**תרגיל חפיפות – מסלול**

בשעה טובה אחרי שתירגלת design הגיע הזמן לעשות קצת data engineering

התרגיל:

עליך לבנות מסלול שקורא מkafka הודעות מסוג Person ומכניס אותן לאורקל ולs3

איך בדיוק?  
- תצטרך לקרוא מkafka הודעות בפורמט person

* תצטרך לכתוב גם מיני סקריפט מחוץ לrepo של הפרויקט שמגנרט הודעות מסוג person ושולח אותן לקפקא שלך (המודל של person בעמוד האחרון)

-לכל הודעה של person יש שדה בשם status – insert או update. אם תקבל person מסוג insert תצטרך לבצע עליו לוגיקת הכנסה לdbs. אם תקבל person מסוג update תצטרך לבצע עליו לוגיקת עידכון על הdbs.

- לוגיקת הכנסה:  
 בעת קבלת person, כל השדות שלו חוץ מaddress צריכים להישמר באורקל בטבלה של persons  
המידע על הכתובת שלו ישורשר לסטרינג אחד עם רווחים, וישמר בs3. בs3 שלך אתה תיצור 10 באקטים בשם person-0…person-9, עבור כל איש, תעשה hash על הid שלו ותחליט לאיזה באקט אתה מכניס אותו ואז תשמור את הקובץ של הכתובת שלו כך שהשם של הקובץ זה הid של האיש

-לוגיקת עדכון:  
אם אתה מקבל status = update תצטרך לבדוק אם אחד השדות הוא לא נאל לפי הסדר הבא:  
1. אם הlastName הוא לא נאל – אז נבצע עידכון לlastname של האיש וזהו  
2. אחרת, אם הphone הוא לא נאל – אז נבצע עידכון לטלפון של האיש וזהו  
3. אחרת, אם הemail הוא לא נאל וגם לאיש קוראים אלכס – אז נבצע עידכון לאימייל וזהו  
שאר השדות לא מתעדכנים (למשל gender)

למה חשוב לשים לב?  
- אתה תקרא מקפקא בבאלקים של 500 הודעות של אנשים ותצטרך להכניס אותם מהר ככל האפשר למאגרי המידע שלך. אם יש מאגר שאפשר לכתוב אליו בבאלקים – אתה צריך לכתוב אליו בבאלקים ולא אחד אחד. המהירות כאן חשובה והודעות לא יכולות להתעכב  
  
- אתה צריך להבין מה אתה עושה כשיש כישלון? אם נניח הצלחת לכתוב לאורקל אבל לא לs3 מה אתה עושה?

-הקוד שלך צריך להיות גנרי ככל האפשר. נניח אל תכתוב פונקציית הכנסה לאורקל שמכניסה רק אובייקטים מסוג person, אם נרצה בהמשך להשתמש בה גם כדי להכניס מידע נוסף – היא צריכה לאפשר את זה

המודל של person: (תעשה שכל השדות הם nullable, שכחתי לבקש מהצאט)

