

## פתרון לשאלה – כיצד לעדכן ממוצעים שעתיים כאשר הנתונים מגיעים בזרימה (stream):

כאשר הנתונים אינם מגיעים מקובץ קבוע אלא בזרימה מתמשכת (stream) נבנה מנגנון שמעדכן בזמן אמת את הממוצע השעתי של ערכים נכנסים. כדי לעשות זאת ביעילות, נשתמש במבנה נתונים מסוג מילון (dictionary) שמאגד את הנתונים לפי שעה עגולה.

שלב הפתרון:

### 1. הגדרת המילון:

ניצור מילון בשם hourly\_data, שבו:  
המפתח (key) הוא הזמן של השעה העגולה (תאריך + שעה ללא דקות ושניות)  
הערך (value) הוא אובייקט או מבנה המכיל 3 שדות:  
Cnt – מספר הערכים שהתקבלו בשעה זו.  
Sum – סכום כל הערכים שהתקבלו בשעה זו.  
Avg – ממוצע הערכים שהתקבלו עד כה בשעה זו.

### 2. עיבוד של כל זוג נתון חדש:

בעת הגעת זוג חדש (timestamp, value), יש לחלץ מתוך ה-timestamp את השעה העגולה.  
נבדוק האם השעה כבר קיימת במילון:  
אם כן – נגדיל את cnt באחד, נוסיף את value ל sum, נחשב מחדש את avg לפי  $sum/cnt$ .  
אם לא – ניצור רשומה חדשה במילון כאשר  $cnt=1$ ,  $sum = value$ ,  $avg = value$ .

### 3. התוצאה:

בסיום העיבוד, המילון hourly\_data יכיל לכל שעה עגולה את הממוצע הנכון של כל הערכים שהתקבלו באותה שעה.

יעילות:

זמן חישוב קבוע ( $O(1)$ ) לכל זוג חדש.

חסכוני בזיכרון – שומר רק את הנתונים המצטברים, לא את כל הערכים.

מתאים לעבודה בזמן אמת – כל עדכון מהיר ופשוט.

אם הנתונים מגיעים במקביל ממספר מקורות (multi-process / multi-threaded) יש להוסיף **מנעול (lock)** סביב העדכון למילון כדי למנוע גישה מקבילה שעלולה לגרום לשגיאות או תוצאות לא עקביות.