Migration von Windows NLB, IIS ARR oder TMG-Erbe auf den LoadMaster



WISSEN

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

› **Kemp LoadMaster**

BERATUNG

NLB-Cluster, ARR-Farm oder TMG-Regelwerk ablösen, ohne dass am Montag die Hälfte der Anwendungen fehlt? Wir machen mit dir Bestandsaufnahme und Schwenkplan.

› **Beratung zum Kemp LoadMaster**

SCHULUNG

Virtual Services, Health Checks und Content Rules im Labor bauen, bevor du die erste URL-Rewrite-Regel in Produktion übersetzt.

› **Kemp-LoadMaster-Schulung**

Es gibt in fast jedem gewachsenen Microsoft-Rechenzentrum einen Server, den keiner anfassen will. Er heißt irgendwas mit „NLB“, „ARR“ oder, wenn du Pech hast, „TMG“. Er läuft seit Jahren, niemand weiß genau warum, und die Person, die ihn eingerichtet hat, arbeitet inzwischen bei einem Mitbewerber oder genießt die Rente. Dokumentation? Ein Screenshot in einem Ticket von 2016. Monitoring? Wenn die Anwenderinnen und Anwender anrufen, ist er kaputt.

Wenn du nach einer windows nlb alternative suchst, hast du den LoadMaster wahrscheinlich schon gekauft oder zumindest im Angebot liegen. Die Frage ist nicht mehr ob, sondern wie: Wie kommen Exchange, SharePoint, ADFS, RDS und die hausgemachte Fachanwendung vom alten Konstrukt auf Virtual Services, ohne dass dabei ein Dienst unbemerkt stirbt? Genau darum geht es hier. Nicht um die Herstellerdoku, die erklärt, welches Feld wo ist, sondern um die Stellen, an denen Migrationen in der Praxis knallen: Port Rules, die mehr tun als gedacht, Rewrite-Regeln mit Nebenwirkungen, TMG-Veröffentlichungen mit Kerberos-Delegation, und ein DNS-Schwenk, der an einer TTL von einem Tag scheitert.

Der Beitrag gehört zur Serie rund um den Kemp LoadMaster. Wenn dir Begriffe wie Virtual Service, Real Server und SubVS noch nicht in Fleisch und Blut übergegangen sind, lies vorher Virtual Service, Real Server, SubVS: das Grundvokabular des LoadMasters. Hier setzen wir voraus, dass du weißt, was ein Virtual Service ist, und kümmern uns darum, was vorher an seiner Stelle stand.

FAKTEN · Das Wichtigste in 30 Sekunden

Microsoft führt Network Load Balancing in der Liste der Funktionen, die ab Windows Server 2025 nicht mehr weiterentwickelt werden, und nennt den Software Load Balancer aus dem SDN-Stack als Alternative. NLB funktioniert noch, bekommt aber keine neuen Fähigkeiten.

Application Request Routing 3.0 erschien 2013 und ist bis heute die letzte Hauptversion. Es hängt am URL-Rewrite-Modul und lebt im IIS des Servers, den es eigentlich schützen soll.

Forefront TMG 2010 ist seit dem 14. April 2020 vollständig aus dem erweiterten Support. Wer noch TMG-Regeln im Kopf oder im Export hat, migriert Wissen, keine Software.

NLB prüft nur, ob der Knoten lebt, nicht ob die Anwendung antwortet. Ein gestoppter IIS-Anwendungspool bekommt weiter seinen Anteil am Verkehr.

Für Exchange-Server, die Mitglied einer DAG sind, ist NLB nicht unterstützt, weil Windows NLB und Failover-Clustering nicht auf demselben Server laufen dürfen.

Der sichere Weg ist immer: Bestand erfassen, Virtual Services parallel auf neuer IP bauen, testen, per DNS mit kurzer TTL schwenken, Altsystem warm halten, erst dann abbauen.

Warum Windows NLB und IIS ARR in Rente gehören

Bevor du Zeit in eine Migration steckst, solltest du dem Management und dir selbst erklären können, warum. „Das neue Gerät ist schöner“ reicht selten. Die gute Nachricht: Die Gründe sind handfest, und keiner davon ist Geschmackssache.

Windows NLB: Lastverteilung ohne Ahnung von der Anwendung

Network Load Balancing ist ein Filtertreiber, der auf jedem Knoten sitzt. Alle Knoten sehen jedes Paket an die Cluster-IP, und ein Hash-Verfahren entscheidet unabhängig auf jedem Knoten, wer antwortet und wer das Paket verwirft. Das ist elegant, weil es keinen zentralen Verteiler braucht, und tückisch, weil es auch keinen zentralen Verstand gibt. Die Knoten tauschen Heartbeats aus. Fällt ein Knoten aus, merkt das der Rest und konvergiert neu. Fällt aber nur die Anwendung aus, etwa weil ein IIS-Anwendungspool nach einem Update nicht mehr startet, bleibt der Knoten im Cluster und bekommt weiter seinen Anteil an Verbindungen. Die zugehörigen Nutzerinnen und Nutzer sehen eine Fehlerseite, der Rest arbeitet, und dein Helpdesk rätselt, warum es „nur manchmal“ hakt.

Dazu kommt das Verhältnis zu Microsofts eigenen Produkten. Exchange-Server, die Mitglied einer Database Availability Group sind, dürfen nicht gleichzeitig NLB nutzen, weil Windows-NLB und Failover-Clustering nicht auf demselben Server kombiniert werden dürfen. Microsoft empfiehlt seit Exchange 2010 ausdrücklich einen Hardware- oder virtuellen Load Balancer. SharePoint- und RDS-Szenarien mit NLB funktionieren zwar, aber Persistence gibt es nur als Client-IP-Affinität, und die versagt in dem Moment, in dem ein Proxy oder eine NAT-Firewall alle Clients hinter einer Adresse versammelt. Dann landet die gesamte Belegschaft auf einem Knoten, und der andere langweilt sich.

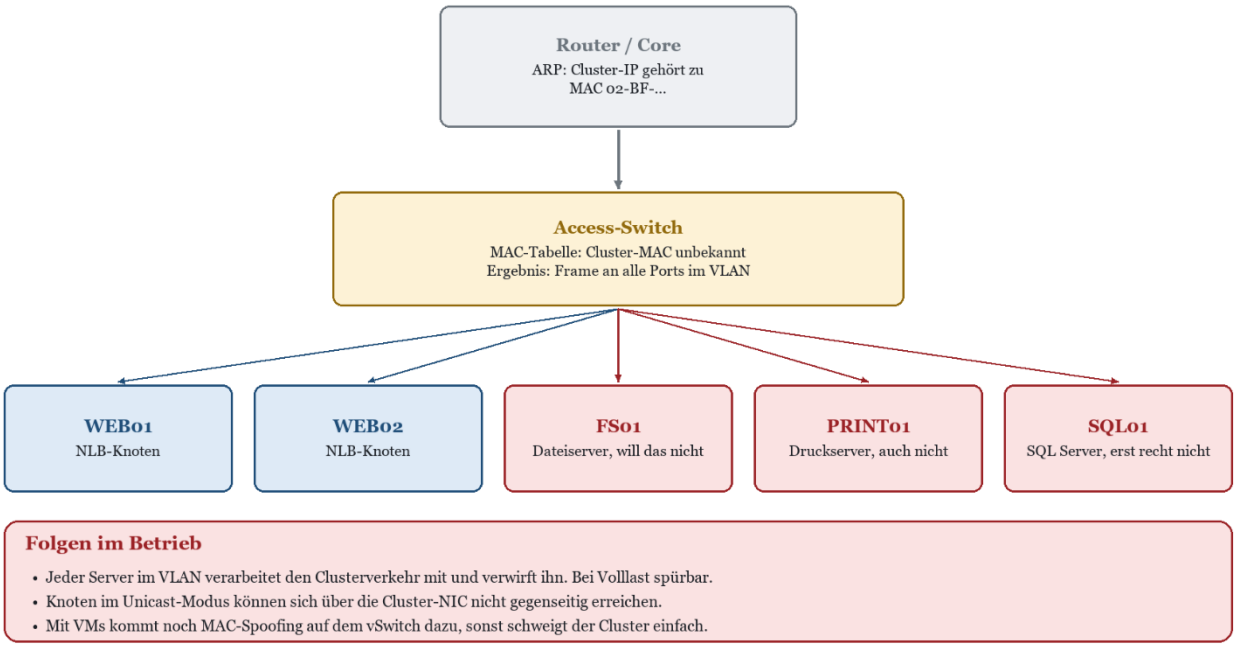
Schließlich die Perspektive: In der Übersicht der Funktionen, die ab Windows Server 2025 nicht mehr weiterentwickelt werden, steht NLB. Als Alternative nennt Microsoft den Software Load Balancer aus dem SDN-Stack, also etwas, das du in einer klassischen Verwaltung mit drei Hyper-V-Hosts eher nicht einführen wirst. Nicht mehr weiterentwickelt heißt nicht abgeschaltet. Es heißt aber: Wenn dir in fünf Jahren eine neue Netzwerkkarte, ein neuer vSwitch oder ein neues Sicherheitsfeature das Leben schwer macht, fragt niemand in Redmond zuerst, ob NLB noch mitkommt.

Unicast, Multicast und das große Fluten

Der Klassiker unter den NLB-Altlasten ist der Unicast-Modus. Er ist der Standard im Assistenten, und genau deshalb läuft er in so vielen Umgebungen. Im Unicast-Modus ersetzt NLB die MAC-Adresse der Cluster-Netzwerkkarte auf allen Knoten durch dieselbe Cluster-MAC. Damit der Switch nicht durcheinanderkommt, maskiert NLB die Absenderadresse ausgehender Frames. Der Switch lernt die Cluster-MAC also nie einem Port zu und tut, was Switches mit unbekannten Zieladressen tun: Er flutet jeden Frame an die Cluster-IP an alle Ports im VLAN. Dateiserver, Druckserver, SQL Server, alles, was im selben Segment hängt, bekommt den Webverkehr der Farm frei Haus und verwirft ihn wieder.

Unicast-NLB: warum der Switch das ganze VLAN beschallt

Alle Knoten tragen dieselbe Cluster-MAC. Der Switch lernt sie nie und flutet jeden Frame



Skizze: vereinfachte Darstellung; Hostnamen frei gewählt

Der Multicast-Modus löst das Fluten nicht, er verlagert es: Die Cluster-IP wird auf eine Multicast-MAC aufgelöst, die viele Router nicht ohne statischen ARP-Eintrag akzeptieren. Mit IGMP-Multicast und IGMP-Snooping am Switch lässt sich das Fluten auf die beteiligten Ports begrenzen, wenn Switch, Router und Netzwerkteam mitspielen. In der Praxis findest du deshalb häufig einen von drei Zuständen: Unicast mit Fluten, das keiner bemerkt hat; Multicast mit statischen ARP-Einträgen auf dem Core-Switch, die keiner dokumentiert hat; oder ein eigenes VLAN nur für den Cluster, das jemand vor Jahren als Notlösung eingerichtet hat. Alle drei sind gute Gründe für den Umzug, und alle drei hinterlassen Spuren, die du nach der Migration aufräumen musst.

WARNUNG · Statische ARP-Einträge überleben ihren Cluster

Hat das Netzwerkteam für einen Multicast-Cluster statische ARP- oder MAC-Einträge auf Router oder Core-Switch angelegt, zeigen die nach der Migration ins Leere. Wird die alte Cluster-IP später neu vergeben, etwa als Virtual-Service-Adresse, schickt der Router den Verkehr stur an die alte Multicast-MAC. Das sieht dann aus wie ein LoadMaster-Fehler, ist aber ein Zettel am Kühlschrank, den keiner abgenommen hat. Frag das Netzwerkteam aktiv nach solchen Einträgen und lass sie im Change mit entfernen.

IIS ARR: der Reverse Proxy, der im eigenen Haus wohnt

Application Request Routing war die Antwort auf die Frage, wie man ohne zusätzliche Hardware einen Reverse Proxy bekommt: Man nimmt einen IIS, installiert URL Rewrite und ARR, legt eine Serverfarm an und lässt Rewrite-Regeln die Anfragen verteilen. Das ging schnell, kostete nichts extra und war für viele Häuser der Ersatz für TMG, als der abgekündigt wurde. Die Version 3.0 erschien 2013. Eine Version 4.0 gibt es nicht.

Was im ersten Jahr praktisch war, ist im zehnten Jahr eine Hypothek. Die Regeln liegen in applicationHost.config oder in web.config-Dateien, teils global, teils auf Sitesebene, teils mit Bedingungen auf Servervariablen, die jemand in einer langen Nacht von einer Foren-Antwort abgeschrieben hat. ARR selbst braucht einen IIS, der gepatcht, überwacht und hochverfügbar gehalten werden will. Für Hochverfügbarkeit des ARR-Servers selbst hat Microsoft früher, Achtung, NLB empfohlen. Du hast also möglicherweise nicht eine Altlast, sondern zwei übereinander.

Dazu kommt Funktionsumfang, der nach heutigen Maßstäben dünn ist: Health Tests prüfen eine URL und optional einen Antwortinhalt, Vorauthentifizierung gibt es nicht, eine Web Application Firewall auch nicht, und TLS-Einstellungen hängen am Betriebssystem statt am Dienst. Für einfache Anwendungen reicht das. Für Exchange, ADFS oder eine Fachanwendung, die im Internet steht, reicht es nicht mehr.

TMG-Erbe: Veröffentlichungen aus einer anderen Zeit

Forefront Threat Management Gateway ist tot, und zwar richtig: Der erweiterte Support für TMG 2010 endete am 14. April 2020. Trotzdem ist TMG in vielen Häusern noch gegenwärtig, nur eben nicht als Server. Es lebt weiter in Firewallregeln, die auf ehemalige TMG-Adressen zeigen, in Exchange-Namespaces, die so geschnitten sind, wie der TMG-Assistent sie damals haben wollte, in Zertifikaten mit genau den Namen, die der Listener brauchte, und in den Köpfen der Kolleginnen und Kollegen, die bis heute von „Weblistener“ und „Veröffentlichungsregel“ sprechen.

Der Kern dessen, was TMG konnte und was Microsoft-Admins vermissen, ist die Vorauthentifizierung: Der Benutzer meldet sich am Rand des Netzes per Formular an, TMG prüft gegen Active Directory und reicht die Anfrage per Kerberos Constrained Delegation an den internen Server weiter. Unauthentifizierter Verkehr erreicht den Exchange-Server gar nicht erst. Genau diese Rolle übernimmt am LoadMaster das Edge Security Pack. Wie das im Detail funktioniert, steht in Edge Security Pack: Vorauthentifizierung, SSO-Domänen und der TMG-Ersatz. Für die Migration reicht hier: Was TMG per Formular und KCD gemacht hat, baust du am LoadMaster mit ESP und SSO-Domäne nach, und die Delegationseinstellungen im Active Directory musst du dabei neu denken, weil das delegierende Konto ein anderes ist.

Kriterium	Windows NLB	IIS ARR	TMG 2010	LoadMaster
Status	nicht mehr weiterentwickelt	letzte Version 3.0 (2013)	Support beendet 2020	aktiv gepflegt
Prüft die Anwendung	nein, nur Knoten	ja, eine URL	ja, einfach	ja, je Virtual Service
Persistence	Client-IP-Affinität	Cookie, Client-IP	Cookie, IP	viele Verfahren
Vorauthentifizierung	nein	nein	ja, Formular und KCD	ja, mit ESP
Nicht-HTTP-Protokolle	ja, Layer 4	nein	eingeschränkt	ja, Layer 4 und 7
Nebenwirkungen im Netz	Fluten, statisches ARP	keine	keine	keine

Mit Exchange-DAG	nicht unterstützt	möglich	historisch üblich	empfohlener Weg

TIPP · Erst entscheiden, ob es überhaupt der LoadMaster sein muss

Nicht jede ARR-Regel braucht einen Hardware-Nachfolger. Eine rein interne Webanwendung mit moderner Authentifizierung ist vielleicht ein Kandidat für Entra Application Proxy, eine Cloud-Anwendung für Azure Application Gateway. Den Entscheidungsbaum dafür findest du in Entscheidungshilfe: LoadMaster, Entra Application Proxy, Azure Application Gateway oder Reverse Proxy?. Für alles mit Kerberos, NTLM, SMTP, RDP oder mehreren Protokollen auf einer Adresse bleibt der LoadMaster die naheliegende Wahl.

Bestandsaufnahme: erst verstehen, dann umziehen

Die Bestandsaufnahme ist der Teil, den alle überspringen wollen, und der Teil, der über Erfolg oder Rückfall entscheidet. Du willst am Ende für jeden Dienst wissen: Unter welchem Namen und welcher IP erreichen ihn die Clients, welche Ports, welche Server stehen dahinter, wie wird Persistence gehalten, woran erkennt man, dass ein Server krank ist, und welche Sonderregeln gibt es. Alles andere ist Hoffnung.

NLB-Cluster per PowerShell ausfragen

Für NLB bringt Windows das Modul NetworkLoadBalancingClusters mit. Get-NlbCluster liefert Cluster-IP, Modus und Cluster-MAC, Get-NlbClusterNode die Knoten mit ihrem Status, Get-NlbClusterPortRule die Port Rules mit Portbereich, Protokoll, Filtermodus und Affinität. Gerade die Port Rules sind spannend, denn die Standardregel des Assistenten deckt die Ports 0 bis 65535 für TCP und UDP ab. Das bedeutet: Der Cluster hat nicht nur 443 verteilt, sondern alles, was je an der Cluster-IP angeklopft hat. Prüf mit den Logs der Firewall oder einer kurzen Netzwerkmitschrift, welche Ports tatsächlich genutzt werden. Sonst stellst du nach dem Schwenk fest, dass irgendein Monitoring-Agent oder eine Altanwendung auf Port 8080 ebenfalls über die Cluster-IP lief.

Notiere außerdem die Affinität je Port Rule. „None“ heißt: keine Persistence. „Single“ entspricht einer Source-IP-Persistence, „Network“ einer Persistence auf das Klasse-C-Netz des Clients. Am LoadMaster bildest du das mit Source-IP-Persistence und gegebenenfalls einer Netzmaske nach, wobei du bei Webanwendungen besser gleich prüfst, ob eine Cookie-basierte Persistence die bessere Wahl ist. Warum, steht in Persistence und Scheduling: wer hängt an welchem Server und warum.

ARR-Farmen und Rewrite-Regeln auslesen

Die ARR-Konfiguration steckt in den Abschnitten webFarms und rewrite der applicationHost.config. Mit appcmd list config -section:webFarms bekommst du die Farmen, ihre Server, den Health Test mit URL, Intervall und erwarteter Antwort sowie die Affinitätseinstellungen. Die Rewrite-Regeln listest du global aus system.webServer/rewrite/globalRules und zusätzlich je Site aus den web.config-Dateien, denn ARR-Farmen werden zwar meist über globale Regeln angesprochen, aber Ausnahmen nisten gern auf Sitesebene.

Jede Regel bekommt in deiner Tabelle eine Zeile mit Muster, Bedingungen, Aktion und, ganz wichtig, dem Zweck. Wenn du den Zweck nicht herausfindest, frag die Fachabteilung oder schalte die Regel im Test ab und schau, was schreit. Typische Funde: Weiterleitung von HTTP auf HTTPS, Umschreiben von Host-Headern, damit die interne Anwendung ihren internen Namen sieht, das Entfernen von Server-Headern, das Umbiegen von Pfaden wie /app auf eine andere Farm, und gelegentlich ausgehende Regeln, die Links im HTML-Body von intern auf extern umschreiben.

Letztere sind die gefährlichsten, weil sie Inhalte verändern und eine Anwendung verdecken, die eigentlich falsch konfiguriert ist.

WICHTIG · Ausgehende Rewrite-Regeln sind ein Symptom

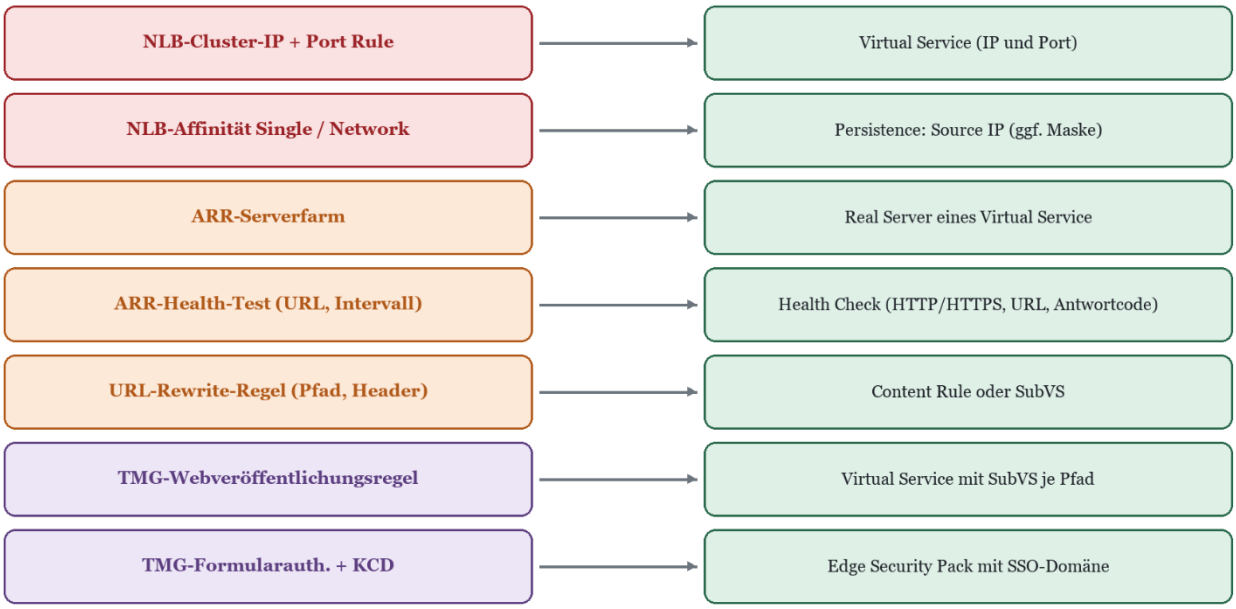
Eine ARR-Regel, die im Antwort-Body http://app01.contoso.local durch https://app.contoso.de ersetzt, repariert eine Anwendung, die ihre eigene externe URL nicht kennt. Am LoadMaster kannst du Antworten ebenfalls per Content Rule umschreiben, aber das kostet Leistung und bricht bei komprimierten Antworten. Besser: die Anwendung so konfigurieren, dass sie die externe URL kennt, etwa über die Alternate Access Mappings in SharePoint oder die Basis-URL der Fachanwendung. Die Migration ist der beste Anlass, solche Pflaster abzuziehen.

TMG-Veröffentlichungen rekonstruieren

Steht der TMG noch als VM im Archiv, exportiere die Konfiguration als XML und nimm dir jede Webveröffentlichungsregel einzeln vor: öffentlicher Name, Pfade, interner Servername, Weiterleitung des Host-Headers, Authentifizierungsdelegation und Listener mit Zertifikat und Authentifizierungsmethode. Ist der TMG längst gelöscht, bleibt Archäologie: Firewall-NAT-Regeln, externe DNS-Einträge, Zertifikats-SANs und die Konfiguration der Nachfolgelösung, die vermutlich ARR war. In beiden Fällen gilt: Du migrierst nicht TMG, du migrierst die Anforderungen, die TMG damals erfüllt hat. Manche davon gibt es nicht mehr, zum Beispiel Outlook Anywhere mit Basic-Authentifizierung von außen.

Was wird woraus? Die Übersetzungstabelle für den Umzug

Jedes Altkonstrukt hat ein Gegenstück am LoadMaster. Nur heißt es anders und denkt anders



Skizze: Orientierung, keine 1:1-Automatik; jede Regel einzeln verstehen, dann neu bauen

CHECKLISTE · Was pro Dienst in die Inventurtabelle gehört

Name im DNS intern und extern, aktuelle IP, TTL des Eintrags, NAT-Regel an der Firewall.

Ports und Protokolle, die tatsächlich genutzt werden, nicht die, die konfiguriert sind.

Real Server mit Namen, Gewichtung und Wartungsstatus.

Persistence-Verfahren und Timeout, Health Check mit URL und erwarteter Antwort.

Zertifikat mit allen SANs, Ablaufdatum, Ort des privaten Schlüssels.

Authentifizierung: anonym, Formular, Kerberos, NTLM, Zertifikat, SAML.

Sonderregeln: Rewrite, Header, Pfadumleitungen, IP-Beschränkungen.

Verantwortliche Person in der Fachabteilung und ein Testfall, der wirklich etwas prüft.

Virtual Services parallel aufbauen

Die goldene Regel jeder Load-Balancer-Migration lautet: niemals die alte IP-Adresse übernehmen, solange das Altsystem noch läuft. Der LoadMaster bekommt für jeden Dienst eine neue Virtual-Service-Adresse. Alt und Neu laufen nebeneinander, auf denselben Real Servern. Der Schwenk passiert durch Umstellen des Namens, nicht durch Umziehen der Adresse. Das hat einen simplen Grund: Solange beide Adressen leben, ist der Rückweg eine DNS-Änderung. Ziehst du die IP um, ist der Rückweg ein Netzwerk-Change mit ARP-Caches, Switch-Tabellen und im NLB-Fall eventuell statischen Einträgen.

Real Server: dieselben Server, zwei Wege

Bei NLB sitzt die Cluster-IP direkt auf den Servern. Das bleibt während der Parallelphase so. Der LoadMaster spricht die Server über ihre eigene, dedizierte IP an, nicht über die Cluster-IP, sonst verteilt NLB hinter dem LoadMaster noch einmal Verbindungen, und du fragst dich, warum dein Health Check von WEB01 eine Antwort von WEB02 bekommt. Prüf außerdem die IIS-Bindungen: Manche Sites sind nur an die Cluster-IP gebunden und antworten auf der Server-IP gar nicht. Dann ergänzt du eine Bindung, bevor du überhaupt einen Real Server anlegst.

Bei ARR ist die Lage umgekehrt: Der ARR-Server ist der Proxy, die Farmserver sind die späteren Real Server. Der LoadMaster spricht sie direkt an, und der ARR-Server wird überflüssig. Achte darauf, dass die Farmserver den LoadMaster als Absender akzeptieren, also IP-Beschränkungen in IIS oder Windows-Firewallregeln, die bisher nur den ARR-Server durchgelassen haben, um die LoadMaster-Adressen ergänzt werden. Wie die echte Client-IP trotz SNAT bei der Anwendung ankommt, steht in Client-IP sichtbar machen: Transparenz, SNAT und X-Forwarded-For.

Health Checks, die mehr können als der Vorgänger

Hier liegt der größte fachliche Gewinn der Migration, und es wäre schade, ihn zu verschenken. Wo NLB gar nicht geprüft hat und ARR nur eine Startseite abgefragt hat, kann der LoadMaster je Virtual Service einen Health Check auf eine URL machen, die wirklich etwas über die Anwendung aussagt. Für Exchange gibt es die healthcheck.htm je virtuellem Verzeichnis, für ADFS den Probe-Endpunkt, für Eigenentwicklungen idealerweise einen Statusendpunkt, der auch die Datenbank prüft. Wie das sauber geht und warum ein offener Port nichts beweist, liest du in Health Checks, die wirklich prüfen: warum „Port 443 offen“ nicht reicht. Für Exchange im Speziellen lohnt sich zusätzlich Exchange Server hinter dem LoadMaster: Templates, Namespaces, Health Checks.

Persistence, TLS und Vorauthentifizierung

Die Persistence übernimmst du nicht blind. Aus „NLB-Affinität Single“ wird nicht automatisch Source-IP-Persistence, sondern du fragst, was die Anwendung braucht. Exchange ab 2016 braucht gar keine, RDS braucht eine bestimmte für das RD Gateway, Webanwendungen mit In-Process-Session brauchen eine, die ein Proxy nicht kaputt macht. Für RDS-Szenarien findest du die Details in Remote Desktop Services: RD Gateway, RD Web und Connection Broker balancieren.

Bei TLS entscheidest du je Dienst zwischen Durchreichen, Offload und Re-Encryption. Unter ARR hat der IIS die Verbindung terminiert und neu aufgebaut, das entspricht Re-Encryption. Das Zertifikat samt privatem Schlüssel muss also auf den LoadMaster, und die Cipher-Einstellungen

solltest du bewusst wählen, statt die des alten Windows Server zu kopieren. Mehr dazu in TLS-Offload, Re-Encryption und Zertifikate: PKI trifft LoadMaster. Und wo TMG vorauthentifiziert hat, kommt ESP ins Spiel, inklusive einer neuen Kerberos-Delegation, die in Kerberos Constrained Delegation am LoadMaster: KCD und RBCD sauber einrichten beschrieben ist.

AUS DER PRAXIS · Die vergessene Port Rule

Eine Verwaltung migriert ihr NLB-Intranet mit zwei IIS-Knoten auf einen Virtual Service für TCP 443. Tests grün, Schwenk am Dienstag, alles ruhig. Am Freitag meldet die Poststelle, dass der Scanner nicht mehr ablegt. Ursache: Der Scanner sprach per FTP mit der Cluster-IP, und die Standard-Port-Rule des NLB-Assistenten hatte FTP all die Jahre brav mitverteilt. Niemand wusste davon, weil es nirgends konfiguriert war. Es war einfach nicht verboten. Lösung: Firewall-Logs der letzten vier Wochen nach Zielports auf die Cluster-IP auswerten, bevor der Change genehmigt wird.

TIPP · Ein Virtual Service je Namespace, SubVS für die Unterschiede

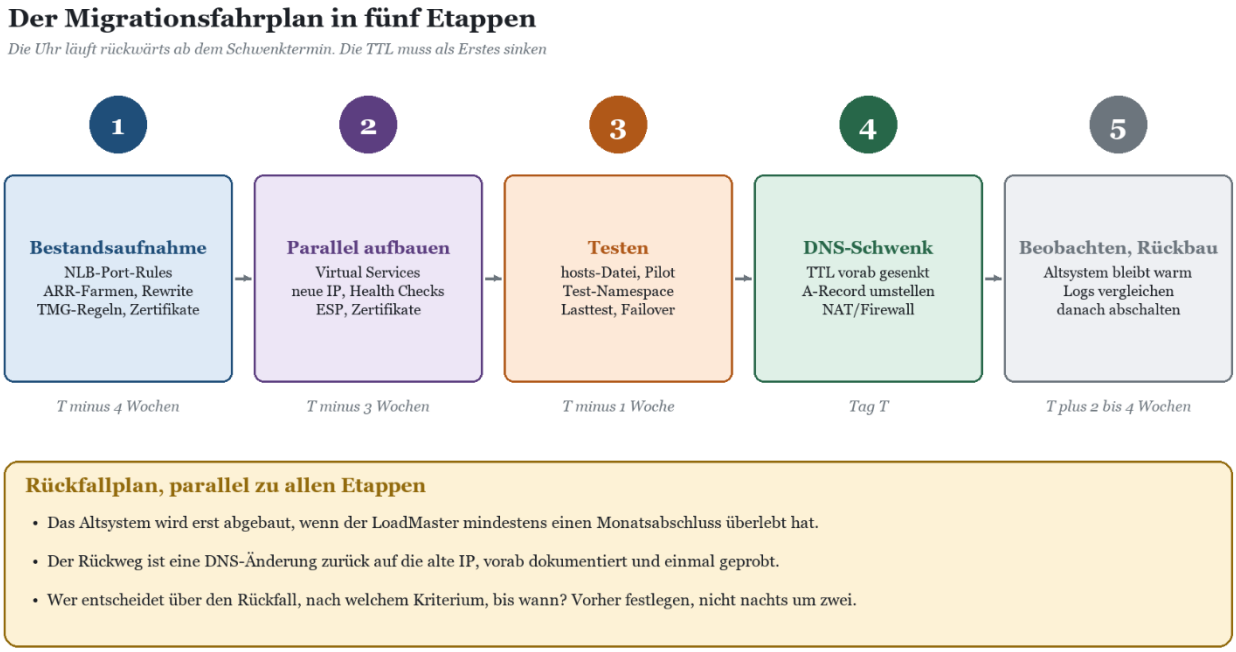
Aus zehn ARR-Regeln mit Pfadbedingungen wird am LoadMaster selten ein Virtual Service mit zehn Content Rules. Lesbarer ist meist ein Virtual Service je öffentlichem Namen und je ein SubVS für Pfade, die andere Server, Health Checks oder Authentifizierung brauchen. Content Rules nutzt du für das, was Content Rules gut können: Header setzen, Pfade umleiten, Antworten verstecken.

Was Content Rules können und wo sie dir das Leben schwer machen, findest du in Content Rules: Header umschreiben, Pfade umleiten, Dinge verstecken. Übersetze ARR-Regeln nie mechanisch, sondern frage bei jeder: Brauche ich das noch, und wenn ja, ist es eine Routing-Entscheidung (SubVS) oder eine Inhaltsänderung (Content Rule)?

Testen, ohne die Produktion anzufassen

Getestet wird gegen die neue Virtual-Service-Adresse, bevor irgendein DNS-Eintrag sich ändert. Drei Wege haben sich bewährt. Erstens ein hosts-Eintrag auf einigen Pilotrechnern, der den produktiven Namen auf die neue IP zeigt. Das ist schnell, aber fehleranfällig, weil hosts-Einträge gern vergessen werden. Zweitens ein Test-Namespace wie test-mail.contoso.de, der vom Zertifikat abgedeckt sein muss und für Kerberos einen eigenen SPN braucht. Drittens bei Exchange und Outlook ein Pilot über eine eigene DNS-Ansicht für ein Testnetz, wenn deine DNS-Infrastruktur das hergibt. Ein Failover-Test gehört dazu: Einen Real Server bewusst stoppen und zusehen, ob der LoadMaster ihn rauswirft und ob die Anwendung das überlebt.

DNS-Schwenk mit kurzer TTL und ein Rückfallplan, der den Namen verdient



Skizze: Zeitangaben als Richtwert für einen mittelgroßen Bestand

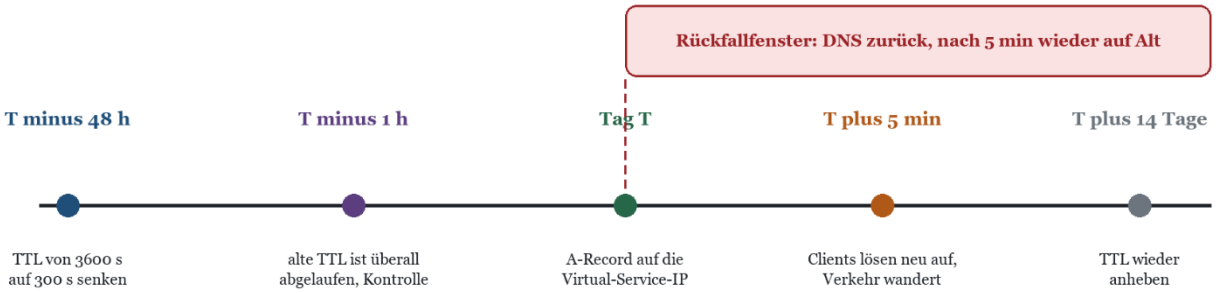
Der Schwenk selbst ist das Unspektakulärste am ganzen Projekt, wenn die Vorarbeit stimmt. Er besteht aus einer DNS-Änderung und gegebenenfalls einer NAT-Regel. Das Spektakuläre passiert, wenn die Vorarbeit fehlt, und am häufigsten fehlt die Vorbereitung der TTL.

Die TTL rechtzeitig senken

Die Time to Live eines DNS-Eintrags legt fest, wie lange Resolver und Clients eine Antwort zwischenspeichern dürfen. Steht dein A-Record mit einer Stunde TTL im Netz, kannst du ihn umstellen, wann du willst: Einige Clients fragen erst in einer Stunde neu. Deshalb senkst du die TTL vorab, mindestens eine alte TTL vor dem Schwenk, besser zwei Tage vorher, auf einen kleinen Wert wie fünf Minuten. In Active-Directory-integrierten Zonen setzt du die TTL je Eintrag in den erweiterten Eigenschaften der DNS-Konsole oder per PowerShell mit Set-DnsServerResourceRecord. Bei externen Zonen beim Provider musst du prüfen, welche Mindest-TTL er überhaupt erlaubt.

DNS-Schwenk mit kurzer TTL: die Zeitachse

Die alte TTL bestimmt, wie früh du anfangen musst. Die neue TTL bestimmt, wie schnell du zurückkannst



Was die TTL nicht regelt

- Bestehende TCP-Verbindungen: Outlook, RDP und Websockets bleiben auf der alten IP, bis sie neu verbinden.
- Clients und Resolver, die TTLs ignorieren oder eigene Mindestwerte setzen. Es gibt sie, leider.
- Externe Namen hinter NAT: Dort schwenkst du oft nicht DNS, sondern die NAT-Regel an der Firewall.
- hosts-Einträge aus der Testphase. Die vergisst jemand garantiert auf genau einem Rechner.

Skizze: Beispielwerte; maßgeblich ist die TTL, die vor der Senkung tatsächlich im Umlauf war

Nach dem Schwenk bleibt die kurze TTL stehen, bis du sicher bist, dass du nicht zurückmusst. Erst nach ein bis zwei Wochen stabilen Betriebs hebst du sie wieder an. Eine dauerhaft sehr kurze TTL erzeugt zwar nur etwas mehr DNS-Last, verschleiert aber auch, wenn mal etwas am Namen falsch ist, weil sich Fehler schneller wieder auswachsen, als jemand sie meldet.

Intern, extern und das Problem mit NAT

Viele Häuser betreiben Split-Brain-DNS: intern zeigt mail.contoso.de auf eine private Adresse, extern auf eine öffentliche, die an der Firewall per NAT auf das Altsystem weitergeleitet wird. Für den internen Teil schwenkst du den A-Record. Für den externen Teil schwenkst du oft gar nicht DNS, sondern das NAT-Ziel an der Firewall von der alten auf die neue interne Adresse. Das wirkt sofort und unabhängig von jeder TTL, trennt aber bestehende Verbindungen, sobald die Firewall ihre Zustandstabelle für die Regel neu aufbaut. Plane das in ein Wartungsfenster und stimme dich mit dem Firewall-Team ab, denn der Rückfall ist dort ebenfalls eine Regeländerung.

WARNUNG · Lange Verbindungen kümmern sich nicht um DNS

Outlook im MAPI-über-HTTP-Modus, RDP-Sitzungen über das RD Gateway, Websockets in Fachanwendungen und Synchronisationsclients halten ihre Verbindungen stunden- oder tagelang. Nach dem DNS-Schwenk laufen sie munter über das Altsystem weiter. Schaltest du das Altsystem zu früh ab, reißen sie alle gleichzeitig ab und verbinden sich dann neu, im besten Fall ohne Rückfrage, im schlechtesten Fall mit einer Anmeldeaufforderung für die gesamte Belegschaft. Lass das Altsystem mindestens bis zum nächsten Arbeitstag aktiv und beobachte die Verbindungszahlen, bevor du es stilllegst.

Der Rückfallplan

Ein Rückfallplan ist kein Satz im Change-Antrag, der „bei Problemen Rückbau“ lautet. Er ist ein Dokument mit drei Antworten: Woran erkennen wir, dass wir zurückmüssen? Wer darf das entscheiden? Und welche Handgriffe in welcher Reihenfolge machen es rückgängig? Das Kriterium sollte messbar sein, etwa „mehr als zehn Störungsmeldungen zu Dienst X in der ersten Stunde“ oder „Health Checks eines Virtual Service länger als 15 Minuten rot“. Die Entscheidung trifft eine

Person, die erreichbar ist und den Mut hat, sie zu treffen. Die Handgriffe schreibst du vorher auf, mit den alten und neuen Werten.

Phase	Altsystem	LoadMaster	Rückweg
Parallelaufbau	produktiv	neue IP, nur Test	nicht nötig
Pilot	produktiv	Pilotclients per hosts oder Test-Namespace	hosts-Eintrag entfernen
Schwenk	läuft weiter, warm	produktiv	A-Record oder NAT zurück, TTL abwarten
Beobachtung	läuft, Verbindungen sinken	produktiv	wie Schwenk, noch problemlos
Rückbau	abgeschaltet, Konfig archiviert	produktiv	nur noch Neuaufbau, daher spät

Wichtig ist die letzte Zeile: Mit dem Rückbau des Altsystems endet dein Rückfallplan. Deshalb kommt der Rückbau nicht in derselben Woche, sondern nach einem Zeitraum, in dem die seltenen Vorgänge einmal stattgefunden haben: Monatsabschluss, Zeugnisdruck, Haushaltsabschluss, das Quartalsreporting oder was auch immer in deinem Haus nur alle paar Wochen über genau diesen Dienst läuft. Danach entfernst du die NLB-Rolle sauber, nimmst die Cluster-IP aus den Netzwerkkarten, lässt statische ARP-Einträge im Netz löschen, deinstallierst ARR und URL Rewrite dort, wo sie nicht mehr gebraucht werden, und gibst die alten IP-Adressen im IPAM frei.

FAKTEN · Reihenfolge für den Rückbau einer NLB-Umgebung

Erst die Knoten aus dem Cluster nehmen, dann den Cluster entfernen, dann das Feature deinstallieren. Wer nur das Feature deinstalliert, kann Netzwerkkarten mit der Cluster-MAC zurücklassen.

Hyper-V: MAC-Spoofing an den vNICs der ehemaligen Knoten wieder abschalten, sofern es nur für NLB aktiviert war.

Netz: statische ARP- und MAC-Einträge für Cluster-IP und Cluster-MAC entfernen lassen, IGMP-Snooping-Ausnahmen prüfen, ein eventuelles Cluster-VLAN zurückbauen.

DNS: alte Einträge für Cluster-Namen löschen, TTLs der migrierten Namen wieder anheben.

FAQ

Kann ich die alte NLB-Cluster-IP für den Virtual Service übernehmen?

Technisch ja, aber nicht während des Parallelbetriebs. Wenn du das willst, schwenkst du zuerst per DNS auf eine neue Adresse, baust NLB vollständig zurück, lässt ARP-Einträge im Netz bereinigen und ziehst dann in einem zweiten Change auf die alte IP um. Meist ist der Aufwand größer als der Nutzen, außer externe Partner haben die IP fest in ihren Firewalls hinterlegt.

Brauche ich nach der Migration noch NLB auf den Servern?

Nein. Die Lastverteilung übernimmt der LoadMaster, die Hochverfügbarkeit des LoadMasters selbst ein HA-Paar. NLB hinter einem Load Balancer ist doppelte Verteilung und macht Fehlersuche unnötig spannend.

Lassen sich ARR-Regeln automatisch in Content Rules übersetzen?

Nicht sinnvoll. Die Regelsprachen sind unterschiedlich, und viele ARR-Regeln haben einen Zweck, der am LoadMaster mit einem SubVS, einer Umleitung oder gar nicht mehr gelöst wird. Übersetze den Zweck, nicht die Syntax.

Was ersetzt am LoadMaster die TMG-Formularanmeldung?

Das Edge Security Pack mit einer SSO-Domäne, die gegen Active Directory authentifiziert, und je nach Ziel einer Weitergabe per Kerberos Constrained Delegation oder Basic. Für Exchange OWA und ECP lässt sich auch moderne Authentifizierung mit Entra ID und MFA anbinden.

Wie kurz sollte die TTL vor dem Schwenk sein?

Fünf Minuten haben sich bewährt. Kürzer bringt selten etwas, weil manche Resolver ohnehin Mindestwerte setzen. Entscheidend ist, dass du sie mindestens eine alte TTL, besser zwei Tage vor dem Schwenk senkst.

Gibt es eine Microsoft-eigene Alternative zu NLB?

Microsoft nennt den Software Load Balancer aus dem SDN-Stack. Der setzt eine SDN-Infrastruktur mit Network Controller voraus und ist für die klassische Umgebung im Mittelstand oder in einer Verwaltung in der Regel überdimensioniert. Für veröffentlichte Webanwendungen kommen außerdem Entra Application Proxy und Azure Application Gateway in Betracht.

Fazit: Rente mit Übergabeprotokoll

Windows NLB, IIS ARR und die Reste von TMG haben ihren Dienst getan, aber jedes dieser Konstrukte hat seine eigene Art, still zu versagen: NLB, weil es die Anwendung nicht sieht und das Netz flutet, ARR, weil es auf dem Server wohnt, den es schützen soll, und TMG, weil es seit Jahren keine Sicherheitsupdates mehr bekommt. Der LoadMaster ist kein Wundermittel, aber er macht diese drei Schwächen zu konfigurierbaren Eigenschaften: Health Checks statt Heartbeat, Content Rules statt Rewrite-Wildwuchs, ESP statt Weblistener.

Die Migration gelingt, wenn du sie langweilig machst. Bestandsaufnahme bis in die letzte Port Rule, Virtual Services auf neuen Adressen parallel zum Altsystem, Tests gegen die neue IP, eine TTL, die zwei Tage vorher gesenkt wurde, und ein Rückfallplan, der aufgeschrieben ist, bevor ihn jemand braucht. Wenn dir für einen der Schritte das Wissen oder die Zeit fehlt, unterstützt dich die Beratung zum Kemp LoadMaster bei Inventur und Schwenkplan, und in der Kemp-LoadMaster-Schulung baust du die Virtual Services vorher im Labor nach. Das Schönste an einer gut gemachten Migration ist ohnehin der Moment danach: Wenn niemand merkt, dass sich etwas geändert hat, und der alte Server endlich das bekommt, was er sich verdient hat. Einen würdevollen Shutdown.

WEITERLESEN · Weiter im Thema

› Virtual Service, Real Server, SubVS: das Grundvokabular des LoadMasters

› Edge Security Pack: Vorauthentifizierung, SSO-Domänen und der TMG-Ersatz

› Entscheidungshilfe: LoadMaster, Entra Application Proxy, Azure Application Gateway oder Reverse Proxy?

› Health Checks, die wirklich prüfen: warum „Port 443 offen“ nicht reicht

› Content Rules: Header umschreiben, Pfade umleiten, Dinge verstecken

› Persistence und Scheduling: wer hängt an welchem Server und warum