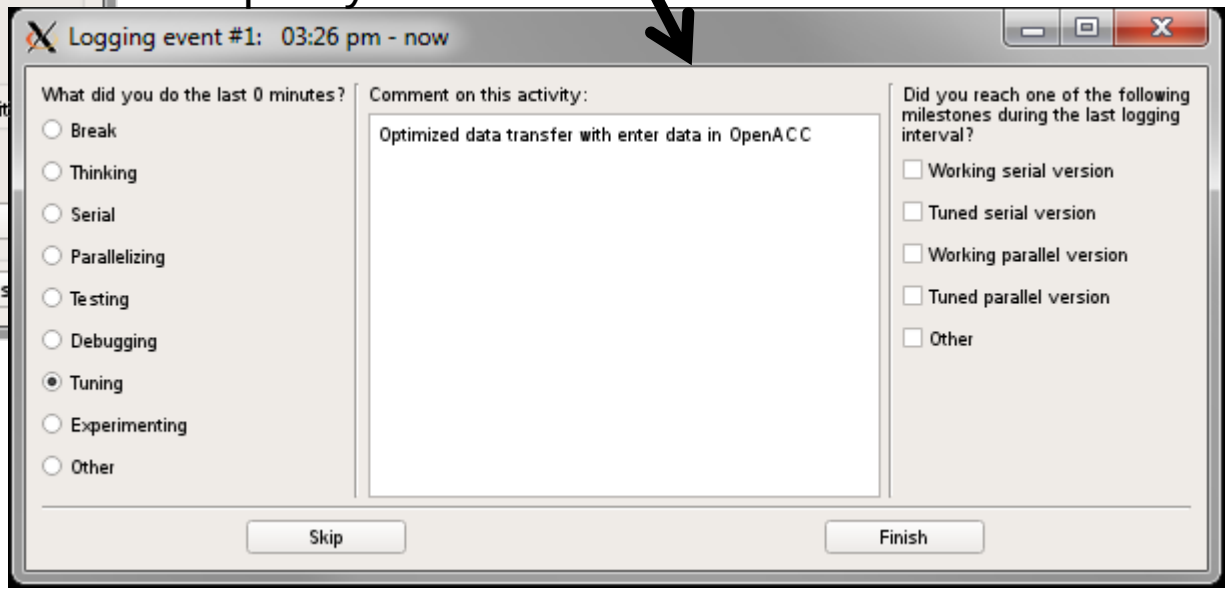


Log Statistic

- Timer
- Manually log entries
- Json files

Log Events

- Specify kind of activity & comment on it
- Specify milestones



EffortLog - Milestones

- *Scratch*: No code base given - just the problem statement
- *Working serial version*: Serially-running version of your code that was tested for correctness, but not tuned for performance, e.g., 1st correct code implementation
- *Tuned serial version*: Serially-running version of your code that was tested for correctness and tuned for performance
- *Working parallel version*: In-parallel running version of your code that was tested for correctness, but not highly-tuned for performance, e.g. 1st correct parallel version of your code
- *Tuned parallel version*: In-parallel running version of your code that was tested for correctness and tuned for performance
- *Other*: Specify this code version in a later step.