חלק א'

סעיף ב'

שאלה 3

כשעובדים עם קובץ, ניתן לקרוא את כל המידע מראש ולחשב את הממוצע לכל שעה באופן מרוכז לאחר קריאת כל הנתונים.   
לעומת זאת, בעבודה עם זרימה (stream) , הנתונים מגיעים באופן רציף ומיידי, ולכן עלינו לעדכן את הממוצע בזמן אמת עבור כל שעה. בנוסף, חשוב שהפתרון יתמודד עם קצב זרימה משתנה, ע"מ לשמור על דיוק העיבוד, ויש למנוע בעיות סינכרוניזציה ע"י הפעלת מערכת אסינכרונית כאשר נתונים מעובדים בו זמנית על ידי מספר מקורות.

שאלה 4

פורמט PARQUET הוא פורמט טורי, דוחס ושומר סכמות, שמתאים במיוחד לעבודה עם כמויות גדולות של נתונים.  
היתרון בו שהוא מאפשר גישה מהירה לעמודות ספציפיות, חיסכון משמעותי בגודל הקובץ, ושיפור ביצועים בעיבוד הנתונים.