

1/6/2021

נתוני עתק עיבוד בענן פרויקט סיום

כללי

מערכות לעיבוד נתוני עתק משמשות כלי רב עוצמה לניהול סיוכונים, הפקת תובנות עסקיות וגיוש אסטרטגיה בארגונים. הדרישה לפיתוח ועיצוב מערכות מסוג זה תופסת תאוצה ע"י שילוב משאבי On-premise עם שירותי ענן מספקים מובילים.

מערכות נתוני עתק מאפיינות חברות דיגיטליות המנצלות את עידן האינטרנט לביצוע משימות ליבה באמצעים שאינם דורשים דיגיטציה מיוחדת לתהליכים, אלא מהווים חלק מלא מהם: לגוגל, פייסבוק וחברות כמותן אין קיום שלא בעולם הסייבר ומלכתחילה כל פעילותן מתועדת וממוחשבת - אי לכך המעבר לניצול נתונים תפעוליים לצורך קבלת החלטות וביצוע Big Data Analytics הוא טבעי, מאחר והתמודדות עם אתגרי נפחים ועומסים היא מראש חלק אינטגרלי מסדר יומם הטכנולוגי.

דוגמאות לשימוש נתוני עתק כבסיס לניתוח וקבלת החלטות כוללים שימוש בלוחות מכוונים (Dashboards) המנטרים סטטוס ערכי מפתח בזמן אמת, וכן ביצוע אנליטיקות על גבי תקופה בכדי לאתר תבניות, חוקיות או קשרים בתוך נתונים אלה.

בפרויקט זה נעצב מערכת המקבלת שטף מסרים (Data Stream), מאחסנת ומעבדת אותם בתבניות ארכיטקטוניות אופייניות ומשתמשת בתשתיות נתוני עתק מקובלות לניתוחם באופן מהיר ואדפטיבי.

מטרה

עיצוב מערכת לוחות מכוונים ואנליטיקות בתפיסת נתוני עתק הרוותמת NoSQL Databases, כלים מבית Hadoop Ecosystem ותפיסה מונחית שירותים בארכיטקטורה טיפוסית ומימוש אתגרי אינטגרציה תשתיות.

דרישות פונקציונליות

- המערכת תאפשר ניהול מקטעי כביש אגרה. לכביש 5 מקטעים (ראו מקטעי כביש 6 לדוגמה בתמונה 1).
- בכל רגע ניתן לדעת מה מספר הרכבים הנמצאים בכל מקטע.
- למקטע נכנסים רכבים ויוצאים רכבים באופן שוטף, סימולטור יחולל אירועים אלה ויגרור עדכונים.
- כל אירוע יכלול נתונים כגון: סוג אירוע (כניסה לכביש\כניסה למקטע\יציאה מהכביש\יציאה ממקטע), מקטע, סוג מכונית (פרטית, מסחרית, משאית), יום בשבוע, שעה, האם יום מיוחד (במקרה של חגים או חריגים).
- המערכת תציג על גבי לוח מכוונים את מספר הרכבים הנוכחי ותאפשר לבקש פירוט רשימת הרכבים.
- המערכת תבנה מודל חיזוי מבוסס עץ החלטה: בהנתן פרטי מקטע כניסה לכביש, סוג רכב, יום בשבוע ושעת כניסה יתבצע חיזוי מאיזה מקטע תצא המכונית.
- ניתן להציג מסך Confusion Matrix ולבקש להתחיל לבצע הערכה למודל: בכל פעם שהסימולטור שולח הודעה על כניסת רכב לכביש יתבצע חיזוי למקטע היציאה, וברגע שהרכב יוצא יעודכן האם צדק או טעה ומדד Accuracy יחושב ויוצג דינאמית (ראו תרשים 1).

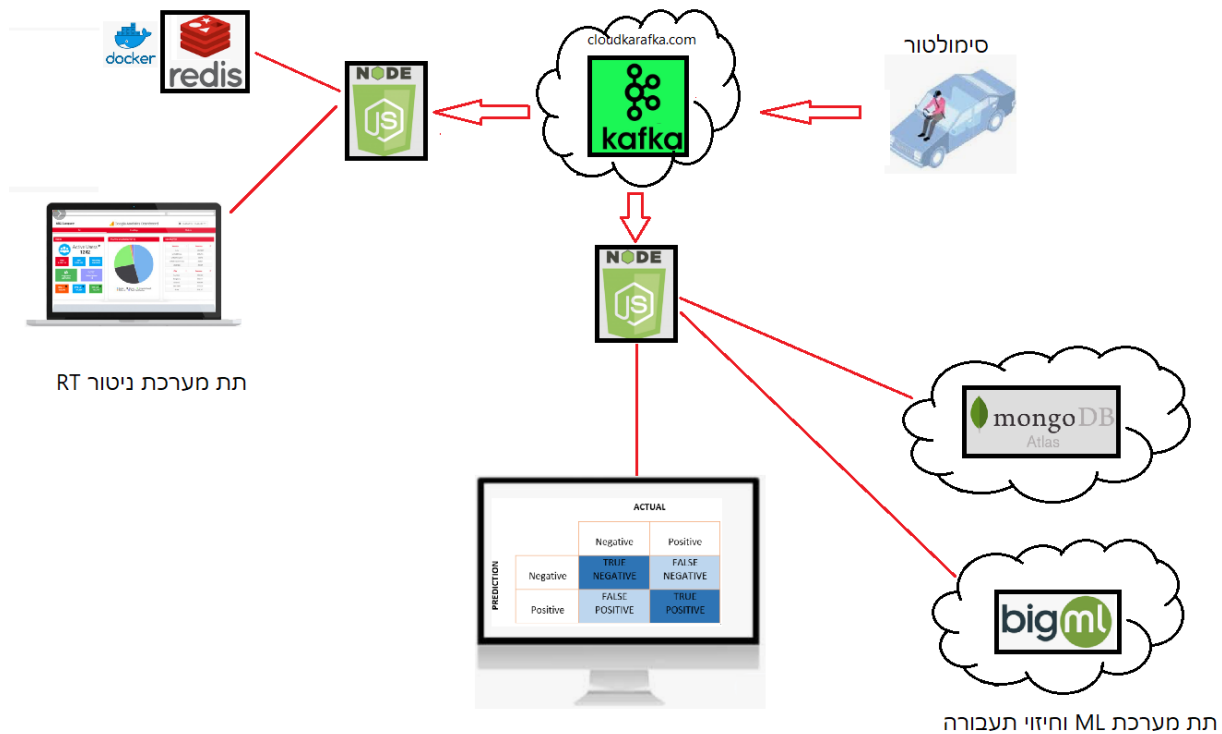
מקטע יציאה חזוי

	1	2	3	4	5
מקטע יציאה בפועל	1	2	3	4	5
1					
2					
3					
4					
5					

תרשים 1: דוגמת מטריצת ערפול

דרישות לא פונקציונאליות

- סימולטור יפיק מסרי הודעות אודות רכבים שצולמו נוסעים במקטע כביש מסוים.
- המערכת תעוצב בגישה מונחית Micro-Services.
- עיצוב המערכת הכללי יתבסס רעיונית על תבנית Lambda architecture.
- ארגון הקוד בשרתי המערכת יתבסס על תבנית MVC.
- הבסיס לשרת האפליקציה ולתצוגה יהיה Node.js וחבילות הרחבה מתאימות (לרבות Express.js).
- המערכת תשתמש בשני מסדי נתונים מסוג NoSQL: Redis & MongoDB. האחד ישמור נתונים מעובדים לשימושים היסטוריים ארוכי טווח, והשני ישמור נתונים מעודכים אד-הוק וזמינים לשאלות מהירות.
- יש להשתמש בשירותי DBaaS למסד נתונים מסוג MongoDB (מומלץ Atlas).
- יש להשתמש ב-Docker Image להפעלת מסד נתונים מסוג Redis (הגרסה הרשמית מומלצת).
- המערכת תשתמש במתווך מסרים Kafka ממשפחת Hadoop Ecosystem (מומלץ ספק שירותי ענן הנ"ל <https://www.cloudkarafka.com>).
- להפקת מודל חיזוי השתמשו בשירותי BigML.com.
- אופציונאלי: אירוח שירותי Node.js שפיתחתם במסגרת heroku.com.
- תרשים 2 מציג מיפוי טכנולוגי של המערכת:



תרשים 2: מיפוי טכנולוגי מוצע למערכת ניטור וניתוח תעבורה במקטעי כביש

הערות:

1. יש להגיש את הפרוייקט בקבוצות של שלושה סטודנטים.
2. סטודנט יציין בהגשה מה היתה אחריותו ומה מבין המודולים השונים הוא המבצע ועליהם יישאל.
3. במהלך ההגשה יישאלו שאלות על החומר הנלמד.
4. יש להרשם בלינק הבא לשעת הגשה (נמצא גם במודל):

<https://docs.google.com/spreadsheets/d/1ewhldDQRKr8Tq-PwVGDXSACFal3AOnuPXf4URHlbg0/edit#gid=0>

5. מועד הגשה: 3.8.21

בהצלחה.



תמונה 1 : דוגמת מקטעי כביש 6