ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Διαχείριση Γνώσης

Dynamo: Amazon's Highly Available Key-value Store

Ερευνητική Εργασία

Εμμανουήλ Παπαδημητρίου
ΑΜ: mcse19021

Επιβλέποντες: Χ. Σκουρλάς Καθηγητής ΠΑΔΑ
Α. Μαρινάγη Καθηγήτρια Γεωπονικού Πανεπιστημίου

Αθήνα, Ιανουάριος 2020

Συζήτηση του άρθρου «Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall and Werner Vogels (2007) Dynamo: Amazon's Highly Available Key-value Store»

Ιστότοπος: https://www.allthingsdistributed.com/files/amazon-dynamo-sosp2007.pdf

Σύνοψη (summary)
Στο άρθρο με τίτλο «Dynamo: Amazon's Highly Available Key-value Store» (2007) οι DeCandia, Hastorun, Jampani, Kakulapati, Lakshman, Pilchin, Sivasubramanian, Vosshall και Vogels τονίζουν τη σημασία της αξιοπιστίας σε τεράστια κλίμακα και παρουσιάζουν τον σχεδιασμό και την υλοποίηση της Dynamo, ένα διαθέσιμο σύστημα που αποτελείται από ένα σύστημα αποθήκευσης κλειδιά-αξίες (key-value). Για να μπορεί να έχει μεγάλα ποσοστά διαθεσιμότητας η Dynamo, θυσιάζει τη συνέπεια κάτω από ορισμένα σενάρια αποτυχίας. Κύρια πτυχή είναι η απουσία σχεσιακού μοντέλου, καθώς παρατηρήθηκε ότι ένα σημαντικό μέρος των υπηρεσιών της Amazon μπορεί να λειτουργήσει ένα απλό query. Η Dynamo έχει στόχο τις εφαρμογές που χρειάζεται να αποθηκεύσουν αντικείμενα που είναι σχετικά μικρά (λιγότερο από 1 MB). Κύριο χαρακτηριστικό είναι, ότι προτεραιότητα έχει η διαθεσιμότητα της βάσης και σε δεύτερο χρόνο η σταθερότητα της. Ακόμη, σημαντικό στοιχείο της υλοποίησης αυτής είναι η δυνατότητα απεριόριστης κλιμάκωσης. Το παρόν άρθρο, είχε μεγάλη επιρροή και ενέπνευσε την δημιουργία πολλών NoSQL βάσεων δεδομένων όπως Apache Cassandra (που αρχικά δημιουργήθηκε από το Facebook) και βάσεις δεδομένων της Amazon όπως SimpleDB και DynamoDB.

1. Εισαγωγή

Οι DeCandia, Hastorun, Jampani, Kakulapati, Lakshman, Pilchin, Sivasubramanian, Vosshall και Vogels εντοπίζουν το πρόβλημα στην έλλειψη ενός κατανεμημένου συστήματος δεδομένων, που θα έχει ως προτεραιότητα την διαθεσιμότητα έναντι της συνέπειας, να επιτρέπει για έλεγχο application-specific της απόδοσης, της ανθεκτικότητας και της σταθερότητας και θα μπορεί να λειτουργήσει σε κλίμακα Amazon (εκατομμύρια checkout, χιλιάδες ταυτόχρονους χρήστες). Οι κύριες ιδέες της υλοποίησης είναι το data store να είναι πάντα εγγράψιμο, δηλαδή ένα data store πάντα διαθέσιμο, να είναι πάντα σταδιακά κλιμακωτή, να υπάρχει ισότητα μεταξύ των κόμβων (συμμετρία), να έχουν δηλαδή το ίδιο σύνολο ευθυνών και επίλυση διενέξεων σε επίπεδο εφαρμογής (application-level conflict resolution). Η αρχιτεκτονική του συστήματος της Dynamo που αποφάσισαν να υλοποιήσουν, αποτελείται από κλιμάκωση (scaling), διαμέριση (partitioning), αναπαραγωγή (replication), versioning, membership και χειρισμός σφαλμάτων. Η Dynamo είναι ένα αποκεντρωμένο (decentralized) σύστημα με ελάχιστη ανάγκη για χειροκίνητη διαχείριση. Αποφάσισαν ότι είναι αναγκαίος ο ορισμός κάποιων ελάχιστων απαιτήσεων για το σύστημα αποθήκευσης, οι απαιτήσεις είναι: query μοντέλο που περιέχει απλές read και write λειτουργίες σε ένα στοιχείο δεδομένων που αναγνωρίζεται με ένα κλειδί, ACID(Atomicity, Consistency, Isolation, Durability) ιδιότητες που εγγυώνται ότι οι συναλλαγές σε αυτή τη βάση δεδομένων πραγματοποιούνται με αξιοπιστία, αποτελεσματικότητα(efficiency), λοιπές απαιτήσεις όπως ότι η Dynamo θα χρησιμοποιηθεί μόνο από το εσωτερικό σύστημα της Amazon το οποίο είναι ασφαλές και δεν έχει απαιτήσεις για αυθεντικοποίηση και εξουσιοδότηση.

2. Βασικές Έννοιες

2.1 Διαθεσιμότητας έναντι Σταθερότητας

Έχει τονιστεί στο άρθρο πολλές φορές, ότι ο σημαντικός στόχος που έχει η υλοποίηση της Dynamo, είναι να πετύχει μεγάλο ποσοστό διαθεσιμότητας. Προτιμάται δηλαδή, να υπάρχει μικρότερη σταθερότητα (consistency). Τονίζεται ότι μια δυνατή σταθερότητα (strong consistency) δεν είναι το ζητούμενο σε όλες τις περιπτώσεις, και κάποιες φορές μπορούμε να συμβιβαστούμε με τελική συνέπεια (eventual consistency) που σημαίνει ότι όλοι οι χρήστες θα δουν τα ίδια δεδομένα. Η Dynamo χρησιμοποιεί απλές πολιτικές όπως “last write win” για την επίλυση των conflicts. Ακόμα, η υλοποίηση αυτή εγγυάται την επιστροφή των αποτελεσμάτων έγκαιρα με την τεχνική Service Level Agreement (SLA). Γίνεται αναφορά σε συστήματα που είναι ευάλωτα σε αποτυχίες διακομιστή και δικτύου και η αύξηση της διαθεσιμότητας σε αυτή τη περίπτωση, γίνεται με την αύξηση χρήσης τεχνικών αναπαραγωγής.

2.2 Απεριόριστη Κλιμάκωση

Οι ερευνητές έχουν καταλήξει στο συμπέρασμα ότι είναι αναγκαία η ύπαρξη της απεριόριστης κλιμάκωσης τους συστήματος χωρίς να υπάρξουν αρνητικές επιπτώσεις στις επιδόσεις του συστήματος. Αυτή η πτυχή είναι αποτέλεσμα της χαλάρωσης των σχεσιακών και σταθερών περιορισμών από τις προηγούμενες βάσεις δεδομένων. Τονίζουν ότι, κατά την κλιμάκωση του συστήματος μπορούν είτε να κλιμακώσουν κάθετα (με μεγαλύτερο server με περισσότερους CPU και RAM) ή να κλιμακώσουν οριζόντια διαιρώντας τα δεδομένα σε πολλαπλές μηχανές. Η οριζόντια κλιμάκωση είναι πιο φθηνή αλλά πιο δύσκολη στην υλοποίηση. Στην Dynamo γίνεται προτίμηση στην οριζόντια κλιμάκωση, με την μείωση των περιορισμών που απαιτούνται. Χρησιμοποιείται συνεκτικός κατακερματισμός για την εξάπλωση των δεδομένων σε έναν αριθμό κόμβων. Η Dynamo υποστηρίζει τη βαθμιαία κλιμάκωση του συστήματος, που είναι σε θέση να κλιμακώσει έναν κόμβο κάθε φορά. Επιπλέον, όλοι οι κόμβοι είναι συμμετρικοί με την έννοια ότι έχουν το ίδιο σύνολο ευθυνών.

3. Υλοποίηση και Συμπεράσματα

Οι ερευνητές έχουν υλοποιήσει τη μελέτη τους και έχουν εξάγει συμπεράσματα και γραφήματα, όπως βλέπουμε στις σελίδες (213-214-215-216-217).

Οι ερευνητές, τονίζουν ότι μετά από τις μελέτες των αποτελεσμάτων των υλοποιήσεων τους, έχουν το επιθυμητό αποτέλεσμα. Ένα σύστημα με μεγάλη διαθεσιμότητα και κλιμακωτό data store. Η Dynamo είναι σταδιακά κλιμακωτή και επιτρέπει στους service owners να κάνουν κλιμάκωση προς τα πάνω ή προς τα κάτω ανάλογα με το τρέχον φορτίο αιτήσεων τους (request load). Επιτρέπει ακόμα στους διαχειριστές, να προσαρμόσουν το σύστημα αποθήκευσης για να έχουν τις επιθυμητές επιδόσεις.

4. Συζήτηση του άρθρου στο πλαίσιο της Διαχείρισης Γνώσης και του μαθήματος

Τα αποτελέσματα είναι χρήσιμα για όλες τις εταιρείες και όλους τους διαχειριστές βάσεων δεδομένων που επιθυμούν την γνώση να είναι υψηλά διαθέσιμη. Αρχικά παρατηρούμε ότι έχει γίνει αποτύπωση γνώσης με δέντρα αποφάσεων (decision trees), τα οποία παρουσιάζουν τον τρόπο με τον οποίο γίνεται το data versioning της διαθέσιμης γνώσης. Ακόμη μία περίπτωση που γίνεται χρήση των δέντρων αποφάσεων είναι στην παρουσίαση της αρχιτεκτονικής που ακολουθεί η Amazon και από εκεί γίνεται εξαγωγή της γνώσης για την κατανόηση των απαιτήσεων που πρέπει να ικανοποιεί η Dynamo υλοποίηση. Επιπλέον, οι ερευνητές κάνουν χρήση διαγραμματικών τεχνικών για την εξαγωγή συμπερασμάτων, γνώσης και αποτελεσμάτων από την υλοποίηση τους. Σε αυτά τα διαγράμματα υπάρχει επεξήγηση της λειτουργίας των κόμβων (nodes) που είναι κομμάτι της κλιμάκωσης του συστήματος. Επιπροσθέτως, παρουσιάζεται το σύνολο της απόδοσης των λειτουργιών του διαβάσματος και εγγραφής του συστήματος, σύγκριση μεταξύ buffered και non-buffered εγγραφών, καμπύλης συμπεριφοράς των κόμβων που δεν έχουν την ζητούμενη συμπεριφορά (out-of-balance). Ακόμη, παρουσιάζονται οι στρατηγικές τμηματοποίησης (partitioning) των κλειδιών που γίνεται σε τρία στάδια και σε ένα ακόμη διάγραμμα γίνεται η σύγκριση της αποδοτικότητας των σταδίων τμηματοποίησης.

Problem	Technique	Advantage
Partitioning	Consistent Hashing	Incremental Scalability
High Availability for writes	Vector clocks with reconciliation during reads	Version size is decoupled from update rates.
Handling temporary failures	Sloppy Quorum and hinted handoff	Provides high availability and durability guarantee when some of the replicas are not available.
Recovering from permanent failures	Anti-entropy using Merkle trees	Synchronizes divergent replicas in the background.
Membership and failure detection	Gossip-based membership protocol and failure detection.	Preserves symmetry and avoids having a centralized registry for storing membership and node liveness information.

Πίνακας 1. Σύνοψη των τεχνικών που χρησιμοποιούνται από την Dynamo