Πανεπιστήμιο Δυτικής Μακεδονίας
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Designing a High Performance System Cluster (HPC)

Εργαστήριο "Ψηφιακών Συστημάτων και Αρχιτεκτονικής Υπολογιστών"

Επιμέλεια: Φώτιος Μακρίδης

Επιβλέπων καθηγητής: Μηνάς Δασυγένης

https://arch.icte.uowm.gr

Περιεχόμενα (1/2)

Περιεχόμενα (2/2)

Εισαγωγή – Πρώτες χρήσεις των υπερυπολογιστών (1/9)

- Οι υψηλών επιδόσεων υπολογιστικές ή πιο απλά οι υπερυπολογιστές, έχουν μία σπουδαία ιστορία τα τελευταία 90 χρόνια και κυρίως από τη δεκαετία του 1940 και μετά, όταν και έκαναν την εμφάνισή τους στις Ηνωμένες πολιτείες της Αμερικής.

- Οι πρώτες χρήσεις των εφαρμογών των υπερυπολογιστών εκείνη την εποχή ήταν για στρατιωτικούς σκοπούς ενώ κατά τη δεκαετία του 1960, οι εφαρμογές αυτές άρχισαν να βρίσκουν υλοποίηση και σε προβλήματα της καθημερινής ζωής.

- Οι πρώτοι ανταγωνιστές των Ηνωμένων πολιτειών στον τομέα των HPC ήταν αρχικά οι Ιάπωνες, με τα δύο αυτά κράτη να ανταγωνίζονται στον τομέα της τεχνολογικής ανάπτυξης των υπερυπολογιστών.

Εισαγωγή – Πρώτες χρήσεις των υπερυπολογιστών (2/9)

- Οι Ιάπωνες εκείνη την εποχή είχαν εστιάσει στο λεγόμενο vector-based supercomputing ή αλλιώς στην υπερυπολογιστική που βασίζεται στα διανύσματα, καταφέρνοντας έτσι να ολοκληρώσουν, όπως λέγεται, την προσομοίωση της γης, δημιουργώντας έτσι έναν ολοκληρωμένο ψηφιακό χάρτη, ο οποίος απεικόνιζε τη γη. [15]

- Από την άλλη, οι Ηνωμένες πολιτείες καινοτόμησαν στο λεγόμενο parallel supercomputing ή αλλιώς στην παράλληλη υπερυπολογιστική.

- Είναι εύκολα αντιληπτό πως από τη δεκαετία του 1960 μέχρι σήμερα οι αλλαγές στους υπερυπολογιστές είναι τεράστιες, κυρίως όσον αφορά την απαιτούμενη υπολογιστική ισχυ και με τον τρόπο με τον οποίο οι υπερυπολογιστές δημιουργούνται.

[15] https://www.nap.edu/read/11148/chapter/12

Εισαγωγή – Ο ρόλος του διαδικτύου (3/9)

- Το ARPANET (το πρώτο δίκτυο μεταγωγής πακέτου, που συνέθετε το internet), το οποίο πρωτοεμφανίστηκε μερικές δεκαετίες μετά την πρώτη δημιουργία των υπερυπολογιστών, έχει πλέον πάνω από 50 χρόνια που δημιουργήθηκε και το World Wide Web (γνωστό σε όλους και ως WWW), υπάρχει πλέον για πάνω από 20 χρόνια.

- Αυτές οι δύο τεχνολογίες έχουν φέρει τα πάνω-κάτω στην εξέλιξη της ανθρωπότητας, της μετάδοσης της πληροφορίας και της προόδου του βιομηχανικού τομέα, χωρίς να υπάρχει κάποιο σημάδι ότι θα επιβραδυνθεί αυτή η κατάσταση.

- Η εμφάνιση μερικών διάσημων υπηρεσιών διαδικτύου, όπως το βασισμένο στο ίντερνετ email, η διαδικτυακή αναζήτηση και τα κοινωνικά δίκτυα, μαζί με την ολοένα και αυξανόμενη ανάγκη για παγκόσμια διαθεσιμότητα για γρήγορα σύνδεση στο ίντερνετ, έχουν φέρει στο προσκήνιο το cloud computing.

Εισαγωγή (4/9)

- Επιπλέον, οι υπολογιστικές διαδικασίες και η αποθήκευση «περνάνε» με γοργούς ρυθμούς από τους επιτραπέζιους υπολογιστές σε μικρότερες φορητές συσκευές, συνδυασμένες με διαδικτυακές υπηρεσίες.

- Ενώ αρχικά πολλές υπηρεσίες διαδικτύου ήταν μόνο πληροφοριακού τύπου, σήμερα, πολλές εφαρμογές διαδικτύου περιλαμβάνουν υπηρεσίες με περιέχομενο email, φωτογραφιών, απόθηκευση βίντεο και διαφόρων τύπου εφαρμογές.

- Η στροφή στο cloud computing δεν οδηγείται μόνο από την ανάγκη για βελτίωση της εμπειρίας χρήσης (UI –User experience) των χρηστών, αλλά και από τα πλεονεκτήματα που προσφέρει στους χρήστες. Το λογισμικό σαν υπηρεσία επιτρέπει γρηγορότερη εξέλιξη εφαρμογών, επειδή είναι απλούστερο για τους προγραμματιστές να κάνουν αλλαγές και βελτιώσεις.

Εισαγωγή (5/9)

- Αντί λοιπόν να υποχρεώνονται εκατομμύρια χρήστες, να κατεβάζουν πολλές και περίεργες τροποποιήσεις λογισμικού και υλικού, οι πωλητές μπορούν να δημιουργήσουν δικό τους κέντρο δεδομένων – ή αλλιώς datacenter - και έτσι να περιορίσουν την διανομή του υλικού σε μόνο μερικές, δοκιμασμένες συσκευές.

- Η νέα τάξη πραγμάτων όσον αφορά τα κέντρα δεδομένων και οι χαμηλότερες – σε σχέση με το παρελθόν - πλέον τιμές δημιουργίας τους, επιτρέπουν πολλές υπηρεσίες εφαρμογών να τρέχουν με ένα χαμηλό κόστος ανά χρήστη.

- Για παράδειγμα, οι διακομιστές μπορούν να διαμοιραστούν ανάμεσα σε χιλιάδες ενεργούς χρήστες, βελτιώνοντας έτσι την εμπειρία χρήσης.

Εισαγωγή (6/9)

- Μερικές εφαρμογές έχουν απαιτήσεις που χρειάζονται τόσο πολύ διαθέσιμη υπολογιστική ισχύ, που είναι υλοποιήσιμες μόνο από υποδομές τεράστιας υπολογιστική ισχύος παρά από κάποια υπολογιστική υποδομή μορφής client-side (λειτουργίες που γίνονται στην πλευρά του πελάτη).

- Η τάση προς το cloud computing και η έκρηξη της δημοφιλότητας των υπηρεσιών διαδικτύου, δημιούργησε μία νέα τάξη πραγμάτων στα υπολογιστικά συστήματα και έφερε στο προσκήνιο τους λεγόμενους HPC (High Performance Computing) clusters.

- Το όνομα αυτό χρησιμοποιείται για να ορίσει το πιο διακριτό χαρακτηριστικό αυτών των μηχανών: την τεράστια κλιμάκωση του υποδομής του λογισμικού, τις αποθήκες δεδομένων και την πλατφόρμα του υλικού.

Εισαγωγή (7/9)

- Με τον ερχομό των κέντρων δεδομένων στην αγορά, επιλύονται διάφορα ζητήματα, όπως το ότι θεωρούσαμε πως ένα πρόγραμμα τρέχει σε μία μόνο μηχανή ή συσκευή.

- Στο HPC, το πρόγραμμα τρέχει σαν υπηρεσία διαδικτύου, η οποία μπορεί να αποτελείται από πολλά διαφορετικά προγράμματα τα οποία αλληλεπιδρούν με πολλές τελικές υπηρεσίες χρήστη όπως το email, η αναζήτηση και οι χάρτες.

- Τα προγράμματα αυτά δημιουργούνται και συντηρούνται από διαφορετικές ομάδες μηχανικών, πιθανώς και από εταιρίες οι οποίες έχουν τελείως διαφορετική οργάνωση, υποδομή και βρίσκονται σε εντελώς διαφορετικά γεωγραφικά όρια.

Εισαγωγή (8/9)

- Το υλικό από μια τέτοια πλατφόρμα αποτελείται από χιλιάδες υπολογιστικούς κόμβους, με τα αντίστοιχα υποσυστήματα διαδικτύου και αποθήκευσης, την κατανομή ενέργειας και τα εκτεταμένα συστήματα ψύξης.

- Το αποτέλεσμα από τέτοια συστήματα είναι στην ουσία ένα δωμάτιο με διακομιστές, το οποίο είναι πανομοιότυπο με μία αποθήκη.

- Έτσι λοιπόν, καθώς η υπολογιστική περνάει στην cloud εποχή, οι υπολογιστικές πλατφόρμες δεν μοιάζουν πλέον με ορθογώνιο κουτί, σαν ψυγείο, αλλά με ένα τεράστιο χώρο, ο οποίος είναι γεμάτος με υπολογιστές.

Εισαγωγή (9/9)

- Αυτά τα νέα μεγάλα κέντρα δεδομένων είναι αρκετά διαφορετικά από τις παραδοσιακές εγκαταστάσεις φιλοξενίας υπολογιστών των προηγούμενων ετών και σε καμία περίπτωση δεν μπορούν να θεωρηθούν ώς μία συλλογή παρόμοιων διακομιστών.

- Μεγάλα τμήματα από υλικό και λογισμικό πρέπει να δουλέψουν αρμονικά μεταξύ τους για να προσφέρουν καλές επιδόσεις όσον αφορά το κομμάτι των υπηρεσιών του ίντερνετ, κάτι το οποίο μπορεί να συμβεί μόνο αν προηγηθεί σωστή μελέτη και δημιουργία του κέντρου δεδομένων.

Τι είναι το High-Performance computing (HPC)

- High-performance computing (HPC) είναι η χρήση της παράλληλης επεξεργασίας (Parallel processing), προκειμένου να «τρέξουν» προηγμένα προγράμματα αποδοτικά, αξιόπιστα και γρήγορα. Αυτός ο όρος ισχύει για συστήματα τα οποία μπορούν να εκτελέσουν λειτουργίες πάνω του ενός teraflop (προέρχεται από το flop, που είναι μονάδα μέτρησης των επιδόσεων ενός υπολογιστή) ή αλλιώς σε 10^{12} λειτουργίες το δευτερόλεπτο.

- Ο όρος HPC συχνά χρησιμοποιείται και ως συνώνυμο για το supercomputing.

- Μερικοί υπερυπολογιστές δουλεύουν σε ρυθμούς επεξεργασίας οι οποίοι είναι λίγο παραπάνω από ένα petaflop ή 10^{15} floating-point λειτουργίες το δευτερόλεπτο (1000 τρισεκατομμύρια υπολογισμούς), ή αλλιώς 10^{3} γρηγορότερα από τα συστήματα που εκτελούν λειτουργίες του ενός teraflop.

Εισαγωγή στο HPC(1/2)

- Στη σύγχρονη τεχνολογία δεν περιλαμβάνονται μόνο οι επιτρεπέζιοι υπολογιστές (PC – Personal Computer), οι φορητοί υπολογιστές (Laptops), τα έξυπνα κινητά (smartphones), οι ταμπλέτες (tablets) και οι διάφορα συσκευές - gadgets. Περιλαμβάνονται και τα συστήματα υψηλής απόδοσης, που ξεπερνούν κατά εκατοντάδες φορές ακόμα και τα πιο ισχυρά PC της λιανικής αγοράς.

- Τα συστήματα αυτά ανήκουν στην κατηγορία του High Performance Computing (HPC).

- Οι υπολογιστές υψηλής ισχύος (HPC), καθίστανται ολοένα και πιο απαραίτητοι τη σύγχρονη εποχή για τις επιχειρήσεις και τους οργανισμούς, συμβάλλοντας τα μέγιστα στην καινοτομία και στην επιτυχία τους.

Εισαγωγή στο HPC(2/2)

- Παρόλα αυτά, πολλοί οργανισμοί δεν διαθέτουν την απαραίτητη τεχνογνωσία, για να διαμορφώσουν, να συνθέσουν και να εγκαταστήσουν ένα σύστημα HPC, χωρίς να παρεκκλίνουν από το κύριο έργο τους, είτε αυτό ανήκει στον τομέα της επιστήμης, της κατασκευής ή της ανάλυσης.

- Για παράδειγμα, σύμφωνα με το Εθνικό Κέντρο Βιομηχανικών Επιστημών, το 98% όλων των προϊόντων θα σχεδιάζονται ψηφιακά μέχρι το 2020, ωστόσο το 95% των 300.000 κατασκευαστικών εταιρειών, που είναι εγγεγραμμένες στο Κέντρο, έχουν περιορισμένη έως καθόλου τεχνογνωσία στον τομέα των συστημάτων HPC. [1]

https://www.zougla.gr/technology/article/dell-simantikes-kenotomies-ke-ekseliksis-sta-sistimata-high-performance-computing [1]

Cluster: Τι είναι(1/2)

- Ένας υπολογιστής υψηλής απόδοσης, κατάλληλος για μικρές ή μεσαίου μεγέθους επιχειρήσεις, δημιουργείται από πολλούς κοινούς υπολογιστές συνδεδεμένους μαζί, σε ένα δίκτυο, ο οποίος διευθύνεται κεντρικά από κάποιο ειδικό λογισμικό.

- Επειδή οι υπολογιστές είναι από φυσικής πλευράς πολύ κοντά μεταξύ τους, ο κοινός όρος για έναν high performance computer είναι ο cluster.

- Όταν γίνεται αναφορά στον όρο HPC cluster, συνήθως εννοείται το πόσοι επεξεργαστές ή πυρήνες διαθέτει. Συνήθως, είναι γνωστό το τι είναι ένας επεξεργαστής αλλά πολλές φορές προκαλείται σύγχηση με τον όρο πυρήνες. Από την αρχή της ύπαρξης των υπολογιστών, οι εταιρίες οι οποίες κατασκεύαζαν επεξεργαστές, δημιουργούσαν κυκλώματα με έναν πυρήνα. Πλέον, οι σύγχρονοι επεξεργαστές, διαθέτουν περισσότερους του ενός πυρήνα, όπως για παράδειγμα 2, 4, 6, 8, 16.

Cluster: Τι είναι(2/2)

- Τα τελευταία χρόνια, η Intel έφερε στο προσκήνιο τους επεξεργαστίες i9, όπου υπάρχουν στον επεξεργαστή 10, 18 ή ακόμα και 28 πυρήνες. Ουσιαστικά αυτή η κατηγορία της intel, αποτέλεσε απάντηση στην AMD η οποία είχε προωθήσει στην αγορά επεξεργαστές με 16 πυρήνες.

- Οι πυρήνες αποτελούν ουσιαστικά τα «μυαλά» του κυκλώματος (chip), το οποίο μπαίνει στις υποδοχές (sockets) της μητρικής πλακέτας (motherboard). Ανάλογα με τις απαιτήσεις και τις ανάγκες, υπάρχουν διάφορες λύσεις όσον αφορά το σασί του cluster. Αρχής γενομένης με τα personal-sized clusters (θα αναλυθεί παρακάτω), μπορούμε να ανεβούμε κλίμακα απόδοσης, ανάλογα με τον αριθμό των υπολογιστών, το φυσικό χώρο και την υπολογιστική δύναμη που θα απαιτηθεί.

HPC Clusters(1/2)

- Ένας HPC cluster αποτελεί μία ομάδα από διακομιστές συνδεδεμένους με ένα ειδικό δίκτυο υψηλής ταχύτητας.

- Παράδειγμα HPC cluster αποτελεί ο Mesabi, ο οποίος είναι ένας κατανεμημένος HP Linux cluster, ο οποίος βρίσκται στο MSI (Minessota SuperComputing Institute). Μερικά από τα χαρακτηριστικά του είναι:
  - Αποτελείται από 741διακομιστές (ή κόμβους).
  - Είναι κατασκευασμένος από την HP με δίκτυο Infiniband της Mellanox.
  - Αποτελείται από 12 πυρήνες Haswell αναπτυγένοι από την Intel, με συχνότητα που φτάνει τα 2.5 GHz.
  - Υπάρχουν 2 υποδοχές ανά κόμβο.
  - Φτάνει ακόμα και τα 960 GFLOP/s ανά κόμβο μέγιστης θεωρητικής απόδοσης.
  - 711 TFLOP/S θεωρητική μέγιστη απόδοση συστήματος.

HPC Clusters(2/2)

- Τύποι δικτύου που χρησιμοποιούνται:
  - Infiniband. Το InfiniBand (IB), είναι ένα πρότυπο δικτύου υπολογιστών που χρησιμοποιείται στο high-performance computing.
    - Χαρακτηριστικά του είναι η χαμηλή καθυστέρηση και η υψηλή διεκπεραιωτική ικανότητα (throughput).
    - Χρησιμοποιεί την τοπολογία switched, όπως είχε εμφανιστεί στο παρελθόν και στο Ethernet.
    - Φτάνει μέχρι και τα 4000 Gbps εύρους ζώνης (bandwidth).
  - Ethernet:
    - Είναι το συνηθέστερο χρησιμοποιούμενο πρότυπο δίκτυο υπολογιστών ενσύρματης σύνδεσης.
    - Φτάνει μέχρι και 10 Gbps εύρος ζώνης.
    - Φθηνότερο από το Infiniband αλλά και πολύ αργότερο επίσης.
  - Custom δίκτυα εταιριών:
    - Cray, IBM, Fujitsi, όλες αυτές οι εταιρίες έχουν custom δίκτυα.

Personal-sized clusters

- Ένα από τα πιο κατάλληλα μεγέθη για τα μικρά cluster είναι τα επιτραπέζια. Τα επιτραπέζια clusters ουσιαστικά είναι ένα διαμορφωμένο σασί, το οποίο μπορεί βρίσκεται σε κάποιο ράφι που είναι καρφωμένο σε τοίχο. Το σασί μπορεί να κρατήσει ένα μικρό σχετικά αριθμό υπολογιστών.

- Ο όρος «μικρό σχετικά αριθμό υπολογιστών» έχει σχέση με το μέγιστο μέγεθος των cluster σήμερα, ο οποίος μπορεί να περιλαμβάνει μερικούς εκατοντάδες επεξεργαστές – ένας αριθμός σίγουρα ικανός για κάποιον που θα ήθελε να κατασκευάσει ένα HPC cluster για την επιχείρησή του.

- Οι επιτρεπέζιοι clusters για εταιρίες όπως η HP, SGI, Cray, μπορούν να περιέχουν μέχρι μερικές χιλιάδες επεξεργαστών, ένα μέγεθος το οποίο είναι πολύ μεγάλο για μικρές επιχειρήσεις.

Ανεβαίνοντας επίπεδο: Σύνδεση μεταξύ των clusters

- Πολλές εταιρίες οι οποίες προσφέρουν επιτραπέζιους clusters, προσφέρουν και τη δυνατότητα να συνδεθούν 2 ή περισσότεροι μεταξύ τους, για να φτιάξουν ένα μεγαλύτερο, μεσαίου μεγέθους cluster. Ουσιαστικά, τα συνδέουν μεταξύ τους με περισσότερες συνδέσεις δικτύου, για να μπορέσουν να δημιουργήσουν ένα μεγαλύτερο cluster.

- Σε άλλη περίπτωση, οι εταιρίες προσφέρουν ένα μεγαλύτερο σασί, το οποίο μπορεί να κρατήσει παραπάνω υπολογιστές, συμπεριλαμβανομένων των μεσαίου μεγέθους ράφια (τα λεγόμενα racks), τα οποία θα ήταν ακόμα κατάλληλα για ένα περιβάλλον γραφείου. Αρκετά συχνά, συναντούμε racks τα οποία μπορούν να συμπεριλάβουν μέχρι και 800 πυρήνες, αριθμός κατάλληλος για περιβάλλον γραφείου.

Κατανοώντας τις βασικές αρχές των clusters(1/4)

- Τα συστήματα HPC είναι κατασκευασμένα από πολλά κομμάτια. Υπάρχουν κάποια κοινά κομμάτια, αλλά οι clusters μπορεί και να αποτελούνται από ένα μεγάλο εύρος λογισμικού και υλικού.

- Για να μπορέσει να κατασκευαστεί μία τέτοια μονάδα, πρέπει να γίνουν αντιληπτές βασικές έννοιες και πληροφορίες, προκειμένου το αποτέλεσμα να είναι το καλύτερο δυνατό και το πιο αποδοτικό όσον αφορά τα κόστη και την απόδοση.

- Κατά τη δημιουργία ενός cluster, είναι απαραίτητη η στροφή σε άτομα με την κατάλληλη υποδομή γνώσεων και τεχνικής κατάρτισης, ούτως ώστε να αποφευχθούν βιαστικές και μη-αποδοτικές λύσεις.

Κατανοώντας τις βασικές αρχές των clusters(2/4)

- Ο γενικός στόχος των HPC είναι είτε να «τρέχει» εφαρμογές γρηγορότερα είτε να λύνει προβλήματα τα οποία δεν μπορούν να τρέξουν σε έναν μοναδικό διακομιστή.

- Για να γίνει αυτό, χρειάζεται να «τρέξουν» παράλληλες εφαρμογές μεταξύ ξεχωριστών κόμβων. Ενώ γίνεται να χρησιμοποιηθεί ένας μονός κόμβος και στη συνέχεια δύο ξεχωριστές εικονικές μηχανές (Virtual Machines), είναι σημαντικό να γίνει κατανοητό πώς οι εφαρμογές «τρέχουν» με φυσικό τρόπο μεταξύ διαφορετικών διακομιστών και πώς γίνεται η διαχείριση του συστήματος.

Κατανοώντας τις βασικές αρχές των clusters(3/4)

- Με αυτόν το στόχο στο μυαλό, πρέπει να διευκρινιστούν οι απαιτήσεις. Εάν για παράδειγμα υπάρχει ενδιαφέρον για παράλληλη υπολογιστική χρησιμοποιώντας πολλαπλούς κόμβους, τότε χρειάζονται τουλάχιστον δύο κόμβοι, ο καθένας με δικό του λογισμικό.

- Για να τρέχουν τα πράγματα απλά, το λογισμικό και στις δύο πλευρές πρέπει να είναι παρόμοιο. Εάν γίνει εγκατάσταση στον κόμβο No.1, τότε χρειάζεται να εγκατασταθεί και στον δεύτερο κόμβο.

- Κάτι τέτοιο θα βοηθήσει όταν θα χρειαστεί να γίνει αποσφαλμάτωση (debug) στο σύστημα.

Κατανοώντας τις βασικές αρχές των clusters(4/4)

- Το δεύτερο πράγμα που χρειάζεται ο cluster είναι μία σύνδεση διαδικτύου μεταξύ των κόμβων, έτσι ώστε να επικοινωνούν και να διαμοιράζονται δεδομένα και πληροφορίες, και να διαδίδονται εύκολα διαδικασίες επίλυσης προβλημάτων.

- Το διαδίκτυο μπορεί θεωρητικά να είναι οτιδήποτε επιτρέπει την επικοινωνία μεταξύ των κόμβων, αλλά η ευκολότερη λύση είναι το Ethernet.

- Στο παρακάτω σκέλος της συγκεκριμένης εργασίας, θα αναλυθεί η δημιουργία ενός HPC.

Αρχιτεκτονική(1/12)

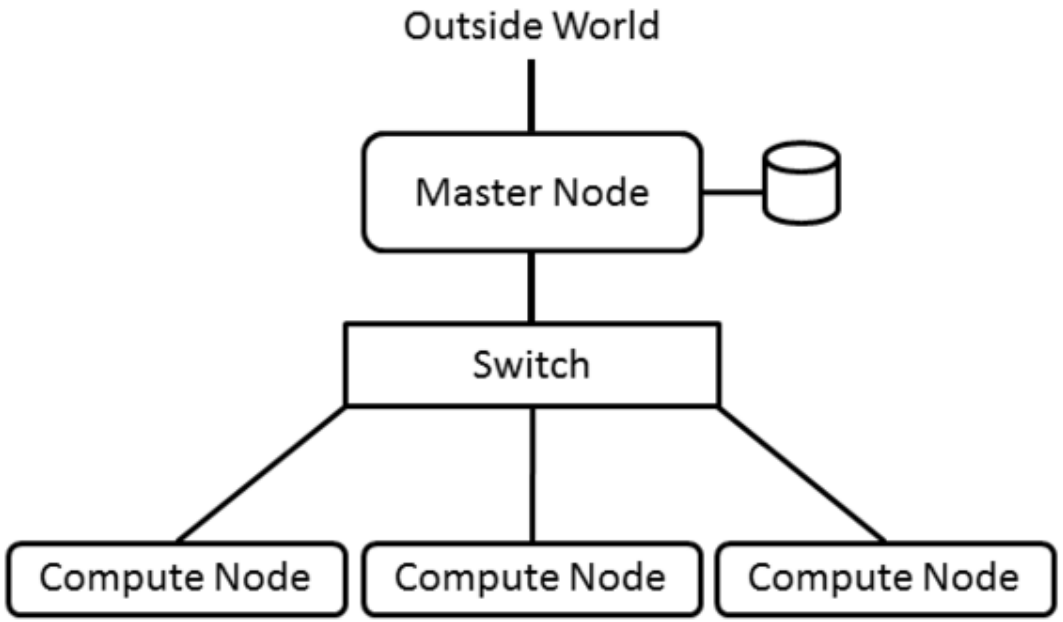
- Η αρχιτεκτονική ενός cluster είναι πολύ ξεκάθαρη. Υπάρχουν μερικοί διακομιστές οι οποίοι εξυπηρετούν αρκετούς ρόλους σε έναν cluster και είναι συνδεδεμένοι με ένα είδος δικτύου.

- Ουστιαστικά αυτή είναι η διασύνδεση. Οι κόμβοι μπορούν να είναι είτε παρόμοιοι μεταξύ τους είτε τελείως διαφορετικοί.

- Ωστόσο, προτείνεται να είναι όσο το δυνατόν ομοιότεροι, καθώς θα είναι πιο εύκολες οι διαδικασίες κατά την κατασκευή και υλοποίηση του HPC, αλλά και στη συνέχεια θα εξυπηρετήσει επίλυση προβλήματων, δημιουργία προγραμμάτων και ευκολότερη αποσφαλμάτωση.

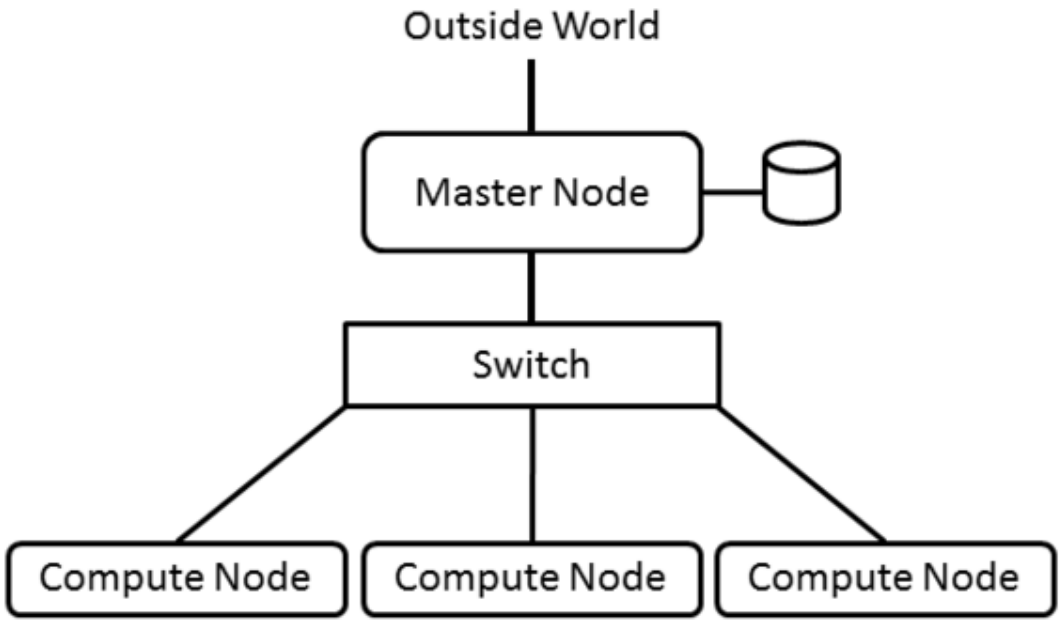
Αρχιτεκτονική(2/12) (Σχήμα)

- Σχήμα: Γενικό περίγραμμα ενός cluster.
- Σχεδόν πάντα υπάρχει ένας «master κόμβος», δηλαδή κεντρικός κόμβος (επίσης καλείται και «head κόμβος»). Ο κεντρικός κόμβος είναι ο διαχειριστής κόμβος για τον cluster. Ελέγχει και επιβλέπει όλο το σύστημα και πολλές φορές είναι ο κόμβος εισόδου για τους χρήστες, προκειμένου να τρέξουν τις εφαρμογές.
- Για μικρότερους κόμβους, ο κεντρικός κόμβος μπορεί να χρησιμοποιηθεί για υπολογιστική διαχείριση, αλλά όσο μεγαλώνει ο cluster, ο κεντρικός κόμβος γίνεται πιο λειτουργικός και δεν χρησιμοποείται μόνο για υπολογιστική.



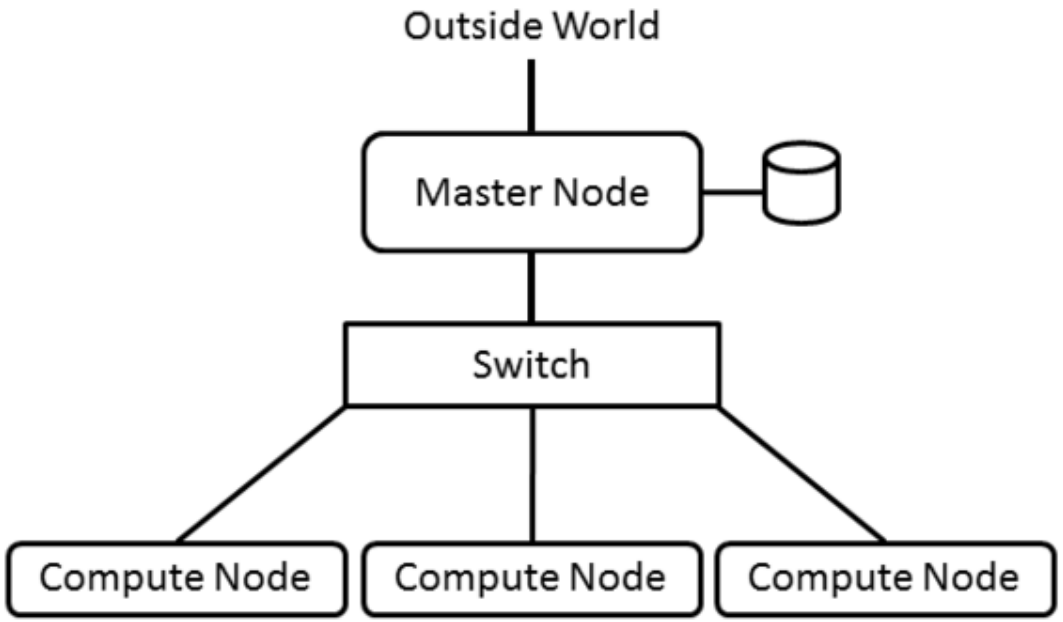
Αρχιτεκτονική(3/12) (Σχήμα)

- Άλλοι κόμβοι στον cluster αποτελούν τους κόμβους υπολογισμού, όπου περιγράφουν τη λειτουργία του.
- Ουσιαστικά, οι κόμβοι υπολογισμού δεν έχουν λειτουργία όσον αφορά τη διαχείριση του cluster, αλλά απλώς κάνουν υπολογισμούς. Οι κόμβοι υπολογισμού είναι συνήθως συστήματα τα οποία εκτελούν τις ελάχιστες απαιτήσεις από τα λειτουργικά συστήματα.
- Αυτό σημαίνει πως οι μη απαραίτητες εφαρμογές, λειτουργίες και πακέτα είναι απενεργοποιημένα και επιπλέον χρησιμοποιούν το ελάχιστο δυνατό υλικό.



Αρχιτεκτονική(4/12) (Σχήμα)

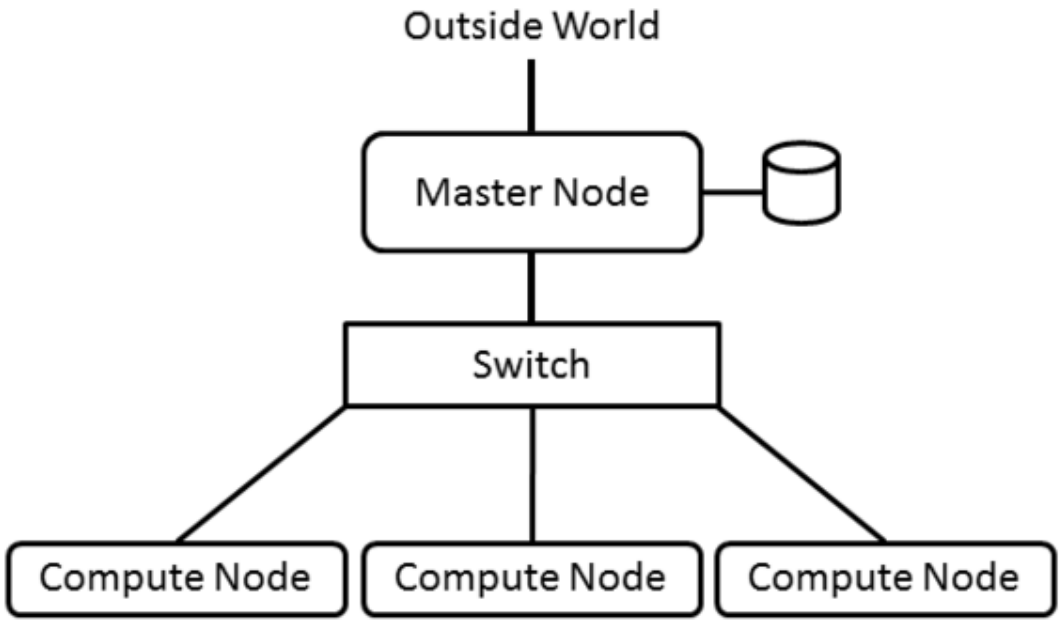
- Όσο μεγαλώνει ο cluster, πληθαίνουν και οι ρόλοι που χρειάζεται να ανατεθούν, με αποτέλεσμα να απαιτούνται επιπλέον κόμβοι.
- Για παράδειγμα, μπορούν να προστεθούν στον cluster κόμβοι δεδομένων. Αυτοί οι κόμβοι δεν τρέχουν εφαρμογές, αλλα αποθηκεύουν και προσφέρουν δεδομένα στον υπόλοιπο cluster. Οι επιπρόσθετοι κόμβοι μπορούν να παρέχουν δυνατότητες εικονοποίησης δεδομένων μέσα στον cluster (συνήθως απομακρυσμένη εικονοποίηση), αλλιώς οι μεγάλοι cluster θα χρειαστούν κόμβους ειδικούς στο να παρακολουθούν τον cluster ή τους χρήστες που συνδέονται και τρέχουν τις εφαρμογές.



Αρχιτεκτονική(5/12) (Σχήμα)

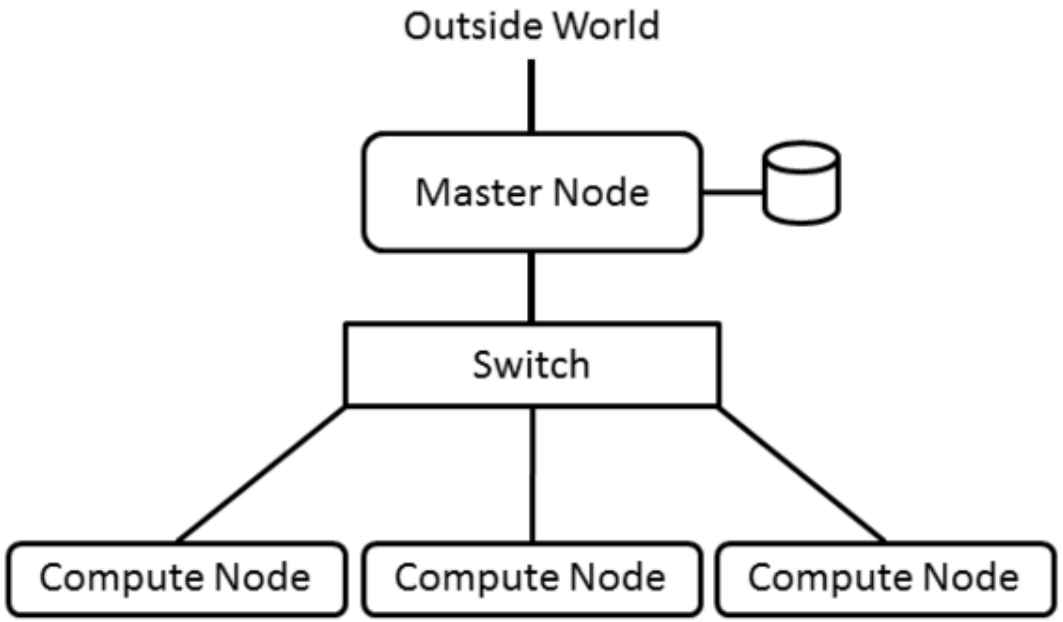
- Για έναν απλό cluster, αποτελούμενο από δύο κόμβους, ο οποίος θα χρησιμοποιηθεί ως αρχικός στο HPC, μπορούν να συνδυαστούν ένας κεντρικός κόμβος και ένας κόμβος υπολογισμών. Ωστόσο, επειδή υπάρχουν μόνο δύο κόμβοι, οι εφαρμογές πολύ πιθανόν θα «τρέξουν» και στους δύο.

- Το δίκτυο που συνδέει τους κόμβους του cluster μπορεί να είναι οποιασδήποτε τεχνολογίας διαδικτύου, αλλά το σημείο εκκίνησης είναι με καλωδιωμένο Ethernet, το οποίο έχει εύρος από 100 Mbps μέχρι 56 Gbps. Συνήθως επιλέγεται το πιο κοινό και γρήγορο Ethernet (100 Mbps) ή το Gigabit Ethernet (1000 Mbps).



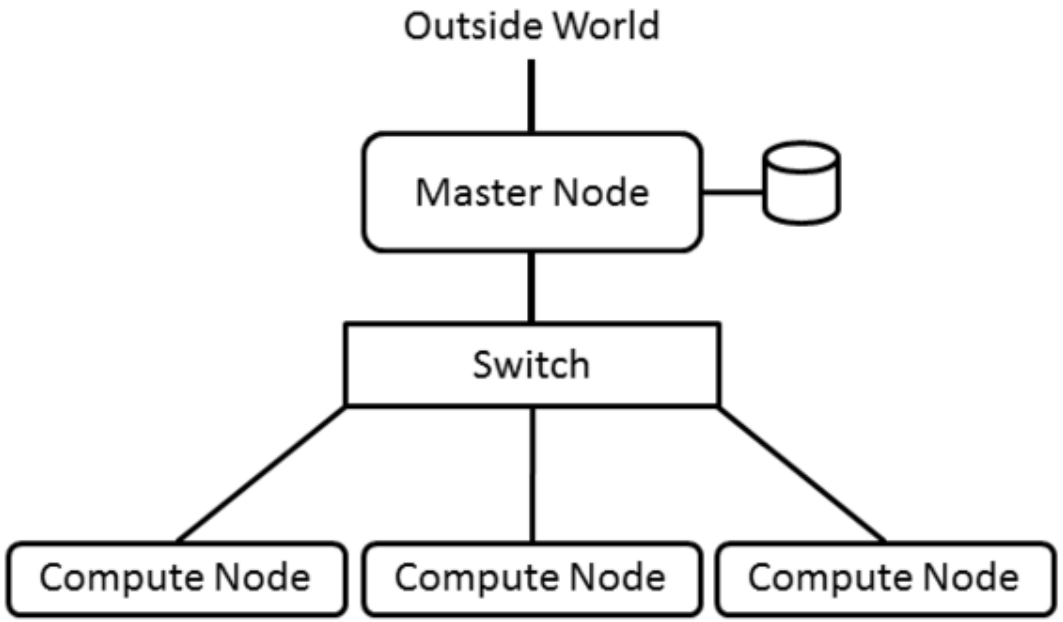
Αρχιτεκτονική(6/12) (Σχήμα)

- Η τοπολογία διαδικτύου που χρησιμοποιείται στους clusters είναι πολύ σημαντική, καθώς έχει επίδραση στην απόδοση της οποιαδήποτε εφαρμογής.
- Ένα απλό περίγραμμα δικτύου έχει ένα μοναδικό (εναλλάκτη) switch, με όλους τους κόμβους να είναι συνδεδεμένοι σε αυτό (όπως φαίνεται και στο σχήμα). Αυτή η εγκατάσταση, καλείται ένα χοντρό δέντρο τοπολογίας, έχει μόνο ένα επίπεδο και είναι απλό και αποδοτικό, ειδικότερα όταν χτίζονται μικρότερα συστήματα.
- Όσο τα συστήματα μεγαλώνουν, μπορούμε να παραμείνουμε στην παραπάνω τοπολογία, αλλά ίσως χρειαστεί να υπάρχουν παραπάνω επίπεδα από switches. Άμα γίνεται επαναχρήση των ήδη υπάρχοντων εναλλακτών, χρειάζεται προσεκτικός σχεδιασμός της τοπολογίας προκειμένου να μην υπάρχουν εμπόδια.



Αρχιτεκτονική(7/12) (Σχήμα)

- Για μικρότερα συστήματα, τα Ethernet switches είναι αρκετά φθηνά, κοστίζοντας μόνο μερικά ευρώ ανά κομμάτι.
- Οι εναλλάκτες ως επιλογές πρόκειται να είναι καλύτεροι από ότι ένα δίκτυο Ethernet, αλλά άμα υπάρχει μόνο μία θύρα - hub, τότε μπορεί να χρησιμοποιηθεί το Ethernet.



Αρχιτεκτονική(8/12)

- Μια πολύ καλή λύση είναι η σύνδεση του cluster σε ένα δημόσιο δίκτυο και ο τρόπος για να γίνει αυτό είναι όποτε ο cluster βρίσκεται σε ιδιωτικό δίκτυο, να προστεθεί ένας δεύτερος ελεγκτή διεπαφής δικτύου (Nic – network interface controller) στον κεντρικό κόμβο.

- Αυτό το δίκτυο, θα έχει δημόσια IP διεύθυνση και θα επιτρέπει τη σύνδεση στον cluster. Μόνο ο κεντρικός κόμβος θα έχει την δημόσια διεύθυνση, καθώς δεν υπάρχει λόγος για τους κόμβους υπολογισμού να έχουν δύο διευθύνσεις.

- Για παράδειγμα, μπορούμε να κάνουμε τη δημόσια διεύθυνση για τον master κόμβο κάτι σαν 72.x.x.x και την ιδιωτική κάτι σαν 10.x.x.x. Η σειρά των διαδικτυακών διεπαφών δεν κάνουν τεράστια διαφορά, αλλά πρέπει να δοθεί προσοχή καθώς γίνονται εγκατάσταση στο λειτουργικό σύστημα.

Αρχιτεκτονική(9/12)

- Είναι δυνατό να δοθούν στον κεντρικό κόμβο δύο διευθύνσεις άμα βρισκόμαστε πίσω από έναν μεταφραστή διευθύνσεων διαδικτύου (NAT). Αυτή η τροποποίηση είναι πολύ κοινή ακόμα και στα router του σπιτιού, τα οποία είναι επίσης NAT συσκευές.

- Για παράδειγμα, στο τοπικό μας δίκτυο πολλές φορές έχουμε ένα δρομολογητή internet (router), ο οποίος είναι στην πραγματικότητα ένα NAT, το οποίο μετατρέπει πακέτα από ένα ιδιωτικό δίκτυο, όπως το 192.168.x.x, στην διεύθυνση του δρομολογητή (internet) και αντίθετα. Οι απλοί clusters έχουν έναν κεντρικό κόμβο με δημόσια IP 192.168.x.x, και έχουν μία δεύτερη NIC (κάρτα δικτύου – Network interface controller) με διεύθυνση 10.x.x.x, η οποία είναι το ιδιωτικό δίκτυο του cluster.

Αρχιτεκτονική(10/12)

- Μία άλλη λειτουργία κλειδί της αρχιτεκτονικής ενός βασικού cluster είναι ο διαμοιρασμένος φάκελος μεταξύ των κόμβων. Στην πραγματικότητα, κάτι τέτοιο δεν είναι απαραίτητο αλλά χωρίς αυτό, κάποιες εφαρμογές MPI (**Message Passing Interface)** δεν θα μπορέσουν να τρέξουν.

- Το MPI είναι ένα πρότυπο αποστολής μηνυμάτων, το οποίο χρησιμοποείται από πολλές αρχιτεκτονικές παράλληλης επεξεργασίας.

- Για αυτό το λόγο, είναι πολύ καλή ιδέα να χρησιμοποιηθεί ένα διαμοιρασμένο σύστημα αρχείων στον cluster. Το NFS (Network File System – **σύστημα αρχείων που επιτρέπει στον χρήστη να έχει πρόσβαση σε αρχεία σε ένα δίκτυο**) είναι το ευκολότερο να χρησιμοποιηθεί, επειδή και ο διακομιστής και ο εξυπηρετητής είναι σε επίπεδο πυρήνα (**το χαμηλότερο επίπεδο κώδικα - θεμέλιο τμήμα ενός λειτουργικού συστήματος**), και ο διαμοιρασμός θα έχει τα εργαλεία για τροποποίηση και παρακολούθηση του NFS.

Αρχιτεκτονική(11/12)

- Η κλασική προσέγγιση NFS σε έναν διαμοιρασμένο κατάλογο είναι να δημιουργήσει μία έξοδο ως κατάλογο από τον κεντρικό κόμβο, προς τους κόμβους υπολογισμού.

- Μπορούν να χρησιμοποιηθούν οποιοιδήποτε κατάλογοι για έξοδο, αλλά πολλές φορές, οι άνθρωποι απλά μοιράζουν τον κατάλογο /home από τον κεντρικό κόμβο, παρόλο που μερικές φορές θα εξάγουν έναν νέο κατάλογο, όπως για παράδειγμα τον /shared. Οι κόμβοι υπολογισμού επίσης τοποθετούν τον διαμοιρασμένο φάκελο ως /home. Για αυτό, εάν κάτι στον /home είναι τοπικό σε κάθε κόμβο, δεν θα είναι διαθέσιμο προς χρήση.

Αρχιτεκτονική(12/12)

- Εννοείται πως μπορούν να υπάρξουν και πιο περίπλοκες ή ίσως πιο καλές προσεγγίσεις, και υπάρχουν λόγοι για αυτό, αλλά ίσως χρειάζεται να υιοθετηθεί ο γενικός κανόνας KISS (Keep it simple Silly). Αυτό σημαίνει ευκολότερη αντιμετώπιση σε προβλήματα αποσφαλμάτωσης και ευκολότερη τροποποίηση στον cluster, εφόσον χρειαστεί. Με την αρχιτεκτονική σταθερή, θα μπορέσουμε να χρησιμοποιήσουμε και το λογισμικό που θέλουμε.

Επεξεργαστές και κόμβοι(1/6)

- Ο επεξεργαστής είναι η κινητήρια μονάδα του cluster. Το κλειδί της καλής απόδοσης, είναι η συνεχής απασχόληση αυτής της κινητήριας μονάδας. Τα παράλληλα προγράμματα διανέμονται συνήθως μεταξύ πολλών κόμβων του cluster.

- Ωστόσο, οι πολλοι-πύρηνοι επεξεργαστές έχουν αλλάξει αυτήν την κατάσταση αρκετά. Οι cluster-κόμβοι μπορούν να έχουν 8 ή ακόμα και 16 πυρήνες ανά κόμβο (για παράδειγμα, ο Sun Fire x4440 server με 4 τετραπύρηνους AMD Opteron επεξεργαστές.)

- Είναι δυνατόν ολόκληρες HPC εφαρμογές να χωρέσουν σε ένα μονό κόμβο cluster. Κάποιες φορές αυτό βοηθάει στην απόδοση του συστήματος και κάποιες φορές την βλάπτει.

Επεξεργαστές και κόμβοι(2/6)

- Η επιλογή του επεξεργαστή είναι πολύ σημαντική, γιατί οι εγκαταστάσεις cluster, πολλές φορές βασίζονται στην κλιμακωτή απόδοση του επεξεργαστή. Οι εξελίξεις στην αρχιτεκτονική x86, όπως η ταυτόχρονη 32/64 διαδικασία των bit, οι ενσωματωμένοι χειριστές της μνήμης και οι τεχνολογίες παρόμοιες με την τεχνολογία AMD HyperTransport, έχουν προωθήσει εμπορικούς επεξεργαστές στο προσκήνιο των HPC.

- Για παράδειγμα, σε μία πρόσφατη έρευνα που έγινε, τα εργαστήρια Lawrence Livermore, Los Alamos, και Sandia National Labs, επέλεξαν cluster τα οποία είναι βασισμένα σε επεξεργαστές AMD Opteron.

- Οι επεξεργαστές Opteron ήταν οι πρώτοι που υποστήριζαν αρχιτεκτονική AMD64 (ουστιαστικά τεχνολογία x86-64). Χρησιμοποιούνται σε διακομιστές και σταθμούς εργασίας και ουσιαστικά είναι το αντίπαλο δέος των Intel Xeon (επεξεργαστές κατάλληλοι για σταθμούς εργασίας, διακομιστές και ενσωματωμένα συστήματα, με προηγμένες λειτουργίες).

Επεξεργαστές και κόμβοι(3/6)

- Ανάλογα με τη σχεδίαση του cluster, οι κόμβοι μπορεί να είναι χοντροί (πολλοί πυρήνες, δίσκοι και μνήμη), λεπτοί (μικρός αριθμός πυρήνων και μνήμης) ή κάτι ενδιάμεσο.

- Μερικές εφαρμογές δουλεύουν και στους δύο τύπους σχεδίασης και μερικές ταιριάζουν σε συγκεκριμένη διαμόρφωση.

- Γενικότερα, είναι σημαντικό να προσεχθεί το μέγεθος της μνήμης. Ουσιαστικά, πολλοί πυρήνες ανά κόμβο σημαίνει περισσότερη μνήμη ανά κόμβο, καθώς κάθε πυρήνας θα πρέπει να τρέξει ένα εντελώς διαφορετικό πρόγραμμα. Πολλοί πυρήνες HPC απαιτούν ένα μεγάλο ποσό μνήμης.

- Αυτό μπορεί να διαπιστωθεί και από την καθημερινή χρήση των υπολογιστών. Αρκεί να ελέγξουμε τις απαιτήσεις μνήμης των εφαρμογών και το μέγεθος των κόμβων κατάλληλα. Όταν γίνεται αναβάθμιση επεξεργαστή από διπύρηνο σε τετραπύρηνο, και δεν αναβαθμιστεί η μνήμη, αυτομάτως μειώνεται το ποσό της μνήμης ανά πυρήνα, κάτι το οποίο μπορεί να οριοθετήσει τα πλεονεκτήματα ενός αναβαθμισμένου επεξεργαστή.

Επεξεργαστές και κόμβοι(4/6)

- Κάτι άλλο το οποίο πρέπει να ληφθεί υπόψιν για τους κόμβους, είναι η διαχείριση του συστήματος. Η έξυπνη διεπαφή διαχείρισης πλατφόρμας (Intelligent Platform Management Interface-IPMI) ορίζει ένα σύνολο από κοινές διεπαφές σε ένα υπολογιστικό σύστημα. Οι διαχειριστές συστήματος μπορούν να χρησιμοποιήσουν το IPMI για να ελέγξουν την κατάσταση του συστήματος και να το διαχειριστούν.

- Συχνά, οι λειτουργίες του IPMI μπορούν να γίνουν διαχειρίσιμες από ένα δίκτυο διεπαφής, κάτι το οποίο είναι πολύ σημαντικό για ένα μεγάλο αριθμό από κόμβους, συχνά τοποθετημένους σε έναν απομακρυσμένο κέντρο δεδομένων.

Επεξεργαστές και κόμβοι(5/6)

- Πίσω στις μέρες των υπερυπολογιστών, υπήρχαν προϊόντα τα οποία ονομαζόντουσαν επεξεργαστές μητρώου, τα οποία ήταν κατασκευασμένα για να κάνουν συγκεκριμένες μαθηματικές λειτουργίες πολύ γρήγορα. Πρόσφατα, υπήρξε μία επαναφορά αυτού του μοντέλου, μέσα από τις Γενικού Σκοπού Γραφικές Επεξεργαστικές Μονάδες (General Purpose Graphical Processing Units), ή αλλιως τις κάρτες γραφικών (video cards).

- Αυτή η τάση δεν είναι νέα στους HPC και δεν είναι καθόλου ασυνήθιστο να χρησιμοποιούνται νέες τεχνολογίες με καινοτόμους τρόπους, προκειμένου να λύθούν περίπλοκα προβλήματα.

Επεξεργαστές και κόμβοι(6/6)

- Όπως και με τους επεξεργαστές μητρώου, έτσι και οι GPU μπορούν να κάνουν συγκεκριμένες μαθηματικές λειτουργίες. Στην περίπτωση των GP-GPU ωστόσο, η δυνατότητα τους να πωληθούν και για εμπορική χρήση, τις καθιστά μία χαμηλού κόστους λύση.

- Οι ήδη υπάρχουσες παράλληλες εφαρμογές χρειάζονται επαναπρογραμματισμό για να χρησιμοποιήσουν τέτοιου είδους συσκευές και για να συμβεί αυτό ήδη έχουν διευθετηθεί λύσεις για το πρόβλημα.

- Για παράδειγμα, το OpenCL, το οποίο είναι μια ανοιχτή και ελεύθερη cross-πλατφόρμα (που χρησιμοποιείται δηλαδή από πολλά και διαφορετικά λογισμικά), είναι μία λύση που ήδη χρησιμοποιείται. Παρόλο που η υιοθέτηση αυτού του τύπου υπολογιστικής έχει αλλάξει σε μερικούς τομείς, είναι δεδομένο πως θα γίνει ένα βασικό εργαλείο στους HPC, με πολλές εταιρείες να δουλεύουν ήδη πάνω σε αυτό για να το καταφέρουν.

Επικοινωνία: Διασυνδέσεις(1/8)

- Προκειμένου να κρατηθεί ένας κόμβος απασχολημένος, μία καλή διασύνδεση (μία θύρα, η οποία προσκολάει μία συσκευή σε μία άλλη) ειναι απαραίτητη. Όπως πάντα, όλα εξαρτώνται από την εφαρμογή που τίθεται σε λειτουργία. Γενικότερα, τα περισσότερα high performance cluster συστήματα, χρησιμοποιούν μία ειδική και γρήγορη διασύνδεση.

- Διασυνδέσεις υψηλής απόδοσης είναι συνήθως βαθμονομημένες από την καθυστέρηση (latency), που είναι ο γρηγορότερος χρόνος στον οποίο ένα μοναδικό byte μπορεί να σταλθεί (μετρημένο σε nanoseconds ή microseconds), και από το εύρος ζώνης (bandwidth), που είναι ο μέγιστος ρυθμός δεδομένων (μετρημένος σε Megabytes ή Gigabytes ανά δευτερόλεπτο). Υπάρχουν άλλοι αριθμοί οι οποίοι βοηθάνε επίσης, όπως το N/2 μέγεθος πακέτου (εξηγείται παρακάτω).

Επικοινωνία: Διασυνδέσεις(2/8)

- Αυτός ο αριθμός (N/2) είναι το μέγεθος του πακέτου, το οποίο φτάνει το μισό της μονής κατεύθυνσης του εύρους ζώνης της διασύνδεσης (μία μέτρηση του πόσο γρήγορα η διεκπεραιωτική ικανότητα μεγαλώνει). Όσο μικρότερος είναι ο αριθμός, τόση περισσότερη ταχύτητα εύρους ζώνης έχουμε από τα μικρά πακέτα.

- Ένας τελευταίος αριθμός που χρήζει προσοχής, είναι ο ρυθμός ανταλλαγής μηνυμάτων. Αυτός ο αριθμός, δείχνει πόσα μηνύματα ανα δευτερόλεπτο μπορεί να στείλει μία διασύνδεση και είναι σημαντικός για πολλοί-πύρηνους κόμβους, καθώς πολλοί πυρήνες πρέπει να μοιραστούν τη διασύνδεση.

Επικοινωνία: Διασυνδέσεις(3/8)

- Παρόλο που οι αριθμοί που αναφέρθηκαν είναι αντιπροσωπευτικοί για να μετρηθεί μία διασύνδεση, η απόλυτη δοκιμασία για τη διασύνδεση ειναι η εκάστοτε εφαρμογή ή οι εφαρμογές. Στις περισσότερες των περιπτώσεων, η εφαρμογή θα επικοινωνήσει μέσω των βιβλιοθηκών MPI (Message Passing Interface). Αυτό το λογισμικό είναι ένα στρώμα επικοινωνίας στην κορυφή του υλικού. Η εισαγωγή του MPI διαφοροποιείται και το πραγματικό τεστ είναι να τρέξουν μερικές συγκρίσεις.

- Οι γρήγορες διασυνδέσεις επίσης απαιτούν εναλλάκτες, έτσι ώστε κάθε κόμβος να μπορεί να επικοινωνήσει με τον άλλο. Ένας τρόπος με τον οποίο οι ακριβοί και καλοί δρομολογητές αξιολογούνται, είναι το εύρος ζώνης διχοτόμησης (bi-section bandwidth). Αυτό βαθμολογεί πόσο καλά ο εναλλάκτης υποστηρίζει πολλές επικοινωνίες ταυτόχρονα. Ένα καλό bi-sectional εύρος ζώνης θα επιτρέψει σε όλους τους κόμβους να επικοινωνούν ταυτόχρονα.

Επικοινωνία: Διασυνδέσεις(4/8)

- Όσον αφορά τη διαθέσιμη τεχνολογία, τα υψήλής απόδοσης δίκτυα υπολογιστικής, αναπτύσονται χρησιμοποιώντας δύο σημαντικές τεχνολογίες: το InfinityBand (IB) και το 10 Gigabit Ethernet. Εάν οι εφαρμογές δεν απαιτούν ένα μεγάλο ποσοστό επικοινωνίας κόμβο σε κόμβο (node-to-node), τότε η standard Gigabit Ethernet τεχνολογία αποτελεί μία καλή λύση (GigE). Βρίσκεται συνήθως στην μητρική πλακέτα και υπάρχουν μεγάλης πυκνότητας/απόδοσης switch ήδη διαθέσιμα.

- Το InfinityBand αναπτύχθηκε επιτυχώς και για τα μεγάλα αλλά και για τα μικρά clusters, χρησιμοποιώντας λεπίδες και 1U διακομιστές. (θα αναλυθούν παρακάτω). Έχει ένα ελεύθερο προς διανομή λογισμικό και θεωρείται από πολλούς ότι βρίσκεται ανάμεσα στις καλύτερες διασυνδέσεις για clustering, λόγω της χαμηλής καθυστέρησης και της υψηλής διεκπεραιωτικής ικανότητας. Ένα άλλο σημείο κλειδί του InfinityBand είναι η διαθεσιμότητα από μεγάλα multi-port switches, όπως το Sun Data Center Switch 3456, το οποίο έχει χαμηλή καθυστέρηση.

- Μία μεγάλη θύρα πυκνότητας επιτρέπει μικρότερη καλωδίωση ανά λιγότερους υπο-εναλλάκτες (sub-switches) όταν δημιουργούνται μεγάλα clusters.

Επικοινωνία: Διασυνδέσεις(5/8)

- Πέρα από αυτές τις δύο τεχνολογίες που βρίσκονται στην κορυφή των προτιμήσεων, υπάρχει και η επιλογή της ιδιόκτητης διασύνδεσης, που σημαίνει ότι μεγάλη κατασκευαστές δημιουργούσαν δικιές του διασυνδέσεις, προκειμένου να συνδέσουν τα δικά τους μηχανήματα. Παραδείγματα εταιρειών που χρησιμοποιούσαν ιδιόκτητες διασυνδέσεις είναι:
  - Η IBM, με το IBM BlueGene και το IBM p775
  - Η Myrinet, η οποία ουσιαστικά δημιούργησε έναν τύπο LAN δικτύου για επικοινωνία μεταξύ των σταθμών εργασίας
  - Η NEC, με το NEC SX-9
  - Το Εθνικό Κέντρο Υπερυπολογιστών της Κίνας, με τον Tianhe-1
  - Η Fujitsu, με το K
  - Η Cray, με την ομώνυμη διασύνδεσή της

[11] https://www.systems.ethz.ch/sites/default/files/file/Spring2013_Courses/AdvCompNetw_Spring2013/13-hpc.pdf

Επικοινωνία: Διασυνδέσεις(6/8) Βασικές αρχές διασυνδέσεων

- Τα δίκτυα συνδέουν κόμβους επεξεργασίας, μνήμες, συσκευές εισόδου / εξόδου, συσκευές αποθήκευσης - Το σύστημα περιλαμβάνει κόμβους δικτύου (όπως switches, routers) και υπολογιστικούς κόμβους.
- Τα δίκτυα διεισδύουν στο σύστημα, σε όλα τα επίπεδα, δημιουργώντας:
  - Δίκτυο μνήμης
  - Δίκτυο μνήμης για σύνδεση πυρήνων σε προσωρινές μνήμες και ελεγκτές μνήμης
  - Δίαυλο PCIe για σύνδεση συσκευών I / O, αποθήκευσης, επιταχυντών
  - Δίκτυο αποθήκευσης για σύνδεση με συστοιχίες δίσκων
  - Τοπικό δίκτυο για διαχείριση και "εξωτερική" συνδεσιμότητα

Επικοινωνία: Διασυνδέσεις(7/8) Βασικές πτυχές σχεδιασμού

- Τοπολογία: Κανόνες που καθορίζουν τον τρόπο με τον οποίο συνδέονται οι υπολογιστικοί κόμβοι και οι κόμβοι του δικτύου. Τα μοτίβα διασύνδεσης μπορούν να περιγραφούν με απλές αλγεβρικές εκφράσεις.

- Δρομολόγηση: Κανόνες που καθορίζουν πώς να πάμε από έναν κόμβο Α σε έναν κόμβο Β. Επειδή οι τοπολογίες είναι κανονικές και γνωστές, οι αλγόριθμοι δρομολόγησης μπορούν να σχεδιαστούν εκ των προτέρων.

- Έλεγχος ροής: Κανόνες που ελέγχουν τη διασταύρωση των συνδέσεων. Σκοπός είναι η αποφυγή της αδράνειας.

Επικοινωνία: Διασυνδέσεις(8/8)

- Η κατανόηση των απαιτήσεων για την επικοινωνία των εφαρμογών, επιτρέπει την βελτιστοποίηση της επίδοσης προς το κόστος (price-to-performance). Ο προϋπολογισμός για τη δημιουργία τέτοιου δικτύου θα πρέπει να είναι ένας ισορροπημένος αριθμός "κόμβων δικτύου".

- Ένα γρήγορο (και ακριβό) δίκτυο, σημαίνει πως μπορεί να χρειαστεί να αγοραστούν λιγότεροι κόμβοι στο μέλλον, κάτι το οποίο σημαίνει γενικότερα χαμηλότερη υπολογιστική ισχύς.

- Ένα αργό (και φθηνότερο) δίκτυο σημαίνει πως μπορεί να χρειαστεί να αγοραστούν περισσότεροι κόμβοι για να αποκτηθεί και μεγαλύτερη υπολογιστική ισχύς.

- Η ισορροπία μεταξύ αυτών των δύο πλευρών πρέπει να είναι βασισμένη στις ανάγκες των εφαρμογών.

Πρωτόκολλα επικοινωνίας(1/6)

- Παραδοσιακά πρωτόκολλα: TCP (Transmission Control Protocol) και UDP (User Datagram Protocol)
- Ειδικά σχεδιασμένα:
  - Ενεργά μηνύματα,
  - VMMC (Virtual Memory Mapped Communication)
  - BIP (Basic Interface for Parallelism)
  - VIA (Virtual Interface Architecture)

Πρωτόκολλα επικοινωνίας TCP και UDP(2/6)

- Πρώτες βιβλιοθήκες ανταλλαγής μηνυμάτων που χρησιμοποιήθηκαν:
  - Το TCP είναι αξιόπιστο
  - Το UDP δεν είναι
- Πλεονεκτήματα:
  - Είναι σταθερά πρωτόκολλα και αρκετά γνωστά
- Μειονεκτήματα:
  - Χρειάζεται συνεχής επίβλεψη (ειδικά για γρήγορα δίκτυα)
  - Πολύ μεγάλη αλληλεπίδραση με το λογισμικό
  - Πολλές παραλλαγές

Πρωτόκολλα επικοινωνίας Ενεργά μηνύματα – Active Messages(3/6)

- Βιβλιοθήκη επικοινωνίας χαμηλής καθυστέρησης.
- Πρωτόκολλο μηδενικής αντιγραφής.
- Τα μηνύματα αντιγράφονται απευθείας, τόσο από και προς στο δίκτυο όσο και από και προς το χώρο της διεύθυνσης του χρήστη.
- Δε χρειάζεται να γίνει κάποια ενέργεια παραλαβής του μηνύματος.

Πρωτόκολλα επικοινωνίας
VMMC - Virtual Memory Mapped Communication (4/6)

- Επικοινωνία χαρτογραφημένης εικονικής μνήμης.
- Προβολή μηνυμάτων ως αναγνωσμένα και εγγραφή τους στη μνήμη (παρόμοια με τη διανεμημένη μνήμη κοινής χρήσης).
- Κάνει αντιστοιχία μεταξύ της πλευράς λήψης και της πλευράς αποστολής της εικονικής σελίδας.

Πρωτόκολλα επικοινωνίας BIP - Basic Interface for Parallelism(5/6)

- Βασική διεπαφή για παραλληλισμό. Χαμηλό επίπεδο στρώματος μηνυμάτων, κατάλληλο για τη διασύνδεση Myrinet. Χρησιμοποιεί αρκετά πρωτόκολλα, ανάλογα με το μέγεθος του μηνύματος

- Προσπαθεί να πετύχει μηδενικές αντιγραφές ενώ χρησιμοποιείται από προγραμματιστές MPI

Πρωτόκολλα επικοινωνίας VIA - Virtual Interface Architecture(6/6)

- Αρχιτεκτονική εικονικής διασύνδεσης.
- Το πρώτο πρότυπο που προωθείται από τη βιομηχανία. Συνδυάζει τα καλύτερα χαρακτηριστικά των ακαδημαϊκών έργων.
- Η διεπαφή είναι σχεδιασμένη για άμεση χρήση από προγραμματιστές. Ωστόσο, πολλοί προγραμματιστές θεωρούν ότι είναι σε πολύ χαμηλό επίπεδο. Για αυτό το λόγο, αναμένονται διεπαφές διασύνδεσης μεγαλύτερου επιπέδου.
- Προτείνεται η χρήση της κάρτας δικτύου μαζί με το πρωτόκολλο VIA.

Πρωτόκολλα αποθήκευσης (1/3)

- Οι δύο βασικές αρχιτεκτονικές αποθήκευσης σήμερα είναι ένα δίκτυο χώρου αποθήκευσης (SAN – Storage Area Network) ή μια λύση προσαρμοσμένης αποθήκευσης δικτύου (NAS – Network Attached Storage).

- Παραδοσιακά, ένα SAN παρέχει Block storage (δηλαδή τα δεδομένα αποθηκεύονται σε μεγάλους όγκους, γνωστούς ως μπλοκ), ενώ ένα NAS παρέχει αποθήκευση σε επίπεδο αρχείου. Αξίζει επίσης να σημειωθεί ότι η εσωτερική αποθήκευση διακομιστών επιστρέφει με την εμφάνιση αποθηκευτικού χώρου που έχει καθοριστεί από το λογισμικό (SDS – Safety Data Sheets Software).

Πρωτόκολλα αποθήκευσης (2/3)

- Τα πέντε πρωτεύοντα πρωτόκολλα αποθήκευσης που υπάρχουν σήμερα στην αγορά είναι:
  - Το Fibre Channel (FC)
  - το Internet Interface Small Computer System (iSCSI)
  - το Fibre Channel Over Ethernet (FCoE)
  - το Network File System (NFS)
  - και το Common Internet File System (CIFS).
- Επιπλέον, ένα άλλο πρότυπο επικοινωνίας – αποθήκευσης, το InfiniBand (IB), ταιριάζει ακόμα περισσότερο σε κέντρα δεδομένων και τους υπερυπολογιστές, λόγω της υψηλής απόδοσής του και της χαμηλής του καθυστέρησης.

Πρωτόκολλα αποθήκευσης (3/3)

Χαρακτηριστικά	Infiniband	Fibre Channel	FCoE	iSCSI
Εύρος Ζώνης (Gbps)	2.5/5/10/14/25/50	8/16/32/128	10/25/40/100	10/25/40/100
Καθυστέρηση προσαρμογέα	25 us	50 us	200 us	Wide range
Καθυστέρηση διακόπτη	100-200 ns	700 ns	200 ns	200 ns
Προσαρμογέας	HCA - host channel adapter	HBA -host bus adapter	CNA - converged network adapter	NIC-network interface card
Μάρκα Διακόπτη	Mellanox, Intel	Cisco, Brocade	Cisco, Brocade	HPE, Cisco, Brocade
Μάρκα Κάρτας Διεπαφής	Mellanox, Intel	Qlogic, Emulex	Qlogic, Emulex	Intel, Qlogic

Πίνακας που συγκρίνει μερικά στοιχεία των πρωτοκόλλων αποθήκευσης

Πρωτόκολλα μεταφοράς δεδομένων (1/2)

- Τα πρωτόκολλα μεταφοράς δεδομένων που μπορούν να αξιοποιηθούν προκειμένου να μεταφερθούν δεδομένα στους υπερυπολογιστές, είναι τα ακόλουθα:
  - Το GridFtp, το οποίο είναι μια επέκταση του τυπικού πρωτοκόλλου μεταφοράς αρχείων (FTP) για υψηλή ταχύτητα, αξιοπιστία και ασφαλή μεταφορά δεδομένων
  - Το sftp, το οποίο κρυπτογραφεί τα δεδομένα προτού τα στείλει στο διαδίκτυο, ενώ επίσης μπορεί να συνεχίσει μεταφορές που έχουν διακοπεί και να διαγράψει απομακρυσμένα αρχεία και φακέλους.
  - Το scp, το οποίο χρησιμοποιεί το Secure Shell (SSH) για τη μεταφορά δεδομένων και χρησιμοποιεί τους ίδιους μηχανισμούς για τον έλεγχο ταυτότητας, διασφαλίζοντας έτσι την αυθεντικότητα και την εμπιστευτικότητα των δεδομένων κατά τη μεταφορά
  - Το Wildcards, όπου χρησιμοποιείται για μεταφορές πολλαπλών αρχείων (όπως για παράδειγμα όλων των αρχείων που είναι επέκτασης .dat
  - Το rsync, το οποίο είναι ένα γρήγορο και εξαιρετικά ευέλικτο εργαλείο αντιγραφής αρχείων. Συνδέει αρχεία και καταλόγους μεταξύ δύο διαφορετικών τοποθεσιών (ή διακομιστών). Το Rsync αντιγράφει μόνο τις διαφορές των αρχείων που έχουν πραγματικά αλλάξει.

Πρωτόκολλα μεταφοράς δεδομένων Εργαλεία μεταφοράς δεδομένων (2/2)

- Μερικά από τα εργαλεία που μας επιτρέπουν να χρησιμοποιήσουμε τα πρωτόκολλα είναι:
  - Για μηχανήματα με λογισμικό Windows:
    - WinSCP – Secure File Transfer Protocol (Sftp)
    - MobaXterm – Για πρωτόκολλα SSH, SFTP, RDP
    - FileZilla – Για πρωτόκολλα FTP και SFTP
    - PSFTP – Secured FTP του εργαλείου Putty
  - Για μηχανήματα με λογισμικό Mac:
    - Cyberduck
    - FileZilla
  - Για μηχανήματα με λογισμικό Linux:
    - Linux Cyberduck με OpenSSH, άλλα εργαλεία που είναι συμβατά με scp

Πρωτόκολλα διαχείρισης (1/4)

- Δύο από τα πρωτόκολλα διαχείρισης των clusters είναι το Ευφυές πρωτόκολλο διεπαφής διαχείρισης πλατφόρμας (IMPI - Intelligent Platform Management Interface Protocol) και το απλό πρωτόκολλο διαχείρισης δικτύου (SNMP - Simple Network Management Protocol)

- Τα πρωτόκολλα αυτά επιτρέπουν στους διακομιστές που είναι τοποθετημένοι εξ αποστάσεως να ενεργοποιούνται, να απενεργοποιούνται και να επανεκκινούνται.

Πρωτόκολλα διαχείρισης (2/4)

- Ένα άλλο πρωτόκολλο διαχείρισης είναι το SSH.

- Το **SSH** (Secure Shell) είναι ένα ασφαλές δικτυακό πρωτόκολλο, το οποίο επιτρέπει τη μεταφορά δεδομένων μεταξύ δύο υπολογιστών. Το SSH όχι μόνο κρυπτογραφεί τα δεδομένα που ανταλλάσσονται κατά τη συνεδρία, αλλά προσφέρει ένα ασφαλές σύστημα αναγνώρισης καθώς και άλλα χαρακτηριστικά όπως ασφαλή μεταφορά αρχείων (SSH File Transfer Protocol, SFTP), κλπ.

- Το SSH παρέχει ένα ασφαλές κανάλι μέσω ενός μη ασφαλούς δικτύου σε μια αρχιτεκτονική πελάτη-διακομιστή, συνδέοντας μια εφαρμογή πελάτη SSH με ένα διακομιστή SSH. Η προδιαγραφή πρωτοκόλλου διακρίνει δύο μεγάλες εκδόσεις, που αναφέρονται ως SSH-1 και SSH-2. Η τυπική θύρα TCP για SSH είναι 22. Το SSH χρησιμοποιείται αρκετά για πρόσβαση σε συστήματα που είναι βασισμένα στο Unix, όπως το Linux, το λογισμικό των HPC clusters.

Πρωτόκολλα διαχείρισης (3/4)

- Επιπροσθέτως, ένα άλλο σημαντικό πρωτόκολλο διαχείρισης είναι το Border Gateway Protocol (BGP).

- Το Border Gateway Protocol (BGP) είναι ένα τυποποιημένο πρωτόκολλο εξωτερικής δρομολόγησης που επιτρέπει την δρομολόγηση πακέτων και την ανταλλαγή πληροφοριών προσβασιμότητας μεταξύ αυτόνομων συστημάτων (AS) στο διαδίκτυο.

- **Αυτόνομο Σύστημα** είναι ένα σύνολο συνδεδεμένων Internet Protocol (IP) δικτύων υπό τον έλεγχο ενός ή περισσότερων διαχειριστών του δικτύου που παρουσιάζει μια κοινή, σαφώς καθορισμένη πολιτική δρομολόγησης στο διαδίκτυο.

Πρωτόκολλα διαχείρισης (4/4)

- Το BGP ανήκει στην κατηγορία των πρωτοκόλλων διανύσματος μονοπατιού (Path Vector) και οι αποφάσεις δρομολόγησης βασίζονται στα διαθέσιμα μονοπάτια δρομολόγησης, στις πολιτικές που ακολουθούνται από κάθε Αυτόνομο Σύστημα καθώς και τους κανόνες που εφαρμόζονται τοπικά από τους διαχειριστές κάθε αυτόνομου συστήματος για τη διαχείριση της εισερχόμενης και εξερχόμενης ροής δικτύου. Το πρωτόκολλο BGP τρέχει επάνω από το TCP, και έτσι κατατάσσεται στα πρωτόκολλα επιπέδου εφαρμογής.

- Αυτό σημαίνει πως ουσιαστικά το λογισμικό που υλοποιεί το BGP λαμβάνει αποφάσεις δρομολόγησης στο επίπεδο δικτύου αλλά χρησιμοποιείται και για την κατασκευή των πινάκων δρομολόγησης (routing tables) που στη συνέχεια θα χρησιμοποιήσουν οι δρομολογητές (routers) για την δρομολόγηση του δικτυακού φορτίου.

- Το BGP μονοπωλεί το Διαδίκτυο όσον αφορά τα εξωτερικά πρωτόκολλα δρομολόγησης. Η μεγάλη αποδοχή του BGP οφείλεται στην επεκτασιμότητα (scalability) του για μεγάλο αριθμό δικτυακών πληροφοριών και στην υποστήριξη πολιτικών δρομολόγησης (routing policies). Το BGP επιτρέπει σε ένα αυτόνομο σύστημα να αναγγείλει την ύπαρξή του στο υπόλοιπο Internet.

Συστήματα αρχείων (1/9)

- Η αποθήκευση είναι συχνά ένας ξεχασμένος τομέας των HPC cluster. Σχεδόν όλοι οι clusters απαιτούν κάποια μορφή υψηλής απόδοσης αποθήκευσης. Η απλούστερη μέθοδος είναι να χρησιμοποιηθεί ένα ανώτερός κόμβος σαν NFS (Network File System) server.

- Ακόμα και αυτή η απλή λύση απαιτεί ότι ο κεντρικός κόμβος έχει κάποιου είδους RAID τεχνολογία υποσυστήματος (**Redundant Array of Independent Disks –** αποτελεί τεχνολογία εικονοποίησης και αποθήκευσης δεδομένων). Συχνά, το NFS δεν χρειάζεται να κάνει κλιμάκωση σε μεγάλο αριθμό κόμβων. Για αυτόν το λόγο, υπάρχουν εναλλακτικοί μέθοδοι σχεδιασμού αποθήκευσης διαθέσιμοι για τους clusters.

- Οι περισσότεροι σχεδιασμοί είναι βασισμένοι σε ένα παράλληλο σύστημα αρχείων το οποίο ορίζει τον τύπο αποθήκευσης του υλικού που θα απαιτηθεί. Ένα καλό παράδειγμα είναι το Lustre παράλληλο σύστημα αρχείων. Το Lustre έχει δοκιμαστεί σε λύσεις λογισμικού ανοιχτού κώδικα, οι οποίες προσφέρουν κλιμακωτή είσοδο/έξοδο στους clusters.

Συστήματα αρχείων (2/9)

- Οι εφαρμογές HPC δημιουργούν μεγάλα ποσά δεδομένων και επιβεβαιώνουν την ανάγκη για ένα καλό και αξιόπιστο σύστημα αρχειοθέτησης, το οποίο πρέπει να είναι διαθέσιμο στα κέντρα δεδομένων.

- Χρησιμοποιώντας για τη μεταφορά δεδομένων τις ήδη δοκιμασμένες τεχνολογίες αντιγράφων ασφαλείας (όπως τις μαγνητικές ταινίες), έχει γίνει ήδη ένα σημαντικό βήμα για να προστατευθεί μία HPC επένδυση.

- Ένα καλό σύστημα αρχειοθέτησης, θα μεταφέρει αυτόματα δεδομένα από μία αποθηκευτική συσκευή σε μία άλλη, βασισμένη στις πολιτικές που έχουν οριστεί από τον χρήστη.

Συστήματα αρχείων (3/9)

- Μία άλλη περιοχή που πρέπει να προσεχθεί είναι η Flash αποθήκευση. Τα τμήματα Flash μπορούν να είναι ενσωματωμένα απευθείας στις μητρικές, όπως στην περίπτωση της Sun Blade X6240, ή ενσωματωμένα χρησιμοποιώντας το δίαυλο PCIe (PCI-Express), ο οποίος χρησιμεύει στην ένωση πρόσθετων καρτών μονάδων με την ομάδα των κεντρικών ολοκληρωμένων κυκλωμάτων στην μητρική πλακέτα .

- Χρειάζεται να προσεχθεί ότι ένα ανεπαρκές υποσύστημα αποθήκευσης μπορεί να επιβραδύνει τον cluster, ίσως και περισσότερο, από μία αργή διασύνδεση.

- Το NFS σύστημα δεν είναι ικανό να λύσεις όλες τις ανάγκες αποθήκευσης στους υπερυπολογιστές.

Συστήματα αρχείων(4/9)

- Σχεδόν όλοι οι clusters χρησιμοποιούν το συμβατικό NFS σύστημα αρχείων για να μοιραστούν τις πληροφορίες ανάμεσα στους κόμβους.

- Αποτελεί μία καλή λύση, ωστόσο, το NFS δεν σχεδιάστηκε για παράλληλη πρόσβαση αρχείου (για παράδειγμα, πολλαπλες επεξεργασίες διαβάσματος και καλωδίωσης στο ίδιο αρχείο).

- Αυτός ο περιορισμός μπλόκαρε τα HPC συστήματα. Για αυτό το λόγο, ξεκίνησε η ανάπτυξη των παράλληλων συστημάτων αρχείου.

Συστήματα αρχείων(5/9)

- Μία από τις περιοχές όπου η ανοιχτή προσέγγιση GNU/Linux εξυπηρέτησε την κοινότητα HPC, είναι τα συστήματα αρχείου.
- Υπάρχει μία πληθώρα επιλογών, από τις οποίες όλες εξαρτώνται από τις απαιτήσεις της εφαρμογής.
- Τα συστήματα αρχειών HPC συχνά καλούνται παράλληλα συστήματα αρχείου, επειδή επιτρέπουν μία μέση (πολλή-κομβική) είσοδο και έξοδο.

Συστήματα αρχείων(6/9)

- Αντί να συγκεντρώνουν όλο το χώρο αποθήκευσης σε μία μονή συσκευή, τα παράλληλα συστήματα αρχείων απλώνουν όλο το φόρτο ανάμεσα σε πολλές και ξεχωριστές συσκευές αποθήκευσης.

- Τα παράλληλα συστήματα αρχείων καλούνται συχνά "σχεδιασμένα", καθώς χρειάζεται να είναι αντιστοιχησμένα με ένα συγκεκριμένο cluster.

- Ένα παράδειγμα δημοφιλούς και ελεύθερου παράλληλου συστήματος αρχείων είναι το Lustre από την Sun Microsystems. Το Lustre είναι ένα παλιό, μεγάλων επιδόσεων παράλληλο σύστημα αρχείων.

Συστήματα αρχείων(7/9)

- Άλλες επιλογές περιλαμβάνουν το PVFS2, το οποίο είναι σχεδιασμένο να δουλεύει με το MPI. Τα συστήματα αρχείων cluster καλύπτουν μία πολύ μεγάλη περιοχή.

- Επιπρόσθετα με τις μεγάλες σειρές δεδομένων εισόδου και σημείων ελέγχου, πολλές εφαρμογές HPC έχουν ως εξοδο μεγάλες σειρές δεδομένων εξόδου δεδομένα εξόδου τα οποία αργότερα οπτικοποιούνται σε ειδικά συστήματα.

- Κάτι το οποίο χρήζει προσοχής είναι το pNFS (NFS έκδοση 4.1), το οποίο είναι κατασκευασμένο για παράλληλη NFS πρόσβαση.

- Πολλά από τα υπάρχοντα συστήματα αρχείων επιχειρούν να υποστηρίξουν το χαρακτηριστικό pNFS και να φέρουν μία τυποποίηση στην περιοχή των παράλληλων συστημάτων αρχείων.

Συστήματα αρχείων(8/9)

- Ένα άλλο πρότυπο, το ZFS, ένα σύστημα αρχείων το οποίο κατασκευάστηκε από τη Sun, προσφέρει μερικές συναρπαστικές προοπτικές βελτίωσης της λειτουργίας ενός HPC , καθώς είναι το πρώτο 128 bit σύστημα αρχείων.

- Με πολλές προηγμένες λειτουργίες (όπου με 128 bit δεν υπάρχει περίπτωση να τίθεται θέμα με τα όρια του αποθηκευτικού χώρου), το ZFS είναι η προτεινόμενη διόρθωση και η λύση για τη σιωπηλή καταστροφή αρχείων.

Συστήματα αρχείων(9/9)

- Τα συνηθισμένα συστήματα αρχείων δεδομένων, μπορεί να έχουν ένα ανώτατο όριο για το μέγεθος του αρχείου, τον αριθμό των αρχείων ή το συνολικό χώρο αποθήκευσης. Τα συστήματα αρχείων που χρησιμοποιούνται στους υπερυπολογιστές έχουν τη δυνατότητα να επεκτείνονται, να μεταφέρουν γρήγορα μεγάλο όγκο δεδομένων και να είναι προσβάσιμα ταυτόχρονα από όλες τις νησίδες κόμβων.

- Για αυτό το λόγο, δημιουργήθηκε και το General Parallel File System (GPFS) της IBM, το οποίο μπορεί να προσφέρει PetaBytes (1.000.000 Gigabytes) αποθηκευτικού χώρου στους χρήστες του.

Racking and stacking(1/7)

- Οι κόμβοι υπολογιστών μπορούν να πάρουν τη μορφή των 1U διακομιστών ή των συστημάτων blade. Η απόφαση δεν είναι και η ευκολότερη.

- Τα blade συστήματα έχουν συχνότερα υψηλότερο κόστος απόκτησης αλλά προσφέρουν πολύ καλύτερη διαχείριση, καλύτερη πυκνότητα και μερική μειώση ενέργειας και ψύξης.

- Ανεξαρτήτως από την επιλογή που θα γίνει στο στήσιμο των κόμβων, πρέπει να είναι σίγουρο πως όλος ο εξοπλισμός του cluster είναι "rack mountable", δηλαδή μπορεί να τοποθετηθεί σε ράφια του rack.

Racking and stacking(2/7)

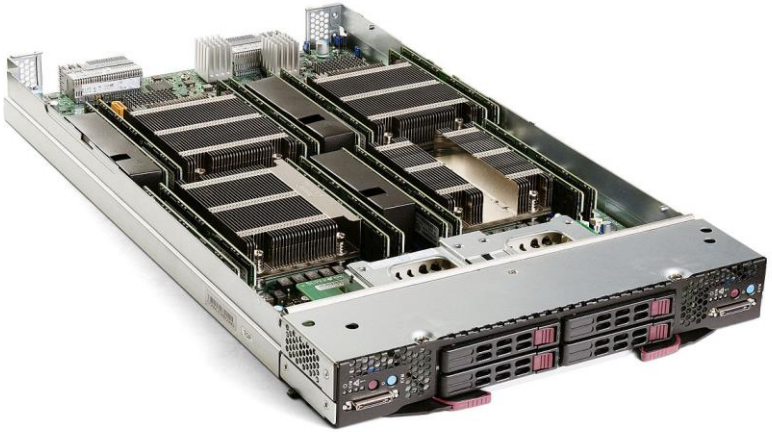
- Το κλασσικό rack σασί μπορεί να φιλοξενήσει εξοπλισμό που ισοδυναμεί με 42U(1U ισοδυναμεί με 1.75 ίντσες ή 44.45 χιλιοστά). Το U, ή αλλιώς RU, αποτελεί την μονάδα μέτρησης των racks, των εξυπηρετητών και των σκληρών δίσκων. Επομένως, ένα 42U rack έχει ύψος 42 x 44,45 = 1.8669 χιλιοστά = 1,866 μέτρα.

- Οι εξυπηρετητές που βρίσκονται σε ράφια των racks και άλλες συσκευές που κατασκευάζονται για να τοποθετηθούν στα ράφια, κατασκευάζονται σε πολλαπλάσια των 1.75 ιντσών και καθορίζονται ως πολλαπλές μονάδες από racks, δηλαδή 1U, 2U, 4U, 7U.

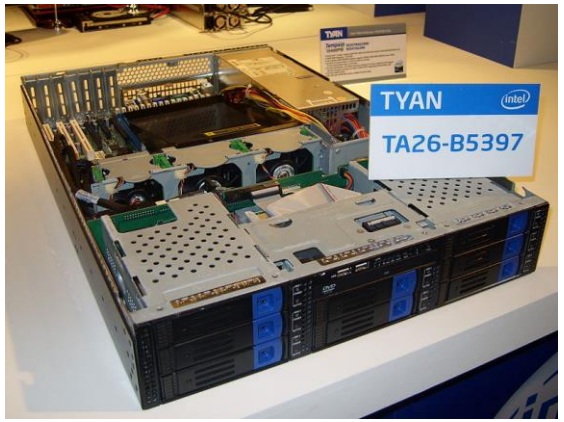
- Τα racks κατασκευάζονται για να μπορούν να κρατήσουν τα μεγέθη που προαναφέρθηκαν. Οι μονάδες μέτρησης rack για εξοπλισμό, θεωρούνται οι μέγιστες διαστάσεις, δηλαδή η μέγιστη διάσταση για 1U είναι 1.75 ίντσες, για 2U 3.5 ίντσες κ.ο.κ. Στην πράξη, πολλές συσκευές κατασκευάζονται σε λίγο μικρότερες διαστάσεις από το καθορισμένο U, προκειμένου να υπάρχει λίγος διαθέσιμος χώρος. Για παράδειγμα, μία συσκευή που είναι 2U, μπορεί στην πραγματικότητα να είναι 3.44 ίντσες και όχι 3.5 ίντσες, όπου είναι το πολλαπλάσιο του 1.75 ίντσες (1U). [10]

[10] https://whatis.techtarget.com/definition/rack-unit

Racking and stacking(3/7)



1U Server



2U Server

4U Server

7U Server

Racking and stacking(4/7)

27U rack
(1,2 μέτρα)

42U Server + 42u rack
(1,866 μέτρα)

Racking and stacking(5/7)

- Είναι πολύ χρήσιμο να χαρτογραφηθεί ο χώρος του σασί του rack για όλο τον εξοπλισμό, με σκοπό να περισσεύει χώρος για μελλοντικές αναβαθμίσεις.

- Πρέπει να είναι κατανοητό το γεγονός ότι όσο καλύτερη είναι η χωροθέτηση του σασί του rack, τόσο καλύτερη θα είναι και η δυνατότητα ψύξης του, κάτι το οποίο είναι ιδιαίτερα σημαντικό για τη συνολική λειτουργία ενός HPC Cluster.

Racking and stacking(6/7)

- Σε κάθε περίπτωση, μία ομάδα από απλούς διακομιστές, μπορεί να τοποθετηθεί σε rack ή blade που έχουν μορφή 1U και να συνδεθούν χρησιμοποιώντας ένα τοπικό εναλλάκτη Ethernet. (Μεταγωγέας όπου αποτελείται από ένα συνδυασμό του επαναλήπτη (Hub) και της γέφυρας (bridge)).

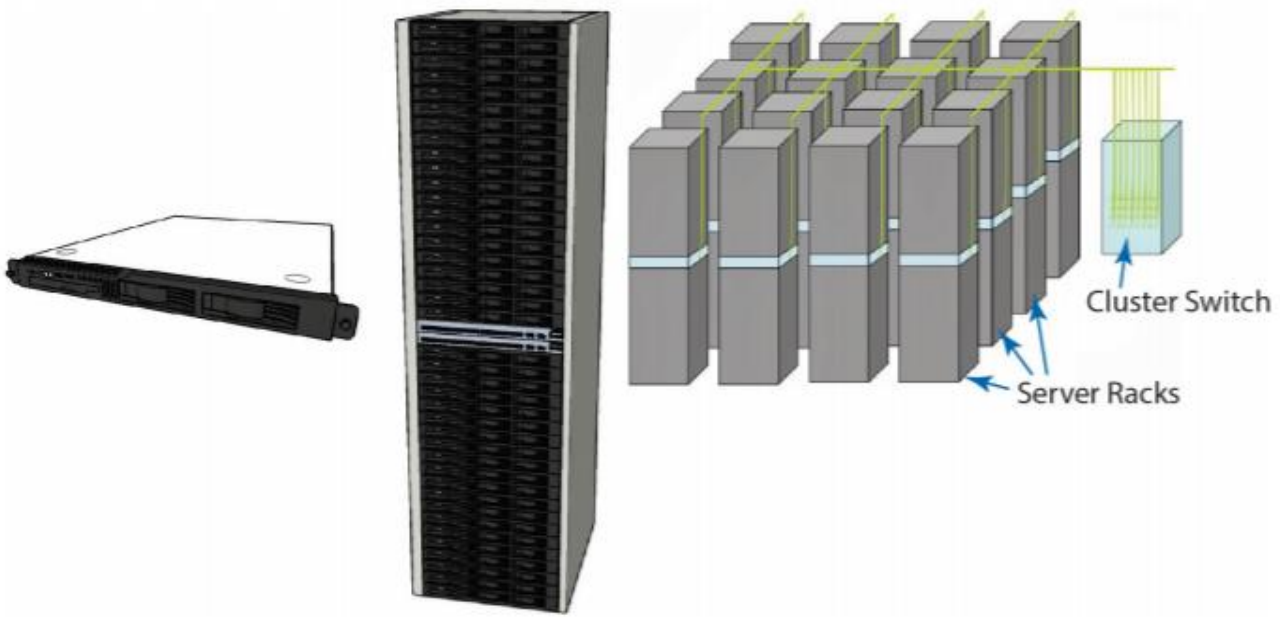
- Αυτά τα switch, τα οποία μπορούν να χρησιμοποιήσουν 1 ή 10 Gbps γραμμές, έχουν έναν αριθμό από uplink συνδέσεις σε έναν ή περισσότερους– επιπέδου cluster – εναλλάκτες Ethernet.

- Αυτή η δεύτερου επιπέδου σύνδεση μπορεί πιθανότατα να προσφέρει επέκταση σε πάνω από 10 χιλιάδες ξεχωριστούς διακομιστές.

Racking and stacking(7/7)

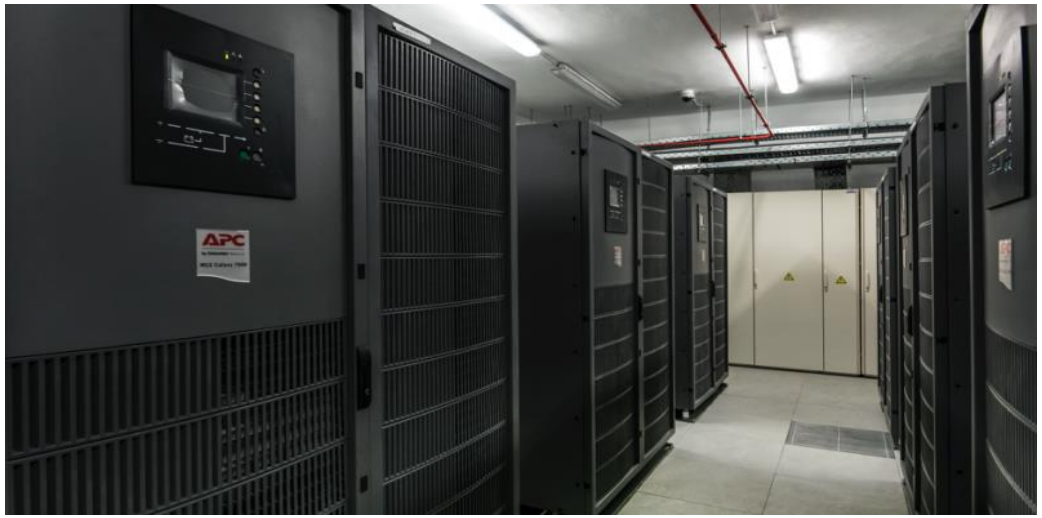
Στο διπλανό σχήμα, παρατηρούμε μερικά τυπικά κομμάτια ενός συστήματος HPC:

- Στα αριστέρα της εικόνας εμφανίζεται ένας server 1U μορφής.
- Στη μέση, ένα ράφι με εναλλάκτη Ethernet.
- Τέλος, δεξιά της εικόνας, απεικονίζεται η εικόνα ενός datacenter HPC, αποτελούμενο από clusters και συνδεδεμένους με Ethernet switch/router επιπέδου cluster.



Παραδείγματα HPC Datacenter (Εικόνες 1/2)

Κέντρο δεδομένων της Microsoft στο Boydton της Virginia



Κέντρο Δεδομένων ΕΔΕΤ (Εθνικό Δίκτυο Έρευνας και Τεχνολογίας) στον ποταμό Λούρο

Παραδείγματα HPC Datacenter (Εικόνες 2/2)

Κέντρο δεδομένων της Facebook στο Pineville

Κέντρο δεδομένων της Switch, του υπερ-υπολογιστή SUPERNAP στο SUPERNAP Campus, στο Las Vegas

Ενέργεια και ψύξη(1/6)

- Η κατανάλωση ενέργειας και η ψύξη έχουν γίνει μία από μία διαδικασία η οποία παραβλεπόταν και παραμεριζόταν αρκετά, σε έναν παράγοντα ο οποίος είναι αρκετά κρίσιμος κατά τη διάρκεια της κατασκευής ενός cluster.
- Ένας γενικός κανόνας για την πρόβλεψη του κόστους της ενέργειας και ψύξης, είναι ότι το ετήσιο κόστος για να κρατηθεί ένας cluster ενεργός και ψυχώμενος ισούται μετά βίας με το 1/3 της τιμής ολόκληρου του cluster.
- Για αυτό το λόγο, η αναζήτηση μίας πράσινης λύσης, κάνει ταυτόχρονα και περιβαλλοντική αλλά και οικονομική αίσθηση.

Ενέργεια και ψύξη(2/6)

- Για παράδειγμα, επιλέγοντας ένα cluster ο οποίος χρησιμοποιεί 45nm Quad-Core επεξεργαστή, και έχει κατανάλωση που ισοδυναμεί με 55-watt ACP (Average CPU Power) επίδοση, κάνουμε μία καλή επιλογή.

- Διαιρώντας με έως και 20 τοις εκατό μικρότερη αδρανή ενέργεια σε σχέση με παρόμοια υπολογιστικά συστήματα, το καθιστά μία πολύ καλή κίνηση.

- Όσον αφορά την ψύξη του συστήματος, τα τελευταία χρόνια η ψύξη με υγρό έχει εξελιχθεί από καινοτομία, σε κάτι κοινότυπο για τα HPC.

Ενέργεια και ψύξη(3/6)

- Υπάρχουν αρκετοί λόγοι που συμβαίνει κάτι τέτοιο:
  - Υπάρχει υποστήριξη για μεγαλύτερες CPU, με σκοπό την καλύτερη υπολογιστική ισχύ
  - Τρομερή πυκνότητα
  - Μειωμένος θόρυβος
  - Μειωμένη ισχύς άρα και μειωμένα έξοδα.

- Καθώς τα HPC προοδεύουν προς την υπολογιστική exascale (τουλάχιστον ένα exaFLOP), αυτοί οι παράγοντες θα συνεχίσουν να επηρεάζουν την διαδικασία δημιουργίας του cluster.

Ενέργεια και ψύξη(4/6)

- Δυστυχώς, δεν μπορεί όλος ο εξοπλισμός του κέντρου δεδομένων να ψυχθεί με το νερό, οπότε αρκετές γνωστές εταιρείες όπως η LRZ και η Lenovo, σε συνεργασία με την Intel, βρίσκονται ήδη στη διαδικασία εναλλακτικής λύσης.

- Η διαδικασία αυτή περιλαμβάνει την μετατροπή του "άχρηστου" ζεστού νερού σε κρύο νερό, το οποίο μπορεί να ξαναχρησιμοποιηθεί για να βοηθήσει στην ελάττωση της θερμοκρασίας όλου του κέντρου δεδομένων.

- Αυτή η διαδικασία, "απορροφά" το υγρό σε ένα ψυγείο, το οποίο παίρνει το ζεστό νερό, από 100 ράφια για παράδειγμα, και το περνάει μέσα από φύλλα με ένα ειδικό υγρό από διοξείδιου του πυριτίου, το οποίο εξατμίζει το νερό, ψυχραίνοντάς το.

Ενέργεια και ψύξη(5/6)

- Από εκεί, το εξατμισμένο νερό πηγαίνει μέσα σε ένα υγρό, το οποίο είναι είτε συνδεδεμένο με τα racks του HPC, είτε βρίσκονται σε έναν πισινό εναλλάκτη θερμότητας για racks αποθήκευσης και διαδικτυακού εξοπλισμού, τα οποία δεν ψύχονται από το νερό.

- Αυτή η διαδικασία είναι ικανή να προσφέρει περισσότερο κρύο νερό από ότι το ίδιο το κέντρο δεδομένων μπορεί στην πραγματικότητα να καταναλώσει. Αυτή η προσέγγιση στα κέντρα δεδομένων είναι πιθανή, καθώς το νερό το οποίο μεταφέρεται από τα ψυγεία είναι αρκετά ζεστό για να κάνει τη διαδικασία να κυλήσει ομαλά.

- Η στενή σύνδεση και η ανεξαρτησία μεταξύ του εξοπλισμού του διακομιστή και της δομής του κέντρου δεδομένων έχει υψηλές προοπτικές.

Ενέργεια και ψύξη(6/6)

- Τα ψυχώμενα μέσω υγρού HPC (iquid-cooled) είναι πλέον μία σταθερή εναλλακτική στα παραδοσιακά ψυχώμενα μέσω αέρα (air-cooled) συστήματα.

- Καθώς ο πυρήνας λειτουργεί, η απαγωγή θερμότητας και η κατανάλωση ενέργειας ολοένα και αυξάνονται, οπότε οι περιστάσεις θα αναγκάσουν τους ιδιοκτήτες HPC να κάνουν την αλλαγή προς το liquid cooling.

- Η στροφή προς αυτή την τεχνολογία, μπορεί να εξοικονομήσει επιπλέον χώρο στα ράφια, ενέργεια, χρήματα αλλά ταυτόχρονα μπορεί να επιτρέψει στο υπολογιστικό σύστημα να αποδίδει το ίδιο καλά, με την ίδια υπολογιστική ισχύ, ακόμα για τις πιο απαιτητικές εργασίες για τα HPC.

Γεννήτριες ενέργειας Μονάδες διανομής ενέργειας(1/8)

- Οι σύγχρονοι clusters απαιτούν σημαντικά ποσά ενέργειας για να λειτουργήσουν. Κάθε rack πολλές φορές συνδέεται με τα λεγόμενα PDUs (Power Distribution Units – Μονάδες διανομής ενέργειας), τα οποία προσφέρουν ενέργεια σε έναν ή περισσότερους κόμβους. Κάθε μονάδα PDU, είναι σημαντική για την παροχή ενέργειας στους διακομιστές.
- Κανονικά, ένα ή περισσότερα PDUs εγκαθίστανται στο πίσω μέρος του rack, πίσω από την υποδομή τοποθέτησης των κόμβων και έτσι δεν επηρεάζουν το διαθέσιμο χώρο του cluster και επιτρέπουν την τοποθέτηση και άλλων εξαρτημάτων. Επιπλέον, οι διαχειριστές των cluster μπορούν να χρησιμοποιήσουν απομακρυσμένα τα PDUs, προκειμένου να απενεργοποιούν κάθε σύστημα στον cluster, εφόσον χρειαστεί ή υπάρχει πρόβλημα.

Γεννήτριες ενέργειας
Μονάδες διανομής ενέργειας(2/8)

- Επιπροσθέτως, υπάρχουν PDUs τα οποία επιτρέπουν στους διαχειριστές των clusters να μπορούν να ελέγξουν την κατάσταση των κυκλωμάτων και το φορτίο τους (Metered PDUs). Τα περισσότερα racks, διαθέτουν τριφασικά PDUs, στα 208Volts (30 Ampere).

- Η ολοένα και αυξανόμενη ανάγκη για υπολογιστή ισχύ και οι περιορισμοί στο φυσικό χώρο, έχουν οδηγήσει σε πιο πυκνά συσκευασμένα racks. Και καθώς οι αριθμοί των διακομιστών που βρίσκονται στα ράφια, οι blade εξυπηρετητές, τα switches διαδικτύου και οι δρομολογητές αυξάνονται, υπάρχει ανάγκη για περισσότερη ισχύ.

- Συνήθως, τα PDUs συνδέονται κατευθείαν με εξόδους στον τοίχο, κάτω από το πάτωμα ή πάνω από την υποδομή των rack. Αξίζει να σημειωθεί επίσης, πως μπορούν να συνδεθούν με μονάδες UPS (Uninterruptible Power Supply – Αδιάκοπης Παροχής Ενέργειας).

Γεννήτριες ενέργειας Συστήματα Αδιάκοπης Παροχής Ενέργειας(3/8)

- Σίγουρα είναι γνωστή η χρήση των UPS, καθώς τροφοδοτούν τις συσκευές με αδιάκοπη τροφοδοσία ρεύματος σε περίπτωση κάποιου προβλήματος. Μια μονάδα UPS είναι απαραίτητη για να διασφαλιστεί ότι δεν θα υπάρξει διακοπή ρεύματος, σε περίπτωση βλάβης του ηλεκτρικού δικτύου ή κάποιου άλλου προβλήματος. Συνήθως, ένα UPS μπορεί από μόνο του να διατηρήσει ένα σύστημα για 30 λεπτά, αλλά άμα χρησιμοποιηθεί συνδυασμός UPS ή μαζί με μία γεννήτρια εγκαταστάσεων, μπορεί να συνεχίσει να τροφοδοτεί τα συστήματα σε παρατεταμένες διακοπές ρεύματος.

- Είναι πολύ σημαντικό να προστατευτούν κρίσιμα μέρη ενός συστήματος, όπως κύριοι κόμβοι, αποθηκευτικά μέσα, επεξεργαστές, με τη χρήση UPS, καθώς έτσι εξασφαλίζεται η προστασία από ξαφνικές διακοπές ρεύματος (ή ολική απώλεια ισχύος), χαμηλές τάσεις, οι οποίες μπορούν να προκαλέσουν καταστροφές σε εξαρτήματα.

Γεννήτριες ενέργειας Συστήματα Αδιάκοπης Παροχής Ενέργειας(4/8)

- Σε περίπτωση παρατεταμένης διακοπής ρεύματος, το λογισμικό παρακολούθησης μπορεί είτε να περιορίσει τη λειτουργία των εξαρτημάτων σε χαμηλότερη ισχύ, προκειμένου να παραταθεί ο χρόνος παροχής ενέργειας των UPS, είτε να γίνει ένας κανονικός τερματισμός λειτουργίας των κόμβων, προκειμένου να μην υπάρξει ζημιά στα εξαρτήματα και να διευκολυνθεί η διαδικασία επανεκκίνησης του συστήματος αργότερα.

- Σημαντική είναι επίσης η προσθήκη επιπλέον μπαταριών στο σύστημα UPS, προκειμένου να παραταθεί ο χρόνος τροφοδοσίας το σύστημα. Ένα κοινό UPS για τους clusters, μπορεί να παρέχει ισχύ ίση με 6000VA, 8000VA, 10000VA, 20000VA (Volt-Ampere, δεν είναι ίδια με τα Watt) και τάση στα 230Volt. Προφανώς και η ισχύς αυτή είναι πολύ μεγαλύτερη από αυτή των κοινών UPS για του επιτραπέζιους υπολογιστές (που ξεκινάνε από 200VA), καθώς οι clusters έχουν τεράστιες απαιτήσεις σε ισχύ.

Γεννήτριες ενέργειας Συστήματα Αδιάκοπης Παροχής Ενέργειας(5/8)

- Χρησιμοποιώντας όμως τα UPS, τίθεται ένα σημαντικό ζήτημα εξοικονόμησης ενέργειας και χρημάτων. Από τη στιγμή που οι μονάδες UPS έχουν ένα σημαντικό κόστος αγοράς, δεν είναι και πολύ έξυπνη η αδιάκοπη χρήση τους, χωρίς την παροχή ενέργειας.

- Μία πιο έξυπνη τακτική θα ήταν η χρήση συσκευών μετατροπής ισχύος (power inventers).
  - Ένας μετατροπέας ισχύος με δυνατότητα αποθήκευσης ενέργειας, μπορεί να χρησιμοποιηθεί ως μία απευθείας πηγή ενέργειας για λιγότερο σημαντικές παροχές ενέργειας, όπως ο φωτισμός.
  - Τα συστήματα UPS μπορούν να μείνουν συνδεδεμένα κατά τη διάρκεια μίας εκτενούς διακοπής ρεύματος, φορτίζοντας τις μπαταρίες τους με την έξοδο που προσφέρουν οι μετατροπείς ισχύος.

Γεννήτριες ενέργειας Συστήματα Αδιάκοπης Παροχής Ενέργειας(6/8)

- Οι μετατροπείς ισχύος, με τις ενσωματωμένες μπαταρίες συνεχούς ρεύματος (direct current – dc) προσφέρουν αδιάκοπη παροχή ενέργειας.

- Μετατρέπουν το συνεχές ρεύμα σε εναλλασσόμενο ρεύμα (dc to ac). Έχουν ως είσοδο 12 Volt συνεχούς ρεύματος και το μετατρέπουν σε εναλλασσόμενο ρεύμα τάσης 220 Volt – 230 Volt.

Εικόνα που δείχνει έναν μετατροπέα ισχύος για UPS με μπαταρία, 12V dc σε 220V ac

Γεννήτριες ενέργειας (Εικόνες 7/8)

Εικόνα με PDUs τοποθετημένα πάνω στο rack του server

Εικόνα με PDUs τοποθετημένα πάνω στο rack του server

Γεννήτριες ενέργειας (Εικόνες 8/8)

Εικόνες με την εμπρός και την πίσω όψη ενός UPS της εταιρείας ARC προδιαγραφών 2000VA, 100 - 127V, μεγέθους 4U

Εικόνες ενός UPS της εταιρείας APC, προδιαγραφών 5000VA, 230V, το οποίο είναι τοποθετημένο στο rack

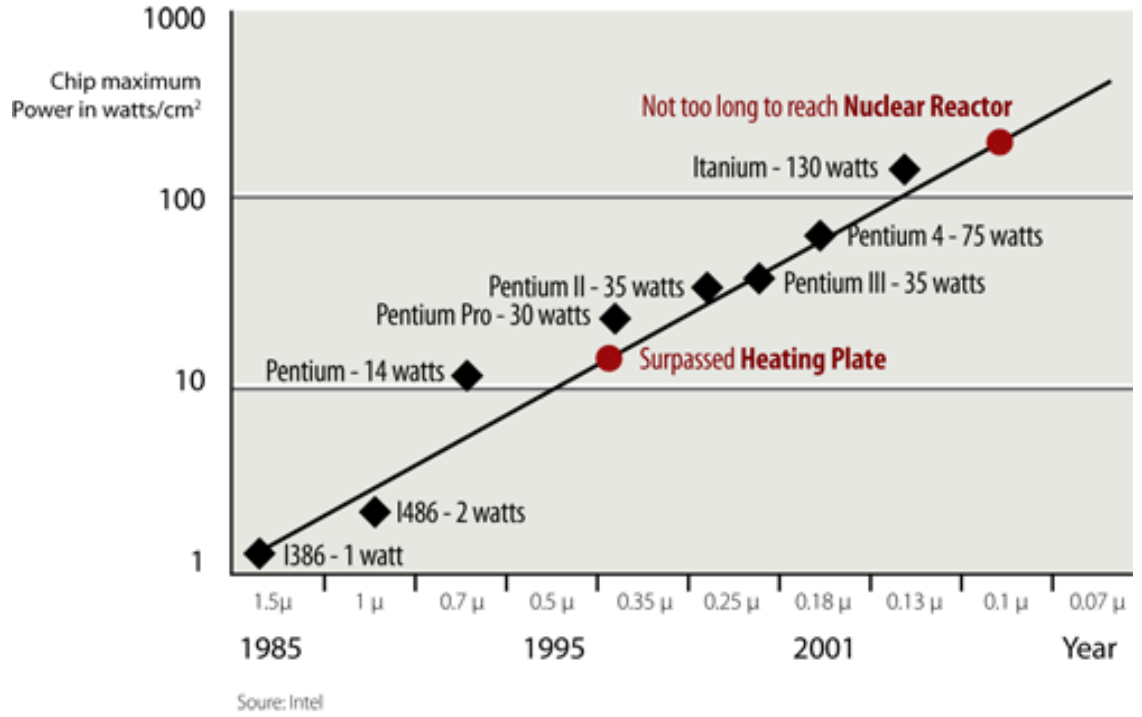
Κατανάλωση ενέργειας στα Hpc συστήματα (1/10)

- Τα μοντέρνα Hpc συστήματα και οι clusters συνήθως αποτελούνται από multicore συστήματα.
- Η αύξηση της ταχύτητας επιτυγχάνεται κυρίως αυξάνοντας τους πυρήνες σε κάθε κόμβο, μειώνοντας την συχνότητα του ρολογιού σε κάθε πυρήνα.
- Το κύριο πρόβλημα τα τελευταία χρόνια, είναι πως οι επεξεργαστές σχεδιάζονται ακολουθώντας τον νόμο του Moore, το οποίο δεν συνεπάγεται εξοικονόμηση ενέργειας.
- Προκειμένου να αντιμετωπιστεί η κατανάλωση ενέργειας, αυξάνεται ολοένα και περισσότερο η χρήση των GPUs, καθώς προσφέρουν καλύτερη απόδοση flop ανά watt.
- Η επιστημονική κοινότητα θεωρεί πως ο νόμος του Moore θα πάψει εξολοκλήρου να ισχύει το 2020.

Κατανάλωση ενέργειας στα Hpc συστήματα (2/10)

- Γίνεται πλέον αντιληπτό , πως υπάρχει η ανάγκη για νέα τεχνολογία σε σχεδιασμό CPUs , αφού οι υπερ-υπολογιστές δεν μπορούν πλέον να επωφεληθούν του νόμου του Moore για να αυξήσουν την θεωρητική μέγιστη απόδοσή τους.
- Σύμφωνα με την IBM , η μεγαλύτερη εξοικονόμηση ενέργειας προέρχεται από την χρήση ενός νέου επεξεργαστή βελτιστοποιημένου ως προς την ενεργειακή απόδοση και όχι την απόδοση ανά νήμα. Πρέπει λοιπόν οι υπερ-υπολογιστές να αποτελούνται από ενεργειακά αποδοτικούς επεξεργαστές.

Κατανάλωση ενέργειας στα Hpc συστήματα (3/10)



Εικόνα που αποτυπώνει τη θεωρητική μέγιστη απόδοση σε Watts των επεξεργαστών, σε σχέση με το μέγεθος των κυκλωμάτων και το πέρασμα των χρόνων.

Κατανάλωση ενέργειας στα Hpc συστήματα (4/10)

- Πέρα από τους επεξεργαστές, υπάρχουν και άλλες πήγες κατανάλωσης ενέργειας στα σύγχρονα Hpc συστήματα, όπως:
  - Η μνήμη
  - Οι συσκευές αποθήκευσης
  - Οι συσκευές επικοινωνίας
  - Οι μονάδες επεξεργασίας γραφικών
- Όλες αυτές οι συσκευές συνδυασμένες, καταναλώνουν πολύ μεγάλα ποσά ενέργειας.

Κατανάλωση ενέργειας στα Hpc συστήματα (Παράδειγμα κατανάλωσης ενέργειας 5/10)

- Ένας από τους γρηγορότερους υπερ-υπολογιστές σήμερα, ο K computer του RAKEN advanced institute of Computational Science (AICS) που βρίσκεται στην Ιαπωνία, χρησιμοποιεί RISC αρχιτεκτονική διαθέτοντας SPARC64 επεξεργαστές , έχει απόδοση 8.62 petaflops το δευτερόλεπτο

- Όπως αναφέρθηκε προηγουμένως, ένα petaflop ισοδυναμεί με 1000 τρισεκατομμύρια υπολογισμούς

- Σύμφωνα με υπολογισμούς, το σύστημα αυτό καταναλώνει 9.89 μεγαβατώρες.

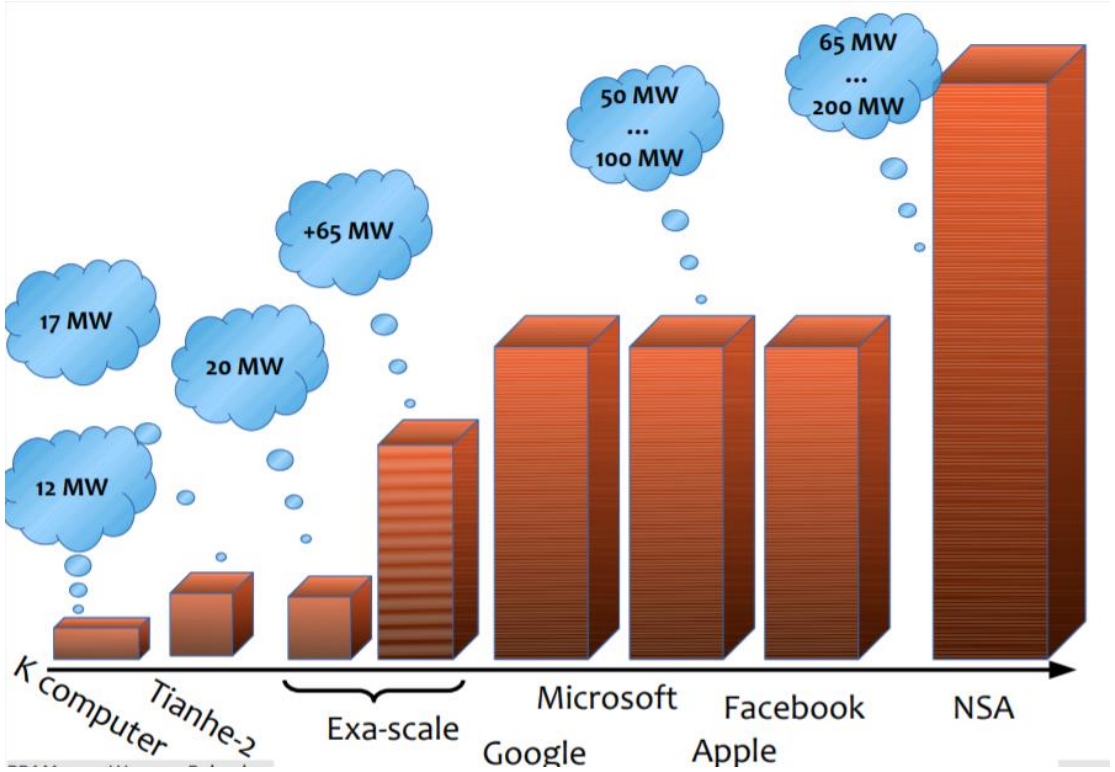
Κατανάλωση ενέργειας στα Hpc συστήματα (Παράδειγμα κατανάλωσης ενέργειας 6/10)

- Μία άλλη πολύ γρήγορη μηχανή είναι ο Tianhe-1A του εθνικού κέντρου υπερ-υπολογιστών στην Tianjin της Κίνας.
- Είναι μια υβριδική μηχανή η οποία είναι σε θέση να επιτύχει ταχύτητα 2.56 petaflops ανά δευτερόλεπτο με συνέπεια να καταναλώνει 4.04 μεγαβατώρες.
- Αυτό επιτυγχάνεται με συνδυασμό δύο βασικών κατηγοριών επεξεργαστών, intel XEON και nVIDIA GPU
- Η διαφορά κατανάλωσης στα δύο παραδείγματα, έγκειται στο γεγονός ότι η δεύτερη μηχανή χρησιμοποιεί GPU

Κατανάλωση ενέργειας στα Hpc συστήματα (Παράδειγμα κατανάλωσης ενέργειας 7/10)

- Σύμφωνα με το IBM Systems Magazine, η μέση κατανάλωση των server ανά κατηγορία είναι η εξής:
  - 1U server 300W-350W
  - 2U server 350W-400W
  - 4U server 600W-1000W
  - Ένα ολοκληρωμένο rack, ξοδεύει κατά μέσο όρο 4500W. Οι μετρήσεις εξαρτώνται από το hardware και τα κομμάτια που χρησιμοποιούνται
- Για παράδειγμα, άμα χρησιμοποιούνται rack που κατά μέσο όρο καταναλώνουν 4500W (4.5 kWh), πολλαπλασιάζοντας με το κόστος 0,17588€ που κοστίζει η kWh στην Ελλάδα και αξιοποιώντας τον 24/7, προκύπτει ένα κόστος για το rack που ισοδυναμεί με 6.933,18€ το χρόνο.

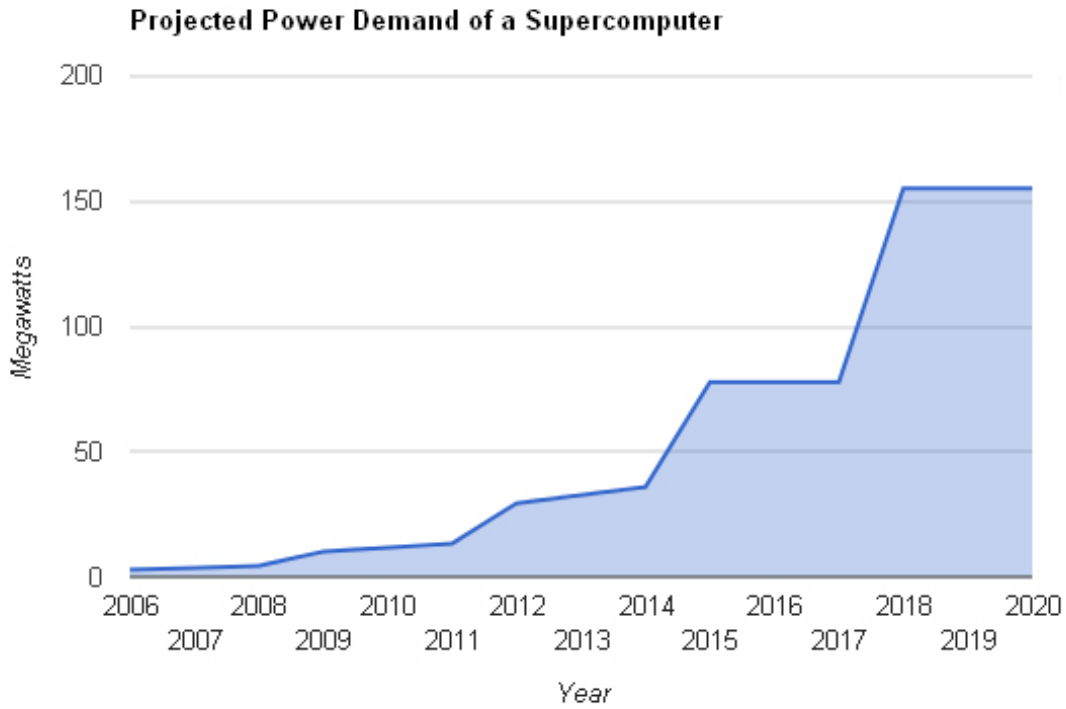
Κατανάλωση ενέργειας στα Hpc συστήματα (8/10)



Πηγή:
http://www.icl.utk.edu/~luszczek/conf/ppam2013_energy_power_trends/energy_power_trends_hpc.pdf

Εικόνα που αποτυπώνει τις ενεργειακές απαιτήσεις διάφορων Hpc συστημάτων εταιρειών-κολοσσών, εκφραζόμενες σε μεγαβατώρες

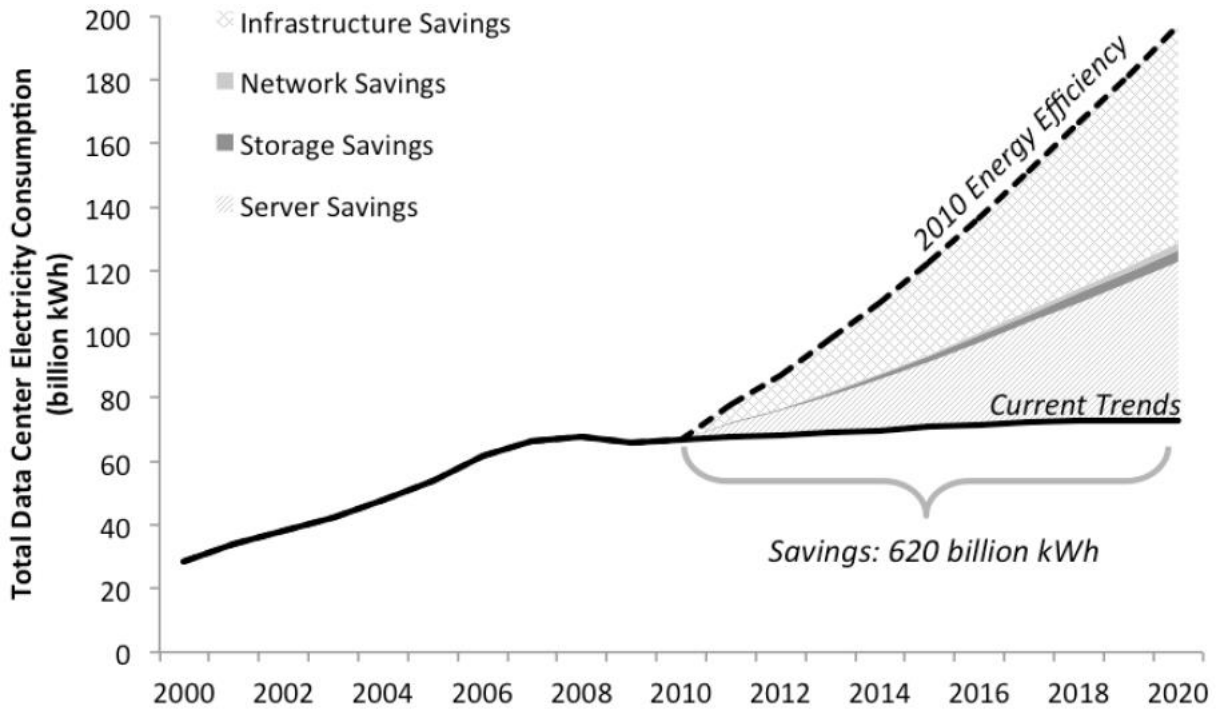
Κατανάλωση ενέργειας στα Hpc συστήματα (9/10)



Πορεία και προβλεπόμενες απαιτήσεις κατανάλωσης ενέργειας ενός υπερ-υπολογιστή από το 2006 μέχρι το 2020, εκφραζόμενες σε μεγαβατώρες

Πηγή: https://ww2.kqed.org/quest/2011/06/24/supercomputing-draft/

Κατανάλωση ενέργειας στα Hpc συστήματα (10/10)



Πορεία και προβλεπόμενες εκτιμήσεις ενεργειακών απαιτήσεων στα κέντρα δεδομένων των Ηνωμένων Πολιτειών της Αμερικής, από το 2000 μέχρι το 2020, εκφραζόμενες σε δισεκατομμύρια kWh

Πηγή: https://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume

Βρίσκοντας το κατάλληλο λογισμικό

- Για να μπορέσουμε να κάνουμε το σύστημα HPC να δουλέψει και να τρέξει, χρειάζεται να βρούμε το κατάλληλο λογισμικό.

- Το πιο κοινό λογισμικό μέχρι στιγμής για τα HPC είναι το linux, καθώς οι χρήστες δείχνουν να το προτιμάνε περισσότερο από τα υπόλοιπα και να δουλεύουν καλύτερα με αυτό.

- Άλλες επιλογές είναι το solaris, το οποίο είναι «ανοιχτό», και έχει πλήρη υποστήριξη από συμβατότητα με τις βιβλιοθήκες linux και το Microsoft HPC Διακομιστή.

Λειτουργικά συστήματα(1/4)

- Όπως προαναφέρθηκε, το επικρατέστερο λογισμικό για HPC είναι το Linux. Εξαιτίας του ερχομού του Linux, η αγορά των HPC ή supercomputing, χρησιμοποιούσε αποκλειστικά UNIX.

- Το linux αναπαριστά μία plug-and-play εναλλακτική (σύνδεσε και παίξε) και δεν προσθέτει επιπλέον τέλη χορήγησης άδειας για τους κόμβους (οι οποίοι μπορεί να είναι πολλοί σε αριθμό).

- Επιπροσθέτως με τον πυρήνα του Linux, πολλά από τα σημαντικά λογισμικά υποστήριξης, έχουν αναπτυχθεί σε περιβάλλον GNU (το οποίο είναι ένα UNIX-like ελεύθερο λογσμικό).

Λειτουργικά συστήματα(2/4)

- Το λογισμικό πυρήνα του GNU/Linux είναι λογισμικό ανοιχτού κώδικα και μπορεί να αντιγραφεί και να χρησιμοποιηθεί από τον καθένα. Υπάρχουν, ωστόσο, απαιτήσεις που ασφαλίζουν το πηγαίο κώδικα που μοιράζεται.

- Η ελευθερία και η διαμοιραστικότητα του GNU/Linux, το έχει κάνει ένα ιδανικό λογισμικό για τα HPC συστήματα. Επιτρέπει στους HPC προγραμματιστές να δημιουργήσουν εφαρμογές, να φτιάξουν οδηγούς (drivers), και να κάνουν αλλαγές οι οποίες κανονικά δεν θα μπορούσαν να γίνουν από λογισμικό κλειστού κώδικα.

- Στην πραγματικότητα, όλες οι εγκαταστάσεις Linux γίνονται με ένα εμπορικό ή δωρεάν πακέτου διανομής λογισμικού.

Λειτουργικά συστήματα(3/4)

- Ενώ η εμπορική διαθεσιμότητα του "ελεύθερου" λογισμικού μπορεί να είναι περίπλοκη, οι περισσότεροι προμηθευτές ελεύθερου λογισμικού χρησιμοποιούν ένα support-based (βασισμένο στην υποστήριξη) μοντέλο. Δίνεται η δυνατότητα της ελευθερίας της χρήσης του ανοιχτού πηγαίου κώδικα, αλλά άμα χρειαστεί υποστήριξη, χρειάζεται να ξοδευτούν χρήματα.

- Οι χρήστες μπορεί να αναγνωρίζουν μερικά από τα πακέτα διανομής όπως το Red Hat, SUSE, και άλλα. Υπάρχουν και εκδόσεις υποστήριξης από το community επίσης.

- Τα Red Hot Fedora, Open SUSE, και CentOS είναι παραδείγματα αυτής της προσέγγισης. Χρειάζεται να επισημανθεί ότι παρόλο που αυτά τα πακέτα διανομής είναι αρκετά καλά από μόνα τους, δεν διαθέτουν όλο το λογισμικό που χρειάζεται ένα HPC cluster για να υποστηριχθεί.

- Οι διανομείς Cluster είναι αυτή που γεμίζουν το κενό αυτό.

Λειτουργικά συστήματα(4/4) Software stack

- Συνοπτικά, το λογισμικό HPC πρέπει να είναι ικανό να κάνει τα παρακάτω:
  1. Να εγκαθιστά Linux στους κόμβους μέσω του διαδικτύου,
  2. Να προσθέτει, αφαιρεί, ή να αλλάζει κόμβους,
  3. Να καταγράφει τους κόμβους,
  4. Να τρέχει απομακρυσμένες εντολές μεταξύ τους κόμβους ή τα γκρουπ κόμβων,
  5. Να παίρνει ανταπόκριση από τις παραπάνω εντολές,
  6. Να καταγράφει κόμβους και εφαρμογές,
  7. Να καταγράφει τη CPU, μνήμη, και τη χρήση του συστήματος,
  8. Να τρέχει αυτοματοποιημένες ανταποκρίσεις όταν ένα γεγονός συμβαίνει στον cluster.

Λειτουργικά συστήματα: Cluster software(1/5)

- Υπάρχουν αρκετοί τύποι από διεργασίες λογισμικού όπου χρειάζεται να τρέχει ένας επιτυχημένος cluster. Αυτές οι διεργασίες περιλαμβάνουν τη διαχείριση, τον προγραμματισμό, την αποσφαλμάτωση, και τον προγραμματισμό των εργασιών.

- Από την πλευρά ενός χρήστη, ο προγραμματισμός του cluster είναι μία από τις πιο σημαντικές διεργασίες. Το πιο σημαντικό εργαλείο HPC για προγραμματισμό είναι ίσως το MPI (Message Passing Interface). Το MPI επιτρέπει στα προγράμματα να επικοινωνούν μεταξύ τους πάνω στα δίκτυα του cluster.

- Χωρίς αυτό το λογισμικό, η δημιουργία παράλληλων προγραμμάτων θα ήταν, όπως και ήταν στο παρελθόν, μία δύσκολη, custom (και λογικά αρκετά χρονοβόρα) διαδικασία.

Λειτουργικά συστήματα: Cluster software(2/5)

- Σήμερα, υπάρχουν και ανοιχτές και εμπορικές MPI εκδόσεις. Οι δύο πιο διάσιμες ανοιχτές MPI εκδόσεις είναι το MPICH2 από το Argonne Lab και το Open MPI project.

- Επιπροσθέτως, οι προγραμματιστές χρειάζονται μεταγλωττιστές και προφίλ. Το GNU software περιλαμβάνει πολύ καλούς μεταγλωττιστές και άλλα προγραμματισιτκά εργαλεία.

- Ωστόσο, πολλοί χρήστες προτιμούν να χρησιμοποιήσουν επαγγελαμτικά πακέτα μεταγλωττιστή/αποσφαλματωτή/προφίλ, όπως αυτά που προσφέρονται από το Sun Microsystems (Sun Studio 12 for Linux), το Portland Group (PGI) και την Intel. Όλοι οι πωλητές προμηθεύουν τα εργαλεία τους και τον cluster με stack λογισμικό (αναλύεται στην επόμενη διαφάνεια).

Λειτουργικά συστήματα: Cluster software(3/5)

- Stack είναι το λογισμικό, το οποίο περιλαμβάνει μία ομάδα προγραμμάτων τα οποία λειτουργούν μεταξύ τους το ένα πίσω από το άλλο, προκειμένου να επιτύχουν έναν κοινό στόχο.

- Είναι φυσικό και επόμενο οι τελικοί χρήστες να απαιτούν από τους προγραμματιστές να τρέχουν αρμονικά τα προγράμματα, με αποτέλεσμα οι διαχειριστές να καλούνται να κρατήσουν τον cluster σε εγρήγορση και συνεχή λειτουργικότητα.

- Υπάρχουν εργαλεία που βοηθούν στη διαχείριση της εγκατάστασης του cluster και στην παρακολούθηση των εργασιών του cluster.

Λειτουργικά συστήματα: Cluster software(4/5)

- Προκειμένου να χρησιμοποιηθεί η συνδυασμένη ισχύς ενός HPC cluster, το σύστημα πρέπει να είναι παράλληλο. Το ίδιο γίνεται και για τα μηχανήματα με πολλούς πυρήνες.

- Ένα τυπικό πρόγραμμα υπολογιστή είναι γραμμένο για να τρέξει σε ένα μονό επεξεργαστή ή πυρήνα. Δεν θα χρησιμοποιήσει αυτόματα επιπλέον πυρήνες ή κόμβους στον cluster. Για να δουλέψει παράλληλα, πρέπει να αλλαχθούν οι εσωτερικές λειτουργίες του προγράμματος.

- Υπάρχουν αρκετοί τρόποι για να επιτευχθεί αυτή η διαδικασία. Άμα θέλει κάποιος να τρέξει το πρόγραμμα τού σε πολλούς πυρήνες, τότε η χρησιμοποίηση των pthreads (σύνολο προγραμαμτιστικών τύπως δεδομένων και κλήσεων γραμμένων στη γλώσσα C) ή OpenMP είναι μία καλή λύση.

Λειτουργικά συστήματα: Cluster software(5/5)

- Εάν από την άλλη πλευρά θέλει να τρέξει ανάμεσα σε πολλούς κόμβους (και τους πολλαπλούς πυρήνες σε έναν κόμβο), τότε η χρήση του MPI (Message Passing Interface) μπορεί να είναι η κατάλληλη λύση.

- Σε άλλες περιπτώσεις, ο συνδυασμός των δύο είναι η κατάλληλη λύση. Διαφορετικά, ανάλογα με το επίπεδο των απαιτήσεων, χρησιμοποιείται η κατάλληλη τεχνική.

- Πολλοί εμπορικοί ISV (Independent Software Vendor) κωδικοί είναι ήδη παράλληλοι και δουλεύουν εξώ από τους clusters. Το ίδιο ισχύει και για πολλές εφαρμογες ανοιχτού κώδικα, παρόλο που συχνά ο χρήστης είναι υπεύθυνος για την ανάπτυξη των εφαρμογών ανοιχτού κώδικα.

Επίπεδο 1: Ρύθμιση λογισμικού (1/5)

- Το πρώτο επίπεδο λογισμικού περιέχει το ελάχιστο λογισμικό για να υλοποιηθεί ο παράλληλος προγραμματισμός. Προφανώς, το πρώτο πράγμα που θα χρειαστεί είναι το λογισμικό και μία τυπική εγκατάσταση είναι αυτή που συνίσταται.

- Όπως προαναφέρθηκε, το επικρατέστερο λογισμικό είναι το linux. Το επόμενο βήμα που χρειάζεται είναι να ρυθμιστούν οι βιβλιοθήκες MPI, όπως το Open MPI ή το MPICH. Αυτές οι βιβλιοθήκες χρησιμοποιούνται για να δημιουργηθεί ο παράλληλος προγραμματισμός και να τρέξουν στο cluster.

- Κάθε κόμβος έχει τις ίδιες βιβλιοθήκες για τις MPI εφαρμογές.

Επίπεδο 1: Ρύθμιση λογισμικού (2/5)

- Στη συνέχεια, έχουμε τα εξής βήματα: τη δημιουργία (build), την εγκατάσταση, και τη ρύθμιση των κατάλληλων μονοπατιών για τις βιβλιοθήκες, στον διαμοιρασμένο φάκελο. Επίσης, μπορεί να γίνει εγκατάσταση των βιβλιοθηκών σε κάθε κόμβο ξεχωριστά.

- Στο επόμενο βήμα, χρειαζόμαστε την εγκατάσταση του SSH. Συγκεκριμένα, χρειάζεται να μπορεί να γίνει SSH από και προς κάθε κόμβο χωρίς κωδικό, επιτρέποντας τις εφαρμογές MPI να τρέξουν εύκολα. (Το **SSH** (Secure Shell) είναι ένα ασφαλές δικτυακό πρωτόκολλο το οποίο επιτρέπει τη μεταφορά δεδομένων μεταξύ δύο υπολογιστών.)

- Ωστόσο, πρέπει να γίνει βέβαιο ότι η εγκατάσταση του SSH είναι δυνατή.

Επίπεδο 1: Ρύθμιση λογισμικού (3/5)

- Επιπροσθέτως, χρειάζεται ο αριθμός των χρηστών και των ομάδων στους κόμβους. Επειδή πρέπει να δημιουργηθεί ο ίδιος χρήστης σε κάθε κόμβο, αυτή η διαδικασία θα μπορούσε να είναι χρονοβόρα εάν υπάρχουν χιλιάδες κόμβοι.

- Η εκτέλεση των εφαρμογών υπό το καθεστός SSH δεν είναι δύσκολη, επειδή οι κόμβοι έχουν ένα διαμοιρασμένο κατάλογο. Επιπλέον, πρέπει να επισημανθεί ότι υπάρχουν περισσότεροι από ένας διαμοιρασμένοι φάκελοι.

- Ας υποθέσουμε ότι δημιουργείται μία εφαρμογή MPI στον κύριο κόμβο στον home κατάλογο. (π.χ. *home/laytonjb/bin/<app>*, όπου *<app>* είναι το εκτελέσιμο αρχείο.

Επίπεδο 1: Ρύθμιση λογισμικού (4/5)

- Ο κατάλογος /home, μπορεί να διαμοιραστεί από το cluster, έτσι ώστε να μπορεί να προορίσει στην εφαρμογή τα ίδια αρχεία εισόδου-εξόδου (Πιθανώς τα αρχεία εισόδου και εξόδου να είναι στο ίδιο διαμοιρασμένο κατάλογο).

- Καθώς ξεκινά η εφαρμογή, το SSH χρησιμοποιείται για να επικοινωνήσει στις βαθμίδες MPI (τις MPI διαδικασίες).

- Επειδή μπορούμε να κάνουμε SSH χωρίς κωδικούς πρόσβασης, η εφαρμογή μπορεί να τρέξει χωρίς προβλήματα.

Επίπεδο 1: Ρύθμιση λογισμικού (5/5)

- Οι λεπτομέρειες από το τρέξιμο της MPI εφαρμογής, εξαρτάται από τη βιβλιοθήκη MPI, η οποία τυπικά παρέχει ένα απλό script ή μικρά εκτελέσιμα αρχεία προκειμένου να τρέξει η εφαρμογή.

- Αυτή η ρύθμιση του λογισμικού είναι μόλις η ελάχιστη για να επιτρέψει στον cluster να τρέξει τις εφαρμογές.

- Στη συνέχεια γίνεται αναφορά στην αρχιτεκτονική και στα εργαλεία του λογισμικού, με τα οποία θα είναι δυνατό να λειτουργήσει αρμονικά ο cluster.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (1/15)

- Το επόμενο επίπεδο του λογισμικού, προσθέτει τα εργαλεία τα οποία βοηθάνε στη μείωση των προβλημάτων του cluster και κάνει ευκολότερη τη διαδικασία της διαχείρισης στον διαχειριστή.

- Χρησιμοποιώντας το βασικό λογισμικό το οποίο αναφέρθηκε προηγουμένως, μπορούμε να έχουμε παράλληλο προγραμματισμό, αλλά επίσης μπορεί να βιώσουμε μερικές δυσκολίες όσο κλιμακώνουμε το σύστημα, οι οποίες έχουν να κάνουν με:

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (2/15)

1. Το τρέξιμο των εντολών σε κάθε κόμβο (parallel shell).
2. Τη διαμόρφωση των παρόμοιων κόμβων (παραποίηση του πακέτου).
3. Τη διατήρηση του ίδιου χρόνου σε κάθε κόμβο (NTP – NETWORK TIME PROTOCOL SERVER).
4. Το τρέξιμο παραπάνω της μίας διεργασίας (προγραμματιστής δουλειών/διαχειριστής πηγών).

- Αυτά τα θέματα προκύπτουν καθώς κλιμακώνεται ο cluster, αλλά ακόμα και σε έναν cluster δύο κόμβων, μπορούν να προκληθούν τέτοιου είδους προβλήματα.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (3/15)

- Αρχικά, χρειάζεται να είμαστε ικανοί να τρέξουμε την ίδια εντολή σε κάθε κόμβο. Μία λύση θα ήταν να γραφτεί ένα απλό shell script (σενάριο) το οποίο θα παίρνει το σύνολο των γραμμών εντολών σαν την βασική «εντολή» και θα τρέχει την εντολή σε κάθε κόμβο ξεχωριστά, χρησιμοποιώντας SSH.

- Ωστόσο, τι θα γίνει στην περίπτωση που θέλουμε να τρέξουμε την εντολή σε ένα υποσύνολο των κόμβων;

- Αυτό που χρειαζόμαστε, είναι αυτό που καλείται παράλληλο shell.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (4/15)

- Στο λογισμικό linux είναι διαθέσιμος ένας αριθμός από παράλληλα εργαλεία shell_, με το πιο κοινό να είναι το pdsh, το οποίο επιτρέπει να τρέχουμε την ίδια εντολή σε κάθε κόμβο.

- Ωστόσο, έχοντας απλά ένα παράλληλο shell, δεν σημαίνει ότι ο cluster μαγικά θα λύσει όλα τα προβλήματα. Οπότε, χρειάζεται να εξελιχθούν κάποιες λειτουργίες και διαδικασίες.

- Συγκεκριμένα, μπορεί να χρησιμοποιηθεί ένα παράλληλο shell για να ξεπεραστεί το δεύτερο θέμα (η διαμόρφωση των παρόμοιων κόμβου - παραποίηση του πακέτου).

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (5/15)

- Η παραποίηση του πακέτου μπορεί να δημιουργήσει πολλά προβλήματα στους διαχειριστές του HPC.
- Εάν υπάρχει μία εφαρμογή η οποία τρέχει τη μία μέρα πολύ καλά, αλλά την άλλη δεν τρέχει, χρειάζεται να ανατρέξουμε στο λόγο που γίνεται αυτό.
- Πιθανώς, κατά τη διάρκεια μία ημέρας, ένας κόμβος ο οποίος τερμάτισε, να ήρθε ξαφνικά πίσω στη λειτουργία και να ξεκινήσαν κατευθείαν να τρέχουν εφαρμογές σε αυτόν.
- Αυτός ο κόμβος λογικά δεν έχει τα ίδια πακέτα ή τις ίδιες εκδόσεις του λογισμικού, όπως οι άλλοι κόμβοι.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (6/15)

- Ως αποτέλεσμα, οι εφαρμογές μπορούν να αποτύχουν. Χρησιμοποιώντας ένα παράλληλο shell, μπορούμε να ελέγξουμε ότι κάθε κόμβος έχει το πακέτο εγκατεστημένο και ότι οι εκδόσεις ταιριάζουν.

- Για να βοηθηθούμε με την παραποίηση των πακέτων, προτείνεται μετά την πρώτη δημιουργία του cluster και την εγκατάσταση του παράλληλου shell, να ξεκινήσει η εξέταση των βασικών σημείων κλειδιών της εγκατάστασης, όπως:

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (7/15)

1. Την έκδοση glibc,
2. Την έκδοση GCC,
3. Την έκδοση Gfortran,
4. Την έκδοση SSH,
5. Την έκδοση Kernel,
6. Την έκδοση της IP,
7. Τις βιβλιοθήκες MPI,
8. Το NIC MTU χρησιμοποιώντας ifconfig
9. Το BogoMips – Παρόλο που ένας αριθμός είναι ανούσιος, πρέπει να είναι ο ίδιος ανάμεσα στους κόμβους άμα χρησιμοποιούμε το ίδιο υλικό.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (8/15)

- Για να ελέγξουμε αυτόν τον αριθμό, πληκτρολογούμε:
  - cat /proc/cpuinfo | grep bogomips
- Για να ελέγξουμε αν οι κόμβοι έχουν την ίδια μνήμη πληκτρολογούμε:
  - cat /proc/meminfo | grep MemTotal

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (9/15)

- Επίσης είναι απαραίτητο να υπάρχει κάποιο κρατημένο αντίγραφο ασφαλείας σε περίπτωση που κάποιος κόμβος «πέσει» όταν γίνεται εγκατάσταση ή αναβάθμιση πακέτων ή ακόμα και σε περίπτωση φυσικής καταστροφής.

- Διαφορετικά, δε θα μπορούμε να επαναφέρουμε τις διαδικασίες πίσω στην κανονικότητα, όταν επιστρέψει και ο κόμβος πίσω. Έτσι, θα ξεκινήσει παραποίηση όλων των πακέτων στους κόμβους και πολλά επακόλουθα προβλήματα.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (10/15)

- Το τρίτο θέμα που καλούμαστε να υπερβούμε, είναι να κρατήσουμε τον ίδιο χρόνο σε κάθε κόμβο. Το πρωτόκολλο συγχρονισμού χρόνου συγχρονίζει τα ρολόγια του συστήματος.

- Οι περισσότερες κατανομές εγκαθιστούν το ntp εξ ορισμούκαι το ενεργοποιούν, αλλά χρειάζεται να ελεγχθεί ότι εγκαθίσταται σε κάθε cluster – και επίσης να ελεγχθεί η έκδοση του ntpd επίσης.

- Χρησιμοποιούμε το chkconfig, εάν η κατανομή έχει αυτό το πακέτο, για να ελεγχθεί ότι το ntp τρέχει. Ωστόσο, χρειάζεται να ρίξουμε μία ματιά στις διαδικασίες που τρέχουν στους κόμβους, για να δούμε εάν το ntpd είναι καταχωρημένο (χρησιμοποιούμε το parallel shell).

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (11/15)

- Στον κύριο κόμβο, πρέπει να σιγουρευτεί ότι η διαμόρφωση NTP δείχνει στους εξωτερικούς διακομιστές (έξω από τον cluster) και ότι ο κύριος κόμβος μπορεί να επιλύσει αυτά τα URL (με το nslookup για παράδειγμα).

- Για κόμβους όπου βρίσκονται σε ιδιωτικό δίκτυο και δεν έχουν πρόσβαση στον διαδίκτυο, το NTP χρειάζεται να διαμορφωθεί με τέτοιον τρόπο ώστε να χρησιμοποιείται ως χρονόμετρο.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (12/15)

- Κάτι τέτοιο μπορεί να συμβεί τροποποιώντας το /etc/htp.conf και αλλάζοντας τους NTP διακομιστές έτσι ώστε να δείχνει στην IP διεύθυνση του κύριου κόμβου.

- Ουσιαστικά, πρέπει να μοιάζει με την πρώτη καταχώρηση. Η IP διεύθυνση του κύριου κόμβου είναι 10.1.0.250. Χρειάζεται να διασφαλιστεί ότι οι κόμβοι υπολογισμού μπορούν να κάνουν ping (έλεγχος διαθεσιμότητας) τη διεύθυνση. Επιπλέον, το ntp πρέπει να ξεκινάει όταν οι κόμβοι φορτώνουν.

```
[root@test1 etc]# more ntp.conf
# For more information about this file, see the man pages
# ntp.conf(5), ntp_acc(5), ntp_auth(5), ntp_clock(5), ntp_misc(5), ntp_mon(5).

#driftfile /var/lib/ntp/drift

restrict default ignore
restrict 127.0.0.1
server 10.1.0.250
restrict 10.1.0.250 nomodify
```

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (13/15)

- Το τελευταίο θέμα που χρήζει επίλυσης είναι αυτή του προγραμματιστή των εργασιών (όπως επίσης καλείται και διαχειριστής πόρων).
- Είναι ένα στοιχείο κλειδί του HPC και μπορεί να χρησιμοποιηθεί ακόμα για μικρούς clusters.
- Ένας προγραμματιστής εργασιών θα τρέξει εφαρμογές εκ μέρους μας και θα περιμένει τον cluster να ελευθερωθεί πριν τρέξουμε τις εφαρμογές.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (14/15)

- Σε ένα σενάριο, ξεκαθαρίζουμε τις πηγές που χρειαζόμαστε, όπως τον αριθμό των κόμβων ή τον αριθμό των πυρήνων, και δίνουμε στον προγραμματιστή εργασιών την εντολή η οποία τρέχει την εφαρμογή:
  - Mpirun –np 4 <executable>

- Ανάμεσα στις διαθέσιμες πηγές, πολλές είναι ανοιχτής πηγής, και πολλές φορές δεν είναι πολύ δύσκολο να εγκατασταθούν και να ρυθμιστούν. Ωστόσο, πρέπει να σιγουρευτεί ότι έχουν ακολουθηθεί οι οδηγίες σύμφωνα με τον οδηγό.

Επίπεδο 2: Αρχιτεκτονική και εργαλεία (15/15)

- Παραδείγματα διαχειριστών πηγών είναι:
  - OpenLava
  - Slurm
  - Torque
  - SGE – Son of Grid Engine
  - OGE – Open Grid Engine.

- Με αυτά τα θέματα λυμένα, διαθέτουμε έναν πολύ λειτουργικό cluster, με μερικά διαχειριστικά εργαλεία. Μπορεί να μην είναι τέλειος, αλλά δουλεύει αρκετά καλά. Για περισσότερο εμπλουτισμό, υπάρχουν εργαλεία και στο 3ο επίπεδο, τα οποία θα αναλυθούν αμέσως παρακάτω.

Επίπεδο 3: Καλύτερη διαχείριση (1/12)

- Το τρίτο επίπεδο των εργαλείων μας πηγαίνει σε μονοπάτια που χρειάζονται μεγαλύτερη ανάλυση όσον αφορά διαχείριση HPC. Τα εργαλεία τα οποία θα αναλυθούν είναι:
  - Εργαλεία διαχείρισης cluster.
  - Εργαλεία παρακολούθησης (πως συμπεριφέρονται οι κόμβοι).
  - Modules περιβάλλοντος.
  - Πολλαπλά διαδίκτυα.

Επίπεδο 3: Καλύτερη διαχείριση (2/12) Εργαλεία διαχείρισης cluster

- **Ένα εργαλείο διαχείρισης cluster** αυτοματοποιεί τη διαμόρφωση, την παρουσίαση και τη διαχείριση των κόμβων υπολογισμού από τον κύριο κόμβο (ή από κάποιον κόμβο που σχεδιάστηκε ένας κύριος).

- Σε μερικές περιπτώσεις, το εργαλείο θα εγκαταστήσει τον κύριο κόμβο για εμάς. Ένας αριθμός εργαλείων διαχείρισης ανοιχτού κώδικα είναι διαθέσιμα, περιλαμβάνοντας τα παρακάτω:
  - Warewulf
  - Xcat
  - ROCKS
  - Oscar
  - onesies

Επίπεδο 3: Καλύτερη διαχείριση (3/12)

- Τα εργαλεία ποικίλουν ανάλογα με την προσέγγιση, αλλά τυπικά μας επιτρέπουν να υπολογίσουμε τους κόμβους υπολογισμού, οι οποίοι είναι μέρος του cluster.
- Αυτό μπορεί να συμβεί από τις εικόνες, στις οποίες μία ολοκληρωμένη εικόνα προωθείται στον κόμβο υπολογισμού, ή μέσα από πακέτα, τα οποία είναι εγκατεστημένα στους κόμβους υπολογισμού.
- Ο τρόπος με τον οποίο επιτυγχάνεται αυτό εξαρτάται από εργαλείο σε εργαλείο, οπότε πρέπει να είμαστε σίγουροι πριν τα εγκαταστήσουμε.

Επίπεδο 3: Καλύτερη διαχείριση (4/12)

- Ένα ενδιαφέρον κομμάτι σε αυτά τα εργαλεία είναι ότι διαγράφουν την – καθόλου - ενδιαφέρουσα δουλειά της εγκατάστασης και διαχείρισης των κόμβων υπολογισμού. Ακόμα και σε έναν cluster τεσσάρων κόμβων, δεν χρειάζεται να συνδέσουμε σε όλους για να τους πειράξουμε.

- Η δυνατότητα να τρέξουμε μία μονή εντολή και η επανεγκατάσταση των παρόμοιων υπολογιστικών κόμβων, μπορεί να εξαλείψει πολλά προβλήματα καθώς γίνεται διαχείριση του cluster.

Επίπεδο 3: Καλύτερη διαχείριση (5/12) Εργαλεία παρακολούθησης cluster.

- Πολλά από τα διαχειριστικά εργαλεία του cluster, περιλαμβάνουν επίσης εργαλεία για **παρακολούθηση** του cluster.

- Για παράδειγμα, η δυνατότητα να γνωρίζουμε ποιοι κόμβοι είναι πάνω ή κάτω ή ποιοι κόμβοι χρησιμοποιούν μεγάλο ποσοστό της επεξεργαστικής ισχύος (και ποια όχι) είναι σημαντικές πληροφορίες για τους διαχειριστές των HPC.

- Η παρακολούθηση των διαφόρων πληροφοριών των cluster, συμπεριλαμβάνοντας στατιστικές από τη χρησιμοποίηση του cluster, μπορούν να χρησιμοποιηθούν όταν ζητούνται πληροφορίες για επιπλέον υλικό ή πόσο πολύ χρησιμοποείται ο cluster.

Επίπεδο 3: Καλύτερη διαχείριση (6/12) Εργαλεία παρακολούθησης cluster.

- Πολλά εργαλεία παρακολούθησης είναι κατάλληλα για HPC clusters, αλλά ένα καθολικό εργαλείο είναι το Ganglia. Μερικά εργαλεία cluster είναι προ-διαρυθμισμένα με το Ganglia, και κάποια όχι, απαιτώντας έτσι εγκατάσταση.

- Εξ εξορισμού, το Ganglia έρχεται με προρυθμισμένα προγράμματα και αρκετές μετρήσεις, οπότε το εργαλείο είναι πολύ ευέλικτο και επιτρέπει να γραφτεί απλός κώδικας προκειμένου να εξασφαλιστούν συγκεκριμένες ρυθμίσεις που επιθυμούμε από τους κόμβους μας.

Επίπεδο 3: Καλύτερη διαχείριση (7/12) Εργαλεία παρακολούθησης cluster.

- Σε αυτό το σημείο, έχουμε τα ίδια εργαλεία εξέλιξης, τους ίδιους μεταγλωττιστές, τις ίδιες βιβλιοθήκες MPI και τις ίδιες βιβλιοθήκες εφαρμογής, εγκατεστημένες όλες στους κόμβους μας.

- Για να δοκιμάσουμε διαφορετικές εκδόσεις μία συγκεκριμένης βιβλιοθήκης, χρειάζεται να τερματίσουμε όλες τις διεργασίες στον cluster, να κάνουμε εγκατάσταση τις βιβλιοθήκες που χρειαζόμαστε, να σιγουρέψουμε ότι βρίσκονται στο φάκελο που ορίστηκαν, και τότε να ξαναρχίσουμε όλες τις διεργασίες.

- Αυτή η διαδικασία μοιάζει με ένα ατύχημα το οποίο περιμένουμε να γίνει. Η προφύλαξη από αυτό καλείται environmental module.

Επίπεδο 3: Καλύτερη διαχείριση (8/12) Εργαλεία παρακολούθησης cluster.

- Αρχικά, οι περιβαλλοντικές μεταβλητές εξελίχθηκαν για να διευθετήσουν το πρόβλημα της ύπαρξης εφαρμογών με διαφορετικές απαιτήσεις σε βιβλιοθήκες ή μεταγλωττιστές.

- Αυτό συμβαίνει μέσα από τη δυναμική διαμόρφωση του περιβάλλοντος του χρήστη, με τμήματα αρχείων. Μπορούμε να φορτώσουμε ένα τμήμα αρχείου.

- Μετά τη δημιουργία της εφαρμογής με την αξιοποίηση αυτών των εργαλείων και βιβλιοθηκών, εάν τρέξουμε μία εφαρμογή η οποία χρησιμοποιεί διαφορετικά τμήματα εργαλείων, μπορούμε να ξεφορτώσουμε το πρώτο τμήμα του αρχείου και να εισάγουμε ένα καινουργιο τμήμα αρχείου.

Επίπεδο 3: Καλύτερη διαχείριση (9/12) Εργαλεία παρακολούθησης cluster.

- Το Lmod, το οποίο είναι ένα Lua based τμήμα συστήματος, είναι ένα εργαλείο το οποίο χειρίζεται με ευκολία τα προβλήματα ιεραρχίας του MODULEPATH.
- Ουσιαστικά, αποτελεί μία νέα έκδοση των τμημάτων του περιβάλλοντος, το οποίο διευθετεί την ανάγκη για ιεραρχία των τμημάτων.
- Με αυτόν τον τρόπο, γίνεται φόρτωση της μονής μεταβλητής «load», η οποία με τη σειρά της μπορεί να φορτώσει μία σειρά από τμήματα.
- Προς το παρόν, το Lmod βρίσκεται κάτω από πολύ ενεργή εξέλιξη.

Επίπεδο 3: Καλύτερη διαχείριση (10/12) Εργαλεία παρακολούθησης cluster.

- Προς το παρόν, υποθέτουμε ότι όλη η κίνηση (traffic) στον cluster, περιλαμβάνοντας τη διαχείριση, την αποθήκευση, και την υπολογιστική, χρησιμοποιούν το ίδιο δίκτυο.

- Για βελτιωμένη υπολογιστική απόδοση ή βελτιωμένη απόδοση της αποθήκευσης, ίσως χρειαστεί να χωρίσουμε την κίνηση σε συγκεκριμένα δίκτυα.

- Για παράδειγμα, μπορεί να θεωρήσουμε ένα ξεχωριστό δίκτυο για την κίνηση της διαχείρισης και την κίνηση της αποθήκευσης, έτσι ώστε ο κόμβος να έχει δύο ιδιωτικά δίκτυα: ένα για υπολογισμό και ένα για διαχείριση και αποθήκευση.

Επίπεδο 3: Καλύτερη διαχείριση (11/12) Εργαλεία παρακολούθησης cluster.

- Το να διαχωρίσουμε την κίνηση είναι πολύ εύκολο, με το να δίνουμε κάθε διεπαφή διαδικτύου (NIC) μέσα στον κόμβο μία IP διεύθυνση με διαφορετικό εύρος διεύθυνσης.

- Για παράδειγμα, ή eth0 μπορεί να είναι σε ένα δίκτυο 10.0.1.x, και η eth1 σε ένα δίκτυο 10.0.2.x. Παρόλο που θεωρητικά μπορούμε να δώσουμε σε όλες τις διεπαφές μία διεύθυνση στο ίδιο εύρος IP, διαφορετικά εύρη IP κάνουν τη διαχείριση πιο εύκολη.

- Όταν τρέχουμε MPI εφαρμογές, χρησιμοποιούμε διευθύνσεις της μορφής 10.0.1.x.

Επίπεδο 3: Καλύτερη διαχείριση (12/12) Εργαλεία παρακολούθησης cluster.

- Για το NFS και κάθε κίνηση διαχειριστή, πρέπει να χρησιμοποιούμε διευθύνσεις σε μορφή 10.0.2.x.
- Με αυτόν τον τρόπο, απομονώνουμε την υπολογιστική κίνηση (traffic).
- Το πλεονέκτημα της απομόνωσης της κίνησης είναι το επιπλέον εύρος ζώνης στα δίκτυα.
- Το μειονέκτημα είναι οι διπλάσιες θύρες, διπλάσια καλώδια και το μεγαλύτερο κόστος.
- Ωστόσο, εάν το κόστος και η πολυπλοκότητα δεν είναι τόσο σημαντικά, η χρήση δύο διαδικτύων είναι και η προτεινόμενη.

Χρονοδρομολογητές κατανεμημένων συστημάτων HPC(1/4)

- Οι clusters συνήθως έχουν πολλούς πυρήνες και πολλούς χρήστες. Ο διαμοιρασμός αυτών των πληροφοριών δεν είναι ένα ασήμαντο θέμα.

- Ευτυχώς, η κατανομή των πυρήνων γίνεται από τους προγραμματιστές λογισμικού και όχι από τους χρήστες. Ανάλογα με την εφαρμογή, ένας χρονοδρομολογητής κατανεμημένου συστήματος μπορεί να διανέμει τις πληροφορίες σε λίγους πυρήνες, ή να τους διανέμει τυχαία μέσα στον cluster.

- Για παράδειγμα, να τους κρατήσει όλους σε έναν μικρό αριθμό από κόμβους.

Χρονοδρομολογητές κατανεμημένων συστημάτων HPC(2/4)

- Ένας χρονοδρομολογητής κατανεμημένου συστήματος λειτουργεί ως εξής:
  - Όπως και στο παρελθόν, όλοι οι χρήστες πρέπει να υποβάλλουν τις εργασίες τους σε μία ουρά.
  - Ως κομμάτι των διαδικασιών της υποβολής, ο χρήστης πρέπει να κάνει συγκεκριμένες τις απαιτήσεις της εργασίας (για παράδειγμα πόσους πυρήνες, πόση μνήμη, πόσο χρόνο θα χρειαστεί κ.ο.κ.).
  - Ο χρονοδρομολογητής τότε ανάλογα με τους διαθέσιμους πόρους, κατανέμει τις εργασίες του προγράμματος. Η διαδικασία της κατανομής εξαρτάται από τη διαθεσιμότητα των πόρων αλλά και από τις τεχνικές τους προδιαγραφές.

Χρονοδρομολογητές κατανεμημένων συστημάτων HPC(3/4)

- Ο χρονοδρομολογητής είναι ένα σημαντικό κομμάτι για τον cluster, επειδή θα ήταν απίθανο να μοιραστούν προγράμματα με την ίδια μορφή, όπως το κάνει το εργαλείο ισορροπίας εκφόρτωσης.

- Μία σημαντική λειτουργία του προγραμματιστικού επιπέδου επιτρέπει στους διαχειριστές να διαχειριστούν τα προγράμματα σε κατάσταση εκτός λειτουργίας, είτε για επιδιόρθωση είτε για ανάβαθμιση.

- Οι χρήστες συνήθως δεν είναι "σοφοί" - σπάνια επιλέγουν σε ποιους κόμβους θα αναθέσουν τις δουλειές.

Χρονοδρομολογητές κατανεμημένων συστημάτων HPC(4/4)

- Επιπροσθέτως, εάν κάποιος κόμβος αποτύχει, οι εργασίες οι οποίες τρέχουν μπορούν και αυτές να αποτύχουν, όμως οι υπόλοιποι κόμβοι θα συνεχίζουν να δουλεύουν κανονικά και γύρω από τον κόμβο που απέτυχε.

- Υπάρχουν αρκετοί διάσημοι και δωρεάν χρονοδρομολογητές. Μία διάσημη επιλογή είναι το Sun Grid Engine από τη Sun Microsystems. Άλλες επιλογές είναι Torque, Lava και Maul.

- Στο εμπορικό κομμάτι, υπάρχουν αρκετές εκδόσεις που υποστηρίζουν τα Sun Grid Engine, Moab, Univa UD UniCluster και Platform LSF.

Ασφάλεια λογισμικού στα κέντρα HPC (1/10)

- Η ασφάλεια μίας υποδομής HPC είναι μία πολύ σημαντική διαδικασία. Υπάρχει ένας μεγάλος αριθμός απειλών ασφάλειας, οι οποίοι προέρχονται από το διαδίκτυο και από τα εσωτερικά δίκτυα και παρόλο που το κόστος διασφάλισης των HPC μπορεί να είναι μεγάλο, είναι σημαντική η προστασία των συστημάτων, γιατί το τίμημα από την απώλεια δεδομένων πληρώνεται πολύ ακριβότερα.
- Όπως και τα περισσότερα υπολογιστικά συστήματα, έτσι και οι υπερ-υπολογιστές χρειάζονται ασφάλεια από κάθε είδους απειλή, προκειμένου να προστατευτούν και να διατηρηθεί η ακεραιότητα των δεδομένων που έχουν. Είναι αντιληπτό, ακόμα και από τον ορισμό "high-performance clusters – υψηλής απόδοσης υπολογιστές", πως οι παραδοσιακοί τρόποι ασφάλειας στα πληροφοριακά συστήματα δεν είναι αποτελεσματικοί στους clusters, καθώς είναι σχεδιασμένοι για συστήματα χαμηλότερων επιδόσεων και προδιαγραφών.

Ασφάλεια λογισμικού στα κέντρα HPC (2/10)

- Ωστόσο, οι επιθέσεις που μπορούν να δεχθούν οι clusters είναι ίδιες με αυτές των απλούστερων συστημάτων. Για παράδειγμα, μπορεί να είναι από κάποιον εργαζόμενο ο οποίος προσπαθεί να υποκλέψει πληροφορίες μίας εταιρείας μέχρι μία επίθεση άρνησης υπηρεσίας (Distributed Denial-of-Service-Attack – γνωστή και ως DDoS), η οποία πραγματοποιείται από πολλούς εξυπηρετητές.
- Το γεγονός ότι οι clusters συνδυάζουν διαφορετικούς εξυπηρετητές, υπηρεσίες και εφαρμογές, καθιστά πολύ δύσκολη την προστασία τους και ο συνδυασμός των λύσεων προστασίας που μπορούν να εφαρμοστούν σίγουρα δεν μπορεί να είναι ποτέ τέλειος, διότι είτε ανακαλύπτονται συνεχώς νέα κενά ασφαλείας στα λογισμικά, είτε οι λύσεις ασφαλείας που εφαρμόζονται μπορούν να περιέχουν bugs.

Ασφάλεια λογισμικού στα κέντρα HPC (3/10)

- Για αυτό το λόγο, οι προσπάθειες βελτιστοποίησης της προστασίας των κέντρων HPC είναι συνεχώς σε εξέλιξη. Μερικές από τις τεχνολογίες που χρησιμοποιούνται για την προστασία των κέντρων δεδομένων, όπως αυτές παρουσιάζονται από την PRACE (Partnership For Advanced Computing in Europe) [7][8], η οποία ουσιαστικά είναι η συμμαχία για την προηγμένη υπολογιστική στην Ευρώπη, είναι οι παρακάτω:

[7] http://www.prace-ri.eu

[8] http://www.prace-ri.eu/IMG/pdf/wp79.pdf

Ασφάλεια λογισμικού στα κέντρα HPC (4/10)

1. Τείχος προστασίας για τοπικά δίκτυα: Ένα τείχος προστασίας είναι η βασική άμυνα ενάντια στις απειλές που έρχονται τόσο εντός, όσο και εκτός του δικτύου. Επειδή όμως αφορά ολόκληρο το HPC κέντρο δεδομένων και χρειάζεται προσοχή στην οργάνωση της ασφάλειάς του, μία από τις καλύτερες τεχνικές προφύλαξης είναι ο πλεονασμός. Αντί να υπάρχει δηλαδή μόνο ένα τείχος προστασίας που να προστατεύει το δίκτυο, είναι συνηθισμένη πρακτική να χρησιμοποιούνται δύο ή περισσότερες συσκευές τείχους προστασίας που συνεργάζονται μεταξύ τους. Αυτό σημαίνει ότι όταν το ένα τείχος προστασίας δε λειτουργεί, το άλλο θα ενεργοποιείται και θα αρχίσει να επεξεργάζεται ολόκληρη την κίνηση στο δίκτυο. Το πιο απλό τείχος προστασίας είναι ένα φίλτρο πακέτων, ενώ αυτό που φέρει την τελευταία τεχνολογία είναι τα τείχη εφαρμογών (application firewalls), τα οποία καλύπτουν όλα τα επίπεδα του μοντέλου ISO/OSI (τα εφτά επίπεδα αρχιτεκτονικής που διέπουν ένα σύστημα επικοινωνιών). [9]

[9] https://www.studytonight.com/computer-networks/complete-osi-model

Ασφάλεια λογισμικού στα κέντρα HPC (5/10)

2. Λογισμικό προστασίας από ιούς – Antivirus software: Οι απειλές ασφαλείας ενδέχεται επίσης να έρθουν σε μορφή κακόβουλου λογισμικού (malware software), όπως λογισμικά που χρησιμοποιούν οι hackers για να διακόψουν τη λειτουργία του υπολογιστή, για να συλλέξουν ευαίσθητες πληροφορίες ή για να αποκτήσουν πρόσβαση σε συστήματα υπολογιστών. Τα κακόβουλα λογισμικά μπορεί να είναι ιοί, λογισμικά παρακολούθησης, λογισμικά καταγραφής πληκτρολογίων, σκουλήκια (worms) ή τα Trojan horses. Ένα λογισμικό καταπολέμησης κακόβουλου λογισμικού αποκαλείται λογισμικό προστασίας από ιούς, παρά το γεγονός ότι εντοπίζει και εξουδετερώνει τα περισσότερα είδη κακόβουλου λογισμικού. Για αυτό το λόγο, η συνεχής ενημέρωσή τους είναι ζωτικής σημασίας, καθώς συνεχώς δημιουργούνται νέα κακόβουλα λογισμικά και αυτός είναι και ο λόγος για τον οποίο δεν θα προστατεύουν ποτέ πλήρως τα υπολογιστικά συστήματα. Επίσης, η λειτουργία προστασίας σε πραγματικό χρόνο, όπως υποδηλώνει και το όνομα, μπορεί να αποκλείσει απειλές κακόβουλου λογισμικού.

Ασφάλεια λογισμικού στα κέντρα HPC (6/10)

3. Συστήματα τοπικής ανίχνευσης εισβολών / πρόληψης εισβολής - Local Intrusion Detection/Intrusion Prevention Systems (IDS): Λόγω του γεγονότος ότι τα τείχη προστασίας και τα συστήματα προστασίας από ιούς δεν επαρκούν για να εξασφαλίσουν την αξιοπιστία στις υποδομές HPC, χρειάζονται να εφαρμοστούν πρόσθετα μέτρα ασφάλειας. Μία από τις λύσεις που έχουν σχεδιαστεί για να βοηθήσουν και να υποστηρίξουν τη διαχείριση της ασφαλείας είναι τα συστήματα ανίχνευσης (IDS). Τα IDS συστήματα αναλαμβάνουν την παρακολούθηση, σε πραγματικό χρόνο, κρίσιμων τμημάτων της υποδομής, με σκοπό να ανακαλυφθούν τυχόν εισβολές ή προσπάθειες εισβολής, ενημερώνοντας του διαχειριστές του συστήματος άμα υπάρχει κάποια παραβίαση. Υπάρχουν δύο βασικές κατηγορίες IDS:
   - Τα συστήματα ανίχνευσης εισβολής τα οποία βασίζονται σε κεντρικό υπολογιστή (Host-based intrusion detection system - HIDS)
   - Τα συστήματα ανίχνευσης εισβολής με βάση το διαδίκτυο (Network-based intrusion detection system - NIDS)

Ασφάλεια λογισμικού στα κέντρα HPC (7/10)

4. Προστασία από επιθέσεις άρνησης υπηρεσίας - Distributed Denial of Service protection: Μια επίθεση άρνησης υπηρεσίας (DDoS) είναι μια επίθεση που διεξάγεται ταυτόχρονα από πολλές διαφορετικές τοποθεσίες. Υπάρχουν διάφοροι τύποι επιθέσεων DDoS: από μια πολύ απλή και ξεπερασμένη (γνωστή ως Ping of Death) επίθεση, σε μία πιο σύνθετη και εξελιγμένη, την οποία απαρτίζουν μια σειρά από μεθόδους επίθεσης, οι οποίες χρησιμοποιούν τεχνικές ενίσχυσης. Η προστασία από επιθέσεις DDoS δεν είναι εύκολη υπόθεση. Ενώ μια επίθεση μικρής κλίμακας μπορεί να μετριαστεί με τη χρήση λογισμικού, όπως τα iptables ή και ακόμα υιοθετώντας εξειδικευμένες λύσεις υλικού (π.χ. firewall με 10Gb / s διεπαφές δικτύου), το πρόβλημα γίνεται πολύ πιο περίπλοκο με την αύξηση του όγκου της επίθεσης. Σε περίπτωση επίθεσης, χρειάζεται να συμμετέχουν οι πιο ισχυρές συσκευές δικτύου, προκειμένου να προστατευτεί η εσωτερική υποδομή, με κύριο στόχο να σταματήσει η επίθεση στην περίμετρο του δικτύου, χωρίς να σπαταληθούν πόροι του εσωτερικού δικτύου. Επιπλέον, τα HPC συστήματα μπορούν να επωφεληθούν από το γεγονός ότι είναι διασυνδεδεμένοι και γεωγραφικά διαχωρισμένοι, κάτι το οποίο μπορεί να βοηθήσει στην αποτελεσματική διαχείριση των πόρων τους.

Ασφάλεια λογισμικού στα κέντρα HPC (8/10)

5. "Μαγνήτες" – Honeypots: Είναι ένας αμυντικός μηχανισμός ασφάλειας ο οποίος ουσιαστικά αποτελεί παγίδα για τους επιτιθέμενους, με σκοπό να ανακαλύψει και να απωθήσει μη-εξουσιοδοτημένες προσπάθειες χρήσης των συστημάτων. Γενικά, αποτελείται από έναν εξυπηρετητή ο οποίος φαίνεται ως μέρος του δικτύου, αλλά στην πραγματικότητα είναι απομονωμένος και παρακολουθείται.

Ασφάλεια λογισμικού στα κέντρα HPC (9/10)

6. Λογισμικό για τη διαφύλαξη των δεδομένων από διαρροή ή απώλεια (Data Loss Prevention / Data Leakage Prevention software – DLP): Το DLP είναι ένας μηχανισμός που έχει σχεδιαστεί για τον έλεγχο των μεταφορών δεδομένων από ένα προστατευμένο σύστημα, σε κάποιο εξωτερικό (δημόσια). Είναι ειδικό στον εντοπισμό και, σε ορισμένες περιπτώσεις, στην αποτροπή ενδεχόμενης διαρροής δεδομένων. Υπάρχουν διάφοροι τύποι συστημάτων DLP. Ορισμένοι από αυτούς λειτουργούν σε επίπεδο διακομιστή και άλλοι σε επίπεδο δικτύου. Τα DLP σε επίπεδο δικτύου κινείται με βάση την ανάλυση της κυκλοφορίας των δεδομένων: όλα τα δεδομένα που μεταφέρονται μέσω του συστήματος DLP συγκρίνονται με τα πρότυπα δεδομένων που καταγράφονται στο σύστημα DLP, έτσι ώστε να ανακαλυφθούν τα ευαίσθητα δεδομένα μέσα στο μεταφερόμενο μήνυμα.

Ασφάλεια λογισμικού στα κέντρα HPC(10/10)

7. Εξουσιοδότηση – authentication: Η πιο συνηθισμένη τακτική εξουσιοδότησης, είναι αυτής της χρήσης συνθηματικού και κωδικού, η οποία όμως δεν είναι και η πιο ασφαλής. Μία από τις πιο καινούργιες τεχνολογίες που εφαρμόζονται είναι αυτή της μίας χρήσης κωδικού, ο οποίος είναι έγκυρος μόνο για μία συνεδρία. Αυτό αποτρέπει επαναλαμβανόμενες επιθέσεις, καθώς ακόμα και εάν ο επιτιθέμενος αποκτήσει πρόσβαση, ο κωδικός δεν θα είναι έγκυρος και δεν θα μπορεί να χρησιμοποιηθεί. Η δημιουργία κωδικού, απαιτεί ειδική συσκευή ή λογισμικό που θα παράγει τα κλειδιά. Μία άλλη πολύ γνωστή τεχνική είναι η ασύμμετρη κρυπτογραφία, σύμφωνα με την οποία ο χρήστης χρειάζεται ένα δημόσιο και ένα ιδιωτικό κλειδί για την εξουσιοδότηση.

Φυσική ασφάλεια (1/6)

- Τα Κέντρα Δεδομένων & Υπηρεσιών χρειάζεται να διαθέτουν συστήματα πυρανίχνευσης, πυρόσβεσης, ελέγχου πρόσβασης, συναγερμού και βιντεοεπιτήρησης που εγγυώνται την προστασία του προσωπικού ελέγχου και επιτήρησης, και του εξοπλισμού.

- Η πυρανίχνευση των χώρων μπορεί να γίνει με συστήματα πρόωρης ανίχνευσης πυρκαγιάς μέσω αναρρόφησης και δειγματοληψίας αέρα από τον χώρο, μέσω ειδικών σωληνώσεων που διατρέχουν τους χώρους (ψευδοπάτωμα, ψευδοροφή και κυρίως χώρος).

Φυσική ασφάλεια (2/6)

- Για την πυρόσβεση προτείνεται χρήση αναγνωρισμένου από διεθνείς οργανισμούς πιστοποίησης κατασβεστικού υλικού (NOVEC 1230), κατάλληλου για χώρους όπου υπάρχει παρουσία εργαζομένων και ηλεκτρολογικού/μηχανολογικού εξοπλισμού μηχανογράφησης. Είναι ασφαλές για τον άνθρωπο, καθώς δεν εκτονώνεται σε υψηλές πιέσεις και δεν μειώνει σημαντικά την περιεκτικότητα σε οξυγόνο του δωματίου.

Φυσική ασφάλεια (3/6)

- Επίσης, για την παρακολούθηση των χώρων προτείνεται μηχανισμός απομακρυσμένης πρόσβασης και παρακολούθησης από βλαβοληπτικό κέντρο, το οποίος λειτουργεί 24×7.
- Το σύστημα παρακολούθησης χρειάζεται να είναι εφοδιασμένο με κάμερες παρακολούθησης, προκειμένου να υπάρχει συνεχόμενη εποπτεία στους χώρους του κέντρου δεδομένων.
- Ενδιάμεσα στις κάμερες και την οθόνη υπάρχει μια ψηφιακή συσκευή καταγραφής (Digital Video Recorder ή DVR) η οποία αναλαμβάνει την καταγραφή της εικόνας και την αποθήκευση της. Τα καταγραφικά είναι εξοπλισμένα με ειδικό σκληρό δίσκο σαν αυτόν που διαθέτουν οι ηλεκτρονικοί υπολογιστές για την αποθήκευση την εικόνας.

Φυσική ασφάλεια (4/6)

- Τα σύγχρονα DVR παρέχουν την δυνατότητα μετάδοσης της εικόνας μέσω ιντερνέτ επιτρέποντας μας να παρακολουθούμε τις κάμερες από απόσταση (από οποιοδήποτε σημείο υπάρχει πρόσβαση στο ιντερνέτ) μέσω υπολογιστή, τάμπλετ ή ακόμα και κινητού τηλεφώνου.

- Οι τύποι κάμερας που μπορούν να χρησιμοποιηθούν είναι κάμερες Θόλου (Dome), κάμερες Σφαίρα (Bullet) και κάμερες νυκτός (Infra Red κάμερες – υπέρυθρος φωτισμός ή μη ορατός φωτισμός).

Φυσική ασφάλεια (5/6)

- Για τον έλεγχο πρόσβασης μέσα στους φυσικούς χώρους, υπάρχουν συστήματα Ελέγχου Πρόσβασης – Access Control.

- Σκοπός ενός συστήματος Ελέγχου Πρόσβασης είναι η διαβάθμιση ενός ή περισσοτέρων φυσικών ή λογικών χώρων, όπου η είσοδος/πρόσβαση επιτρέπεται μόνο σε διαβαθμισμένα άτομα τα οποία το σύστημα Ελέγχου Πρόσβασης αναγνωρίζει και πιστοποιεί

- Η διαδικασία της πιστοποίησης ατόμων από ένα σύστημα Ελέγχου Πρόσβασης επιτυγχάνεται με διάφορους τρόπους, όπως:
  - Έλεγχος Πρόσβασης με Χρήση Κωδικού PIN
  - Έλεγχος Πρόσβασης με Χρήση Έξυπνης Κάρτας
  - Έλεγχος Πρόσβασης με Χρήση Βιομετρικών Χαρακτηριστικών του ατόμου - Βιομετρία. (Δακτυλικό αποτύπωμα, Ίριδα, Φωνή, Αναλογίες Προσώπου)
  - Έλεγχος Πρόσβασης με συνδυασμό των παραπάνω

Φυσική ασφάλεια (6/6)

- Τα **συστήματα ελέγχου πρόσβασης** - **access control**, έχουν χρήση όταν θέλουμε νά μην έχει πρόσβαση ο οποιοσδήποτε σε κάποιο χώρο, πού δεν θέλουμε εμείς, η να μπορούμε να έχουμε παρουσία των ατόμων εισόδου και εξόδου. Αυτό σημαίνει ότι μπορούμε μέσω κάποιων στοιχείων προσωπικών ανά χρηστή (πχ κωδικός η κάρτα πρόσβασης) να παρέχουμε πρόσβαση σε κάποιους χώρους μόνο και σε κάποιους άλλους χώρους να μην μπορούν να εισέρχονται. Επίσης, να μπορούμε να έχουμε σύστημα καταγραφής της στιγμής και του ατόμου που εισέρχεται και εξέρχεται από ένα χώρο. Αυτό είναι ιδιαίτερα χρήσιμο σε χώρους όπως οι χώροι εγκατάστασης των κέντρων δεδομένων, για να έχουμε ένα τρόπο χωρομέτρησης παρουσίας του κάθε άτομου - εργαζομένου.

Υποστήριξη και ανάπτυξη(1/2)

- Υποθέτουμε ότι θέλουμε να αγοράσουμε ένα HPC cluster. Από πού ξεκινάμε; Υπάρχει έμπειρο τεχνικό επιτελείο να το υποστηρίξει; Υπάρχουν οι συνθήκες για το συντηρίσουμε; Είναι αρκετά τα λεφτά για να αγοραστεί;

- Στις περισσότερες περιπτώσεις, κάποια από αυτά δεν τα έχουμε. Όσον αφορά το τεχνικό κομμάτι, χρειάζεται συμβουλή ή υποστήριξη από κάποιες εναλλακτικές εκπαίδευσης και καθοριστική είναι η συμβολή του διαδικτύου.

- Υπάρχει μία πληθώρα επιλογών και πληροφοριών για τα HPC clusters στο διαδίκτυο.

Υποστήριξη και ανάπτυξη(2/2)

- Επιπλέον, είναι αναγκαία η υποστήριξη από διάφορες κοινότητες και λίστες email, όπου μπορούν να απαντηθούν πολλά ερωτήματα και να συζητηθούν πολλά θέματα.

- Επιπλέον, είναι σημαντική η επικοινωνία και με πωλητές λογισμικού, οι οποίοι προσφέρουν HPC Clusters και αρκετές λύσεις.

- Οι συγκεκριμένοι πωλητές μπορεί να αποδειχθούν αρκετά βοηθητικοί, καθώς σε πολλές περιπτώσεις διαθέτουν δικές τους υλοποιήσεις λογισμικών ή cluster.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το λογισμικό(1/4)

- Η επιλογή του κατάλληλου λογισμικού εξαρτάται από το εάν θα είναι εμπορικό ή ελεύθερο.
- Σε περίπτωση που είναι εμπορικό, μέσω των ISV (ανεξάρτητων πωλητών λογισμικού) μπορούν να υπάρξουν πολλές διευκολύνσεις, καθώς σίγουρα θα είχαν ή θα έχουν πελάτες οι οποίοι θα αντιμετώπισαν προβλήματα όπως ενός πρωτοεισερχόμενου σε αυτόν τον τομέα και θα διαθέτουν έτοιμες λύσεις.
- Επιπροσθέτως, είναι πολύ πιθανό να διαθέτουν λύσεις όσον αφορά την επιλογή του τύπου του υλικού και του λογισμικού που ταιριάζει καλύτερα στον cluster.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το λογισμικό(2/4)

- Με τόσα πολλά κομμάτια να συνθέτουν τον cluster, η συμβατότητα παίζει καθοριστική σημασία.
- Στην περίπτωση του ανοιχτού λογισμικού, ίσως είναι αναγκαίο να γίνει εμβάθυνση στην έρευνα, με την έννοια ότι πρέπει να διευκρινιστεί τι τύπους σχεδίασης λειτουργεί καλύτερα για μία εφαρμογή λογισμικού.
- Είναι σίγουρο πως οι λύσεις θα είναι αρκετές και θα έρχονται σε σύγκρουση πολλές φορές μεταξύ τους, όμως η σωστή κρίση και ο συνδυασμός των κομματιών του cluster μπορούν να δώσουν τη σωστή λύση.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το λογισμικό(3/4)

- Επιπρόσθετα με το λογισμικό εφαρμογής, χρειάζεται να εξεταστεί και η δομή του λογισμικού. Σε πολλές περιπτώσεις θα γίνει χρήση των Linux.
- Είναι σχετικά εύκολο να εγκατασταθεί σε έναν cluster λογισμικό Linux το οποίο περιλαμβάνει λογισμικό clustering στους κόμβους.
- Η επιτυχημένη εγκατάσταση μπορεί να έρθει αμέσως, όμως σε βάθος χρόνου μπορεί να δημιουργηθούν αρκετά προβλήματα εξαρτήσεων. Οπότε, ο σωστός προγραμματισμός από την αρχή μέχρι το τέλος είναι ζωτικής σημασίας.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το λογισμικό(4/4)

- Υπάρχουν άλλα θέματα τα οποία προκύπτουν όταν χρησιμοποιείται do-it-yourself (κάντο μόνος σου) λογισμικό συστήματος στους clusters.
- Αρχικά, πόσο καλά δουλεύει σε όλες τις κλίμακες; Για παράδειγμα, κάτι το οποίο δουλεύει για 8 διακομιστές, μπορεί να αποτύχει για 128 διακομιστές.
- Δεύτερον, πρέπει να ληφθεί υπόψιν πως θα αναβαθμιστεί και θα διατηρηθεί το λογισμικό σε ολόκληρο τον cluster.
- Παρόλο που αυτά είναι βασικά θέματα για έναν διαχειριστή συστήματος, είναι εκπληκτικό πως πολλοί διαχειριστές καταφέρνουν και τα φέρνουν εις πέρας με όλα αυτά τα προβλήματα που έρχονται στην επιφάνεια στους clusters.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το υλικό(1/2)

- Αφού υπάρξει πρώτα μία γενική έρευνα για τον τύπου το υλικού, χρειάζεται να ξεκινήσει η σκέψη για τα υλικά που θα χρειαστούν.

- Και σε αυτήν την περίπτωση, είναι καταλυτική η βοήθεια των πωλητών υλικού για εξέλιξη των μονάδων, προκειμένου να υπάρξει και μια σαφέστερη άποψη που θα βοηθήσει στην αξιολόγηση από τον ίδιο τον υποψήφιο αγοραστή.

- Η διαδικασία της σωστής αξιολόγησης όλων των προβλημάτων και των υποθέσεων, είναι αναγκαία πριν την οποιαδήποτε επένδυση χρημάτων στο υλικό.

Υποστήριξη και ανάπτυξη: Αξιολογώντας το υλικό(2/2)

- Αμέσως μετά, είναι σημαντική μία συνοπτική και συγκεκριμένη εισήγηση, η οποία θα αφορά τις ανάγκες του αγοραστή. Οι ανάγκες αυτές θα αξιολογηθούν και θα αναλυθούν από υποψήφιους πωλητές οι οποίοι θα αναλάβουν την υποστήριξη της εκπλήρωσης των στόχων της δημιουργίας ενός HPC.

- Μερικές εταιρίες, μπορούν επίσης να υποστηρίξουν την διευθέτηση των επιχειρηματικών αναγκών του αγοραστή.

- Για παράδειγμα, η Sun Microsystems, έχει κέντρα επίλυσης ανά τον κόσμο για να βοηθήσεις σε τέτοια προβλήματα. Περισσότερα, στο www.sun.com/solutionscenters/index.jsp

Εκτιμώμενα κόστη(1/5)

- Οι τιμές δημιουργίας ενός HPC ποικίλουν, ανάλογα το συνδυασμό των υλικών μερών και του λογισμικού που θα χρησιμοποιηθεί. Οι τιμές μπορούν να ξεκινήσουν από 5.000€ (απλοί servers) και να φτάσουν ακόμα και σε ποσά που αγγίζουν το 1.000.000€.

- Ανάλογα με τις απαιτήσεις της κάθε εταιρείας που χρειάζεται έναν υπερ-υπολογιστής, προκύπτουν και τα ανάλογα ποσά.

- Παρακάτω θα αναλυθούν μερικά διαφορετικά παραδείγματα clusters, αποτελούμενα από διαφορετικά κομμάτια και τελικές τιμές αγοράς.

Εκτιμώμενα κόστη(2/5)

Παράδειγμα 1° (**ACTserv x2210)**:

Base system: Dual socket Xeon SP 2U system with 8x 2.5" drive bays

Processors: 2x Intel 12-Core Xeon Silver 4214 2.2GHz - 85W

Memory: 96GB - 6x 16GB DDR4 2933MHz

Storage configuration: 8x 2.5" SATA drives (software RAID only)

Boot / OS drive location: Installed inside the system, not externally accessible

Networking: 2x RJ45 10Gb ethernet ports

OCP networking expansion: None

GPU configuration: 2x GPUs

Management: Remote iKVM with in-band management

Power supply: Dual 1300W PSU (redundant)

Power cables: 2x No power cords

Warranty: 1 year standard warranty

A/C: Inventor V5MFI32-60/V5MFO32-6, 49817 BTU

Τελική τιμή: 8,113.00€

Πηγή: https://www.advancedclustering.com/act_systems/actserv-x2210/

Στο συγκεκριμένο παράδειγμα, χρειάζεται να συνυπολογίσουμε και το προσωπικό το οποίο χρειάζεται για τη συντήρηση του server. Αν υποθέσουμε ότι χρειαζόμαστε έναν υπεύθυνο μηχανικό για την απόδοση του συστήματος και του λογισμικού και έναν υπεύθυνο μηχανικό για τη λειτουργία του διαδικτύου, τότε χρειαζόμαστε τουλάχιστον επιπλέον κεφάλαια της τάξης των 2.000€ – 2.500€ τον μήνα για κάθε μηχανικό. Δηλαδή σύνολο 4.000€ - 5.000€ κάθε μήνα + τα έξοδα ασφάλισης.

Εκτιμώμενα κόστη(3/5)

Παράδειγμα 2ο **(ACTblade x210):**

2U enclosure with room for 4x independent Compute Block nodes

Storage:1x 2.5" drive per node (4 drives in enclosure)

Power supply:Redundant 2130W PSU (requires > 200V input power)

Nodes:4x Nodes

Base system:Dual socket Xeon SP blade system with 1x 2.5" drive bay

Processor:2x Intel 12-Core Xeon Silver 4214 2.2GHz - 85W

Memory:96GB - 6x 16GB DDR4 2933MHz

Storage configuration:1x 2.5" drive bay, 1x 42mm M.2, 1x 80mm M.2

Networking:2x RJ45 10Gb ethernet ports

Expansion slot:Empty PCI-e slot

Management:Remote iKVM with in-band managementWarranty:

1 year standard warranty

A/C: Daikin FVQ71C / RZQG71L9V1, 22178 BTU

Τελική τιμή: 18,777€

Πηγή:https://www.advancedclustering.com/act_systems/actblade-x210/

Και σε αυτό το παράδειγμα χρειάζεται να συνυπολογιστεί ένα επιπλέον ποσό, το οποίο είναι ανεξάρτητο της τιμής αγοράς του rack και του server. Δεδομένου ότι το μέγεθος είναι όπως και του προηγούμενου παραδείγματος, τότε χρειαζόμαστε έναν υπεύθυνο μηχανικό για την απόδοση του συστήματος και του λογισμικού και έναν υπεύθυνο μηχανικό για τη λειτουργία του διαδικτύου, τότε χρειαζόμαστε τουλάχιστον επιπλέον κεφάλαια της τάξης των 2.000€ – 2.500€ τον μήνα για κάθε μηχανικό. Δηλαδή σύνολο 4.000€ - 5.000€ κάθε μήνα + τα έξοδα ασφάλισης.

Εκτιμώμενα κόστη(4/5)

- Στην περίπτωση της δημιουργίας ενός ολόκληρου κέντρου δεδομένων, τα ποσά ξεφεύγουν αρκετά. Χρειαζόμαστε να συνυπολογίσουμε τους εξής παράγοντες:
  - Το κτίριο που θα στεγάζεται το κέντρο δεδομένων (100.000€ - 300.000€)
  - Τους server (cpu, gpu) και τις αποθηκευτικές μονάδες (ανάλογα με τους κόμβους, 100.000€ για περίπου 10 κόμβους)
  - Τις άδειες για τα λογισμικά και τις εφαρμογές (10.000€)
  - Την απαραίτητη ενέργεια (Η τιμή της κιλοβατώρας για κατανάλωση ισχύος άνω των 2000 κιλοβατώρων είναι 0,21985€ το τετράμηνο. Επομένως, η 1 μεγαβατώρα κοστίζει 219,85€ το τετράμηνο και 659,55€ το έτος).
  - Τη συνδεσιμότητα στο διαδίκτυο (Ένα χιλιόμετρο οπτικών ινών κοστίζει μέχρι και 250.000€)
  - Την ασφάλεια των χώρων (10.000€)
  - Την ψύξη των χώρων (10.000€)
  - Το προσωπικό που είναι απαραίτητο, όπου σε ένα ολόκληρο κέντρο δεδομένων μπορεί να είναι απαραίτητο να καλυφθούν 13 διαφορετικές ειδικότητες με μηνιαίο μισθό + κόστη ασφάλισης που ξεπερνάνε τα 3.000€ το άτομο.

Εκτιμώμενα κόστη(5/5)

- Έτσι λοιπόν, τα ποσά που είναι απαραίτητα για τη δημιουργία ενός κέντρο δεδομένου μπορεί να αγγίξουν ακόμα και το 1.000.000€ + μηνιαία κόστη για μισθούς και ασφαλίσεις του προσωπικού.

- Όλα είναι εξαρτώμενα από τις επιλογές που θα γίνουν για τη δημιουργία του κέντρου δεδομένων. Είναι σημαντικό όμως να καλυφθούν τουλάχιστον οι παράγοντες που αναλύθηκαν προηγουμένως στα κόστη (διαφάνεια 176)

Συντήρηση(1/2)

- Αφού δημιουργηθεί ο cluster, χρειάζεται να υπάρχει και η πρόβλεψη για τη συντήρηση και τη διατήρηση της ακεραιότητας του. Τι θα γίνει σε περίπτωση αστοχίας κάποιου εξαρτήματος, όπως π.χ. άμα καταστραφεί ο δίσκος που αποθηκεύονται τα δεδομένα ή χαλάσει κάποιος από τους πυρήνες;

- Αρχικά, ο αγοραστής ενός cluster πρέπει να διασφαλίσει την εγγύηση του πωλητή και την αδιάκοπη υποστήριξη τόσο του συστήματος, όσο και στην επίλυση των προβλημάτων που προκύπτουν από την πλευρά του πωλητή .

- Από την πλευρά του ο αγοραστής, πρέπει συνεχώς να επιβλέπει το κέντρο δεδομένων του, να επαληθεύει ότι όλα λειτουργούν σωστά και να απομακρύνει τυχόν σκόνες και σκουπίδια που μπορούν βρίσκονται πάνω στα συστήματα και να δυσκολέψει τη λειτουργία τους (π.χ. η μη απομάκρυνση της σκόνης από τον επεξεργαστή μπορεί να οδηγήσει σε υπερθέρμανση).

Συντήρηση(2/2)

- Ωστόσο, καλή επίβλεψη προκειμένου να υπάρχει σωστή συντήρηση σημαίνει ότι πρέπει να γίνει πρόσληψη του κατάλληλου προσωπικού για την αντιμετώπιση τέτοιων προβλημάτων.

- Η επένδυση σε προσωπικό συντήρησης και αποκατάστασης των clusters είναι σημαντική διαδικασία. Από τη στιγμή που υπάρχουν τόσα εξαρτήματα που συνθέτουν τους clusters και συνολικά τα κέντρα δεδομένων (επεξεργαστές, μνήμες, αποθηκευτικά μέσα, λογισμικά, τροφοδοτικά, racks, λογισμικά προστασίας, δικτύωση), χρειάζεται επίβλεψη από πολλά άτομα και πρόβλεψη για τυχόν αστοχίες.

- Άμα κάποιος δίσκος δυσλειτουργήσει, πρέπει να υπάρχουν αντίγραφα ασφαλείας, άμα καταστραφεί κάποιος πυρήνας, χρειάζεται να δουλέψουν οι υπόλοιποι, σε περίπτωση διακοπής ρεύματος, να υπάρχουν γεννήτριες οι οποίες θα είναι ικανές να υποστηρίξουν όλη την υποδομή του κέντρο δεδομένων.

- Άλλωστε, η πρόληψη είναι και η καλύτερη θεραπεία, καθώς έτσι αποφεύγονται απώλειες δεδομένων, καταστροφές εξαρτημάτων και μειώνεται το κόστος συντήρησης.

Προσωπικό που απασχολείται και αναγκαίες δεξιότητες(1/5)

- Σύμφωνα με την αγορά εργασίας και τις απαιτήσεις ενός τέτοιου τομέα, που ασχολείται με τους υπερυπολογιστές, υπάρχουν αρκετές θέσεις στις οποίες μπορεί κάποιος να κάνει αίτηση προκειμένου να εργαστεί.

- Οι περισσότερες εταιρίες και εργοδότες, αναζητούν μηχανικούς ηλεκτρονικών υπολογιστών, με προϋπηρεσία μεγαλύτερη των 5 (πέντε) χρόνων και με πολλές δεξιότητες και γνώσεις.

- Οι ευθύνες ανάληψης ενός τέτοιου πόστου είναι μεγάλες, επομένως οι εργοδότες κάνουν προσεκτική επιλογή του εργατικού τους δυναμικού.

Προσωπικό που απασχολείται και αναγκαίες δεξιότητες(2/5)

- Οι θέσεις οι οποίες χρειάζονται συνήθως κάλυψη είναι μία από τις παρακάτω [13]:
  - Υπεύθυνος μηχανικός για την απόδοση του συστήματος (Performance Engineer)
  - Μηχανικός Έρευνας και Ανάπτυξης (Research and Development Engineer)
  - Μηχανικός Μοντελοποίησης της απόδοσης του συστήματοςο (Performance Modeling Engineer)
  - Μηχανικός Διαδικτύων (Network Engineer)
  - Μηχανικός λογισμικού (Software Engineer)
  - Υπεύθυνος Μηχανικός Υπολογιστικής Υψηλών Επιδόσεων (High Performance Computing Engineering Manager)
  - Υπεύθυνος Μηχανικός Απόδοσης Εφαρμογών (Application Performance Engineer)
  - Μηχανικός Ισχύος και Απόδοσης (Power and Performance Engineer)
  - Υπεύθυνος Μηχανικός Αναλύσεων Απόδοσης (Performance Analytics Manager)
  - Ηλεκτρολόγος Μηχανικός (Electrical Engineer)
  - Μηχανολόγου Μηχανικού (Mechanical Engineer)
  - Υπεύθυνος Μηχανικός απόδοσης της Υποδομής (Infrastructure Performance Engineer)

[13] https://www.indeed.com/jobs?q=High+Performance+Computing+Engineer

Προσωπικό που απασχολείται και αναγκαίες δεξιότητες(3/5)

- Όπως φάνηκε και από την προηγούμενη διαφάνεια, υπάρχει ένα μεγάλο φάσμα από επιλογές θέσεων εργασίας, οι οποίες είναι όμως απαιτητικές ως προς τις δεξιότητες και τις γνώσεις. Ενδεικτικά, συνήθως απαιτείται:
  - Εμπειρία στη διαχείριση των συστημάτων Debian και Red Hat Linux, συμπεριλαμβανομένης της εγκατάστασης, διαμόρφωσης και συντήρησης των υπηρεσιών Linux (DNS, DHCP, LDAP, VPN, κ.λπ.) και δικτύωσης Linux (πρωτόκολλα TCP/IP).
  - Δυνατότητα αντιμετώπισης προβλημάτων υλικού και λογισμικού (ακόμα και σε μεγάλη κλίμακα, όπως σε περιβάλλοντα υψηλής υπολογιστικής απόδοσης).
  - Δεξιότητες στην ανάλυση και αντιμετώπιση προβλημάτων
  - Εμπειρία scripting σε perl / python / bash και προγραμματισμό σε C/ C++ / Fortran / CUDA
  - Γνώση παράλληλου προγραμματισμού
  - Εμπειρία με τους πυρήνες (kernel) Linux, κυρίως για τις GPU της NVIDIA και τις κάρτες Mellanox InfiniBand
  - Εμπειρία χρήσης των MPI και openMP, όπως επίσης και εξοικείωση και αντιμετώπιση προβλημάτων των μεταγλωττιστών της Intel και του λειτουργικού Linux

Προσωπικό που απασχολείται και αναγκαίες δεξιότητες(4/5)

- Εμπειρία με την εγκατάσταση και τη διαχείριση των συστημάτων αρχείων Luster και ZFS
- Γνώσεις όσον αφορά την αρχιτεκτονική των υπολογιστών σε μικρή κλίμακα, όπως για παράδειγμα σε υποσυστήματα επεξεργαστών και μνημών
- Εμπειρία στη μετατροπή ισχύος από AC σε DC και το αντίστροφο
- Εμπειρία στο σχεδιασμό του συστήματος παροχής ισχύος του κέντρου δεδομένων, στις μεθοδολογίες συστημάτων διανομής ισχύος και στις απαιτήσεις σχεδιασμού ασφάλειας ισχύος.
- Εμπειρία στα μεγάλα κλίμακας δίκτυα, στο σχεδιασμό και την υλοποίηση δικτύων μεγάλης κλίμακας και στη διαχείριση υποδομών επικοινωνιών.

Προσωπικό που απασχολείται και αναγκαίες δεξιότητες(5/5)

- Αυτές είναι μερικές από τις απαιτήσεις που υπάρχουν στην αγορά εργασίας. Είναι προφανές ότι χρειάζονται στο ελάχιστο μεταπτυχιακές σπουδές σε αντίστοιχα τμήμα μηχανικών υπολογιστών και δικτύων, ηλεκτρολόγων μηχανικών και μηχανολόγων μηχανικών.

- Επιπλέον, υπάρχουν μεγάλες απαιτήσεις στην προϋπηρεσία των υποψηφίων εργαζομένων, ενώ θεωρούνται δεδομένες δεξιότητες ομαδικής δουλειάς, επικοινωνίας με άλλους και ηγετικές ικανότητες, για ανάληψη έργων.

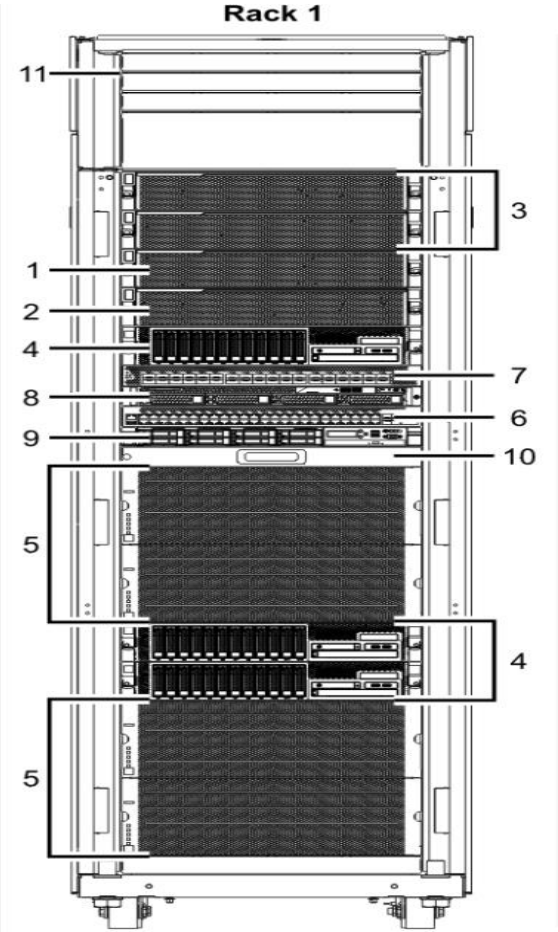
[14] https://www.icts.res.in/opportunities/hpc-2016-12-16

Παράδειγμα χωροθέτησης Hpc cluster (1/3)

- Το παράδειγμα που θα ακολουθήσει στις επόμενες διαφάνειες, παρουσιάζει τη χωροθέτηση ενός IBM HPC Cluster σε πολλαπλά Rack.

- Ο cluster αποτελείται από τρία rack, τα οποία συνδέονται μεταξύ τους και συνθέτουν ολόκληρο τον υπερ-υπολογιστή.

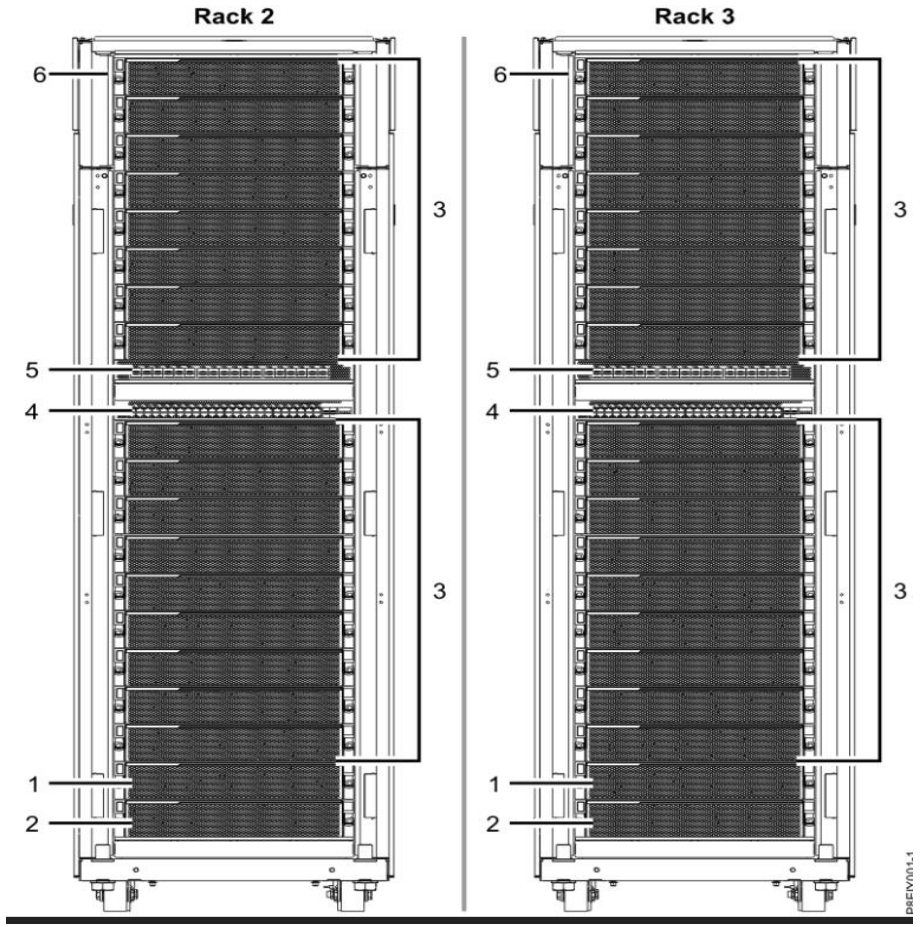
Παράδειγμα χωροθέτησης Hpc cluster (2/3)



Αριθμός	Περιγραφή
1	Κόμβος Σύνδεσης
2	Κόμβος Διαχείρισης
3	Κόμβος Υπολογισμών
4	Διακομιστής αποθήκευσης
5	Συσκευές αποθήκευσης
6	Ethernet Switch (RJ45 θύρες με 1 Gb + τέσσερις SFP+ θύρες με 10 Gb υποστήριξη
7	InfiniBand switch (100 Gb)
8	Ενιαίος Διαχειριστής δομής
9	Κονσόλα Διαχείρισης Hardware
10	Κονσόλα του rack
11	Επαγγελματικό rack

Πίνακας που περιγράφει τη χωροθέτησης των εξαρτημάτων στο Rack τύπου 1

Παράδειγμα χωροθέτησης Hpc cluster (3/3)



Αριθμός	Περιγραφή
1	Κόμβος Σύνδεσης
2	Κόμβος Διαχείρισης
3	Κόμβος Υπολογισμών
4	Ethernet Switch (RJ45 θύρες με 1 Gb + τέσσερις SFP+ θύρες με 10 Gb υποστήριξη
5	InfiniBand switch (100 Gb)
6	Επαγγελματικό rack

Πίνακας που περιγράφει τη χωροθέτησης των εξαρτημάτων στα Rack τύπου 2 και τύπου 3

Πηγή:
https://www.ibm.com/support/knowledgecenter/en/P8HPC/p8eiy/p8eiy_example_multi_rack_max_compute_config.htm

ARIS – Το καμάρι της Ελλάδας (1/8)

- Το **ARIS (Advanced Research Information System)** είναι το ισχυρότερο υπολογιστικό σύστημα στην Ελλάδα για επιστημονικές εφαρμογές. Τέθηκε σε λειτουργία τον Ιούλιο του 2015 από την ΕΔΕΤ Α.Ε. προσφέροντας ένα ισχυρό εργαλείο έρευνας στην Ελληνική επιστημονική κοινότητα.

- Το σύστημα κατά την έναρξη λειτουργίας του συμπεριλήφθηκε στη λίστα με τους 500 ισχυρότερους υπολογιστές του κόσμου (top500.org) και έβαλε την Ελλάδα στο παγκόσμιο χάρτη των συστημάτων υψηλών επιδόσεων. Το υπολογιστικό σύστημα ARIS σήμερα έχει μέγιστη θεωρητική υπολογιστική ισχύ 444 TFlop/s (μπορεί δηλαδή να εκτελεί 444 τρισεκατομμύρια μαθηματικές πράξεις το δευτερόλεπτο) και προσφέρει πολλαπλές δυνατότητες επεξεργασίας δεδομένων.

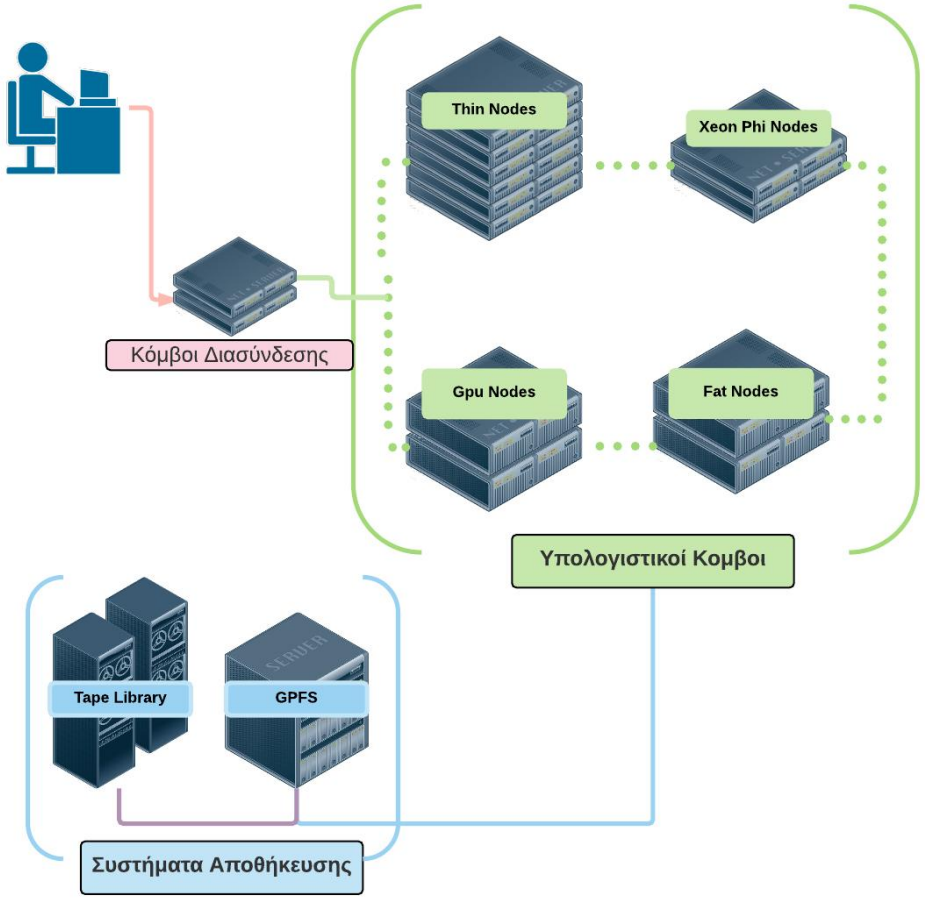
ARIS – Το καμάρι της Ελλάδας Αρχιτεκτονική του συστήματος (2/8)

- Ο ARIS συνδυάζει 4 διαφορετικές αρχιτεκτονικές διαμοιρασμένες σε αντίστοιχες "νησίδες κόμβων".
- Αναλυτικά, η υποδομή αποτελείται από:
  - Μία νησίδα η οποία διαθέτει 426 υπολογιστικούς κόμβους (**thin nodes**). Κάθε κόμβος διαθέτει δύο επεξεργαστές και κάθε επεξεργαστής περιέχει 10 επεξεργαστικούς πυρήνες προσφέροντας έτσι συνολικά 8.520 πυρήνες (CPU cores). Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές υψηλής παραλληλίας που μπορούν να σπάσουν τα δεδομένα τους σε πολλά μικρά κομμάτια πριν τα επεξεργαστούν.
  - Μια νησίδα κόμβων μεγάλης μνήμης (**fat nodes**) που αποτελείται από 44 κόμβους. Κάθε κόμβος προσφέρει 4 επεξεργαστές, 40 πυρήνες και 512 GB κεντρικής μνήμης ανά κόμβο. Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές που χρειάζονται πολύ μεγάλη κεντρική μνήμη και όχι τόσο για υψηλή κλιμάκωση.

ARIS – Το καμάρι της Ελλάδας Αρχιτεκτονική του συστήματος (3/8)

- Μια νησίδα κόμβων επιταχυντών GPU (**gpu nodes**) που αποτελείται από 44 κόμβους. Κάθε κόμβος περιέχει 2 επεξεργαστές με 10 πυρήνες ανά επεξεργαστή, 64 GB μνήμης και 2 κάρτες γραφικών GPU NVidia K40. Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές που υλοποιούν υπολογιστικές πράξεις που μπορούν να αξιοποιήσουν τις κάρτες γραφικών ως συνεπεξεργαστές για επιτάχυνση των υπολογισμών.
- Μια νησίδα κόμβων επιταχυντών Xeon Phi (**phi nodes**) που αποτελείται από 18 κόμβους, καθένας εκ των οποίων περιέχει 2 επεξεργαστές με 10 πυρήνες, 64 GB μνήμης και 2 συνεπεξεργαστές Intel Xeon Phi 7120P. Είναι κατάλληλη για παράλληλες εφαρμογές που αξιοποιούν την τεχνολογία συνεπεξεργαστών της Intel Xeon Phi.

ARIS – Το καμάρι της Ελλάδας
Αρχιτεκτονική του συστήματος (4/8)



Εικόνα η οποία δείχνει την επικοινωνία των τεσσάρων αρχιτεκτονικών των "νησίδων κόμβων". Οι υπολογιστικοί κόμβοι συνδέονται απομακρυσμένα με τους χρήστες (επιστήμονες, φοιτητές) μέσω κόμβων διασύνδεσης (χρησιμοποιούνται πρωτόκολλα για τη μεταφορά αρχείων scp ή sftp), ενώ επίσης συνδέονται με τα συστήματα αποθήκευσης του υπερυπολογιστή. Ο ARIS, για το σύστημα αρχείων του υλοποιεί την τεχνολογία General Parallel File System (GPFS) της IBM προσφέροντας 2 PetaBytes αποθηκευτικού χώρου στους χρήστες του. Επίσης, διατηρείται στην υποδομή ένα μαγνητικό σύστημα αποθήκευσης ταινίας και συγκεκριμένα το IBM TS3500, το οποίο προσφέρει αποθηκευτικό χώρο ίσο με 2 PetaBytes για την αρχειοθέτηση δεδομένων για μεγάλες χρονικές περιόδους.

ARIS – Το καμάρι της Ελλάδας Προφίλ εφαρμογών που τρέχουν στο ARIS (5/8)

- Κατάλληλες εφαρμογές για το σύστημα ARIS είναι αυτές που μπορούν να υλοποιηθούν υιοθετώντας κάποιο μοντέλο παράλληλης επεξεργασίας. Μια παράλληλη εφαρμογή κατά την εκτέλεσή της διαχωρίζεται σε εκατοντάδες ή και χιλιάδες επιμέρους διεργασίες, οι οποίες εκτελούνται ταυτόχρονα και συνεργατικά επιλύουν ένα κοινό πρόβλημα.

- Οι διεργασίες αυτές για να επιλύσουν το πρόβλημα πρέπει να έχουν πρόσβαση στα ίδια δεδομένα και να επικοινωνούν μεταξύ τους ανταλλάσσοντας αποτελέσματα. Ανάλογα με τον τρόπο που επιτυγχάνεται αυτή η επικοινωνία μπορούμε να διαφοροποιήσουμε τον τρόπο που σχεδιάζονται και αναπτύσσονται οι εφαρμογές αυτές και τα εργαλεία προγραμματισμού που πρέπει να χρησιμοποιηθούν.

- Οι γλώσσες προγραμματισμού που χρησιμοποιούνται για την ανάπτυξη παράλληλων εφαρμογών είναι στις περισσότερες περιπτώσεις, οι συνηθισμένες γλώσσες γενικού σκοπού, όπως η C/C++ ή Fortran που επεκτείνονται με εξειδικευμένες βιβλιοθήκες ώστε να υποστηρίζουν την παραλληλία MPI, OpenMP, CUDA. Οι προγραμματιστές πρέπει να σχεδιάσουν την εφαρμογή τους με τέτοιο τρόπο ώστε να λειτουργούν σε μορφή παράλληλων συνεργατικών διεργασιών χρησιμοποιώντας τις δυνατότητες που προσφέρουν οι βιβλιοθήκες MPI και OpenMP.

ARIS – Το καμάρι της Ελλάδας
Πρόσβαση και χρήση του συστήματος (6/8)

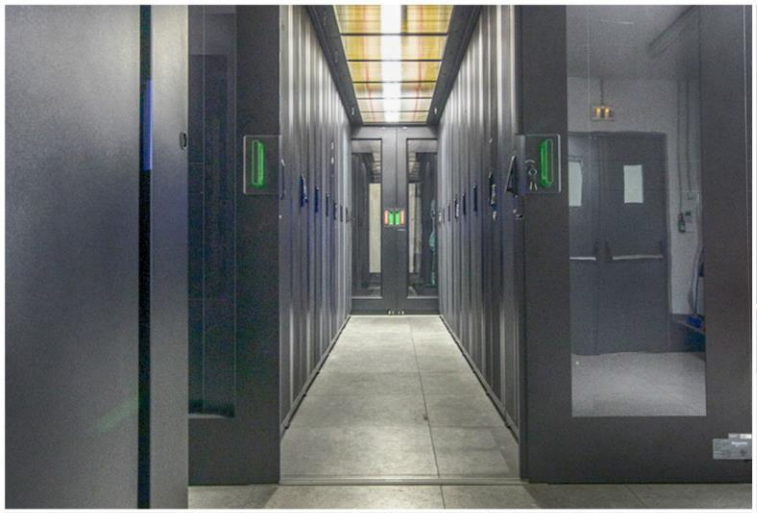
- Η πρόσβαση στο ARIS είναι ανοιχτή για όλους τους επιστήμονες και ερευνητές, οι οποίοι εργάζονται σε Ελληνικά εκπαιδευτικά ή/και ερευνητικά ιδρύματα, και γίνεται μέσω προσκλήσεων πρότασης έργου. Το ΕΔΕΤ για να εξασφαλίσει την ισότιμη και ανοιχτή πρόσβαση στο σύστημα έχει καθιερώσει την Πολιτική Πρόσβασης και Χρήσης του Εθνικού Υπερυπολογιστικού συστήματος https://hpc.grnet.gr/access/access_policy/.

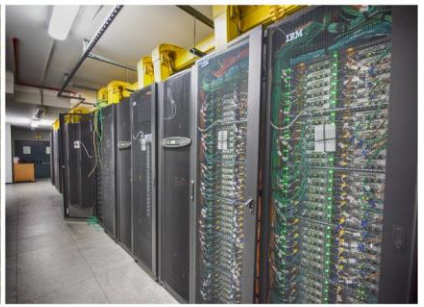
- Η πολιτική πρόσβασης ακολουθεί τα διεθνή πρότυπα και τις καλές πρακτικές που ακολουθούν τα περισσότερα υπερυπολογιστικά κέντρα του κόσμου τα οποία όπως και το ΕΔΕΤ προσφέρουν υπολογιστικούς πόρους σε επιστημονικές ομάδες.

ARIS – Το καμάρι της Ελλάδας
Πρόσβαση και χρήση του συστήματος (7/8)

- Οι ενδιαφερόμενοι επιστήμονες που δραστηριοποιούνται σε Ελληνικά ιδρύματα και χρειάζονται πρόσβαση στο ARIS θα πρέπει να καταθέσουν μια πρόταση στο πλαίσιο σχετικής πρόσκλησης του ΕΔΕΤ. Οι προσκλήσεις έργων ομαδοποιούνται σε δύο κατηγορίες:
- Τα **Έργα Παραγωγής**, τα οποία έχουν την τεχνική αρτιότητα για να εκμεταλλευτούν τους διαθέσιμους πόρους και έχουν αξιολογηθεί θετικά από ομότιμους αξιολογητές (“peer review”).
  - Πρόσκληση για την κατάθεση προτάσεων: κάθε 6 μήνες
  - Διάρκεια έργων: 12 μήνες.
  - Αξιολόγηση: Τεχνική / Επιστημονική μέσω διαδικασίας peer-review
- Τα **Έργα Προετοιμασίας/Ανάπτυξης**, τα οποία έχουν περάσει το στάδιο ανάπτυξης και χρειάζονται επικύρωση της τεχνικής αρτιότητας (π.χ. ικανότητα κλιμάκωσης), ή έχουν σκοπό τη μεταφορά και βελτιστοποίηση του κώδικά τους ώστε να λειτουργεί αποδοτικά στο σύστημα ARIS με σκοπό να μπορέσουν να προχωρήσουν σε κατάσταση παραγωγής.
  - Συνεχής Πρόσκληση για την κατάθεση προτάσεων
  - Διάρκεια έργων: 2- 4 μήνες
  - Τεχνική Αξιολόγηση
- Η κατάθεση των προτάσεων γίνεται online μέσω του ιστοχώρου της υπηρεσίας HPC του ΕΔΕΤ: https://hpc.grnet.gr/access/

ARIS – Το καμάρι της Ελλάδας
Εικόνες (8/8)






Απεικόνιση μερικών χώρων του κέντρου δεδομένων του ARIS. Διακρίνονται τα δωμάτια που περιέχουν τα rack με τους servers, αρκετές καλωδιώσεις και σκληροί δίσκοι οι οποίοι αναλαμβάνουν την αποθήκευση των δεδομένων.

Ποιος πραγματικά θα βοηθήσει;(1/5)

- Τις πρώτες μέρες του cluster computing, τα περισσότερα συστήματα ήταν do-it-yourself (DIY) τεχνικών. Ακόμα και σήμερα, η τεχνική αυτή είναι ιδιαίτερα διαδεδομένη λόγω του ιδιαίτερα μικρού κόστους.

- Πλέον όμως δεν είναι η ενδεδειγμένη λύση για να επιτευχθούν οι στόχοι δημιουργίας ενός HPC (ειδικότερα εάν γίνεται τώρα είσοδος στον κόσμο του HPC).

- Μία άλλη επιλογή είναι όπως προαναφέρθηκε, η αλληλεπίδραση με πωλητές – είτε υλικού είτε λογισμικού – προκειμένου να υπάρχει υποστήριξη για όλον τον cluster.

Ποιος πραγματικά θα βοηθήσει;(2/5)

- Στο σημερινό επιχειρηματικό κόσμο, ένας καλός HPC cluster έχει πολλά πλεονεκτήματα και αξίζει τα επιπλέον κόστη και έξοδα.
- Ένας πωλητής ο οποίος μπορεί να έχει πρόσβαση και εξειδίκευση σε ένα μεγάλο φάσμα υλικού και λογισμικού, μπορεί να είναι καθοριστικός για την δημιουργία του HPC συστήματος.
- Για παράδειγμα, ένας πωλητής ο οποίος μπορεί να προσφέρει ένα ενσωματωμένο και ελεγμένο παράλληλο σύστημα αρχείων, θα προτιμηθεί από κάποιον ο οποίος απλά το αγοράζει από άλλον πωλητή.

Ποιος πραγματικά θα βοηθήσει;(3/5)

- Επιπλέον, η μεγάλης διάρκειας υποστήριξη είναι επίσης σημαντική. Κατά τη διάρκεια χρήσης ενός cluster, είναι πολύ πιθανόν να υπάρχουν απορίες και δυσκολίες οι οποίες θα χρειάζονται να αντιμετωπιστούν.

- Τα προβλήματα αυτά μπορεί να έχουν να κάνουν με την αναβάθμιση, την τροποποίηση, και την επέκταση του cluster. Ένας καλός πωλητής, θα καταφέρει να εξυπηρετήσει πολύ καλά σε αυτήν την περίπτωση.

Ποιος πραγματικά θα βοηθήσει;(4/5)

- Ένα ακόμα θέμα, το οποίο πολλές φορές παραβλέπεται, είναι η τοπική ενσωμάτωση. Σχεδόν όλοι οι clusters, χρειάζεται να χωρέσουν μία ήδη υπάρχουσα υποδομή επεξεργασίας δεδομένων.

- Η διασφάλιση ότι ο πωλητής έχει τη δυνατότητα να υποστηρίξει τους πελάτες με αυτήν τη δυνατότητα, είναι πολύ σημαντική.

- Παρόμοιο με το "last mile" θέμα με την πρόσβαση στο διαδίκτυο, η τοπική ενσωμάτωση ίσως χρειάζεται παραπάνω δουλειά από ότι υπολογίζεται.

- Με τον όρο "last mile" εννοείται το τελευταίο στάδιο στις τηλεπικοινωνίες, όπου η υπηρεσία φτάνει στον τελικό χρήστη (καταναλωτή).

Ποιος πραγματικά θα βοηθήσει;(5/5)

- Για αυτούς και άλλους λόγους, επιλέγοντας μία κορυφαία εταιρεία ως HPC partner, όπως η Sun Microsystems, μπορεί να αποδειχθεί καθοριστική.
- Εταιρείες με μία ολοκληρωμένη γραμμή λειτουργίας υψηλών επιδόσεων, χρησιμοποιώντας επεξεργαστές τελευταίας τεχνολογίας, έχουν προωθηθεί για τη χρήση των HPC Cluster.
- Η ήδη υπάρχουσα εμπειρία στα HPC και οι συνεισφορές στο λογισμικό, έχουν βοηθήσει την αγορά HPC να προχωρήσει μπροστά.

Αναφορά στη χρήση της Fortran(1/3)

- Όταν οι άνθρωποι ρωτάνε σχετικά με το λογισμικό, με έκπληξη διαπιστώσουν πως πολλά μέρη κώδικα των HPC είναι γραμμένα σε Fortran.
- Ενώ από πολλούς θεωρείται πως η Fortran είναι μία «αρχαία» γλώσσα, στην πραγματικότητα χρησιμοποιείται αρκετά από την κοινότητα του HPC.
- Ο λόγος που γίνεται η χρήση της δεύτερης παλαιότερης γλώσσας είναι αρκετά ιστορικός.

Αναφορά στη χρήση της Fortran(2/3)

- Πολλά από τα HPC προγράμματα γράφτηκαν αρχικά σε Fortran και οι χρήστες είναι απρόθυμοι να την αλλάξουν.
- Στην πραγματικότητα, πολλά HPC προγράμματα έχουν στη σύνθεση τους πάνω από 1 εκατομμύριο γραμμές πηγαίου κώδικα, και πολλές από αυτές τις γραμμές έχουν εκατοντάδες ή και χιλιάδες ώρες δουλείας εξέλιξης.
- Η μετατροπή όλων αυτών των γραμμών σε μία άλλη, πιο καινούρια και μοντέρνα γλώσσα, δεν θα ήταν καλή ιδέα.

Αναφορά στη χρήση της Fortran(3/3)

- Επιπροσθέτως, οικονομικά δεν είναι εφικτό. Εργαλεία λογισμικού όπως οι μεταγλωττιστές, μπορούν να μεταφράσουν αρκετά καλά τη Fortran και είναι πολύ καλοί στην βελτιστοποίηση του πηγαίου κώδικα.

- Οι C και C++ είναι επίσης δημοφιλείς γλώσσες προγραμματισμού για τα HPC.

- Όπως η Fortran, οι προαναφερθείσες γλώσσες προγραμματισμού, είναι σχετικά κοντά με τη γλώσσα μηχανής του επεξεργαστή και μπορούν να προσφέρουν στο χρήση μέγιστη απόδοση στις εφαρμογές HPC.

Προαπαιτούμενες δράσεις(1/4)

- **Α)Δημιουργία ενός έργου πλάνου**, το οποίο θα καλύπτει όλες τις πτυχές για το τι χρειάζεται να αποκτηθεί για τη δημιουργία ενός cluster (τελικοί χρήστες, διαχειριστές, υποδομές, προσωπικό).

- Χρειάζεται να δοθεί έμφαση στα κόστη των υποδομών και να βρεθεί ο κατάλληλος χώρος πριν αγοραστεί. Ο χώρος και η ψύξη ενός HPC Cluster είναι από τα σημαντικότερα προαπαιτούμενα, καθώς καθορίζουν σε μεγάλο βαθμό τη λειτουργία του HPC center.

Προαπαιτούμενες δράσεις(2/4)

- **Β)Αναζήτηση ενός αξιόπιστου HPC συνεταίρου**. Αξιοποιώντας έναν πωλητή cluster ο οποίος έχει εμπειρία στην δημιουργία και υποστήριξη ενός HPC μπορεί να αποβεί πολύ ωφέλιμη.

- Ειδικότερα, εάν δοθεί έμφαση στην σχέση μεταξύ των συνεταίρων, μπορεί να υπάρξει μια πολύ καλή συνεργασία και όχι απλά μία συναλλαγή πωλητή-πελάτη.

- Ο καθένας μπορεί να πουλήσει έναν rack server, αλλά λίγοι θα εγγυηθούν για παράδειγμα ότι μεταφέρουν έναν συγκεκριμένο αριθμό TFLOPS ή ότι βοηθάνε στα θέματα και στις ερωτήσεις που έχουνε οι πελάτες των HPC clusters.

Προαπαιτούμενες δράσεις(3/4)

- Κάτι που δεν μπορεί να παραληφθεί, είναι το πόσο σημαντικό είναι να υπάρχει επιφυλακτικότητα όσον αφορά τις παγίδες που προκύπτουν από του χαμηλού κόστους κατασκευές υλικού.

- Η βιομηχανία είναι κορεσμένη με πολλές καταστροφικές ιστορίες που αφορούν την αγορά φθηνού υλικού με σκοπό την εξοικονόμηση μερικών δολαρίων. Το φθηνό υλικό μπορεί να αποκτηθεί ευκολότερα, αλλά σε πολλές περιπτώσεις υστερεί σε ποιότητα και αποτελεσματικότητα.

- Τέτοιο παράδειγμα αποτελεί η αγορά μνήμης ή κομματιών υλικού διαδικτύου.

- Όποτε υπάρχει αμφιβολία, η ποιότητα ενός καλού πωλητή HPC θα ενισχύσει αρκετά τη γνώμη του πελάτη και θα τον βοηθήσει στο στήσιμο του cluster.

Προαπαιτούμενες δράσεις(4/4)

- **Γ)Να σιγουρευτεί ότι θα δημιουργηθεί σύμφωνα με τις απαιτήσεις του αγοραστή.** Παρόλο που ο πωλητής μπορεί να δημιουργήσει έναν cluster ανάλογα με την εμπειρία και τις γνώσεις του, πρέπει να σιγουρευτεί ότι θα είναι σύμφωνα με τις απαιτήσεις του αγοραστή.

- Είναι σημαντικό να υπάρχει συνεχόμενη εξέταση των κομματιών που συνθέτουν το HPC cluster και ότι είναι πλήρως λειτουργικό από την πρώτη στιγμή.

Τι πρέπει να αποφευχθεί(1/5)

- Α)**Δεν πρέπει να κατασκευαστεί ένας HPC Cluster μόνο από τις πληροφορίες ενός φυλλαδίου τεχνικών χαρακτηριστικών.** Τα φυλλάδια αυτά μπορούν να δώσουν πολλές πληροφορίες, μπορεί να είναι είτε παρωχημένα είτε ανακριβή. Η άμεση επικοινωνία με εξειδικευμένους και καταρτισμένους πωλητές αποτελεί μοναδική λύση

- Οι clusters έχουν αρκετές τεχνικές λεπτομέρειες και απαιτούν αρκετές εργατοώρες προκειμένου να γίνουν κατανοητοί και να μελετηθούν. Σίγουρα, η χρήση επιστημονικών περιοδικών ή φυλλάδων παρέχουν λεπτομέρειες σχετικά με την αιχμή της τεχνολογίας και με τις προτεινόμενες λύσεις, αλλά πρέπει να λαμβάνονται υπόψιν μόνο ως συμπληρωματική ενημέρωση και όχι ως τεχνικοί σύμβουλοι.

Τι πρέπει να αποφευχθεί(2/5)

- **Β) Η χρήση ενός υλικού κομματιού που δεν έχει χρησιμοποιηθεί ακόμα μπορεί να μην είναι και η σωστή τακτική.**

- Μία καινούργια μητρική, ένας νέος επεξεργαστή, ή μία διασύνδεση μπορεί να μοιάζουν αρκετά ελκυστικά και να έχουν και πολύ καλή απόδοση.

- Ωστόσο, γνωρίζοντας το πόσο καλά μπορεί να λειτουργήσουν αρμονικά και σε συνθήκες 24/7, είναι πάντα ένας προβληματισμός. Επομένως συνίσταται να προσεχθεί η λειτουργικότητα των νέων υλικών κατά την περίοδο της δοκιμής τους ή των πρώτων μηνών της κυκλοφορίας τους.

Τι πρέπει να αποφευχθεί(3/5)

- Οι μεγαλύτεροι πωλητές συνεχώς προηγούνται των μικρότερων, όσον αφορά την προμήθεια νέων μερών υλικού, αλλά υπάρχει μία καλή εξήγηση για αυτή τη διαφορά. Οι μεγάλοι πωλητές μπορούν να επιβεβαιώσουν το υλικό και να εγγυηθούν ότι λειτουργεί όπως αναμενόταν.

- Δεν υπάρχει για παράδειγμα κάτι χειρότερο από μερικά ράφια από διακομιστές τα οποία αποτυγχάνουν να λειτουργήσουν και έναν πωλητή ο οποίος αδιαφορεί για τα προβλήματα των πελατών του ή αδυνατεί να επιλύσει τα ανερχόμενα προβλήματα και τις δυσκολίες τους.

Τι πρέπει να αποφευχθεί(4/5)

- **Γ)Να μην χρησιμοποιηθεί ένας και μόνος πυρήνας για δοκιμές επιδόσεων σε πραγματικές συνθήκες.**

- Εφόσον πρόκειται να τρέξουμε σε ένα περιβάλλον πολλών πυρήνων, τότε γιατί να μην υπάρχει ένα δοκιμαστικό περιβάλλον που περιλαμβάνει πολλούς πυρήνες;

- Οι αριθμοί συγκρίσεων με βάσει έναν μοναδικό πυρήνα δεν έχουν ουσία εάν υπάρχει η σκέψη το περιβάλλον να βασιστεί σε πολλούς πυρήνες.

Τι πρέπει να αποφευχθεί(5/5)

- Πολύ συχνά, οι χρήστες είναι απογοητευμένοι με τις επιδόσεις σε πραγματικό περιβάλλον μετά την αγορά του υλικού, καθώς δεν χρησιμοποιήθηκαν σωστές συγκρίσεις για τα χαρακτηριστικά του υλικού που επιθυμούν να προμηθευτούν.

- Αυτή η διαφορά γίνεται για παράδειγμα πιο εμφανής όταν κάποιος χρησιμοποιεί επεξεργαστές αρχιτεκτονικής x86, και αυτό συμβαίνει διότι ένας επεξεργαστής 32-bit μπορεί να υποστηρίξει μνήμη που φτάνει μέχρι τα 4Gb. Στις μέρες μας, μία τέτοια μνήμη δεν είναι ποτέ αρκετή.

- Αρκεί να φανταστεί κανείς πως ένας κοινός ηλεκτρονικός υπολογιστής στις μέρες, χρειάζεται στο ελάχιστο 8Gb μνήμης ram για να μπορέσει να ανταποκριθεί στις απαιτήσεις του λογισμικού. Αυτομάτως, η επιλογής του επεξεργαστή πρέπει να είναι υποχρεωτικά 64-bit.

HPC Community(1/5)

- Το να κατανοήσουν οι οργανισμοί το ελεύθερο και ανοιχτό λογισμικό είναι ένα εμπόδιο. Ο όρος «ελεύθερο» πολλές φορές παραποιείται και το περιεχόμενο της λέξης εννοείται όπως στο «ελεύθερο γεύμα»

- Για παράδειγμα, δεν χρειάζεται να πληρώσουμε κάποιο αντίτιμο για την άδεια του αυτού του λογισμικού.

- Όταν οι περισσότεροι επαγγελματίες ανοιχτού λογισμικού αναφέρονται στον όρο «ελεύθερο», το εννοούν σαν τον όρο «ελεύθερος λόγος». Η διάκριση είναι σημαντική, και στην περίπτωση των HPC, ένα σημαντικό ανταγωνιστικό πλεονέκτημα.

HPC Community(2/5)

- Ελεύθερο λογισμικό σημαίνει ότι ο χρήστης έχει την άδεια να καταλάβει και να αλλάξει το λογισμικό, όπως τον εξυπηρετεί. Στα HPC, αυτό προσφέρει τα παρακάτω πλεονεκτήματα.

1. Αρχικά, το λογισμικό μπορεί να τροποποιηθεί με πολλούς τρόπους, επιτρέποντας υποστήριξη σε πολλά συστήματα αρχείων, διασυνδέσεις, περιφερειακά, και άλλα έργα «μικρής αγοράς».

2. Επιπροσθέτως, προσφέρει ασφάλεια ενάντια σε μη-προγραμματισμένες παρωχημένες ενέργειες.

HPC Community(3/5)

- Στο παρελθόν, ήταν πολύ συνηθισμένο για ένα μεγάλο υπολογιστικό σύστημα να χάσει την υποστήριξη λογισμικού, καθώς πολλοί πωλητές δεν τους υποστήριζαν, τα συστήματα ήταν παρωχημένα ή ο προϋπολογισμός εξαντλειόταν.

- Με το ανοιχτό λογισμικό, οι χρήστες έχουν την επιλογή να συνεχίσουν να χρησιμοποιούν αυτά τα συστήματα, με την υποστήριξη ουσιαστικά να παρέχεται από τους ίδιους.

- Επιπλέον, οι πηγές αρχείων ανοιχτού κώδικα επιτρέπουν υψηλού επιπέδου βελτιστοποίηση.

HPC Community(4/5)

- Οι εγκαταστάσεις λογισμικού μπορούν να ελαχιστοποιηθούν έτσι ώστε μόνο το σημαντικό λογισμικό του HPC να είναι λειτουργικό.

- Για παράδειγμα, δεν είναι απαραίτητοι οι drivers για μία γραφική διεπαφή χρήστη ή για μία κάρτα ήχου στους κόμβους του cluster.

- Οι πελάτες αρέσκονται στο να επιλέγουν λογισμικό υψηλού κόστους ανάπτυξης καθώς θεωρούν ότι υπάρχει αρκετή ασφάλεια στο λογισμικό, μεγάλη υποστήριξη, αρμονική λειτουργία και αξιοποίηση έμπειρων και καλών προγραμματιστών.

HPC Community(5/5)

- Η ειλικρίνεια επίσης ενισχύει την κοινότητα και την ομαδικότητα. Στην περίπτωση των HPC, υπάρχει ένα μεγάλο οικοσύστημα από χρήστες, πωλητές και προγραμματιστές από πολλές εταιρίες.
- Μεταξύ αυτού του οικοσυστήματος, ανταλλάσονται πολλές ιδέες ελεύθερα, όπως επίσης και λογισμικά, χωρίς την ανάγκη για νόμιμες συμφωνίες.
- Αυτή η κοινότητα, προσφέρει μία τεράστια και ανοιχτή βάση ιδεών και γνώσης, η οποία συνεισφέρει με πολλές καλές πρακτικές και λύσεις όπου βοηθάνε όλους όσους εμπλέκονται.
- Στα HPC, η ιδέα του «ανοιχτού» λογισμικού δουλεύει πολύ καλά και ακολουθείται ως φιλοσοφία.

Πως λειτουργεί το ελεύθερο λογισμικό στα HPC(1/4)

- Στην αγορά, υπάρχουν πολλοί τύποι από άδειες ελεύθερου λογισμικού. Για παράδειγμα, η άδεια GNU από τον οργανισμό ελεύθερου λογισμικού (Free Software Foundation), έχει μερικές απαιτήσεις στο πως πρέπει να συμπεριληφθεί ο πηγαίος κώδικας σε κάποιο έργο.

- Άλλοι, μπορεί να απαιτούν μόνο απόδοση και αναγνώριση των πνευματικών δικαιωμάτων όταν οι κώδικες διανέμονται.

- Σε κάθε περίπτωση, ο στόχος του ελεύθερου λογισμικού είναι αρχικά η χρησιμοποίηση του δωρεάν λογισμικού και στη συνέχεια η συνεισφορά στην ανάπτυξη.

Πως λειτουργεί το ελεύθερο λογισμικό στα HPC(2/4)

- Κατά μία έννοια, μερικά έργα, όπως η δημιουργία και η συντήρηση ενός λογισμικού συστήματος, είναι πολύ μεγάλες δουλειές και είναι πολύ λογικό να υπάρχει διαμοιρασμός πληροφοριών και της εξέλιξης ανάμεσα σε εταιρίες.

- Το ίδιο συμβαίνει και στην αγορά HPC!

- Ενώ υπάρχουν πολλά επιχειρήματα υπέρ και κατά των ελεύθερων λογισμικών, η αγορά HPC αναπαριστά ένα χώρο στον οποίο συνυπάρχουν και λειτουργούν ταυτόχρονα τα ελεύθερα ή μη λογισμικά.

Πως λειτουργεί το ελεύθερο λογισμικό στα HPC(3/4)

- Δεν είναι καθόλου ασυνήθιστο για έναν ολόκληρο cluster να αποτελείται από ελεύθερο λογισμικό, με μερικές προσθήκες από εμπορικές εφαρμογές.
- Για χρήστες οι οποίοι μεταγλωττίζουν τους δικούς τους κώδικες, ένας εμπορικός μεταγλωττιστής, ένας αποσφαλματωτής, ένα προφίλ, είναι εργαλεία τα οποία σχεδόν πάντα χρησιμοποιούνται στις διαδικασίες μεταγλωττισμού και αποσφαλμάτωσης.
- Επίσης, πολλές ομάδες κυκλοφορούν προγραμματιστές χρονοδιαγράμματος, οι οποίοι διανέμουν πηγές στον cluster (ή ομάδες από clusters) σε έναν οργανισμό.

Πως λειτουργεί το ελεύθερο λογισμικό στα HPC(4/4)

- Τελικά, μπορεί το λογισμικό να είναι δωρεάν και να διανέμεται ελεύθερα, όμως υπάρχει μία σειρά από μερικούς οργανισμούς οι οποίοι προσφέρουν εμπορική υποστήριξη για το λογισμικό του cluster.

- Εν κατακλείδι, ο συνδυασμός των ελεύθερων ή μη λειτουργιών ενός HPC οικοσυστήματος-λογισμικού, φαίνεται να δουλεύει αρκετά καλά και κυρίως αποτελεσματικά.

6 συμπεράσματα που εξάγουμε(1/7)

- Συνοψίζοντας, η είσοδος στον κόσμο του HPC computing δεν είσαι όσο δύσκολη μπορεί να ήταν πριν από μερικές δεκαετίες ή χρόνια, όμως ακόμα απαιτείται καλή πληροφόρηση, σωστές κινήσεις και συνεργασίας με τα κατάλληλα άτομα ώστε να έρθει η επιτυχία, κάτι που σημαίνει ότι δεν είναι ακατόρθωτη.

- Τα θέματα που πρέπει να αναλυθούν, όπως φάνηκε και από την παρούσα παρουσίαση, είναι αρκετά και σημαντικά. Για αυτό είναι σημαντικό να γίνουν κατανοητά τα παρακάτω 6 συμπεράσματα που προκύπτουν από την ανάλυση των HPC.

6 συμπεράσματα που εξάγουμε(2/7)

1. Τα HPC μπορούν να γίνουν η πηγή ενός καινούργιου, ανταγωνιστικού πλεονεκτήματος:

- Αποτελούν ένα ανταγωνιστικό πλεονέκτημα για πολλές εταιρίες και επιχειρηματικούς κλάδους. Πολλοί οργανισμοί, ακόμα και οι ανταγωνιστές, χρησιμοποιούν HPC υπολογιστές για να μειώσουν το κόστος, να σχεδιάσουν νέα προϊόντα και διαδικασίες, να επιλύσουν ζητήματα και να αυξήσουν το κέρδος. Αυτή η τάση δεν είναι κάτι πρωτόγνωρο.

6 συμπεράσματα που εξάγουμε(3/7)

2. Οι επιλογές που θα γίνουν μπορούν να φέρουν βέλτιστο αποτέλεσμα:

- Εξαιτίας της ανάπτυξης των υλικών, ο τομέας του HPC έχει μετατραπεί από μία ακριβή αγορά σε μία αποτελεσματική και προσβάσιμη τεχνολογία, με συνέπεια να μεταφερθεί η αιχμή της τεχνολογίας στην αγορά. Πολλοί από τους υπερυπολογιστές, οι οποίοι κάποτε κατασκευάζονταν μόνο από κολοσσούς και θεωρούνταν ακατόρθωτοι στη δημιουργία, δημιουργούνται πλέον και από μικρές εταιρίες και οργανισμούς. Έτσι, με την αύξηση πελατών HPC, αυξάνεται η ανατροφοδότηση και πολλά προβλήματα λύνονται ευκολότερα και αποτελεσματικότερα.

6 συμπεράσματα που εξάγουμε(4/7)

3. Είναι αναγκαίο να δοκιμάσουμε πριν αγοράσουμε:

- Τα clustered HPC συστήματα μπορούν να δημιουργηθούν από υλικά διαθέσιμα στο εμπόριο για γενική χρήση. Το κόστος για να δοκιμαστούν οι τροποποιήσεις μπορεί να είναι χαμηλό, αλλά το τίμημα για να δημιουργηθεί ένα σωστό HPC μπορεί να είναι αρκετά υψηλό. Για να γίνει κατανοητό πως μπορούν να επιταχυνθούν οι διαδικασίες κατασκευής ενός HPC, χρειάζεται η εξερεύνηση σε πηγές πληροφοριών, όπως για παράδειγμα σε εκδοτικούς οίκους εξειδικευμένους σε θέματα δημιουργίας HPC ή στην επιστημονική κοινότητα HPC. Αυτές οι πηγές, μπορούν να επιτρέψουν να γίνουν γνωστές τόσο υλοποιήσιμες λύσεις όσο και μοναδικά μοντέλα και μέθοδοι που έχουν αναπτυχθεί. Επιπλέον, αναγκαία και χρήσιμη είναι η βοήθεια των έμπειρων πωλητών HPC συστημάτων.

6 συμπεράσματα που εξάγουμε(5/7)

4. Οι υπερυπολογιστές πάντα βρισκόντουσαν στην κορυφών των τεχνολογικών αιχμών στον τομέα των υπολογιστών:

- Οι υπερυπολογιστές και γενικότερα και η τεχνολογία που προσέφεραν ήταν πάντα ζωτικής σημασίας, καθώς συνέβαλαν πάντα σε σημαντικούς τομείς της καθημερινής ζωής, της πολιτικής και των επιστημών, βοηθώντας να διευθετηθούν σημαντικά κοινωνικά θέματα. Τη σημερινή εποχή, οι υπερυπολογιστές χρησιμοποιούνται για να επιλύσουν ζητήματα όπως η περιβαλλοντική αλλαγή, η τελειοποίηση των μέσων μαζικής μεταφοράς και η ενίσχυσης της στρατιωτικής δυναμικότητας των κρατών. Ο ρόλος των υπερυπολογιστών είναι σημαντικός και για αυτό το λόγο είναι σημαντική η συνέχιση της διάδοσής τους στην παγκόσμια αγορά και η χρησιμοποίησή τους προς όφελος της ανθρωπότητας.

6 συμπεράσματα που εξάγουμε(6/7)

5. Δραστηριοποίηση στις κοινότητες hpc:

- Το ελεύθερο λογισμικό και τα λειτουργικά συστήματα Linux, είναι λογισμικά που χρησιμοποιούνται στους HPC clusters. Υπάρχει μία τρομερή αξία στη χρήση λογισμικού παραδομένου από την κοινότητα και η εξειδίκευση βοηθάει στην εξυπηρέτηση των στόχων, είτε των πελατών, είτε των ανεξάρτητων πωλητών, είτε των μεγάλων εταιρίων. Σημαντικό γεγονός αποτελεί επίσης η ενεργός συμμετοχή πολλών κορυφαίων εταιριών κατασκευής HPC ως μέλη της κοινότητας των Hpc.

6 συμπεράσματα που εξάγουμε(7/7)

6. **Ενασχόληση με προοπτικές:**

- Το HPC είναι ένας πολύ καλός τρόπος να προωθηθούν και να δοκιμαστούν νέες και ενδιαφέρουσες ιδέες. Το κόστος για την εισαγωγή στον κόσμο των HPC είναι προς το παρόν το χαμηλότερο δυνατό μέχρι σήμερα και υπάρχουν πολλές πηγές που μπορούν να βοηθήσουν στο στήσιμο και στην εξέλιξη του εξοπλισμού.

- Ωστόσο, χρειάζεται να λαμβάνουμε σοβαρά υπόψιν όλες τις παραμέτρους που προαναφέρθηκαν, προκειμένου να έχουμε το καλύτερο δυνατό αποτέλεσμα με όσο το δυνατόν χαμηλότερο κόστος. Αυτό δεν σημαίνει ότι πρέπει να καταφεύγουμε στις πιο φθηνές λύσεις, καθώς πολλές φορές η επιλογή ενός μικρού budget, γίνεται εις βάρος της ποιότητας και της αποδοτικότητας.

Βιβλιογραφία (1/3)

[1] https://www.zougla.gr/technology/article/dell-simantikes-kenotomies-ke-ekseliksis-sta-sistimata-high-performance-computing

[2] http://hpc.fs.uni-lj.si/sites/default/files/HPC_for_dummies.pdf

[3] http://www.admin-magazine.com/HPC/Articles/Building-an-HPC-Cluster

[4] https://insidehpc.com/hpc101/intro-to-hpc-whats-a-cluster/

[5] https://learn.scientificprogramming.io/introduction-to-high-performance-computing-hpc-clusters-9189e9daba5a

[6] http://web.eecs.umich.edu/~mosharaf/Readings/DC-Computer.pdf

[7] http://www.prace-ri.eu/

[8] http://www.prace-ri.eu/IMG/pdf/wp79.pdf

[9] https://www.studytonight.com/computer-networks/complete-osi-model

Βιβλιογραφία (2/3)

[10] https://whatis.techtarget.com/definition/rack-unit

[11] https://www.systems.ethz.ch/sites/default/files/file/Spring2013_Courses/AdvCompNetw_Spring2013/13-hpc.pdf

[12] https://www.aspsys.com/solutions/infrastructure/power/

[13] https://www.indeed.com/jobs?q=High+Performance+Computing+Engineer

[14] https://www.indeed.com/jobs?q=High%20Performance%20Computing%20Engineer&start=30&vjk=0a4f32564df143d6

[15] https://arch.icte.uowm.gr/docs/High_Performance_Computing_Energy_Issues_GreenICT2014.pdf

[16] https://www.switch.com/switch-ranked-1-top-10-cloud-campus-ecosystems/

[17] https://icl.utk.edu/ctwatch/quarterly/print.php%3Fp=13.html

[18] https://www.vertatique.com/average-power-use-server

[19] https://www.ny-engineers.com/blog/do-you-need-a-ups-or-an-inverter

[20] http://www.sphomerun.com/data-center-sales-and-marketing-blog/what-does-it-cost-to-build-a-data-center

Βιβλιογραφία (3/3)

- [21] https://edgeoptic.com/storage-protocols-comparison-fibre-channel-fcoe-infiniband-iscsi/
- [22] https://public.confluence.arizona.edu/display/UAHPC/Transferring+Files
- [23] https://www.computerweekly.com/tip/Introduction-to-Border-Gateway-Protocol-BGP
- [24] https://hpc.grnet.gr/supercomputer/
- [25] https://www.nap.edu/read/11148/chapter/12