מגישה: שפרה שיינר

תשובה לשאלה 3

אם הנתונים מגיעים בזרימה במקום מקובץ, אז צריך לשמור את הנתונים המגיעים במקום זמני, וכל כמה זמן כאשר יש מאסה גדולה מספיק (לדוגמא במקרה שלנו כל יום, כי חילקנו את הנתונים לימים) מריצים את הפונקציה שמחשבת את הממוצעים השעתיים, ומוסיפים את השורות החדשות לקובץ הסופי.

תשובה לשאלה 4

היתרונות של שימוש בparquet הם שהקובץ הוא יותר מכווץ ומהיר ולכן הוא יותר מתאים לביג דאטה. שמתי לב במהלך העבודה שהוא רץ מהר יותר מקובץ הexcel. מצד שני, הוא לא נוח לשימוש כי הוא לא קריא לבני אדם. אי אפשר לערוך אותו ישירות או אפילו לקרוא אותו בלי לעבד באמצעות הקוד.