**PYSPARK DATAFRAMES**

**Q1.** Load the above CSV files (transactions.csv and accounts.csv) into PySpark DataFrames with header support.

**Q2.** Find the total number of transactions and distinct customers.

**Q3.** Calculate the **total credit and debit amount** separately.

**Q4.** Find the **top 5 accounts by highest debit transactions**.

**Q5.** Join the two DataFrames and create a report showing customer\_name, branch, transaction\_date, transaction\_type, amount.

**Q6.** Find customers who have done **transactions greater than 50,000 INR**.

**Q7.** For each branch, calculate the **net transaction amount (credits - debits)**.

**Q8.** Identify the **top 3 merchants** by total transaction volume.

**Q9.** Detect **suspicious accounts**: accounts with more than 3 transactions above 1,00,000 INR.

**Q10.** Calculate the **average transaction amount per account\_type** (savings vs current).

**Complex Queries**

1. From the transactions DataFrame, calculate each account’s **running balance** over time (ordered by transaction\_date).
2. Identify the **top 2 highest-value transactions per branch**, along with customer\_name, branch, and merchant.
3. Find all accounts where the **total debit amount in a month exceeds the account balance** recorded in accounts.
4. Create a **pivot table** showing, for each branch, the **total credit and total debit amounts**, side by side.
5. Detect **“bursty” accounts**: accounts with **3 or more transactions within a 5-minute window**.
6. For each merchant, calculate the **average transaction amount and standard deviation**, and flag transactions that are **3x above the standard deviation** as “anomalous”.
7. Generate a report of **inactive accounts**: customers who did not perform any transaction in the last 90 days.
8. Identify the **branch with the highest net outflow** (debits > credits), and show the top 5 customers contributing most to that outflow.
9. For each account, find the **largest gap (in days)** between two consecutive transactions.
10. Calculate the **year-over-year growth rate in total credit transactions** for each branch.