



KubeCon



CloudNativeCon

Europe 2026

MUNI

Masaryk
University



Ctrl-X, Ctrl-V Your Pods

WG Checkpoint Restore in Kubernetes

Viktória Spišáková, Radostin Stoyanov, Adrian Reber, Peter Hunt



Red Hat



DevZero



Viktória Spišaková
@viktoriaas



Radostin Stoyanov
@rst0git



Adrian Reber
@adrianreber



Peter Hunt
@haircommander

Announcing the Checkpoint/Restore Working Group
Kubernetes Blog – Jan 21, 2026

Kingdom Come Deliverance II

New Game

Save Game

Load Game

Settings

Help

DLCs & Extras

Credits

Quit

Save & Quit

Resume

version: LL1-11377-release_1.1

PAUSE MENU

RESUME

LOAD CHECKPOINT

SETTINGS

MILESTONES

RESET MISSION

EXIT TO MAIN MENU

EXIT TO DESKTOP



FLATLINED

コードチェック 0.00052

LOAD LAST CHECKPOINT

LOAD GAME

SETTINGS

EXIT TO MAIN MENU



KubeCon



CloudNativeCon

Europe 2026

Checkpoint/Restart (mostly) in User Space

[Post link](#), July 19, 2011

- Exciting for debuggers, tracers, profilers
- Turn OOM killer into OOM dumper
- Security kernel upgrades with limited downtime

LCE: Checkpoint/Restore in User Space: are we there yet?

[Post link](#), November 20, 2012

"Although many groups are interested in having checkpoint/restore functionality, an implementation that works with the mainline kernel has taken a long time in coming. When one looks into the details and realize how complex the task is, it is perhaps unsurprising that it has taken so long."

Motivation and Use Cases

Many companies are creating their own software stacks around Checkpoint/Restore

Increased focus on GPU resources

- Expensive
- May be scarce (in a cluster)
- In high demand

Enable upstream Kubernetes-native support for Checkpoint/Restore

Accelerating Application Startup

- Useful for containers that have a long startup time
- With C/R it is possible to transparently save initialized state for future starts
 - One or more copies of the container can be created without the need to reinitialize
- Already used in combination with AI Agents, persistent memory systems and Java apps
 - AI systems are market drivers

[Zeroing and the semantic gap between host and guest](#). FOSDEM(2024)

[Engine-Agnostic Model Hot-Swapping for Cost-Effective LLM Inference](#). CANOPIE-HPC (2025)

... research ideas finding their ways into production-ready inference serving frameworks

[Dynamo: Introducing ChReK \(Checkpoint Restore in K8s\)](#)

Fault Tolerance

- Periodic checkpointing for long-running workloads
 - Avoid loss of already done work
 - Avoid recomputations
- Not all applications implement recovery mechanisms or interim state saving
 - All sorts of workloads benefit from infrastructure-provided fault tolerance
- Integrations with planned maintenance, advance-detection mechanisms, spot instances

[The Party Must Go on-Resume Pods After Spot Instance Shutdown](#). KubeCon EU (2024)

[Just-In-Time Checkpointing: Low Cost Error Recovery from DL Training Failures](#). EuroSys (2024)

[Transparent, Infra-Level Checkpoint and Restore for Resilient AI/ML Workloads at Scale](#). KubeCon EU (2025)

Optimizing Resource Utilization

- Inefficient CPU and GPU utilization is a persistent issue of cloud infrastructures
- Checkpoint/Restore as a primitive for
 - Load balancing
 - Non-disruptive preemptions of lower-priority workloads
 - Resource reclamation from (idle) stateful workloads
- Integration with cluster autoscaler and scheduler

[Task Migration at Scale Using CRIU](#). LPC (2018)

[Singularity: Planet-Scale, Preemptive and Elastic Scheduling of AI Workloads](#). arXiv (2022)

[Optimizing Resource Utilization for Interactive GPU Workloads With Transparent Container Checkpointing](#). FOSDEM (2025)

[Kubernetes Scheduling with Checkpoint/Restore: Challenges and Open Problems](#). JSSPP (2025)

Checkpoint/Restore in Kubernetes

Kubelet Checkpoint API

 **FEATURE STATE:** Kubernetes v1.30 [beta](enabled by default)

kubernetes.io/docs/reference/node/kubelet-checkpoint-api

Container Checkpointing in Kubernetes

KEP-2008: Forensic Container Checkpointing

- 2020 - 2021: Initial KEP and prototype created
- Aug 2021: Introduced phased implementation plan forensics user story
- Sep 2021: Narrowed scope & added ContainerCheckpoint feature gate
- May 2022: Added CRI API & restore RPC removed (v1.25 target)
- Feb 2024: Graduated to Beta (v1.30)

KubeCon Europe

- 2024: Enabling Coordinated Checkpointing for Distributed HPC Applications
- 2025: Efficient Transparent Checkpointing of AI/ML Workloads in Kubernetes
- 2026: Optimizing Error Recovery for Cost-Efficient Distributed AI Model Training

KEP-5823: Pod-Level Checkpoint/Restore

Initial Proposal

[WIP] Add --checkpoint to drain #97194

Closed adrianreber wants to merge 29 commits into `kubernetes:master` from `adrianreber:2020-12-10-drain--checkpoint`

Conversation 26 · Commits 29 · Checks 0 · Files changed 291

adrianreber commented on Dec 10, 2020

What type of PR is this?

/kind feature

What this PR does / why we need it:

This PR implements `--checkpoint` on drain. This means that Pods can be checkpointed instead of being killed when draining a node, so that the Pods can be restored later with the same state. This is especially, maybe only, interesting for stateful containers which need a long time to start up (java), need to load a lot of data from storage (database), or for other stateful containers without a storage backend.

The basic steps to use this would look something like this:

- `kubectl drain 127.0.0.1 --checkpoint`
- `reboot`
- Restart kubelet (and restore all checkpointed containers)

This PR does not yet implement the restore part. The corresponding CRI-O implementation to support checkpoint/restore already implements checkpointing and restoring of Pods, but it has not been added to this PR, yet.

Add 'checkpoint' command to kubectl #120898

Closed adrianreber wants to merge 1 commit into `kubernetes:master` from `adrianreber:2023-09-26-kubectl-checkpoint`

Conversation 25 · Commits 1 · Checks 0 · Files changed 27

adrianreber commented on Sep 26, 2023 · edited

What type of PR is this?

/kind feature

What this PR does / why we need it:

Kubernetes 1.25 introduced the possibility to checkpoint a container.

For details please see the KEP 2008 Forensic Container Checkpointing [kubernetes/enhancements#2008](#)

The initial implementation only provided a kubelet API endpoint to trigger a checkpoint. The main reason for not extending it to the API server and kubectl was that checkpointing is a completely new concept.

Although the result of the checkpointing, the checkpoint archive, is only accessible by root it is important to remember that it contains all memory pages and thus all possible passwords, private keys and random numbers. With the checkpoint archive being only accessible by root it does not directly make it easier to access this potentially confidential information as root would be able to retrieve that information anyway.

Now, at least three Kubernetes releases later, we have not heard any negative feedback about the checkpoint archive and its data. There were, however, many questions to be able to create a checkpoint via kubectl and not just via the kubelet API endpoint.

This commit adds 'checkpoint' support to kubectl. The 'checkpoint' command is heavily influenced by the code of the 'exec' and 'logs' command. The checkpoint command is only available behind the 'alpha' sub-command as the "Forensic Container Checkpointing" KEP is still marked as Alpha.

- `POST /api/v1/namespaces/{ns}/pods/{name}/checkpoint`
- `POST /api/v1/namespaces/{ns}/pods/{name}/restore`

KEP-5823: Pod-Level Checkpoint/Restore

Goal: Introduce support for transparent checkpoint/restore of running Pods

Preliminary proposal scope

- CheckpointPod & RestorePod CRI APIs and kubelet APIs
- Namespaced PodCheckpoint and PodRestore Objects
- PodCheckpoint and PodRestore Controllers (TBD)

Out of scope: Lifecycle Management , Storage Backends, Shared Resources, IP Migration

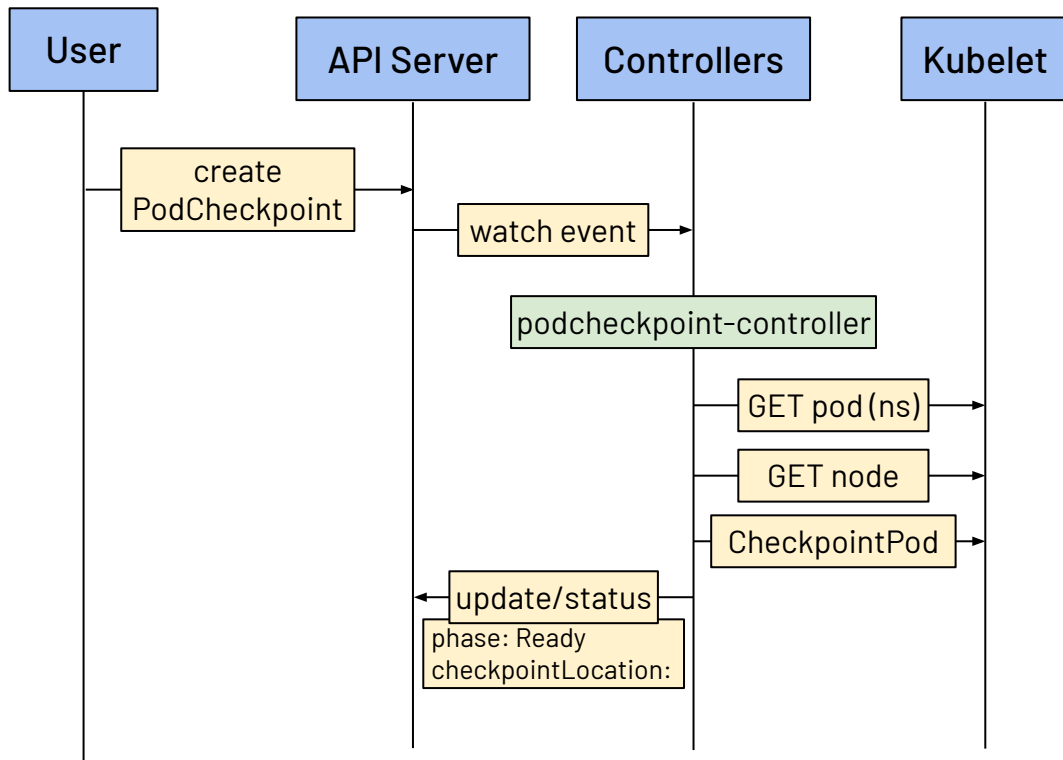
Proof of Concept Implementations:

- github.com/rst0git/pod-snapshot-controller
 - github.com/adrianreber/kubernetes/commits/2026-03-09-declarative-podrestore-api
- Working group development fork of Kubernetes: github.com/checkpoint-restore/kubernetes

KEP-5823: Pod-Level Checkpoint/Restore

Provisional PodCheckpoint Object

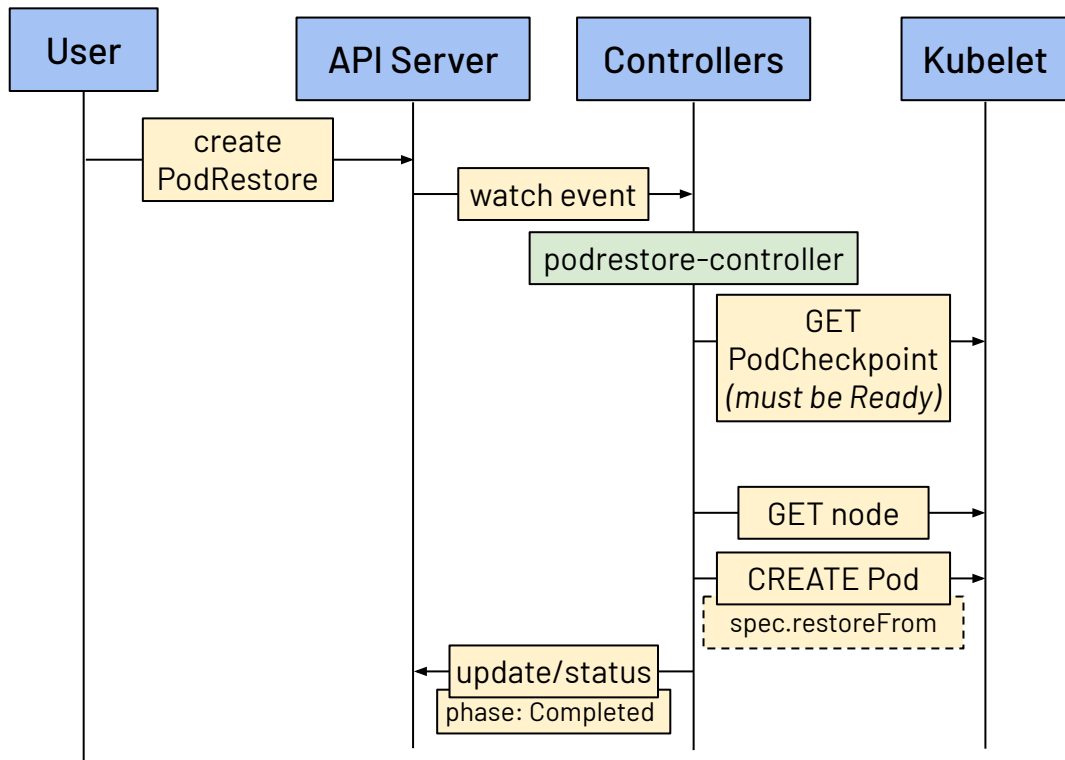
```
apiVersion: checkpoint.k8s.io/v1alpha1
kind: PodCheckpoint
metadata:
  name: my-checkpoint
  namespace: default
spec:
  sourcePodName: my-pod
status:
  phase: Ready
  nodeName: worker-1
  checkpointLocation: foo-bar
```



KEP-5823: Pod-Level Checkpoint/Restore

Provisional Restore Request (PodRestore) Object

```
apiVersion: checkpoint.k8s.io/v1alpha1
kind: PodRestore
metadata:
  name: restore-1
  namespace: default
spec:
  checkpointName: my-checkpoint
  podTemplate:
    metadata:
      labels:
        app: restored
    spec:
      containers:
        - name: app
          image: ignored-on-restore
status:
  phase: Completed
  restoredPodName: restored-pod-xyz
```



KEP-5823: Pod-Level Checkpoint/Restore

Provisional Restore Request (PodRestore) Object - Initial Prototype

```
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl apply -f counter-pod.yaml
pod/counter-app created
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl get pods
NAME          READY   STATUS    RESTARTS   AGE
counter-app   1/1     Running   0           5s
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl logs counter-app
count: 0
count: 1
count: 2
count: 3
count: 4
count: 5
count: 6
count: 7
count: 8
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl apply -f checkpoint.yaml
podcheckpoint.checkpoint.k8s.io/counter-checkpoint created
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl get podcheckpoints.checkpoint.k8s.io
NAME          SOURCE POD PHASE    NODE
counter-checkpoint counter-app Ready    legion
[radostin:~/Projects/ ... 5823/pod-snapshot-controller/examples] main ± kubectl delete -f counter-pod.yaml
pod "counter-app" deleted from default namespace
```

[Pod-Level Checkpoint/Restore: Architecture and Design](#)

[RBAC & API Workflow Walkthrough](#)



Photo by Outward Bound Costa Rica on Unsplash

Open Questions: WG Checkpoint Restore

- What use cases?
 - Live Migration
 - Warm Start
 - Suspend/Resume

Open Questions: SIG Node

- How closely should a checkpoint be related to a pod?
- How do we cover multiple checkpointing implementations?
- Are there future integrations with evictions?

Open Questions: SIG Security

- Who may checkpoint a pod?
- Who may restore a checkpoint?
- What granularity should policy engines have access to?

Open Questions: SIG Storage

- How do we move checkpoints between nodes
- How do we specify paths to keep them?
- Who should do the moving?
- How do we store to different storage mediums?

Open Questions: SIG Apps

- How do we checkpoint higher level objects?
- What about objects with special checkpointing behavior
 - E.g. kubevirt, pytorch

Open Questions: SIG Network

- How can we save IP Address?

Get Involved / Learn More

Weekly Meetings on Thursdays at 17:00 (UTC)

Mailing list: kubernetes.io/g/wg-checkpoint-restore

Slack channel: [#wg-checkpoint-restore](#)

Development Repository: github.com/checkpoint-restore/kubernetes

CRIU Docs: criu.org/Kubernetes



Optimizing Error Recovery for Cost-Efficient Distributed AI Model Training with Kubernetes - Radostin Stoyanov, University of Oxford & Andrey Velichkevich, Apple; Viktória Spišáková, Masaryk University



Click here to remove from My Schedule.

📅 Thursday March 26, 2026 14:30 - 15:00 CET

📍 Elicium 2



KubeCon



CloudNativeCon

Europe 2026



This session's feedback form

