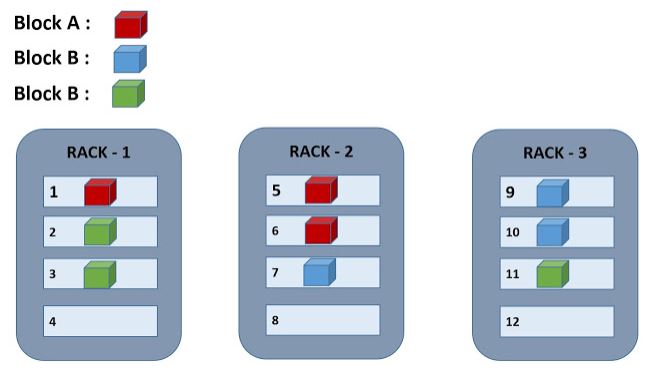
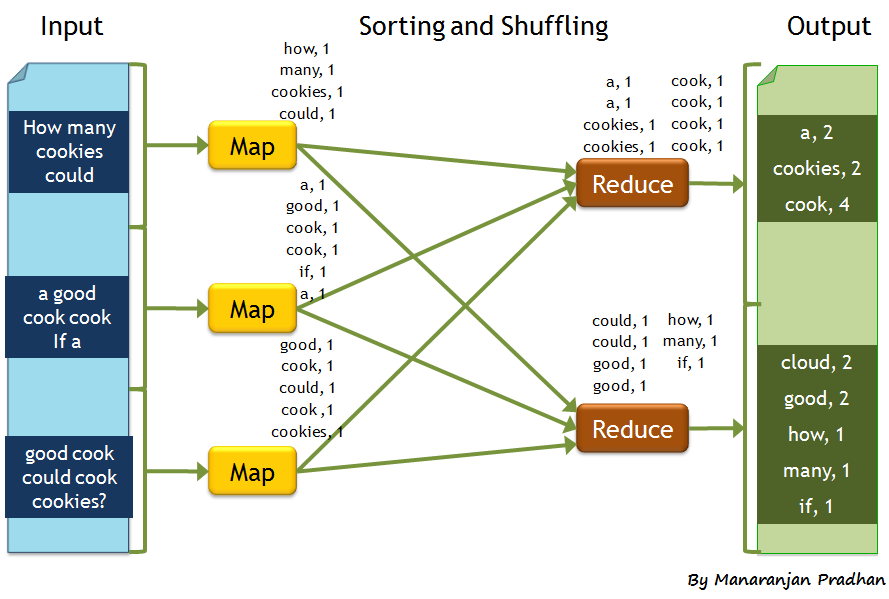
1. אני מתחייב לקבל אחריות שלא להיעזר על ידי אחרים ולא לעזור לאחרים

**Hadoop and MapReduce Overview**

1. נניח שהאיור המופיע למטה (נלקח מהמצגת) מתאר שמירה של קובץ שלם בגודל 192 מגה ב- HDFS. איזה מהמשפטים הבאים נכון?



1. גודל הבלוק הוא 128 מגה.
   1. Default in hadoop 2.0, but not necessarily the configuration
2. הקובץ מורכב מ- תשע בלוקים שונים.-- 3, as stated in the top
3. הקוביות באותו צבע (כגון אדום) מתייחסים לאותם הנתונים.
   1. Correct, these are replications for data resiliency
4. הקובץ שמור על שלושה דיסקים קשיחים בלבד.(depends on config)
5. מחשבים עם מספרים 4, 8 ו- 12 אינם חלק מהאשכול.
6. הקונפיגורציה באיור זה מונעת לחלוטין אפשרות של איבוד נתונים.
7. איזה מהמשפטים הבאים נכון על תהליך MapReduce באיור הבא הנלקח מהמצגת?



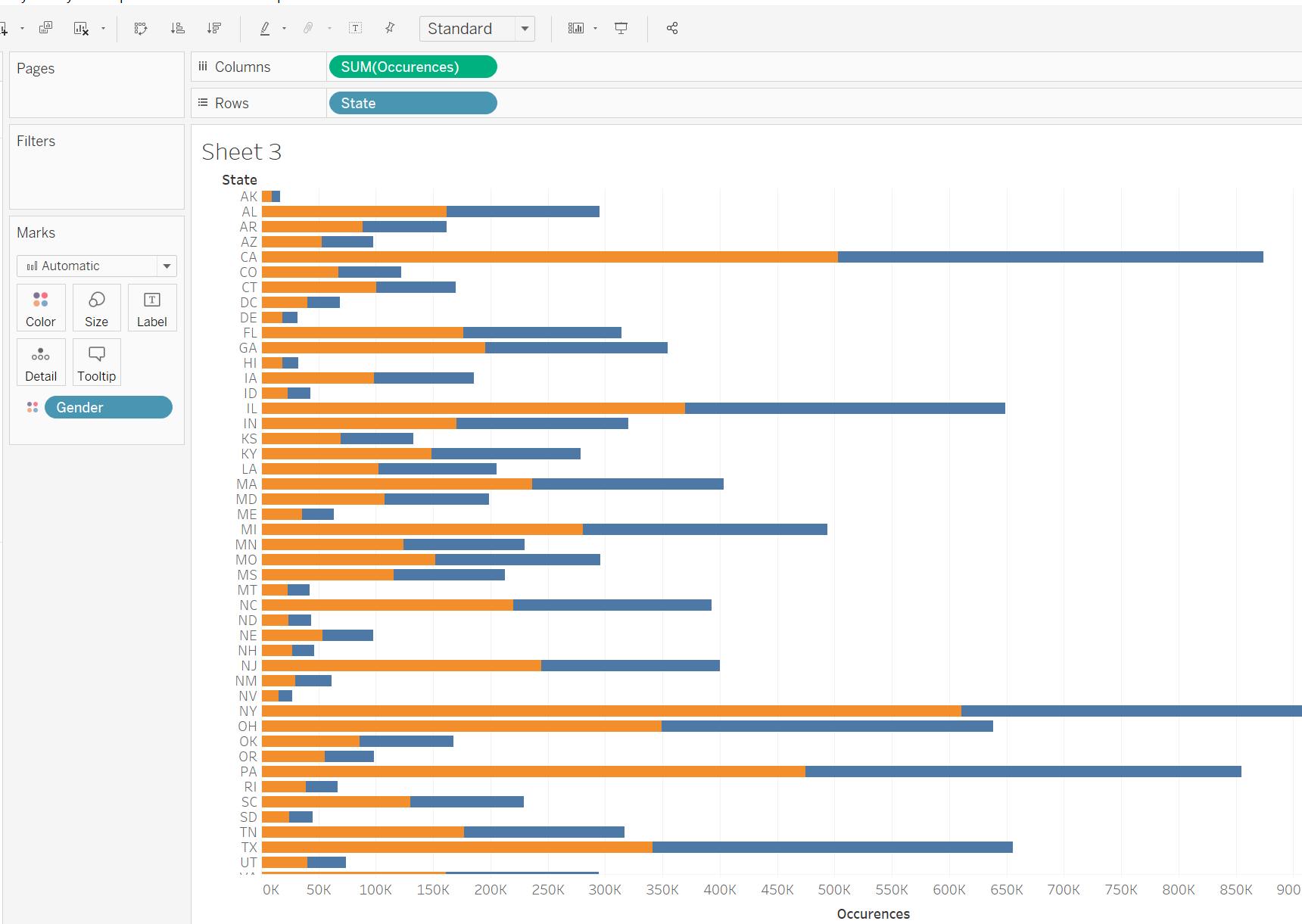
1. הפלט יופיע בשני קבצים שונים.
2. איור זה יכול להיות נכון רק אם יש נתונים בקלט שאינם מופיעים באיור.
3. העובדה ש- Reducer אחד קיבל שלוש מפתחות ואילו Reducer אחר קיבל חמש מפתחות מעידה על כתיבה לא נכונה של Mapper.
4. איור זה אינו נכון, כי cook אמור להופיע לפני a ולפני cookies בפלט, כי המילה cook תדירה יותר בקלט.
5. הוספה של 1 אחרי כל מילה ע"י Mapper היא רק כדי לקיים דרישה שפלט של Mapper צריך להיות מילון. אין ל- אחדים האלה תפקיד אחר בחישוב.
6. עבור הקלט הספציפי באיור, אין צורך בשלב של Sorting & Shuffling.

**RegEx**

1. נניח שרוצים למצוא בתוך קובץ את כל המשפטים שמתחילים ב The ומסתיימים ב Apple?
2. ^The.\*Apple$
3. ^The.\*Apple$\*
4. The.\*Apple
5. T\*e$
6. ^The\s{n}Apple$
7. ^The\s\*Apple
8. נניח שרוצים למצוא בתוך קובץ את כל השורות שמתחילות ב-A אחת לפחות, באמצע יש רק ספרות ולא כל הספרות אלא 0,1,2 בלבד ומסתיימות באות B. דוגמה של מחרוזת: AAA22200B. איזו מהביטויים הרגולריים עשוי לבצע זאת?
9. ^A+[\d]+[B]$
10. ^A+[\d]+B$
11. $A+[0-2]+B^
12. ^A+[012]+B$
13. ^A+[\0\1\2]+B$
14. ^A\*[012]\*[B]$
15. ^\a[012]+[^b]$

**TABLEAU**





בתמונה רואים תרשים שנתקבל מתוך Tableau. על סמך קובץ השמות הנפוצים מתרגיל הבית. איזו מהטענות הבאות נכונה?

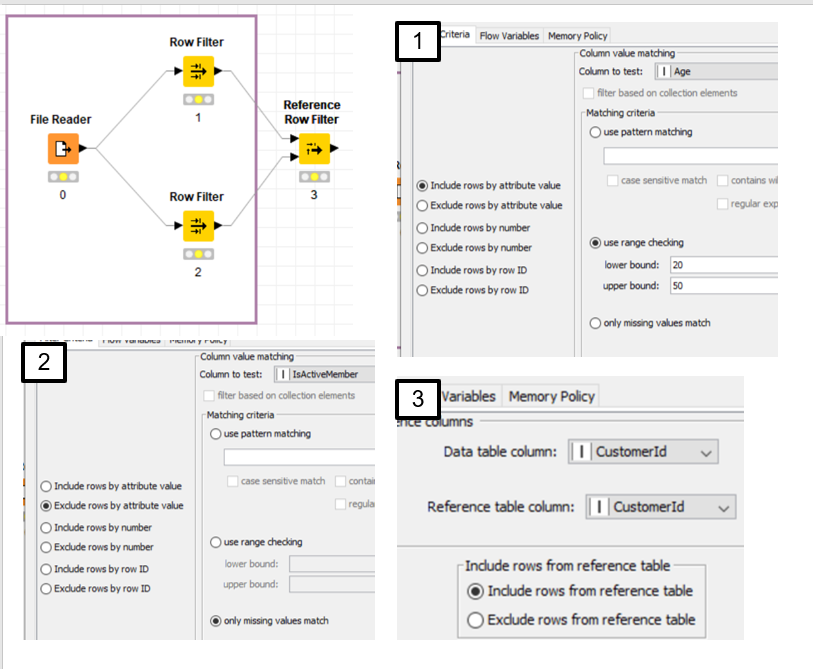
(Occurrences – מופעים, Gender – מגדר, State – מדינה)

1. מטרת הצביעה בתרשים היא כדי להבדיל בין שמות נפוצים לאלו שפחות נפוצים.
2. ניתן לראות בתרשים את מספר המופעים של השמות הנפוצים שיש בכל מדינה, הצבע מציין את רמת הנפוצות של השם.
3. התרשים הזה שגוי כי בניגוד לתרשימים שראינו יש כאן היפוך של כיוון העמודות.
4. הצביעה מדגימה את החלוקה של האוכלוסייה בין המגדרים השונים בכל מדינה.
5. לו רצינו לראות את כל המדינות יחד היינו גוררים את הממד State ל Detail.
6. על מנת לשנות את הצביעה של התרשים, חייבים לשנות את התוכן של Columns ו Rows שמעל התרשים.

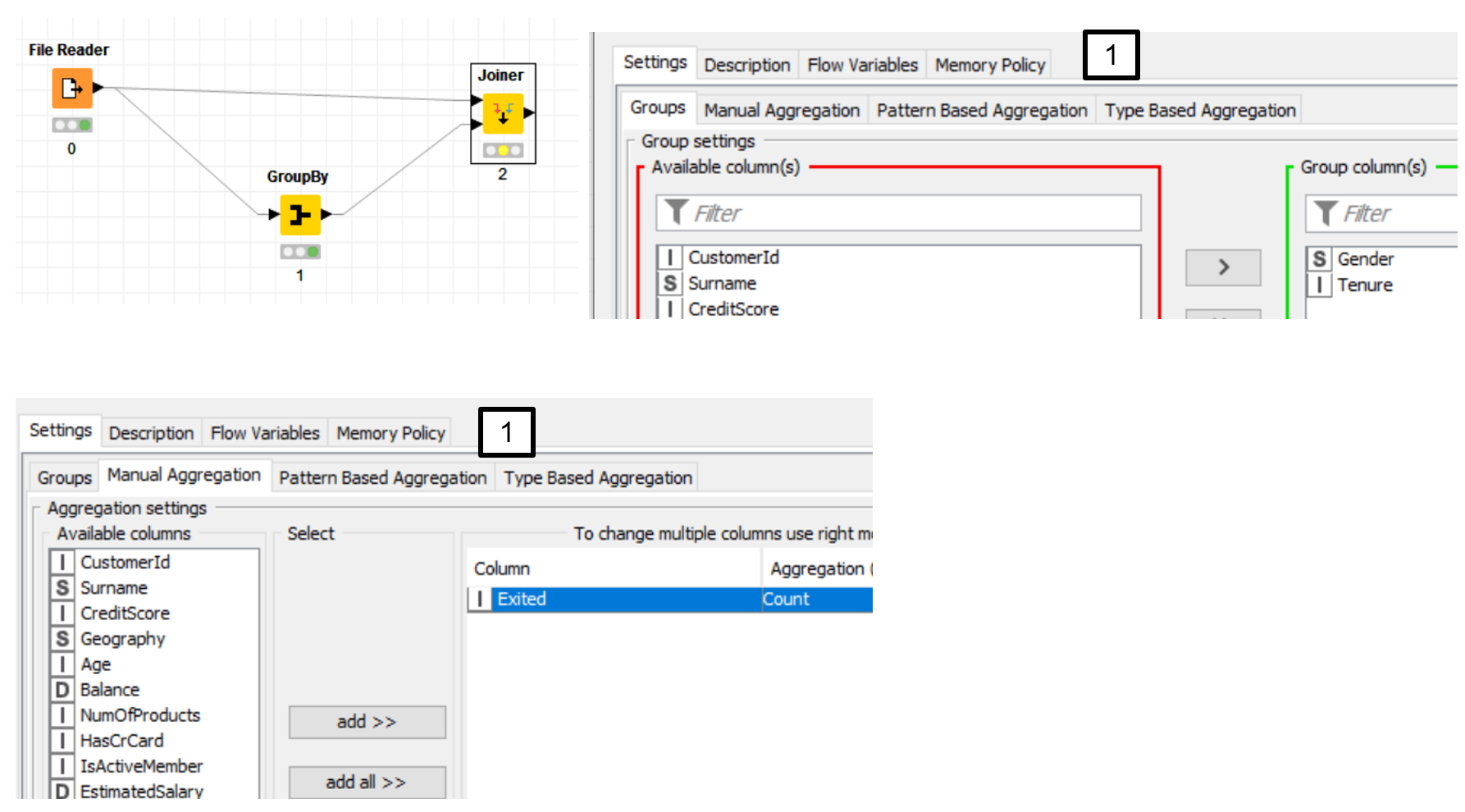
**KNIME**

1. בעבודה ב Knime איזו מהטענות הבאות נכונה?
2. קישור בין הקודקודים נועד כדי לתאר את סדר החשיבות שלהם.
3. כאשר מריצים (מבצעים, מפעילים) קודקוד ירוצו (יתבצעו, יופעלו) באופן אוטומטי ולפי הצורך כל הקודקודים שהוא תלוי בפלט שלהם.
4. כל הקודקודים בפרויקט אחד חייבים להיות בגיליון אחד.
5. כל הקודקודים הם מאותו הסוג ומשתמשים ב- configure כדי להגדיר עבורם פעולות שונות.
6. ניתן לראות את הנתונים הקיימים כרגע בקודקוד רק לפני ההפעלה שלו.
7. ביטויים רגולריים יכולים לעזור ב Knime אבל הם דורשים ייצור של קודקוד מיוחד בשם "RegEx".
8. על פי המוצג בתרשים Knime הבא עבור אנליזה של נתונים על לקוחות בנק, איזו מהטענות הבאות נכונה?

שימו לב שכל חלון של מאפיינים מופיע עם מספר של הקודקוד המתאים לו



1. לאחר הפעלת קודקוד 3 נקבל שורות עבור לקוחות בעלי גיל 20 או 50 ושהשדה IsActiveMember מראה שהם פעילים.
2. לאחר הפעלת קודקוד 3 נקבל שורות עבור לקוחות בעלי גיל 20 או 50 ושהשדה IsActiveMember מראה שהם לא פעילים.
3. לאחר הפעלת קודקוד 2 נקבל עמודות עבור לקוחות שיש להם ערך בשדה CustomerId והגיל הוא 20 עד 50.
4. לאחר הפעלת קודקוד 3 נקבל שורות עבור לקוחות בעלי גיל 20 עד 50 ושיש להם ערך בשדה IsActiveMember.
5. לא ניתן לדעת מה צפוי להתקבל כי כל הנורות צהובות וזה מעיד על תקלה בהגדרות.
6. יש כאן טעות בתרשים כי הריבוע הסגול אינו מקיף את כל הקודקודים ולכן לא נוכל להפעיל את הקודקודים.
7. על פי המוצג בתרשים Knime הבא איזו מהטענות הבאות נכונה? שימו לב ששני החלונות של מאפיינים מתייחסים לקודקוד עם מספר 1.



1. התרשים של ה KNIME לא יכול להיות נכון כיוון ש- קודקוד עם 2 כניסות אינו חוקי, זו הסיבה שהקודקוד מופיע עם אור צהוב.
2. בהנחה שהפעלנו את קודקוד 1 עם הפרמטרים בחלונות הקונפיגורציה שמופיים כאן, נוכל לעשות Join בקודקוד 2 עם תנאי צירוף טבעי לפי עמודות של Gender ו Tenure.
3. בהנחה שהפעלנו את קודקוד 1 עם הפרמטרים בחלונות הקונפיגורציה שמופיים כאן, נוכל לעשות Join בקודקוד 2 לפי עמודות של Gender ו Exited.
4. בהנחה שהפעלנו את קודקוד 1 עם הפרמטרים בחלונות הקונפיגורציה שמופיעים כאן, נקבל טבלה שיש לה עמודה אחת של Exited.
5. יש שגיאה בתרשים כי קודקוד 2 חייב לקבל להיות אחרי שני קודקודים מאותו הסוג.
6. בהנחה שהפעלנו את קודקוד 1 עם הפרמטרים בחלונות הקונפיגורציה שמופיעים כאן, נוכל לעשות Join בקודקוד 2 עם תנאי צירוף טבעי לפי סטטוס העזיבה (העמודה של Exited). Join זה יהיה בעל משמעות פרקטית.

**מבין 4 השאלות הבאות יש לבחור 2**

**Advanced MapReduce and Spark**

1. נגיד שיצרנו RDD של Spark בעזרת המשפט הבא ב- PySpark:

>>> rdd = sc.parallelize([('A', 10), ('B', 20), ('C', 30), ('D', 40)])

כל אופציה למטה מציגה משפט ופלט. באיזו מהאופציות הפלט יכול להיות מתאים למשפט? (תזכורת: פונקציית len מחזירה אורך הרשימה שהיא מקבלת בתור ארגומנט)

>>> len(rdd.glom().collect())

8

>>> len(rdd.map(lambda el: el[1]).collect())

8

>>> rdd.map(lambda el: el[0] if el[1]>20 else '').collect()

['C', 'D']

>>> rdd.filter(lambda el: el[1] > 20).collect()

[30, 40]

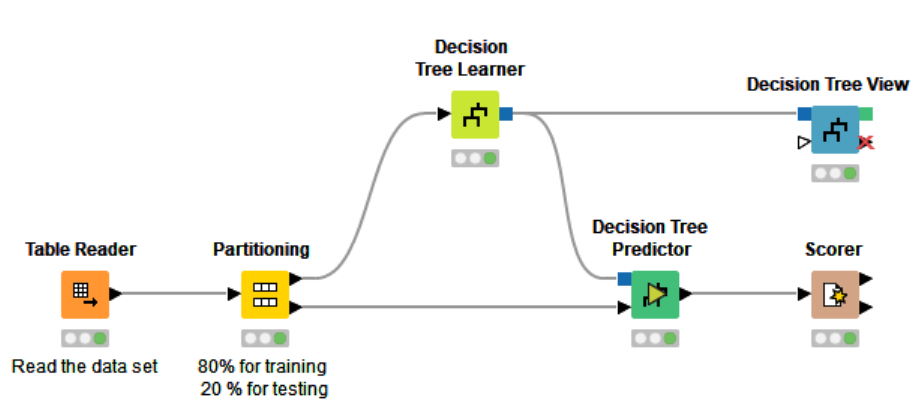
>>> len(rdd.groupByKey().collect())

1

>>> len(rdd.flatMap(lambda el: el).collect())

8

1. בעבודה ב Tableau איזו מהטענות הבאות נכונה?
2. סוג של Dimension (ממד) קובע באופן ייחודי את סוג התרשים (תרשים עוגה – pie chart, תרשים עמודה – bar chart, מפה וכדו').
3. לא ניתן להעביר מאפיינים מDimensions (ממדים) ל Measures (מדדים) ולהפך.
4. הנתונים ש Tableau יודע לעבוד איתם חייבים להגיע מבסיס נתונים יחסי.
5. מבחן AB יכול לבדוק איך ממדים מסוג מחזורת משפיעים על התרשים.
6. שימוש ב Alias יכול להיות נוח על מנת לתת לערכים מספריים שם בעל משמעות.
7. כאשר רושמים ב Colums שני ממדים נקבל שני תרשימים נפרדים.
8. ב Knime איזו מהטענות הבאות נכונה בהתייחס לאיור הנתון?



1. Knime הוא כלי שמיועד לעבודה על מבנה נתונים. התרשים הזה אמנם דומה ל Knime אבל לקוח מכלי אחר.
2. בתרשים רואים תהליך של כריית מידע, Knime אינו יכול לבצע את הכרייה עצמה, אבל התרשים משמש לאפיון התהליך.
3. בתרשים רואים תהליך של כריית מידע, Knime מבצע את החלקים של הקריאה וחלוקת הנתונים אבל לצורך ההמשך צריך חבילות מיוחדות ויקרות של Knime שלא יכלנו לנסות אותם בקורס.
4. Knime מאפשר ביצוע של כריית מידע כמתואר בציור ואפילו ניתן לצפות בעץ ההחלטה שהתקבל.
5. התרשים שמופיע כאן הוא רק לצורך הדגמה של פיתוח עתידי בו WEKA ו Knime יחוברו.
6. אם היו הנחות שביצעת עקב אי בהירות, רשום באיזו שאלה הייתה הבעיה ומה הנחת.