

חלק א' – תשובות :

סעיף ב':

3. אם הנתונים מגיעים ע"י זרימה

אני אצור defaultdict ש key זה השעה ותאריך והערך הוא הסכום שנוצר עד עכשיו ומספר התאריכים שנוספו

כל תאריך חדש שנוסף אני יוסיף למילון את הערך למה שקיים ויקדם את ה count וכאשר ירצו לשלוח את הנתונים מהמילון אני יחלק את הסכום למספר העברים שנוצרו וכך יהיה לי ממוצע בזמן אמת

בסוף הקובץ sectionB יש בהערה את המימוש לזה ...

4. היתרונות של parquet:

- דוחס אוטומטית את הנתונים לקובץ קטן יותר לאומת csv ואחרים ...

- מהיר יותר בקריאה וכתיבה אליו – הקריאה מתבעת לפי טורים ולא לפי שורות -שאם רוצים רק מידע על עמודה אחת הוא לא טוען לזיכרון את כל הקובץ רק את העמודה

- שומר גם את המידע על הטיפוסים שבו (כמו תאריך , מספר , וכו..) וכך לא צריך להמיר כמו ב csv ששומר כמחרוזת ואת צריך להמיר לטיפוס המתאים ..