Getting Familiar with IBM Streams, Stream Studio and Streams Processing Language (SPL)

Chan-Ching Hsu

Purpose

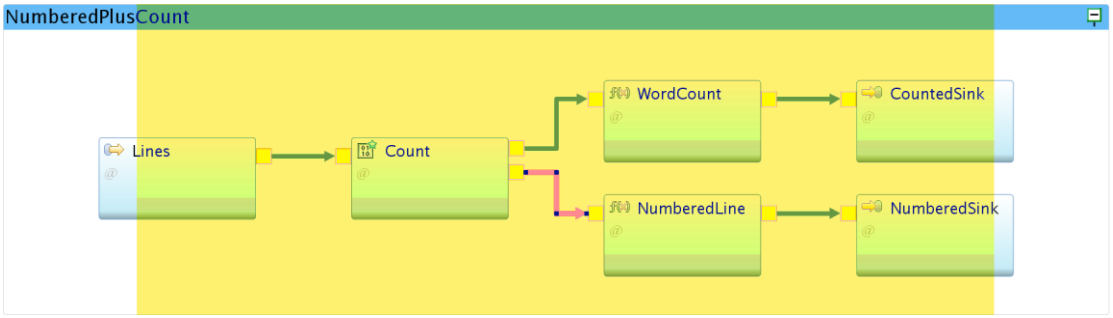
Data streams are continuously flowing data, e.g., sensor data or network packets. IBM Streams can be used for real time analysis of streams, using Streams Processing Language, e.g., analysis of streams when they arrive, without storing them in the system. More information can be found in the IBM Streams official document at:

https://www.ibm.com/support/knowledgecenter/SSCRJU_4.2.0/com.ibm.streams.welcome.doc/doc/kc-homepage.html

Task 1

Write an application reads a text file and outputs the same text file with each line prefixed with the corresponding line number. The input is the text file "big.txt". The output is directed to a file "bigresults.txt". Also output the number of occurrence of the words "history" and "adventure" in a file called "count.txt".

A screen snapshot of the graph of the operators is as follows. Please see the code for how the program works.



Task 2

An online shopping website gives a 1% discount to customers who make a minimum purchase of $250, 2% discount for a minimum purchase of $500 and 5% discount for a minimum purchase of $1000.

The input data stream has tuples of the form:

<Transaction ID, First Name, Last Name, Purchase Amount ($)>

Process the input stream to generate an output of the following form:
<Transaction ID, Customer Name, Purchase Amount ($), Discount ($)>, where "Discount" is the amount in dollars which is computed using the above given logic.

The data is saved in a file "shopping.csv". The output is directed to a file named "CustomerStatus.txt".

A screen snapshot of the graph of the operators is as follows. Please see the code for how the program works.