

9.4.23

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.

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

בפרויקט זה נעצב מערכת הניזונה ממסרים המתריעים על אירועים אסטרונומיים, מציגה אותם בזמן אמת (NRT) או ברמת עדכון יומית, מאחסנת ומאפשרת תחקור וחיפוש, כמו גם איתור תבניות באמצעות למידת מכונה כבונה.

רקע:

השראה למערכות מסוג זה ניתן למצוא במאמר ובלינק הבא :

<https://iopscience.iop.org/article/10.1088/1538-3873/aae904/meta>

<https://github.com/growth-astro/ztfrest>

- אירועים אסטרונומיים מחוללים אותות פיסיקליים כגון גלי כבידה או קרינה אלקטרו מגנטית לסוגיה (גלי רדיו, מיקרוגל, אינפרא אדום, אור נראה, אולטרא סגול, קרינת רנטגן וגמא) בעקבות אירועים כגון התנגשות חורים שחורים, סופרנובות, פולסארים והבהקי כוכבים (flares).
- אירועים אלו מכונים "אירועים חולפים" (Transients) והינם אבני בניין למחקרים אסטרונומיים מבוססי זמן.
- קיימת חשיבות רבה להפיץ הודעה למצפים ברחבי העולם ולמפעילי טלסקופים, מיד עם גילוי אחד מאירועים אלה, בכדי שיפנו אותם לכיוון ההתרחשות כך שניתן יהיה לאסוף את מרבית הנתונים הרלוונטים למחקר.
- סוג נוסף של נתונים הינו רשימת אובייקטים העומדים לעבור בקרבת כדור הארץ ביממה הקרובה (Near Earth Objects).
- כמובן תחזית שעתית/יומית אודות פעילות מגנטית וסערות שמש חשובה למפעילי לווינים ותשתיות לאומיות (כגון חברת חשמל) בעקבות הסכנות הטמונות ברמות גבוהות של פעילות זו.

מטרה:

עיבוד ומימוש מערכת התרעות ואנליטיקות נתוני עתק הוותמת תשתיות NoSQL Databases, כלים מבית Hadoop Ecosystem ותפיסה מונחית שירותים (לוקאליים ומבוססי ענן) באמצעות תבנית ארכיטקטונית טיפוסית לסביבת מחשוב ענן היברידית.

דרישות פונקציונאליות:

- למערכת סוג משתמש אחד ואין צורך בניהול משתמשים.
- המערכת תשמור לוקאלית עותק של קטלוג אובייקטים שמיימים בהירים ([Bright Star Catalog](#))
- המערכת תקלוט מסרים מסימולטור אודות אירועים אסטרונומיים.
- כל מסר יכלול:
 - תאריך ושעה (UTC)
 - גורם מודיע (מתוך רשימה של [בעשרה מצפים](#) או טלסקופים לוויניים).
 - מיקום ביחידות [DEC - I RA](#)
 - סוג האירוע:
 - התפרצות קרינת גאמה (GRB)
 - עליה משמעותית ברמת הארה נצפית (Apparent Brightness Rise)
 - עליה משמעותית בקרינת UV (UV Rise)
 - עליה משמעותית בקרינת רנטגן (X-Ray Rise)
 - גילוי כוכב שביט (Comet)
 - רמת דחיפות התייחסות (1-5)
- המערכת תציג רשימת נתוני האירועים האחרונים במסגרת Dashboard, אירוע אחרון יוצג בנפרד עם פירוט מלא אודות המקור (נניח פרטי הכוכב) ואם רמת הדחיפות היא 4 ומעלה הוא יודגש וימומשו אפקטי UX למשיכת תשומת לב המשתמש.
- המערכת תאפשר חיפוש ואיתור:
 - כל המסרים בהם מעורב גוף שמימי כלשהו בטווח זמנים.
 - כל המסרים, או מסרים מסוג מסוים, בטווח זמנים.
 - כל המסרים שמקורם ממצפה/טלסקופ מסוים בטווח זמנים.
- המערכת תציג את רשימת הגופים האמורים לעבור בקרבת כ"א ב-24 השעות הקרובות.
- המערכת תציג את תחזית פעילות השמש לשעות הקרובות
- בנוסף: המערכת תחפש תבניות המחוברות בין מאפייני מזג האוויר\פגעי טבע לפעילות השמש.

דרישות לא פונקציונאליות:

- עיצוב המערכת הכללי יתבסס רעיונית על תבנית Lambda Architecture.
- שירותים יפותחו על בסיס תבנית Microservices.
- יש לבסס את שרת האפליקציה המרכזי על Node.js/Express אך ניתן לשלב גם שירותי FastApi/Python.
- נתוני המקורות הפוטנציאליים לאירועים אסטרונומיים (כוכבים, ערפיליות) יועתקו מקטלוג אוניברסיטת הרווארד מהלינק הבא, ויונהלו על ידי שרת Redis (לוקאלית או ע"ג Docker):

<https://github.com/aduboisforge/Bright-Star-Catalog-JSON>

- נתונים אודות אסטרואידים הנמצאים בקרבת כ"א ניתן למצוא בלינק ראשון (תצוגה) ולינק שני (שירות Asteroids- NeoWs):

[NEO Earth Close Approaches \(nasa.gov\)](#)

[NASA Open APIs](#)

- שרת Kafka (בענן) יקבל מסרים ויפיץ אותם למנוע חיפושים Elastic Search (במיכל Docker).
- נתוני פעילות השמש ניתן למצוא בלינק המצורף (או כל מקור אחר) – את הנתונים ניתן למצות באמצעות Web Scrapping:

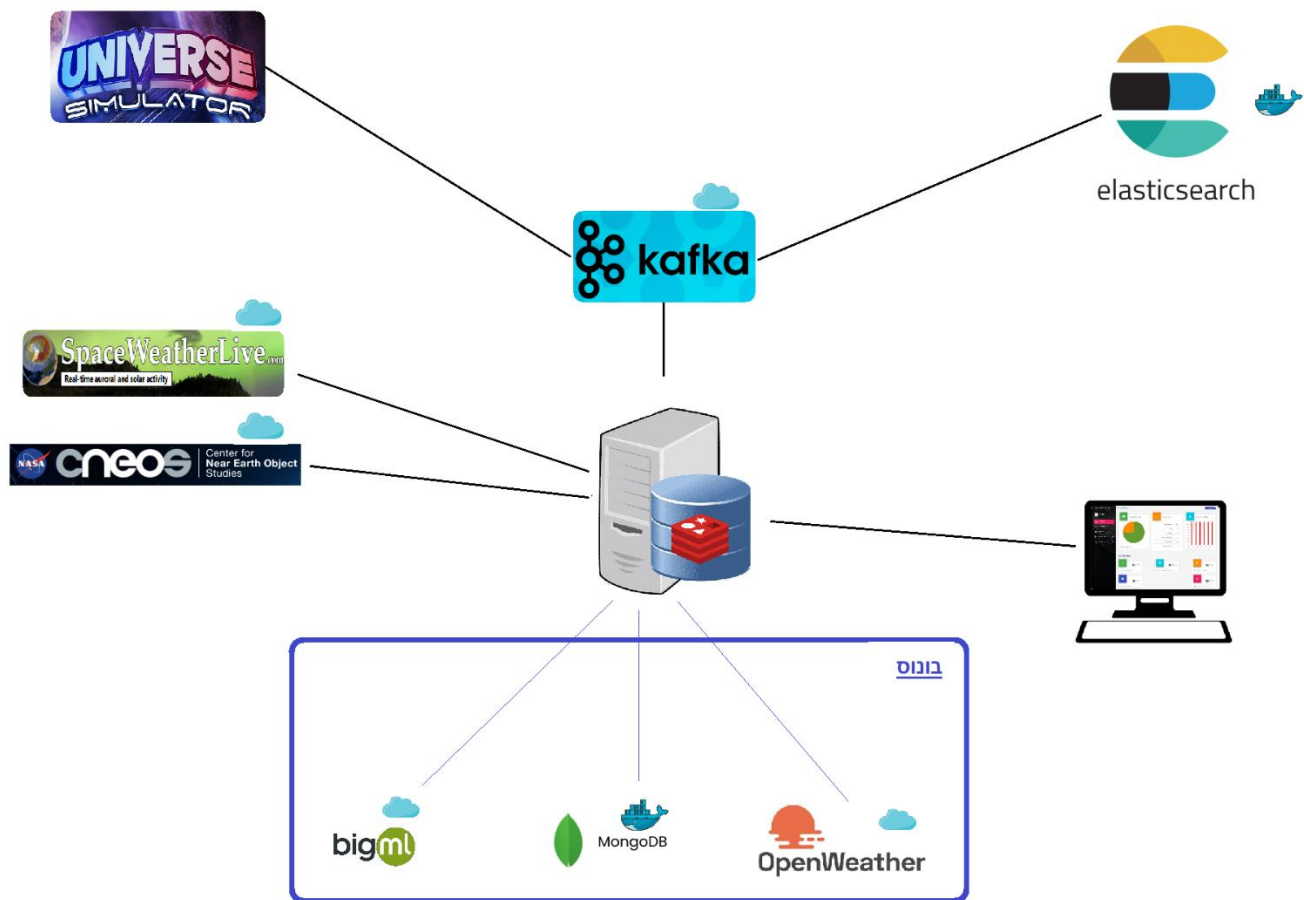
[Solar activity | SpaceWeatherLive.com](#)

- שרת מבוסס Node.js (לוקאלי) ישתמש במסדי הנתונים הבאים:
 - Elastic Search (Docker): ביצוע שאילתות לגבי נתוני אירועים אסטרונומיים כמפורט בדרישות לעיל.
 - Redis (לוקאלי או Docker): יישום כ-Cache לאיתור מהיר של נתוני כוכבים וגופים נוספים.
 - MongoDB (בענן) בבנוס : אחסון נתוני פעילות שמש לצורך ביצוע למידת מכונה.
- שרת מבוסס Node.js (לוקאלי) יצור דשבורד מבוסס Cards ובו מדדים מסכמים המציגים את מספר האירועים מכל סוג שדווחו ביממה האחרונה, ובמידה ומגיעה הודעה דחופה יציג חלונית מהבהבת (באמצעות פרוטוקול WS).
- בנוסף שרת מבוסס Node.js (לוקאלי) יציג:
 - התפלגות סוגי האירועים בשבוע האחרון (בחתך דחיפות 1-5)
 - גראף התפלגות אסטרואידים שעברו בקרבת כ"א בחודש האחרון (בחתך גודל)
 - טבלת NEO העתידים לעבור בקרבת כ"א ב-24 השעות הקרובות ונתונים
 - גראף פעילות השמש בשעתיים אחרונות (רמות קרינת X)
 - הדמיית הפעילות ומראה פני השמש בזמן הנוכחי
- פרטים אודות כוכב\אובייקט שמימי ניתן לקבל ממגוון שירותים או באמצעות קריאת http או ע"י שימוש בנתוני פרויקט גאיה וחבילת פייתון מתאימה (https://allendowney.github.io/AstronomicalData/01_query.html) או קטלוג SIMBAD ([How to query SIMBAD ? \(u-strasbg.fr\)](https://simbad.u-strasbg.fr/SIMBAD/))
- **בנוס (חובה למגישים ברביעיה):**
 - שמירת נתוני פעילות שמש בשרת נתונים MongoDB בענן.
 - קריאה וארגון נתוני מזג אויר משירות כדוגמת [OpenWeather](https://openweathermap.org/) או תופעות טבע משירות [EONET](https://eonet.co.il/) של נאס"א, בשרת נתונים מסוג MongoDB בענן.
 - בניית מודל חוקי אסוציאציה בין מאפייני פעילות שמש למאפייני מזג אויר (ממוצעים או באיזור מסוים) באמצעות שירותי Bigml.com.
- **בנוס:** אירוח השרת הלוקאלי בספק שירותי ענן (ניתן למצוא חינוכיים לחודש).
- 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 המתארים באופן כללי את מבנה התצוגה (התעלמו מכותרות).

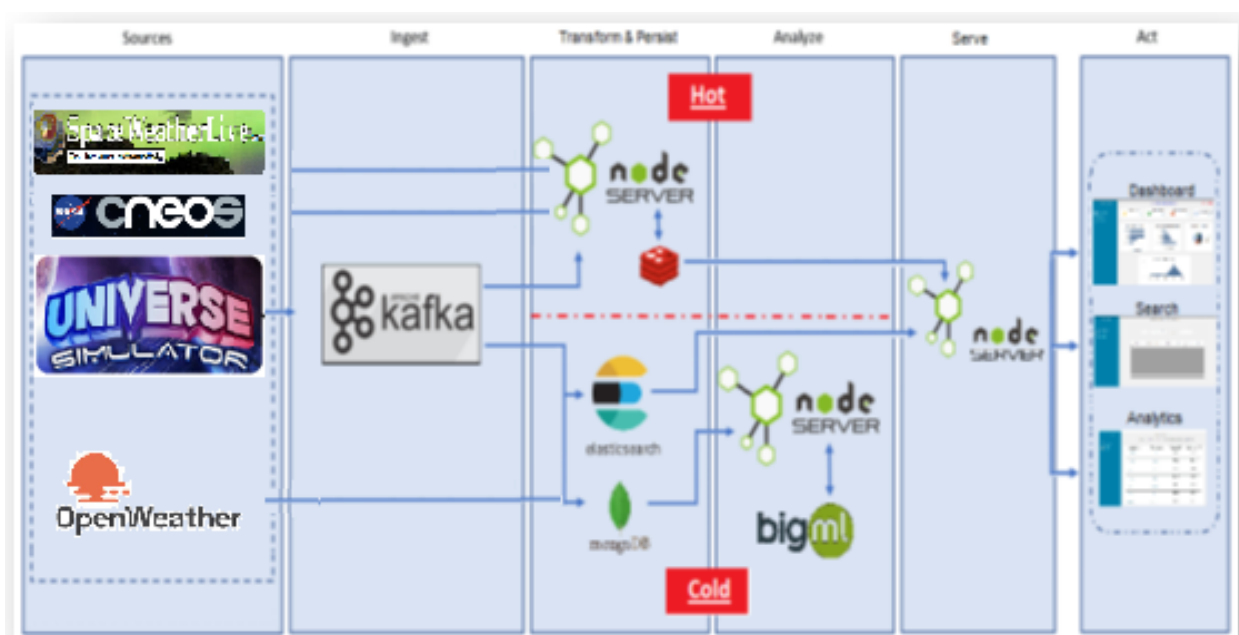
הגשה:

- ניתן להגיש את הפרויקט בקבוצות של ארבע – רביעיה נדרשת לממש את חלק הבנוס.
- כל חבר בקבוצה חייב לקבל אחריות על חלק מוגדר במערכת ועליה ייבחן.
- ציון הפרויקט הוא בסיס משותף + ציון פרטני לכל חבר בקבוצה.
- תאריך הגשה (עשוי להשתנות): 31.7.23
- נא להרשם בלינק הבא (גם לבחון) :

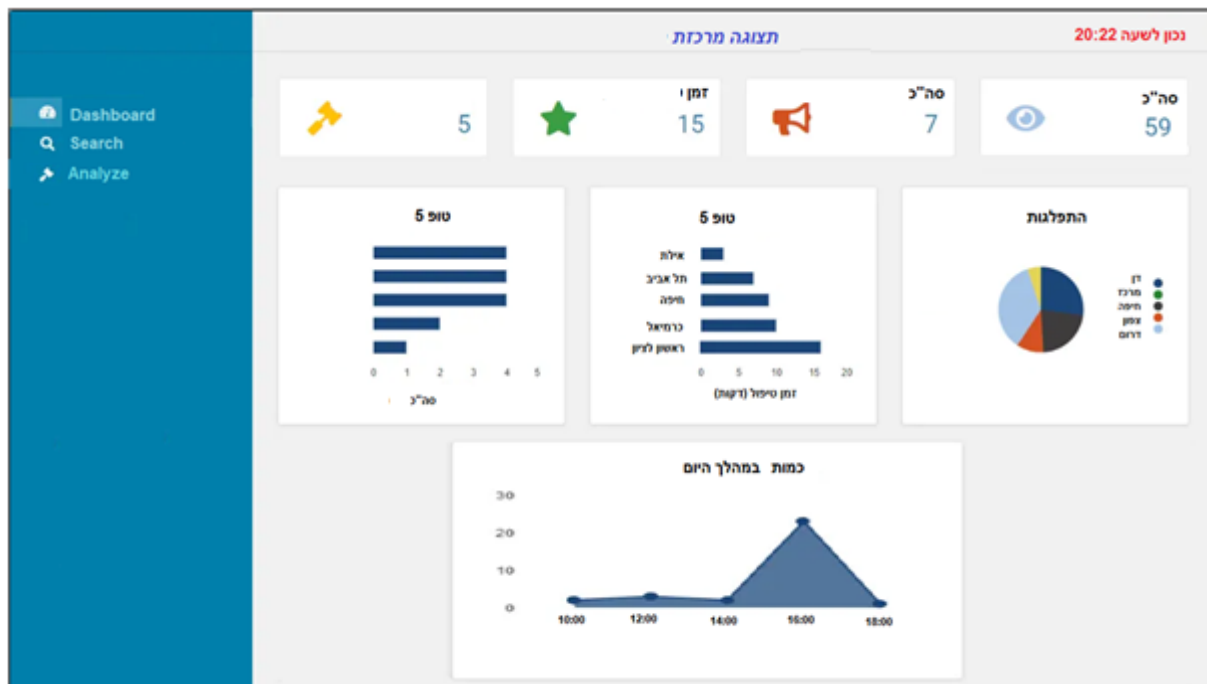
<https://docs.google.com/spreadsheets/d/1x4DenGYwCkS1s3R-4wEhnlQI5Qp0MF2J0-XLOCH00kE/edit#gid=392570041>



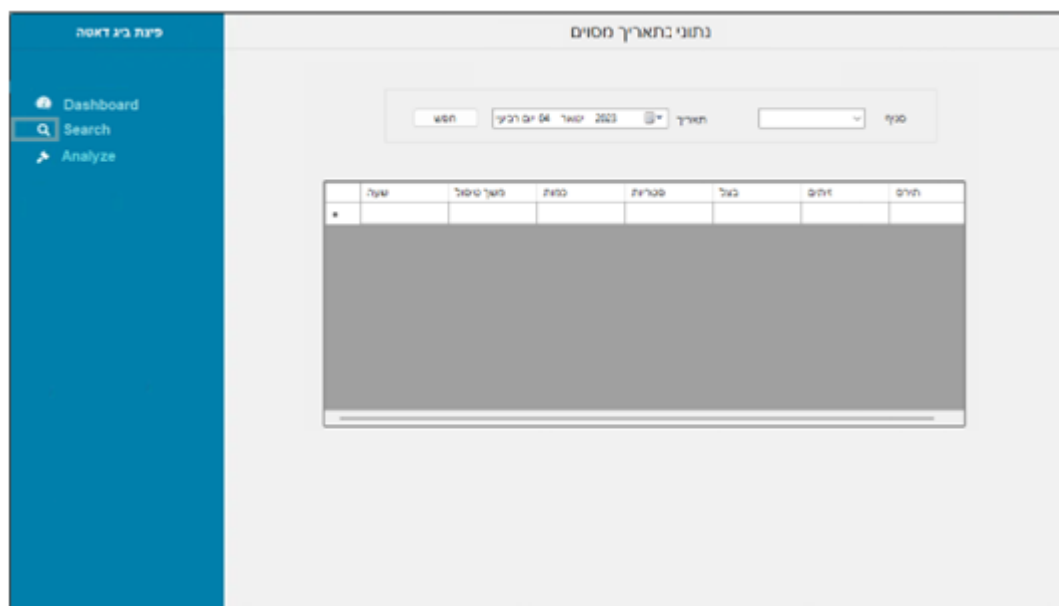
דיאגרמה 1: מרכיבי מערכת ניטור, חיפוש ואנליטיקות



דיאגרמה 2 : Data Pipeline Stages



מסך 1 : דשבורד מרכז נתוני NRT





Antecedent	Consequent	Support (%)	Confidence (%)
רמת קרינת X	טמפרטורה נמוכה	26.69	38.58
מספר בתמי שמש	טמפרטורה גבוהה	23.00	79.43
מספר flares	טמפרטורה נמוכה	22.18	76.06
		18.48	76.27
		15.61	73.79
		14.78	75.79
		13.35	76.47

מסך 3 : אימון ומציאת חוקי אסוציאציה

קישורי עזר:

1. הסברים אודות חוקי אסוציאציה לניתוח בין מוצרים (שקול לאירועים) :

[Visualizing Market Basket Analysis \(analyticsvidhya.com\)](https://analyticsvidhya.com/)

[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 המחבר בין קפקא ל-ES:

[Kafka Connect Elasticsearch Sink Connector \(kloia.com\)](https://kloia.com/)

4. מתאם המחבר בין קפקא ל- MongoDB :

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

בהצלחה.

