

Ενσωμάτωση μηχανισμών caching,
observability και self-healing στην
υποδομή διαχείρισης δεδομένων IoT της
πλατφόρμας Nostradamus

Δημήτριος Ντέντας

5 Απριλίου 2025

Περιεχόμενα

1	Επισκόπηση της Ερευνητικής Περιοχής	1
2	Θεωρία	4
	Βιβλιογραφία	6

1

Επισκόπηση της Ερευνητικής Περιοχής

Στη σύγχρονη γεωργία, ένας δυσλειτουργικός αισθητήρας που περνά απαρατήρητος για ώρες μπορεί να οδηγήσει σε αποτυχημένο πότισμα ή καταστροφή καλλιεργειών. Αυτό το ενιαίο σημείο αποτυχίας αναδεικνύει την ανάγκη για αυτόνομα, ανθεκτικά και παρατηρήσιμα συστήματα σε κλίμακα. Η ανάγκη αυτή εναρμονίζεται με τον οδικό χάρτη (Roadmap) προς ανθεκτικά Κυβερνοφυσικά Συστήματα (CPS) που περιγράφεται από τους Ratasich et al. [1], οι οποίοι δίνουν έμφαση στην ανίχνευση ανωμαλιών κατά τη διάρκεια λειτουργίας, στην απομόνωση σφαλμάτων και στην αυτοϊαση σε δυναμικά Internet of Things (IoT) περιβάλλοντα.

Σε κρίσιμες εφαρμογές, όπως η αγροδιατροφή, η διατάραξη στη ροή δεδομένων ή στη λήψη των αποφάσεων μπορεί να οδηγήσει σε σπατάλη τροφίμων, οικονομική απώλεια ή επισιτιστική ανασφάλεια. Σύμφωνα με τους Callo και Mansouri [2], η ανθεκτικότητα των παγκόσμιων δικτύων διανομής τροφίμων εξαρτάται από την δυνατότητα των πληροφοριακών συστημάτων να ανταπεξέρχονται σε γεωπολιτικές ή υγειονομικές κρίσεις μέσω μηχανισμών ευελιξίας και προσαρμογής. Αυτό σημαίνει πως σε ότι αφορά σε αντίστοιχα συστήματα, στα οποία βασίζεται ο πλυθησμός (και η οικονομία) μιας χώρας, η αρχιτεκτονική που χρησιμοποιείται θα πρέπει να είναι άρτια, μελετημένη, αλλά και να εξασφαλίζει την ομαλή και συνεχή λειτουργία τους.

Οι σύγχρονες cloud-native τεχνολογίες παρέχουν τα εργαλεία και τα πρότυπα για την υλοποίηση τέτοιων χαρακτηριστικών, και την πρόληψη των προαναφερθέντων επιπτώσεων. Μέσω συνεχόμενης παρατήρησης ορισμένων μετρικών (*Observability*), αλλά και ειδοποιήσεων σε κρίσιμες περιπτώσεις (*Alerting*), μπορεί είτε να επανορθωθεί η ομαλή λειτουργία του συστήματος αυτόματα, είτε να ενημερωθεί ο αρμόδιος για την επανόρθωση χειριστής. Η ενσωμάτωση εργαλείων όπως το Prometheus σε cloud-native μικροϋπηρεσίες έχει ήδη αποδειχθεί αποτελεσματική σε εφαρμογές πραγματικού χρόνου, παρέχοντας κρίσιμη τηλεμετρία και αυτόματες ενέργειες ανάδρασης [3].

Βάσει της μελέτης των Sharma et al. [4], αναδεικνύεται η ολοένα αυξανόμενη σημασία της χρήσης IoT τεχνολογιών στον τομέα της γεωργίας ακριβείας, όπου η αξιοποίηση αισθητήρων για την παρακολούθηση παραμέτρων όπως η υγρασία, η θερμοκρασία, η αγωγιμότητα και τα επίπεδα θρεπτικών συστατικών του εδάφους (NPK), επιτρέπει την ακριβή λήψη αποφάσεων σε πραγματικό χρόνο. Η προσέγγιση αυτή ευθυγραμμίζεται με την ανάγκη για ευέλικτες, αποκεντρωμένες υποδομές λήψης αποφάσεων, οι οποίες υποστηρίζονται από μηχανισμούς edge analytics και cloud ενορχήστρωσης. Στο πλαίσιο αυτό, η μελέτη των Akhtar et al. [5] τονίζει τη σημασία της ενσωμάτωσης του edge computing για την επεξεργασία δεδομένων από αισθητήρες σε πραγματικό χρόνο, επιτρέποντας την αξιολόγηση του εδάφους και την παρακολούθηση ρύπων με μεγαλύτερη αποτελεσματικότητα. Υποστηρίζει, επιπλέον, πως η ενσωμάτωση τεχνικών μηχανικής μάθησης στην αλυσίδα επεξεργασίας των δεδομένων, όπως η αυτόματη πρόβλεψη της καταλληλότητας του εδάφους για συγκεκριμένες καλλιέργειες, ενισχύει την αξία των IoT συστημάτων και απαιτεί αρχιτεκτονική σχεδίαση που υποστηρίζει δυναμική ανάλυση και διαλειτουργικότητα. Τα ευρήματα αυτά ενισχύουν την άποψη ότι η αρχιτεκτονική ενός ανθεκτικού IoT συστήματος στον αγροδιατροφικό τομέα πρέπει να υποστηρίζει συνεχές monitoring, on-device προεπεξεργασία και ασφαλή, επεκτάσιμη μετάδοση δεδομένων, προκειμένου να επιτευχθεί πλήρης αυτοματοποίηση και ευστάθεια λειτουργίας σε ετερογενή, διασυνδεδεμένα περιβάλλοντα πεδίου.

Με στόχο, λοιπόν, την αρχιτεκτονική ενίσχυση του συστήματος *Nostradamus* προτείνεται, ως απόρροια των παραπάνω μελετών, η ενσωμάτωση μηχανισμών caching, observability και self-healing. Παρότι υπάρχουν επιμέρους τεχνολογίες που προσφέρουν τέτοιες δυνατότητες, η υιοθέτησή τους σε πραγματικές πολυεπίπεδες IoT πλατφόρμες συνοδεύεται από σημαντικές προκλήσεις. Στα κατανεμημένα συστήματα, η προσθήκη μιας νέας υπηρεσίας ή λειτουργικότητας συνοδεύεται συχνά από μη προβλέψιμες επιπτώσεις στην απόδοση του συστήματος, όπως

αναφέρει επανειλημμένα ο Kleppmann [6]. Αν και οι μηχανισμοί όπως το caching και το observability αποσκοπούν στην βελτίωση της εμπειρίας και της διαχειρισιμότητας, μπορούν, ως παράπλευρη συνέπεια, να οδηγήσουν σε αύξηση του latency ή της κατανάλωσης πόρων. Το φαινόμενο αυτό αποδίδεται στην πολυπλοκότητα των αλληλεπιδράσεων μεταξύ μικροϋπηρεσιών και την έλλειψη πλήρους visibility σε runtime επίπεδο.

Κατ' επέκταση, η ερευνητική συνεισφορά εστιάζει όχι μόνο στην επιλογή σχετικών τεχνολογιών, αλλά κυρίως στη συνεκτική και δυναμική ενορχήστρωσή τους, με στόχο την επίτευξη αυτονομίας και διαχειρισιμότητας σε cloud-native κατανεμημένα περιβάλλοντα υψηλής πολυπλοκότητας, για την υποστήριξη αυτού του συστήματος πραγματικού χρόνου στο χώρο του *food security*.

2

Θεωρία

Η ανάγκη για αποδοτική ροή δεδομένων μεταξύ μικροϋπηρεσιών οδηγεί συνήθως στην υιοθέτηση τεχνολογιών όπως το *Apache Kafka*, ένα - πλέον - *de facto* πρότυπο για την υλοποίηση ανθεκτικών και επεκτάσιμων messaging υποδομών, χάρη στη δυνατότητά του να αποθηκεύει τα μηνύματα σε μορφή καταγραφής (log-based) [7]. Η αρχιτεκτονική του Kafka ενδείκνυται ιδιαίτερα για περιβάλλοντα που απαιτούν real ή near-real time επεξεργασία γεγονότων, καθώς παρέχει υψηλή διαθεσιμότητα, εγγυημένη διανομή μηνυμάτων και δυνατότητα οριζόντιας κλιμάκωσης. Συγκριτικές μελέτες [8] επιβεβαιώνουν ότι το Kafka παρουσιάζει σημαντικό πλεονέκτημα σε όρους throughput και fault tolerance σε σχέση με άλλες προσεγγίσεις, γεγονός που το καθιστά κατάλληλο για απαιτητικές IoT εφαρμογές με αυξημένο όγκο και ταχύτητα δεδομένων. Στο πλαίσιο αυτό, το Kafka συνιστά θεμελιώδες δομικό στοιχείο για αρχιτεκτονικές που στοχεύουν στην ημι-πραγματική επεξεργασία δεδομένων, όπως το σύστημα *Nostradamus*.

Η αποθήκευση μεγάλου όγκου δεδομένων σε κατανεμημένα συστήματα βασίζεται - κυρίως - σε βάσεις δεδομένων τύπου *wide-column*, με το *Apache Cassandra* να αποτελεί ένα από τα πιο διαδεδομένα και ώριμα συστήματα σε αυτόν τον χώρο. Το Cassandra ακολουθεί το μοντέλο *eventual consistency*, υποστηρίζει κατανεμημένη αποθήκευση με replication και partitioning, και έχει σχεδιαστεί για *write-heavy* εφαρμογές [9]. Σε περιβάλλοντα IoT, όπου οι συσκευές παράγουν συνεχώς δεδομένα τηλεμετρίας (π.χ. θερμοκρασία, τάση, ρεύμα) με υψηλή συ-

χνότητα, η Cassandra μπορεί να λειτουργήσει ως backend αποθήκευσης για ροές δεδομένων σχεδόν σε πραγματικό χρόνο, διασφαλίζοντας υψηλή διαθεσιμότητα και συνεχή εγγραφή με χαμηλό latency. Η προσέγγιση αυτή έχει εφαρμοστεί επιτυχώς σε σενάρια όπως η παρακολούθηση φωτοβολταϊκών μονάδων μέσω Raspberry Pi [10], όπου το σύστημα συλλέγει και αποθηκεύει μετρήσεις αισθητήρων κάθε 15 λεπτά για περαιτέρω ανάλυση και βελτιστοποίηση απόδοσης. Αντίστοιχες αρχιτεκτονικές υποδεικνύουν τη σημασία της επιλογής αποθηκευτικού συστήματος που να ανταποκρίνεται τόσο σε επιχειρησιακές ανάγκες χαμηλής καθυστέρησης όσο και σε απαιτήσεις αξιοπιστίας και επεκτασιμότητας.

Σε ότι αφορά την ανάγνωση των αποθηκευμένων δεδομένων αυτού του όγκου, αλλά και δεδομένης της συνεχόμενης άφιξης τους, τα προσωρινά δεδομένα σε μνήμη (*in-memory caching*) αποτελούν κρίσιμο μηχανισμό βελτιστοποίησης της απόδοσης. Δύο από τις πιο διαδεδομένες τεχνολογίες στον τομέα αυτό είναι τα συστήματα *Memcached* [11] και *Redis* [12], τα οποία υλοποιούν μηχανισμούς προσωρινής αποθήκευσης ζευγών κλειδιού-τιμής, με στόχο τη μείωση του χρόνου και του κόστους προσπέλασης σε βάσεις δεδομένων. Η βασική διαφορά τους έγκειται κυρίως στην αρχιτεκτονική του ολικού συστήματος, καθώς και στον τρόπο διαχείρισης της ταυτόχρονης εκτέλεσης - με χρήση *multithreading* (*Memcached*) ή *event loop* (*Redis*).

Η ανάγκη για παρατηρησιμότητα, αποδοτική διαχείριση προσωρινών δεδομένων και σταθερή ροή μηνυμάτων σε περιβάλλοντα με κατανεμημένη υπολογιστική λογική, καθιστά την πλατφόρμα *Kubernetes* - συχνά - απαραίτητο δομικό στοιχείο. Η χρήση της επιτρέπει την υλοποίηση μηχανισμών αυτόματης αποκατάστασης, *autoscaling*, και επιτήρησης σε επίπεδο υπηρεσίας. Σε συνδυασμό με συστήματα όπως το *Prometheus* (και τα συνεργατικά του υποσυστήματα *Grafana* και *AlertManager*) [13] - κοινώς τη *lingua franca* του *observability* - συμβάλλει στον σχεδιασμό και στην υλοποίηση μεγάλων υποδομών, πληρώνοντας τα κριτήρια μιας παραγωγικής (*production-grade*) αρχιτεκτονικής.

Βιβλιογραφία

- [1] Denise Ratasich, Faiq Khalid, Florian Geissler, Radu Grosu, Muhammad Shafique, and Ezio Bartocci. A roadmap toward the resilient internet of things for cyber-physical systems. *IEEE Access*, 2019.
- [2] Anthony Callo and Mo Mansouri. Food security in global food distribution networks: A systems thinking approach. In *2024 IEEE International Systems Conference (SysCon)*, 2024.
- [3] Ashima Chawla, Pradeep Babu, Trushnesh Gawande, Erik Aumayr, Paul Jacob, and Sheila Fallon. Intelligent monitoring of iot devices using neural networks. In *2021 24th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, 2021.
- [4] Dr. Rashmi Sharma, Vishal Mishra, and Suryansh Srivastava. Enhancing crop yields through iot-enabled precision agriculture. In *2023 International Conference on Disruptive Technologies (ICDT)*, 2023.
- [5] Mohammad Nishat Akhtar, Abdurrahman Javid Shaikh, Ambareen Khan, Habib Awais, Elmi Abu Bakar, and Abdul Rahim Othman. Smart sensing with edge computing in precision agriculture for soil assessment and heavy metal monitoring: A review. *Agriculture*, "2021".
- [6] Martin Kleppmann. *Designing Data-Intensive Applications*. O'Reilly Media, 2017.
- [7] Rishika Shree, Tanupriya Choudhury, Subhash Chand Gupta, and Praveen Kumar. Kafka: The modern platform for data management and analysis in big data domain. In *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, 2017.
- [8] Chung Lai Deryck Ho, Chung-Horng Lung, and Zhengding Mao. Comparative analysis of real-time data processing architectures:

- Kafka versus mqtt broker in iot. In *2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, 2024.
- [9] Avinash Lakshman and Prashant Malik. Cassandra: a decentralized structured storage system. *SIGOPS Oper. Syst. Rev.*, 2010.
- [10] Arbër Sh. Perçuku, Daniela V. Minkovska, Lyudmila Y. Stoyanova, and Arta E. Abdullahu. Iot using raspberry pi and apache cassandra on pv solar system. In *2020 XXIX International Scientific Conference Electronics (ET)*, 2020.
- [11] Rajesh Nishtala, Hans Fugal, Steven Grimm, Marc Kwiattkowski, Herman Lee, Harry C. Li, Ryan McElroy, Mike Paleczny, Daniel Peek, Paul Saab, David Stafford, Tony Tung, and Venkateshwaran Venkataramani. Scaling memcache at facebook. In *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*. USENIX Association, 2013.
- [12] Josiah Carlson. *Redis in Action*. Manning Publications, 2013.
- [13] Sneha M. Pragathi B.C., Hrithik Maddirala. Implementing an effective infrastructure monitoring solution with prometheus and grafana. *International Journal of Computer Applications*, 2024.