Popper: Practical Reproducible Evaluation of Systems
Ivo Jimenez, Michael Sevilla, Noah Watkins, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau, Jay Lofstead, Carlos Maltzhan, Kathryn Mohror

Reproducibility as a DevOps Problem
• Independently validating experimental results is challenging.
• Recreating experimental setup is often difficult to impossible.
• Software engineers deal with reproducibility all the time:
  - Bug A can be reproduced in version X on platform Y using input Z.
• Manage an academic article as a software project!

Import Existing Projects
$ cd mypaper-repo
$ popper init
  -- Initialized Popper repo mypaper-repo
$ popper search
  -- available templates
  ceph-rados proteusntn epi-com man adjun sirius cond-opem
  cloverleaf gasysfs cling bwa unum-py cudnn-deaplim
  spark-stand torpor malacology geneve mantle rita-ide
  hadoop-yarn kubsched alg-encycl macrob domvisor obfuscdata
$ popper add gasysfs
  -- Added gasysfs experiment to mypaper-repo

Project Structure

Scaffolding
$ popper init exp1
  -- initialized exp1 pipeline.
  $ ls -1 pipelines/exp1/
total 20K
  +rw-r--r--  1 ivo ivo README.md
  +rw-r--r--  1 ivo ivo run.sh
  +rw-r--r--  1 ivo ivo setup.sh
  +rw-r--r--  1 ivo ivo teardown.sh
  +rw-r--r--  1 ivo ivo validate.sh

Validation
$ popper check exp1
  Popper check started
  Stage: setup.sh ............
  Stage: run.sh ..............
  Stage: teardown.sh ....
  Popper check finished
  Status: SUCCESS

Codified Validations:
WHEN
  NOT network_saturated AND run_nodes==
  EXPECT
  system_throughput >> (baseline_throughput * 0.9)

Benefits and Challenges
Pros:
• Experiments can be falsifiable with minimal re-execution effort.
• Facilitates collaboration by following the OSS model for sharing.
• Investing time in DevOps skills quickly pays off.

Challenges:
• Big cultural change; new experimentation paradigm.

The Convention:
1. Pick a DevOps tool for each stage of the scientific experimentation workflow.
2. Put all associated scripts (experiment and manuscript) in version control, in order to provide a self-contained repository.
3. Document changes as experiment evolves, in the form of version control commits.

Popper: Take a common generic experimentation workflow (above) and apply a DevOps practice used in the development of open source software (OSS) projects (below).

Our Approach

Popper Compliance

Tools: Generate referenceable assets (associate unique IDs to binaries, data, configuration and infrastructure state); usable from scripts/CLI and capable of acting upon IDs.

Experiment: Provide all necessary assets in a single repository (self-contained), including experiment code, orchestration logic, data dependencies, results and validation criteria.

Article: Provide full text and figures of article, as well as all popper-compliant experiments.