

BabelFish: Fusing Address Translations for Containers

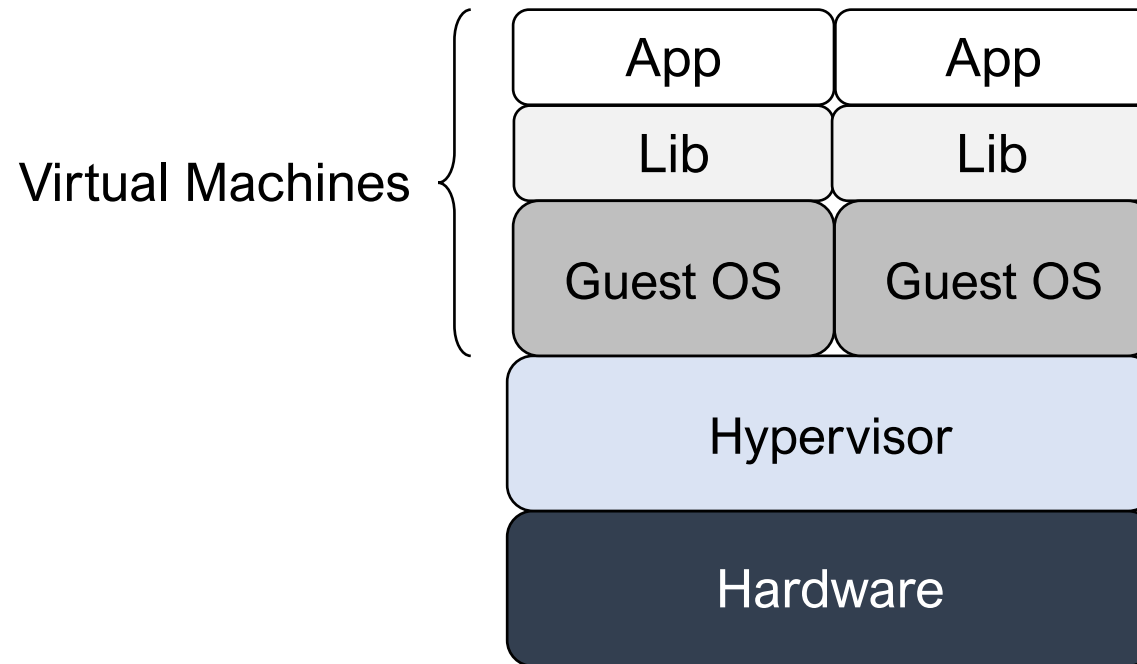
Dimitrios Skarlatos, Umur Darbaz, Bhargava Gopireddy,
Nam Sung Kim, and Josep Torrellas

University of Illinois at Urbana-Champaign

skarlat2@illinois.edu

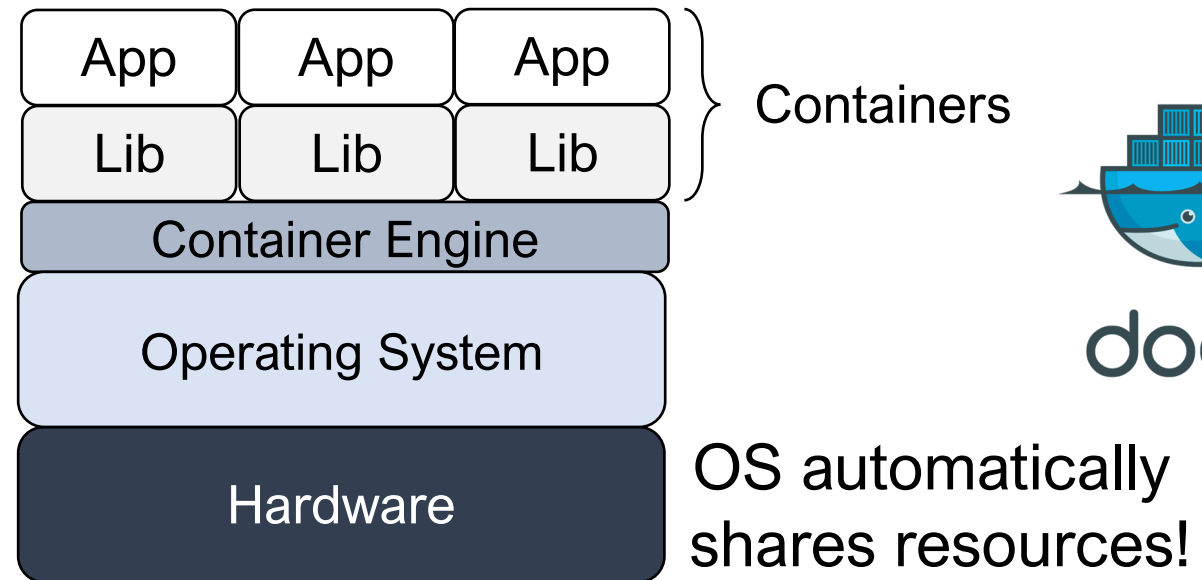
ISCA 2020

Conventional Cloud Computing



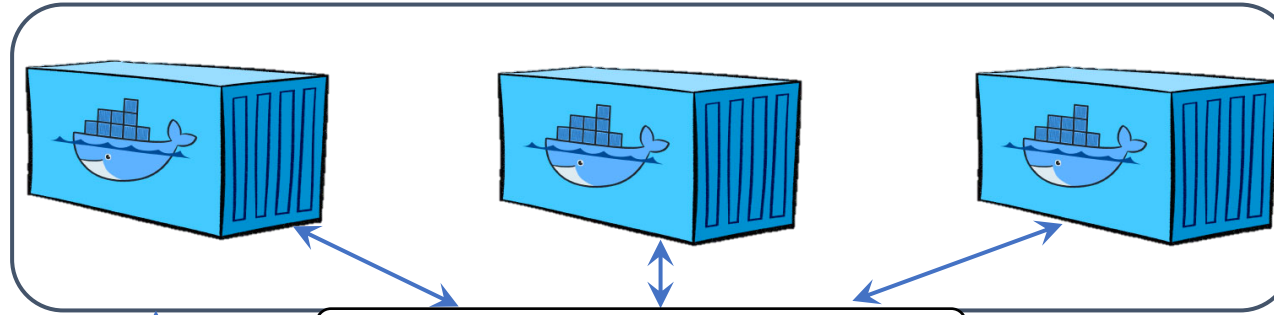
New Era in Cloud Computing

1. Lightweight
2. Faster bringup
3. Higher consolidation

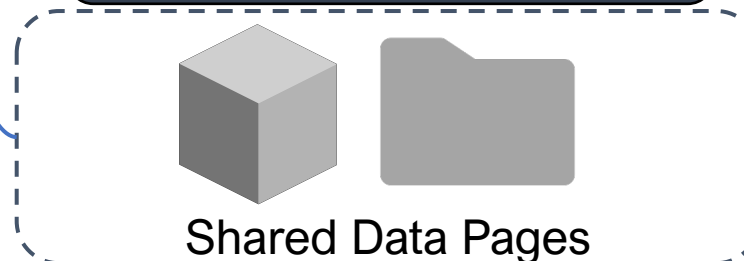
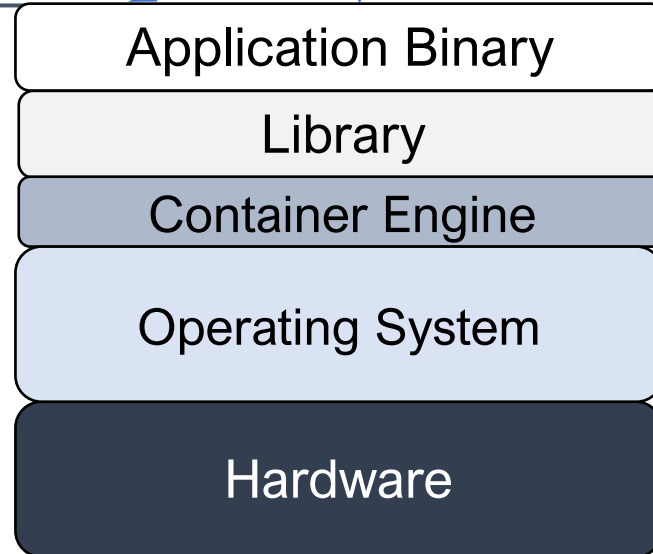


Containers in the Cloud

Pod



kubernetes



Replicated Containers

Scalability

Fault Tolerance

Load Balancing

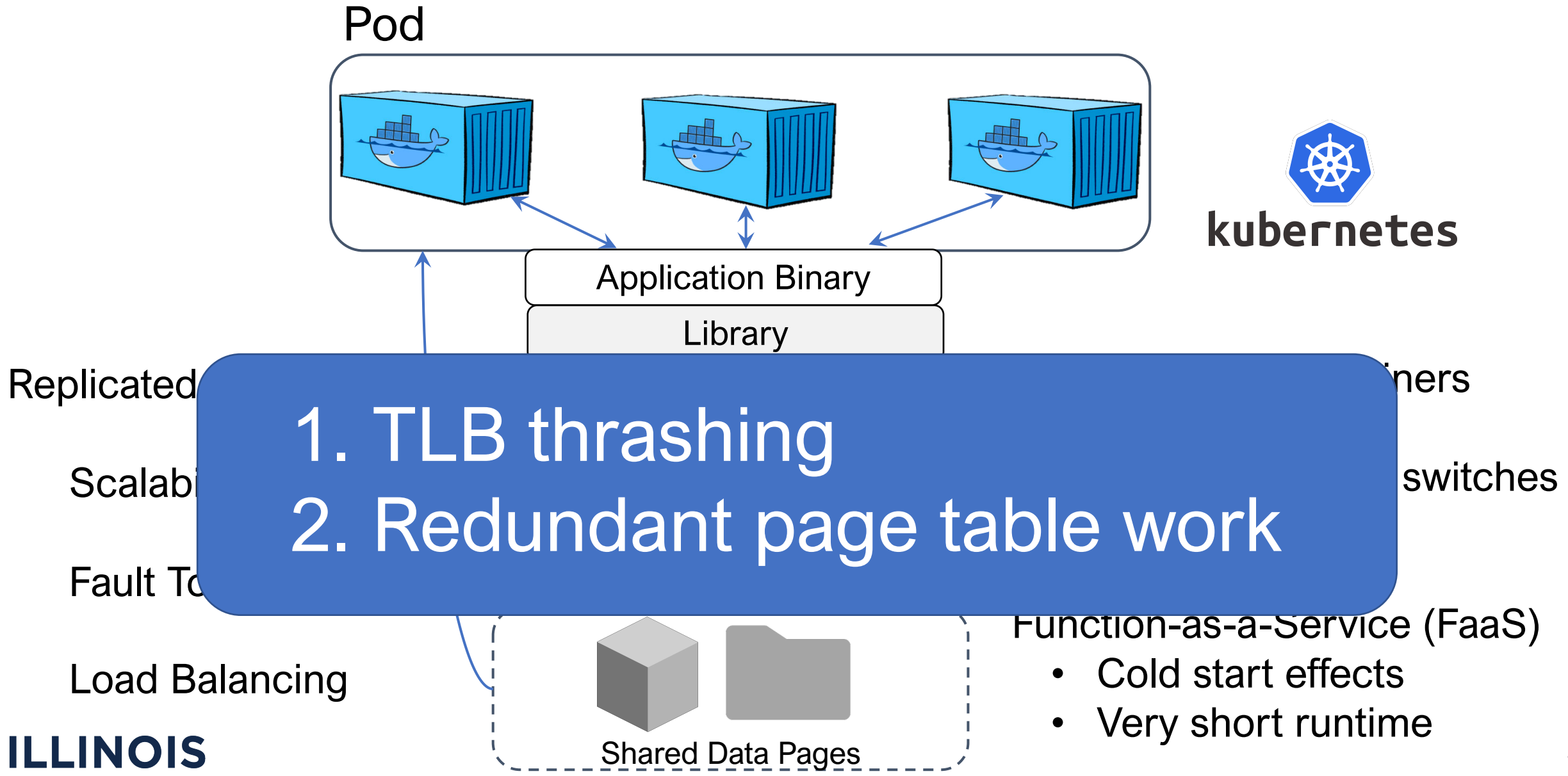
High number of containers

Very frequent context switches

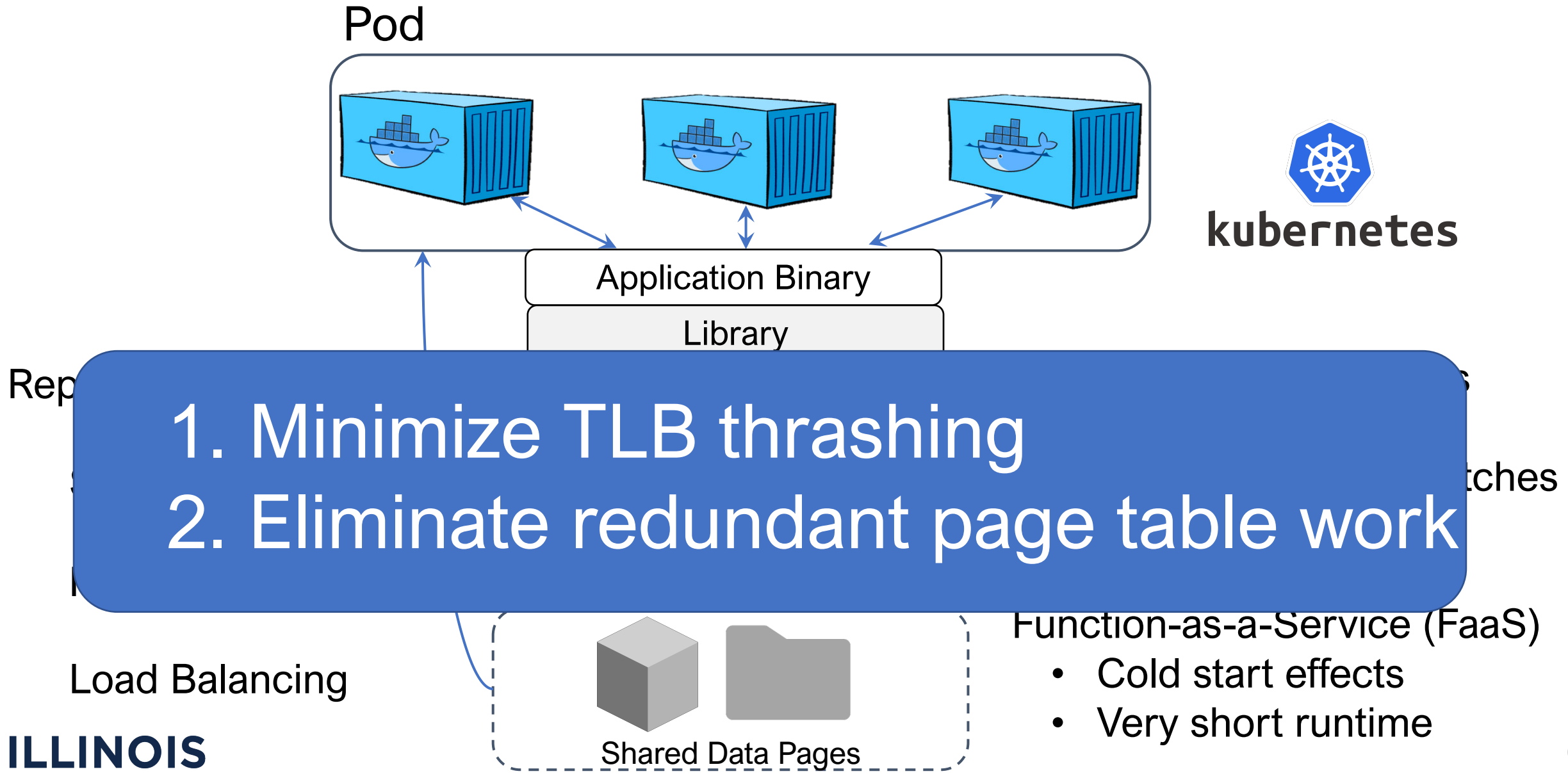
Serverless &
Function-as-a-Service (FaaS)

- Cold start effects
- Very short runtime

Problem: Replicated VA→PA Translations



Solution: Share VA→PA Translations



Contribution: BabelFish

HW and OS support to share translations across containers

1. Introduce Container Context IDentifiers (CCID)
2. Extend TLB design
3. Share page tables

Performance improvement

Data-serving: 11%-18%

Compute: 11%

Function-as-a-Service: 10-55%

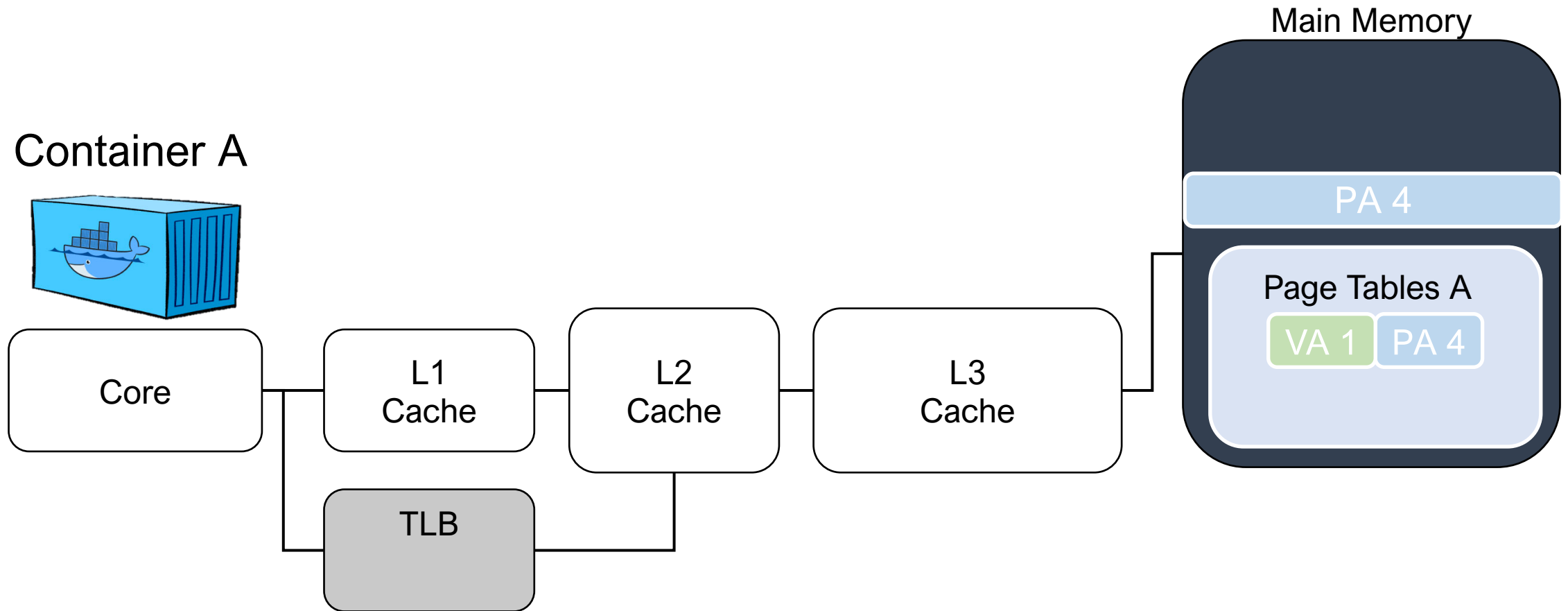
Container bring-up: 8%



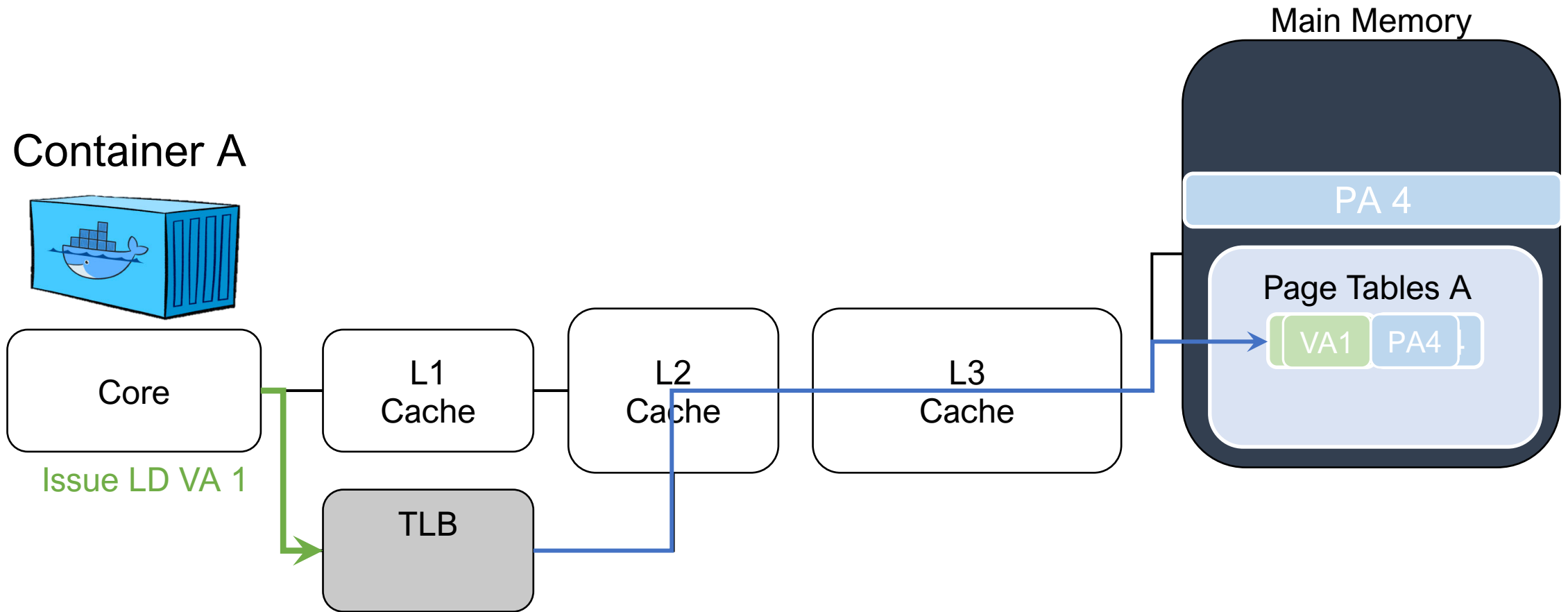
“The Babel fish is small, yellow, leech-like, and probably the oddest thing in the Universe. The practical upshot of all this is that if you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language.”

The Hitchhiker's Guide to the Galaxy – Douglas Adams

Problem: TLB Bloat

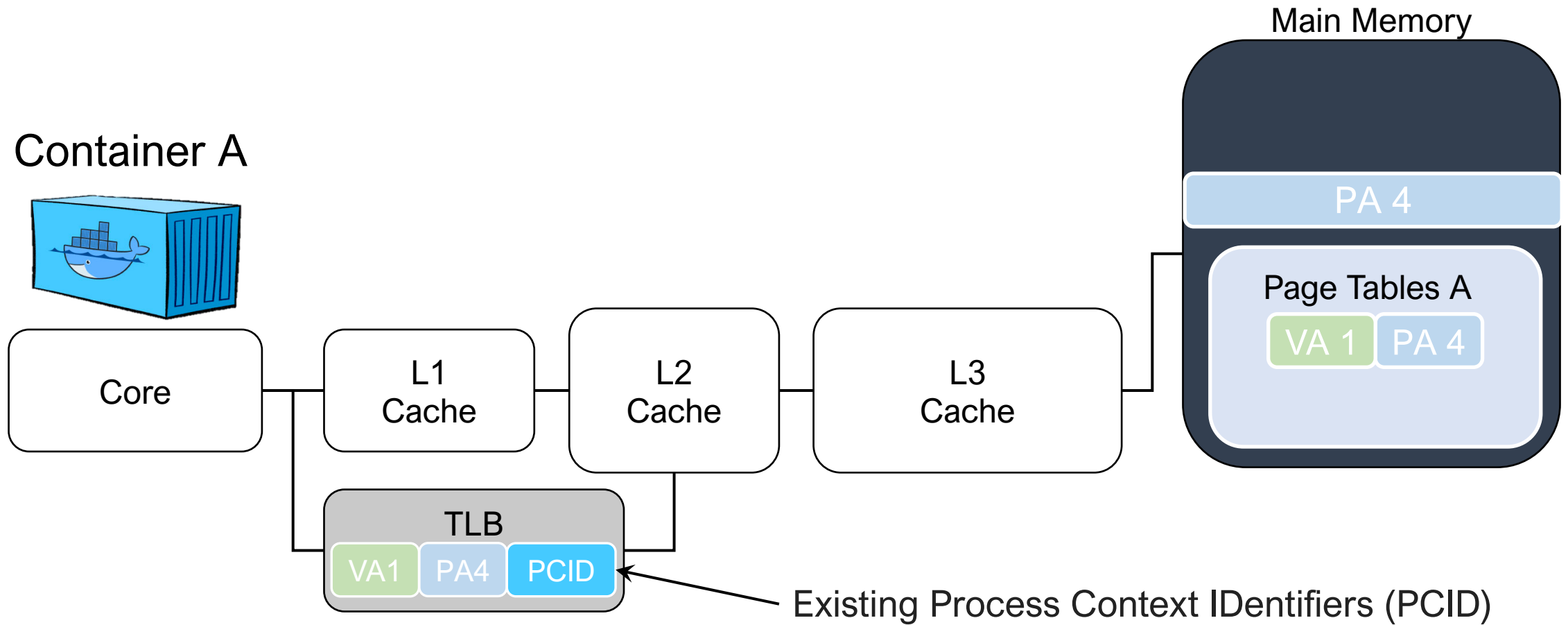


Problem: TLB Bloat

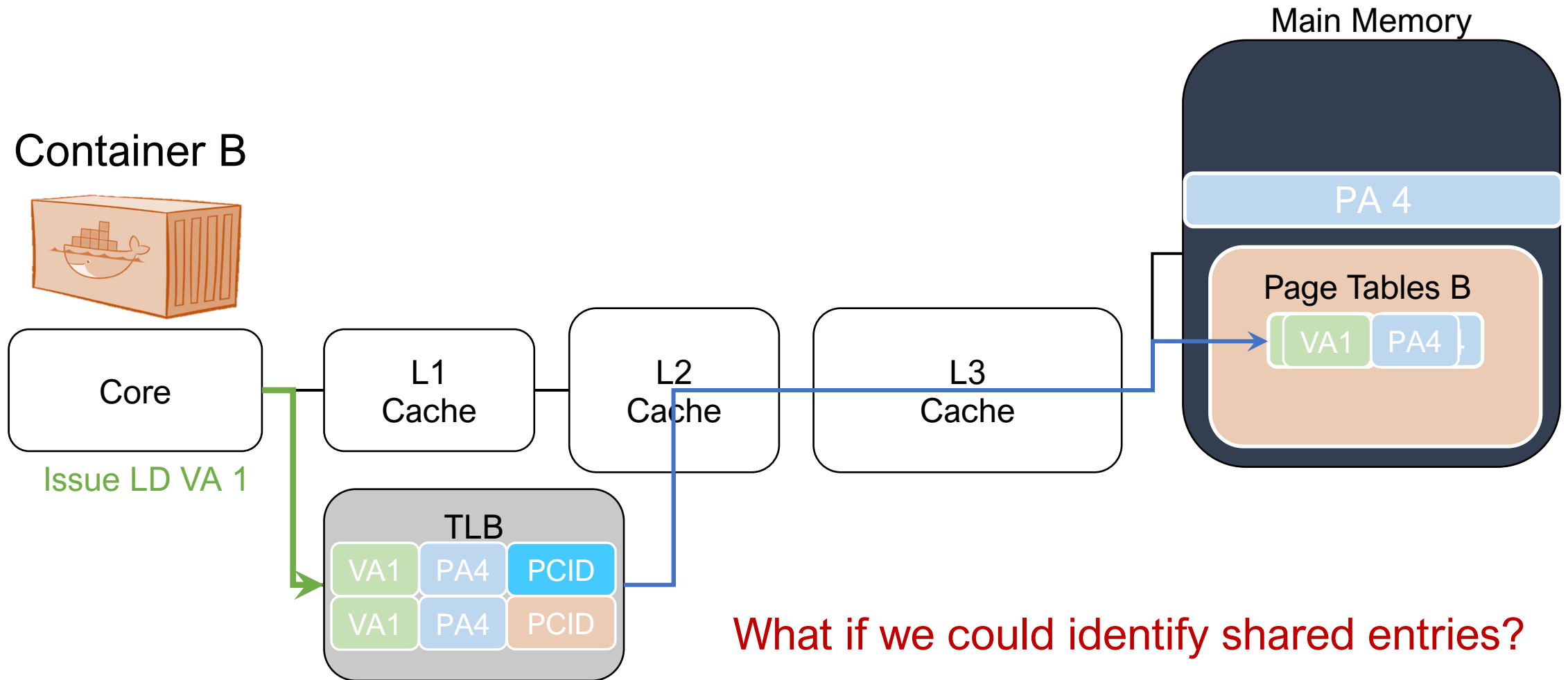


TLB Miss → "Page Walk" = Fetch entry from page table

Problem: TLB Bloat



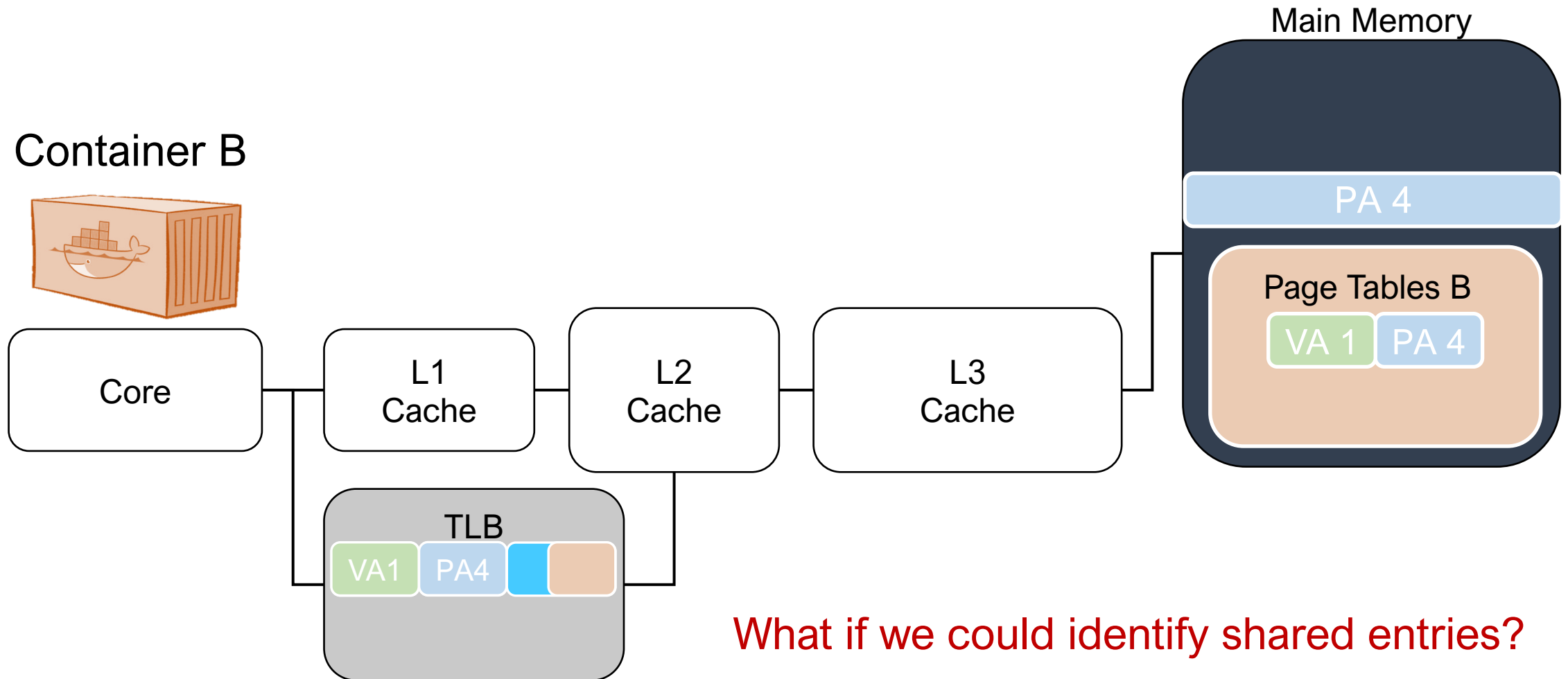
Problem: TLB Bloat



What if we could identify shared entries?

TLB Miss → "Page Walk" = Fetch entry from page table

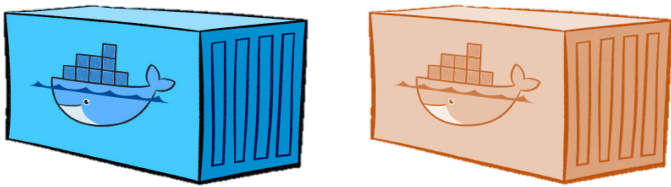
Problem: TLB Bloat



Solution: BabelFish TLB

Container Context Identifier (CCID)

Container A Container B



Core

L1
Cache

L2
Cache

L3
Cache

TLB

VA1

PA4

CCID

Main Memory

PA 4

Page Tables B

VA 1

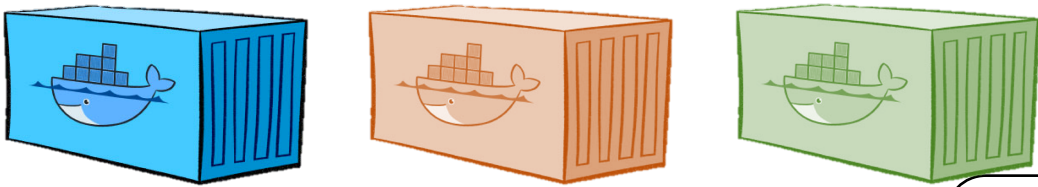
PA 4

Unlimited sharing of translations
among any number of containers!

Solution: BabelFish TLB

Container Context Identifier (CCID)

Container A Container B Container C



Core

L1
Cache

L2
Cache

L3
Cache

TLB

VA1 PA4 CCID PC

Main Memory

PA 4

Page Tables B

VA 1 PA 4

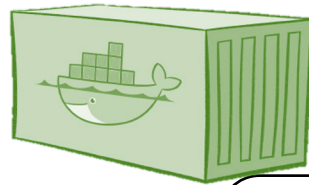
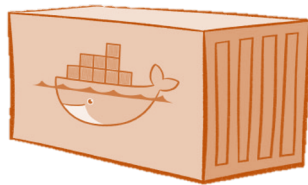
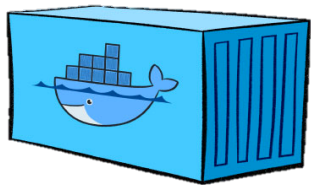
PrivateCopy (PC)

VA→PA shared for some containers
& private for others

Solution: BabelFish TLB

Container Context Identifier (CCID)

Container A Container B Container C



Core

L1
Cache

L2
Cache

L3
Cache

TLB

VA1	PA4	CCID	PC
VA1	PA6	CCID	PC

Main Memory

PA 4

Page Tables B

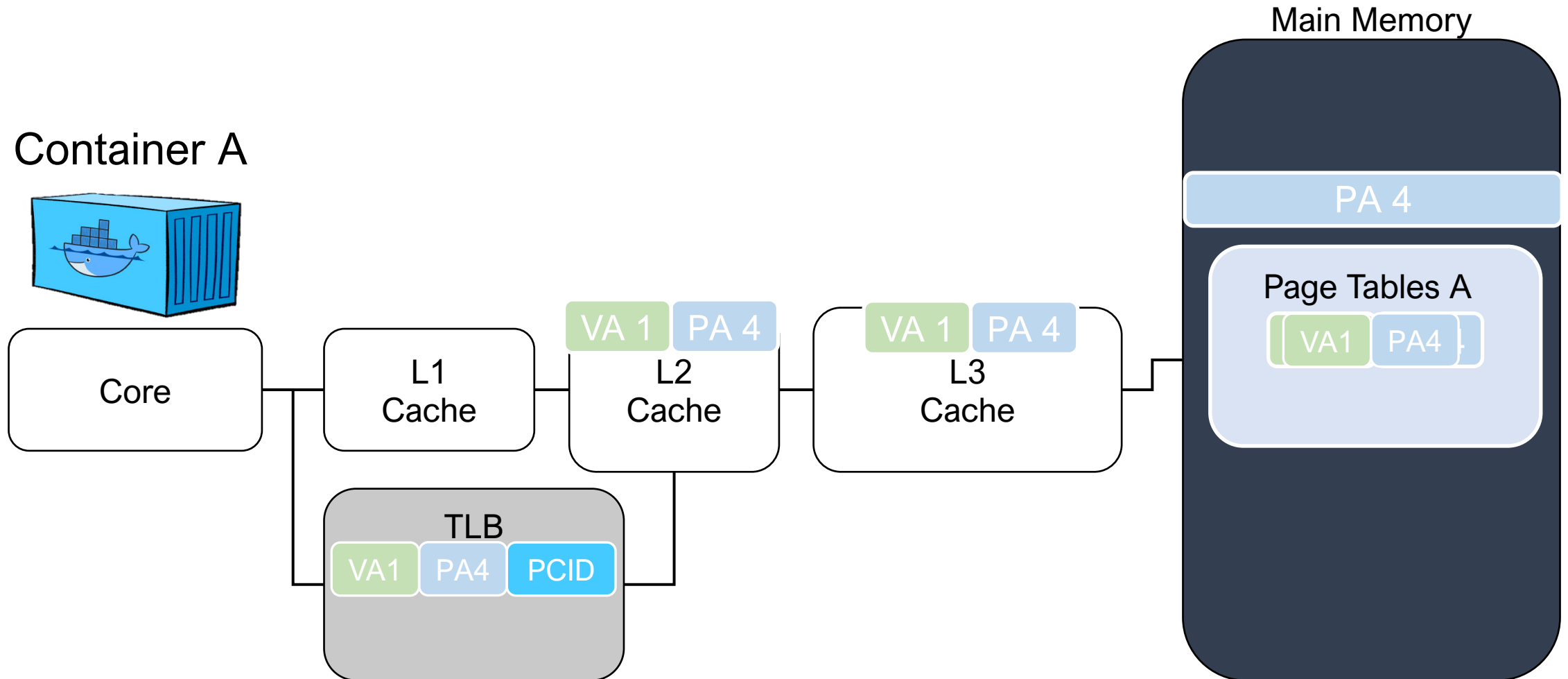
VA 1

PA 4

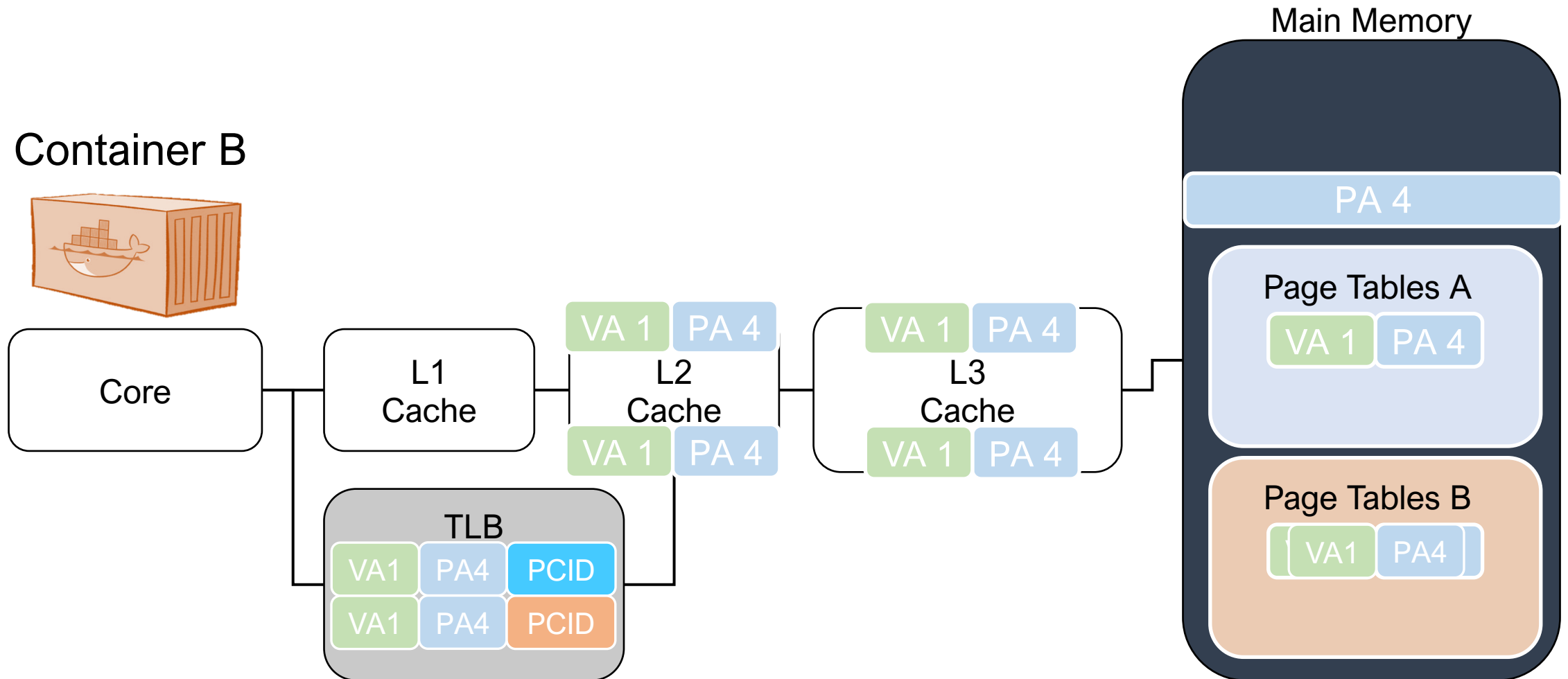
PrivateCopy (PC)

VA→PA shared for some containers
& private for others

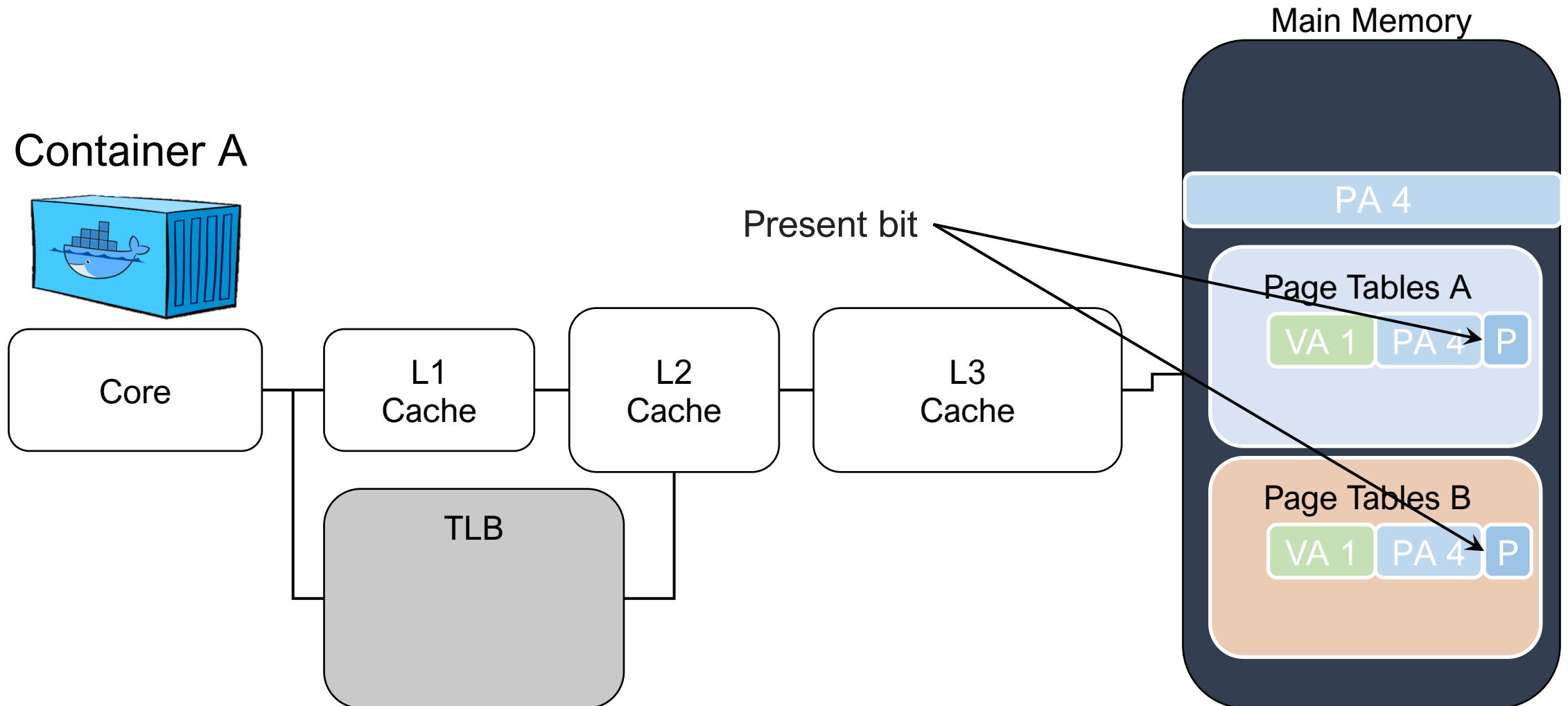
Problem: Page Table Entry Bloat



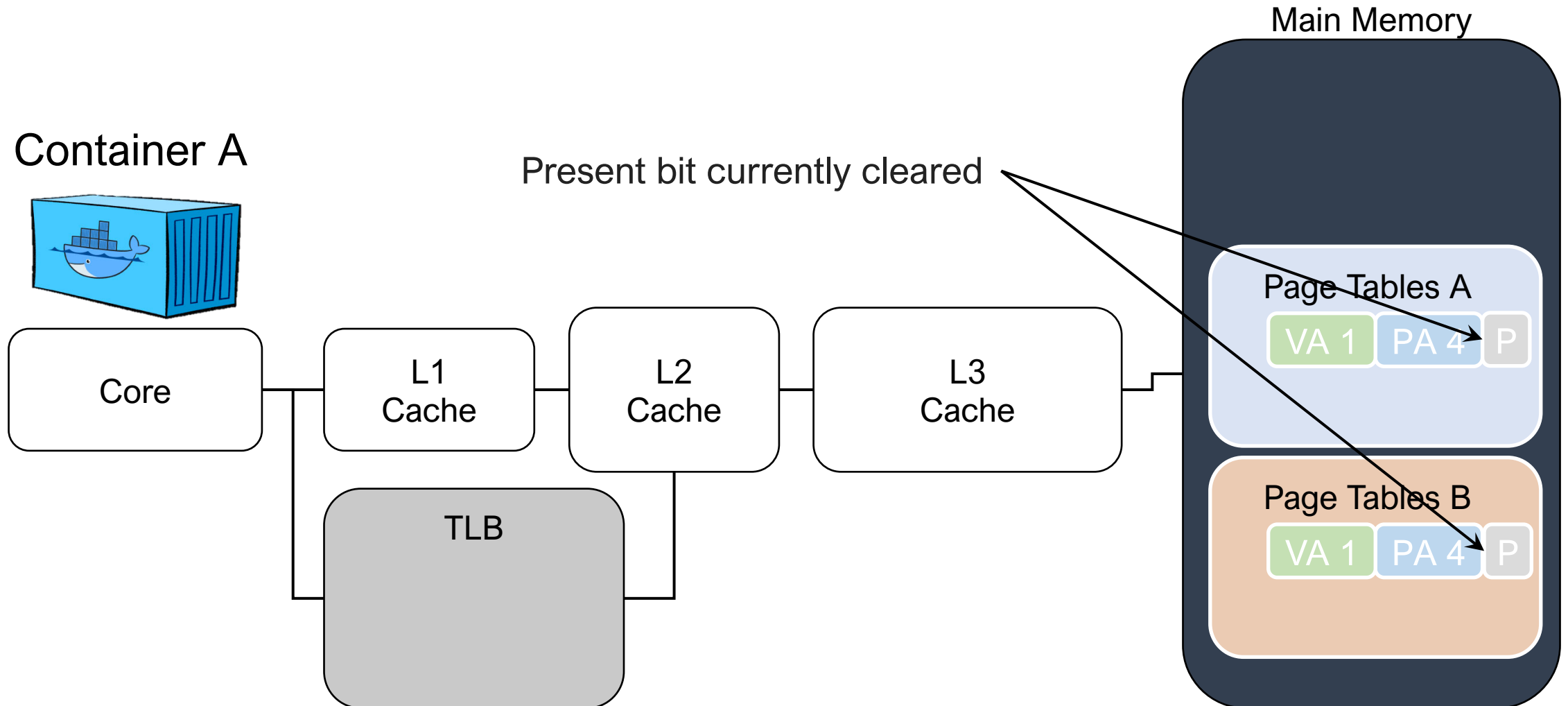
Problem: Page Table Entry Bloat



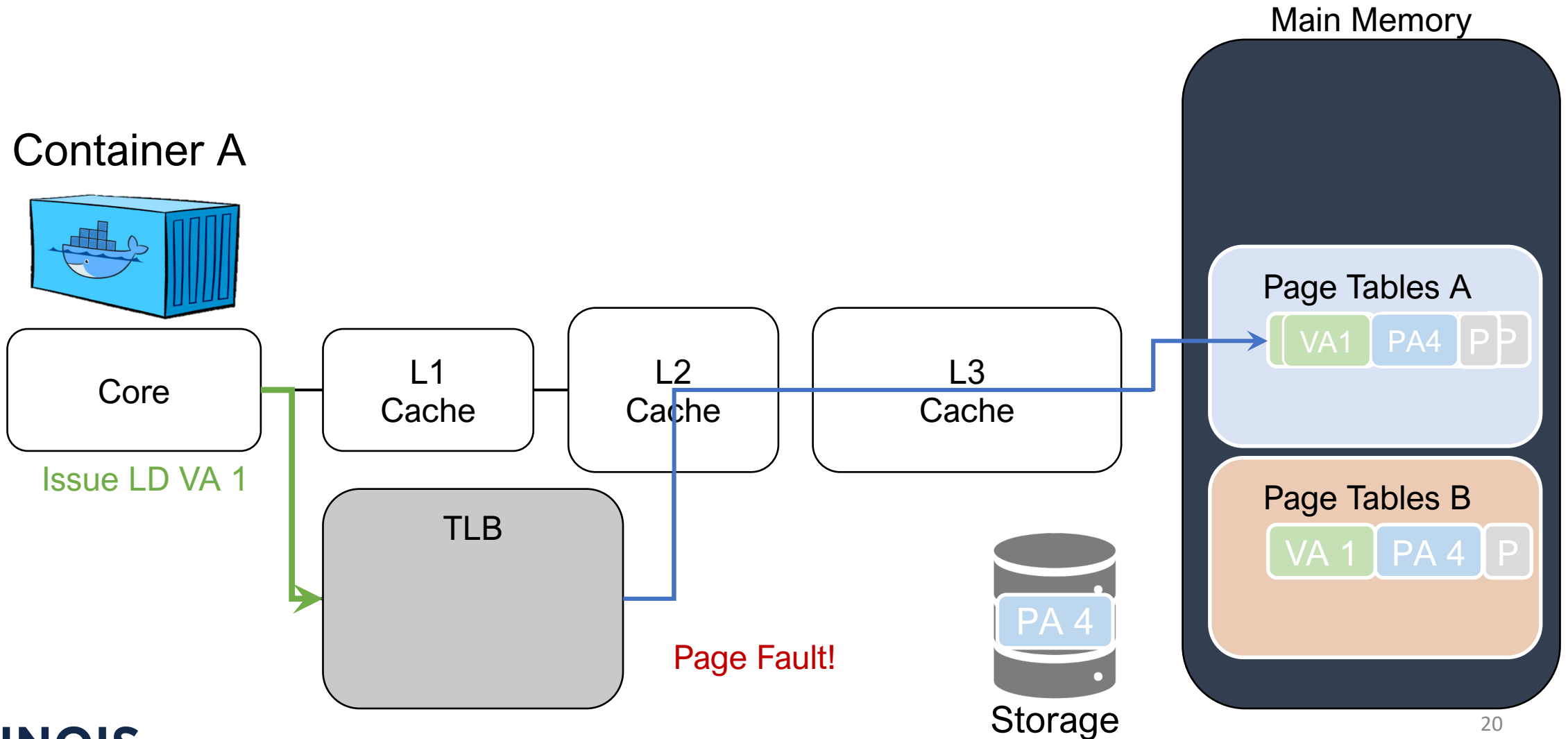
Problem: Lazy Page Table Management



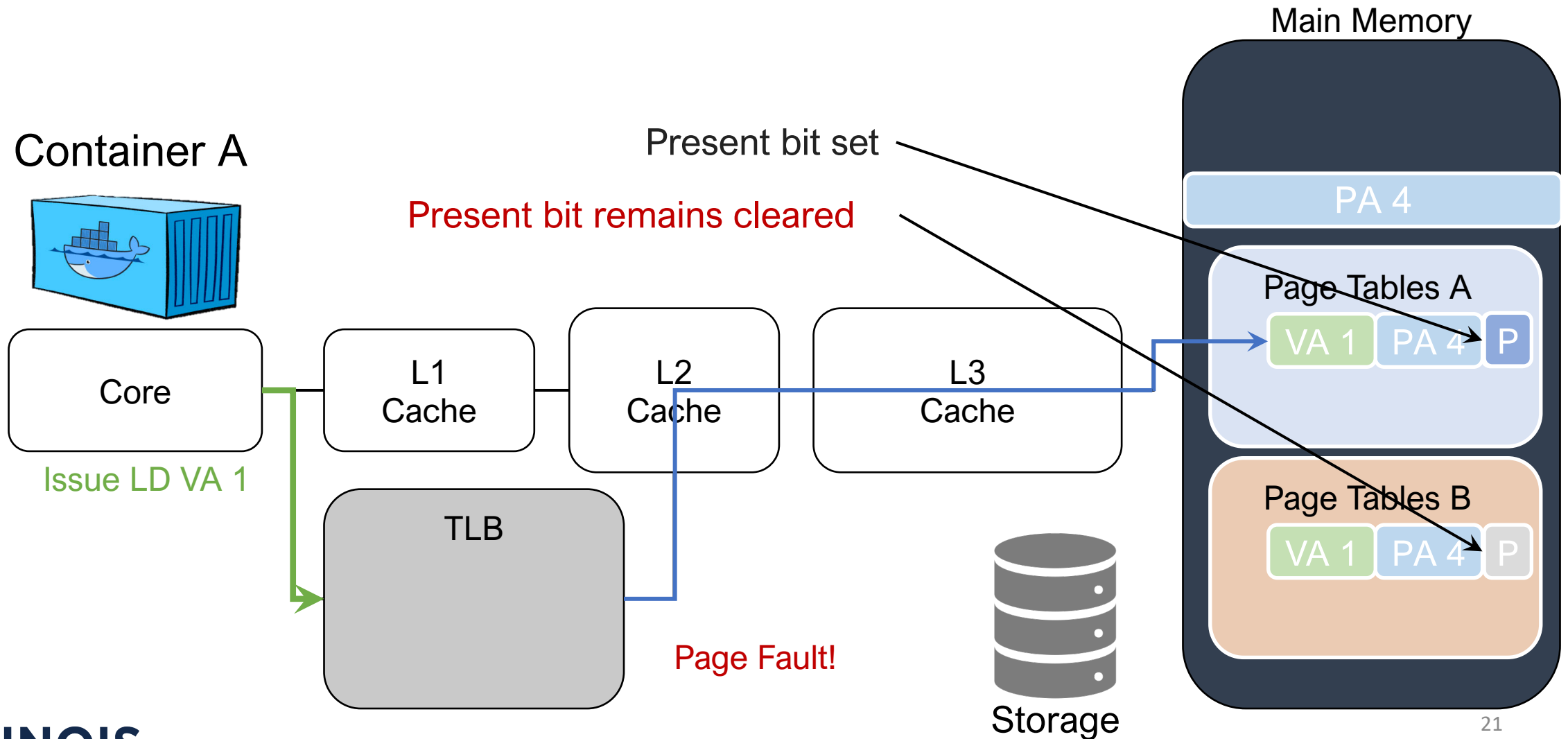
Problem: Lazy Page Table Management



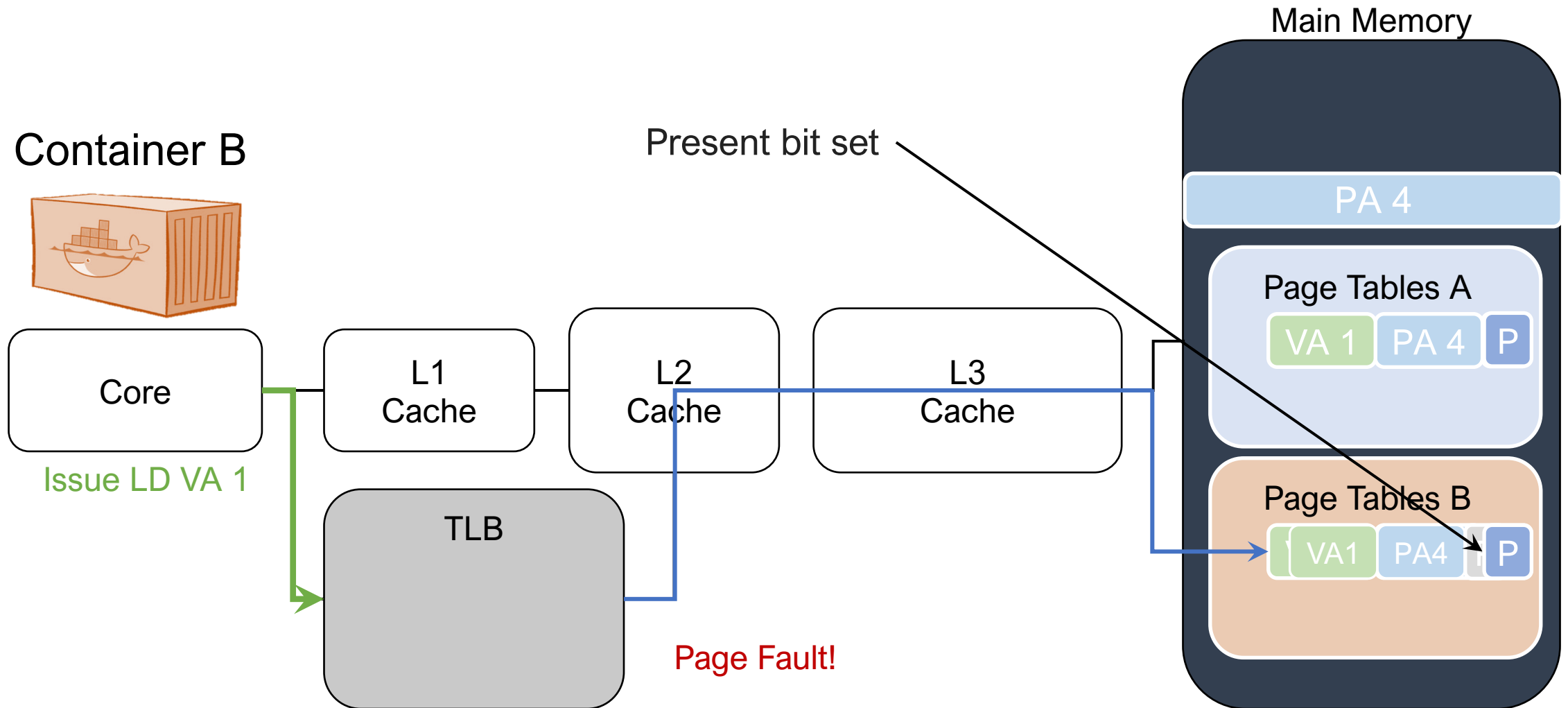
Problem: Lazy Page Table Management



Problem: Lazy Page Table Management

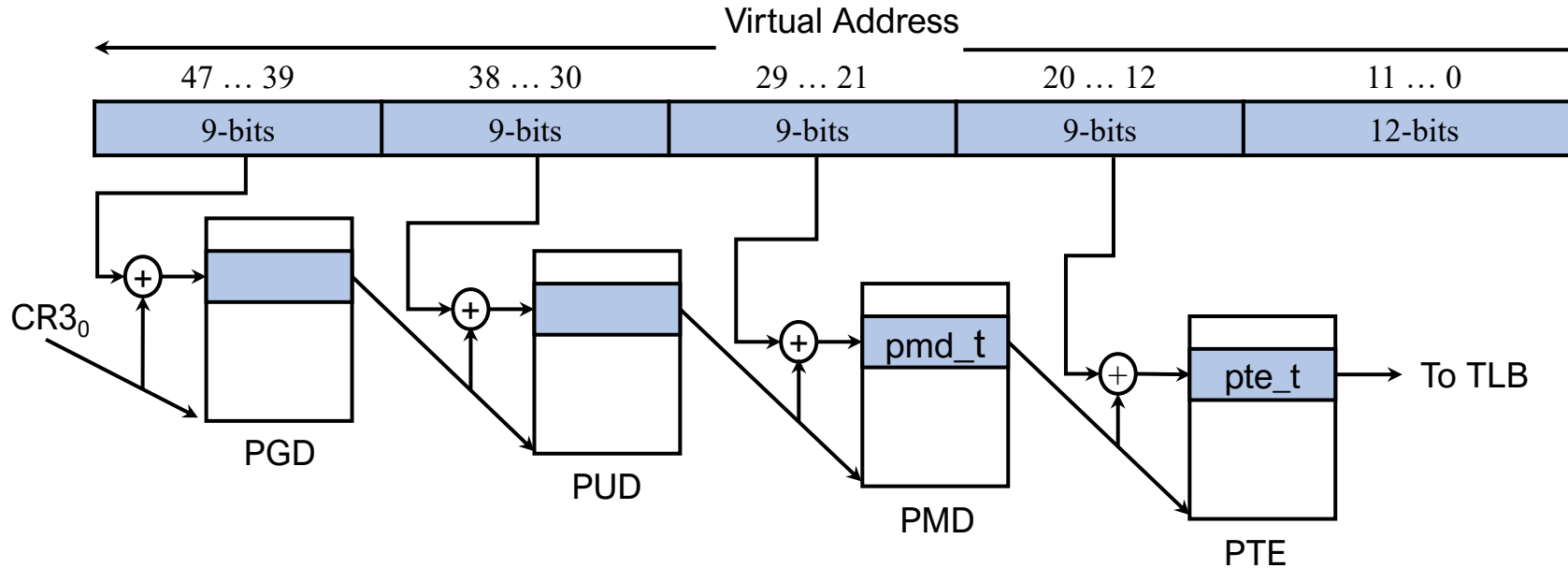
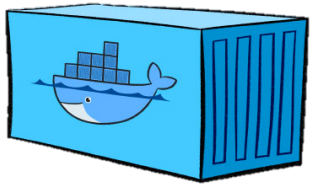


Problem: Lazy Page Table Management



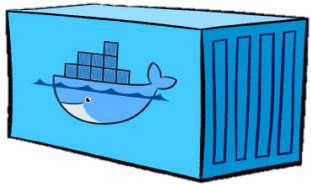
Conventional Page Tables

Container A

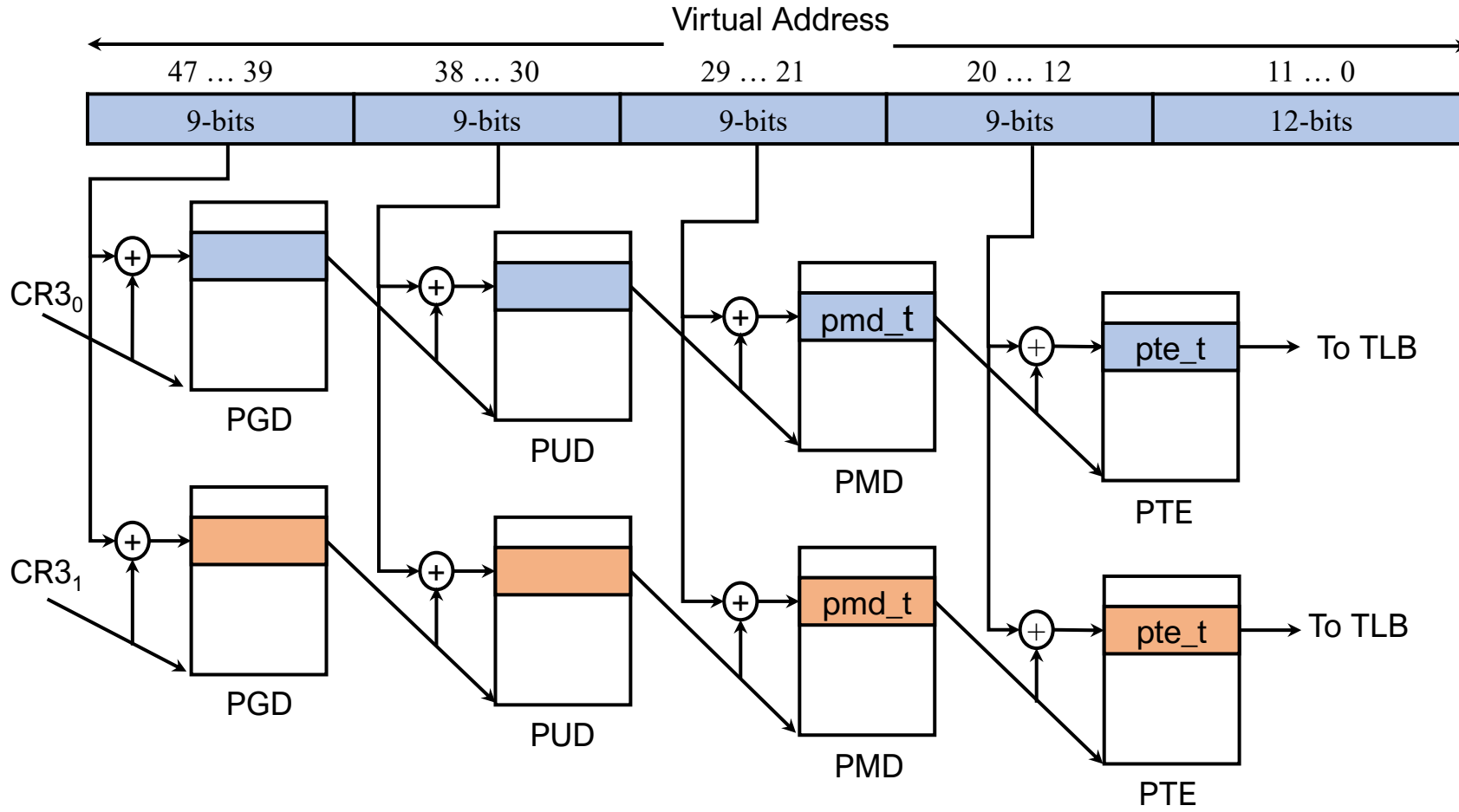
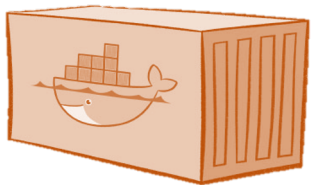


Conventional Page Tables

Container A

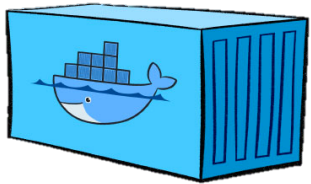


Container B

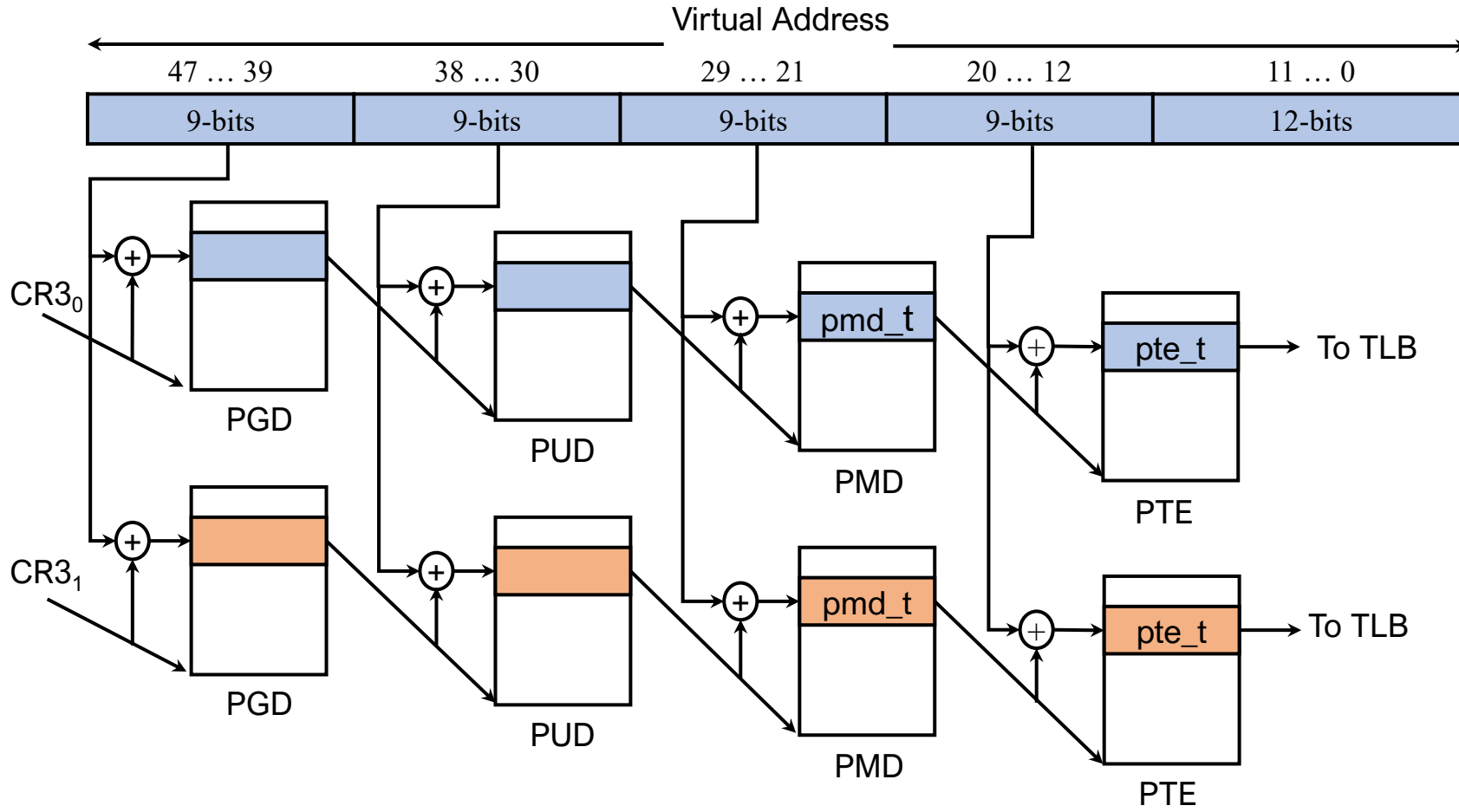
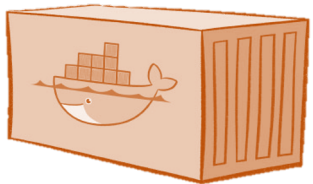


Conventional Page Tables

Container A

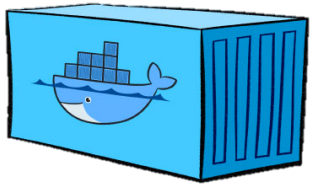


Container B

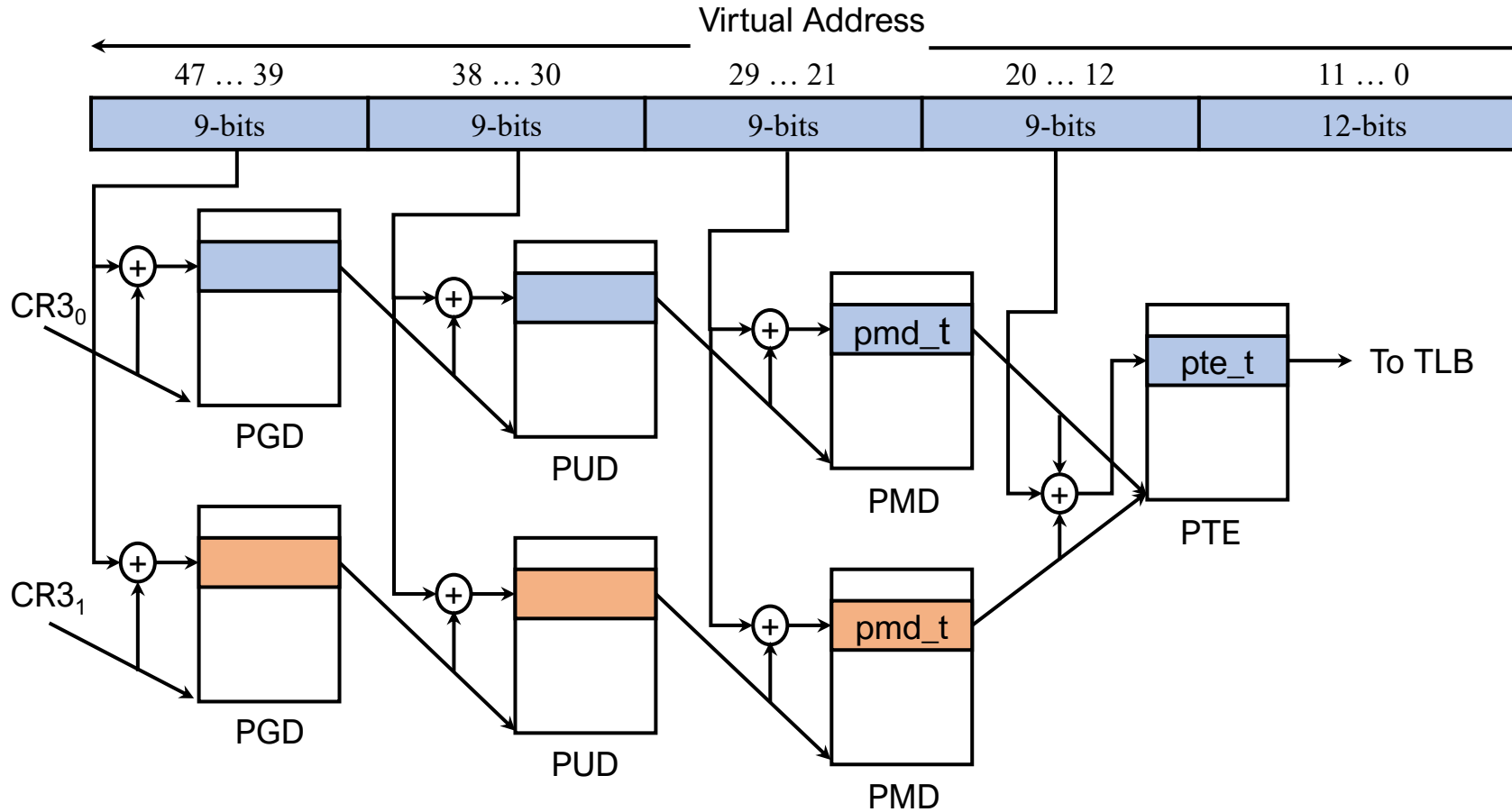
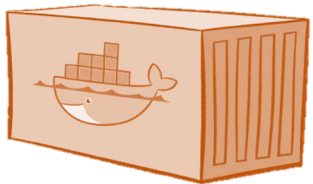


Solution: BabelFish Page Table Sharing

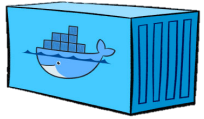
Container A



Container B



Example



Container A: Runs on Core 0, requests $VA \rightarrow PA$

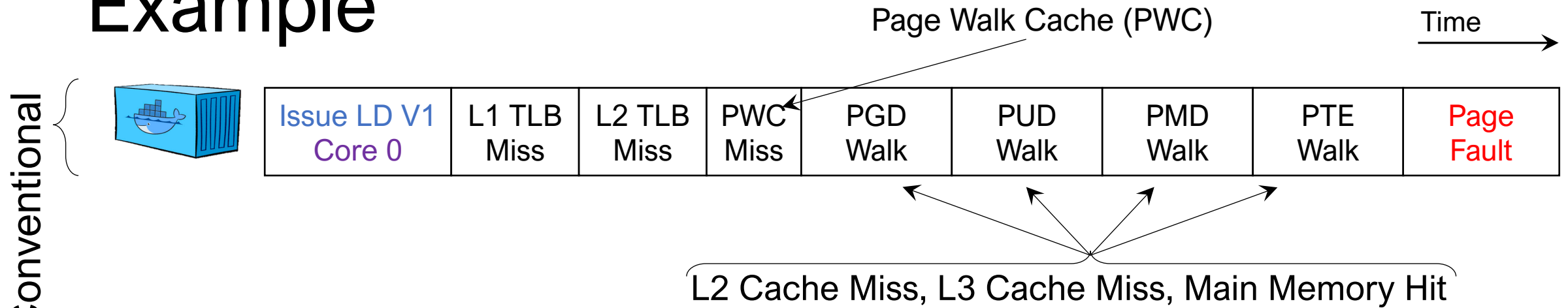


Container B: Runs on Core 1, requests $VA \rightarrow PA$



Container C: Runs on Core 0, requests $VA \rightarrow PA$

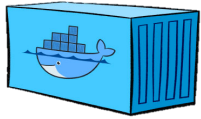
Example



Example

Time →

Conventional



Issue LD V1
Core 0

L1 TLB
Miss

L2 TLB
Miss

PWC
Miss

PGD
Walk

PUD
Walk

PMD
Walk

PTE
Walk

Page
Fault



Issue LD V1
Core 1

L1 TLB
Miss

L2 TLB
Miss

PWC
Miss

PGD
Walk

PUD
Walk

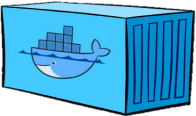
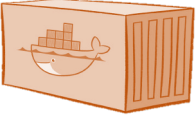
PMD
Walk

PTE
Walk

Page
Fault

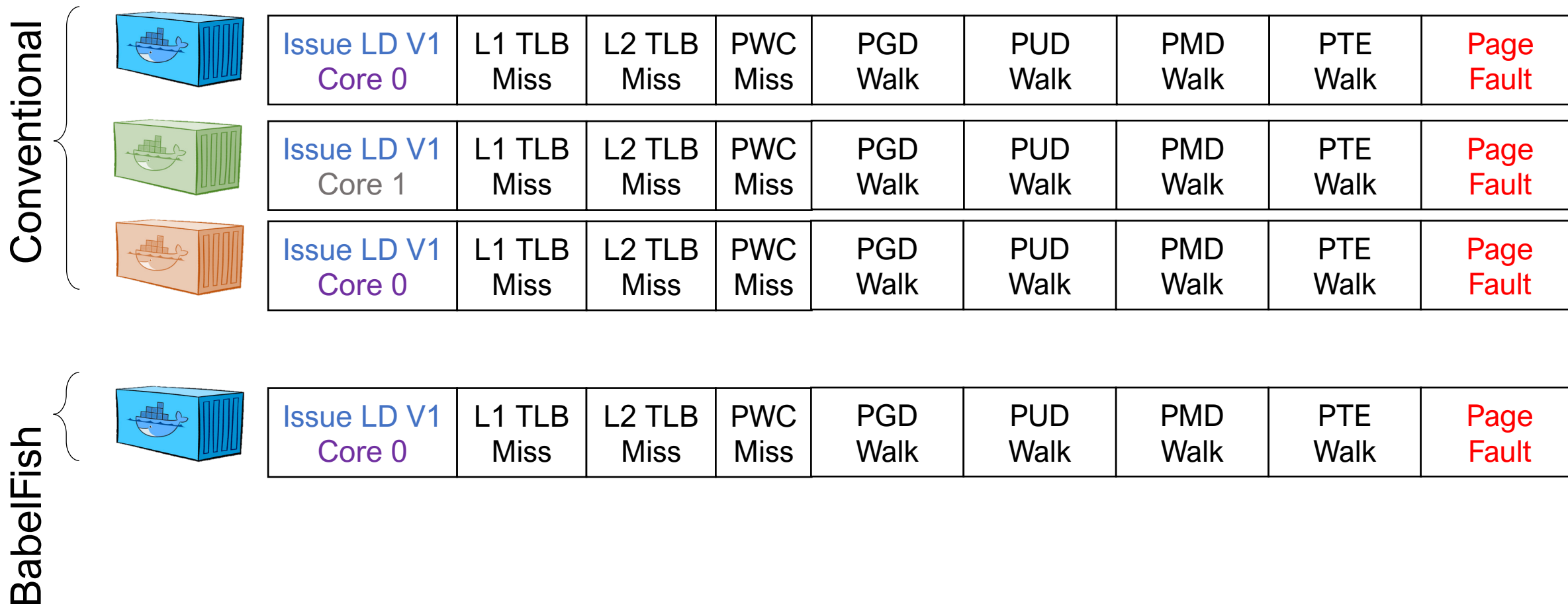
Example

Time →

Conventional		Issue LD V1 Core 0	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault
		Issue LD V1 Core 1	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault
		Issue LD V1 Core 0	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault

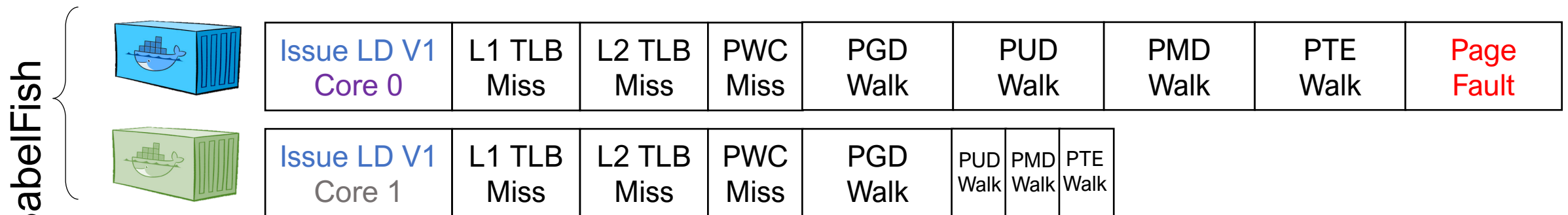
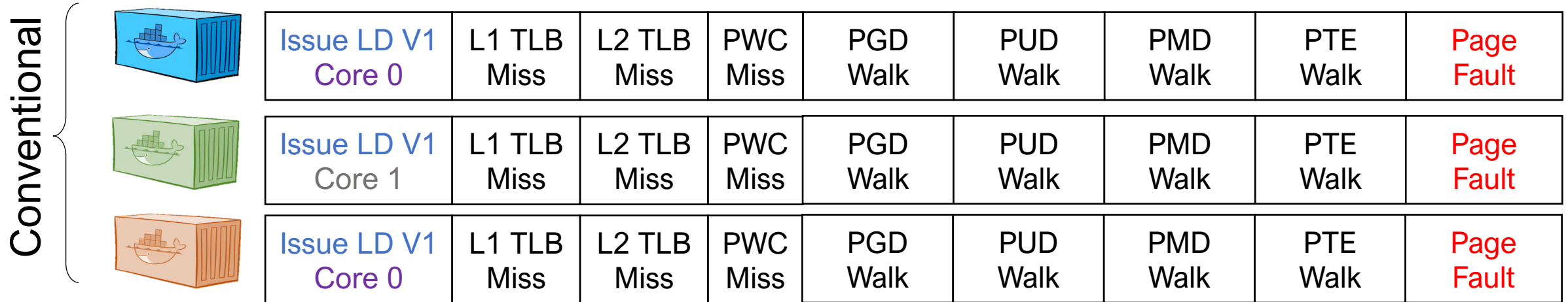
Example

Time →



Example

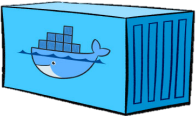
Time →

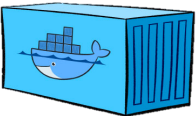


L2 Cache Miss, L3 Cache Hit

Example

Time →

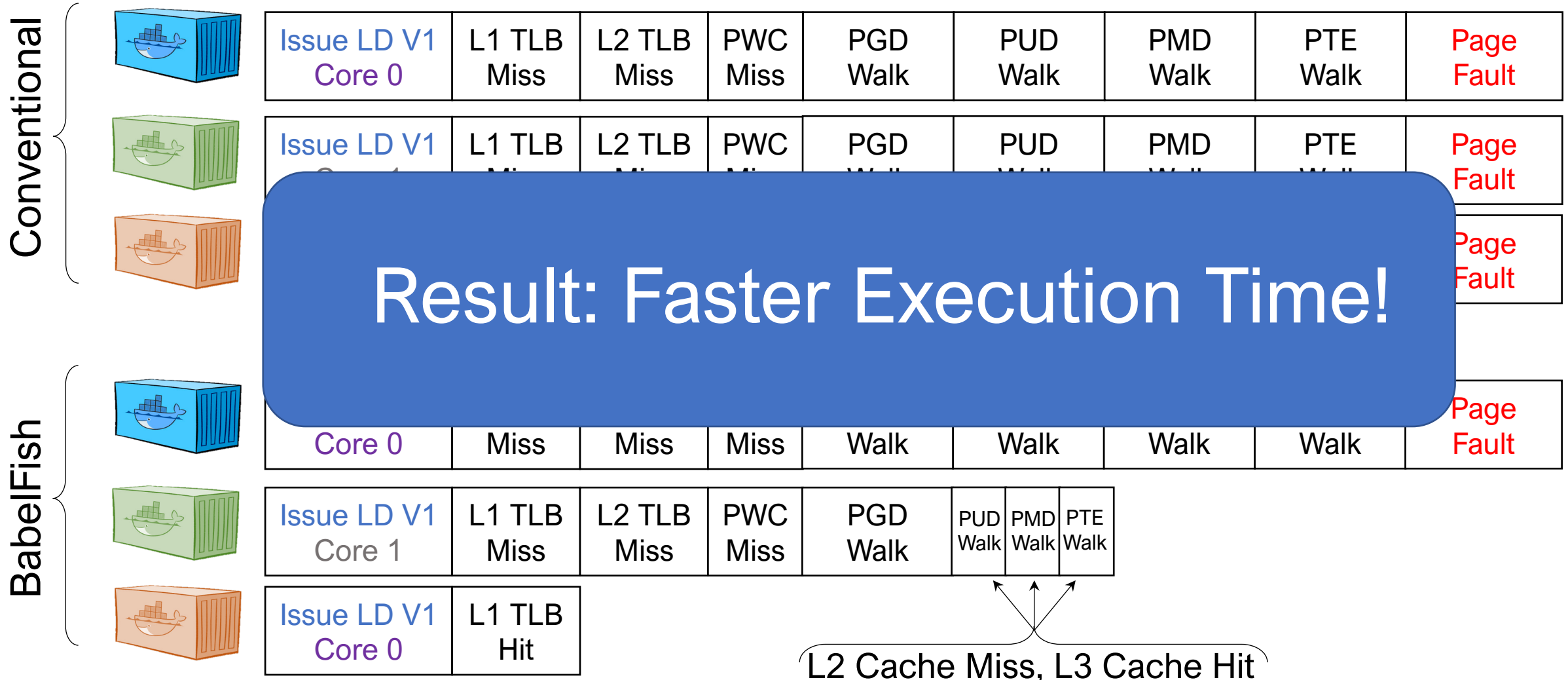
Conventional		Issue LD V1 Core 0	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault
		Issue LD V1 Core 1	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault
		Issue LD V1 Core 0	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault

BabelFish		Issue LD V1 Core 0	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	Page Fault
		Issue LD V1 Core 1	L1 TLB Miss	L2 TLB Miss	PWC Miss	PGD Walk	PUD Walk	PMD Walk	PTE Walk	
		Issue LD V1 Core 0	L1 TLB Hit							

L2 Cache Miss, L3 Cache Hit

Example

Time →



Workloads & Methodology

Workloads:

Data serving: ArangoDB, MongoDB, HTTPd
Compute: GraphChi, FIO

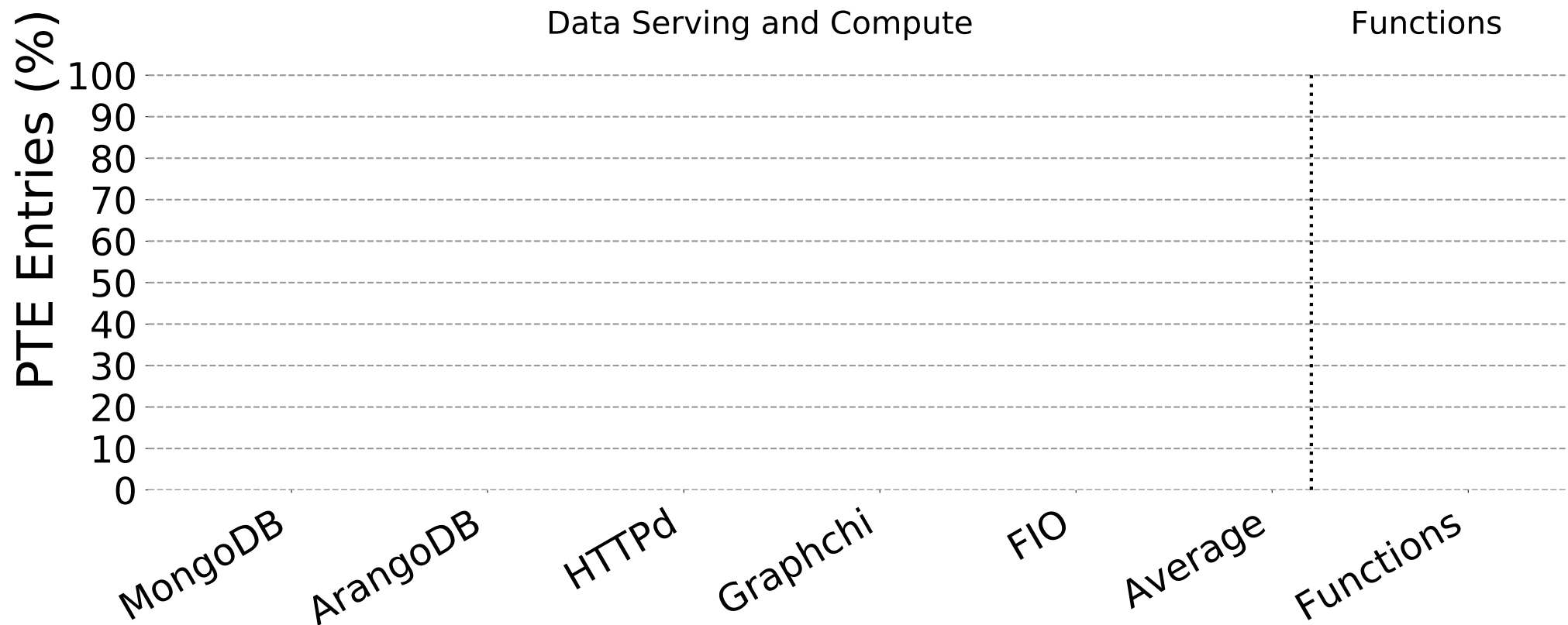
} 2 containers per core

FaaS: Parse, Hash, Marshal (both dense and sparse)

} 3 containers per core

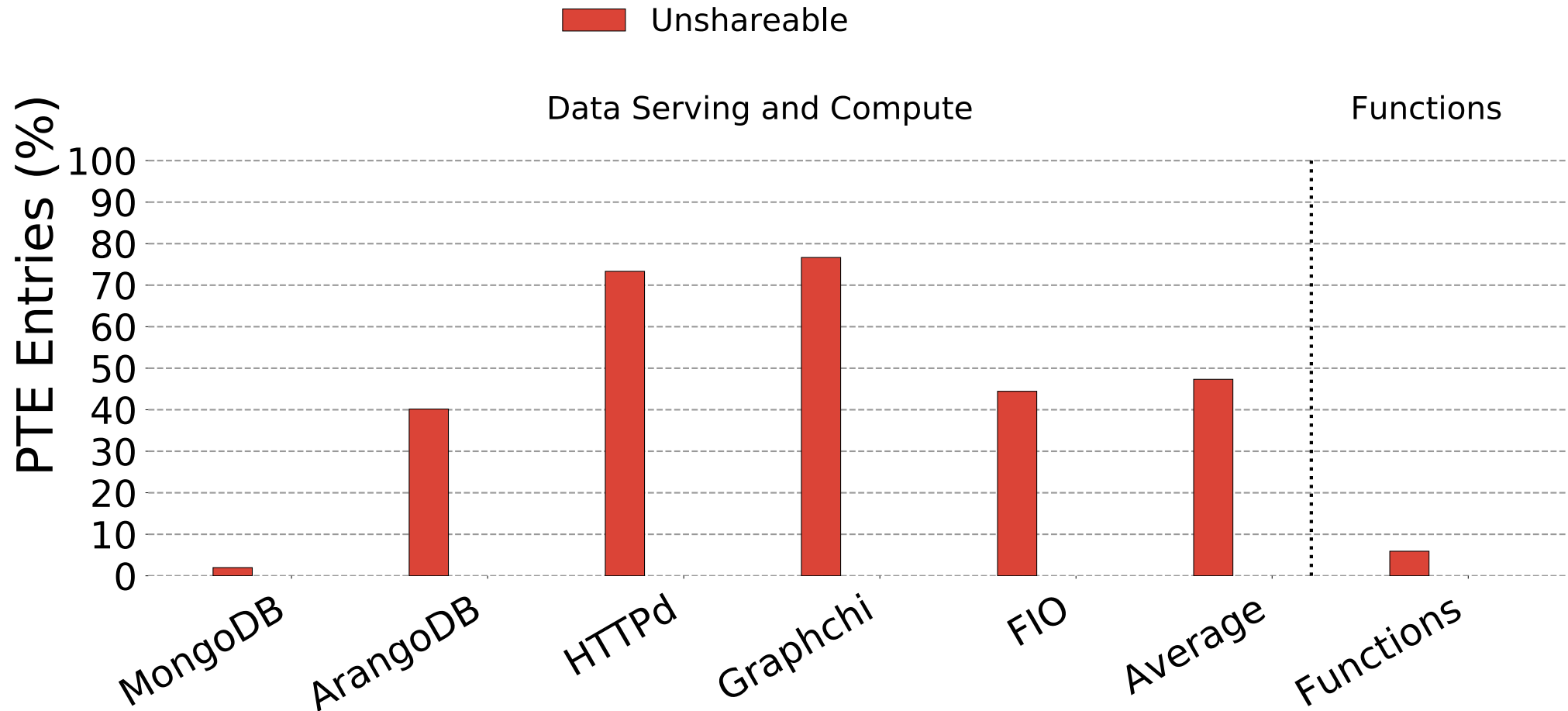
Full system simulations with Simics + SST

Page Table Sharing for Two Containers



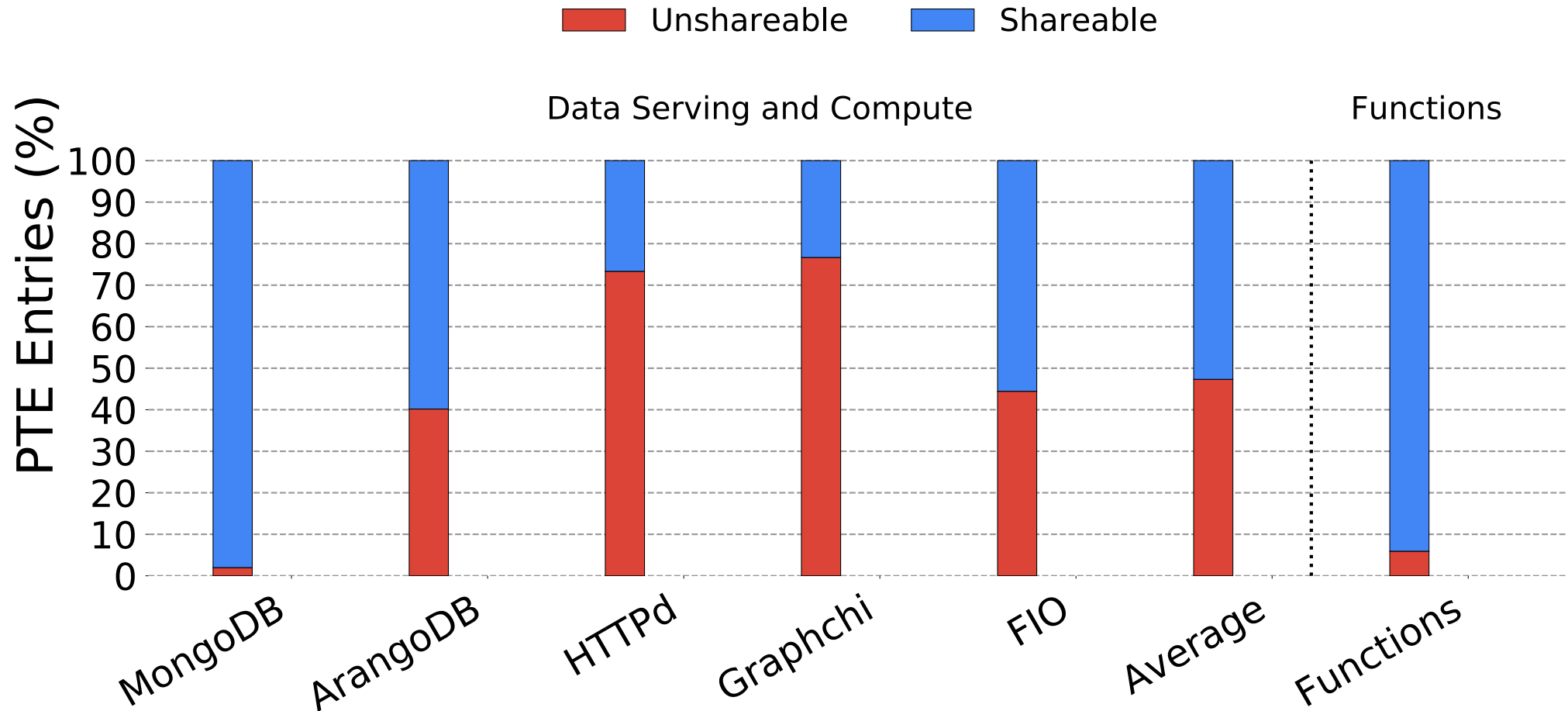
5-minute run on a real machine

Page Table Sharing for Two Containers



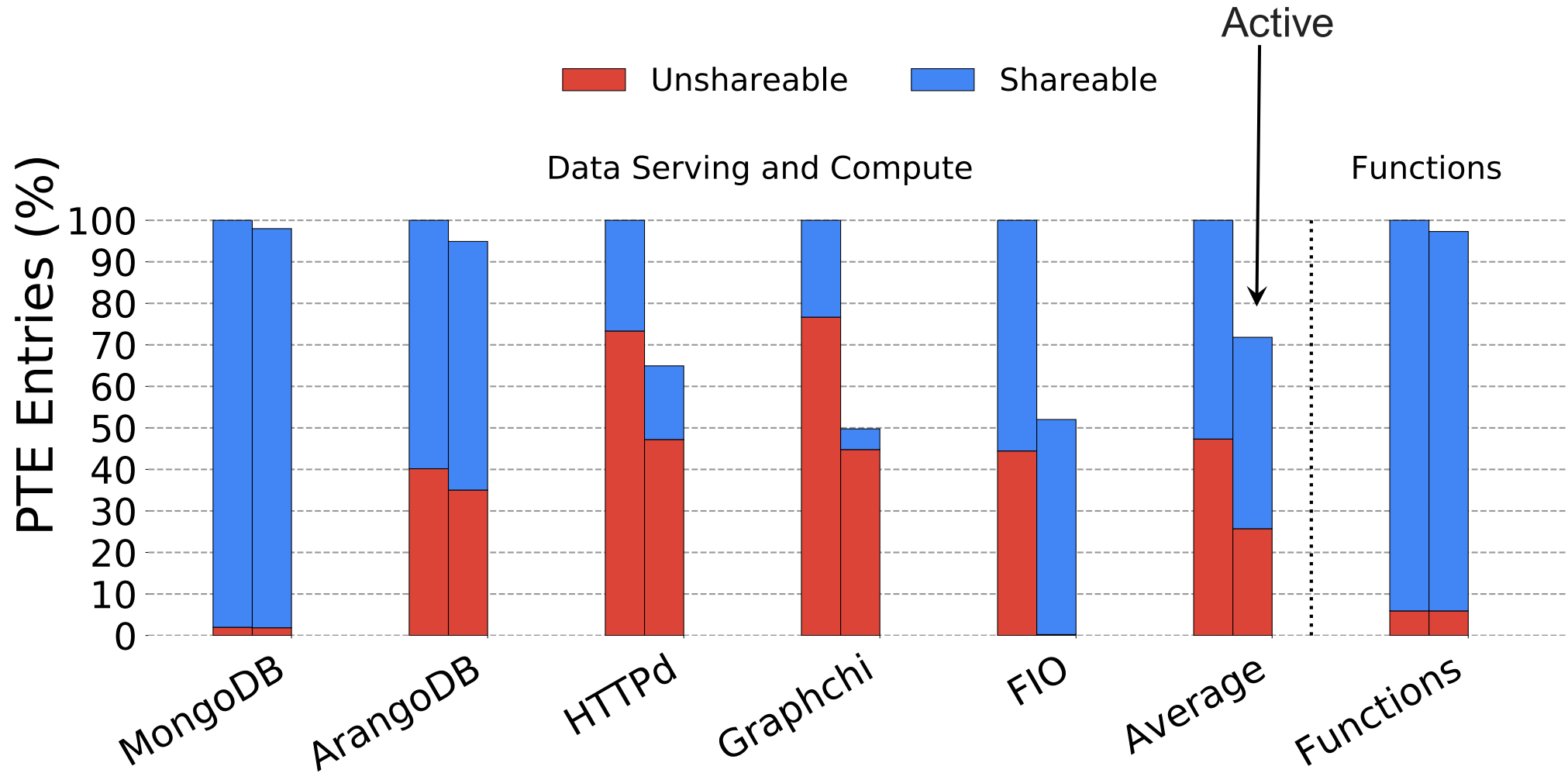
5-minute run on a real machine

Page Table Sharing for Two Containers



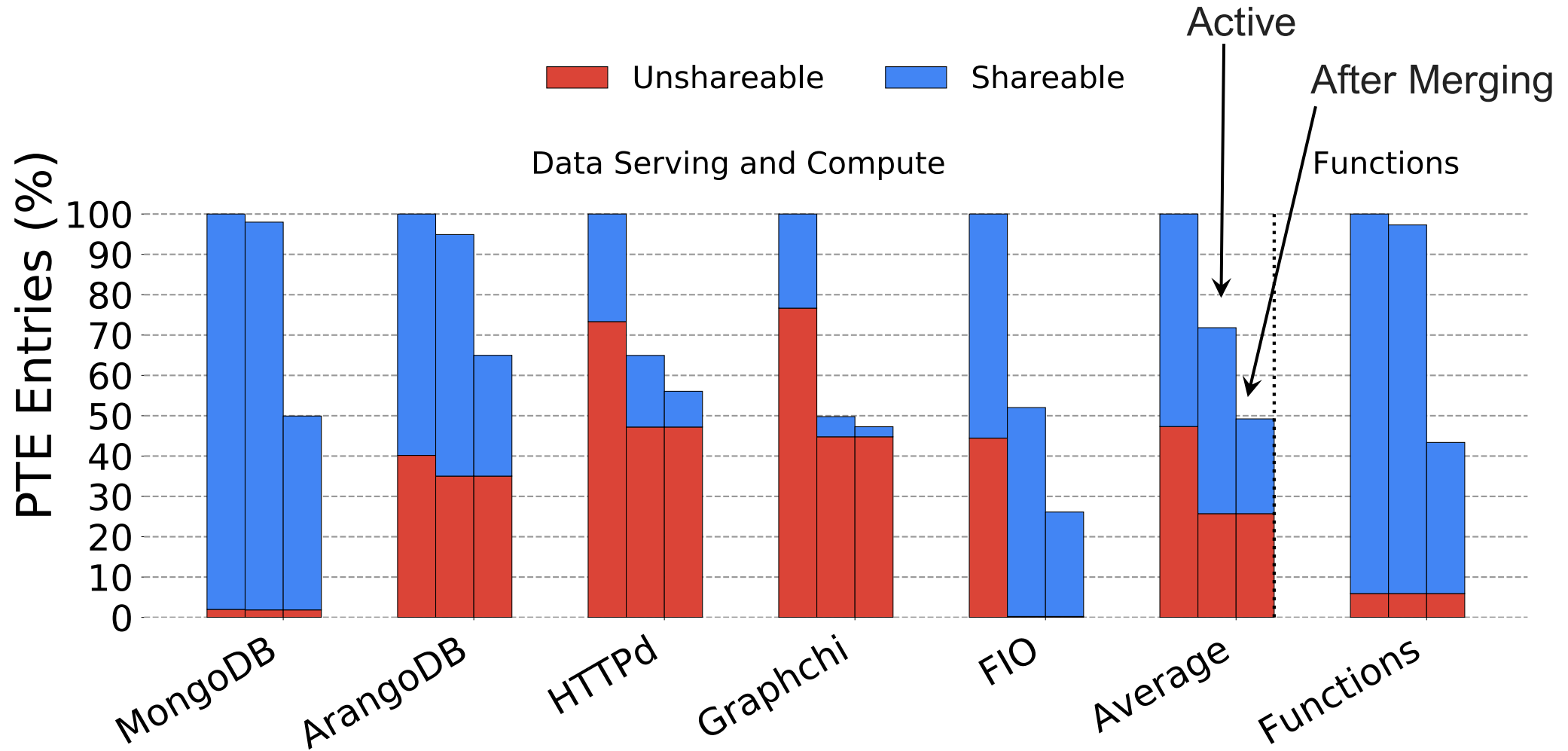
5-minute run on a real machine

Page Table Sharing for Two Containers



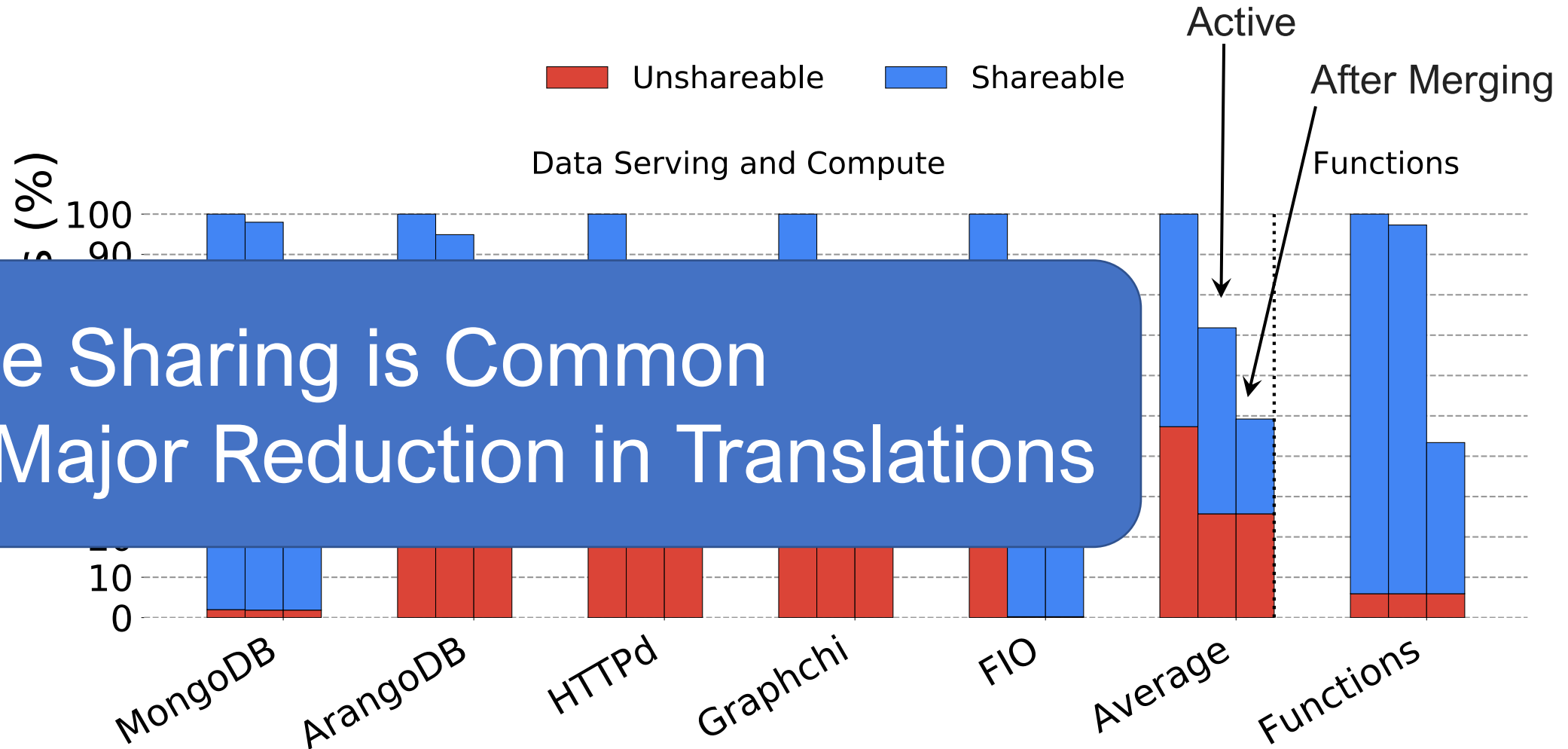
5-minute run on a real machine

Page Table Sharing for Two Containers



5-minute run on a real machine

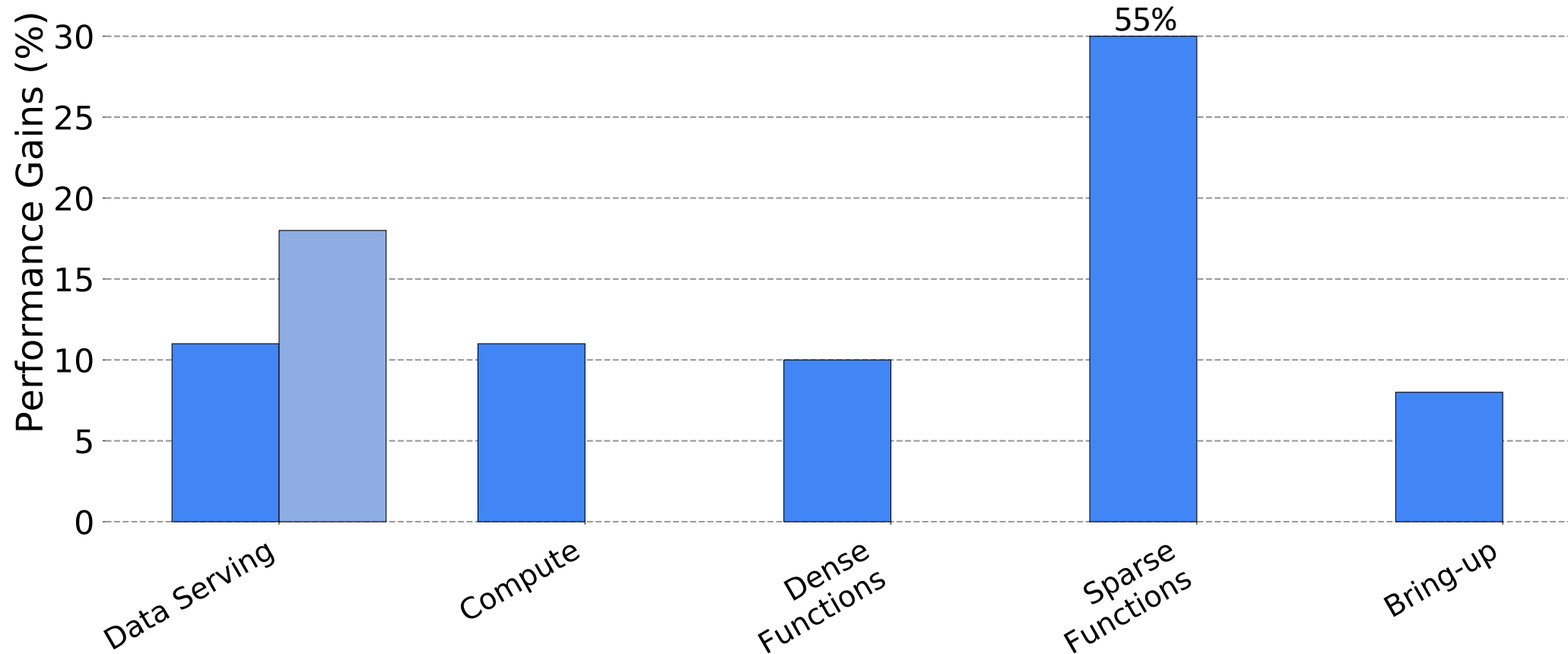
Page Table Sharing for Two Containers



5-minute run on a real machine

Performance

Major Performance Gains
→ Both from TLB and page tables



More in the Paper

Copy-on-Write pages

Address Space Layout Randomization (ASLR) for security

Other security considerations

More evaluation

- Area, energy, sharing, comparison vs larger TLB

Takeaway: BabelFish

Sharing Translations in TLB and Page Tables

Introduces Container Context ID (CCID) at OS and HW

TLB extensions & Page Table sharing support

Substantial speedup in application execution and container bring-up

Future work → CCID in OS and HW

- Virtualized environments and nested page tables
- Other resources

BabelFish: Fusing Address Translations for Containers

Dimitrios Skarlatos, Umur Darbaz, Bhargava Gopireddy,
Nam Sung Kim, and Josep Torrellas

University of Illinois at Urbana-Champaign

skarlat2@illinois.edu

ISCA 2020