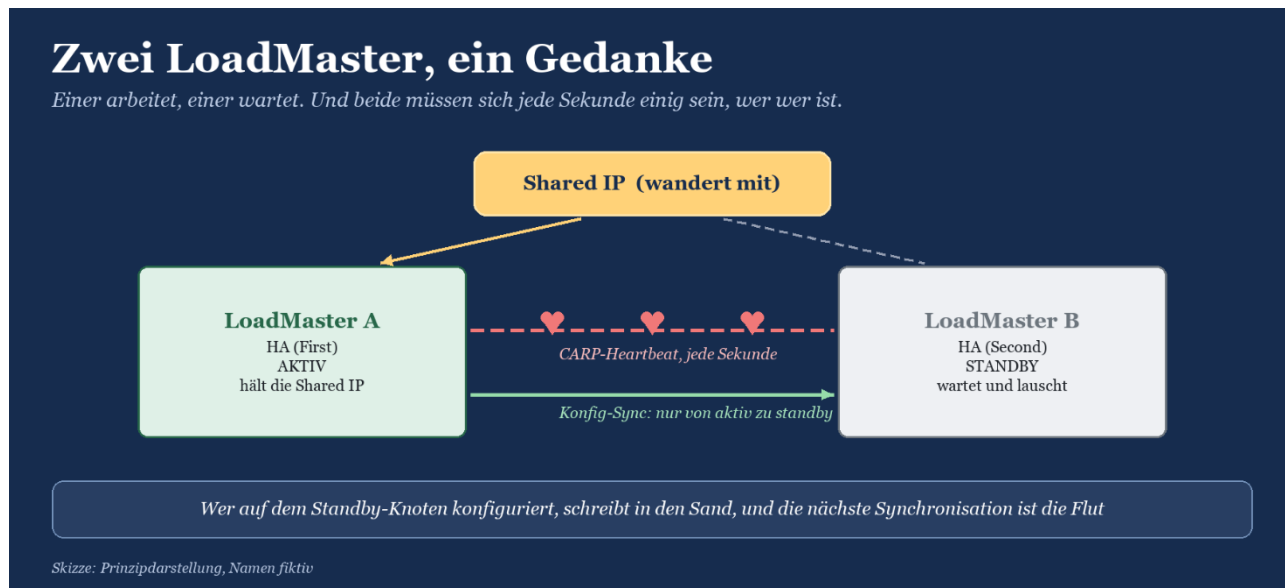


HA-Paar, Failover und Konfig-Sync: zwei LoadMaster, ein Gedanke



WISSEN

Alles zum Kemp LoadMaster mit Microsoft-Brille: Exchange, SharePoint, ADFS, RDS und Kerberos an einem Ort.

› [Kemp LoadMaster](#)

BERATUNG

HA-Paar geerbt und nie getestet? Wir spielen den Ausfall gemeinsam durch, bevor er sich selbst ankündigt.

› [Beratung zum Kemp LoadMaster](#)

SCHULUNG

Grundlagen und Praxis für dein Admin-Team: vom ersten Virtual Service bis zum kontrollierten Failover.

› [Kemp-LoadMaster-Schulung](#)

Ein einzelner Load Balancer vor einer Exchange-Organisation ist ein bemerkenswertes Konstrukt. Du hast zwei, drei oder vier Mailboxserver, eine Database Availability Group, redundante Domänencontroller, und dann schickst du den gesamten Verkehr durch genau ein Gerät. Das ist ungefähr so, als würdest du für dein Rechenzentrum zwei Notstromaggregate kaufen und beide an dieselbe Steckdosenleiste hängen. Funktioniert prima, bis jemand mit dem Staubsauger kommt.

Deshalb steht der LoadMaster in ernst gemeinten Umgebungen als Paar da: zwei Knoten, einer aktiv, einer im Standby, verbunden über einen Heartbeat, der jede Sekunde „ich lebe noch“ ruft. Dieser Beitrag dreht sich um alles, was beim kemp loadmaster ha in der Praxis zählt: wie Active/Standby mit Shared IP funktioniert, wofür die Heartbeat-Interfaces da sind, ob ein bevorzugter Knoten eine gute Idee ist, was die Konfigurationssynchronisation abgleicht und, wichtiger, was sie NICHT abgleicht. Dazu kommen Split-Brain, also der Moment, in dem sich beide Knoten für den Chef halten, die Fallstricke unter vSphere und Hyper-V und eine Geschichte, die fast jede Administratorin und fast jeder Administrator einmal erlebt: Die Einstellung „springt zurück“.

Die Herstellerdokumentation listet dir die Parameter auf, Microsoft Learn erklärt, was Exchange, ADFS oder die Remotedesktopdienste von einem Load Balancer erwarten. Was dazwischen fehlt, ist die Erklärung, warum ein HA-Paar so tickt, wie es tickt, und an welchen Stellen es in einer typischen Microsoft-Umgebung knallt. Den Überblick über alle Themen rund um den LoadMaster findest du im Wissensbereich [Kemp LoadMaster: der Wissensbereich](#).

FAKTEN • Das HA-Paar in Zahlen

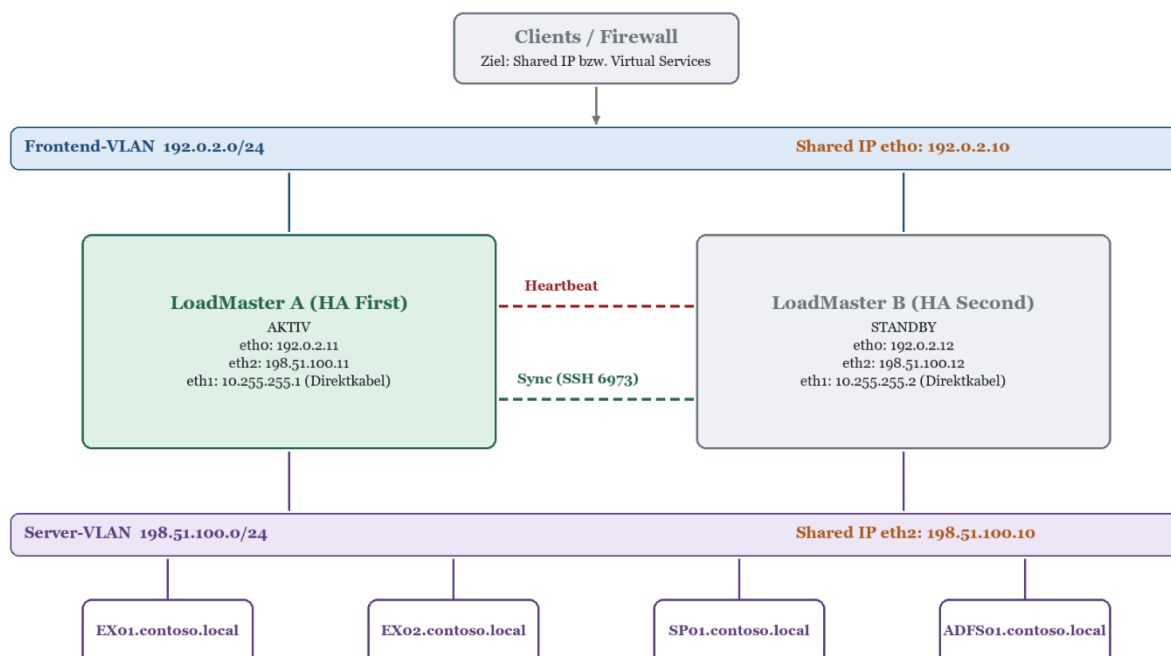
Zwei identische LoadMaster, Rollen HA (First) und HA (Second). Pro Netz eine Shared IP, die nur der aktive Knoten hält, plus je eine eigene Adresse pro Knoten. Heartbeat über CARP, ein Paket pro Sekunde, standardmäßig an die Multicast-Adresse 224.0.0.18. HA Timeout zwischen 3 und 15 Sekunden in 3-Sekunden-Schritten, Standard 9 Sekunden. Konfigurationsabgleich über SSH auf Port 6973, regulär alle zwei Minuten, auf Knopfdruck sofort.

Active/Standby verstehen: ein Paar, eine Adresse, ein Herzschlag

Kemp setzt beim klassischen HA-Paar auf Active/Standby. Das heißt: Genau ein Knoten verarbeitet den Verkehr, der andere langweilt sich professionell. Active/Active im Sinne von „beide balancieren gleichzeitig dieselben Virtual Services“ gibt es in diesem Modell nicht. Wer Last auf mehrere Geräte verteilen will, landet bei Clustering oder bei GSLB über mehrere Standorte, und das ist ein anderes Thema mit anderen Problemen.

Anatomie eines HA-Paars (Dual-Arm)

Jeder Knoten hat eigene Adressen, gemeinsam teilen sie sich eine Shared IP pro Netz



Adressen aus Dokumentationsbereichen, Hostnamen fiktiv. Gateway der Server bei Transparenz: die Shared IP, nie eine Knoten-IP.

Shared IP und Knoten-IPs

Jedes Interface, das am HA-Geschehen teilnimmt, braucht drei Adressen: eine für Knoten A, eine für Knoten B und die Shared IP, die immer beim aktiven Knoten wohnt. Die Virtual Services liegen zusätzlich auf eigenen Adressen, die ebenfalls mit dem aktiven Knoten umziehen. Bei einer Dual-Arm-Konfiguration mit getrenntem Frontend- und Server-Netz brauchst du also zwei Shared IPs, eine pro Seite.

Warum das wichtig ist, merkst du spätestens bei transparenten Virtual Services. Wenn die Real Server die echte Client-IP sehen sollen, muss ihr Antwortverkehr wieder über den LoadMaster laufen, meistens weil er ihr Default Gateway ist. Trägst du dort die Knoten-IP von A ein statt der Shared IP, funktioniert alles wunderbar, bis A ausfällt. Dann schicken die Exchange-Server ihre

Antworten an ein Gerät, das gerade nicht mehr existiert. Mehr zu Transparenz und SNAT steht in [Client-IP sichtbar machen: Transparenz, SNAT und X-Forwarded-For](#).

Heartbeat und CARP

Der Heartbeat basiert auf CARP, dem Common Address Redundancy Protocol. Der aktive Knoten sendet jede Sekunde ein Lebenszeichen, standardmäßig per Multicast. Bleibt es länger aus als das eingestellte HA Timeout, übernimmt der Standby-Knoten die Shared IP und alle Virtual-Service-Adressen und verkündet das per Gratuitous ARP im Netz. Die Heartbeats laufen über alle Interfaces, bei denen die HA-Prüfung aktiviert ist. Das ist sinnvoll: Fällt nur das Server-Netz eines Knotens weg, soll das Paar das merken, nicht erst der Benutzer.

Ein dediziertes Direktkabel zwischen beiden Knoten, oft auf eth1, ist bei Hardware-Geräten eine gute Ergänzung. Es ist aber kein Ersatz für Heartbeats auf den produktiven Netzen. Die Logik dahinter: Das Direktkabel beweist nur, dass beide Geräte noch Strom haben. Ob der aktive Knoten seine Clients erreicht, sagt es nicht.

WICHTIG · Gleiches Subnetz, gleicher Standort

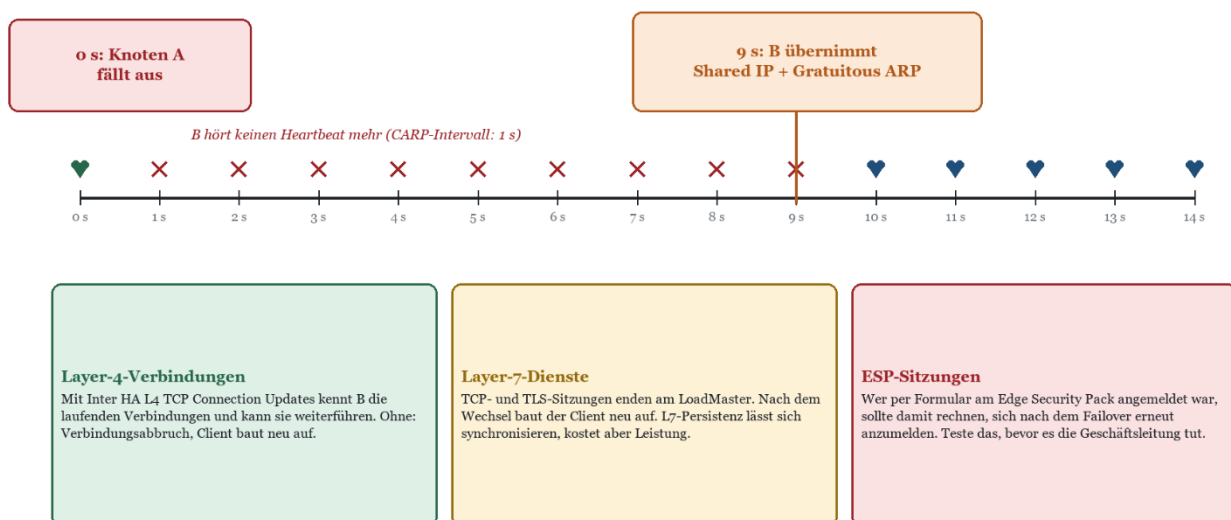
Beide Knoten müssen im selben Layer-2-Segment hängen, dasselbe Default Gateway nutzen und am selben Standort stehen. Ein HA-Paar über zwei Rechenzentren mit geroutetem Link dazwischen ist kein Hochverfügbarkeitskonzept, sondern eine Wette auf die WAN-Leitung. Für mehrere Standorte ist GSLB das richtige Werkzeug: pro Standort ein eigenes Paar, die Verteilung darüber.

Wie du zwei Rechenzentren mit je einem Paar sauber verbindest, beschreibt [Kemp 360 Central und Global Server Load Balancing für mehrere Standorte](#).

Was ein Failover für Verbindungen bedeutet

Ein Failover in Zeitlupe

HA Timeout 9 Sekunden (Standard): was in dieser Zeit passiert und was danach



Prinzipdarstellung. HA Timeout einstellbar von 3 bis 15 Sekunden in 3-Sekunden-Schritten.

Neun Sekunden klingen nach wenig. Für einen Outlook-Client, der per MAPI over HTTP mit Exchange redet, sind sie eine Ewigkeit, aber eine verzeihliche: Outlook baut die Verbindung neu auf, der Benutzer sieht im schlimmsten Fall kurz „Verbindung wird hergestellt“. Anders sieht es bei allem aus, was Sitzungszustand am LoadMaster hält. Layer-7-Virtual-Services terminieren TCP

und meist auch TLS am LoadMaster. Diese Verbindungen existieren nach dem Failover schlicht nicht mehr und müssen neu aufgebaut werden.

Was	Ohne Zustandsabgleich	Mit Zustandsabgleich
Layer-4-Verbindungen	Abbruch, Client baut neu auf	Inter HA L4 TCP Connection Updates: laufende Verbindungen können weiterlaufen
Layer-7-Persistenz	Persistenztabelle leer, Client landet evtl. auf anderem Real Server	Inter HA L7 Persistency Updates: Zuordnung bleibt erhalten, kostet aber spürbar Leistung
TLS-Sitzungen am LoadMaster	Neuer Handshake	Neuer Handshake, daran ändert kein Abgleich etwas
ESP-Anmeldungen	Benutzer muss sich in der Regel neu anmelden	Im Einzelfall testen
Health-Check-Status	Neuer Knoten prüft selbst	Neuer Knoten prüft selbst

Beide Abgleich-Optionen brauchen ein Interface mit funktionierendem Multicast. Die L7-Variante kann laut Hersteller die Leistung deutlich beeinträchtigen. Bei Exchange, SharePoint und ADFS ist sie selten nötig, weil diese Dienste ohnehin ohne Persistenz oder mit unkritischer Persistenz auskommen. Warum das so ist, erklärt [Persistence und Scheduling: wer hängt an welchem Server und warum](#).

Bevorzugter Knoten: Prefer First HA oder lieber nicht?

Die Option „Switch to Preferred Server“ legt fest, ob nach einem Failover wieder zurückgeschwenkt wird. Mit „Prefer First HA“ reißt Knoten A die Führung an sich, sobald er wieder da ist. Mit „No Preferred Host“, der Voreinstellung, bleibt der Knoten aktiv, der gerade aktiv ist.

Einstellung	Verhalten	Einschätzung
No Preferred Host	Kein Zurückschwenken, wer aktiv ist, bleibt aktiv	Empfehlung des Herstellers für den Betrieb. Ein Ausfall, ein Umschalten.
Prefer First HA	A übernimmt nach jedem Neustart wieder	Zweites Umschalten, zweiter Verbindungsabbruch. Bei flatterndem Knoten: Pingpong.
Prefer Second HA	Wie oben, nur mit B	Gleiche Nachteile, zusätzlich verwirrend.

Der Reflex „ich will aber wissen, welcher Knoten aktiv ist“ ist verständlich. Die Antwort darauf ist Monitoring, nicht ein bevorzugter Knoten. Ein Gerät, das wegen eines halbdefekten Netzteils alle zwanzig Minuten neu startet und dabei jedes Mal die Führung zurückfordert, erzeugt genau die Ausfälle, die HA verhindern soll. Den Status des Paares holst du per SNMP oder Syslog ins Monitoring, wie in [Monitoring, Syslog und SNMP: den LoadMaster ins Monitoring holen](#) beschrieben.

TIPP · Preferred Host nur zum Testen

Für Failover-Tests im Wartungsfenster ist ein bevorzugter Knoten praktisch, weil du reproduzierbar hin- und herschalten kannst. Danach stellst du wieder auf No Preferred Host. Schreib dir das in die Checkliste, sonst steht die Testeinstellung drei Jahre später noch da.

Konfig-Sync: was abgeglichen wird und was nicht

Der Abgleich läuft immer in eine Richtung: vom aktiven zum Standby-Knoten. Der aktive Knoten schiebt seine Konfiguration regelmäßig, laut Hersteller alle zwei Minuten, über das HA Update Interface per SSH auf Port 6973 zum Partner. Mit „Force Partner Update“ kannst du den Abgleich sofort anstoßen, sofern sich beide Knoten sehen.

Das klingt banal, hat aber eine Konsequenz, die man sich auf die Stirn tätowieren sollte: Der Standby-Knoten hat keine eigene Meinung. Alles, was dort an synchronisierten Objekten steht, ist eine Kopie, und Kopien werden überschrieben.

Was wandert und was bleibt

Bereich	Abgleich	Was du tun musst
Virtual Services, SubVS, Real Server, Health Checks	Ja	Nur auf dem aktiven Knoten pflegen
Content Rules, WAF-Regeln, ESP-Einstellungen	Ja	Nur auf dem aktiven Knoten pflegen
Zertifikate und Zwischenzertifikate	Ja	Nach dem Import am Standby-Knoten kontrollieren, ob sie angekommen sind
Knoteneigene Interface-Adressen	Nein, per Definition pro Knoten	Beim Aufbau sauber dokumentieren
HA-Rolle (First/Second)	Nein	Auf jedem Knoten einzeln, niemals beide gleich
Kennwort des Kontos bal	Nein	Auf beiden Knoten identisch setzen und gemeinsam ändern
Firmware-Version	Nein	Knoten einzeln aktualisieren, Standby zuerst
Lizenz	Nein	Jeder Knoten braucht seine eigene
Default Gateway (LoadMaster in Azure)	Nein	Pro Knoten setzen

Die Liste ist bewusst praxisorientiert und nicht erschöpfend. Je nach Firmware-Version gibt es weitere knotenindividuelle Werte. Die sichere Faustregel lautet: Alles, was den einzelnen Knoten als Gerät beschreibt, bleibt lokal. Alles, was den Dienst beschreibt, wandert. Wer unsicher ist, vergleicht nach einer Änderung beide Knoten, idealerweise per Skript über die REST-API, wie in [Automatisierung per REST-API und PowerShell](#) gezeigt.

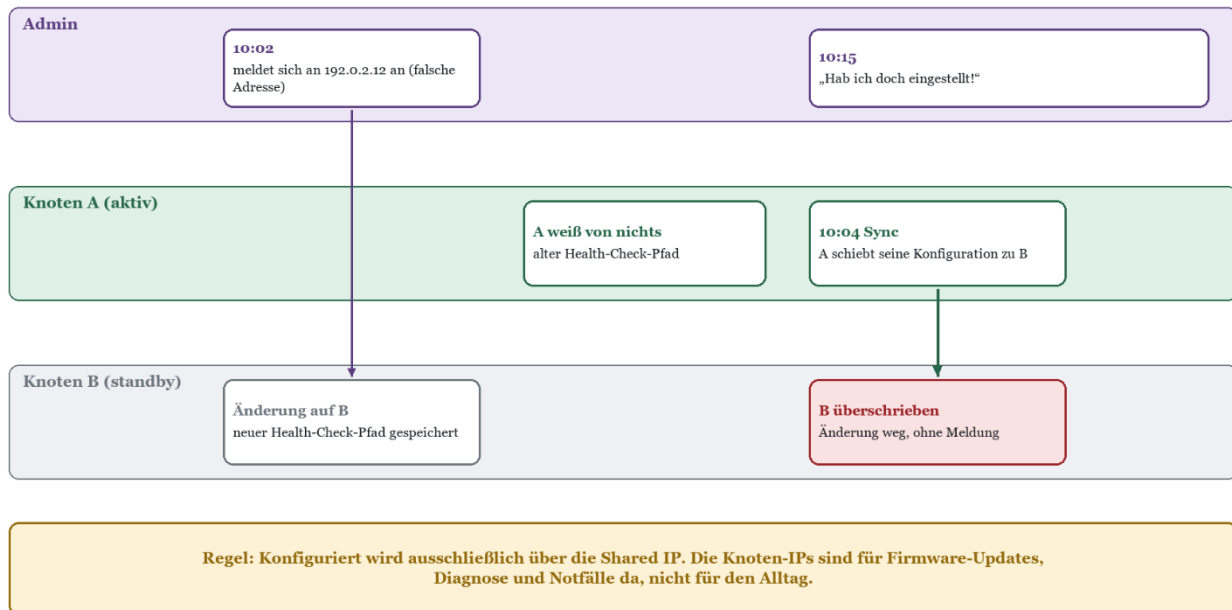
WARNUNG · Das bal-Kennwort wird nicht synchronisiert

Wenn du das Kennwort des Administratorkontos bal auf dem aktiven Knoten änderst, gilt auf dem Standby-Knoten weiterhin das alte. Das fällt erst auf, wenn du nach einem Failover oder beim nächsten Firmware-Update dringend auf B musst und im Kennwortsafe nur das neue steht. Ändere es immer auf beiden Knoten im selben Arbeitsgang, und noch besser: Nutze für den Alltag benannte Konten statt bal.

Praxis: Die Einstellung springt zurück

Warum deine Einstellung „zurückspringt“

Ein Dienstagvormittag in drei Bahnen: Admin, aktiver Knoten, Standby-Knoten



Die Geschichte geht immer gleich. Jemand hat ein Lesezeichen auf die Weboberfläche des LoadMasters, und zwar auf die Knoten-IP, die bei der Einrichtung als erste funktionierte. Seit dem letzten Failover ist das aber der Standby-Knoten. Die Änderung wird gespeichert, die Oberfläche zeigt sie brav an, und zwei Minuten später ist sie weg, weil der aktive Knoten seine Sicht der Dinge herüberschiebt. Kein Fehler, keine Warnung im Gesicht, nur ein Admin, der an seinem Verstand zweifelt.

Je nach Firmware-Version und Bereich verhindert der Standby-Knoten manche Änderungen von vornherein oder zeigt nur ein eingeschränktes Menü. Darauf verlassen solltest du dich nicht. Die Regel ist einfach und gehört in jede Betriebsdokumentation:

- **Konfiguration nur über die Shared IP.** Sie zeigt immer auf den aktiven Knoten, egal wie oft das Paar seit der Einrichtung umgeschaltet hat.
- **Knoten-IPs nur für Knotenarbeit.** Firmware-Update, Diagnose, Neustart eines einzelnen Knotens, sonst nichts.
- **Lesezeichen aufräumen.** Im Browser des Teams und in der Admin-Dokumentation stehen nur der DNS-Name der Shared IP, etwa `lm.contoso.local`, und die beiden Knotennamen klar beschriftet.
- **Status prüfen, bevor du klickst.** Die WUI zeigt den HA-Status jedes Knotens an. Ein kurzer Blick, ob du auf dem aktiven Knoten bist, spart die Viertelstunde Selbstzweifel.
- **Skripte gegen die Shared IP.** Wer per REST-API automatisiert, adressiert die Shared IP. Ein Skript gegen eine Knoten-IP produziert denselben Effekt, nur zuverlässiger.

AUS DER PRAXIS · Der Health Check, der dreimal eingestellt wurde

In einer Kommunalverwaltung sollte der Health Check eines SharePoint-Virtual-Service von „Port offen“ auf eine echte HTTP-Prüfung umgestellt werden. Der Kollege speicherte die Änderung, testete, war zufrieden. Am nächsten Morgen war die alte Einstellung wieder da. Er stellte sie erneut um, diesmal mit Bildschirmfoto als Beweis. Wieder weg. Beim dritten Mal fiel jemandem auf, dass in der Adressleiste die Knoten-IP von B stand, der seit einem Stromausfall im Vorjahr Standby war. Die

Lösung bestand aus einem neuen Lesezeichen und einer Notiz im Betriebshandbuch. Die Bildschirmfotos werden bis heute in der Abteilung herumgereicht.

Wie ein Health Check aussehen sollte, der nicht nur offene Ports zählt, beschreibt [Health Checks, die wirklich prüfen: warum „Port 443 offen“ nicht reicht](#).

Zertifikate im HA-Paar

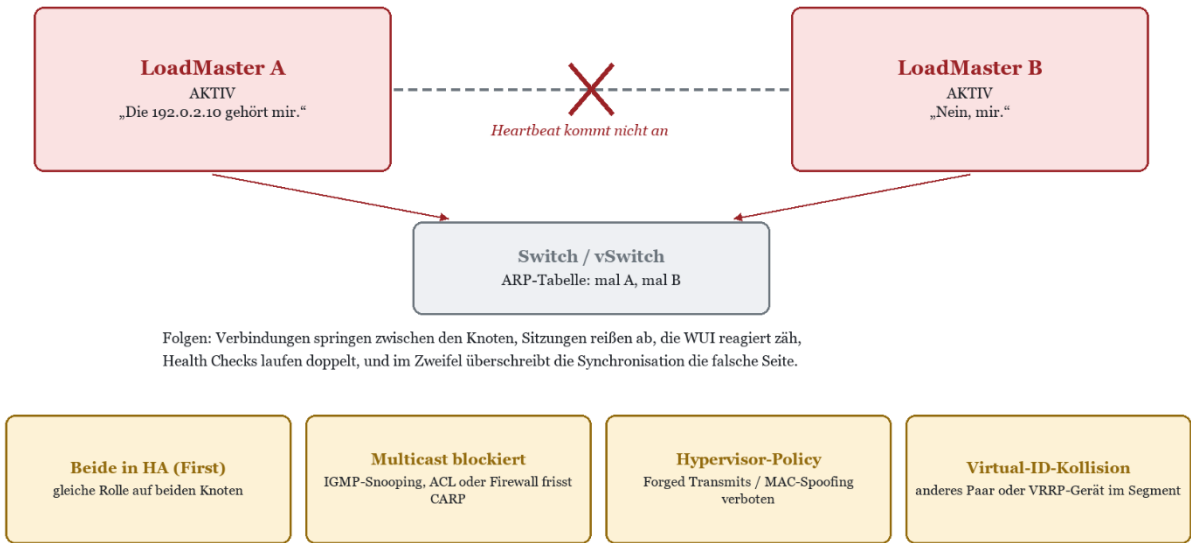
Zertifikate werden mitsynchronisiert, und genau deshalb lohnt sich ein Kontrollblick. Nach einem Zertifikatswechsel am aktiven Knoten wartest du den Abgleich ab oder stößt ihn an und prüfst auf dem Standby-Knoten, ob das neue Zertifikat und die Zwischenzertifikate angekommen sind. Sonst erlebst du beim nächsten Failover einen zweiten Zertifikatswechsel, diesmal unfreiwillig und rückwärts. Die Details zum Zertifikatsprozess stehen in [TLS-Offload, Re-Encryption und Zertifikate: PKI trifft LoadMaster](#).

Split-Brain: wenn beide Knoten Chef sein wollen

Split-Brain ist der Zustand, in dem beide Knoten aktiv sind. Das passiert, wenn der Standby-Knoten keinen Heartbeat mehr hört, der aktive Knoten aber quicklebendig ist. Aus Sicht von B ist A tot, also übernimmt B. Aus Sicht von A hat sich nichts geändert. Ab jetzt beanspruchen zwei Geräte dieselben Adressen, und der Switch entscheidet nach Tagesform, wer welches Paket bekommt.

Split-Brain: zwei Chefs, ein Büro

Beide Knoten halten sich für aktiv und beanspruchen dieselbe Shared IP



Die vier häufigsten Ursachen. Fast nie ist es der LoadMaster selbst.

Die üblichen Verdächtigen

Ursache	Woran du es erkennst	Abhilfe
Beide Knoten in HA (First)	Beide aktiv direkt nach der Einrichtung oder nach einem Restore	Ein Knoten auf HA (Second), Neustart, dann sauber neu synchronisieren
Multicast wird gefiltert	Paar läuft lange stabil, nach Switch-Tausch oder neuer ACL plötzlich doppelt aktiv	IGMP-Snooping und ACLs für 224.0.0.18 prüfen, alternativ CARP per Broadcast
Hypervisor verwirft	Funktioniert, solange beide VMs auf	Forged Transmits und MAC-

Pakete	einem Host laufen	Änderungen bzw. MAC-Spoofing erlauben
HA Virtual ID doppelt	Seltsames Verhalten, sobald ein zweites Paar oder ein VRRP-Gerät ins Segment kommt	Eindeutige Virtual ID pro Segment, Werte über 10
Partner Shared Secret abweichend	Knoten ignorieren sich gegenseitig	Shared Secret auf beiden Knoten identisch setzen
HA-Prüfung auf Interface aus	Ein Netzausfall auf einer Seite wird nicht bemerkt	HA-Prüfung auf allen produktiven Interfaces aktivieren

FAKTEN · CARP und VRRP teilen sich mehr, als man denkt

CARP nutzt wie VRRP die IP-Protokollnummer 112 und standardmäßig die Multicast-Adresse 224.0.0.18. Steht im selben Segment ein Router- oder Firewall-Paar mit VRRP, kollidieren die virtuellen IDs im ungünstigen Fall. Seit Firmware 7.2.36 wählt der LoadMaster die HA Virtual ID automatisch aus dem letzten Oktett der ersten Shared IP. Das ist besser als die frühere feste Voreinstellung, ersetzt aber nicht den Blick auf das, was sonst noch im Segment funkt.

Split-Brain vermeiden statt aufräumen

- **Rollen dokumentieren und prüfen.** A ist HA (First), B ist HA (Second). Ein Restore eines Backups von A auf B ist der Klassiker, der beide zu „First“ macht.
- **Heartbeat auf mehreren Wegen.** HA-Prüfung auf allen produktiven Interfaces plus, wo möglich, eine direkte Verbindung. Ein einzelner Pfad ist ein einzelner Fehlerpunkt.
- **Netzwerkteam einbinden.** Wer am Switch IGMP-Snooping umstellt oder ACLs verschärft, sollte wissen, dass dort ein HA-Paar Multicast braucht. Ein Satz im Netzwerkplan genügt.
- **Timeout nicht zu knapp.** Ein HA Timeout von 3 Sekunden klingt sportlich, macht das Paar aber empfindlich für kurze Hänger, etwa bei einem vMotion oder einer Spanning-Tree-Neuberechnung. Die 9 Sekunden der Voreinstellung sind für Microsoft-Workloads meist der vernünftige Kompromiss.
- **HA Initial Wait Time beachten.** Nach einem Neustart wartet ein Knoten eine Weile, bevor er sich für aktiv erklärt. Bei Switches, die ihre Ports langsam hochfahren, verhindert ein passender Wert, dass ein frisch gestarteter Knoten vorschnell übernimmt.
- **Failover regelmäßig testen.** Ein HA-Paar, das nie umgeschaltet hat, ist eine Hypothese.

WARNUNG · Im Split-Brain nicht hektisch konfigurieren

Wenn beide Knoten aktiv sind, ist die WUI oft zäh oder gar nicht erreichbar, und jede Änderung landet auf irgendeinem der beiden. Erst die Ursache beheben, dann einen Knoten gezielt in Standby bringen, notfalls durch Trennen der Netzwerkverbindung, und anschließend prüfen, welcher Konfigurationsstand der richtige ist. Sonst synchronisiert der falsche Knoten seinen Stand auf den richtigen, und du suchst danach nach Änderungen, die es nie gegeben hat.

Wenn du beim Eingrenzen tiefer graben musst, etwa mit einem Mitschnitt auf dem Interface, um zu sehen, ob CARP-Pakete überhaupt ankommen, hilft [Troubleshooting am LoadMaster: Logs, L7-Debug und TCP-Dump](#).

Hypervisor-Fallstricke: vSphere-Portgruppe und Hyper-V MAC-Spoofing

Virtuelle LoadMaster sind im Mittelstand und in Verwaltungen der Normalfall. Sie sind schnell bereitgestellt, sauber gesichert und lassen sich zwischen Hosts verschieben. Sie haben nur ein

Problem: Ein virtueller Switch ist ein sehr misstrauisches Stück Software. Er weiß, welche MAC-Adresse zu welcher VM gehört, und wenn eine VM plötzlich Pakete mit fremden Adressen schickt oder Adressen übernimmt, hält er das für einen Angriff. Ein HA-Failover sieht für ihn genau so aus.

Virtuelle LoadMaster: die Netzwerk-Hausaufgaben

Ohne diese Einstellungen funktioniert HA im Labor, aber nicht beim ersten echten Ausfall

VMware vSphere

Portgruppe der LoadMaster-Interfaces:

- MAC-Adressänderungen: Akzeptieren
- Gefälschte Übertragungen (Forged Transmits): Akzeptieren
- Neuere vSphere-Versionen stehen hier standardmäßig auf Ablehnen
- Eigene Portgruppe nur für die LoadMaster, damit die Lockerung nicht für alle VMs gilt
- Notify Switches: Nein, wenn die Knoten auf verschiedenen Hosts laufen
- DRS-Anti-Affinitätsregel: A und B nie auf demselben Host

Microsoft Hyper-V

Netzwerkkarten der LoadMaster-VM:

- Erweiterte Features: Spoofing von MAC-Adressen aktivieren
- Für jede vNIC, die eine Shared IP trägt, nicht nur für eine
- Im Failovercluster: Anti-Affinität per AntiAffinityClassNames
- Gleiche Einstellung auf allen Hosts, sonst funktioniert HA nur bis zur nächsten Live-Migration
- Switch Embedded Teaming und Port-ACLs auf Multicast prüfen

Beide Knoten auf einem Host ist kein HA-Paar, sondern ein Paar, das gemeinsam stirbt.

VMware vSphere

Für die Portgruppen, an denen die LoadMaster-Interfaces hängen, müssen in der Sicherheitsrichtlinie MAC-Adressänderungen und gefälschte Übertragungen (Forged Transmits) auf Akzeptieren stehen. Neuere vSphere-Versionen setzen beides standardmäßig auf Ablehnen. Das Tückische: Solange beide Knoten auf demselben Host laufen, merkst du davon oft nichts, weil der Verkehr den Host nicht verlässt. Der Fehler zeigt sich erst, wenn DRS die VMs sinnvollerweise verteilt.

Weil diese Lockerung die Schutzwirkung gegen MAC-Spoofing aufweicht, legst du dafür eine eigene Portgruppe an, in der nur die LoadMaster hängen. Die Datenbankserver im selben VLAN brauchen die Ausnahme nicht. Laufen die Knoten auf verschiedenen Hosts, empfiehlt der Hersteller außerdem, Notify Switches auf Nein zu stellen, damit beim Einschalten einer VM keine RARP-Pakete die Switches durcheinanderbringen. Und eine DRS-Regel, die A und B auf getrennten Hosts hält, versteht sich von selbst.

Microsoft Hyper-V

Unter Hyper-V heißt der entscheidende Schalter „Spoofing von MAC-Adressen aktivieren“ und sitzt in den erweiterten Features jeder Netzwerkkarte der VM. Er muss für jede vNIC aktiv sein, die eine Shared IP oder Virtual-Service-Adressen trägt. Per PowerShell erledigt das Set-VMNetworkAdapter mit dem Parameter -MacAddressSpoofing On. Im Failovercluster sorgst du über die Clustereigenschaft AntiAffinityClassNames dafür, dass die beiden LoadMaster-VMs möglichst nicht auf demselben Knoten landen.

Thema	vSphere	Hyper-V
Fremde Absender-MAC erlauben	Forged Transmits: Akzeptieren	MAC-Spoofing aktivieren
MAC-Änderung erlauben	MAC-Adressänderungen: Akzeptieren	MAC-Spoofing aktivieren
Eingrenzen	Eigene Portgruppe nur für LoadMaster	Nur an den LoadMaster-vNICs setzen
Knoten trennen	DRS-Anti-Affinitätsregel	AntiAffinityClassNames im

		Failovercluster
Typischer Fehler	Läuft nur, solange beide VMs auf einem Host sind	Läuft, bis eine vNIC nachträglich hinzugefügt wird
Virtuelle MAC-Adressen	Option „Use Virtual MAC Addresses“ steht auf virtuellen LoadMastern nicht zur Verfügung	Ebenso

TIPP · Erst den Ausfall simulieren, dann den Betrieb freigeben

Verschiebe nach der Einrichtung einen Knoten bewusst auf einen anderen Host und schalte dann den aktiven Knoten hart ab. Läuft der Verkehr nach dem HA Timeout weiter, stimmt die Hypervisor-Konfiguration. Läuft er nicht, hast du es im Wartungsfenster erfahren statt am Montagmorgen.

Läuft der LoadMaster nicht auf eigenem Blech, sondern in Azure, gelten wieder andere Regeln, weil es dort kein klassisches Layer 2 und kein Gratuitous ARP gibt. Das HA-Modell in Azure arbeitet mit dem Azure Load Balancer davor. Die Unterschiede beschreibt [LoadMaster in Azure und hybriden Umgebungen](#).

Hardware-Geräte und der harte Neustart

Bei Hardware-LoadMastern gibt es die Option „Use Virtual MAC Addresses“: Dann wandert nicht nur die IP, sondern auch eine gemeinsame MAC-Adresse mit, und vorgelagerte Geräte müssen ihre ARP-Tabellen nach einem Failover nicht aktualisieren. Das hilft bei Firewalls, die Gratuitous ARP ignorieren oder lange cachen. Die Umstellung erfordert einen Neustart beider Knoten und ein Leeren der ARP-Caches auf den Nachbargeräten, gehört also ins Wartungsfenster.

Seit Firmware 7.2.53 gibt es zudem „Hard Reboot on link Failure“: Verliert ein Knoten auf einem konfigurierten Interface den Link, startet er neu. Das ist nur ohne bevorzugten Knoten verfügbar und ein recht grobes Werkzeug. Es kann sinnvoll sein, wenn ein Knoten mit halb totem Interface sonst aktiv bliebe. Teste es aber, bevor ein flatternder Switch-Port deinen LoadMaster in eine Neustartschleife schickt.

Betrieb: Parameter, Updates und der Testplan

Die HA-Parameter im Überblick

Parameter	Bedeutung	Empfehlung für Microsoft-Umgebungen
HA Mode	Rolle des Knotens: HA (First), HA (Second) oder kein HA	Unterschiedlich auf beiden Knoten, Änderung erfordert Neustart
HA Timeout	Wie lange der aktive Knoten schweigen darf, 3 bis 15 s	Standard 9 s beibehalten
HA Initial Wait Time	Wartezeit nach dem Booten, bevor ein Knoten aktiv wird	Bei langsamen Switch-Ports erhöhen
HA Virtual ID	Kennung des Paares im Segment, 1 bis 255	Eindeutig, über 10, mit VRRP-Geräten abgleichen
Switch to Preferred Server	Zurückschwenken nach Failover	No Preferred Host
HA Update Interface	Interface für den Konfigurationsabgleich	Ein Interface, auf dem beide Knoten sich zuverlässig erreichen
Inter HA L4 / L7 Updates	Zustandsabgleich für Verbindungen bzw. Persistenz	L4 bei Bedarf, L7 nur mit gutem Grund

Use Broadcast IP address	CARP per Broadcast statt Multicast	Nur, wenn Multicast im Netz nicht durchsetzbar ist
---------------------------------	------------------------------------	--

WICHTIG · Menüpfade ändern sich, Begriffe bleiben

Die HA-Einstellungen findest du in der Systemkonfiguration des LoadMasters im Bereich für Hochverfügbarkeit und Clustering. Die genaue Anordnung hat sich über Firmware-Generationen verschoben. Suche nach den Parameternamen aus der Tabelle, die sind stabil.

Firmware-Updates im Paar

Firmware wird nicht synchronisiert, jeder Knoten wird einzeln aktualisiert. Die bewährte Reihenfolge: Backup ziehen, zuerst den Standby-Knoten über seine eigene Knoten-IP aktualisieren und neu starten, prüfen, dass er wieder als Standby im Paar steht, dann kontrolliert umschalten und den bisher aktiven Knoten aktualisieren. Dazwischen liegt die einzige legitime Gelegenheit, auf einer Knoten-IP zu arbeiten. Den vollständigen Ablauf mit Backup und Notfallplan findest du in [Firmware-Updates, Backup und Notfallplan: LoadMaster-Betrieb ohne Herzrasen](#).

WARNUNG · Backup von A ist kein Backup für B

Ein Backup enthält auch knotenbezogene Einstellungen. Wer das Backup von Knoten A auf einen neuen Knoten B zurückspielt, bekommt im ungünstigen Fall zwei Knoten mit derselben Rolle und denselben Adressen, also Split-Brain auf Bestellung. Ein Ersatzgerät wird als frischer Knoten mit eigener Rolle ins Paar aufgenommen und holt sich die Dienstkonfiguration per Abgleich.

Ein Testplan, der den Namen verdient

Test	Durchführung	Erwartetes Ergebnis
Geplanter Wechsel	Aktiven Knoten über die WUI neu starten	Umschalten nach HA Timeout, Outlook und Browser verbinden sich neu
Harter Ausfall	VM ausschalten bzw. Strom ziehen	Wie oben, ohne Zutun
Netzausfall einseitig	Server-Interface des aktiven Knotens trennen	Paar erkennt den Ausfall, B übernimmt
Hostwechsel	Standby-VM auf anderen Host verschieben, dann Failover	Funktioniert auch über Hostgrenzen
Konfig-Sync	Änderung auf aktivem Knoten, Abgleich anstoßen	Änderung auf B sichtbar
Rückfall	Ausgefallenen Knoten wieder starten	Er wird Standby, kein Zurückschwenken
Anwendungstest	Exchange, SharePoint, ADFS, RD Gateway während des Wechsels nutzen	Kurze Unterbrechung, keine dauerhaften Fehler, keine Zertifikatswarnungen

Besonders aufschlussreich ist der Anwendungstest mit Vorauthentifizierung oder Kerberos Constrained Delegation. Wer ESP oder KCD einsetzt, sollte prüfen, ob Keytabs, SSO-Domänen und Zertifikate auf beiden Knoten identisch funktionieren, siehe [Edge Security Pack: Vorauthentifizierung, SSO-Domänen und der TMG-Ersatz](#) und [Kerberos Constrained Delegation am LoadMaster: KCD und RBCD sauber einrichten](#). Wenn du das HA-Paar gemeinsam mit uns aufbauen oder einen bestehenden Aufbau durchtesten willst, ist das ein typisches Thema für eine [Beratung zum Kemp LoadMaster](#).

FAQ

Kann ein Kemp LoadMaster HA-Paar Active/Active laufen?

Im klassischen HA-Modell nein. Ein Knoten ist aktiv, der andere im Standby. Für echte Lastverteilung über mehrere Geräte gibt es Clustering oder GSLB über mehrere Standorte.

Wie lange dauert ein Failover?

Im Wesentlichen so lange wie das HA Timeout, standardmäßig 9 Sekunden, einstellbar von 3 bis 15 Sekunden. Dazu kommt die Zeit, die Clients für den Neuaufbau ihrer Verbindungen brauchen.

Warum ist meine Änderung am LoadMaster nach ein paar Minuten wieder weg?

Fast immer hast du auf dem Standby-Knoten gearbeitet. Der aktive Knoten synchronisiert seine Konfiguration regelmäßig auf den Partner und überschreibt dabei deine Änderung. Arbeite immer über die Shared IP.

Was wird zwischen den Knoten nicht synchronisiert?

Unter anderem die knoteneigenen Adressen, die HA-Rolle, die Firmware, die Lizenz und das Kennwort des Kontos bal. Alles, was den einzelnen Knoten als Gerät beschreibt, bleibt lokal.

Sollte ich einen bevorzugten Knoten festlegen?

Für den Betrieb nein. Der Hersteller empfiehlt No Preferred Host, weil ein bevorzugter Knoten nach jedem Neustart ein zweites Umschalten und damit einen zweiten Verbindungsabbruch auslöst.

Woran erkenne ich ein Split-Brain?

Beide Knoten melden sich als aktiv, Verbindungen brechen sporadisch ab, die WUI reagiert zäh. Häufige Ursachen sind identische HA-Rollen, gefilterter Multicast und restriktive Hypervisor-Richtlinien.

Welche Einstellungen braucht ein virtueller LoadMaster unter Hyper-V?

MAC-Spoofing muss an jeder Netzwerkkarte aktiviert sein, die eine Shared IP oder Virtual-Service-Adressen trägt. Zusätzlich sollten die beiden VMs über Anti-Affinität auf verschiedenen Hosts laufen.

Und unter VMware vSphere?

In der Portgruppe der LoadMaster-Interfaces müssen MAC-Adressänderungen und Forged Transmits erlaubt sein. Am besten in einer eigenen Portgruppe, damit die Lockerung nicht für andere VMs gilt.

Kann ich die beiden Knoten in verschiedenen Rechenzentren betreiben?

Nicht als HA-Paar. Beide Knoten müssen im selben Subnetz und am selben Standort stehen. Für mehrere Standorte ist ein Paar pro Standort mit GSLB darüber die passende Lösung.

Brauche ich für jeden Knoten eine eigene Lizenz?

Ja, jeder Knoten ist ein eigenständiger LoadMaster. Welche Edition für welchen Einsatz passt, erklärt [Editionen, Lizenzierung und Sizing: welcher LoadMaster für welchen Laden](#).

Fazit: Zwei Geräte, eine Wahrheit

Ein HA-Paar ist schnell eingerichtet und genauso schnell missverstanden. Die Technik dahinter ist überschaubar: eine Shared IP, die mit dem aktiven Knoten wandert, ein CARP-Heartbeat im Sekundentakt und ein Konfigurationsabgleich, der nur in eine Richtung läuft. Die Probleme entstehen fast nie im LoadMaster, sondern drumherum: in Switches, die Multicast verschlucken, in

virtuellen Switches, die einen Failover für einen Angriff halten, und vor dem Bildschirm, wenn jemand auf dem Standby-Knoten konfiguriert.

Wenn du drei Dinge mitnimmst, dann diese: Konfiguriere ausschließlich über die Shared IP. Lass den bevorzugten Knoten weg und hol den HA-Status ins Monitoring. Und teste den Failover regelmäßig, mit echten Anwendungen und über Hostgrenzen hinweg. Ein Paar, das nie umgeschaltet hat, ist nur ein teures Einzelgerät mit Zuschauer. Wer sein Team dafür fit machen will, findet den passenden Rahmen in der [Kemp-LoadMaster-Schulung](#).

WEITERLESEN · Weiter im Thema

- › [Virtual Service, Real Server, SubVS: das Grundvokabular des LoadMasters](#)
- › [Health Checks, die wirklich prüfen: warum „Port 443 offen“ nicht reicht](#)
- › [Firmware-Updates, Backup und Notfallplan: LoadMaster-Betrieb ohne Herzrasen](#)
- › [Monitoring, Syslog und SNMP: den LoadMaster ins Monitoring holen](#)
- › [Troubleshooting am LoadMaster: Logs, L7-Debug und TCP-Dump](#)
- › [Automatisierung per REST-API und PowerShell](#)
- › [Kemp 360 Central und Global Server Load Balancing für mehrere Standorte](#)
- › [LoadMaster in Azure und hybriden Umgebungen](#)