



Πανεπιστήμιο Πελοποννήσου
Σχολή Οικονομίας και Τεχνολογίας
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Γ. Κ. Βλάχου, 22100, Τρίπολη

Διαχείριση και αποθήκευση χρονοσειρών

Ευάγγελος Καρυστινός

Μεταπτυχιακή διατριβή

Τρίπολη, Δεκέμβριος 2025

Διαχείριση και αποθήκευση χρονοσειρών.

Ευάγγελος Καρυστινός.

Μεταπτυχιακή διατριβή.

Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Πελοποννήσου

Copyright © Δεκέμβριος 2025. All Rights Reserved.



Πανεπιστήμιο Πελοποννήσου
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Γ. Κ. Βλάχου, 22100, Τρίπολη
Τηλ. 2710-230177
Φαξ. 2710-372160

Διαχείριση και αποθήκευση χρονοσειρών

Ευάγγελος Καρυστινός

Μεταπτυχιακή διατριβή

Επιβλέπων:

Σπύρος Σκιαδόπουλος, Καθηγητής, Πανεπιστήμιο Πελοποννήσου

Επιτροπή Αξιολόγησης:

Ευχαριστίες

Η ολοκλήρωση αυτής της εργασίας κλείνει τον κύκλο των μεταπτυχιακών μου σπουδών και σηματοδοτεί και την ολοκλήρωση της μεταπτυχιακής έρευνάς μου. Αυτή η έρευνα δε θα ήταν δυνατό να ολοκληρωθεί χωρίς την υποστήριξη του καθηγητή Σπύρου Σκιαδόπουλου, που με καθοδήγησε στην προσπάθειά μου.

Θα ήθελα επίσης να εκφράσω τις ιδιαίτερες ευχαριστίες στην οικογένειά μου για την αγάπη και την ενθάρρυνσή τους σε όλες μου τις προσπάθειες. Το θάρρος τους στη ζωή και η δύναμή τους στις δυσκολίες ήταν ένα επιπλέον κίνητρο για αυτή την προσπάθεια.

Το έργο μου αφιερώνω στην ιερή μνήμη του πατέρα μου *Νικολάου Καρυστινού*

Περίληψη

Σκοπός της εργασίας είναι η διερεύνηση και ανάπτυξη μεθόδων διαχείρισης και αποθήκευσης χρονοσειρών, με στόχο τη βελτιστοποίηση της απόδοσης και τη μείωση του χρόνου λήψης αποφάσεων σε κρίσιμους τομείς, όπως οι πλατφόρμες διαχείρισης χρηματοοικονομικών, η ενέργεια, η υγεία, και η εφοδιαστική αλυσίδα. Η εργασία εξετάζει τις πιο προηγμένες τεχνικές αποθήκευσης και ανάλυσης χρονοσειρών, προσφέροντας προτάσεις για την αποτελεσματική διαχείριση μεγάλων ποσοτήτων δεδομένων και τη βελτίωση της ακρίβειας και της αποτελεσματικότητας στις σχετικές εφαρμογές.

Περιεχόμενα

1	Εισαγωγή στην ανάλυση χρονοσειρών	1
1.1	Ορισμός και σκοπός της ανάλυσης χρονοσειρών	2
1.2	Σημασία της ανάλυσης χρονοσειρών	2
1.3	Ιστορική αναδρομή	4
1.4	Ανάγκες που καλούνται να καλύψουν οι χρονοσειρές	5
1.5	Δομή διατριβής	6
2	Δεδομένα χρονοσειρών	9
2.1	Ορισμός	9
2.2	Βασικές έννοιες και όροι	13
2.3	Αλγόριθμοι ανάλυσης χρονοσειρών	14
2.4	Χαρακτηριστικά των δεδομένων χρονοσειρών	17
2.5	Τομείς εφαρμογής δεδομένων χρονοσειρών	18
2.6	Τεχνικές κανονικοποίησης και κλιμάκωσης	19
2.7	Μέθοδοι αποσύνθεσης δεδομένων	20
2.7.1	Συστατικά χρονοσειρών και η σημασία τους	21
2.8	Μέθοδοι ανάλυσης και πρόγνωσης	22
2.9	Προσεγγίσεις πρόβλεψης	24
2.10	Εργαλεία οπτικοποίησης	25
2.11	Σύνοψη κεφαλαίου	26
3	Αποθήκευση χρονοσειρών	27
3.1	Γενικά συστήματα	27

3.1.1	Σχεσιακά συστήματα	27
3.1.1.1	PostgreSQL	28
3.1.1.2	Microsoft SQL Server	29
3.1.1.3	MySQL	30
3.1.2	Μη σχεσιακές βάσεις δεδομένων	31
3.1.2.1	Cassandra	32
3.1.2.2	Elasticsearch	34
3.1.2.3	OpenTSDB	36
3.2	Εξειδικευμένα συστήματα	39
3.2.1	InfluxDB	39
3.2.1.1	Βασικά σημεία	40
3.2.1.2	Κατακερματισμός δεδομένων	41
3.2.1.3	Μηχανή αποθήκευσης	42
3.2.1.4	Περιπτώσεις χρήσης του InfluxDB	42
3.2.1.5	Πλεονεκτήματα και μειονεκτήματα του InfluxDB	44
3.2.2	TimescaleDB	47
3.2.2.1	Κατανεμημένη αρχιτεκτονική	47
3.2.2.2	Περιπτώσεις χρήσης TimescaleDB	48
3.2.2.3	Πλεονεκτήματα και μειονεκτήματα	48
3.3	Σύγκριση υλοποιήσεων	49
3.3.1	InfluxDB και παραδοσιακές βάσεις δεδομένων	50
3.3.2	Συγκριτική ανάλυση InfluxDB και TimescaleDB	50
3.3.2.1	Σχεδίαση βάσης δεδομένων	51
3.3.2.2	Απόδοση και κλιμακωσιμότητα	51
3.3.2.3	Συμβατότητα και οικοσύστημα	52
3.3.2.4	Διαχείριση και χρήση	52
3.4	Σύνοψη	53
4	Δεδομένα χρονοσειρών που αξιοποιούνται	55
4.1	Κριτήρια επιλογής σταθμών βάσει απαιτήσεων	55

4.2	Δεδομένα μετεωρολογικών σταθμών	58
4.2.1	Αγροτικοί σταθμοί	58
4.2.1.1	Δορυφορικές μονάδες	61
4.2.1.2	Τοποθεσία σταθμών	62
4.2.2	Αστικοί σταθμοί	66
4.2.2.1	Δορυφορικές μονάδες	68
4.2.2.1.1	Μονάδα 68	68
4.2.2.1.2	Μονάδα 69	68
4.2.2.1.3	Μονάδα 70	69
4.2.2.2	Τοποθεσία σταθμών	69
4.3	Σύνοψη κεφαλαίου	70
5	Διαχείριση δεδομένων στην InfluxDB και TimescaleDB	73
5.1	Προετοιμασία της InfluxDB	74
5.1.1	Γενική αρχιτεκτονική	74
5.1.2	Αντιστοίχιση των δεδομένων του Κεφαλαίου 4 σε τύπους της InfluxDB	75
5.1.3	Δημιουργία βάσης και πολιτικών διατήρησης	75
5.1.4	Έλεγχος διεπαφής HTTP API	76
5.1.5	Ανάκτηση και ανάλυση	77
5.2	Προετοιμασία της TimescaleDB	78
5.2.1	Γενική αρχιτεκτονική	79
5.2.2	Αντιστοίχιση δεδομένων σε οντότητες TimescaleDB	79
5.2.3	Δημιουργία σχήματος και υπερπινάκων	80
5.2.4	Ερωτήματα και ανάλυση	82
5.2.5	Πολιτικές συμπίεσης και σύνοψης	83
5.3	Εισαγωγή δεδομένων και αρχική παραμετροποίηση	84
5.3.1	Αρχική παραμετροποίηση	84
5.3.2	Δημιουργία κλάσεων	87
5.3.3	Δημιουργία κλάσης για τη InfluxDB	89

5.3.4	Δημιουργία κλάσης για την TimescaleDB	91
5.3.5	Εισαγωγή δεδομένων	92
5.4	Ανάκτηση δεδομένων και οπτικοποίηση	95
5.4.1	Ανάκτηση δεδομένων από την InfluxDB	95
5.4.2	Οπτικοποίηση δεδομένων μέσω Grafana	96
5.4.3	Ανάκτηση δεδομένων από την TimescaleDB για οπτικοποίηση . .	98
5.5	Σύνοψη κεφαλαίου	99
6	Σύνοψη και συμπεράσματα	101
6.1	Σύνοψη της διατριβής	101
6.2	Κύρια αποτελέσματα και συνεισφορά	103
6.3	Περιορισμοί της εργασίας	104
6.4	Μελλοντικές κατευθύνσεις	104
A'	Απαιτήσεις και προδιαγραφές σταθμών	107
A'.1	Αγρομετεωρολογικός σταθμός	107
A'.2	Σταθμός μέτρησης ατμοσφαιρικών ρύπων και μικροκλίματος	109
	Βιβλιογραφία	113

Λίστα εικόνων

2.1	Ο χρόνος ως βασικός συντελεστής για τα δεδομένα	10
2.2	Κανονικές χρονοσειρές και χρονοσειρές συμβάντων	11
2.3	Αλγόριθμοι ARIMA και SARIMA	14
2.4	Αλγόριθμοι ARIMA και Exponential Smoothing	15
2.5	Αλγόριθμοι ARIMA και PROPHET. Η μπλε γραμμή αποτυπώνει τη μηνιαία πρόβλεψη, η πορτοκαλί την πρόβλεψη του ARIMA και η πράσινη την πρόβλεψη PROPHET	16
2.6	Αλγόριθμοι ARIMA και LSTM	17
2.7	Διαφορές μεθόδων πρόγνωσης χρονοσειρών	23
3.1	Αρχιτεκτονική του OpenTSDB	37
3.2	Αρχιτεκτονική δομή InfluxDB	40
4.1	Γεωγραφική αποτύπωση όλων των αγροτικών σταθμών.	59
4.2	Γεωγραφική αποτύπωση του σταθμού 2736.	62
4.3	Γεωγραφική αποτύπωση του σταθμού 2737.	62
4.4	Γεωγραφική αποτύπωση του σταθμού 2738.	64
4.5	Γεωγραφική αποτύπωση του σταθμού 2741.	64
4.6	Γεωγραφική αποτύπωση του σταθμού 2742.	64
4.7	Γεωγραφική αποτύπωση του σταθμού 2743.	65
4.8	Γεωγραφική αποτύπωση του σταθμού 2744.	65
4.9	Γεωγραφική αποτύπωση του σταθμού 142217.	65
4.10	Γεωγραφική αποτύπωση του σταθμού 2873 (Επιμελητήριο Αρκαδίας). . .	69
4.11	Γεωγραφική αποτύπωση του σταθμού 2874 (Πανεπιστήμιο Πελοποννήσου). .	70

5.1	Ορισμός βάσης και πολιτικής διατήρησης	76
5.2	Αποστολή μιας συγκεκριμένης μέτρησης ενός αισθητήρα στην InfluxDB	77
5.3	Αποστολή μιας συγκεκριμένης μέτρησης ενός αισθητήρα στην InfluxDB	78
5.4	Δημιουργία πινάκων και υπερπινάκων στη TimescaleDB.	82
5.5	Υπολογισμός μέσου όρο τιμών.	83
5.6	Προσθήκη πολιτικής συμπίεσης δεδομένων.	84
5.7	Διάγραμμα κλάσεων του υποσυστήματος αποθήκευσης σε UML.	88
5.8	Παράδειγμα οπτικοποίησης δεδομένων InfluxDB στο Grafana.	97

Κεφάλαιο 1

Εισαγωγή στην ανάλυση χρονοσειρών

Η ανάλυση και διαχείριση δεδομένων χρονοσειρών αποτελεί σήμερα έναν από τους πλέον κρίσιμους τομείς της στατιστικής, της επιστήμης δεδομένων και των πληροφοριακών συστημάτων, καθώς πλήθος σύγχρονων εφαρμογών βασίζονται στη συνεχή συλλογή, αποθήκευση και επεξεργασία δεδομένων που εξελίσσονται στον χρόνο. Η ραγδαία ανάπτυξη αισθητήρων, υποδομών παρακολούθησης και αυτοματοποιημένων διαδικασιών έχει οδηγήσει στη δημιουργία μεγάλων όγκων χρονικά εξαρτώμενων δεδομένων, καθιστώντας απαραίτητη τη χρήση αποδοτικών μεθόδων αποθήκευσης, ανάλυσης και ανάκτησης πληροφορίας.

Στο παρόν κεφάλαιο παρουσιάζονται οι βασικές έννοιες και ο σκοπός της ανάλυσης χρονοσειρών, η σημασία τους σε σύγχρονους επιστημονικούς και επιχειρησιακούς τομείς, καθώς και μια συνοπτική ιστορική αναδρομή της εξέλιξης των σχετικών μεθόδων. Παράλληλα, αναλύονται οι ανάγκες που καλούνται να καλύψουν τα δεδομένα χρονοσειρών, τόσο σε επίπεδο ανάλυσης και πρόβλεψης όσο και σε επίπεδο αποθήκευσης και διαχείρισης μεγάλων όγκων δεδομένων. Το κεφάλαιο αυτό θέτει το απαραίτητο θεωρητικό υπόβαθρο για τα επόμενα κεφάλαια της διατριβής, όπου εξετάζονται σύγχρονες τεχνολογίες αποθήκευσης χρονοσειρών και παρουσιάζονται πρακτικές υλοποιήσεις βασισμένες σε εξειδικευμένα συστήματα βάσεων δεδομένων.

1.1 Ορισμός και σκοπός της ανάλυσης χρονοσειρών

Η ανάλυση χρονοσειρών [?] είναι ένας στατιστικός τομέας που ασχολείται με τη μελέτη των δεδομένων που παράγονται κατά διαστήματα χρόνου. Μια χρονοσειρά είναι μια συλλογή παρατηρήσεων ή μετρήσεων που καταγράφονται σε συγκεκριμένα χρονικά σημεία και συνήθως παρουσιάζουν κάποιου είδους χρονική εξέλιξη.

Ο σκοπός της ανάλυσης χρονοσειρών είναι να κατανοήσει και να προβλέψει τη μελλοντική συμπεριφορά μιας χρονοσειράς με βάση τα δεδομένα που έχουν συλλεχθεί στο παρελθόν. Αυτό μπορεί να γίνει μέσω της ανάλυσης δομικών χαρακτηριστικών της χρονοσειράς, της εξαγωγής συνόλων δεδομένων που περιέχουν πληροφορίες για την τάση, την εποχιακότητα, την αυτοσυσχέτιση και τις τυχαίες διακυμάνσεις, καθώς και μέσω της εφαρμογής μοντέλων πρόβλεψης. Οι αναλύσεις χρονοσειρών εφαρμόζονται σε πολλούς τομείς, όπως οι οικονομικές προβλέψεις, η αποθήκευση δεδομένων, η κλιματολογία, η ιατρική και η τηλεπικοινωνία. Οι μέθοδοι ανάλυσης χρονοσειρών περιλαμβάνουν την απλή περιγραφή και οπτικοποίηση της χρονοσειράς, την ανίχνευση των μοτίβων και των τάσεων, την πρόβλεψη μελλοντικών τιμών και την αξιολόγηση της ακρίβειας των προβλέψεων.

Συνοψίζοντας, ο σκοπός της ανάλυσης χρονοσειρών είναι να παρέχει μια μέθοδο για την κατανόηση και πρόβλεψη των μελλοντικών συμπεριφορών μιας χρονοσειράς με βάση τα παρελθοντικά δεδομένα.

1.2 Σημασία της ανάλυσης χρονοσειρών

Η ανάλυση χρονοσειρών [?] έχει σημασία σε πολλούς τομείς και μπορεί να παράσχει πολύτιμες πληροφορίες και εφαρμογές. Ας παρουσιάσουμε μερικούς από τους κύριους τομείς όπου η ανάλυση χρονοσειρών είναι σημαντική.

Οικονομία και χρηματοοικονομία. Η ανάλυση χρονοσειρών χρησιμοποιείται για την πρόβλεψη οικονομικών μεταβλητών όπως η ανεργία, ο πληθωρισμός, οι χρηματιστηριακές τιμές και τα επιτόκια που συνοδεύουν τις τιμές. Αυτό βοηθά

στη λήψη επενδυτικών αποφάσεων, χρηματοδότηση και σχεδιασμό οικονομικών πολιτικών.

Ενέργεια. Η ανάλυση χρονοσειρών παίζει κρίσιμο ρόλο στη διαχείριση της ενέργειας, επιτρέποντας την πρόβλεψη της ζήτησης και της προσφοράς ενέργειας σε διάφορες χρονικές κλίμακες. Αυτό βοηθά στη βελτιστοποίηση της παραγωγής και της κατανάλωσης ενέργειας, καθώς και στη λήψη αποφάσεων σχετικά με την αποθήκευση ενέργειας και την ανάπτυξη ανανεώσιμων πηγών ενέργειας. Η ικανότητα πρόβλεψης των ενεργειακών αναγκών συμβάλλει στην αποδοτικότερη λειτουργία των δικτύων και στη μείωση των περιβαλλοντικών επιπτώσεων.

Εφοδιαστική Αλυσίδα. Στον τομέα της εφοδιαστικής αλυσίδας, η ανάλυση χρονοσειρών χρησιμοποιείται για την πρόβλεψη της ζήτησης προϊόντων και την αποτελεσματική διαχείριση των αποθεμάτων. Αυτό περιλαμβάνει τη βελτίωση του προγραμματισμού των παραδόσεων, τη μείωση του χρόνου παράδοσης, και την ελαχιστοποίηση των καθυστερήσεων. Με την αξιοποίηση των χρονοσειρών, οι εταιρείες μπορούν να προβλέψουν με μεγαλύτερη ακρίβεια τις ανάγκες της αγοράς, βελτιώνοντας τη συνολική αποδοτικότητα της εφοδιαστικής αλυσίδας.

Ανάλυση κλίματος. Η ανάλυση χρονοσειρών χρησιμοποιείται για τη μελέτη και πρόβλεψη των μετεωρολογικών μοντέλων και φαινομένων, όπως οι θερμοκρασίες, οι βροχοπτώσεις και οι κλιματικές αλλαγές. Αυτό βοηθά στην κατανόηση του κλίματος και στη λήψη μέτρων για την προστασία του περιβάλλοντος.

Υγεία. Η ανάλυση χρονοσειρών χρησιμοποιείται στην παρακολούθηση και πρόβλεψη των επιδημιών, των νοσηλευτικών περιστατικών και της διασποράς ασθενειών. Αυτό μπορεί να βοηθήσει στη λήψη μέτρων δημόσιας υγείας και τη βελτίωση της πρόνοιας υγείας.

Τηλεπικοινωνίες. Η ανάλυση χρονοσειρών χρησιμοποιείται για την πρόβλεψη και βελτιστοποίηση της χρήσης των δικτύων τηλεπικοινωνιών, όπως η κίνηση δεδομένων, οι τηλεφωνικές κλήσεις και οι ροές κίνησης σε δίκτυα κινητής τηλεφωνίας.

Επιστήμη και μηχανική. Η ανάλυση χρονοσειρών χρησιμοποιείται σε πολλές επιστημονικές και μηχανικές εφαρμογές, όπως η ανίχνευση προτύπων σε σήματα, η πρόβλεψη επιδόσεων συστημάτων, η παρακολούθηση διαδικασιών και η πρόβλεψη βλαβών.

Αυτοί είναι μόνο μερικοί από τους τομείς όπου η ανάλυση χρονοσειρών είναι σημαντική. Συνολικά, η ανάλυση χρονοσειρών μπορεί να παρέχει πολύτιμες πληροφορίες για την κατανόηση των χρονικών δεδομένων και τη λήψη αποφάσεων που βασίζονται στην πρόβλεψη και τη μελλοντική συμπεριφορά.

1.3 Ιστορική αναδρομή

Η ανάλυση χρονοσειρών έχει μια μακρά ιστορία και έχει εξελιχθεί μέσα από διάφορες φάσεις και αναπτύξεις. Ας κάνουμε μια σύντομη ιστορική αναδρομή στην ανάλυση χρονοσειρών:

Πρώιμη περίοδος (προ 1900). Οι πρώτες προσπάθειες για την ανάλυση χρονοσειρών είναι συνδεδεμένες με τη μελέτη των αστρονομικών φαινομένων και των κλιματικών μοτίβων. Στην εποχή αυτή, οι μέθοδοι που χρησιμοποιούνταν ήταν κυρίως γραφικές, όπως γραφήματα και γραμμές τάσης.

Ανάπτυξη των αρχών του 20ού αιώνα. Κατά τη διάρκεια αυτής της περιόδου, οι μέθοδοι ανάλυσης χρονοσειρών εξελίχθηκαν περαιτέρω. Πρωταρχικά, εμφανίστηκαν μέθοδοι που βασίζονταν στην παλινδρόμηση, όπως η μέθοδος των ελαχίστων τετραγώνων. Το 1927, ο οικονομολόγος Ιρβίνγκ Φίσερ εισήγαγε τη μέθοδο της ανάλυσης της τιμής-ποσότητας για την ανάλυση χρονοσειρών των οικονομικών δεδομένων.

Ανάπτυξη της μαθηματικής θεωρίας. Κατά τη διάρκεια της δεκαετίας του 1940, η μαθηματική θεωρία της ανάλυσης χρονοσειρών άρχισε να αναπτύσσεται περαιτέρω. Ο Χάρυ Ράνι, ο Γκεόργκ Μποξ και ο Τζων Τούκι παρουσίασαν το μοντέλο ARMA (Autoregressive Moving Average), το οποίο αποτέλεσε τη βάση για πολλές μετέπειτα μεθόδους ανάλυσης χρονοσειρών.

Εποχή των υπολογιστών και των αυτόματων μεθόδων. Με την εμφάνιση των υπολογιστών, η ανάλυση χρονοσειρών επωφελήθηκε από τις αυτόματες μεθόδους. Εμφανίστηκαν αλγόριθμοι όπως το ARIMA, το οποίο συνδυάζει την αυτοπαλίνδρομη και την κινητή μέση ανάλυση. Επίσης, εμφανίστηκαν μέθοδοι βασισμένες σε νευρωνικά δίκτυα, όπως τα LSTM, που έχουν αποδειχθεί αποτελεσματικές στην ανάλυση και πρόβλεψη χρονοσειρών.

Σήμερα, η ανάλυση χρονοσειρών είναι μια ευρέως χρησιμοποιούμενη μέθοδος σε πολλούς τομείς, όπως η οικονομία, οι χρηματοοικονομικές αγορές, η μετεωρολογία, η δημόσια υγεία, η αποτίμηση ρίσκου και πολλοί άλλοι. Οι σύγχρονες τεχνικές και οι προηγμένοι αλγόριθμοι έχουν επιτρέψει την ανάλυση πιο σύνθετων χρονοσειρών και την πρόβλεψη μεγάλης ακρίβειας.

1.4 Ανάγκες που καλούνται να καλύψουν οι χρονοσειρές

Οι χρονοσειρές καλούνται να καλύψουν μια ποικιλία αναγκών και προκλήσεων σε διάφορους τομείς. Ορισμένες από τις βασικές ανάγκες που οι χρονοσειρές προσπαθούν να αντιμετωπίσουν περιγράφονται παρακάτω. [?]

Πρόβλεψη. Μια από τις κύριες ανάγκες είναι η πρόβλεψη μελλοντικών τιμών ή συμπεριφοράς μιας μεταβλητής βάσει των παρελθοντικών παρατηρήσεων. Αυτό καθίσταται ιδιαίτερα χρήσιμο σε πολλούς τομείς, όπως η οικονομία, οι χρηματοοικονομικές αγορές, η παραγωγή, η διανομή, η εφοδιαστική αλυσίδα, η μετεωρολογία και άλλοι.

Παρακολούθηση και έλεγχος. Οι χρονοσειρές μπορούν να χρησιμοποιηθούν για την παρακολούθηση και τον έλεγχο της απόδοσης ή της ανάπτυξης ενός συστήματος - διαδικασίας. Αυτό είναι συχνά σημαντικό σε εταιρείες και βιομηχανίες, καθώς επιτρέπει την ανίχνευση ανωμαλιών, τη βελτίωση της απόδοσης και τον έλεγχο της ποιότητας.

Ανίχνευση προτύπων και αναλυτική κατανόηση. Οι χρονοσειρές μπορούν να αποκαλύψουν προτεραιότητες, τάσεις, κυκλικότητα, εποχικότητα και άλλα πρότυπα στα δεδομένα. Τα πρότυπα αυτά προκύπτουν από τη μοντελοποίηση των δεδομένων που θα αναλύσουμε παρακάτω. Η ανίχνευση αυτή μπορεί να βοηθήσει στην κατανόηση της δυναμικής ενός φαινομένου και να παράσχει σημαντικές πληροφορίες για τη λήψη αποφάσεων.

Μοντελοποίηση και πρόβλεψη αβεβαιότητας. Οι χρονοσειρές μπορούν να αναπτύξουν μοντέλα που λαμβάνουν υπόψη την αβεβαιότητα και την ποικιλία των δεδομένων. Αυτό μπορεί να είναι χρήσιμο για τη διαχείριση κινδύνων, τον σχεδιασμό στρατηγικών και την πρόβλεψη σε περιβάλλοντα με αβεβαιότητα.

Οι βασικές αυτές ανάγκες είναι οι κύριες που προσπαθούν να καλύψουν οι χρονοσειρές. Οι μέθοδοι ανάλυσης είναι σχεδιασμένες για να ανταποκριθούν σε αυτές τις ανάγκες και να παρέχουν αξιόπιστες προβλέψεις, κατανόηση και λήψη αποφάσεων βασισμένες στην πορεία των χρονοσειρών.

1.5 Δομή διατριβής

Στα επόμενα κεφάλαια, παρουσιάζεται βήμα προς βήμα η θεωρητική και πρακτική προσέγγιση για την κατανόηση, αποθήκευση και ανάλυση δεδομένων χρονοσειρών.

Κεφάλαιο 2. Εξετάζει τα βασικά χαρακτηριστικά των δεδομένων χρονοσειρών, τις βασικές έννοιες και τις μεθόδους κανονικοποίησης, αποσύνθεσης και πρόβλεψης. Παρέχονται, επίσης, εργαλεία και τεχνικές ανάλυσης.

Κεφάλαιο 3. Αναλύει τους τρόπους αποθήκευσης χρονοσειρών, τόσο σε σχέση με τις σχεσιακές όσο και τις μη σχεσιακές βάσεις δεδομένων. Παρουσιάζονται λεπτομερώς τόσο η InfluxDB όσο και η TimescaleDB, καθώς και η αρχιτεκτονική και οι δυνατότητές τους.

Κεφάλαιο 4. Αναφέρεται αναλυτικά στη διαδικασία εγκατάστασης, ρύθμισης και χρήσης της InfluxDB με βάση πρακτικά παραδείγματα. Θίγονται ζητήματα όπως οι

πολιτικές διατήρησης δεδομένων (retention policies), η δημιουργία συνεχών ερωτημάτων (continuous queries) και η χρήση της γλώσσας ερωτημάτων InfluxQL.

Κεφάλαιο 5. Ασχολείται με τη διαχείριση δεδομένων μετεωρολογίας. Παρουσιάζονται οι διαδικασίες αποθήκευσης, συγχρονισμού και αυτοματοποίησης της egress των δεδομένων σε βάσεις όπως η TimescaleDB και η InfluxDB.

Το έργο επιδιώκει να προσφέρει μια συνολική κατανόηση της ανάλυσης δεδομένων χρονοσειρών, καλύπτοντας τόσο τη θεωρητική όσο και την εφαρμοσμένη πλευρά και δίνοντας λύσεις για την υλοποίησή της σε πραγματικές καταστάσεις.

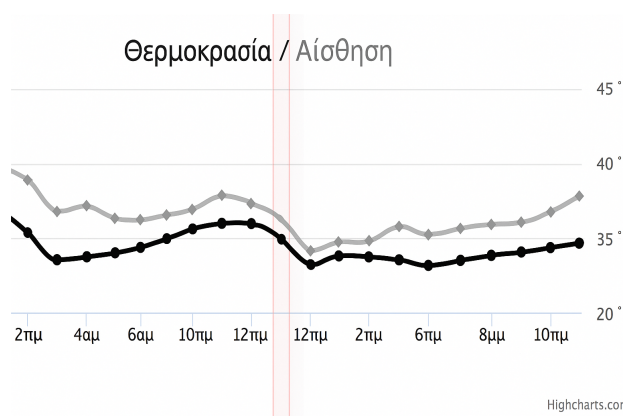
Κεφάλαιο 2

Δεδομένα χρονοσειρών

Η ανάλυση και η κατανόηση δεδομένων χρονοσειρών είναι ένα σημαντικό θέμα που αφορά πολλούς τομείς της επιστήμης και της τεχνολογίας. Στο κεφάλαιο αυτό, γίνεται μια εισαγωγή στις βασικές έννοιες των χρονοσειρών, περιγράφεται η φύση τους, αναλύονται οι κατηγορίες τους και τα χαρακτηριστικά τους, καθώς και παρουσιάζονται οι τρόποι με τους οποίους μπορούμε να τις αναλύσουμε, να τις προβλέψουμε και να τις οπτικοποιήσουμε. Μέσω παραδειγμάτων, θεωρητικής ανάλυσης και αναφοράς σε σύγχρονους αλγόριθμους, αποδεικνύεται η σημασία των χρονοσειρών ως πηγή πληροφοριών για τη λήψη αποφάσεων και την ανάδειξη σημαντικών συμπερασμάτων.

2.1 Ορισμός

Τα δεδομένα χρονοσειρών [17] είναι μία συλλογή παρατηρήσεων μέσω επαναλαμβανόμενων μετρήσεων με την πάροδο του χρόνου. Τέτοια δεδομένα συναντώνται συχνά, καθώς ο χρόνος αποτελεί καίριο στοιχείο σε πολλούς τομείς μέτρησης. Τα τελευταία χρόνια, η ανάπτυξη αισθητήρων και των αντίστοιχων αυτόνομων συστημάτων έχει αυξήσει σημαντικά τη ζήτηση για την καταγραφή και την αποθήκευση τέτοιου είδους δεδομένων. Για παράδειγμα, στο Σχήμα 2.1 απεικονίζονται δύο χρονοσειρές: η θερμοκρασία και η αίσθησή της όπως γίνεται αντιληπτή από τον άνθρωπο, σε χρονική κλίμακα 24 ωρών. Η σύγκριση αυτών των δύο σειρών δεδομένων μας δείχνει πόσο σημαντικό είναι να λαμβάνουμε υπόψη όχι μόνο τις απόλυτες τιμές αλλά και τη χρονική τους δυναμική,



Σχήμα 2.1: Ο χρόνος ως βασικός συντελεστής για τα δεδομένα

όπως φαίνεται και στο Σχήμα 2.1. Τέτοια παραδείγματα χρονοσειρών παρουσιάζονται παρακάτω.

Καταγραφή ημερήσιων εβδομαδιαίων ή μηνιαίων καιρικών δεδομένων.

Καταγραφή αλλαγών στις επιδόσεις εφαρμογών.

Ιατρικές συσκευές που παρουσιάζουν δεδομένα ζωτικής σημασίας σε πραγματικό χρόνο.

Καταγραφή ημερολογίου δικτύου.

Καταγραφή δεδομένων καιρικών συνθηκών.

Καταγραφή οικονομικών δεικτών και μετρήσεων βιομετρικών δεδομένων.

Πέρα από την ανάγκη αποθήκευσης καταγραφών ανά τακτά χρονικά διαστήματα υπάρχουν και οι περιπτώσεις όπου θέλουμε να καταγράψουμε γεγονότα όταν συμβούν. Τέτοια γεγονότα είναι για παράδειγμα τα συμβάντα μητρώου, η επικοινωνία ανάμεσα σε εφαρμογές και λειτουργικά συστήματα και η επεξεργασία μηνυμάτων σε ένα σύστημα τηλεδιάσκεψης.

Οπότε, μπορούμε να πούμε ότι οι χρονοσειρές μπορούν να κατηγοριοποιηθούν στους παρακάτω τύπους.

Μετρήσεις που συγκεντρώνονται ανά τακτά χρονικά διαστήματα.



Σχήμα 2.2: Κανονικές χρονοσειρές και χρονοσειρές συμβάντων

Συμβάντα που συγκεντρώνονται όταν απαιτηθεί ή όταν συμβεί ένα συγκεκριμένο γεγονός.

Στο Σχήμα 2.2 παρουσιάζονται μετρήσεις που καταγράφονται σε σταθερά χρονικά διαστήματα, ενώ στο κάτω γράφημα απεικονίζονται γεγονότα που καταγράφονται μόνο όταν συμβούν, δηλαδή σε μη καθορισμένα χρονικά διαστήματα.

Επιπλέον, οι χρονοσειρές διακρίνονται σε γραμμικές και μη γραμμικές. *Γραμμική χρονοσειρά* λέγεται η συλλογή εκείνη δεδομένων όπου κάθε καταγραφή μπορεί να συσχετιστεί με προηγούμενες ή επόμενες μετρήσεις ή τη διαφορά αυτών, ενώ *μη γραμμικές χρονοσειρές* καλούνται εκείνες που παράγονται από δυναμικά συστήματα, των οποίων οι διεργασίες δεν συσχετίζονται γραμμικά. Σχετικά με τις δύο αυτές κατηγορίες θα πρέπει να τονίσουμε τις παρακάτω ιδιότητες.

- Αν ένα σύστημα δεν ακολουθεί τους κανόνες της γραμμικής μοντελοποίησης τότε μιλάμε για μη γραμμικές χρονοσειρές.
- Οι μη γραμμικές χρονοσειρές μπορεί να περιέχουν τιμές που δεν σχετίζονται μεταξύ τους.

Το σημαντικότερο γνώρισμα των χρονοσειρών σε σύγκριση με άλλα είδη δεδομένων είναι η εξάρτησή τους από το χρόνο: όταν αναζητάμε δεδομένα χρονοσειράς, τα κυριότερα ερωτήματα σχετίζονται με το πότε έγινε η καταγραφή κάθε τιμής.

Ακόμη, ένα ακόμη σημαντικό χαρακτηριστικό είναι ότι τα δεδομένα των χρονοσειρών δεν αλλάζουν, αλλά προστίθενται νέες τιμές αλλά οι παλιές παραμένουν σταθερές. Αυτό τις διαφοροποιεί από τις συμβατικές σχεσιακές βάσεις δεδομένων, όπου οι εγγραφές μπορεί να τροποποιούνται τακτικά.

Αυτή η διαδοχική καταγραφή τιμών, συχνά καθιστά τις χρονοσειρές σειριακά εξαρτημένες, δηλαδή η τιμή σε μια συγκεκριμένη χρονική στιγμή συχνά σχετίζεται με τις τιμές προηγούμενων χρονικών στιγμών. Η σειριακή εξάρτηση αποτελεί βασικό χαρακτηριστικό των χρονοσειρών, ανεξάρτητα από το αν η σχέση αυτή είναι γραμμική ή μη γραμμική.

Επίσης, οι χρονοσειρές μπορεί να απαιτούν ακριβή χρονική επισημείωση (σε επίπεδο δευτερολέπτων ή ακόμη και υποδιαίρέσεων του δευτερολέπτου). Υπάρχουν διάφορες μορφές δεδομένων γνώσεων και δεν είναι πάντα αριθμητικά δεδομένα. Μπορεί να είναι διαφόρων τύπων, όπως `int64`, `float64`, `bool`, ή `string`. Τα παρακάτω χαρακτηριστικά μάς βοηθούν να ξεχωρίσουμε τον τύπο των δεδομένων που έχουμε (χρονοσειρά, εγκάρσια ή πάνελ) και είναι απαραίτητα για να επιλέξουμε τις κατάλληλες μεθόδους ανάλυσης.

1. Όταν διαθέτουμε μόνο το χρονικό στίγμα (timestamp), τότε έχουμε χρονοσειρά, δηλαδή μια ακολουθία μετρήσεων καταγεγραμμένων σε διαφορετικές χρονικές στιγμές.
2. Όταν μελετώνται επιπλέον μεταβλητές που δεν εξαρτώνται έντονα από τον χρόνο, τότε έχουμε εγκάρσια (cross-sectional) δεδομένα, τα οποία χρησιμοποιούνται συνήθως για τη μελέτη διαφορών μεταξύ ατόμων ή ομάδων σε μια συγκεκριμένη χρονική στιγμή.
3. Όταν απαιτείται τόσο χρονικό στίγμα όσο και ταυτοποίηση της οντότητας (π.χ., ένα ID), τότε έχουμε δεδομένα τύπου πάνελ (panel data), τα οποία συνδυάζουν χαρακτηριστικά χρονοσειρών και εγκάρσιων δεδομένων και χρησιμοποιούνται για τη μελέτη της εξέλιξης πολλών οντοτήτων μέσα στον χρόνο.

Όλα τα παραπάνω μπορούν να γίνουν περισσότερο κατανοητά αν εξετάσουμε τον βασικό ορισμό των δεδομένων χρονοσειράς.

2.2 Βασικές έννοιες και όροι

Στην ενότητα αυτή παρουσιάζονται βασικές έννοιες και ορισμοί που σχετίζονται με την ανάλυση χρονοσειρών, όπως η τάση, η εποχικότητα και η σταθερότητα.

Βασικός ορισμός. Δεδομένα χρονοσειράς είναι μια συλλογή μετρήσεων ή παρατηρήσεων που αφορούν μια συγκεκριμένη οντότητα και καταγράφονται σε διαφορετικές χρονικές στιγμές — είτε περιοδικά είτε κατά απαίτηση.

Βασικές έννοιες και όρους που σχετίζονται με την ανάλυση χρονοσειρών:

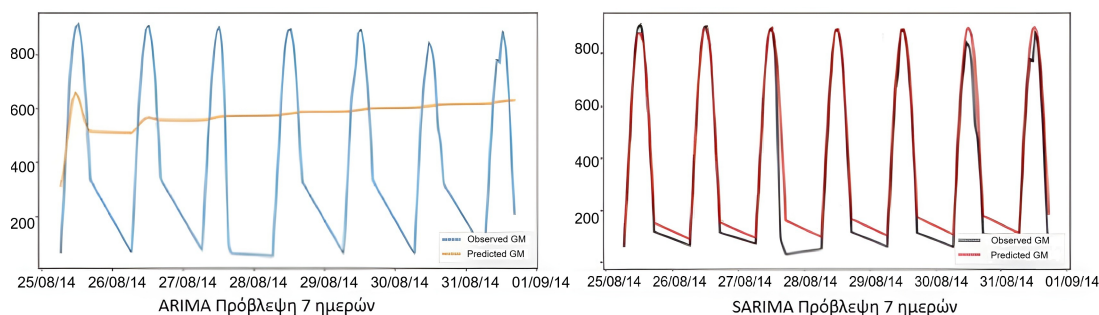
Χρονοσειρά. Σειρά από μετρήσεις που καταγράφονται σε συγκεκριμένα χρονικά σημεία (ημερήσια, μηνιαία, ετήσια, κλπ.).

Τάση. Η τάση αναφέρεται στη μακροπρόθεσμη αύξηση ή μείωση της μέσης τιμής της χρονοσειράς καθώς προχωρά στον χρόνο. Μια ανοδική τάση υποδεικνύει αύξηση των τιμών, ενώ μια καθοδική τάση υποδεικνύει μείωση. Η τάση μπορεί να είναι γραμμική ή μη γραμμική.

Εποχικότητα. Η εποχικότητα αναφέρεται στην ανακατανομή των παρατηρήσεων που επαναλαμβάνονται σε σταθερά χρονικά διαστήματα. Αυτό συμβαίνει συνήθως λόγω εποχιακών παραγόντων, όπως η θερμοκρασία, οι παύσεις μεταξύ μετρήσεων κλπ. Η εποχικότητα μπορεί να επηρεάσει την τάση και να προκαλέσει περιοδικά μοτίβα τιμών στη χρονοσειρά.

Σταθερότητα. Η σταθερότητα αφορά το αν η χρονοσειρά διατηρεί σταθερές στατιστικές ιδιότητες στον χρόνο. Δεν παρουσιάζει συστηματικές αλλαγές στη μέση τιμή, τη διακύμανση ή την αυτοσυσχέτιση με την πάροδο του χρόνου. Η σταθερότητα είναι σημαντική για να μπορέσουμε να εφαρμόσουμε αξιόπιστες μεθόδους ανάλυσης χρονοσειρών.

Η κατανόηση των βασικών όρων είναι σημαντική για την αποτελεσματική ανάλυση και πρόβλεψη των χρονοσειρών.



Σχήμα 2.3: Αλγόριθμοι ARIMA και SARIMA

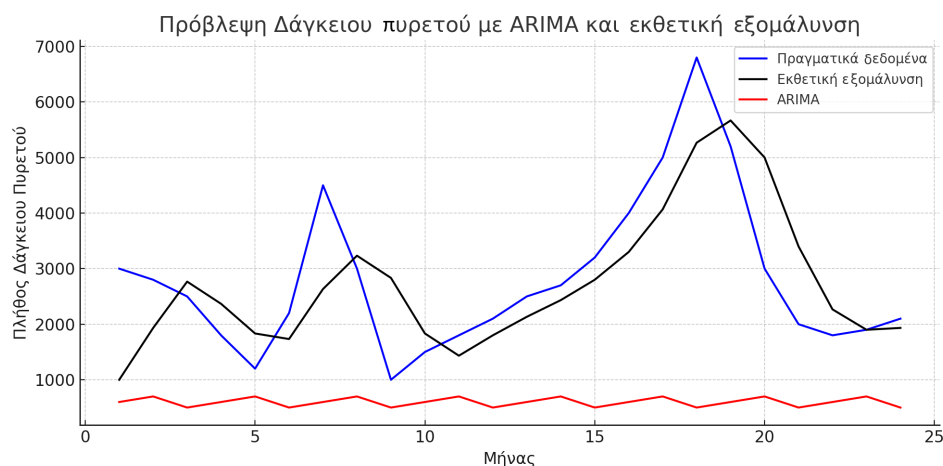
2.3 Αλγόριθμοι ανάλυσης χρονοσειρών

Υπάρχουν πολλοί γνωστοί αλγόριθμοι ανάλυσης χρονοσειρών που χρησιμοποιούνται για την πρόβλεψη, την ανίχνευση προτύπων και την εξαγωγή πληροφοριών από τις χρονοσειρές. Ορισμένοι από αυτούς περιγράφονται παρακάτω.

ARIMA (Autoregressive Integrated Moving Average). Ο αλγόριθμος ARIMA [2] είναι μια μέθοδος πρόβλεψης χρονοσειρών που συνδυάζει την αυτοπαλίνδρομη (autoregressive) και την κινητή μέση (moving average) ανάλυση. Χρησιμοποιείται κυρίως για τη μοντελοποίηση και πρόβλεψη σταθερών χρονοσειρών, ενώ μπορεί επίσης να προσαρμοστεί ώστε να λαμβάνει υπόψη τη μακροπρόθεσμη τάση και την εποχικότητα των δεδομένων.

SARIMA (Seasonal ARIMA). Ο αλγόριθμος SARIMA [2] είναι μια επέκταση του ARIMA που λαμβάνει υπόψη την εποχικότητα σε μια χρονοσειρά. Αυτός ο αλγόριθμος είναι κατάλληλος για την ανάλυση και πρόβλεψη χρονοσειρών που παρουσιάζουν εποχικά μοτίβα.

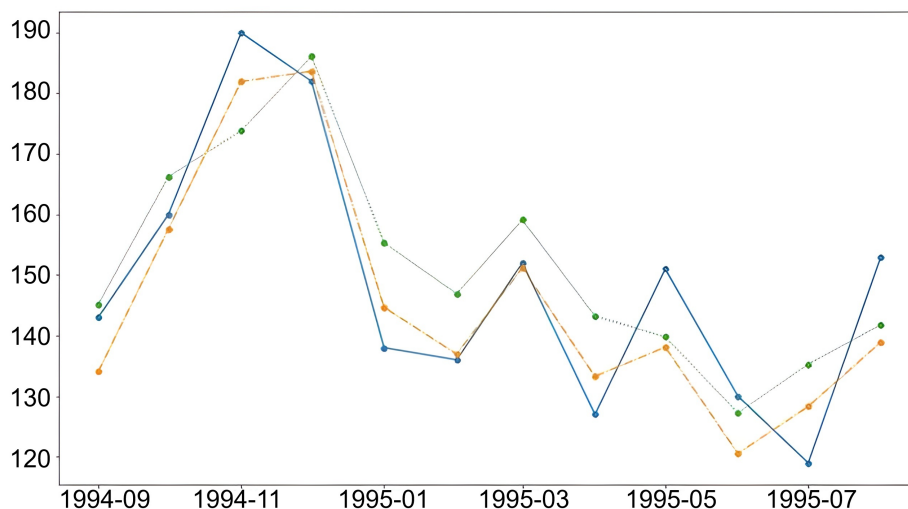
Στο Σχήμα 2.3 απεικονίζεται η σύγκριση δύο μοντέλων πρόβλεψης για χρονοσειρές. Ο αλγόριθμος ARIMA (αριστερά) εφαρμόζεται χωρίς εποχική πληροφορία, με αποτέλεσμα η πρόβλεψή του να αποκλίνει από την επαναλαμβανόμενη μορφή των δεδομένων. Αντίθετα, ο SARIMA (δεξιά) ενσωματώνει εποχικά στοιχεία, προσφέροντας σημαντικά πιο ακριβή πρόβλεψη για χρονοσειρές με περιοδικότητα, όπως φαίνεται από την καλύτερη προσαρμογή του στα κυκλικά μοτίβα της χρονοσειράς.



Σχήμα 2.4: Αλγόριθμοι ARIMA και Exponential Smoothing

Exponential smoothing (εκθετική εξομάλυνση). Η εκθετική εξομάλυνση [6] είναι μια μέθοδος που εφαρμόζεται στην πρόβλεψη χρονοσειρών. Χρησιμοποιείται για την εκτίμηση της τάσης και της εποχικότητας στις χρονοσειρές και μπορεί να προσαρμοστεί για να λάβει υπόψη διάφορες παραλλαγές του μοντέλου. Στο Σχήμα 2.4 βλέπουμε μια σύγκριση της αποτελεσματικότητας των αλγορίθμων ARIMA και εκθετικής εξομάλυνσης στην πρόβλεψη κρουσμάτων Δάγκειου πυρετού. Η μπλε γραμμή αντιπροσωπεύει την πρόβλεψη με τη μέθοδο της εκθετικής εξομάλυνσης, η κόκκινη γραμμή δείχνει την πρόβλεψη με το μοντέλο ARIMA, ενώ η μαύρη γραμμή αντιστοιχεί στα πραγματικά δεδομένα. Παρατηρείται ότι η εκθετική εξομάλυνση προσεγγίζει καλύτερα τις πραγματικές τιμές, ιδιαίτερα κατά τις κορυφώσεις των κρουσμάτων.

Prophet. Ο Prophet [15] είναι ένας αλγόριθμος ανάλυσης χρονοσειρών που αναπτύχθηκε από το Facebook. Χρησιμοποιεί μια γραμμική μοντελοποίηση με την προσθήκη της εποχικότητας και της τάσης, ώστε να μπορεί να αντιμετωπίσει χρονοσειρές με διακοπτόμενη εμφάνιση ή απουσία δεδομένων. Στο Σχήμα 2.5 βλέπουμε μια σύγκριση προβλέψεων των αλγορίθμων ARIMA και Prophet σε δεδομένα χρονοσειρών. Η μπλε γραμμή αντιπροσωπεύει την πραγματική χρονοσειρά, η πορτοκαλί την πρόβλεψη του μοντέλου ARIMA και η πράσινη την πρόβλεψη του Prophet. Και οι δύο αλγόριθμοι ακολουθούν ικανοποιητικά τη γενική τάση των



Σχήμα 2.5: Αλγόριθμοι ARIMA και PROPHET. Η μπλε γραμμή αποτυπώνει τη μηνιαία πρόβλεψη, η πορτοκαλί την πρόβλεψη του ARIMA και η πράσινη την πρόβλεψη PROPHET

δεδομένων, ωστόσο το μοντέλο Prophet εμφανίζει ελαφρώς καλύτερη ακρίβεια στην προσέγγιση των πραγματικών τιμών, ειδικά σε σημεία αιχμής και μεταβολών της τάσης.

LSTM (Long Short-Term Memory). Το LSTM είναι ένα είδος αλγορίθμου νευρωνικών δικτύων που είναι ιδιαίτερα αποτελεσματικό [22] για την ανάλυση χρονοσειρών. Έχει την ικανότητα να διαπιστώσει μακροπρόθεσμες εξαρτήσεις στα δεδομένα και να προβλέψει με υψηλή ακρίβεια. Στο Σχήμα 2.6 έχουμε μια σύγκριση των αλγορίθμων ARIMA, LSTM Prophet στην πρόβλεψη μηνιαίας παραγωγής μύρας. Η μπλε γραμμή αντιπροσωπεύει τα πραγματικά δεδομένα παραγωγής. Οι προβλέψεις των μοντέλων αποτυπώνονται με διαφορετικά χρώματα: η πορτοκαλί γραμμή δείχνει τις προβλέψεις του ARIMA, η πράσινη του LSTM και η κόκκινη του Prophet. Το μοντέλο LSTM παρουσιάζει μεγαλύτερη ευαισθησία στις μεταβολές των δεδομένων, ενώ τα μοντέλα ARIMA και Prophet ακολουθούν ομαλότερες τάσεις.

Στον Πίνακα 2.1 παρατίθεται συγκεντρωτικά τα πλεονεκτήματα και τα μειονεκτήματα των παραπάνω αλγορίθμων. Οι αλγόριθμοι που αναφέρονται στον Πίνακα 2.1 είναι οι πιο γνωστοί και ευρέως διαδεδομένοι. Βέβαια, υπάρχουν και άλλοι αλγόριθμοι και



Σχήμα 2.6: Αλγόριθμοι ARIMA και LSTM

Αλγόριθμος	Μοντελοποίηση	Πλεονεκτήματα	Μειονεκτήματα
ARIMA	Σταθερές σειρές	Απλός, κλασικός	Όχι καλός σε μη γραμμικές
SARIMA	Σειρές με εποχικότητα	Καλό σε εποχικά μοτίβα	Περίπλοκη ρύθμιση
Prophet	Σειρές με τάση/εποχικότητα	Ανθεκτικός σε κενά δεδομένων	Λιγότερο ακριβής σε σύντομες προβλέψεις
LSTM	Μη γραμμικές/σύνθετες	Πιάνει μακροχρόνιες εξαρτήσεις	Απαιτεί πολλά δεδομένα

Πίνακας 2.1: Σύγκριση αλγορίθμων ανάλυσης χρονοσειρών

μεθοδολογίες που μπορούν να χρησιμοποιηθούν ανάλογα με τη φύση των δεδομένων και τους στόχους της ανάλυσης.

2.4 Χαρακτηριστικά των δεδομένων χρονοσειρών

Τα δεδομένα χρονοσειρών έχουν ορισμένα χαρακτηριστικά που τα διαχωρίζουν από άλλα είδη δεδομένων. Ορισμένα από τα κύρια χαρακτηριστικά των δεδομένων χρονοσειρών περιγράφονται παρακάτω.

Εξάρτηση από τον χρόνο. Κάθε σημείο μιας χρονοσειράς εξαρτάται από τη χρονική του θέση. Τα δεδομένα μιας χρονοσειράς είναι άμεσα συνδεδεμένα με τη χρονική τους σήμανση, αφού η χρονική διάσταση καθορίζει την ύπαρξη τάσεων, των εποχικών παραλλαγών και των επαναλαμβανόμενων μοτίβων. Ενόψει του χρονικού πλαισίου είμαστε σε θέση να ανιχνεύουμε και να ερμηνεύουμε αυτές τις χρονικά εξαρτώμενες συμπεριφορές των χρονοσειρών.

Ακολουθιακή ιδιότητα. Οι παρατηρήσεις μιας χρονοσειράς παρουσιάζονται με συγκεκριμένη σειρά. Η σειρά παίζει σημαντικό ρόλο καθώς κάθε τιμή συνδέεται με τις προηγούμενες παρατηρήσεις της σειράς, επιτρέποντας την ανίχνευση μοτίβων και την πρόβλεψη των μελλοντικών τιμών. Παρόλα αυτά, αυτό ισχύει για τη σειρά των δεδομένων και όχι απαραίτητα για το αν εξαρτώνται οι τιμές από το χρόνο.

Χρονικά μοτίβα. Οι χρονοσειρές συχνά περιέχουν χρονικά μοτίβα, όπως τάσεις, εποχικότητα, κυκλικότητα και αυξανόμενη διακύμανση. Αυτά τα μοτίβα αντικατοπτρίζουν συστηματικά χαρακτηριστικά που επαναλαμβάνονται κατά κάποιο τρόπο στη χρονοσειρά. Η ανίχνευση και η μοντελοποίηση αυτών των μοτίβων μας επιτρέπει να κατανοήσουμε και να εξηγήσουμε τη συμπεριφορά της χρονοσειράς και να προβλέψουμε τις μελλοντικές τιμές.

Αυτά τα χαρακτηριστικά είναι σημαντικά για την ανάλυση χρονοσειρών, καθώς μας επιτρέπουν να αποκωδικοποιήσουμε την πληροφορία που περιέχεται στις χρονοσειρές και να προβλέψουμε μελλοντικές τιμές ή συμπεριφορές.

2.5 Τομείς εφαρμογής δεδομένων χρονοσειρών

Τα δεδομένα χρονοσειρών μπορούν να χρησιμοποιηθούν σε διάφορους τομείς, λόγω των ιδιαίτερων χαρακτηριστικών τους και της πληροφορίας που παρέχουν. Παρακάτω παρουσιάζονται ορισμένοι σημαντικοί τομείς εφαρμογής τους, συνοδευόμενοι από παραδείγματα χρήσης.

Στην εξόρυξη γνώσης από δεδομένα (Data Mining), της αναγνώρισης προτύπων και της μηχανικής μάθησης. Η ανάλυση δεδομένων χρονοσειρών

χρησιμοποιείται ευρέως για να γίνει ομαδοποίηση των δεδομένων, να γίνει ταξινόμηση, να ανιχνευθούν ανωμαλίες και φυσικά για να γίνουν προβλέψεις. Πολλές είναι οι εφαρμογές που μπορεί να μας δώσει αυτή η τεχνική όπως είναι οι προβλέψεις για τη ζήτηση προϊόντων, η ανάλυση των ιατρικών δεδομένων για την ανίχνευση ασυνήθιστων περιπτώσεων και η αναγνώριση των μοτίβων συμπεριφοράς των χρηστών στο διαδίκτυο.

Στη μετα-επεξεργασία σήματος, την αυτόματη ρύθμιση (control engineering) και τη μηχανική επικοινωνιών (communication engineering). Οι χρονοσειρές αποτελούν πηγή πληροφοριών για τον εντοπισμό (signal detection) και τον υπολογισμό (estimation) των σημάτων. Κλασικά παραδείγματα περιλαμβάνουν την ανάλυση των μετρήσεων από τους αισθητήρες σε συστήματα ελέγχου, το monitoring των τηλεπικοινωνιακών συνδέσεων και την επεξεργασία των ηχητικών σημάτων.

Στην οικονομετρία και στην ποσοτική χρηματοοικονομία. Η ανάλυση δεδομένων χρονοσειρών χρησιμοποιείται ευρέως στην πρόβλεψη οικονομικών μεγεθών όπως μετοχικές τιμές, επιτόκια, συναλλαγματικές ισοτιμίες και άλλους χρηματοοικονομικούς δείκτες. Οι μέθοδοι αυτές συμβάλλουν στην εκτίμηση και διαχείριση κινδύνων, καθώς και στη λήψη στρατηγικών αποφάσεων επένδυσης και οικονομικού σχεδιασμού.

2.6 Τεχνικές κανονικοποίησης και κλιμάκωσης

Η κανονικοποίηση και η κλιμάκωση των δεδομένων [6] χρονοσειρών είναι δύο διαδικασίες που χρειάζονται για να γίνει αποτελεσματική ανάλυση και για να μπορέσουμε να κάνουμε ακριβείς προβλέψεις, ειδικά όταν τα δεδομένα έχουν μεγάλη διακύμανση, δεν είναι σταθερά ή δεν έχουν ομοιόμορφη κλίμακα. Αυτές οι τεχνικές μας βοηθούν να συγκρίνουμε και να ερμηνεύουμε καλύτερα τα δεδομένα, καθώς αφαιρούν ή μειώνουν τις επιπτώσεις της διαφορετικής κλίμακας και των ακραίων τιμών. Οι πιο δημοφιλείς τεχνικές κανονικοποίησης και αλλαγής κλίμακας που χρησιμοποιούνται για τη διαχείριση των χρονοσειρών περιγράφονται παρακάτω.

Κανονικοποίηση στο εύρος $[0, 1]$. Με αυτόν τον τρόπο, οι τιμές της χρονοσειράς απεικονίζονται στο διάστημα από 0 έως 1. Αυτός ο μετασχηματισμός γίνεται

αφαιρώντας την ελάχιστη τιμή και διαιρώντας με τη μεγαλύτερη τιμή που μπορεί να πάρει η σειρά. Χρησιμοποιούμε αυτόν τον τρόπο όταν μας ενδιαφέρει να δούμε τις αλλαγές μεταξύ των τιμών, αντί για τα απολύτως μεγέθη τους.

Κανονικοποίηση με Z-score (τυποποίηση). Με αυτή την τεχνική, διαμορφώνουμε τις τιμές έτσι ώστε η μέση τιμή του dataset να γίνει 0 και η τυπική απόκλιση 1. Για να το πετύχουμε αυτό, αφαιρούμε τη μέση τιμή των δεδομένων από κάθε τιμή και έπειτα διαιρούμε με την τυπική τους απόκλιση. Η μέθοδος Z-score είναι πολύ χρήσιμη όταν θέλουμε να αφαιρέσουμε την επίδραση των μεγάλων κύκλων και των τάσεων από τα δεδομένα μας.

Κλιμάκωση Min-Max (ελάχιστο – μέγιστο). Με αυτόν τον τρόπο, οι τιμές προσαρμόζονται έτσι ώστε να κινούνται εντός ενός συγκεκριμένου διαστήματος (π.χ., $[0, 1]$ ή $[-1, 1]$). Γίνεται αυτή η μετατροπή αφαιρώντας την ελάχιστη τιμή από κάθε παρατήρηση και διαιρώντας το αποτέλεσμα με το μήκος του διαστήματος (διαφορά μεταξύ μέγιστης και ελάχιστης τιμής). Έτσι, διατηρείται η σχετική θέση των τιμών και αντιμετωπίζεται η επίδραση των ακραίων τιμών.

Οι παραπάνω μέθοδοι κανονικοποίησης και κλιμάκωσης χρησιμοποιούνται ανάλογα με τον τύπο των δεδομένων και τους σκοπούς της ανάλυσης χρονοσειρών. Είναι σημαντικό να επιλεγεί η κατάλληλη τεχνική, καθώς μπορεί να βελτιώσει την ακρίβεια, την απόδοση και την ερμηνευσιμότητα των αποτελεσμάτων της ανάλυσης.

Αυτές οι τεχνικές κανονικοποίησης και κλιμάκωσης μπορούν να εφαρμοστούν ανάλογα με τη φύση των δεδομένων και τους στόχους της ανάλυσης χρονοσειρών. Η επιλογή της κατάλληλης τεχνικής μπορεί να βελτιώσει την ερμηνεία, την απόδοση και την ακρίβεια των αποτελεσμάτων.

2.7 Μέθοδοι αποσύνθεσης δεδομένων

Οι μέθοδοι αποσύνθεσης δεδομένων χρονοσειράς [6] αναφέρονται σε τεχνικές που σκοπό έχουν να εντοπίσουν τα βασικά συστατικά της. Με αυτόν τον τρόπο, μπορούμε

να αναλύσουμε και να κατανοήσουμε καλύτερα τα επιμέρους μοτίβα που περιέχει. Οι κύριες μέθοδοι αποσύνθεσης περιγράφονται παρακάτω.

Μέθοδος κινούμενων μέσων (moving averages). Αυτή η μέθοδος αποσύνθεσης χρησιμοποιεί ένα κινούμενο μέσο ως εργαλείο για να καθαρίσει τη χρονοσειρά από την εποχικότητά της, αφήνοντας ακέραιη την έκταση και την τυχαιότητα. Αφού αφαιρεθεί η εποχικότητα, μπορεί να εφαρμοστεί μια μέθοδος απλής κινητής μέσης, προκειμένου να διαχωριστεί η τάση.

Αποσύνθεση χρονοσειρών με ανάλυση συντελεστών. Αυτή η μέθοδος χρησιμοποιεί τη μεθοδολογία STL (Seasonal and Trend decomposition using Lowss) για να αποσυνθέσει μια χρονοσειρά σε τρεις βασικές συνιστώσες. Την τάση, την εποχικότητα και την τυχαιότητα. Η μέθοδος εφαρμόζει μια μη γραμμική παρεμβολή LOWSS (Locally Weighted Scatterplot Smoothing) για την απομάκρυνση της τάσης και της εποχικότητας.

Μέθοδος αποσύνθεσης exponential smoothing. Αυτή η μέθοδος αποσύνθεσης χρησιμοποιεί την τεχνική της εξομάλυνσης εκθέτοντας τη διάσπαση μιας χρονοσειράς. Αρχικά, η μέθοδος εφαρμόζει τρεις διαδοχικούς αλγορίθμους (εκθετική εξομάλυνση, γραμμική εξομάλυνση και απλή κίνηση) στην αρχική χρονοσειρά για να εκτιμήσει την τάση, την εποχικότητα και την τυχαιότητά της.

Αυτές είναι μερικές από τις κύριες μεθόδους αποσύνθεσης δεδομένων χρονοσειρών που χρησιμοποιούνται για τη διαίρεση και την εξαγωγή των διάφορων συνιστωσών της χρονοσειράς. Κάθε μέθοδος έχει τα πλεονεκτήματά της και επιλέγεται ανάλογα με τη φύση των δεδομένων και τους στόχους της ανάλυσης.

2.7.1 Συστατικά χρονοσειρών και η σημασία τους

Η ανάλυση δεδομένων χρονοσειρών απαιτεί την κατανόηση των συστατικών δεδομένων που αποτελούν μια χρονοσειρά και τη σημασία τους για την επεξεργασία και την ανάλυσή τους. Τα βασικά συστατικά των δεδομένων χρονοσειρών περιγράφονται παρακάτω.

Χρονική σήμανση (timestamp). Αναφέρεται στην αναπαράσταση του χρόνου που αντιστοιχεί σε κάθε παρατήρηση της χρονοσειράς. Η χρονική σήμανση είναι συνήθως η ανεξάρτητη μεταβλητή σε μια χρονοσειρά και επιτρέπει την ταξινόμηση, την αναζήτηση και την ανάκτηση δεδομένων.

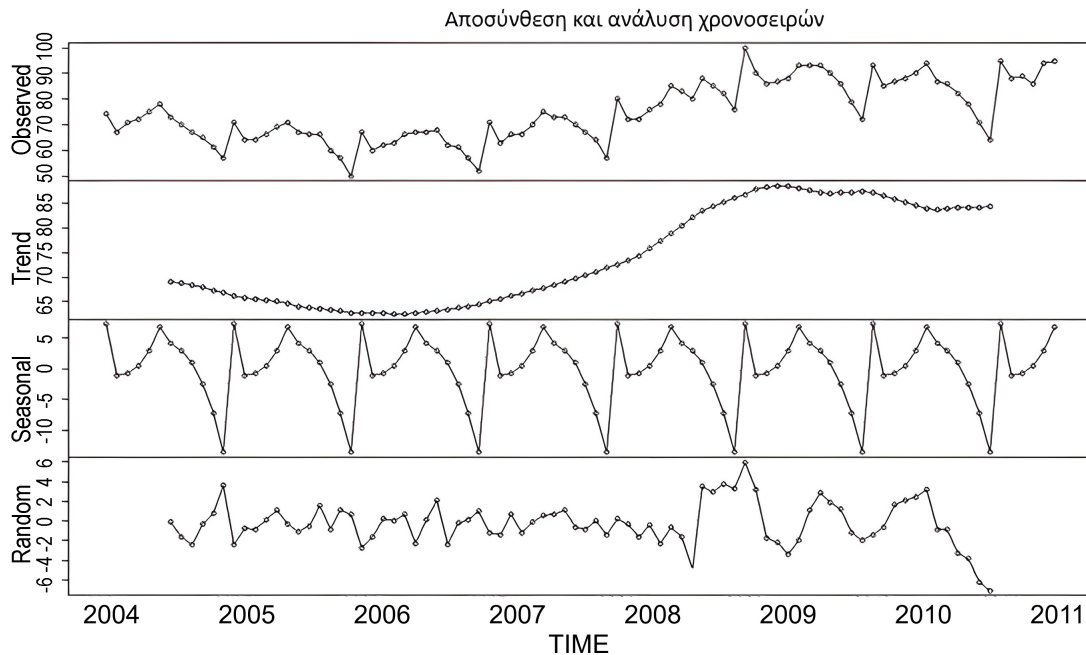
Τιμή (value). Αναφέρεται στη μετρούμενη ποσότητα, παράμετρο ή χαρακτηριστικό που σχετίζεται με την κάθε χρονική σήμανση. Η τιμή μπορεί να αναπαραστήσει οποιοδήποτε είδος δεδομένων, όπως θερμοκρασία, πωλήσεις, πίεση, αριθμό επισκεπτών κ.λπ.

Συμπληρωματικές πληροφορίες (additional information). Αναφέρεται σε πρόσθετα χαρακτηριστικά ή επισημειώσεις που συνοδεύουν την κάθε χρονική σήμανση και την τιμή. Αυτές οι πληροφορίες μπορούν να είναι σε μορφή ετικετών (tags) ή πεδίων (fields) και χρησιμοποιούνται για τον προσδιορισμό ή την ομαδοποίηση των δεδομένων. Για παράδειγμα, ετικέτες μπορεί να περιγράφουν την τοποθεσία ή την κατηγορία των δεδομένων, ενώ τα πεδία μπορεί να περιέχουν πληροφορίες όπως ονόματα αισθητήρων ή κωδικοί κατηγοριοποίησης.

Η σημασία αυτών των συστατικών δεδομένων χρονοσειρών είναι η καίρια όσον αφορά την ανάλυση, πρόβλεψη και ανακάλυψη μοτίβων στα χρονοσειριακά δεδομένα. Η χρονική σήμανση επιτρέπει την εύρεση και την ταξινόμηση των δεδομένων σύμφωνα με τη χρονολογική σειρά τους. Η τιμή αναπαριστά την ποσοτική μέτρηση που έγινε και είναι συνήθως το κύριο ενδιαφέρον για ανάλυση. Οι συμπληρωματικές πληροφορίες παρέχουν επιπλέον λεπτομέρειες και πληροφορίες για την κατηγοριοποίηση και την ομαδοποίηση των δεδομένων, καθιστώντας δυνατή την περαιτέρω ανάλυση και αξιολόγησή τους.

2.8 Μέθοδοι ανάλυσης και πρόγνωσης

Η ανάλυση χρονοσειρών είναι η χρήση στατιστικών μεθόδων για την ανάλυση χρονικών δεδομένων [16] και την εξαγωγή στατιστικών και χαρακτηριστικών που έχουν νόημα ως προς τα δεδομένα. Η ανάλυση βοηθά στην αναγνώριση τάσεων, κύκλων και εποχιακών διακυμάνσεων με κύριο σκοπό την πρόγνωση μελλοντικών γεγονότων. Η ανάλυση



Σχήμα 2.7: Διαφορές μεθόδων πρόγνωσης χρονοσειρών

μπορεί να είναι χρήσιμη ώστε να δούμε πώς μία δεδομένη μεταβλητή αλλάζει με την πάροδο του χρόνου (ενώ σχεδόν πάντα ο χρόνος είναι μία ανεξάρτητη μεταβλητή). Όπως βλέπουμε και στο Σχήμα 2.7, οι αναλύσεις αυτές μπορούν επίσης να χρησιμοποιηθούν για να εξετάσουμε με ποιό τρόπο συμβαίνουν οι αλλαγές σε σχέση με μία συγκεκριμένη παρατήρηση και πώς συγκρίνονται με τις αλλαγές μίας άλλης παρατήρησης. Η πρόγνωση χρονοσειρών χρησιμοποιεί πληροφορίες σε σχέση με τα ιστορικά δεδομένα και τα σχετικά μοτίβα ώστε να προβλέψει μελλοντική δραστηριότητα. Οι μέθοδοι πρόγνωσης χρονοσειρών περιλαμβάνουν τις παρακάτω αναλύσεις.

Η ανάλυση τάσεων (trend analysis) αφορά την αναγνώριση μακροχρόνιων κατευθύνσεων στα δεδομένα.

Η ανάλυση κυκλικών διακυμάνσεων (cyclical fluctuation analysis)

εξετάζει περιοδικές αυξομειώσεις που σχετίζονται με οικονομικούς ή κοινωνικούς κύκλους μεγαλύτερης διάρκειας.

Η ανάλυση εποχικών προτύπων (seasonal pattern analysis)

επικεντρώνεται στις μεταβολές που επαναλαμβάνονται σε σταθερά διαστή-

ματα, όπως μήνες ή εποχές.

Όπως και σε όλες τις μεθόδους πρόβλεψης η επιτυχία δεν είναι εγγυημένη. Για το λόγο αυτό ακόμα συχνά χρησιμοποιείται η τεχνική του machine learning. Για να μπορέσουμε να προβλέψουμε το μέλλον, η μοντελοποίηση των χρονοσειρών περιλαμβάνει δεδομένα βάσει χρονικής σειράς (χρόνια, μήνες ή μέρες) ώστε να αποκρυπτογραφήσουμε κρυμμένα χαρακτηριστικά καθοριστικής σημασίας για τη λήψη αποφάσεων.

2.9 Προσεγγίσεις πρόβλεψης

Οι προσεγγίσεις πρόβλεψης χρονοσειρών περιλαμβάνουν διάφορες τεχνικές που χρησιμοποιούνται για να προβλέψουν τις μελλοντικές τιμές μιας χρονοσειράς βάσει των παρελθοντικών δεδομένων. Οι πιο γνωστές προσεγγίσεις των προβλέψεων αυτών περιγράφονται παρακάτω.

Απλή πρόβλεψη. Αυτή η προσέγγιση περιλαμβάνει τη χρήση της τελευταίας τιμής της χρονοσειράς ως πρόβλεψη για την επόμενη τιμή. Αν και είναι η πιο απλή μέθοδος πρόβλεψης, μπορεί να είναι αποτελεσματική για χρονοσειρές με σταθερή τάση και λίγη εποχικότητα.

Κινητός μέσος όρος. Αυτή η προσέγγιση χρησιμοποιεί έναν κινητό μέσο όρο για την πρόβλεψη μελλοντικών τιμών. Ο κινητός μέσος όρος υπολογίζεται παίρνοντας τον μέσο όρο των τελευταίων N τιμών της χρονοσειράς, όπου N είναι ένας φυσικός αριθμός που συνήθως ορίζεται από τον χρήστη. Αυτή η μέθοδος μπορεί να βοηθήσει στην απομάκρυνση του θορύβου και στην ανίχνευση των τάσεων.

Εκθετική εξομάλυνση. Η εκθετική εξομάλυνση είναι μια προσέγγιση πρόβλεψης που βασίζεται στην εκθετική απόκριση των δεδομένων. Υπάρχουν διάφορες μέθοδοι εκθετικής εξομάλυνσης, όπως η απλή εκθετική εξομάλυνση, η εκθετική εξομάλυνση Holt και η εκθετική εξομάλυνση Holt-Winters. Αυτές οι μέθοδοι λαμβάνουν υπόψη την τάση και την εποχικότητα των δεδομένων για την πρόβλεψη των μελλοντικών τιμών.

Ο κινητός μέσος όρος και η εκθετική εξομάλυνση αποτελούν μερικές από τις κύριες προσεγγίσεις πρόβλεψης χρονοσειρών. Κάθε μέθοδος έχει τα δικά της χαρακτηριστικά και πρέπει να επιλέγεται ανάλογα με τη φύση των δεδομένων και τους στόχους της πρόβλεψης.

2.10 Εργαλεία οπτικοποίησης

Τα εργαλεία αποτύπωσης χρονοσειρών αναφέρονται σε μεθόδους ή τεχνικές που χρησιμοποιούνται για την απεικόνιση και την οπτικοποίηση των χρονοσειρών. Οι απεικονίσεις των χρονοσειρών μπορούν να παρέχουν πολύτιμες πληροφορίες και επιτρέπουν την ανάλυση και την ερμηνεία των δεδομένων. Παρατίθενται οι κύριες κατηγορίες γραφημάτων.

Γραφήματα γραμμής. Τα γραφήματα γραμμής είναι ένα από τα πιο κοινά εργαλεία αποτύπωσης χρονοσειρών. Απεικονίζουν την τιμή της χρονοσειράς σε συνάρτηση με τον χρόνο, με τις χρονικές σημείες να βρίσκονται στον οριζόντιο άξονα και τις τιμές στον κατακόρυφο άξονα.

Γραφήματα διασποράς. Τα γραφήματα διασποράς αναπαριστούν τη διακύμανση ή τη διασπορά των δεδομένων χρονοσειράς. Απεικονίζουν τις τιμές στον κατακόρυφο άξονα και το χρόνο στον οριζόντιο άξονα, με σημεία που αντιστοιχούν στις παρατηρήσεις της χρονοσειράς.

Γραφήματα διασποράς-χρόνου. Τα γραφήματα διασποράς-χρόνου είναι μια ειδική μορφή γραφημάτων όπου οι παρατηρήσεις απεικονίζονται σε ένα διάγραμμα διασποράς, με τον οριζόντιο άξονα να αναπαριστά τον χρόνο και τον κατακόρυφο άξονα να αναπαριστά τη διακύμανση των παρατηρήσεων.

Ιστόγραμμα. Το ιστόγραμμα είναι μια απεικόνιση που δείχνει τη συχνότητα εμφάνισης των διάφορων τιμών σε ένα δεδομένο σύνολο δεδομένων χρονοσειράς. Αυτό μπορεί να βοηθήσει στην ανίχνευση τυχόν μοτίβων και τάσεων στα δεδομένα.

Συνδεδεμένα διαγράμματα. Τα συνδεδεμένα διαγράμματα απεικονίζουν τις παρατηρήσεις της χρονοσειράς με γραμμές που συνδέουν τα σημεία. Αυτό επιτρέπει την απεικόνιση των μεταβολών και των μοτίβων μεταξύ των παρατηρήσεων.

Μπορείτε να χρησιμοποιήσετε τα παραπάνω εργαλεία ή μια συνδυασμένη προσέγγιση για να αποτυπώσετε και να οπτικοποιήσετε τα δεδομένα χρονοσειράς σας, προκειμένου να ανακαλύψετε πρότυπα, τάσεις και άλλες πληροφορίες που ενδέχεται να είναι κρυμμένες στα δεδομένα σας.

2.11 Σύνοψη κεφαλαίου

Στο παρόν κεφάλαιο παρουσιάστηκε μια ολοκληρωμένη εισαγωγή στα δεδομένα χρονοσειρών, ξεκινώντας από τον ορισμό και τη σημασία τους σε διάφορους τομείς. Δόθηκε έμφαση στη βασική τους ιδιαιτερότητα, δηλαδή τη σύνδεσή τους με τον χρόνο, που τις διαφοροποιεί από άλλες μορφές δεδομένων. Παρουσιάστηκαν οι βασικές κατηγορίες χρονοσειρών (γραμμικές και μη γραμμικές) και αναλύθηκαν θεμελιώδεις έννοιες, όπως η τάση, η εποχικότητα και η σταθερότητα. Επιπλέον, έγινε διάκριση ανάμεσα σε άλλους τύπους δεδομένων, όπως τα cross-sectional και τα panel data.

Στη συνέχεια, παρουσιάστηκαν οι πιο γνωστοί αλγόριθμοι ανάλυσης και πρόβλεψης χρονοσειρών, μεταξύ των οποίων οι ARIMA, SARIMA, Prophet και LSTM, με στόχο την εκτίμηση μελλοντικών τιμών μέσα από μαθηματικά μοντέλα. Εξετάστηκαν ακόμη τα χαρακτηριστικά των χρονοσειρών, οι τεχνικές κανονικοποίησης και η σημασία των βασικών τους συστατικών: χρονική σήμανση, τιμή και συμπληρωματικά μεταδεδομένα.

Τέλος, αναδείχθηκαν οι κύριες μέθοδοι απεικόνισης, υπογραμμίζοντας τον ρόλο της οπτικοποίησης για την κατανόηση και ερμηνεία των δεδομένων. Το κεφάλαιο αυτό αποτελεί τη θεωρητική βάση για τα επόμενα, όπου θα παρουσιαστούν πρακτικές εφαρμογές αποθήκευσης, επεξεργασίας και ανάλυσης χρονοσειρών σε πραγματικά περιβάλλοντα.

Κεφάλαιο 3

Αποθήκευση χρονοσειρών

Η αποθήκευση δεδομένων χρονοσειρών αποτελεί κρίσιμο στάδιο στη συνολική διαδικασία διαχείρισής τους. Μετά την ανάλυση των βασικών χαρακτηριστικών των χρονοσειρών στο προηγούμενο κεφάλαιο, σε αυτό το κεφάλαιο εξετάζονται οι τρόποι με τους οποίους μπορούμε να αποθηκεύσουμε αποτελεσματικά τέτοια δεδομένα, ώστε να διασφαλίζεται η ακρίβεια, η ταχύτητα και η δυνατότητα ανάλυσης σε πραγματικό χρόνο.

3.1 Γενικά συστήματα

Πριν περάσουμε στις εξειδικευμένες βάσεις δεδομένων για χρονοσειρές, είναι σημαντικό να παρουσιαστούν τα γενικά συστήματα αποθήκευσης δεδομένων. Αυτά αποτελούν τη βάση για την κατανόηση των διαφορών ανάμεσα στα σχεσιακά και μη σχεσιακά μοντέλα που θα ακολουθήσουν.

3.1.1 Σχεσιακά συστήματα

Τα σχεσιακά συστήματα αποθήκευσης δεδομένων αποτελούν τον θεμέλιο λίθο της τεχνολογίας βάσεων δεδομένων, αφού παρέχουν την πιο βασική και ταυτόχρονα αποτελεσματική οργάνωση και διαχείριση δεδομένων. Το βασικό χαρακτηριστικό των σχεσιακών βάσεων είναι η ικανότητά τους να διαχειρίζονται τα δεδομένα με πίνακες, όπου κάθε σειρά αντιπροσωπεύει μια εγγραφή και κάθε στήλη αναπαριστά ένα συγκεκριμένο χαρακτηριστικό. Η δομή αυτή παρέχει σαφήνεια, αποτελεσματική ανάκτηση δεδομένων

και δυνατότητα εφαρμογής συνθηκών μέσω της δομημένης γλώσσας ερωτημάτων SQL. Η συσχέτιση μεταξύ πινάκων επιτρέπει την αναζήτηση, τη σύνδεση και την επεξεργασία δεδομένων με ακρίβεια. Επιπλέον, η σχεσιακή δομή μας δίνει τη δυνατότητα εφαρμογής περιορισμών για τη διατήρηση της συνοχής και της ακεραιότητας των δεδομένων, κάνοντας τα σχεσιακά συστήματα ιδανική επιλογή για εφαρμογές που απαιτούν αξιόπιστη αποθήκευση και ανάκτηση δεδομένων.

3.1.1.1 PostgreSQL

Πρόκειται για ένα ισχυρό και ανοιχτού κώδικα σύστημα διαχείρισης βάσεων δεδομένων που ξεχωρίζει για την αξιοπιστία και τις προηγμένες δυνατότητες που παρέχει στον απλό χρήστη, αλλά και εταιρικές ανάγκες με μεγάλες δυνατότητες επέκτασης. Ως ένα σύστημα που ανήκει στις σχεσιακές δομές δεδομένων, η PostgreSQL παρέχει πλούσια χαρακτηριστικά που καλύπτουν ευρύ φάσμα εφαρμογών. Η δυνατότητα της να διαχειρίζεται μεγάλο όγκο δεδομένων, η υποστήριξη πολυσύνολων, η ευελιξία του συστήματος και η εξαιρετική απόδοσή της την καθιστούν επιλογή προτίμησης για επιχειρήσεις και προγραμματιστές. Επιπλέον, η PostgreSQL υποστηρίζεται από μια ενεργή κοινότητα προγραμματιστών, προσφέροντας τακτικές ενημερώσεις και βελτιώσεις, ενισχύοντας με αυτό το τρόπο την ασφάλεια και τη συνολική απόδοση του συστήματος. Με τη δυνατότητα επέκτασης μέσω πρόσθετων εργαλείων, η PostgreSQL συνιστά ένα ευέλικτο και ολοκληρωμένο εργαλείο για τη διαχείριση και ανάπτυξη βάσεων δεδομένων.

Μερικά από τα πλεονεκτήματά της είναι η υποστήριξη για πολλαπλές μορφές δεδομένων όπως το JSONB, μια βελτιστοποιημένη δυαδική μορφή αποθήκευσης JSON που επιτρέπει ταχύτερες αναζητήσεις και μικρότερη κατανάλωση χώρου σε σχέση με το απλό JSON. Επιπλέον, η δυνατότητα προσθήκης επεκτάσεων από την κοινότητα, δίνει αμέτρητες νέες λειτουργίες όπως κρυπτογράφηση δεδομένων, ευκολίες στην εξαγωγή και διαχείριση συμπιεσμένων βάσεων και περαιτέρω διασφάλιση της πληροφορίας που βρίσκεται αποθηκευμένη στη βάση δεδομένων.

Φυσικά, υπάρχουν και κάποια βασικά μειονεκτήματα. Η εγκατάσταση και βασική λειτουργία της PostgreSQL είναι σαφώς αρκετά δυσκολότερη από άλλα σχεσιακά συστήματα, ενώ παράλληλα πολλές προηγμένες λειτουργίες απαιτούν εξειδικευμένες γνώσεις

που δεν είναι εύκολο να αποκτηθούν από έναν νέο προγραμματιστή.

3.1.1.2 Microsoft SQL Server

Μερικά από τα βασικά χαρακτηριστικά του Microsoft SQL Server περιγράφονται παρακάτω, δίνοντας μια πιο σαφή εικόνα των σχεσιακών βάσεων δεδομένων, σε συνδυασμό με την περιγραφή της PostgreSQL που παρατίθεται στην προηγούμενη ενότητα. Το SQL Server [?] παρέχει εκτεταμένα εργαλεία διαχείρισης, όπως το SQL Server Management Studio, που επιτρέπουν τη διαχείριση των βάσεων δεδομένων και τη δημιουργία και εκτέλεση ερωτημάτων. Επιπλέον, παρέχει ποικίλα χαρακτηριστικά ασφαλείας, συμπεριλαμβανομένης της κρυπτογράφησης δεδομένων, του διαχωρισμού ρόλων και του ελέγχου πρόσβασης σύμφωνα με τα δικαιώματα του εκάστοτε ρόλου. Επιτρέπει την ανάπτυξη εφαρμογών με χρήση πολλών γλωσσών προγραμματισμού όπως Java, Python κ.λπ. Επίσης, υποστηρίζει ενσωμάτωση με άλλα πληροφοριακά συστήματα μέσω διαφόρων πρωτοκόλλων επικοινωνίας. Τέλος, αν και δεν έχει σχεδιαστεί αποκλειστικά για χρονοσειρές, το Microsoft SQL Server υποστηρίζει την αποθήκευση και την ανάκτηση τέτοιων δεδομένων μέσω χρήσης χρονοσφραγίδων και λειτουργιών ανάλυσης που επιτρέπουν τη διαχείριση και σύγκριση τιμών στο χρόνο.

Ίσως, το ισχυρότερο από τα πλεονεκτήματα έναντι κάθε άλλης βάσης, σχεσιακής και μη, είναι οι δυνατότητες επιχειρηματικής νοημοσύνης που παρέχει η συγκεκριμένη πλατφόρμα. Με αυτά τα εργαλεία έχει τη δυνατότητα να πραγματοποιήσει εξόρυξη δεδομένων με σκοπό τη λήψη αποφάσεων. Επιπλέον, αυτής της ιδιότητας, το Microsoft SQL Server προσφέρει συμβατότητα με πολλές πλατφόρμες και εφαρμογές, που έχουν σκοπό την ανάπτυξη λογισμικού, τη διαχείριση και παρουσίαση των δεδομένων. Διατίθενται πολλά εργαλεία που διευκολύνουν τη διαχείριση, συντήρηση και παρακολούθηση της βάσης δεδομένων, ενώ παράλληλα προσφέρονται ανάλογα εργαλεία με σκοπό την προστασία των δεδομένων, μέσω κρυπτογράφησης και μηχανισμών περιορισμού πρόσβασης στη βάση.

Το πρώτο από τα μειονεκτήματα, απορρέει και από το τίτλο της πλατφόρμας. Είναι η έλλειψη επαρκούς συμβατότητας με λειτουργικά συστήματα εκτός Windows. Έχουν δημιουργηθεί εφαρμογές με σκοπό τη λειτουργία και σε συστήματα που χρησιμοποιούν

Linux/Mac αλλά δυστυχώς για όλο το φάσμα των δυνατοτήτων του. Ένα επιπλέον μειονέκτημα είναι το κόστος αγοράς μιας άδειας, αφού για μια μεγάλη επιχείρηση με πολλούς υπολογιστές, μπορεί να απαιτηθούν πολλά χρήματα για να καλυφθούν οι ανάγκες της επιχείρησης. Θα πρέπει επίσης να λάβουμε υπόψιν μας ότι οι απαιτήσεις της εκάστοτε υποδομής είναι πολύ υψηλές, αφού χρειάζεται τεχνικές γνώσεις και υποδομή, ιδιαίτερα για πολύπλοκες εγκαταστάσεις και μεγάλους όγκους δεδομένων. Τέλος, όπως συμβαίνει και σε κάθε άλλη σχεσιακή βάση δεδομένων, η απόδοση του SQL Server εξαρτάται από τον αριθμό των διεργασιών που εκτελούνται ταυτόχρονα, καθώς και από τον συνολικό όγκο των δεδομένων που διαχειρίζεται η βάση.

3.1.1.3 MySQL

Τα βασικά χαρακτηριστικά του MySQL [?] το καθιστούν μία από τις πιο δημοφιλείς σχεσιακές βάσεις δεδομένων, παρουσιάζοντας ομοιότητες με το PostgreSQL αλλά με πιο απλοποιημένη αρχιτεκτονική. Πρόκειται για την πιο διάσημη σχεσιακή βάση δεδομένων στον κόσμο. Σχεδιάστηκε με σκοπό να είναι γρήγορη και εύκολη στη χρήση. Μπορεί να χρησιμοποιηθεί σε διαφορετικές πλατφόρμες, δίνοντας διαφορετικά χαρακτηριστικά ανάλογα την έκδοσή της. Για παράδειγμα, η πλήρης συμβατότητα της MySQL με την παραδοσιακή δομημένη γλώσσα προγραμματισμού SQL επιτυγχάνεται μέσω της χρήσης της μηχανής αποθήκευσης InnoDB. Βασικό χαρακτηριστικό σε κάθε έκδοσή της είναι ότι υποστηρίζει τις ιδιότητες ACID, (Atomicity, Consistency, Isolation, Durability), διασφαλίζοντας την ακεραιότητα και τη συνέπεια των δεδομένων. Οι δυνατότητες επεκτασιμότητάς της περιορίζονται σε μια διαδικασία επέκτασης που ονομάζεται κατακόρυφη κλιμάκωση. Υπάρχει, βέβαια, η δυνατότητα κοινής χρήσης βάσεων δεδομένων ίδιου τύπου· ωστόσο, σε αυτή την περίπτωση παρατηρούνται σημαντικές καθυστερήσεις στις επιδόσεις της πλατφόρμας. Σε γενικές γραμμές, η επεκτασιμότητα της MySQL βασίζεται στη λειτουργία σύνδεσης (JOIN), η οποία επιτρέπει τη συσχέτιση πινάκων ή ακόμη και διαφορετικών βάσεων δεδομένων μεταξύ τους. Η δυνατότητα αυτή αποτελεί σημαντικό πλεονέκτημα, καθώς προσφέρει ευελιξία στη διαχείριση και οργάνωση των δεδομένων, αλλά ταυτόχρονα μπορεί να επιφέρει επιπτώσεις στις επιδόσεις των ερωτημάτων της βάσης.

Το σημαντικότερο από τα πλεονεκτήματα της πλατφόρμας είναι η ίδια της η ουσία, δηλαδή το γεγονός ότι πρόκειται για μια ολοκληρωμένη πλατφόρμα σχεσιακής βάσης δεδομένων. Οι δυνατότητες που προσφέρει η λειτουργία σύνδεσης πινάκων (JOIN) καθιστούν τα ερωτήματα στη βάση ιδιαίτερα ευέλικτα, επιτρέποντας τη δημιουργία σύνθετων ερωτημάτων και την εξαγωγή πολυδιάστατων απαντήσεων. Ωστόσο, λόγω αυτής της αυξημένης ευελιξίας, η MySQL εμφανίζει συμβατότητα με τα πρότυπα SQL, ενώ παράλληλα προσπαθεί να αποτρέπει τις διπλοεγγραφές. Επιπλέον, χάρη στην υποστήριξη των ιδιοτήτων ατομικότητας, συνοχής, απομόνωσης και διάρκειας (ACID), μπορεί να προσφέρει επιπρόσθετες δομικές δυνατότητες ανάλογα με τις ανάγκες της εκάστοτε βάσης δεδομένων.

Το βασικότερο από τα μειονεκτήματα της MySQL αφορά την ευελιξία της επεκτασιμότητάς της. Αν και η MySQL μπορεί να επεκταθεί, η διαδικασία αυτή συχνά συνεπάγεται σημαντικές καθυστερήσεις και αυξημένη κατανάλωση υπολογιστικών πόρων. Η ταχύτητα εκτέλεσης των ερωτημάτων μειώνεται όσο αυξάνεται ο αριθμός των πιθανών συνδυασμών τους. Για παράδειγμα, όταν πραγματοποιείται σύνδεση (JOIN) σε πολλαπλούς πίνακες, παρατηρείται μείωση των χρόνων απόκρισης, είτε πρόκειται για ερωτήματα ανάγνωσης είτε για ερωτήματα εγγραφής. Τέλος, για την πλήρη αξιοποίηση όλων των δυνατοτήτων της MySQL απαιτείται η χρήση κάποιας από τις εμπορικές επί πληρωμή εκδόσεις της, οι οποίες είναι διαθέσιμες με την πληρωμή του ανάλογου αντιτίμου.

3.1.2 Μη σχεσιακές βάσεις δεδομένων

Οι μη σχεσιακές βάσεις δεδομένων, γνωστές και ως NoSQL, αναπτύχθηκαν για να αντιμετωπίσουν τους περιορισμούς των παραδοσιακών σχεσιακών βάσεων στη διαχείριση μεγάλου όγκου και μη δομημένων δεδομένων. Προσφέρουν υψηλή επεκτασιμότητα, ταχύτητα και ευελιξία, καθιστώντας τις ιδανικές για σύγχρονες εφαρμογές μεγάλης κλίμακας.

Σε αντίθεση με τις σχεσιακές βάσεις, τα συστήματα NoSQL δε βασίζονται σε πίνακες, αλλά σε διαφορετικά μοντέλα αποθήκευσης όπως ζεύγη κλειδιού-τιμής, έγγραφα ή στήλες. Έτσι επιτυγχάνεται μεγαλύτερη προσαρμοστικότητα και αποδοτικότητα. Στις επόμενες ενότητες παρουσιάζονται ενδεικτικά μερικά από τα πιο γνωστά συστήματα

NoSQL, με έμφαση στα χαρακτηριστικά και τα πλεονεκτήματά τους.

3.1.2.1 Cassandra

Το Apache Cassandra [?] είναι ένα από τα πιο γνωστά συστήματα μη σχεσιακών βάσεων δεδομένων ανοιχτού κώδικα, που σχεδιάστηκε αρχικά από το Facebook (<https://www.facebook.com>) και στη συνέχεια εξελίχθηκε υπό την αιγίδα του ιδρύματος Apache. Δημιουργήθηκε με στόχο να δώσει λύση σε ένα από τα μεγαλύτερα προβλήματα των παραδοσιακών βάσεων: τη δυσκολία διαχείρισης τεράστιου όγκου δεδομένων σε πραγματικό χρόνο, χωρίς να επηρεάζεται η απόδοση και η διαθεσιμότητα του συστήματος. Για τον λόγο αυτό, το Cassandra βασίζεται σε μία πλήρως κατανεμημένη αρχιτεκτονική χωρίς κεντρικό διακομιστή, όπου όλοι οι κόμβοι του συστήματος είναι ισότιμοι και συνεργάζονται μεταξύ τους για την αποθήκευση, ανάκτηση και ενημέρωση των δεδομένων.

Η επικοινωνία μεταξύ των κόμβων πραγματοποιείται μέσω ενός ειδικού πρωτοκόλλου, γνωστού ως Gossip, το οποίο επιτρέπει στους κόμβους να ανταλλάσσουν συνεχώς πληροφορίες για την κατάστασή τους και να ενημερώνουν το σύστημα σε περίπτωση βλάβης ή αποσύνδεσης κάποιου κόμβου. Με αυτόν τον τρόπο, το Cassandra μπορεί να προσαρμόζεται αυτόματα σε αλλαγές του δικτύου, να παραμένει λειτουργικό ακόμη και σε περιπτώσεις απώλειας κόμβων, και να διατηρεί υψηλή διαθεσιμότητα χωρίς ανθρώπινη παρέμβαση. Επιπλέον, η ύπαρξη μηχανισμών αυτόματης αντιγραφής των δεδομένων σε πολλαπλούς κόμβους προσφέρει σημαντική ανθεκτικότητα, καθώς κάθε πληροφορία υπάρχει σε περισσότερα από ένα σημεία του συστήματος, εξασφαλίζοντας την ακεραιότητα των δεδομένων ακόμη και σε συνθήκες αστοχίας υλικού.

Ένα ακόμη βασικό χαρακτηριστικό του Cassandra είναι η δυνατότητά του να επεκτείνεται οριζόντια με πολύ απλό τρόπο. Η προσθήκη νέων κόμβων στο δίκτυο δεν απαιτεί επανεκκίνηση ή τροποποίηση της υπάρχουσας δομής, γεγονός που επιτρέπει την αύξηση της απόδοσης και της χωρητικότητας όσο αυξάνονται τα δεδομένα. Αυτό το χαρακτηριστικό το καθιστά ιδανικό για συστήματα που αναπτύσσονται σταδιακά ή που χρειάζονται να υποστηρίξουν συνεχώς αυξανόμενο όγκο πληροφορίας, όπως εφαρμογές τηλεμετρίας, αισθητήρων και του διαδικτύου των πραγμάτων (internet of things – IoT).

Η ευελιξία της αρχιτεκτονικής του Cassandra επιτρέπει επίσης τη ρύθμιση του βαθμού συνέπειας ανάλογα με τις ανάγκες της εφαρμογής. Ο διαχειριστής μπορεί να επιλέξει αν μία εγγραφή θεωρείται επιτυχής μόλις αποθηκευτεί σε έναν κόμβο ή μόνο όταν έχει επιβεβαιωθεί από την πλειοψηφία των κόμβων. Έτσι, επιτυγχάνεται ισορροπία μεταξύ ταχύτητας και αξιοπιστίας, ανάλογα με το πόσο κρίσιμα είναι τα δεδομένα. Για παράδειγμα, σε εφαρμογές που προτεραιότητα έχει η ταχύτητα (όπως οι μετρήσεις αισθητήρων), μπορεί να προτιμηθεί χαμηλότερος βαθμός συνέπειας, ενώ σε εφαρμογές οικονομικών συναλλαγών προτιμάται υψηλότερος βαθμός συνέπειας για απόλυτη ακρίβεια των δεδομένων.

Ένα ακόμη πλεονέκτημα του Cassandra είναι ότι δεν απαιτεί σταθερό σχήμα δεδομένων. Δηλαδή, δε χρειάζεται να είναι προκαθορισμένα τα πεδία κάθε εγγραφής, όπως συμβαίνει στις σχεσιακές βάσεις. Αυτό δίνει τη δυνατότητα στους προγραμματιστές να προσαρμόζουν τη δομή των δεδομένων ανάλογα με τις ανάγκες της εφαρμογής, να προσθέτουν νέα πεδία χωρίς αναδιοργάνωση των πινάκων και να αποθηκεύουν ετερογενή δεδομένα στον ίδιο πίνακα. Αυτή η ελευθερία καθιστά το Cassandra ιδιαίτερα εύελικο και εύχρηστο σε περιβάλλοντα όπου τα δεδομένα αλλάζουν συνεχώς ή δεν έχουν ενιαία μορφή.

Παρά τα πλεονεκτήματα αυτά, η χρήση του συνοδεύεται και από ορισμένες προκλήσεις. Η εγκατάσταση και ρύθμιση του συστήματος απαιτεί προσεκτικό σχεδιασμό και τεχνική γνώση, καθώς η απόδοση εξαρτάται άμεσα από τη σωστή επιλογή κλειδιών κατακερματισμού και από τον τρόπο που ορίζονται τα ερωτήματα. Επιπλέον, η εκμάθηση των βασικών αρχών λειτουργίας του Cassandra μπορεί να είναι πιο απαιτητική σε σχέση με πιο απλές βάσεις δεδομένων, καθώς χρειάζεται κατανόηση της κατανεμημένης λογικής και των μηχανισμών συγχρονισμού μεταξύ κόμβων. Ωστόσο, αυτό δεν αποτελεί σημαντικό εμπόδιο, καθώς η κοινότητα υποστήριξης του έργου είναι ιδιαίτερα ενεργή και διαθέτει πλούσια τεκμηρίωση και παραδείγματα χρήσης.

Σε πρακτικές εφαρμογές, το Cassandra έχει αποδειχθεί εξαιρετικά αποτελεσματικό στη διαχείριση δεδομένων που συλλέγονται συνεχώς και σε μεγάλο όγκο, όπως δεδομένα μετρητών, δικτυακής κίνησης ή καταγραφών (logs). Χάρη στην αρχιτεκτονική του, μπορεί να επεξεργάζεται εκατομμύρια εγγραφές ανά δευτερόλεπτο χωρίς να επιβαρύνεται

σημαντικά η απόδοση, ενώ η διαδικασία ανάκτησης δεδομένων παραμένει γρήγορη και αξιόπιστη ακόμη και όταν το μέγεθος της βάσης αυξάνεται σημαντικά.

Συνοψίζοντας, το Apache Cassandra αποτελεί μία από τις πιο αξιόπιστες και αποδοτικές λύσεις για τη διαχείριση μεγάλου όγκου δεδομένων χρονοσειρών. Η σταθερότητα, η ευκολία επέκτασης, η αντοχή σε αποτυχίες και η δυνατότητα αυτόματης ισορρόπησης του φόρτου το καθιστούν ιδανικό εργαλείο για κάθε σύγχρονη υποδομή που απαιτεί συνεχή λειτουργία και υψηλή απόδοση. Παρόλο που η αρχική εξοικείωση με το περιβάλλον του μπορεί να απαιτεί χρόνο, η συνολική ευκολία συντήρησης και η αποτελεσματικότητά του δικαιολογούν πλήρως τη δημοφιλία του ως βασικό εργαλείο αποθήκευσης δεδομένων σε κατακευματισμένα συστήματα.

3.1.2.2 Elasticsearch

Το Elasticsearch [?] είναι μία από τις πιο διαδεδομένες πλατφόρμες αναζήτησης και ανάλυσης δεδομένων παγκοσμίως. Δημιουργήθηκε το 2010 και βασίζεται στη μηχανή Lucene της Apache, προσφέροντας έναν εξαιρετικά αποδοτικό τρόπο αναζήτησης και ανάλυσης κειμένου και δεδομένων σε πραγματικό χρόνο. Αρχικά σχεδιάστηκε για τη διαχείριση μεγάλων ποσοτήτων εγγράφων, αλλά σήμερα χρησιμοποιείται σε πλήθος εφαρμογών, από μηχανές αναζήτησης και συστήματα παρακολούθησης μέχρι πλατφόρμες ανάλυσης δεδομένων και καταγραφής συμβάντων. Χρησιμοποιείται από εκατομμύρια χρήστες και μεγάλες εταιρείες διεθνώς, λόγω της σταθερότητας και της υψηλής απόδοσής του.

Η αποθήκευση δεδομένων στο Elasticsearch βασίζεται στην έννοια των δεικτών (indexes), οι οποίοι λειτουργούν παρόμοια με τους πίνακες στις σχεσιακές βάσεις δεδομένων. Κάθε δείκτης περιέχει οργανωμένα δεδομένα, όπως εγγραφές χρηστών, συναλλαγές ή πληροφορίες συστημάτων. Το ιδιαίτερο χαρακτηριστικό του είναι ότι κάθε δείκτης μπορεί να χωριστεί σε μικρότερα τμήματα, τμήματα της βάσης (shards) τα οποία μπορούν να αποθηκευτούν σε διαφορετικούς κόμβους ενός συστήματος. Αυτή η αρχιτεκτονική επιτρέπει στο Elasticsearch να διαχειρίζεται πολύ μεγάλα σύνολα δεδομένων, εξασφαλίζοντας ταυτόχρονα γρήγορη πρόσβαση και ευκολία στην επεκτασιμότητα. Η προσθήκη νέων κόμβων σε μία συστάδα υπολογιστών (cluster) γίνεται εύκολα και αυτό-

ματα, καθώς το ίδιο το σύστημα φροντίζει για τη διανομή και την αντιγραφή των shards, ώστε να διατηρείται η ισορροπία και η διαθεσιμότητα.

Η υψηλή ταχύτητα του Elasticsearch οφείλεται στη χρήση της τεχνικής των ανεστραμμένων ευρετηρίων, τα οποία οργανώνουν τις πληροφορίες με τέτοιο τρόπο ώστε να μπορούν να ανακτηθούν αμέσως, ακόμη και όταν το σύστημα περιέχει εκατομμύρια εγγραφές. Η τεχνική αυτή επιτρέπει στις αναζητήσεις να εκτελούνται όχι μόνο κατά λέξη, αλλά και κατά φράση, συνδυαστικά ή ακόμη και με ασαφή κριτήρια, όπως η αναζήτηση ομοίων λέξεων. Επιπλέον, το Elasticsearch διαθέτει έναν ακόμη αποδοτικό μηχανισμό, τις αποθηκευμένες τιμές εγγράφων (doc values), που βελτιώνουν σημαντικά την απόδοση στις συγκεντρωτικές αναζητήσεις και στα στατιστικά ερωτήματα, καθώς αποθηκεύουν τις τιμές με τρόπο που επιτρέπει την ταχύτερη ταξινόμηση και ομαδοποίηση των δεδομένων.

Ένα σημαντικό πλεονέκτημα της πλατφόρμας είναι η ευελιξία στη διαχείριση πολλών διαφορετικών τύπων δεδομένων. Μπορεί να δεχτεί πληροφορίες από πλήθος πηγών, όπως αρχεία καταγραφής, αισθητήρες, εφαρμογές ή βάσεις δεδομένων, και να τα ενοποιήσει σε ένα κοινό σύστημα αναζήτησης. Η δυνατότητα αυτή, σε συνδυασμό με την ανθεκτικότητα του συστήματος σε βλάβες, καθιστά το Elasticsearch ιδανικό για εφαρμογές όπου η ταχύτητα και η αξιοπιστία είναι κρίσιμες, όπως για παράδειγμα στην ανάλυση δικτυακής δραστηριότητας, στη διαχείριση ασφάλειας ή στην παρακολούθηση υποδομών πληροφορικής.

Ωστόσο, η ευελιξία αυτή συνοδεύεται και από ορισμένες τεχνικές προκλήσεις. Η απόδοση του Elasticsearch κατά την εισαγωγή δεδομένων μπορεί να μειωθεί όταν υπάρχουν πολλές ταυτόχρονες ροές από διαφορετικές πηγές ή όταν τα δεδομένα έχουν ετερογενή μορφή. Η σωστή διαχείριση απαιτεί υπολογιστικούς πόρους, επαρκή μνήμη και προσεκτική παραμετροποίηση του συστήματος, ιδίως σε εγκαταστάσεις με μεγάλο όγκο δεδομένων. Επιπλέον, η συντήρηση και η λειτουργία μίας μεγάλης συστάδας υπολογιστών μπορεί να είναι σύνθετη, καθώς απαιτεί καλή κατανόηση των μηχανισμών κατανομής και συγχρονισμού των τμημάτων της βάσης, καθώς και παρακολούθηση της κατάστασης των κόμβων.

Ιδιαίτερη προσοχή χρειάζεται και στο θέμα της ασφάλειας. Αν και οι νεότερες εκδό-

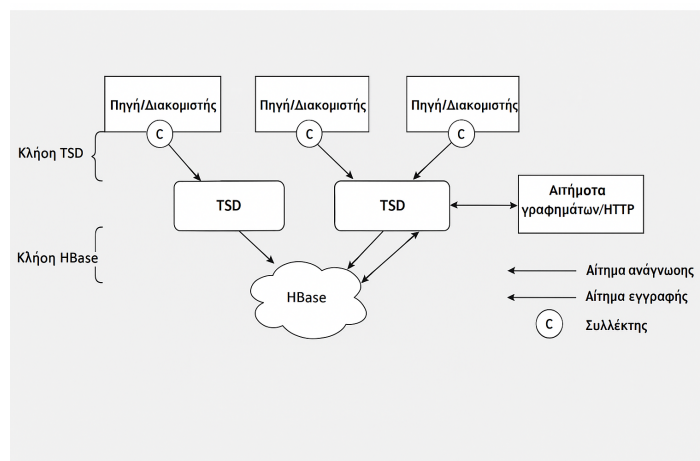
σεις του Elasticsearch προσφέρουν εργαλεία ελέγχου πρόσβασης και κρυπτογράφησης, η σωστή ρύθμισή τους απαιτεί εμπειρία, ώστε να αποτραπούν τυχόν διαρροές ή μην εξουσιοδοτημένες προσβάσεις. Επιπλέον, σε περιπτώσεις μικρών έργων ή πειραματικών εγκαταστάσεων, η χρήση του Elasticsearch μπορεί να θεωρηθεί υπερβολική, καθώς απαιτεί σημαντικούς πόρους σε σχέση με απλούστερες λύσεις. Παρόλα αυτά, για μεσαίες και μεγάλες εγκαταστάσεις, η απόδοσή του και η δυνατότητα επεξεργασίας δεδομένων σε πραγματικό χρόνο αντισταθμίζουν πλήρως αυτή την πολυπλοκότητα.

Συνοψίζοντας, το Elasticsearch αποτελεί μια ισχυρή και ευέλικτη πλατφόρμα που συνδυάζει αναζήτηση, ανάλυση και αποθήκευση δεδομένων με τρόπο που λίγες άλλες τεχνολογίες μπορούν να προσφέρουν. Η ταχύτητα, η επεκτασιμότητα και η ικανότητά του να επεξεργάζεται πολύπλοκα δεδομένα το καθιστούν εξαιρετικό εργαλείο για τη διαχείριση χρονοσειρών, την επεξεργασία καταγραφών και τη δημιουργία εφαρμογών ανάλυσης σε πραγματικό χρόνο, αρκεί να υποστηρίζεται από σωστό σχεδιασμό και επαρκή τεχνική υποδομή.

3.1.2.3 OpenTSDB

Το OpenTSDB (Open Time Series Database) [?] αποτελεί μία από τις πρώτες πλατφόρμες που σχεδιάστηκαν αποκλειστικά για τη διαχείριση και αποθήκευση δεδομένων χρονοσειρών σε μεγάλη κλίμακα. Αναπτύχθηκε από το ίδρυμα Apache πάνω στην τεχνολογία του αλγορίθμου HBase, με κύριο στόχο να προσφέρει μία αξιόπιστη και αποδοτική λύση για την αποθήκευση τεράστιου όγκου δεδομένων που συλλέγονται από διάφορες πηγές, όπως αισθητήρες, μετρητές ή συστήματα παρακολούθησης. Το OpenTSDB συνδυάζει την ευκολία χρήσης μιας βάσης δεδομένων υψηλού επιπέδου με τη δύναμη ενός κατανεμημένου συστήματος, παρέχοντας εξαιρετικά γρήγορες λειτουργίες εγγραφής και ανάγνωσης δεδομένων.

Η αρχιτεκτονική του βασίζεται στην αποθήκευση των δεδομένων σε τέσσερις κύριους πίνακες του HBase, γνωστούς ως TSDB, TSDB-UID, TSDB-META και TSDB-TREE. Η επικοινωνία με αυτούς τους πίνακες πραγματοποιείται μέσω των λεγόμενων διεργασιών χρονοσειρών (TSD), οι οποίοι είναι οι βασικές διεργασίες της πλατφόρμας, όπως φαίνεται και στο Σχήμα 3.1. Κάθε τέτοια διεργασία λειτουργεί ανεξάρτητα, χωρίς να



Σχήμα 3.1: Αρχιτεκτονική του OpenTSDB

απαιτείται κάποιος κεντρικός κόμβος ελέγχου ή κοινόχρηστη διεργασία. Αυτή η αποκεντρωμένη δομή επιτρέπει την οριζόντια επεκτασιμότητα του συστήματος και εξασφαλίζει ότι η πλατφόρμα μπορεί να συνεχίσει να λειτουργεί ακόμη και αν κάποιοι κόμβοι τεθούν εκτός λειτουργίας. Επιπλέον, η δομή του σχήματος δεδομένων είναι σχεδιασμένη ώστε να ελαχιστοποιεί τον χώρο που καταλαμβάνουν τα δεδομένα και να επιτρέπει ταχύτατα συγκεντρωτικές λειτουργίες (aggregations), κάτι που είναι ιδιαίτερα χρήσιμο σε εφαρμογές που απαιτούν ανάλυση μεγάλων χρονικών σειρών.

Οι χρήστες δε χρειάζεται να έρθουν σε άμεση επαφή με τη δομή των πινάκων του HBase. Η πρόσβαση στα δεδομένα γίνεται μέσω διεπαφών φιλικών προς τον χρήστη, όπως μια διεπαφή προγραμματισμού εφαρμογών μέσω HTTP (HTTP API) ή ένα απλό πρωτόκολλο τύπου telnet, ενώ διατίθεται και γραφικό περιβάλλον για οπτικοποιημένη διαχείριση. Ο μηχανισμός συλλογής δεδομένων είναι υπεύθυνος για τη συγκέντρωση των πληροφοριών από τις πηγές και τη μεταφορά τους στις διεργασίες χρονοσειρών, οι οποίες αναλαμβάνουν τη διαδικασία εγγραφής στη βάση. Όλες οι επικοινωνίες πραγματοποιούνται μέσω της ίδιας θύρας, καθώς οι διεργασίες αυτές μπορούν να αναγνωρίσουν αυτόματα το πρωτόκολλο του πελάτη από τα πρώτα δεδομένα που λαμβάνουν.

Ένα από τα βασικά χαρακτηριστικά του OpenTSDB είναι η ικανότητά του να επεκτείνεται πρακτικά χωρίς όριο. Μπορεί να επωφεληθεί από τη δυνατότητα κατακεντρωμένης αποθήκευσης και να διαχειριστεί τεράστιους όγκους δεδομένων από εκατοντάδες ή και χιλιάδες διαφορετικές πηγές. Παράλληλα, η αρχιτεκτονική του επιτρέπει υψηλές ταχύτη-

τες ανάγνωσης και εγγραφής, κάτι που το καθιστά ιδανικό για εφαρμογές όπου οι ρυθμοί ροής δεδομένων είναι εξαιρετικά μεγάλοι, όπως στη βιομηχανική παρακολούθηση, στα συστήματα IoT ή στην ανάλυση επιδόσεων υπολογιστικών υποδομών. Η ευελιξία στη διαχείριση του σχήματος των δεδομένων επιτρέπει την εύκολη προσθήκη νέων τύπων μετρήσεων χωρίς να απαιτούνται αλλαγές στη βασική δομή της βάσης, προσφέροντας ένα ιδιαίτερα προσαρμόσιμο περιβάλλον.

Το OpenTSDB υποστηρίζει δεδομένα που φέρουν χρονική σήμανση, κάτι που το καθιστά ιδανικό για την αποθήκευση και επεξεργασία δεδομένων χρονοσειρών. Κάθε μέτρηση συνοδεύεται από χρονική σφραγίδα και ετικέτες, οι οποίες επιτρέπουν την ταξινόμηση και την εύκολη ανάκτηση των δεδομένων. Αυτή η δομή διευκολύνει τις συγκεντρωτικές αναλύσεις, καθώς τα δεδομένα μπορούν να ομαδοποιηθούν ή να φιλτραριστούν με βάση τις ετικέτες, επιτρέποντας την εξαγωγή πολύτιμων συμπερασμάτων σε σύντομο χρόνο. Παράλληλα, η υποστήριξη πολλών διαφορετικών τύπων δεδομένων καθιστά το OpenTSDB κατάλληλο για σύνθετες εφαρμογές, όπου απαιτείται η ανάλυση μετρήσεων από διαφορετικές πηγές.

Ωστόσο, αν και εξαιρετικά αποδοτικό στη συλλογή και αποθήκευση δεδομένων, παρουσιάζει κάποιους περιορισμούς στις δυνατότητες ανάλυσης. Οι ενσωματωμένες λειτουργίες του για αναζητήσεις και προβλέψεις είναι σχετικά βασικές, ενώ για πιο σύνθετες αναλύσεις απαιτείται η χρήση εξωτερικών εργαλείων ή η ανάπτυξη πρόσθετου κώδικα. Επιπλέον, η απουσία προηγμένων εργαλείων οπτικοποίησης καθιστά απαραίτητη τη διασύνδεση του OpenTSDB με άλλες πλατφόρμες, όπως το Grafana, προκειμένου να παρουσιαστούν τα δεδομένα με πιο κατανοητό και δυναμικό τρόπο.

Ακόμη, παρόλο που διαθέτει μια ενεργή κοινότητα χρηστών και προγραμματιστών, το οικοσύστημα του OpenTSDB δεν είναι τόσο εκτεταμένο όσο εκείνο άλλων βάσεων δεδομένων χρονοσειρών, όπως του InfluxDB ή του TimescaleDB. Αυτό σημαίνει ότι η διαθεσιμότητα βιβλιοθηκών, παραδειγμάτων ή ενσωματωμένων λύσεων μπορεί να είναι περιορισμένη, καθιστώντας την αρχική εκμάθηση και εγκατάσταση πιο απαιτητική, ειδικά για νέους χρήστες. Παρά ταύτα, η σταθερότητα, η αξιοπιστία και η δυνατότητα επεξεργασίας τεράστιων ποσοτήτων δεδομένων αντισταθμίζουν σε μεγάλο βαθμό αυτή την πολυπλοκότητα.

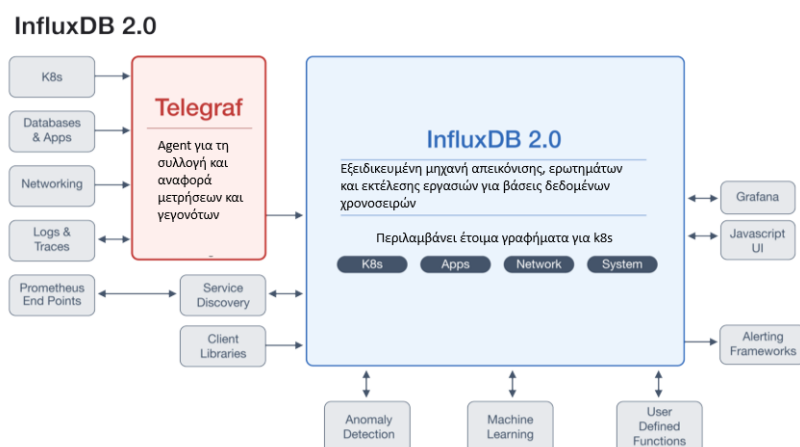
Συνολικά, το OpenTSDB αποτελεί μια ώριμη και σταθερή λύση για την αποθήκευση και διαχείριση δεδομένων χρονοσειρών σε περιβάλλοντα μεγάλης κλίμακας. Ο συνδυασμός της ισχυρής αρχιτεκτονικής του, της ευκολίας επεκτασιμότητας και της υψηλής ταχύτητάς του, το καθιστά ιδιαίτερα ελκυστικό για οργανισμούς που χρειάζονται αξιόπιστη και αποδοτική διαχείριση δεδομένων σε πραγματικό χρόνο. Αν και η έλλειψη εξελιγμένων εργαλείων ανάλυσης και οπτικοποίησης μπορεί να θεωρηθεί μειονέκτημα, η συνολική του απόδοση και η σταθερότητα που παρέχει το καθιστούν μία από τις πιο αξιόπιστες πλατφόρμες για την επεξεργασία χρονοσειρών σε βιομηχανικό ή επιχειρησιακό επίπεδο.

3.2 Εξειδικευμένα συστήματα

Σε αυτή την ενότητα εξετάζουμε εξειδικευμένες βάσεις δεδομένων χρονοσειρών (TSDBs), δηλαδή συστήματα που έχουν σχεδιαστεί εξ αρχής για υψηλούς ρυθμούς εγγραφής μετρήσεων, αποδοτική συμπίεση, πολιτικές διατήρησης και αρχικοποίησης, καθώς και ταχεία ανάκτηση με βάση ετικέτες και προκαθορισμένα ευρετήρια. Εστιάζουμε σε δύο αντιπροσωπευτικές προσεγγίσεις: την InfluxDB (η οποία λειτουργεί με δικό της μοντέλο δεδομένων και μηχανή αποθήκευσης) και την TimescaleDB (αποτελεί επέκταση της PostgreSQL, ειδικά κατασκευασμένη για χρονοσειρές). Για καθεμία παρουσιάζονται αρχιτεκτονική, βασικές έννοιες δεδομένων, μηχανισμοί κλιμάκωσης και συμπίεσης, συνήθεις περιπτώσεις χρήσης, πλεονεκτήματα και περιορισμοί. Η σύγκριση τους γίνεται με κριτήρια που είναι κρίσιμα για αγρομετεωρολογικά δεδομένα: ρυθμός εγγραφής, καθυστέρηση ερωτημάτων, λόγος συμπίεσης, ευκολία ενσωμάτωσης δεδομένων καθώς και πρακτικές λειτουργικές δυνατότητες (παρακολούθηση δεδομένων κ.α.).

3.2.1 InfluxDB

Η InfluxDB [?] είναι μια ανοικτού κώδικα μη σχεσιακή βάση δεδομένων που δημιουργήθηκε αποκλειστικά για τον σκοπό της αποθήκευσης χρονοσειρών. Είναι γραμμένη σε γλώσσα προγραμματισμού GO και έχει κατασκευαστεί ώστε να μπορεί να διαχειριστεί τα δεδομένα χρονοσειρών με το βέλτιστο τρόπο. Παρέχει μία γλώσσα προγραμματισμού



Σχήμα 3.2: Αρχιτεκτονική δομή InfluxDB

παρόμοια με τη SQL. Στην έκδοση ανοικτού κώδικα παρέχει μία πλατφόρμα δεδομένων χρονοσειρών με διάφορες υπηρεσίες συμπεριλαμβανομένου το InfluxDB Core και μπορεί να εγκατασταθεί είτε στο υπολογιστικό νέφος (cloud) είτε σε ένα μοναδικό υπολογιστικό κόμβο. Οι εκδόσεις που δεν είναι ανοικτού κώδικα, οι οποίες ονομάζονται Influx Enterprise και Influx Cloud, προσφέρουν περαιτέρω λειτουργίες και ικανότητες όπως υψηλή διαθεσιμότητα, υπηρεσίες δημιουργίας και επαναφοράς αντιγράφων ασφαλείας, υψηλή επεκτασιμότητα και μπορούν να εγκατασταθούν σε υποδομές υπολογιστικού νέφους ή σε μεμονωμένους κόμβους. Στο Σχήμα 3.2 απεικονίζεται η σύνοψη της βασικής αρχιτεκτονικής.

3.2.1.1 Βασικά σημεία

Η σημαντικότερη μεταβλητή στο InfluxDB είναι ο χρόνος [?]. Κάθε βάση δεδομένων στην πλατφόρμα περιλαμβάνει υποχρεωτικά μια στήλη με την ονομασία χρόνος, η οποία κρατάει διακριτά χρονικά σημεία που αντιστοιχούν σε συγκεκριμένα δεδομένα. Τα πεδία δεδομένων αποτελούν τα ρεαλιστικά δεδομένα των χρονοσειρών κι έχει πάντα μία χρονοσήμανση συνδεδεμένη μαζί τους. Η κάθε τιμή ενός πεδίου ονομάζεται τιμή πεδίου και το μοναδικό της στοιχείο πάντα συνοδεύεται από μία χρονοσήμανση. Το ζεύγος κλειδί πεδίου - τιμή πεδίου ονομάζεται σετ πεδίου. Το πού ανήκουν αυτά τα σύνολα, καθώς και οι σημάνσεις που λειτουργούν ως φίλτρα στα ερωτήματα SQL, ονομάζονται ετικέτες. Κάθε ετικέτα αποτελείται από ένα κλειδί και μία τιμή. Το πλήρες σύνολο από

κλειδί - τιμή ετικέτας ονομάζεται σετ ετικέτας. Η μεγάλη διαφορά των ετικετών από τα πεδία είναι ότι οι ετικέτες γίνονται ευρετήρια. Αυτό έχει ως αποτέλεσμα τα ερωτήματα που βασίζονται σε ετικέτες να είναι πολύ πιο γρήγορα συγκριτικά με τα ερωτήματα που βασίζονται σε πεδία. Επιπλέον, το πρωτεύον κλειδί κάθε εγγραφής στην InfluxDB αποτελείται από τη χρονοσήμανση και ένα σετ ετικέτας. Ένας ακόμη σημαντικός όρος στο InfluxDB είναι η μέτρηση, που καθορίζει το είδος ή την κατηγορία των δεδομένων που περιέχονται σε ένα πεδίο. Μία μέτρηση μπορεί να συσχετίζεται με μία ή περισσότερες πολιτικές διατήρησης. Οι πολιτικές διατήρησης καθορίζουν για πόσο χρονικό διάστημα θα διατηρούνται τα δεδομένα στην InfluxDB, καθώς και πόσα αντίγραφα των δεδομένων θα αποθηκεύονται στο σύμπλεγμα κόμβων.

Η προκαθορισμένη πολιτική διατήρησης, γνωστή ως *autogen*, κρατά τα δεδομένα για πάντα χωρίς να δημιουργεί αντίγραφα. Τέλος, η πιο σημαντική έννοια στην InfluxDB είναι η έννοια της σειράς. Μία σειρά είναι ένα σύνολο δεδομένων που έχουν την ίδια πολιτική διατήρησης, μέτρηση και σετ ετικέτας. Κάθε εγγραφή σε μια συγκεκριμένη σειρά δεδομένων διαθέτει τη δική της χρονοσφραγίδα, η οποία υποδεικνύει τη χρονική στιγμή συλλογής ή εισαγωγής του δεδομένου.

3.2.1.2 Κατακερματισμός δεδομένων

Ο κατακερματισμός δεδομένων είναι η διαδικασία κατάτμησης των δεδομένων σε μικρότερα τμήματα οριζόντια, έτσι ώστε να είναι ευκολότερα στη διαχείριση. Τα τμήματα αυτά ονομάζονται τεμάχια ή κατακερματισμένα τμήματα. Στο InfluxDB, τα δεδομένα αποθηκεύονται σε ομάδες κατακερματισμένων τμημάτων, οι οποίες ονομάζονται ομάδες τεμαχίων. Η οργάνωση αυτών των ομάδων γίνεται με βάση την πολιτική διατήρησης και περιλαμβάνουν χρονικά σημεία για ένα συγκεκριμένο χρονικό διάστημα. Οι προκαθορισμένες διάρκειες των ομάδων τμημάτων στην InfluxDB καθορίζονται παρακάτω.

1 ώρα, όταν η πολιτική διατήρησης είναι μικρότερη των 2 ημερών.

1 ημέρα, όταν η πολιτική διατήρησης κυμαίνεται από 2 ημέρες έως 6 μήνες.

7 ημέρες, όταν η πολιτική διατήρησης υπερβαίνει τους 6 μήνες.

Η κατάλληλη ρύθμιση της διάρκειας των ομάδων τεμαχίων αποτελεί κρίσιμο παράγοντα για την αποδοτική λειτουργία και τη συνολική απόδοση του συστήματος αποθήκευσης, καθώς καθορίζει τον ρυθμό εγγραφής, ανάκτησης και καθαρισμού των δεδομένων.

3.2.1.3 Μηχανή αποθήκευσης

Το InfluxDB χρησιμοποιεί μια ενσωματωμένη δομή δεδομένων που ονομάζεται δέντρο συγχώνευσης με χρονική σειρά (TSM Tree). Ουσιαστικά, αυτή η δομή έχει σχεδιαστεί ειδικά για την αποθήκευση δεδομένων που ακολουθούν μια χρονολογική σειρά, και αποτελεί το βασικό σύστημα αποθήκευσης στις πιο πρόσφατες εκδόσεις της πλατφόρμας. Σημαντικό επίσης είναι το γεγονός ότι η InfluxDB αρχικά στηρίχθηκε στο σύστημα LevelDB — ένα σύστημα διαχείρισης βάσης δεδομένων που βασίζεται σε δέντρα συγχώνευσης αρχείων. Το LevelDB χρησιμοποιήθηκε κυρίως στις πρώτες εκδόσεις της InfluxDB ως η κύρια μέθοδος αποθήκευσης των δεδομένων. Παρόλα αυτά, λόγω των περιορισμών του συγκεκριμένου συστήματος — ιδιαίτερα της ανεπαρκούς υποστήριξης για δημιουργία αντιγράφων ασφαλείας — η InfluxDB άρχισε σταδιακά να χρησιμοποιεί άλλες, πιο αποδοτικές λύσεις, όπως το RocksDB και το HyperLevelDB. Οι νέες αυτές δομές προσέφεραν καλύτερες επιδόσεις και τους επέτρεψαν να αποθηκεύουν και να διαχειρίζονται με αποτελεσματικότερο τρόπο μεγάλο όγκο δεδομένων που ακολουθούν χρονολογική σειρά.

3.2.1.4 Περιπτώσεις χρήσης του InfluxDB

Το InfluxDB έχει ξεπεράσει τις 200.000 ενεργές εγκαταστάσεις και αποτελεί ένα από τα πιο δημοφιλή εργαλεία για παρακολούθηση και ανάλυση συστημάτων. Χρησιμοποιείται ευρέως για παρακολούθηση υποδομών, εφαρμογών, υπηρεσιών υπολογιστικού νέφους, συσκευών IoT, καθώς και για ανάλυση δεδομένων σε πραγματικό χρόνο. Μεταξύ των πελατών του συγκαταλέγονται οι εταιρείες AXA, Cisco, eBay και IBM.

Εποπτεία διαδικασιών DevOps. Τα προϊόντα της IBM® Trusteer® προσφέρουν δυνατότητες ανίχνευσης και παρεμπόδισης επιθέσεων σε ένα ευρύ φάσμα περιπτώσεων, που περιλαμβάνουν διαδικτυακές εφαρμογές, εφαρμογές κινητών συσκευών,

καθώς και περιβάλλοντα με πολλά κανάλια επικοινωνίας. Για τον σκοπό αυτό, η πλατφόρμα Trusteer πρέπει να διατηρεί υψηλή διαθεσιμότητα. Η ομάδα που είναι υπεύθυνη για την προσέγγιση ανάπτυξης λειτουργιών (DevOps) χρησιμοποιεί τα εργαλεία της InfluxDB. Κυριότερα εργαλεία είναι το Telegraf, ένα ισχυρό προϊόν του οικοσυστήματος του InfluxData και το Grafana, ένα λογισμικό ανοικτού κώδικα, κατάλληλο για ανάλυση χρονοσειρών. Πιο συγκεκριμένα, το Telegraf χρησιμοποιείται για τη συλλογή δεδομένων, το InfluxDB για την αποθήκευσή του και το Grafana για την ανάλυση και την οπτικοποίησή τους.

Εποπτεία IoT συστημάτων. Ένα παράδειγμα παρακολούθησης IoT είναι η περίπτωση του Spiio [?]. Το συγκεκριμένο λογισμικό επιτρέπει την παρακολούθηση πράσινων τοίχων κατακόρυφης φύτευσης και θερμοκηπίων υψηλών προδιαγραφών, μέσω ενός δικτύου αισθητήρων που συλλέγουν και αποστέλλουν πληροφορίες στους υπεύθυνους κηπουρούς. Οι πράσινοι τοίχοι γνωρίζουν ολοένα και μεγαλύτερη διάδοση, ιδιαίτερα στα αστικά κέντρα, καθώς φέρνουν τη φύση πιο κοντά στην καθημερινή ζωή της πόλης. Ωστόσο, η συντήρηση πολλών τέτοιων εγκαταστάσεων μέσα σε μία πόλη μπορεί να αποδειχθεί ιδιαίτερα απαιτητική για τους επαγγελματίες του κλάδου. Σε αυτό το σημείο επεμβαίνει το λογισμικό, αξιοποιώντας τα δεδομένα των αισθητήρων για να αξιολογεί την απόδοση των εγκαταστάσεων και να εντοπίζει έγκαιρα περιπτώσεις που απαιτούν διορθωτικές ενέργειες. Το Spiio προσπάθησε να χρησιμοποιήσει διάφορες λύσεις πριν το InfluxDB, συγκεκριμένα σε πλατφόρμες IoT, όπως είναι τα AWS Greengrass και Azure IoT, σχεσιακές βάσεις δεδομένων (MySQL, Cassandra, Elasticsearch), ακόμα και βάσεις δεδομένων χρονοσειρών, όπως το OpenTSDB. Ωστόσο, καμία λύση δεν πρόσφερε την ολιστική προσέγγιση του InfluxData. Αυτή τη στιγμή το Spiio χρησιμοποιεί το InfluxDB για την αποθήκευση δεδομένων, το Kapacitor για αναπαραγωγή των δεδομένων σε πραγματικό χρόνο και το Chronograf για την παρουσίασή τους στο χρήστη. Και τα δύο προγράμματα (Kapacitor, Chronograf) είναι μέρος του οικοσυστήματος InfluxData, τα οποία παρέχουν στους πελάτες του Spiio τη δυνατότητα πρόσβασης, παρακολούθησης αλλά και διαμοιρασμού των δεδομένων των αισθητήρων.

Δεδομένα ανάλυσης πραγματικού χρόνου. Το πιο χαρακτηριστικό παράδειγμα αυτής της κατηγορίας, είναι εκείνο του Ebay (<https://www.ebay.com>). Πολλές ομάδες της συγκεκριμένης εταιρείας χρησιμοποιούν το οικοσύστημα των InfluxDB και InfluxData κυρίως για τις δυνατότητες αυτοματοποίησης διαδικασιών DevOps και ανάλυσης δεδομένων σε πραγματικό χρόνο. Πιο συγκεκριμένα, η ομάδα δοκιμών και πειραματισμών της εταιρείας Ebay παρέχει σημαντικές αναλύσεις και διαπιστώσεις, σχετικά με κρίσιμα εμπορικά ερωτήματα. Ωστόσο, τα πειράματα που διεξάγουν για να εξάγουν αυτά τα συμπεράσματα συχνά αντιμετωπίζουν δυσκολίες, όπως είναι η αλλοίωση της ροής δεδομένων, και εδώ είναι που επεμβαίνουν άμεσα οι δυνατότητες του InfluxDB. Οι αλλοιώσεις αυτές εντοπίζονται καθημερινά χρησιμοποιώντας τους αλγορίθμους Anomaly/Traffic Prediction, αποθηκεύονται στο InfluxDB και απεικονίζονται με τη βοήθεια του περιβάλλοντος απεικόνισης, Grafana. Η ανάλυση αυτών των δεδομένων σε μορφή χρονοσειρών είναι το κλειδί για τη δυνατότητα απεικόνισης τους στο Grafana Dashboard. Ο συνδυασμός αυτός επιτρέπει στη συγκεκριμένη ομάδα να λειτουργεί κλιμακωτά και να διαθέτει αυτο-επάρκεια σε κάθε απαίτηση.

3.2.1.5 Πλεονεκτήματα και μειονεκτήματα του InfluxDB

Παρακάτω αναλύουμε τα βασικά πλεονεκτήματα.

Προσαρμοσμένο για δεδομένα χρονοσειρών. Το InfluxDB είναι σχεδιασμένο να χειρίζεται το συγκεκριμένο τύπο δεδομένων ώστε να είναι πιο αποδοτικό, με προηγμένη εγγραφή και ανάγνωση των δεδομένων.

Προσφέρει λύσεις για κάθε ανάγκη. Το λογισμικό ως υπηρεσία (software as a service – SaaS) είναι ένα μοντέλο διάθεσης λογισμικού στο οποίο η εφαρμογή εκτελείται σε υποδομές του παρόχου και προσφέρεται στους χρήστες ως υπηρεσία μέσω διαδικτύου, εξαλείφοντας την ανάγκη για τοπική εγκατάσταση ή διαχείριση. Ο πυρήνας του InfluxDB είναι τύπου ανοικτού κώδικα και περιλαμβάνει τα InfluxDB, Kapacitor, Telegraf και Chronograf, δωρεάν. Παρέχει επίσης και μία SaaS πρόταση, που ονομάζεται Influx Cloud και προσφέρει υψηλή δια-

θεσιμότητα, προηγμένες λειτουργίες ασφαλούς ανάκτησης δεδομένων καθώς και λειτουργίες επαναφοράς δεδομένων, για χρήστες που έχουν μεγαλύτερες ανάγκες και μικρότερες δυνατότητες φυσικών εγκαταστάσεων. Η συγκεκριμένη πρόταση έχει διακομιστές σε Αμερική, Καναδά και Ευρώπη. Υπάρχει επίσης και μία πρόταση που λέγεται Influx Enterprise και απευθύνεται σε πελάτες που θέλουν να κάνουν χρήση των υποδομών τους και των υπηρεσιών στο υπολογιστικό νέφος.

Ολιστική λύση. Το InfluxDB σχεδιάστηκε να λειτουργεί άψογα σε συνδυασμό με τα υπόλοιπα προϊόντα του InfluxData οικοσυστήματος, όπως είναι τα Kapacitor, Telegraf και Chronograf. Υπό αυτό το πρίσμα, παρέχει πολύ περισσότερες δυνατότητες απ'ότι μια απλή βάση δεδομένων. Πρόκειται για μια ολιστική λύση που παρέχει δυνατότητες απόκτησης, αποθήκευσης, ανάλυσης και παρουσίασης των δεδομένων μας.

Διαφορετικές επεκτάσεις. Οι δυνατότητες εισαγωγής δεδομένων δεν περιορίζονται σε μία - δύο μεθόδους όπως συμβαίνει σε άλλες βάσεις δεδομένων χρονοσειρών. Εδώ προσφέρεται πληθώρα διαθέσιμων επεκτάσεων λογισμικού και μάλιστα δωρεάν. Πέρα από το κλασικό HTTP API προσφέρονται επεκτάσεις όπως τα UDP, Graphite (επιτρέπει εισαγωγή για το γραμμικό πρωτόκολλο Graphite), CollectD, OpenTSDB (είτε τα δεδομένα εισαχθούν μέσω Telnet είτε μέσω του OpenTSDB). Αυτό πρακτικά σημαίνει ότι το InfluxDB μπορεί να αντικαταστήσει ή και να λειτουργήσει παράλληλα με ένα σύστημα OpenTSDB.

Υποστήριξη του Grafana. Πρόκειται για το πλέον κατάλληλο εργαλείο ανάλυσης και απεικόνισης χρονοσειρών. Διαθέτει την ενσωματωμένη δυνατότητα να δέχεται ως πηγή δεδομένων το InfluxDB, και χρησιμοποιείται κυρίως για τη διαχείριση του κεντρικού πάνελ, στο γραφικό του περιβάλλον.

Υποστήριξη για πολλές γλώσσες προγραμματισμού. Το InfluxDB υποστηρίζει πολλές γλώσσες προγραμματισμού, μεταξύ των οποίων είναι .Net, Java, Perl, PHP, Python, R, Ruby, Scala κ.α.

Ερωτήματα παρόμοια με SQL. Η γλώσσα διαχείρισης δεδομένων του InfluxDB

μοιάζει πολύ με τη SQL, κάνοντας εύκολη την εκμάθησή της από τον απλό χρήστη που έχει χρησιμοποιήσει στο παρελθόν Mysql, PostgreSQL, SQL Server ή SQLite.

Εύκολη εγκατάσταση. Μεταγλωττίζοντας ένα μόνο εκτελέσιμο αρχείο χωρίς καμία εξάρτηση από εξωτερικά πακέτα (Linux ή Windows) έχουμε έτοιμο το διακομιστή του InfluxDB.

Απεριόριστα πεδία. Εδώ να τονίσουμε ότι και άλλοι τρόποι διαχείρισης χρονοσειρών μπορούν να προσθέσουν πεδία χωρίς περιορισμό, αλλά η διαφορά είναι ότι εδώ ο αριθμός των πεδίων δεν έχει αντίκτυπο στις επιδόσεις των ερωτημάτων, άρα δεν περιορίζονται με κανένα τρόπο.

Προεγκατεστημένο περιβάλλον διαχείρισης. Όπως και στην SQL, το InfluxDB παρέχει ένα όμορφο γραφικό περιβάλλον, κυρίως για τους χρήστες που δεν έχουν εξοικείωση με τη γραμμή εντολών.

Εκτεταμένη βιβλιογραφία. Η διαδικτυακή πλατφόρμα του InfluxDB παρέχει αναλυτική βιβλιογραφία για όλες τις πιθανές ενέργειες, από τη βασική εγκατάσταση έως και σύνθετα ερωτήματα προς τη βάση.

Παρακάτω παρουσιάζουμε τα βασικά μειονεκτήματα.

Περιορισμένη επεκτασιμότητα. Το InfluxDB δυστυχώς ανακοίνωσε από την αρχή της έκδοσής του ότι η δυνατότητα της συσταδοποίησης (clustering) δε θα περιλαμβάνεται στην ανοικτού κώδικα έκδοση της πλατφόρμας. Αυτό έχει ως αποτέλεσμα οι χρήστες να αναζητούν εναλλακτικούς τρόπους βελτίωσης της επεκτασιμότητας, όπως είναι η τμηματοποίηση (sharding). Τμηματοποίηση σημαίνει τον διαχωρισμό των δεδομένων σε μικρότερα, ανεξάρτητα τμήματα (shards), τα οποία αποθηκεύονται ή διαχειρίζονται σε διαφορετικούς κόμβους ή βάσεις δεδομένων, έτσι ώστε να κατανέμεται το φορτίο και να επιτυγχάνεται καλύτερη απόδοση. Ωστόσο, επειδή το InfluxDB δεν προσφέρει ενσωματωμένη υποστήριξη για συσταδοποίηση, η χρήση της τμηματοποίησης απαιτεί πρόσθετο σχεδιασμό και διαχείριση, κάτι που δημιουργεί δυσκολίες, ιδιαίτερα για νεοφυείς εταιρείες. Παρ’

όλα αυτά, το κόστος δεν είναι απαγορευτικό, αφού ξεκινά από 149 δολάρια τον μήνα για την εκτός σύνδεσης στο διαδίκτυο πλατφόρμα και φτάνει τα 249 δολάρια τον μήνα για την έκδοση που βρίσκεται στο υπολογιστικό νέφος.

Προβλήματα της κοινότητας InfluxDB. Δεδομένου του γεγονότος ότι πρόκειται για νέα πλατφόρμα, η κοινότητά της είναι ιδιαίτερα μικρή σε σύγκριση για παράδειγμα με την πλατφόρμα της Cassandra. Αυτό σημαίνει ότι μία απλή αναζήτηση στην πλατφόρμα του Google πιθανότατα δε θα μας παρέχει τη λύση στο πρόβλημα που ψάχνουμε. Ωστόσο, η δημοτικότητα της πλατφόρμας αυξάνει συνεχώς και αναμένεται να μεγαλώσει πολύ στο εγγύς μέλλον.

3.2.2 TimescaleDB

Το TimescaleDB είναι μια επέκταση της PostgreSQL [?], από την οποία κληρονομεί όλα τα θετικά χαρακτηριστικά μιας σχεσιακής βάσης δεδομένων, προσθέτοντας μια δομή δεδομένων κατάλληλη για χρονοσειρές. Η δομή αυτή είναι σχεδιασμένη ώστε να βελτιστοποιείται η απόδοση ανάκτησης και επεξεργασίας των δεδομένων αυτού του τύπου. Επιπλέον, λειτουργούν αυτόματα μηχανισμοί συμπίεσης για τη μείωση του χώρου αποθήκευσης και τη βελτίωση των ερωτημάτων του χρήστη. Τέλος, υπάρχει ευελιξία στο τρόπο που αποθηκεύονται τα δεδομένα, δίνοντας και τη δυνατότητα προσαρμογής των σχημάτων των δεδομένων.

3.2.2.1 Κατανεμημένη αρχιτεκτονική

Το TimescaleDB [?] είναι σχεδιασμένο για να λειτουργεί σε ένα δίκτυο από κόμβους, επιτρέποντας την αποτελεσματική και κλιμακούμενη διαχείριση μεγάλων όγκων δεδομένων χρονοσειρών. Ως αποτέλεσμα, έχουμε τη δυνατότητα διεύρυνσης της βάσης δεδομένων χρονοσειρών μας, ανάλογα με τις ανάγκες που προκύπτουν όταν για παράδειγμα αυξάνουν οι σταθμοί παροχής δεδομένων. Η διάσπαση των δεδομένων γίνεται διαιρώντας σε τμήματα (chunks) με βάση το χρόνο, επιτρέποντας την αποθήκευσή τους σε διαφορετικούς κόμβους. Στους κόμβους αυτούς αναπαριστώνται και αντιγράφονται τα δεδομένα ως διπλοεγγραφές, με σκοπό την αντοχή σε αποτυχίες και επιτυγχάνοντας

αξιόπιστη λειτουργία.

3.2.2.2 Περιπτώσεις χρήσης TimescaleDB

Το TimescaleDB χρησιμοποιείται κατά κύριο λόγο σε εφαρμογές διαδικτύου των πραγμάτων (IoT) [?], όπως έξυπνα σπίτια, βιομηχανικοί αισθητήρες ή συστήματα παρακολούθησης. Τα δεδομένα που παράγονται από αισθητήρες αποθηκεύονται σε πραγματικό χρόνο, με αποτέλεσμα να μπορούν να παρουσιαστούν επίσης σε πραγματικό χρόνο. Στον τομέα των πληροφοριακών συστημάτων, μπορεί να χρησιμοποιηθεί για την παρακολούθηση της απόδοσης συστημάτων, όπως δικτυακά συστήματα, servers ή εφαρμογές. Επιπλέον, χρησιμοποιείται για συστήματα καταγραφής και ανάλυσης δεδομένων καθώς και σε όλες τις εφαρμογές που η δομή τους ακολουθεί το μοτίβο των δεδομένων χρονοσειράς.

3.2.2.3 Πλεονεκτήματα και μειονεκτήματα

Παρακάτω αναφέρονται τα βασικότερα πλεονεκτήματα του TimescaleDB.

Εξειδίκευση σε δεδομένα χρονοσειρών. Είναι σχεδιασμένο ειδικά για τη διαχείριση μεγάλου όγκου χρονοσειριακών δεδομένων, προσφέροντας βελτιστοποιημένες επιδόσεις για αυτού του είδους τα δεδομένα.

Συμβατότητα με PostgreSQL. Μας δίνει τη δυνατότητα να κάνουμε μια μεικτή βάση δεδομένων, με δεδομένα χρονοσειρών και σχεσιακά δεδομένα.

Ευελιξία και επεκτασιμότητα. Μπορεί να επεκταθεί εύκολα για να ανταποκριθεί σε αυξημένες ανάγκες χωρητικότητας ή απόδοσης.

Αυτοματοποιημένη διαχείριση χώρου αποθήκευσης. Όπως και στο InfluxDB, έχουμε και εδώ τη δυνατότητα να διαχειρίζεται αυτόματα τη συμπίεση δεδομένων και να βελτιστοποιεί τον χώρο αποθήκευσης.

Ως προς τα μειονεκτήματα χρήσης του TimescaleDB, αναφέρονται παρακάτω τα βασικότερα.

Εξοικείωση με τα χαρακτηριστικά του. Απαιτείται χρόνος και ιδιαίτερη εκπαίδευση για την εκμάθηση σωστής χρήσης όλων των δυνατοτήτων του TimescaleDB.

Μειωμένη υποστήριξη για άλλους τύπους δεδομένων. Σε περίπτωση που θέλουμε να εισάγουμε και δεδομένα τα οποία δεν ακολουθούν το μοτίβο των χρονοσειρών, θα πρέπει να στραφούμε στην PostgreSQL, αφού οι δυνατότητες του TimescaleDB είναι πολύ περιορισμένες σε τέτοιες περιπτώσεις.

Διαχείριση και κατανόηση του αποθηκευτικού χώρου. Σε μεγάλης κλίμακας δεδομένα, η διαχείριση του χώρου αποθήκευσης μπορεί να απαιτεί προσεκτική παρακολούθηση και επεμβάσεις ώστε να αποδίδει το TimescaleDB τις μέγιστες επιδόσεις του.

3.3 Σύγκριση υλοποιήσεων

Σε αυτήν την ενότητα πραγματοποιούμε μια στοχευμένη σύγκριση των βασικών προσεγγίσεων αποθήκευσης χρονοσειρών, με έμφαση τόσο στις παραδοσιακές σχεσιακές βάσεις (MySQL, PostgreSQL, Elasticsearch) όσο και μεταξύ των εξειδικευμένων πλατφορμών για δεδομένα χρονοσειρών InfluxDB και TimescaleDB. Η αξιολόγηση γίνεται με βάση το σύνολο των αγρομετεωρολογικών δεδομένων με τη χρήση ισοδύναμων σχημάτων, ώστε να διασφαλίζεται όσο το δυνατόν η ορθή σύγκριση. Τα κύρια κριτήρια είναι: ρυθμός εισαγωγής (ingestion throughput), ο χρόνος απόκρισης ερωτημάτων (latency) για τυπικά σενάρια αναλύσεων, ο λόγος συμπίεσης προς τις απαιτήσεις χώρου, η κατανάλωση πόρων (επεξεργαστική ισχύ, προσωρινή μνήμη, αριθμός εισόδων - εξόδων), η ευκολία διαχείρισης των δεδομένων, οι δυνατότητες κλιμάκωσης και η διατήρηση της ακεραιότητας των δεδομένων. Τα αποτελέσματα ερμηνεύονται με βάση τις ιδιαιτερότητες των χρονοσειρών (υψηλή συχνότητα, ετικέτες, μεταδεδομένα) και τυχόν πρακτικές λειτουργικές ανάγκες.

3.3.1 InfluxDB και παραδοσιακές βάσεις δεδομένων

Οι παραδοσιακές βάσεις δεδομένων όπως οι MySQL και PostgreSQL κ.λπ. χρησιμοποιούνται ευρέως για γενικές εφαρμογές. Έχουν συνήθως σχεδιαστεί για γενικού τύπου δεδομένα και ερωτήματα που δεν είναι εξειδικευμένα σε δεδομένα χρονοσειρών. Το InfluxDB είναι σχεδιασμένο ειδικά για την αποθήκευση και την ερώτηση χρονοσειρικών δεδομένων. Η δομή του επιτρέπει γρήγορη αποθήκευση και ανάκτηση δεδομένων με χρονολογικό περιεχόμενο, σε αντίθεση με τις παραδοσιακές βάσεις δεδομένων, οι οποίες δεν είναι βέλτιστες όταν εξυπηρετούν τέτοιου τύπου δεδομένα. Το InfluxDB έχει σχεδιαστεί για να παρέχει υψηλή απόδοση στην εισαγωγή, ανάκτηση και επεξεργασία χρονοσειρικών δεδομένων. Καθώς η βάση δεδομένων αυξάνει το μέγεθός της, το InfluxDB συνήθως διατηρεί υψηλή απόδοση, ενώ παράλληλα προσφέρει καλή δυνατότητα κλιμάκωσης. Χρησιμοποιεί ειδικές δομές δεδομένων όπως τα μέτρα, τα πεδία και τις ετικέτες, για την οργάνωση των δεδομένων επιτρέποντας έτσι τη γρήγορη και αποδοτική ανάκτηση πληροφοριών. Επιπλέον, το InfluxDB προσφέρει συχνά ευκολία χρήσης σε σχέση με παραδοσιακές βάσεις δεδομένων. Η εύκολη εγκατάσταση, η χρήση της γλώσσας ερωτημάτων και η δυνατότητα εύκολης διαχείρισης χρονοσειρών το καθιστούν ελκυστικό για αρχάριους χρήστες και εφαρμογές δεδομένων χρονοσειρών. Όλα τα παραπάνω, σε συνδυασμό με τις πληροφορίες στον Πίνακα 3.1, παρέχουν τη γενική εικόνα και το λόγο όπου το InfluxDB, όταν πρόκειται για δεδομένα χρονοσειρών, είναι ασφαλώς προτεινόμενο σε σχέση με τις παραδοσιακές βάσεις δεδομένων.

3.3.2 Συγκριτική ανάλυση InfluxDB και TimescaleDB

Στην ενότητα αυτή πραγματοποιείται μια συστηματική σύγκριση των δύο βάσεων δεδομένων ως προς τη σχεδίαση, την απόδοση, την δυνατότητα κλιμάκωσης, τη συμβατότητα και τη διαχείριση, με στόχο να αναδειχθούν τα πλεονεκτήματα και οι περιορισμοί κάθε προσέγγισης στο πλαίσιο αποθήκευσης και ανάλυσης δεδομένων χρονοσειρών. Ο Πίνακας 3.2 συνοψίζει τις διαφορές που αναλύονται παρακάτω.

Χαρακτηριστικό	InfluxDB	MySQL	PostgreSQL	Elasticsearch
Μοντέλο δεδομένων	Εστιασμένο σε χρονοσειρές	Σχεσιακό	Σχεσιακό	Μη-σχεσιακό
Απόδοση ερωτημάτων	Εξαιρετική για χρονοσειρές	Καλή	Καλή	Καλή
Κλιμακωσιμότητα	Οριζόντια κλιμακωσιμότητα	Οριζόντια	Οριζόντια	Οριζόντια
Ευελιξία σχήματος	Σταθερό σχήμα για χρονοσειρές	Σταθερό	Σταθερό	Ευέλικτο
Ευκολία χρήσης	Σχεδιασμένο για χρονοσειρές	Μέτρια	Μέτρια	Μέτρια
Χρήσεις	Δεδομένα χρονοσειρών, IoT	Γενική	Γενική	Αναζήτηση, ανάλυση
Συνέπεια	Ενδεχομένως συνέπεια	ACID συμμόρφωση	ACID συμμόρφωση	Ενδεχομένως συνέπεια
Υποστήριξη κοινότητας	Αυξανόμενη κοινότητα	Εδραιωμένη	Εδραιωμένη	Ισχυρή
Καμπύλη μάθησης	Μέτρια καμπύλη μάθησης	Μέτρια	Μεγάλη	Μέτρια

Πίνακας 3.1: Διαφορές InfluxDB με παραδοσιακές βάσεις δεδομένων

3.3.2.1 Σχεδίαση βάσης δεδομένων

InfluxDB. Έχει ειδικά σχεδιαστεί για χρονοσειρικά δεδομένα με μια δομή που εστιάζει σε μέτρα, πεδία και ετικέτες.

TimescaleDB. Επεκτείνει τη λειτουργικότητα της PostgreSQL για χρονοσειρές μέσω ενός επεκτάσιμου μηχανισμού που χρησιμοποιεί σχεσιακά δεδομένα και πίνακες.

3.3.2.2 Απόδοση και κλιμακωσιμότητα

InfluxDB. Εξασφαλίζει υψηλή απόδοση για χρονοσειρές δεδομένων και είναι εξειδικευμένο σε αυτόν τον τομέα. Η κλιμακωσιμότητά του είναι καλή για μεγάλα σύνολα δεδομένων.

TimescaleDB. Επωφελείται από τη δύναμη και την δυνατότητα κλιμάκωσης της PostgreSQL, προσφέροντας παρόμοια απόδοση για χρονοσειρές.

Χαρακτηριστικό	TimescaleDB	InfluxDB
Μοντέλο δεδομένων	Χρονοσειρές στη βάση δεδομένων PostgreSQL	Εστιασμένο σε χρονοσειρές
Απόδοση ερωτημάτων	Καλή για χρονοσειρές και SQL ερωτήματα	Εξαιρετική για χρονοσειρές
Κλιμακωσιμότητα	Οριζόντια κλιμακωσιμότητα με τον διαχωρισμό δεδομένων	Οριζόντια κλιμακωσιμότητα
Ευελιξία σχήματος	Συμβατό με τις λειτουργίες του PostgreSQL	Σταθερό σχήμα για χρονοσειρές
Ευκολία χρήσης	Οικείο στους χρήστες PostgreSQL, γνώση SQL	Σχεδιασμένο για χρονοσειρές, μέτρια καμπύλη μάθησης
Χρήσεις	Αναλυτικά δεδομένα, παρακολούθηση, IoT	Δεδομένα χρονοσειρών, IoT
Συνέπεια	ACID συμμόρφωση	Ενδεχομένως συνέπεια
Υποστήριξη κοινότητας	Ισχυρή υποστήριξη κοινότητας	Αυξανόμενη κοινότητα
Καμπύλη μάθησης	Οικεία για χρήστες PostgreSQL, μέτρια καμπύλη μάθησης	Μέτρια καμπύλη μάθησης

Πίνακας 3.2: Διαφορές μεταξύ InfluxDB και TimescaleDB

3.3.2.3 Συμβατότητα και οικοσύστημα

InfluxDB. Διαθέτει ένα ευρύ οικοσύστημα για εφαρμογές IoT, εφαρμογές παρακολούθησης πόρων και αισθητήρων, με πολλά εργαλεία και οδηγούς ενσωμάτωσης.

TimescaleDB. Ως επέκταση της PostgreSQL, έχει πρόσβαση σε μια ευρεία γκάμα εργαλείων και βιβλιοθηκών που την υποστηρίζουν.

3.3.2.4 Διαχείριση και χρήση

InfluxDB. Η διαχείριση είναι σχετικά απλή, ειδικά όταν πρόκειται για χρονοσειριακά δεδομένα, αλλά η εκμάθηση του InfluxQL μπορεί να απαιτήσει χρόνο για να εκμεταλλευτείτε πλήρως τις δυνατότητές του.

TimescaleDB. Η χρήση είναι πιο οικεία για χρήστες PostgreSQL και SQL γενικά, με επιπλέον δυνατότητες όπως οι δυνατότητες SQL για πλήρη ανάλυση των δεδομένων.

3.4 Σύνοψη

Μετά την ανάλυση όλων των διαθέσιμων σχεσιακών και μη βάσεων δεδομένων, καταλήγουμε ότι οι πιο αποδοτικές πλατφόρμες για τη διαχείριση δεδομένων χρονοσειρών είναι το InfluxDB και το TimescaleDB. Για το λόγο αυτό, θα επιλέξουμε και τις δύο πλατφόρμες προς εγκατάσταση για τα αγρομετεωρολογικά μας δεδομένα και θα συγκρίνουμε τη μεταξύ τους απόδοση.

Κεφάλαιο 4

Δεδομένα χρονοσειρών που αξιοποιούνται

Στο παρόν κεφάλαιο παρουσιάζονται αναλυτικά τα δεδομένα χρονοσειρών που αξιοποιήθηκαν στο πλαίσιο της παρούσας διατριβής. Η ανάλυση επικεντρώνεται στην προέλευση, τη φύση και τα χαρακτηριστικά των δεδομένων, καθώς και στον τρόπο με τον οποίο αυτά συλλέγονται και αξιοποιούνται από τους μετεωρολογικούς σταθμούς. Η κύρια πηγή δεδομένων είναι οι μετεωρολογικοί σταθμοί που λειτουργούν στην Περιφέρεια Πελοποννήσου, τόσο σε αγροτικές όσο και σε αστικές περιοχές. Οι σταθμοί αυτοί είναι εξοπλισμένοι με πλήθος αισθητήρων για την καταγραφή περιβαλλοντικών παραμέτρων όπως θερμοκρασία, υγρασία, ατμοσφαιρική πίεση, βροχόπτωση και ηλιακή ακτινοβολία [25]. Τα δεδομένα συλλέγονται σε μορφή χρονοσειρών με σταθερή συχνότητα δειγματοληψίας και αποστέλλονται σε κεντρικό σύστημα επεξεργασίας και αποθήκευσης.

4.1 Κριτήρια επιλογής σταθμών βάσει απαιτήσεων

Η επιλογή των μετεωρολογικών σταθμών που αξιοποιήθηκαν στη διατριβή δεν έγινε τυχαία, αλλά βασίστηκε σε ένα σαφώς καθορισμένο σύνολο τεχνικών απαιτήσεων. Η προσέγγιση αυτή υιοθετήθηκε ώστε τα δεδομένα χρονοσειρών που συλλέγονται να είναι

συγκρίσιμα, αξιόπιστα και κατάλληλα για αυτοματοποιημένη επεξεργασία, μειώνοντας την αβεβαιότητα που εισάγουν ετερογενείς εγκαταστάσεις ή ασυνεπείς μετρήσεις. Με άλλα λόγια, η διαδικασία επιλογής στηρίχθηκε σε κριτήρια που εξασφαλίζουν ότι το σύνολο των σταθμών μπορεί να λειτουργήσει ως ενιαίο δίκτυο συλλογής δεδομένων και να τροφοδοτήσει ομοιόμορφα τα στάδια εισαγωγής, καθαρισμού και αποθήκευσης.

Οι βασικοί στόχοι ήταν:

- (α) Η εξασφάλιση της αξιοπιστίας των μετρήσεων.
- (β) Η επάρκεια και καταλληλότητα των αισθητήρων ως προς τις μεταβλητές ενδιαφέροντος.
- (γ) Η λειτουργική αυτονομία και η ανθεκτικότητα σε περιβαλλοντικές συνθήκες.
- (δ) Η δυνατότητα τηλεμετρικής αποστολής δεδομένων με τρόπο που να υποστηρίζει σταθερή συχνότητα δειγματοληψίας.

Οι απαιτήσεις που συλλέξαμε αποτελούν το σημείο αναφοράς για το κατά πόσο ένας σταθμός μπορεί να υποστηρίξει τη συλλογή χρονοσειρών υψηλής ανάλυσης (π.χ. ανά 10 λεπτά), να περιορίσει τα κενά δεδομένων και να ενσωματωθεί ομαλά στο σύστημα αποθήκευσης και επεξεργασίας.

Στο Παράρτημα [Α'](#) παρατίθενται αναλυτικά οι προδιαγραφές και οι απαιτήσεις που καθορίσαμε και αφορούν:

Πλήρως εξοπλισμένους αγρομετεωρολογικούς σταθμούς.

Πλήρως εξοπλισμένα συστήματα μέτρησης ατμοσφαιρικών ρύπων και μικροκλίματος.

Για λόγους σαφήνειας, οι απαιτήσεις μπορούν να ομαδοποιηθούν στις παρακάτω κατηγορίες:

Απαιτήσεις ποιότητας μετρήσεων και συγκρισιμότητας. Αφορούν την ακρίβεια/ανάλυση των αισθητήρων, τη σταθερότητα των μετρήσεων και τη δυνατότητα βαθμονόμησης ή επαναβαθμονόμησης στο πεδίο. Η ύπαρξη αντίστοιχων αισθητήρων ανά σταθμό εξασφαλίζει ότι οι χρονοσειρές είναι ομοειδείς και επιτρέπουν συγκρίσεις μεταξύ περιοχών και χρονικών περιόδων.

Απαιτήσεις διαθεσιμότητας και ανθεκτικότητας. Περιλαμβάνουν προστασία περιβλήματος, εύρος θερμοκρασιών λειτουργίας, υλικά κατάλληλα για εξωτερική εγκατάσταση (αντοχή σε ακτινοβολία και διάβρωση), καθώς και μηχανισμούς ειδοποίησης για τυχόν δυσλειτουργίες. Οι απαιτήσεις αυτές μειώνουν τις πιθανότητες απώλειας δεδομένων λόγω περιβαλλοντικών καταπονήσεων ή βλαβών.

Απαιτήσεις ενεργειακής αυτονομίας. Περιλαμβάνουν χαμηλή κατανάλωση ισχύος, αυτονομία λειτουργίας πολλών ημερών και υποστήριξη φωτοβολταϊκής τροφοδοσίας με έλεγχο φόρτισης. Η ενεργειακή αυτονομία είναι κρίσιμη ιδίως σε απομακρυσμένες αγροτικές περιοχές, ώστε να διατηρείται αδιάλειπτη η συλλογή μετρήσεων.

Απαιτήσεις επικοινωνίας και τηλεμετρίας. Αφορούν τη δυνατότητα τηλεμετάδοσης και τοπικής αποθήκευσης, την ύπαρξη modem κινητής τηλεφωνίας, καθώς και τηλεμετρικό έλεγχο (διάγνωση κατάστασης αισθητήρων, ειδοποιήσεις σε περίπτωση αποσύνδεσης ή βλάβης). Οι λειτουργίες αυτές υποστηρίζουν την έγκαιρη ανίχνευση προβλημάτων και τον περιορισμό των χρονικών κενών στις χρονοσειρές.

Απαιτήσεις λειτουργικής ολοκλήρωσης με το σύστημα επεξεργασίας. Η δυνατότητα προγραμματισμού του ρυθμού μέτρησης και αποστολής, καθώς και η συστηματική παραγωγή χρονοσφραγίδων είναι καθοριστικές για την ομαλή λειτουργία της αυτοματοποιημένης ETL διαδικασίας. Στη συγκεκριμένη μελέτη, η ύπαρξη σταθερού ρυθμού δειγματοληψίας και τυποποιημένης δομής δεδομένων επιτρέπει την αυτοματοποιημένη εισαγωγή, τον έλεγχο πληρότητας και την αποθήκευση σε βάσεις δεδομένων χρονοσειρών χωρίς χειροκίνητες παρεμβάσεις.

Με βάση τις απαιτήσεις που καθορίσαμε, επιλέχθηκαν σταθμοί που καλύπτουν κρίσιμες λειτουργίες:

- (α) τηλεμετάδοση και τοπική αποθήκευση μετρήσεων.
- (β) αυτονομία ισχύος (μέσω φωτοβολταϊκής τροφοδοσίας).
- (γ) προστασία περιβάλλοντος και ανθεκτικότητα εγκατάστασης.

(δ) υποστήριξη των απαιτούμενων αισθητήρων (βροχόπτωση, άνεμος, ακτινοβολία, θερμοκρασία/υγρασία, βαρομετρική πίεση και, για το αστικό περιβάλλον, ρύποι και αιωρούμενα σωματίδια).

Συνεπώς, το δίκτυο των δέκα σταθμών που παρουσιάζεται στις επόμενες ενότητες ικανοποιεί τις προδιαγραφές που παρουσιάζονται στο Παράρτημα Α' και μπορεί να θεωρηθεί ως σύνολο εγκαταστάσεων που ανταποκρίνεται στις ανάγκες της μελέτης, εξασφαλίζοντας υψηλή διαθεσιμότητα, συγκρίσιμες μετρήσεις και αξιόπιστα δεδομένα χρονοσειρών.

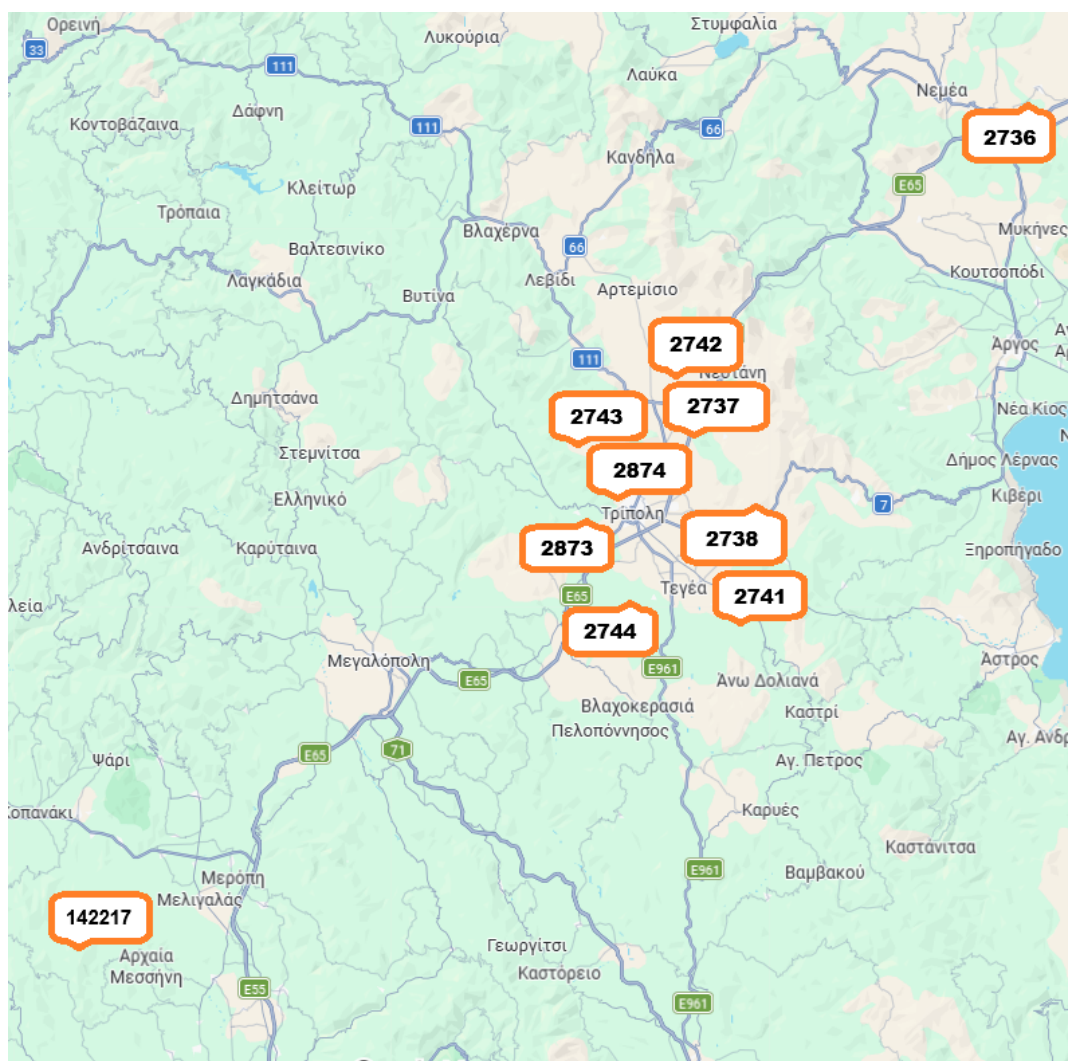
4.2 Δεδομένα μετεωρολογικών σταθμών

Στο πλαίσιο της διατριβής αξιοποιούνται τα δεδομένα των δέκα μετεωρολογικών σταθμών οι οποίοι βρίσκονται εγκατεστημένοι σε διαφορετικά σημεία της Περιφέρειας Πελοποννήσου. Κάθε σταθμός αποτελείται από μία κεντρική μονάδα και ένα ή περισσότερα δορυφορικά υποσυστήματα. Οι δορυφορικές μονάδες επικοινωνούν αποκλειστικά με την κεντρική μονάδα, η οποία αναλαμβάνει τη συγκέντρωση όλων των μετρήσεων και την αποστολή τους στο δίκτυο. Οι μετρήσεις πραγματοποιούνται σε τακτά χρονικά διαστήματα, δημιουργώντας έτσι σύνολα χρονοσειρών υψηλής ανάλυσης. Συνολικά λειτουργούν οκτώ αγροτικοί και δύο αστικοί σταθμοί, οι οποίοι παρουσιάζονται αναλυτικά στις επόμενες ενότητες.

4.2.1 Αγροτικοί σταθμοί

Οι αγροτικοί σταθμοί είναι εγκατεστημένοι σε περιοχές γεωργικού ενδιαφέροντος και έχουν ως στόχο την παρακολούθηση κρίσιμων περιβαλλοντικών μεταβλητών που επηρεάζουν άμεσα την ανάπτυξη, την υγεία και την παραγωγικότητα των καλλιεργειών. Η γεωγραφική τους κατανομή φαίνεται στο Σχήμα 4.1.

Κάθε αγροτικός σταθμός περιλαμβάνει μία κεντρική μονάδα (με Αναγνωριστικό 17) και δύο δορυφορικές μονάδες (με Αναγνωριστικά 1 και 2). Η κεντρική μονάδα συγκεντρώνει δεδομένα υψηλής ακρίβειας από ένα πλήρες σετ αισθητήρων, το οποίο παρουσιάζεται αναλυτικά παρακάτω:



Σχήμα 4.1: Γεωγραφική αποτύπωση όλων των αγροτικών σταθμών.

Μετρητής υγρασίας αέρα (%): Μετρά τη σχετική υγρασία του αέρα, δηλαδή το ποσοστό των υδρατμών σε σχέση με τη μέγιστη ποσότητα που μπορεί να συγχρατήσει ο αέρας στη συγκεκριμένη θερμοκρασία. Η παράμετρος αυτή επηρεάζει σημαντικά τη διαπνοή των φυτών, τη φωτοσύνθεση και την πιθανότητα ανάπτυξης μυκητολογικών ασθενειών.

Μετρητής θερμοκρασίας αέρα (°C): Καταγράφει τη θερμοκρασία του περιβάλλοντος αέρα. Οι τιμές της χρησιμεύουν για την εκτίμηση της ημερήσιας και εποχικής μεταβολής της θερμότητας, που καθορίζει τη διάρκεια του βλαστικού κύκλου των καλλιεργειών.

Μετρητής θερμοκρασίας σημείου δρόσου (°C): Αντιπροσωπεύει τη θερμοκρασία στην οποία ο αέρας φτάνει σε κορεσμό και αρχίζει να σχηματίζεται δρόσος.

Η παράμετρος αυτή είναι σημαντική για την πρόβλεψη συνθηκών συμπύκνωσης πάνω στα φύλλα και για την εκτίμηση κινδύνου ασθενειών που ευνοούνται από την υγρασία.

Μετρητής θερμοκρασίας εξάτμισης (wet-bulb) (°C): Αντιστοιχεί στη χαμηλότερη θερμοκρασία που μπορεί να επιτευχθεί με εξάτμιση νερού από την επιφάνεια ενός αισθητήρα. Η τιμή αυτή σχετίζεται με τη δυνατότητα ψύξης του αέρα μέσω εξάτμισης και χρησιμοποιείται σε υπολογισμούς εξατμισοδιαπνοής, δηλαδή της ποσότητας νερού που χάνεται από το έδαφος και τα φυτά.

Μετρητής ύψους βροχής (mm): Καταγράφει την ποσότητα βροχόπτωσης σε χιλιοστά ανά χρονική περίοδο. Ο βροχομετρικός αισθητήρας λειτουργεί με μηχανισμό tipping bucket, όπου κάθε ανατροπή αντιστοιχεί σε 0.22 mm νερού και το όργανο μηδενίζεται αυτόματα κάθε 10 λεπτά. Οι μετρήσεις αυτές είναι θεμελιώδεις για την ανάλυση του υδάτινου ισοζυγίου των καλλιεργειών.

Μετρητής ατμοσφαιρικής πίεσης (hPa): Καταγράφει τη βαρομετρική πίεση της ατμόσφαιρας. Οι διακυμάνσεις της πίεσης βοηθούν στην αναγνώριση μεταβολών του καιρού και στη συσχέτισή τους με τις συνθήκες βροχόπτωσης ή ανέμων στις αγροτικές περιοχές.

Μετρητής ηλιακής ακτινοβολίας (W/m^2): Μετρά την ολική ηλιακή ακτινοβολία (άμεση και διάχυτη) που δέχεται το σημείο μέτρησης. Η παράμετρος αυτή καθορίζει τη διαθεσιμότητα ενέργειας για τη φωτοσύνθεση, τον ρυθμό εξάτμισης και τη θερμική ισορροπία του εδάφους.

Μετρητής κατεύθυνσης ανέμου (°): Υποδεικνύει τη διεύθυνση προέλευσης του ανέμου ως προς τον γεωγραφικό βορρά. Ο παράγοντας αυτός είναι κρίσιμος για τη διάδοση ρύπων, τη μεταφορά υγρασίας και τη διασπορά φυτοπαθογόνων μικροοργανισμών.

Μετρητής ταχύτητας ανέμου (m/s): Καταγράφει τη μέση ταχύτητα του ανέμου σε συγκεκριμένο χρονικό διάστημα. Η παράμετρος αυτή επηρεάζει την εξαιμισοδιαπνοή, τη μεταφορά θερμότητας και την επικονίαση των καλλιεργειών.

Μετρητής ριπών ανέμου (m/s): Μετρά τις μέγιστες στιγμιαίες τιμές ταχύτητας του ανέμου (ριπές). Οι ριπές μπορούν να προκαλέσουν μηχανικές καταπονήσεις στα φυτά και χρησιμοποιούνται για την ανάλυση δυναμικών φαινομένων στον αγρό, όπως οι στροβιλισμοί και η αστάθεια της ατμόσφαιρας.

Η συλλογή και αποστολή των δεδομένων πραγματοποιείται σε τακτά χρονικά διαστήματα, συνήθως ανά 10 λεπτά, προς τον κεντρικό εξυπηρετητή του συστήματος. Τα δεδομένα αξιοποιούνται τόσο για την παρακολούθηση πραγματικών συνθηκών όσο και για τη βαθμονόμηση μοντέλων πρόβλεψης απόδοσης και άρδευσης [1].

4.2.1.1 Δορυφορικές μονάδες

Οι δύο δορυφορικές μονάδες κάθε αγροτικού σταθμού διαθέτουν πιο περιορισμένο αλλά εξειδικευμένο εξοπλισμό. Κάθε μία περιλαμβάνει τους παρακάτω αισθητήρες.

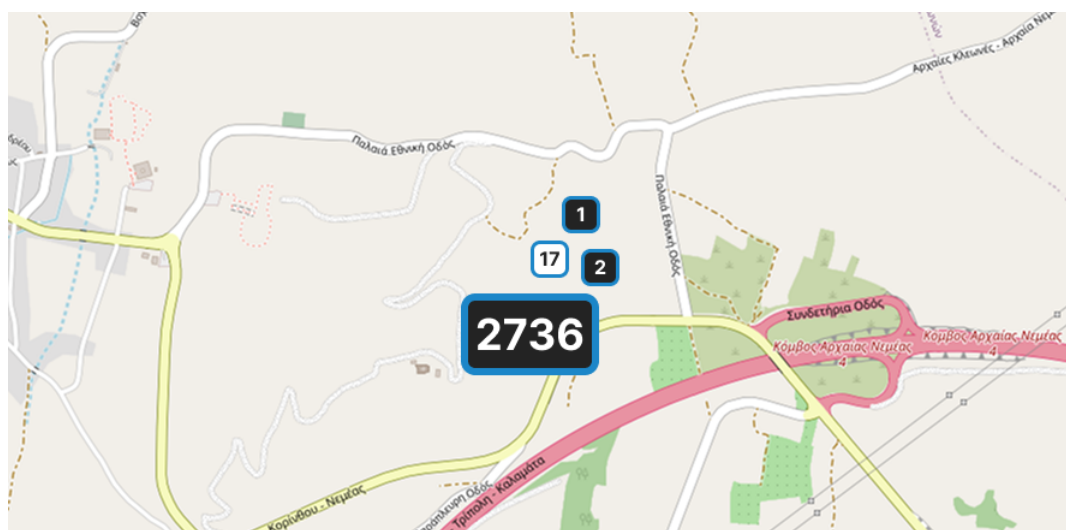
Μετρητής υγρασίας αέρα (%): Όπως περιγράφηκε παραπάνω.

Μετρητής υγρασίας φύλλου (eleaf) (%): Μετρά την υγρασία στην επάνω και κάτω επιφάνεια ενός τεχνητού φύλλου (αισθητήρας Eleaf), παρέχοντας δύο τιμές ανά μέτρηση.

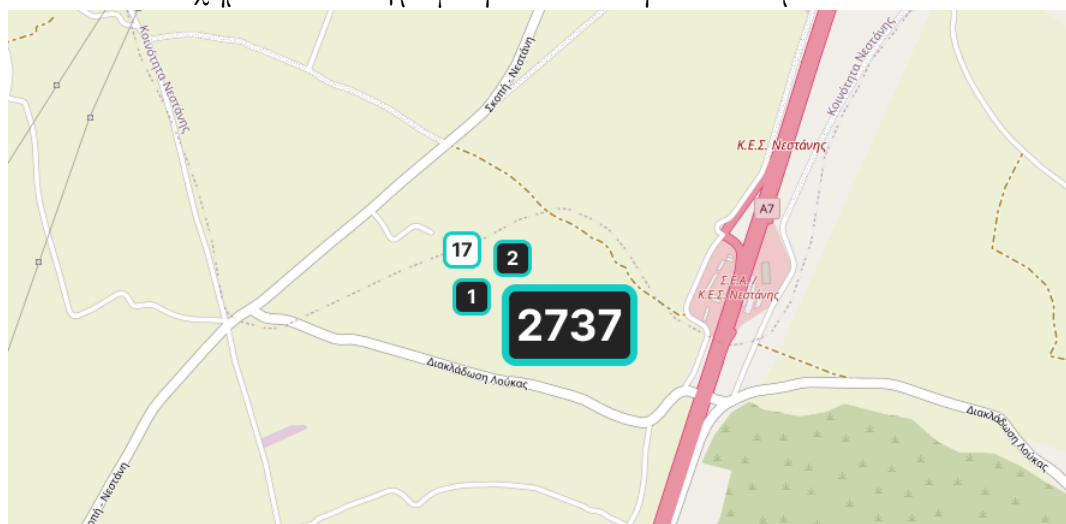
Μετρητής θερμοκρασίας αέρα και σημείου δρόσου (°C): Πανομοιοτύποι με τους αισθητήρες της κεντρικής μονάδας.

Μετρητής τάσης μπαταρίας 3.6V (V): Παρακολουθεί την τάση της μπαταρίας κάθε δορυφορικής μονάδας.

Μετρητής εσωτερικής θερμοκρασίας (°C): Μετρά τη θερμοκρασία στο εσωτερικό περίβλημα της μονάδας.



Σχήμα 4.2: Γεωγραφική αποτύπωση του σταθμού 2736.



Σχήμα 4.3: Γεωγραφική αποτύπωση του σταθμού 2737.

4.2.1.2 Τοποθεσία σταθμών

Οι επτά αγροτικοί σταθμοί διαθέτουν τα εξής αναγνωριστικά και θέσεις που αναλύονται ακολούθως.

Σταθμός 2736: Έχει εγκατασταθεί στην Αρχαία Νεμέα στο οινοποιείο Σπυρόπουλου (Σχήμα 4.2). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.80558	22.72306
Δορυφορική	1	37.80686	22.72398
Δορυφορική	2	37.80504	22.72367

Σταθμός 2737: Έχει εγκατασταθεί στη Μαντίνεια (Σχήμα 4.3). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.58095	22.42309
Δορυφορική	1	37.57956	22.42389
Δορυφορική	2	37.58100	22.42463

Σταθμός 2738: Έχει εγκατασταθεί στο Ζευγολατιό (Σχήμα 4.4). Η κεντρική μονάδα είναι στην παρακάτω θέση.

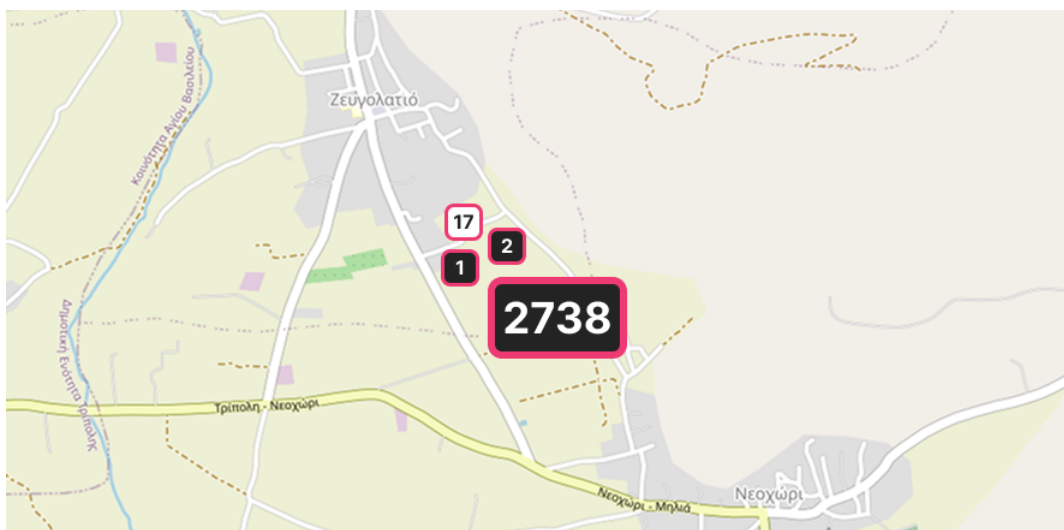
Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.51466	22.44363

Σταθμός 2741: Έχει εγκατασταθεί στην Τεγέα (Σχήμα 4.5). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

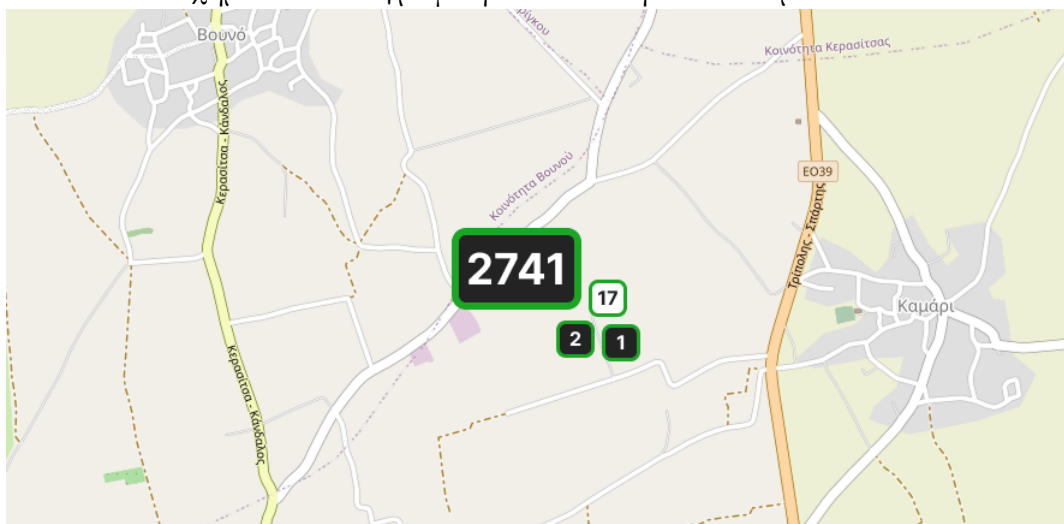
Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.43980	22.40298
Δορυφορική	1	37.44001	22.40231
Δορυφορική	2	37.44019	22.40338

Σταθμός 2742: Έχει εγκατασταθεί στην Κάψια (Σχήμα 4.6). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

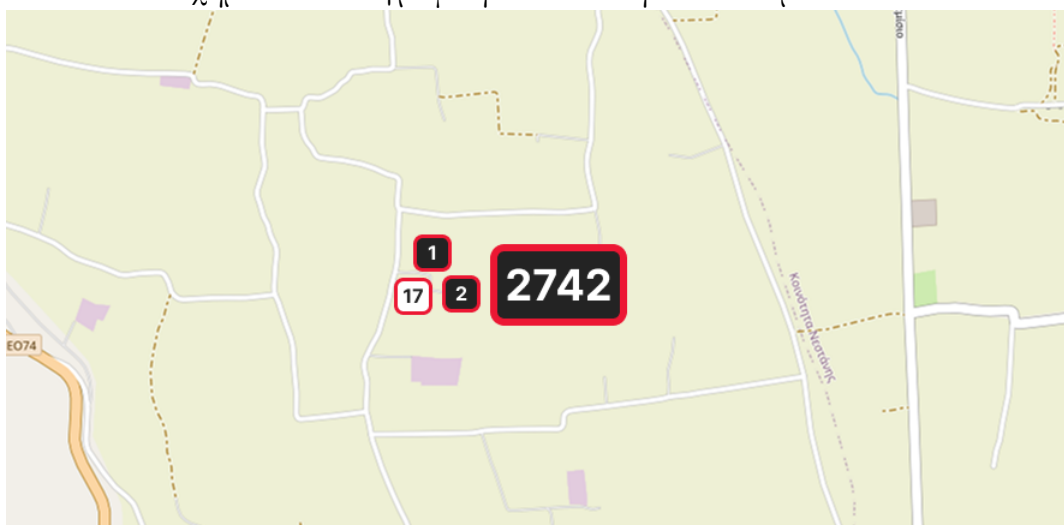
Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.60740	22.37400
Δορυφορική	1	37.60736	22.37428
Δορυφορική	2	37.60734	22.37452



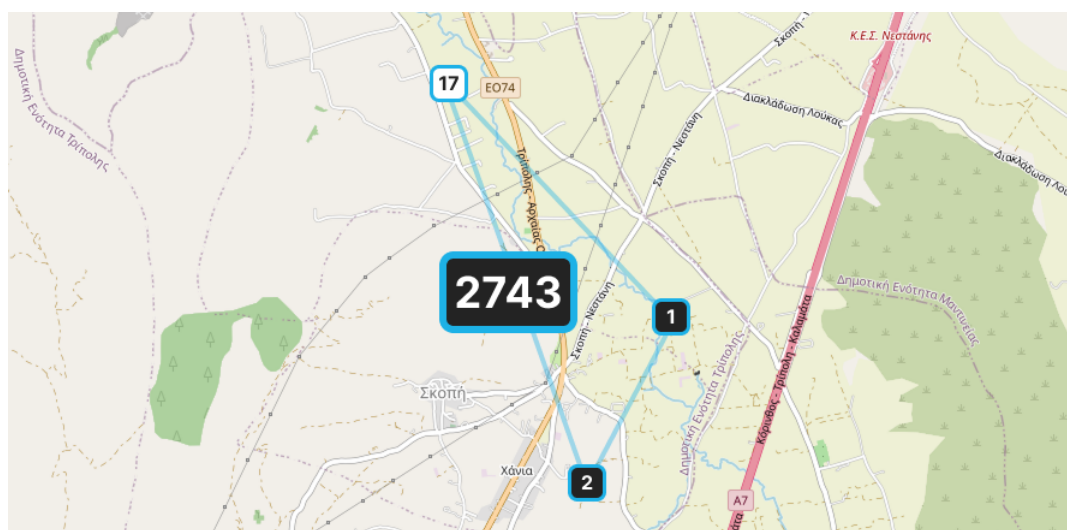
Σχήμα 4.4: Γεωγραφική αποτύπωση του σταθμού 2738.



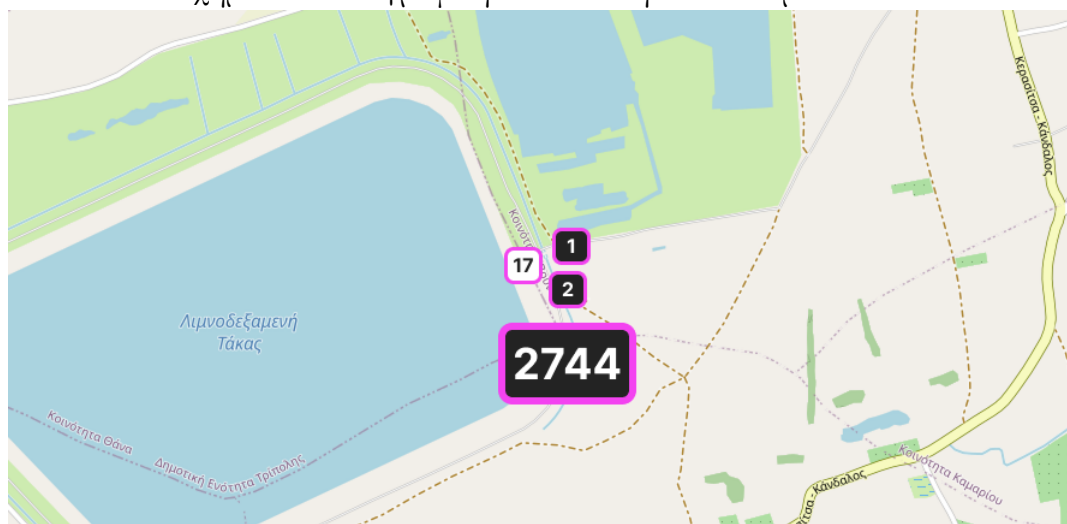
Σχήμα 4.5: Γεωγραφική αποτύπωση του σταθμού 2741.



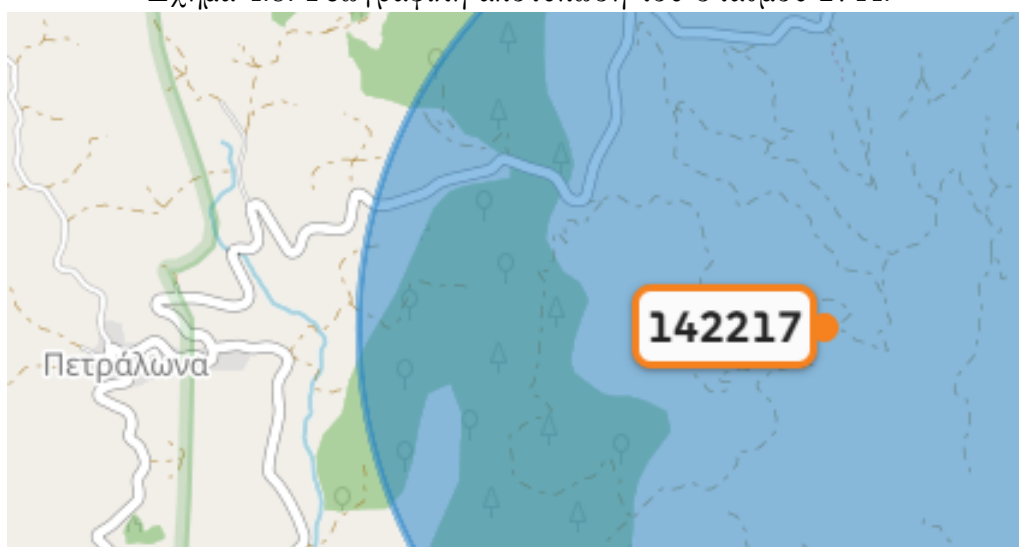
Σχήμα 4.6: Γεωγραφική αποτύπωση του σταθμού 2742.



Σχήμα 4.7: Γεωγραφική αποτύπωση του σταθμού 2743.



Σχήμα 4.8: Γεωγραφική αποτύπωση του σταθμού 2744.



Σχήμα 4.9: Γεωγραφική αποτύπωση του σταθμού 142217.

Σταθμός 2743: Έχει εγκατασταθεί στη Σκοπή, κοντά στο οινοποιείο Τρουπή (Σχήμα 4.7). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.57919	22.39147
Δορυφορική	1	37.56221	22.41221
Δορυφορική	2	37.55049	22.40447

Σταθμός 2744: Έχει εγκατασταθεί στη λιμνοδεξαμενή Τάχα (Σχήμα 4.8). Η κεντρική μονάδα και οι δορυφορικές μονάδες είναι στις παρακάτω θέσεις.

Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.43451	22.37876
Δορυφορική	1	37.48015	22.34849
Δορυφορική	2	37.41370	22.40363

Σταθμός 142217: Έχει εγκατασταθεί στην περιοχή Μαγγανιακός (Σχήμα 4.9). Η κεντρική μονάδα μαζί με τη μετεωρολογική μονάδα βρίσκονται στις παρακάτω θέσεις.

Μονάδα	Κωδικός	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα	17	37.17743	21.90868
Μετεωρολογική	1	37.17743	21.90868

Οι τοποθεσίες των σταθμών συνοψίζονται στον χάρτη του Σχήματος 4.1, ενώ κάθε επιμέρους σχήμα αποτυπώνει τη γεωγραφική διάταξη των επιμέρους μονάδων.

4.2.2 Αστικοί σταθμοί

Οι δύο αστικοί σταθμοί είναι εγκατεστημένοι σε περιοχές εντός του αστικού ιστού, με σκοπό την παρακολούθηση περιβαλλοντικών παραμέτρων που επηρεάζουν τη ρύπανση και τη μικροκλιματική συμπεριφορά. Κάθε σταθμός αποτελείται από μία κεντρική μονάδα (κωδικός 17) και τρεις δορυφορικές μονάδες (κωδικοί 68, 69, 70). Οι σταθμοί είναι εξοπλισμένοι με αισθητήρες ακριβείας, ικανούς να μετρούν την ποιότητα του αέρα, τη

θερμοκρασία, την υγρασία και άλλα κρίσιμα ατμοσφαιρικά χαρακτηριστικά, παρέχοντας συνεχή ροή δεδομένων σε πραγματικό χρόνο.

Η κεντρική μονάδα κάθε αστικού σταθμού διαθέτει τους εξής αισθητήρες:

Μετρητής υγρασίας αέρα (%): Όπως περιγράφηκε παραπάνω.

Μετρητής θερμοκρασίας αέρα (°C): Όπως περιγράφηκε παραπάνω.

Μετρητής θερμοκρασίας σημείου δρόσου (°C): Όπως περιγράφηκε παραπάνω.

Μετρητής ύψους βροχής (mm): Όπως περιγράφηκε παραπάνω.

Μετρητής διοξειδίου του άνθρακα (ppm): Μετρά τη συγκέντρωση CO₂ στην ατμόσφαιρα. Το CO₂ αποτελεί κύριο δείκτη ανθρωπογενούς δραστηριότητας και χρησιμοποιείται για την εκτίμηση της ποιότητας του αέρα, ειδικά σε περιοχές με αυξημένη κυκλοφορία ή ανθρώπινη παρουσία.

Μετρητής ατμοσφαιρικής πίεσης (hPa): Όπως περιγράφηκε παραπάνω.

Μετρητής κατεύθυνσης ανέμου (°): Όπως περιγράφηκε παραπάνω.

Μετρητής ταχύτητας και ριπής ανέμου (m/s): Όπως περιγράφηκε παραπάνω.

Μετρητής υπεριώδους ακτινοβολίας (MED): Μετρά την ένταση της υπεριώδους ακτινοβολίας (UV) που φτάνει στο έδαφος. Οι τιμές εκφράζονται σε MED (Minimal Erythema Dose) και ο αισθητήρας αυτός είναι εγκατεστημένος μόνο στον σταθμό 2873. Συμβάλλει στην αξιολόγηση της ηλιακής έκθεσης και των πιθανών κινδύνων για το δέρμα.

Μετρητής ηλιακής ακτινοβολίας (W/m²): Όπως περιγράφηκε παραπάνω.

4.2.2.1 Δορυφορικές μονάδες

Οι δορυφορικές μονάδες των αστικών σταθμών (68, 69, 70) συμπληρώνουν τη λειτουργία της κεντρικής μονάδας, καταγράφοντας πιο εξειδικευμένες ατμοσφαιρικές παραμέτρους που σχετίζονται κυρίως με τη ρύπανση και τα αιωρούμενα σωματίδια.

4.2.2.1.1 Μονάδα 68

Η δορυφορική μονάδα με κωδικό 68, διαθέτει τους ακόλουθους αισθητήρες:

Μετρητής τάσης μπαταρίας 12V (V): Παρακολουθεί την τάση της μπαταρίας της δορυφορικής μονάδας, εξασφαλίζοντας τη συνεχή λειτουργία του εξοπλισμού και την αξιοπιστία των μετρήσεων.

Μετρητής συγκέντρωσης σωματιδίων PM1, PM2.5, PM10 ($\mu\text{g}/\text{m}^3$):

Καταγράφει τη συγκέντρωση αιωρούμενων σωματιδίων με διάμετρο μικρότερη των 1, 2.5 και 10 μικρομέτρων αντίστοιχα. Αυτά τα σωματίδια έχουν σημαντική επίδραση στην ανθρώπινη υγεία και αποτελούν κρίσιμους δείκτες ατμοσφαιρικής ρύπανσης.

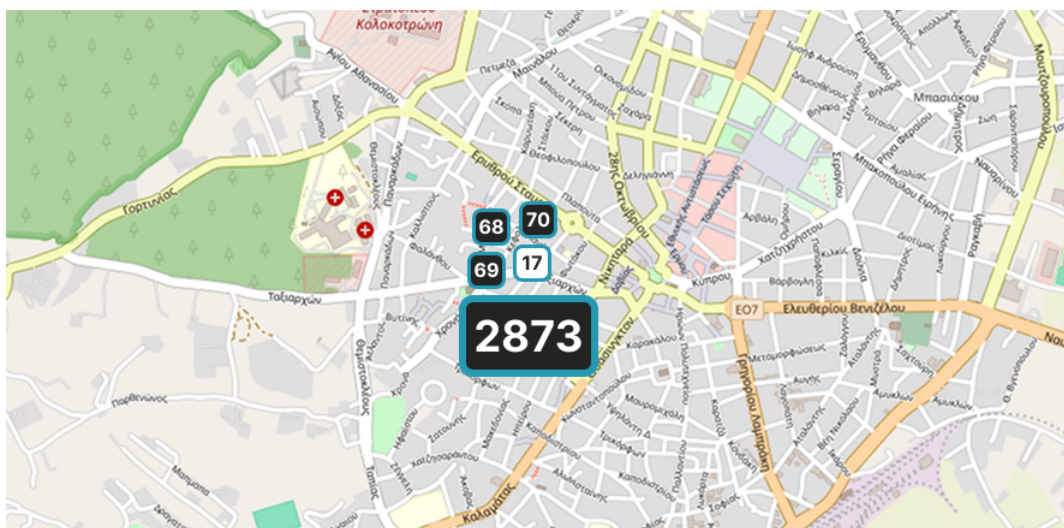
4.2.2.1.2 Μονάδα 69

Η δορυφορική μονάδα με κωδικό 69, διαθέτει τους ακόλουθους αισθητήρες:

Μετρητής όζοντος (O_3): Μετρά τη συγκέντρωση όζοντος στην ατμόσφαιρα σε μέρη ανά δισεκατομμύριο (ppb). Το όζον είναι δευτερογενής ρύπος που σχηματίζεται από φωτοχημικές αντιδράσεις, και οι τιμές του επηρεάζονται έντονα από την ηλιακή ακτινοβολία και τη θερμοκρασία.

Μετρητής διοξειδίου του αζώτου (NO_2): Καταγράφει τη συγκέντρωση NO_2 , ενός βασικού δείκτη κυκλοφοριακής ρύπανσης. Η παρουσία του συνδέεται με καύσεις οχημάτων και βιομηχανικές δραστηριότητες.

Μετρητής Κοβαλτίου (Co): Παρακολουθεί τη συγκέντρωση του στοιχείου κοβαλτίου στον αέρα. Παρότι σε μικρές ποσότητες είναι ακίνδυνο, σε αυξημένες συγκεντρώσεις μπορεί να προκαλέσει ερεθισμούς ή αναπνευστικά προβλήματα.



Σχήμα 4.10: Γεωγραφική αποτύπωση του σταθμού 2873 (Επιμελητήριο Αρκαδίας).

Οι μετρήσεις γίνονται με ηλεκτροχημικούς αισθητήρες τύπου WE και AE, που ταυτόχρονα καταγράφουν και τη θερμοκρασία λειτουργίας.

4.2.2.1.3 Μονάδα 70

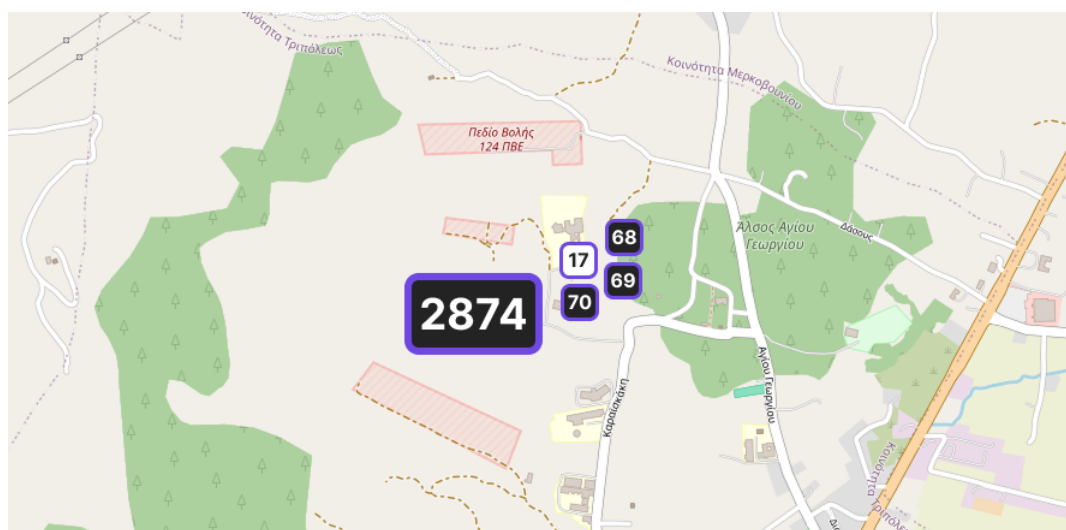
Η δορυφορική μονάδα με κωδικό 70, διαθέτει τον παρακάτω αισθητήρα:

Μετρητής διοξειδίου του θείου (SO_2): Μετρά τη συγκέντρωση του διοξειδίου του θείου (SO_2) στην ατμόσφαιρα, το οποίο προέρχεται κυρίως από καύσεις ορυκτών καυσίμων και βιομηχανικές εκπομπές. Οι αισθητήρες WE και AE καταγράφουν τόσο τη συγκέντρωση όσο και τη θερμοκρασία λειτουργίας, επιτρέποντας ακριβή υπολογισμό σε ppb. Η παρακολούθηση του διοξειδίου του θείου είναι ιδιαίτερα σημαντική, καθώς πρόκειται για ρύπο με ισχυρή οσμή και σημαντικές περιβαλλοντικές επιπτώσεις (όξινη βροχή).

4.2.2.2 Τοποθεσία σταθμών

Οι δύο αστικοί σταθμοί έχουν εγκατασταθεί σε σημεία υψηλού ενδιαφέροντος, ώστε να καταγράφουν αντιπροσωπευτικά δεδομένα για το αστικό περιβάλλον της Τρίπολης.

Σταθμός 2873: Βρίσκεται στο κέντρο της πόλης (κοντά στο επιμελητήριο Αρκαδίας) και καταγράφει δεδομένα που αντικατοπτρίζουν τις συνθήκες ρύπανσης και



Σχήμα 4.11: Γεωγραφική αποτύπωση του σταθμού 2874 (Πανεπιστήμιο Πελοποννήσου).

μικροκλίματος σε πυκνοκατοικημένες περιοχές (Σχήμα 4.10). Η κεντρική μονάδα και οι δορυφορικές 68, 69, 70 βρίσκονται στις παρακάτω συντεταγμένες.

Μονάδα	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα, 68, 69, 70	37.5108556	22.3689361

Σταθμός 2874: Εγκατεστημένος σε πανεπιστημιακό περιβάλλον (εντός του πανεπιστημίου Πελοποννήσου), καταγράφει μικροκλιματικές διαφορές σε σχέση με τον πυρήνα της πόλης, παρέχοντας πολύτιμα δεδομένα για τη σύγκριση αστικού και ημι-αστικού περιβάλλοντος (Σχήμα 4.11). Η κεντρική μονάδα και οι δορυφορικές 68, 69, 70 βρίσκονται στις παρακάτω συντεταγμένες.

Μονάδα	Συντεταγμένη x	Συντεταγμένη y
Κεντρική μονάδα, 68, 69, 70	37.5309556	22.3715667

4.3 Σύνοψη κεφαλαίου

Οι αγροτικοί σταθμοί είναι εγκατεστημένοι σε περιοχές γεωργικού ενδιαφέροντος και αποσκοπούν στην καταγραφή παραμέτρων που επηρεάζουν άμεσα την ανάπτυξη των καλλιεργειών, όπως θερμοκρασία, υγρασία, ατμοσφαιρική πίεση, βροχόπτωση, ηλιακή ακτινοβολία και άνεμος. Κάθε σταθμός αποτελείται από μία κεντρική και δύο δορυφορικές

μονάδες, οι οποίες συλλέγουν δεδομένα σε τακτά χρονικά διαστήματα (ανά 10 λεπτά) και τα αποστέλλουν στον κεντρικό εξυπηρετητή για αποθήκευση και επεξεργασία. Οι δορυφορικές μονάδες διαθέτουν συμπληρωματικούς αισθητήρες, όπως μέτρηση υγρασίας φύλλου, εσωτερικής θερμοκρασίας και τάσης μπαταρίας.

Οι αστικοί σταθμοί, τοποθετημένοι στην πόλη της Τρίπολης (στο Επιμελητήριο Αρκαδίας και στο Πανεπιστήμιο Πελοποννήσου), παρακολουθούν μικροκλιματικές συνθήκες και δείκτες ατμοσφαιρικής ρύπανσης. Διαθέτουν αισθητήρες για την καταγραφή CO₂, όζοντος (O₃), διοξειδίου του αζώτου (NO₂), διοξειδίου του θείου (SO₂), καθώς και αιωρούμενων σωματιδίων (PM1, PM2.5, PM10) [24]. Επιπλέον, οι σταθμοί αυτοί μετρούν βασικές περιβαλλοντικές παραμέτρους, όπως θερμοκρασία, υγρασία, βροχόπτωση, πίεση και ηλιακή ή υπεριώδη ακτινοβολία.

Συνολικά, το κεφάλαιο αυτό τεκμηριώνει την πηγή, την αξιοπιστία και τη χρησιμότητα των δεδομένων που αποτελούν το θεμέλιο της ανάλυσης της διατριβής. Το δίκτυο των σταθμών λειτουργεί ως ολοκληρωμένο σύστημα συλλογής και επεξεργασίας περιβαλλοντικών χρονοσειρών, προσφέροντας δεδομένα υψηλής ακρίβειας για την παρακολούθηση των περιβαλλοντικών συνθηκών και τη βαθμονόμηση προβλεπτικών μοντέλων.

Κεφάλαιο 5

Διαχείριση δεδομένων στην InfluxDB και TimescaleDB

Στο κεφάλαιο αυτό περιγράφεται η πρακτική διαδικασία αποθήκευσης των δεδομένων που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, τα οποία προέρχονται από μετεωρολογικούς σταθμούς, σε δύο διαφορετικές βάσεις δεδομένων χρονοσειρών: την InfluxDB και την TimescaleDB.

Κάθε πλατφόρμα αξιοποιεί διαφορετική φιλοσοφία αποθήκευσης: η InfluxDB βασίζεται σε ένα μοντέλο που δεν απαιτεί τη δημιουργία σχήματος (schema-less), που χρησιμοποιεί μετρήσεις (measurements), ετικέτες (tags), πεδία (fields) και χρονοσφραγίδες (timestamps), ενώ η TimescaleDB, ως επέκταση της PostgreSQL, βασίζεται σε πλήρως ορισμένο σχεσιακό σχήμα με ειδικούς πίνακες (hypertables) και ξένα κλειδιά (foreign keys).

Η διαδικασία ενσωματώνει τα στάδια της διαδικασίας Εξαγωγής-Μετασχηματισμού-Μεταφόρτωσης (Extract-Transform-Loading – ETL) που περιλαμβάνουν την εξαγωγή των δεδομένων από τους σταθμούς, τον μετασχηματισμό και κανονικοποίηση τους, και την τελική αποθήκευση τους στις δύο βάσεις για περαιτέρω ανάλυση επεξεργασία και οπτικοποίηση.

5.1 Προετοιμασία της InfluxDB

Στην παρούσα ενότητα παρουσιάζεται η διαδικασία εισαγωγής και οργάνωσης των δεδομένων χρονοσειρών στην InfluxDB, με εφαρμογή στα μετεωρολογικά δεδομένα που αναλύθηκαν στο Κεφάλαιο 4. Στόχος είναι να αποσαφηνιστεί ο τρόπος με τον οποίο η βάση δεδομένων αξιοποιεί βασικές έννοιες, όπως οι μετρήσεις, οι ετικέτες και τα πεδία, ώστε να διαχειρίζεται αποτελεσματικά μεγάλους όγκους δεδομένων που συνοδεύονται από χρονική σήμανση.

Η ενότητα εξετάζει πρώτα τα κύρια χαρακτηριστικά της αρχιτεκτονικής της InfluxDB, και στη συνέχεια παρουσιάζει τη χαρτογράφηση των δεδομένων των σταθμών στις αντίστοιχες δομές της. Ακολουθεί η ανάλυση των διαδικασιών δημιουργίας βάσης, ορισμού πολιτικών διατήρησης και μηχανισμών εισαγωγής των δεδομένων, συμπεριλαμβανομένης της χρήσης διεπαφών προγραμματισμού για την αποστολή μετρήσεων. Επιπλέον, παρουσιάζονται οι διαθέσιμες δυνατότητες ανάκτησης, επεξεργασίας και οπτικοποίησης που προσφέρει η πλατφόρμα, επιτρέποντας ολοκληρωμένη εποπτεία και αξιοποίηση των αποθηκευμένων χρονοσειρών.

Συνολικά, η ενότητα λειτουργεί ως σύνδεσμος μεταξύ της θεωρητικής παρουσίασης των δεδομένων χρονοσειρών και της πρακτικής υλοποίησής τους σε ένα σύστημα βάσης δεδομένων, προσφέροντας ένα συνεκτικό πλαίσιο κατανόησης της διαδικασίας εισαγωγής τους στην InfluxDB.

5.1.1 Γενική αρχιτεκτονική

Η InfluxDB είναι μια βάση δεδομένων βελτιστοποιημένη για αποθήκευση χρονοσειρών με υψηλό ρυθμό εισαγωγής δεδομένων. Η εγγραφή δεδομένων γίνεται χωρίς προκαθορισμένο σχήμα, καθώς κάθε νέα μέτρηση ή ετικέτα δημιουργείται αυτόματα κατά την πρώτη εισαγωγή.

Η εσωτερική δομή βασίζεται σε τέσσερα κύρια στοιχεία [13]:

Μετρήσεις: Η θεματική κατηγορία των δεδομένων (π.χ., θερμοκρασία, υγρασία).

Ετικέτες: Ζεύγη κλειδιού-τιμής για βελτιστοποίηση αναζήτησης (π.χ., `station_id=ST-002, unit_id=3`).

Πεδίο δεδομένων	Τύπος InfluxDB	Παράδειγμα
Station ID	Ετικέτα	station_id=ST-002
Unit ID	Ετικέτα	unit_id=3
Sensor ID	Ετικέτα	sensor_id=TEMP
Sensor Name	Ετικέτα	sensor_name=temperature_air
Value	Πεδίο	value=21.6
Measurement Unit	Πεδίο	measurement_unit="C"
Timestamp	Χρονοσφραγίδα	1678125000

Πίνακας 5.1: Αντιστοίχιση δεδομένων από τους σταθμούς στα στοιχεία της InfluxDB.

Πεδία: Οι πραγματικές αριθμητικές τιμές (π.χ., value=21.6).

Χρονοσφραγίδες: Ο χρόνος μέτρησης, σε εποχική μορφή (epoch).

5.1.2 Αντιστοίχιση των δεδομένων του Κεφαλαίου 4 σε τύπους της InfluxDB

Τα δεδομένα από τους μετεωρολογικούς σταθμούς του προηγούμενου κεφαλαίου περιλαμβάνουν τιμές όπως θερμοκρασία, υγρασία, πίεση, ταχύτητα και κατεύθυνση ανέμου, καθώς και πληροφορίες ταυτοποίησης σταθμού και μονάδας. Αυτές οι πληροφορίες χαρτογραφήθηκαν στους τύπους της InfluxDB (μετρήσεις, ετικέτες πεδία και χρονοσφραγίδες), όπως παρουσιάζονται στον Πίνακα 5.1.

5.1.3 Δημιουργία βάσης και πολιτικών διατήρησης

Η δημιουργία της βάσης δεδομένων και ο ορισμός πολιτικών διατήρησης αποτελούν θεμελιώδη βήματα για τη σωστή διαχείριση των δεδομένων χρονοσειρών στην InfluxDB. Σε αυτό το στάδιο καθορίζεται το περιβάλλον μέσα στο οποίο θα εισαχθούν οι μετρήσεις, ενώ παράλληλα ορίζονται οι κανόνες με τους οποίους τα δεδομένα θα διατηρούνται ή θα απομακρύνονται μετά από συγκεκριμένο χρονικό διάστημα.

Οι πολιτικές διατήρησης είναι ιδιαίτερα σημαντικές σε βάσεις δεδομένων που δέχονται μεγάλο όγκο συχνών μετρήσεων, καθώς επιτρέπουν τον αυτόματο καθαρισμό παλαιών δεδομένων, βελτιώνοντας την απόδοση και εξασφαλίζοντας ότι η βάση παραμένει αποδοτική και λειτουργική.

```

1 CREATE DATABASE meteo;
2 CREATE RETENTION POLICY rp_180d ON meteo
3 DURATION 180d REPLICATION 1 DEFAULT;

```

Σχήμα 5.1: Ορισμός βάσης και πολιτικής διατήρησης

Οι δύο εντολές που παρουσιάζονται στο Σχήμα 5.1 δημιουργούν το απαραίτητο περιβάλλον αποθήκευσης για τα δεδομένα χρονοσειρών που χρησιμοποιούμε και αναλύσαμε στο Κεφάλαιο 4. Η εντολή δημιουργίας της βάσης (CREATE DATABASE) ορίζει το λογικό χώρο αποθήκευσης μέσα στο οποίο θα αποθηκευτούν οι μετρήσεις, ενώ η εντολή καθορισμού πολιτικής διατήρησης (CREATE RETENTION POLICY) θέτει τον χρονικό ορίζοντα κατά τον οποίο τα δεδομένα παραμένουν προσβάσιμα.^[12]

Η συγκεκριμένη πολιτική διατήρησης, διάρκειας 180 ημερών, επιλέχθηκε επειδή ανταποκρίνεται στις ανάγκες του υπό εξέταση σεναρίου, όπου οι μετεωρολογικοί σταθμοί παράγουν συνεχή ροή δεδομένων μεγάλης ποσότητας. Η διατήρηση των δεδομένων για έξι μήνες επιτρέπει την πραγματοποίηση αναλύσεων εποχικότητας, εντοπισμού τάσεων και σύγκρισης περιόδων, χωρίς να επιβαρύνεται υπερβολικά ο αποθηκευτικός χώρος. Παράλληλα, ο αυτόματος καθαρισμός των παλαιότερων μετρήσεων συμβάλλει στη σταθερή απόδοση της βάσης, διατηρώντας το μέγεθός της σε λειτουργικά επίπεδα.

Με αυτόν τον τρόπο, η πολιτική διατήρησης εξισορροπεί δύο αντικρουόμενους στόχους: την ανάγκη να παραμένουν διαθέσιμα επαρκή ιστορικά δεδομένα για ανάλυση, και την ανάγκη να αποφεύγεται η συσσώρευση υπερβολικού όγκου πληροφοριών που θα επιβράδυνε τη λειτουργία της InfluxDB.

5.1.4 Έλεγχος διεπαφής HTTP API

Η εισαγωγή των μετρήσεων στην InfluxDB πραγματοποιείται μέσω του HTTP API, το οποίο επιτρέπει την αποστολή δεδομένων σε μορφή γραμμικού πρωτοκόλλου ^[13]. Η μέθοδος αυτή είναι ιδανική για αυτοματοποιημένες διαδικασίες συλλογής δεδομένων, καθώς παρέχει ευελιξία και ταχύτητα χωρίς ανάγκη σύνδεσης μέσω εξειδικευμένης εφαρμογής. Κάθε γραμμή δεδομένων περιλαμβάνει το όνομα της μέτρησης, της ετικέτας, του πεδίου και τη χρονοσφραγίδα, χωρισμένα με κόμματα και κενά. Η απλότητα της πα-

```

1 curl -i -XPOST 'http://localhost:8086/write?db=meteo' \
2 --data-binary 'stations,station_id=ST002,unit_id=3,sensor_id=TEMP \
3 value=21.6,measurement_unit="C" 1678125000000000000'

```

Σχήμα 5.2: Αποστολή μιας συγκεκριμένης μέτρησης ενός αισθητήρα στην InfluxDB

ραπάνω αναπαράστασης επιτρέπει την αποστολή δεδομένων ακόμη και από συστήματα με πολύ περιορισμένη υπολογιστική ισχύ, όπως οι μετεωρολογικοί σταθμοί που χρησιμοποιούνται στη μελέτη. Με τον όρο αποστολή δεδομένων εννοούμε ότι ο διακομιστής συλλέγει πληροφορίες είτε από προγράμματα δέσμης που εκτελούνται σε συγκεκριμένα χρονικά διαστήματα και αντλούν δεδομένα από το διαδίκτυο, είτε απευθείας από τους σταθμούς που στέλνουν τις μετρήσεις τους. Σε κάθε περίπτωση, η διαδικασία δεν απαιτεί σημαντική υπολογιστική ισχύ.

Η εντολή που παρουσιάζουμε στο Σχήμα 5.2 χρησιμοποιεί την υπηρεσία εγγραφής της βάσης (`write?db=meteo`) για να μεταφέρει μία εγγραφή με δεδομένα, η οποία περιλαμβάνει το όνομα της μέτρησης, τις αντίστοιχες ετικέτες, τις αριθμητικές τιμές και τη χρονοσφραγίδα της παρατήρησης. Με τον τρόπο αυτό επιτυγχάνεται η άμεση καταχώριση μίας νέας εγγραφής χωρίς τη χρήση ενδιάμεσου λογισμικού ή εφαρμογής, αξιοποιώντας μια απλή κλήση HTTP [11]. Η παραπάνω μέθοδος επιτρέπει τον έλεγχο και την επαλήθευση της διαδικασίας εισαγωγής, ενώ παράλληλα διευκολύνει την αυτοματοποίηση, καθώς μπορεί να ενσωματωθεί εύκολα σε προγραμματισμένες εργασίες ή συστήματα συλλογής δεδομένων.

5.1.5 Ανάκτηση και ανάλυση

Η InfluxDB παρέχει εκτεταμένες δυνατότητες για την ανάκτηση και ανάλυση των αποθηκευμένων χρονοσειρών, επιτρέποντας την εξαγωγή χρήσιμων στατιστικών και τη διερεύνηση τάσεων ή αποκλίσεων στις μετρήσεις. Για παράδειγμα, μπορούμε να υπολογίσουμε τη μέση ωριαία θερμοκρασία των τελευταίων είκοσι τεσσάρων ωρών για κάθε σταθμό.

Το ερώτημα που πραγματοποιούμε στο Σχήμα 5.3 ομαδοποιεί τα δεδομένα με βάση το χρόνο (`GROUP BY time(1h)`) και υπολογίζει αξιοποιεί το μέσο όρο των τιμών τους

```
1 SELECT mean(value)
2 FROM stations
3 WHERE time > now() - 24h
4 AND sensor_id='TEMP'
5 GROUP BY time(1h);
```

Σχήμα 5.3: Αποστολή μιας συγκεκριμένης μέτρησης ενός αισθητήρα στην InfluxDB

`mean(value)`. Με αυτόν τον τρόπο δημιουργείται ένα συνοπτικό προφίλ θερμοκρασίας ανά σταθμό, το οποίο διευκολύνει την ανάλυση των ημερήσιων διακυμάνσεων αλλά και την αναγνώριση πιθανών ανωμαλιών ή απότομων μεταβολών στη συμπεριφορά των μετρήσεων.

Η ομαδοποίηση ανά χρονικά διαστήματα αποτελεί μία από τις πιο χρήσιμες λειτουργίες της InfluxDB, καθώς μειώνει τον όγκο των δεδομένων και παρέχει συγκεντρωτικές μετρήσεις σε πραγματικό χρόνο.

Η δυνατότητα αυτή αξιοποιείται στη συνέχεια για την παραγωγή γραφημάτων και στατιστικών αναφορών στο περιβάλλον του Grafana, διευκολύνοντας τη σύγκριση μεταξύ σταθμών και την παρακολούθηση τάσεων σε βάθος χρόνου.

5.2 Προετοιμασία της TimescaleDB

Η ενότητα αναλύει τον τρόπο με τον οποίο τα δεδομένα που προέρχονται από τους μετεωρολογικούς σταθμούς χαρτογραφούνται στις οντότητες της TimescaleDB, καθώς και τα βήματα που απαιτούνται για τη δημιουργία των απαραίτητων δομών, όπως το σχήμα της βάσης και τους υπερπίνακες. Παρουσιάζονται, επίσης, οι διαθέσιμες μέθοδοι εισαγωγής δεδομένων και τα εργαλεία για την εκτέλεση ερωτημάτων και την ανάλυση των μετρήσεων.

Τέλος, εξετάζονται οι πολιτικές συμπίεσης και σύνοψης, οι οποίες συμβάλλουν στη μείωση του αποθηκευτικού χώρου και στη βελτίωση της απόδοσης, καθιστώντας την TimescaleDB κατάλληλη επιλογή για συστήματα που απαιτούν συνεχή και υψηλής ακρίβειας διαχείριση χρονοσειρών.

5.2.1 Γενική αρχιτεκτονική

Η TimescaleDB αποτελεί επέκταση της PostgreSQL για αποθήκευση χρονοσειρών. Διατηρεί τη σχεσιακή λογική (πίνακες, κλειδιά, σχέσεις) ενώ εισάγει νέες έννοιες όπως οι υπερπίνακες, η τμηματοποίηση και οι συνεχείς συναθροίσεις. Το πλεονέκτημά της είναι ότι επιτρέπει ταυτόχρονα κλασικά SQL ερωτήματα και βελτιστοποιήσεις για δεδομένα χρονοσειρών.

5.2.2 Αντιστοίχιση δεδομένων σε οντότητες TimescaleDB

Η χαρτογράφηση των ίδιων δεδομένων στην TimescaleDB ακολουθεί το σχεσιακό μοντέλο. Πριν την υλοποίηση των υπερπινάκων της TimescaleDB, είναι απαραίτητο να περιγραφεί το εννοιολογικό και λογικό σχήμα των δεδομένων που αποθηκεύονται. Το σύστημα βασίζεται σε ένα σχεσιακό μοντέλο, στο οποίο διακριτές οντότητες αναπαριστούν τους μετεωρολογικούς σταθμούς, τις μονάδες μέτρησης και τους αισθητήρες, καθώς και τις μετρήσεις που παράγονται από αυτούς.

Κάθε σταθμός (station) μπορεί να περιλαμβάνει περισσότερες από μία μονάδες (units), οι οποίες αντιστοιχούν σε φυσικές εγκαταστάσεις ή υποσυστήματα του σταθμού. Κάθε μονάδα διαθέτει πολλαπλούς αισθητήρες (sensors), οι οποίοι μετρούν διαφορετικές φυσικές μεταβλητές, όπως θερμοκρασία, υγρασία ή ταχύτητα ανέμου.

Οι μετρήσεις αποτελούν χρονοσειρές τιμών που παράγονται από έναν συγκεκριμένο αισθητήρα και χαρακτηρίζονται από χρονική σφραγίδα, αριθμητική τιμή και μονάδα μέτρησης. Για τον λόγο αυτό, οι μετρήσεις αποθηκεύονται σε ξεχωριστό πίνακα, ο οποίος συνδέεται με τους αισθητήρες μέσω ξένων κλειδιών.

Το λογικό σχήμα υλοποιείται μέσω πινάκων που αναπαριστούν τις παραπάνω οντότητες και τις μεταξύ τους συσχετίσεις. Ειδικότερα, ο πίνακας `stations_units_sensors` λειτουργεί ως πίνακας συσχέτισης, ο οποίος προσδιορίζει μοναδικά κάθε συνδυασμό σταθμού, μονάδας και αισθητήρα. Ο πίνακας `measurements` αποθηκεύει τις χρονοσειρές και αναφέρεται σε αυτόν τον συνδυασμό μέσω του πεδίου `station_unit_sensor_id`.

Με αυτόν τον τρόπο, το μοντέλο παραμένει πλήρως σχεσιακό, με σαφώς ορισμένες οντότητες και σχέσεις, ενώ η TimescaleDB αξιοποιείται για τη βελτιστοποιημέ-

Πεδίο δεδομένων	Πίνακας / Στήλη	Παράδειγμα
Station ID	stations.id	ST-002
Station Name	stations.name	TRIPOLI-02
Unit ID	units.id	3
Sensor ID	sensors.id	TEMP
Sensor Name	sensors.name	temperature_air
Value	measurements.value	21.6
Measurement Unit	sensors.measurement_unit	C
Timestamp	measurements.timestamp	2023-03-06 18:44:20

Πίνακας 5.2: Χαρτογράφηση λογικού μοντέλου μετεωρολογικών δεδομένων σε πίνακες της TimescaleDB.

νη αποθήκευση και επεξεργασία των χρονοσειρών μέσω της μετατροπής του πίνακα `measurements` σε υπερπίνακα. Η αντιστοίχιση των πληροφοριών των σταθμών αποτυπώνεται στον Πίνακα 5.2.

5.2.3 Δημιουργία σχήματος και υπερπινάκων

Για την αποθήκευση των μετρήσεων στην TimescaleDB απαιτείται πρώτα ο ορισμός του σχεσιακού σχήματος και στη συνέχεια η μετατροπή του κύριου πίνακα σε υπερπίνακα. Ο Πίνακας 5.2 δεν αποσκοπεί στην παρουσίαση ενός πλήρους διαγράμματος οντοτήτων-σχέσεων, αλλά στην αποτύπωση της χαρτογράφησης μεταξύ του λογικού μοντέλου των δεδομένων και της υλοποίησής του σε σχεσιακούς πίνακες της TimescaleDB.

Οι βασικές οντότητες του συστήματος (σταθμοί, μονάδες και αισθητήρες) υλοποιούνται ως ξεχωριστοί πίνακες, ενώ οι μεταξύ τους σχέσεις αποτυπώνονται μέσω πινάκων συσχέτισης. Συγκεκριμένα, η σχέση μεταξύ σταθμών, μονάδων και αισθητήρων υλοποιείται μέσω του πίνακα `stations_units_sensors`, ο οποίος προσδιορίζει μοναδικά κάθε ενεργό αισθητήρα στο σύστημα.

Οι μετρήσεις δεν αποθηκεύονται απευθείας στις οντότητες αυτές, αλλά σε ξεχωριστό πίνακα χρονοσειρών, ο οποίος αναφέρεται στον αντίστοιχο συνδυασμό σταθμού-μονάδας-αισθητήρα μέσω ξένου κλειδιού(*foreign key*). Με αυτόν τον τρόπο διαχωρίζεται καθαρά η στατική πληροφορία του συστήματος από τα δυναμικά δεδομένα χρονο-

σειρών.

Παρότι η TimescaleDB εισάγει την έννοια του υπερπίνακα (hypertable), το μοντέλο δεδομένων παραμένει πλήρως σχεσιακό. Οι πίνακες, τα πρωτεύοντα και ξένα κλειδιά, καθώς και η χρήση της γλώσσας SQL, δεν αλλάζουν σε επίπεδο σχεδιασμού σε σχέση με μία κλασική βάση PostgreSQL.

Η βασική διαφορά εντοπίζεται στο επίπεδο της φυσικής οργάνωσης των δεδομένων. Ένας υπερπίνακας λειτουργεί ως λογική όψη (logical abstraction) ενός συνόλου εσωτερικών πινάκων, οι οποίοι χωρίζονται αυτόματα σε τμήματα με βάση τη χρονική διάσταση. Κάθε τμήμα αποθηκεύει δεδομένα για ένα συγκεκριμένο χρονικό διάστημα και διαχειρίζεται ανεξάρτητα από τα υπόλοιπα.

Σε αντίθεση με έναν συμβατικό σχεσιακό πίνακα, όπου όλες οι εγγραφές αποθηκεύονται σε ενιαία δομή, ο υπερπίνακας επιτρέπει στη βάση να περιορίζει το εύρος των δεδομένων που εξετάζονται κατά την εκτέλεση ερωτημάτων χρονοσειρών. Αυτό έχει ως αποτέλεσμα τη σημαντική βελτίωση της απόδοσης τόσο στις εισαγωγές νέων μετρήσεων όσο και στις αναζητήσεις που αφορούν χρονικά φίλτρα.

Συνεπώς, η χρήση υπερπινάκων δεν τροποποιεί το σχεσιακό σχήμα ή τον τρόπο με τον οποίο ο χρήστης αλληλεπιδρά με τα δεδομένα μέσω SQL, αλλά αποτελεί μια εσωτερική βελτιστοποίηση αποθήκευσης και εκτέλεσης ερωτημάτων, προσαρμοσμένη στις απαιτήσεις των δεδομένων χρονοσειρών.

Ο υπερπίνακας αποτελεί την κεντρική δομή της TimescaleDB για τη διαχείριση χρονοσειρών και υλοποιείται εσωτερικά ως τμήματα κατακερματισμού που χωρίζονται βάσει χρόνου.

Ο κώδικας του Σχήματος 5.4 υλοποιεί τα απαιτούμενα βήματα για τη μετατροπή ενός απλού σχεσιακού πίνακα σε υπερπίνακα χρονοσειρών στην TimescaleDB. Αρχικά, ενεργοποιεί την επέκταση timescaledb, η οποία προσθέτει στη PostgreSQL τις λειτουργίες που απαιτούνται για την αποδοτική διαχείριση χρονοσειρών. Στη συνέχεια δημιουργείται ο πίνακας measurements, όπου κάθε εγγραφή αντιστοιχεί σε μία μέτρηση από έναν συγκεκριμένο αισθητήρα· το πεδίο timestamp χρησιμοποιείται ως χρονική διάσταση, ενώ το station_unit_sensor_id προσδιορίζει ποιος αισθητήρας παρήγαγε τη μέτρηση. Τέλος, η εντολή create_hypertable μετατρέπει τον κεντρικό πίνακα σε

```

1 CREATE EXTENSION IF NOT EXISTS timescaledb;
2 CREATE TABLE measurements (
3     station_unit_sensor_id INT NOT NULL,
4     timestamp TIMESTAMPTZ NOT NULL,
5     value DOUBLE PRECISION NOT NULL,
6     measurement_unit TEXT,
7     PRIMARY KEY (station_unit_sensor_id, timestamp)
8 );
9 SELECT create_hypertable('measurements', 'timestamp',
10     chunk_time_interval => interval '7 day');

```

Σχήμα 5.4: Δημιουργία πινάκων και υπερπινάκων στη TimescaleDB.

υπερπίνακα με τμήματα (chunks) διάρκειας επτά ημερών [23], επιτρέποντας στη βάση δεδομένων να οργανώνει, να αποθηκεύει και να αναχτά μεγάλους όγκους χρονοσειρών με υψηλή απόδοση.

5.2.4 Ερωτήματα και ανάλυση

Η TimescaleDB αξιοποιεί πλήρως τη δύναμη της γλώσσας SQL για την ανάλυση χρονοσειρών, προσφέροντας εξειδικευμένες συναρτήσεις για ομαδοποίηση δεδομένων σε χρονικά διαστήματα. Μέσω τέτοιων ερωτημάτων SQL μπορούμε να υπολογίσουμε στατιστικούς δείκτες (π.χ., μέσους όρους, μέγιστες και ελάχιστες τιμές) και να παρακολουθήσουμε την εξέλιξη των μετρήσεων σε βάθος χρόνου. Η ενσωμάτωση με τις υπόλοιπες οντότητες της βάσης, σταθμούς και αισθητήρες, επιτρέπει πολύπλοκες αναλύσεις, όπως τη σύγκριση μεταξύ διαφορετικών σταθμών ή αισθητήρων, παρέχοντας έτσι ένα ισχυρό εργαλείο επιστημονικής και επιχειρησιακής αξιολόγησης. Το παράδειγμα του Σχήματος 5.5 παρουσιάζει ένα χαρακτηριστικό ερώτημα ανάλυσης χρονοσειρών στην TimescaleDB, όπου οι μετρήσεις θερμοκρασίας ομαδοποιούνται σε ωριαία διαστήματα μέσω της συνάρτησης `time_bucket`. Με αυτόν τον τρόπο υπολογίζεται ο μέσος όρος των τιμών για κάθε ώρα του τελευταίου 24ώρου, επιτρέποντας την εξαγωγή μίας πιο εξομαλυμένης χρονοσειράς που αποτυπώνει τη γενική τάση της μεταβλητής. Η χρήση συναθροιστικών συναρτήσεων σε συνδυασμό με τη χωρική συσχέτιση των πινάκων αισθητήρων και μετρήσεων επιτρέπει την παραγωγή πολύτιμων δεικτών και συνοπτικών στατιστικών, καθιστώντας το σύστημα κατάλληλο για την επιστημονική ανάλυση των

```

1 SELECT time_bucket('1 hour', timestamp) AS hour,
2        avg(value) AS avg_temp
3 FROM measurements m JOIN stations_units_sensors s
4        ON m.station_unit_sensor_id = s.id
5 WHERE s.sensor_id = 'TEMP'
6        AND m.timestamp > now() - interval '1 day'
7 GROUP BY hour
8 ORDER BY hour;

```

Σχήμα 5.5: Υπολογισμός μέσου όρου τιμών.

δεδομένων και την αξιολόγηση των περιβαλλοντικών συνθηκών.

5.2.5 Πολιτικές συμπίεσης και σύνοψης

Οι μηχανισμοί συμπίεσης και σύνοψης της TimescaleDB δεν είναι απαραίτητοι για τη βασική λειτουργία αποθήκευσης και ανάκτησης των δεδομένων χρονοσειρών, ωστόσο παρέχουν πρόσθετες δυνατότητες βελτιστοποίησης σε περιβάλλοντα με μεγάλο όγκο ιστορικών δεδομένων. Στο πλαίσιο της παρούσας εργασίας παρουσιάζονται ενδεικτικά, με σκοπό την ανάδειξη των δυνατοτήτων της πλατφόρμας και όχι ως αναγκαία προϋπόθεση για τη λειτουργία του συστήματος.

Εναλλακτικές προσεγγίσεις για τη διαχείριση μεγάλου όγκου δεδομένων περιλαμβάνουν τη χρήση πολιτικών διατήρησης (retention policies), την εκτέλεση συνθηροιστικών ερωτημάτων κατά τον χρόνο ανάκτησης των δεδομένων, ή τη δημιουργία προϋπολογισμένων συνόψεων σε ξεχωριστούς πίνακες. Οι μηχανισμοί που παρέχει η TimescaleDB, επιτρέπουν την αυτοματοποίηση αυτών των διαδικασιών, μειώνοντας το υπολογιστικό κόστος σε μακροχρόνια χρήση.

Στην παρούσα υλοποίηση, η συμπίεση εφαρμόζεται αποκλειστικά σε παλαιότερα δεδομένα, ώστε να περιοριστεί το αποθηκευτικό αποτύπωμα χωρίς να επηρεάζεται η απόδοση των ερωτημάτων που αφορούν πρόσφατες μετρήσεις.

```

1 ALTER TABLE measurements SET (
2   timescaledb.compress,
3   timescaledb.compress_orderby='timestamp DESC'
4 );
5 SELECT add_compression_policy('measurements', INTERVAL '90 days');

```

Σχήμα 5.6: Προσθήκη πολιτικής συμπίεσης δεδομένων.

5.3 Εισαγωγή δεδομένων και αρχική παραμετροποίηση

Η εισαγωγή δεδομένων πραγματοποιείται ζητώντας τα δεδομένα από το περιβάλλον διεπαφής των σταθμών και στη συνέχεια εισάγοντας τα δεδομένα αυτά στις πλατφόρμες αποθήκευσης. Πριν όμως ζητήσουμε τα δεδομένα, ελέγχουμε τη χρονοσφραγίδα της τελευταίας μας εγγραφής ώστε να ζητήσουμε μόνο τα δεδομένα από εκείνη τη χρονική στιγμή και μέχρι τη στιγμή του ερωτήματος. Ως αποτέλεσμα, εξασφαλίζεται ότι δε θα υπάρχουν διπλοεγγραφές.

Το αρχείο παραμετροποίησης οργανώνεται σε διακριτές ενότητες, οι οποίες περιγράφουν τις πηγές δεδομένων και τις πλατφόρμες αποθήκευσης. Αρχικά, ορίζονται οι παράμετροι των μετεωρολογικών σταθμών, όπως το αναγνωριστικό κάθε σταθμού, η διεπαφή πρόσβασης (API) και οι γεωγραφικές συντεταγμένες των μονάδων μέτρησης.

Στη συνέχεια καθορίζονται οι παράμετροι της διαδικασίας συλλογής δεδομένων (scraping), όπως η χρονική ζώνη και το μέγεθος των δεδομένων που ανακτώνται σε κάθε αίτημα. Τέλος, περιλαμβάνονται οι παράμετροι σύνδεσης για τις πλατφόρμες αποθήκευσης, επιτρέποντας την εύκολη αλλαγή ή επέκταση του συστήματος σε πρόσθετες βάσεις δεδομένων χωρίς τροποποίηση του κώδικα.

Για λόγους αναγνωσιμότητας, στο παρόν κεφάλαιο παρουσιάζονται μόνο ενδεικτικά αποσπάσματα του αρχείου παραμετροποίησης.

5.3.1 Αρχική παραμετροποίηση

Αρχικά, ορίζουμε ένα αρχείο παραμετροποίησης, το οποίο περιέχει τις απαραίτητες πληροφορίες για την πρόσβαση στις πλατφόρμες TimescaleDB και InfluxDB. Επιπλέον,

ον, περιέχονται τα στοιχεία πρόσβασης στη διεπαφή πρόσβασης (API). Τα (ενδεικτικά) περιεχόμενα του αρχείου βλέπουμε στα Σχήματα 5.1, 5.2 και 5.3. Με τον τρόπο αυτό διασφαλίζουμε τη δυνατότητα προσθήκης περαιτέρω πλατφόρμας αποθήκευσης στο μέλλον, αν αυτό απαιτηθεί.

Η χρήση κεντρικού αρχείου ρυθμίσεων επιτρέπει τη σαφή απομόνωση της παραμετροποίησης από τη λογική του κώδικα και διευκολύνει τη συντήρηση και επέκταση του συστήματος.

Αρχικά, στο αρχείο παραμετροποίησης ορίζονται οι παράμετροι των μετεωρολογικών σταθμών, όπως το αναγνωριστικό κάθε σταθμού, το σημείο πρόσβασης στο API και οι μονάδες μέτρησης που περιλαμβάνει. Για κάθε μονάδα καταγράφονται προαιρετικά γεωγραφικές συντεταγμένες και πληροφορίες χαρτογράφησης, οι οποίες αξιοποιούνται σε μεταγενέστερα στάδια ανάλυσης και οπτικοποίησης. Στο Σχήμα 5.1 παρουσιάζουμε την δήλωση των παραμέτρων του πρώτου σταθμού.

```
1 stations = {  
2   1: {  
3     'url': 'https://live.netsens.it/export/xml_export_1A.php',  
4     'username': 'XXXXX',  
5     'password': 'XXXXX',  
6     'min_date': '2022-12-21',  
7     'station_id': 2736,  
8     'units_locs': [  
9       {  
10        'unit_id': 1,  
11        'lat': 37.80686,  
12        'lon': 22.72398,  
13        'google_maps': 'https://www.google.com/maps/d/u/0/edit?mid=  
XXXXX=en&ll=37.806968045011345%2C22.724249982889376&z=18',  
14        'open_street_map': ''  
15      }  
16    ]  
}
```

17 }

Σχήμα 5.1: Δήλωση σταθμών στο αρχείο παραμέτρων.

Στη συνέχεια καθορίζονται οι παράμετροι της διαδικασίας συλλογής δεδομένων (Σχήμα 5.2), όπως η χρονική ζώνη αναφοράς και το μέγεθος των χρονικών παρτίδων δεδομένων που ανακτώνται σε κάθε αίτημα. Η τμηματική ανάκτηση δεδομένων επιτρέπει την αποδοτική διαχείριση μεγάλων χρονικών διαστημάτων χωρίς υπερφόρτωση του εξυπηρετητή ή των πλατφορμών αποθήκευσης.

```

1 scraping_params = {
2     'datetime_format': '%Y-%m-%d %H:%M:%S',
3     'timezone': 'Europe/Rome',
4     'batch': {
5         'days': 3,
6         'hours': 0,
7         'minutes': 0,
8         'seconds': 0,
9     }
10 }
```

Σχήμα 5.2: Διαδικασίας συλλογής δεδομένων στο αρχείο παραμέτρων.

Τέλος, περιλαμβάνονται οι παράμετροι σύνδεσης των πλατφορμών αποθήκευσης, συγκεκριμένα της InfluxDB και της TimescaleDB. Οι παράμετροι αυτές περιγράφουν τα στοιχεία σύνδεσης και τα βασικά χαρακτηριστικά αποθήκευσης, επιτρέποντας στο ίδιο σύστημα εισαγωγής να υποστηρίζει περισσότερες από μία βάσεις δεδομένων με ενιαίο τρόπο και παρουσιάζονται στο Σχήμα 5.3.

```

1 INFLUX = {
2     'url' : 'https://influx.sodasense.uop.gr',
3     'organization' : 'XXXXXXXXXXXXXXXXX',
4     'access_token' : 'XXXXXXXXXXXXXXXXX',
5     'bucket' : 'stations',
6     'measurement' : 'any',
7 }
```

```

8 PG = {
9   'host'      : 'dev.sodasense.uop.gr',
10  'port'      : 'XXXX',
11  'dbname'    : 'stations',
12  'user'      : 'XXXX',
13  'password' : 'XXXX',
14  }

```

Σχήμα 5.3: Στοιχεία σύνδεσης στο αρχείο παραμέτρων.

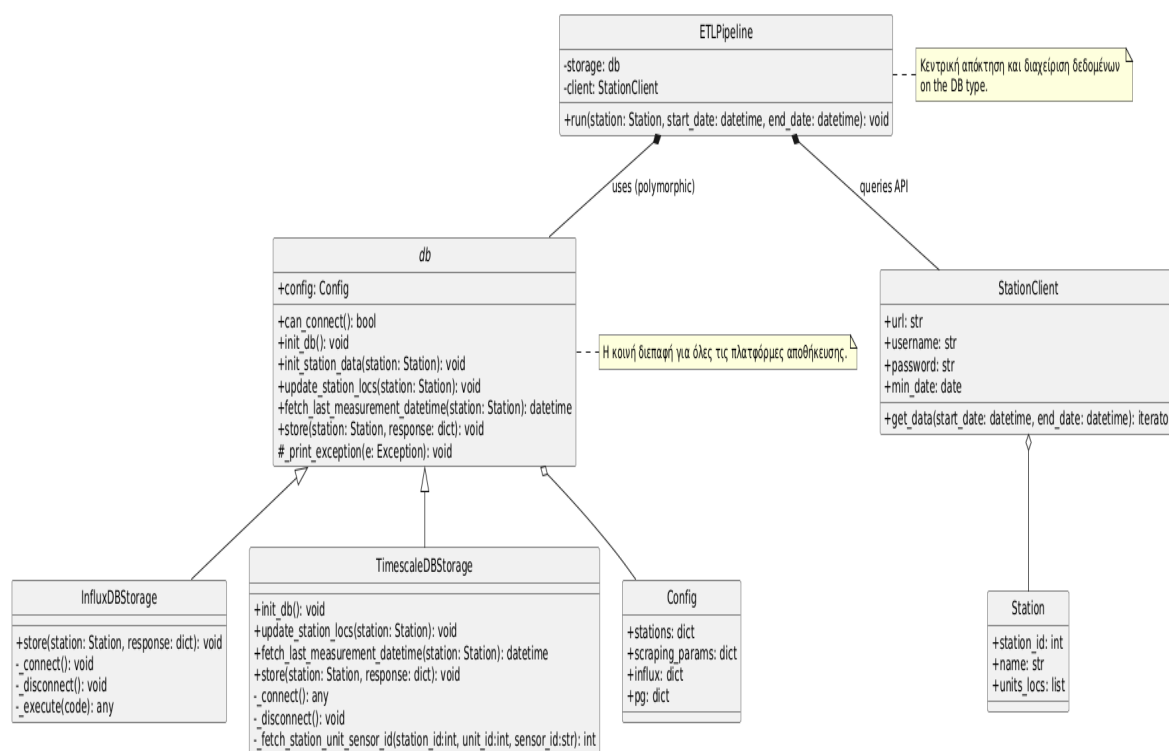
5.3.2 Δημιουργία κλάσεων

Πριν την παρουσίαση των επιμέρους κλάσεων διαχείρισης των βάσεων δεδομένων, εισάγεται μία γενική αφηρημένη κλάση, η οποία ορίζει ένα κοινό σύνολο λειτουργιών για όλες τις πλατφόρμες αποθήκευσης που χρησιμοποιούνται στο σύστημα. Η κλάση αυτή δεν υλοποιεί συγκεκριμένη λογική αποθήκευσης, αλλά λειτουργεί ως ενιαίο σημείο αναφοράς για την αλληλεπίδραση με διαφορετικά συστήματα βάσεων δεδομένων.

Στο Σχήμα 5.7 παρουσιάζεται η δομή των κλάσεων και οι μεταξύ τους σχέσεις σε UML. Μία αφηρημένη κλάση ορίζει το κοινό σύνολο λειτουργιών που απαιτούνται για την αποθήκευση δεδομένων χρονοσειρών, το οποίο υλοποιείται από δύο εξειδικευμένες κλάσεις, μία για τη βάση δεδομένων InfluxDB και μία για τη βάση δεδομένων TimescaleDB. Η ροή της διαδικασίας εξαγωγής–μετασχηματισμού–φόρτωσης συνεργάζεται πολυμορφικά με την εκάστοτε υλοποίηση αποθήκευσης, χωρίς να εξαρτάται από τον τύπο της βάσης δεδομένων. Η άντληση των δεδομένων πραγματοποιείται από ξεχωριστό συστατικό επικοινωνίας με τις διεπαφές των σταθμών, επιτυγχάνοντας σαφή διαχωρισμό μεταξύ συλλογής δεδομένων και μηχανισμού αποθήκευσης.

Συγκεκριμένα, η γενική κλάση ορίζει βασικές μεθόδους όπως τον έλεγχο σύνδεσης, την αρχικοποίηση της βάσης, την ανάκτηση της χρονικής στιγμής της τελευταίας αποθηκευμένης μέτρησης και την αποθήκευση νέων δεδομένων. Κάθε επιμέρους πλατφόρμα (InfluxDB και TimescaleDB) υποχρεούται να υλοποιήσει αυτές τις μεθόδους, προσαρμόζοντάς τις στα ιδιαίτερα χαρακτηριστικά της.

Με αυτόν τον σχεδιασμό επιτυγχάνεται ο διαχωρισμός της επιχειρησιακής λογικής



Σχήμα 5.7: Διάγραμμα κλάσεων του υποσυστήματος αποθήκευσης σε UML.

συλλογής δεδομένων από τις λεπτομέρειες της εκάστοτε βάσης δεδομένων. Η διαδικασία εισαγωγής δεδομένων παραμένει κοινή και ανεξάρτητη από το σύστημα αποθήκευσης, ενώ η προσθήκη νέας πλατφόρμας στο μέλλον μπορεί να πραγματοποιηθεί χωρίς τροποποίηση του βασικού κώδικα συλλογής. Η γενική αυτή κλάση λειτουργεί αντίστοιχα με μια διεπαφή στον αντικειμενοστραφή σχεδιασμό. Στο Σχήμα 5.4 παραθέτουμε τη γενική αρχικοποιημένη κλάση.

```

1 class db():
2     def __init__(self, config):
3         self.config = config
4
5     def can_connect(self, *args, **kwargs):
6         raise Exception("Not implemented method")
7
8     def init_db(self, *args, **kwargs):

```

```

9 raise Exception("Not implemented method")
10
11 def init_station_data(self, station, *args, **kwargs):
12     raise Exception("Not implemented method")
13
14 def update_station_locs(self, station, *args, **kwargs):
15     raise Exception("Not implemented method")
16
17 def fetch_last_measurement_datetime(self, station, *args, **kwargs):
18     raise Exception("Not implemented method")
19
20 def store(self, station, response_dict, *args, **kwargs):
21     raise Exception("Not implemented method")
22
23 def _print_exception(self, e):
24     print(type(e).__name__, e)

```

Σχήμα 5.4: Γενική κλάση διαχείρισης βάσεων δεδομένων.

Στη συνέχεια δημιουργούμε τις εξαρτώμενες κλάσεις για κάθε πλατφόρμα.

5.3.3 Δημιουργία κλάσης για τη InfluxDB

Η κλάση διαχείρισης της InfluxDB υλοποιεί το κοινό σύνολο λειτουργιών που ορίζονται από τη γενική κλάση βάσης δεδομένων, προσαρμοσμένο στο μοντέλο αποθήκευσης χρονοσειρών της InfluxDB. Κύριος ρόλος της είναι η μετατροπή των δεδομένων που λαμβάνονται από τους μετεωρολογικούς σταθμούς στη μορφή που απαιτεί το γραμμικό πρωτόκολλο της InfluxDB και η αποδοτική αποστολή τους μέσω του HTTP API.

Η κλάση αναλαμβάνει τον έλεγχο σύνδεσης με τον διακομιστή, τη διαχείριση των buckets αποθήκευσης και την εισαγωγή των μετρήσεων με κατάλληλες ετικέτες, πεδία και χρονοσφραγίδες. Η επιλογή χρήσης ετικετών για τα αναγνωριστικά σταθμών, μονάδων και αισθητήρων επιτρέπει την αποδοτική φιλτραρισμένη ανάκτηση δεδομένων και τη μετέπειτα οπτικοποίηση τους.

Οι επιμέρους λεπτομέρειες υλοποίησης, όπως η διαχείριση συνδέσεων και η μετατροπή των χρονικών σφραγίδων, δεν είναι κρίσιμες για την κατανόηση της συνολικής αρχιτεκτονικής και παρατίθενται για λόγους πληρότητας. Στο Σχήμα 5.5 παραθέτουμε ένα ενδεικτικό τμήμα την κλάση της InfluxDB.

```

1 class influx(db):
2     def store(self, station, response_dict, *args, **kwargs):
3         def code():
4             points = []
5
6             for unit in response_dict['netsens']['stazione']['unita']:
7                 for sensor in unit['sensore']:
8                     measurements = sensor.get('misura', [])
9                     measurements = [measurements] if isinstance(measurements
, dict) else measurements
10
11                     for m in measurements:
12                         p = Point(measurement_name=self.measurement)
13                         p.time(int(m['timestamp']), WritePrecision.S) # UTC/
GMT
14                         p.field('recorded', float(m['valore']))
15                         p.tag('station_id', response_dict['
netsens']['stazione']['id'])
16                         p.tag('station_name', response_dict['netsens']['
stazione']['nome'])
17                         p.tag('unit_id', unit['id'])
18                         p.tag('unit_name', unit['nome'])
19                         p.tag('sensor_id', sensor['id'])
20                         p.tag('sensor_name', sensor['nome'])
21                         p.tag('sensor_unit', sensor['unita']) # unit of
measurement
22
23                     points.append(p)
24
25             write_api = self.client.write_api()
26             write_api.write(bucket=self.bucket, record=points, write_options

```

```

=SYNCHRONOUS)
24         write_api.close()
25         self._execute(code)
26     def _execute(self, code) -> any:
27         self._connect()
28         result = code()
29         self._disconnect()
30         return result

```

Σχήμα 5.5: Κλάση διαχείρισης για τη InfluxDB.

5.3.4 Δημιουργία κλάσης για την TimescaleDB

Η κλάση διαχείρισης της TimescaleDB υλοποιεί τις ίδιες βασικές λειτουργίες εισαγωγής δεδομένων, ακολουθώντας όμως το σχεσιακό μοντέλο αποθήκευσης της PostgreSQL. Η κλάση είναι υπεύθυνη για την αρχικοποίηση του σχήματος της βάσης, τη διαχείριση των οντοτήτων σταθμών, μονάδων και αισθητήρων, καθώς και την εισαγωγή των μετρήσεων στον υπερπίνακα χρονοσειρών.

Κατά τη διαδικασία εισαγωγής, τα δεδομένα που λαμβάνονται από τους σταθμούς μετασχηματίζονται ώστε να αντιστοιχιστούν σε υπάρχοντες συνδυασμούς σταθμού–μονάδας–αισθητήρα, διατηρώντας τη συνέπεια του σχεσιακού σχήματος. Η αποθήκευση των μετρήσεων πραγματοποιείται με τρόπο που αποτρέπει διπλοεγγραφές και εξασφαλίζει την ορθότητα των χρονοσειρών.

Στο Σχήμα 5.6 παραθέτουμε την αντίστοιχη (ενδεικτική) κλάση για την εισαγωγή δεδομένων στην TimescaleDB.

```

1 def update_station_locs(self, station):
2     self._update_station_units_locations(station)
3
4 def fetch_last_measurement_datetime(self, station):
5     conn = self._connect()
6
7 def _init_station_data(self, response_dict):

```

```

8      station_data = [(
9          response_dict['netsens']['stazione']['id'],
10         response_dict['netsens']['stazione']['nome'],
11         response_dict['netsens']['stazione']['last_measure']
12     )]
13
14 def _update_station_units_locations(self, station):
15     print('\tStation {}. Updating stations_units
16
17 def _fetch_last_measurement_datetime(self, station):
18     conn = self._connect()
19
20 def _fetch_station_unit_sensor_id(self, station_id, unit_id, sensor_id):
21
22 def _update_last_measurements(self, response_dict):
23     station_id = int(response_dict['netsens']['stazione']['id'])
24     station_last_measurement = int(response_dict['netsens']['stazione']['
last_measure'])

```

Σχήμα 5.6: Κλάση διαχείρισης για την TimescaleDB.

5.3.5 Εισαγωγή δεδομένων

Μετά τη δημιουργία των κλάσεων για τη διαχείριση της εκάστοτε πλατφόρμας αποθήκευσης δεδομένων, υλοποιείται το σενάριο απόκτησης των δεδομένων από τους μετεωρολογικούς σταθμούς και η εισαγωγή τους στις βάσεις δεδομένων. Το σενάριο αυτό λειτουργεί ως συνδετικός κρίκος μεταξύ της διαδικασίας συλλογής των δεδομένων και των μηχανισμών αποθήκευσης που παρουσιάστηκαν στις προηγούμενες ενότητες.

Στο Σχήμα 5.7 παρουσιάζεται ενδεικτικό απόσπασμα του κώδικα, μέσω του οποίου πραγματοποιούνται περιοδικά ερωτήματα προς τις διεπαφές των σταθμών. Οι απαντήσεις λαμβάνονται σε μορφή JSON και περιέχουν τις μετρήσεις που πρόκειται να αποθηκευτούν, συμπεριλαμβανομένης της χρονικής σφραγίδας, της αριθμητικής τιμής και της μονάδας

μέτρησης κάθε παρατήρησης.

Η διαδικασία είναι πλήρως αυτοματοποιημένη και ενσωματώνεται στο στάδιο φόρτωσης (loading) της αλυσίδας επεξεργασίας ETL. Ο ρόλος του σεναρίου περιορίζεται στον συντονισμό της ροής των δεδομένων, χωρίς να επηρεάζει τη λογική αποθήκευσης που υλοποιείται στις επιμέρους κλάσεις των βάσεων δεδομένων. Με τον τρόπο αυτό διασφαλίζεται η συνεχής και αξιόπιστη ενημέρωση των βάσεων με νέες μετρήσεις από τους σταθμούς.

```

1 def get_data(self, start_date, end_date):
2
3     batch    = timedelta(**config.scraping_params['batch'])
4     interval = end_date - start_date
5     print('\nStation {}, getting data from {}, to {}...'.format(self.index,
6         start_date, end_date))
7     if start_date > end_date:
8         print("start_date is later than end_date... skipping")
9         return
10    for i in range(math.floor(interval / batch) + 1):
11        start_date_param = (start_date + batch * i)
12        end_date_tmp      = start_date + batch * (i + 1) - timedelta(seconds
13            =1)
14        end_date_param    = end_date_tmp if end_date_tmp < end_date else
15            end_date
16        print("\nBatch {}, from: {}, to: {}".format(i+1, start_date_param,
17            end_date_param))
18        response = requests.get(self.url, params={
19            'start_date': start_date_param,
20            'end_date'   : end_date_param,
21            'username'   : self.username,
22            'password'   : self.password
23        })
24        response.raise_for_status()

```

```

21     response_dict = xmltodict.parse(response.text, attr_prefix="")
22     if response_dict['netsens']['errore'] == '1':
23         print("Response Error! Code: " + response_dict['netsens']['
codice'] + " Message: " + response_dict['netsens']['messaggio'])
24         continue
25     yield response_dict
26
27 def print_measurements_lengths(self, response_dict, verbose=False):
28     total_len = 0
29     for unit in response_dict['netsens']['stazione']['unita']:
30         if verbose: print("Station {}, Unit {} {}".format(self.index, unit['
id'], unit['nome']))
31         unit_measurements_len = 0
32         for sensor in unit['sensore']:
33             misura = sensor.get('misura', [])
34             misura = [misura] if isinstance(misura, dict) else misura
35             measurements_len = len(misura)
36             unit_measurements_len += measurements_len
37             if verbose: print("\tSensor {} {}, measurements: {}".format(
sensor['id'], sensor['nome'], measurements_len))
38             if verbose: print("\tUnit's total: ", unit_measurements_len)
39             total_len += unit_measurements_len
40         print("Station's total measurements: ", total_len)
41
42 def min_datetime(self):
43     return datetime.strptime(self.min_date + ' 00:00:00', config.
scraping_params['datetime_format'])

```

Σχήμα 5.7: Απόκτηση δεδομένων από τη διεπαφή των σταθμών.

5.4 Ανάκτηση δεδομένων και οπτικοποίηση

Στην παρούσα ενότητα παρουσιάζεται η διαδικασία ανάκτησης και οπτικής αποτύπωσης των δεδομένων χρονοσειρών που έχουν αποθηκευτεί στις βάσεις δεδομένων του συστήματος. Στόχος της ενότητας είναι η επιβεβαίωση της επιτυχούς εισαγωγής και ορθής αποθήκευσης των δεδομένων, καθώς και η παρουσίαση του τρόπου με τον οποίο αυτά μπορούν να αξιοποιηθούν για ανάλυση και ερμηνεία. Μέσα από ενδεικτικά ερωτήματα ανάκτησης και παραδείγματα οπτικοποίησης, αποδεικνύεται ότι τα δεδομένα είναι συνεπή, χρονικά ευθυγραμμισμένα και κατάλληλα για περαιτέρω επεξεργασία.

Αρχικά, παρουσιάζονται παραδείγματα ανάκτησης δεδομένων από τη βάση InfluxDB, με σκοπό τον έλεγχο της δομής και της χρονικής οργάνωσης των αποθηκευμένων μετρήσεων. Στη συνέχεια, ακολουθεί η οπτική απεικόνιση των δεδομένων μέσω του εργαλείου Grafana, η οποία επιτρέπει την άμεση κατανόηση της χρονικής εξέλιξης των μετρήσεων και λειτουργεί ως μέσο επαλήθευσης της ποιότητας των δεδομένων. Τέλος, παρουσιάζονται ενδεικτικά ερωτήματα ανάκτησης από τη βάση TimescaleDB, διαμορφωμένα κατά τρόπο ώστε τα αποτελέσματά τους να είναι άμεσα αξιοποιήσιμα για οπτικοποίηση και ανάλυση.

5.4.1 Ανάκτηση δεδομένων από την InfluxDB

Στην παρούσα υποενότητα παρουσιάζονται ενδεικτικά ερωτήματα ανάκτησης δεδομένων από τη βάση InfluxDB, με σκοπό την επαλήθευση της επιτυχούς αποθήκευσης των μετρήσεων και τον έλεγχο της χρονικής τους οργάνωσης. Τα ερωτήματα αυτά αξιοποιούνται για την εξαγωγή συγκεντρωτικών τιμών σε συγκεκριμένα χρονικά διαστήματα, ώστε τα αποτελέσματα να είναι άμεσα κατάλληλα για οπτική απεικόνιση και περαιτέρω ανάλυση.

```
1 SELECT mean(value)
2 FROM stations
3 WHERE sensor_id = 'TEMP'
4 AND time > now() - 24h
```

```
5 GROUP BY time(1h)
```

Σχήμα 5.8: Βασική ανάκτηση δεδομένων της InfluxDB.

```
1 SELECT mean(value)
2 FROM stations
3 WHERE sensor_id = 'TEMP'
4 AND time > now() - 24h
5 GROUP BY time(1h)
```

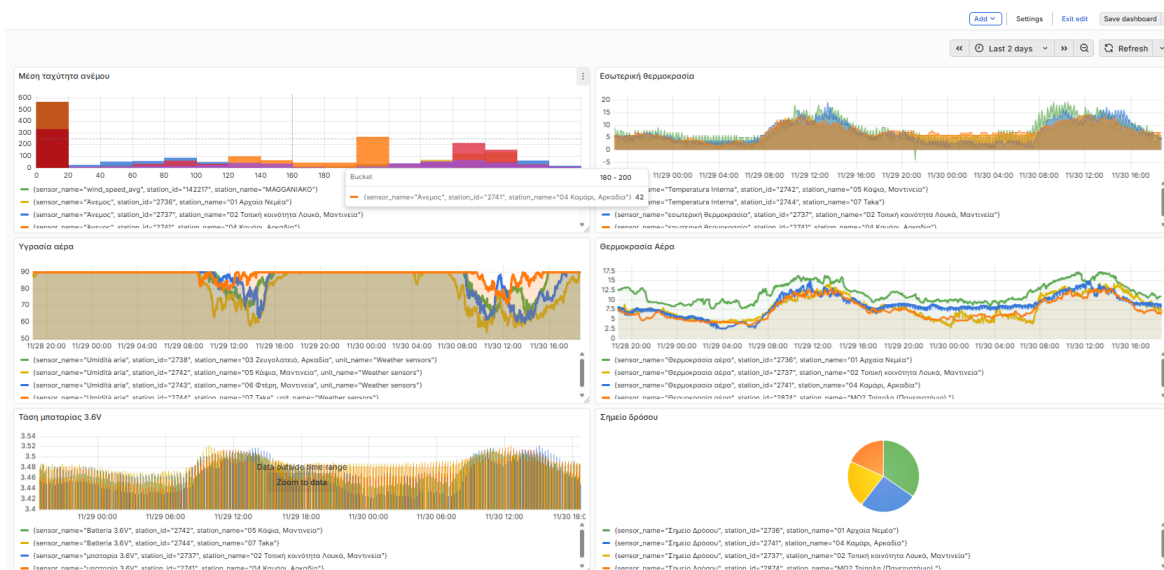
Σχήμα 5.9: Ανάκτηση δεδομένων ανά σταθμό, στην InfluxDB.

Τα αποτελέσματα των ερωτημάτων στα Σχήματα 5.8 και 5.9 επιβεβαιώνουν ότι οι μετρήσεις έχουν αποθηκευτεί ορθά και είναι διαθέσιμες σε διακριτά χρονικά διαστήματα. Η χρήση χρονικής ομαδοποίησης επιτρέπει την απευθείας αξιοποίηση των δεδομένων για οπτικοποίηση, διευκολύνοντας την παρακολούθηση της εξέλιξης των μετρήσεων στον χρόνο.

5.4.2 Οπτικοποίηση δεδομένων μέσω Grafana

Συνδέσαμε την InfluxDB με το εργαλείο οπτικοποίησης δεδομένων Grafana για τη δημιουργία διαδραστικών γραφημάτων (π.χ., μεταβολή θερμοκρασίας, ροδογράμματα ανέμου, ημερήσια μέση τιμή υγρασίας). Το περιβάλλον αυτό επιτρέπει την οπτική παρακολούθηση βασικών μετεωρολογικών μεταβλητών, όπως θερμοκρασία, υγρασία, ταχύτητα και διεύθυνση ανέμου, βροχόπτωση και δείκτες ποιότητας αέρα. Τα επιμέρους γραφήματα απεικονίζουν τη χρονική εξέλιξη κάθε μεταβλητής, επιτρέποντας τον εντοπισμό τάσεων, αιχμών ή ανωμαλιών στα δεδομένα. Με αυτόν τον τρόπο, η InfluxDB δε λειτουργεί μόνο ως αποθήκη χρονοσειρών, αλλά σε συνδυασμό με το Grafana παρέχει ένα ολοκληρωμένο εργαλείο παρακολούθησης και ανάλυσης των μετεωρολογικών μετρήσεων.

Ένα παράδειγμα οπτικοποίησης χρησιμοποιώντας το Grafana που απεικονίζει δεδομένα απευθείας από την InfluxDB παρουσιάζουμε στο Σχήμα 5.8. Στο συγκεκριμένο παράδειγμα οπτικοποίησης παρουσιάζονται σε ένα ενιαίο διαδραστικό πίνακα οι κυριότερες μετεωρολογικές μεταβλητές που συλλέγονται από τους σταθμούς του συστήματος.



Σχήμα 5.8: Παράδειγμα οπτικοποίησης δεδομένων InfluxDB στο Grafana.

Κάθε γράφημα αντλεί σε πραγματικό χρόνο δεδομένα από την InfluxDB και τα αποδίδει σε δυναμικές χρονοσειρές. Στο άνω μέρος του πίνακα εμφανίζονται οι μετρήσεις που σχετίζονται με τον άνεμο, όπως η μέση ταχύτητα ανέμου και η εσωτερική θερμοκρασία των σταθμών, επιτρέποντας την παρατήρηση ημερήσιων διακυμάνσεων και έντονων φαινομένων. Στο κεντρικό τμήμα του διαδραστικού πίνακα απεικονίζονται οι μεταβολές της σχετικής υγρασίας καθώς και η εξωτερική θερμοκρασία αέρα για πολλαπλούς σταθμούς, επιτρέποντας την απευθείας σύγκριση μεταξύ διαφορετικών γεωγραφικών σημείων.

Στο κάτω μέρος περιλαμβάνονται τιμές που συνδέονται με την ενεργειακή κατάσταση και τη λειτουργικότητα του εξοπλισμού, όπως η τάση της μπαταρίας 3.6V σε κάθε σταθμό, ενώ παράλληλα εμφανίζεται και η εξέλιξη του σημείου δρόσου, χρήσιμη μεταβλητή για την εκτίμηση υγρασίας και πιθανών φαινομένων συμπύκνωσης. Η διάταξη των γραφημάτων επιτρέπει την ταυτόχρονη παρακολούθηση πολλών παραμέτρων, την εύκολη ανίχνευση συσχετίσεων και την αναγνώριση εξάρσεων ή ανωμαλιών που απαιτούν περαιτέρω διερεύνηση. Έτσι, ο διαδραστικός πίνακας λειτουργεί ως ένα ολοκληρωμένο εργαλείο εποπτείας των μετεωρολογικών συνθηκών και της επιχειρησιακής κατάστασης των σταθμών μέτρησης.

5.4.3 Ανάκτηση δεδομένων από την TimescaleDB για οπτικοποίηση

Στην παρούσα υποενότητα παρουσιάζονται ενδεικτικά ερωτήματα ανάκτησης δεδομένων από τη βάση TimescaleDB, διαμορφωμένα με στόχο την άμεση αξιοποίησή τους σε διαδικασίες οπτικοποίησης. Η χρήση χρονικής ομαδοποίησης και συγκεντρωτικών συναρτήσεων επιτρέπει τη μετατροπή των πρωτογενών μετρήσεων σε δομημένα σύνολα δεδομένων, τα οποία αποτυπώνουν τη χρονική εξέλιξη των παρατηρούμενων μεγεθών και είναι κατάλληλα για γραφική απεικόνιση.

```

1 SELECT
2   time_bucket('1 hour', time) AS      _,
3   AVG(value) AS                      _
4 FROM measurements
5 WHERE sensor_type = 'TEMP'
6 AND time > NOW() - INTERVAL '24 hours'
7 GROUP BY
8   _;

```

Σχήμα 5.10: Εξαγωγή ωριαίου μέσου όρου, στην TimescaleDB.

```

1 SELECT
2   time_bucket('1 hour', time) AS      _,
3   station_id,
4   AVG(value) AS                      _
5 FROM measurements
6 WHERE sensor_type = 'TEMP'
7 AND time > NOW() - INTERVAL '24 hours'
8 GROUP BY
9   _, station_id
10 ORDER BY
11   _, station_id;

```

Σχήμα 5.11: Εξαγωγή δεδομένων ανά σταθμό, στην TimescaleDB.

Το ερώτημα στο Σχήμα 5.10 αξιοποιεί χρονική ομαδοποίηση με σκοπό τον υπολογισμό της μέσης τιμής των μετρήσεων σε ωριαία διαστήματα. Με τον τρόπο αυτό, οι

πρωτογενείς καταγραφές μετασχηματίζονται σε ένα συγκεντρωτικό σύνολο δεδομένων, το οποίο αποτυπώνει τη γενική τάση της υπό μελέτη μεταβλητής στον χρόνο και είναι άμεσα κατάλληλο για γραφική απεικόνιση.

Το ερώτημα στο Σχήμα 5.11 επεκτείνει τη διαδικασία αυτή, εισάγοντας επιπλέον ομαδοποίηση ανά σταθμό μέτρησης. Η προσέγγιση αυτή επιτρέπει τη σύγκριση πολλαπλών χρονοσειρών στο ίδιο χρονικό πλαίσιο, διευκολύνοντας την ταυτόχρονη απεικόνιση και ανάλυση των μετρήσεων που προέρχονται από διαφορετικές γεωγραφικές θέσεις.

5.5 Σύνοψη κεφαλαίου

Στο κεφάλαιο αυτό παρουσιάστηκαν οι διαδικασίες αποθήκευσης και αξιοποίησης δεδομένων χρονοσειρών στις πλατφόρμες InfluxDB και TimescaleDB. Η InfluxDB ξεχωρίζει για την απλότητα και ταχύτητα εισαγωγής, ενώ η TimescaleDB προσφέρει δομημένη προσέγγιση και πλήρη υποστήριξη της γλώσσας SQL.

Και οι δύο βάσεις αποδείχθηκαν κατάλληλες για την αποθήκευση δεδομένων των μετεωρολογικών σταθμών, με την επιλογή να εξαρτάται από το είδος των αναλύσεων και τη συχνότητα ενημέρωσης που απαιτείται στο εκάστοτε έργο.

Ιδιαίτερη έμφαση δόθηκε στην ανάκτηση και οπτικοποίηση των αποθηκευμένων δεδομένων, με στόχο την επιβεβαίωση της ορθής λειτουργίας της διαδικασίας εισαγωγής και της ποιότητας των δεδομένων. Μέσω ενδεικτικών ερωτημάτων ανάκτησης και παραδειγμάτων οπτικής απεικόνισης, αποδείχθηκε ότι τα δεδομένα είναι συνεπή, χρονικά οργανωμένα και άμεσα αξιοποιήσιμα για ανάλυση.

Κεφάλαιο 6

Σύνοψη, συμπεράσματα και μελλοντικές κατευθύνσεις

Το παρόν κεφάλαιο ολοκληρώνει τη διατριβή, συνοψίζοντας το περιεχόμενο και τα βασικά αποτελέσματα της εργασίας, εξάγοντας τα κύρια συμπεράσματα που προκύπτουν από την υλοποίηση και αξιολόγηση του προτεινόμενου συστήματος, και παρουσιάζοντας ενδεικτικές κατευθύνσεις για μελλοντική επέκταση και βελτίωσή του. Η σύνοψη αποσκοπεί στην ενοποίηση των επιμέρους ενοτήτων της διατριβής σε ένα συνεκτικό σύνολο, αναδεικνύοντας τόσο τη μεθοδολογική προσέγγιση όσο και τη συνεισφορά της εργασίας στον τομέα της διαχείρισης δεδομένων χρονοσειρών.

6.1 Σύνοψη της διατριβής

Σκοπός της παρούσας διατριβής ήταν η μελέτη, ο σχεδιασμός και η υλοποίηση μίας ολοκληρωμένης προσέγγισης για τη διαχείριση δεδομένων χρονοσειρών που προέρχονται από δίκτυο μετεωρολογικών σταθμών στην Περιφέρεια Πελοποννήσου. Η εργασία καλύπτει τόσο το θεωρητικό υπόβαθρο των δεδομένων χρονοσειρών όσο και την πρακτική υλοποίηση συστημάτων αποθήκευσης, ανάκτησης και αξιοποίησής τους.

Αρχικά, παρουσιάστηκε η έννοια των δεδομένων χρονοσειρών, τα βασικά χαρακτηριστικά τους και οι ιδιαιτερότητες που τα διαφοροποιούν από άλλες κατηγορίες δεδομένων, όπως η έντονη χρονική διάσταση, η υψηλή συχνότητα καταγραφής και ο μεγάλος όγκος

πληροφοριών. Στο πλαίσιο αυτό, αναλύθηκαν οι περιορισμοί των παραδοσιακών σχεσιακών βάσεων δεδομένων στη διαχείριση τέτοιων συνόλων δεδομένων και εξετάστηκαν τα πλεονεκτήματα των εξειδικευμένων συστημάτων χρονοσειρών, τα οποία έχουν σχεδιαστεί ειδικά για αποδοτική αποθήκευση, χρονική ομαδοποίηση και ταχεία ανάκτηση μετρήσεων.

Στη συνέχεια, πραγματοποιήθηκε συγκριτική ανάλυση μεταξύ διαφορετικών προσεγγίσεων αποθήκευσης, με έμφαση στη φιλοσοφία σχεδίασης, τα μοντέλα δεδομένων και τις δυνατότητες επεξεργασίας που προσφέρουν τα εξειδικευμένα συστήματα χρονοσειρών σε σύγκριση με τις κλασικές σχεσιακές βάσεις. Η ανάλυση αυτή οδήγησε στην επιλογή δύο αντιπροσωπευτικών πλατφορμών για περαιτέρω μελέτη και υλοποίηση:

- της InfluxDB, η οποία υιοθετεί μοντέλο αποθήκευσης χωρίς προκαθορισμένο σχήμα με χρήση μετρήσεων, ετικετών και πεδίων, και
- της TimescaleDB, ως επέκταση της PostgreSQL, η οποία βασίζεται σε συγκεκριμένο σχεσιακό σχήμα και αξιοποιεί υπερπίνακες για τη βελτιστοποιημένη διαχείριση χρονικών δεδομένων.

Στο πρακτικό σκέλος της διατριβής σχεδιάστηκε και υλοποιήθηκε μία ολοκληρωμένη αλυσίδα εξαγωγής, μετασχηματισμού και μεταφόρτωσης για την περιοδική ανάκτηση δεδομένων από τις διεπαφές των μετεωρολογικών σταθμών, την κανονικοποίησή τους και την αποθήκευσή τους στις δύο πλατφόρμες. Ιδιαίτερη έμφαση δόθηκε στη διατήρηση συνεπούς χρονικής ευθυγράμμισης, στην αποφυγή διπλοεγγραφών και στον σαφή διαχωρισμό μεταξύ της λογικής συλλογής δεδομένων και των μηχανισμών αποθήκευσης.

Τέλος, παρουσιάστηκαν ενδεικτικά ερωτήματα ανάκτησης και παραδείγματα οπτικοποίησης των αποθηκευμένων δεδομένων μέσω του εργαλείου Grafana, με στόχο την επιβεβαίωση της ορθής λειτουργίας της διαδικασίας εισαγωγής και την ανάδειξη της χρηστικότητας των δεδομένων για ανάλυση και ερμηνεία. Μέσα από τη συνδυαστική θεώρηση θεωρίας και υλοποίησης, η διατριβή καταδεικνύει πώς διαφορετικές προσεγγίσεις αποθήκευσης χρονοσειρών μπορούν να υποστηρίξουν αποτελεσματικά σύγχρονα συστήματα παρακολούθησης και ανάλυσης περιβαλλοντικών δεδομένων.

6.2 Κύρια αποτελέσματα και συνεισφορά

Η υλοποίηση της προτεινόμενης αρχιτεκτονικής κατέδειξε ότι και οι δύο πλατφόρμες είναι κατάλληλες για την αποθήκευση και διαχείριση δεδομένων χρονοσειρών από μετεωρολογικούς σταθμούς, καλύπτοντας διαφορετικές ανάγκες και σενάρια χρήσης.

Η InfluxDB προσφέρει απλή και αποδοτική εισαγωγή δεδομένων, φυσική μοντελοποίηση χρονοσειρών και άμεση διασύνδεση με εργαλεία οπτικοποίησης, καθιστώντας την ιδιαίτερα κατάλληλη για συστήματα παρακολούθησης και επιχειρησιακή εποπτεία σε πραγματικό ή σχεδόν πραγματικό χρόνο. Αντίθετα, η TimescaleDB επιτρέπει αυστηρότερο σχεδιασμό δεδομένων, αξιοποίηση της γλώσσας SQL, σαφή ορισμό σχέσεων και περιορισμών, και υποστήριξη σύνθετων αναλυτικών ερωτημάτων, γεγονός που την καθιστά κατάλληλη για περιβάλλοντα όπου απαιτείται δομημένη ανάλυση και ενοποίηση με άλλα πληροφοριακά συστήματα.

Η υιοθέτηση κοινής αφηρημένης διεπαφής για την αποθήκευση των δεδομένων επέτρεψε την εναλλακτική χρήση διαφορετικών συστημάτων βάσεων δεδομένων χωρίς τροποποίηση της βασικής λογικής συλλογής. Επιπλέον, η χρήση μηχανισμών ελέγχου της τελευταίας χρονικής σφραγίδας διασφάλισε την αποφυγή διπλοεγγραφών και τη συνεπή ενημέρωση των βάσεων δεδομένων.

Συνολικά, τα αποτελέσματα δείχνουν ότι η επιλογή πλατφόρμας αποθήκευσης δεν είναι απόλυτη, αλλά εξαρτάται από τις απαιτήσεις του εκάστοτε συστήματος, όπως η ανάγκη για ταχύτητα εισαγωγής, η πολυπλοκότητα των αναλύσεων και ο βαθμός δομημένης πληροφορίας που απαιτείται.

Τα κύρια αποτελέσματα της εργασίας συνοψίζονται ως εξής:

Υλοποιήθηκε πλήρης διαδικασία εξαγωγής, μετασχηματισμού και μεταφόρτωσης για περιοδική ανάκτηση και μεταφόρτωση μετεωρολογικών χρονοσειρών.

Σχεδιάστηκε κοινή διεπαφή αποθήκευσης και υλοποιήθηκαν δύο εναλλακτικά backends (InfluxDB και TimescaleDB), επιτρέποντας διττή χρήση χωρίς αλλαγές στη ροή συλλογής.

Ορίστηκε χαρτογράφηση των δεδομένων των σταθμών σε δομές της

InfluxDB (μετρήσεις/ετικέτες/πεδία) και σε σχεσιακό σχήμα της TimescaleDB (οντότητες, κλειδιά, υπερπίνακες).

Ενσωματώθηκε μηχανισμός αποφυγής διπλοεγγραφών με βάση τη χρονοσφραγίδα της τελευταίας αποθηκευμένης μέτρησης.

Παρουσιάστηκαν ενδεικτικά ερωτήματα ανάκτησης, συναθροίσεων και παραδείγματα οπτικοποίησης, επιβεβαιώνοντας τη συνέπεια και τη χρηστικότητα των δεδομένων.

6.3 Περιορισμοί της εργασίας

Παρότι η εργασία καλύπτει πλήρη κύκλο συλλογής, αποθήκευσης και αξιοποίησης δεδομένων χρονοσειρών, υπάρχουν ορισμένοι περιορισμοί που πρέπει να επισημανθούν. Η αξιολόγηση των συστημάτων επικεντρώθηκε κυρίως στη λειτουργικότητα και την αρχιτεκτονική προσέγγιση και όχι σε εκτεταμένες πειραματικές μετρήσεις απόδοσης υπό διαφορετικά φορτία. Επιπλέον, ζητήματα όπως η μακροχρόνια αξιολόγηση πολιτικών διατήρησης δεδομένων, η πλήρης διαχείριση ποιότητας μετρήσεων και η ανθεκτικότητα σε σφάλματα δεν εξετάστηκαν σε βάθος, καθώς δεν αποτελούσαν βασικό στόχο της παρούσας διατριβής.

6.4 Μελλοντικές κατευθύνσεις

Μελλοντική επέκταση της εργασίας θα μπορούσε να εστιάσει στη βελτίωση της ποιότητας των δεδομένων μέσω ανίχνευσης ανωμαλιών, διαχείρισης ελλειπών μετρήσεων και αξιολόγησης της αξιοπιστίας των αισθητήρων. Επιπλέον, θα μπορούσε να πραγματοποιηθεί συστηματική πειραματική σύγκριση των πλατφορμών αποθήκευσης ως προς την απόδοση εισαγωγής και ανάκτησης δεδομένων σε διαφορετικές κλίμακες.

Άλλες πιθανές κατευθύνσεις περιλαμβάνουν τη μετάβαση σε αρχιτεκτονικές συνεχούς ροής δεδομένων, την ενσωμάτωση προηγμένων μεθόδων ανάλυσης και πρόβλεψης χρονοσειρών, καθώς και την ενίσχυση θεμάτων ασφάλειας, διαχείρισης πρόσβασης και διακυβέρνησης δεδομένων σε περιβάλλοντα παραγωγής.

Συνολικά, η εργασία θέτει ένα λειτουργικό και επεκτάσιμο πλαίσιο διαχείρισης δεδομένων χρονοσειρών, το οποίο μπορεί να αποτελέσει βάση για μελλοντικές επεκτάσεις σε ερευνητικά ή επιχειρησιακά περιβάλλοντα.

Παράρτημα Α΄

Απαιτήσεις και προδιαγραφές σταθμών

Το παρόν παράρτημα περιλαμβάνει συνοπτικά τις απαιτήσεις που χρησιμοποιήθηκαν ως κριτήρια επιλογής των σταθμών.

Α΄.1 Αγρομετεωρολογικός σταθμός

Προδιαγραφή	Απαίτηση
Πλήρης αγρομετεωρολογικός σταθμός	NAI
Πλήθος μονάδων	8
Να αναφερθεί μοντέλο και εταιρεία κατασκευής	NAI
Ψηφιακή μονάδα καταγραφής και τηλεμετάδοσης των μετρήσεων	NAI
Οι συνδέσεις με τους αισθητήρες γίνονται μέσω εσωτερικών συνδέσμων	NAI
Χαμηλή κατανάλωση ισχύος, αυτονομία τουλάχιστον 5 ημερών	NAI
Δυνατότητα τοπικής αποθήκευσης	NAI
Εσωτερικός αισθητήρας εσωτερικής θερμοκρασίας και βαρομετρικής πίεσης	NAI
Καταγραφή events βροχής	NAI
Πλήρης τηλεμετρικός έλεγχος της μονάδος	NAI
Ενσωματωμένο modem κινητής τηλεφωνίας	NAI

Προδιαγραφή	Απαίτηση
Αποσπώμενη κεραία omni directional	NAI
Ενσωματωμένη μπαταρία	NAI
Ενσωματωμένο σύστημα ελέγχου φόρτισης μπαταρίας	NAI
Προγραμματισμός ρυθμού μέτρησης/αποστολής (5 λεπτά έως 2 ώρες)	NAI
Σύνδεση και τροφοδοσία από ηλιακό πάνελ	NAI
Ειδοποίηση για πιθανό πρόβλημα ηλιακού πάνελ	NAI
Ειδοποίηση για πιθανό πρόβλημα αισθητήρα	NAI
Ειδοποίηση για πιθανή δυσλειτουργία	NAI
Προστασία IP56	NAI
Εύρος λειτουργίας θερμοκρασίας αέρα -10°C έως 60°C	NAI
Δυνατότητα προσθήκης επιπλέον αισθητήρων	NAI
Τα υλικά κατασκευής δεν πρέπει να είναι μεταλλικά	NAI
Φωτοβολταϊκός συλλέκτης	NAI
Βροχόμετρο	NAI
– Ανάλυση	0.1 mm
– Επιφάνεια συλλογής	214 cm ²
– Δυνατότητα επαναβαθμονόμησης στο πεδίο	NAI
– Πλαστικό φίλτρο κατακράτησης φύλλων/υλικών (καλύπτει είσοδο)	NAI
– Μέτρηση σε ένταση >10 mm/min	NAI
– Υλικό ανθεκτικό σε UV/διάβρωση (UV-Stabilized ABS)	NAI
Αισθητήρας διύγρανσης φύλλων	NAI
– Ανάλυση	1 mm
– Επιφάνεια συλλογής	50 cm ²
– Έξοδος σε λεπτά (μέτρηση σε λεπτά της ώρας)	NAI
– Κατάλληλος για εισαγωγή εντός κόμης χωρίς ζημιά	NAI
– Όρια 230–380 kΩ (αν ως αντίσταση)	NAI
Αισθητήρας σχετικής υγρασίας και θερμοκρασίας αέρα	NAI
– Προστατευτικός κλωβός με θερμοπλαστικές πλάκες	NAI
– Κατανάλωση ισχύος	<2 mA
– Περιοχή υγρασίας	0–100%

Προδιαγραφή	Απαίτηση
– Ακρίβεια υγρασίας (0–90%)	3%
– Ακρίβεια θερμοκρασίας	<0.5 °C
– Περιοχή θερμοκρασίας	-25 °C έως +65 °C
– Καλωδίωση ανθεκτική σε UV	NAI
– Θερμοκρασία λειτουργίας	-50 °C έως +100 °C
– Υγρασία λειτουργίας	0–100%
Αισθητήρας ολικής ηλιακής ακτινοβολίας	NAI
– Γραμμική περιοχή	0–1800 W/m ²
– Φασματικό εύρος	400–1100 nm
– Θερμοκρασία λειτουργίας	-40 °C έως 65 °C
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας ταχύτητας και διεύθυνσης ανέμου	NAI
– Τύπος υπερήχων	NAI
– Εύρος μετρήσεων	0–40 m/s
– Ριπές έως	78 knots
– Θερμοκρασία λειτουργίας	-15 °C έως 55 °C
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας βαρομετρικής πίεσης	NAI
– Σφάλμα <0.05 hPa για -10 έως 50 °C	NAI
– Περιοχή μέτρησης	200–1100 hPa
– Θερμοκρασία λειτουργίας	-40 °C έως 85 °C

Α'.2 Σταθμός μέτρησης ατμοσφαιρικών ρύπων και μικροκλίματος

Προδιαγραφή	Απαίτηση
Συστήματα μέτρησης ατμοσφαιρικών ρύπων και μικροκλίματος	NAI
Πλήθος μονάδων	2

Προδιαγραφή	Απαίτηση
Ψηφιακή μονάδα καταγραφής και τηλεμετάδοσης	NAI
Οι συνδέσεις με τους αισθητήρες γίνονται μέσω εσωτερικών συνδέσμων	NAI
Χαμηλή κατανάλωση ισχύος, αυτονομία τουλάχιστον 5 ημερών	NAI
Δυνατότητα τοπικής αποθήκευσης	NAI
Αισθητήρας θερμοκρασίας	NAI
Πλήρης τηλεμετρικός έλεγχος	NAI
Ενσωματωμένο modem κινητής τηλεφωνίας	NAI
Ενσωματωμένη μπαταρία	NAI
Ενσωματωμένο σύστημα φόρτισης μπαταρίας	NAI
Προγραμματισμός ρυθμού μέτρησης/αποστολής	NAI
Τροφοδοσία από ηλιακό πάνελ	NAI
Θερμοκρασία λειτουργίας -30°C έως 70°C	NAI
Ειδοποίηση αποσύνδεσης από ηλιακό πάνελ	NAI
Ειδοποίηση αποσύνδεσης αισθητήρα	NAI
Ειδοποίηση δυσλειτουργίας	NAI
Προστασία IP56	NAI
Εύρος λειτουργίας θερμοκρασίας αέρα -10°C έως 60°C	NAI
Δυνατότητα προσθήκης επιπλέον αισθητήρων	NAI
Υλικά κατασκευής όχι μεταλλικά	NAI
Επικοινωνία με transmitter για PM1, PM2.5, PM10	NAI
– Μονάδες μέτρησης	ppm
– Μέγεθος particles τουλάχιστον	18 μm
– Χαμηλή κατανάλωση ισχύος	NAI
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας PM1	NAI
– Περιοχή μέτρησης 0.000 έως 10000 ppm	NAI
– Interface	SPI/USB
Αισθητήρας PM2.5	NAI
– Περιοχή μέτρησης 0.000 έως 10000 ppm	NAI
– Interface	SPI/USB

Προδιαγραφή	Απαίτηση
Αισθητήρας PM10	NAI
– Περιοχή μέτρησης 0.000 έως 10000 ppm	NAI
– Interface	SPI/USB
Επικοινωνία με transmitter για NO ₂	NAI
– Δυναμική δειγματοληψία	NAI
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας NO ₂	NAI
– Περιοχή μέτρησης	0–20 ppm
– Ευαισθησία	1 ppb
– Θερμοκρασία λειτουργίας	-30 °C έως 50 °C
– Χρόνος απόκρισης	30 s
Επικοινωνία με transmitter για SO ₂	NAI
– Δυναμική δειγματοληψία	NAI
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας SO ₂	NAI
– Περιοχή μέτρησης	0–50 ppm
– Ευαισθησία	1 ppb
– Θερμοκρασία λειτουργίας	-30 °C έως 50 °C
– Χρόνος απόκρισης	30 s
Επικοινωνία με transmitter για O ₃	NAI
– Δυναμική δειγματοληψία	NAI
– Καλωδίωση ανθεκτική σε UV	NAI
Αισθητήρας O ₃	NAI
– Περιοχή μέτρησης	0–0.5 ppm
– Ελάχιστο ανιχνεύσιμο όριο	<0.01 ppm
– Ανάλυση	4 ppb
– Χρόνος απόκρισης	<60 s
– Θερμοκρασία λειτουργίας	-30 °C έως 40 °C
Πλήρης μετεωρολογικός σταθμός για μέτρηση (άνεμος, θερμοκρασία/υγρασία, πίεση, βροχή, ακτινοβολία)	NAI

Προδιαγραφή	Απαίτηση
Ενιαίος σταθμός, ψηφιακή έξοδος (Serial/USB/LAN)	NAI
– Υπεριώδης ακτινοβολία 0–199 MED (ανάλυση 1 MED, ακρίβεια $\pm 5\%$)	NAI
– Βροχόπτωση 0–250 mm/h (ανάλυση 0.1 mm, ακρίβεια $\pm 5\%$)	NAI
– Σχετική υγρασία 0–100% (ανάλυση 0.1%RH, ακρίβεια $\pm 3\%$ RH)	NAI
– Θερμοκρασία αέρα -25 έως 65 °C (ανάλυση 0.1 °C)	NAI
– Βαρομετρική πίεση 200–1100 hPa (ανάλυση 0.1 hPa)	NAI
– Ταχύτητα ανέμου 0–40 m/s (ανάλυση 0.1 m/s, ακρίβεια 4% ή 2 kts)	NAI
– Διεύθυνση ανέμου 0–359° (ανάλυση 1°, ακρίβεια $\pm 4^\circ$)	NAI
– Κλίση ιστού/εγκατάστασης -90° έως +90° (ανάλυση 0.1°, ακρίβεια $\pm 1^\circ$)	NAI

Βιβλιογραφία

- [1] Allen, Richard G., Luis S. Pereira, Dirk Raes, and Martin Smith: *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements (FAO Irrigation and Drainage Paper 56)*. Food and Agriculture Organization of the United Nations (FAO), 1998. <http://www.fao.org/docrep/X0490E/x0490e00.htm>.
- [2] Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel: *Time Series Analysis: Forecasting and Control*. Wiley, 5th edition, 2015.
- [3] Chebotko, Artem, Andrey Kashlev, and Shiyong Lu: *A big data modeling methodology for apache cassandra*. In *Proceedings of the IEEE International Congress on Big Data*, 2015.
- [4] Dunning, Ted: *Time Series Databases: New Ways to Store and Access Data*. O'Reilly Media, 2015.
- [5] Hermans, Kris: *Mastering InfluxDB*. Packt Publishing, 2023.
- [6] Hyndman, Rob J. and George Athanasopoulos: *Forecasting: Principles and Practice*. OTexts, 3rd edition, 2021.
- [7] InfluxData: *Customers – influxdata*. <https://www.influxdata.com/customers/>. Accessed: 2025-12-01.
- [8] InfluxData: *Influxdb 3.0 system architecture*. <https://www.influxdata.com/blog/influxdb-3-0-system-architecture/>, Accessed: 2025-10-12.
- [9] InfluxData: *Influxdb enterprise v1 frequently asked questions: Relationship between shard group durations and retention policies*. <https://docs.influxdata.com/>

- [enterprise_influxdb/v1/troubleshooting/frequently-asked-questions/](#). Accessed: 2025-11-23.
- [10] InfluxData: *Influxdb storage engine evolution (leveldb, rocksdb, hyperleveldb → tsm tree)*. https://archive.docs.influxdata.com/influxdb/v1.2/concepts/storage_engine/. Accessed: 2025-11-26.
- [11] InfluxData: *Http api: Writing data (influxdb v1.1)*. Online documentation, 2016. <https://archive.docs.influxdata.com/influxdb/v1.1/tools/api/#write-http-endpoint>, visited on 2026-01-04.
- [12] InfluxData: *Influxql: Create retention policy (influxdb v1.2)*. Online documentation, 2017. https://archive.docs.influxdata.com/influxdb/v1.2/query_language/database_management/, visited on 2026-01-04.
- [13] InfluxData: *Line protocol reference (influxdb v1.2)*. Online documentation, 2017. https://archive.docs.influxdata.com/influxdb/v1.2/write_protocols/line_protocol_reference/, visited on 2026-01-04.
- [14] InfluxData: *Customer case study: Spiio uses influxdb for iot monitoring of green walls*, 2021. <https://www.influxdata.com/customer/customer-case-study-spiio/>, Accessed: 2025-10-12.
- [15] Jha, Bineet Kumar and Shilpa Pande: *Time series forecasting model for supermarket sales using fb-prophet*. In *Proceedings of the International Conference on Computing Methodologies and Communication (ICCMC)*, 2021.
- [16] Mills, Terence C.: *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*. Academic Press, 2019.
- [17] Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci: *Introduction to Time Series Analysis and Forecasting*. Wiley, 2015.
- [18] Petkovic, Dusan: *Microsoft SQL Server 2019: A Beginner's Guide*. McGraw-Hill, 2019.
- [19] Salo, Kristian: *Analysis and future development of the sensor data storage of an iot application*. Master's thesis, Aalto University, 2020.

-
- [20] Stefancova, Elena: *Evaluation of the timescaledb postgresql time series extension*. Technical report, CERN, 2018.
- [21] Tahaghoghi, Seyed M. M. and Hugh E. Williams: *Learning MySQL*. O'Reilly Media, 2019.
- [22] Taslim, Deddy Gunawan and Made Murwantara: *A comparative study of arima and lstm in forecasting time series data*. In *Proceedings of the 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2022.
- [23] TigerData (Timescale): *create_hypertable()*. Online documentation (API Reference), 2025. https://www.tigerdata.com/docs/api/latest/hypertable/create_hypertable/, visited on 2026-01-04.
- [24] World Health Organization: *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. World Health Organization, 2021. <https://apps.who.int/iris/bitstream/handle/10665/345329/9789240034228-eng.pdf>.
- [25] World Meteorological Organization: *Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)*. World Meteorological Organization, 2008. <https://www.weather.gov/media/epz/mesonet/CWOP-WM08.pdf>, Seventh edition.
- [26] Zamfir, Vlad Andrei, Mihai Carabas, and Costin Carabas: *Systems monitoring and big data analysis using the elasticsearch system*. In *Proceedings of the International Conference on Control Systems and Computer Science (CSCS)*, 2019.