**מס' תלמיד:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**קמפוס:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

למילוי ע"י הסטודנט

פתרון

* שם המרצה: ד"ר מאיר גולדנברג, ד"ר אריאלה ריכרדסון
* תאריך הבחינה: 09/07/2021
* משך הבחינה (בדקות): 90 דקות
* חומר עזר מותר לשימוש: דף A4 כתוב משני צידיו
* מחשבון: ללא
* המבחן כולל סה"כ 10 שאלות, יש לענות על כל השאלות.
* את התשובות יש לכתוב **ע"ג השאלון**, דפי הטיוטא לא ייבדקו.
* פירוט ניקוד: ישנם 10 שאלות, בכל שאלה 4 סעיפים.
  + לכל סעיף יינתן ניקוד של 2.5 (סה"כ 100 נקודות)
  + על כל שאלה יש לענות נכון/לא נכון.

**תלמיד יקר,**

1. **נוהל הבחינות של המרכז האקדמי לב מחייב אותך**, באחריותך לקוראו ולהכירו - בחינה עלולה להיפסל על כל חריגה מהנוהל.
2. אם אינך מבין את כוונת המרצה בשאלה כלשהי, עליך לכתוב בראש התשובה כיצד הינך מבין את השאלה ולפתור בהתאם. המרצה ישקול האם יש מקום להבנה זו ואז ינקד בהתאם.
3. לידיעתך, תורדנה נקודות לא רק על שגיאות, אלא גם על תוספות לא רלוונטיות, העדר נימוק הולם לתשובה, חוסר סדר ותשובה דו-משמעית, כאשר נדרשת תשובה חד משמעית.

***בהצלחה רבה !***

# לינוקס

1. נתונה הפקודה:

cat ./spar\* | grep "some"

סמן עבור כל משפט נכון/לא נכון:

* 1. הפלט יהיה מורכב משמות קבצים (אם יש קבצים מתאימים).

נכון לא נכון

* 1. ייתכן שהפלט יכלול תוכן חלקי של מספר קבצים.

נכון לא נכון

* 1. grep תקבל קלט מהמשתמש.

נכון לא נכון

* 1. תוצאת הפקודה תשתנה אם נוריד את הנקודה והלוכסן (סלש) שאחרי הרווח הראשון בפקודה.

נכון לא נכון

# Hadoop ו- MapReduce.

1. ראינו את התמונה הבאה שממחישה איך קובץ שמור ב- HDFS:

Graphical user interface

Description automatically generated

סמן עבור כל משפט נכון/לא נכון:

* 1. בתמונה זאת מצוירים שלושה מחשבים עם ארבעה מעבדים בכל מחשב.

נכון לא נכון

* 1. NameNode חייב לשמור מידע על הימצאות של כל בלוק.

נכון לא נכון

* 1. מחשבים באותה שורה של הציור (כגון 1, 5, 9) נמצאים בקרבה גאוגרפית.

נכון לא נכון

* 1. HDFS אינה אחראית על חלקות משבאים חישוביים באשכול Hadoop.

נכון לא נכון

1. סמן עבור כל משפט נכון/לא נכון:
   1. נניח קובץ קלט ששומר מחירון:

Apple 5

Banana 7

Potato 3

Pepper 6

Broccoli 20

וכו'

ניתן להשתמש ב- MapReduce כדי לקבוע, עבור כל אות מ A עד Z, את ממוצע המחירים של פריטים ששמם מתחיל באותה האות.

נכון לא נכון

* 1. עבור קובץ הקלט למעלה, ניתן לחשב את כמות ה- Mapper-ים ככמות השורות חלקי 26 (שזה מספר האותיות השונות).

נכון לא נכון

* 1. אם Reducer מגדיר מערך, המערך הזה יהיה מפוזר ביו זיכרונות של מחשבים שונים באשכול.

נכון לא נכון

* 1. תהליך MapReduce עבור קובץ קלט מהשאלה הקודמת ישתמש בשלושה Reducers.

נכון לא נכון

1. נניח קובץ קלט השומר מחירון פירות וירקות, ביחד אם אינדקס של ערך תזונתי כלשהו בעמודה האחרונה:

Apple 5 23.45

Banana 7 25.9

Potato 3 13.23

Pepper 6 18.46

Broccoli 20 50.1

וכו'

סמן עבור כל משפט נכון/לא נכון:

* 1. שימו לב ש- מי שקונה Apple קונה בשקל אחד ערך תזונתי שווה 23.45/5=4.69. נניח שאנחנו רוצים לחשב ערך תזונתי מקסימלי שנקנה בשקל אחד (כלומר, מספר אחד). כדאי להשתמש בתהליך MapReduce אחד, אף שלא נגיע לתשובה הסופית אם נשתמש בכמה Reducer-ים.

נכון לא נכון

* 1. כדי לפתור את השאלה באופציה הקודמת, צריך להשתמש בשיטת שימוש במידע על ההקשר (using context information) כמו שעשינו בדוגמא של מכפלה סקלרית של וקטורים.

נכון לא נכון

* 1. נניח שקיים קובץ קלט שני השומר דעה אחרת על ערך תזונתי:

Apple 13.45

Banana 15.9

Potato 23.23

Pepper 28.46

Broccoli 30.1

וכו'

נניח שאנחנו רוצים לחשב רשימת פירות וירקות שלפי הדעה השנייה הערך שלהם גדול מהערך שלהם לפי דעת המופיעה במחירון (הרשימה יכולה להיות מחולקת למספר קבצים). אם אנחנו רוצים להשתמש בשיטת MapReduce, אז עלינו להשתמש בשרשור תהליכי MapReduce.

נכון לא נכון

* 1. סביר להניח שניתן לבצע את החישוב באופציה הקודמת מהר יותר ללא שימוש ב- Hadoop.

נכון לא נכון

# Spark

1. סמן עבור כל משפט נכון/לא נכון:
   1. אחד ההבדלים בין Spark ל- MapReduce הוא שב- Spark כותבים Mapper ו- Reducer ביחד בקובץ אחד. זהו גם ההבדל העיקרי.

נכון לא נכון

* 1. ייתכן שאלה שדרשה שימוש בשיטת שרשור תהליכי MapReduce, ואפשר לפתור אותה בעזרת תהליך Spark בודד.

נכון לא נכון

* 1. אין דרך לדעת איך Spark החליטה לחלק את הנתונים לבלוקים (partitions).

נכון לא נכון

* 1. ייתכן שביצוע חישוב בעזרת Spark יצריך מחשבים יקרים יותר מביצוע של אותו החישוב בעזרת MapReduce.

נכון לא נכון

1. נתונות שורות קוד הבאות:

Text

Description automatically generated with medium confidence

סמן עבור כל משפט נכון/לא נכון:

* 1. הטרנספורמציה בשורה 2 לא תגרום להעברת נתונים דרך הרשת.

נכון לא נכון

* 1. ביצוע של שורה 3 תגרום (בפרט) לביצוע של הטרנספורמציה בשורה 2.

נכון לא נכון

* 1. ייתכן שהפלט בשורה 5 הוא: [(2, 'a'), (1, 'a')]

נכון לא נכון

* 1. שינוי של הארגומנט השני לפונקציית parallelize בשורה 1 מ- 2 ל- 1 תגרום בוודאות לשינוי של פלט הקוד.

נכון לא נכון

# **RegEx**

1. נניח שנתון קובץ המכיל שמות, כתובת (רחוב מספר ועיר - מופרדים ע"י פסיק), ומיקוד. העמודות מופרדות ע"י תווי \t, אך ייתכנו שגיאות.

Text, letter

Description automatically generated

סמן עבור כל משפט נכון/לא נכון:

1. הביטוי הרגולרי הבא:

^C[^\t]+\t[^\t]+\t

יחזיר תוצאות משורות 1 ו 3 בלבד.

נכון לא נכון

1. הביטוי הרגולרי הבא:

^[^\t]+\t[^\t]+\t[\d]{5}

יחזיר בין השאר תוצאות משורה 4.

נכון לא נכון

1. הביטוי הרגולרי הבא:

[0-9]\*

מתאים את כל תוכן הקובץ כי בכל שורה יש ספרות.

נכון לא נכון

1. ניתן לתקן את שורה 5 ע"י ההחלפה של

([\w]\*,[\w]\*,[\w]\*),

ב

\1\t

נכון לא נכון

# **KNIME**

1. סמן עבור כל משפט נכון/לא נכון:
   1. לכל קודקוד יש לפחות כניסה (קלט) אחת.

נכון לא נכון

* 1. ישנו סוג קודקוד יחיד וההבדל בין הקודקודים נקבע רק לפי הפרמטרים שמגדירים באופציות (configuration) שלו.

נכון לא נכון

* 1. כאשר משנים את ההגדרות של קודקוד, אין צורך להפעיל מחדש את הקודקוד, השינויים מופעלים אוטומטית.

נכון לא נכון

* 1. ישנם סוגים שונים של קודקודים המחולקים לפי קטגוריות ועבור כל קודקוד ניתן גם להגדיר במדויק איך הוא יפעל בעזרת configure.

נכון לא נכון

1. בתרשים Knime הבא מוצגת אנליזה של נתונים על לקוחות בנק. ניתן להניח שכל ההגדרות שמופיעות הן חוקיות.

סמן עבור כל משפט נכון/לא נכון:

(בכותרת של חלון המאפיינים בסוגריים מופיע המספר של הקודקוד שהוא מתייחס אליו)

A picture containing indoor, wall, toilet, tiled

Description automatically generated

Graphical user interface, text, application

Description automatically generated

1. יתכן שחלק מהפלט של קודקוד 2 הוא:

Graphical user interface

Description automatically generated with medium confidence

נכון לא נכון

1. יתכן שחלק מהפלט של קודקוד 2 הוא:

A picture containing graphical user interface

Description automatically generated

נכון לא נכון

1. אין דרך לדעת מה יהיה הפלט של קודקוד 2, כי התמונה של הקודקודים חתוכה ולא רואים מה קורה מחוץ לתמונה.

נכון לא נכון

1. הפעולה שמתבצעת בקודקוד 2 תגרום למספר השורות להיות קטן או שווה למספר השורות שהיו בפלט של קודקוד 1.

נכון לא נכון

# **TABLEAU**

**Chart

Description automatically generated**  **Chart, bar chart

Description automatically generated**

בתמונה רואים שני תרשימים שנתקבלו מתוך Tableau. על סמך קובץ הנטישה (Churning) מהבנק שראינו בהרצאה.

סמן עבור כל משפט נכון/לא נכון:

1. ייתכן שהצביעה השונה בין הגרף הימני לשמאלי נגרמה על ידי גרירה של אטריבוטים שונים ל Color.

נכון לא נכון

1. שימו לב שהגילאים של הלקוחות המופיעים על ציר ה X הם עגולים. זה בוודאי נגרם מהעובדה שהגילאים הרשומים בקובץ הם מספרים עגולים.

נכון לא נכון

1. השדה שמופיע ב Rows בשני התרשימים הוא שדה מחושב.

נכון לא נכון

1. כדי ליצור את התרשימים הנ"ל לא עשינו שימוש ב Label.

נכון לא נכון