**תיעוד: יתרונות פורמט Parquet**

**Parquet** הוא פורמט אחסון טורי (columnar) אשר פותח על ידי Apache. להלן היתרונות המרכזיים שלו:

1. **דחיסה יעילה** – Parquet מבצע דחיסה טורית, כך שהוא צורך פחות מקום אחסון ביחס ל-CSV.
2. **ביצועים גבוהים** – קריאה/כתיבה מהירה במיוחד בטבלאות גדולות (אפשר לקרוא רק טורים מסוימים).
3. **תמיכה בטיפוסים מורכבים** – כמו רשימות, מילונים וכו', מה שאין ב-CSV.
4. **תמיכה באחזור חלקי** – מאפשר קריאה של קטעים מסוימים של הקובץ בלי לטעון את כולו לזיכרון.
5. **אינטגרציה עם מערכות ביג דאטה** – מתאים לעבודה עם Spark, Hive, AWS Athena ועוד.

**לעומת זאת:**

* CSV הוא פורמט פשוט, קריא אנושית, ונפוץ יותר – אך הוא איטי יותר, צורך יותר מקום, ואינו תומך בטיפוסים מורכבים.