

**SEAWAY Edge**

**内核热升级用户指南**

中科海微（北京）科技有限公司

Seaway Technologies Co. Ltd

## **目录**

[目录 1](#_Toc395252437)

[文档修订目录 1](#_Toc840788488)

[文档版本 1](#_Toc689720556)

[Version3.2 1](#_Toc16658186)

[第一章 安装与部署 4](#_Toc801257992)

[1.1软硬件要求 4](#_Toc2020604854)

[1.2环境准备 4](#_Toc2146871167)

[1.3安装内核热升级工具 4](#_Toc443466875)

[1.4部署内核热升级工具 5](#_Toc1579513035)

[1.5使能内核热升级工具 6](#_Toc1830218678)

[第二章 使用方法 7](#_Toc2077045165)

[2.1 命令用法 7](#_Toc1551406170)

[2.2 使用限制 7](#_Toc1884540963)

[2.3 加速特性说明及使用 8](#_Toc243655538)

[2.4 产生的日志信息 8](#_Toc2014795984)

[第三章 常见问题与解决方法 9](#_Toc1153957192)

# 文档修订目录

## 文档版本

## Version3.2

|  |  |  |  |
| --- | --- | --- | --- |
| 文档版本号 | 修订日期 | 修订内容 | 修订人 |
| V3.1 | 2023/04/29 | 初订 | 陈海鑫 |

**前言**

在使用本手册之前，请您认真阅读以下使用许可协议，只有在同意以下使用许可协议的情况下方能使用本手册中介绍的产品。

**版权声明**

中科海微（北京）科技有限公司版权所有，并保留对本文档及本声明的最终解释权和修改权。本文档中出现的任何文字叙述、文档格式、插图、照片、方法、过程等内容，除另有特别注明外，其著作权或其他相关权利均属于中科海微（北京）科技有限公司。未经中科海微（北京）科技有限公司书面同意，任何人不得以任何方式或形式对本手册内的任何部分进行复制、摘录、备份、修改、传播、翻译成其它语言、将其全部或部分用于商业用途**。**

**免责条款**

本文档依据现有信息制作，其内容如有更改，恕不另行通知。中科海微（北京）科技有限公司在编写该文档的时候已尽最大努力保证其内容准确可靠，但中科海微（北京）科技有限公司不对本文档中的遗漏、不准确、或错误导致的损失和损害承担责任。

**技术支持与信息反馈**

如果您在使用我们的产品时遇到问题,或者您认为我们的产品有某些功能缺陷,请访问我们的官网www.haiwei.tech联系我们的客服,我们将为您解决问题和反馈;或者需要技术支持指导以及有任何宝贵意见，也请您通过官网或者电话联系我们：

**内核热升级简介**

本文档介绍SeawayEdge系统内核热升级特性的安装部署和使用方法，SeawayEdge的内核热升级特性通过快速重启内核和程序热迁移实现，我们提供了一个用户态工具以自动化这一过程。

本文档适用于使用SeawayEdge系统并希望了解和使用内核热升级的开发者。使用人员需要具备基础的Linux操作系统知识。

内核热升级的目标，是实现在秒级的端到端时延下，实现进程运行现场的保存和恢复。使用场景通常符合以下两个条件：

* 内核由于漏洞修复，版本更新等原因，需要重新启动
* 运行在内核之上的业务能够在内核重启后快速恢复状态

**内核热升级用户指南**

本文档主要适用于使用SeawayEdge并需要安装容器的用户。用户需要具备以下经验和技能：

* 熟悉Linux基本操作
* 对容器有一定了解

# 安装与部署

## 1.1软硬件要求

* + 1. 硬件要求

当前仅支持arm64架构

* + 1. 软件要求

操作系统：SeawayEdge-v3.3

## 1.2环境准备

安装SeawayEdge系统

安装内核热升级工具需要root权限

## 1.3安装内核热升级工具

安装内核热升级工具的操作步骤如下：

1. 挂载SeawayEdge的iso文件。



1. 配置本地yum源



配置内容如下所示：



1. 将RPM数字签名GPG公钥导入系统



1. 安装内核热升级工具



1. 验证是否安装成功，命令和回显如下，则表示安装成功



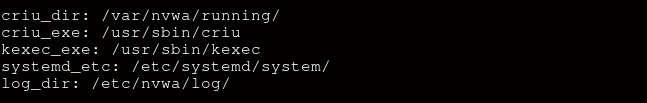
## 1.4部署内核热升级工具

* 配置介绍

内核热升级工具的配置文件位于/etc/nvwa，配置文件包括：nvwa-restore.yaml该配置文件用于指导内核热升级工具在内核热升级过程中如何保存和恢复现场，具体配置如下：1. pidspids用于指明nvwa热升级过程中需要保留和恢复的进程，此处的进程通过进程号(pid)进行标识，需要注意的是，nvwa管理的进程在nvwa服务启动后，会被自动恢复。2. servicesservices用于指明nvwa热升级过程中需要保留和恢复的服务。与pids的区别在于，内核热升级工具可以直接保存和恢复进程的状态，对于服务，内核热升级工具则需要依赖systemd进行相关操作。此处的服务名称，应该使用systemd中使用的服务名称。需要注意的是，对于nvwa管理的服务，是否要在nvwa启动时自动恢复，取决于systemd中有没有使能该服务，且当前支持的服务类型只有notify和oneshot。3. restore\_netrestore\_net用于指明是否需要内核热升级工具保存和恢复网络配置，如果网络配置有误，有可能导致恢复后网络不可用，默认关闭。4. enable\_quick\_kexecenable\_quick\_kexec用于指明是否需要使能quick kexec特性，quick kexec是nvwa社区推出的，加速内核重启过程的一个特性。使用该特性，需要在cmdline中，加入"quickkexec=128M"。128指分配给quick kexec特性的内存大小，该内存将用于在升级过程中加载kernel和initramfs，因此大小需要大于升级过程中涉及到的kernel，initramfs大小之和。该特性默认关闭。5. enable\_pin\_memoryenable\_pin\_memory用于指明是否需要使能pin memory特性，pin memory是nvwa社区推出的，加速进程保存恢复过程的一个特性。使用该特性，需要在cmdline中，加入"max\_pin\_pid\_num=10 redirect\_space\_size=2M pinmemory=200M@0x640000000"。其中，max\_pin\_pid\_num代表支持pin memory恢复的最大进程数目，redirect\_space\_size代表pin memory过程中重定向物理页所需要的预留内存空间，建议配置为pin memory总预留内存的1/100，pinmemory指明这段内存的起点和大小。从0x640000000开始的200M空间，是pin memory使用的全部内存空间，这段空间不应该被其他程序使用。nvwa-restore.yaml的配置示例：



nvwa-server.yaml该文件包含了内核热升级工具运行过程中，需要使用到的配置信息，具体如下：criu\_dir用于指明内核热升级工具在保存现场过程中，存储产生的信息文件夹路径。需要注意的是，这些信息可能会占用较大的磁盘空间。criu\_exe用于指明内核热升级工具使用的criu可执行文件路径，除非是对criu进行调测，一般不建议修改。kexec\_exe用于指明内核热升级工具使用的kexec可执行文件路径，除非是对kexec进行调测，一般不建议修改。systemd\_etc用于指明覆盖systemd配置过程中，使用到的文件夹路径。该路径由systemd决定，一般不需要修改。log\_dir存放内核热升级工具产生的log信息，log模块当前未启用。内核热升级工具日志信息的查看，参考使用方法-产生的日志信息。nvwa-server.yaml的配置示例：



## 1.5使能内核热升级工具

内核热升级工具的运行依赖配置文件，配置文件修改后应该重新运行内核热升级工具程序。安装成功后，可以通过systemd的相关命令来操作内核热升级工具。

* 使能nvwa。



* 启动nvwa。



* 查看nvwa服务状态及相关信息。



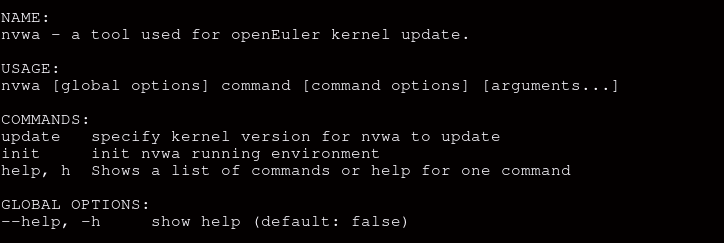
* 更多用法参考systemd用法。

# 第二章 使用方法

## 2.1 命令用法

* nvwa help

打印帮助信息，打印的信息如下:



* nvwa update

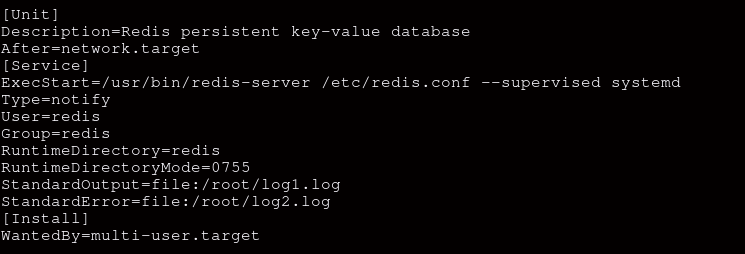
热升级到内核某一版本，nvwa会去/boot目录下寻找内核镜像和ramfs。kernel的命名格式需为vmlinuz-<kernel version>，rootfs命名格式需为initramfs-<kernel version>.img。需要注意的是，升级过程有可能会失败，如果失败，部分被dump的进程或者服务，将停止运行。

* nvwa init

清除nvwa产生的现场信息以及对systemd的配置修改，用于nvwa执行前或者执行失败后，对现场进行清理。

## 使用限制

* 对于需要通过nvwa保存的service，其配置中需要设置标准输出(StandardOutput)和错误输出(StandardError)，以redis为例:



* 使用加速特性需要修改cmdline以及分配合适的内存
* 运行过程中需要关闭SELINUX

理论上，仅需要在执行nvwa update之后和系统重启nvwa恢复现场这段时间前关闭。稳妥起见，建议全程关闭SELINUX。

## 加速特性说明及使用

* cpu park(加速内核重启过程)cpu park，是在使用kexec过程，使cpu进入一种忙等的状态，更快的响应主核发送的中断请求，减少状态的变化。使用cpu park，需要在cmdline中加入"cpuparkmem=0x200000000"，其中0x200000000是一段未被其他程序使用的内存起始地址，cpuparkmem将占用从该地址开始，size为1M左右的内存空间。需要注意的是，在内存允许的情况下，此处的地址选择，建议范围在4G(0x100000000)之后，前4G通常被系统各组件预留，容易冲突。
* quick kexec(加速内核启动过程)quick kexec，是对kexec加载镜像过程中的一种加速。使用quick kexec，需要在配置文件中使能相关选项，更多信息参考<<安装与部署>>(配置介绍)。
* pin\_memory(加速现场保存恢复过程)pin memory，是对criu进行现场保存恢复过程中的一种加速。使用pin memory，需要在配置文件中使能相关选项，更多信息参考<<安装与部署>>(配置介绍)。

## 产生的日志信息

内核热升级工具产生的日志分为两部分:、

* 运行过程产生的日志

通过service nvwa status查看

* 保留现场过程中产生的日志

日志位于criu\_dir指定的路径对应命名的进程/服务文件夹中

# 第三章 常见问题与解决方法

* 执行nvwa update后未升级

原因：保留现场或者内核替换过程中出现错误。解决方法：查看日志，找出错误原因。

* 开启加速特性后，nvwa执行命令失败

原因：nvwa提供了诸多加速特性，包括quick kexec，pin memory，cpu park等等。这些特性都涉及到cmdline的配置和内存的分配，在选取内存时，通过cat /proc/iomem确保选取的内存没有与其他程序冲突。必要时，通过dmesg查看使能特性后是否存在错误日志。

* 热升级后，相关现场未被恢复

原因：首先检查nvwa服务是否运行，运行情况下，可能存在两种情况，一种是服务恢复失败，一种是进程恢复失败。解决方法：通过service nvwa status查看nvwa的日志，如果是服务启动失败，首先确认是否使能了该服务，再通过systemd查看对应服务的日志。进一步的日志，去criu\_dir指定的路径对应命名的进程/服务文件夹中。其中dump.log为保存现场产生的日志，restore.log为恢复现场产生的。

* 恢复失败，日志显示Can't fork for 948: File exists

原因：内核热升级工具在恢复程序过程中，发现程序的pid已经被占用。解决方法：当前内核没有提供保留pid的机制，相关策略正在开发，预计会在将来的内核版本中解决这一限制，当前仅能手动重启相关进程。

* 使用nvwa去保存和恢复简单程序(hello world)，显示失败或者程序未在执行

原因: criu使用存在诸多限制解决办法：查看nvwa的日志，如果显示是criu相关的错误，去相应的目录下检查dump.log或者restore.log，criu相关的使用限制，可以参考wiki。