Big data: architectures and data analytics

Spark - Exercises

2

Exercise #51

- High stock price variation identification in real-time
- Input:
  - A streaming of stock prices
    - Each input record has the format
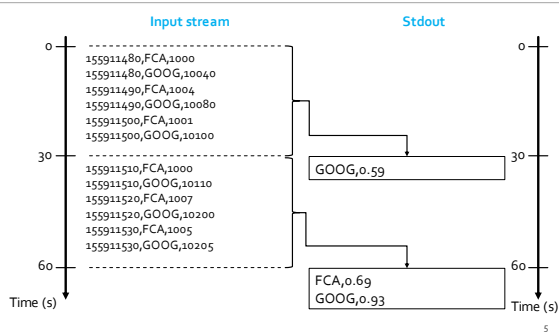   - Timestamp,StockID,Price

3

Exercise #51

- Output:
  - Every 30 seconds print on the standard output, and store in the output folder, the **StockID and the price variation (%) in the last 30 seconds** of the stocks with a **price variation greater than 0.5%** in the last 30 seconds
    - Given a stock, its prices variation during the last 30 seconds is:

$$\frac{max(price)-min(price)}{max(price)}$$

4

Exercise #51- Example

Input stream	Stdout
0	0
155911480,FCA,1000	
155911480,GOOG,10040	
155911490,FCA,1004	
155911490,GOOG,10080	
155911500,FCA,1001	
155911500,GOOG,10100	
30	30
155911510,FCA,1000	GOOG,0.59
155911510,GOOG,10110	
155911520,FCA,1007	
155911520,GOOG,10200	
155911530,FCA,1005	
155911530,GOOG,10205	
60	60
Time (s)	FCA,0.69
	GOOG,0.93
	Time (s)

5

Exercise #52

- Anomalous stock price identification in real-time
- Input:
  - A textual file containing the historical information about stock prices in the last year
    - Each input record has the format
      - Timestamp,StockID,Price
  - A real time streaming of stock prices
    - Each input record has the format
      - Timestamp,StockID,Price

6

Exercise #52

- Output:
  - Every 1 minute print on the standard output, and store in the output folder, the StockIDs of the stocks that satisfy one of the following conditions
    - price of the stock (received on the real-time input data stream) < historical minimum price of that stock (based only on the historical file)
    - price of the stock (received on the real-time input data stream) > historical maximum price of that stock (based only on the historical file)
  - If a stock satisfies the conditions multiple times in the same batch, return the stockId only one time for each batch
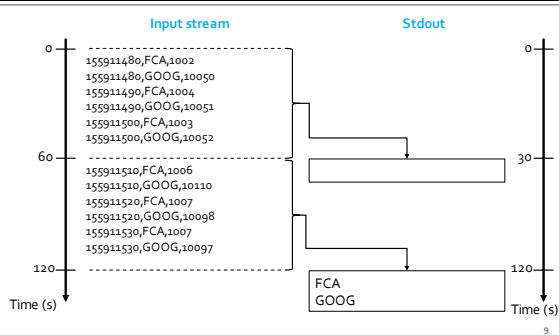
7

Exercise #52- Example

- Textual file containing the historical information about stock prices in the last year
  - 130000000,FCA,1000
  - 130000000,GOOG,10040
  - 130000060,FCA,1004
  - 130000060,GOOG,10080
  - 130000120,FCA,1001
  - 130000120,GOOG,10100

8

Exercise #52- Example

9