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

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

למילוי ע"י הסטודנט

פתרון

* שם המרצה: ד"ר מאיר גולדנברג, ד"ר אריאלה ריכרדסון
* תאריך הבחינה: 06/08/2021
* 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:

ls ../spar\* | grep "some"

סמן עבור כל משפט נכון/לא נכון:

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

נכון לא נכון

* 1. הפלט יהיה ריק, כי המחרוזת some לא כלולה במחרוזת spar.

נכון לא נכון

* 1. שתי הנקודות מתייחסות לתיקיית הורה של תיקיית הבית של המשתמש.

נכון לא נכון

* 1. התפקיד של הקו המאונך הוא להבדיל בין שתי הפקודות.

נכון לא נכון

# Hadoop ו- MapReduce.

1. סמן עבור כל משפט נכון/לא נכון:
   1. כאשר רוצים לפתוח קובץ השמור ב HDFS, מציינים בפקודה באיזה מחשב של האשכול הקובץ נמצא.

נכון לא נכון

* 1. הסיבה שלא משתמשים במערכת קבצים HDFS במחשבים פרטיים היא שקשה להתקין אותה וללמוד להשתמש בה.

נכון לא נכון

* 1. ל- HDFS תפקיד חשוב בהתאוששות מתקלות בתהליך MapReduce.

נכון לא נכון

* 1. "MapReduce" – שם של כלי תוכנה ב- Hadoop.

נכון לא נכון

1. סמן כל משפט נכון/לא נכון:
   1. אם Mapper מגדיר מערך, המערך הזה יהיה מפוזר ביו זיכרונות של מחשבים שונים באשכול.

נכון לא נכון

* 1. נניח שקובץ קלט מחולק ל 100 בלוקים וש HDFS שומרת שלושה עותקים של כל בלוק. אז MapReduce יריץ 300 Mappers.

נכון לא נכון

* 1. נניח קובץ קלט השומר שמות בני אדם:

Adam Grace

Joe Samson

Rich Grey

Joe Pierson

Joe Field

Moe Grace

אם ה- Mapper יפלוט את הקלט שלו כמו שהוא, רק יחליף כל רווח ב- Tab (\t), אז לא ייתכן ששמות משפחה Pierson ו- Field יגיעו ל- Reducer-ים שונים.

נכון לא נכון

* 1. לפי מה שלמדנו, אם רצוננו לדעת כמה בני אדם יש עם כל שם משפחה, אז על ה Mapper לפלוט את שמות המשפחה בלבד.

נכון לא נכון

1. נניח קובץ קלט השומר שמות בני אדם עם תאריך הלידה שלהם:

Adam Grace 22.07.1977

Joe Samson 9.08.1998

Rich Grey 15.01.1969

Joe Pierson 14.02.1988

Joe Field 13.10.1987

Moe Grace 12.11.2003

וכו'

סמן עבור כל משפט נכון/לא נכון:

* 1. ניתן להשתמש בתהליך אחד של MapReduce כדי לחשב, עבור כל שנה שמופיע בקובץ קלט, את מספר אנשים שנולדו בשנה ההיא.

נכון לא נכון

* 1. נניח שברצוננו לחשב, עבור כל שם פרטי המופיע בקובץ קלט, מי הוא האדם הוותיק ביותר שנושא את השם הפרטי הזה. אך אנחנו רוצים לחשב גם, עבור כל שם **משפחה** המופיע בקובץ קלט, מי הוא האדם הוותיק ביותר שנושא את שם המשפחה הזה. עלינו להשתמש בשרשור תהליכי MapReduce.

נכון לא נכון

* 1. אם ה- Mapper יפלוט את הקלט שלו כמו שהוא, רק יחליף כל רווח שאחרי שם פרטי ב- Tab (\t), אז אנחנו משתמשים בשיטת ערך מורכב.

נכון לא נכון

* 1. נניח שקיים קובץ קלט שני, בו עבור כל שנה שמור מספר האנשים שנולדו באותה השנה:

1960 57429

1961 58923

1962 60234

2021 78123

וכו'

נניח שעלינו לחשב, עבור כל שם פרטי וכל שנה, מה הוא אחוז האנשים שניתן להם השם הפרטי הזה בשנה ההיא מתוך כל האנשים שנולדו בשנה ההיא. אז יעזור לנו לבצע צירוף טבעי של שני הקבצים ע"י בחירת שנת לידה כמפתח עבור פלט ה- Mapper של תהליך MapReduce הראשון.

נכון לא נכון

# Spark

1. סמן עבור כל משפט נכון/לא נכון:
   1. Spark משתמש בשיטת חישוב עצלני (lazy evaluation).

נכון לא נכון

* 1. הפונקציה collect אינה גורמת להעברת נתונים דרך הרשת.

נכון לא נכון

* 1. כאשר משתמשים בשיטת Spark, הנתונים נשמרים בדיסק קשיח אחרי כל טרנספורמציה וזה מאפשר התאוששות מנפילות.

נכון לא נכון

* 1. ל- Spark יכולת לקרוא נתונים השמורים ב- HDFS.

נכון לא נכון

1. נתונות שורות קוד הבאות:

Text, letter

Description automatically generated

סמן עבור כל משפט נכון/לא נכון:

1. אחרי ביצוע של שורה 1, הנתונים יהיו מבוזרים.

נכון לא נכון

1. הפלט של שורה 3 הוא 2.

נכון לא נכון

1. הפלט של שורה 5 הוא (1, 2, 3)

נכון לא נכון

1. שורה 8 תגרום לשגיאת ריצה.

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.

סמן עבור כל משפט נכון/לא נכון:

Text, letter

Description automatically generated

1. הביטוי הרגולרי

^Cohen

יחזיר 3 תוצאות.

נכון לא נכון

1. הביטוי הרגולרי הבא:

[A-z]+,[A-z]+

יעזור לנו לזהות את העובדה שחסר מספר בית בשורה 2.

נכון לא נכון

1. עבור הקובץ הנתון, שני הביטויים הרגולריים יחזירו את אותן התשובות.

\d{3}$

\t\d{3}$

נכון לא נכון

1. עבור הקובץ הנתון, שני הביטויים הרגולריים יחזירו את אותן התשובות.

^[A-z]+

^[A-z]\*

נכון לא נכון

# KNIME

1. סמן עבור כל משפט נכון/לא נכון:
   1. כאשר מפעילים קודקוד, כל הקודקודים המצוירים מצד שמאל שלו יופעלו גם כן.

נכון לא נכון

* 1. ניתן לצפות בנתונים בפלט של כל קודקוד.

נכון לא נכון

* 1. בכל פרויקט חייב להופיע בדיוק FileReader אחד.

נכון לא נכון

* 1. הנתונים שמנקים אותם בעזרת Knime צריכים לעבור ניקוי בעזרת ביטויים רגולריים לפני שיהיה ניתן לטעון אותם לקודקוד.

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)

Chart

Description automatically generated with medium confidence

Graphical user interface, application, email

Description automatically generated

1. לאחר הפעלת קודקוד 3 יהיו בטבלה 3 עמודות.

נכון לא נכון

1. יתכן שהשורות הראשונות בטבלה לאחר הפעלת קודקוד 3 הם:

Table

Description automatically generated

נכון לא נכון

1. אם היינו רוצים לסנן עמודות ולאחר מכן לסנן שורות היה אפשר לחבר את היציאה של קודקוד 3 לכניסה של קודקוד 2.

נכון לא נכון

1. אין דרך לדעת מה יהיה מוצג אחרי הפעלת קודקוד 3, כי ההגדרות של קודקוד 2 לא מופיעות.

נכון לא נכון

# TABLEAU

1. בתמונה רואים תרשים שנתקבל מתוך Tableau. על סמך קובץ הנטישה (Churning) מהבנק שראינו בהרצאה.

Graphical user interface, text

Description automatically generated with medium confidence

סמן עבור כל משפט נכון/לא נכון:

1. אם נוריד מMarks (מסומן בעיגול) את Exited נקבל תרשים שכולל רק את החלק הכהה של העמודות.

נכון לא נכון

1. ייתכן שהToolTip (מסומן בחץ) הופיע כיוון שריחפנו עם העכבר על העמודה הימנית בתרשים.

נכון לא נכון

1. ייתכן שהשלב הראשון ביצירת התרשים היה על ידי גרירת Gender (מגדר) לעמודות (Columns) וגרירת שדה מחושב לשורות.

נכון לא נכון

1. ניתן לשכפל את התרשים ולייצר ממנו תרשים דומה עבור מדינות ע"י החלפת השדה Gender ב Geography בעמודות (Columns).

נכון לא נכון