

Charliecloud's Git-based Cache is Competitive

Z Noah Hounshel

North Carolina State University
zhounshel@gmail.com

Ashlyn Lee

Colorado State University
ashlynrlee@gmail.com

Benjamin Stormer

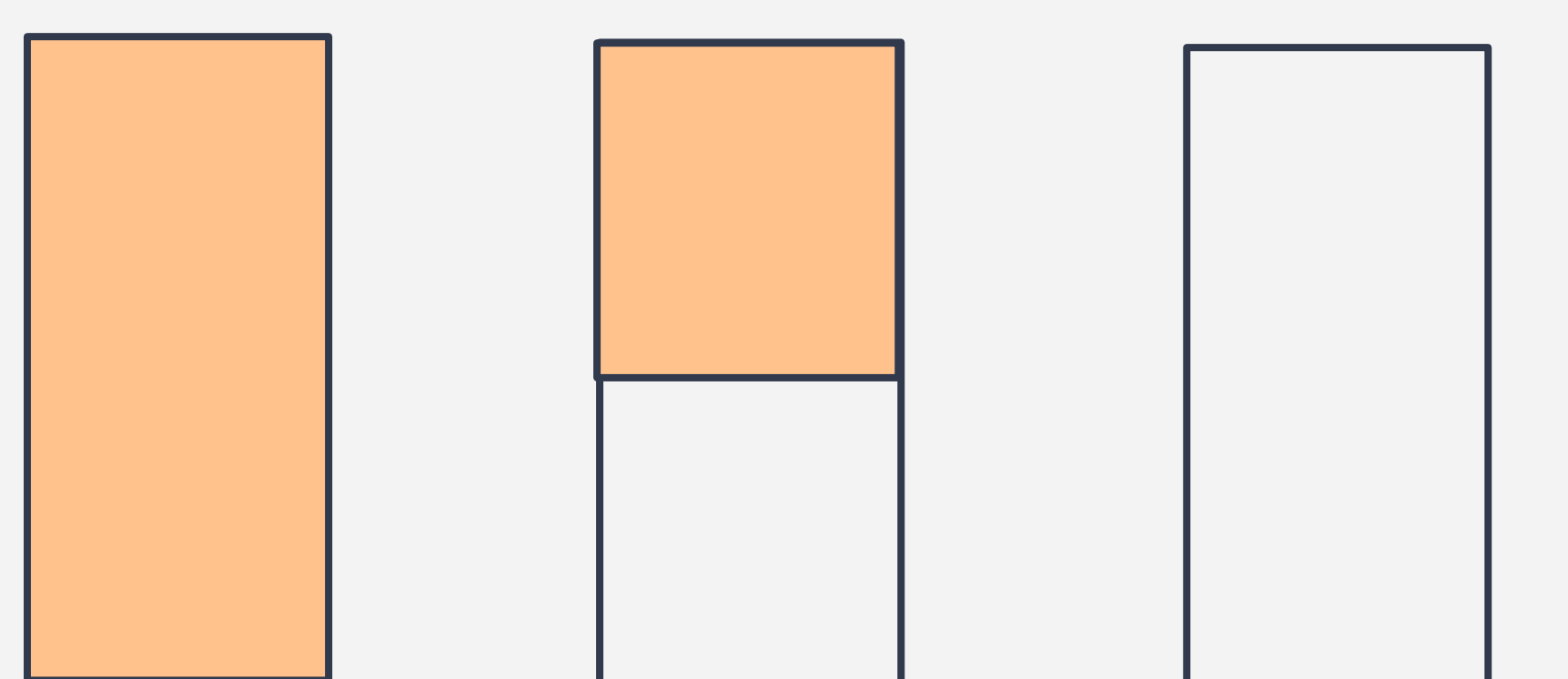
The University of Texas at Austin
benjaminstormer1@gmail.com

Mentors: Jordan Ogas, Reid Priedhorsky

Project Overview

Containers provide flexibility by encapsulating complete software stacks. Container implementations often use build caches to speed up build time. **Charliecloud** is a container implementation that uses a novel **Git-based cache** and is fully unprivileged. Other implementations for building containers are **Docker** and **Podman**, which require varying levels of **privilege** to run and by default use a File System based cache. We measured the build speed of images for each implementation to determine the practicality of Charliecloud's Git-based build cache.

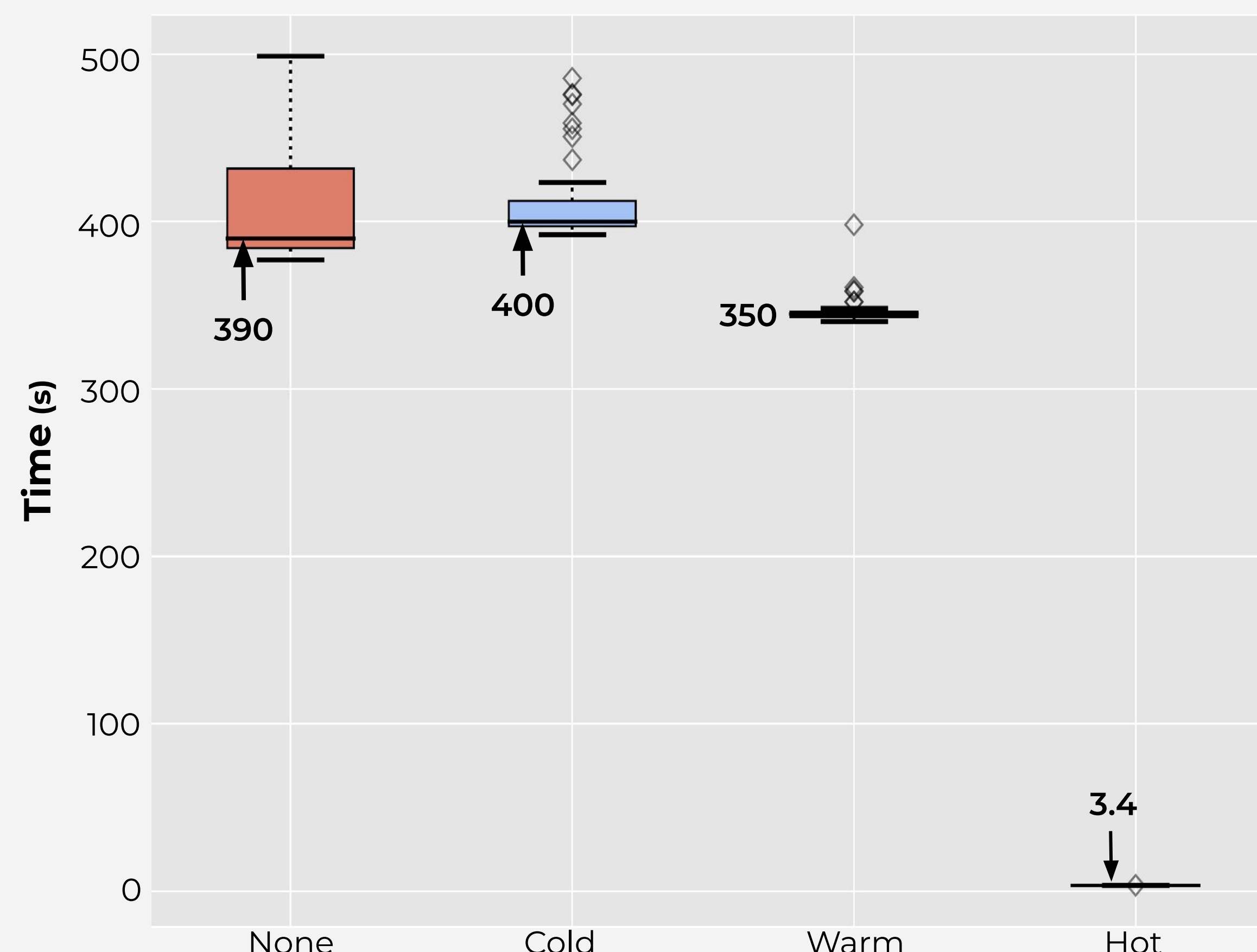
Cache tests



Should Charliecloud continue to develop its build cache?

- Is the Git-based cache:
 - A meaningful improvement over no cache?
 - Competitive with existing container implementations?

OpenMPI - All Caches - Charliecloud



Charliecloud

- LANL-developed container build system
- Uses Git-based caching
- Fully unprivileged**



Testing setup and process

- AMD EPYC 7502 32-Core 128GB
 - Master, 250GB SATA SSD, xfs
 - 10 Compute, storageless, tmpfs

Results

Docker Time/Charliecloud Time (s)

Container Image	Cold		Warm		Hot	
	Docker	Charliecloud	Docker	Charliecloud	Docker	Charliecloud
OpenMPI	440	410	380	350	0.11	3.4
Nvidia	330	430	310	400	0.33	18
Almalinux	110	140	4.5	6.3	0.32	2.8
Debian	7.4	10	6.4	9.3	0.33	0.61
SL	0.66	9.6			0.32	0.98
Atlassian	0.4	9			0.32	4.4

- Results from master node
- Podman/Charliecloud trends are similar
- Podman failed to collect some data

Charliecloud Faster

1±0.5x

1.5-5x

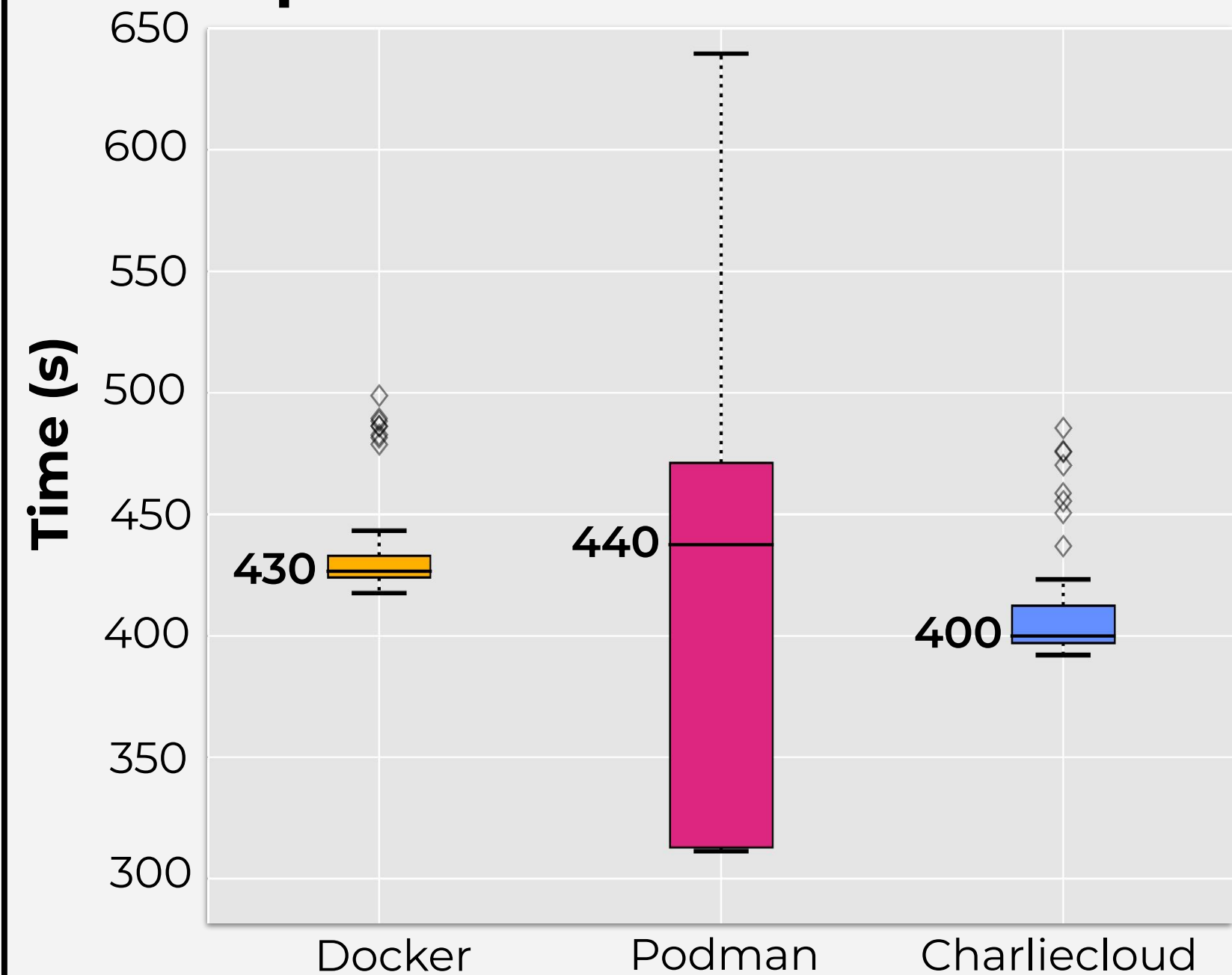
5-10x

>10x

Docker Faster

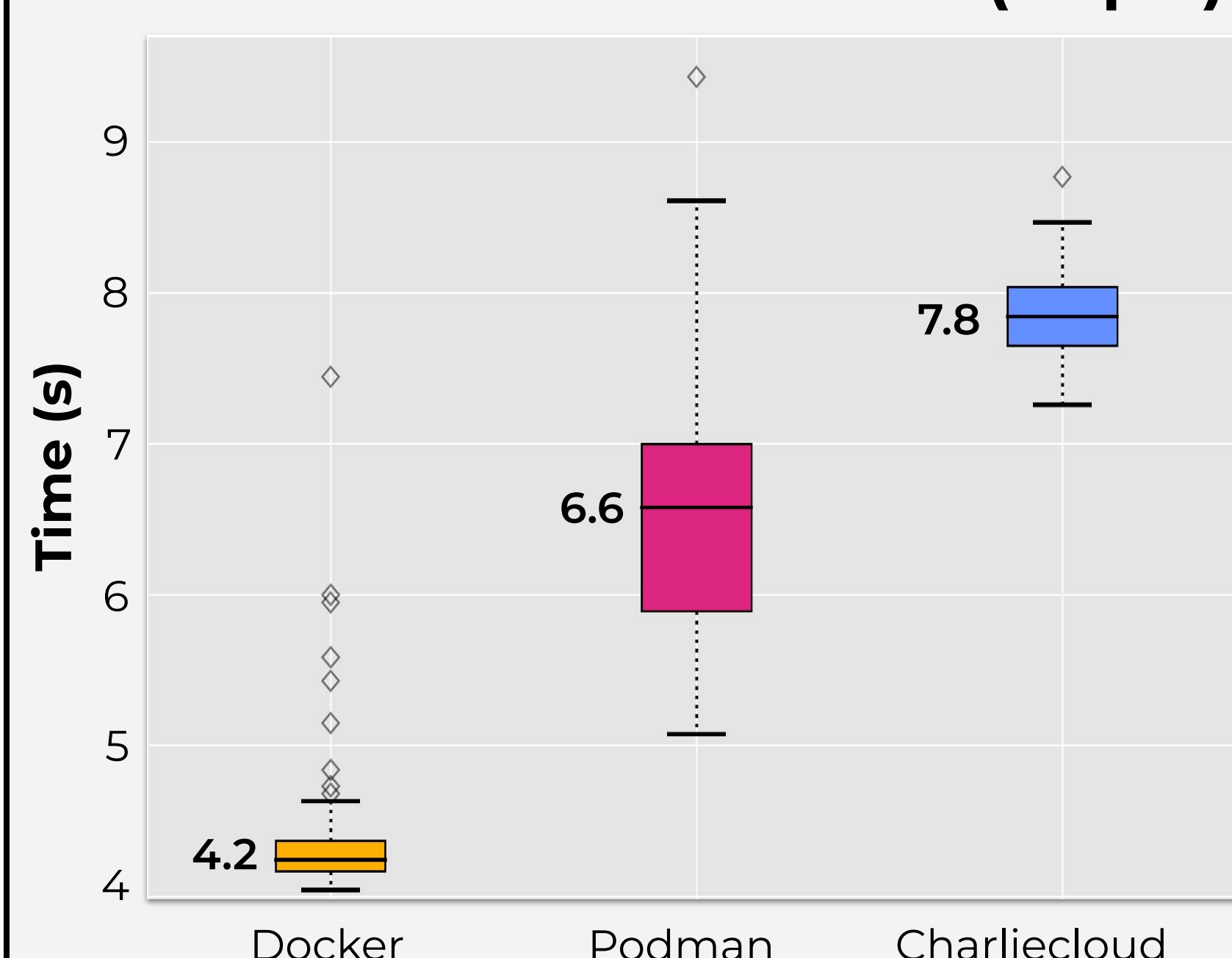
Charliecloud is fastest

OpenMPI - Cold - All Builders



Most common trend

Debian - Cold All Builders (tmpfs)



Non-ideal image for Git cache

Nvidia - All Caches - Charliecloud



Git Caching Works

Assertions

- Charliecloud is slower, but fast enough
- Podman struggles to have consistent results and had frequent crashes.
- Charliecloud is about ~30% slower than docker in longer tests, and only a few seconds slower in hot cache tests.

Future Work

- Why is one particular system faster or slower?
- What parts of the build process are slow or need improvement?
- How does Charliecloud compare to container systems other than Podman and Docker?