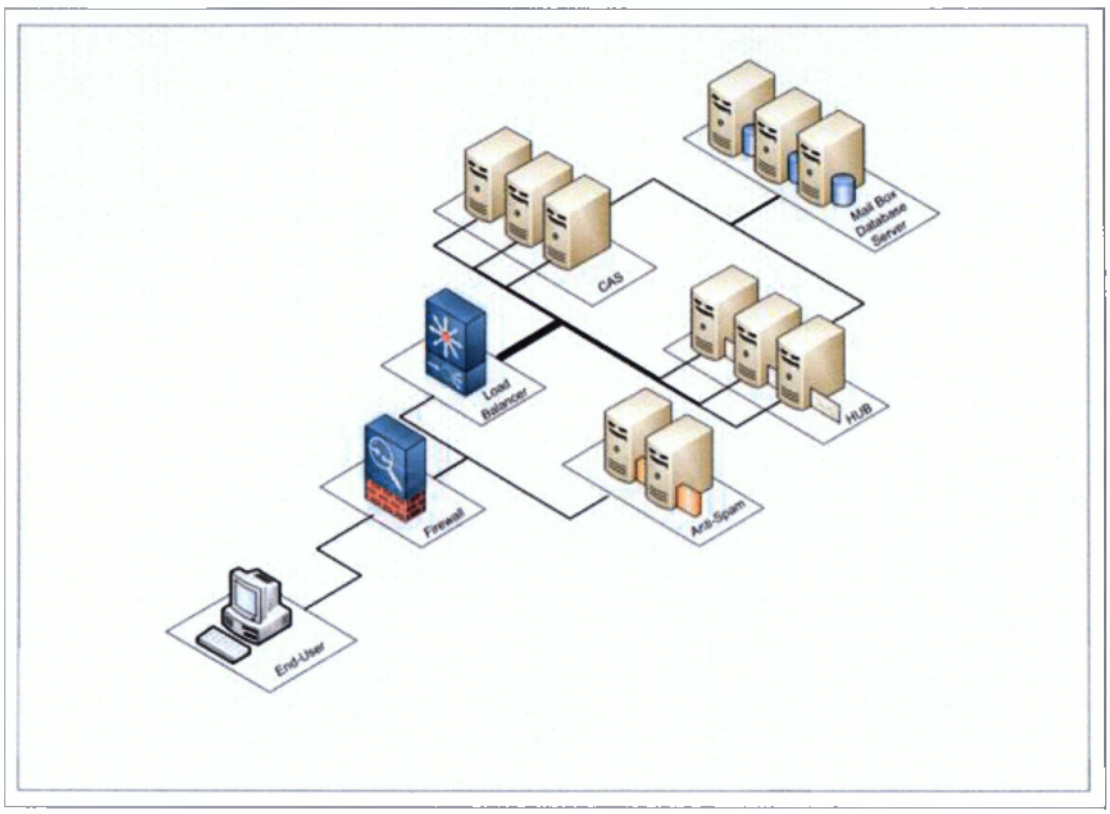
Τ.Ε.Ι. ΠΕΛΟΠΟΝΝΗΣΟΥ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.



ΤΙΤΛΟΣ:

ΤΕΧΝΟΛΟΓΙΕΣ ΑΝΑΚΑΤΑΝΟΜΗΣ ΦΟΡΤΟΥ "LOAD BALANCING"ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΜΗΧΑΝΙΣΜΩΝ ΑΠΟΤΥΧΙΑΣ "FAIL OVER" ΓΙΑ ΕΠΙΤΕΥΞΗ ΥΨΗΛΗΣ ΔΙΑΘΕΣΙΜΟΤΗΤΑΣ ΣΕ ΟΜΑΔΕΣ "CLUSTERS" ΕΞΥΠΗΡΕΤΗΤΩΝ

ΟΝΟΜΑ ΕΠΙΒΛΕΠΟΝΤΑ : ΜΠΑΡΔΗΣ ΓΕΩΡΓΙΟΣ
ΟΝΟΜΑ ΣΠΟΥΔΑΣΤΗ : ΜΥΤΑΡΑΣ ΓΕΩΡΓΙΟΣ (2008067)
ΟΝΟΜΑ ΣΠΟΥΔΑΣΤΗ : ΜΠΑΓΙΑΡΤΑΚΗΣ ΝΙΚΟΛΑΟΣ (2008032)

Σπάρτη, 2013

Ευχαριστίες:

Θα θέλαμε να ευχαριστήσουμε ιδιαιτέρως τον καθηγητή μας Μπάρδη Γιώργο, η συμβολή του οποίου υπήρξε καθοριστική για την ολοκλήρωση της πτυχιακής μας μελέτης, καθώς επίσης και τις οικογένειες μας για την υποστήριξή τους καθ' όλη τη διάρκεια ολοκλήρωσης των σπουδών μας.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Εισαγωγή

Πολλές εταιρείες διαθέτουν σημαντικό αριθμό υπολογιστών σε λειτουργία τόσο σε μικρή απόσταση όσο και σε μεγάλες αποστάσεις μεταξύ τους. Έστω ότι σε κάθε υποκατάστημα υπάρχει υπολογιστής για την τήρηση των στοιχείων της αποθήκης, για την παρακολούθηση των λογαριασμών των πελατών, για την εξαγωγή της μισθοδοσίας του προσωπικού και άλλα.[1]

Από την δικτύωση η εταιρεία θα κέρδιζε άμεση ενημέρωση για κεντρική λήψη αποφάσεων, χαμηλότερο κόστος προμηθειών και λειτουργίας και σωστή κατανομή του ανθρώπινου δυναμικού.

Πιο αναλυτικά, η παρούσα μελέτη πρόκειται να ασχοληθεί με το ζήτημα των τεχνολογιών ανακατανομής φόρτου "Load Balancing" και υλοποίησης μηχανισμών αποτυχίας προκειμένου να επιτευχθούν υψηλά επίπεδα διαθεσιμότητας σε ομάδες εξυπηρετητών.[2]

[1] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[2] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

Κεφάλαιο 1

Τεχνολογίες Διαδικτύου

1.1.Αρχιτεκτονική υπολογιστικών συστημάτων

Η αρχιτεκτονική υπολογιστών έχει σημειώσει μεγάλη άνοδο και εξέλιξη σε σχέση με τις εφαρμογές της. Η απλούστερη μορφή ήταν η αρχιτεκτονική υπερυπολογιστή (Mainframe Architecture) όπου όλες οι διαδικασίες εμπεριέχονταν στον κεντρικό υπολογιστή (mainframe).[3]

Οι χρήστες επικοινωνούσαν με τον κεντρικό Η/Υ μέσα από τους τερματικούς σταθμούς που μεταδίδονταν οι εντολές και εμφανίζονταν τα αποτελέσματα σε αυτόν που το χρησιμοποιούσε. Οι εν λόγω εφαρμογές, ήταν κάπως αργές και εξαιτίας της ανάγκης για μετάδοση εντολών στον βασικό υπολογιστή. [4]

Με την εισαγωγή του (PC), ο οποίος έχει υπολογιστική ισχύ υπάρχει η δυνατότητα εξέλιξης και σε σχέση με την εξάπλωση των δικτυωμένων συστημάτων, φτάσαμε στον 2ο τύπο αρχιτεκτονικής υπολογιστικών συστημάτων της ¨κοινής χρήσης αρχείων¨(File sharing).[5]

Η εν λόγω αρχιτεκτονική εργάζεται σωστά όταν η κοινή χρήση είναι σε μειωμένα επίπεδα, τα αιτήματα ενημερώσεων είναι λίγα και ο όγκος των στοιχείων που μεταφέρεται είναι μικρός. Βέβαια, γρήγορα έγινε σαφές ότι η κοινή χρήση αρχείων ¨έπνιξε¨ τα δίκτυα όταν αυτά μεγάλωσαν σε έκταση, ταυτόχρονα με τις εφαρμογές να απαιτούν όλο και μεγαλύτερες ποσότητες στοιχείων που πρέπει να διαβιβάζονται και στις δύο κατευθύνσεις. [6]

---

[3] Quagga, a software routing suite, http://www.quagga.net

[4] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[5] Quagga, a software routing suite, http://www.quagga.net

[6] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

Η δυσκολία στο χειρισμό τεράστιου όγκου πληροφοριών στην αρχιτεκτονική κοινής χρήσης αρχείων, κατεύθυνε προς στην ανάπτυξη της αρχιτεκτονικής πελάτη-διακομιστή (client-server) στις αρχές του 1980, που αντικαταστάθηκε από τον διακομιστή βάσης στοιχείων (database server)[7]

Η απάντηση σε συγκεκριμένο αίτημα μειώθηκε έντονα με τη χρήση του δικτύου, πράγμα που ενέκρινε την ανάπτυξη εφαρμογών, σε συγκεκριμένη βάση στοιχείων. Η επικοινωνία μεταξύ του πελάτη-διακομιστή γινόταν σε δομημένη γλώσσα αναζητήσεων (SQL) ή μέσω κλήσεων Απομακρυσμένων Διεργασιών (RPCS). [8]

1.1.1.Παγκόσμιος ιστός

Το 1990 ο Τιμ Μπέρνερς Λι κατασκεύασε τον παγκόσμιο ιστό (world wide web) εκ μέρους του Ευρωπαϊκού Οργανισμού Πυρηνικής Έρευνας (CERN), που ήταν μια εφαρμογή του διαδικτύου (internet). Το πακέτο περιελάμβανε το πρωτόκολλο επικοινωνίας HTTP, την γλώσσα HTML, το πρόγραμμα του περιηγητή (web browser) μέσα στο οποίο περιελαμβανόταν η λειτουργία συντάκτη (editor) και το λογισμικό εξυπηρετητή ιστού (http server). [9]

1.1.2.Λειτουργία παγκόσμιου ιστού

Σήμερα ο παγκόσμιος ιστός είναι η πιο γνωστή υπηρεσία του ίντερνετ λόγω της εύκολης χρήσης της και του πλήθους των πληροφοριών που μας παρέχει. Η διαδικασία αποτύπωσης της ιστοσελίδας στον περιηγητή αρχίζει όποτε ο χρήστης έχει εισάγει την διεύθυνση (URL) η οποία αποτελείται από τα εξής πεδία : [10]

---

[7] Quagga, a software routing suite, http://www.quagga.net

[8] Quagga, a software routing suite, http://www.quagga.net

[9] Quagga, a software routing suite, http://www.quagga.net

[10] High Availability, http://www.linux-ha.org

scheme://domain:port/path?query_string#fragment_id

α) scheme είναι το πρωτόκολλο επικοινωνίας π.χ. http,https,ftp κτλ

β) domain είναι το όνομα χώρου. Το πεδίο είναι δυνατόν να δεχθεί και την IP διεύθυνση π.χ. google.com ή 72.14.207.99

γ) port είναι ο αριθμός θύρας. Το πεδίο είναι προαιρετικό. Ανάλογα με την υπηρεσία ο περιηγητής χρησιμοποιεί την προεπιλεγμένη θύρα της υπηρεσίας π.χ. 80 για http, 443 για https,21 για ftp κτλ [11]

δ) path θεωρείται η διαδρομή μέσω της οποίας βρίσκουμε τα αιτούμενα δεδομένα.

ε) query string θεωρείται τα δεδομένα που στέλνουμε στην εφαρμογή που "τρέχει" στον εξυπηρετητή, που είναι δυνατόν να είναι ζευγάρια ονόματος-τιμής. π.χ ?view=sections&alias=cars

στ) fragment identifier είναι το αναγνωριστικό ορίου το οποίο μας οδηγεί σε συγκεκριμένο τομέα στο κείμενο HTML. [12]

Το αίτημα κατευθύνεται από τον περιηγητή στον διακομιστή ακολουθώντας με τα εξής βήματα : α) αν το αίτημα περιλαμβάνει 5 ονόματα του χώρου αυτό μεταφράζεται σε διεύθυνση IP μέσω της υπηρεσίας DNS, β) το αίτημα στην περίπτωση μας HTTP αποστέλλεται στον διακομιστή και γ) ο διακομιστής ανάλογα με το αίτημα πραγματοποιεί όλες τις διαδικασίες που απαιτούνται για την επιστροφή του εξαγόμενου αποτελέσματος στον περιηγητή[13]

[11] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[12] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[13] Quagga, a software routing suite, http://www.quagga.net

1.1.3.Αρχιτεκτονική πελάτη-διακομιστή δύο επιπέδων

Η αρχιτεκτονική καταναλωτή και διακομιστή υλοποιείται μέσω του λογισμικού του περιηγητή που αποστέλλει και λαμβάνει δεδομένα από τον διακομιστή. [14]

Τα αιτήματα πηγάζουν από το λογισμικό εξυπηρέτησης ιστού (Apache Http Server).

1.1.4.Αρχιτεκτονική πελάτη-διακομιστή τριών επιπέδων

Εφόσων οι πόροι της συσκευής του εξυπηρετητή δεν επαρκούν εξαιτίας της μεγάλης χρήσης από το λογισμικό, οι δύο υπηρεσίες (λογισμικό εξυπηρέτησης ιστού, βάσης στοιχείων) χωρίζονται σε δύο διαφορετικές συσκευές. [15]

Το πέρασμα από την προηγούμενη αρχιτεκτονική στην εν λόγω αρχιτεκτονική, υλοποιείται χωρίς ουσιαστικές αλλαγές στην εφαρμογή που εκτελείται από το λογισμικού εξυπηρέτησης ιστού (σε γλώσσα PHP). Η μοναδική διαφοροποίηση που γίνεται είναι στον πελάτη της ΣΒΔΜ (Σχεσιακής Βάσης Στοιχείων) στην γλώσσα PHP όπου αλλάζει την διεύθυνση δικτύου στην νέα διεύθυνση της συσκευής του εξυπηρετητή βάσης στοιχείων (Database Server). [16]

1.1.5.Αρχιτεκτονική πελάτη-διακομιστή ν επιπέδων

Σε αρκετές περιπτώσεις η αρχιτεκτονική 3 επιπέδων δεν είναι αρκετή για την εξυπηρέτηση μεγάλου αριθμού πελατών. Η πρόσθεση συσκευών είναι

---

[14] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[15] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[16] Quagga, a software routing suite, http://www.quagga.net

συνήθως οριζόντια για την κατανομή του φορτίου σε περισσότερες συσκευές εξυπηρετητών. [17]

1.1.6.Εξισορρόπηση φορτίου

Το λογισμικό που είναι σε χρήση (squid) εισάγει τις λειτουργίες της αντίστροφης λειτουργίας μεσολαβητή (reverse proxy) και την προσωρινή μνήμη αυτού (reverse proxy cache). Όποτε δεχθεί ένα αίτημα από τον πελάτη αποστέλει το δικό του αίτημα σε κάποιον από τους εξυπηρετητές εφαρμογής μέσα από τον αλγόριθμο χρονοπρογραμματισμού εξυπηρέτησης εκ περιτροπής (round robin). Αν η διαδικασία πραγματοποιηθεί, τότε και η σελίδα αποθηκεύεται στην προσωρινή μνήμη ώστε να χρησιμοποιηθεί αργότερα για το ίδιο αίτημα. Στην εν λόγω λειτουργία είναι 4 παράμετροι σχετικές με τις σελίδες που σώζονται στην προσωρινή μνήμη και παίζουν καθοριστικό ρόλο. [18] [19]

1.2.Εξυπηρετητής εφαρμογής

Το λογισμικό προσωρινής μνήμης memcache, χρησιμοποιείται για την αποθήκευση αποτελεσμάτων από την βάση στοιχείων. Η λειτουργία του στηρίζεται σε μία κοινή εικονική μνήμη (pool) με αποτέλεσμα τη μνήμη 3 εξυπηρετητών με ποσότητα του κάθε ένα 4GB, σε σύνολο 3x4=12GB, όπου έχουν πρόσβαση και οι τρεις εξυπηρετητές. [20]

Κάτι τέτοιο χρησιμεύει στην εν λόγω αρχιτεκτονική ως εξής: π.χ. Έχουμε δύο αιτήματα που αφορούν την ίδια σελίδα. Το πρώτο αίτημα κατευθύνεται από τον εξισορροπητή φορτίου στον πρώτο εξυπηρετητή

[17] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[18] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[19] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[20] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

εφαρμογής και το δεύτερο αίτημα στον δεύτερο εξυπηρετητή. Ο αρχικός αναζητά το αποτέλεσμα της κλήσης της βάσης στοιχείων, καταρχήν στην εικονική μνήμη (μέσω της εφαρμογής) και έπειτα αν δεν υπάρχει, πραγματοποιεί την κλήση και αποθηκεύει το αποτέλεσμα στην μνήμη. Ο 2ος εξυπηρετητής ψάχνει το αποτέλεσμα στην εικονική μνήμη.

Ως αποτέλεσμα είναι η ελάττωση των αιτημάτων στην βάση στοιχείων και η ταχύτατη απόκριση των αιτημάτων, λόγω της αποθήκευσης των αποτελεσμάτων στην μνήμη RAM. [21]

1.3.Εξυπηρετητές βάσης στοιχείων

Η λειτουργία αντιγράφων ονομάζεται η τεχνολογία που πραγματοποιεί την αντιγραφή στοιχείων σε αρκετούς εξυπηρετητές βάσης στοιχείων. Ο τρόπος με τον οποίο υλοποιείται αυτό είναι ο εξής: ένας εξυπηρετητής ορίζεται ως βασικός (master) και αποτυπώνει κάθε εκτέλεση αιτήματος σε ένα αρχείο ιστορικού (binary log). Οι υπόλοιποι εξυπηρετητές λειτουργούν ως υποκείμενοι (slaves) και αιτούνται πληροφορίες εκτέλεσης από το αρχείο ιστορικού στον κύριο εξυπηρετητή. Ο βασικός εξυπηρετητής δεν γνωρίζει πόσοι υποκείμενοι εξυπηρετητές υπάρχουν, απλά επιστρέφει απαντήσεις αιτημάτων στους εξυπηρετητές πού έχουν το δικαίωμα να τις πραγματοποιούν. Η ενέργεια των υποκείμενων εξυπηρετητών είναι ως εξής:[22] τα αιτήματα ιστορικού που έχουν μεταδοθεί από τον κύριο εξυπηρετητή, σώζονται στο ¨μεταβιβασμένο¨ αρχείο ιστορικού (relay log) από το οποίο εκτελούνται οι διεργασίες στην βάση στοιχείων του εξυπηρετητή. Διεργασίες αιτημάτων στην βάση, οι οποίες είναι μεγάλες σε χρόνο, μπορούν να έχουν σαν αποτέλεσμα, η αναπαραγωγή στους υποκείμενους εξυπηρετητές να

[21] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

[22] RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html

υστερεί έναντι του κύριου εξυπηρετητή. Η παραπάνω λειτουργία αφορά την ασύγχρονη σχέση κυρίου-υποκείμενων εξυπηρετητών. [23]

Στη βάση στοιχείων Mysql υποστηρίζεται η ημισύγχρονη (semi-synchronous) λειτουργία αντιγράφων. Σε αυτή την λειτουργία ο κύριος εξυπηρετητής δεν επιστρέφει αποτέλεσμα για το αίτημα στην βάση στοιχείων, το οποίο έχει πραγματοποιηθεί, αν τουλάχιστον ένας υποκείμενος εξυπηρετητής δεν έχει αποθηκεύσει στο αρχείο ιστορικού του την διεργασία αυτή. Στις πιο πολλές περιπτώσεις χρειαζόμαστε μεγαλύτερη ταχύτητα απόκρισης, σε σχέση με την ποσότητα των αιτημάτων και αυτό επιτυγχάνεται με την διασπορά των αιτημάτων ανάγνωσης σε αρκετούς υποκείμενους εξυπηρετητές. [24]

1.4.Συστοιχίες μεγάλης διαθεσιμότητας

Συστοιχίες μεγάλης διαθεσιμότητας καλούνται οι ομάδες εξυπηρετητών που είναι δυνατόν να χρησιμοποιηθούν σε ελάχιστο χρόνο μη διαθεσιμότητας. Υπάρχουν εξυπηρετητές που ενεργοποιούνται σε λειτουργία όταν κάποια μέρη του συστήματος αποτυγχάνουν. Όποτε σ' ένα σύστημα έχουμε αποτυχία υλικού ή λογισμικού το σύστημα δυσλειτουργεί ή αποτυγχάνει. Η πιο μικρή σε έκταση συστοιχία μεγάλης διαθεσιμότητας είναι των δύο κόμβων, αλλά πολλές συστοιχίες διαθέτουν αρκετούς περισσότερους κόμβους. Ανάλογα με τον τρόπο λειτουργίας κατηγοριοποιούνται ως εξής [25]:

α) ενεργός προς ενεργό κόμβο (active/active).

β) Ενεργός προς παθητικό κόμβο (active/passive).

---

[23] Ldirectord, http://www.vergenet.net/linux/ldirectord/

[24] Quagga, a software routing suite, http://www.quagga.net

[25] Quagga, a software routing suite, http://www.quagga.net

γ) N+1. Παρέχει μόνο έναν επιπλέον κόμβο ο οποίος αναλαμβάνει τον ρόλο του κόμβου που αποτυγχάνει.

δ) N+M. Στις περιπτώσεις που η συστοιχία διατηρεί πολλές υπηρεσίες το μοντέλο N+1 δεν είναι αρκετό για να παρέχει υψηλή διαθεσιμότητα.

Οι βασικοί εξυπηρετητές βάσης στοιχείων συνδέονται μεταξύ τους με ζεύξη τύπου DRBD η οποία έχει το ρόλο της λειτουργίας αντιγράφων. Το λογισμικό της βάσης στοιχείων είναι ενεργό μόνο στον ενεργό εξυπηρετητή. Αν υπάρξει αποτυχία ο παθητικός κόμβος γίνεται ενεργός και ξεκινά η υπηρεσία της βάσης στοιχείων. Οι βασικοί εξυπηρετητές βάσης στοιχείων λειτουργούν σύμφωνα με το μοντέλο ενεργού παθητικού κόμβου ενώ οι υποκείμενοι σύμφωνα με το μοντέλο N προς N. [26]

1.4.1.Αποθήκευση στοιχείων σε συστοιχία μεγάλης διαθεσιμότητας

Προκειμένου να υπάρχει κατοχύρωση σε δεδομένα και εξυπηρετητές, υπάρχει η τεχνολογία συστοιχίας ανεξάρτητων δίσκων (RAID). Μέσα από λογισμικά οι σκληροί δίσκοι κρατούν στη μνήμη αντίγραφα των πληροφοριών σε αρκετές τοποθεσίες. Όταν υπάρχει αστοχία υλικού του εξυπηρετητή οι πληροφορίες δεν είναι προσβάσιμες. Η τεχνολογία που υπάρχει για να υποθηκεύσει δεδομένα μεγάλης διαθεσιμότητας είναι το κατανεμημένο σύστημα αρχείων (Distributed File System). [27]

1.4.2.Hadoop

Το εν λόγω σύστημα αρχείων είναι κατανεμημένο, σε γλώσσα προγραμματισμού JAVA. Οι κόμβοι στοιχείων (data nodes) αποθηκεύουν δεδομένα σε τμήματα (blocks), ανάλογα με τις οδηγίες του κόμβου

[26] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[27] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

ονοματοδοσίας. Το κατάλληλο μέγεθος είναι πολλαπλάσιο των 64MB και υποστηρίζει συμπίεση στοιχείων τύπου bzip2. [28]

1.4.3.MogileFS

Η συστοιχία μεγάλης διαθεσιμότητας του συστήματος αρχείων MogileFS αποτελεί κατανεμημένο σύστημα αρχείων, και όπως και το σύστημα hadoop αποθηκεύει τα δεδομένα σε τουλάχιστον τρεις κόμβους στοιχείων, για κατοχύρωση, προωθεί τη συμπίεση των στοιχείων bzip2 και είναι σχεδιασμένο για την αποθήκευση στοιχείων μεγάλης χωρητικότητας (τμήματα των 128MB) χωρίς να χρειάζεται η χρησιμοποίηση συστοιχίας ανεξάρτητων δίσκων[29]

1.4.4.GlusterFS

Καλείται ένα επεκτάσιμο σύστημα αρχείων που χρησιμοποιείται ή για ταχύτητα μετάδοσης στοιχείων , ή για κατοχύρωση διατηρώντας πολλαπλά αντίγραφα (replicating) στους κόμβους αποθήκευσης. Ουσιαστικά ο κάθε κόμβος αποθήκευσης εξάγει το τοπικό σύστημα αρχείων σαν τόμο (volume). Οι τόμοι αποτελούν τελικό τόμο που είναι το άθροισμα του συνόλου των τόμων των κόμβων αποθήκευσης. [30]

1.4.5.Lustre

Η μεγάλη διαθεσιμότητα των κόμβων αποθήκευσης στοιχείων στοιχείων δεν υλοποιείται μέσω του συστήματος, αλλά με την βοήθεια π.χ. του συστήματος DRBD. Στους κόμβους αποθήκευσης είναι δυνατόν να χρησιμοποιηθεί η δυνατότητα κατανεμημένων αντιγράφων για την αύξηση της ταχύτητας απολαβής των στοιχείων, όμως την κατοχύρωση των στοιχείων

---

[28] Failover Clustering in Windows Server 2008 R2, *Microsoft White Paper*, April 2009

[29] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[30] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

πρέπει να αναλάβει σύστημα αποτυχίας δίσκων (disk failover). [31]

Κεφάλαιο 2

Σύγκριση του PowerPath με τις ενδογενείς λύσεις

2.1.Εξισορρόπηση φορτίου

2.1.1.Άδεια PowerPath Multipathing

Το PowerPath υποστηρίζει έως και 32 διαδρομές από πολλαπλά HBA (iSCSI TOEs ή FCoE CNAs) σε πολλαπλές θύρες αποθήκευσης όταν εφαρμόζεται η άδεια για δρομολόγηση πολλαπλών διαδρομών (multipathing). Χωρίς την άδεια για multipathing, το PowerPath θα χρησιμοποιήσει μόνο μία θύρα ενός προσαρμογέα. Αυτό είναι γνωστό ως "PowerPath SE" ή "χωρίς άδεια PowerPath". Σε αυτή τη λειτουργία, η μόνη ενεργή θύρα μπορεί να χωριστεί σε ζώνες κατ' ανώτατο όριο σε δύο θύρες αποθήκευσης. Η ρύθμιση αυτή παρέχει μόνο ανακατεύθυνση της θύρας αποθήκευσης και όχι εξισορρόπηση φορτίου με βάση τον κεντρικό υπολογιστή (host) ή ανακατεύθυνση με βάση τον κεντρικό υπολογιστή (host).

Αυτή η ρύθμιση υποστηρίζεται, αλλά δεν συνιστάται αν ο πελάτης θέλει πραγματική εξισορρόπηση φορτίου συσκευών I/O (Εισόδου/Εξόδου) στον host και ανακατεύθυνση HBA. Η έξυπνη δρομολόγηση των I/O του PowerPath και η ανακατεύθυνση HBA μπορούν να επιτευχθούν μόνο όταν το λογισμικό έχει άδεια.

Σημείωση: το PowerPath/VE για το vSphere δεν υποστηρίζεται για λειτουργία χωρίς άδεια

2.2.Εξισορρόπηση φορτίου PowerPath

Το PowerPath εξισορροπεί το φορτίο I/O σε μια βάση host-by-host.

Διατηρεί τα στατιστικά στοιχεία για όλα τα I/O σε όλες τις διαδρομές. Για κάθε αίτηση I/O, το PowerPath επιλέγει με έξυπνο τρόπο, με βάση τις στατιστικές και τις ευρετικές και την πολιτική εξισορρόπησης φορτίου και της ανακατεύθυνσης σε ισχύ, την πιο υποχρησιμοποιούμενη διαθέσιμη διαδρομή.

Οι βελτιστοποιημένοι αλγόριθμοι του PowerPath δεν έχουν σχεδιαστεί για να ισορροπούν τέλεια το φορτίο σε όλες τις διαδρομές σε ένα σύνολο διαδρομών. Λόγω του διαφορετικού I/O φορτίου και της καθοδικής δραστηριότητας SAN, είναι απίθανο οι διαδρομές να έχουν ίσο φορτίο. Το PowerPath θα συνεχίσει να χρησιμοποιεί την ίδια διαδρομή με ελαφρά διατηρημένα φορτία, γι' αυτό μπορεί να φαίνεται ότι υπάρχουν I/O ανισορροπίες στις διαδρομές. Απλά εξισώνοντας το φορτίο σε όλες τις διαθέσιμες διαδρομές δεν αποτελεί πάντα την επωφελή λύση για τον host, επειδή όλες οι διαδρομές δεν είναι κατ' ανάγκη ίσες με τις καθημερινές δραστηριότητες SAN. Το PowerPath στοχεύει στη μεγιστοποίηση της απόδοσης και της διαθεσιμότητας σε όλα τα κανάλια. Αυτό σημαίνει ότι η χρήση της ίδιας διαδρομής αποτελεί βελτιστοποίηση που είναι προτιμότερη από την εξίσωση. Η χρησιμοποίηση μια πολιτικής Round Robin μπορεί να επιφέρει ένα λιγότερο επιθυμητό αποτέλεσμα ως προς τη συνολική απόδοση του host, διότι η πολιτική αυτή επιδιώκει να εξισώσει και να μην βελτιστοποιήσει, με αποτέλεσμα την μείωση της απόδοσης του λειτουργικού συστήματος. Το PowerPath λαμβάνει υπόψη του όλη την ικανότητα επεξεργασίας I/O και διαύλου όλων των διαδρομών. Μία διαδρομή δεν πρέπει ποτέ να υπερφορτωθεί και να είναι αργή, ενώ άλλες διαδρομές είναι σε αδράνεια. Επίσης, μία κατειλημμένη διαδρομή ποτέ δεν αντιμετωπίζεται ως ίση με μία αδρανή διαδρομή.

Το PowerPath έχει έναν αυτοματοποιημένο μηχανισμό διαμόρφωσης πολιτικής εξισορρόπησης φορτίου. Όταν ξεκινά το λειτουργικό σύστημα και ξεκινά και το PowerPath, πληροφορίες συγκεκριμένες με την συσκευή στα διαχειριζόμενα LUN διαβάζονται από το PowerPath. Η βελτιστοποιημένη πολιτική εξισορρόπησης φορτίου ρυθμίζεται αυτόματα σε μια βάση ανά LUN. Το PowerPath υποστηρίζει το EMC και ορισμένες διατάξεις μη - EMC. Αν πολλές διατάξεις είναι προσβάσιμες στο SAN και χωρισμένες σε ζώνες που προορίζονται για την εξισορρόπηση φορτίου EMC PowerPath και την

ανακατεύθυνση PowerPath στον host, σε κάθε LUN που είναι καλυμμένο στον server θα ανατεθεί η πολιτική που παρέχει τον καλύτερο αλγόριθμο για εξισορρόπηση φορτίου.

Το EMC έχει αναπτύξει πολλαπλές πολιτικές εξισορρόπησης φορτίου που προορίζονται να στηρίξουν τις διάφορες απαιτήσεις των πελατών σε περιβάλλον παραγωγής και δοκιμής. Το PowerPath συνιστάται για απλές και σύνθετες διαμορφώσεις αποθήκευσης. Όλοι οι τύποι των πελατών μπορούν να επωφεληθούν από τις βελτιστοποιημένες πολιτικές. Οι πολιτικές Symmetrix Optimized και το CLARiiON Optimized είναι οι προεπιλεγμένες πολιτικές για τις συσκευές Symmetrix, VNX™ και CLARiiON, αντίστοιχα. Η χρήση άλλων πολιτικών θα πρέπει να πραγματοποιείται μόνο υπό την καθοδήγηση της Υποστήριξης Πελατών EMC. Οι πολιτικές αυτές χρησιμοποιούνται από σχεδόν όλους τους χρήστες της EMC για τις διατάξεις Symmetrix, VNX και CLARiiON.

Το PowerPath έχει πολλαπλούς αλγόριθμους. Οι συγκρίσεις σε αυτήν την εργασία με το εγγενές multipathing εστιάζουν στις λειτουργίες Symmetrix και CLARiiON Optimized.

Υπάρχει ένας αποκλειστικός αλγόριθμος Symmetrix Optimized - EMC που έχει σχεδιαστεί για τα παλιά και τα σημερινά μοντέλα Symmetrix. Η χρήση αυτού συνιστάται για όλες τις εφαρμογές Symmetrix.

Ο αποκλειστικός αλγόριθμος CLARiiON Optimized - EMC έχει σχεδιαστεί για τα μοντέλα VNX και CLARiiON. Η χρήση αυτού συνιστάται για όλες τις εφαρμογές VNX και CLARiiON. Οι λειτουργίες Asynchronous Logical Unit Access (Alua) και μη – Alua υποστηρίζονται. Αυτή η πολιτική αναγνωρίζει τη διαφορά μεταξύ των βελτιστοποιημένων και μη βελτιστοποιημένων διαδρομών, όπως ορίζεται στον σχεδιασμό Alua.

Οι άλλες πολιτικές χρησιμοποιούνται σπάνια. Αυτές είναι:

Adaptive, Least Block, Least I/O (επίσης γνωστή ως Least Queued I/O), Request, Round Robin, Streamlo, και Basic Failover

Λόγω του αποκλειστικού σχεδιασμού και της πατέντας του PowerPath, ο ακριβής αλγόριθμος για τις πολιτικές αυτές δεν μπορεί να αναλυθεί εδώ. Ωστόσο, η παρούσα εργασία θα εξηγήσει τις υψηλού επιπέδου λειτουργίες των πολιτικών και τις αλληλεπιδράσεις τους που συμβάλλουν στην

προηγμένη δυνατότητα εξισορρόπησης φορτίου.

Η προεπιλεγμένη πολιτική του PowerPath για τους διάφορους τύπους αποθήκευσης είναι επίσης η βέλτιστη πολιτική. Αυτό σημαίνει ότι οι διαχειριστές δεν χρειάζεται να αλλάξουν ή να τροποποιήσουν τις παραμέτρους. Αυτό επιτρέπει στον διαχειριστή να αναλώνει τον χρόνο του σε άλλες δραστηριότητες και όχι στην διαμόρφωση των επιλογών multipathing.

Το PowerPath επιλέγει μία διαδρομή για κάθε I/O, σύμφωνα με την πολιτική εξισορρόπησης φορτίου και ανακατεύθυνσης για την εν λόγω λογική συσκευή. Το PowerPath επιλέγει την καλύτερη διαδρομή ανάλογα με τον επιλεγμένο αλγόριθμο. Η επέκταση Multipathing εξετάζει όλες τις πιθανές διαδρομές για το I/O και επιλέγει την καλύτερη.

Για κάθε I/O, συσχετίζεται ένα βάρος με όλες τις διαθέσιμες και έγκυρες (όχι νεκρές) διαδρομές για το I/O. Οι νεκρές διαδρομές για τη συσκευή δεν λαμβάνονται υπόψη. Όταν ληφθούν υπόψη όλες οι διαδρομές, το PowerPath επιλέγει τη διαδρομή με το χαμηλότερο βάρος. Όσο χαμηλότερο είναι το βάρος, τόσο μεγαλύτερη είναι η πιθανότητα να επιλεγεί η διαδρομή. Ακολουθεί λεπτομερέστερη εξέταση των βελτιστοποιημένων πολιτικών του PowerPath.

Τα Symmetrix Optimized και CLARiiON Optimized χρησιμοποιούν μια σειρά από παράγοντες στους υπολογισμούς τους όταν επιλέγουν την βέλτιστη διαδρομή για ένα I/O, υπό κανονικές, υποβαθμισμένες και ελαττωματικές συνθήκες στο SAN. Το βάρος για την επιλογή της διαδρομής βασίζεται στα ακόλουθα:

I/O εν αναμονή στην διαδρομή – Το PowerPath λαμβάνει υπόψη το ποσό των εκκρεμών αιτήσεων ανάγνωσης ή εγγραφής που αναμένουν πρόσβαση στην διάταξη της αποθήκευσης, η οποία επηρεάζεται άμεσα από το βάθος της ουράς αναμονής.

Το μέγεθος των I/Os – Οι φυσικοί και ιδιαίτερα οι εικονικοί hosts θα μπορούσαν να έχουν πολλαπλές εφαρμογές να τρέχουν με διαφορετικά χαρακτηριστικά I/O. Ανάλογα με την εφαρμογή, το μέγεθος I/O μπορεί να ποικίλει και να είναι από μικρό έως μεγάλο. Οι βελτιστοποιημένοι αλγόριθμοι PowerPath φυσικά αποφεύγουν τα μικρά I/O από το να κολλούν πίσω από μεγάλες ουρές I/Os γιατί θα επιλεγεί η λιγότερο φορτωμένη διαδρομή.

Τύποι I/Os – Λαμβάνονται υπόψη οι εκκρεμείς αναγνώσεις και εγγραφές. Το PowerPath ζυγίζει, διαβάζει και γράφει διαφορετικά, επειδή λαμβάνει υπόψη μια τις ικανότητες που σχετίζονται συγκεκριμένα με την διάταξη σε σχέση με το προφίλ του I/O.

Διαδρομές που χρησιμοποιήθηκαν πιο πρόσφατα – Το PowerPath θα επιχειρήσει να χρησιμοποιήσει την ίδια διαδρομή και πάλι, αν δεν υπάρχει μια λιγότερο σταθμισμένη διαδρομή διαθέσιμη. Αυτό αποφεύγει την ανάγκη για συνεχή εναλλαγή της διαδρομής για κάθε I/O.

Σε μια λειτουργία VNX ή CLARiiON στην Alua ή μια ενεργή/παθητική λειτουργία, το PowerPath θα εξισορροπήσει το φορτίου I/O μεταξύ των κατεχόμενων θυρών επεξεργασίας της αποθήκευσης. Σε μια διαμόρφωση Alua, ένα LUN θα ανήκει είτε στο SPA ή στο SPB. Κάθε SP έχει πολλαπλές θύρες. (Ο ακριβής αριθμός των θυρών εξαρτάται από το μοντέλο VNX ή CLARiiON). Οι κατεχόμενες θύρες SP θεωρούνται ως οι βέλτιστες θύρες, επειδή αυτές οι διαδρομές έχουν χαμηλότερη λανθάνουσα κατάσταση. Εάν το κατεχόμενο SP έχει τέσσερις θύρες, τότε το PowerPath θα φορτώσει την ισορροπία σε αυτές τις τέσσερις βελτιστοποιημένες διαδρομές. Το PowerPath πιθανότατα δεν θα χρησιμοποιήσει τις μη επεξεργασμένες (ή μη βελτιστοποιημένες) διαδρομές εκτός εάν υπάρχει κάποια ακραία υποβάθμιση σε όλες τις βελτιστοποιημένες διαδρομές. Στην πράξη, οι μη βελτιστοποιημένες διαδρομές δεν χρησιμοποιούνται εκτός κι αν υπάρχει καταπάτηση.

Οι κεντρικοί υπολογιστές (hosts) του PowerPath δεν επικοινωνούν μεταξύ τους. Ως εκ τούτου, οι αποφάσεις διαδρομής για έναν μόνο host δεν επηρεάζονται από τις αποφάσεις διαδρομής σε άλλους hosts. Ωστόσο, οι δραστηριότητες των άλλων hosts λαμβάνονται έμμεσα υπόψη. Για παράδειγμα, εάν μία διαδρομή υποβαθμίζεται λόγω της άλλης δραστηριότητας του host ή επειδή έχει μια πιο αργή ικανότητα, τα I/O θα παραμείνουν για περισσότερο χρόνο σε ουρά. Οι βελτιστοποιημένοι αλγόριθμοι του PowerPath λαμβάνουν υπόψη τους αυτήν την έκταση της ουράς και επιλέγουν την επόμενη καλύτερη διαδρομή.

Οι διαχειριστές της αποθήκευσης και του server έχουν επηρεαστεί από ένα πλήθος γεγονότων στο SAN που επηρέασαν και την απόδοση των

εφαρμογών. Οι διαχειριστές βιώνουν μια σειρά από συνθήκες σε κανονικά και ελαττωματικά περιβάλλοντα που επηρεάζουν την διαμεταγωγή I/O.

Τα παραδείγματα περιλαμβάνουν:

- Υπερκαλλυμένες διατάξεις θυρών αποθήκευσης
- Ανισορροπία στο φορτίο I/O στην διάταξη της αποθήκευσης που οφείλεται σε μια ποικιλία hosts με διαφορετικά προφίλ εφαρμογής I/O που είναι κατανεμημένα σε ζώνες στις ίδιες θύρες αποθήκευσης.
- Η έλλειψη buffer-to-buffer credit αναγκάζει σε επιβράδυνση τον χρόνο απόκρισης του SAN, αυξάνοντας έτσι την έκταση της ουράς
- Η μη σωστή λειτουργία του κώδικα του οδηγού HBA που προκαλεί επαναλαμβανόμενες fabric συνδέσεις έχει ως αποτέλεσμα την μειωμένη διαμεταγωγή της εφαρμογής σε ορισμένες θύρες
- Η απώλεια ISL μεταξύ των διακοπτών Fibre Channel στο Fabric A προκαλεί μείωση του εύρους ζώνης μεταξύ των διακοπτών και οδηγεί σε ανισορροπία της διαμεταγωγής I/O σε σχέση με το Fabric B.
- Η υποβάθμιση της υποδομής των οπτικών ινών (για παράδειγμα, κατεστραμμένα καλώδια οπτικών ινών ή σάπια οπτικά λέιζερ), προκαλούν αστάθεια στην διαδρομή
- Και πολλά άλλα ...

Οι προηγμένοι αλγόριθμοι εξισορρόπησης φορτίου του PowerPath έχουν σχεδιαστεί για να αξιοποιήσουν τα μέγιστα την υψηλής διαθεσιμότητας και πλεονάζουσα υποδομή που έχει φτιαχτεί από τους διαχειριστές της αποθήκευσης και του server. Τα SAN διαθέτουν φυσική δυναμική, η οποία αναπόφευκτα καταλήγει σε μια μεταβολή στην αξιοποίηση διαδρομή. Οι αλλαγές στην εφαρμογή, η πρόσθεση ή η αφαίρεση hosts και τα λάθη μπορούν να προκαλέσουν ασύμμετρες διαδρομές I/O, οδηγώντας σε υποβάθμιση των επιδόσεων. Με τη χρήση των βελτιστοποιημένων αλγορίθμων εξισορρόπησης φορτίου, το PowerPath μπορεί να αντισταθμίσει και να προσαρμοστεί για τις δυναμικές αλλαγές σε φυσικά και εικονικά

περιβάλλοντα. Σε αντίθεση με άλλες λύσεις εξισορρόπησης φορτίου, ο διαχειριστής μπορεί να χρειαστεί να ρυθμίσει ξανά τις διαδρομές, να ρυθμίσει τις παραμέτρους διαχείρισης των διαδρομών και να συνεχίζει να τις επαναδιαμορφώνει καθώς η κυκλοφορία των I/O μεταξύ του host και της αποθήκευσης αλλάζει ως απάντηση στις αλλαγές χρήσης.

2.3.Μειονεκτήματα Round Robin

Το Round Robin είναι η προεπιλεγμένη και η πιο συχνά χρησιμοποιούμενη πολιτική εξισορρόπησης φορτίου για τις περισσότερες ενδογενείς λύσεις MPIO στα λειτουργικά συστήματα. Αποτελεί έναν πολύ απλό αλγόριθμο εξισορρόπησης φορτίου. Η πολιτική αυτή έχει πλεονεκτήματα σε σχέση με την βασική πολιτική ανακατεύθυνσης, επειδή όλες οι διαδρομές είναι πλέον σε θέση να υποστηρίξουν τα I/O στην συσκευή ταυτόχρονα. Για τους διαχειριστές, η δυνατότητα χρησιμοποίησης όλων των διαδρομών στην εξισορρόπηση φορτίου και ανακατεύθυνσης παρουσιάζει τεράστιο όφελος.

Ωστόσο, όλες οι διαδρομές δεν είναι ίσες λόγω των προβλημάτων στον host, του δικτύου και την διάταξη της αποθήκευσης. Για τη βελτιστοποίηση των διαθέσιμων διαδρομών, οι διαχειριστές θα πρέπει να ασχοληθούν με το πώς χρησιμοποιούνται αυτές οι διαδρομές.

Το Round Robin χρησιμοποιεί όλες τις διαδρομές σε μια στατική ρύθμιση. Ωστόσο, τα SANs είναι δυναμικά από τον σχεδιασμό τους. Οι servers, οι εφαρμογές, τα εξαρτήματα δικτύωσης και οι διατάξεις αποθήκευσης προστίθενται, αφαιρούνται και τροποποιούνται εκ προθέσεως και μερικές φορές λανθασμένα. Μια πολιτική εξισορρόπησης φορτίου θα πρέπει να είναι σε θέση να δράσει και να αντιδράσει σε αυτό το είδος περιβάλλοντος. Το Round Robin:

Δεν πραγματοποιεί δυναμική αναδρομολόγηση της κίνησης των συσκευών I/O

Δεν εντοπίζει τις τυχόν αλλαγές στη δραστηριότητα του δικτύου

Δεν εξετάζει τυχόν αλλαγές στη διάταξη αποθήκευσης

Δεν λαμβάνει υπόψη τις μοναδικές ιδιότητες της διάταξης

Δεν παρακολουθεί το βάθος της ουράς του HBA

Δεν λαμβάνει υπόψη το μέγεθος του I/O

Το Round Robin δεν διαθέτει νοημοσύνη πίσω από την δρομολόγηση της διαδρομής. Εφ' όσον μία διαδρομή θεωρείται ότι είναι διαθέσιμη, δίδει την ίδια βαρύτητα σε όλες τις άλλες διαδρομές, εκτός εάν η έκδοση MPIO διαθέτει ρυθμίσεις που μπορούν να γίνουν από τον χρήστη για την στάθμιση των διαδρομών. Ωστόσο, όπως αναφέρθηκε και παραπάνω, οι διαδρομές δεν είναι πάντοτε ίσες. Δεν υπάρχει κανένα τέλειο ή εντελώς στατικό SAN. Κάθε κέντρο δεδομένων αλλάζει και κάθε κέντρο δεδομένων είναι διαφορετικό.

Κάποια θα έχουν περισσότερες διακοπές από άλλα. Κάποια θα πρέπει να σχεδιαστούν με τις βέλτιστες πρακτικές κατά νου, ενώ άλλα όχι. Υπό κανονικές συνθήκες λειτουργίας χωρίς σφάλματα, όπου το I/O είναι ομοιόμορφα κατανεμημένο σε όλο το δίκτυο και την διάταξη, το Round Robin θα λειτουργήσει αρκετά καλά για να ισορροπήσει το I/O. Δεν έχει σημασία πόσο καλά έχει σχεδιαστεί το SAN, σε κάποιο σημείο θα προκύψουν προβλήματα ούτως ή άλλως.

Οι διαχειριστές αποθήκευσης και του server θέλουν μια λύση εξισορρόπησης φορτίου που να λειτουργεί καλύτερα και κάτω από όλες τις συνθήκες και όχι μόνο όταν το περιβάλλον είναι το βέλτιστο. Η χρήση όλων των διαδρομών εξίσου δεν είναι το ίδιο με τη βελτιστοποίηση αυτών των διαδρομών. Το PowerPath αξιοποιεί την υψηλής διαθεσιμότητας πλεονάζουσα υποδομή του SAN για την μεγιστοποίηση της διαμεταγωγής I/O πραγματοποιώντας έξυπνα και δυναμικά τη δρομολόγηση της κυκλοφορίας της εφαρμογής χωρίς την ανάγκη για παρέμβαση από τον χρήστη. Οι σχεδιαστές SAN σχεδιάζουν περιβάλλοντα που είναι ανθεκτικά, αξιόπιστα και προβλέψιμα. Τα χειρότερα σενάρια λαμβάνονται πάντα υπόψη στο σχεδιασμό. Κανείς δεν θέλει να συμβούν, αλλά θα πρέπει να υπολογιστούν και αυτά. Έχοντας αυτό κατά νου, οι σχεδιαστές δεν σχεδιάζουν περιβάλλοντα που είναι απλώς «αρκετά καλά».

Είτε το SAN είναι απλό ή σύνθετο, οι διαχειριστές θα πρέπει να ασχοληθούν με τις αναμενόμενες, αλλά και τις απροσδόκητες αλλαγές σε μεγαλύτερο ή μικρότερο βαθμό, ανάλογα με το περιβάλλον. Με όλες τις μεταβλητές στο SAN, η χρήση του PowerPath παρέχει ένα επίπεδο

βελτιστοποίησης της διαδρομής και της διαθεσιμότητας I/O που δεν είναι εφικτό με τις λύσεις Round Robin.

2.4.Εξισορρόπηση φορτίου στα Windows

Στα Windows 2003, η Microsoft ανέπτυξε το πλαίσιο MPIO, επιτρέποντας στους τρίτους κατασκευαστές συστημάτων αποθήκευσης να εγγράψουν κώδικα λογισμικού για να χρησιμοποιήσουν αυτό το πλαίσιο. Αυτή όμως δεν αποτέλεσε μία ανεξάρτητη λύση εξισορρόπησης φόρτου. Τα DSM είναι plug-ins σε αυτό το πλαίσιο. Η EMC ανέπτυξε ένα DSM με βάση το πλαίσιο MPIO. Το PowerPath είναι μία λύση που επιτρέπει στους πελάτες να έχουν τα χαρακτηριστικά του PowerPath στο πλαίσιο που παρέχεται από την Microsoft.

Στα Windows 2008, η Microsoft ανέπτυξε ένα εγγενές προϊόν multipathing που χρησιμοποιεί το ίδιο πλαίσιο με βασικές πολιτικές εξισορρόπησης φορτίου και ανακατεύθυνσης. Στα Windows 2008 R2, οι βελτιώσεις στο multipathing περιλαμβάνουν περισσότερες πολιτικές. Η Microsoft επιτρέπει στον διαχειριστή να επιλέξει μία από τις ακόλουθες πολιτικές[31]:

Fail Over Only - Μία πολιτική η οποία δεν εκτελεί εξισορρόπηση φορτίου. Η πολιτική αυτή χρησιμοποιεί μία μόνο ενεργή διαδρομή και το υπόλοιπο των διαδρομών αποτελούν διαδρομές αναμονής. Η ενεργή διαδρομή χρησιμοποιείται για την αποστολή όλων των I/O. Εάν η ενεργή διαδρομή αποτύχει, τότε μία από τις διαδρομές αναμονής θα χρησιμοποιηθούν.

Round Robin – Αυτή η πολιτική εξισορρόπησης φορτίου χρησιμοποιεί όλες τις διαθέσιμες διαδρομές με έναν ισορροπημένο τρόπο. Αυτή είναι η προεπιλεγμένη πολιτική που επιλέγεται όταν ο ελεγκτής της αποθήκευσης ακολουθεί την πρακτική active/active. Αυτή θα είναι η προεπιλεγμένη πολιτική για το Symmetrix.

[31] Microsoft Windows 2008 R2 MPIO policies information can be found at http://technet.microsoft.com/en-us/library/dd851699.aspx.

Round Robin with Subset – Η πολιτική αυτή είναι παρόμοια με την πολιτική του Round Robin. Η εξισορρόπηση φορτίου καθορίζει μια σειρά από διαδρομές που πρέπει να χρησιμοποιούνται με τρόπο Round Robin και με μια σειρά από διαδρομές αναμονής. Το DSM χρησιμοποιεί μια διαδρομή σε αναμονή μόνο όταν όλες οι κύριες διαδρομές έχουν αποτύχει. Η πολιτική αυτή προορίζεται για την υποστήριξη διατάξεων Alua. Αυτή θα είναι η προεπιλεγμένη πολιτική για τα VNX και CLARiiON.

Least Queue Depth – Αυτή η πολιτική εξισορρόπησης φορτίου στέλνει I/O στην διαδρομή με τις λιγότερες εκείνη την στιγμή εκκρεμείς αιτήσεις I/O. Η πολιτική αυτή είναι παρόμοια με την πολιτική Least I/O του PowerPath.

Weighted Paths – Αυτή η πολιτική εξισορρόπησης φορτίου αποδίδει ένα βάρος σε κάθε διαδρομή. Το βάρος υποδεικνύει την σχετική προτεραιότητα μιας δεδομένης διαδρομής. Όσο μεγαλύτερος ο αριθμός, τόσο χαμηλότερα είναι αναλόγως και η προτεραιότητα. Το DSM επιλέγει την λιγότερο σταθμισμένη διαδρομή ανάμεσα από τις διαθέσιμα διαδρομές.

Least Blocks – Αυτή η πολιτική εξισορρόπησης φορτίου στέλνει I/O στην διαδρομή με τον λιγότερο αριθμό ομάδων δεδομένων που εκείνη την στιγμή βρίσκονται στο στάδιο της επεξεργασίας. Η πολιτική αυτή είναι παρόμοια με την πολιτική Least Block του PowerPath.

2.5.Εξισορρόπηση φορτίου RHEL

Το Device Mapper Multipathing (DM-MPIO) είναι η προεπιλεγμένη λύση multipathing για το Red Hat. Περιλαμβάνει υποστήριξη για τις πιο κοινές διατάξεις αποθήκευσης που υποστηρίζουν το DM-Multipath. Αυτό περιλαμβάνει τις διατάξεις EMC. Οι υποστηριζόμενες συσκευές μπορούν να βρεθούν στο αρχείο multipath.conf.defaults. Οι προεπιλεγμένες ρυθμίσεις σε αυτό το αρχείο θα ρυθμίσουν αυτόματα τους συγκεκριμένους τύπους σειρών. Για παράδειγμα, ένα VNX θα τεθεί σε λειτουργία Alua. Το Round Robin είναι η προεπιλεγμένη εξισορρόπηση φορτίου για τα Symmetrix VMAX™ και VNX. Αν η διάταξη αποθήκευσης υποστηρίζει DM-Multipath και δεν έχει ρυθμιστεί από προεπιλογή σε αυτό το αρχείο, μπορεί να χρειαστεί η προσθήκη μίας

διάταξης στο DM - Multipath αρχείο ρυθμίσεων multipath.conf. θα πρέπει να ληφθεί υπόψη ότι οι προεπιλεγμένες ρυθμίσεις δεν είναι απαραίτητα οι βέλτιστες ρυθμίσεις για το περιβάλλον κάθε πελάτη.

Για τα RHEL5 και RHEL6, ο προεπιλεγμένος αλγόριθμος είναι Round Robin, ο οποίος έχει μία παράμετρο που καθορίζει τον αριθμό των αιτήσεων I/O για δρομολόγηση σε μια διαδρομή πριν από τη μετάβαση στην επόμενη διαδρομή στην τρέχουσα ομάδα διαδρομών. Ο προεπιλεγμένος αριθμός είναι 1000, αλλά μπορεί να τροποποιηθεί. Το προφίλ I/O του host (για παράδειγμα, μικρό ή μεγάλο μπλοκ, τυχαία ή διαδοχικά, βαριά ως προς την ανάγνωση ή την εγγραφή κλπ) καθορίζει πώς πρέπει να καθοριστεί αυτή η τιμή. Ο διαχειριστής καθορίζει την τιμή.

Το RHEL6 παρέχει δύο νέους αλγορίθμους επιλογής διαδρομής που καθορίζουν ποια διαδρομή μπορεί να χρησιμοποιηθεί για την επόμενη λειτουργία I/O:

Μήκος ουράς (Queue-length): Αλγόριθμος που ελέγχει τον όγκο των εκκρεμών I/O στις διαδρομές για να προσδιορίσει ποια διαδρομή θα χρησιμοποιηθεί στη συνέχεια.

Χρόνος εξυπηρέτησης (Service-time): Αλγόριθμος που ελέγχει τον όγκο των εκκρεμών I/O και την σχετική διαμεταγωγή των διαδρομών για να προσδιορίσει ποια διαδρομή να χρησιμοποιηθεί στη συνέχεια.

Ακόμα κι αν οι νέοι αλγόριθμοι στο RHEL6 παρέχουν περισσότερες επιλογές για τους διαχειριστές, το Round Robin εξακολουθεί να αποτελεί την προεπιλογή. Χωρίς μια αυτοματοποιημένη και βελτιστοποιημένη πολιτική εξισορρόπησης φορτίου ή χωρίς τεκμηριωμένες βέλτιστες πρακτικές από τον προμηθευτή του λειτουργικού συστήματος, η διαμόρφωση είναι πιο δύσκολη.

Η δοκιμή της διαδρομής στο RHEL γίνεται από το multipathd daemon. Οι διαχειριστές μπορούν να επιλέξουν ανάμεσα από πολλαπλούς τρόπους δοκιμής. Ο τρόπος «Direct IO» είναι προεπιλεγμένος. Ωστόσο, υπάρχουν και πολλοί άλλοι τρόποι ανάμεσα στους οποίους μπορεί να γίνει η επιλογή ανάλογα με τον τύπο διάταξης. Η δοκιμή της διαδρομής γίνεται κάθε πέντε δευτερόλεπτα ως προεπιλογή. Αυτό είναι επίσης διαμορφώσιμο από τον χρήστη.

Το DM-MPIO δεν αναγνωρίζει τη διαφορά μεταξύ των βέλτιστων και των

μη-βέλτιστων διαδρομών με διατάξεις ALUA. Ωστόσο, οι προεπιλεγμένες ρυθμίσεις για τα βάρη της διαδρομής έχουν ως αποτέλεσμα την χρήση των μη - βέλτιστων διαδρομών. Η χρήση των μη - βέλτιστων διαδρομών αυξάνει την λανθάνουσα κατάσταση των I/O και τους χρόνους ολοκλήρωσης.

Κεφάλαιο 3

Βιβλιογραφική ανασκόπηση

3.1.Σύμπλεγμα Ανακατεύθυνσης Διακομιστή Windows

3.1.1.Windows Server Failover Clustering

Το Σύμπλεγμα Ανακατεύθυνσης του Διακομιστή Windows (Windows Server Failover Clustering - WSFC) αποτελεί τον διάδοχο της Υπηρεσίας Συμπλέγματος Διακομιστή της Microsoft (Microsoft Cluster Service - MSCS). Το WSFC και ο προκάτοχός του MSCS, προσφέρουν υψηλή διαθεσιμότητα κρίσιμων εφαρμογών όπως ηλεκτρονικό ταχυδρομείο, βάσεις στοιχείων και εφαρμογές line-of-business μέσω της εφαρμογής ενός εφεδρικού συμπλέγματος διακομιστών των Windows που παρέχουν μία ενιαία εικόνα του συστήματος στους χρήστες.[32]

Το MSCS αποτελεί την λύση της Microsoft στην οικοδόμηση μεγάλης διαθεσιμότητας συμπλεγμάτων διακομιστών Windows, από την πρώτη στιγμή που παρουσιάστηκε στα Windows NT Server 4.0. Το MSCS έχει ενισχυθεί σημαντικά και έχει απλουστευθεί και μετονομάστηκε σε WSFC με την κυκλοφορία των Windows Server 2008. Το WSFC για τα Windows Server 2008 R2 έχει δεχθεί ακόμη περισσότερες βελτιώσεις σε σχέση με το σύμπλεγμα (clustering) των Windows.[33]

Η υπηρεσία συμπλέγματος WSFC παρακολουθεί την εύρυθμη λειτουργία του συμπλέγματος και μετακινεί αυτόματα τις εφαρμογές από έναν αποτυχημένο κόμβο στους επιζώντες κόμβους, επιφέροντας υψηλή διαθεσιμότητα σε κρίσιμες εφαρμογές. Το WSFC επιφέρει επίσης υψηλή διαθεσιμότητα στις υπηρεσίες Hyper-V virtualization της Microsoft.

---

[32] Failover Clustering in Windows Server 2008 R2. *Microsoft White Paper*, April 2009

[33] Failover Clustering in Windows Server 2008 R2. *Microsoft White Paper*, April 2009

Το WSFC φέρνει πολλές βελτιώσεις στις υπηρεσίες συμπλέγματος MSCS. Με το WSFC, έως και δεκαέξι Windows servers μπορούν να οργανωθούν σε ένα πολλαπλών περιοχών, γεωγραφικά διάσπαρτων συμπλεγμάτων, με περιοχές συμπλέγματος που διαχωρίζονται από εκατοντάδες χιλιόμετρα. Ένα βολικό εργαλείο διαχειριστή GUI με την υποστήριξη πολλών οδηγών καταργεί την ανάγκη για την ύπαρξη ενός ειδικού συμπλεγμάτων για την διαμόρφωση και διαχείριση του συμπλέγματος.[34]

Το WSFC πραγματοποιεί την τεχνολογία συμπλεγμάτων ακόμη πιο ελκυστική για τις μικρές επιχειρήσεις καθώς και για τις μεγάλες επιχειρήσεις.

Η Microsoft ορίζει το σύμπλεγμα ως εξής[35]:

«Ένα σύμπλεγμα ανακατεύθυνσης είναι μία ομάδα ανεξάρτητων υπολογιστών, ή κόμβων, οι οποίοι είναι συνδεδεμένοι με ένα τοπικό δίκτυο (LAN) ή ένα δικτύου ευρείας περιοχής (WAN) και τα οποία συνδέονται μέσω προγραμματισμού με ένα λογισμικό συμπλέγματος. Η ομάδα των κόμβων αντιμετωπίζεται ως ένα ενιαίο σύστημα και μοιράζεται ένα κοινό namespace. Η ομάδα συνήθως περιλαμβάνει πολλαπλές συνδέσεις δικτύου και αποθήκευσης στοιχείων που συνδέονται με τους κόμβους μέσω δικτύων αποθήκευσης (SAN). Το σύμπλεγμα ανακατεύθυνσης λειτουργεί με την κίνηση των πόρων μεταξύ των κόμβων για την παροχή υπηρεσιών, στην περίπτωση αποτυχίας της λειτουργίας των στοιχείων του συστήματος».

Οι κόμβοι ενός συμπλέγματος των Windows είναι διακομιστές Windows που διασυνδέονται μεταξύ τους φυσικά, με ένα εφεδρικό ιδιωτικό δίκτυο για την παρακολούθηση των κόμβων και της ανακατεύθυνσης. Οι κόμβοι έχουν πρόσβαση σε ένα κοινό σύνολο εφεδρικών πόρων του δίσκου μέσα από ένα δίκτυο αποθήκευσης (SAN). Η υπηρεσία συμπλέγματος είναι το λογισμικό που συνδέει μέσω προγραμματισμού τους κόμβους με το σύμπλεγμα και παρέχει μία ενιαία προβολή του συστήματος στους πελάτες που χρησιμοποιούν το σύμπλεγμα.

Οι πελάτες δεν γνωρίζουν ότι έχουν να κάνουν με ένα σύμπλεγμα. Το

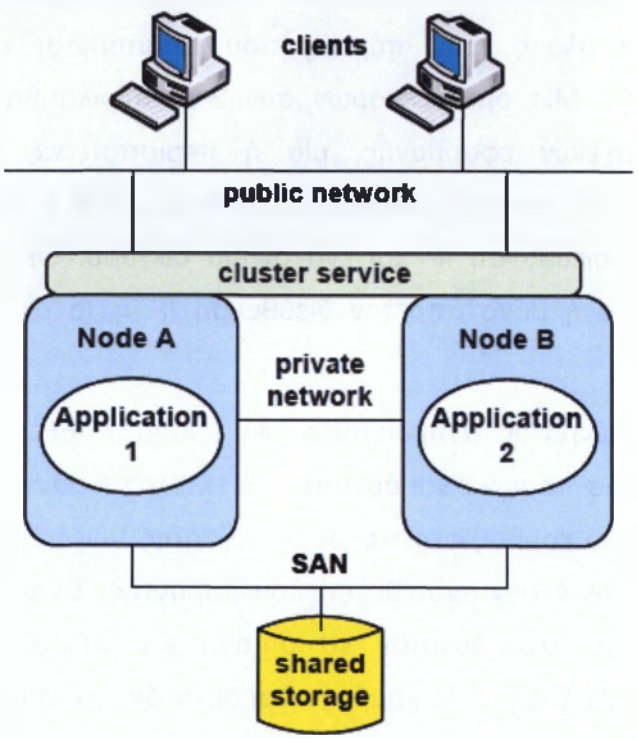
[34] Failover Clustering in Windows Server 2008 R2, *Microsoft White Paper*, April 2009

[35] Failover Clustering in Windows Server 2008 R2, *Microsoft White Paper*, April 2009

σύμπλεγμα σε αυτούς φαίνεται να είναι ένας μόνο διακομιστής των Windows. Στην πραγματικότητα, η εφαρμογή εκτελείται σε έναν εικονικό διακομιστή.

Μία εφαρμογή τρέχει σε έναν μόνο κόμβο κάθε φορά. Ωστόσο, η εφεδρεία (redundancy) που είναι ενσωματωμένη στο σύμπλεγμα, παρέχει προστασία έναντι κάθε αποτυχημένης λειτουργίας του κάθε στοιχείου. Σε περίπτωση που αποτύχει ένας διακομιστής, ένας σύνδεσμος επικοινωνίας, ένας σύνδεσμος αποθήκευσης, ή μια εφαρμογή, η βλάβη εντοπίζεται αυτόματα από την υπηρεσία συμπλέγματος, η οποία θα μεταφέρει την αποτυχημένη εφαρμογή σε έναν επιζώντα κόμβο. Οι χρήστες ενδέχεται να αντιμετωπίσουν προσωρινά μειωμένη απόδοση του συστήματός τους, αλλά δεν θα χάσουν εντελώς την πρόσβαση στις εφαρμογές τους.[36]

Εικόνα 1: Windows Server Failover Clustering[37]



[36] Quagga, a software routing suite, http://www.quagga.net

[37] Quagga, a software routing suite, http://www.quagga.net

3.1.2.Ομάδες Πόρων (Resource Groups)

Θεμελιώδους σημασίας για τη λειτουργία ενός συμπλέγματος είναι η έννοια των πόρων και των ομάδων των πόρων[38]. Ένας πόρος είναι ένα στοιχείο του υλικού εξοπλισμού ή του λογισμικού το οποίο διαχειρίζεται η υπηρεσία του συμπλέγματος. Οι πόροι περιλαμβάνουν εκτελέσιμα αρχεία εφαρμογών, δίσκους, λογικές μονάδες αποθήκευσης, IP διευθύνσεις, ονόματα δικτύων και κάρτες διασύνδεσης δικτύων (NIC). Κάθε πόρος διαθέτει έναν παρακολουθητή πόρου (resource monitor) που του επιτρέπει να αναφέρει την κατάστασή του στην υπηρεσία συμπλέγματος και αυτό επιτρέπει στην υπηρεσία συμπλέγματος να ελέγχει την κατάσταση του πόρου και να αποστέλλει οδηγίες προς τους πόρους για να καταστούν online ή offline. Η πιο σημαντική λειτουργία του παρακολουθητή πόρου είναι να παρακολουθεί την εύρυθμη λειτουργία του πόρου και να αναφέρει τις αλλαγές στην λειτουργία στην υπηρεσία συμπλέγματος.[39]

Μία ομάδα πόρων αποτελεί την ομάδα των πόρων που συνθέτουν μια εφαρμογή. Περιλαμβάνει όλους τους πόρους που απαιτούνται για την εκτέλεση μίας εφαρμογής. Μία ομάδα πόρων συνήθως περιλαμβάνει ένα σύνολο εκτελέσιμων αρχείων εφαρμογής, μία ή περισσότερες λογικές μονάδες αποθήκευσης (που προσδιορίζονται μέσω των LUNs ή λογικούς αριθμούς μονάδων), μία διεύθυνση IP και ένα όνομα δικτύου. Οι πελάτες αναγνωρίζουν την εφαρμογή μόνο από την διεύθυνση IP ή το όνομα του δικτύου.

Η υπηρεσία συμπλέγματος αντιμετωπίζει μία ομάδα πόρων ως μία ατομική μονάδα. Μία ομάδα πόρων είναι δυνατόν να εκτελείται μόνο σε έναν κόμβο κάθε φορά. Αυτός ο κόμβος φέρεται ως ο ιδιοκτήτης των πόρων μιας ομάδας πόρων στον οποίον έχουν ανατεθεί επί του παρόντος. Ένας κόμβος είναι δυνατόν να λειτουργεί (είναι δυνατόν να του ανήκουν) πολλές ομάδες πόρων ανά πάσα στιγμή. Δηλαδή, ένας κόμβος είναι δυνατόν να υποστηρίζει

[38] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003
[39] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

διάφορες εφαρμογές ταυτόχρονα.

3.1.3.Ανακατεύθυνση (Failover)

Σε περίπτωση που μία εφαρμογή επηρεαστεί από ένα σφάλμα του υλικού του συστήματος ή του λογισμικού, η υπηρεσία συμπλέγματος είναι δυνατόν να αναλάβει μία από πολλές δράσεις:

• Είναι δυνατόν να προσπαθήσει να επανεκκινήσει την εφαρμογή στον κόμβο που αυτή ανήκει.

• Είναι δυνατόν να μετακινήσει την ομάδα πόρων σε άλλον κόμβο του συμπλέγματος.

• Εάν το πρόβλημα αποτελεί μία αποτυχημένη λειτουργία του κόμβου, είναι δυνατόν να μετακινήσει όλες τις ομάδες πόρων που ανήκουν στον κόμβο εκείνη την στιγμή, σε άλλους κόμβους του συμπλέγματος.

Κάθε εφαρμογή είναι δυνατόν να διαθέτει μία λίστα προτίμησης που να δείχνει σε ποιον κόμβο προτιμά να εκτελείται και σε ποιους κόμβους θα πρέπει να ανακατευθυνθεί κατά σειρά προτίμησης. Κατηγοριοποιεί επίσης και τις εξαρτήσεις, αναφέροντας για κάθε πόρο ποιοι άλλοι πόροι θα πρέπει να είναι πρώτα διαθέσιμοι. Όταν η υπηρεσία συμπλέγματος εντοπίσει μία αποτυχημένη λειτουργία, καθορίζει σε ποιον κόμβο να μετακινήσει την αποτυχημένη ομάδα πόρων με βάση διάφορους παράγοντες, όπως τον φόρτο του κόμβου και την προτίμηση. Οι πόροι της ομάδας πόρων που μετακινούνται στη συνέχεια ενεργοποιούνται στον νέο κόμβο με τη σειρά που καθορίζεται από τις εξαρτήσεις της ομάδας πόρων.[40]

Όταν αποκαθίσταται η λειτουργία ενός κόμβου και επανασυνδέεται με το σύμπλεγμα, όλες οι ομάδες πόρων που έχουν καθορίσει τον ανακτημένο κόμβο ως τον προτιμώμενο κόμβο τους επιστρέφουν σε αυτόν τον κόμβο.

Δεδομένου ότι οι πελάτες αναγνωρίζουν την εφαρμογή τους μόνο από τη διεύθυνση της IP ή το όνομα του δικτύου και δεδομένου ότι αυτοί είναι οι

[40] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, : March 2003

πόροι που μεταφέρονται στον νέο κόμβο σε περίπτωση βλάβης, οι βλάβες των στοιχείων του συμπλέγματος είναι ορατές στους πελάτες. Απλά συνεχίσουν να κάνουν χρήση της εφαρμογής, ακόμη και αν τώρα εκτελείται σε έναν διαφορετικό κόμβο. Ένας περιορισμός είναι ότι η κατάσταση της συνεδρίας και η κατάσταση της εφαρμογής στη μνήμη θα χαθεί μετά από μια ανακατεύθυνση. Ως εκ τούτου, ένα σύμπλεγμα παρέχει υψηλή διαθεσιμότητα, αλλά δεν είναι ανεκτικό στα σφάλματα.[41]

Μία ομάδα πόρων είναι δυνατόν να ανήκει σε έναν κόμβο κάθε φορά. Αυτό σημαίνει ότι στα LUN είναι δυνατόν να έχει πρόσβαση μόνο ένας κόμβος κάθε φορά. Ωστόσο, όλοι οι κόμβοι πρέπει να έχουν μία σύνδεση με όλα τα LUN που είναι δυνατόν να πρέπει να έχουν μετά από μια αποτυχημένη λειτουργία. Η απαίτηση αυτή ικανοποιείται έχοντας όλα τα LUN στο κοινόχρηστο χώρο αποθήκευσης που παρέχεται από το SAN.

Αν και εφαρμογές συνήθως δεν χρειάζεται να τροποποιηθούν για να τρέξουν σε ένα σύμπλεγμα, οι «cluster-aware» εφαρμογές συχνά επωφελούνται από τις επιπλέον παροχές που βρίσκονται στους παρακολουθητές των πόρων σχετικά με την εκτεταμένη υψηλή διαθεσιμότητα και τα χαρακτηριστικά επεκτασιμότητας.[42]

3.1.4.Απαρτία (Quorum)

Εκτός από τους πόρους που είναι δυνατόν να ανήκουν σε κόμβους, ένα σύμπλεγμα διαθέτει έναν πολύ σημαντικό κοινό πόρο - την απαρτία. Η απαρτία είναι μία βάση στοιχείων διαμόρφωσης παραμέτρων του συμπλέγματος που φιλοξενείται στον κοινόχρηστο αποθηκευτικό χώρο και είναι ως εκ τούτου είναι προσβάσιμη σε όλους τους κόμβους. Η βάση στοιχείων διαμόρφωσης περιλαμβάνει πληροφορίες όπως το ποιοι διακομιστές περιλαμβάνονται στο σύμπλεγμα την δεδομένη στιγμή, ποιοι

[41] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

[42] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

πόροι είναι εγκατεστημένοι στο σύμπλεγμα και ποια είναι η τρέχουσα κατάσταση κάθε πόρου. Ένας κόμβος είναι δυνατόν να συμμετέχει σε ένα σύμπλεγμα μόνο εάν είναι δυνατόν να επικοινωνήσει με το σύμπλεγμα απαρτίας.[43]

Η απαρτία διαθέτει δύο κύριες λειτουργίες:

Συνοχή

Η απαρτία αποτελεί έναν ορισμένο κατάλογο όλων των πληροφοριών ρύθμισης παραμέτρων που σχετίζονται με το σύμπλεγμα. Παρέχει σε κάθε φυσικό διακομιστή μια συνεπή εικόνα για το πώς είναι ρυθμισμένο το σύμπλεγμα. Παρέχει επίσης τις πληροφορίες διαμόρφωσης που απαιτούνται από έναν κόμβο που επέστρεψε στο σύμπλεγμα ή από έναν νέο κόμβο που προστίθεται στο σύμπλεγμα.[44]

Διαιτησία

Όπως σε κάθε εφαρμογή του δικτύου πολλαπλών κόμβων, ένα σύμπλεγμα υπόκειται στο σύνδρομο του split-brain. Σε περίπτωση που μία βλάβη του δικτύου σπάσει το σύμπλεγμα, ώστε να προκύψουν δύο ή περισσότερες απομονωμένες ομάδες κόμβων, και εάν δεν αναληφθεί δράση, κάθε μία από τις απομονωμένες ομάδες θα μπορούσε να αποτελεί το επιζών απομεινάρι του συμπλέγματος και θα αναλάβει την ιδιοκτησία των ομάδων πόρων που ανήκουν στους κόμβους που θεωρεί ότι έχουν αποτύχει. Οι ομάδες πόρων τώρα ανήκουν σε πολλαπλούς κόμβους του συμπλέγματος, οδηγώντας στην καταστροφή της βάσης στοιχείων καθώς εκτελούνται ανεξάρτητες και ασυντόνιστες ενημερώσεις στις βάσεις στοιχείων των επηρεασμένων εφαρμογών.[45]

Η λειτουργία split-brain πρέπει να αποφεύγεται. Αυτή είναι μία

---

[43] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

[44] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

[45] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

λειτουργία που παρέχεται από την απαρτία. Θα ανιχνεύσει ότι το σύμπλεγμα έχει σπάσει και θα επιλέξει το επιζών σύμπλεγμα, σύμφωνα με την πλειοψηφία. Πλειοψηφία σημαίνει ότι η επιζώσα ομάδα κόμβων που επελέγη για την συνέχιση των λειτουργιών του συμπλέγματος, θα πρέπει να περιέχει περισσότερους από τους μισούς κόμβους που έχουν διαμορφωθεί για το σύμπλεγμα. Εάν υπάρχουν n κόμβοι, η επιζώσα ομάδα θα πρέπει να περιέχει τουλάχιστον n/2+1 κόμβους. Όλοι οι άλλοι κόμβοι θα πρέπει να αφαιρεθούν από το σύμπλεγμα και αυτή η νέα ρύθμιση θα σημειωθεί στην βάση στοιχείων της διαμόρφωσης της απαρτίας. Η επιζώσα ομάδα λέγεται ότι έχει μία «απαρτία». Εάν καμία ομάδα δεν έχει απαρτία, το σύμπλεγμα θεωρείται ότι έχει «πέσει» και θα πρέπει να περιμένει τους κόμβους για να επανασυνδεθεί με το σύμπλεγμα.

Αυτό αφήνει το πρόβλημα ενός συμπλέγματος με άρτιο αριθμό κόμβων. Εάν το σύμπλεγμα διαιρεθεί ομοιόμορφα, καμία ομάδα δεν έχει απαρτία και το σύμπλεγμα θεωρείται «πεσμένο». Για να αποφευχθεί αυτό, η βάση στοιχείων της απαρτίας είναι δυνατόν να δεχθεί μία «ψήφο» έτσι ώστε να υπάρχει αποτελεσματικά ένας μονός αριθμός κόμβων, επιτρέποντας την δημιουργία της απαρτίας.

3.2.Η Υπηρεσία Συμπλέγματος

Η υπηρεσία συμπλέγματος αποτελεί μία συλλογή στοιχείων λογισμικού που τρέχει σε κάθε κόμβο και εκτελεί δραστηριότητα σχετική με το σύμπλεγμα. Τα στοιχεία της υπηρεσίας συμπλέγματος αλληλεπιδρούν μεταξύ τους μέσω του ιδιωτικού δικτύου, διασυνδέοντας τους κόμβους του συμπλέγματος. Τα στοιχεία περιλαμβάνουν τα ακόλουθα:[46]

[46] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

3.3.Διαχειριστής Κόμβου

Ο Διαχειριστής Κόμβου τρέχει σε κάθε κόμβο και διατηρεί μια λίστα με όλους τους κόμβους που ανήκουν στο σύμπλεγμα. Παρακολουθεί την εύρυθμη λειτουργία των κόμβων με την αποστολή μηνυμάτων heartbeat σε κάθε κόμβο. Εάν δεν λάβει απάντηση σε ένα μήνυμα hearbeat μετά από μια σειρά από προσπάθειες, εκπέμπει προς πολλαπλούς αποδέκτες σε ολόκληρο το σύμπλεγμα ένα μήνυμα ζητώντας από κάθε μέλος του συμπλέγματος να επαληθεύσει την εικόνα του για την τρέχουσα κατάσταση των μελών του συμπλέγματος. Οι ενημερώσεις της βάσης στοιχείων διακόπτονται έως ότου ο αριθμός των μελών του συμπλέγματος σταθεροποιηθεί.

Εάν ένας κόμβος δεν ανταποκριθεί, έχει τεθεί εκτός υπηρεσίας και οι ενεργές ομάδες των πόρων του, μεταφέρονται σε άλλους λειτουργικούς κόμβους σύμφωνα με τις προτιμήσεις κάθε ομάδας πόρων.[47]

3.4.Διαχειριστής Βάσης Στοιχείων

Ο Διαχειριστής βάσης στοιχείων τρέχει σε κάθε κόμβο και διατηρεί την βάση στοιχείων ρύθμισης παραμέτρων του συμπλέγματος. Αυτή η βάση στοιχείων περιέχει πληροφορίες για όλες τις φυσικές και λογικές οντότητες στο σύμπλεγμα, όπως το σύμπλεγμα το ίδιο, την συμμετοχή μελών στον κόμβο, τα είδη των πόρων και τις περιγραφές τους και τις ομάδες πόρων. Αυτές οι πληροφορίες χρησιμοποιούνται για την παρακολούθηση της τρέχουσας κατάστασης του συμπλέγματος και για τον προσδιορισμό της επιθυμητής κατάστασης.

Οι διαχειριστές βάσεων στοιχείων συνεργάζονται για να εξασφαλίσουν την διατήρηση μιας συνεκτικής εικόνας του συμπλέγματος σε κάθε κόμβο. Ο διαχειριστής βάσης στοιχείων δρα στον κόμβο πραγματοποιώντας μια αλλαγή

[47] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

διαμόρφωσης, ενεργοποιεί την αντιγραφή της ενημέρωσής του στους άλλους κόμβους. Η αντιγραφή είναι ατομική και σειριακή και χρησιμοποιεί ένα κομμάτι μίας φάσης. Εάν ένας κόμβος δεν είναι δυνατόν να πραγματοποιήσει μία ενημέρωση, τίθεται εκτός λειτουργίας.

Οι αλλαγές είναι επίσης αποθηκευμένες στην απαρτία ως ημερολόγιο για τον κόμβο ανάκτησης .[48]

3.5.Διαχειριστής Ανακατεύθυνσης

Οι Διαχειριστές Ανακατεύθυνσης, οι οποίοι τρέχουν σε κάθε κόμβο, συνεργάζονται για να διαιτητεύσουν στην κυριότητα των ομάδων των πόρων μετά από μία αποτυχία σε κάποιο στοιχείο του συμπλέγματος. Είναι αρμόδιοι για την ενεργοποίηση της ανακατεύθυνσης των ομάδων πόρων και για την έναρξη και τη διακοπή των πόρων, σύμφωνα με τις εξαρτήσεις που καθορίζονται από κάθε ομάδα πόρων.

Εάν ένας πόρος αποτύχει, ο Διαχειριστής Ανακατεύθυνσης σε αυτόν τον κόμβο είναι δυνατόν να προσπαθήσει να σταματήσει και να επανεκκινήσει τον πόρο. Εάν αυτό δεν διορθώσει το πρόβλημα, ο Διαχειριστής Ανακατεύθυνσης παύει τον πόρο προκαλώντας μία ανακατεύθυνση της αποτυχημένης ομάδας πόρων σε άλλον κόμβο. Σε περίπτωση μετακίνησης μίας ομάδας πόρων, ο διαχειριστής ανακατεύθυνσης θα ενημερώσει τη βάση στοιχείων ρύθμισης παραμέτρων μέσω του διαχειριστή βάσεων στοιχείων.

Η ανακατεύθυνση είναι δυνατόν να ενεργοποιηθεί σε απάντηση μιας απρογραμμάτιστης αποτυχίας υλικού του συστήματος ή εφαρμογής, ή είναι δυνατόν να ενεργοποιηθεί χειροκίνητα από τον διαχειριστή του συμπλέγματος, έτσι ώστε να είναι δυνατόν να αναβαθμιστεί ένας κόμβος. Στην τελευταία περίπτωση, ο τερματισμός είναι ομαλός. Αν η ανακατεύθυνση

[48] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

πυροδοτείται από μία αποτυχημένη λειτουργία ενός στοιχείου του συμπλέγματος, η διακοπή είναι δυνατόν να είναι ξαφνική και να προκαλέσει αναστάτωση. Σε περίπτωση που συμβεί αυτό, απαιτούνται επιπλέον βήματα για να αξιολογηθεί η κατάσταση του συμπλέγματος και η ακεραιότητα της βάσης στοιχείων της εφαρμογής πριν οι αποτυχημένες ομάδες πόρων επιστραφούν στην υπηρεσία σε έναν επιζώντα κόμβο.[49]

Όταν ένας κόμβος τεθεί και πάλι σε λειτουργία και επανασυνδεθεί με το σύμπλεγμα, ο διαχειριστής ανακατεύθυνσης διαχειρίζεται την αποτυχία επαναφοράς των ομάδων πόρων. Αποφασίζει ποιες ομάδες πόρων θα μετακινηθούν στον ανακτημένο κόμβο με βάση τις προτιμήσεις. Η μετακίνηση των ομάδων πόρων σε έναν ανακτημένο κόμβο, είναι δυνατόν να περιορίζεται για ορισμένες ώρες ώστε να αποτρέπονται μαζικές μετακινήσεις κατά τις ώρες αιχμής της δραστηριότητας.[50]

Ο διαχειριστής ανακατεύθυνσης ευθύνεται επίσης για τη δημιουργία αντιγράφων ασφαλείας και την επαναφορά των αρχείων καταγραφής απαρτίας και άλλων κρίσιμων αρχείων. [51]

3.6.Συμπλέγματα πολλαπλών περιοχών

Αν και τα μεγάλης διαθεσιμότητας συμπλέγματα μειώνουν τις επιπτώσεις των αποτυχημένων λειτουργιών ενός στοιχείου, το σύμπλεγμα εξακολουθεί να είναι ευάλωτο σε καταστροφές περιοχής-τοποθεσίας, όπως πυρκαγιές και πλημμύρες. Η μοναδική προστασία από τις καταστροφές στην περιοχή είναι να υπάρχει ένα άλλο σύμπλεγμα που θα βρίσκεται αρκετά μακριά και όπου θα καθίσταται απίθανο για μία καταστροφή να επηρεάσει και

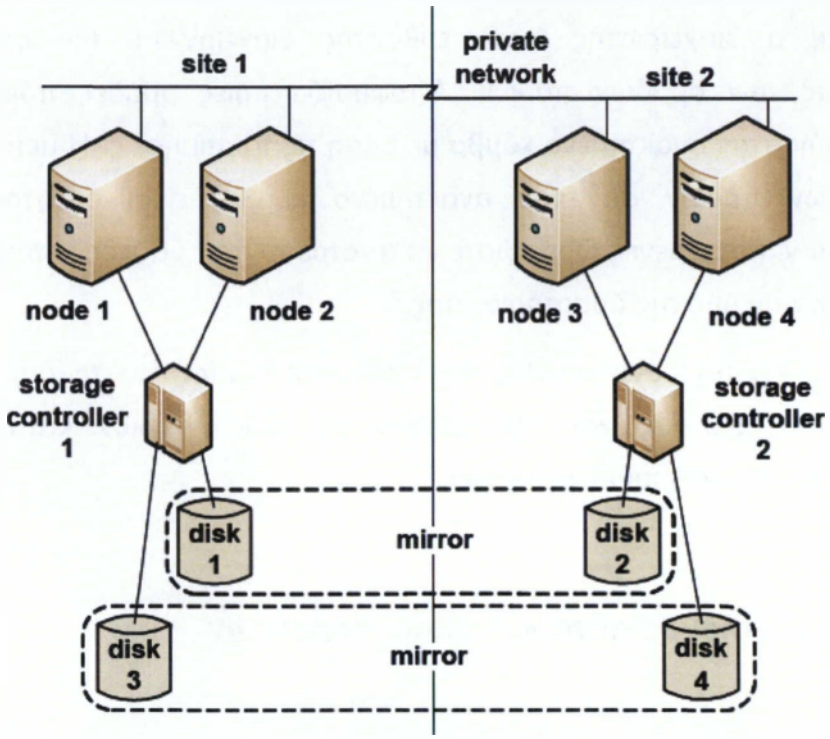
[49] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

[50] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

[51] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

τις δύο περιοχές του συμπλέγματος. Αυτό είναι δυνατόν να επιτευχθεί μέσω γεωγραφικής διασποράς συμπλεγμάτων σε πολλαπλές περιοχές[52], στα οποία τα διασυνδεόμενα συμπλέγματα βρίσκονται σε δύο ή περισσότερες γεωγραφικά διαφορετικές περιοχές.

Εικόνα 2: Συμπλέγματα πολλαπλών περιοχών



Σε μια διάταξη σε πολλαπλές περιοχές, η αντιγραφή στοιχείων πρέπει να χρησιμοποιηθεί για να εξασφαλιστεί ότι και οι δύο περιοχές έχουν μία ενημερωμένη εικόνα όλων των αρχείων και βάσεων στοιχείων. Οι δίσκοι στοιχείων μπορούν προαιρετικά να αντικατοπτρίζονται ασυγχρονικά ή συγχρονικά, αν και οι εφαρμογές θα πρέπει να είναι σε θέση να αντιμετωπίσουν μερική απώλεια στοιχείων, εάν χρησιμοποιηθεί ασύγχρονη αντιγραφή. Ωστόσο, ο δίσκος απαρτίας πρέπει να αντιγραφεί συγχρονικά για να εξασφαλιστεί μια συνεκτική εικόνα του διανεμημένου συμπλέγματος ανά πάσα στιγμή. Εάν χρησιμοποιηθεί ασύγχρονη αντιγραφή των βάσεων

[52] Geographically Dispersed Clusters. *Microsoft TechNet*; 2010

στοιχείων της εφαρμογής, η απόσταση δεν αποτελεί πρόβλημα, δεδομένου ότι ο χρόνος απόκρισης μιας ενημέρωσης της απαρτίας δεν επηρεάζει άμεσα την απόδοση των εφαρμογών. Τα συμπλέγματα μπορούν να διαχωρίζονται με εκατοντάδες χιλιόμετρα. Εάν οι βάσεις στοιχείων της εφαρμογής αντιγράφονται συγχρονικά, η απόσταση που χωρίζει τις περιοχές είναι περιορισμένη.[53]

Η αντιγραφή είναι δυνατόν να είναι είτε βασισμένη στο λογισμικό σε επίπεδο host ή βασισμένη στο hardware στο επίπεδο ελεγκτή SAN. Ωστόσο, αν χρησιμοποιηθεί μπλοκάρισμα της αντιγραφής SAN, θα πρέπει να εξασφαλιστεί ότι η σειρά των εγγραφών θα σωθεί για να διατηρηθεί η συνοχή της βάσης στοιχείων .

Η υπηρεσία συμπλέγματος WSFC και οι εφαρμογές του συμπλέγματος αγνοούν τον γεωγραφικό διαχωρισμό. Όλες οι λειτουργίες του συμπλέγματος πραγματοποιούνται με τον ίδιο τρόπο ανεξάρτητα από το πού βρίσκονται τα μέλη του συμπλέγματος. Η Microsoft δεν παρέχει προϊόν αντιγραφής για συμπλέγματα πολλαπλών περιοχών. Θα πρέπει να χρησιμοποιούνται προϊόντα τρίτων, όπως το Neverfail ClusterProtector[54], το οποίο παρέχει συγχρονική αντιγραφή, ανίχνευση σφαλμάτων και υπηρεσίες ανακατεύθυνσης απομακρυσμένης περιοχής.[55]

3.7.Βελτιώσεις του WSFC έναντι του MSCS

Το WSFC έχει βελτιωθεί σημαντικά σε σχέση το MSCS σε αρκετούς τομείς.

---

[53] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

[54] NeverFail ClusterProtector, http://extranet.neverfailgroup.com/download/DS-cluster-08-09-4page-lo.pdf.

[55] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

3.7.1.Διαχείριση Συμπλέγματος

Μία σημαντική ιστορική πρόκληση σχετικά με τα συμπλέγματα ήταν η πολυπλοκότητα της κατασκευής, της διαμόρφωσης και τη διαχείρισης των συμπλεγμάτων. Το WSFC κρύβει τα "nuts and bolts" του συμπλέγματος πίσω από μια νέα διεπαφή GUI, η διαχείριση του συμπλέγματος ανακατεύθυνσης συμπληρώνει την κονσόλα διαχείρισης της Microsoft. Η Microsoft ισχυρίζεται ότι ένας ειδικός επί των συμπλεγμάτων δεν θεωρείται πλέον απαραίτητος για την επιτυχή ανάπτυξη και τη διατήρηση ενός συμπλέγματος. Οι λειτουργίες αυτές μπορούν πλέον να πραγματοποιηθούν από έναν τεχνικό με γενικές γνώσεις Τεχνολογίας της Πληροφορικής.

Το εργαλείο διαχείρισης του συμπλέγματος ανακατεύθυνσης είναι προσανατολισμένο σε κάποιο έργο και όχι στους πόρους και απλοποιεί τη διαχείριση μέσω διαφόρων νέων οδηγών. Για παράδειγμα, με το MSCS, προκειμένου να δημιουργηθεί ένας μεγάλης διαθεσιμότητας διαμερισμός αρχείων, ο διαχειριστής έπρεπε να δημιουργήσει μία ομάδα, έναν δίσκο πόρου, μία διεύθυνση IP, ένα όνομα δικτύου, να ρυθμίσει τα μηνύματα heartbeat, να δημιουργήσει μία λίστα προτιμώμενων κόμβων και καθορίσει τις εξαρτήσεις πόρων. Με το WSFC, όλα όσα πρέπει να πραγματοποιεί ο διαχειριστής είναι να ορίσει ένα όνομα δικτύου. Ο High Availability Wizard πραγματοποιεί τα υπόλοιπα.

Με τη διαχείριση συμπλέγματος ανακατεύθυνσης, ένας διαχειριστής είναι δυνατόν να διατηρήσει πολλαπλά συμπλέγματα στην οργάνωση. Τα συμπλέγματα μπορούν να διαχειρίζονται εξ αποστάσεως μέσω των εργαλείων απομακρυσμένης διαχείρισης διακομιστή.[56]

Για τους έμπειρους διαχειριστές συμπλεγμάτων που θέλουν να ρυθμίσουν περαιτέρω την διαμόρφωση του συμπλέγματος, οι MSC εντολές Cluster.exe εξακολουθούν να είναι διαθέσιμες και να επιτρέπουν την πλήρη πρόσβαση σε όλες τις MSCS διαχειριστικές δυνατότητες. Ωστόσο, οι εντολές Cluster.exe θα αντικατασταθούν με νέα cmdlet του PowerShell των Windows

[56] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

σε νεότερες εκδόσεις.

3.7.2.Επεκτασιμότητα

Σύμφωνα με το MSCS, ο μέγιστος αριθμός των κόμβων που θα μπορούσαν να υπάρχουν σε ένα σύμπλεγμα ήταν οκτώ. Το WSFC έχει αυξήσει αυτό το όριο σε δεκαέξι x64 κόμβους σε ένα σύμπλεγμα.[57]

3.7.3.Κατοχύρωση

Το Kerberos τώρα χρησιμοποιείται για την ταυτοποίηση του χρήστη. Όλη η επικοινωνία μεταξύ των κόμβων διαθέτει υπογραφή και είναι δυνατόν να κρυπτογραφηθεί.

Η Microsoft ενθαρρύνει σθεναρά τη χρήση των εφεδρικών, ξεχωριστά και ευδιάκριτα δρομολογημένων δικτύων για την παροχή ανθεκτικότητας σφαλμάτων στο ιδιωτικό δίκτυο που συνδέει τους κόμβους. Αν αυτό δεν παρέχεται, το WSFC θα δημιουργήσει ένα μήνυμα προειδοποίησης και το σύμπλεγμα είναι δυνατόν να μη γίνει αποδεκτό από τη Microsoft για την υποστήριξη.

Στην περίπτωση των εφεδρικών δικτύων, το ταχύτερο δίκτυο θα λάβει προτεραιότητα στην εσωτερική κίνηση. Σύμφωνα με το MSCS, η μέγιστη επιτρεπόμενη καθυστέρηση πάνω από το ιδιωτικό δίκτυο ήταν 500 χιλιοστά του δευτερολέπτου. Αυτό ίσχυε λόγω των περιορισμών των heartbeat. Τα διαστήματα Heartbeat δεν είχαν ρυθμιστεί. Στο WSFC, οι παράμετροι των heartbeat επιδέχονται ρύθμισης και ο περιορισμός της καθυστέρησης έχει αφαιρεθεί. Επιπλέον, αντί της μετάδοση heartbeats, το WSFC τώρα χρησιμοποιεί το πρωτόκολλο συνδέσεων TCP/IP για τη βελτίωση της αξιοπιστίας των heartbeat. Υποστηρίζονται επίσης και τα IPv6 και IPv4.

---

[57] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*, March 2003

3.7.4.Δίκτυα

Σύμφωνα με το MSCS, τα μέλη των συμπλεγμάτων και στις δύο περιοχές σε ένα γεωγραφικά διάσπαρτο σύμπλεγμα πολλαπλών περιοχών θα έπρεπε να βρίσκεται στο ίδιο υποδίκτυο. Αυτό σήμαινε ότι το ιδιωτικό δίκτυο που διασυνδέει τις δύο περιοχές έπρεπε να είναι ένα VLAN (Virtual LAN) που απλώνεται πάνω από ένα σύνδεσμο επικοινωνιών WAN. Ο περιορισμός αυτός έχει αφαιρεθεί από το WSFC. Τα μέλη των συμπλεγμάτων σε κάθε χώρο περιοχή τώρα μπορούν να βρίσκονται σε διαφορετικά δευτερεύοντα δίκτυα που συνδέονται με μία απλή (εφεδρική) σύνδεση WAN. Δεν χρειάζεται να δημιουργηθεί VLAN.[58]

3.7.5.Ενσωμάτωση Hyper-V

Το WSFC είναι ενσωματωμένο με υπηρεσίες Hyper -V virtualization της Microsoft. Κάθε κόμβος του συμπλέγματος είναι δυνατόν να φιλοξενήσει εικονικές μηχανές και οι εικονικές μηχανές μπορούν να αποτύχουν κατά τη διάρκεια μεμονωμένα ή μαζικά. Η πραγματική μεταφορά των εικονικών μηχανών κατ' εντολήν του διαχειριστή προκύπτει σε κλάσματα του δευτερολέπτου χωρίς αντιληπτό downtime και χωρίς απώλεια συνδέσεων.

3.7.6.Επικύρωση

Η επικύρωση ενός οδηγού διαμόρφωσης είναι δυνατόν να τρέξει για να εξασφαλιστεί ότι μια ρύθμιση των παραμέτρων του συμπλέγματος θα υποστηρίζεται από τη Microsoft. Επικυρώνει όλα τα στοιχεία του υλικού του συστήματος (hardware) έναντι της λίστας συμβατότητας υλικού της Microsoft και επικυρώνει τη διαμόρφωση του συμπλέγματος. Είναι χρήσιμο όχι μόνο όταν δημιουργείται ένα σύμπλεγμα, αλλά είναι δυνατόν επίσης να χρησιμοποιηθεί και για την περιοδική επικύρωση της διαμόρφωσης του

[58] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

συμπλέγματος.[59]

3.7.7.Αναβαθμίσεις Rolling

Οι κόμβοι μπορούν να αναβαθμιστούν αφαιρώντας έναν κάθε φορά από το σύμπλεγμα, αναβαθμίζοντάς τον και στη συνέχεια τοποθετώντας τον πάλι στο σύμπλεγμα. Ωστόσο, η μεταφορά από συμπλέγματα τύπου MSCS δεν υποστηρίζεται ειδικώς. Το Migration Wizard είναι διαθέσιμο για να βοηθήσει αυτές τις μεταφορές.[60]

3.8.Η Τυχαία Μονοεκπομπή (Anycast) ως χαρακτηριστικό εξισορρόπησης φόρτου εργασίας

Ο οργανισμός Χ ασχολείται με την τεχνολογία της πληροφορίας και αποτελείται από πολλές υπο-ομάδες που η κάθε μία παρέχει μία υπηρεσία όπως DNS, LDAP, HTTP proxy και ούτω καθ' εξής. Κάθε μία έχει αναπτυχθεί σε παγκόσμιο επίπεδο, με τους δικούς της μηχανισμούς αναπαραγωγής. Η ομάδα μας παρέχει υπηρεσίες εξισορρόπησης φόρτου εργασίας (load balancing) και ανακατεύθυνσης (failover) με έναν τρόπο που οι άλλες ομάδες μπορούν να χρησιμοποιήσουν χωρίς να πρέπει να διαχειριστούν την υποκείμενη τεχνολογία. Πρόσφατα προστέθηκε η τυχαία μονοεκπομπή (Anycast) ως υπηρεσία που προσφέρει σε άλλες ομάδες που πρέπει να είναι σε θέση να πραγματοποιήσουν failover μεταξύ των εξισορροπητών φόρτου εργασίας (Load Balancers). Ενώ η εφαρμογή Anycast είναι ένα πολύπλοκο και μυστηριώδες στοιχείο για αρκετούς διαχειριστές συστημάτων, η αρχιτεκτονική μας παρέχει την υπηρεσία με

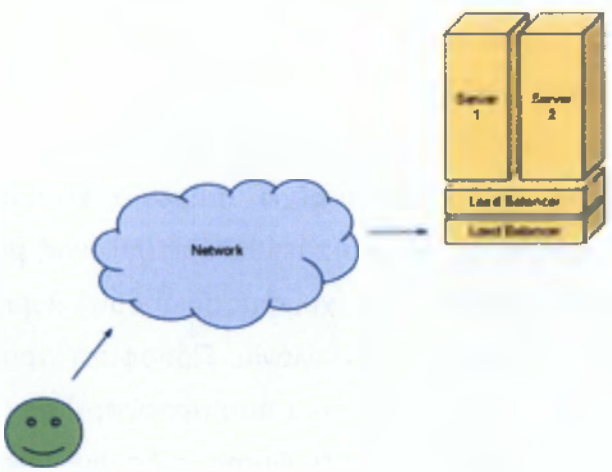
[59] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

[60] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

έναν τρόπο που οι άλλες ομάδες δεν χρειάζεται να ανησυχούν για τις λεπτομέρειες. Παρέχουν απλώς την υπηρεσία πίσω από τους Load Balancers που χρησιμοποιούν επί του παρόντος, με μία επιπλέον εικονική διεύθυνση IP. Αυτή η μελέτη περιγράφει τον τρόπο με τον οποίο λειτουργεί η Anycast, τα οφέλη του και την αρχιτεκτονική που χρησιμοποιήσαμε για την παροχή του Anycast failover ως υπηρεσία.

Ο πιο απλός τρόπος για να σκεφτεί κανείς την εξισορρόπηση φόρτου, είναι να τοποθετήσει όσα περισσότερα αντίγραφα υπηρεσιών χρειάζονται στο δωμάτιο των διακομιστών (server room) και να έχει έναν Load Balancer να κατανέμει τον φόρτο εργασίας μεταξύ τους. Για να αυξηθεί η αξιοπιστία, οι Load Balancers συνήθως αναπτύσσονται σε ζεύγη μεγάλης διαθεσιμότητας. [61]

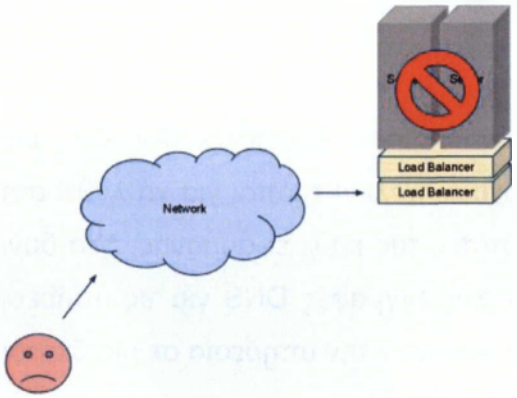
Εικόνα 3: Σύνηθες Σενάριο Load Balancing



[61] Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003

Το παραπάνω αποτελεί ήδη μία βελτίωση στην αξιοπιστία, αλλά είναι δυνατόν να προχωρήσει και περαιτέρω. Φανταστείτε ένα σενάριο καταστροφής, όπου οι χρήστες είναι ακόμα ενεργοί και ζητούν την υπηρεσία και έτσι είναι και ο εξισορροπητής φόρτου, αλλά όλα τα συστήματα υποστήριξης για μια συγκεκριμένη υπηρεσία δεν είναι. Η λύση αυτή από μόνη της δεν θα λύσει το πρόβλημα [62]

Εικόνα 4: Αποτυχημένη λειτουργία

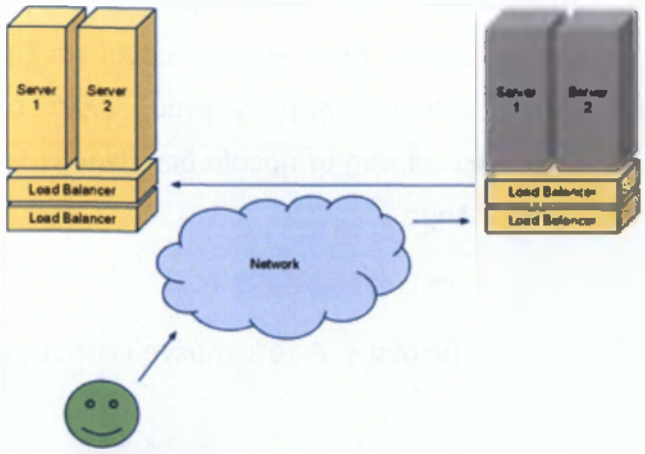


Ένα καλύτερο σχέδιο θα ανακατευθύνει αυτόματα όλους εκείνους τους πελάτες σε μια άλλη θέση (ή server room), καθιστώντας τη διαδικασία όσο το δυνατόν διαφανέστερη. Ένας τρόπος για να επιτευχθεί αυτό είναι να εντοπιστεί η πλησιέστερη δευτερεύουσα θέση και να ρυθμιστεί ο εξισορροπητής φόρτου στον διακομιστή μεσολάβησης (proxy) ή να ανακατευθυνθούν εκεί όλες οι αιτήσεις των χρηστών, μέχρι να αποκατασταθεί η τοπική υπηρεσία. Τα περισσότερα προϊόντα εξισορρόπησης φόρτου προσφέρουν αυτόματη ανακατεύθυνση[63]

[62] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*; March 2003

[63] Server Clusters: Architecture Overview for Windows Server 2003. *Microsoft White Paper*; March 2003

Εικόνα 5: Απομακρυσμένο failover

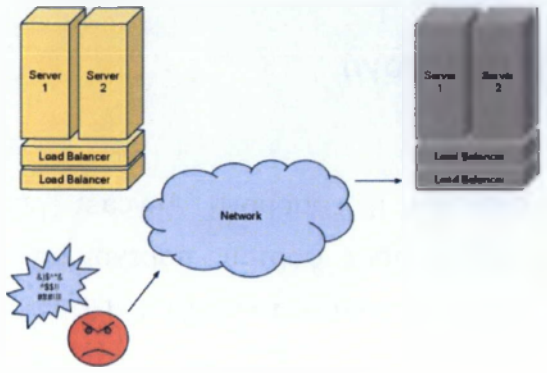


Αλλά τι γίνεται αν οι Load Balancers δεν είναι επίσης διαθέσιμοι; (Βλέπε: Σχέδιο 4) Υπάρχουν πολλοί τρόποι για να λυθεί αυτό το πρόβλημα, ανάλογα με τις ιδιαιτερότητες της κάθε εφαρμογής. Μία δυνατότητα θα ήταν να γίνει μία ενημέρωση στις εγγραφές DNS για τις υπηρεσίες, έτσι ώστε οι χρήστες να μπορούν να φθάσουν την υπηρεσία σε μια διαφορετική θέση.

Ενδεχομένως, αυτή η αναβάθμιση DNS είναι δυνατόν να είναι αυτοματοποιημένη, αλλά θα πρέπει να υπάρχει ένας μηχανισμός για να ελέγχει την κατάσταση της υπηρεσίας σε άλλες θέσεις και να παρακολουθεί την κατάστασή τους, έτσι ώστε το σύστημα να ξέρει πού να στείλει τους χρήστες σε περίπτωση αποτυχίας. Λαμβάνοντας υπόψη ότι οι υπηρεσίες συχνά αναπτύσσονται σε εκατοντάδες περιοχές, δεν θα ήταν αποτελεσματικό να υπάρχει μία κεντρική θέση που θα συλλέγει όλες τις πληροφορίες σχετικά με τις υπηρεσίες, έτσι ώστε ο μηχανισμός ενημέρωσης του DNS θα πρέπει να διανεμηθεί σε τόσες θέσεις όσες θα έχει αναπτυχθεί η υπηρεσία. Αυτό θεωρείτε μία μη βέλτιστη λύση, αφού υπάρχει η δυνατότητα να ενσωματωθεί η παρακολούθηση και το αυτόματο failover στην υπάρχουσα υποδομή εξισορρόπησης φόρτου.[64]

[64] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

Εικόνα 6: Απομακρυσμένο failover – αποτυχημένη λειτουργία



3.9.Βασικά στοιχεία του Anycast

Το DNS TTL είναι δυνατόν επίσης να αποτελεί βάρος. Μερικές φορές δεν είναι δυνατόν να χρησιμοποιηθούν πολύ μικρές TTL και ο χρόνος που χρειάζεται για να διαδοθούν οι DNS αλλαγές θα εξακολουθήσει να αποτελεί downtime από τη σκοπιά των χρηστών. Μόλις επανέλθει το σύστημά σας, παρουσιάζεται και πάλι η ανάγκη για ενημέρωση των DNS εγγραφών ώστε να κατευθυνθούν οι χρήστες πίσω στην αρχική θέση.

Η εφαρμογή Anycast είναι μία τεχνική δρομολόγησης δικτύου, όπου πολλές συσκευές χρηστών (hosts) έχουν ακριβώς την ίδια διεύθυνση IP. Οι πελάτες που προσπαθούν να φθάσουν σε αυτήν την διεύθυνση IP δρομολογούνται στον πλησιέστερο server. Εάν αυτοί οι διπλοί servers παρέχουν όλοι την ίδια υπηρεσία, οι πελάτες λαμβάνουν απλά την υπηρεσία από τον host που είναι τοπολογικά πλησιέστερος.

Η εφαρμογή Anycast δεν διαθέτει πληροφορίες για την ειδική κατάσταση της εύρυθμης λειτουργίας της υπηρεσίας, η οποία θα μπορούσε να οδηγήσει σε αιτήματα που αποστέλλονται σε τοποθεσίες που η κατάστασή της υπηρεσίας που εκτελείται δεν είναι εύρυθμη. Εάν μια συγκεκριμένη υπηρεσία έχει περίπου 200 διαφορετικές καταστάσεις, η διαχείριση των ελέγχων της εύρυθμης λειτουργίας και η διαμόρφωση του

Border Gateway Protocol 4 (BGP) για κάθε μία από αυτές τις καταστάσεις είναι δυνατόν να είναι πολύ περίπλοκη.[65]

3.10.Η εφαρμογή

Χρησιμοποιείται η εφαρμογή Anycast για το failover μεταξύ των συστάδων εξισορρόπησης φόρτου, παρέχοντας τα οφέλη της Anycast σε οποιαδήποτε υπηρεσία πίσω από τους Load Balancers μας.

Αυτό μειώνει πολύ την πολυπλοκότητα του περιβάλλοντος του δικτύου, δεδομένου του μειωμένου αριθμού των μηχανών διαφήμισης διαδρομών. Η λύση μας χρησιμοποιεί BGP, διότι επιτρέπει τη δημιουργία μιας ιεραρχίας για την διαφήμιση της διαδρομής, αλλά και άλλα πρωτόκολλα λειτουργούν εξίσου καλά. Χρησιμοποιώντας την Anycast, δεν υπάρχει πλέον ανάγκη για απομακρυσμένο failover με χρήση proxies, παρέχεται μία πιο καθαρή λύση δεδομένου ότι ο πελάτης συνδέεται άμεσα με την τοποθεσία failover, ενώ η διαμεσολάβηση σε πραγματοποιεί συνήθως να χάνεις τις πληροφορίες ταυτοποίησης του πελάτη. Επίσης αποθηκεύει το proxy overhead μεταξύ των διακομιστών και των χρηστών.

Η ομάδα μας παρέχει την Εξισορρόπηση Φόρτου Εργασίας ως υπηρεσία, καθιστώντας την εντελώς ξεχωριστή από τη συγκεκριμένη εγκατάσταση υπηρεσιών. Πολλαπλές υπηρεσίες μπορούν να επωφεληθούν από την ίδια υποδομή εξισορρόπησης φόρτου και η αύξηση του αριθμού των αντιγράφων μιας υπηρεσίας δεν θα αυξήσει την πολυπλοκότητα του σχεδιασμού του δικτύου, δεδομένου ότι υπάρχει ένας ελεγχόμενος αριθμός των διαφημιστών διαδρομής.

Ένα άλλο πλεονέκτημα των Load Balancers ως ζεύγη Anycast είναι η μείωση του αριθμού των αλλαγών δρομολόγησης, γιατί ο Load Balancer συνδυάζει πολλαπλές καταστάσεις μιας υπηρεσίας σε ένα VIP. Αυτό

[65] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

αποτελούσε μία από τις ανησυχίες σχετικά με την ανάπτυξη του Anycast. Έχοντας τον Load Balancer να αντιμετωπίζει τους ειδικούς ελέγχους της εύρυθμης λειτουργίας της υπηρεσίας καθίσταται δυνατή η εγκατάσταση της Anycast όχι μόνο για υπηρεσίες που βασίζονται στο UDP, αλλά και για τις υπηρεσίες που βασίζονται στο TCP.

Διαμορφώσαμε το περιβάλλον δικτύου ώστε να υπάρχει ένα δευτερεύον δίκτυο που προορίζεται για όλες τις εικονικές IPs (VIPs) της Anycast. Οι δρομολογητές έχουν ρυθμιστεί ώστε να δέχονται / 32 διαφημίσεις διαδρομής σε αυτό το υποδίκτυο από τους Load Balancers. Αυτό επιτρέπει την εφαρμογή προστασίας από την εσφαλμένη διαμόρφωση μέσω της χρήσης των ACL που επιτρέπουν μόνο τα δρομολόγια από το καθορισμένο δευτερεύον δίκτυο, αποτρέποντας την τυχαία κατάληψη χώρου IP.

Οι Anycast VIPs μπορούν να ρυθμιστούν πέρα από τις συνήθεις VIPs στους ίδιους Load Balancers.[66]

3.11.Λογισμικό που χρησιμοποιήθηκε για την εφαρμογή

Όλοι οι Load Balancers μας αναπτύχθηκαν σε ζεύγη μεγάλης διαθεσιμότητας για την προστασία από ενιαία μηχανική βλάβη. Για το σκοπό αυτό χρησιμοποιήσαμε το Heartbeat, από το Linux-HA πρότζεκτ[67], το οποία είναι λογισμικό διαχείρισης συστάδας πόρων. Το Heartbeat φέρνει τις διεπαφές δικτύου και το backend λογισμικό διαχείρισης πάνω και κάτω. Αυτά όλα είναι διαμορφωμένα ως πόροι Heartbeat.[68]

Για την παρακολούθηση backend και το failover, χρησιμοποιείται το

[66] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

[67] High Availability, http://www.linux-ha.org

[68] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

ldirectord[69]. Το Ldirectord πραγματοποιεί του ελέγχους της κατάστασης της λειτουργίας έναντι των backends για κάθε μία από τις VIPs μας, προσθέτοντας ή αφαιρώντας από την δεξαμενή εξισορρόπησης φόρτου των καταστάσεων της υπηρεσίας που αλλάζει την κατάσταση της εύρυθμης λειτουργίας. Είναι δυνατόν επίσης να ανακατευθύνει όλες τις συνδέσεις σε διαφορετική τοποθεσία σε περίπτωση αποτυχίας σε όλα τα backends, χρησιμοποιώντας την εναλλακτική επιλογή.

Προσθέσαμε ένα χαρακτηριστικό στο ldirectord, εφαρμόζοντας μία εναλλακτική εντολή διαμόρφωσης: όταν το τελευταίο από τα τοπικά backends σταματήσει, ενεργοποιείται αυτή η εντολή. Αυτό χρησιμοποιείται για να ανέβει και να κατέβει η Anycast διεύθυνση IP με βάση την κατάσταση του backend.

Το Ldirectord επικοινωνεί απευθείας με το ifconfig (για να φέρει τις IP πάνω και κάτω) και την ip_vs (μέσω ipvsadm) για να προσθέσει και να αφαιρέσει υπηρεσίες backends από την δεξαμενή της εξισορρόπησης φόρτου.

Το ip_vs είναι ένα κύκλωμα (module) του πυρήνα Linux για την εξισορρόπηση του φόρτου εργασίας[70]. Υποστηρίζει το tunnelling, το μισό της NAT και λειτουργίες Direct Routing (DR). Στην εγκατάστασή μας, όλες οι VIPs έχουν ρυθμιστεί χρησιμοποιώντας DR.

Το Quagga[71] είναι πακέτο λογισμικού δρομολόγησης δικτύου, που επιτρέπει σε ένα GNU / Linux σύστημα να συμμετάσχει στα δικτυακά πρωτόκολλα δρομολόγησης. Στη λύση, χρησιμοποιείται η εφαρμογή BGP για να ενεργοποιηθούν οι Load Balancers να διαφημίσουν τις BGP διαδρομές προς τις συσκευές του δικτύου.

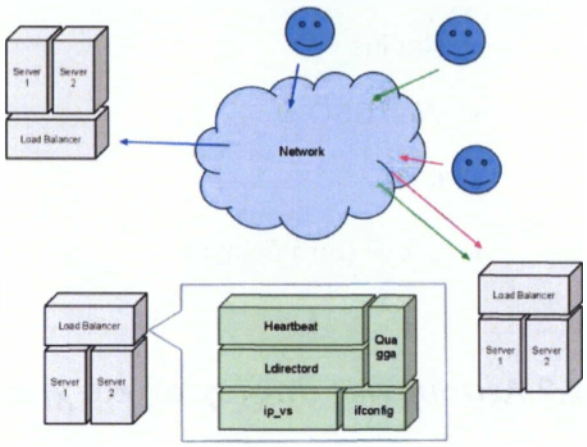
Διατίθεται μια IP για κάθε υπηρεσία και δημιουργείται το πρότυπο διαμόρφωσης για το LVS για να ανέβουν οι VIPs και χρησιμοποιείται το χαρακτηριστικό που προστίθεται στο ldirectord για να ανέβει ή να κατέβει η διασύνδεση του δικτύου της Anycast IP, ανάλογα με την κατάσταση των backends. Εάν όλα τα backends είναι κατεβασμένα, το ldirectord θα

[69] Ldirectord, http://www.vergenet.net/linux/ldirectord/
[70] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[71] Quagga, a software routing suite, http://www.quagga.net

κατεβάσει την IP από εκείνη την VIP και το Quagga αμέσως θα αφήσει τους δρομολογητές να το μάθουν, ώστε να αποσύρουν τη διαδρομή προς εκείνη την VIP.[72]

Εικόνα 7: Εξισορρόπηση Φόρτου Εργασίας με αρχιτεκτονική λογισμικού Anycast



Η εικόνα στα δεξιά δείχνει την αρχιτεκτονική του δικτύου, καθώς και λεπτομέρειες σχετικά με την εγκατάσταση του λογισμικού.

3.12.Προσθήκη νέων υπηρεσιών στην εγκατάσταση

Υπηρεσίες μπορούν να προστεθούν στην Anycast απλώς ρυθμίζοντας τα backends τους σε μία VIP πάνω σε μια ενεργοποιημένη Anycast εξισορρόπησης φόρτου. Στην περίπτωσή μας, αυτό σημαίνει διαμόρφωση των Heartbeat και ldirectord.

Η διαμόρφωση του δικτύου θα είναι ήδη σε ισχύ, μειώνοντας σημαντικά το εμπόδιο εισόδου για νέες υπηρεσίες. Η επέκταση μίας υπηρεσίας σε νέα/ες τοποθεσία/ες γίνεται με την ίδια ακριβώς διαδικασία, η

[72] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

δρομολόγηση της Anycast θα φροντίσει για την αποστολή της κυκλοφορίας του χρήστη στην νέα, πιο κοντινή εγκατάσταση εξισορρόπησης φόρτου.

3.12.1.Υπηρεσίες που χρησιμοποιούν αυτή τη ρύθμιση

- DNS
- HTTP proxy
- Radius
- Web SSO
- NTP
- LDAP (υπό δοκιμή)

3.12.2.Λειτουργίες αποτυχίας και χρόνοι ανάκτησης

Όλοι οι χρόνοι ανάκτησης που αναφέρονται λαμβάνουν υπόψη την συγκεκριμένη εγκατάσταση της Anycast μας. Τα περιβάλλοντα με διαφορετικά χρονικά όρια και παραμέτρους διαμόρφωσης μπορούν να έχουν διαφορετικούς χρόνους απόκρισης και ανάκτησης. Ο χρόνος διάδοσης της διαδρομής διαρκεί λιγότερο από 1 δευτερόλεπτο και υπάρχει θέσει ένα "dead timer" 30 δευτερολέπτων για τους δρομολογητές ώστε να αντιληφθούν τα νεκρά ζεύγη BGP.[73]

Η καθαρή διακοπή της υπηρεσίας BGP διασύνδεσης δημιουργεί μία διακοπή διάρκειας λιγότερο από 1 δευτερόλεπτο για τις υπηρεσίες καθώς ενημερώνονται οι διαδρομές. Στην περίπτωση που όλες οι υπηρεσίες backend καταστούν μη διαθέσιμες, θα χρειαστεί ο χρόνος του healthcheck της υπηρεσίας πλέον του <1s καθυστέρησης της διάδοσης της διαδρομής για την ανάκτηση.

[73] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

Σε μια ξαφνική αποτυχία του δικτύου ή διακοπή ρεύματος στην τοποθεσία, για την ανάκτηση θα χρειαστεί χρόνος για το "dead timer" ώστε να λήξει πλέον της μικρής καθυστέρησης στην διάδοση της διαδρομής.[74]

Η μετακίνηση της δρομολόγησης του Anycast σε μια υπηρεσία διαχείρισης της εξισορρόπησης φόρτου εργασίας, ελαχιστοποιεί το έργο και την πολυπλοκότητα που απαιτείται για τη διαμόρφωση των υπηρεσιών, παρέχοντας ένα γρήγορο, αυτόματο και ενήμερο της απόστασης failover.

Βοηθά επίσης στην μείωση του φόρτου και της πολυπλοκότητας της υποδομής του δικτύου συγκεντρώνοντας τις διαφημίσεις των υπηρεσιών σε ένα σημείο διασύνδεσης ανά περιοχή και μειώνοντας τον βαθμό των αλλαγών δρομολόγησης για την ολοκλήρωση μόνο των αποτυχιών της τοποθεσίας.

3.12.3.Αρχιτεκτονική πελάτη-διακομιστή ν επιπέδων

Σε αρκετές περιπτώσεις η αρχιτεκτονική τριών επιπέδων δεν επαρκεί για την εξυπηρέτηση αυξημένου αριθμού πελατών. Η πρόσθεση κόμβων (συσκευών) γίνεται τις περισσότερες περιπτώσεις οριζόντια για την κατανομή του φορτίου σε περισσότερες συσκευές εξυπηρετητών.

3.12.4.Εξισορρόπηση φορτίου

Η συσκευή εξυπηρετητή που λειτουργεί ως εξισορροπητής φορτίου (load balancer) για τους εξυπηρετητές εφαρμογής (Application Servers). Το λογισμικό που χρησιμοποιείται (squid) ενσωματώνει τις λειτουργίες της αντίστροφης λειτουργίας μεσολαβητή (reverse proxy) και την προσωρινή μνήμη αυτού (reverse proxy cache).

Υπάρχουν τέσσερις παράμετροι που αφορούν τις σελίδες που σώζονται στην προσωρινή μνήμη και παίζουν καθοριστικό ρόλο.

α) Πότε έχει τροποποιηθεί τελευταία φορά η σελίδα (last modified)

[74] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

β) λήξη δημοσιοποίησης (expire) χρησιμοποιείται για την διαγραφή από την προσωρινή μνήμη

γ) έλεγχος προσωρινής μνήμης (cache-control) χρησιμοποιείται για την αποθήκευση ή όχι στην προσωρινή μνήμη

3.13.Εξυπηρετητής εφαρμογής

Οι κόμβοι είναι οι εξυπηρετητές εφαρμογής (τρεις στον αριθμό). Ο κάθε ένας διατηρεί λογισμικό εξυπηρέτησης ιστού (Apache web server), με τα απαραίτητα συστατικά για την υποστήριξή της γλώσσας προγραμματισμού της εκάστοτε εφαρμογής. Ένα ακόμη συστατικό είναι το λογισμικό προσωρινής μνήμης memcache , το οποίο χρησιμοποιείται για την αποθήκευση αποτελεσμάτων από την βάση στοιχείων . Η λειτουργία του βασίζεται σε μία κοινή εικονική μνήμη (pool) όπου ο κάθε εξυπηρετητής προσθέτει την δική του μνήμη (RAM). Το αποτέλεσμα είναι μια μνήμη τριών εξυπηρετητών με ποσότητα του κάθε ένα 4GB, σε σύνολο 3x4=12GB, όπου έχουν πρόσβαση και οι τρεις εξυπηρετητές. [75]

Ο πρώτος αναζητά το αποτέλεσμα της κλήσης της βάσης στοιχείων , καταρχήν στην εικονική μνήμη (μέσω της εφαρμογής) και έπειτα αν δεν υπάρχει, πραγματοποιεί την κλήση και αποθηκεύει το αποτέλεσμα στην μνήμη. Ο δεύτερος εξυπηρετητής αναζητά το αποτέλεσμα στην εικονική μνήμη και χωρίς να πραγματοποιήσει κλήση στην βάση στοιχείων βρίσκει το αποθηκευμένο αποτέλεσμα από τον πρώτο εξυπηρετητή. Το αποτέλεσμα είναι η ελάττωση των αιτημάτων στην βάση στοιχείων και η ταχύτατη απόκριση των αιτημάτων, λόγω της αποθήκευσης των αποτελεσμάτων στην μνήμη RAM.

[75] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

3.14.Εξυπηρετητές βάσης στοιχείων

Η λειτουργία αντιγράφων είναι η τεχνολογία που επιτρέπει την αντιγραφή στοιχείων σε αρκετούς εξυπηρετητές βάσης στοιχείων . Ο τρόπος με τον οποίο υλοποιείται αυτό είναι ο εξής : ένας εξυπηρετητής ορίζεται ως κύριος (master) και αποτυπώνει κάθε εκτέλεση αιτήματος σε ένα αρχείο ιστορικού (binary log). Οι υπόλοιποι εξυπηρετητές λειτουργούν ως υποκείμενοι (slaves) και αιτούνται πληροφορίες εκτέλεσης από το αρχείο ιστορικού στον κύριο εξυπηρετητή. Ο κύριος εξυπηρετητής δεν γνωρίζει πόσοι υποκείμενοι εξυπηρετητές υπάρχουν, απλά επιστρέφει απαντήσεις αιτημάτων στους εξυπηρετητές πού έχουν το δικαίωμα να τις πραγματοποιούν. Από την έκδοση 5.5 της βάσης στοιχείων Mysql υποστηρίζεται η ημισύγχρονη (semi-synchronous) λειτουργία αντιγράφων. Σε αυτή την λειτουργία ο κύριος εξυπηρετητής δεν επιστρέφει αποτέλεσμα για το αίτημα στην βάση στοιχείων , το οποίο έχει πραγματοποιηθεί, αν τουλάχιστον ένας υποκείμενος εξυπηρετητής δεν έχει αποθηκεύσει στο αρχείο ιστορικού του την διεργασία αυτή. Στις περισσότερες περιπτώσεις χρειαζόμαστε μεγαλύτερη ταχύτητα απόκρισης, σε σχέση με την ποσότητα των αιτημάτων και αυτό επιτυγχάνεται με την διασπορά των αιτημάτων ανάγνωσης σε αρκετούς υποκείμενους εξυπηρετητές. [76]

3.15.Συστοιχίες μεγάλης διαθεσιμότητας

Συστοιχίες μεγάλης διαθεσιμότητας (High Availability clusters) είναι ομάδες εξυπηρετητών που μπορούν να χρησιμοποιηθούν αξιόπιστα με ελάχιστο χρόνο μη διαθεσιμότητας. Χρησιμοποιούνται εξυπηρετητές ή ομάδες εξυπηρετητών που συνεχίζουν την λειτουργία όταν ορισμένα μέρη του συστήματος αποτυγχάνουν. Όταν σ' ένα σύστημα έχουμε αποτυχία υλικού ή λογισμικού το σύστημα δυσλειτουργεί ή αποτυγχάνει. Στην περίπτωση της

[76] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

συστοιχίας μεγάλης διαθεσιμότητας, αυτό αντιμετωπίζεται με τον μηχανισμό αποτυχίας (failover) όπου η αποτυχία ανιχνεύεται και ενεργοποιείται η συγκεκριμένη υπηρεσία σε άλλο διακομιστή. Ανάλογα με τον τρόπο λειτουργίας κατηγοριοποιούνται ως εξής :

α) ενεργός προς ενεργό κόμβο (active/active). Η κυκλοφορία που προορίζεται για τον κόμβο που αποτυγχάνει, μεταβιβάζεται σε άλλο ενεργό κόμβο ή διαμοιράζεται μέσω του εξισορροπητή φορτίου σε άλλους κόμβους. [77]

Αυτό είναι μόνο δυνατό όταν οι κόμβοι χρησιμοποιούν ομοιογενές λογισμικό. β) Ενεργός προς παθητικό κόμβο (active/passive). Παρέχει για κάθε κόμβο ένα κόμβο αποτυχίας ο οποίος ενεργοποιείται όταν ο πρωτεύον αποτύχει. Είναι η πιο απαιτητική σε υλικό συστοιχία διότι χρειαζόμαστε τούς διπλάσιους κόμβους για να λειτουργήσει.

γ) N+1. Παρέχει μόνο έναν επιπλέον κόμβο ο οποίος αναλαμβάνει τον ρόλο του κόμβου που αποτυγχάνει. Στην περίπτωση που οι συστοιχία διατηρεί ετερογενές λογισμικό, ο επιπλέον κόμβος είναι σε θέση να παρέχει τις υπηρεσίες κάθε κόμβου που αποτυγχάνει. Στην περίπτωση που η συστοιχία διατηρεί ομοιογενές λογισμικό το μοντέλο είναι όμοιο με το β.

δ) N+M. Στις περιπτώσεις που η συστοιχία διατηρεί πολλές υπηρεσίες το μοντέλο N+1 δεν είναι αρκετό για να παρέχει υψηλή διαθεσιμότητα. Στην περίπτωση αυτή περισσότεροι από ένας κόμβοι (M) λειτουργούν ως εναλλακτικοί. Ο αριθμός των εναλλακτικών κόμβων έχει αντίστροφη σχέση μεταξύ κόστους και αξιοπιστίας.

ε) N προς 1. Λειτουργία στην οποία ο κόμβος αναμονής αποτυχίας (failover standby) γίνεται ενεργός προσωρινά έως ότου αποκατασταθεί ή ενεργοποιηθεί ο κόμβος προς 1. Λειτουργία στην οποία ο κόμβος αναμονής αποτυχίας (failover standby) γίνεται ενεργός προσωρινά έως ότου αποκατασταθεί ή ενεργοποιηθεί ο κόμβος από τον κόμβο που έχει αποτύχει στους υπόλοιπους ενεργούς κόμβους. Στο μοντέλο αυτό χρειάζεται οι κόμβοι

[77] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

που είναι ενεργοί να διαθέτουν επιπλέον χωρητικότητα, για την κάλυψη των αναγκών σε περίπτωση αποτυχίας κάποιου κόμβου.

Για την διαχείριση των συμβάντων χρησιμοποιείται λογισμικό διαχείρισης πόρων συστοιχίας (Cluster Resource Manager). Αυτό είναι δυνατόν να δημιουργήσει αλλοίωση των στοιχείων στους κόμβους βάσεων στοιχείων . Ν προς Ν. Οι βασικοί εξυπηρετητές βάσης στοιχείων συνδέονται μεταξύ τους με ζεύξη τύπου DRBD η οποία αναλαμβάνει την λειτουργία αντιγράφων. Το λογισμικό της βάσης στοιχείων είναι ενεργό μόνο στον ενεργό εξυπηρετητή. Στην περίπτωση αποτυχίας ο παθητικός κόμβος γίνεται ενεργός και ξεκινά η υπηρεσία της βάσης στοιχείων . Οι βασικοί εξυπηρετητές βάσης στοιχείων λειτουργούν σύμφωνα με το μοντέλο ενεργού παθητικού κόμβου ενώ οι υποκείμενοι σύμφωνα με το μοντέλο Ν προς Ν. [78]

3.16.Αποθήκευση στοιχείων σε συστοιχία μεγάλης διαθεσιμότητας

Για την κατοχύρωση των στοιχείων στους εξυπηρετητές, χρησιμοποιείται η τεχνολογία συστοιχίας ανεξάρτητων δίσκων (RAID). Μέσω υλικού ή λογισμικού οι σκληροί δίσκοι διατηρούν αντίγραφα των πληροφοριών σε πολλές τοποθεσίες. Όταν όμως έχουμε αστοχία υλικού του εξυπηρετητή η πληροφορίες δεν είναι προσβάσιμες. Η τεχνολογία που χρησιμοποιείται για αποθήκευση στοιχείων μεγάλης διαθεσιμότητας είναι το κατανεμημένο σύστημα αρχείων (Distributed File System). Στις ακόλουθες παραγράφους παρουσιάζονται κατανεμημένα συστήματα αρχείων με παραδείγματα χρήσης. [79]

[78] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[79] Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355

3.16.1.Hadoop

Το σύστημα αρχείων Hadoop είναι κατανεμημένο, επεκτάσιμο, με δυνατότητες προσαρμογής, γραμμένο σε γλώσσα προγραμματισμού JAVA. Οι κόμβοι στοιχείων (data nodes) αποθηκεύουν τα δεδομένα σε τμήματα (blocks), σύμφωνα με τις οδηγίες του κόμβου ονοματοδοσίας, σε πολλαπλές θέσεις (συνήθως τρεις) για θέμα κατοχύρωση ς χωρίς να χρειάζεται η χρησιμοποίηση της τεχνολογία συστοιχίας ανεξάρτητων δίσκων. Είναι σχεδιασμένο να αποθηκεύει αρχεία μεγάλης χωρητικότητας. Το ιδανικό μέγεθος είναι πολλαπλάσιο των 64MB και υποστηρίζει συμπίεση στοιχείων τύπου bzip2.

3.16.2.MogileFS

Είναι κατανεμημένο σύστημα αρχείων, υψηλών επιδόσεων και επεκτάσιμο. Τα δεδομένα σώζονται στους κόμβους αποθήκευσης (Storage Node) σύμφωνα με τις οδηγίες των ιχνηλατών (tracker nodes). Όπως και το σύστημα hadoop αποθηκεύει τα δεδομένα σε τουλάχιστον τρεις κόμβους στοιχείων , για κατοχύρωση , υποστηρίζει συμπίεση των στοιχείων bzip2 και είναι σχεδιασμένο για την αποθήκευση στοιχείων μεγάλης χωρητικότητας (τμήματα των 128MB) χωρίς να χρειάζεται η χρησιμοποίηση συστοιχίας ανεξάρτητων δίσκων. Τα δεδομένα της διαχείρισης των αρχείων σώζονται από τους ιχνηλάτες σε βάση στοιχείων Mysql.

3.16.3.GlusterFS

Είναι κατανεμημένο, επεκτάσιμο σύστημα αρχείων το οποίο χρησιμοποιείται είτε για ταχύτητα διαμεταγωγής στοιχείων , χρησιμοποιώντας κατανεμημένα δεδομένα (striping) στους κόμβους αποθήκευσης είτε για κατοχύρωση διατηρώντας πολλαπλά αντίγραφα (replicating) στους κόμβους αποθήκευσης. Κάθε κόμβος αποθήκευσης εξάγει το τοπικό σύστημα αρχείων σαν τόμο (volume). Οι τόμοι συνθέτουν ένα τελικό τόμο ο οποίος είναι το άθροισμα του συνόλου των τόμων των κόμβων αποθήκευσης. Ο πελάτης είναι δυνατόν εκτός από την αποθήκευση αρχείων να εκτελέσει και

εφαρμογές στον τόμο αποθήκευσης λόγω του ότι το σύστημα αρχείων είναι συμμορφωμένο με τα πρότυπα λειτουργικών unix (posix compliant). [80]

3.16.4.Lustre

Στους κόμβους αποθήκευσης είναι δυνατόν να χρησιμοποιηθεί η δυνατότητα κατανεμημένων αντιγράφων για την αύξηση της ταχύτητας απολαβής των στοιχείων όμως την κατοχύρωση των στοιχείων πρέπει να αναλάβει σύστημα αποτυχίας δίσκων (disk failover). Η διαδικασίες των στοιχείων όπως αποθήκευση, αντιγραφή, μετονομασία κ.τ.λ πραγματοποιούνται από τον πελάτη στους κόμβους αποθήκευσης υπό την εποπτεία του κόμβου στοιχείων. Ο πελάτης είναι δυνατόν να συνδέσει την συστοιχία αποθήκευσης, ως τοπικό ή απομακρυσμένο σύστημα αρχείων, το οποίο είναι το σύνολο του αποθηκευτικού χώρου της συστοιχίας. Το σύστημα είναι δυνατόν να διαχειριστεί μέχρι 2 δισεκατομμύρια αρχεία και χωρητικότητα 32 Petabytes.[81]

[80] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

[81] The Linux Virtual Server Project, http://www.linuxvirtualserver.org

Κεφάλαιο 4

Αλγόριθμοι Load Balancing

4.1.NETWORK ADDRESS TRANSLATION (NAT)

Σχετικά με το NAT πρέπει να γνωρίζουμε ότι η μετάφραση διευθύνσεων μπορεί να γίνει στατικά ή δυναμικά. Στην πρώτη περίπτωση, η ανάθεση του NAT-IPs είναι σαφής, στην τελευταία περίπτωση δεν είναι. Στην static-NAT μια καθορισμένη αρχική IP μεταφράζεται στην ίδια IP NAT, ανά πάσα στιγμή, και καμία άλλη IP δεν μεταφράζεται στο ίδιο NAT-IP, ενώ στο δυναμικό NAT το IP NAT εξαρτάται από διάφορες συνθήκες εκτέλεσης και μπορεί να είναι εντελώς διαφορετική για κάθε μεμονωμένη σύνδεση.[82]

Στα ακόλουθα τμήματα m, n ορίζονται ως εξής:

m: αριθμός των IPs που πρέπει να μεταφραστούν (original IP)

n: αριθμός των IPs που διατίθενται για τη μετάφραση (NAT IPs)

Static Network Address Translation

Η στατική μετάφραση διευθύνσεων μπορεί να μεταφράσει μεταξύ των δικτύων IP που έχουν το ίδιο μέγεθος (περιέχουν τον ίδιο αριθμό των IPs). Μια ειδική περίπτωση είναι όταν τα δύο δίκτυα περιέχουν μόνο μία IP, δηλαδή η μάσκα δικτύου είναι 255.255.255.255. Αυτή η στρατηγική NAT είναι εύκολο να εφαρμοστεί, δεδομένου ότι η όλη διαδικασία της μετάφρασης μπορεί να γραφτεί ως μία γραμμή που να περιέχει μερικούς απλούς λογικούς μετασχηματισμούς:new-address = νέο δίκτυο ή (old-address AND (NOT netmask) [83]

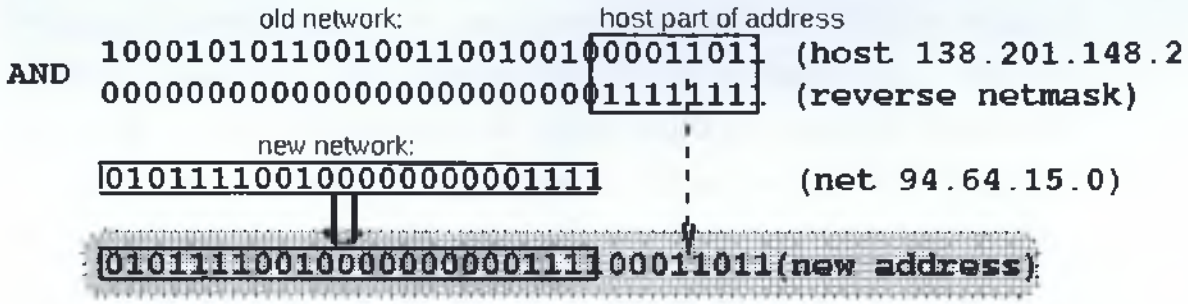
Επιπλέον, δεν υπάρχουν πληροφορίες σχετικά με την κατάσταση των συνδέσεων που μεταφράζονται , κοιτάζοντας κάθε πακέτο IP ξεχωριστά.

[82] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002

[83] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002

Συνδέσεις από το εξωτερικό δίκτυο προς τα μέσα του δικτύου δεν είναι πρόβλημα, το μόνο που φαίνεται να έχουν είναι διαφορετική IP από ό, τι στο εσωτερικό, έτσι η στατική NAT είναι (σχεδόν) transparent.[84]

Εικόνα 8: Static Network Address Translation



Dynamic Address Translation

Η Δυναμική μετάφραση διευθύνσεων είναι απαραίτητη όταν ο αριθμός των IPs για μετάφραση δεν ισούται με τον αριθμό των IPs που μεταφράζονται, αλλά για κάποιο λόγο δεν είναι επιθυμητό να υπάρχει μια στατική χαρτογράφηση. Ο αριθμός των hosts που επικοινωνούν γενικά περιορίζεται από τον αριθμό των IPs NAT που είναι διαθέσιμα . Όταν όλες οι IPs NAT χρησιμοποιούνται τότε καμία άλλη σύνδεση δεν μπορεί να μεταφραστεί και ως εκ τούτου πρέπει να απορριφθεί από το δρομολογητή NAT ,στέλνοντας « host unreachable " .[86]

Η Dynamic NAT είναι πιο περίπλοκη από ό, τι η στατική NAT , διότι πρέπει να παρακολουθούμε την επικοινωνία που φιλοξενεί και ενδεχομένως και των συνδέσεων που προϋποθέτει την εξέταση του TCP πληροφοριών σε πακέτα .Όπως αναφέρθηκε ανωτέρω, η δυναμική NAT μπορεί επίσης να είναι χρήσιμη όταν υπάρχουν αρκετά NAT IPs, δηλαδή όταν το m = n. Μερικοί χρησιμοποιούν αυτό ως μέτρο ασφαλείας, είναι αδύνατο για κάποιον έξω από ένα δίκτυο να πάρει χρήσιμα πακέτα IP για να συνδεθεί με hosts πίσω από ένα NAT router που κάνει δυναμική μετάφραση διευθύνσεων κοιτάζοντας τις

[84] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002

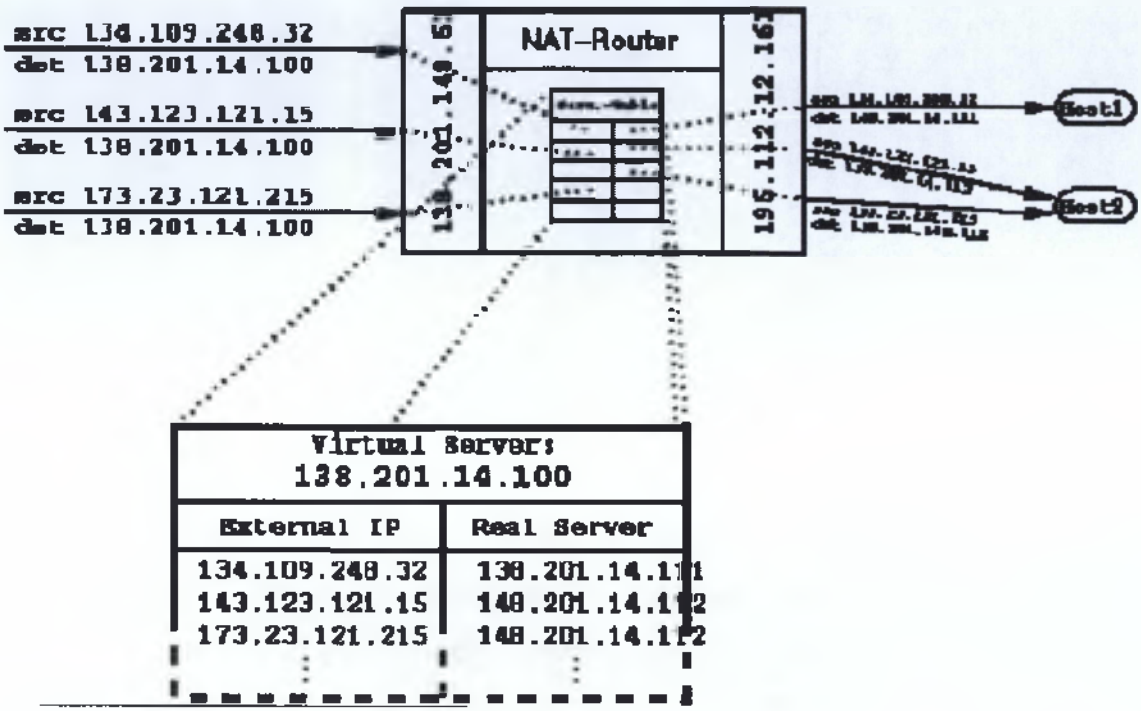
συνδέσεις που λαμβάνουν χώρα , δεδομένου ότι την επόμενη φορά μπορεί να συνδεθείτε χρησιμοποιώντας μια εντελώς διαφορετική IP.[86]

Συνδέσεις από έξω είναι δυνατόν μόνο όταν ο host που πρέπει να φθάσει εξακολουθεί να έχει μια NAT -IP εκχωρημένη , δηλαδή εάν εξακολουθεί να έχει μια καταχώρηση στο δυναμικό πίνακα NAT , όπου ο δρομολογητής NAT παρακολουθεί την εσωτερική IP που αντιστοιχίζεται με το NAT IP . Για παράδειγμα ,στις non-passive FTP συνεδρίες , όπου ο διακομιστής επιχειρεί να δημιουργήσει τα δεδομένα καναλιών , δεν είναι πρόβλημα (για το πρωτόκολλο) , δεδομένου ότι , όταν ο server στέλνει τα πακέτα του στο FTP -client υπάρχει ήδη μια καταχώρηση για τον πελάτη στο NAT - table , και είναι εξαιρετικά πιθανό να εξακολουθεί να περιέχει τον ίδιο client - IP στη χαρτογράφηση NAT - IP που υπήρχαν όταν ο client αρχίσει το κανάλι FTP-control, εκτός εάν η σύνοδος FTP έχει μείνει αδρανής για περισσότερο από το χρονικό όριο της εισόδου.[86]

Ωστόσο , εάν ένας "ξένος" θέλει να δημιουργήσει μια σύνδεση με ένα συγκεκριμένο host στο εσωτερικό σε αυθαίρετο χρόνο υπάρχουν δύο δυνατότητες. Ο εσωτερικός host που δεν έχει μια καταχώρηση στον πίνακα NAT και είναι , ως εκ τούτου απρόσιτος , ή έχει μια καταχώρηση , εκτός, αν η IP για να συνδεθεί είναι γνωστή, επειδή η εσωτερική υποδοχή επικοινωνεί με το εξωτερικό. Στην τελευταία περίπτωση, όμως, μόνο η NAT-IP είναι γνωστή, αλλά δεν είναι η εσωτερική IP του host, και αυτή η γνώση είναι έγκυρη μόνο όταν η ανακοίνωση της εσωτερικής υποδοχής λαμβάνει χώρα καθώς το χρονικό όριο της εισόδου στον πίνακα δρομολογητών της NAT.[85]

[85] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002

Εικόνα 9: Dynamic Address Translation



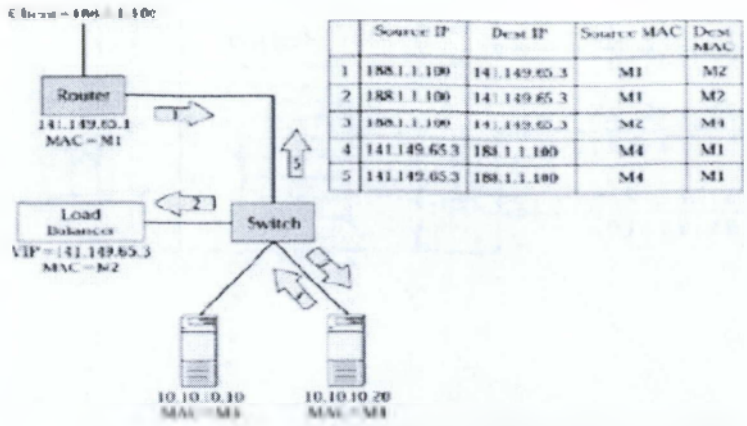
4.2.DIRECT SERVER RETURN (DSR)

Ο Direct Server Return σε αντίθεση με τις άλλες τεχνικές load balancing "μιλάει" κατευθείαν με τον client χωρίς να παρεμβαίνει ο load balancer.Αυτό έχει σαν αποτέλεσμα την μείωση των πακέτων και την αποφυγή συμφόρησης του load balancer.

Αυτό γίνετε, επειδή ο loadbalancer έχει ρυθμιστεί να κάνει μετάφραση της MAC address του παραλήπτη και η IP διεύθυνση του παραμένει ίδια με την VIP του load balancer.Είναι σκόπιμο όλοι οι server μέσα στο σύμπλεγμα να είναι στο ίδιο επίπεδο 2 προκειμένου οι αιτήσεις να φτάσουν στον παραλήπτη μόνο βάσης της MAC address.Σε κάθε server μέσα στο σύμπλεγμα ρυθμίζουμε την loopback IP να είναι ίδια με την VIP address του loadbalancer ,αυτό γίνετε για να μην απορρίψει ο server το εισερχόμενο πακέτο του παραλήπτη.[86]

[86] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002.

Εικόνα 10: Direct server return



	Source IP	Dest IP	Source MAC	Dest MAC
1	188.1.1.100	141.149.65.3	M1	M2
2	188.1.1.100	141.149.65.3	M1	M2
3	188.1.1.100	141.149.65.3	M2	M4
4	141.149.65.3	188.1.1.100	M4	M1
5	141.149.65.3	188.1.1.100	M4	M1

Όπως φαίνεται ο load balancer δεν αλλάζει την IP address με την VIP address αλλά αλλάζει την MAC προορισμού ίδια με την MAC του server που θα δεκτή την αίτηση. Καθώς το switch λειτουργεί στο Layer 2 απλά προωθεί το πακέτο στον server με την επιλεγμένη MAC διεύθυνση, ο server δέχεται το πακέτο και η VIP διεύθυνση ορίζεται ως loopback IP. Όταν ο server απαντήσει στη αίτηση η VIP γίνεται source IP και η διεύθυνση του client γίνεται destination IP,το πακέτο προωθείται από το switch στον router και έπειτα στον client χωρίς την ανάγκη χρήσης NAT και επιπλέον παρακάμψαμε επιτυχώς τον load balancer χωρίς να τον επιβαρύνουμε με επιπλέον εργασία.[87]

4.3.ROUND ROBIN (RR)

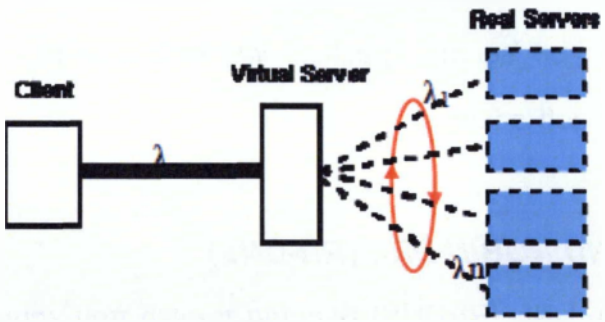
Ο Round Robin αλγόριθμος είναι ο πιο απλός και "δίκαιος", αυτό το πετυχαίνει με τον απλό τρόπο λειτουργίας του. Τις αιτήσεις μέσα στο σύμπλεγμα τις στέλνει κυκλικά, δηλαδή, αρχίζει στέλνοντας την πρώτη αίτηση στον πρώτο server,την δεύτερη αίτηση στον δεύτερο server θεωρώντας τους ίσους σε ισχύει και σε χρόνο απόκρισης. Αν σε περίπτωση έχουμε και άλλες αιτήσεις τις στέλνει κυκλικά.[87]

---

[87] Καταμερισμός φορτίου σε εξυπηρετητές Web server Load-balancing,Σπυρίδων Σ. Παπαδάκης,2007

Ο συγκεκριμένος αλγόριθμος είναι κατάλληλος στα ομογενή συστήματα όπου λ1= λ2= λ3=..... =λn =λ / n (λ= αιτήσεις και n= συνολικός αριθμός server) επιπλέον αν οι αιτήσεις φτάνουν ομοιόμορφα και κάθε αίτηση επεξεργάζεται στο ίδιο χρονικό διάστημα τότε θα έχουμε τον καλύτερο καταμερισμό φόρτου εργασίας και επιπλέον τον καλύτερο καταμερισμό της υπολογιστικής ισχύς των server, δηλαδή Pi = P/ n.[88]

Εικόνα 11: Round Robin



4.3.1 WEIGHTED ROUND ROBIN (WRR)

Αν μέσα στο σύμπλεγμα υπάρχουν server που δεν έχουν τις ίδιες δυνατότητες μεταξύ τους(RAM,CPU,κλπ.) υπάρχει ο αλγόριθμος weighted round robin που είναι μια πιο προχωρημένη έκδοση του αλγόριθμου (RR).Στον αλγόριθμο αυτό ο κάθε server αντιπροσωπεύεται από ένα w(weight) που επιδεικνύει τις δυνατότητες κάθε server.Πιο κάτω βλέπουμε τον ψευδοκώδικα του αλγορίθμου:[88]

```
// calculate number of packets to be served each round by connections


for each flow f

  f.normalized_weight = f.weight / f.mean_packet_size


min = findSmallestNormalizedWeight
```

---

[88] http://en.wikipedia.org/wiki/Weighted_round_robin

```
for each flow f
  f.packets_to_be_served = f.normalized_weight / min

// main loop
loop
  for each non-empty flow queue f
    min(f.packets_to_be_served, f.packets_waiting).times do
      servePacket f.getPacket
```

4.3.2 ROUND ROBIN DNS (RR-DNS)

Round Robin DNS (RRDNS) είναι μια τεχνική που χρησιμοποιείται για την εξισορρόπηση φορτίου κίνησης σε οποιαδήποτε ιστοσελίδα ή FQDN με την ανάγκη της πραγματικής εξισορρόπησης φορτίου του υλικού ή επιπλέον εξοπλισμού. Περιλαμβάνει τη χρήση των DNS servers για να κατανέμουν την κυκλοφορία σε διαφορετικούς φυσικούς servers που μπορεί να ρυθμιστεί ώστε να εξυπηρετεί web, mail ή οποιοδήποτε άλλο είδος της κίνησης . Η τεχνική αυτή χρησιμοποιείται κυρίως σε μεγάλα δίκτυα , όπου το επίπεδο της κυκλοφορίας δεν είναι διαχειρίσιμη από μία μόνο μηχανή . Round Robin DNS εξαρτάται σε μεγάλο βαθμό από την TTL (Time to Live) τιμές που καθορίζονται για τις εγγραφές DNS.

Κάθε εγγραφή DNS έχει πολλές διευθύνσεις IP που της έχουν ανατεθεί . Κάθε φορά που μια αίτηση DNS γίνεται για την ιστορία ενός από τα Ips επιστρέφεται ως αποτέλεσμα σε μια μόδα Round Robin . Αυτό επιτρέπει την κατανομή της κυκλοφορίας μεταξύ των πολλαπλών Ips . Όσο χαμηλότερο είναι το TTL ,πιο γρήγορα οι διευθύνσεις IP περιστρέφεται . Το μειονέκτημα της χρήσης ενός TTL είναι ότι αυξάνει το φορτίο στο διακομιστή DNS. RRDNS είναι πολύ χρήσιμο για τις περιοχές που έχουν πολύ μεγάλη κυκλοφορία και έχουν γεωγραφικά διασκορπισμένο ακροατήριο, καθώς και γεωγραφικά διάσπαρτες web servers. Με τη βοήθεια του RRDNS και άλλη προηγμένη

τεχνική που ονομάζεται ως geolocation , αυτές οι περιοχές είναι σε θέση να ανακατευθύνει το κοινό τους από συγκεκριμένες χώρες / ηπείρους στις τοπικές περιοχές τους .

Round Robin DNS χρησιμοποιείται επίσης για υπηρεσίες όπως μηνύματα, ftp και ακόμη IIRC. Οι περισσότερες εταιρικές περιοχές έχουν servers πολλαπλών ταχυδρομείου έχει ρυθμιστεί σε λειτουργία robin γύρο για να χειριστεί το τεράστιο ποσό της κυκλοφορίας ηλεκτρονικού ταχυδρομείου που παίρνουν. Χρησιμοποιούμε Round Robin DNS ως βασική ιδέα με τους διακομιστές μας . Όντας ένας πάροχος υπηρεσιών που προσφέρει μόνο τα συμπλέγματα και τις ίδιες ρυθμίσεις που φιλοξενεί υψηλή επισκεψιμότητα.

4.3.3 OPTIMIZED WEIGHTED ROUND ROBIN (OWRR)

Ο αλγόριθμος Optimized Weighted Round Robin είναι ίδιος με τον αλγόριθμο Weighted Round Robin,με την μόνη διαφορά ότι στον αλγόριθμο αυτό οι servers καθορίζουν μόνοι τους το w ανάλογα κάθε φορά με το φόρτο εργασίας.

4.4.PORT ADDRESS TRANSLATION (PAT)

Port Address Translation (PAT), είναι μια επέκταση του δικτύου για μετάφραση διευθύνσεων (NAT), που επιτρέπει πολλαπλές συσκευές σε ένα τοπικό δίκτυο (LAN) να χαρτογραφηθούν σε μια ενιαία δημόσια διεύθυνση IP. Ο στόχος της PAT είναι η διατήρηση διευθύνσεις IP .

Τα περισσότερα οικιακά δίκτυα χρησιμοποιούν PAT . Σε ένα τέτοιο σενάριο , η υπηρεσία παροχής Internet (ISP) εκχωρεί μια διεύθυνση IP στο δρομολογητή του οικιακού δικτύου . Όταν έχουμε Χ υπολογιστές σε ένα cluster στο Διαδίκτυο , ο δρομολογητής εκχωρεί στον πελάτη έναν αριθμό θύρας , το οποίο επισυνάπτεται στην εσωτερική διεύθυνση IP . Αυτό, στην πραγματικότητα, δίνει στο Computer Χ μια μοναδική διεύθυνση. Αν ο υπολογιστής Ζ συνδεθεί στο διαδίκτυο την ίδια στιγμή , ο δρομολογητής

εκχωρεί την ίδια τοπική διεύθυνση IP με διαφορετικό αριθμό θύρας . Παρά το γεγονός ότι και οι δύο υπολογιστές που μοιράζονται την ίδια δημόσια διεύθυνση IP και την πρόσβαση στο Internet ταυτόχρονα , ο δρομολογητής γνωρίζει σε ποιόν ακριβώς υπολογιστή να στείλει τα ειδικά πακέτα επειδή κάθε κάθε υπολογιστής έχει μια μοναδική εσωτερική διεύθυνση .[89]

Η τεχνική αυτή, μεταφράζει τον αριθμό πόρτας στο TCP/UDP πακέτο σε ένα προκαθορισμένο αριθμό πόρτας (port). Πιο αναλυτικά, ρυθμίζουμε την πόρτα 80 (port 80) του load balancer να είναι ταυτισμένη (bind) με την πόρτα 1000 των server μέσα στο σύμπλεγμα, έτσι κάθε πακέτο / αίτησης που φτάνει στην πόρτα 80 του load balancer προωθείται στην πόρτα 1000 κάποιου server και όχι στην αντίστοιχη πόρτα 80 όπως γίνεται συνήθως.[90]

4.5.IP TUNNELING

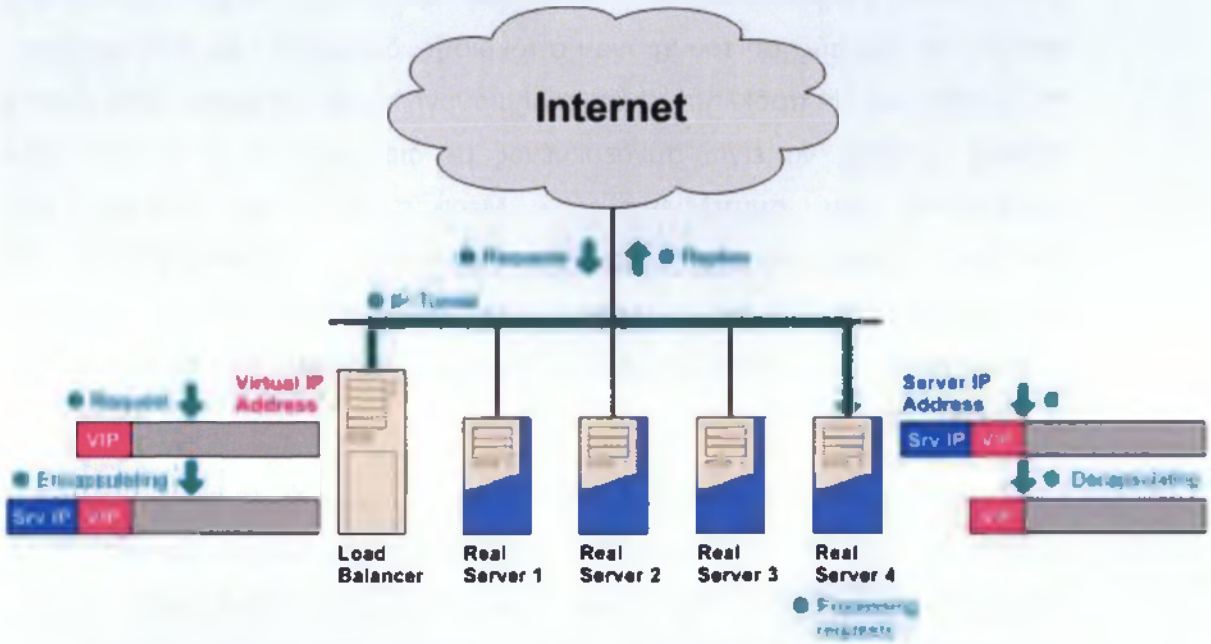
Η μέθοδος διάνοιξης σηράγγων απαιτεί από όλους τους πραγματικούς servers να έχουν διασυνδέσεις σηράγγων που έχει συσταθεί με τη διεύθυνση IP VIP. Μια διασύνδεση της εξισορρόπησης φορτίου έχει το VIP που της έχουν ανατεθεί, επίσης. Δεδομένου ότι οι διασυνδέσεις σήραγγας δεν ανταποκρίνονται στα αιτήματα ARP, η διεύθυνση MAC μιας διεπαφής της εξισορρόπησης φορτίου θα είναι στον πίνακα ARP του router που έχει συνδέθει το σύστημα με το διαδίκτυο. Αυτός είναι ο λόγος για τον οποίο τα εισερχόμενα αιτήματα φτάνουν στο load balancer, τα οποία συμπυκνώνουν σε ένα πακέτο IP με τη διεύθυνση IP προορισμού ενός από τους real servers. Υπάρχει η αίτηση decapsulated και επεξεργάζονται από το λογισμικό διακομιστή. Οι απαντήσεις μπορεί να περάσουν πίσω στον πελάτη χωρίς καμία τροποποίηση από τον εξισορροπητή φορτίου, επειδή ο πραγματικός διακομιστής συμπληρώνει το VIP ως διεύθυνση πηγής των πακέτων απόκρισης, δεδομένου ότι αποδίδεται στο περιβάλλον της σήραγγας.[90]

---

[89] Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002.

[90] http://www7.informatik.uni-erlangen.de/~ksjh/research/cluster/

Εικόνα 12: Βήματα IP tunneling



Αυτή η μέθοδος εξισορρόπησης φορτίου λειτουργεί χωρίς την επιβάρυνση σηράγγων και δίνει την υψηλότερη απόδοση των τριών αυτών μηχανισμών που περιγράφονται. Δεν υπάρχει πραγματικό πακέτο, μόνον η μετάφραση των διευθύνσεων IP σε διευθύνσεις MAC. Ένα μειονέκτημα είναι ότι όλοι οι server πρέπει να είναι στο ίδιο φυσικό δίκτυο.

4.6.APPLICATION GATEWAY SYSTEM (AGS)

Το Application Gateway System (AGS) είναι ένα κατανεμημένο σύστημα διακομιστή που αποτελείται από πολλαπλούς ετερογενείς διακομιστές που εμφανίζονται ως ένα ενιαίο σύστημα υψηλής απόδοσης στο Διαδίκτυο . Μπορεί να παρέχει υπηρεσίες stand-alone ή πρόσβαση back- end δυνατότητες , όπως βάσεις δεδομένων ή αποθήκες.[91]

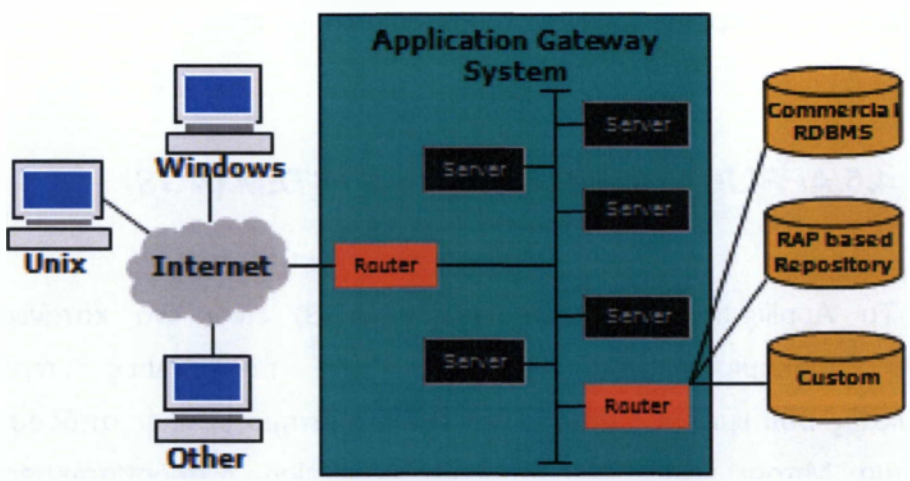
---

[91] http://www.cnri.reston.va.us/AGS/problem.html

Οι cluster των server χρησιμοποιούνται συχνά για να παρέχουν μια ενιαία λογική υπηρεσία στο Internet . Αυτός είναι ένας οικονομικά αποδοτικός τρόπος για την αύξηση του χρόνου απόκρισης, διακίνηση και διαθεσιμότητα της υπηρεσίας . Η πρόκληση είναι να δημιουργήσει την υπηρεσία, έτσι ώστε ο τελικός χρήστης να είναι συνδεδεμένος με διαφάνεια σε έναν από τους διακομιστές του συμπλέγματος . Μερικές από τις τεχνικές που χρησιμοποιούνται συνήθως που χαρτογραφούν τα μηχανήματα σε ένα σύμπλεγμα με το ίδιο όνομα υπηρεσίας περιλαμβάνουν DNS robin , πακέτο επανεγγραφής και προσαρμοσμένο λογισμικό πελάτη . Κάθε τεχνική έχει ισχυρά και αδύνατα σημεία.[92]

Το Application Gateway System (AGS) εξισορροπεί το φορτίο σε ένα σύμπλεγμα διακομιστών. Χρησιμοποιεί ένα προσαρμοστικό αλγόριθμο εξισορρόπησης φορτίου που αποφεύγει να εισάγει ένα ενιαίο σημείο αποτυχίας. Ο αλγόριθμος προσαρμόζεται γρήγορα στις αλλαγές στις ρυθμίσεις ή στο πραγματικό φορτίο των server.[92]

Εικόνα 13: Application Gateway Systems



Ο AGS load balancing αλγόριθμος απαιτεί κάποιο προσαρμοσμένο client code, αλλά είναι πολύ πιο ευέλικτο από ό,τι το custom client (π.χ., στο Netscape browsers ειδικά την περίπτωση της αρχικής σελίδας του Netscape). Σε σύγκριση με το round-robin DNS, παρέχει γρήγορη, αυτόματη προσαρμογή στις αλλαγές του φορτίο. Σε σύγκριση με της Cisco

LocalDirector, μια εμπορικά διαθέσιμη συσκευή πακέτων επανεγγραφής, η πλήρως κατανεμημένη υλοποίηση AGS δεν διαθέτει κεντρικό σημείο αποτυχίας και προσφέρει πιο άμεσες μετρήσεις του φορτίου του server.[92]

4.7.CONNECTION ALGORITHMS

4.7.1.LEAST CONNECTIONS (LC)

Ο load balancer, με χρήση του αλγόριθμου " least connections " θα κατευθύνει τα αιτήματα στον εξυπηρετητή με τον ελάχιστο αριθμό των υφιστάμενων συνδέσεων. για παράδειγμα, η επιλογή leastconns Predictor της Cisco κατευθύνει τις συνδέσεις δικτύου στο διακομιστή με τουλάχιστον τον αριθμό των ανοιχτών συνδέσεων. Δεδομένου ότι ο server με τις λιγότερες συνδέσεις θα πάρει περισσότερο νέες αιτήσεις και εκείνοι με περισσότερη αναμονή θα πάρουν λιγότερο, την πάροδο του χρόνου κάθε server θα έχει περίπου τον ίδιο αριθμό ανοικτών συνδέσεων πέραν ορισμένων κυκλοφοριακών φορτίων. Με άλλα λόγια, κάθε διακομιστής θα έχει την ίδια "ουρά μήκος ".[93]

Χρησιμοποιώντας το μοντέλο αναμονής, «το ίδιο μήκος ουράς σε κάθε server" συνεπάγεται ότι:

$$\frac{\lambda_i}{\lambda_j} = \frac{P_i}{P_j} = \frac{P_i / P}{P_j / P} \qquad \text{for} \qquad \forall i, j$$

[94]

Ως εκ τούτου, το καθαρό αποτέλεσμα του αλγόριθμου " least connections " είναι να κατευθύνει τις αιτήσεις κατά τέτοιο τρόπο ώστε η αναλογία των ποσοστών αφίξεων στους servers να είναι η ίδια με τις αξιολογήσεις των επιδόσεων των εξυπηρετητών. Με άλλα λόγια, ένας διακομιστής που είναι δύο φορές πιο ισχυρός από έναν άλλον server θα

[92] http://www.cnri.reston.va.us/AGS/problem.html

[93] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

πάρει περίπου διπλάσιες συνδέσεις ανά δευτερόλεπτο. Επιπλέον, συνεπάγεται επίσης ότι η αναλογία μεταξύ του επιτοκίου της αίτησης και της απόδοσης ενός διακομιστή είναι σταθερή για όλους τους εξυπηρετητές, δηλαδή : [94]

$$\frac{\lambda_1}{P_1} = \frac{\lambda_2}{P_2} = \cdots = \frac{\lambda_n}{P_n} = \frac{\lambda}{P}$$

Με αυτό, μπορούμε να εξάγουμε το μέσο χρόνο απόκρισης του συμπλέγματος διακομιστών από το μοντέλο αναμονής για τον " least connections " αλγόριθμο:[94]

$$r_{LC} = \frac{ns_1P_1}{P - \lambda s_1P_1}$$

Σημειώνουμε ότι το LC r είναι ανεξάρτητο από την βαθμολόγηση των επιδόσεων των servers στο σύμπλεγμα διακομιστών. Αυτό είναι ένα πολύ επιθυμητό χαρακτηριστικό του αλγόριθμου "least connections". Σε συμπλέγματα διακομιστών όπου υπάρχουν μεγάλες διαφορές στην ικανότητα των διαφόρων εξυπηρετητών, ο χρόνος απόκρισης είναι πιο συνεπής και είναι συνήθως καλύτερο από τον αλγόριθμο Round-Robin.[95]

Για ένα ομοιογενές server farm όπου υπάρχει συλλογή των servers με παρόμοιες επιδόσεις ο least connections load balancer θα στείλει τα ίδια ποσότητα των αιτήσεων σε κάθε server πάνω χρόνο. Ως εκ τούτου, έχει το ίδιο αποτέλεσμα με αυτό της Round-Robin ή του Uniform-Balancing αλγόριθμο για να εξομαλύνει την εισερχόμενη κίνηση προς τους servers.[95]

Για μια ετερογενή φάρμα διακομιστών, ο “least connections” load-balancing παρέχει έναν απλό μηχανισμό για να στείλει περισσότερες αιτήσεις σε περισσότερους ισχυρούς servers.Μολονότι δεν ελαχιστοποιείται ο συνολικός μέσος χρόνος απόκρισης του server farm έτσι εκτελείται αρκετά καλά, ειδικά όταν οι servers δεν είναι δραστικά διαφορετικοί στην απόδοση αξιολογήσεις. Προκειμένου να ελαχιστοποιηθεί η μέση απόκριση, με βάση την

[94] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

κατάσταση , η αναλογία αίτησης μεταξύ του διακομιστή i και του διακομιστή j πρέπει να ικανοποιεί αυτό.[95]

$$\frac{\lambda_i}{\lambda_j} = \frac{\frac{P_i}{P} - (prob.\_farm\_idle)\left(\frac{\sqrt{P_i}}{\sum_{k=1}^{n}\sqrt{P_k}}\right)}{\frac{P_j}{P} - (prob.\_farm\_idle)\left(\frac{\sqrt{P_j}}{\sum_{k=1}^{n}\sqrt{P_k}}\right)}$$

Που δεν είναι ακριβώς το ίδιο με τον όρο

$$\frac{\lambda_i}{\lambda_j} = \frac{P_i}{P_j} = \frac{P_i / P}{P_j / P} \qquad \text{for} \qquad \forall i, j \qquad 95$$

που χρησιμοποιείται από τον αλγόριθμο LC, εκτός, φυσικά, αν ο κάθε server έχει την ίδια βαθμολογία απόδοσης. Για την ελαχιστοποίηση του χρόνου απόκρισης στο σύμπλεγμα διακομιστών, η αναλογία αιτήσεων οποιωνδήποτε από τους δύο διακομιστές του συμπλέγματος διακομιστών πρέπει να ικανοποιούν τη συνθήκη

$$\frac{\lambda_i}{\lambda_j} = \frac{\frac{P_i}{P} - (prob.\_farm\_idle)\left(\frac{\sqrt{P_i}}{\sum_{k=1}^{n}\sqrt{P_k}}\right)}{\frac{P_j}{P} - (prob.\_farm\_idle)\left(\frac{\sqrt{P_j}}{\sum_{k=1}^{n}\sqrt{P_k}}\right)} \qquad 96$$

Εάν ικανοποιείται η συνθήκη αυτή τότε ο βέλτιστος χρόνος απόκρισης για τον least connections load balancer είναι μικρότερο από ή ίσο με :

[95] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

$$r_{LC} = \frac{ns_1P_1}{P - \lambda s_1P_1}$$

Άρα συνεπάγεται ότι:

$$r_{min} = \frac{1}{\lambda}\left[\frac{\left(\sum_{k=1}^{n}\sqrt{P_k}\right)^2}{P - \lambda s_1P_1} - n\right] \le \frac{1}{\lambda}\left[\frac{n\sum_{k=1}^{n}P_k}{P - \lambda s_1P_1} - n\right]$$

$$= \frac{ns_1P_1}{P - \lambda s_1P_1} = r_{LC} \quad 96$$

4.7.2 WEIGHTED LEAST CONNECTIONS (WLC)

Η διαφορά μεταξύ του Uniform Weighted Balancer (Σταθμισμένος Round Robin) και του Weighted Least Connections (WLC) load balancer είναι ότι η πρώτη βασίζεται σχετικά με το ποσοστό αίτησης , ενώ η τελευταία σχετικά με τον αριθμό των συνδέσεων. Τώρα καθορίζουμε τα βέλτιστα βάρη για το WLC .[97]

Ο αριθμός των συνδέσεων και ο ρυθμός αίτησης σχετίζονται . Μπορούμε να βρούμε τα βέλτιστα weights για το WLC από το βέλτιστο ποσοστό αίτησης :[97]

$$w_i(WLC) = \frac{\lambda_i s_i}{1 - \lambda_i s_i}\bigg|_{\lambda_i = optimal\_request\_rate\_to\_server\_i}$$

όπου Si είναι ο χρόνος εξυπηρέτησης του server i, αυτά τα βάρη μπορούν επίσης να εκπροσωπήσουν το ποσοστό των συνολικών συνδέσεων:[97]

[96] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

$$w_i(WLC) \leftarrow \frac{w_i(WLC)}{\sum_{k=1}^{n} w_k(WLC)}$$

4.7.3 LOCALITY-BASED LEAST-CONNECTION SCHEDULING

Διανέμει τα περισσότερα αιτήματα σε διακομιστές με τις λιγότερες ενεργές συνδέσεις σε σχέση με IPs προορισμού τους. Αυτός ο αλγόριθμος έχει σχεδιαστεί για χρήση σε ένα σύμπλεγμα διακομιστή μεσολάβησης-cache. Δρομολογεί τα πακέτα για μια διεύθυνση IP στο διακομιστή για την εν λόγω διεύθυνση, εκτός αν ο διακομιστής είναι φορτωμένος και έχει ένα διακομιστή στο μισό φορτίο του, στην περίπτωση αυτή εκχωρεί τη διεύθυνση IP στο λιγότερα φορτωμένο διακομιστή.[97]

4.7.4 LOCALITY-BASED LEAST-CONNECTION SCHEDULING REPLICATION SCHEDULING

Διανέμει τα περισσότερα αιτήματα σε διακομιστές με τις λιγότερες ενεργές συνδέσεις σε σχέση με την IP προορισμού . Αυτός ο αλγόριθμος έχει επίσης σχεδιαστεί για τη χρήση σε σύμπλεγμα διακομιστών μεσολάβησης-cache. Διαφέρει από την Locality-Based Least-Connection Scheduling με τη χαρτογράφηση της διεύθυνσης IP σε ένα υποσύνολο πραγματικών κόμβων server. Οι αιτήσεις θα οδεύουν προς το διακομιστή στο υποσύνολο με το χαμηλότερο αριθμό συνδέσεων. Εάν όλοι οι κόμβοι για την IP προορισμού είναι υπεράνω της δυναμικότητας, αναπαράγει ένα νέο διακομιστή για αυτήν τη διεύθυνση IP προορισμού με τις λιγότερο συνδέσεις από το συνολικό απόθεμα των πραγματικών servers για το υποσύνολο των πραγματικών διακομιστών για τη συγκεκριμένη IP προορισμού. Ο πιο φορτωμένος κόμβος στη συνέχεια θα πέσει από το πραγματικό υποσύνολο των server για να αποφευχθεί η αντιγραφή.[98]

[97] http://www.centos.org/docs/5/html/Virtual_Server_Administration/s2-lvs-sched-VSA.html

4.7.5.MAXIMUM CONNECTIONS (MC)

Ο αλγόριθμος Maximum Connections load balancer θέτει ένα περιορισμό σχετικά με τις μέγιστες συνδέσεις που θα δεχτεί ένας διακομιστής. Οι servers θα μπορούσαν να αποφύγουν την υπέρβαση από τα κατώτατα όρια των επιδόσεων τους, χρησιμοποιώντας αυτό το μηχανισμό.

Αν Li είναι το μέγιστο όριο των συνδέσεων στο διακομιστή i. Μπορούμε να υποδείξουμε το MC load balancer ως ένα σύστημα εξυπηρέτησης με ελάχιστο χώρο αναμονής και να υπολογίσουμε τη μέση απόκριση του χρόνου για την εξυπηρέτηση κάθε αιτήματος του i διακομιστή, $r_i(L_i)$. Ως ειδική περίπτωση, όταν ο server μπορεί να λάβει μόνο μία αίτηση σε ένα χρόνο, δηλαδή, Li = 1, ο χρόνος απόκρισης είναι ίσος με την υπηρεσία χρόνου ($r_i(L_i)=s_i$).[98]

4.8.DESTINATION HASHING SCHEDULING

Διανέμει τις αιτήσεις στο σύμπλεγμα των πραγματικών servers αναζητώντας την IP προορισμού σε έναν στατικό πίνακα κατακερματισμού. Αυτός ο αλγόριθμος έχει σχεδιαστεί για την χρήση σε ένα σύμπλεγμα διακομιστή proxy-cache.[99]

4.9.SOURCE HASHING SCHEDULING

Διανέμει τις αιτήσεις στο σύμπλεγμα των πραγματικών servers αναζητώντας την source IP σε έναν στατικό πίνακα κατακερματισμού. Αυτός ο αλγόριθμος έχει σχεδιαστεί για LVS δρομολογητές με πολλαπλά τείχη προστασίας.[100]

---

[98] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

[99] http://www.centos.org/docs/5/html/Virtual_Server_Administration/s2-lvs-sched-VSA.html

4.10.SHORTEST EXPECTED DELAY SCHEDULING

Ο shortest expected delay scheduling αλγόριθμος εκχωρεί συνδέσεις δικτύου στο διακομιστή με τη μικρότερη αναμενόμενη καθυστέρηση. Η αναμενόμενη καθυστέρηση θα είναι (Ci + 1) / Ui αν αποστέλλονται στο διακομιστή i, όπου Ci είναι ο αριθμός των συνδέσεων στο διακομιστή i και Ui είναι το σταθερό επιτόκιο των υπηρεσιών (κατά βάρος) του διακομιστή i.[100]

4.11.NEVER QUEUE SCHEDULING

Ο never queue scheduling αλγόριθμος υιοθετεί ένα μοντέλο δύο ταχυτήτων. Όταν υπάρχει ένας αδρανής server διαθέσιμος, η εργασία θα σταλεί στον αδρανή διακομιστή, αντί να περιμένουν για έναν γρήγορο server. Όταν δεν υπάρχει αδρανής διακομιστής διαθέσιμος, η εργασία θα πρέπει να σταλλεί στον διακομιστή που έχει τον μικρότερο χρόνο απόκρισης.[101]

4.12.SERVER RESPONSE TIME (SRT)

Ο "Server Response Time" load balancer επικεντρώνεται σε ένα βασικό δείκτη απόδοσης που σχετίζεται με το SLO : χρόνος απόκρισης . Κατευθύνει τις αιτήσεις στο διακομιστή που έχει τον γρηγορότερο χρόνο απόκρισης . Διακομιστές Web φαίνεται να ανταποκρίνονται κατηγορηματικά σε ένα σημείο , και στη συνέχεια σε ένα ορισμένο σημείο υπάρχει απότομη και δραματική αύξηση του χρόνου απόκρισης. Σε αυτές τις καταστάσεις , ο load balancer θα έχει την τάση να υπερφορτώσει τον συγκεκριμένο διακομιστή που

[100] Job Scheduling Algorithms in Linux Virtual Server

[101] Job Scheduling Algorithms in Linux Virtual Server

τον θεωρεί ως τον πιο γρήγορο πριν από τη μετάβαση σε άλλον. Ας χρησιμοποιήσουμε ένα απλό μοντέλο αναμονής να εξηγήσουμε το γιατί.[102]

Base Server Utilization (%)	Response Time Increase
10	1.1
20	2.6
30	4.5
40	7.1
50	11.1
60	17.6
70	30.4
80	66.7
90	900

Πίνακας 1: Server Response Time SRT

Ας υποθέσουμε ότι " ο ταχύτερος server" είναι u * 100% απασχολημένος . Όταν αυξάνεται η βασική χρήση 10 % από u σε 1.1u , ο χρόνος απόκρισης θα αυξηθεί περίπου 10u / (1-1.1u)%.[103]

Για παράδειγμα αν η βασική χρήση είναι στο 20% τότε η αύξηση 10% στην χρήση του διακομιστή θα αυξήσει 2.6% τον χρόνο απόκρισης του server. Από την άλλη πλευρά, εάν η χρήση του διακομιστή είναι ήδη στο 90%, τότε η αύξηση 10% θα αυξήσει τον χρόνο απόκρισης του server κατά 900%.Ο πίνακας παρακάτω δείχνει μερικά παραδείγματα.[103]

[102] Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software

4.13.LOWEST CPU UTILIZATION ALGORITHM

Ο Lowest cpu utilization αλγόριθμος στέλνει τις αιτήσεις στον διακομιστή με την μικρότερη χρήση της Κεντρικής Μονάδας Επεξεργασίας (CPU). Ο αλγόριθμος αυτός είναι κατάλληλος για αιτήσεις με δυναμικό περιεχόμενο.

4.14.SOURCE IP ADDRESS ALGORITHM

Ο source ip address αλγόριθμος στέλνει τις αιτήσεις σύμφωνα με την IP προορισμού, είναι απλός αλγόριθμος και δουλεύει πολύ καλά σε συμπλέγματα μικρών εταιρικών δικτύων.

4.15.RESPONSE TIME ALGORITHM

Ο αλγόριθμος Χρόνος Απόκρισης χρησιμοποιεί μια λίστα με τους χρόνους απόκρισης των servers προκειμένου να καθοριστεί ποιος server έχει το λιγότερο φορτίο . Για κάθε διεργασία διακομιστή στο σύμπλεγμα εξισορρόπησης φορτίου , ο load balancing spawner διατηρεί μια ταξινομημένη λίστα των servers και τον μέσο χρόνο απόκρισης τους . Κάθε φορά που ο spawner λαμβάνει μια αίτηση πελάτη, ανακατευθύνει τον πελάτη με τη διαδικασία του διακομιστή στην κορυφή της λίστας. Ο αποθέτης ενημερώνει τους χρόνους απόκρισης των servers περιοδικά . Μπορούμε να καθορίσουμε τη συχνότητα ενημέρωσης για το χρόνο απόκρισης (χρόνος απόκρισης ανανέωσης) στα με τα δεδομένα για το σύμπλεγμα εξισορρόπησης φόρτου. Ο αλγόριθμος Χρόνος Απόκρισης υποστηρίζει μόνο Stored Process

Servers.[103]

Ο αλγόριθμος χρόνος απόκρισης χρησιμοποιεί τις ακόλουθες παραμέτρους:[104]

-Respone refresh rate: καθορίζει τη διάρκεια της περιόδου σε χιλιοστά του δευτερολέπτου που to load balancing spawner θα χρησιμοποιήσει τους τρέχοντες χρόνους απόκρισης . Στο τέλος αυτής της περιόδου ο αλγόριθμος ενημερώνει τους χρόνους απόκρισης για όλους τους διακομιστές του συμπλέγματος και στη συνέχεια επαναταξινομεί τη λίστα των servers .

-Maximum Clients: καθορίζει τον μέγιστο αριθμό των πελατών που ένας διακομιστής μπορεί να έχει . Όταν ένας διακομιστής φθάσει το μέγιστο αριθμό των πελατών του , ο αλγόριθμος δεν θα ανακατευθύνει κανένα αίτημα στο διακομιστή μέχρι κάποιος client να αποσυνδέεται .[104]

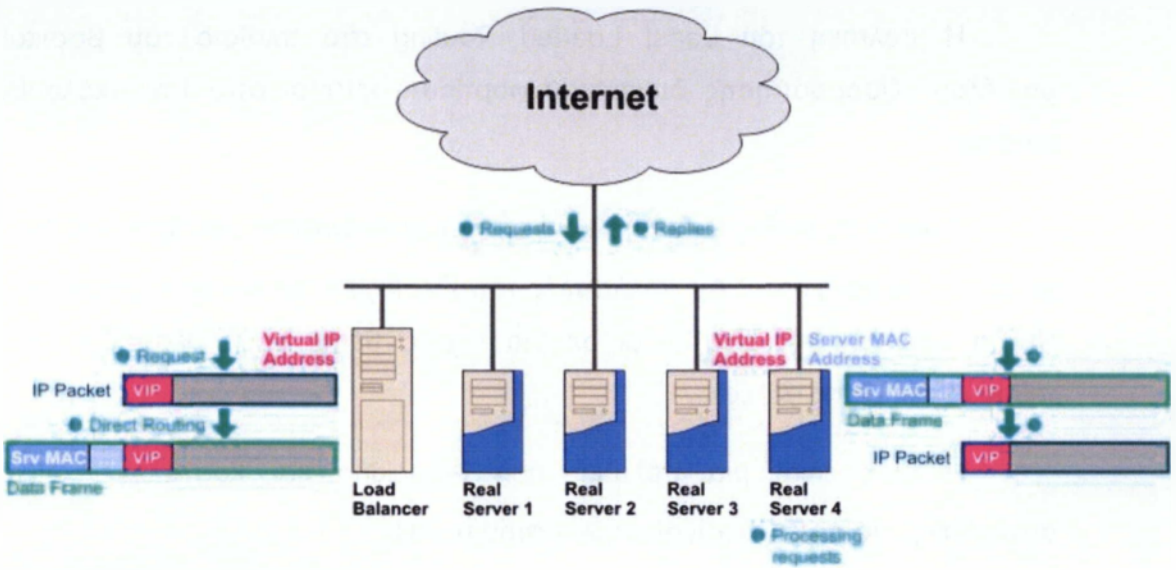
4.16.DIRECT ROUTING (OR DIRECT PATH ROUTING)

Η VIP έχει εκχωρηθεί σε ένα interface του load balancer και σε συσκευές alias για τις συσκευές δικτύου των πραγματικών servers. Οι διασυνδέσεις alias δεν πρέπει να απαντήσουν στα αιτήματα ARP. Για το σκοπό αυτό, ένα ειδικό patch πρέπει να εφαρμοστεί για τους σύγχρονους πυρήνες Linux. Δεδομένου ότι η διεύθυνση MAC interface του load balancer είναι ο ένας πίνακας ARP εισόδου για το VIP στο δρομολογητή διαδικτύου, τα πακέτα αιτήσεων φθάνουν στον load balancer πρώτα. Ο load balancer δεν έχει ούτε να ξαναγράψει ούτε να ενσωματώσει το πακέτο. Περνάει την απάντηση στο πραγματικό server που καθορίζεται από έναν αλγόριθμο κάνοντας την διεύθυνση MAC αυτού του πραγματικού διακομιστή ως διεύθυνση αποστολής. Όταν το πακέτο φτάνει στο πραγματικό server, γίνετε αποδεκτή επειδή το alias φέρει την VIP. Οι απαντήσεις που δημιουργούνται από το λογισμικό διακομιστή μπορεί να αποστέλλονται στον πελάτη,

[103] http://support.sas.com/ Understanding the Load-Balancing Algorithms

χρησιμοποιώντας το VIP ως διεύθυνση πηγής, χωρίς να χρειάζεται να περάσουν από τον load balancer για δεύτερη φορά.[104]

Εικόνα 14: Direct Routing



Αυτή η μέθοδος εξισορρόπησης φορτίου λειτουργεί χωρίς την επιβάρυνση tunneling και δίνει την υψηλότερη απόδοση των τριών αυτών μηχανισμών που περιγράφονται. Δεν υπάρχει πακέτο επανεγράψιμο, μόνον η μετάφραση των διευθύνσεων IP σε διευθύνσεις MAC δυναμικά. Ένα μειονέκτημα είναι ότι όλα τα μηχανήματα πρέπει να είναι στο ίδιο τμήμα φυσικού δικτύου.[105]

4.17.LAST VISITED ROUTING

Ο last visited routing αλγόριθμος δρομολογεί τις αιτήσεις του συμπλέγματος στο διακομιστή που εξυπηρέτησε την αρχική αίτηση του συμπλέγματος. Για την επιλογή του διακομιστή αυτού υπάρχουν πολλοί αλγόριθμοι.

[104] http://www7.informatik.uni-erlangen.de/~ksjh/research/cluster/

4.18.LEAST LOADED ROUTING

Η πολιτική του Least Loaded Routing στο πλαίσιο του βασικού μοντέλου εξισορρόπησης δυναμικού φορτίου ορίζεται από τον ακόλουθο κανόνα:

Όταν ένας τύπος u καταναλωτή φτάνει, ανατίθεται σε μια θέση $v \in N(u)$ με τον ελάχιστο φόρτο. Αν πολλαπλές τοποθεσίες επιτελούν το ελάχιστο στο N(u),ο καταναλωτής ανατίθεται σε μία τυχαία θέση, κάθε θέση έχει ίσες πιθανότητες με τις άλλες.

Η LLR είναι μια πολιτική non repacking και κοστολογεί $|N(u)|$ συγκρίσεις ανά άφιξη καταναλωτών τύπου $u \in U$.

Ένα άλλο χαρακτηριστικό του LLR είναι ότι μπορεί να εφαρμοστεί σε ένα κατανεμημένο τρόπο χρησιμοποιώντας ένα ανεξάρτητο παράγοντα εκχώρησης ανά τύπο καταναλωτή. Κάθε άφιξη μπορεί να ανατεθεί σε μια θέση με βάση τις πληροφορίες σχετικά με την κατάσταση του δικτύου. Επιπλέον ο LLR είναι ένας ισχυρός αλγόριθμος με τη ζήτηση αιτημάτων στο δίκτυο.[105]

4.19.PORT-BOUND SERVERS

Όταν ορίζουμε έναν εικονικό διακομιστή, πρέπει να καθορίζουμε το πρωτόκολλο TCP και UDP που χειρίζεται αυτόν τον εικονικό διακομιστή. Ωστόσο, εάν ρυθμίσουμε το NAT του συμπλέγματος διακομιστών, μπορούμε επίσης να ρυθμίσουμε το port-bound server.Το Port-bound server επιτρέπει μια εικονική διεύθυνση IP του διακομιστή να αντιπροσωπεύει ένα σύνολο πραγματικών εξυπηρετητών για μία υπηρεσία, όπως HTTP, και ένα

[105] Analysis of Simple Algorithms for Dynamic Load Balancing Murat Alanyali and Bruce Hajek

διαφορετικό σύνολο πραγματικών εξυπηρετητών για μια άλλη υπηρεσία, όπως το Telnet.[106]

4.20.CLIENT-ASSIGNED LOAD BALANCING

Ο αλγόριθμος Client-assigned load balancing επιτρέπει να περιορίζουμε την πρόσβαση σε έναν εικονικό διακομιστή, καθορίζοντας τη λίστα των IP που επιτρέπεται να χρησιμοποιούν τον εικονικό διακομιστή. Με αυτήν τη δυνατότητα, μπορούμε να ορίσουμε ένα σύνολο υποδικτύων IP του client (όπως εσωτερικά υποδίκτυα) που συνδέονται σε μια εικονική διεύθυνση IP σε ένα σύμπλεγμα διακομιστών, και να ορίσουμε ένα άλλο σύνολο πελατών (όπως εξωτερικούς πελάτες) σε ένα διαφορετικό σύμπλεγμα διακομιστών που να μην έχει την δυνατότητα σύνδεσης.[107]

4.21.STICKY CONNECTIONS

Ο sticky connections αλγόριθμος , καταγράφει τις νέες αιτήσεις των client από ποιόν server εξυπηρετήθηκαν, αν σε περίπτωση ξανά λάβει μια αίτηση από τον ίδιο client τότε την αίτηση την ξανά στέλνει στον server που την είχε εξυπηρετήσει στην αρχή. Αυτό γίνεται ως εξής, ο load balancer δημιουργεί sticky objects για να παρακολουθεί τις αναθέσεις των πελατών. Αυτά τα sticky objects παραμένουν στην βάση δεδομένων, μετά από την τελευταία σύνδεση διαγράφονται για μια περίοδο που ορίζεται από ένα διαμορφώσιμο sticky timer. Εάν ο χρονοδιακόπτης έχει ρυθμιστεί σε έναν εικονικό διακομιστή , οι νέες συνδέσεις από έναν πελάτη αποστέλλονται στον ίδιο πραγματικό server που χειρίστηκε την προηγούμενη σύνδεση πελάτη , υπό την προϋπόθεση μία από τις ακόλουθες συνθήκες να είναι αληθής:

[106] http://www.cisco.com/ Configuring Server Load Balancing

[107] http://www.cisco.com/ Configuring Server Load Balancing

• Μια σύνδεση για τον ίδιο πελάτη να υπάρχει ήδη .

• Το χρονικό διάστημα μεταξύ του τέλους της προηγούμενης σύνδεσης από τον πελάτη και την έναρξη της νέας σύνδεσης να είναι εντός της διάρκειας του χρονοδιακόπτη.

Οι Sticky συνδέσεις επιτρέπουν επίσης τη σύζευξη των υπηρεσιών που διακινούνται από περισσότερους από έναν virtual servers . Αυτό επιτρέπει την υποβολή αιτήσεων σύνδεσης για τις σχετικές υπηρεσίες να χρησιμοποιούν το ίδιο πραγματικό server . Για παράδειγμα , ο διακομιστής Web (HTTP) συνήθως χρησιμοποιεί τη θύρα TCP 80 , και το HTTP Secure Socket Layer (HTTPS) χρησιμοποιεί τη θύρα 443 . Εάν το HTTP και το HTTPS είναι σε συνδυασμό , οι συνδέσεις για τις θύρες 80 και 443 από την ίδια διεύθυνση IP υποδικτύου πελάτη μπορούν να ανατεθούν στον ίδιο πραγματικό διακομιστή.[108]

4.22.DELAYED REMOVAL OF TCP CONNECTION CONTEXT

Λόγω των ανωμαλιών των πακέτων IP στην παραγγελία, ο load balancer θα μπορούσε να "δει" το κλείσιμο μιας σύνδεσης TCP .Αυτό το ζήτημα παρουσιάζεται συνήθως όταν υπάρχουν πολλαπλές διαδρομές που τα πακέτα σύνδεσης TCP μπορούν να ακολουθήσουν. Για να ανακατευθύνει σωστά τα πακέτα που φτάνουν μετά, τερματίζει την σύνδεση. Ο load balancer διατηρεί τις πληροφορίες σύνδεσης TCP, ή το πλαίσιο, για ένα καθορισμένο χρονικό διάστημα. Το χρονικό διάστημα που το κείμενο διατηρείται μετά την αποσύνδεση ελέγχεται από ένα ρυθμιζόμενο χρονοδιακόπτη καθυστέρησης.[109]

[108] http://www.cisco.com/ Configuring Server Load Balancing

4.23.RANDOM

Ο αλγόριθμος αυτός είναι πολύ εύκολος και κατανοητός, δουλεύει ως εξής:

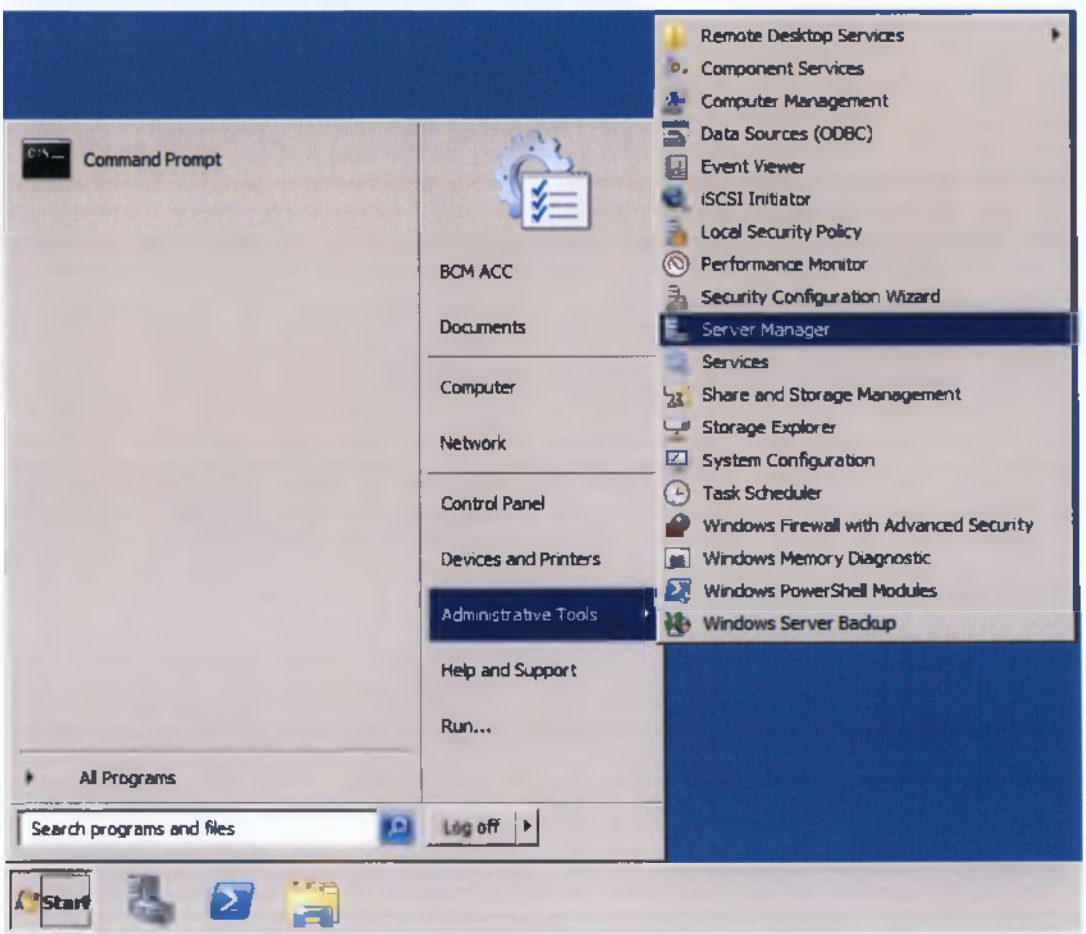
Έστω οτι έχουμε 5 διακομιστές να εξυπηρετήσουν ένα αίτημα, αυτό το αίτημα έχει τις ίδιες πιθανότητες να εξυπηρετηθεί από τους servers για τον λόγο ότι ο server που θα εξυπηρετήσει το αίτημα επιλέγετε τυχαία.

ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ

Εγκατάσταση λειτουργίας Failover

Ρυθμίσεις στο Server A

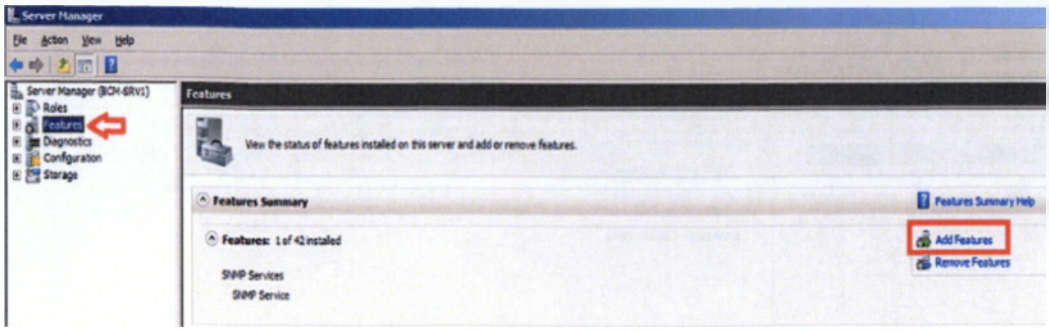
Για να εγκαταστήσουμε τη λειτουργία Failover, ανοίγουμε τον Server Manager, και επιλέγουμε **Start > Administrative Tools > Server Manager**[109]



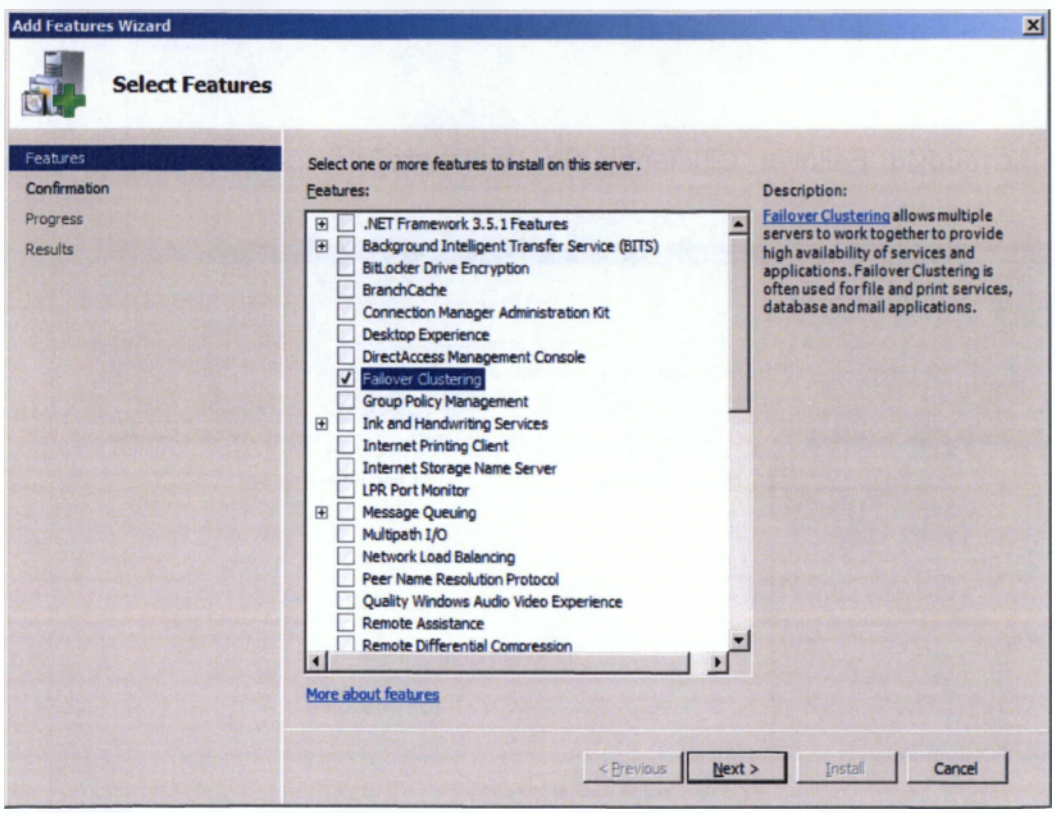
109

http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx

Ανοίγουμε το **Features**, και έπειτα επιλέγουμε **Add Feature**.



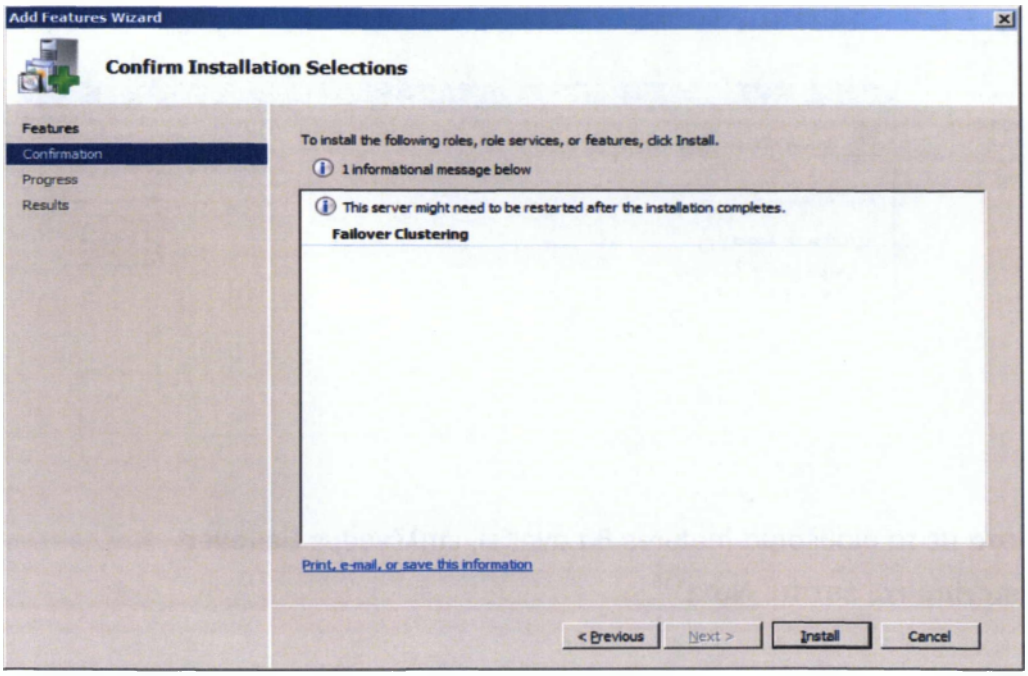
Η λίστα με τα διαθέσιμα features θα ανοίξει, επιλέγουμε **Failover Clustering** και έπειτα **Next**[110]



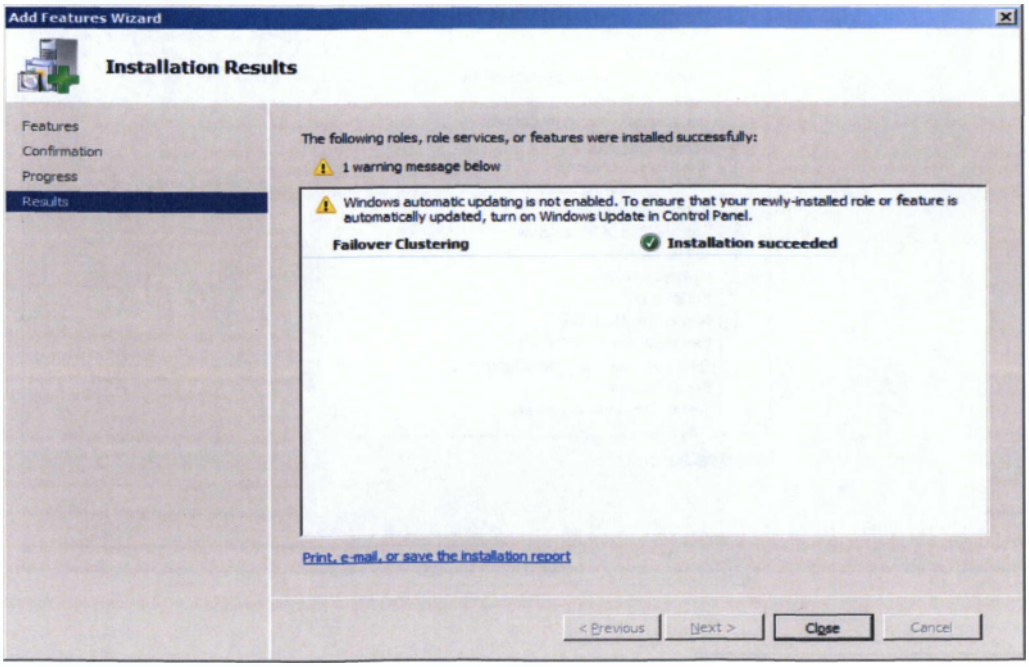
110

http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx

Επιλέγουμε Install

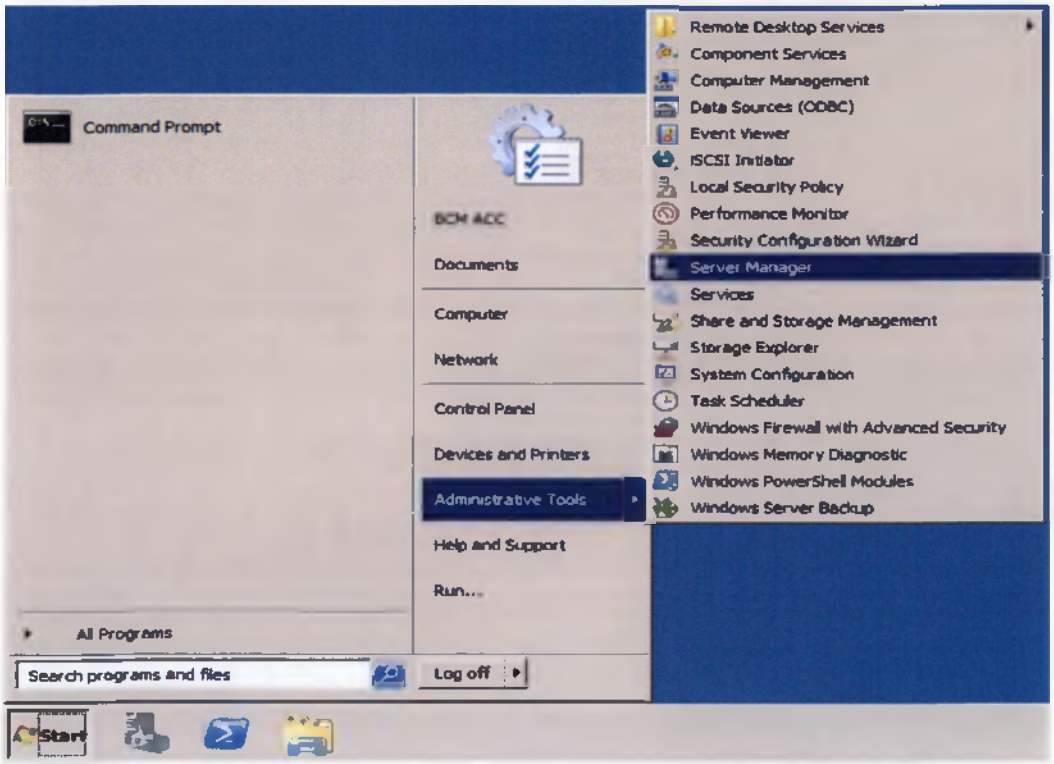


Η λειτουργία Failover Clustering θα εγκατασταθεί, και επιλέγουμε Close

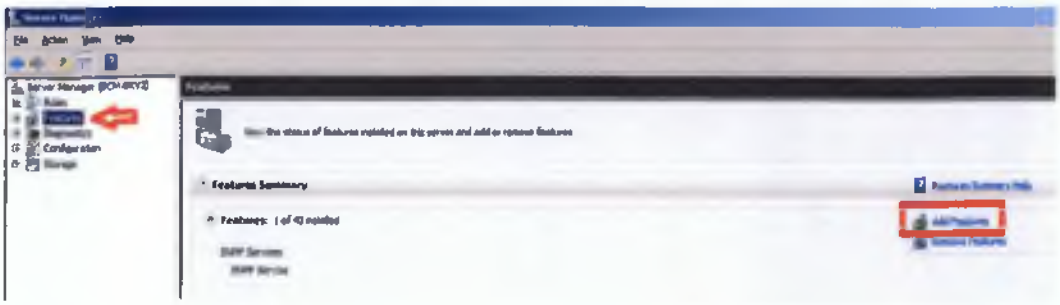


Ρυθμίσεις στο Server B

Παρομοίως στο Server B, πρέπει να εγκαταστήσουμε τη λειτουργία Failover, οπότε επιλέγουμε **Start > All Programs > Administrative Tools > Server Manager**[111]



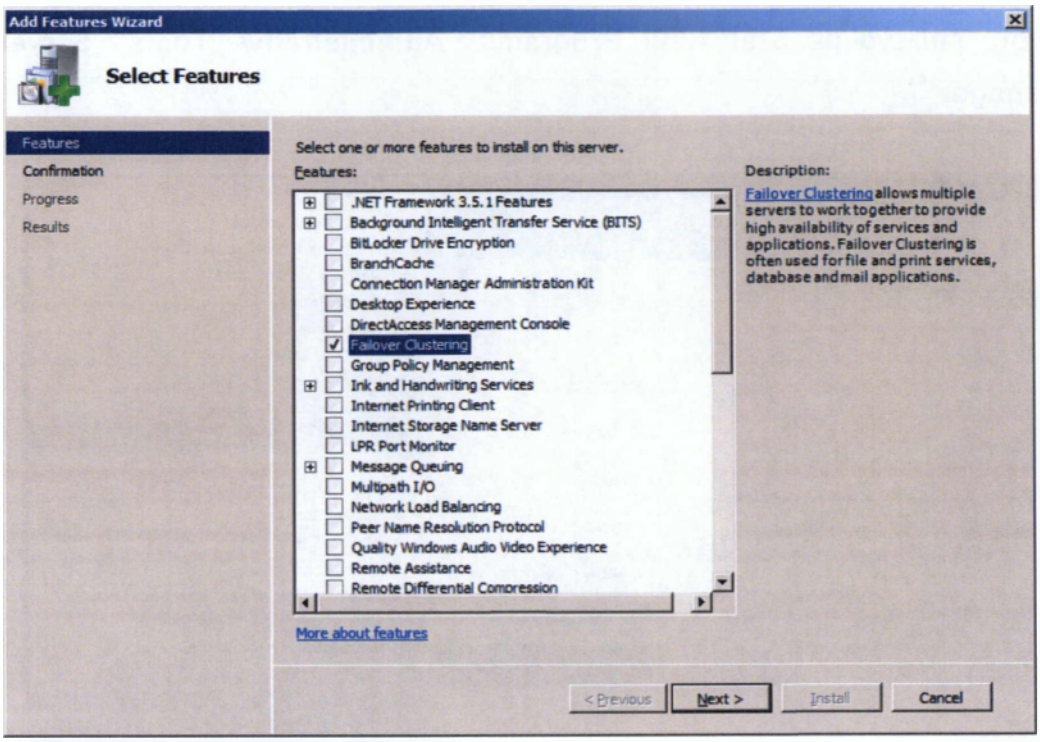
Ανοίγουμε το **Features**, και έπειτα επιλέγουμε **Add Feature**.



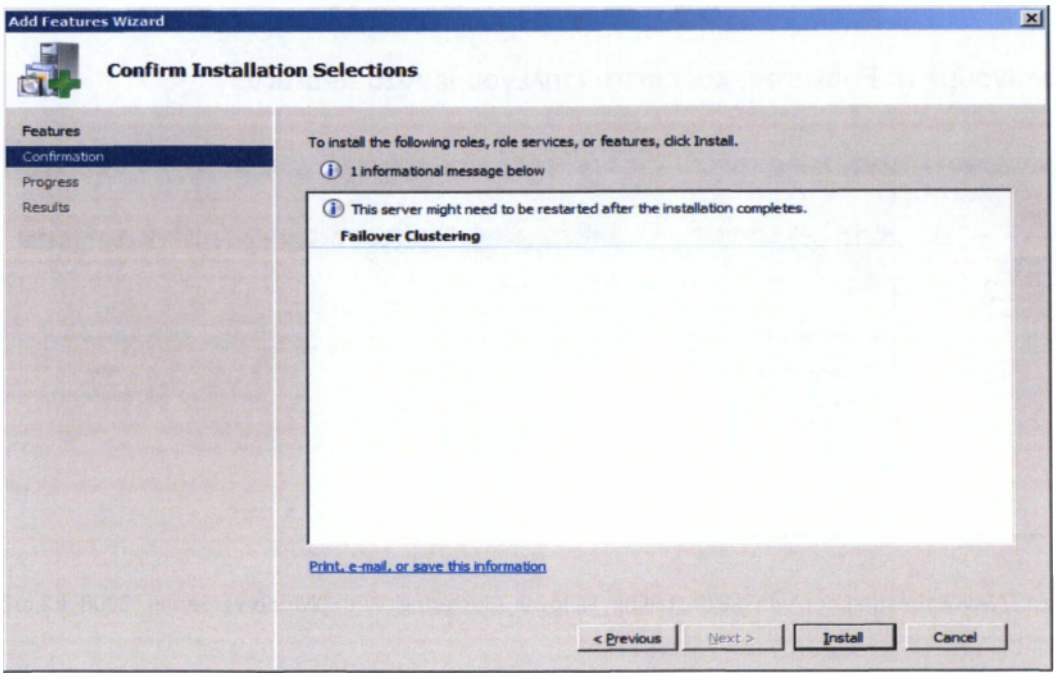
111

http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.asp
x

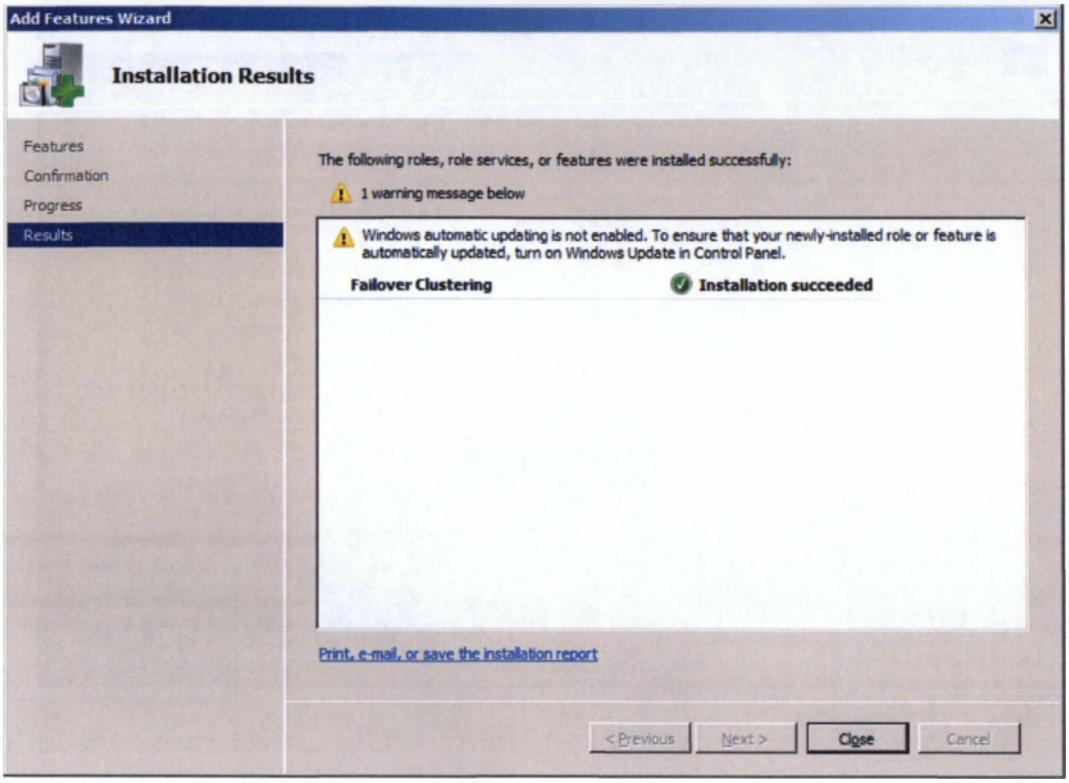
Η λίστα με τα διαθέσιμα features θα ανοίξει, επιλέγουμε **Failover Clustering** και επιλέγουμε **Next**



Επιλέγουμε **Install**



Η λειτουργία Failover Clustering θα εγκατασταθεί, και επιλέγουμε Close

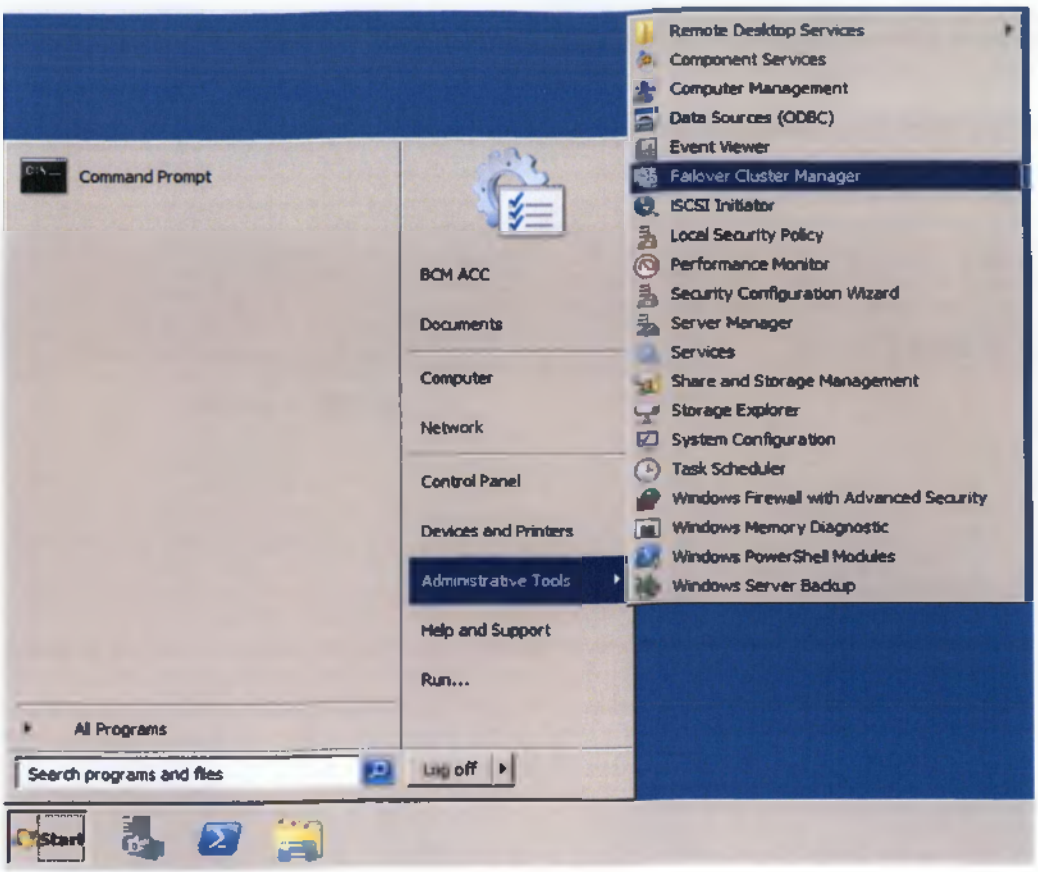


Τώρα που έχουμε εγκαταστήσει και στους δύο server τη λειτουργία Failover Clustering, μπορούμε να δημιουργήσουμε ένα cluster σε ένα από αυτούς τους server και να "εισάγουμε" τον άλλο στον ίδιο cluster. Τώρα, πρέπει να ρυθμίσουμε το όνομα του cluster, την IP και τους κόμβους.

Για να ανοίξουμε τη λειτουργία Failover Clustering, επιλέγουμε **Start > Administrative Tools > Failover Cluster Manager**[112]

112

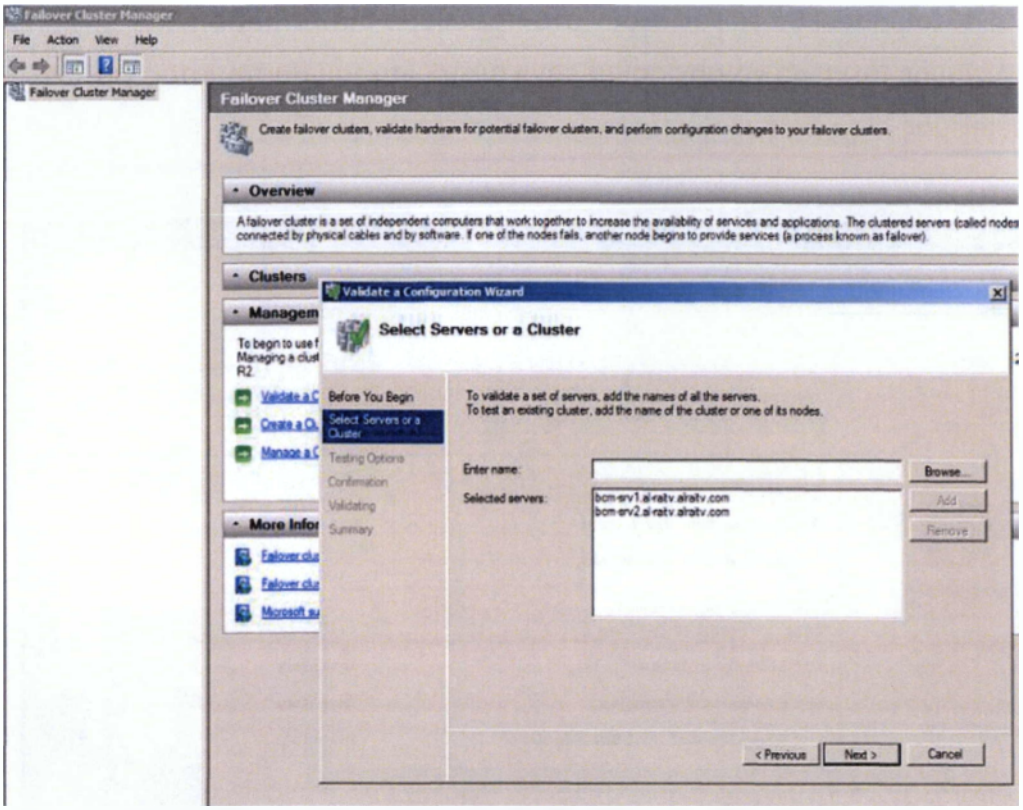
http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx



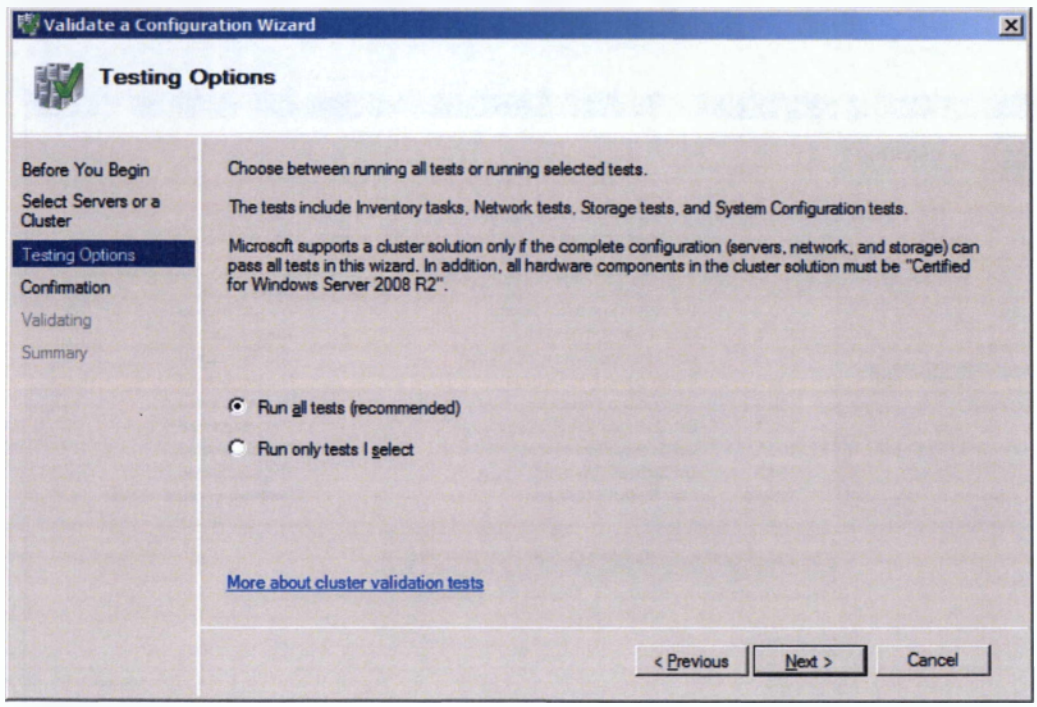
1. Το πρώτο βήμα είναι για να δημιουργήσουμε ένα failover clustering, είναι με την επικύρωση του συστήματος. Αυτό θα γίνει επιλέγοντας **Validate a Configuration**

Με την επιλογή **Validate a Configuration**, θα χρειαστεί να προσθέσουμε Cluster nodes, αυτοί είναι οι servers που θα περιλαμβάνει το cluster μας, και έπειτα επιλέγουμε **Next**[113]

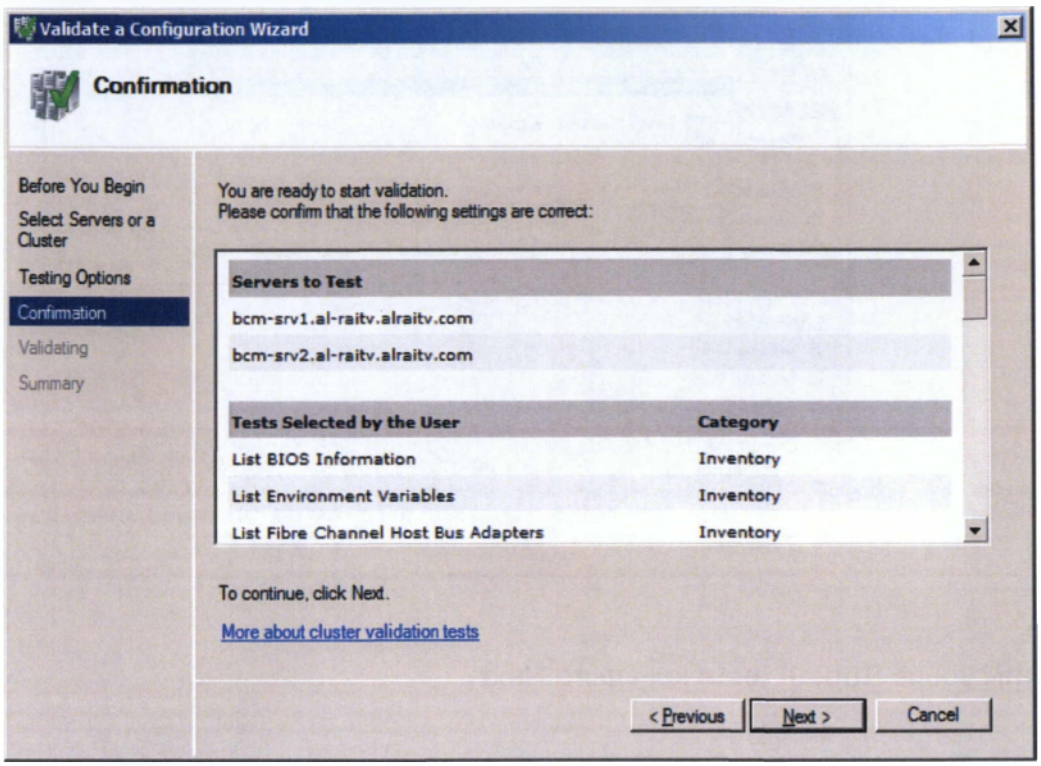
[113] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

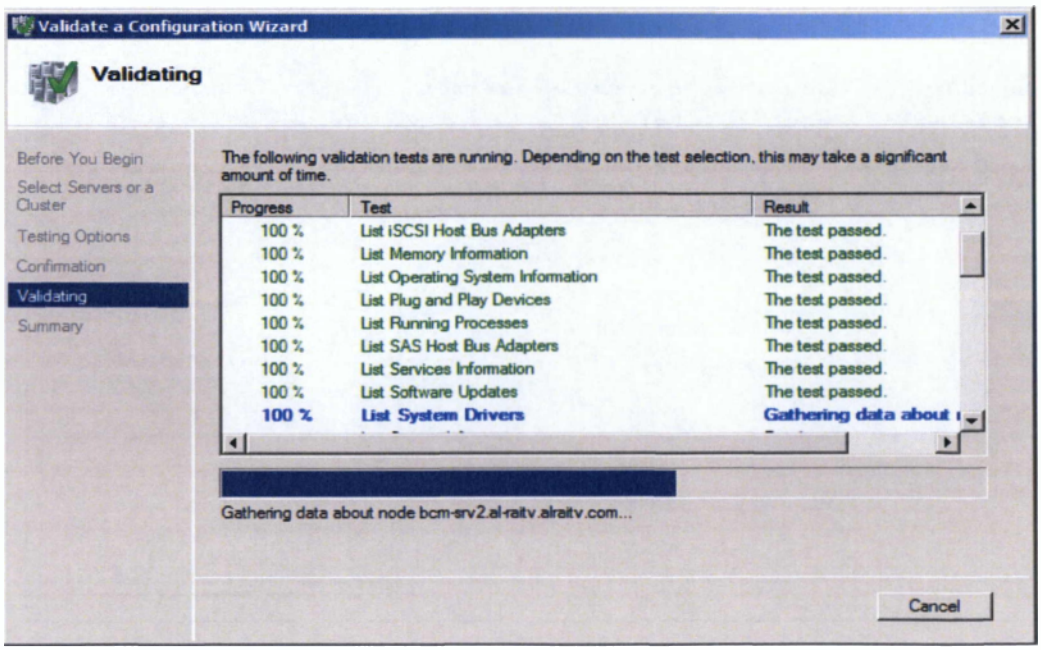


Επιλέγουμε **Run all tests** και έπειτα **Next**

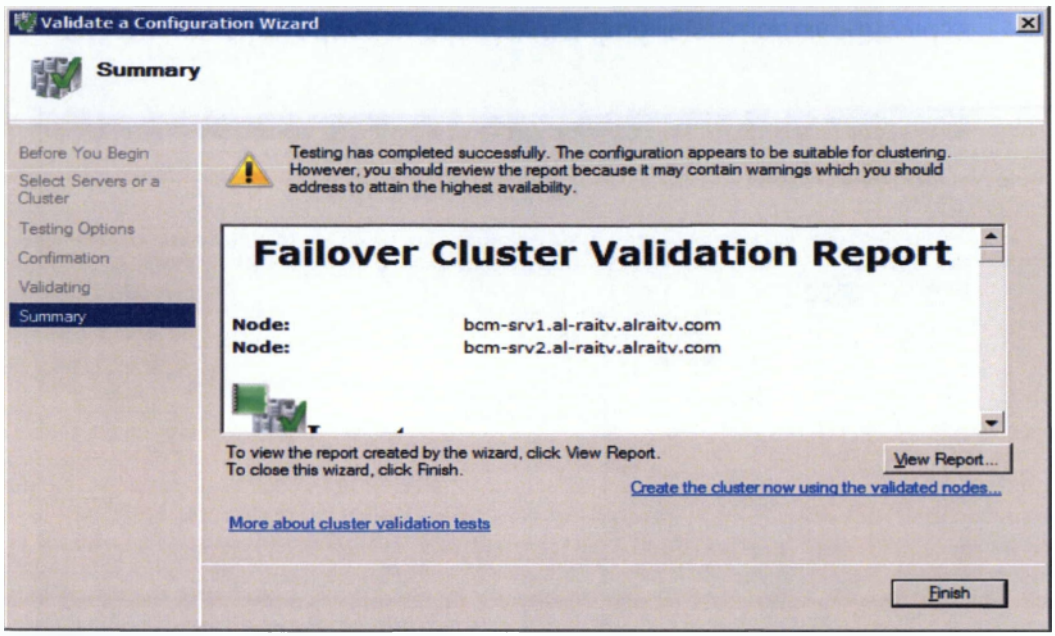


Τα διαθέσιμα test θα εμφανιστούν στο παράθυρο **confirmation**, και επιλέγουμε **Next** για να αρχίσουμε την επικύρωση του συστήματος για τον cluster **μας**

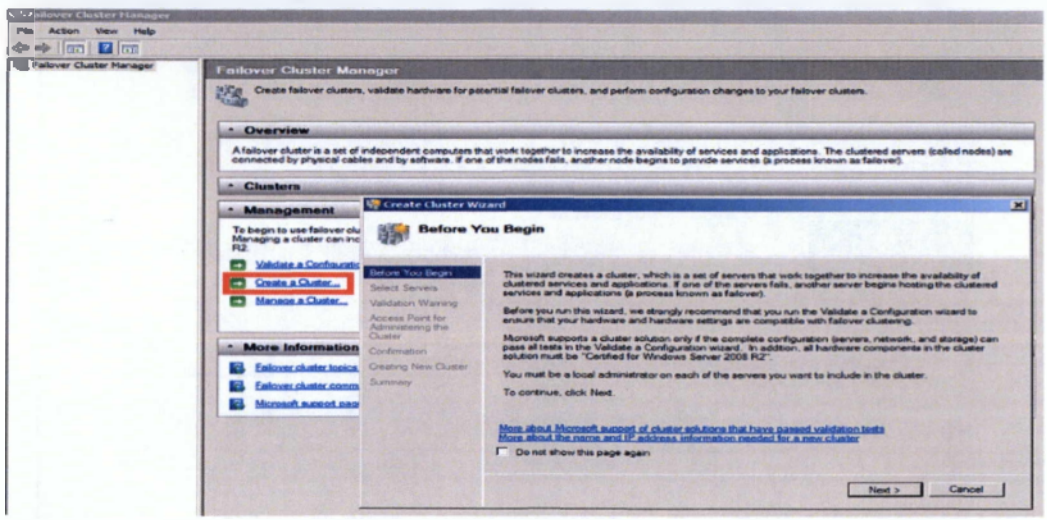




Τέλος επιλέγουμε Finish



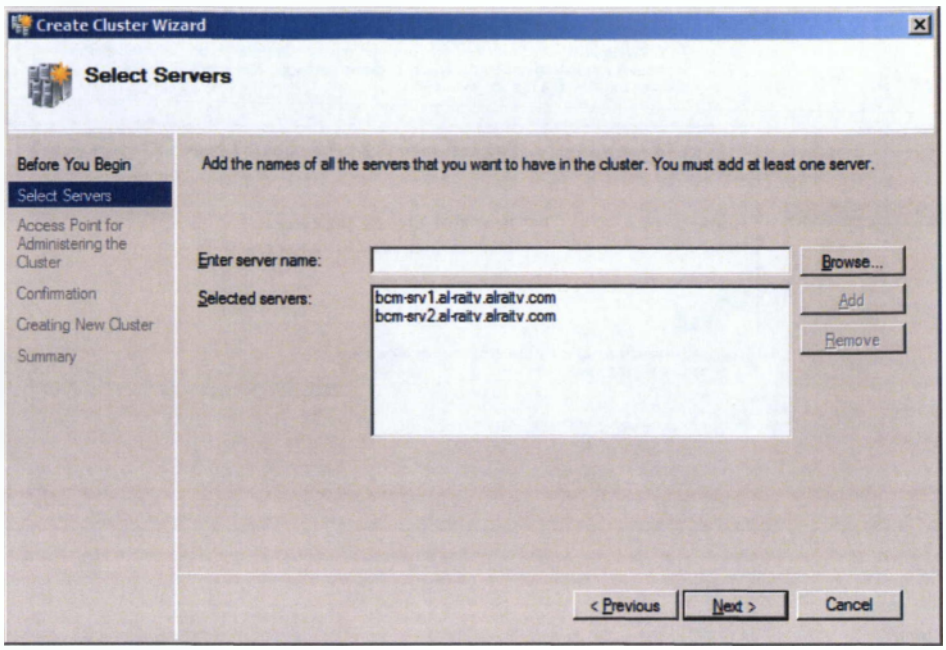
2. Τώρα που η διαμόρφωση του συστήματος μας έχει επικυρωθεί μπορούμε να κάνουμε τις απαραίτητες ρυθμίσεις στο cluster μας. Επιλέγουμε τη δεύτερη επιλογή, **Create a Cluster**, και έπειτα **Next**[114]



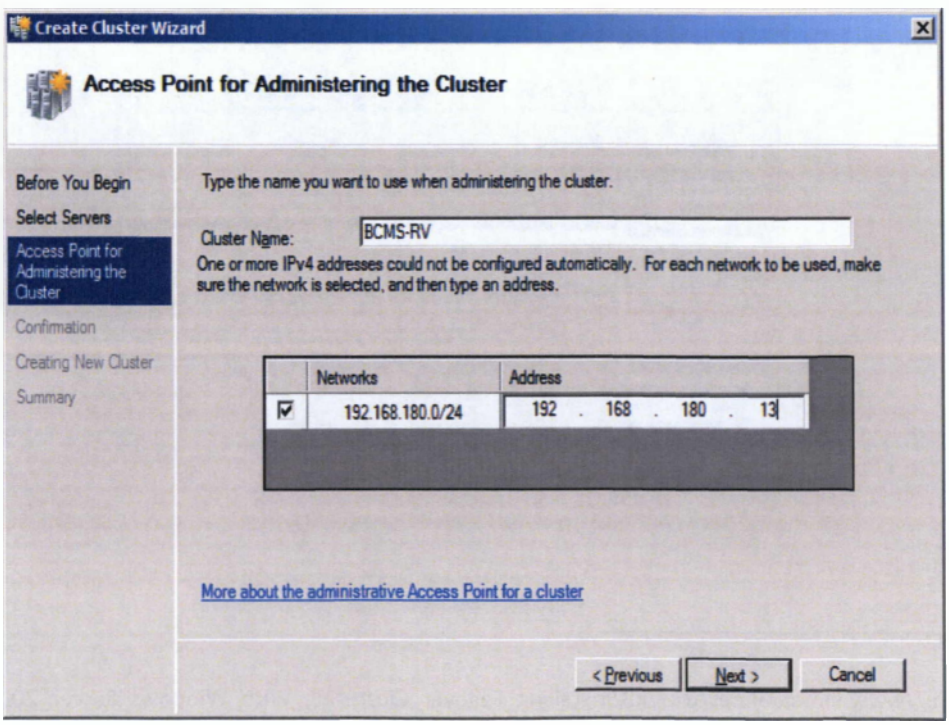
114

http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx

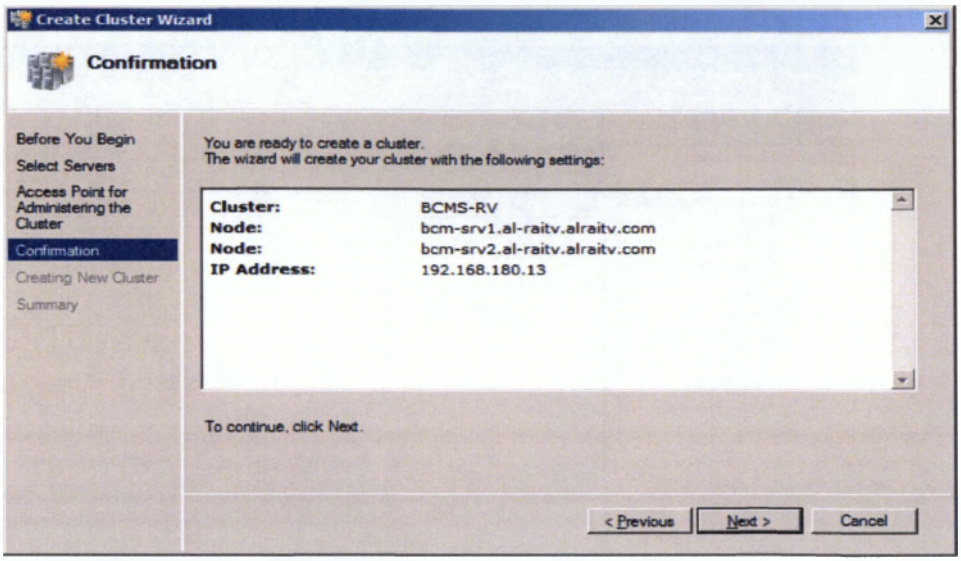
3. Τώρα πρέπει να εισάγουμε τα ονόματα των server που θέλουμε να περιλαμβάνει το cluster μας. Επιλέγουμε τα ονόματα και έπειτα **Next**



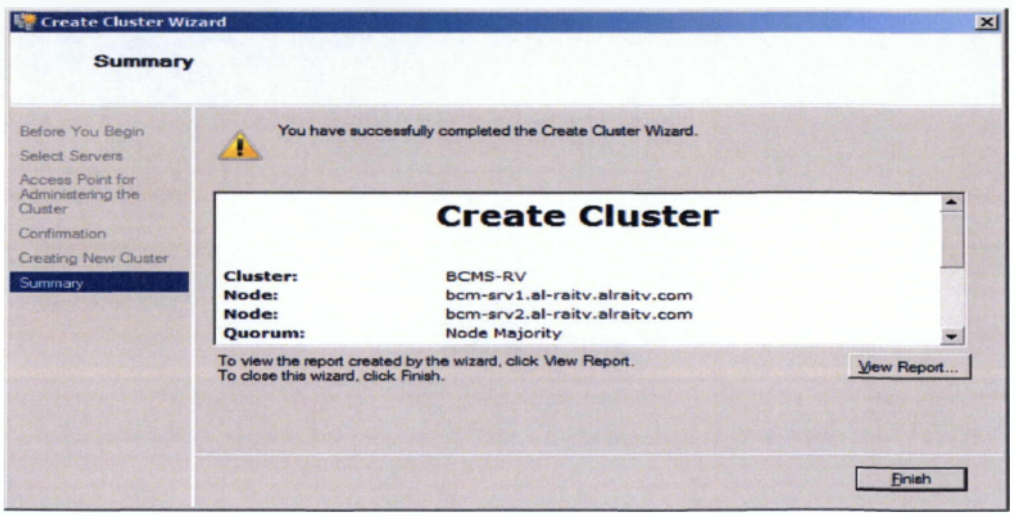
4. Εφόσον έχουμε επιλέξει τους server, πρέπει να δώσουμε ένα όνομα και μια IP στον cluster μας



5. Στο παράθυρο Confirmation, θα πρέπει να εμφανίζονται το όνομα του cluster, η IP διεύθυνση του και τα ονόματα των server. Εφόσον είναι όλα σωστά, επιλέγουμε **Next.**[115]



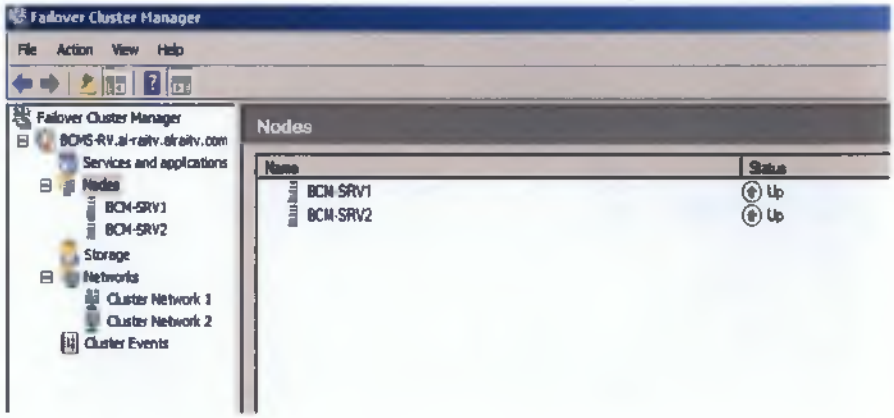
6. Στο παράθυρο summary εμφανίζεται η αναφορά ότι ρυθμίσαμε επιτυχώς το cluster μας και επιλέγουμε finish [116]



[115] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

[116] http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx

7. Τέλος αν ανοίξουμε το **Failover Cluster Manager** μπορούμε να βεβαιώσουμε ότι οι server μας ανταποκρίνονται στη λειτουργία **Failover**.



Εγκατάσταση Λειτουργίας Loadbalancing

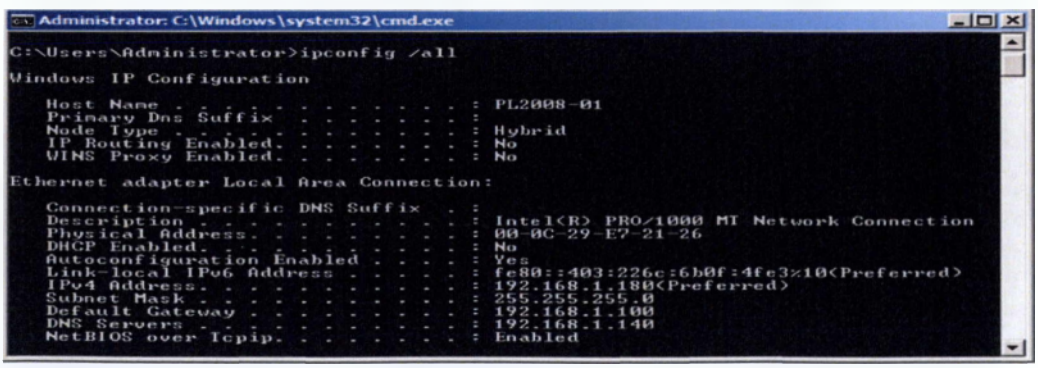
Συνδεόμαστε και στους δύο servers και εκτελούμε ipconfig / all από τη γραμμή εντολών. Χρειαζόμαστε το όνομα, το domain και τη διεύθυνση IP του κάθε server που θα είναι στο NLB Cluster. Θα πρέπει επίσης να υπάρχει ένα πρόσθετο όνομα για το cluster σε αυτό το παράδειγμα θα χρησιμοποιήσουμε SERVER-LB για το όνομα του εικονικού cluster.

Οι δυο servers που είμαστε load balancing είναι οι PL2008-01 και PL2008-02. Το όνομα του εικονικού cluster θα είναι PL2008-V. Έτσι, αν αυτό ήταν ένας χρήστης του web server θα πάει στο http://PL2008-V, ανάλογα με το πώς θα ρυθμίσουμε το NLB είτε PL2008-01, PL2008-02 ή και τα δύο servers, θα εξυπηρετήσει το web αίτημα.[117]

SERVER NAME	IP ADDRESS	TYPE
PL2008-01 pintolake net	192 168 1 180	Server 1
PL2008-02 pintolake net	192 168 1 181	Server 2
PL2008-V pintolake net	192 168 1 182	Virtual cluster name and IP address of Servers 1/2

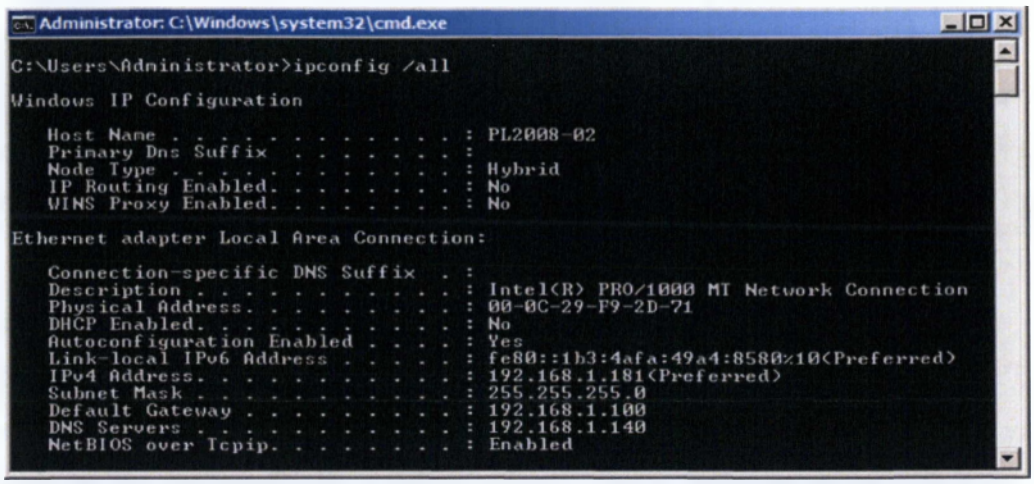
Εκτελώντας την εντολή **ipconfig / all** και στους δύο servers έχουμε:

PL2008-01



[117] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

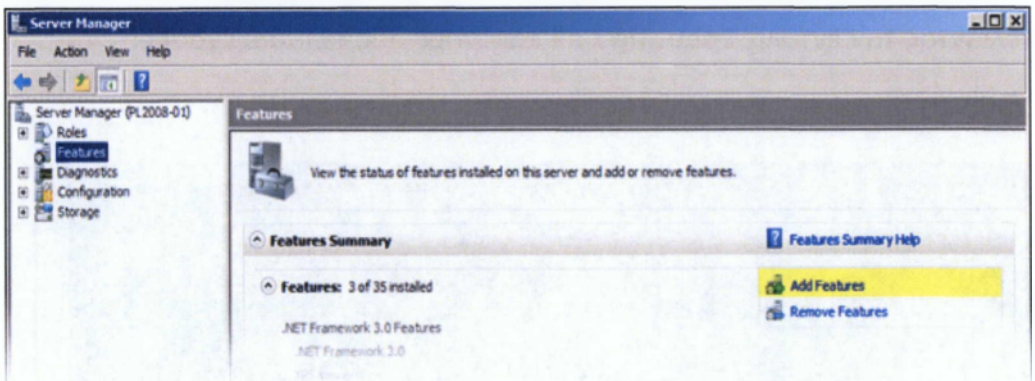
PL2008-02



Εγκατάσταση NLB feauter σε όλους τους NLB κόμβους

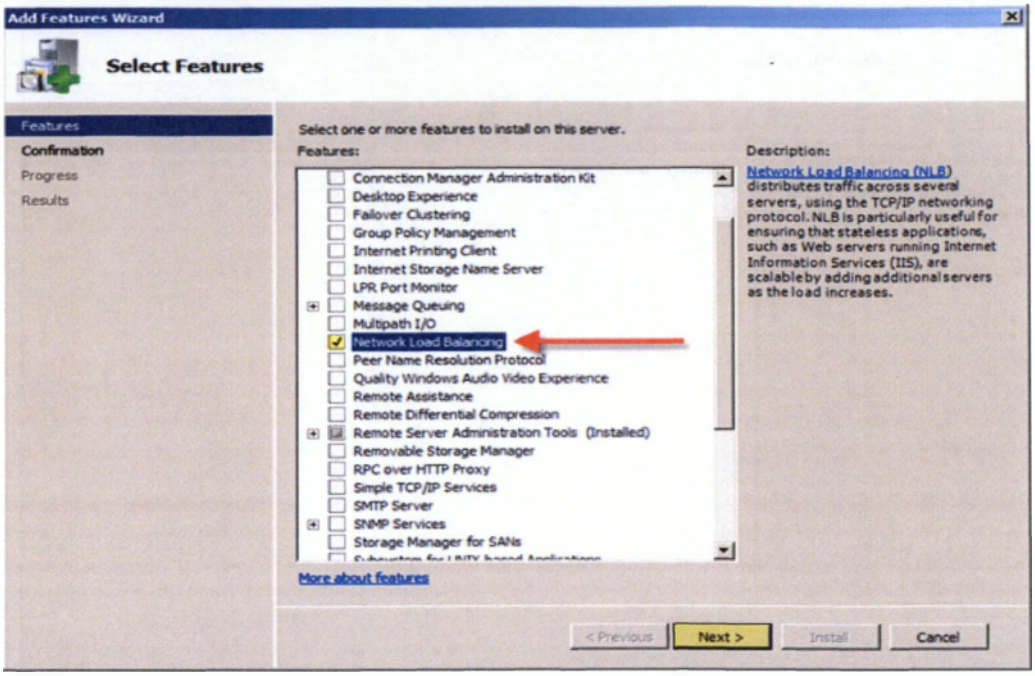
Αυτό πρέπει να γίνει σε όλους τους κόμβους του συμπλέγματος NLB. Σε αυτή την περίπτωση εκτελούμαι την εγκατάσταση και στους δύο servers.

- Επιλέγουμε " Features " από το μενού Server Manager στα αριστερά
- Επιλέγουμε "Add Features"[118]

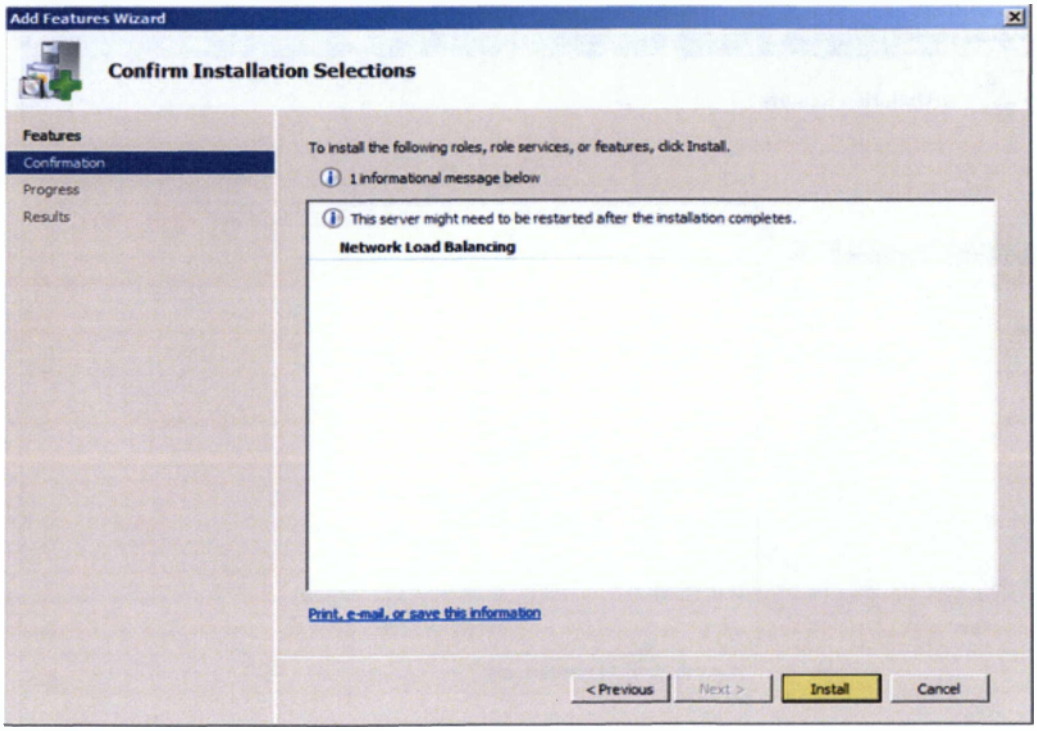


[118] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

- Επιλέγουμε το checkbox δίπλα στο "Network Load Balancing"
- Επιλέγουμε "Next"

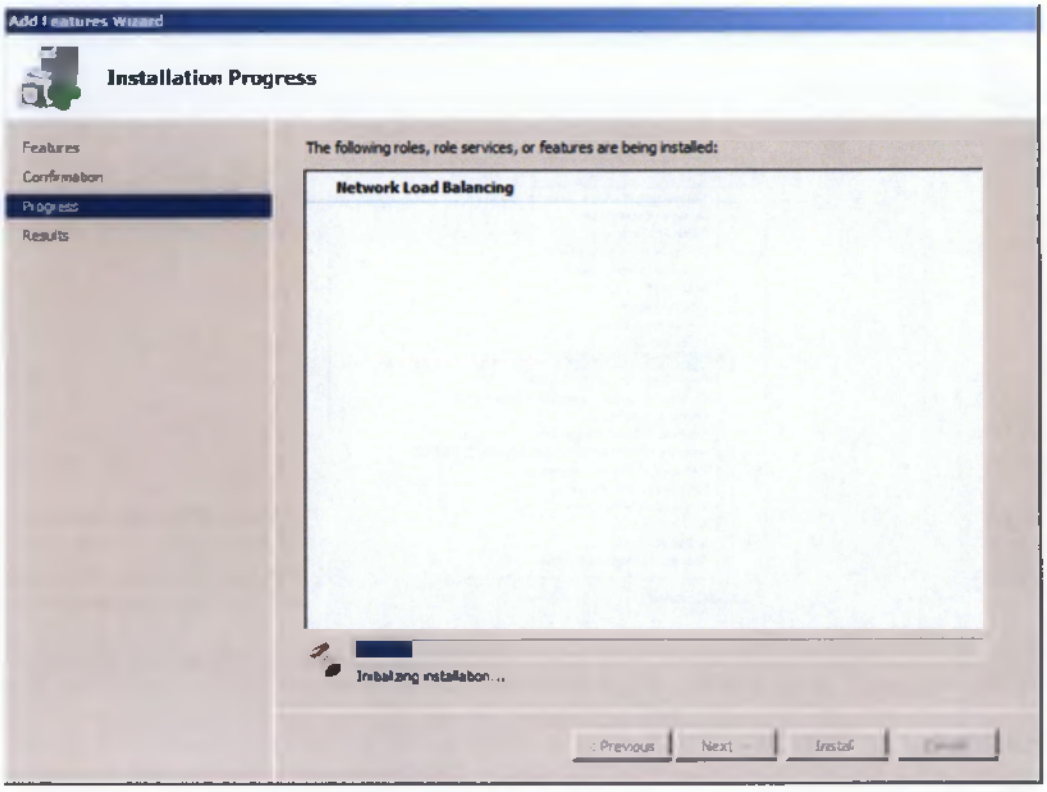


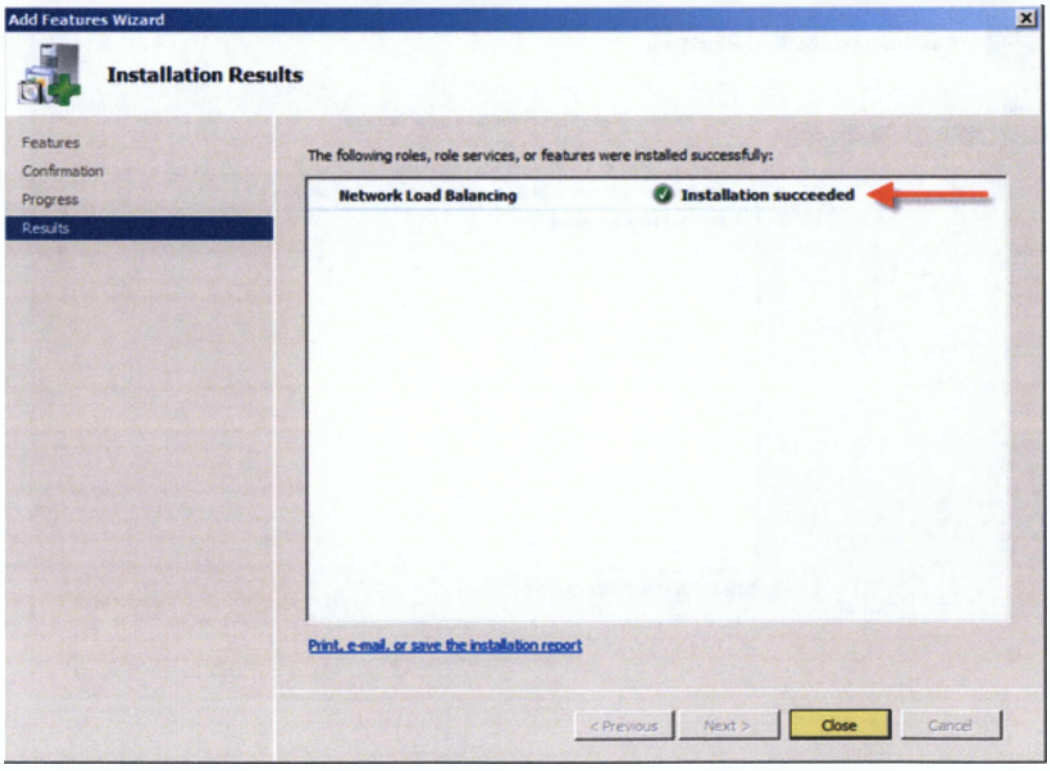
- Επιλέγουμε "Install"[119]



[119] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

Η εγκατάσταση θα προχωρήσει για να εγκαταστήσει τα απαραίτητα στοιχεία





Η εγκατάσταση έχει ολοκληρωθεί

Συνιστάται ιδιαίτερα να επαναλάβουμε αυτή τη διαδικασία για όλους τους κόμβους του συμπλέγματος NLB σε αυτό το σημείο πριν συνεχίσουμε με τη διαμόρφωση[120]

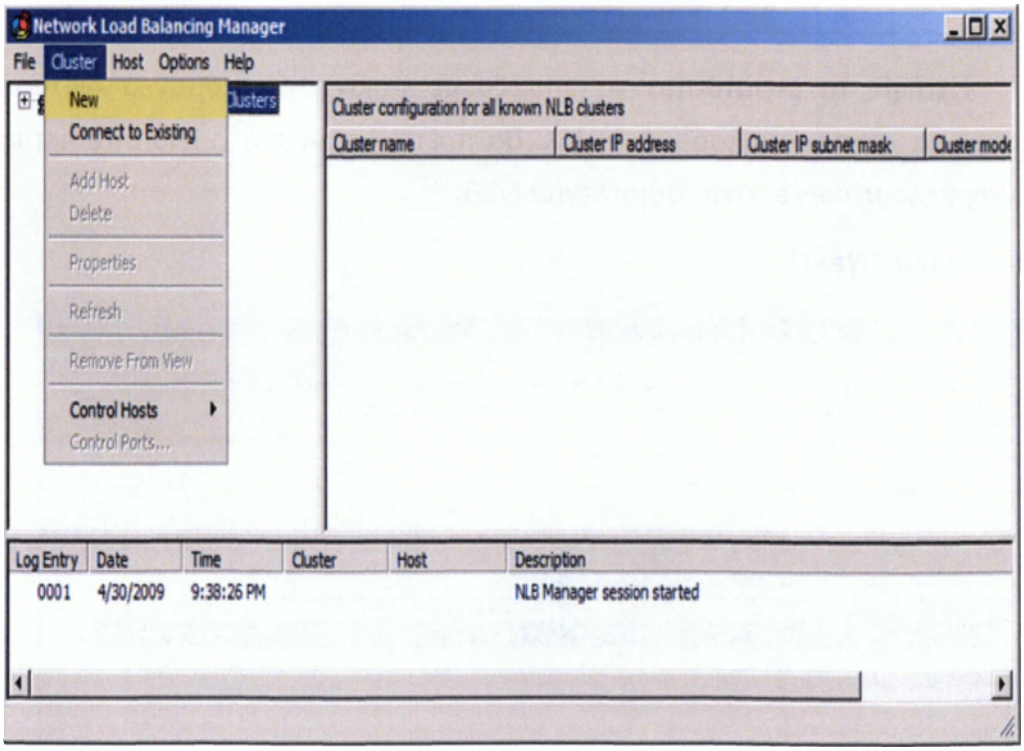
• Επιλέγουμε Close

Διαμόρφωση NLB

Διαμόρφωση NLB στον κόμβο 1 (PL2008-01)

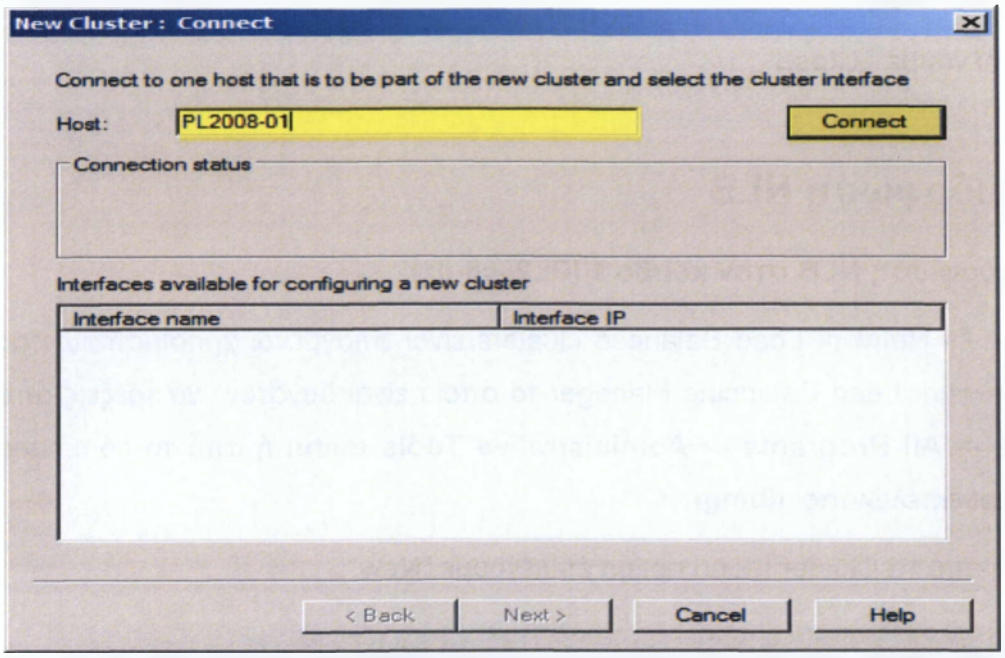
Οι Network Load Balanced clusters είναι φτιαγμένοι χρησιμοποιώντας το Network Load Balancing Manager το οποίο είναι δυνατόν να τρέξεις από **Start -> All Programs -> Administrative Tools menu** ή από το command prompt εκτελώντας **nlbmgr**

Κάτω από το Cluster Menu option επιλέγουμε **"New"**



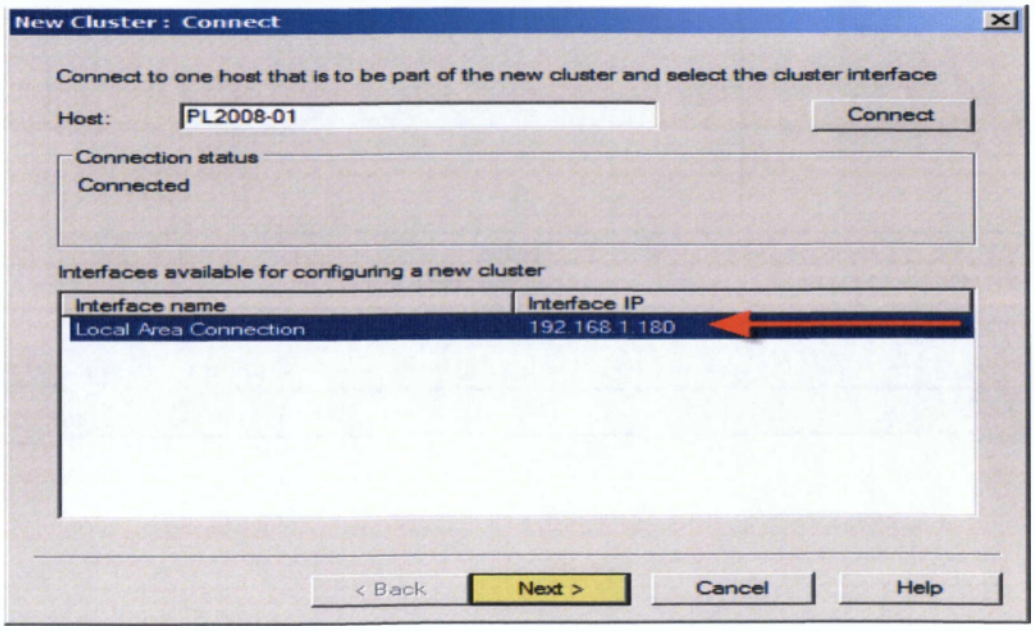
[120] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

- Εισάγουμε τον πρώτο κόμβο του συμπλέγματος που είναι PL2008-01
- Πατάμε **"Connect"**



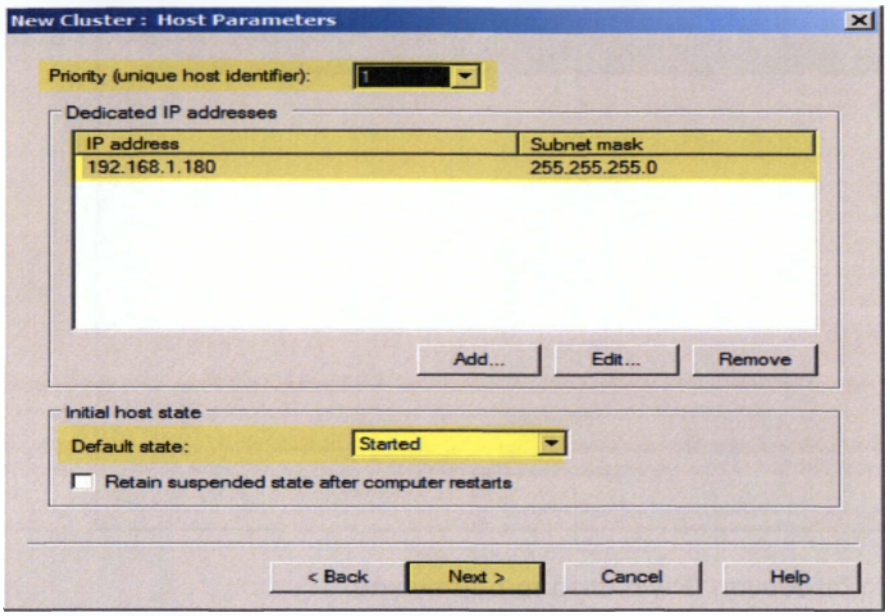
Έχουμε τη δυνατότητα να επιλέξουμε ποιόν προσαρμογέα δικτύου θέλουμε να χρησιμοποιήσουμε, το NIC θα πρέπει να είναι στο ίδιο υποδίκτυο με τους άλλους servers στο σύμπλεγμα NLB.[121]

- Επιλέγουμε **"Next"**

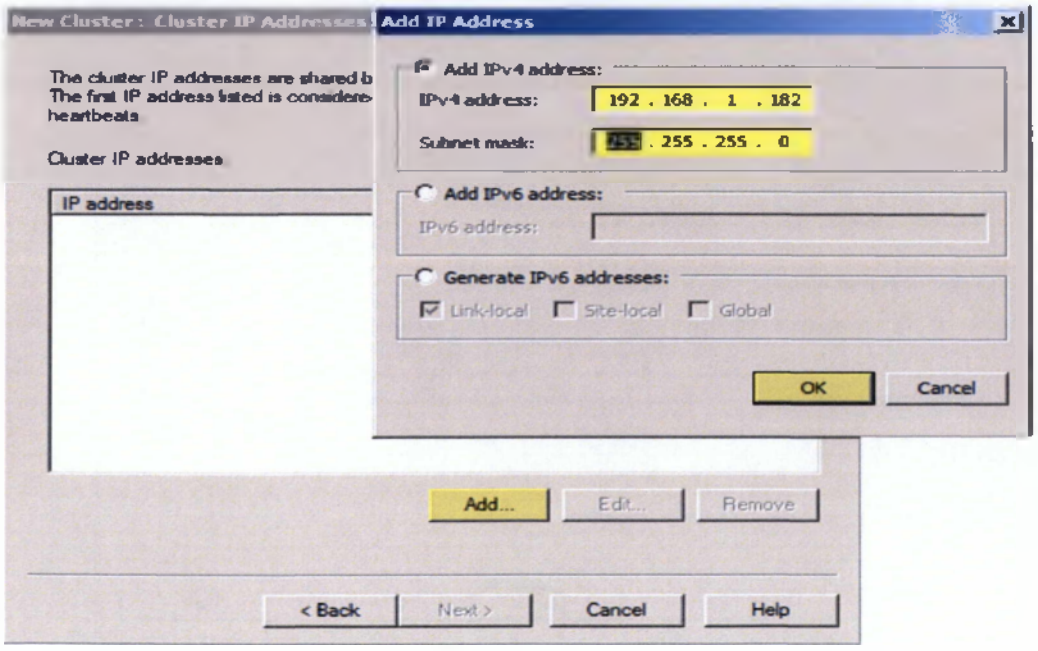


[121] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

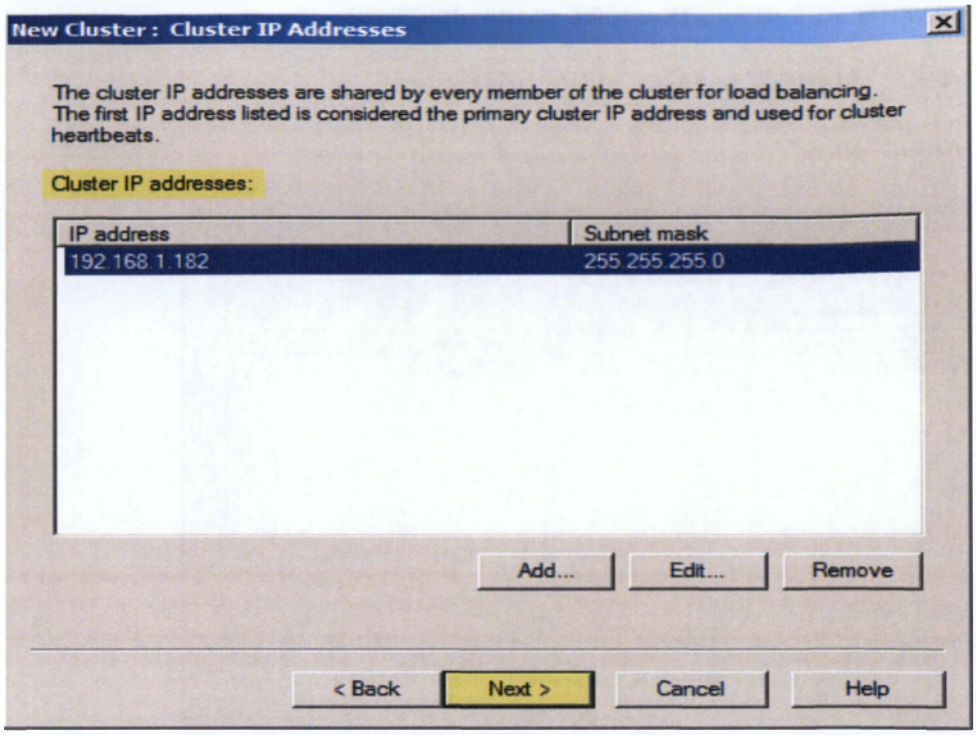
- Πληκτρολογούμε το Priority ID όπως, 1 (κάθε κόμβος του συμπλέγματος NLB θα πρέπει να έχει ένα μοναδικό αναγνωριστικό)
- Επιλέγουμε **"Started"** για την **"Initial host state"**
- Επιλέγουμε **"Next"**



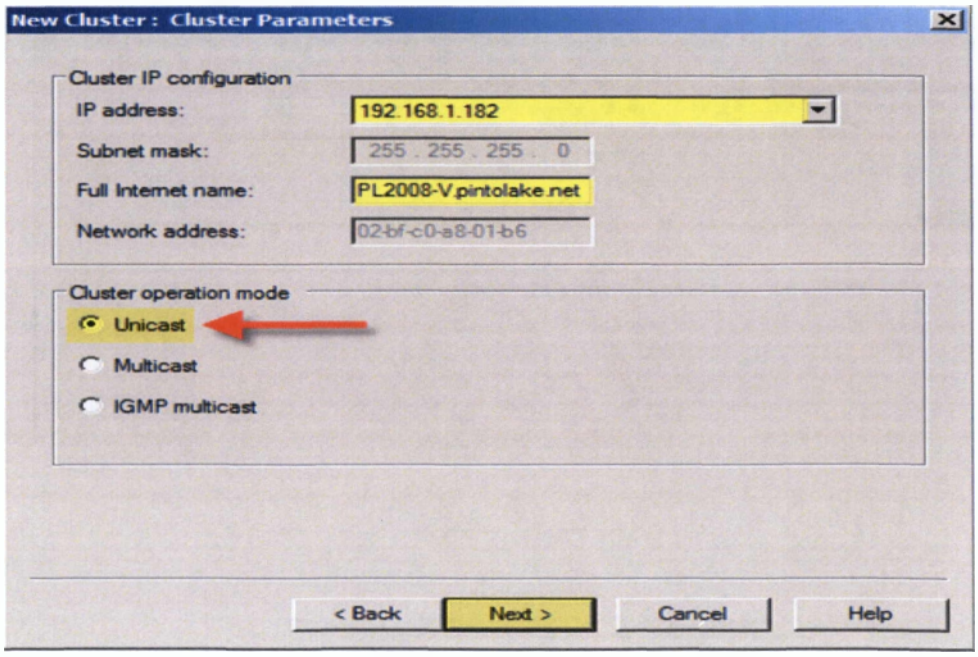
- Επιλέγουμε **"Add"**
- Εισάγουμε τη διεύθυνση IP cluster και τη μάσκα υποδικτύου
- Επιλέγουμε "OK"



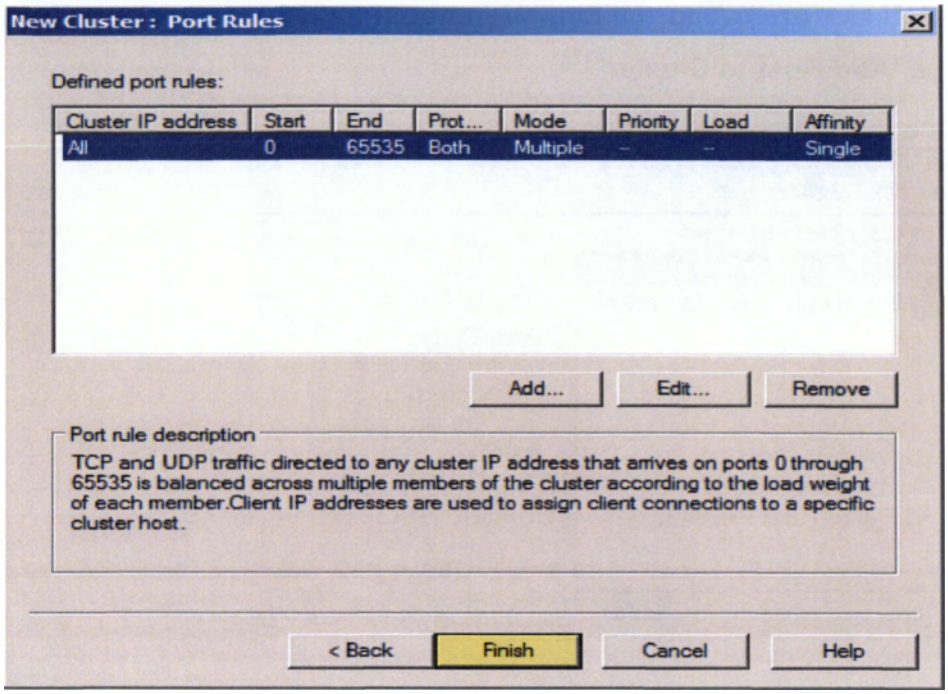
- Επιλέγουμε **"Next"**



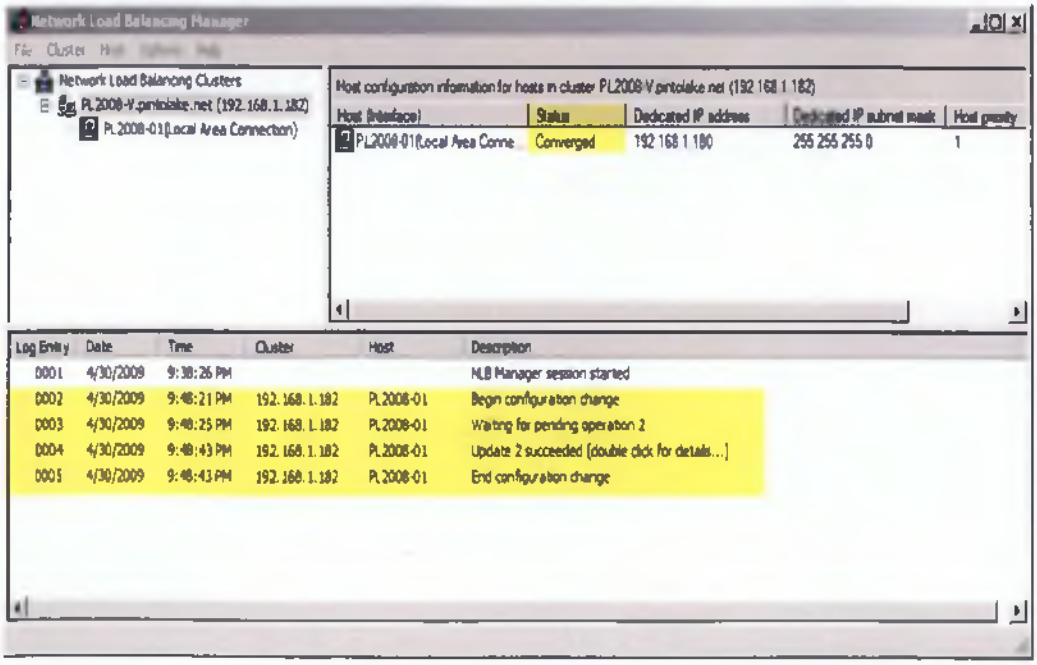
- Επιλέγουμε τη διεύθυνση IP για αυτό το σύμπλεγμα
- Εισάγουμε τη διεύθυνση NLB "PL2008-V.pintolake.net"
- Βάζουμε "Unicast" για το "Cluster operation mode"
- Επιλέγουμε **"Next"**



- Επιλέγουμε "**Finish**"

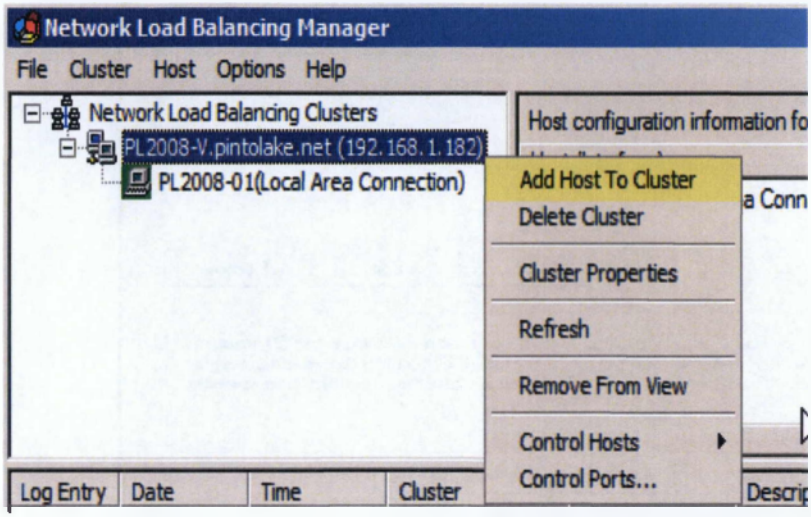


- Θα δούμε ότι αλλάζει η κατάσταση του κόμβου σε "**Converged**"
- Θα δούμε ένα "**succeeded**" στο παράθυρο καταγραφής

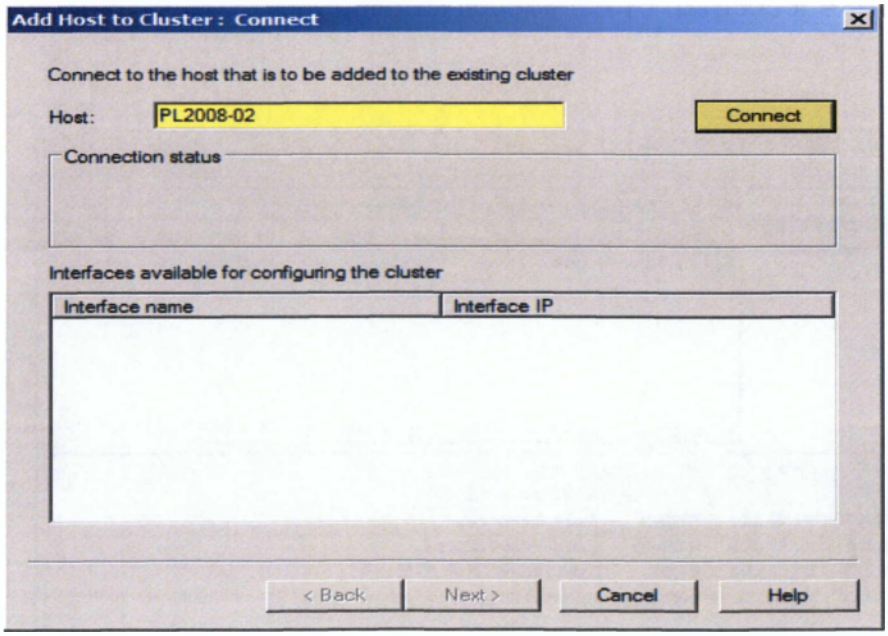


Διαμόρφωση NLB στον κόμβο 2 (PL2008-02)

Κάνουμε δεξί κλικ στο όνομα του συμπλέγματος "PL2008-V.pintolake.net" και επιλέγουμε **"Add Host to Cluster"**[122]



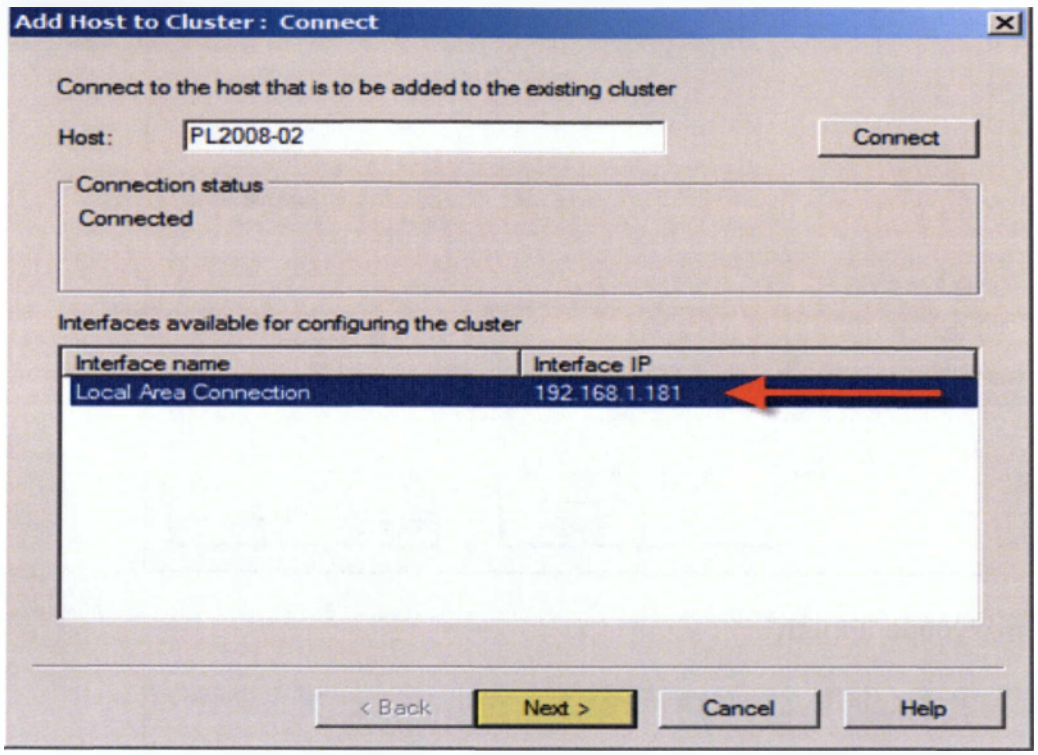
Εισάγουμε PL2008-02 και πατάμε το πλήκτρο **"Connect"**



[122] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

Μια λίστα προσαρμογέων δικτύου θα εμφανιστούν

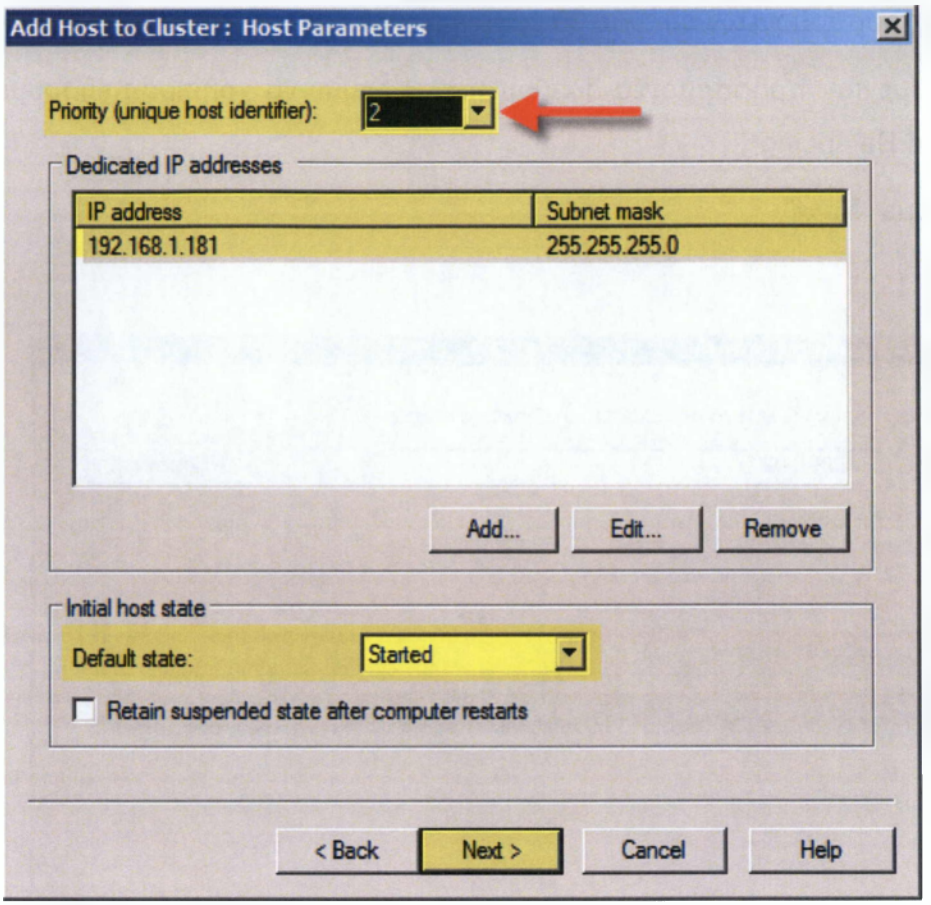
- Επιλέγουμε τον προσαρμογέα δικτύου που θέλουμε να χρησιμοποιήσουμε για το Load Balancing
- Επιλέγουμε "**Next**"



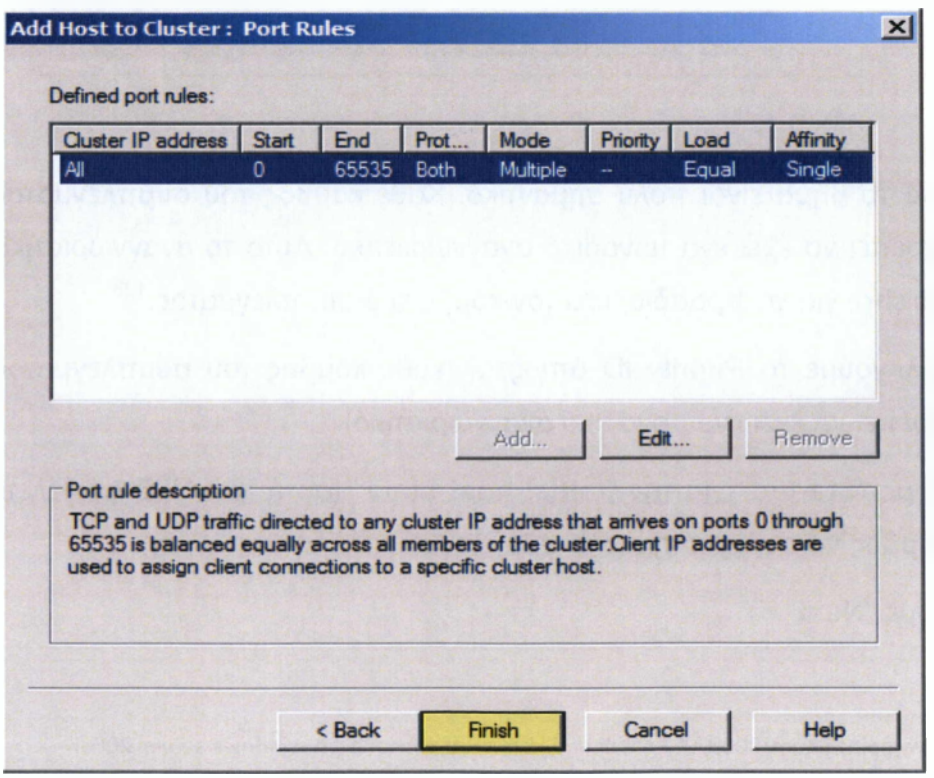
Αυτό το βήμα είναι πολύ σημαντικό. Κάθε κόμβος του συμπλέγματος NLB θα πρέπει να έχει ένα μοναδικό αναγνωριστικό. Αυτό το αναγνωριστικό χρησιμοποιείται για να προσδιορίσει τον κόμβο του συμπλέγματος.[123]

- Πληκτρολογούμε το Priority ID όπως, 2 (κάθε κόμβος του συμπλέγματος NLB θα πρέπει να έχει ένα μοναδικό αναγνωριστικό)
- Επιλέγουμε **"Started"** για την "**Initial host state**" (αυτό λέει NLB αν θέλετε αυτός ο κόμβος να συμμετάσχει στο σύμπλεγμα κατά την εκκίνηση)
- Επιλέγουμε "**Next**"

[123] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

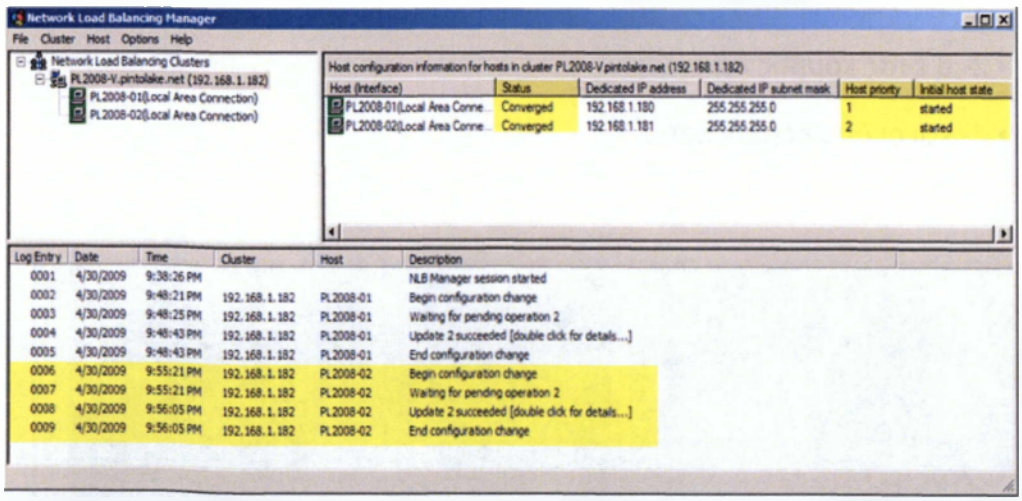


- Επιλέγουμε "Finish"



Θα πρέπει να δούμε μερικά πράγματα στο Διαχειριστή NLB

- Θα δούμε ότι αλλάζει η κατάσταση κάθε κόμβου σε **"Converged"**
- Θα δούμε επίσης ότι κάθε κόμβος έχει ένα μοναδικό **"host priority"** ID
- Επίσης κάθε κόμβος είναι **"started"** υπό τον τίτλο **"initial host state"**
- Θα δούμε ένα **"succeeded"** στο παράθυρο καταγραφής για το δεύτερο κόμβο



Δοκιμές

Πηγαίνουμε στη γραμμή εντολών και πληκτρολογούμε **"wlbs query"** , όπως μπορούμε να δούμε HOST 1 και HOST 2 συνέκλιναν με επιτυχία στο σύμπλεγμα. Αυτό σημαίνει ότι τα πράγματα λειτουργούν καλά.

- Κάνουμε ping σε κάθε server τοπικά και απομακρυσμένα
- Κάνουμε ping στον virtual IP τοπικά και απομακρυσμένα - θα πρέπει να το κάνουμε αυτό τρεις φορές από κάθε θέση.
    - 1- και οι δύο κόμβοι επάνω
    - 2- ο ένας κόμβος κάτω
    - 3- και οι δύο κόμβοι κάτω[124]

[124] http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/

Επίλογος

Οι τεχνολογίες "failover" και "load balancing" έχουν εξελιχθεί πολύ στις μέρες μας, και έχουν γίνει ένα αναπόσπαστο αλλά και ζωτικής σημασίας κομμάτι για την υποστήριξη των server για αυτό χρησιμοποιούνται ευρέος από μικρές έως και πολύ μεγάλες επιχειρήσεις, οργανισμούς και εκπαιδευτικά ιδρύματα.

Σύμφωνα λοιπόν με όσα αναφέραμε, οι παραπάνω τεχνολογίες είναι μοναδικές καθώς προσφέρουν διαθεσιμότητα όπου σε περίπτωση που κάποιος server δεν ανταποκριθεί θα βγει εκτός cluster αναλαμβάνοντας οι υπόλοιποι να απαντήσουν στις αιτήσεις μέχρι να λυθεί το πρόβλημα και να επανέλθει στο cluster. Όπως επίσης επεκτασιμότητα καθώς ανάλογα με την υπολογιστική ισχύ που χρειαζόμαστε απλά μπορούμε να προσθέσουμε ακόμη ένα server στο δίκτυο αντί να γίνει αντικατάσταση του ήδη υπάρχοντος server με ένα μεγαλύτερης ισχύος. Ασφάλεια επίσης είναι ένα από τα μεγαλύτερα πλεονεκτήματα που μας παρέχετε καθώς ο χρήστης δεν μπορεί να δει τη διεύθυνση από του server μας, παρά μόνο την εικονική διεύθυνση του cluster μας, όπου αυτό μας εξασφαλίζει ανοσία σε κακόβουλες επιθέσεις.

Τέλος με τις καινούργιες απαιτήσεις που έχουν δημιουργηθεί στο κλάδο των δικτύων αλλά και τις αυξημένες ανάγκες των εφαρμογών για αξιοπιστία, ασφάλεια, και ανθεκτικότητα, η ανάγκη για μηχανισμούς αποτυχίας και εξισορρόπηση φορτίου είναι μεγαλύτερη από ποτέ.

Βιβλιογραφία

- Condos,C.,James A, Every P., Terry Simpson T.(2010)" Ten usability principles for the development of effective WAP and m-commerce services, Aslib Proceedings Vol. 54 No. 6, pp. 345-355
- Failover Clustering in Windows Server 2008 R2, *Microsoft White Paper*, April 2009
- Geographically Dispersed Clusters, *Microsoft TechNet*; 2010
- High Availability, http://www.linux-ha.org
- High Availability, http://www.linux-ha.org
- Ldirectord, http://www.vergenet.net/linux/ldirectord/
- Ldirectord, http://www.vergenet.net/linux/ldirectord/
- Microsoft Windows 2008 R2 MPIO policies information can be found at http://technet.microsoft.com/en-us/library/dd851699.aspx.
- NeverFail ClusterProtector, http://extranet.neverfailgroup.com/download/DS-cluster-08-09-4page-lo.pdf.
- Quagga, a software routing suite, http://www.quagga.net
- Quagga, a software routing suite, http://www.quagga.net
- RFC1771 - A Border Gateway Protocol 4, http://www.faqs.org/rfcs/rfc1771.html
- Server Clusters: Architecture Overview for Windows Server 2003, *Microsoft White Paper*, March 2003
- The Linux Virtual Server Project, http://www.linuxvirtualserver.org
- Chandra Kopparapu "Load Balancing Servers, Firewalls, and Caches", John Wiley & Sons, 2002
- Καταμερισμός φορτίου σε εξυπηρετητές Web server Load-balancing,Σπυρίδων Σ. Παπαδάκης,2007
- http://en.wikipedia.org/wiki/Weighted_round_robin

- http://www7.informatik.uni-erlangen.de/~ksjh/research/cluster/
- http://www.cnri.reston.va.us/AGS/problem.html
- Yiping Ding "Performance Impact of Load Balancers on Server Farms" BMC Software
- http://www.centos.org/docs/5/html/Virtual_Server_Administration/s2-lvs-sched-VSA.html
- Job Scheduling Algorithms in Linux Virtual Server
- http://support.sas.com/ Understanding the Load-Balancing Algorithms
- Analysis of Simple Algorithms for Dynamic Load Balancing Murat Alanyali and Bruce Hajek
- http://www.cisco.com/ Configuring Server Load Balancing
- http://www.elmajdal.net/win2k8/Installing_Failover_Clustering_With_Windows_Server_2008_R2.aspx
- http://www.jppinto.com/2009/05/install-and-configure-wlbs-nlb-on-windows-server-2008/