ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΛΥΣΗ ΤΟΥ 2018 CLUSTER TRACE ΤΗΣ ALIBABA

Διπλωματική Εργασία

ΜΑΡΙΟΣ-ΕΡΡΙΚΟΣ ΚΥΡΙΚΛΙΔΗΣ

Επιβλέπων: ΧΡΗΣΤΟΣ Δ. ΑΝΤΩΝΟΠΟΥΛΟΣ

Βόλος 2019

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΛΥΣΗ ΤΟΥ 2018 CLUSTER TRACE ΤΗΣ ALIBABA

Διπλωματική Εργασία

ΜΑΡΙΟΣ-ΕΡΡΙΚΟΣ ΚΥΡΙΚΛΙΔΗΣ

Επιβλέπων: ΧΡΗΣΤΟΣ Δ. ΑΝΤΩΝΟΠΟΥΛΟΣ

Βόλος 2019

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

ANALYSIS OF THE 2018 ALIBABA'S CLUSTER TRACE

Diploma Thesis

MARIOS-ERRIKOS KIRIKLIDIS

Supervisor: CHRISTOS D. ANTONOPOULOS

Volos 2019

ΕΥΧΑΡΙΣΤΙΕΣ

Με την παρούσα διπλωματική εργασία ολοκληρώνεται ο κύκλος των σπουδών μου στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών. Αρχικά, θα ήθελα να ευχαριστήσω τους καθηγητές του τμήματός για την μεταλαμπάδευση των γνώσεων τους καθώς και για την αφοσίωση τους στο να διατηρείται το τμήμα σε υψηλό επίπεδο παρά τις όποιες δυσκολίες. Ειδικότερα, θα ήθελα να ευχαριστήσω τον κ. Αντωνόπουλο Χρήστο για την βοήθεια και την καθοδήγηση που μου προσέφερε, αλλά κυρίως για το χρόνο του και την υπομονή του κατά τη συγγραφή της παρούσας διπλωματικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για τη στήριξη που μου παρέχει αδιαλείπτως σε όλα τα επίπεδα και ήταν δίπλα μου όλα αυτά τα χρόνια.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

«Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής».

Ο Δηλών

(Υπογραφή)

ΚΥΡΙΚΛΙΔΗΣ ΜΑΡΙΟΣ-ΕΡΡΙΚΟΣ

08/07/2019

ΠΕΡΙΛΗΨΗ

Στα σύγχρονα κέντρα δεδομένων παροχής υπηρεσιών cloud, η τεχνική της "συν-τοποθέτησης" (co-locating) των online εργασιών και των εργασιών συστήματος, ανάγεται σε μία αποτελεσματική τεχνική αξιοποίησης πόρων. Η Alibaba, η οποία και αποτελεί την μεγαλύτερη Κινεζική εταιρεία παροχής υπηρεσιών cloud και μία από τις μεγαλύτερες παγκοσμίως, για να διευκολύνει την κατανόηση των αλληλεπιδράσεων των εργασιών αυτών και την ανάλυση των πραγματικών τους απαιτήσεων σε πόρους, δημοσίευσε το 2ο "Σύνολο Φόρτου Εργασίας Συστήματος" (*cluster-trace),* διάρκειας 8 ημερών, για το έτος 2018. Στην διπλωματική αυτή εργασία, αναλύεται ο βαθμός αξιοποίησης των πόρων CPU και MEMORY του cluster, καθώς επίσης και επισημαίνονται χωρικές και χρονικές ανισορροπίες στον χρονοπρογραμματισμό των εργασιών στους πόρους. Επιπλέον εντοπίζεται ότι το workload είναι memory bound καθώς λόγω εσφαλμένου over-estimation των εργασιών οι πόροι μνήμης σχεδόν εξαντλούνται. Τέλος, παρατηρείται μία πολιτική προτεραιότητας ως προς τις online εργασίες κατά τη διάρκεια της ημέρας, με πρόφαση την διαφύλαξη των SLA συμφωνιών, η οποία επίσης οδηγεί σε μη αποδοτική κατανομή και αξιοποίηση πόρων.

ABSTRACT

Co-locating online services and offline batch jobs, in modern cloud data center, is an efficient approach to improve datacenter utilization. Alibaba is the largest cloud provider companies in China, and one of the largest cloud providers worldwide. The company has recently released their 2nd, 8-day-long cluster usage and co-located workload dataset (*cluster-trace-v2018)*, in order to facilitate the understanding of interaction among the co-located workloads and their real-world operational demands. In this diploma thesis, the levels of CPU and MEMORY utilization across the cluster are analyzed, and so, spatial and temporal imbalances of resource utilization are spotted. Moreover, workload is identified as memory bounded, because of the false over-estimation of jobs, in need of memory resources. Finally, in an attempt to satisfy SLA agreements, a priority policy in favor of the online service jobs is spotted, which leads to further non-efficient resource utilization.

Περιεχόμενα

Περιεχόμενα Εικόνων

1) ΕΙΣΑΓΩΓΗ

Η τεχνολογία του συστήματος Cloud (Νεφελοειδές Υπολογιστικό Σύστημα) εφαρμόζεται με πολύ μεγάλη επιτυχία στις μέρες μας σε επιχειρήσεις και επιστημονικούς τομείς, όπως το Ηλεκτρονικό Εμπόριο (e-commerce), την Ηλεκτρονική Διακυβέρνηση (E-government), τις Υπηρεσίες Μηχανικού Σχεδιασμού (Engineering Design) και Ανάλυσης, την Οικονομία (Finance), την Υγειονομική Περίθαλψη (Healthcare), τη Φιλοξενία Ιστοσελίδων (Web Hosting), την online Κοινωνική Δικτύωση (Social Networks) κτλ. Πιο συγκεκριμένα, η τεχνολογία αυτή κατάφερε να προσφέρει οικονομικά αποδοτικές και κλιμακούμενες λύσεις, χάρη στην ελαστικότητα της παροχής πόρων και της χρήσης προηγμένων μηχανισμών Εικονικοποίησης (Virtualization) και μηχανισμών Χρονο-προγραμματισμού (Scheduling Mechanisms) [1] [2].

Ο Φόρτος εργασίας (Workload) του cloud αποτελείται από μια συλλογή από πολλές διαφορετικές εφαρμογές και υπηρεσίες που χαρακτηρίζονται από τις δικές τους επιδόσεις, από αιτήσεις πόρων, και από τους περιορισμούς που καθορίζονται στη μορφή μιας Συμφωνίας Επιπέδου Υπηρεσιών SLA (Service Level Agreements). Παράλληλα, ένας μεγάλος αριθμός παραγόντων επηρεάζουν την απόδοση των συστημάτων cloud, μεταξύ των οποίων, 1) η μεταβλητότητα στην διαθεσιμότητα των πόρων, 2) οι συνθήκες δικτύου και 3) η εξαιρετικά δυναμική φύση του φόρτου εργασίας, η ένταση της οποίας μπορεί αυξάνεται ή συρρικνώνεται ως συνέπεια των αλληλεπιδράσεων του χρήστη. Ειδικότερα, η χρήση εικονικών, διαμοιραζόμενων χρονικά πόρων (virtualized time-shared resources) θα μπορούσε να οδηγήσει σε υποβιβασμό της απόδοσης (performance degradation). Ο υποβιβασμός αυτός οφείλεται κυρίως στις παρεμβάσεις και στις συγκρούσεις για χρήση πόρων (interference and resource contention) που απορρέουν από τη συνύπαρξη ετερογενών workload σε μία κοινή φυσική υποδομή και από το κόστος overhead που εισάγεται από τις πολιτικές διαχείρισης πόρων που έχουν υιοθετηθεί [3]. Αυτά τα θέματα επιδόσεων θα μπορούσαν να γίνουν ακόμη πιο κρίσιμα σε περιβάλλοντα πολλαπλών cluster, όπου ο φόρτος εργασίας μπορεί του cloud μπορεί να είναι κατανεμημένος σε διαφορετικές υποδομές.

Παρά τη σπουδαιότητά του, ο χαρακτηρισμός και η πρόβλεψη (characterization and forecasting) του φόρτου εργασίας σε ένα σύστημα cloud, αντιμετωπίστηκε στη βιβλιογραφία, σε μάλλον περιορισμένο βαθμό και κυρίως στο επίπεδο των VMs, χωρίς να λαμβάνονται υπόψη τα χαρακτηριστικά των επιμέρους στοιχείων φόρτου εργασίας που εκτελούνται από τα ίδια τα VMs. Από τις δημοσιεύσεις αυτές [4] [5] [6] προκύπτει το γεγονός ότι οι εγκαταστάσεις των Cloud συστημάτων λειτουργούν σε χαμηλά επίπεδα αξιοποίησης. Πιο συγκεκριμένα, τα παγκόσμια επίπεδα απόδοσης των παρόχων-διακομιστών (servers) πριν από μερικά χρόνια άγγιζαν μόλις το 12%, ενώ ακόμα και με την χρήση του Virtualization το ποσοστό αυτό δεν ξεπερνάει το 17%. Η τοποθέτηση (co-locating) αμφότερων των Εργασιών Online Υπηρεσιών (Online Service Jobs) και των Offline Εργασιών Συστήματος (Offline Batch Jobs) σε κοινά συμπλέγματα, αποδεικνύεται ότι είναι μια αποτελεσματική προσέγγιση για τη βελτίωση των επιπέδων απόδοσης των παρόχων, σύγχρονα κέντρα δεδομένων (datacenters) [7] [8].

Συνοψίζοντας λοιπόν, η πλατφόρμα cloud μπορεί να παρέχει μεγάλη ευελιξία και οικονομία απόδοσης, τόσο στους πελάτες-χρήστες όσο και στους παρόχους των υπηρεσιών cloud. Ωστόσο η χαμηλή αξιοποίηση των πόρων στα σύγχρονα datacenters, προκαλεί μεγάλες σπατάλες πόρων και εγείρει την ανάγκη να μελετηθούν και να προσδιοριστούν οι παράγοντες που επηρεάζουν αρνητικά την εύρυθμη λειτουργία και την απόδοση τους. Στόχος λοιπόν αυτής της διπλωματικής εργασίας είναι να παρέχει μια ανάλυση των κύριων ζητημάτων που σχετίζονται με την διαχείριση του φόρτου εργασίας σε έναν ολόκληρο κύκλο ζωής, μέσα σε περιβάλλοντα cloud και πιο συγκεκριμένα την ανάλυση του workload της εταιρείας παροχής cloud υπηρεσιών Alibaba, η οποία και χρησιμοποιεί την τεχνική co-locating. Τα δεδομένα για την παρούσα ανάλυση, τα δημοσίευσε η εταιρεία με το 2ο της "Σύνολο Φόρτου Εργασίας του Συστήματος" (*cluster-trace)* το έτος 2018, όπου και εμπεριέχονται πληροφορίες από 4000 μηχανήματα,online υπηρεσίες,εργασίες συστήματος, συγκεντρωμένα όλα σε διάρκεια 8 ημερών. Η διαφορά σε σχέση με το προηγούμενο Alibaba *cluster-trace* του 2017, από την ανάλυση του οποίου εμπνευστήκαμε την εργασία αυτή [9] (12 ώρες διάρκεια /1300 μηχανήματα), είναι ότι καινούριο trace πέρα του όγκου των δεδομένων παρουσιάζει νέα χαρακτηριστικά και τάσεις από τα οποία μπορούν να εξαχθούν συμπεριφορές και τρόποι αντιμετώπισης διαφόρων καταστάσεων του workload. Έτσι, από την μελέτη αυτή προέκυψαν α) φαινόμενα χρονικής και χωρικής ανισορροπίας

(temporal & spatial imbalance) της διαχείρισης του φόρτου εργασίας, β) ανισορροπία στην ζήτηση και στην αξιοποίηση των πόρων (resource demands & utilization imbalance), γ) η αξιοποίηση της μνήμης ανάγεται ως σημείο συμφόρησης (bottleneck) και περιορίζει την αποδοτικότητα των πόρων σε ένα datacenter της Alibaba και δ) οι εργασίες συστήματος batch, αντιμετωπίζονται με διαφορετικά κριτήρια από τις online εργασίες, ώστε να εξασφαλιστούν οι SLA συμφωνίες.

Το υπόλοιπο της εργασίας είναι οργανωμένο ως εξής: Στο κεφάλαιο 2 γίνεται περιγραφή του θεωρητικού υπόβαθρου του υπολογιστικού συστήματος νέφους. Στο κεφάλαιο 3 γίνονται αναφορές στην υπάρχουσα βιβλιογραφία, ενώ στο κεφάλαιο 4 περιγραφή των χαρακτηριστικών του dataset. Τέλος στο κεφάλαιο 5 παρουσιάζονται τα διαγράμματα από την επεξεργασία των δεδομένων και στο κεφάλαιο 6 εξάγονται τα συμπεράσματα της εργασίας αυτής.

2) ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

2.1 Θεωρητικό Υπόβαθρο

Την τελευταία δεκαετία, οι τεχνολογικές εξελίξεις στον τομέα των virtualization τεχνολογιών και γενικότερα στον τομέα των υπολογιστών, έχουν καταστήσει εφικτή την δημιουργία μεγάλων, αποδοτικών Data Centers ευρείας κλίμακας, τα οποία και επεξεργάζονται σημαντικό τμήμα των σημερινών εφαρμογών του διαδικτύου.
Επιπλέον, η καθιέρωση των οικονομιών κλίμακας, δηλαδή της αναλογικής εξοικονόμησης του κόστους, η οποία επιτυγχάνεται μέσω του αυξημένου επίπεδου παραγωγής, επέτρεψε τη μίσθωση της υποδομής των κέντρων δεδομένων (data centers) σε τρίτους.

Εξελίχθηκε λοιπόν η τεχνολογία του cloud computing (υπολογιστικά συστήματα νέφους), όπου γίνεται κοινή χρήση ενός συνόλου υπολογιστικών πόρων, μεταξύ εφαρμογών που είναι διαδικτυακά προσβάσιμες. Κατά συνέπεια, εξελίχθηκε σε έναν δημοφιλή όρο-εργαλείο όχι μόνο για την τεχνολογική κοινότητα αλλά και για το ευρύ κοινό, αφού χρησιμοποιείται ταυτόχρονα και από εφαρμογές που παρέχονται μέσω του

διαδικτύου, αλλά και από λογισμικά υλικού και συστήματος που χρησιμοποιούνται εντός των κέντρων δεδομένων που φιλοξενούν αυτές τις εφαρμογές.

Με μία πιο γενική περιγραφή θα ειπωθεί ότι το cloud computing είναι μία κατά παραγγελία παροχή υπηρεσιών πληροφορικής - από εφαρμογές μέχρι αποθήκευση και επεξεργαστική ισχύ -συνήθως μέσω του διαδικτύου και με βάση την πολιτική pay-as-you-go.

2.2 Λειτουργία της Τεχνολογίας σε Γενικές Γραμμές

Αρχικά υλοποιείται μία συμφωνία επιπέδου υπηρεσιών SLA (service-level agreement), αποτελώντας την δέσμευση μεταξύ ενός παρόχου υπηρεσιών και ενός πελάτη. Στην συμφωνία συμπεριλαμβάνονται τα ιδιαίτερα χαρακτηριστικά της υπηρεσίας: α) η απαιτούμενη ποιότητα υπηρεσιών, β) η απαιτούμενη διαθεσιμότητα και γ) οι ευθύνες-απαιτήσεις του 1ου και του 2ου μέλους αντίστοιχα.
Η υπηρεσία cloud περιλαμβάνει: τόσο α) την παροχή πόρων σε τρίτους, σε μία ενοικιαζόμενη-ανάλογη με την χρήση βάση, και β) σε ιδιωτικές υποδομές που διατηρούνται από μεμονωμένες εταιρείες. Η πρώτη περίπτωση είναι γνωστή ως δημόσιο (public) cloud και η δεύτερη ως ιδιωτικό (private) cloud.
Εναλλακτικά μπορεί μια εταιρεία-πελάτης να επεκτείνει το Ιδιωτικό της cloud με επιπλέον ενοικίαση πόρων από το δημόσιο cloud (cloud bursting) και να προκύψει έτσι μια τρίτη κατηγορία παροχής υπηρεσιών, το υβριδικό (hybrid) cloud. Τέλος υπάρχει και μια 4η κατηγορία παροχής υπηρεσιών, το κοινωτικό (community) cloud, όπου οι πόροι παρέχονται από πολλές εταιρείες / οργανώσεις και η διαχείριση είναι αποκεντρωμένη.

Το δημόσιο (public) cloud, που έχει αναφερθεί και παραπάνω, είναι το πιο κλασικό μοντέλο cloud computing, όπου οι χρήστες μπορούν να έχουν πρόσβαση σε ένα μεγάλο δίκτυο υπολογιστικής ισχύος μέσω του διαδικτύου. Η τεράστια αυτή κλίμακα δίνει την δυνατότητα στους παρόχους να έχουν πλεονάζοντες πόρους και να μπορούν εύκολα να αντεπεξέλθουν, εάν κάποιος συγκεκριμένος πελάτης χρειάζεται περισσότερους πόρους. Για αυτό και χρησιμοποιείται συχνά για λιγότερο ευαίσθητες εφαρμογές που

απαιτούν κλιμακούμενους πόρους. Η ταξινόμηση που συναντάται σε περιβάλλοντα δημοσίου cloud είναι η εξής:

1. Η υποδομή ως υπηρεσία IaaS (Infrastructure as a service) αναφέρεται στα θεμελιώδη δομικά στοιχεία των υπολογιστών που μπορούν να ενοικιαστούν: φυσικοί ή εικονικοί servers, αποθήκευση (storage) και δικτύωση (networking). Το γεγονός αυτό κάνει το IaaS ελκυστικό για τους πελάτες που θέλουν να χτίσουν οι ίδιοι τις εφαρμογές τους και θέλουν επίσης να ελέγχουν όλα σχεδόν τα στοιχεία γύρω από αυτές.

2. Η πλατφόρμα ως υπηρεσία (PaaS) ή η πλατφόρμα εφαρμογών ως υπηρεσία (aPaaS) είναι μια κατηγορία υπηρεσιών cloud computing που παρέχει στους πελάτες μια πλατφόρμα, επιτρέποντας τους να αναπτύσσουν, να τρέχουν και να διαχειρίζονται τις εφαρμογές τους, χωρίς την πολυπλοκότητα της οικοδόμησης και συντήρησης κάποιας φυσικής υποδομής.

3. Το λογισμικό ως υπηρεσία (SaaS) είναι ένα μοντέλο διανομής λογισμικού στο οποίο ένας πάροχος φιλοξενεί εφαρμογές και τις καθιστά διαθέσιμες στους πελάτες μέσω του διαδικτύου. Το θεμελιώδης υλικό και το λειτουργικό σύστημα δεν έχουν σημασία για τον τελικό χρήστη, ο οποίος θα έχει πρόσβαση στην υπηρεσία μέσω ενός προγράμματος περιήγησης ή μιας εφαρμογής στο διαδίκτυο. Συχνά αγοράζεται ανά χρήση ή ανά χρήστη.

2.3 Το Χρονολόγιο του Cloud

Η ιστορία πίσω από τον όρο cloud computing ανάγεται στις γύρω στις αρχές της δεκαετίας του 2000, αλλά η έννοια του computing ως υπηρεσία, έχει ξεκινήσει από τη δεκαετία του 1960, όταν γραφεία ηλεκτρονικών υπολογιστών επέτρεψαν σε εταιρείες να νοικιάσουν χρόνο σε κάποιο μηχάνημα, αντί να αγοράζουν ένα οι ίδιοι. Αυτές οι υπηρεσίες "χρονομεριστικής μίσθωσης" ξεπεράστηκαν σε μεγάλο βαθμό από την άνοδο του προσωπικού υπολογιστή (P.C.) που κατέστησε πολύ πιο οικονομικό το να

είναι κάποιος ιδιοκτήτης ενός υπολογιστή, και στη συνέχεια από την άνοδο των εταιρικών κέντρων δεδομένων όπου οι επιχειρήσεις θα μπορούν πλέον να αποθηκεύουν τεράστιες ποσότητες δεδομένων.

Ωστόσο, η έννοια της μίσθωσης της πρόσβασης σε υπολογιστική ισχύ επανεμφανίστηκε στα τέλη της δεκαετίας του 1990 και στις αρχές της δεκαετίας του 2000 όταν αυξήθηκε μαζικά η ανάγκη-ζήτηση για λογισμικά ως υπηρεσία (SaaS) και γενικά για περισσότερες εταιρείες που παρέχουν υπολογιστικούς πόρους σε επιχειρήσεις ή άτομα.

2.4 Παράμετροι που Επηρεάζουν την Λειτουργία ενός Συστήματος Cloud

Ι) ΚΑΤΑΝΟΜΗ ΠΟΡΩΝ (RESOURCE ALLOCATION)

Στα συστήματα cloud, η κατανομή πόρων (resource allocation) είναι η διαδικασία εκχώρησης διαθέσιμων πόρων στις αιτούμενες εφαρμογές του cloud μέσω του διαδικτύου. Η κακή-αλόγιστη κατανομή πόρων, όταν δεν γίνεται με ακρίβεια μπορεί να οδηγήσει σε εξάντληση των πόρων. Η προοικονομία πόρων (resource provisioning) λύνει αυτό το πρόβλημα επιτρέποντας στους παρόχους υπηρεσιών να διαχειρίζονται τους πόρους για κάθε μεμονωμένο κόμβο.

Η στρατηγική κατανομής πόρων RAS (resource allocation strategy) αφορά τις δραστηριότητες του παρόχου, για τη χρήση και την κατανομή χρήσιμων πόρων μέσα στα όρια του συστήματος, έτσι ώστε να ικανοποιούνται οι ανάγκες όλων των αιτήσεων του cloud. Για την βέλτιστη στρατηγική RAS απαιτούνται διάφορες παράμετροι εισόδου όπως ο τύπος, η ποσότητα των πόρων που απαιτούνται, ο χρόνος και η σειρά με την όποια πρέπει να ολοκληρωθεί μία οποιαδήποτε εργασία.
Παράλληλα τα κριτήρια για το τι πρέπει να αποφεύγει μια τέτοια πολιτική είναι:

1. Η σύγκρουση αιτούμενων πόρων (resource contention) δημιουργείται όταν δύο εφαρμογές προσπαθούν ταυτόχρονα να έχουν πρόσβαση στον ίδιο πόρο.

2. Η έλλειψη πόρων (scarcity of resources) δημιουργείται όταν υπάρχουν περιορισμένοι πόροι.
3. Ο κατακερματισμός των πόρων (resource fragmentation) ανακύπτει όταν υπάρχουν απομονωμένοι πόροι, που δεν είναι όμως ικανοί να καλύψουν την αιτούμενη ζήτηση των εφαρμογών.
4. Η υπερεκτίμηση των αναγκών για πόρους (over-provisioning of resources), που προκύπτει, όταν οι αιτήσεις λαμβάνουν πλεόνασμα πόρων από τον πραγματικά απαιτούμενο.
5. Η υποεκτίμηση των αναγκών για πόρους (under-provisioning of resources) που προκύπτει, όταν οι αιτήσεις λαμβάνουν λιγότερους πόρους από την πραγματική ζήτηση.

ΙΙ) ΤΕΧΝΟΛΟΓΙΑ VIRTUALIZATION

Σε ένα σύμπλεγμα cloud computing, οι πόροι του συστήματος μπορούν να αξιοποιηθούν πιο αποτελεσματικά με την χρήση της εικονικοποίησης. Στους διάφορους πελάτες, μπορεί να εγγυηθούν απαιτητικές SLA συμφωνίες με την αντιστοίχιση όλων των απαιτούμενων υπηρεσιών σε μια εικονική μηχανή και στη συνέχεια με την χαρτογράφηση (mapping) σε ένα φυσικό διακομιστή (server).

Η τεχνολογία εικονικοποίησης, είναι το βασικό χαρακτηριστικό του cloud computing που επιτρέπει την συνύπαρξη πολλαπλών εικονικών μηχανών μέσα σε ένα φυσικό μηχάνημα καθώς και τη μετανάστευση (migration) των VMs. Η εικονικοποιημένη υποδομή (virtualized vnfrastructure) δημιουργεί ένα επίσης εικονικό σύνολο πόρων (virtualized pool of resources), ενοποιεί τους διακομιστές διαχείρισης, αποθήκευσης και δικτύων. Έτσι οι απαιτούμενοι πόροι από το σύνολο πόρων, μπορούν να δεσμεύονται ανά πάσα στιγμή σύμφωνα με την ζήτηση.

Διαφορετικές εφαρμογές, με διαφορετικές απαιτήσεις πόρων μπορούν να εκτελούνται ταυτόχρονα στο ίδιο φυσικό μηχάνημα πράγμα που ίσως οδηγήσει σε μη ισορροπημένο φόρτο εργασίας. Με τη χρήση της τεχνολογίας του virtualization, η ανακατανομή των πόρων του συστήματος μπορεί να πραγματοποιηθεί δυναμικά, με την χαρτογράφηση (re-mapping) των εικονικών και φυσικών μηχανών, σύμφωνα με τη μεταβολή του φορτίου [10]. Σε συνδυασμό, με την ενοποίηση (consolidation) των

VMs στον ελάχιστο αριθμό ενεργών διακομιστών και την απενεργοποίηση (ή την λειτουργία σε sleep mode) των κατάλληλων κόμβων επιτυγχάνεται βελτιστοποίηση χρήσης των πόρων και μείωσης της απαιτούμενης ενέργειας.

Ωστόσο, προκειμένου να επιτευχθεί η καλύτερη απόδοση, οι εικονικές μηχανές πρέπει να αξιοποιήσουν πλήρως τις υπηρεσίες και τους πόρους της τεχνολογίας αυτής, προσαρμοζόμενες στο περιβάλλον του cloud. Για να αξιοποιηθεί πλήρως λοιπόν η αξιοποίηση των πόρων [11], πρέπει να επιτυγχάνεται παράλληλα και ο σωστός χρόνο-προγραμματισμός διεργασιών (task scheduling).

ΙΙΙ) ΧΡΟΝΟ-ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΔΙΕΡΓΑΣΙΩΝ (TASK SCHEDULING)

Ο χρόνο-προγραμματισμός διεργασιών είναι η διαδικασία με την οποία γίνεται προσπάθεια να αναθέτονται "ιδανικά" οι διεργασίες, των πελατών ή και του ίδιου του συστήματος, στους διαθέσιμους πόρους, έτσι ώστε να βελτιωθεί συνολικά η διαχείριση-αύξηση της χρήσης των πόρων [12].

Δεδομένου ότι η κατανομή των πόρων του cloud βασίζεται σε κάποια SLA συμφωνία, το κόστος εκτέλεσης εργασιών θεωρείται μία από τις κύριες παραμέτρους επιδόσεων του αλγόριθμου προγραμματισμού εργασιών [13].

Υπάρχουν πολλές παράμετροι που πρέπει να ληφθούν υπόψη για την ανάπτυξη ενός task scheduling αλγόριθμου. Ορισμένες από αυτές τις παραμέτρους είναι σημαντικές για την προοπτική του χρήστη του cloud (π.χ. ο χρόνος επεξεργασίας των εργασιών, το κόστος και ο χρόνος απόκρισης). Άλλες παράμετροι είναι σημαντικές για την προοπτική του παρόχου cloud (π.χ. η χρήση πόρων, η ανεκτικότητα σφαλμάτων και η κατανάλωση ενέργειας) [14].

Παραδείγματα τέτοιων scheduling αλγορίθμων που έχουν διερευνηθεί σε διάφορες δημοσιεύσεις είναι οι Min-Min, Max-Min, X-Sufferage, Γενετικοί Αλγόριθμοι, Particle Swarm Optimization (βελτιστοποίηση σμήνους σωματιδίων) κ.λ.π..

IIV) ΑΣΦΑΛΕΙΑ

Η ασφάλεια στα πλαίσια του cloud computing είναι ένα ιδιαίτερα ανησυχητικό ζήτημα εξαιτίας του γεγονότος ότι οι συσκευές που χρησιμοποιούνται για την παροχή υπηρεσιών δεν ανήκουν στους ίδιους τους χρήστες,. Οι χρήστες δεν έχουν συνεπώς έλεγχο, ούτε γνώση του τι θα μπορούσε να συμβεί στα δεδομένα τους. Το γεγονός αυτό προκαλεί μεγάλη ανησυχία σε περιπτώσεις όπου υπάρχουν αποθηκευμένες πολύτιμες, απόρρητες και προσωπικές πληροφορίες.

Οι χρήστες δεν είναι διατεθειμένοι να θέσουν σε κίνδυνο την ιδιωτικότητα τους, και συνεπώς οι πάροχοι υπηρεσιών πρέπει να διασφαλίζουν ότι οι πληροφορίες των πελατών τους είναι ασφαλείς. Αυτό, ωστόσο, γίνεται ολοένα και πιο δύσκολο, καθώς όσο και να εξελίσσονται οι μηχανισμοί και οι πολιτικές ασφάλειας, πάντα φαίνεται να υπάρχει κάποιος που θα προσπαθήσει να τις παρακάμψει και να επωφεληθεί από τις πληροφορίες των χρηστών. Θέματα ασφαλείας και πιθανοί κίνδυνοι [15] που απασχολούν αμφότερους πελάτες και παρόχους ενός συστήματος Cloud είναι: η ταυτότητα (identity), η υποδομή (infrastructure), η ιδιωτικότητα (privacy), η μετάδοση δεδομένων (data transmission) κ.α..

V) ΦΟΡΤΟΣ ΕΡΓΑΣΙΑΣ (WORKLOAD)

Ως φόρτος εργασίας μπορεί να οριστεί ως η ποσότητα εργασίας που συσχετίζεται με έναν πελάτη, ή έναν διακομιστή ή γενικά με ένα σύστημα για κάποια δεδομένη χρονική περίοδο. Τα είδη της εργασίας αυτής μπορούν να προέρχονται από ιστότοπους (websites), ηλεκτρονικές επεξεργασίες συναλλαγών (online transaction processing), ηλεκτρονικό εμπόριο (e-commerce), υπολογιστικές διεργασίες, διεργασίες ανάλυσης δεδομένων και ούτω καθεξής.

Η διαχείριση του φόρτου εργασίας αποτελεί έναν από τους σημαντικότερους παράγοντες για την επίτευξη υψηλών επιδόσεων στα συστήματα cloud. Συνεπώς, η ανάλυση και ο χαρακτηρισμός του είναι ιδιαίτερα απαιτητικά, λόγω της ετερογένειας του υπάρχοντος υλικού σε ένα ή περισσότερα data centers, του VM consolidation και του migration που μπορεί να συμβαίνουν ανά τακτά χρονικά διαστήματα. Τα παραπάνω

σε συνδυασμό με τα θέματα εμπιστευτικότητας, τις απαιτήσεις ανοχής σε σφάλματα, το cloud uptime, την ανεπάρκεια πόρων κτλ. παράγουν πιο πολύπλοκα και σύνθετα workloads.

Η κατανόηση των διαφορετικών χαρακτηριστικών και περιορισμών του φόρτου εργασίας στο cloud είναι ευεργετική τόσο για τους προμηθευτές υπηρεσιών cloud όσο και για τους ερευνητές. Για τους προμηθευτές, θα συμβάλει στην αναβάθμιση των μηχανισμών διαχείρισης πόρων έτσι ώστε να αυξηθεί η παραγωγικότητα και η ποιότητα της υπηρεσίας του συστήματος τους (quality of service). Για τους ερευνητές, θα επιτρέψει την αξιολόγηση των θεωρητικών μηχανισμών που υποστηρίζονται από τα χαρακτηριστικά των κέντρων δεδομένων cloud μέσω προσομοιωτών όπως cloudsim, cloudanalyst, greencloud κ.λπ., τα οποία αναπαράγουν τη συμπεριφορά των υποδομών και την κατανάλωση ενέργειας ενός συστήματος.

Ένα workload περιλαμβάνει ένα σύνολο από διεργασίες (task), όπου κάθε διεργασία ανήκει σε μια εργασία (job). Μία εργασία μπορεί να αποτελείται από πολλές διεργασίες, και αντίστοιχα μία διεργασία να αποτελείται από πολλά στιγμιότυπα (instances). Προκειμένου να κατανοήσουμε καλύτερα, να περιγράψουμε και να βελτιώσουμε την λειτουργία του cloud, η ανάλυση των tasks είναι στοιχείο μείζονος σημασίας. Άρα και στόχος της ανάλυσης ενός workload, είναι να εξεταστούν οι διάφορες πτυχές και τα χαρακτηριστικά ενός συνόλου δεδομένων, ακολουθώντας ορισμένους περιορισμούς όσον αφορά τον φόρτο εργασίας [16]:

- Ένα workload μπορεί να έχει ένα συγκεκριμένο χρόνο έναρξης (begin time) ή ένα ελαστικό χρόνο έναρξης (elastic begin time).

- Ένα workload μπορεί να χρειαστεί να τελειώσει μέσα ένα συγκεκριμένο χρονικό διάστημα.

- Ένα workload μπορεί να είναι χρονικά περιορισμένο (time bound) ή χρονικά απεριόριστο (time unbound).

- Ένα workload μπορεί να έχει κάποιο χαμηλότερο απαιτούμενο όριο των πόρων που χρειάζεται (resource needs).

3) ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Η Google είναι η πρώτη εταιρεία-πάροχος υπηρεσιών cloud η οποία δημοσίευσε τα δεδομένα της υποδομής του cloud της (Cluster Trace Data) [17] το 2011. Τα δεδομένα αυτά, ανήκουν σε 12.500 ετερογενείς μηχανές και αφορούν 25 εκατομμύρια διεργασίες-tasks, σε ένα χρονικό διάστημα 29 ημερών. Πάνω σε αυτή την δημοσίευση της Google έχουν διεξαχθεί αρκετές μελέτες από διαφορετικές οπτικές γωνίες. Οι Reiss *et al.* [18], μελέτησαν την ετερογένεια και τη δυναμική των ιδιοτήτων του workload της Google. Οι Zhang *et al.* επικεντρώθηκαν στο χαρακτηρισμό της χρήσης των πόρων CPU, μνήμης και δίσκου, στο χρόνο εκτέλεσης (run-time) για κάθε διεργασία [19]. Η ομάδα των Liu *et al.* ασχολήθηκε με την συχνότητα και το μοτίβο των συμβάντων συντήρησης των μηχανημάτων (machine maintenance events), όπως επίσης και με τη συμπεριφορά των εργασιών (jobs) και των διεργασιών (tasks), και γενικά την αξιοποίηση των συνολικών πόρων του συστήματος [20]. Τέλος η ομάδα των Di *et al*. μελέτησαν τα φορτία εργασιών και μηχανών και συνέκριναν τις διαφορές μεταξύ ενός datacenter της Google και άλλων εγκαταστάσεων Grid / HPC [21].

Η Alibaba ακολουθώντας το παράδειγμα της Google, δημοσίευσε και αυτή για πρώτη φορά το 2017 το δικό της Cluster Trace Data, που περιέχει δεδομένα για τον τρόπο με τον οποίο αλληλεπιδρούν τα containers και τα batch workloads, σε κοινούς φυσικούς πόρους. Οι Lu *et al.* [9] ανέλυσαν το workload της Alibaba και επικεντρώθηκαν στην ανισορροπία χρήσης-αξιοποίησης των πόρων του συστήματος. Η ομάδα των Cheng *et al.* [22] ασχολήθηκε με τον τρόπο που συνυπάρχουν και αλληλεπιδρούν μεταξύ τους διάφορα workloads σε κοινούς φυσικούς πόρους. Με την ελαστικότητα που έχει η κατανομή πόρων στα συνυπάρχοντα συστατικά των workloads της Alibaba ασχολήθηκαν οι Liu *et al.* [23]. Ενώ άλλες έρευνες έχουν αξιοποιήσει τα δεδομένα του workload, κάνοντας σε αυτά ανάλυση αξιοπιστίας, όπως πρότυπα αποτυχίας εξόρυξης (mining failure patterns) [24], πρόβλεψη αποτυχίας (failure prediction) [25] κτλ..

Σε αντίστοιχη μελέτη, οι Cortez *et al.* [26] παρουσίασαν ένα εκτεταμένο χαρακτηρισμό του φόρτου εργασίας των VMs από την Microsoft Azure. Η έρευνα τους εξετάζει τις κατανομές της διάρκειας ζωής των VM (VMs' lifetime), το μέγεθος ανάπτυξης (VMs'deployment size) και την αξιοποίηση των πόρων, καθώς επίσης κάνει και μια πρόβλεψη της μελλοντικής συμπεριφοράς, βασιζόμενη στην συνέπεια των συμπεριφορών των Vms.

Άλλη μια παρόμοια μελέτη είναι αυτή των Wolski και Brevik [27], οι οποίοι περιγράφουν, αναλύουν και μοντελοποιούν το workload της Eucalyptus IaaS cloud και επικεντρώνονται κυρίως στη μετανάστευση του φόρτου εργασίας (workload migration) και στο χαρακτηρισμό του φόρτου εργασίας των datacenters.

Τέλος, οι Cooper *et al.* [28] μελετούν το workload της Yahoo (*Yahoo Cloud Serving Benchmark (YCSB))* με στόχο τη σύγκριση απόδοσης των νέων συστημάτων cloud δεδομένων.

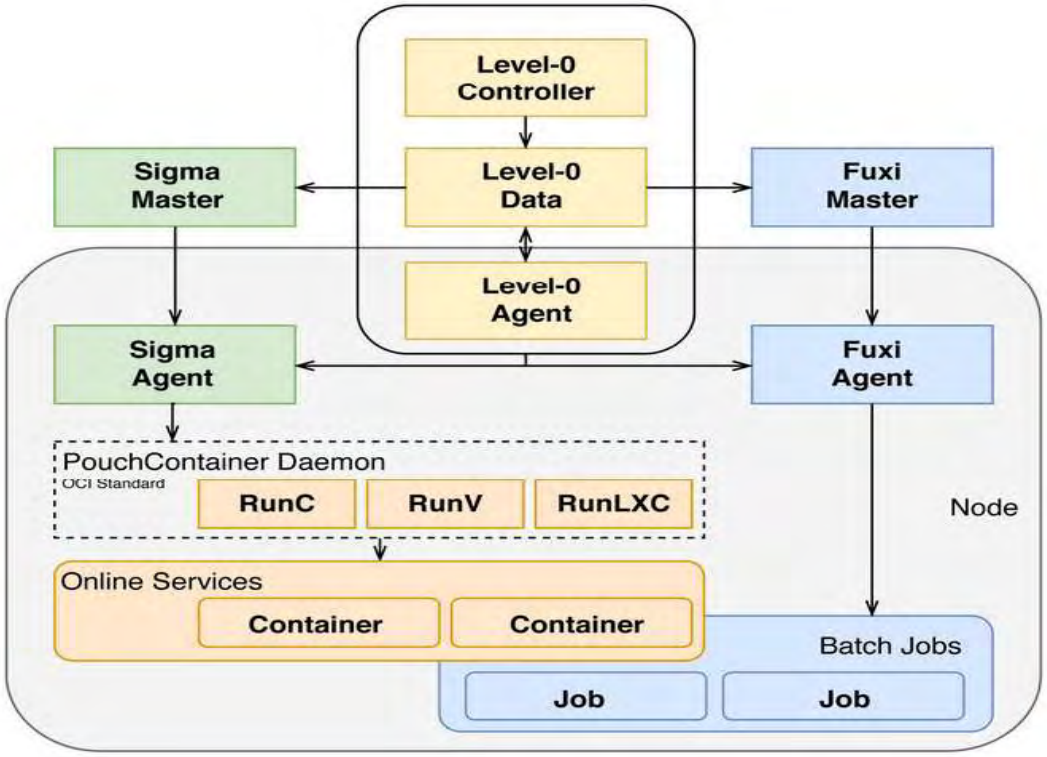
4) ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ (Dataset)

Στα πλαίσια του προγράμματος δημοσίευσης δεδομένων από συστήματα παροχής υπηρεσιών cloud της Alibaba (Alibaba Open Cluster Trace Program), η εταιρεία κυκλοφόρησε το 2ο της σύνολο δεδομένων για το 2018. Το trace, που προέρχεται από ένα από τα συγκροτήματα παραγωγής (production clusters), περιέχει την συνύπαρξη (co-locating) των εργασιών τόσο των online υπηρεσιών (online service jobs) όσο και των offline εργασιών συστήματος (offline batch jobs) σε κοινά φυσικά συμπλέγματα, όπως και το σύνολο δεδομένων του 2017. Ωστόσο, στο trace του 2018 ο βαθμός συνύπαρξης είναι αρκετά ενισχυμένος, γεγονός που αποφέρει βελτίωση στη συνολική χρήση των πόρων.

Αναλυτικότερα, στο trace, το οποίο και είναι μεγέθους 280 GB, υπάρχουν καταγραφές από 4000 μηχανήματα σε μία περίοδο 8 ημερών και το dataset είναι οργανωμένο σε 6 πίνακες σύμφωνα με το ακόλουθο μοντέλο:

- machine_meta.csv: περιέχει καταγραφές για τον αριθμό μηχανής(machine_id), την κατάσταση(status, μπορεί να είναι USING ή IMPORT_INSTALLING), τις προδιαγραφές(specifications) και το επίπεδο αποτυχίας(disaster level).
- machine_usage.csv: περιέχει καταγραφές για το χρόνο εκτέλεσης και την πραγματική χρήση πόρων, όπως CPU, μνήμη, δίκτυο κτλ.
- container_meta.csv: περιέχει στατικές πληροφορίες των containers για τους ζητούμενους πόρους όπως η κατάσταση(status), ο αριθμός του host στον οποίο τρέχει το container κ.α. Η κατάσταση ενός container μπορεί να είναι *Allocated,Started,Stopped,Unknown* όπου και σημαίνει αντίστοιχα ότι έχει τοποθετηθεί,ξεκίνησε,σταμάτησε και κάποιο πρόβλημα προέκυψε, για κάθε περίπτωση.
- container_usage.csv: οι πληροφορίες για την πραγματική χρήση πόρων για τα γεγονότα των containers (CPU, μνήμη, δίκτυο κτλ.)
- batch_task.csv: καταγράφονται οι πληροφορίες για τους ζητούμενους πόρους όπως χρόνοι έναρξης και λήξης(start-finish time), DAG εξαρτήσεις, ο αριθμός των instances κ.α.
- batch_instance.csv: καταγράφονται οι πληροφορίες εκτέλεσης και η πραγματική χρήση πόρων συμπεριλαμβανομένης και της τελικής κατάστασης (*Interrupted,Ready,Running,Terminated*)

Για να βελτιστοποιηθεί η αξιοποίηση πόρων, η Alibaba εφάρμοσε την συνύπαρξη των εργασιών των online υπηρεσιών και των offline εργασιών συστήματος. Η συνύπαρξη και ο συντονισμός (coordination) των 2 ειδών εργασιών γίνεται υπό τον έλεγχο 2 χρονοπρογραμματιστών (schedulers), οι οποίοι είναι ο SIGMA για τις online service και ο FUXI για τις offline batch εργασίες [22]. Οι SIGMA και FUXI έχουν ο καθένας τους δικούς τους διαθέσιμους πόρους (resource pool). Συγκεκριμένα, ο SIGMA είναι υπεύθυνος για τις μακροπρόθεσμες εργασίες υπηρεσιών που εκτελούνται στα containers, ενώ ο FUXI για τις εργασίες συστήματος που εκτελούνται σε φυσικές μηχανές. Για να επιτευχθεί ο συντονισμός μεταξύ των 2 χρονοπρογραμματιστών και να ληφθούν καλύτερες αποφάσεις κατανομής πόρων, την κεντρική διαχείριση, την επικοινωνία και την μεταφορά δεδομένων μεταξύ SIGMA και FUXI την επιβλέπει ένας κεντρικός προγραμματιστής επιπέδου 0 (Level-0 controller). Στην Εικόνα 1 περιγράφεται η αρχιτεκτονική της συνύπαρξης των SIGMA και FUXI:



Εικόνα 1. Αρχιτεκτονική Συνύπαρξης των Schedulers SIGMA και FUXI. Πηγή:https://github.com/alibaba/clusterdata/blob/master/cluster-trace-v2018/trace_2018.md

Για να περιγραφεί πλήρως μία εργασία συστήματος batch μπορεί να χρησιμοποιηθεί το 3πτυχο "job-task-instance". Δηλαδή, μία εργασία (job) μπορεί να αποτελείται από πολλές διεργασίες (tasks) και μια διεργασία μπορεί να αποτελείται από πολλά στιγμιότυπα (instances). Τα Instances, που αποτελούν την μικρότερη μονάδα χρονοπρογραμματισμού εντός του workload, εκτελούν τον ίδιο ακριβώς δυαδικό κώδικα, με ταυτόσημες απαιτήσεις πόρων, αλλά μπορούν να επεξεργάζονται διαφορετικές-ξεχωριστές ποσότητες δεδομένων.

Διάφοροι τύποι εργασιών batch, εμπεριέχονται στο Cluster-trace-v2018 της Alibaba. Υπάρχουν ορισμένες ανεξάρτητες batch εργασίες οι οποίες έχουν ονόματα με τυχαίους χαρακτήρες (π.χ.Task_Nzg3ODAwNDgzMTAwNTc2NTQ2Mw==), ενώ οι περισσότερες από αυτές ακολουθούν την περιγραφή εξαρτήσεων DAG (Directed Acyclic Graph).

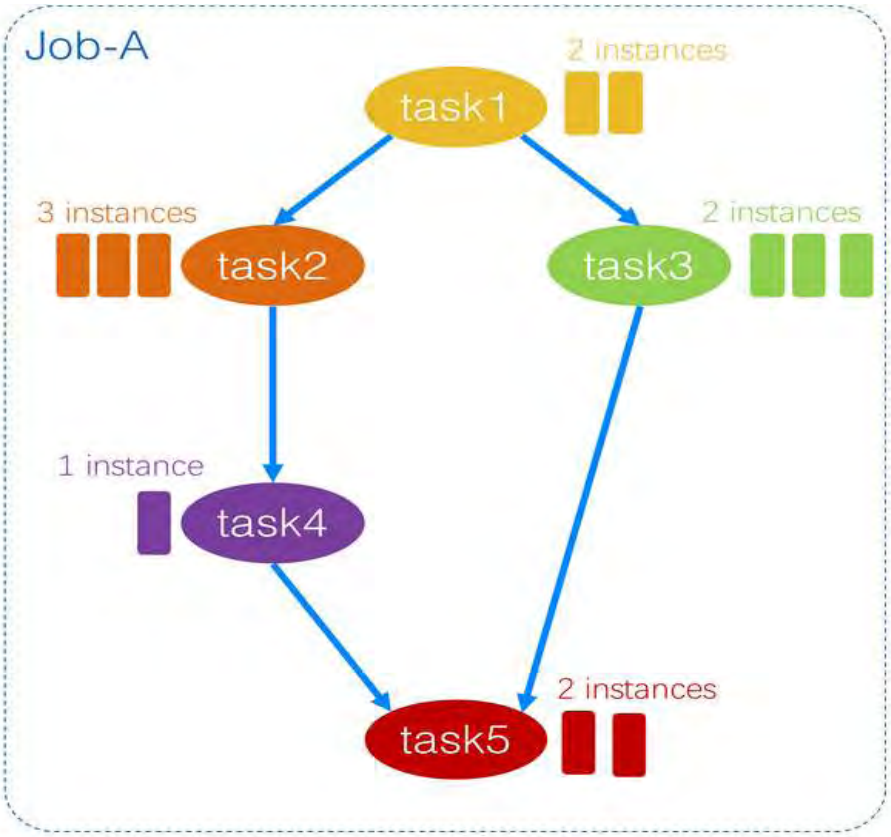
Έτσι λοιπόν σύμφωνα με το DAG, κάθε διεργασία task, που μπορεί να έχει πολλά στιγμιότυπα (instances), μπορεί να θεωρηθεί ως "ολοκληρωμένη" μόνο όταν ολοκληρωθούν όλα τα στιγμιότυπα της. Δηλαδή εάν η διεργασία-2 εξαρτάται από την διεργασία-1, δεν μπορεί να ξεκινήσει οποιοδήποτε στιγμιότυπο της διεργασίας-2 πριν ολοκληρωθούν όλα τα στιγμιότυπα της διεργασίας-1.

Ακολούθως, το DAG στο επίπεδο των εργασιών, μπορεί να συναχθεί από το πεδίο task_name όλων των διεργασιών (tasks) αυτής της εργασίας (job) και εξηγείται με το ακόλουθο παράδειγμα (Εικόνα 2):

Το DAG του ο Job-A αποτελείται από 5 διεργασίες με κάποιες εξαρτήσεις, εκφραζόμενες από το task_name τους.

- Το όνομα M1 του task1: σημαίνει ότι η task1 είναι μια ανεξάρτητη διεργασία και μπορεί να ξεκινήσει χωρίς να περιμένει οποιαδήποτε άλλη διεργασία.
- Το όνομα M2_1 του task2: σημαίνει ότι η task2 εξαρτάται από την ολοκλήρωση της task1.
- Το όνομα M3_1 του task3: σημαίνει ότι η task3 εξαρτάται από την ολοκλήρωση της task1.
- Το όνομα R4_2 του task4: σημαίνει ότι η task4 εξαρτάται από την ολοκλήρωση της task2.
- Το όνομα M5_3_4 του task5: σημαίνει ότι η task5 εξαρτάται τόσο από την task3 όσο και από την task4, δηλαδή ότι η task5 δεν μπορεί να ξεκινήσει πριν ολοκληρωθούν όλες οι περιπτώσεις task3 και task4.

(Σημειώνεται ότι μόνο ο αριθμητικός δείκτης στο task_name έχει σημασία, ενώ ο πρώτος χαρακτήρας (π.χ. M, R στο παράδειγμα) δεν έχει καμία σχέση με την εξάρτηση.)



Εικόνα 2. Το DAG (Directed Acyclic Graph) για την Εργασία-Α. Πηγή:https://github.com/alibaba/clusterdata/blob/master/cluster-trace-v2018/trace_2018.md

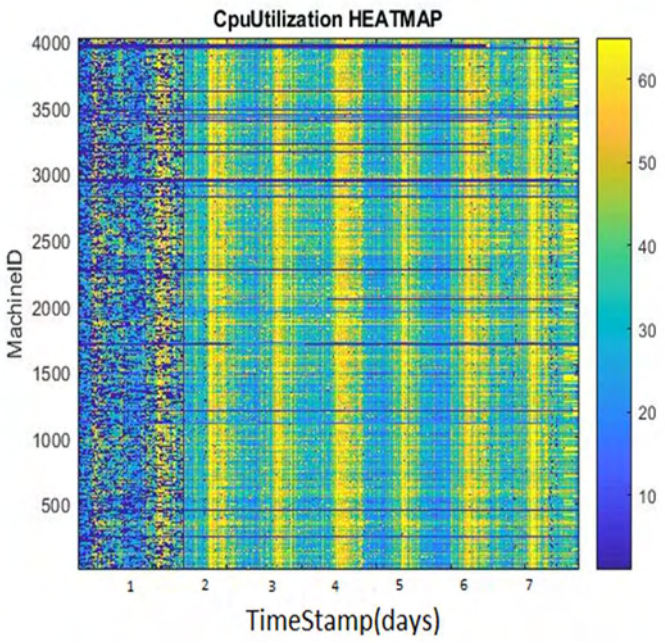
5) ΑΝΑΛΥΣΗ

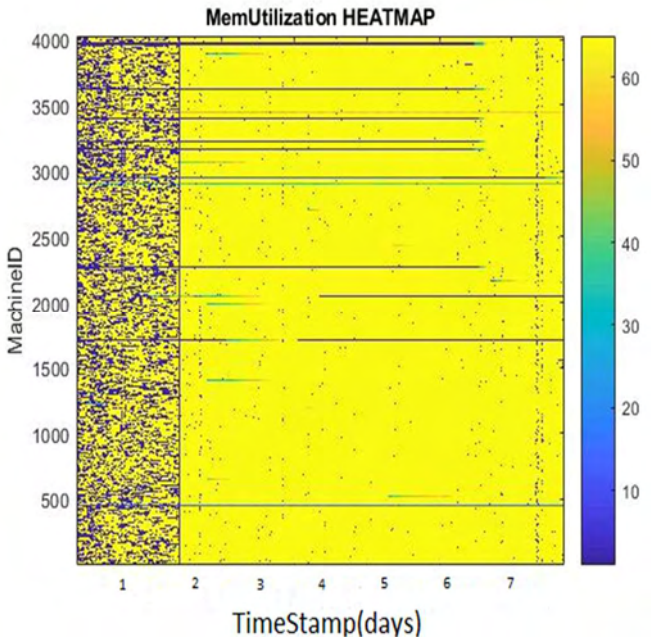
Η ανάλυση των δεδομένων πραγματοποιήθηκε με τη χρήση του Matlab R2017 σε Intel(r) Core(tm) i7-2630qm cpu @ 2.00ghz και 10 GB RAM MEMORY. Λόγω του μεγάλου όγκου των δεδομένων (280 GB), για την επεξεργασία στο Matlab, τα .csv αρχεία αντιμετωπίστηκαν ως Databases και επεξεργάστηκαν τμηματικά. Σε περιπτώσεις που ήταν αδύνατο να διαβαστούν από τη μνήμη τα επιθυμητά δεδομένα, λήφθηκε ο μέσος όρος ανά μικρές ομάδες σε τιμές των καταχωρήσεων. Επιπλέον, για την έλεγχο ποιότητας των δεδομένων, έγινε data filtering σε μη έγκυρες τιμές, οι οποίες αντικαταστάθηκαν από την ακριβώς προηγούμενη τους έγκυρη τιμή.

5.1) Ανισορροπία Αξιοποίησης Πόρων (Machine Imbalance)

Στην Εικόνα 3 αποτυπώνεται κανονικοποιημένη η αξιοποίηση των πόρων CPU (Εικόνα 3α) και MEMORY (Εικόνα 3β) για όλο το εύρος του δοθέντος χρόνου και για όλο το εύρος των συστημάτων που αποτελούν την υποδομή cloud. Τα δεδομένα της τα οποία προέρχονται από το αρχείο "machine_usage", απαρτίζουν μια πλήρη περιγραφή των όλων γεγονότων του trace στις μηχανές, για 8 περίπου μέρες.

Η απεικόνιση αυτή, γίνεται με τη χρήση διαγράμματος Heatmap. Για κάθε μηχανή (Machine_id), σε κάθε χρονική στιγμή (time_stamp) σε μορφή συντενταγμένων, υπάρχει η τιμή της 3ης μεταβλητής (CPU/MEMORY), η οποία εκφράζεται με μια κλίμακα χρωμάτων.



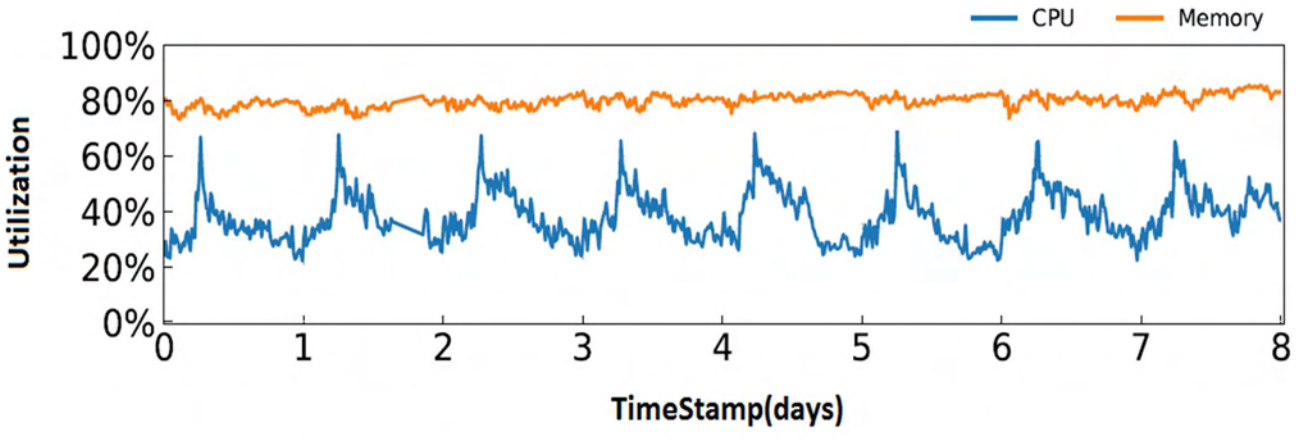


Εικόνα 3α

Εικόνα 3β

Εικόνα 3. Απεικόνιση με διάγραμμα Heatmap της αξιοποίησης των πόρων CPU (Εικόνα 3α) και MEMORY (Εικόνα 3β) για όλο το εύρος του δοθέντος χρόνου (8 ημέρες).

Πιο συγκεκριμένα, από το διάγραμμα της Εικόνα 3α παρατηρείται ένα επαναλαμβανόμενο μοτίβο αυξομείωσης στην αξιοποίηση (utilization) των CPU. Επίσης, προκύπτει ότι υπάρχουν "περιοχές" μηχανημάτων (Οριζόντιες μπλε γραμμές, π.χ. Μηχανές με ID κοντά στο 250,500,1250, κτλ.) με ελάχιστη ή καθόλου χρήση των CPU και αντίστοιχα κάποιες περιοχές μηχανημάτων με αυξημένη χρήση των CPU (Οριζόντιες κίτρινες γραμμές, π.χ. Μηχανές με ID κοντά στο 450,550,1900, κτλ.). Τέλος, παρατηρείται μια ασαφή γενικά κατάσταση λόγω και των ελλιπών στοιχείων του trace στο συγκεκριμένο διάστημα, μέχρι περίπου τη 2η ημέρα. Αναλυτικότερα, παρατηρείται ότι πριν το τέλος της 1ης ημέρας υπάρχει μία πτώση στο CPU utilization σε όλα τα μηχανήματα(έντονη κάθετη μπλε γραμμή).Λίγο μετά τις 36 ώρες παρατηρείται άλλη μια απότομη πτώση (στενή κάθετη μπλε γραμμή) στο CPU utilization, ενώ μετά τη 2η ημέρα παρατηρείται καθαρά ένα επαναλαμβανόμενο μοτίβο, όπου στην αρχή κάθε 24ωρου υπάρχει αυξημένη χρήση πόρων CPU η οποία πέφτει σταδιακά προς τέλος κάθε ημέρας. Το επαναλαμβανόμενο μοτίβο, όπως αναφέρεται και παρακάτω (Εικόνα 9), οφείλεται σε 2 γεγονότα που συμβαίνουν κατά τις νυχτερινές ώρες: α) την αύξηση των αφίξεων περισσότερων batch εργασιών και β) την συμπεριφορά του FUXI να προγραμματίζει τις εργασίες αυτές τις νυχτερινές ώρες (Εικόνα 4) λόγω μειωμένης ζήτησης πόρων από τις online εργασίες των πελατών.

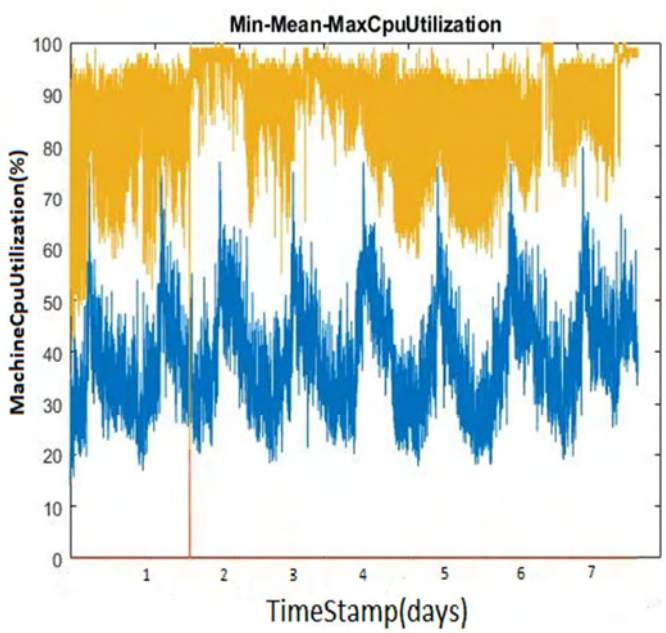


Εικόνα 4. CPU και MEMORY utilization σε όλη τη διάρκεια των 8 ημερών του trace.

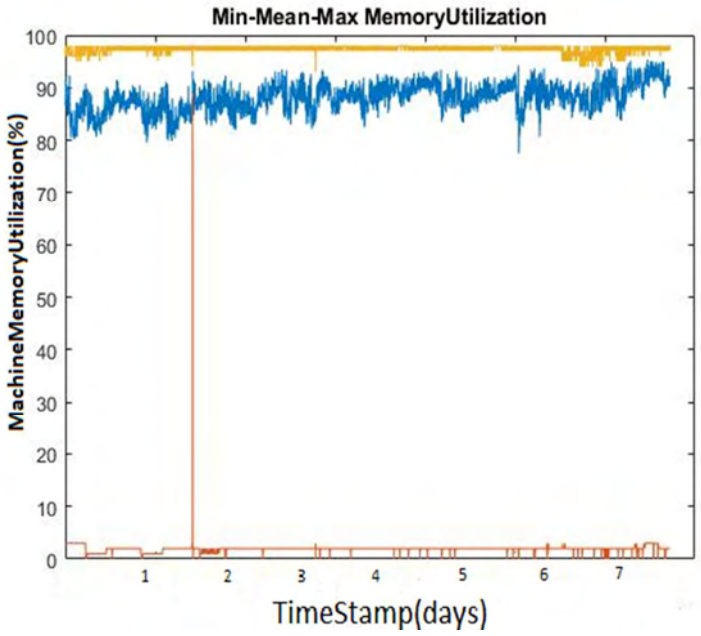
Σε αντίθεση με τις πολλές αυξομειώσεις που συμβαίνουν στην αξιοποίηση των CPU, στην Εικόνα 3β παρατηρείται συνολικά μία πιο αυξημένη χρήση της μνήμης (memory utilization) σε σύγκριση με το CPU utilization, καθόλη τη διάρκεια των 8 ημερών. Υπάρχουν όμως πάλι κάποιες "περιοχές" μηχανημάτων με ελάχιστη εώς καθόλου αξιοποίηση της μνήμης (οριζόντιες μπλε γραμμές, π.χ. μηχανές με ID κοντά στο 500,1750,2300 κτλ.). Παράλληλα συνεχίζει να υπάρχει και σε αυτήν την αποτύπωση η πιο ασαφής περιοχή, μέχρι τη χρονική στιγμή λίγο μετά τις 36 ώρες, όπου και πάλι γίνεται μία απότομη αύξηση στο memory utilization.

Προκύπτει λοιπόν το συμπέρασμα ότι στο trace υπάρχει χωρική ανισορροπία (spatial imbalance), αφού εντοπίζεται ετερογενής αξιοποίηση πόρων στο σύνολο των μηχανημάτων, καθώς επίσης και το συμπέρασμα ότι υπάρχει και χρονική ανισορροπία (temporal imbalance) που εντοπίζεται από την κυμαινόμενη χρήση στο χρόνο, των πόρων σε κάθε μηχάνημα. Επιπλέον, όπως επιβεβαιώνεται τόσο από την Εικόνα 3β όσο και από την Εικόνα 5β, το workload δείχνει να είναι memory bound. Ότι δηλαδή, η αποδοτικότητα των πόρων στο datacenter, περιορίζεται από τη μνήμη η οποία και ανάγεται σε σημείο συμφόρησης (bottleneck), ενώ την ίδια στιγμή υπάρχουν διαθέσιμοι πόροι CPU.

Μία πιο λεπτομερής ανάλυση της χρήσης των πόρων αποτυπώνεται στην Εικόνα 5, όπου συνοψίζονται η μέγιστη (Max), η μέση (Mean) και η ελάχιστη (Min) αξιοποίηση πόρων σε ένα διάγραμμα, για καθένα από τα CPU και MEMORY.



Εικόνα 5α



Εικόνα 5β

Εικόνα 5. Μέγιστη (Max), Μέση (Mean) και Ελάχιστη (Min) αξιοποίηση πόρων στο σύνολο των μηχανημάτων σε ένα διάγραμμα, για τα CPU (Εικόνα 5α) και MEMORY (Εικόνα 5β).

Όπως φαίνεται από την Εικόνα 5α, η Μέση CPU Utilization κάθε χρονική στιγμή κυμαίνεται μεταξύ των 20-80%. Επιβεβαιώνεται επίσης, το επαναλαμβανόμενο μοτίβο που παρατηρήθηκε στην Εικόνα 3, καθώς επίσης και η ασταθής συμπεριφορά μέχρι τις 36+ ώρες. Επιπλέον φαίνεται ότι η Μέγιστη CPU utilization, που έχει μεγάλες διακυμάνσεις στην αρχή (από 45 μέχρι 95%) φτάνει προς το τέλος σε υψηλές τιμές κοντά στο 100%. Ενώ η Ελάχιστη CPU utilization είναι σχεδόν παντού 0%, εκτός της χρονικής στιγμής λίγο μετά τις 36 ώρες, όπου και είναι λίγο πάνω από 20%.

Ομοίως από την Εικόνα 5β φαίνεται ότι η μέση MEMORY utilization κυμαίνεται μεταξύ του 80 και του 95%, έχοντας προς το τέλος μια αυξητική τάση. Η μέγιστη MEMORY utilization έχει τιμές γύρω από το 95%. Ενώ αξιοσημείωτη είναι η συμπεριφορά της ελάχιστης MEMORY utilization, που ενώ είναι σχεδόν σε όλη τη διάρκεια του trace κάτω από 5%, στις 36+ ώρες έχει μία εκρηκτική άνοδο και φτάνει το 90%.

Συγκρίνοντας τις μεγάλες διαφορές μεταξύ των Max, Mean, Min CPU utilizations των μηχανημάτων, γίνεται και πάλι εμφανής η χωρική ανισορροπία (spatial imbalance) και παράλληλα η ανάγκη για χρονοπρογραμματιστές με καλύτερη απόδοση στα "σημεία" όπου η διαχείριση του workload επιδέχεται βελτίωση, ώστε να κατανέμεται πιο ομαλά στο σύνολο των μηχανημάτων του cluster. Τέτοια είναι τα σημεία όπου η Max CPU utilization είναι σχετικά χαμηλή και άρα τα μηχανήματα δουλεύουν σε πιο χαμηλή αξιοποίηση από το επιθυμητό.

Σε αντίθεση με το CPU utilization, το MEMORY utilization έχει πιο σταθερή-καλή συμπεριφορά, καθώς η διαφορά μεταξύ των Max και Mean τιμών είναι μικρή. Άρα η βελτιστοποίηση των χρονοπρογραμματιστών για καλύτερο MEMORY utilization είναι μία πρόκληση, αφού οι schedulers λειτουργούν σχετικά κοντά στο επιθυμητό. Παράλληλα, σε περίπτωση αλλαγής της πολιτικής των προγραμματιστών για το CPU utilization μπορεί να μειωθούν οι δεσμεύσεις για βελτίωση της απόδοσης της μνήμης και να υπάρξει υποβάθμιση απόδοσης (performance degradation) των ίδιων των μηχανημάτων, πράγμα που σίγουρα δεν είναι επιθυμητό.

5.2) Ανισορροπίες στο Workload

Ι) ΠΟΣΟΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Όπως έχει αναφερθεί και πιο πάνω, στο workload υπάρχουν 2 είδη εργασιών. Τις μακροπρόθεσμες online εργασίες υπηρεσιών (online service jobs) και τις βραχυπρόθεσμες offline εργασίες συστήματος (offline batch jobs). Κάθε στιγμιότυπο (instance) μιας online εργασίας υπηρεσιών "τρέχει" εντός ενός Linux container και έχει συγκεκριμένη χρονοσφραγίδα (time_stamp). Η τιμή 0 σε μία χρονοσφραγίδα σημαίνει ότι η χρονική στιγμή που περιγράφεται είναι πιο πριν ή πιο μετά από την διάρκεια των 8+ ημερών του trace. Ο συνολικός αριθμός των service instances (container_id εντός του trace) ανέρχεται στις 71467, ενώ έχει παρατηρηθεί ότι οι online service jobs τρέχουν και στα 4000 μηχανήματα.

Από την άλλη, κάθε στιγμιότυπο (instance) μιας offline εργασίας συστήματος ανήκει σε μία διεργασία (task). Κατά μέσο όρο προκύπτει ότι μία διεργασία έχει 98

στιγμιότυπα, ενώ το εύρος των στιγμιοτύπων σε μία διεργασία είναι από 1 το ελάχιστο μέχρι 99583 το μέγιστο. Παράλληλα, οι batch εργασίες μπορούν να χαρακτηριστούν χρονικά από τα start_time & end_time, με τη μέγιστη διάρκεια να είναι 691142 sec (δηλαδή περίπου 8 μέρες), μέση διάρκεια 1,76 sec και ελάχιστη τα 0 sec. Με τέτοιες διαφορές τόσο στον αριθμό όσο και στην διάρκεια του χρόνου εκτέλεσης των online service jobs και των offline batch jobs, η ανισορροπία στο workload πρέπει να αναλυθεί για κάθε μία από τις παραπάνω κατηγορίες ως προς τις απαιτήσεις των πόρων που απαιτούν (resource demands).

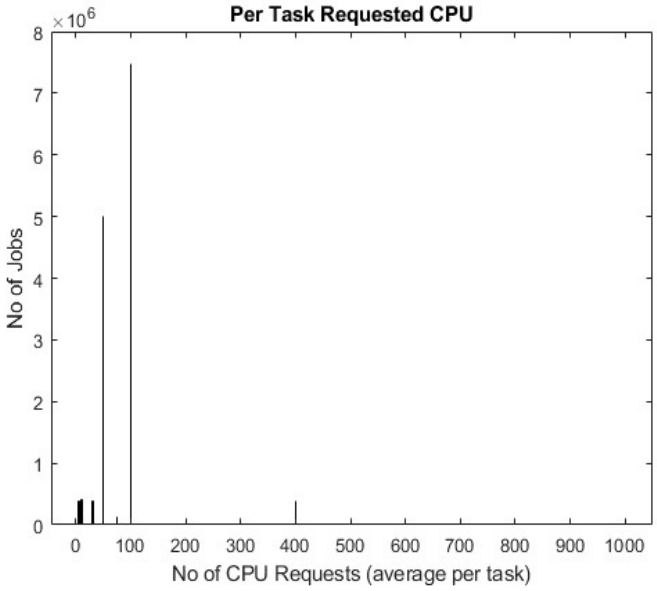
ΙΙ) ΑΠΑΙΤΗΣΕΙΣ ΠΟΡΩΝ ΤΩΝ ΕΡΓΑΣΙΩΝ ΣΥΣΤΗΜΑΤΟΣ (BATCH JOBS)

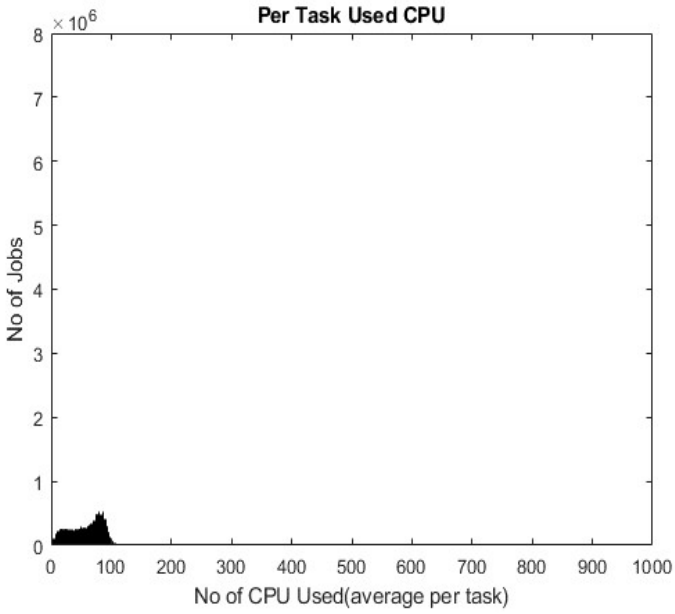
Για να γίνει εμφανές το πρόβλημα όσον αφορά την απαίτηση και την αξιοποίηση των ζητούμενων πόρων, για κάθε εργασία (job), συνοψίστηκε ο μέσος όρος των αιτούμενων (requested) και των χρησιμοποιούμενων (used) πόρων CPU ανά διεργασία στα 3 διαγράμματα της Εικόνα 6. Αναλυτικότερα, τα δεδομένα αιτούμενων πόρων αντλήθηκαν από το αρχείο "batch_task", ενώ τα δεδομένα χρησιμοποιούμενων πόρων από το αρχείο "batch_instance". Για να εξαχθούν τα δεδομένα από το αρχείο "batch_instance", για κάθε διεργασία μιας εργασίας, ελήφθησαν οι εκάστοτε τιμές των στιγμιοτύπων που ανήκουν στην κάθε διεργασία.

Από το διάγραμμα της Εικόνα 6α, προκύπτει ότι 7,4M Batch εργασίεςς αιτούνται κατά μέσο όρο 1 CPU ανά διεργασία (Η τιμή 100 στο trace ισοδυναμεί με 1 CPU). Ένας αριθμός περίπου 5M Batch εργασιών αιτούνται κατά μέσο όρο 0,5 CPU, ενώ όπως είναι αναμενόμενο, υπάρχουν και αρκετές εργασίες που αιτούνται περισσότερες από 1 CPUs.

Την ίδια στιγμή, από το διάγραμμα της Εικόνα 6β και Εικόνα 6γ, παρατηρείται ότι οι περισσότερες batch εργασίες χρησιμοποιούν 0,8 CPUs. Εξάγονται λοιπόν 2 συμπεράσματα από την σύγκριση των 2 διαγραμμάτων. Το πρώτο είναι ότι η πλειοψηφία των εργασιών αιτούνται 1 CPU και χρησιμοποιούν ανάμεσα σε 0,7-0,9 CPUs, άρα αφαιρετικά μπορούμε να πούμε η σπατάλη για ένα μεγάλο ποσοστό εργασιών είναι μικρή. Προκύπτει ωστόσο από το διάγραμμα της Εικόνα 6β και Εικόνα 6γ, ότι οι τιμές ανάμεσα στο διάστημα 0,6-1είναι αθροιστικά περισσότερες από τις 100+K του διαγράμματος της Εικόνα 6α, πράγμα που σημαίνει ότι συνεχίζουν να υπάρχουν πάρα πολλές εργασίες οι οποίες αιτούνται μεγαλύτερο αριθμό πόρων CPU

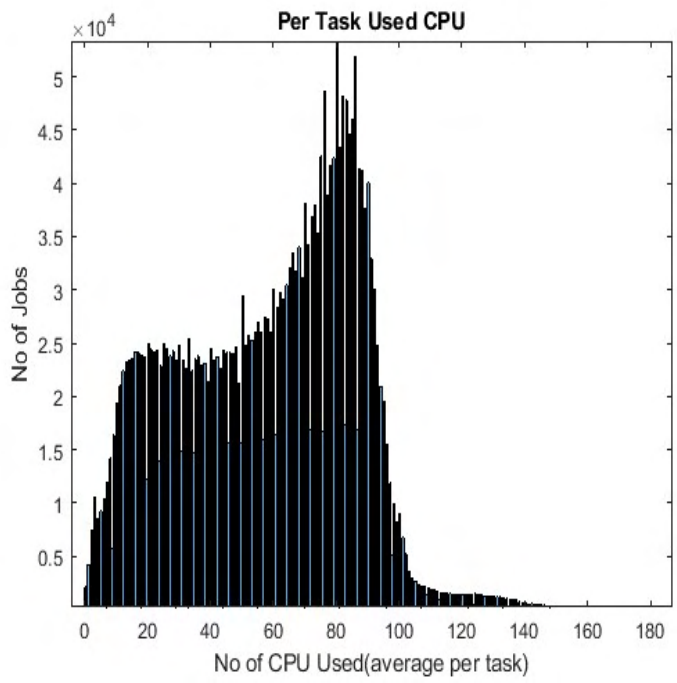
από αυτόν που πραγματικά χρησιμοποιούν. Για παράδειγμα μια εργασία που αιτήθηκε 1 CPU μπορεί να χρησιμοποίησε 0,4 CPUs ή άλλη περίπτωση με μεγαλύτερη σπατάλη, μια εργασία που αιτήθηκε 4 CPUs μπορεί να χρησιμοποίησε μόνο 1,4 CPUs κτλ. Η κλίμακα στα 3 διαγράμματα είναι διαφορετική ώστε να είναι ευκρινής η διαφορά στον παρατηρητή.





Εικόνα 6α

Εικόνα 6β



Εικόνα 6γ

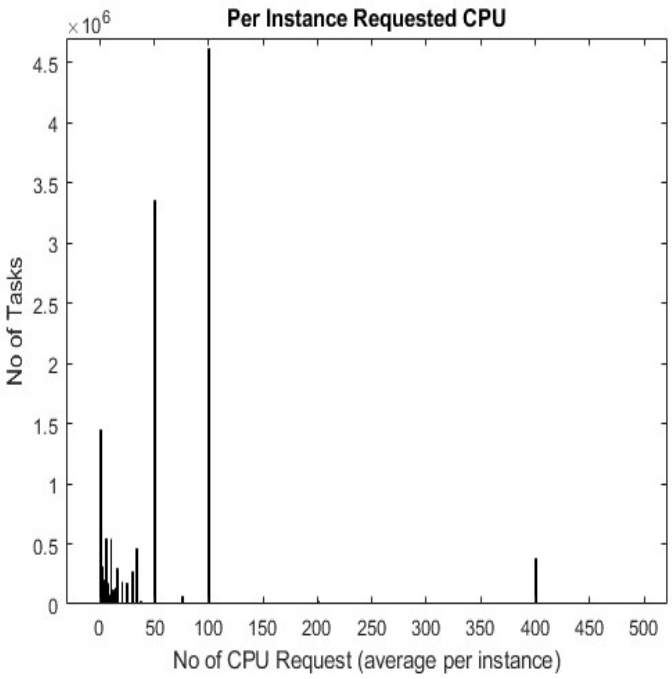
Εικόνα 6. Ο μέσος όρος για κάθε Εργασία (Job) των Αιτούμενων (Requested) (Εικόνα 6α) και των Χρησιμοποιούμενων (Used) πόρων CPU ανά Διεργασία (Εικόνα 6β).Το διάγραμμα της Εικόνας 6γ είναι ίδιο με το διάγραμμα της Εικόνας 6β σε διαφορετική κλίμακα.

Επιπλέον, παρατηρήθηκε ότι ενώ πολλές batch εργασίες αιτούνται προκαταβολικά περισσότερους πόρους από αυτούς που πραγματικά χρειάζονται. Έτσι, υπάρχουν εργασίες στο workload που περιμένουν (έχουν στο status: *Ready)* αφού δεν υπάρχουν ελεύθεροι πόροι. Άρα το συμπέρασμα που προκύπτει, είναι ότι ο scheduler δέχεται το overestimation των πόρων που ζητάει μία εργασία χωρίς να λαμβάνει υπόψιν του την πραγματική χρήση πόρων.

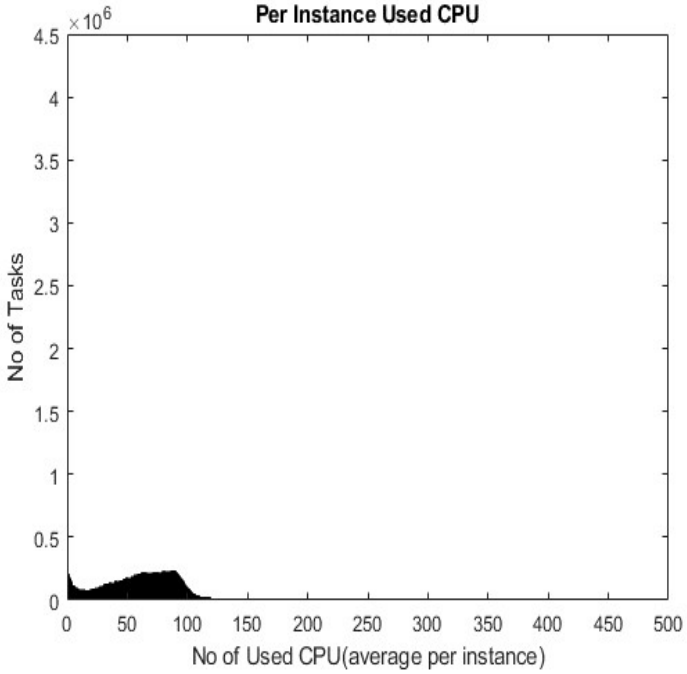
Στις Εικόνα 7 και Εικόνα 8 συνοψίζεται για κάθε διεργασία (task), ο μέσος όρος των αιτούμενων (requested) και των χρησιμοποιούμενων (used) πόρων CPU και MEMORY, ανά στιγμιότυπο (instance) (Η τιμή 100 στο trace είναι κανονικοποιημένη και ισοδυναμεί με 1 CPU). Για να γίνει αυτό εφικτό, στην περίπτωση των requested από το αρχείο "batch_task" χρησιμοποιήθηκαν για κάθε task_name, ο αριθμός των στιγμιοτύπων (instance_num) και τα αιτούμενα CPU & MEMORY (plan_cpu &plan_mem). Έτσι η τιμή που αναπαριστάται στις Εικόνα 7α και Εικόνα 8α είναι το πηλίκο της διαίρεσης των αιτούμενων πόρων προς τον αριθμό των στιγμιοτύπων εντός της κάθε διεργασίας. Αντίστοιχα για να αναπαρασταθούν οι τιμές των Εικόνα 7β και Εικόνα 8β, χρησιμοποιήθηκαν από το αρχείο "batch_instance" για κάθε Task_Name & Instance_Name ο αριθμός των αιτούμενων Cpu_avg και Mem_avg στην κάθε περίπτωση. Οι τιμές οι οποίες αναπαρίστανται στις Εικόνα 7β και Εικόνα 8β είναι ο μέσος όρος ανά 100, των στοιχείων των Cpu_avg και Mem_avg. Η επιλογή αυτή έγινε σύμφωνα με τα ποσοτικά χαρακτηριστικά που εξετάστηκαν παραπάνω και συγκεκριμένα σύμφωνα με την παρατήρηση ότι μία διεργασία έχει κατά μέσο όρο περίπου 100 στιγμιότυπα (για την ακρίβεια 98 Instances).

Από την ανάλυση των δεδομένων της Εικόνα 7, προκύπτει ότι ενώ η πλειοψηφία των αιτήσεων πόρων για CPU είναι είτε 1 είτε 0,5, η πλειοψηφία των πραγματικά

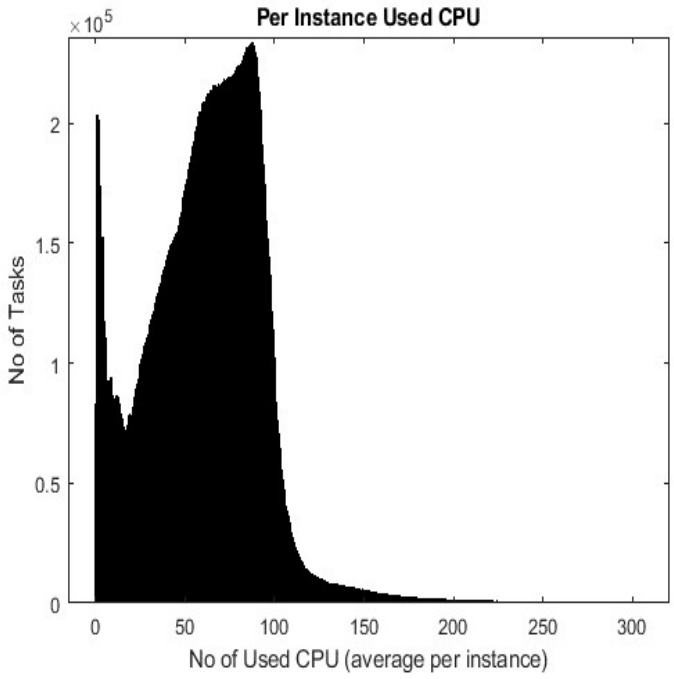
χρησιμοποιούμενων πόρων είναι ανάμεσα στις τιμές 1 και 0,5 ή στο κοντά στο 0. Παρατηρείται δηλαδή, σημαντική σπατάλη πόρων σε διάφορες περιπτώσεις, όπως για παράδειγμα μπορεί να υπάρχει διεργασία που αιτήθηκε 1 CPU και χρησιμοποίησε 0,5 ή και 0 CPUs, ή και περίπτωση στην οποία Διεργασία αιτήθηκε 4 CPUs και τελικά χρησιμοποίησε κοντά στις 2 CPUs. Η κλίμακα στα 3 διαγράμματα είναι και πάλι διαφορετική για να είναι ευκρινής η διαφορά στον παρατηρητή.



Εικόνα 7α



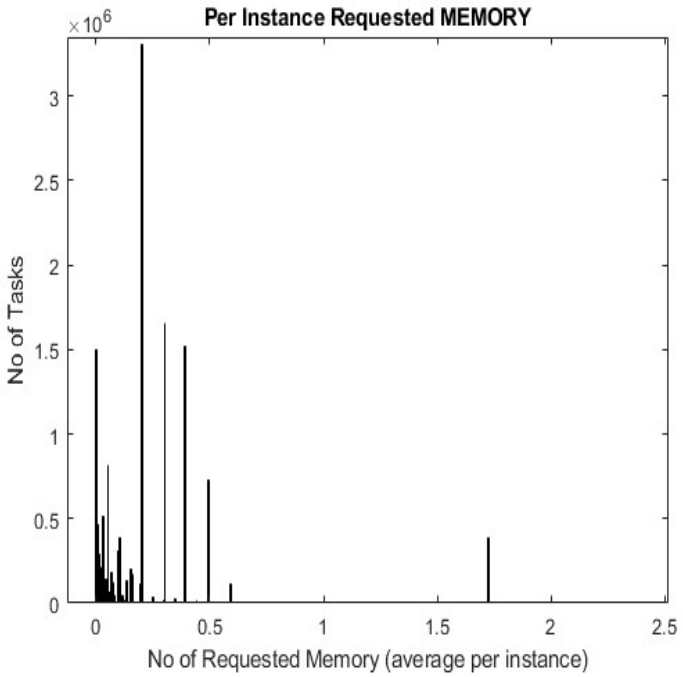
Εικόνα 7β



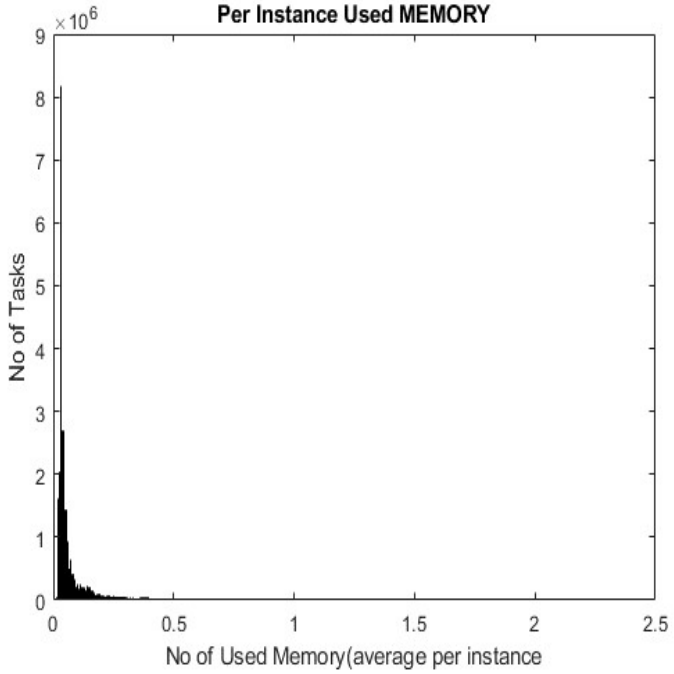
Εικόνα 7γ

Εικόνα 7. Ο μέσος όρος για κάθε Στιγμιότυπο (Instance) των Αιτούμενων (Requested) (Εικόνα 7α) και των Χρησιμοποιούμενων (Used) πόρων CPU ανά Διεργασία (Εικόνα 7β).Το διάγραμμα της Εικόνας 7γ είναι ίδιο με το διάγραμμα της Εικόνας 7β σε διαφορετική κλίμακα.

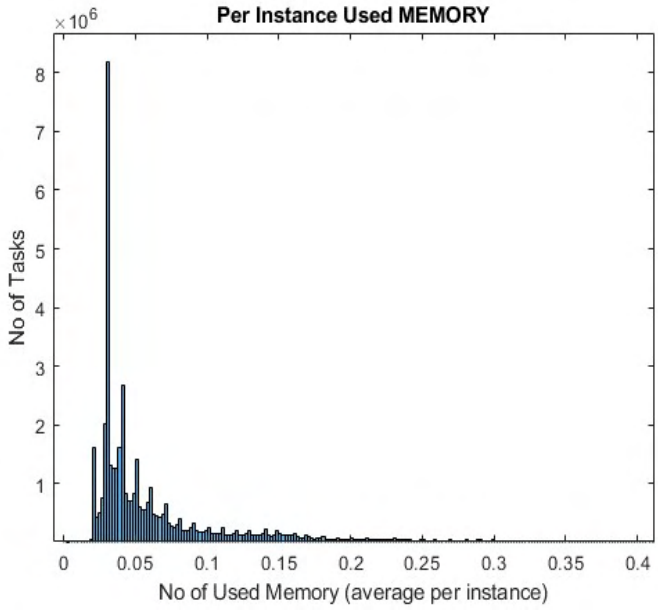
Στο trace της Alibaba, όλα τα μηχανήματα έχουν κανονικοποιημένη 1 μονάδα μνήμης. Έτσι, από τα δεδομένα της Εικόνα 8 προκύπτει και πάλι ότι ενώ η πλειοψηφία των αιτήσεων για πόρους μνήμης είναι ανάμεσα στο 0,25 και 0,5, η πλειοψηφία στην πραγματική χρήση πόρων φτάνει μόλις στο 0,025 με 0,05. Υπάρχουν επίσης, σε αντιστοιχία με την χρήση πόρων CPU, και πιο ακραίες περιπτώσεις σπατάλης πόρων, όπως για παράδειγμα η περίπτωση σημαντικού αριθμού αιτήσεων για 1,75 πόρους μνήμης, που όπως φαίνεται δεν αξιοποιούνται καθόλου σε αυτό το βαθμό.



Εικόνα 8α



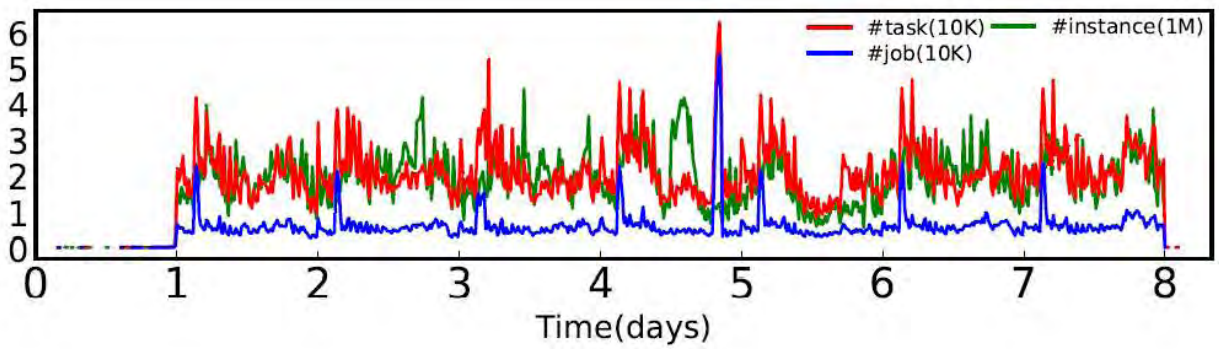
Εικόνα 8β



Εικόνα 8γ

Εικόνα 8. Ο μέσος όρος για κάθε Στιγμιότυπο (Instance) των Αιτούμενων (Requested) (Εικόνα 8α) και των Χρησιμοποιούμενων (Used) πόρων MEMORY ανά Διεργασία (Εικόνα 8β κ 8γ).

Τέλος, όπως και σε άλλες έρευνες [29] έχει παρατηρηθεί το μοτίβο ημερήσιας και νυχτερινής δραστηριότητας στην άφιξη-εξυπηρέτηση των αιτήσεων των batch και των service εργασιών. Έτσι και στο trace_2018 της Alibaba εντοπίστηκε παρόμοια συμπεριφορά. Λόγω της ημερήσιας δραστηριότητας των χρηστών και της διασφάλισης της ποιότητας υπηρεσιών (quality of service), οι online υπηρεσίες είναι περισσότερες και εξυπηρετούνται κατά προτεραιότητα την ημέρα εις βάρος των batch εργασιών που έχουν προκαθορισμένους διαθέσιμους πόρους. Την νύχτα όπου υπάρχουν λιγότερες online υπηρεσίες, χρονοπρογραμματίζονται από τον FUXI οι batch εργασίες συστήματος. Το γεγονός αυτό επιβεβαιώνει το διάγραμμα της Εικόνα 9, όπου παρατηρείται αύξηση στην άφιξη των εργασιών batch κάθε ξημέρωμα στις 2:00-4:00.

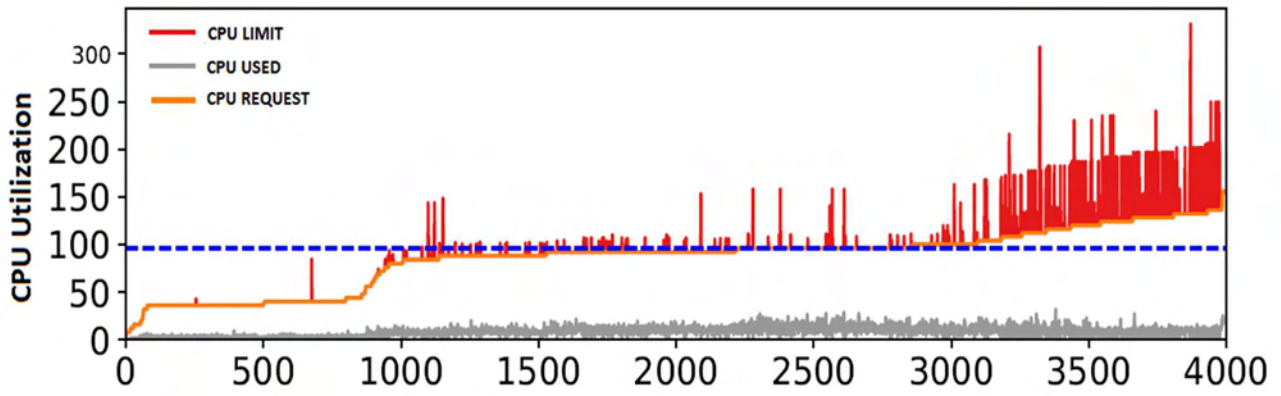


Εικόνα 9. Ο αριθμός των Batch εργασιών (Jobs), των διεργασιών (Tasks) και των στιγμιοτύπων (Instances) τους στο χρόνο τον οποίο χρονοπρογραμματίζονται από τον FUXI.

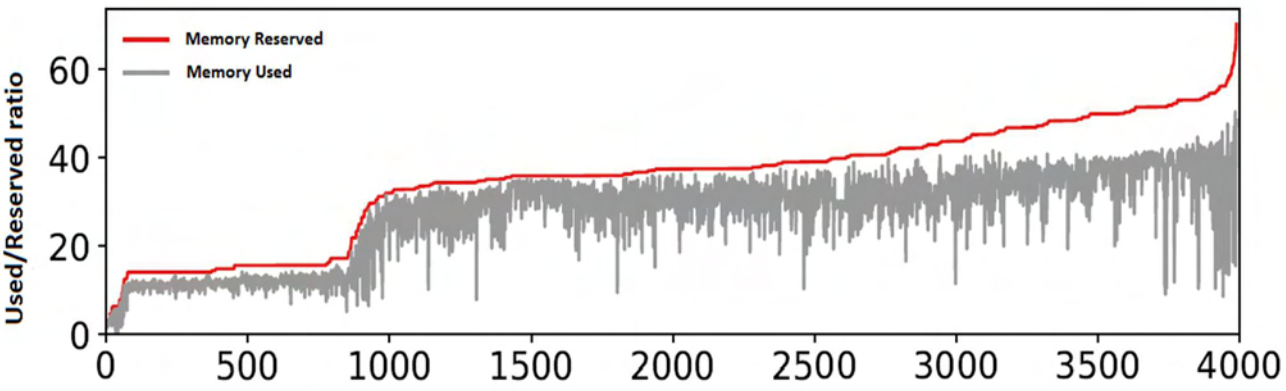
III) ΑΠΑΙΤΗΣΕΙΣ ΠΟΡΩΝ ΤΩΝ ΕΡΓΑΣΙΩΝ ONLINE ΥΠΗΡΕΣΙΩΝ (SERVICE JOBS)

Για την διαχείριση των πόρων στις online υπηρεσίες, ο SIGMA ακολουθεί μια συντηρητική πολιτική για να αναθέσει τους πόρους στα διάφορα containers. Αυτό γίνεται εμφανές στα διαγράμματα α και β της Εικόνα 10, όπου για κάθε μηχάνημα υπολογίζονται αθροιστικά (CDF) οι πόροι που απαιτούνται σε κάθε container. Οι τιμές που αποτυπώνονται στο διάγραμμα είναι ο αριθμός πόρων (κανονικοποιημένος στα 100 για CPU), το όριο, δηλαδή ο μέγιστος αριθμός πόρων που μπορούν να χρησιμοποιήσουν τα container και τέλος η πραγματική χρήση πόρων. Προκύπτει

λοιπόν, ότι ενώ το άθροισμα των αιτούμενων πόρων CPU από τις online υπηρεσίες είναι σχεδόν ίσο με τους συνολικούς πόρους του cluster, κατά μέσο όρο πάνω από 90% (Εικόνα 10α) και οι πόροι μνήμης που δεσμεύονται είναι λίγο κάτω από 40% (Εικόνα 10β), η πραγματική χρήση της CPU είναι κατά μέσο όρο κάτω από 10% και η πραγματική χρήση μνήμης περίπου 80%. Ο λόγος που συμβαίνει αυτή η μεγάλη προ-δέσμευση πόρων(over-provisioning) είναι η ανάγκη να εξασφαλιστούν οι συνθήκες ποιότητας (quality of service) και οι SLA συμφωνίες από την πλευρά του παρόχου cloud υπηρεσιών.



Εικόνα 10α



Εικόνα 10β

Εικόνα 10. Για κάθε μηχάνημα υπολογίζονται αθροιστικά (CDF) οι πόροι που απαιτούνται σε κάθε container. Στην Εικόνα 10α αποτυπώνονται η αίτηση, το όριο και η πραγματική χρήση CPU στα containers. Στην Εικόνα 10β η δεσμευμένη και η χρησιμοποιούμενη μνήμη στα containers.

6) ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ανάλυση και η διαχείριση του φόρτου εργασίας των datacenter αποτελεί σημαντικό ζήτημα για μία εταιρεία παροχής υπηρεσιών cloud, καθώς μέχρι πρόσφατα ο βαθμός αξιοποίησης των πόρων σε αυτά, ήταν χαμηλός. Με την εργασία αυτή, μελετήθηκε το τελευταίο δημοσιευμένο trace_2018 της εταιρείας Alibaba, ώστε να εντοπιστούν οι ανισορροπίες του workload και να αξιολογηθεί ο βαθμός αξιοποίησης των πόρων με την τεχνική Co-locating που χρησιμοποιεί η εταιρεία. Συγκεκριμένα, για τα 4000 μηχανήματα του trace, μελετήθηκαν τα CPU και Memory utilization καθόλη την διάρκεια των 8 ημερών και προέκυψαν χωρικές και χρονικές ανισορροπίες. Συνεπώς, εγείρεται η ανάγκη για καλύτερο χρονοπρογραμματισμό εργασιών στους πόρους των cluster. Επιπλέον, με την ανάλυση του τρόπου με τον οποίο κατανέμονται οι πόροι στα μηχανήματα, παρατηρήθηκε ότι το workload αυτό της Alibaba είναι memory bound. Πράγμα το οποίο συμβαίνει καθώς η εξάντληση των πόρων μνήμης, εξαιτίας της ζήτησης περισσοτέρων από τους πραγματικά αναγκαίους, αποτρέπει την ανάθεση επιπλέον εργασιών στα μηχανήματα ενώ ταυτόχρονα υπάρχουν διαθέσιμοι πόροι CPU. Επιπρόσθετα, από τις αναλύσεις των απαιτήσεων πόρων, τόσο των online υπηρεσιών όσο και των εργασιών συστήματος προέκυψαν τα εξής συμπεράσματα: α) Οι online εργασίες, με σκοπό την διαφύλαξη των SLA συμφωνιών και της ποιότητας υπηρεσίας (QoS), επιδίδονται στο να αιτούνται πόρους, περισσότερους από αυτούς που πραγματικά θα χρησιμοποιήσουν. β) Οι εργασίες συστήματος έχουν μειωμένη προτεραιότητα και περιορισμένους πόρους κατά τη διάρκεια της ημέρας σε σχέση με τις online εργασίες, και πάλι με σκοπό την διαφύλαξη των SLA συμφωνιών με τον πελάτη και του QoS.

Τέλος, αξίζει να αναφερθεί, ότι από την ανάλυση δεδομένων του trace, όπως τα 'disk_io_percent' στο *machine_usage* και στο *container_usage*, σε συνδυασμό με το μοτίβο ημερήσιας-νυχτερινής δραστηριότητας που παρατηρήθηκε στο workload, θα μπορούσαν μελλοντικά να εξαχθούν συμπεράσματα και για τις garbage collection στρατηγικές που χρησιμοποιούνται από την Alibaba στα κέντρα δεδομένων της. Επίσης, ενδιαφέρουσα μελλοντική έρευνα θα ήταν η ανάλυση των εξαρτήσεων τόσο των batch όσο και των online εργασιών, ώστε να εντοπιστεί το βάθος των γράφων εξαρτήσεων και να προσδιοριστούν οι τρόποι κατανομής τους στους πόρους από τους FUXI και SIGMA.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] M. Armbrust, A. Fox, . R. Grith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia, «A View of Cloud Computing,» *Commununications of the ACM,* vol. 53, No. 4, pp. 50-58, 2010.

[2] R. Buyya, C. Yeo, S. Venugopal, J. Broberg and I. Brandic, «Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th Utility,» *Future Generation Computer Systems,* vol. 25, No. 6, pp. 599-616, 2009.

[3] J. Schad, J. Dittrich and J. A. Quiane-Ruiz, «Runtime Measurements in the Cloud: Observing, Analyzing, and Reducing Variance,» *Proceedings of the VLDB Endowment,* vol. 3, No. 1-2, pp. 460-471, 2010.

[4] C. Delimitrou and C. Kozyrakis, «Quasar: resource-efficient and qos-aware cluster management,» In *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems*, 2014.

[5] S. A. Jyothi, C. Curino, I. Menache, S. M. Narayanamurthy, . A. Tumanov, J. Yaniv, I. Goiri, S. Krishnan, J. Kulkarni and S. Rao, «Morpheus: Towards automated slos for enterprise clusters,» In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, 2016.

[6] K. Rajan, D. Kakadia, C. Curino and S. Krishnan, «Perforator: eloquent performance models for resource optimization,» In *Proceedings of the 7th ACM Symposium on Cloud Computing*, 2016.

[7] Y. Yan, Y. Gao, Y. Chen, Z. Guo, B. Chen and T. Moscibroda, «Tr-spark: Transient computing for big data analytics,» In *Proceedings of the 7th ACM Symposium on Cloud Computing*, 2016.

[8] W. Chen, J. Rao and X. Zhou, «Preemptive, low latency datacenter scheduling via lightweight virtualization,» In *USENIX Annual Technical Conference (USENIX ATC 17)*, 2017.

[9] C. Lu, K. Ye, G. Xu, C. Xu and T. Bai, «Imbalance in the cloud: An analysis on Alibaba cluster trace,» *In IEEE International Conference on Big Data (Big Data)*, 2017.

[10] B. Sotomayor, K. Keahey, I. Foster and T. Freeman, «nabling cost-effective resource leases with virtual machines,» In *ACM/IEEE International Symposium on High*, 2007.

[11] L. Cherkasova, D. Gupta and A. Vahdat, «When virtual is harder than real: Resource allocation challenges in virtual machine based it environments,» In *Technical Report HPL 2007-25*, 2007.

[12] R. Kaur and S. Kinger, «Enhanced Genetic Algorithm based Task Scheduling in Cloud Computing,» *International Journal of Computer Applications,* τόμ. 101, 2014.

[13] J. W. Ge and Y. S. Yuan, «Research of cloud computing task scheduling algorithm based on improved genetic algorithm,» *Applied Mechanics and Materials,* pp. 2426-2429, 2013.

[14] «Coarse-grained Workload Categorization in Virtual Environments using the Dempster-Shafer Fusion,» In *2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design*, 2015.

[15] A. Bakshi and Y. B. Dujodwala, «Securing cloud from DDoS Attacks using Intrusion Detection System in Virtual Machine,» In *ICCSN '10 Proceeding of the 2010 Second Internationa Conference on Communication Software and networks, IEEE Computer Society*, 2010.

[16] S. Singh, «Metrics based Workload Analysis Technique for IaaS Cloud, Inderveer Chana,» In *International Conference on Next Generation Computing and Communication Technologies*, 2014.

[17] Google, «[n. d.]. Google cluster workload traces,» [online]. Available: h.ps://github.com/google/cluster-data..

[18] C. Reiss, A. Tumanov, G. R. Ganger, . R. H. Katz and M. A. Kozuch, «Heterogeneity and dynamicity of clouds at scale: Google trace analysis.,» *In .e .ird ACM Symposium on Cloud Computing*, 2012.

[19] Q. Zhang, J. L. Hellerstein and R. Boutaba, «Characterizing task usage shapes in google compute clusters,» In *In Large Scale Distributed Systems and Middleware Workshop*, 2011.

[20] Z. Liu and S. Cho, «Characterizing machines and workloads on a google cluster,» In *41st International Conference on Parallel Processing Workshops*, 2012.

[21] S. Di, D. Kondo and W. Cirne, «Characterization and comparison of cloud versus grid workloads,» In *IEEE International Conference on Cluster Computing*, 2012.

[22] Y. Cheng, Z. Chai and A. Anwar, «Characterizing Co-located Datacenter Workloads: An Alibaba Case Study.,» [online]. Available: h.ps://arxiv.org/abs/1808.02919.

[23] Q. Liu and Z. Yu, «e Elasticity and Plasticity in Semi-Containerized Colocating Cloud Workload: a View from Alibaba Trace,» In *Proceedings of ACM Symposium on Cloud Computing*, 2018.

[24] X. Fu, R. Ren, J. Zhan, W. Zhou, Z. Jia and G. Lu, «LogMaster: Mining Event Correlations in Logs of Large-Scale Cluster Systems,» In *IEEE 31st Symposium on Reliable Distributed Systems*, 2012.

[25] Y. Watanabe, H. Otsuka, . M. Sonoda, S. Kikuchi and Y. Matsumoto, «Online failure prediction in cloud datacenters by real-timemessage pa.ern learning,» *In International Conference on Cloud Computing Technology and Science.*, 2012.

[26] E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura and R. Bianchini, «Resource Central: Understanding and Predicting Workloads for Improved Resource Management in Large Cloud Platforms,» In *SOSP '17 Proceedings of the 26th Symposium on Operating Systems Principles*, Shanghai, 2017.

[27] R. Wolski and J. Brevik, «Using Parametric Models to Represent Private Cloud Workloads,» *UCSB Computer Science Technical Report No. 2013-05*, 2013.

[28] B. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan and R. Sears, «Benchmarking Cloud Serving Systems with YCSB,» *Yahoo Research*, CA, 2010.

[29] G. Amvrosiadis, J. Park, G. R. Ganger, G. A. Gibson, E. Baseman and N. D. Bardeleben, «On the diversity of cluster workloads and its impact on research results,» *In 2018 {USENIX} Annual Technical Conference*, 2018.