



# Compas conference, Toulouse, Jul. 2018 KVM Live and Incremental Checkpointing

K. Pouget, A. Rigo, D. Raho

Virtual Open Systems



**PARTNERS**

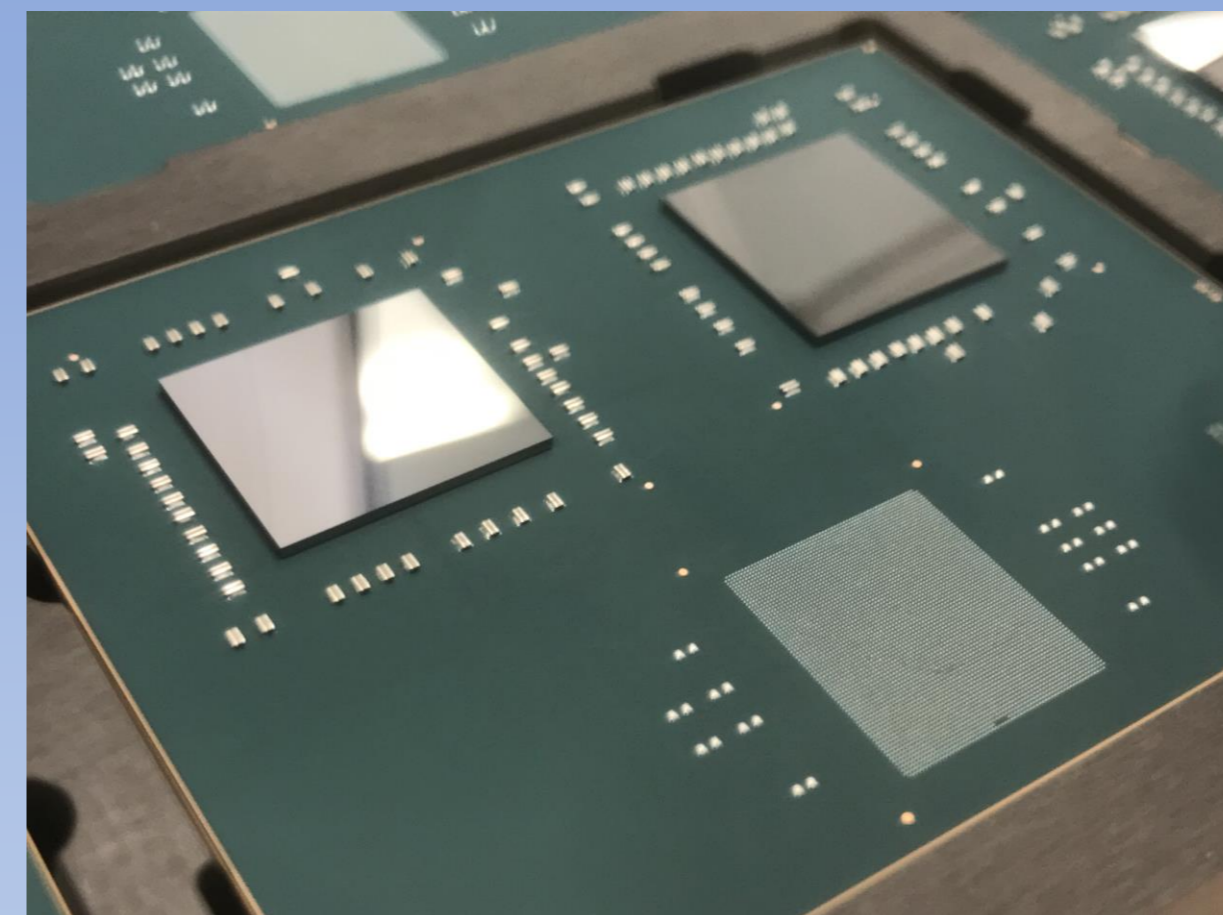


web: [www.exanode.eu](http://www.exanode.eu)



European Exascale Processor  
Memory Node Design

## ExaNoDe Multi-Chip Module



### HPC COMPUTE NODE

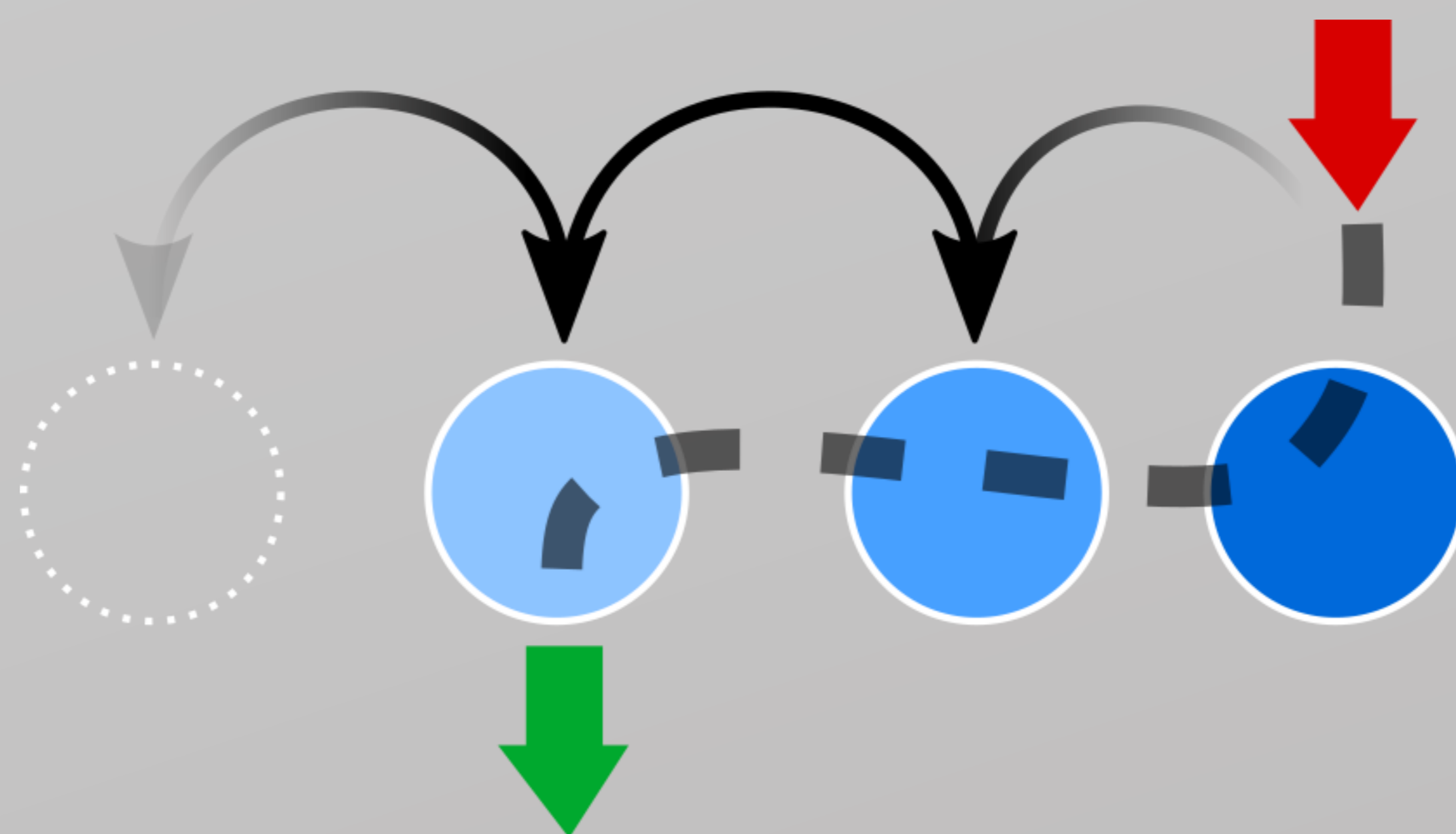
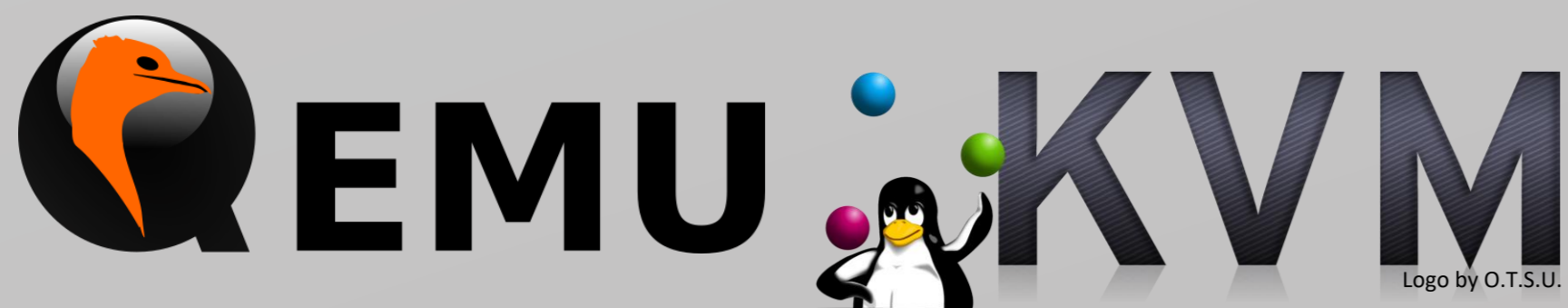
- 32 ARMv8 cores on 4 cm<sup>2</sup>
- UNIMEM shared-memory
- FPGA compute accelerator

Scalability  
+  
Energy efficiency

## KVM CHECKPOINTING

### VIRTUALIZATION

- ✓ **Manageability** for admins
  - pause/unload/migrate VMs
- ✓ **Flexibility** for users
  - full control of SW eco-system
    - kernel, libraries, rootfs
- ✓ **Resiliency** to hardware failures
  - periodic state checkpointing



- Periodically save the state of the VM
- After a crash, reload the last checkpoint
- ... but limit the performance impact!

### LIVE

Keep the VM running during the checkpointing

- ✓ *less guest downtime*

### INCREMENTAL

Save only the memory pages modified since the previous checkpoint

- ✓ *less guest downtime*
- ✓ *lower disk footprint*

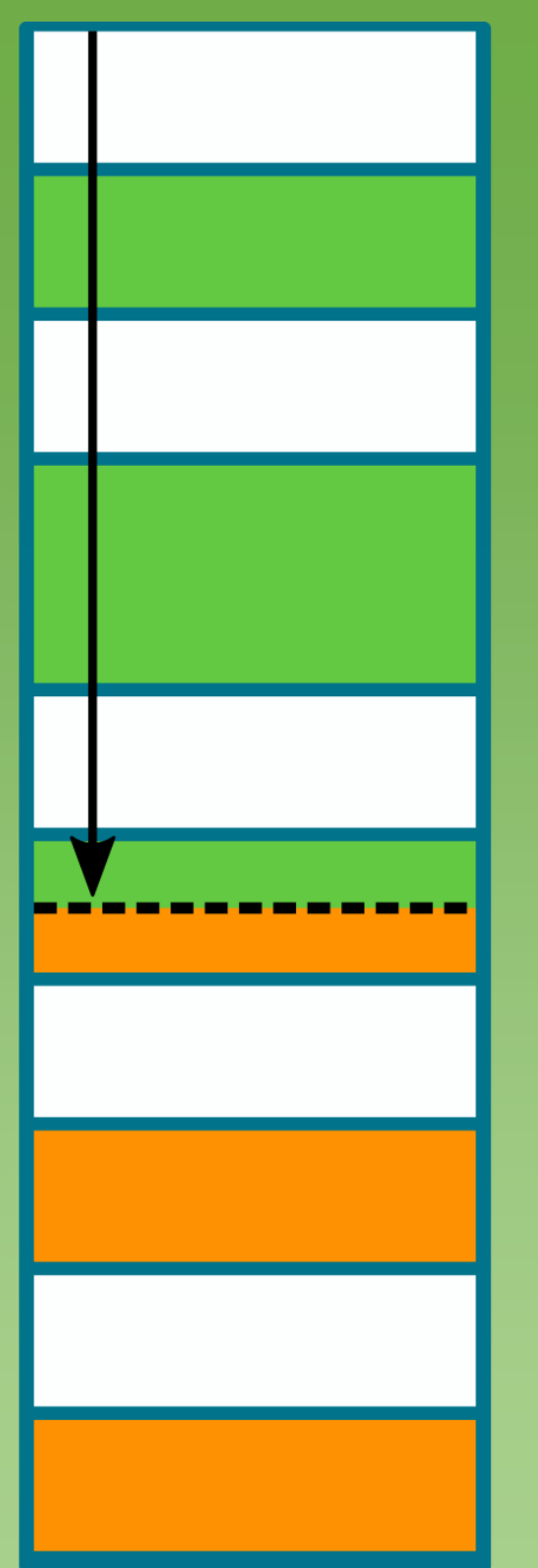
## USERFAULTFD: PAGE FAULT IN USERLAND

On checkpoint request:

- write protect the VM RAM
- start page-saving thread

On write-protection faults:

- copy to shadow memory
- mark for next checkpoint
- remove write protection



Pages not dirty    Dirty pages already saved    Dirty pages not yet saved

## MULTI-VM MULTI-HOST

- For distributed applications
- Running over hosts connected via UNIMEM

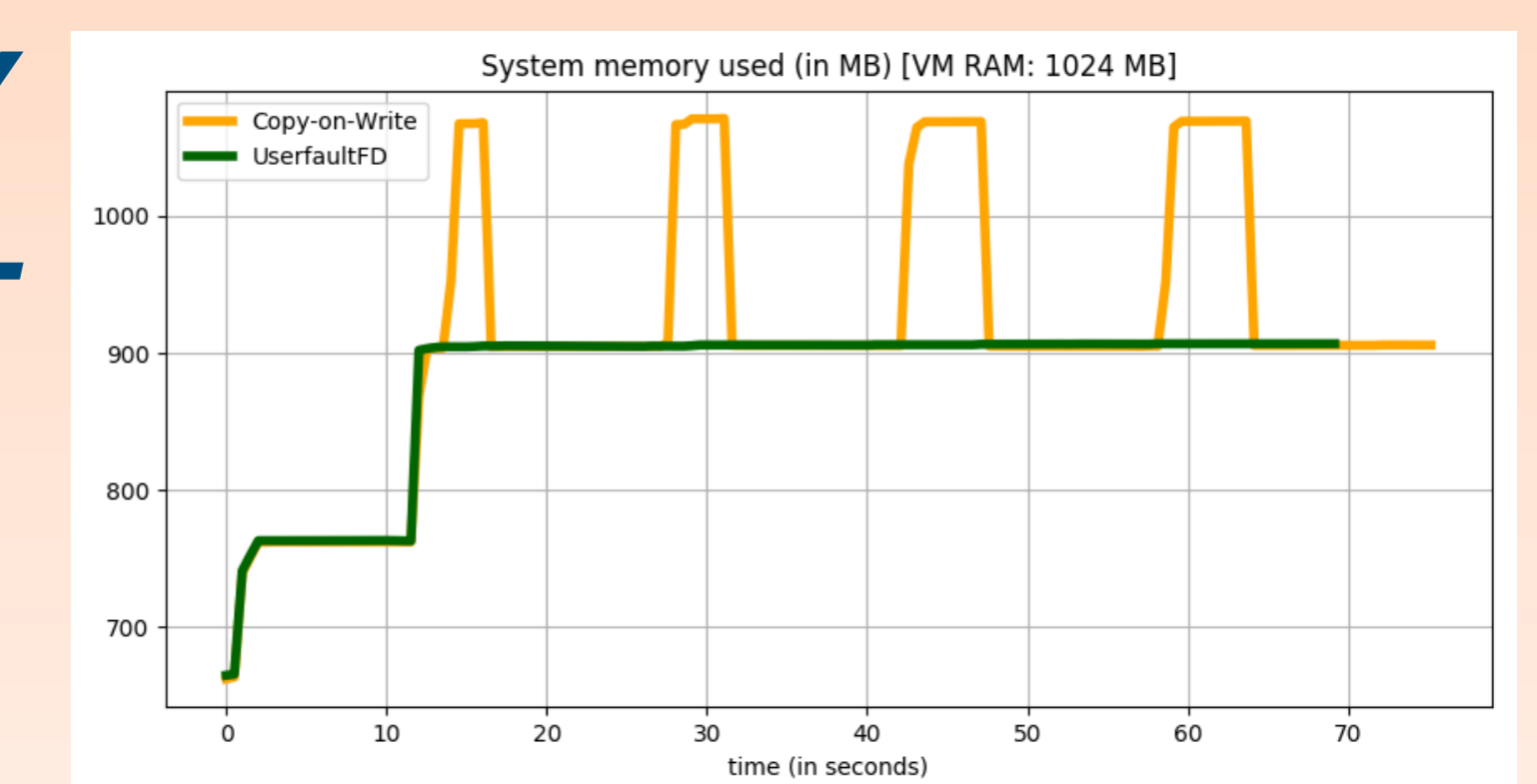
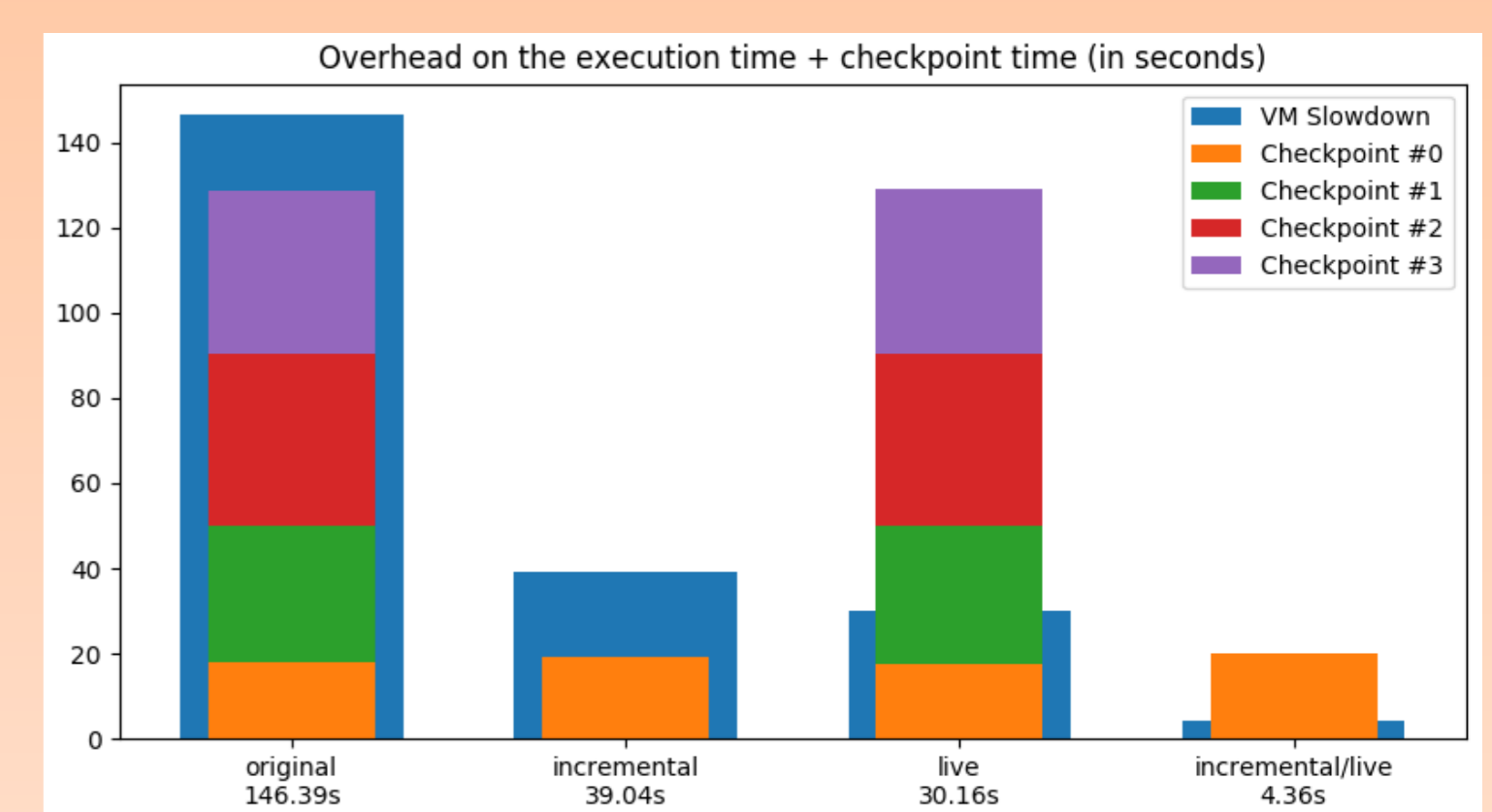
VM CHECKPOINT/RESTART controlled by the application

fine-grained checkpointing

**FAST**

**MEMORY EFFICIENT**

Compared against Copy-on-Write



## IMPLEMENTATION

## RESULTS

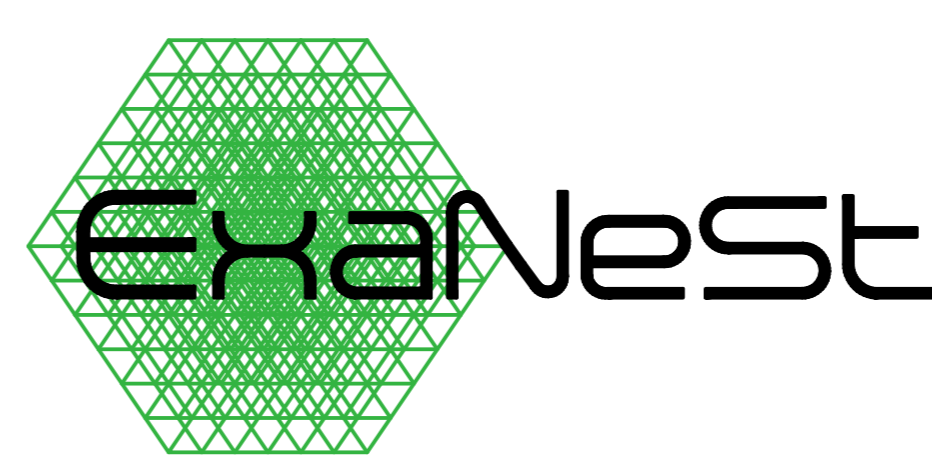
## ExaNoDe as part of a global strategy



[www.euroserver-project.eu](http://www.euroserver-project.eu)



[www.ecoscale.eu](http://www.ecoscale.eu)



[www.exanest.eu](http://www.exanest.eu)



[euroexa.eu](http://euroexa.eu)

web: [www.exanode.eu](http://www.exanode.eu)

<https://twitter.com/ExanodeProject>

<https://www.facebook.com/Exanode-1669383456699997/>

This ExaNoDe research project is supported by the European Commission

under the "Horizon 2020 Framework Programme"

with grant number 671578

