

# Complex event processing on streaming sensor data using Esper and Storm

Διπλωματική εργασία στα πλαίσια του ΠΜΣ «Προηγμένα Πληροφοριακά Συστήματα

Φοιτητές: Νομικού Σοφία, Γιαννούτσος Βασίλειος  
Επιβλέπων Καθηγητής: Χατζηευθυμιάδης Ευστάθιος



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
Εθνικό και Καποδιστριακό  
Πανεπιστήμιο Αθηνών

# Δομή παρουσίασης

- Το σύστημα που χρησιμοποιήθηκε
- Η μέθοδος επίλυσης του προβλήματος
- Η λειτουργία των επαναλαμβανόμενων συμβάντων
- Μελλοντικές βελτιώσεις
- Συμπεράσματα



# Το σύστημα που χρησιμοποιήθηκε

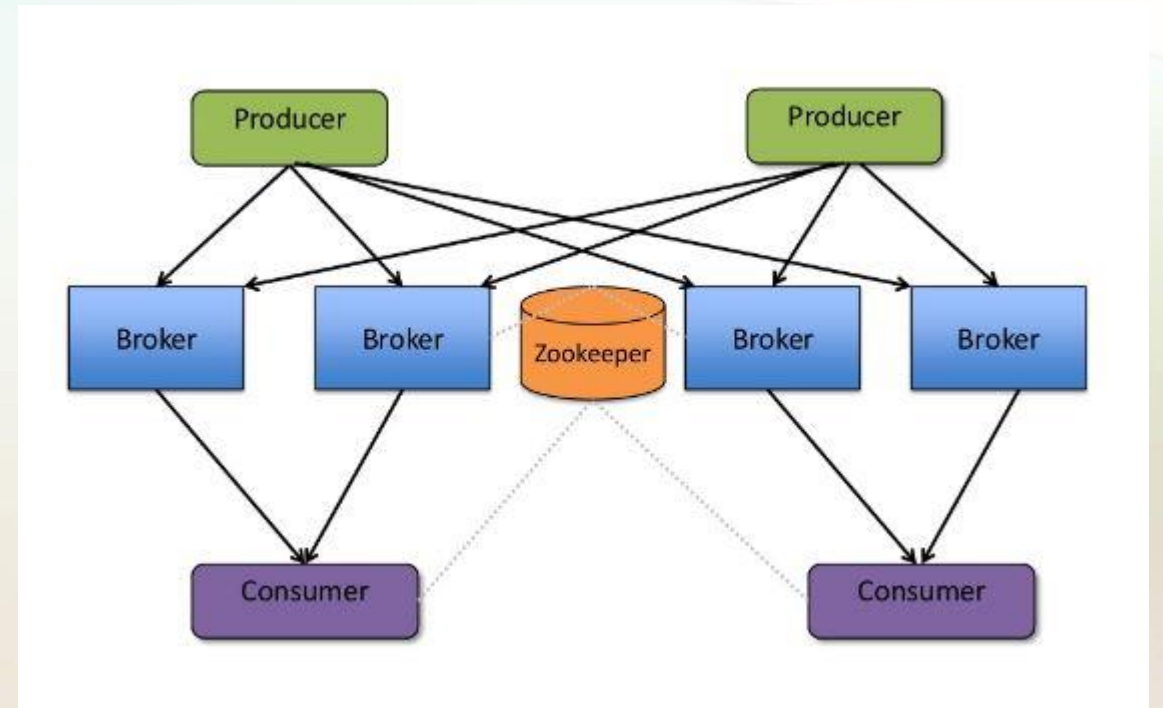
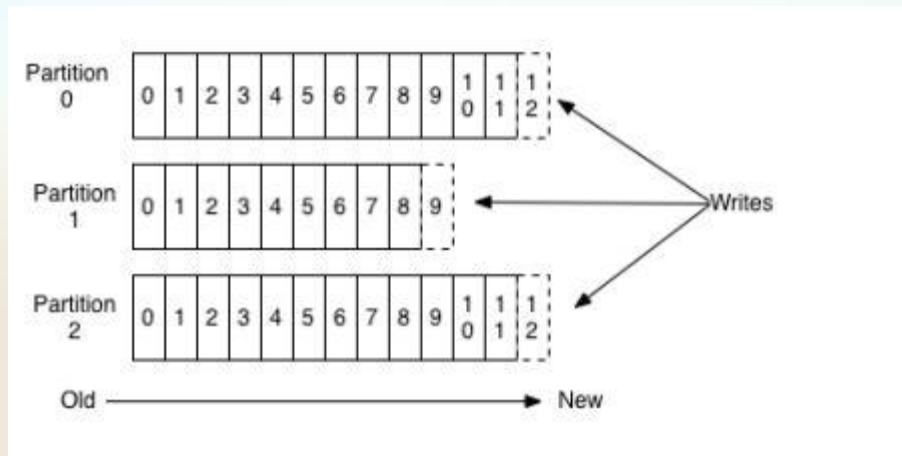
## Apache Kafka

- Μεγάλη ταχύτητα για μεγάλο όγκο δεδομένων
- Υψηλή δυνατότητα κλιμάκωσης
- Ανοχή σε σφάλματα
- Ο producer δημοσιεύει δεδομένα
- Τα δεδομένα οργανώνονται σε topics
- Τα topics διαμοιράζονται στα partitions
- Οι brokers διαχειρίζονται τα partitions



# Το σύστημα που χρησιμοποιήθηκε

## Apache Kafka



# Το σύστημα που χρησιμοποιήθηκε

## Apache Storm

- Ομοιότητες με το σύστημα Hadoop για αποτυχίες εκτέλεσης.
- Τοπολογίες αντί για διεργασίες
- Το “Nimbus” τρέχει από το master node και διαμοιράζει τον κώδικα στο cluster, αναθέτει εργασίες στα ξεχωριστά μηχανήματα και επιβλέπει
- Το “Supervisor” daemon τρέχει από το worker node και ξεκινά και διακόπτει διεργασίες βασιζόμενο στις εντολές του “Nimbus”



# Το σύστημα που χρησιμοποιήθηκε

## Apache Storm

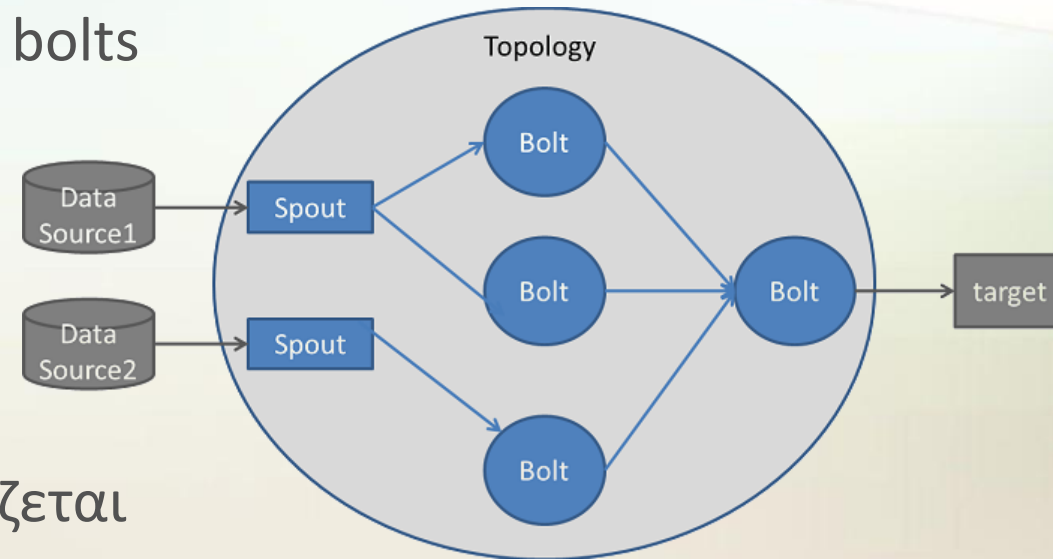
- Το Zookeeper συντονίζει το Nimbus με τα Supervisors
- Βασικό στοιχείο το stream
- Το stream αποτελείται από tuples
- Κατανεμημένος και αξιόπιστος τρόπος εφαρμογής μετασχηματισμών σε ρεύμα δεδομένων



# Το σύστημα που χρησιμοποιήθηκε

## Apache Storm

- Η τοπολογία αποτελείται από spouts και bolts
- Η τοπολογία εκτελείται συνεχώς
- Το spout αποτελεί μία πηγή από streams
- Το bolt δέχεται δεδομένα από τα spouts
- Η επεξεργασία και η λογική που εφαρμόζεται στα δεδομένα πραγματοποιούνται στα bolts



# Το σύστημα που χρησιμοποιήθηκε

## Esper

- Δυνατότητα εκτέλεσης επερωτημάτων σε ροή δεδομένων
- Data windows για τη διαχείριση και λήξη συμβάντων με αποδοτικό τρόπο
- Named windows για το διαμοιρασμό των συμβάντων μεταξύ των επερωτημάτων
- Ανάλυση σειρών από γεγονότα με match-recognize και patterns





# Το σύστημα που χρησιμοποιήθηκε

## ElasticSearch

- Σύμπλεγμα που αποτελείται από διάφορους κόμβους οι οποίοι συνεργάζονται για να παραχθεί ένα αποτέλεσμα
- Ομαδοποίηση εγγράφων σε ευρετήρια
- Ένα έγγραφο αποτελείται από ζεύγη κλειδιών-τιμών
- Το ευρετήριο διαχωρίζεται σε ανεξάρτητα δευτερεύοντα ευρετήρια για επίτευξη καλύτερης απόδοσης
- Υπάρχει αντίγραφο των εγγράφων για ανάκτηση δεδομένων σε περίπτωση σφάλματος
- Ανεστραμμένο ευρετήριο για γρήγορη αναζήτηση σε μεγάλο όγκο δεδομένων



# Το σύστημα που χρησιμοποιήθηκε

## Kibana

- Πλατφόρμα οπτικοποίησης και εξερεύνησης δεδομένων
- Βασισμένο στο Elasticsearch
- Εξειδικευμένο για την απεικόνιση μεγάλων όγκων δεδομένων ροής σε πραγματικό χρόνο



# Δομή παρουσίασης

- Το σύστημα που χρησιμοποιήθηκε
- **Η μέθοδος επίλυσης του προβλήματος**
- Η λειτουργία των επαναλαμβανόμενων συμβάντων
- Μελλοντικές βελτιώσεις
- Συμπεράσματα



# Η μέθοδος επίλυσης του προβλήματος

## Δεδομένα

- Δεδομένα από αισθητήρες τοποθετημένοι σε δωμάτια
- Τα δεδομένα αποτελούνται από πλειάδες όπου σε κάθε μία αναφέρεται το id του χρήστη, το id του αισθητήρα και το timestamp

### *Παράδειγμα των δεδομένων*

<0x34-0x0-0x35-0xe0-0x17-0x2-0x7-0x0-0x0-0x0-0x0-0x72-0x0c-0x4-0x97-0x0c, 5, 10:42:05.038>

<0x34-0x0-0x35-0xe0-0x17-0x2-0x7-0x0-0x0-0x0-0x0-0x72-0x0f-0xcf-0x98-0x86, 5, 10:42:05.089>

<0x34-0x0-0x35-0xe0-0x17-0x2-0x7-0x0-0x0-0x0-0x0-0x72-0x54-0x22-0x90-0x6c, 2, 10:42:05.861>



# Η μέθοδος επίλυσης του προβλήματος

## Η Storm τοπολογία

- Ορίζεται στο Flux
- Αποτελείται από τα παρακάτω bolts:
  - TxtTupleValidatorBolt
  - LoggingBolt
  - CoordinatorBolt
  - PrepareEsBolt
  - EsIndexBolt



# Η μέθοδος επίλυσης του προβλήματος

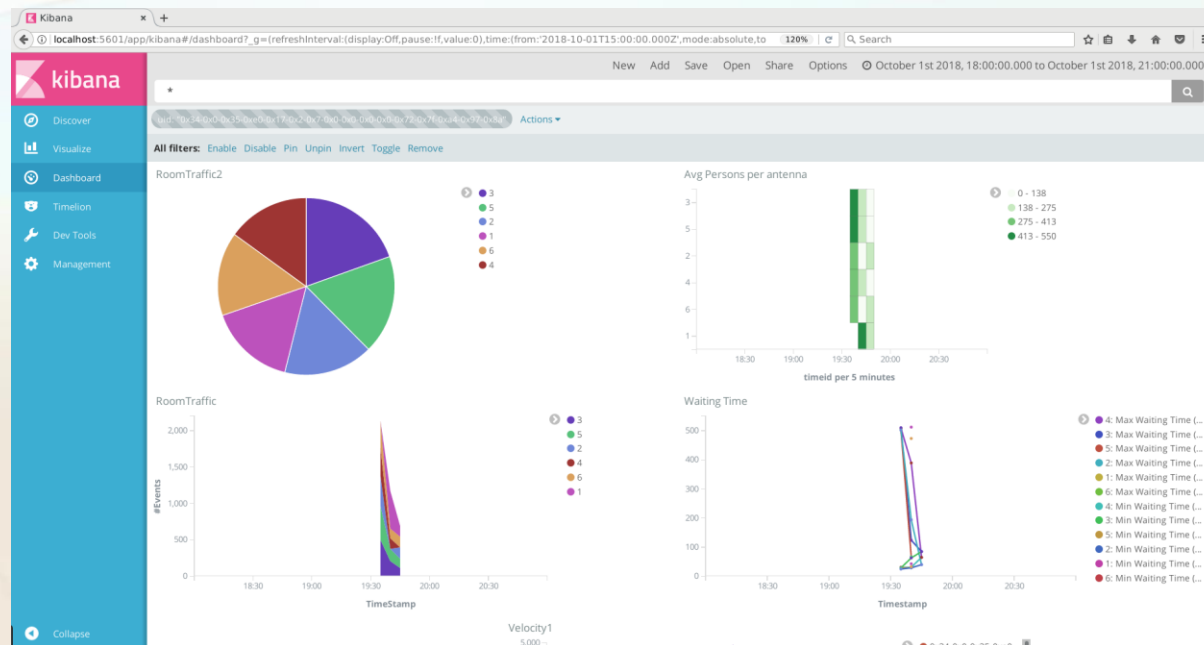
## Οι μετρικές που εφαρμόστηκαν

- Κίνηση του κάθε δωματίου
- Μέσος χρόνος αναμονής σε κάθε δωμάτιο
- Μέσος αριθμός επισκέψεων σε κάθε δωμάτιο
- Μέση ταχύτητα των χρηστών



# Η μέθοδος επίλυσης του προβλήματος

## Απεικόνιση των αποτελεσμάτων



# Δομή παρουσίασης

- Το σύστημα που χρησιμοποιήθηκε
- Η μέθοδος επίλυσης του προβλήματος
- **Η λειτουργία των επαναλαμβανόμενων συμβάντων**
- Μελλοντικές βελτιώσεις
- Συμπεράσματα





# Η λειτουργία των επαναλαμβανόμενων συμβάντων

## Ορισμός της επαναληψιμότητας

- Εντοπισμός της περιοδικότητας στα συμβάντα
- Πρόληψη μελλοντικών συμβάντων
- Υψηλή πολυπλοκότητα



# Η λειτουργία των επαναλαμβανόμενων συμβάντων

## Η ενσωμάτωση στο σύστημα

- Εντοπισμός μεταβάσεων χρήστη στο χρόνο
  - Ταυτισμός των χρονικών στιγμών για την ίδια μετάβαση για συγκεκριμένο χρήστη με τη χρήση δύο παράλληλων streams
  - Ομαδοποίηση των παρόμοιων μεταβάσεων για συγκεκριμένο χρονικό διάστημα και εύρεση αυτών με μεγάλη σχετικά συχνότητα



# Δομή παρουσίασης

- Το σύστημα που χρησιμοποιήθηκε
- Η μέθοδος επίλυσης του προβλήματος
- Η λειτουργία των επαναλαμβανόμενων συμβάντων
- **Μελλοντικές βελτιώσεις**
- Συμπεράσματα



# Μελλοντικές βελτιώσεις

- Σύστημα ροής δεδομένων
- Ενσωμάτωση στατικής πληροφορίας στο σύστημα
- Ενσωμάτωση κρυφής μνήμης
- Λειτουργίες μηχανικής μάθησης



# Δομή παρουσίασης

- Το σύστημα που χρησιμοποιήθηκε
- Η μέθοδος επίλυσης του προβλήματος
- Η λειτουργία των επαναλαμβανόμενων συμβάντων
- Μελλοντικές βελτιώσεις
- **Συμπεράσματα**



# Συμπεράσματα

- Θετικά αποτελέσματα για πολλά δεδομένα
  - Μικρός χρόνος επεξεργασίας
  - Γρήγορη ανάλυση
  - Άμεση απεικόνιση αποτελεσμάτων επεξεργασίας
- Ολοκληρωτικά αυτοματοποιημένη διαδικασία από πλευράς μοντελοποίησης
- Σύστημα ικανό να υποστηρίξει περισσότερες λειτουργίες από αυτές του προβλήματος
  - Επεξεργασία πολλών περισσότερων δεδομένων από αυτά του προβλήματος
  - Πολυάριθμες πηγές δεδομένων
- Οι ανάγκες για επίλυση προβλημάτων ανάλυσης δεδομένων υψηλής κλίμακας καθιστούν το εν λόγω σύστημα ως σημείο αναφοράς

