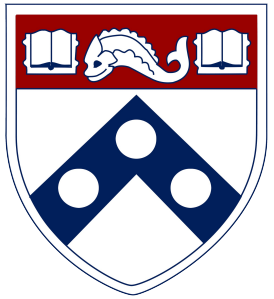
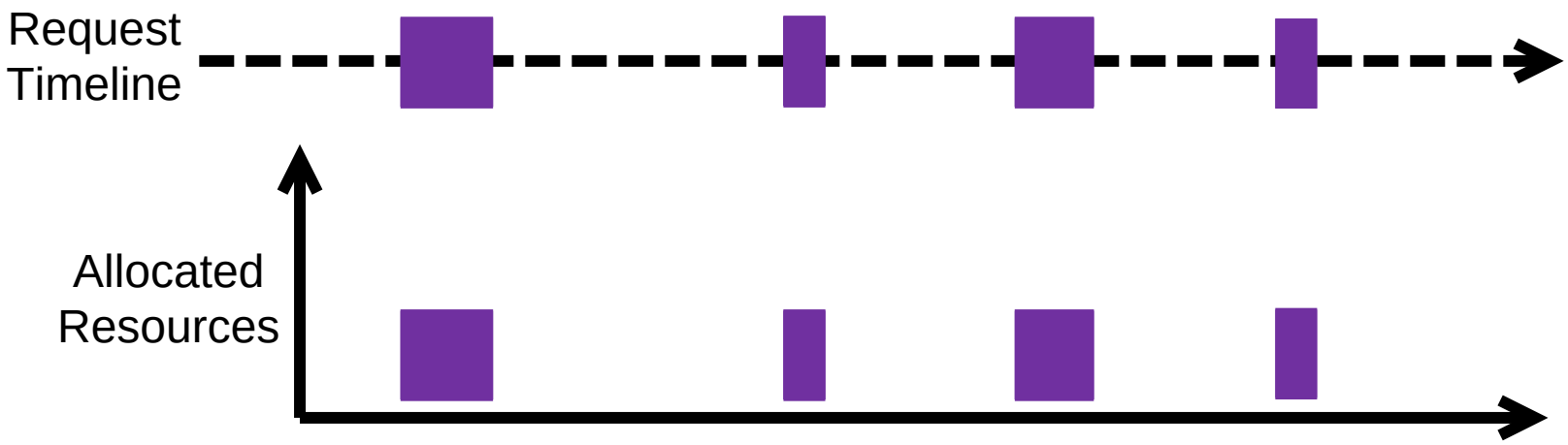
Rethinking Process Snapshots for Near-Warm Serverless Cold Starts

Ben Holmes, Baltasar Dinis, Lana Honcharuk, Adam Belay, and Joshua Fried

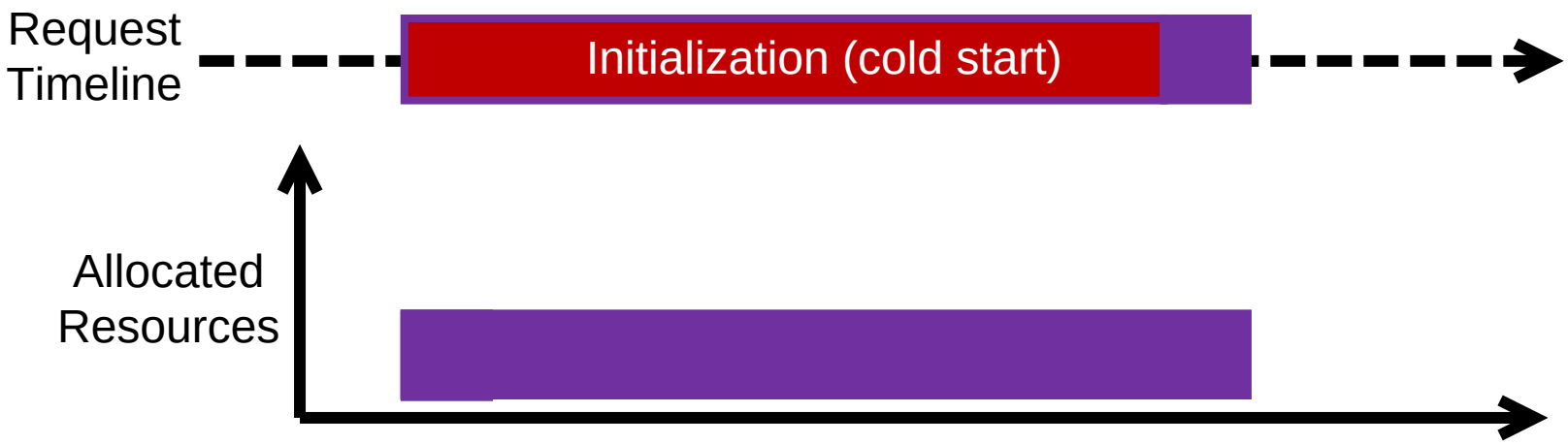




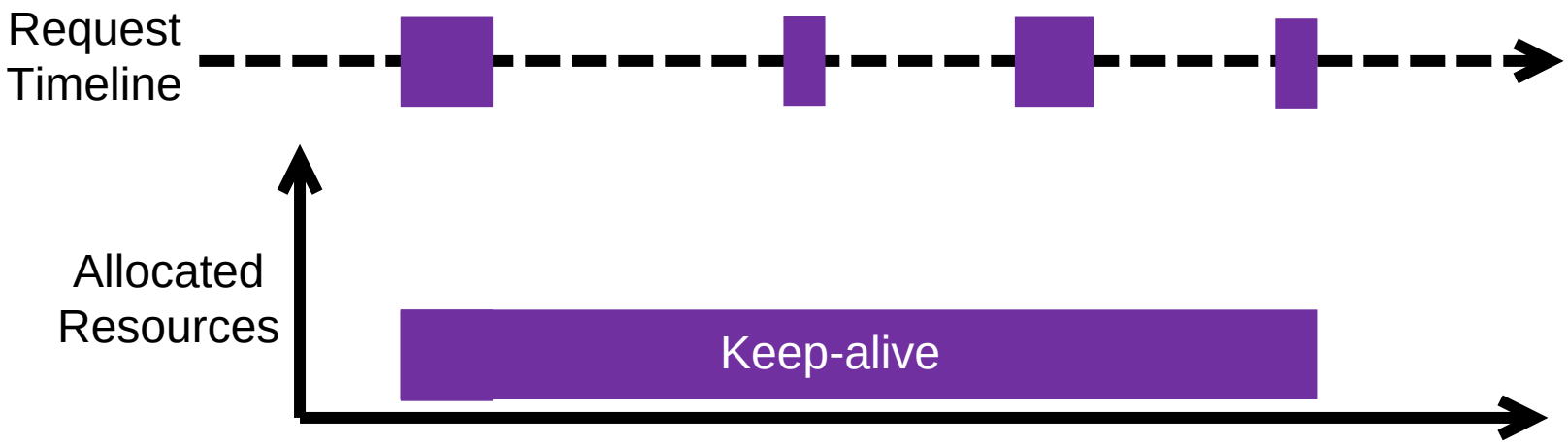
Serverless allocates resources on demand



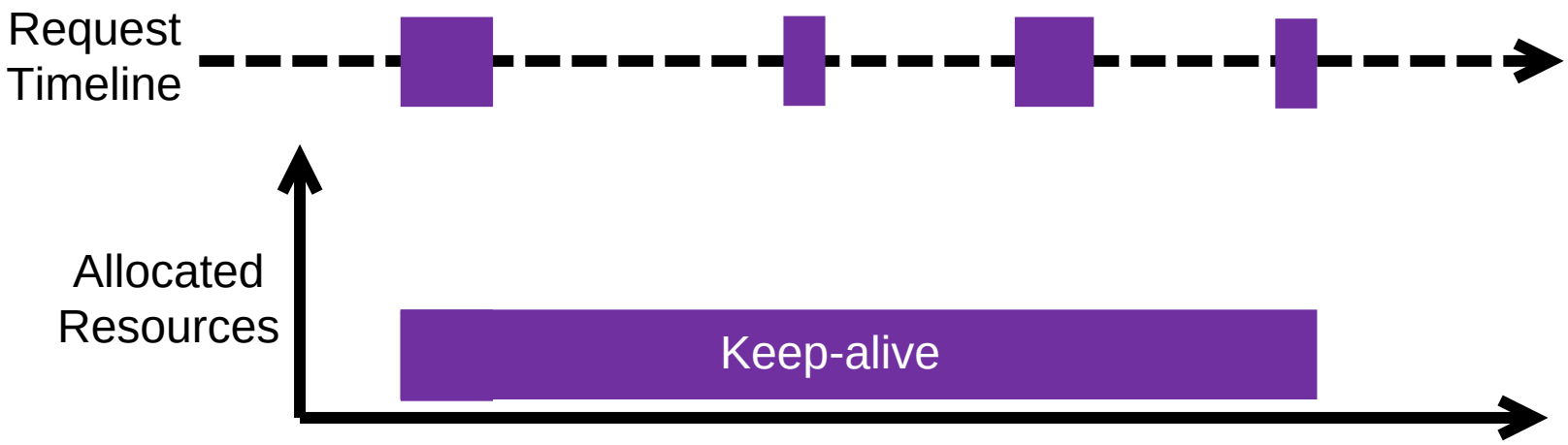
Starting applications from scratch is slow



Starting applications from scratch is slow

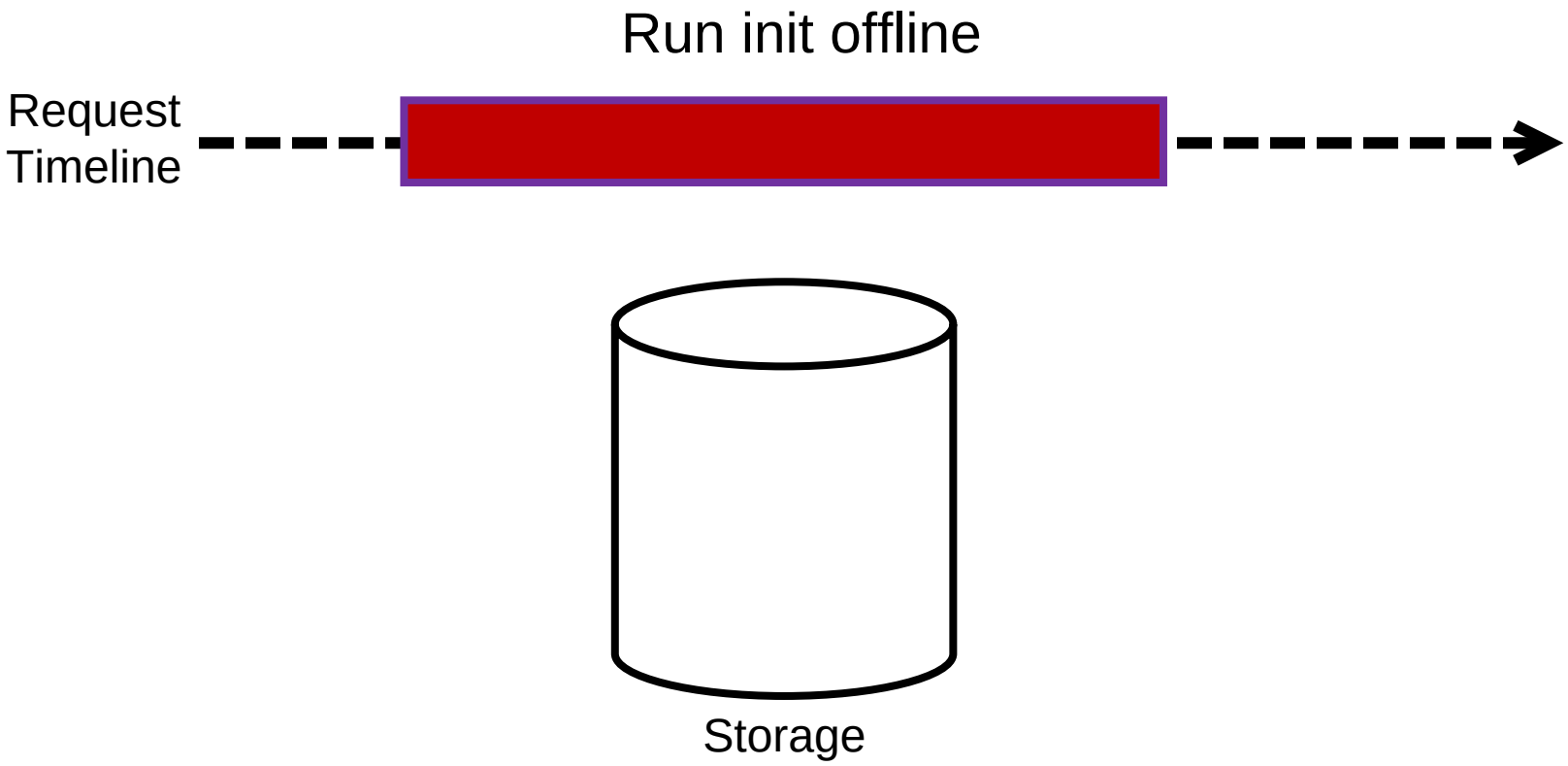


Starting applications from scratch is slow

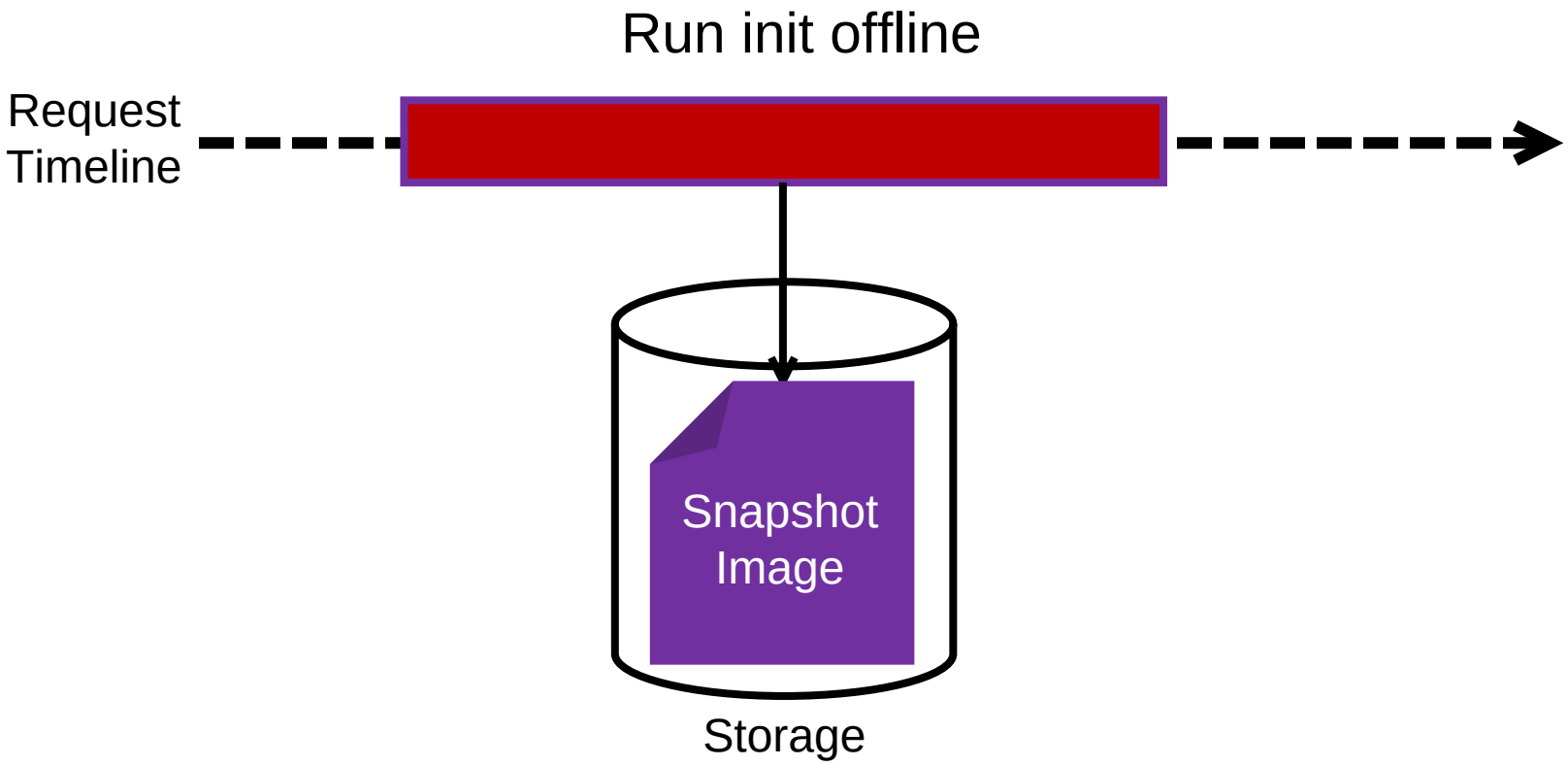


Reducing cold start time is imperative to reclaiming efficiency

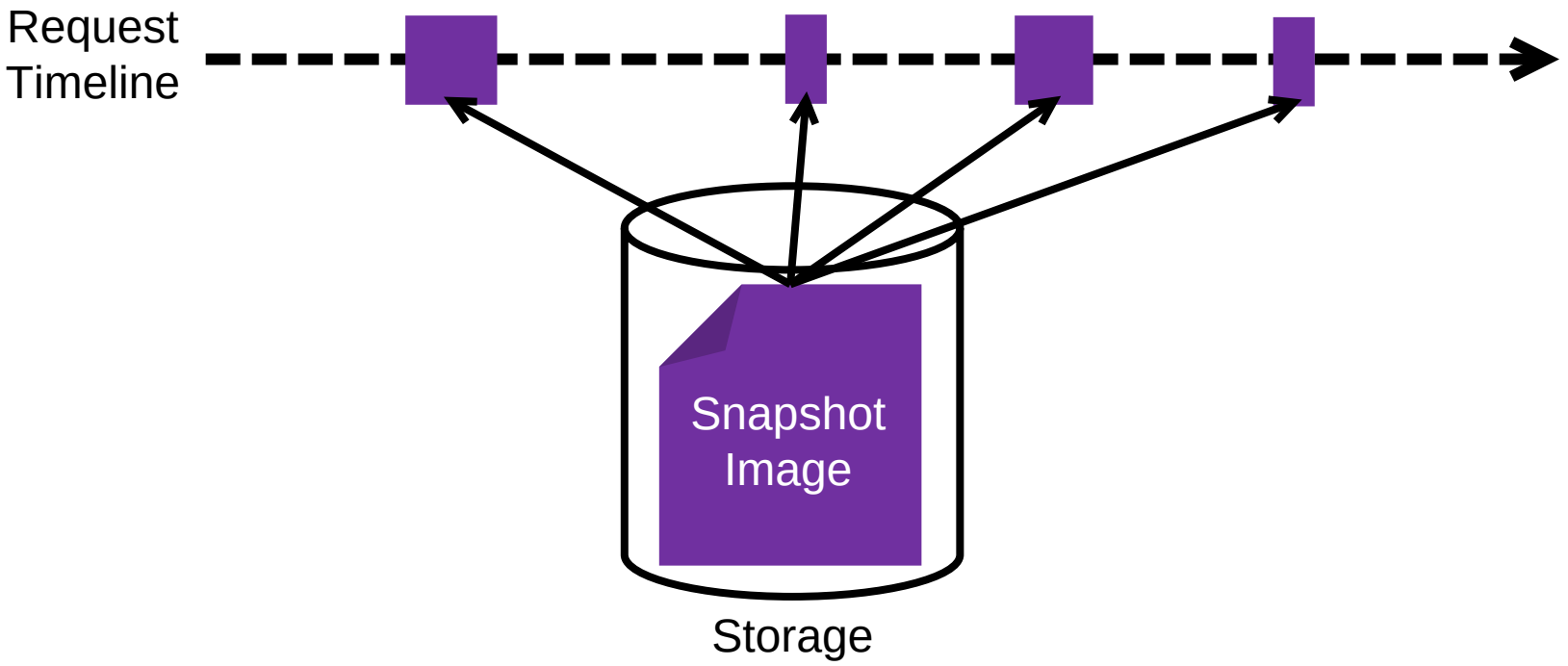
Snapshots bypass initialization



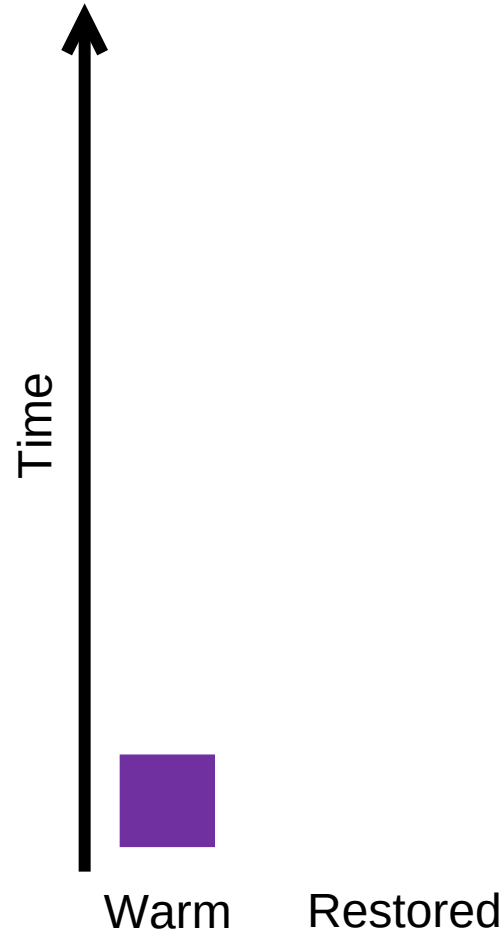
Snapshots bypass initialization



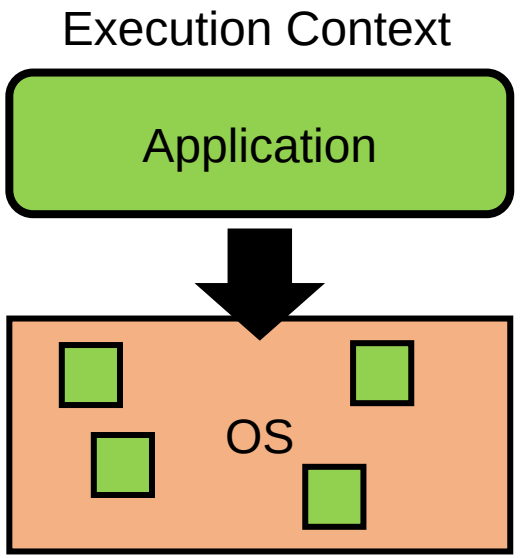
Snapshots bypass initialization

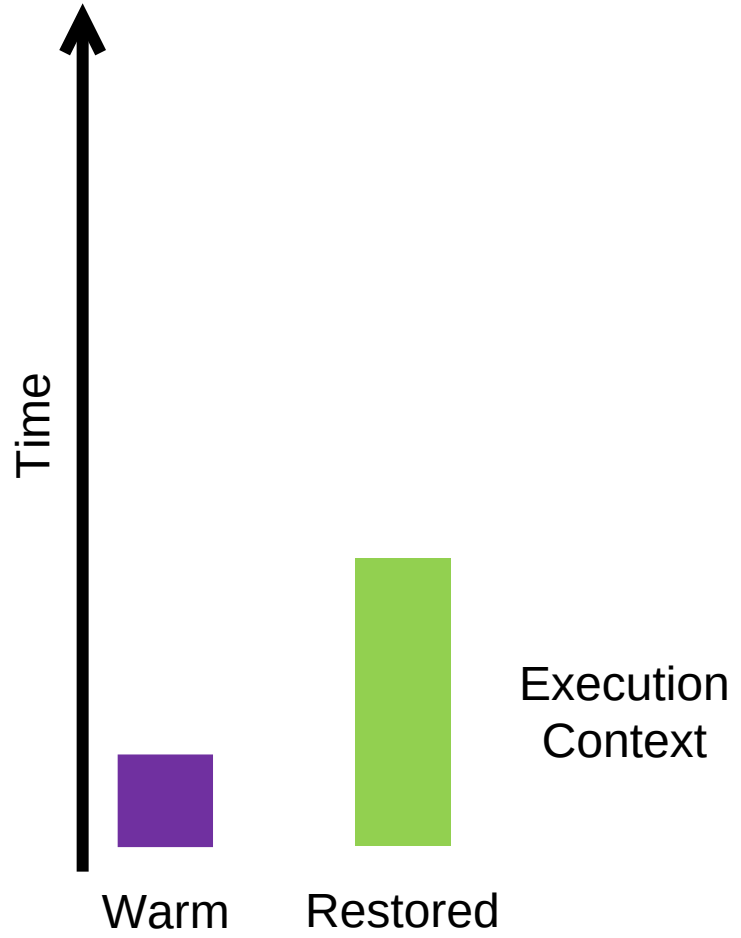


Startup costs are still high

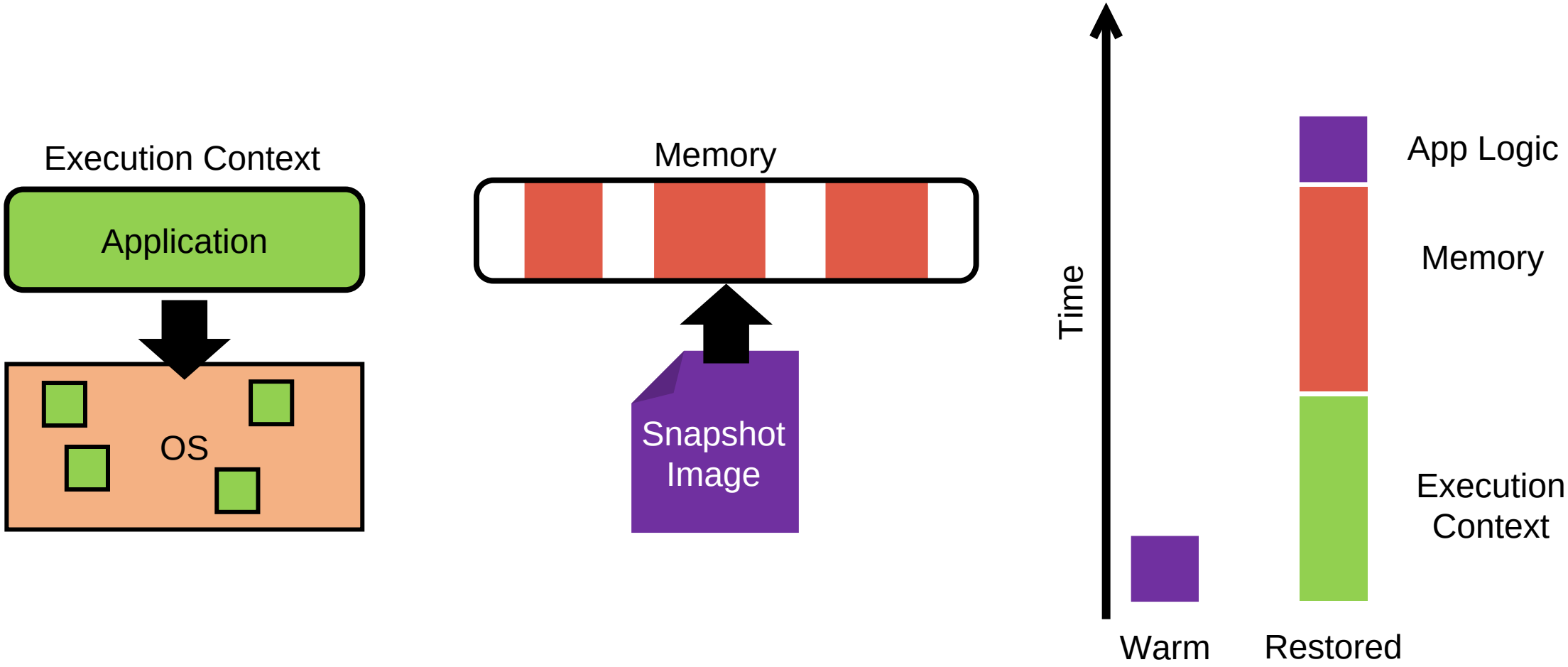


Startup costs are still high

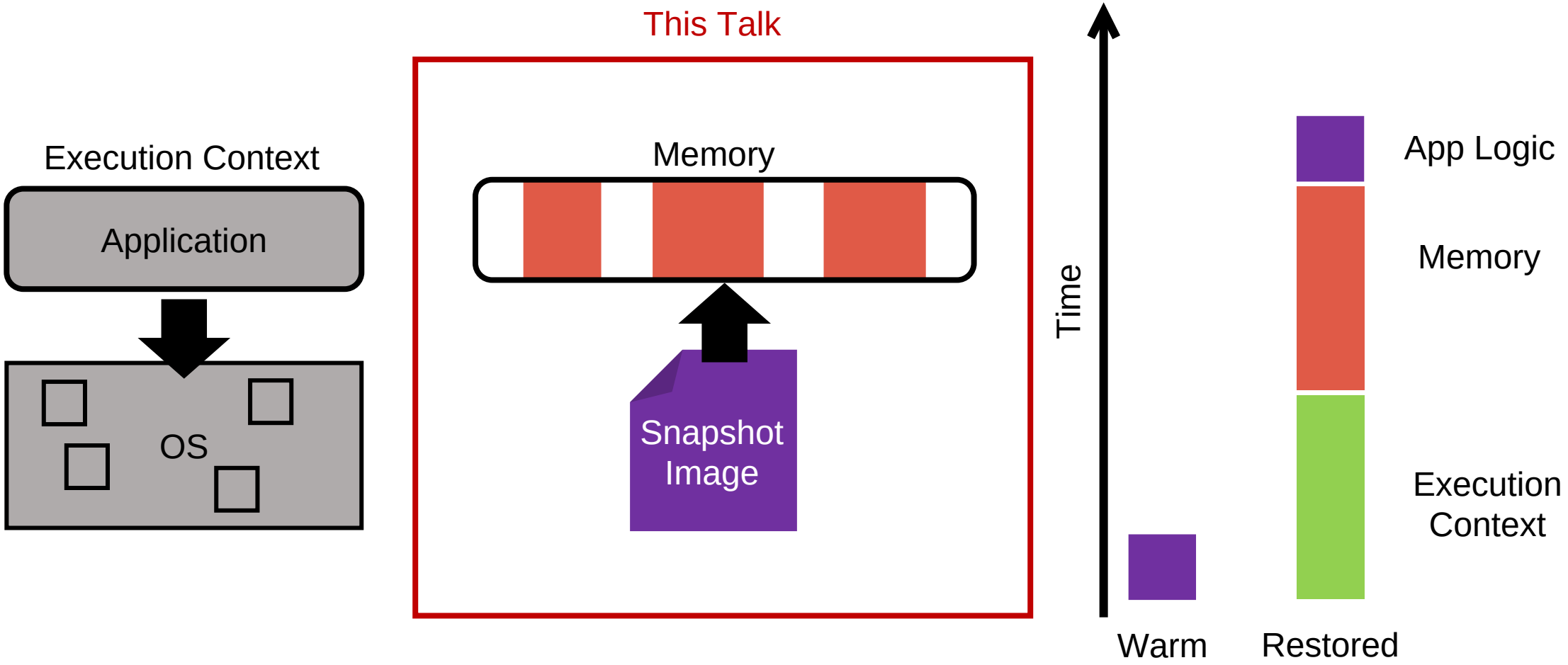




Startup costs are still high



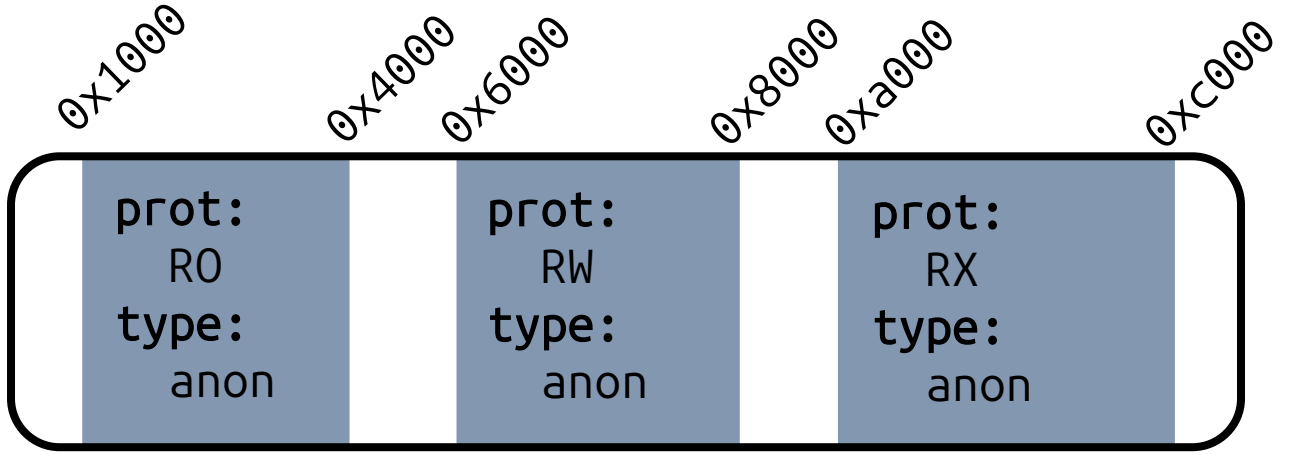
Startup costs are still high



Background: restoring memory

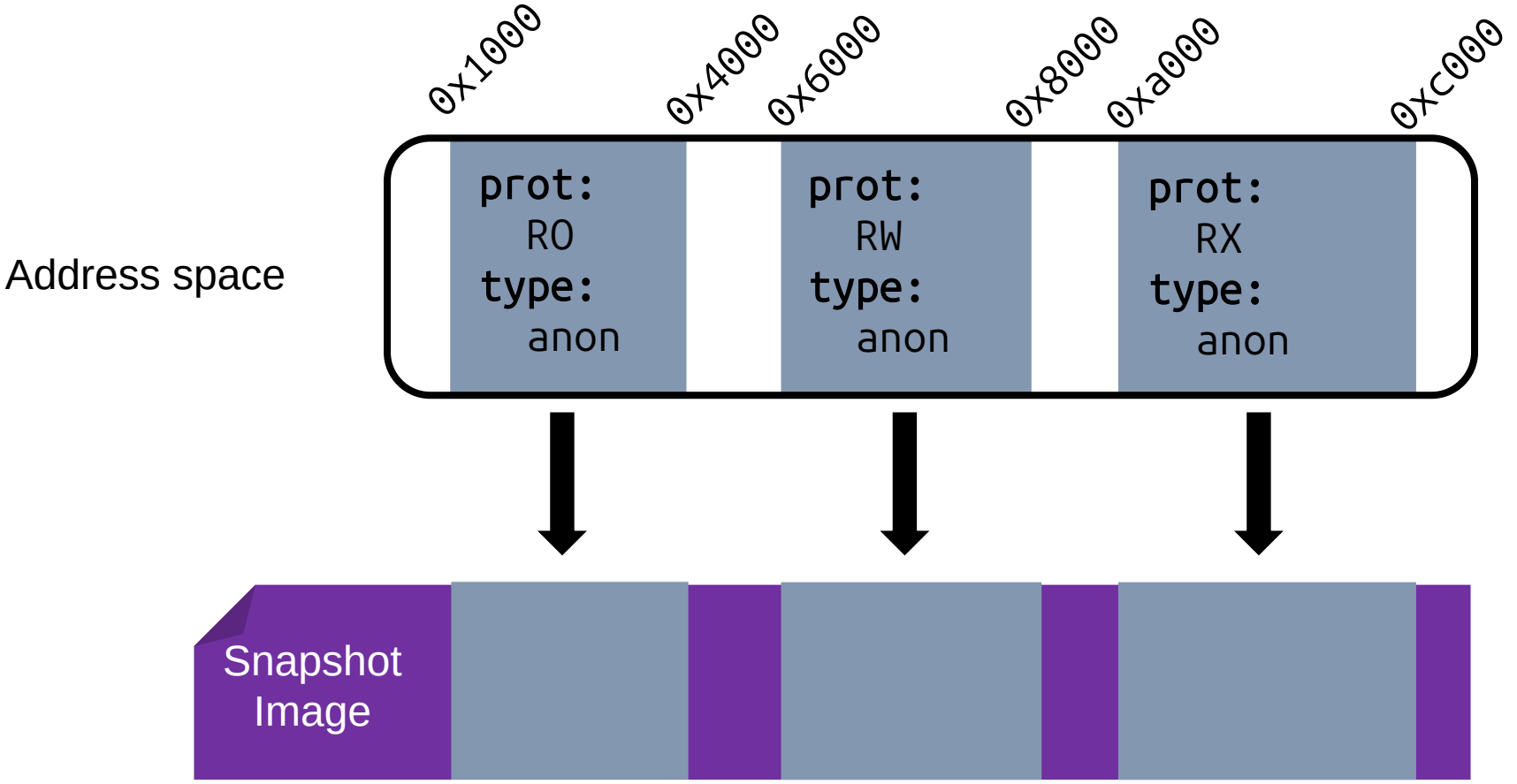
- Memory is organized by VMAs

Address space



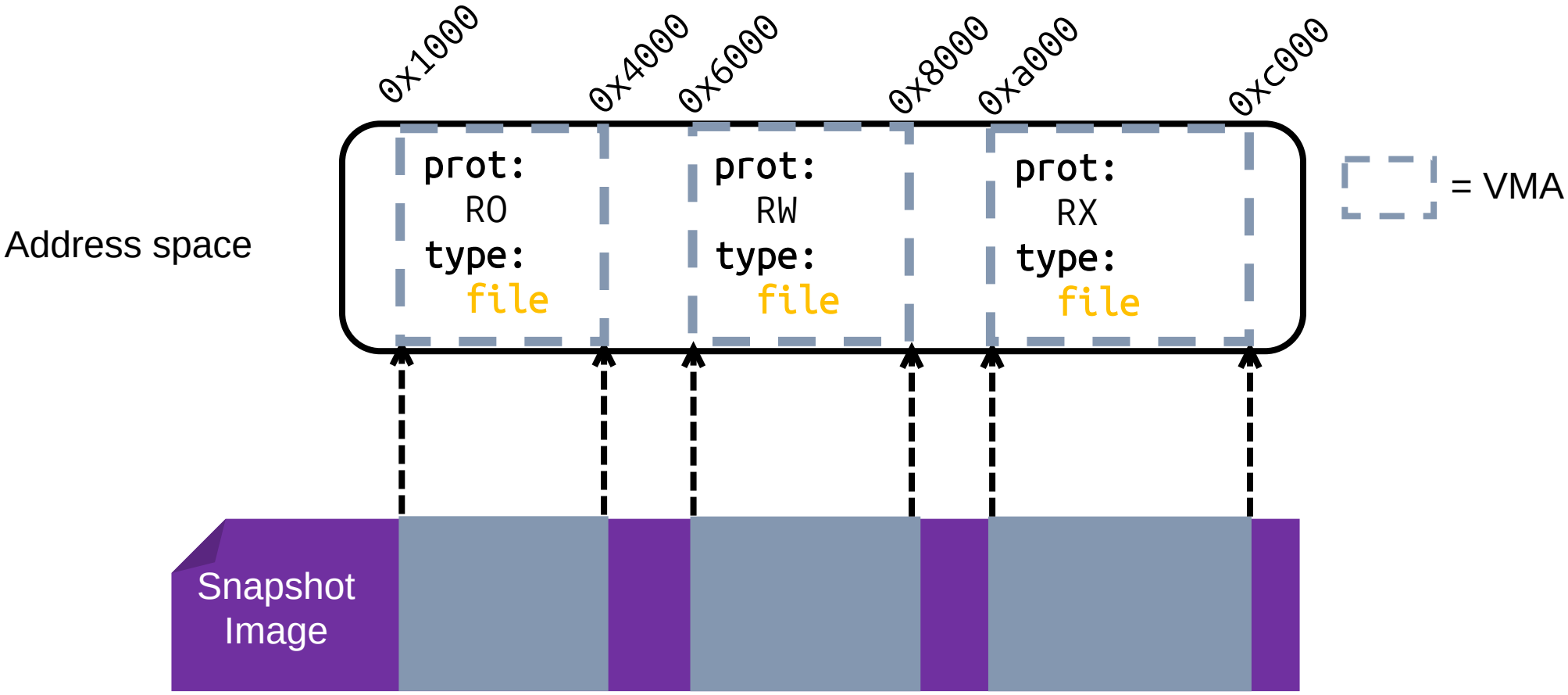
Background: restoring memory

- Dump VMAs as chunks to a file



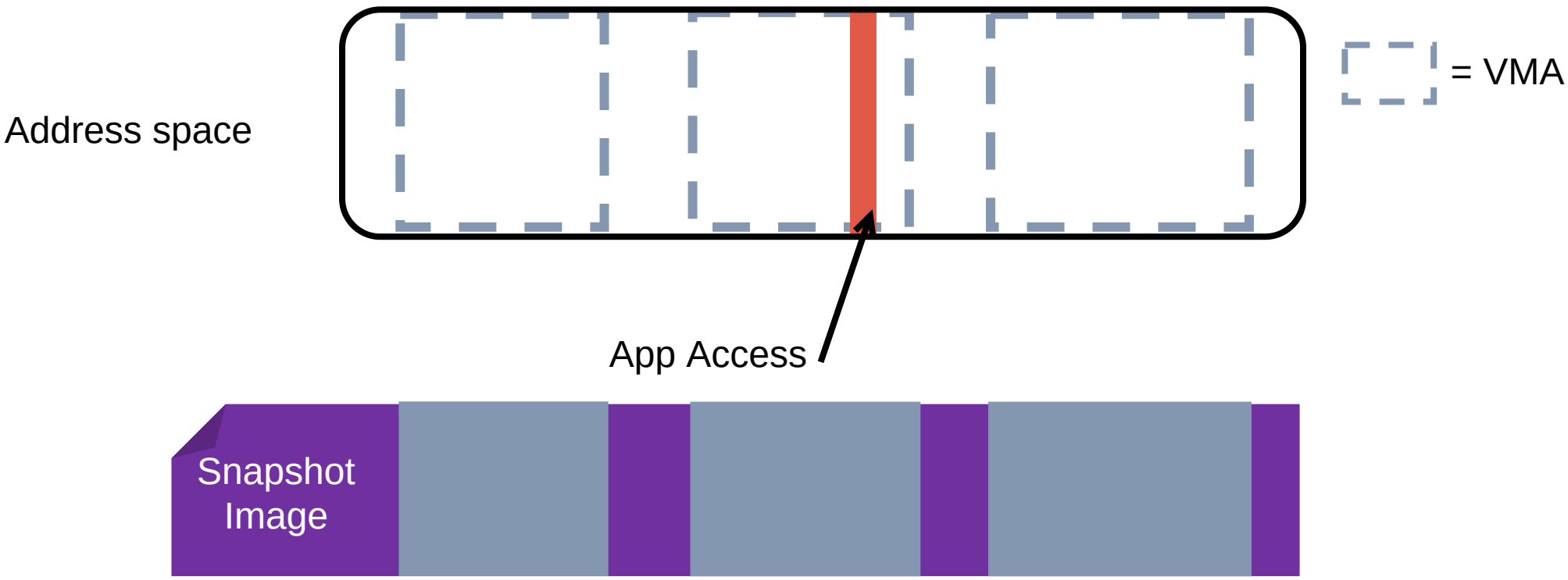
Background: restoring memory

- Map chunks from the file on restore



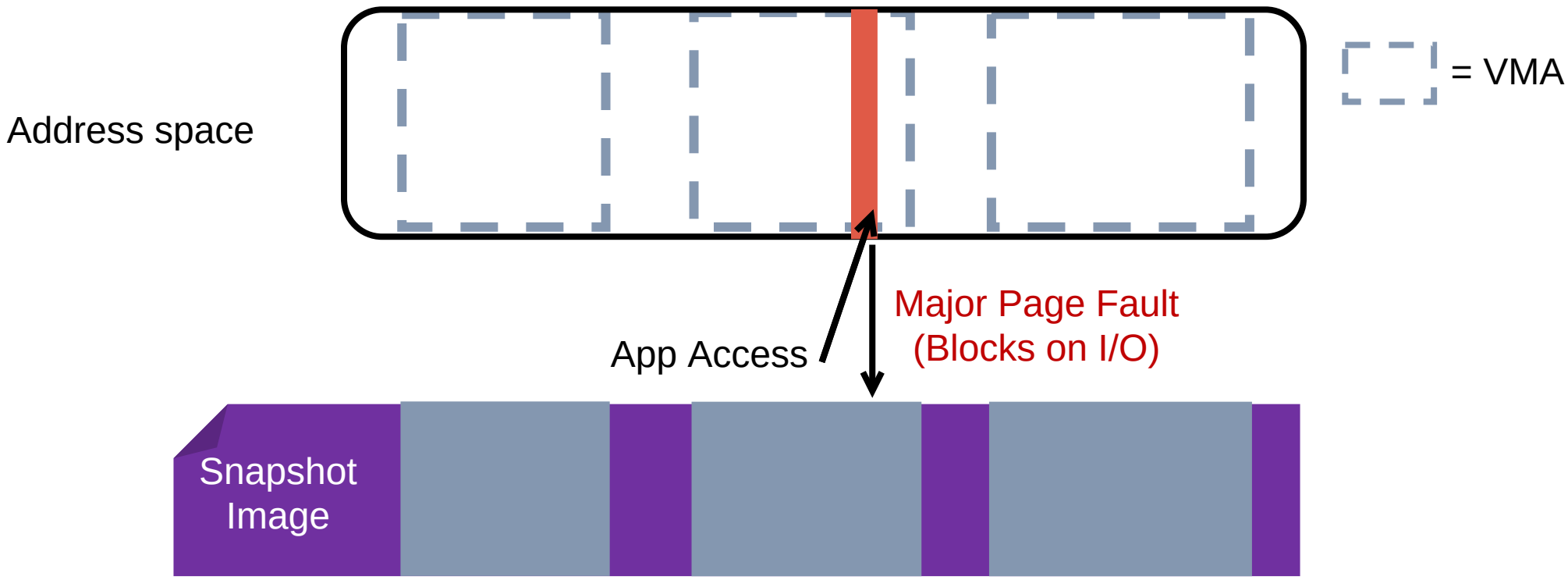
Background: restoring memory

- Still need the actual bytes from the file



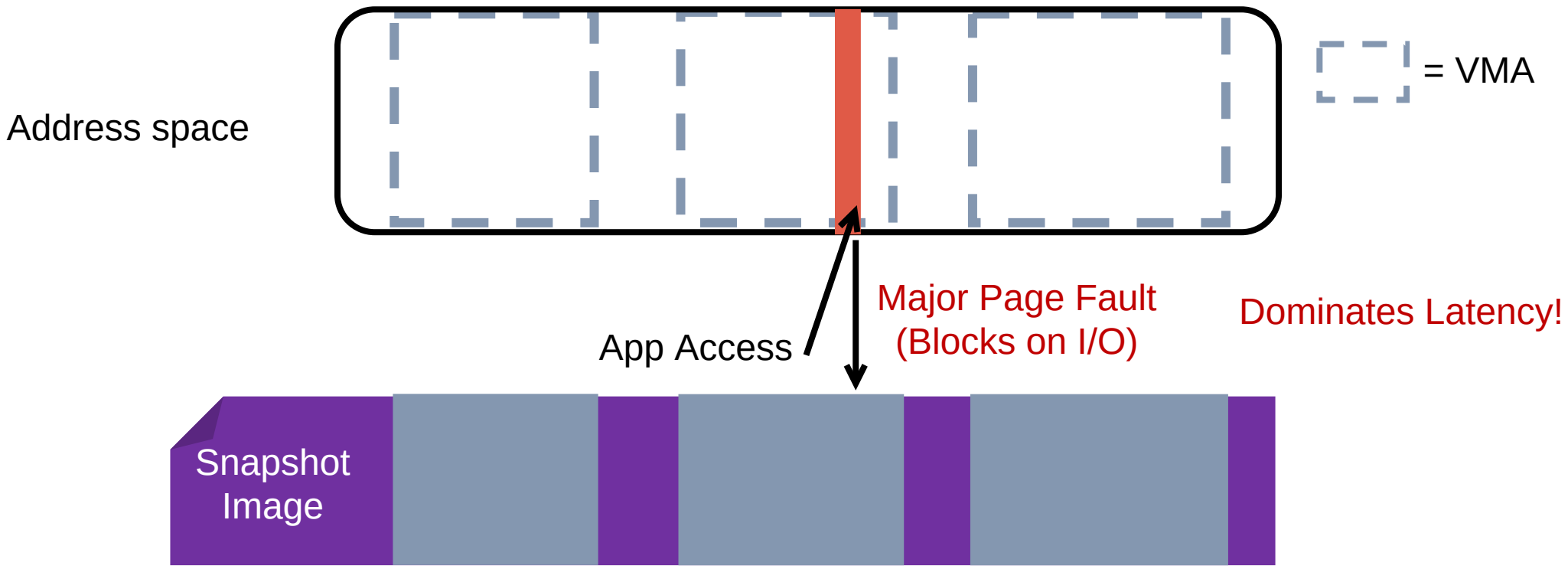
Background: restoring memory

- Still need the actual bytes from the file



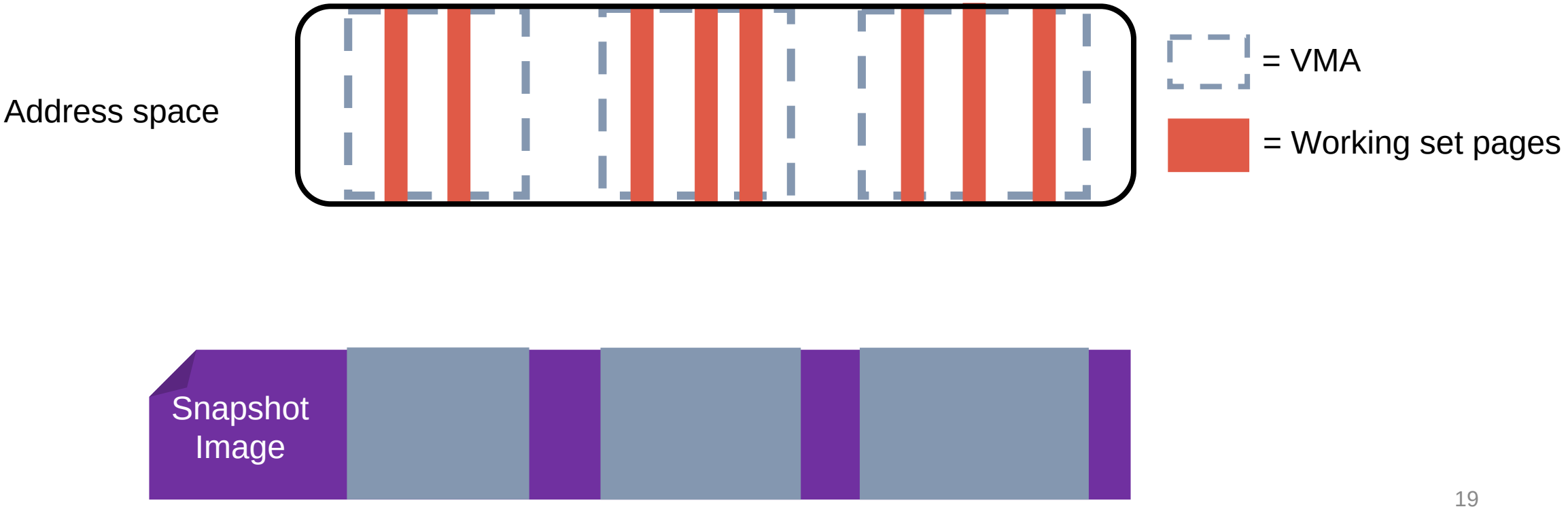
Background: restoring memory

- Still need the actual bytes from the file



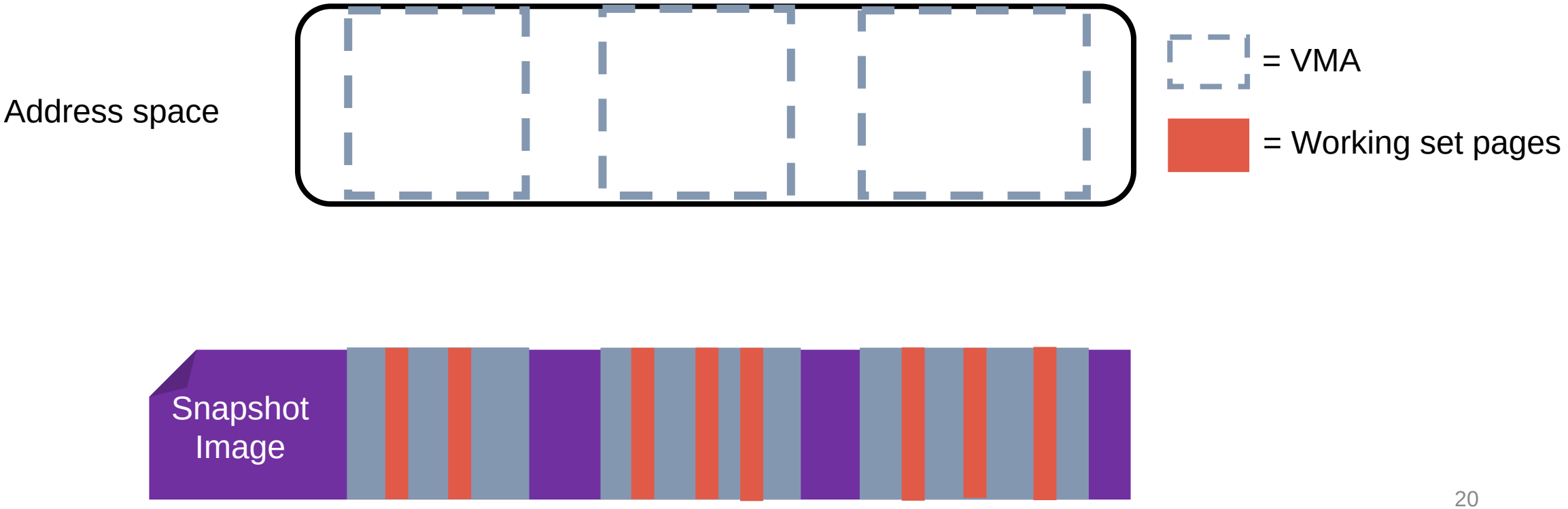
Background: restoring memory

- Working set is a predictable subset



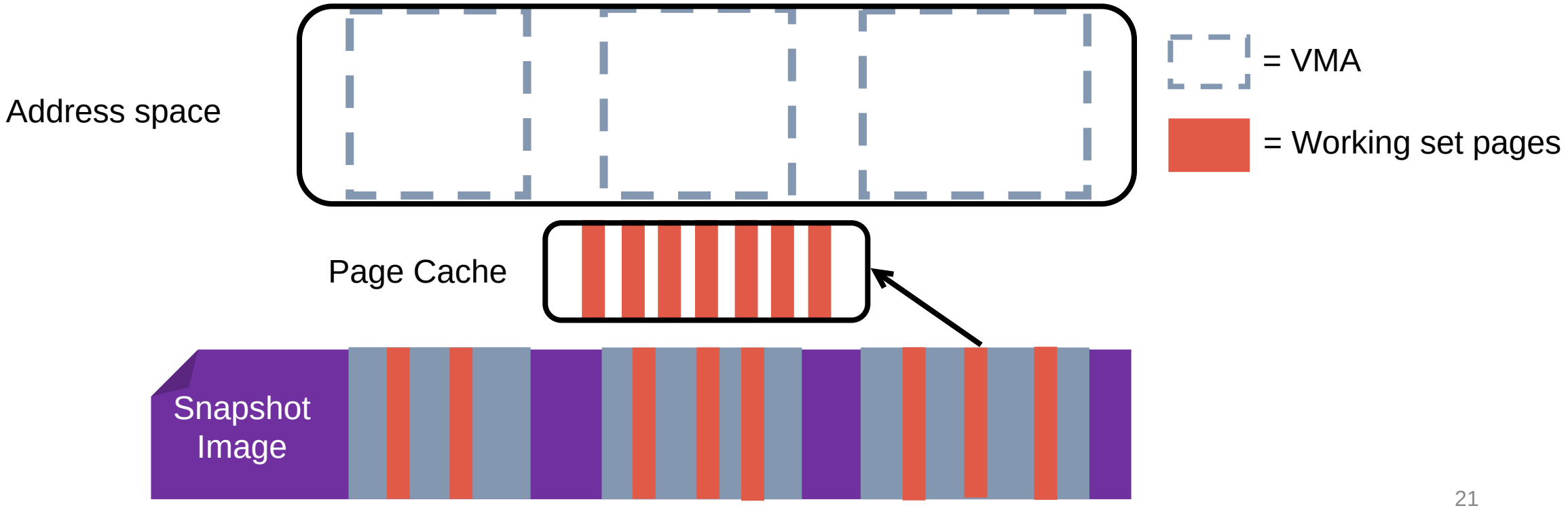
Background: restoring memory

- Record accessed pages



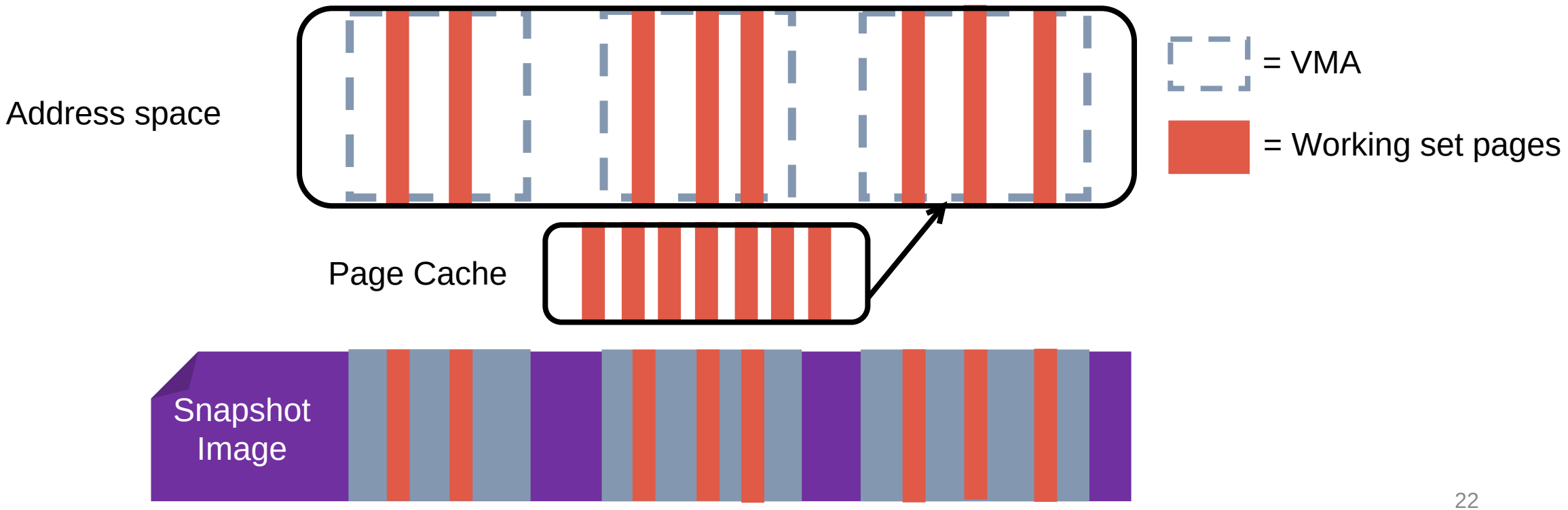
Background: restoring memory

- **Prefetch** the working set

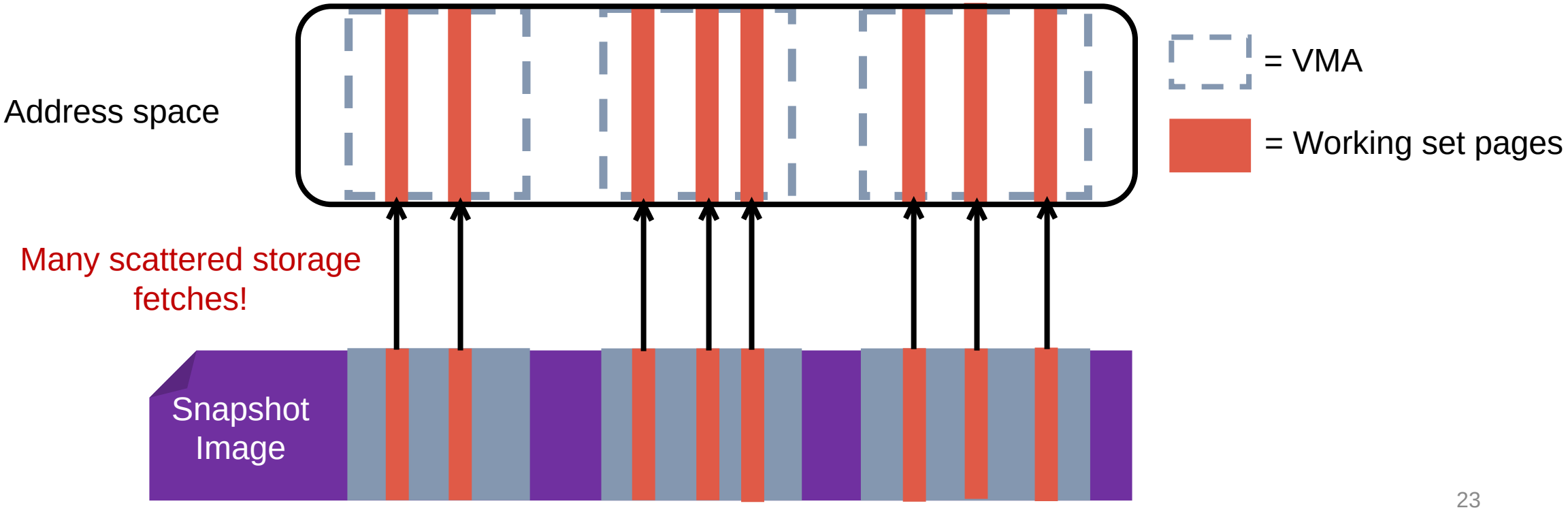


Background: restoring memory

- **Prefetch** the working set

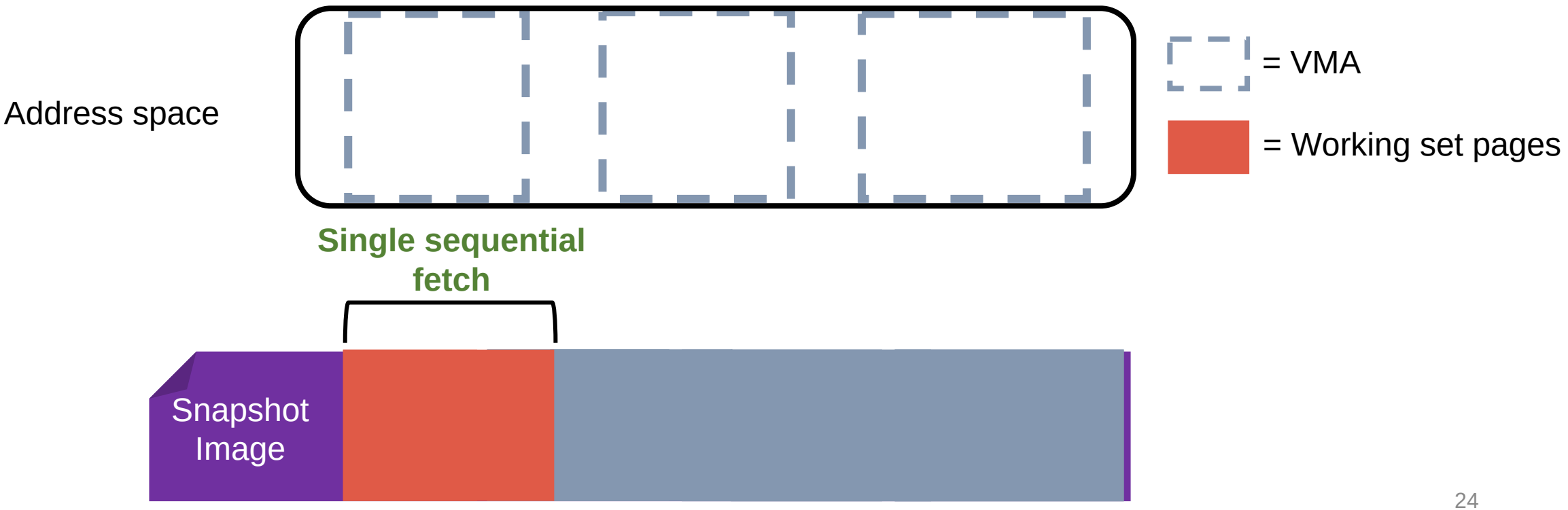


Problem #1: file layout



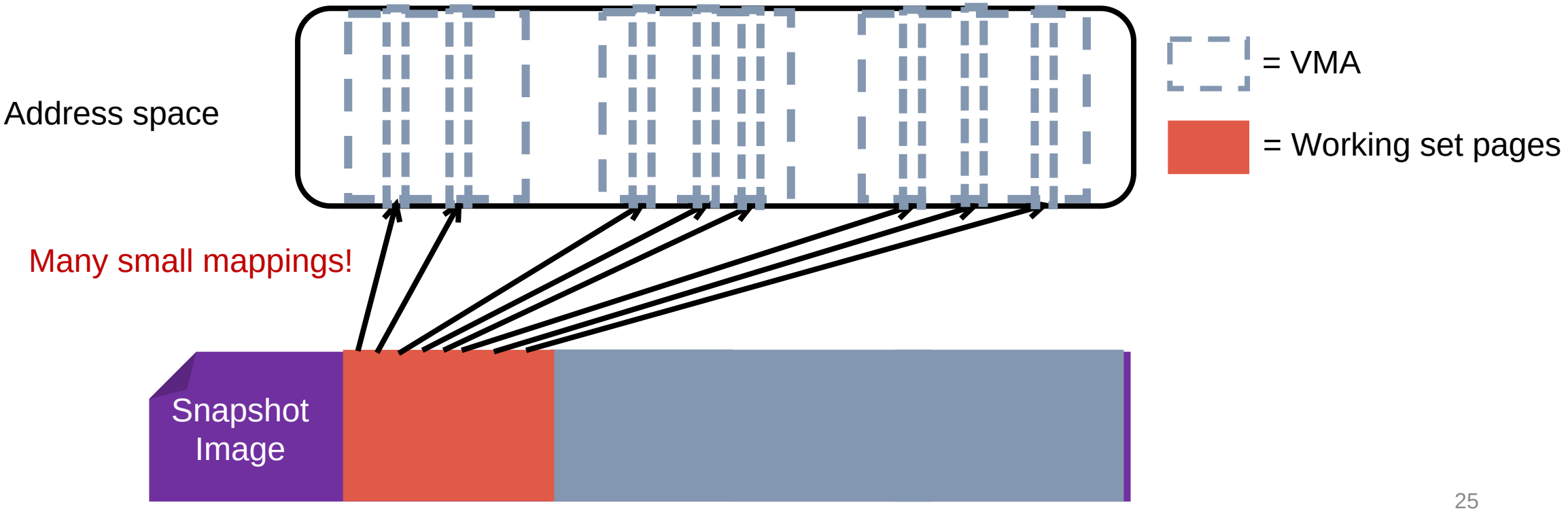
Problem #1: file layout

- **Group the working set together**



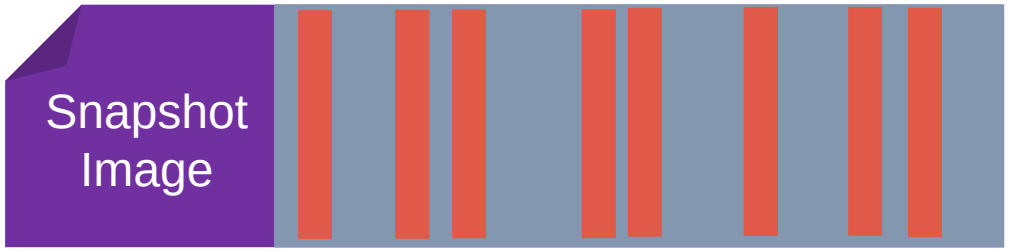
Problem #1: file layout

- **Group the working set together**



No way to resolve the tension

one mapping
many fetches

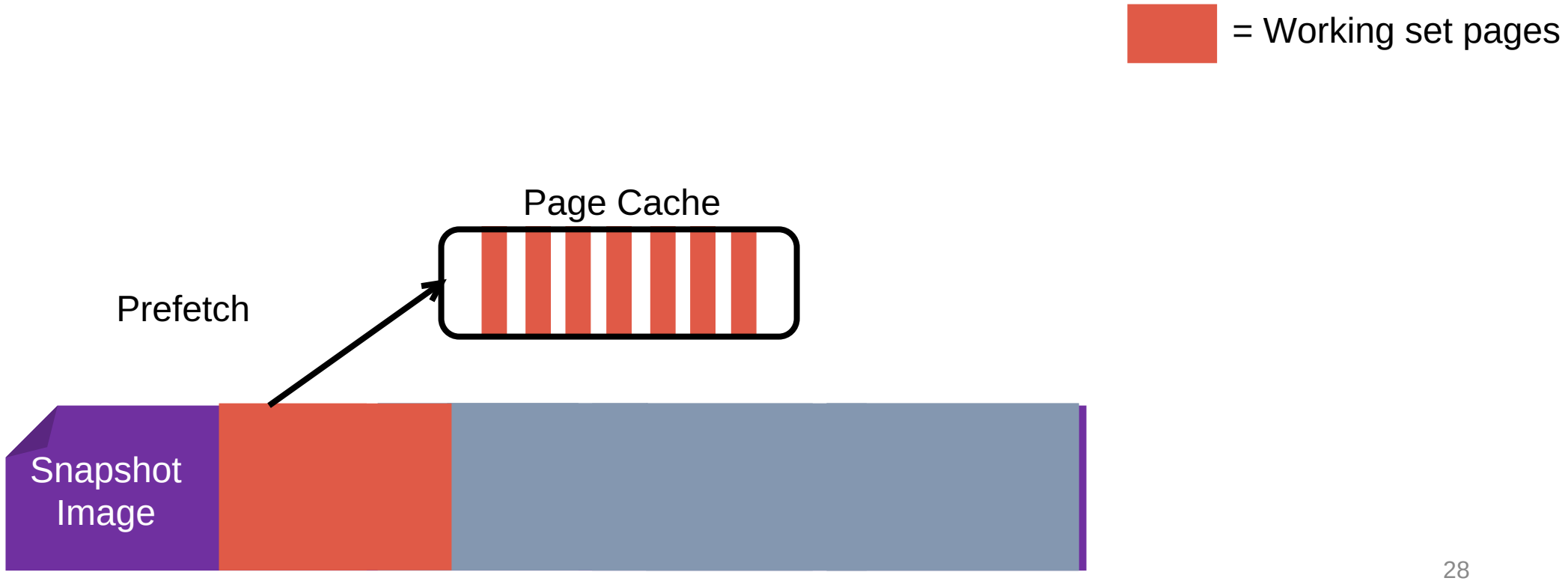


one fetch
many mappings

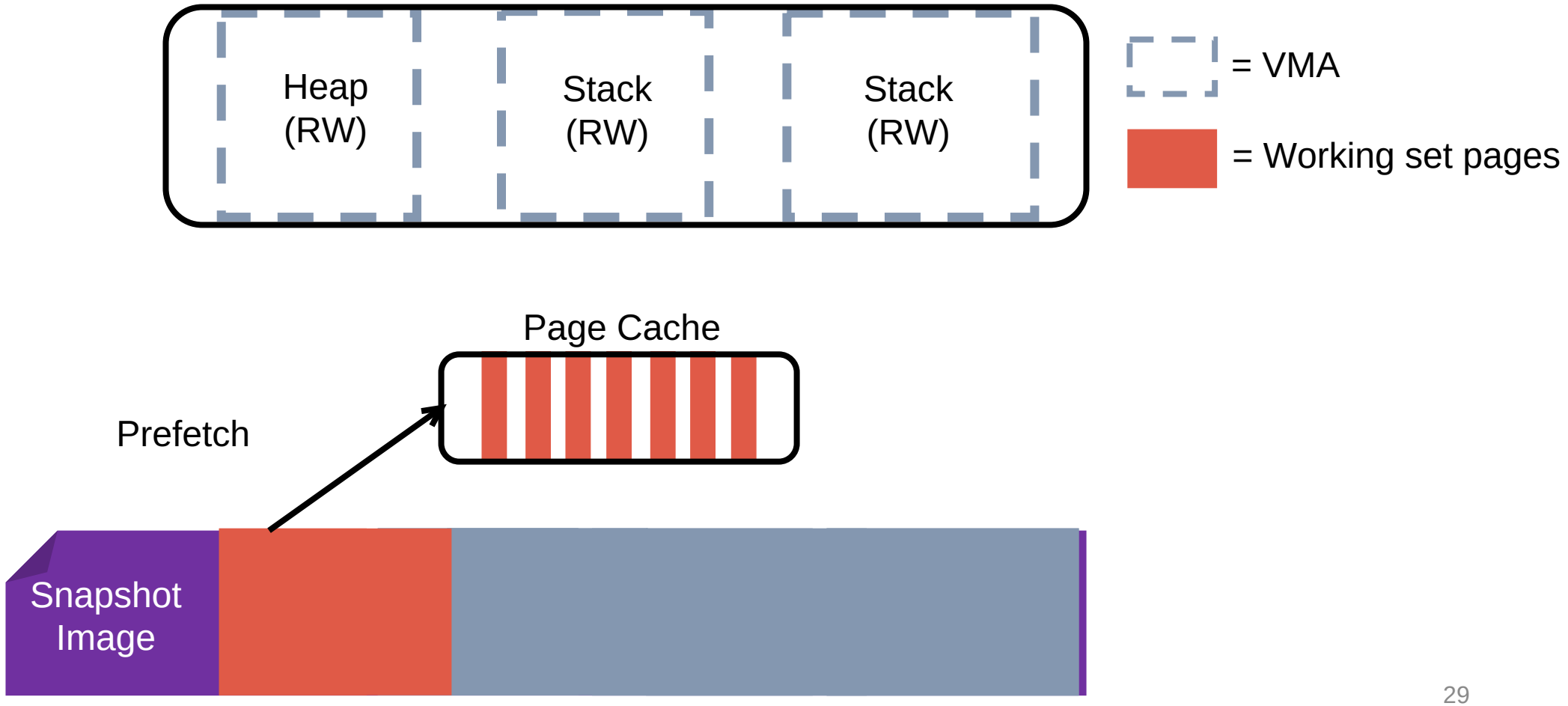


Problem #2: Unnecessary CoW

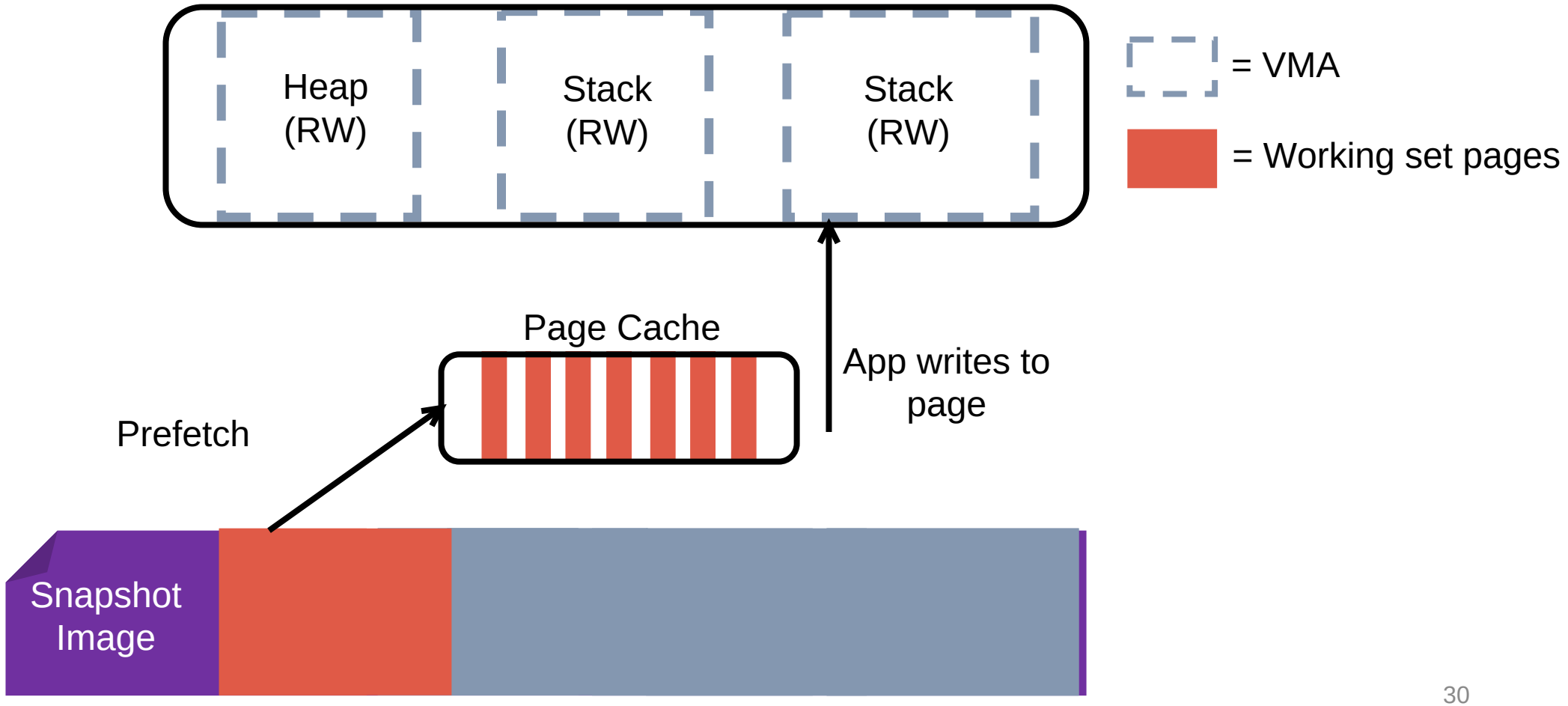
Problem #2: Unnecessary CoW



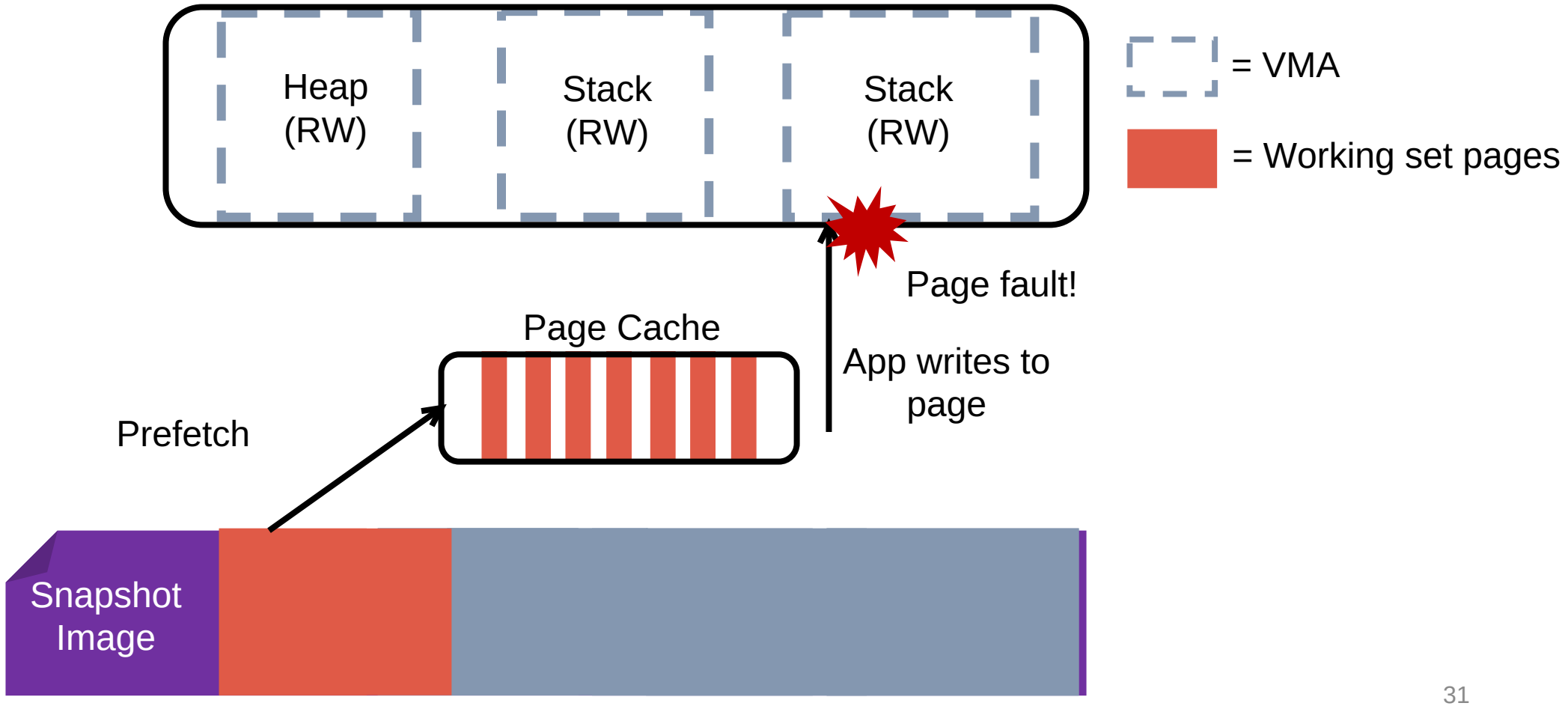
Problem #2: Unnecessary CoW



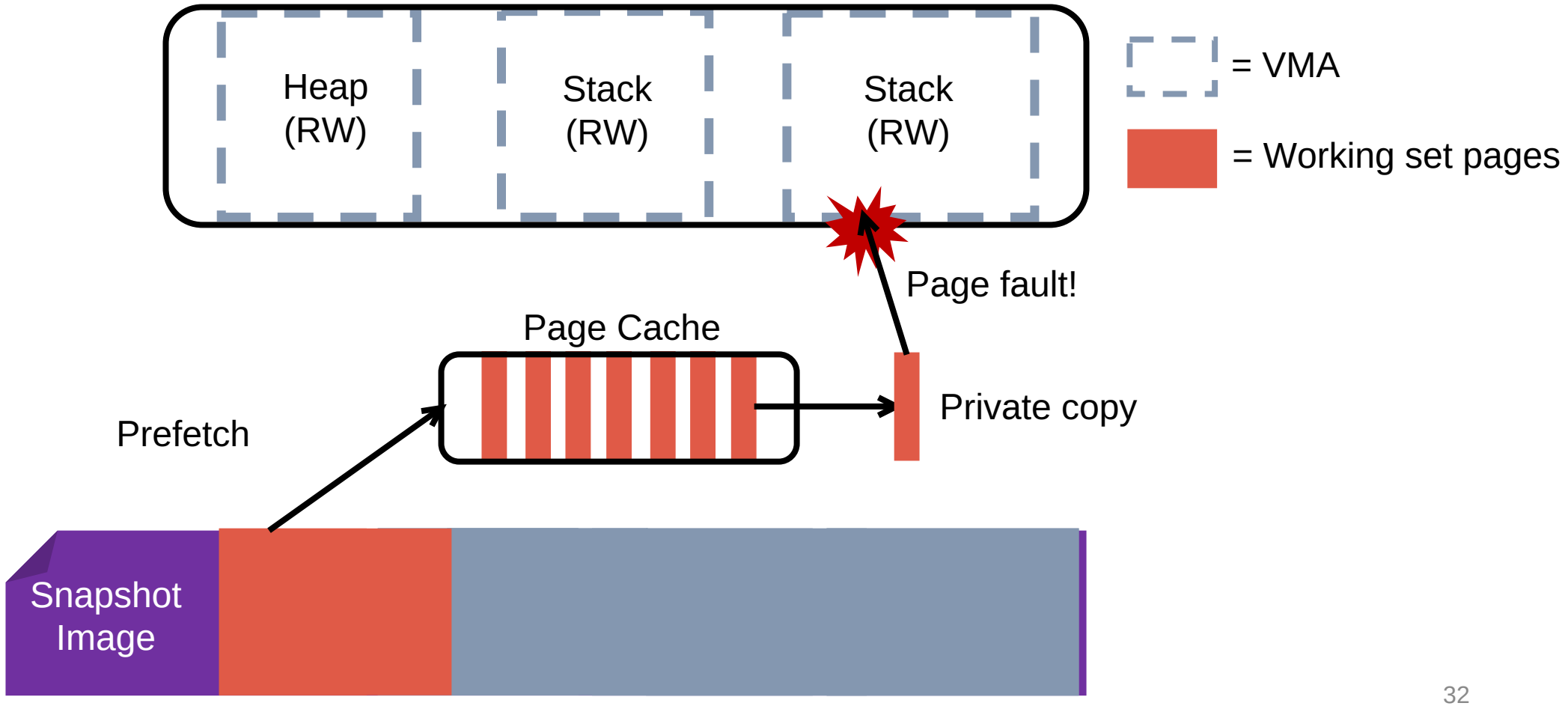
Problem #2: Unnecessary CoW



Problem #2: Unnecessary CoW



Problem #2: Unnecessary CoW



File-backed mappings are a poor fit

File-backed mappings are a poor fit

Starting a process from a file



Simple linear mappings



Useful to share code/data pages

File-backed mappings are a poor fit

Starting a process from a file



Simple linear mappings



Useful to share code/data pages

Restoring a process from a file



Address space layout != file layout



Pages are not shared, likely written to

File-backed mappings are a poor fit

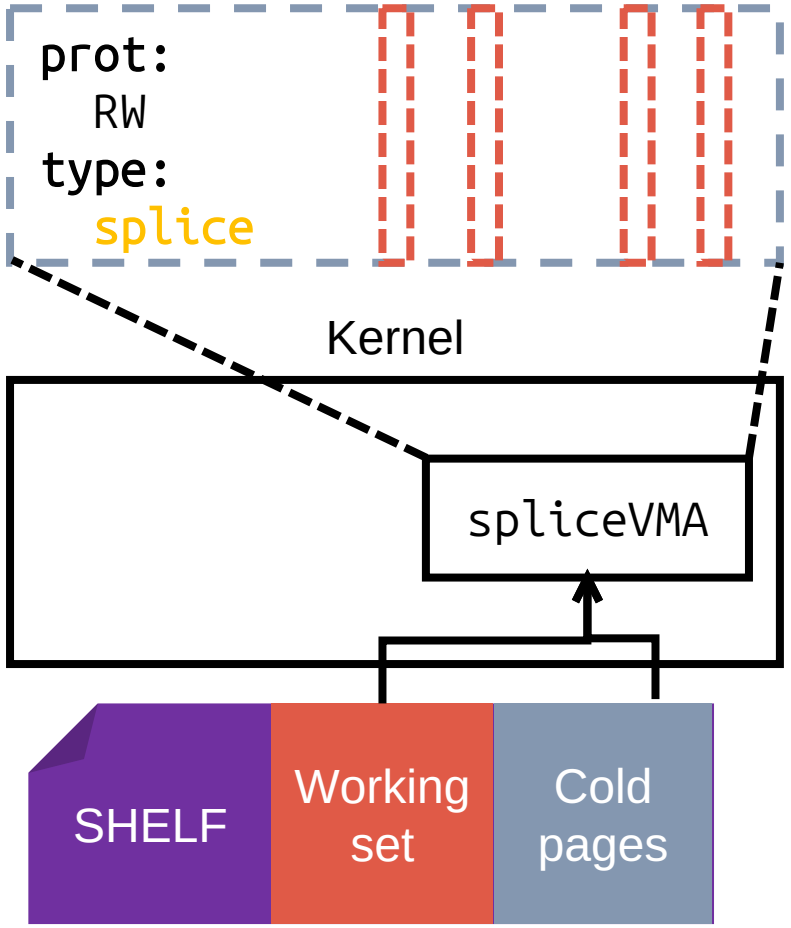
Starting a process from a file	**Restoring a process from a file**
 Simple linear mappings	 Address space layout != file layout
 Useful to share code/data pages	 Pages are not shared, likely written to

We need first-class OS support for restoring an address space from a file

Spice: first-class OS support for restore

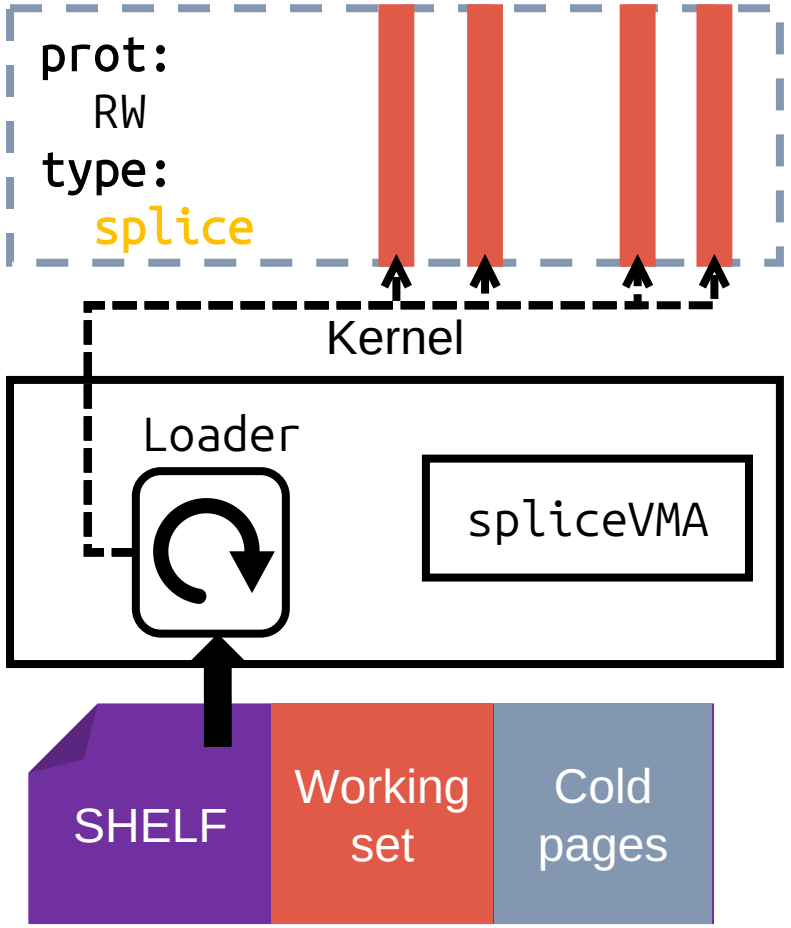
Spice: first-class OS support for restore

- One-shot address space restore
  - spliceVMA: new mapping type
  - Snapshot Hybrid ELF (SHELF): object format mapped by spliceVMAs



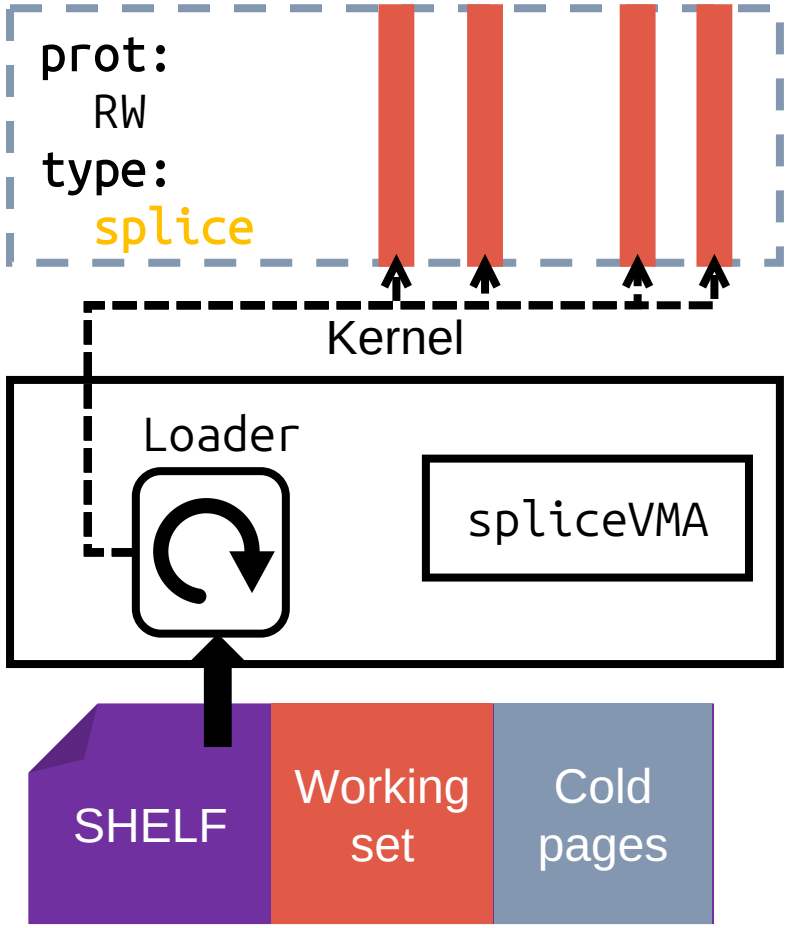
Spice: first-class OS support for restore

- One-shot address space restore
  - spliceVMA: new mapping type
  - Snapshot Hybrid ELF (SHELF): object format mapped by spliceVMAs
- Object format loader in the kernel
  - Handles prefetching and pre-filling page table entries



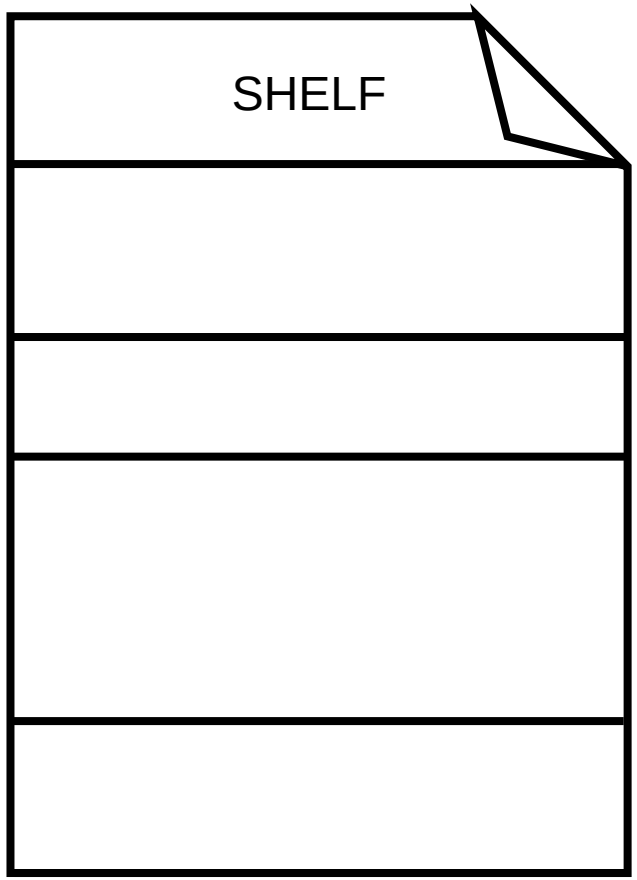
Spice: first-class OS support for restore

- One-shot address space restore
  - spliceVMA: new mapping type
  - Snapshot Hybrid ELF (SHELF): object format mapped by spliceVMAs
- Object format loader in the kernel
  - Handles prefetching and pre-filling page table entries
- Cold start with near-warm latency
  - Reduces cold start latency by 14-96%



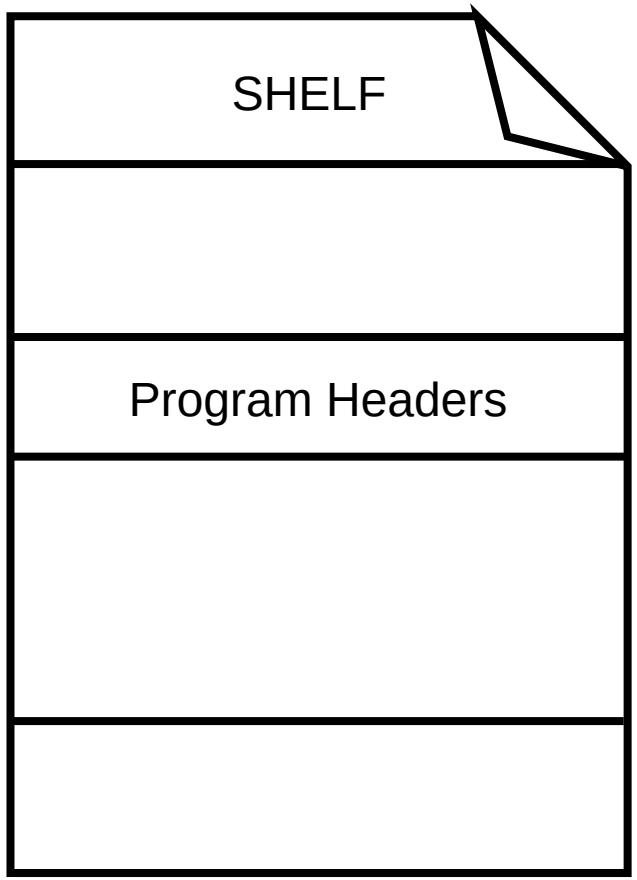
Snapshot Hybrid ELF (SHELF)

- Inspired by ELF; format for bootstrapping an address space from a file



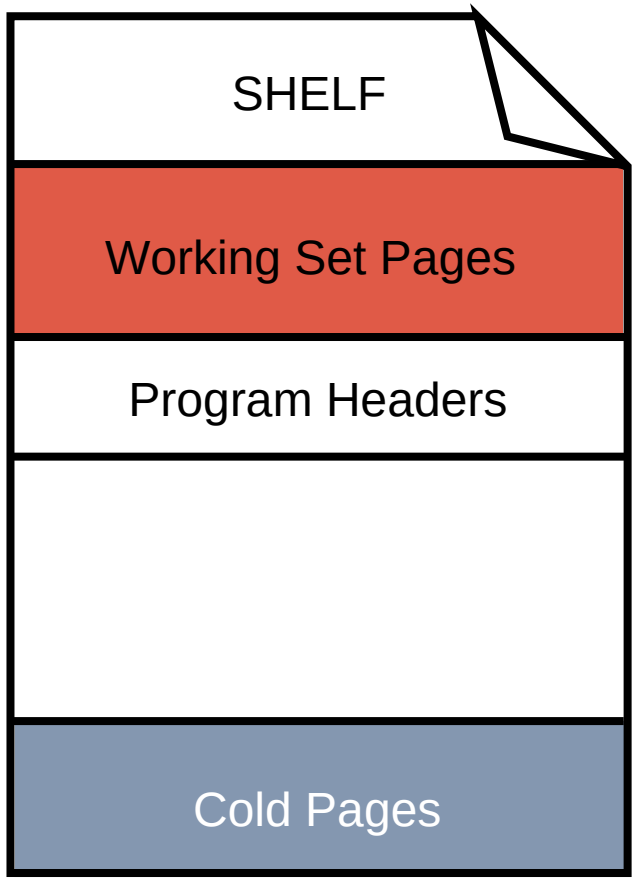
Snapshot Hybrid ELF (SHELF)

- Inspired by ELF; format for bootstrapping an address space from a file
- Program headers describe ranges to map



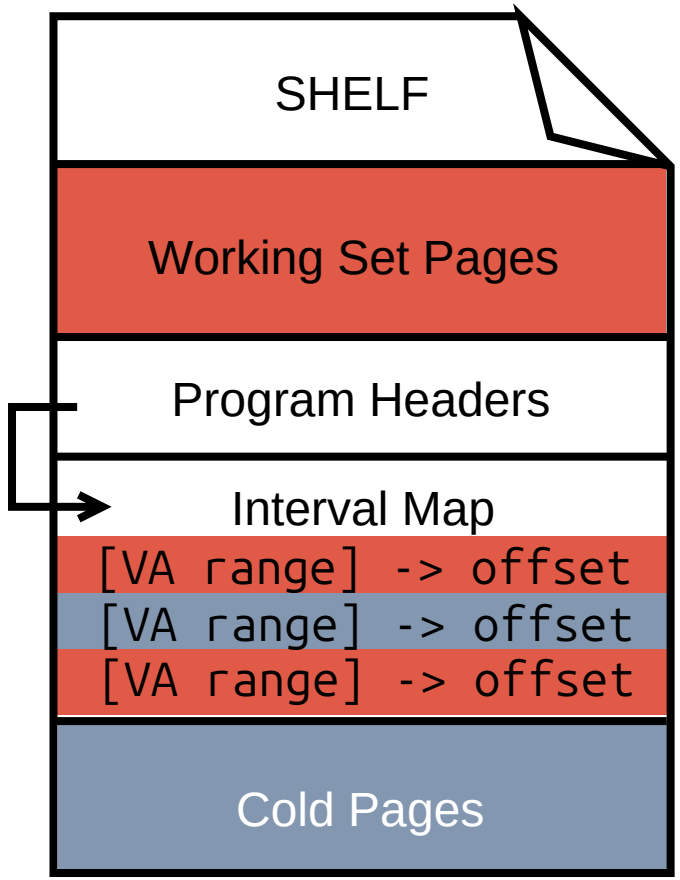
Snapshot Hybrid ELF (SHELF)

- Inspired by ELF; format for bootstrapping an address space from a file
- Program headers describe ranges to map
- Separate hot pages from cold



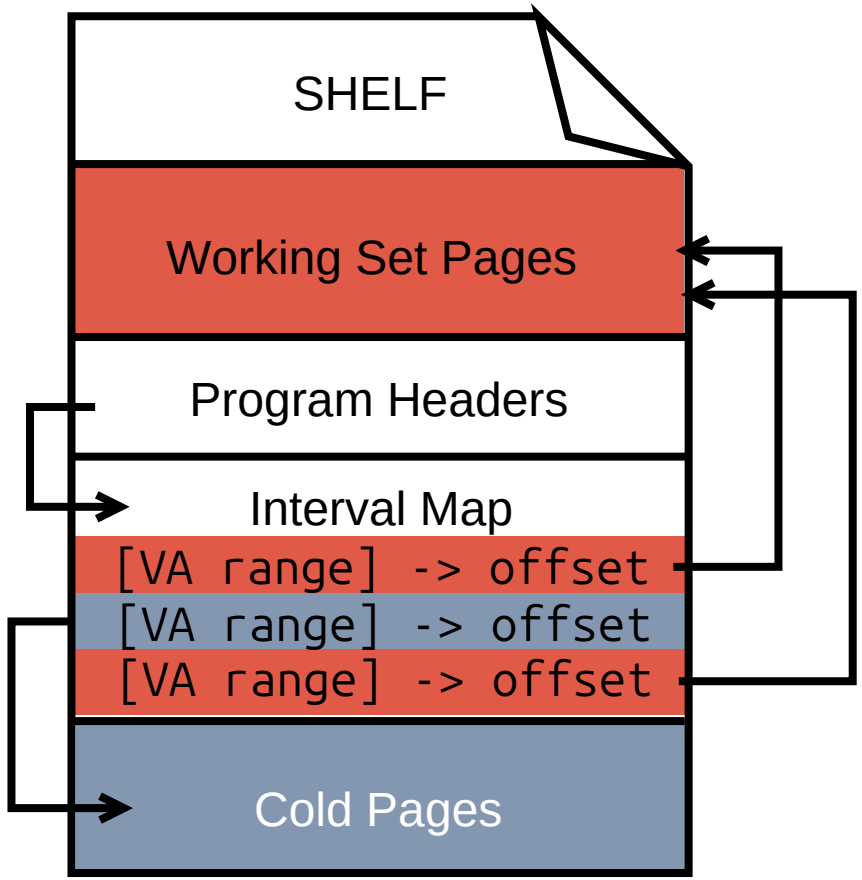
Snapshot Hybrid ELF (SHELF)

- Inspired by ELF; format for bootstrapping an address space from a file
- Program headers describe ranges to map
- Separate hot pages from cold
- Interval map manages indirection

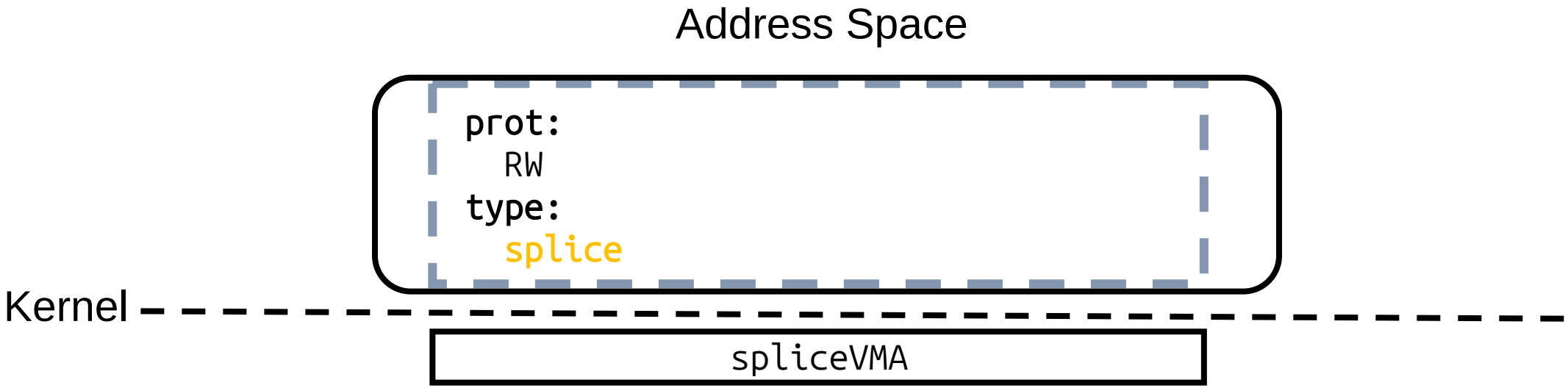


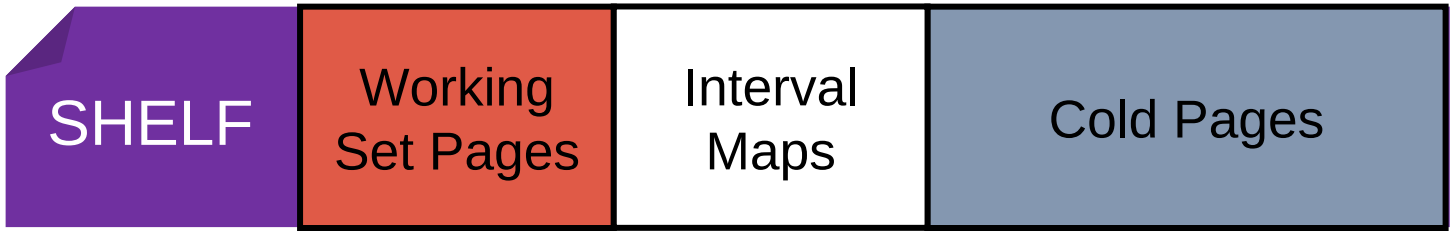
Snapshot Hybrid ELF (SHELF)

- Inspired by ELF; format for bootstrapping an address space from a file
- Program headers describe ranges to map
- Separate hot pages from cold
- Interval map manages indirection

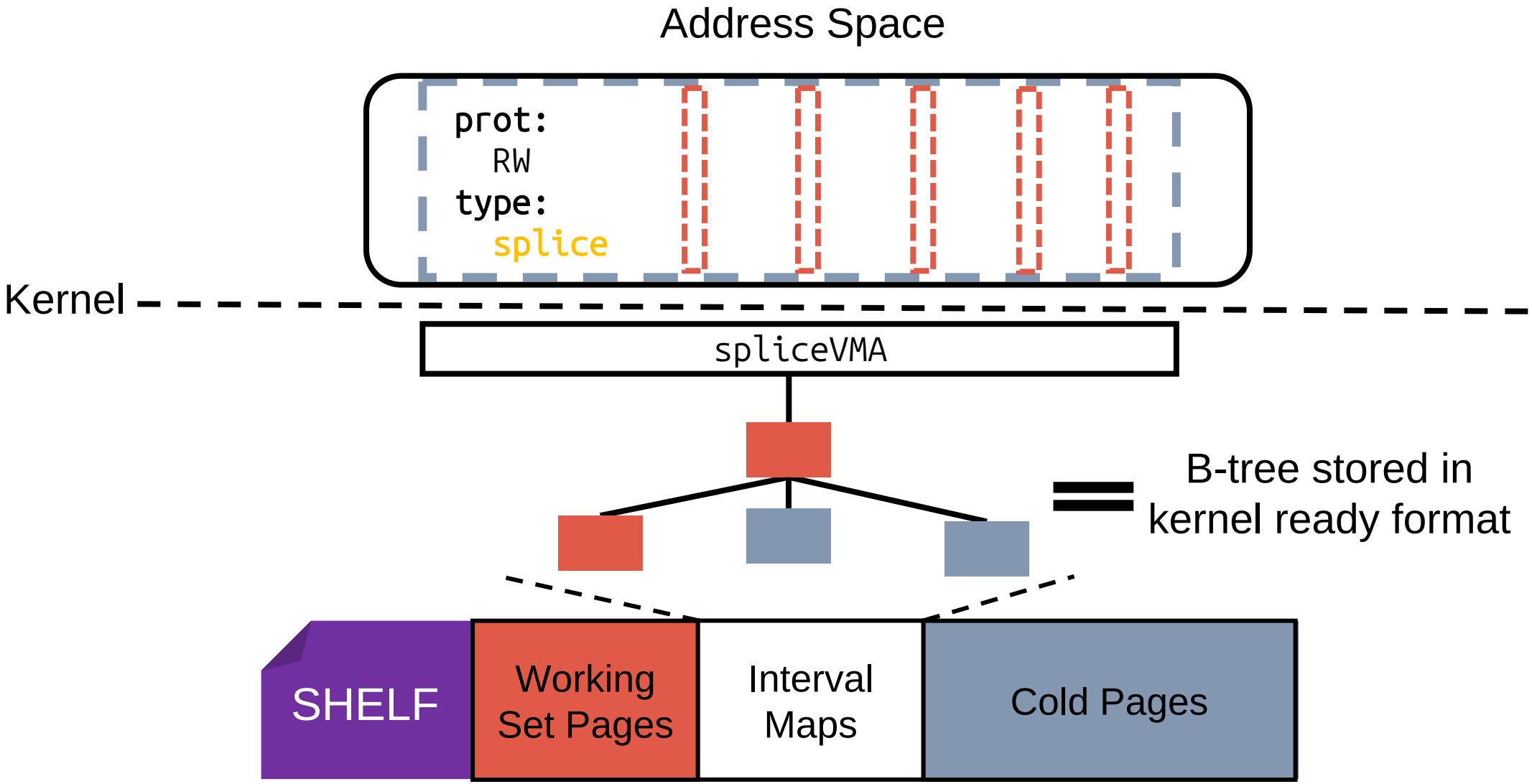


SpliceVMAs

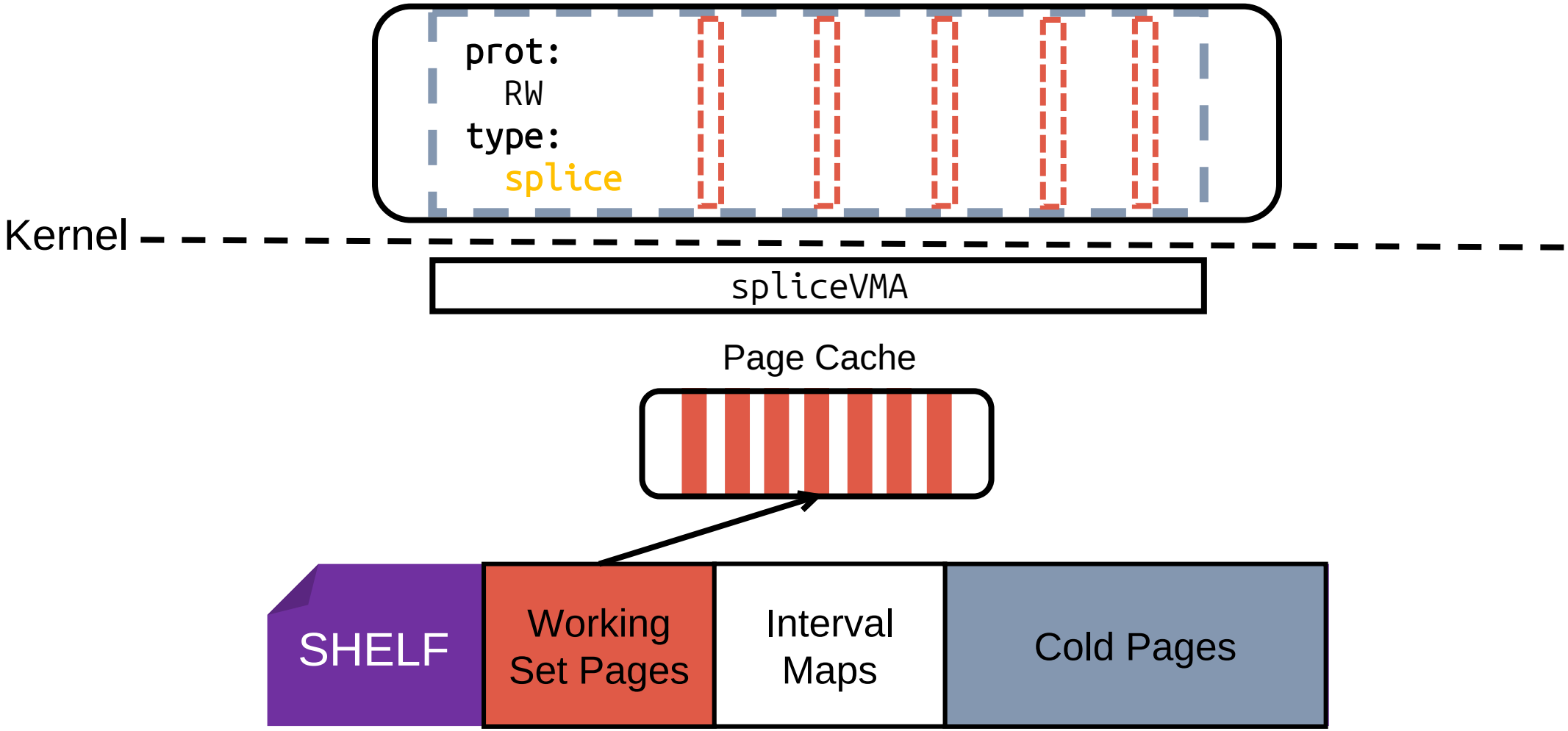

Address Space
prot:
RW
type:
splice
Kernel
spliceVMA


SHELF
Working Set Pages
Interval Maps
Cold Pages

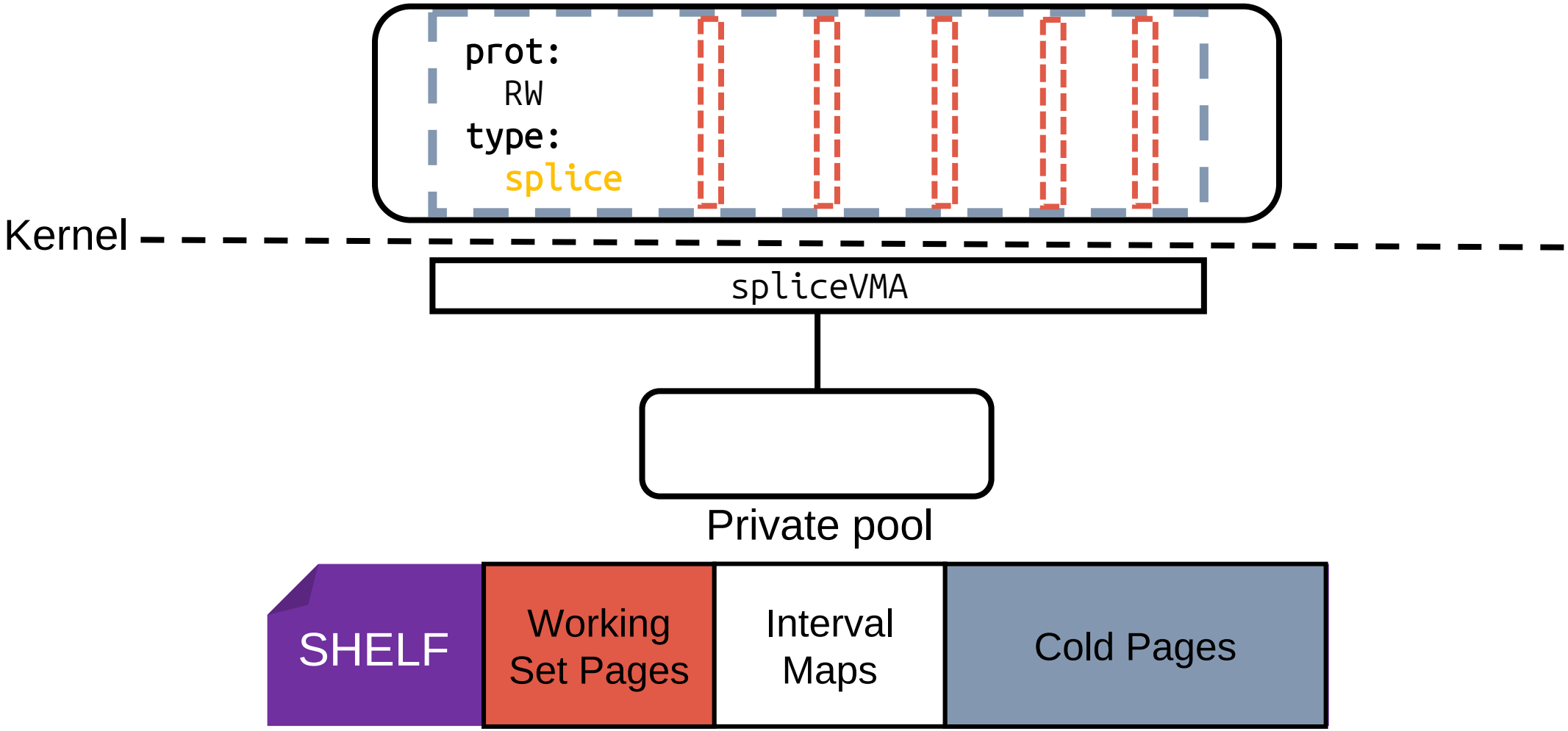
SpliceVMAs


Address Space
prot:
RW
type:
splice
Kernel
spliceVMA
B-tree stored in
kernel ready format
SHELF
Working
Set Pages
Interval
Maps
Cold Pages

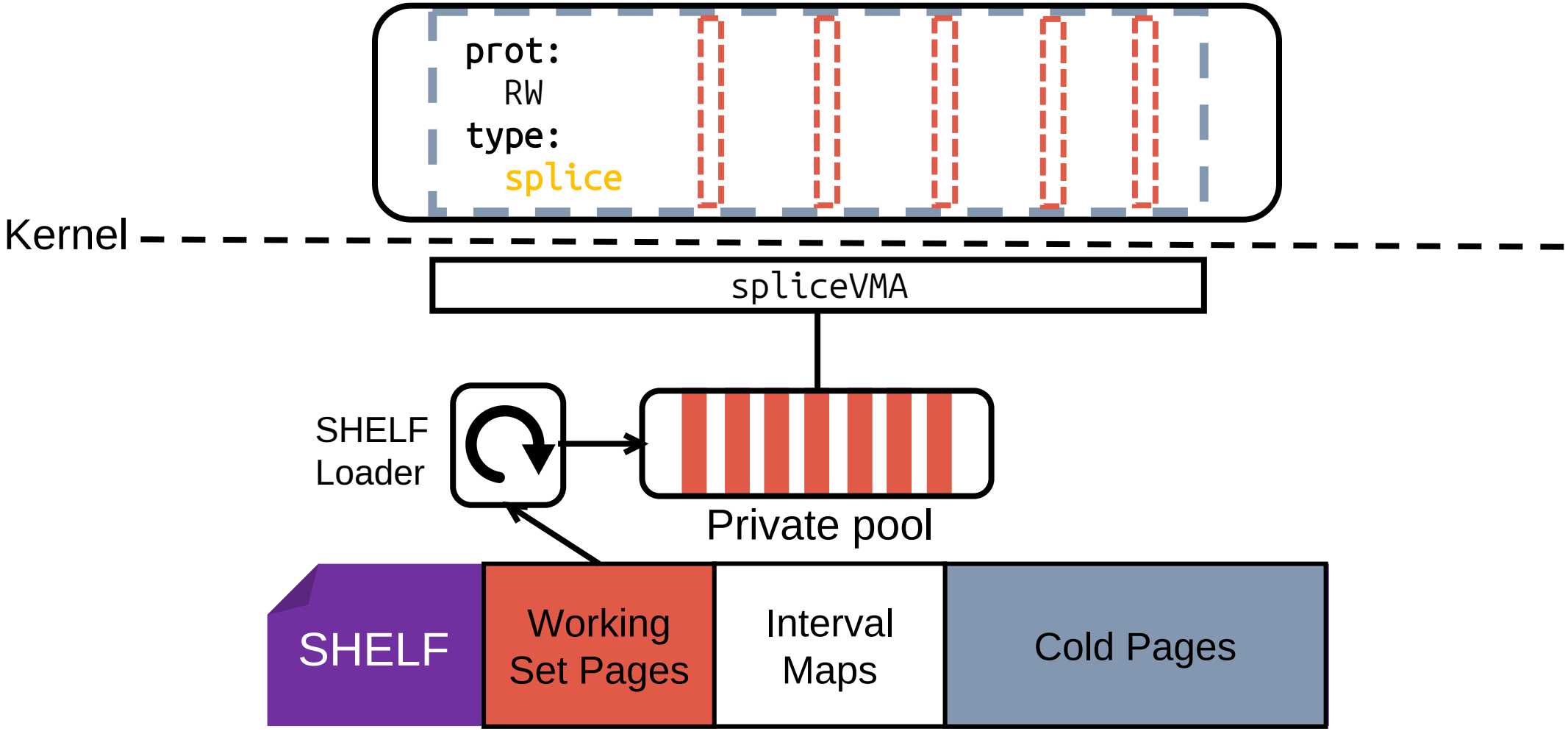
Prefetching: page cache bypass



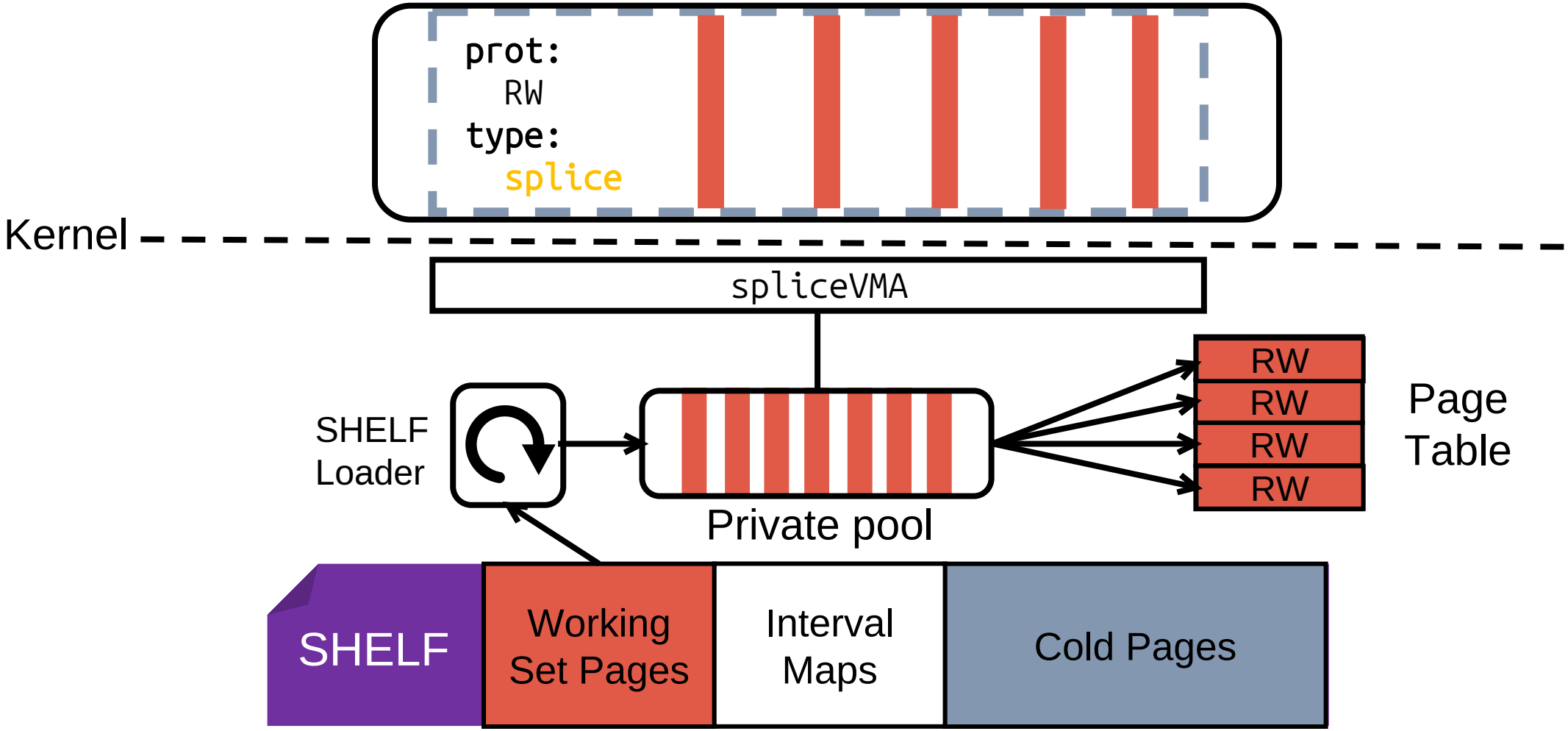
Prefetching: page cache bypass



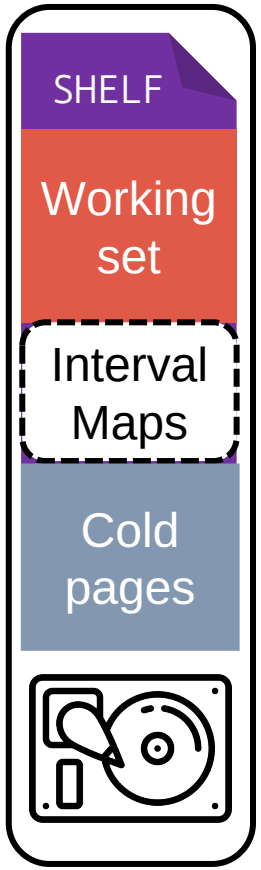
Prefetching: page cache bypass



Prefetching: page cache bypass

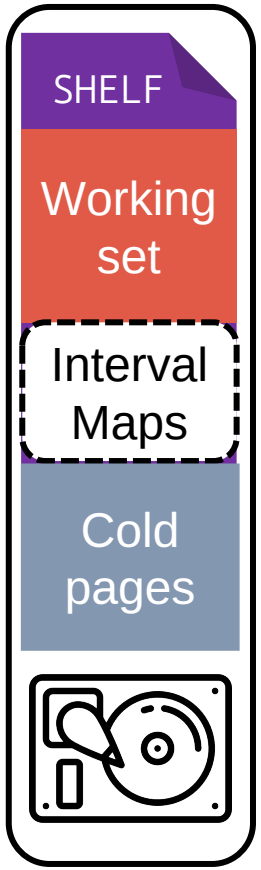


`reexec()`: Memory restore workflow



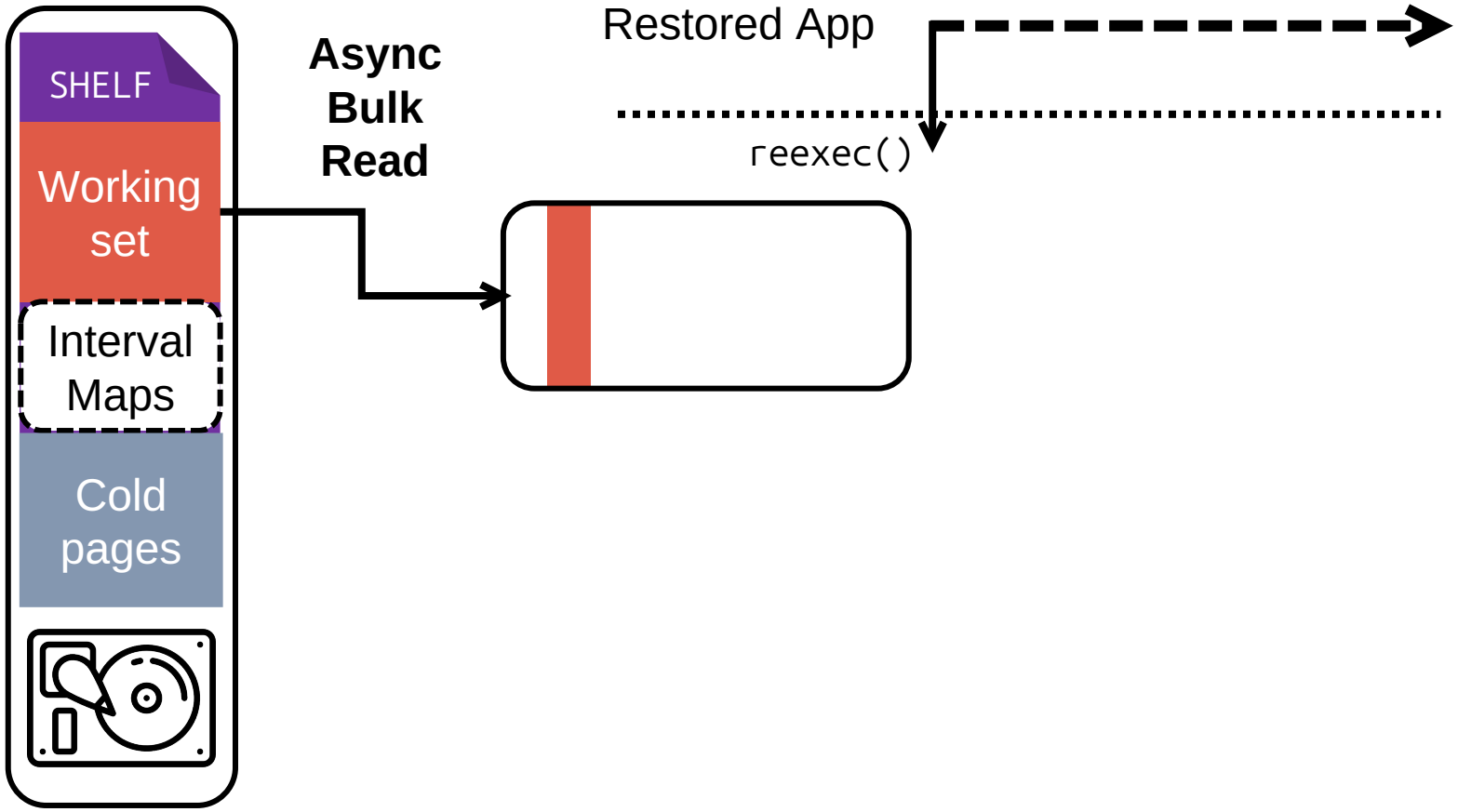
Restored App

reexec(): Memory restore workflow

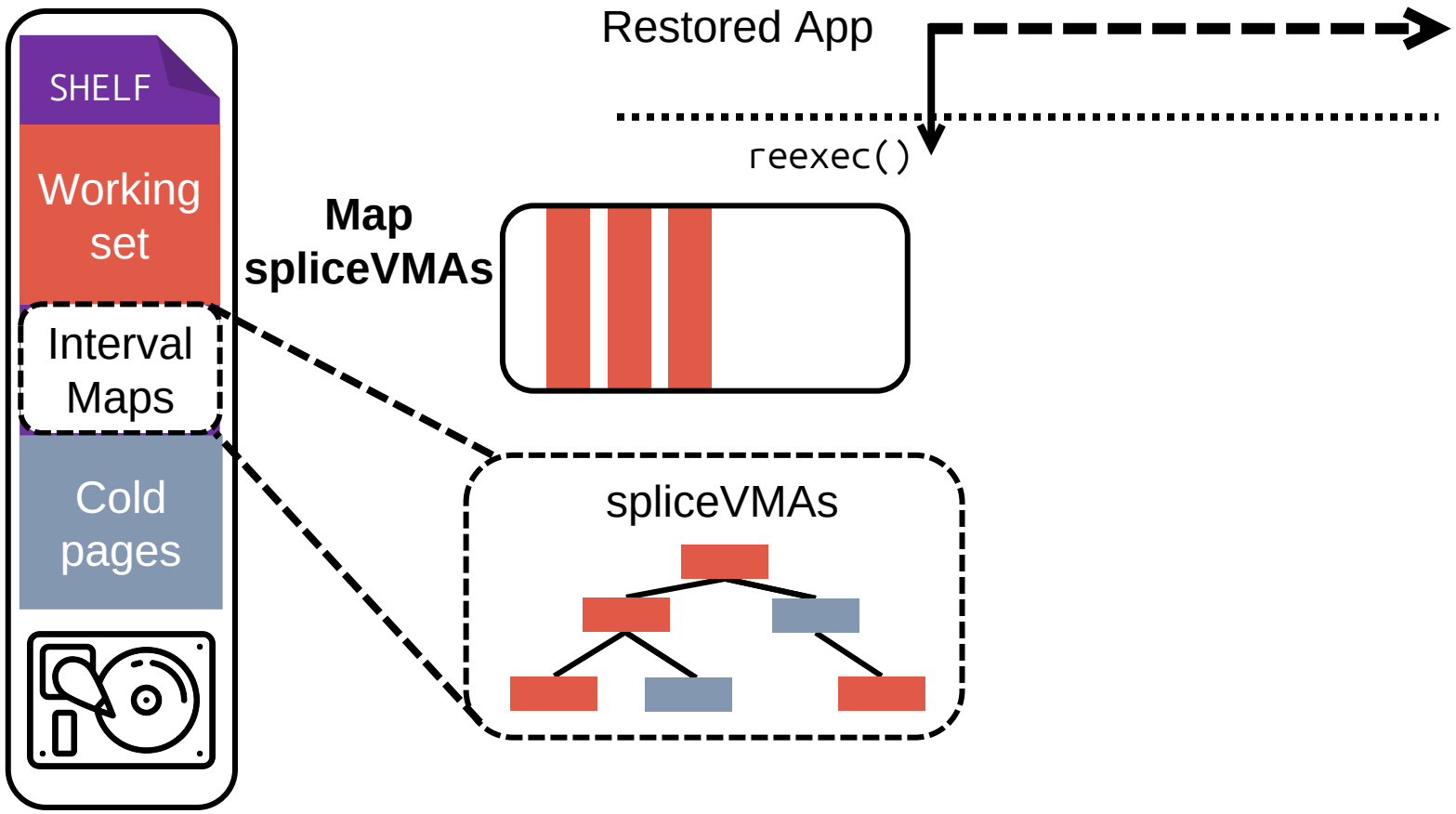




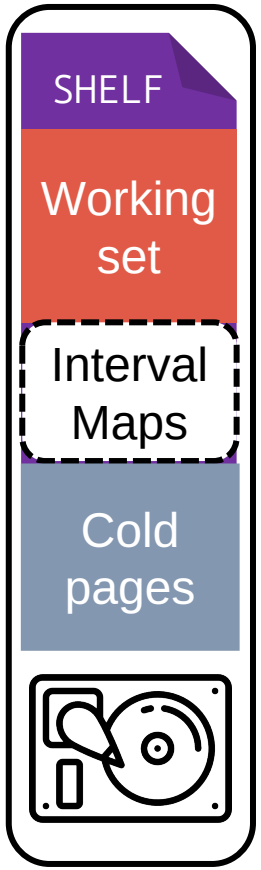
reexec(): Memory restore workflow

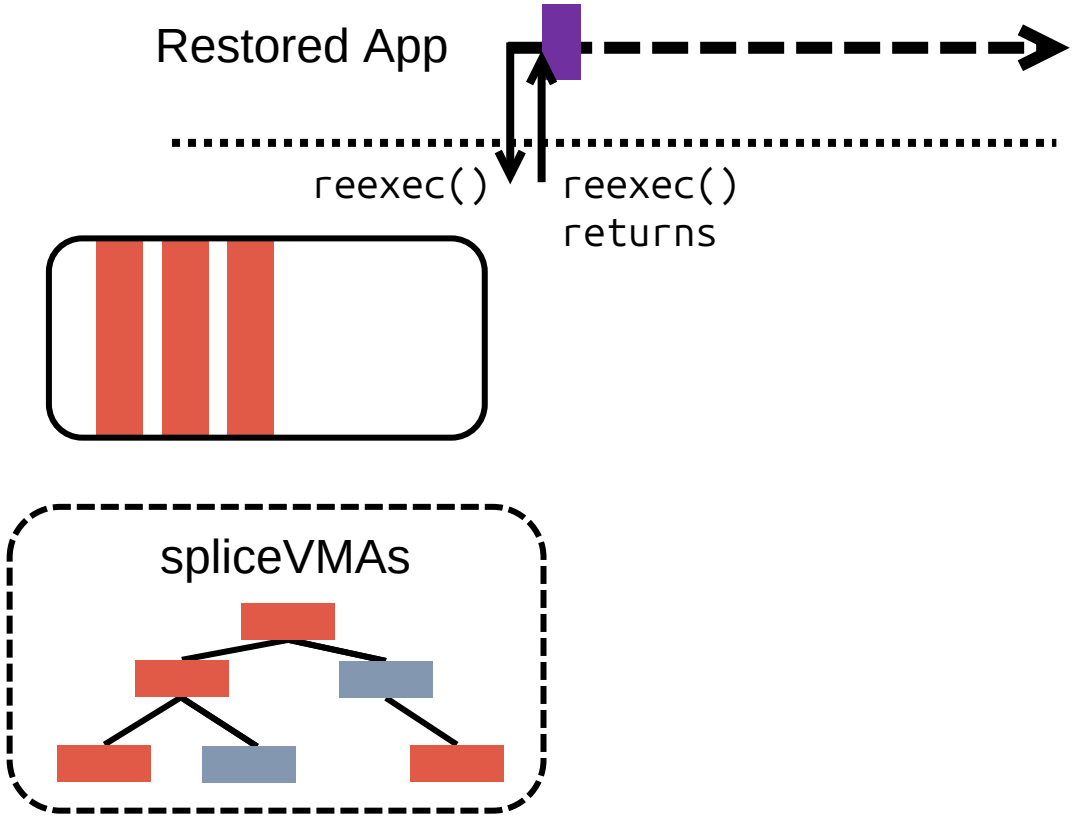


reexec(): Memory restore workflow

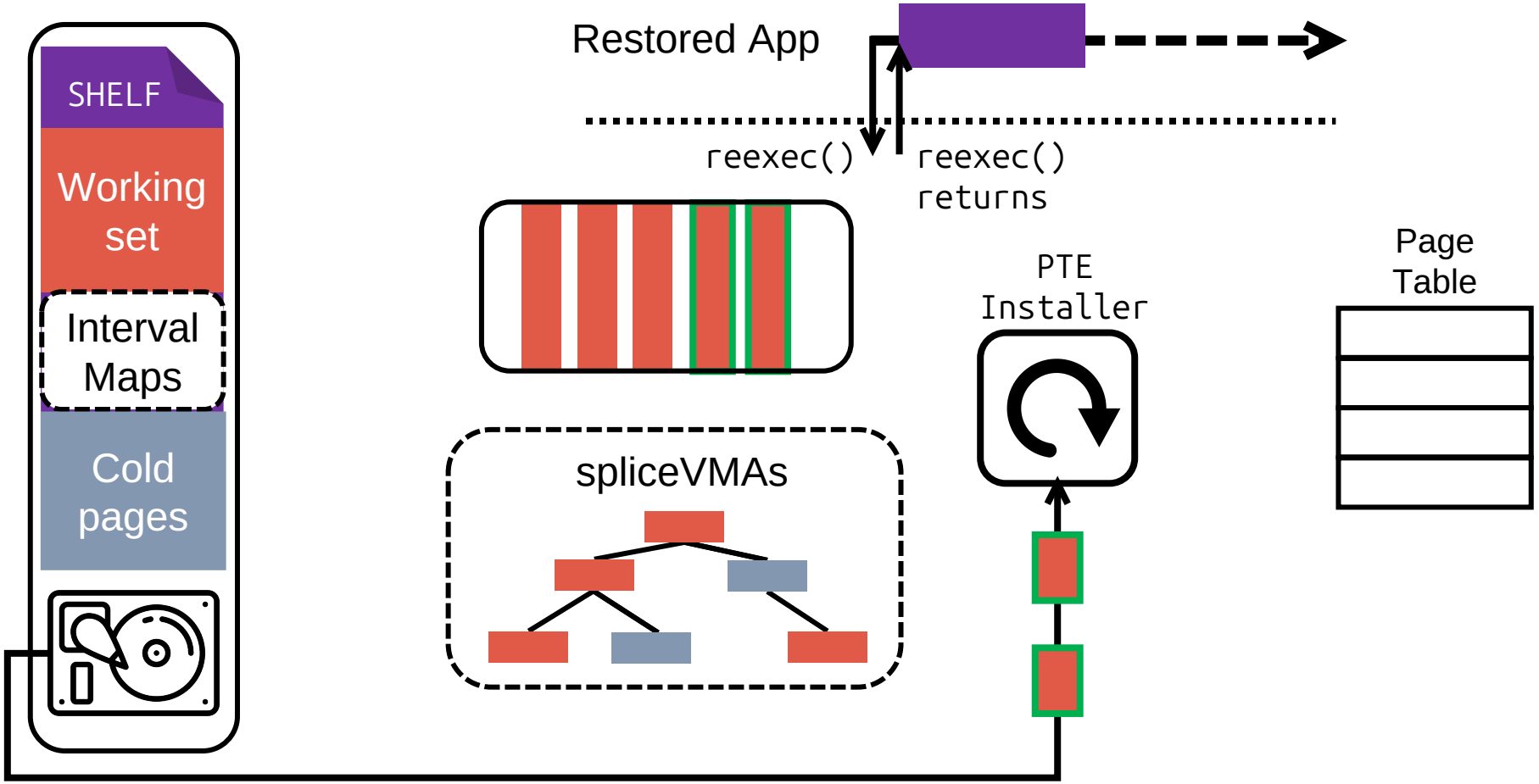


`reexec()`: Memory restore workflow

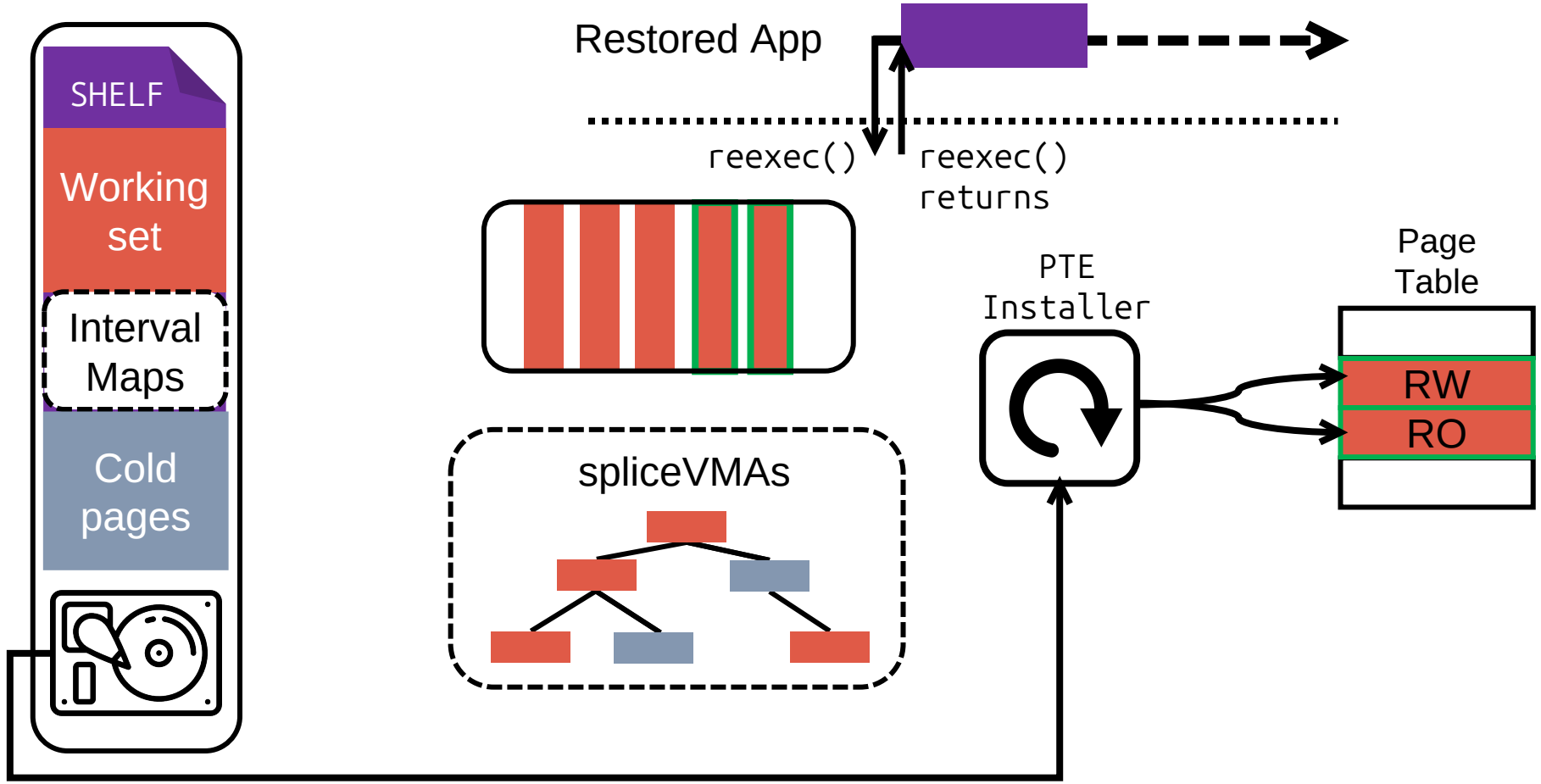




`reexec()`: Memory restore workflow



reexec(): Memory restore workflow



Implementation

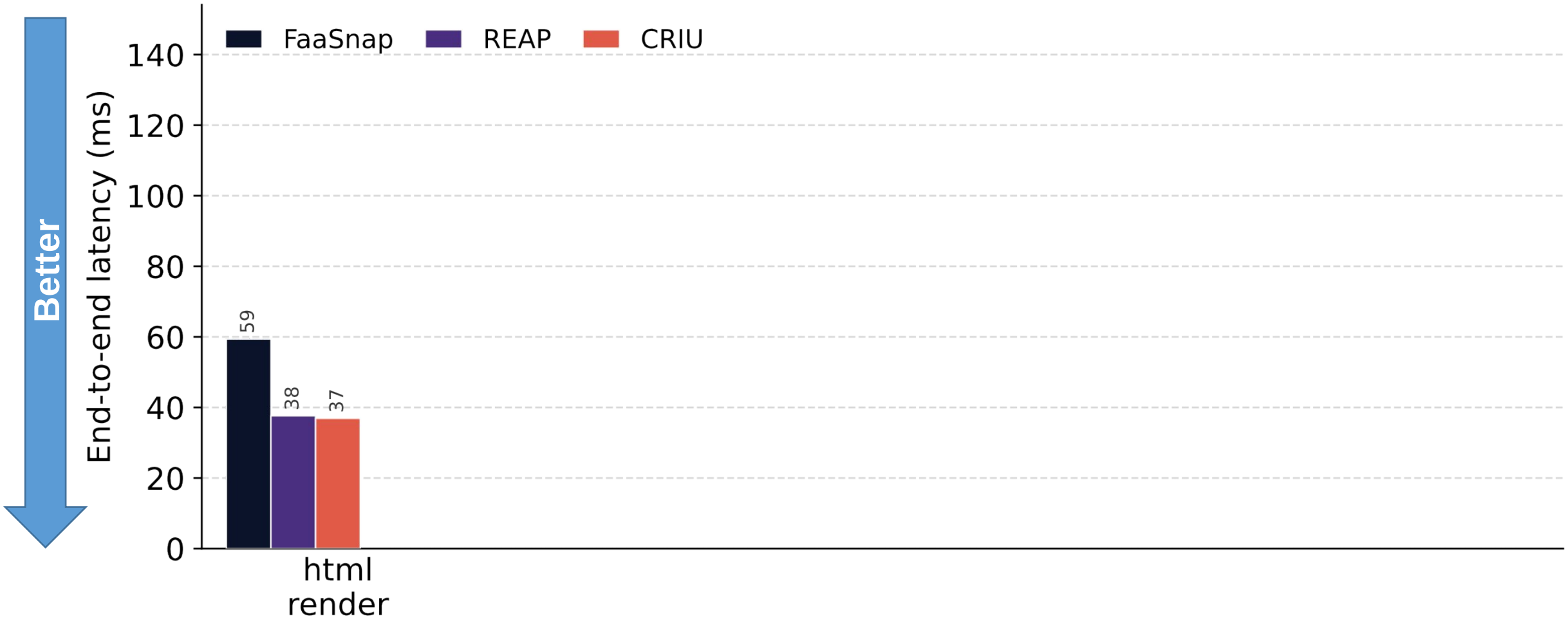
- Memory management: 7k LoC Linux 6.5 kernel module
  - SpliceVMAs
  - SHELF loader (additional optimizations in the paper)

- Process restore
  - Process state (de)serialization implemented in Junction libOS [NSDI ‘24]
  - Requires no modification to applications

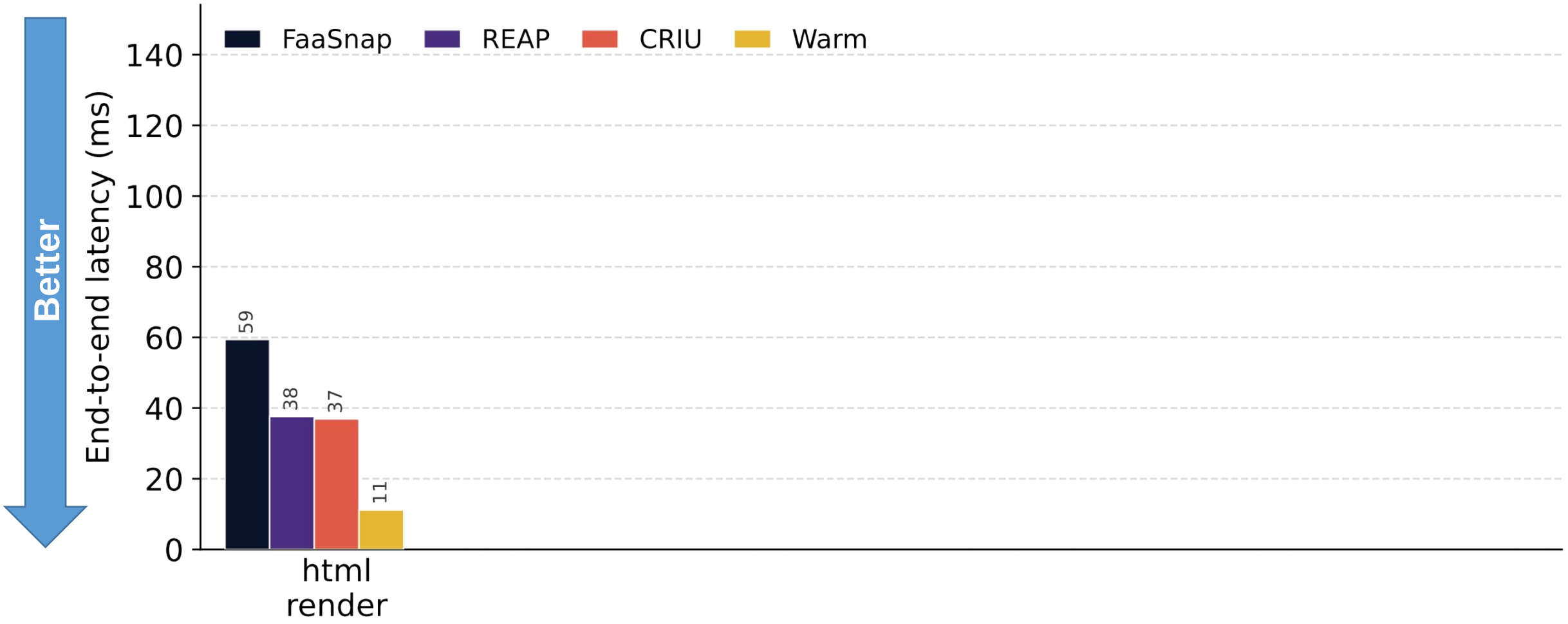
Evaluation

- **Evaluation metric:**
  - End-to-end latency of function invocations from **FunctionBench**
- **Systems evaluated:**
  - FaaSnap [EuroSys ‘22]
  - REAP [ASPLOS ‘21]
  - CRIU

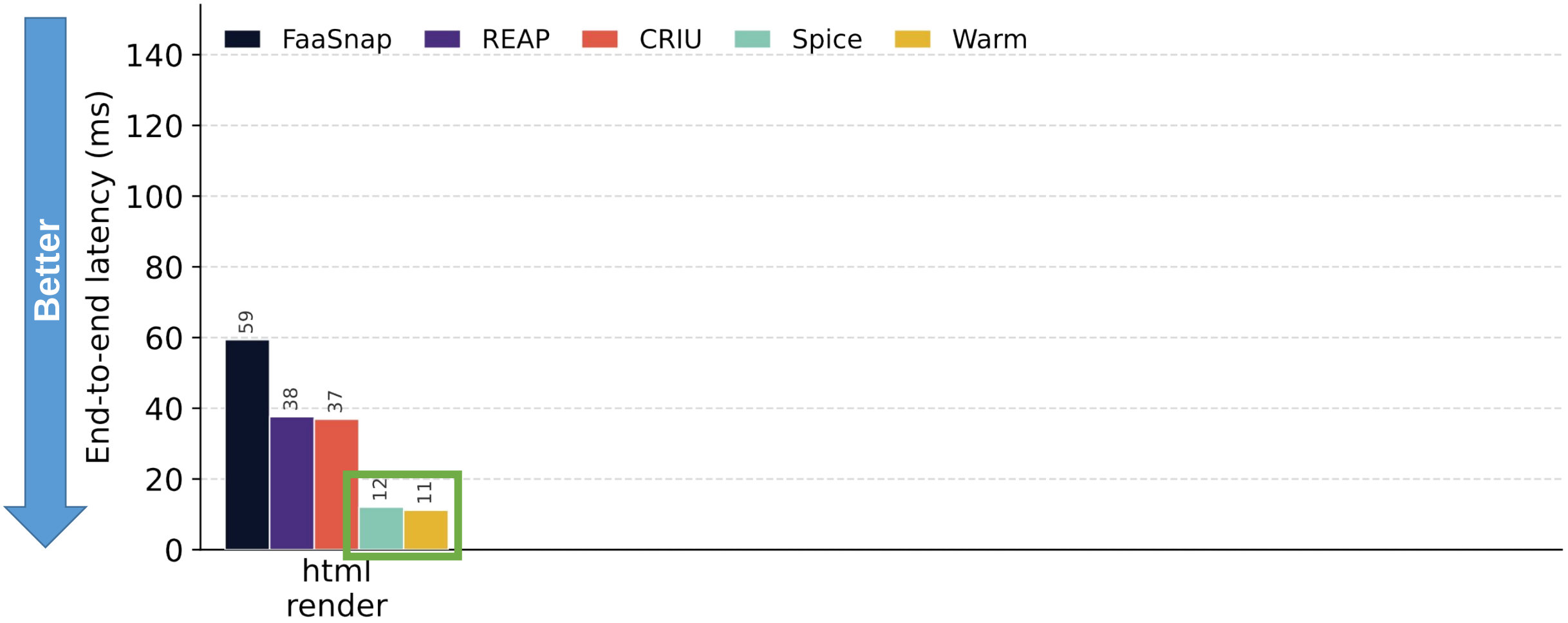
Invocation Latency



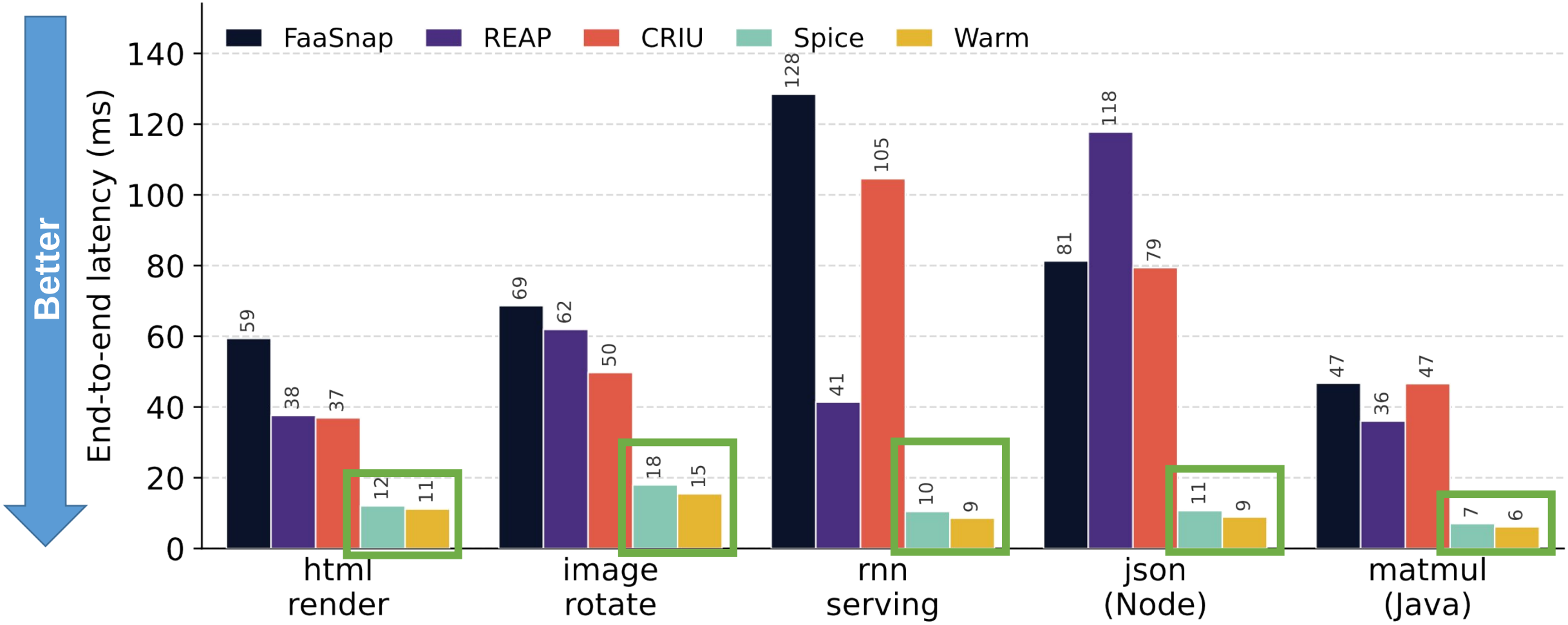
Invocation Latency



Invocation Latency



Invocation Latency



Conclusion

First-class kernel support for restoring processes enables near-warm latency for on-demand execution

Spice is open source!



Conclusion

First-class kernel support for restoring processes enables near-warm latency for on-demand execution

This unlocks potential for aggressively reclaiming resources that serverless promises

Spice is open source!



Conclusion

First-class kernel support for restoring processes enables near-warm latency for on-demand execution

This unlocks potential for aggressively reclaiming resources that serverless promises

Spice is open source!



Thank you!