

# BenchCloud

A Platform for Scalable Performance Benchmarking

---

Dirk Beyer, **Po-Chun Chien**, and Marek Jankola

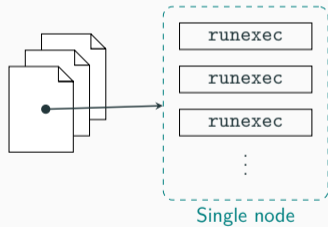


ASE 2024 @ Sacramento  
SoSy-Lab, LMU Munich



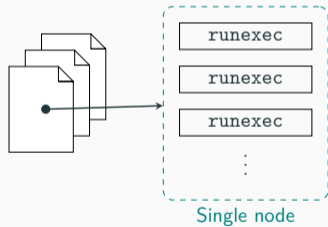
# Motivation

BENCHEXEC [5]

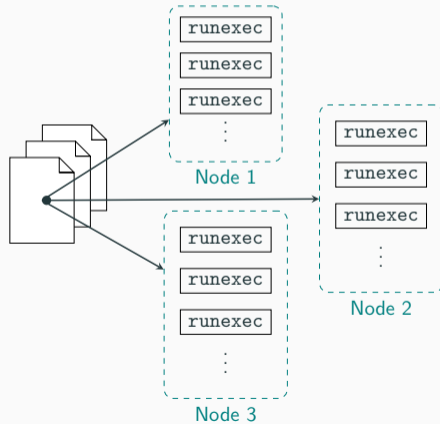


# Motivation

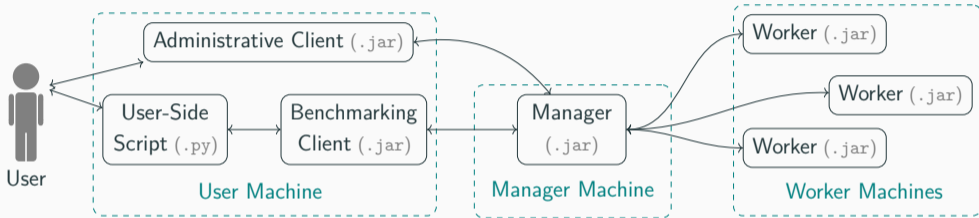
BENCHEXEC [5]



BENCHCLOUD

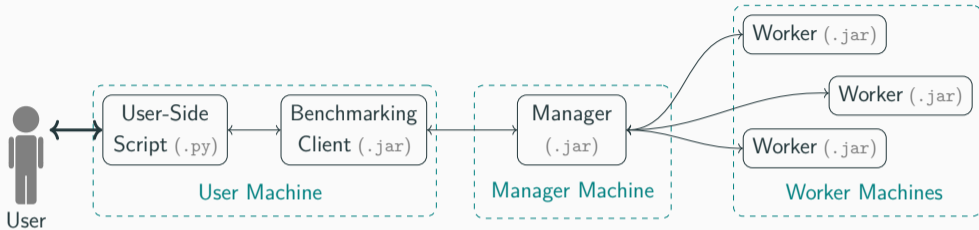


# Overview of BenchCloud



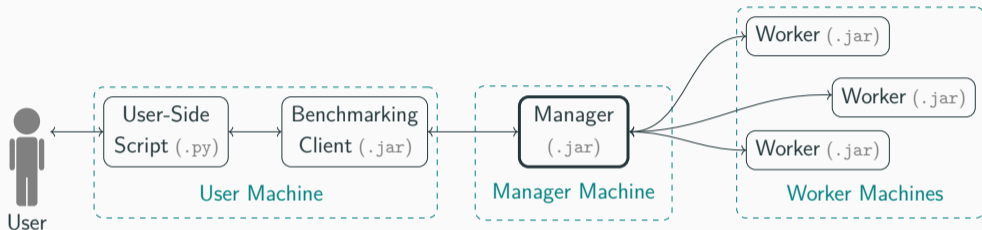
- Reliable and scalable performance benchmarking
- Seamless integration with `BENCHEXEC`
- Used by SMT-COMP [7], SV-COMP [2], and Test-Comp [1]

# Benchmarking: Inputs and Outputs



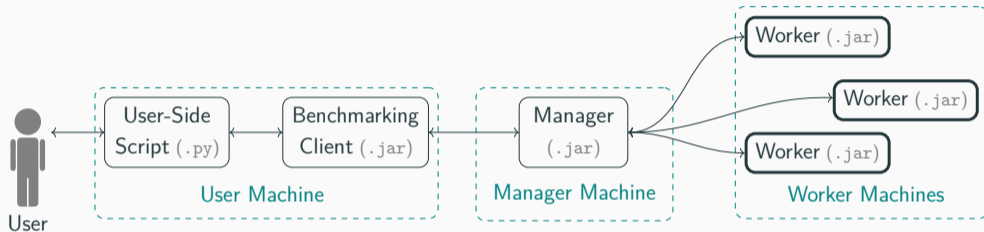
- Inputs:
  - Benchmark definitions: resource limits, CPU specs, tasks, tool configs
  - Precompiled tool executables
- Outputs:
  - Tool outputs and resource measurements
  - Postprocessing and visualization with `table-generator`

# Benchmarking: BenchCloud Manager



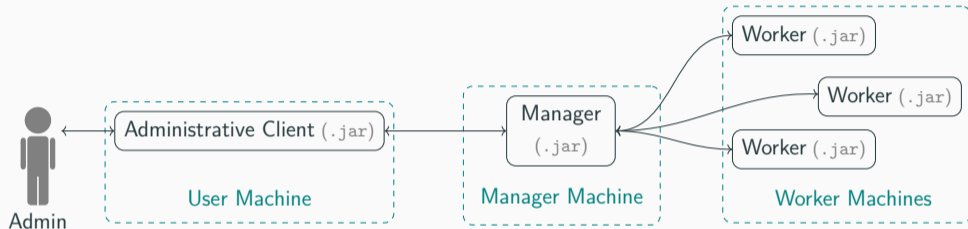
- Distribute runs to nodes aligning with user's specifications
- Schedule runs according to *priorities*
- Aggregate benchmarking results from the workers

# Benchmarking: Workers



- Start each benchmark run with `runexec`
- Precise resource management via `cgroups`
- Isolated and containerized execution via `namespaces`

# System Administration



- Add/remove worker nodes
- Monitor worker status
- Renice/cancel benchmark runs



# Use Cases

- International tool competitions in automated reasoning: SMT-COMP [7], SV-COMP [2], and Test-Comp [1]
- CPAchecker's [4] nightly regression tests



- 4 running instances at European research institutes

# Conclusion

- Facilitate large-scalable benchmarking
- Easy to use and manage
- Running system:  
<https://benchcloud.sosy-lab.org/>
- Source code available on Zenodo [3]
- Future plan: support tool execution in customizable containers [6]



## References i

- [1] Beyer, D.: Software testing: 5th comparative evaluation: Test-Comp 2023. In: Proc. FASE. pp. 309–323. LNCS 13991, Springer (2023).  
[https://doi.org/10.1007/978-3-031-30826-0\\_17](https://doi.org/10.1007/978-3-031-30826-0_17)
- [2] Beyer, D.: State of the art in software verification and witness validation: SV-COMP 2024. In: Proc. TACAS (3). pp. 299–329. LNCS 14572, Springer (2024).  
[https://doi.org/10.1007/978-3-031-57256-2\\_15](https://doi.org/10.1007/978-3-031-57256-2_15)
- [3] Beyer, D., Chien, P.C., Jankola, M.: BENCHCLOUD release 1.1. Zenodo (2024).  
<https://doi.org/10.5281/zenodo.13742756>
- [4] Beyer, D., Keremoglu, M.E.: CPACHECKER: A tool for configurable software verification. In: Proc. CAV. pp. 184–190. LNCS 6806, Springer (2011).  
[https://doi.org/10.1007/978-3-642-22110-1\\_16](https://doi.org/10.1007/978-3-642-22110-1_16)

## References ii

- [5] Beyer, D., Löwe, S., Wendler, P.: Reliable benchmarking: Requirements and solutions. *Int. J. Softw. Tools Technol. Transfer* **21**(1), 1–29 (2019).  
<https://doi.org/10.1007/s10009-017-0469-y>
- [6] Beyer, D., Wachowitz, H.: FM-WECK: Containerized execution of formal-methods tools. In: *Proc. FM. LNCS 14934*, Springer (2024).  
[https://doi.org/10.1007/978-3-031-71177-0\\_3](https://doi.org/10.1007/978-3-031-71177-0_3)
- [7] Weber, T., Conchon, S., Déharbe, D., Heizmann, M., Niemetz, A., Reger, G.: The SMT competition 2015-2018. *J. Satisf. Boolean Model. Comput.* **11**(1), 221–259 (2019).  
<https://doi.org/10.3233/SAT190123>