On the reproducibility of experiments in the Cloud

Damien Saucez with A. Quereilhac, W. Dabbous, H. Soni, T. Turletti…

Inria Sophia Antipolis

Today's talks

- 17 had an evaluation part.
  - 3 used analytical study.

How many of them can John Doe repeat?

  - 8 used testbed or real deployment.
  - 1 used an unknown technique.

Evaluation! Why?

- Validate concepts and ideas,
- understand complex systems,
- compare solutions.

How to evaluate?

- Analytical methods
  - e.g., mathematical models
- Empirical methods
  - e.g., simulations, emulations, deployments

Tools to support evaluation

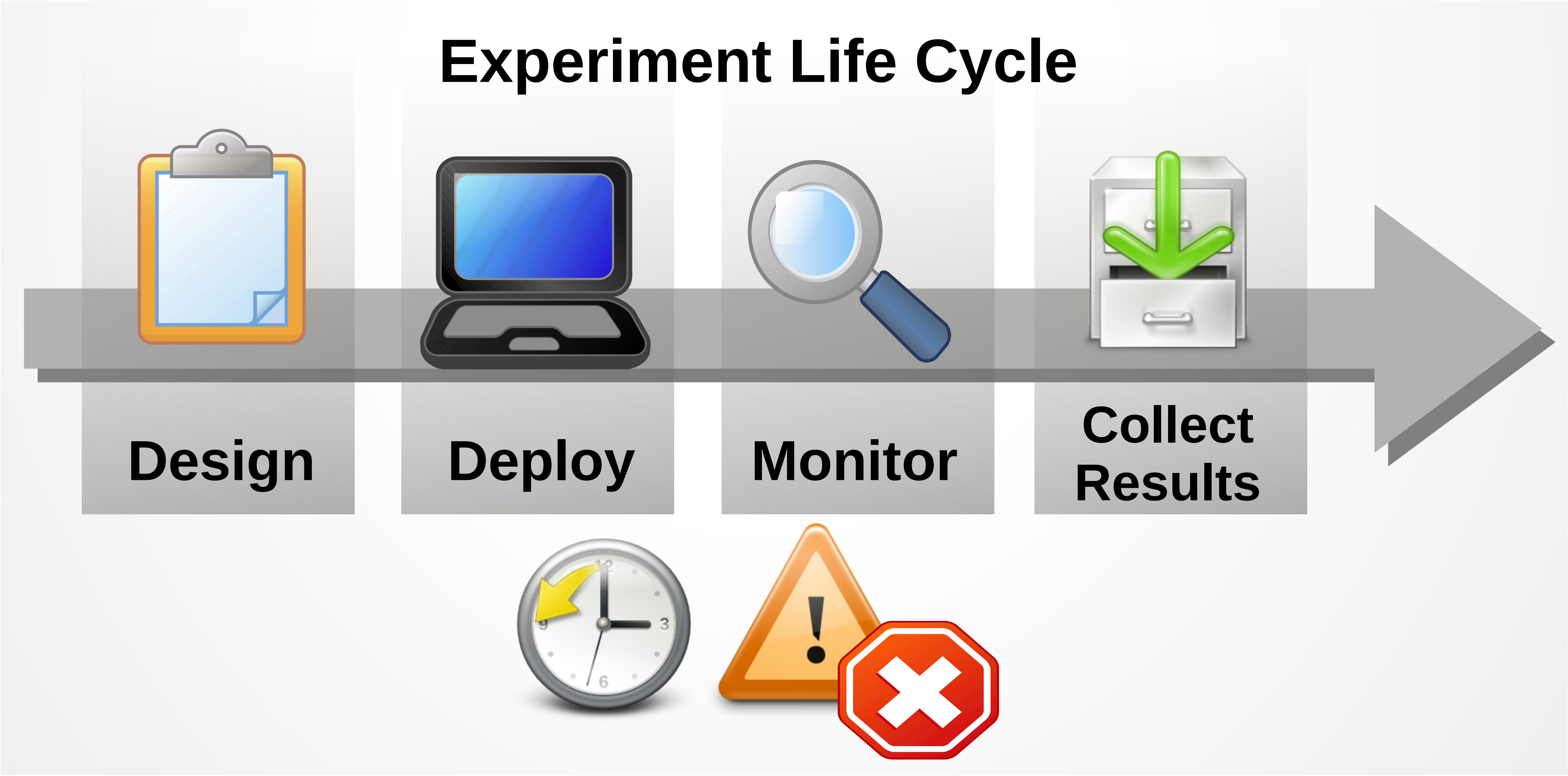
- Theory.
- Simulators
  - imitate the behaviour of real systems.
- Emulators
  - combine system models with real system components.
- Testbeds
  - dedicated infrastructures to conduct experiments under live conditions.
- Live deployment
  - experiment in the targeted environment.

3R: Repeatability, Reproducibility, Replicability

- Repeat: clone the experiments results on a given platform.
- Reproduce: clone the procedure to conduct an experiment on a given platform.
- Replicate: clone the experiment scenario on a different (similar) platform.

Automate evaluations is essential

- Rigorous experimentation is multifold and challenging



NEPI to automate

- NEPI, Network Experiment Programming Interface, is a framework to automate network experiments
  - that abstracts components behind a common interface: the resource
  - to automate experimentation steps.
- Runs locally, no need to modify the experiment facility
  - e.g., ns-3, PlanetLab, Grid’5000.
- http://nepi.inria.fr

And the Cloud?

The fundamentals of the Cloud

- Service sharing is the foundation of cloud computing.

How to reach the 3 Rs?

  - Partial isolation.
  - Elasticity.

Observations

- Cloud services are deployed in data centers.
  - Easy to rent resources in the infrastructure
  - but hard to modify the infrastructure (hard or soft).
- Data centers are expensive to build.
- Each data center is different.

Observations (contd.)

- Research community has easy access to grid computing infrastructures.
  - Enormous processing power.
  - High speed network.
  - High quality storage.
- e.g., Grid’5000.

Data centers in the Grid - DiG

- Emulate data center topologies in a Grid
  - respecting computing and network resource constraints
  - with performance guarantees
- to run cloud applications and algorithms.

The 3 steps of DiG

- Experimental Network Embedding
- Configuration Generation
- Deployment

Experimental Network Embedding

- Target network and physical grid infrastructure modelled in DOT language
  - using labels to represent workload and constraints.
- Resolution of the Virtual Network Embedding (VNE) problem
  - with ALEVIN [BLF+14]
- to generate a node mapping file.

[BLF+14] M. Beck, C. Linnhoff-Popien, A. Fischer, F. Kokot, and H. de Meer, "A simulation framework for virtual network embedding algorithms," in Telecommunications Network Strategy and Planning Symposium (Networks), 2014 16th International, Sept 2014, pp. 1–6.

Configuration Generator for Experimental Network

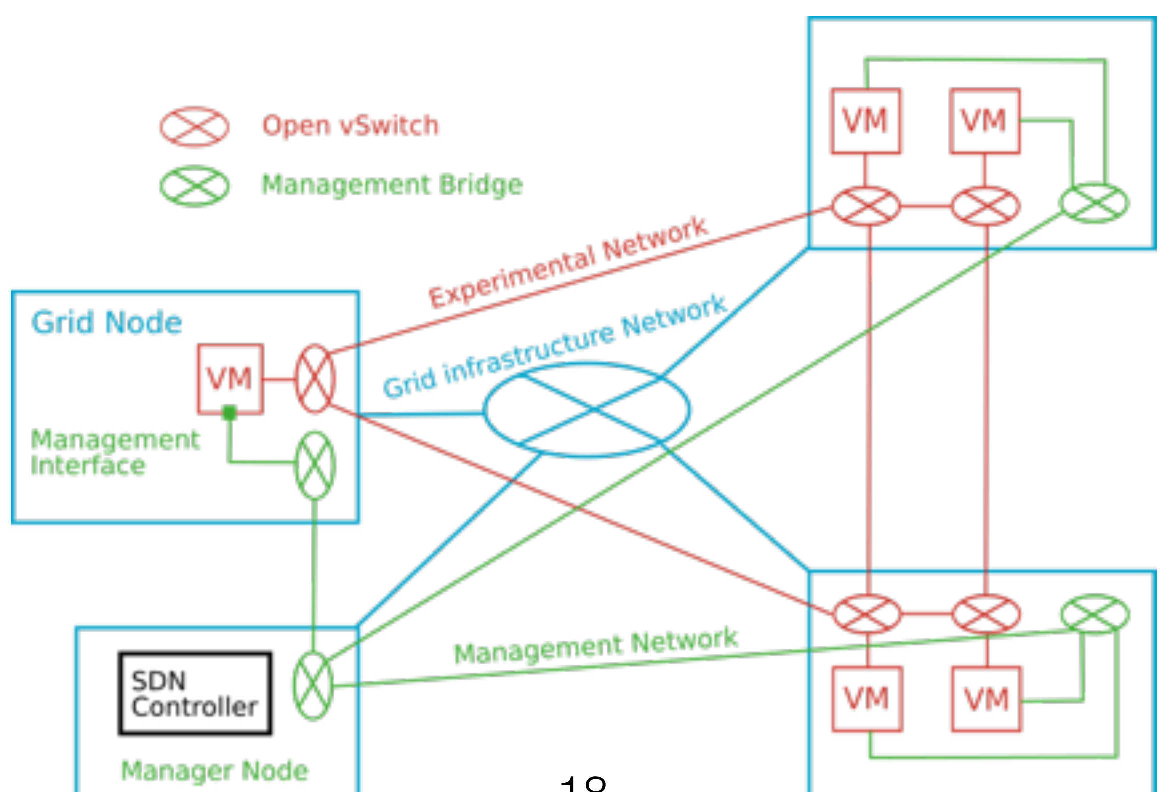
- Prepare configurations to be put on each resource of the grid infrastructure.
  - Softwares to install.
  - Docker and VMs.
  - Resource limitations, CPU affinity.

Deployment

- Build a L2 overlay network in the grid to carry the experiment
  - OVS switches
  - L2TPv3 tunnels
  - network conditions controlled with *tc*.
- Deploy the experiment on top of the overlay based on the configurations.

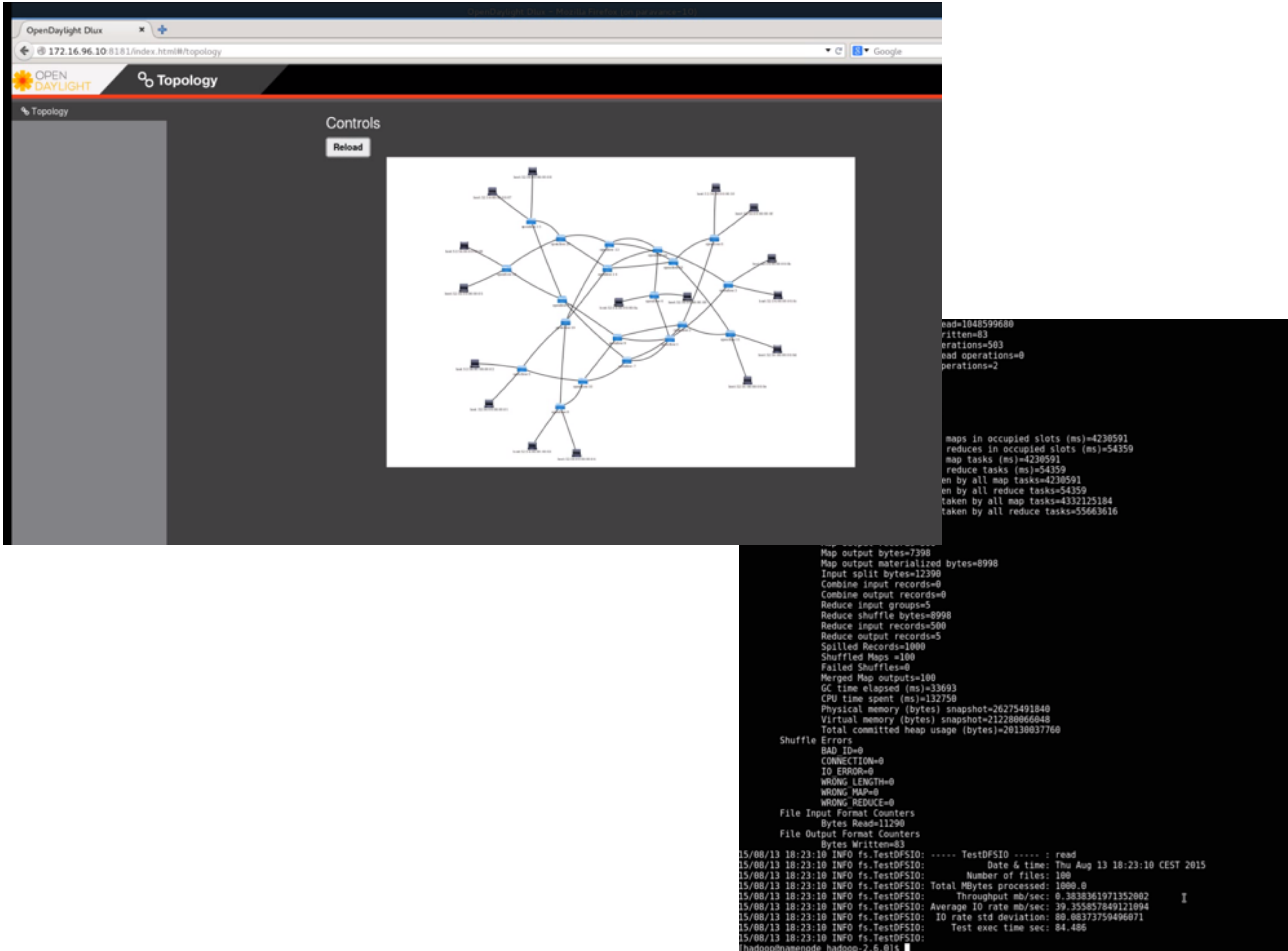
Deployment (contd.)

- Centralised management of the experiment.
  - Dedicated node in the grid.
  - Dedicated (isolated) virtual network.



Demo

- Hadoop benchmark suite in a fat-tree data center on Grid 5000
- https://www.youtube.com/watch?v=zikbQN8B7OE


OpenDaylight Dlux
172.16.96.10:8181/index.html#/topology
Google
OPEN DAYLIGHT
Topology
Topology
Controls
Reload
ead=1048599680
ritten=83
erations=503
ead operations=0
perations=2
maps in occupied slots (ms)=4230591
reduces in occupied slots (ms)=54359
map tasks (ms)=4230591
reduce tasks (ms)=54359
en by all map tasks=4230591
en by all reduce tasks=54359
taken by all map tasks=4332125184
taken by all reduce tasks=55663616
Map output bytes=7398
Map output materialized bytes=8998
Input split bytes=12390
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=8998
Reduce input records=500
Reduce output records=5
Spilled Records=1000
Shuffled Maps =100
Failed Shuffles=0
Merged Map outputs=100
GC time elapsed (ms)=33693
CPU time spent (ms)=132750
Physical memory (bytes) snapshot=26275491840
Virtual memory (bytes) snapshot=212280066048
Total committed heap usage (bytes)=20130037760
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=11290
File Output Format Counters
Bytes Written=83
15/08/13 18:23:10 INFO fs.TestDFSIO: ----- TestDFSIO ----- : read
15/08/13 18:23:10 INFO fs.TestDFSIO: Date & time: Thu Aug 13 18:23:10 CEST 2015
15/08/13 18:23:10 INFO fs.TestDFSIO: Number of files: 100
15/08/13 18:23:10 INFO fs.TestDFSIO: Total MBytes processed: 1000.0
15/08/13 18:23:10 INFO fs.TestDFSIO: Throughput mb/sec: 0.38383619713520O2
15/08/13 18:23:10 INFO fs.TestDFSIO: Average IO rate mb/sec: 39.355857849121094
15/08/13 18:23:10 INFO fs.TestDFSIO: IO rate std deviation: 80.08373759496071
15/08/13 18:23:10 INFO fs.TestDFSIO: Test exec time sec: 84.486
15/08/13 18:23:10 INFO fs.TestDFSIO:
[hadoop@namenode hadoop-2.6.0]$

Take away message

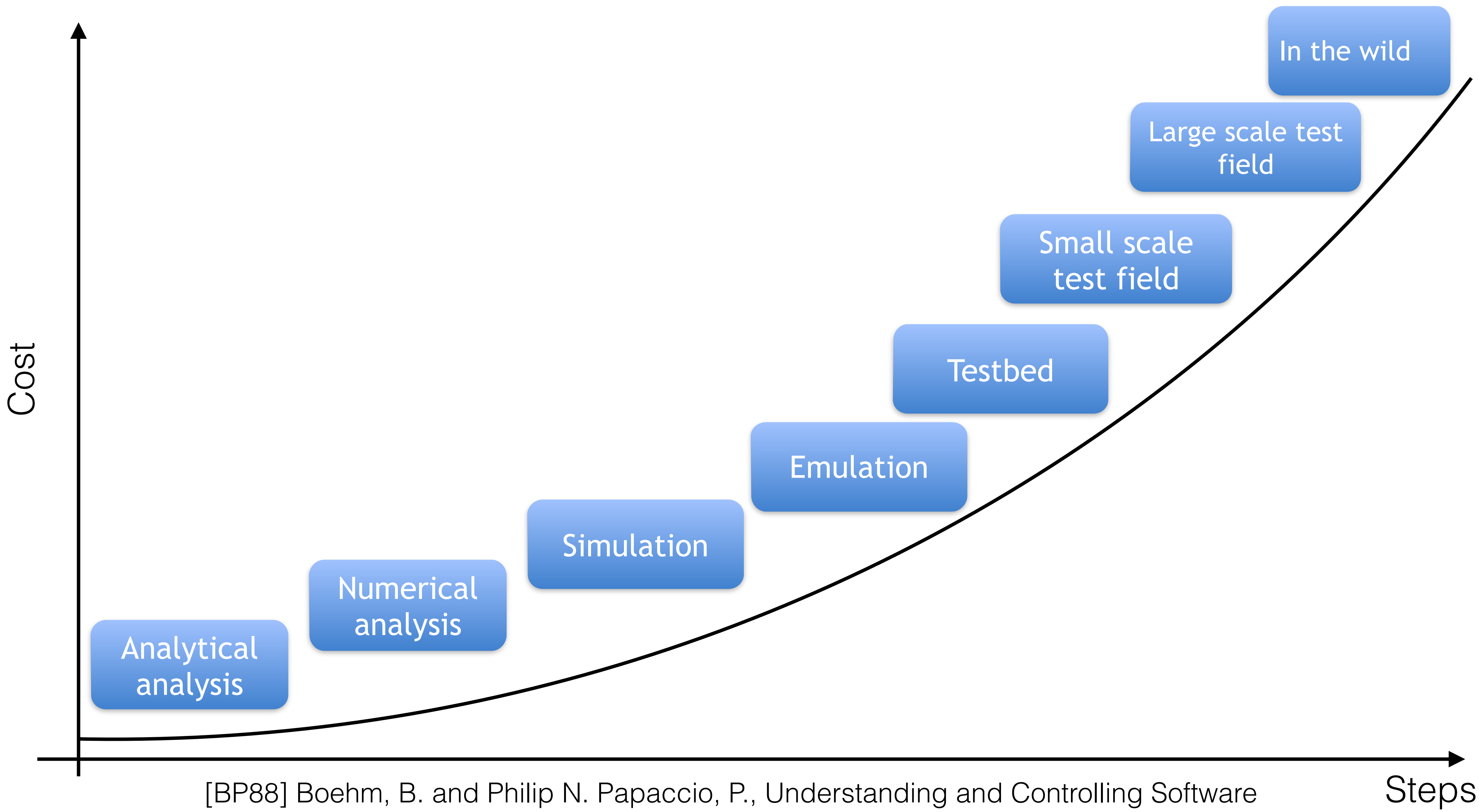
- Cloud incurs large variability.
- A lot of “hidden” hard technical work.
  - ➡ Hard to reproduce experiments.
  - ➡ Hard to compare solutions.
- Let’s keep the 3Rs in mind while performing evaluations.

Question?

Backup

Network solutions evaluation costs

- The cost to fix a defect increases exponentially [BP88].



[BP88] Boehm, B. and Philip N. Papaccio, P., Understanding and Controlling Software Costs, IEEE Transactions on Software Engineering, October 1988

Modules interactions