

5/1/23

נתוני עתק ומחשוב ענן

פרויקט סיום

כללי:

מערכות עיבוד נתוני עתק נמצאות בחזית האתגרים הטכנולוגיים בשנים האחרונות. בעולמות העסקיים היכולת להפוך את מאסות הנתונים הנאגרים תפעולית מנטל לנכס המאפשר ניהול סיכונים וקבלת החלטות מבוססת תובנות הנלמדות מנתונים אלה הינה משימה מרכזית בניהול ארגוני מודרני.

גם בתחומים מדעיים כגון מחקר רפואי או אסטרונומי, תחומים המאופיינים בצבירת נתונים ומידע בקצבים שהם בסדרי גודל מיכולת עיבודם, נדרשות שיטות טכנולוגיות ותשתיות המתמודדות עם האתגרים באופן המאפשר תמיכה בצורכי משתמשים וחוקרים ביעילות ובמהירות.

בפרויקט זה נעצב מערכת הניזונה מטרנסאקציות מסחריות בהיקף נרחב, מאחסנת אותם ומאפשרת הן ניטור סטטוס NRT, הן תחקור וחיפוש, והן איתור תבניות באמצעות למידת מכונה.

מטרה:

עיצוב ומימוש מערכת אנליטיקות נתוני עתק הרוחמת תשתיות NoSQL Databases, כלים מבית Hadoop Ecosystem ותפיסה מונחית שירותים (לוקאליים ומבוססי ענן) באמצעות תבנית ארכיטקטונית טיפוסית לסביבת מחשוב ענן היברידי.

דרישות פונקציונאליות:

- המערכת תקלוט טרנסאקציות הזמנת פיצות מרשת בעלת עשרות סניפים.
- המערכת תקלוט מסר אודות פתיחה או סגירת סניף.
- נתוני טרנסאקציות הזמנה תכלול לכל הפחות את הנתונים הבאים:
 - מספר מזהה הזמנה ייחודי
 - מספר מזהה סניף ייחודי
 - שם הסניף
 - איזור (צפון, חיפה, מרכז, דן, דרום)
 - תאריך ושעת הזמנה
 - סטטוס הזמנה (בביצוע, הושלמה)
 - רשימת תוספות להזמנת פיצה
- המערכת תציג באמצעות דשבורד את הנתונים העדכניים הבאים:
 - מספר סניפים פתוחים
 - זמן טיפול ממוצע כללי ברשת
 - סה"כ הזמנות פתוחות כללי ברשת
 - סה"כ הזמנות שמולאו היום
 - חמשת התוספות שהוזמנו הכי הרבה
 - חמשת הסניפים עם זמן טיפול הכי קצר
 - התפלגות הזמנות שמולאו לפי איזור
 - פריסת כמות ההזמנות במהלך היום
- המערכת תאפשר חיפוש בנתוני הזמנות לסניף בתאריך מסוים
 - * ניתן להוסיף פילטרים אלטרנטיבים לחיפוש ממוקד יותר
- המערכת תאפשר מציאת חוקי קשר בין סוגי התוספות מתוך נתוני טווח תאריכים ותאפשר בחירה של הצגתם בטבלה או באמצעות תרשים גרפי לבחירתכם

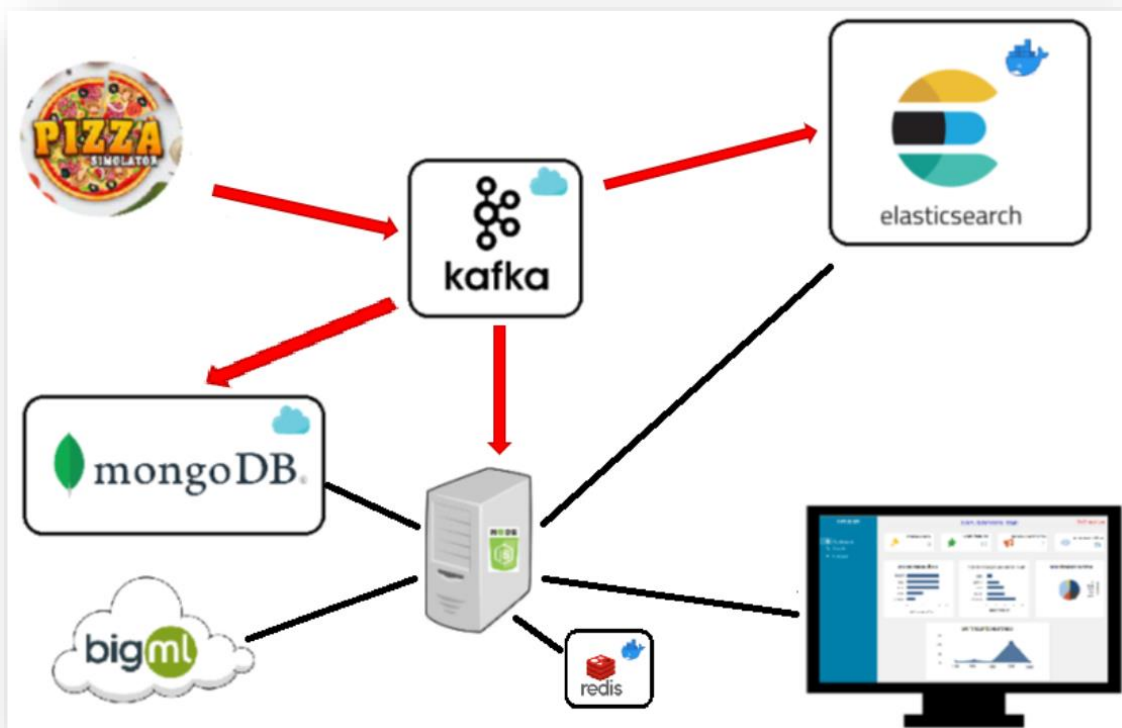
דרישות לא פונקציונליות:

- עיצוב המערכת הכללי יתבסס רעיונית על תבנית Lambda Architecture.
- שירותים יפותחו על בסיס תבנית Microservices.
- סימולטור ייצור נתוני טרנסאקציות להזמנות ואירועי פתיחה/סגירה של סניף. ניתן לכתוב את הסימולטור בכל שפה או תשתית.
- הסימולטור מפקח הזמנות פיצות בטיפול לרבות שעת הזמנה, באיזה סניף ואילו תוספות התבקשו.
- לאחר אינטרוול זמן אקראי הסימולטור ידווח כי ההזמנה הושלמה.
- שרת Kafka (בענן) יקבל מסרים ויפיץ אותם למנוע חיפושים Elastic Search (במיכל Docker), למד נתונים מסוג MongoDB (בענן)
- שרת מבוסס Node.js (לוקאלי) ישתמש במסדי הנתונים הבאים:
 - MongoDB (בענן): שליפת נתונים לצורך אימון מודל חוקי קשר באמצעות שירות ענן (BigML, AWS או כל שירות מתאים אחר).
 - Elasticsearch (במיכל Docker): ביצוע חיפוש נתוני הזמנות בטווח תאריכים.
 - Redis (במיכל Docker): שמירת סטטוס נתונים מרכזיים נכון לרגע זה ומוצגים בדשבורד
- שרת מבוסס Node.js (לוקאלי) יצור דשבורד מבוסס Cards ובו מדדים מסכמים:
 - סה"כ ההזמנות הנמצאות בטיפול כרגע
 - סה"כ ההזמנות שטופלו ביום זה
 - כמה סניפים פתוחים
 - זמן טיפול ממוצע בהזמנה
- בנוסף שרת מבוסס Node.js (לוקאלי) יציג ארבעה גרפים:
 - התפלגות הזמנות לפי אזור ביום זה
 - התפלגות הזמנות ביום זה (על ציר זמן)
 - חמשת הסניפים עם זמן טיפול הקצר ביותר ביום זה
 - חמשת התוספות המבוקשות ביותר ביום זה
- Redis שומר סטטוס כל הנתונים המוצגים בדשבורד ומעדכן באמצעות פרוטוקול WS.
- שרת מבוסס Node.js (לוקאלי) ישתמש בשירותי Bigml.com לבניית מודל אסוציאציות.
- ראו דיאגרמות 1-2 לתאור כללי של מבנה המערכת ושלבי זרימת הנתונים (חיצים אדומים מתארים זרימת מידע חדש, קווי אסוציאציה שחורים מתארים אינטראקציה עם שרת Node.js)
- ראו תאור מסכים המצורפים להלן (1-3)

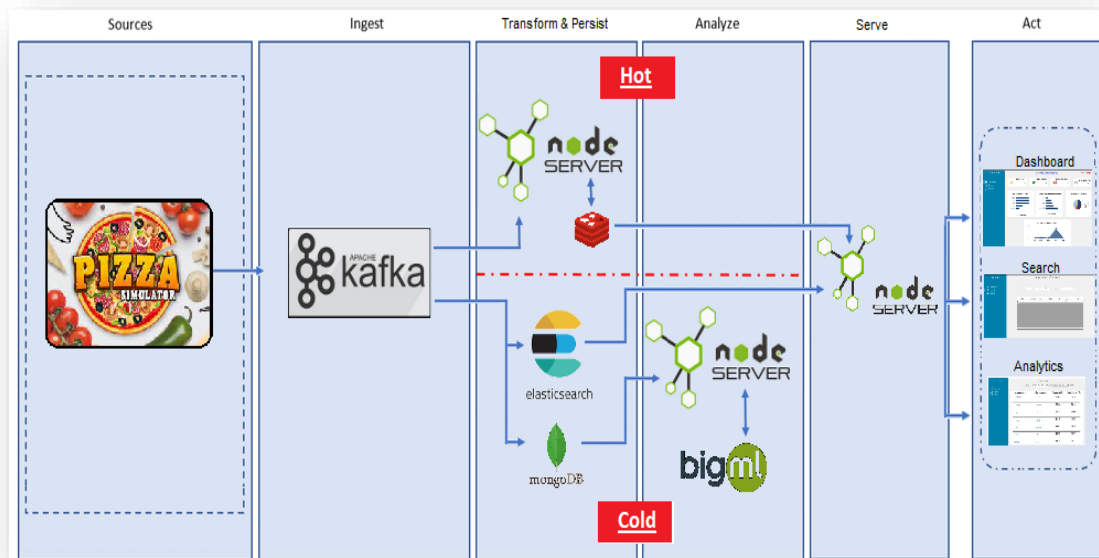
הגשה:

- ניתן להגיש את הפרויקט בקבוצות של ארבע.
- כל חבר בקבוצה חייב לקבל אחריות על חלק מוגדר במערכת ועליה יבחן.
- ציון הפרויקט הוא בסיס משותף + ציון פרטני לכל חבר בקבוצה.
- תאריך הגשה (עשוי להשתנות): 7.3.23
- נא להרשם בלינק הבא (גם לבוחן):

<https://docs.google.com/spreadsheets/d/1bEM1cJc0OeMkkNy43mVzSuBmYiNiiwam7ctxAZtdEU/edit?usp=sharing>

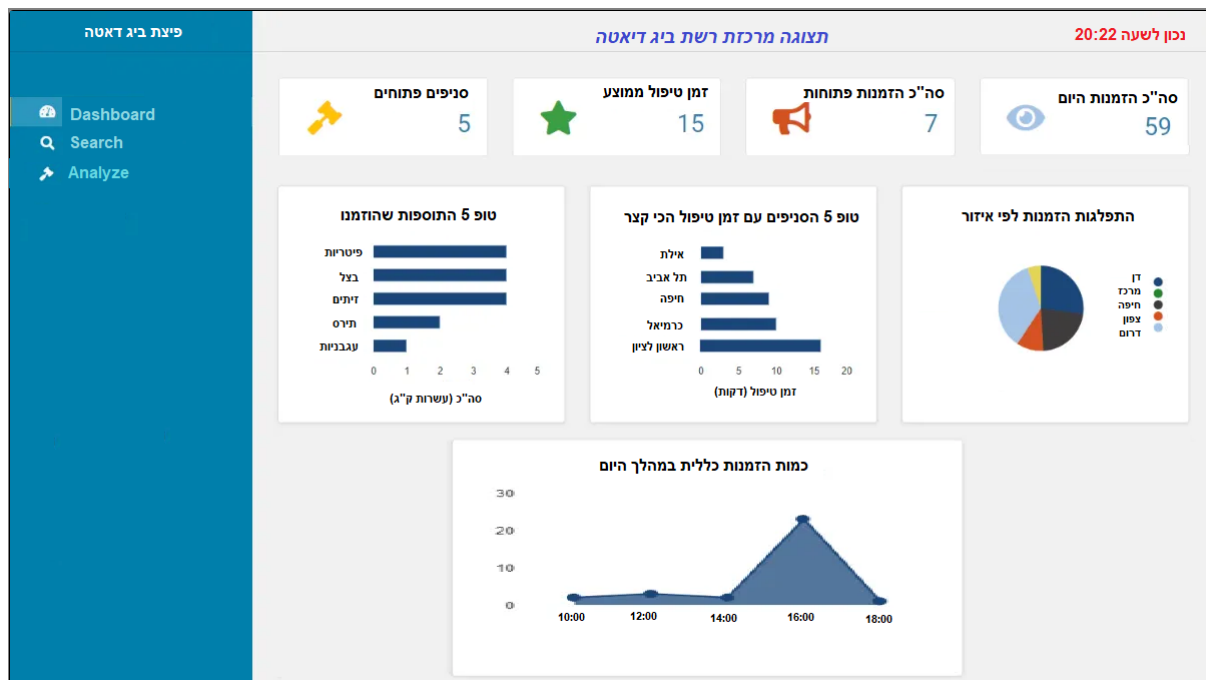


דיאגרמה 1 : מרכיבי מערכת חיפוש ואנליטיקות

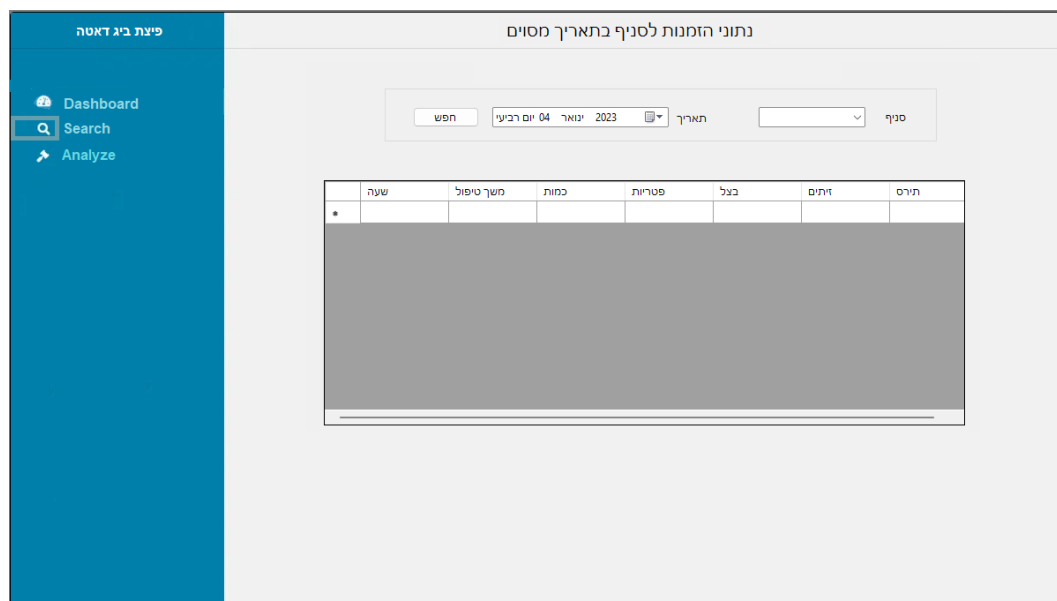


דיאגרמה 2 : Data Pipeline Stages

אוניברסיטת אריאל בשומרון
המחלקה למדעי המחשב



מסך 1 : דשבורד מרכז נתוני NRT



מסך 2 : חיפוש נתונים בחתר תאריכים

פיצת ביג דיאטה

מציאת חוקי קשר

צור מודל מתאריך 04 ינואר 2023 לתאריך 04 ינואר 2023

Antecedent	Consequent	Support (%)	Confidence (%)
פטרייות	בצל	26.69	38.58
פטרייות	זיתים	23.00	79.43
בצל	זיתים	22.18	76.06
זיתים	תירס	18.48	76.27
תירס	פיטריות	15.61	73.79
חצילים	בצל	14.78	75.79
עגבניות	תירס	13.35	76.47

מסך 3 : אימון ומציאת חוקי אסוציאציה

קישורי עזר:

1. הסברים אודות חוקי אסוציאציה לניתוח סל קניות :

[Visualizing Market Basket Analysis \(analyticsvidhya.com\)](https://analyticsvidhya.com/visualizing-market-basket-analysis/)

[Simple Market Basket Analysis with Association Rules Mining | by Chi Nguyen | Towards Data Science](#)

2. ויזואליזצית קשרים באמצעות ספריית D3:

[RPods - Market Basket Analysis with D3 JS library](#)

3. מתאם המחר בין קפקא ל-ES:

[Kafka Connect Elasticsearch Sink Connector \(kloia.com\)](https://kloia.com/kafka-connect-elasticsearch-sink-connector/)

4. מתאם המחר בין קפקא ל- MongoDB :

[Getting started with the MongoDB Connector for Apache Kafka and MongoDB Atlas | MongoDB](#)