

S

1. Tuplelization: read from a file line by line and then split each into words
2. Calculate the reconfiguration time (from graph 3-6 blow), the time between stream stop and stream resume.

Sink

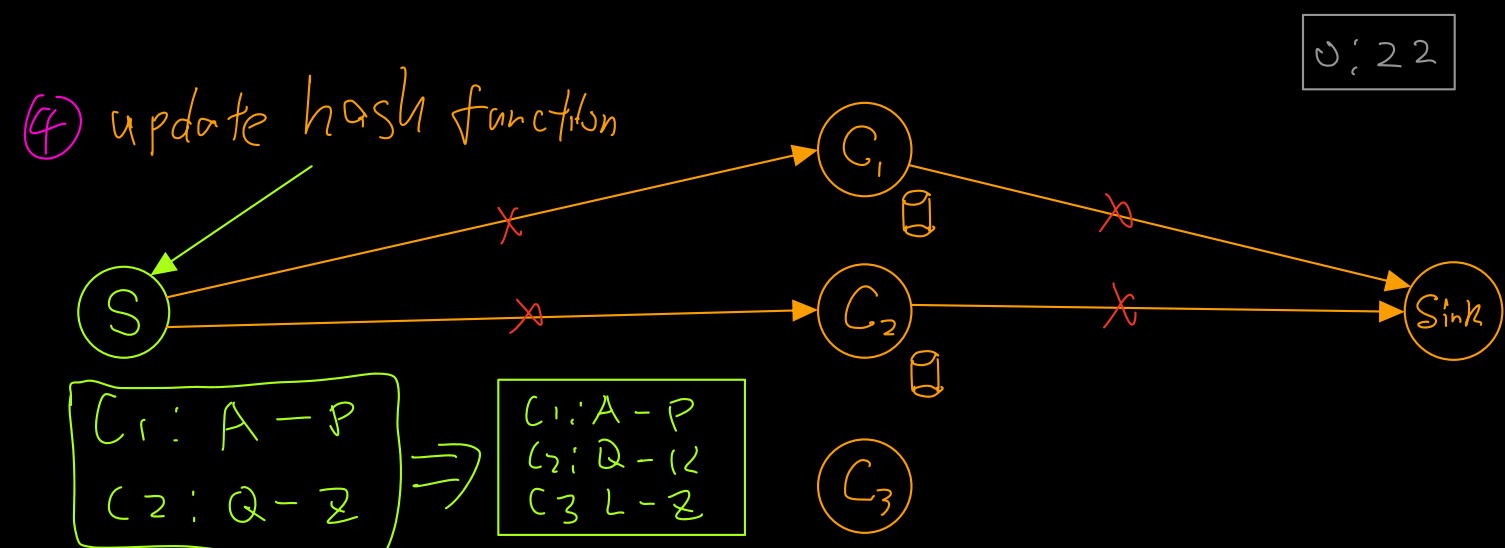
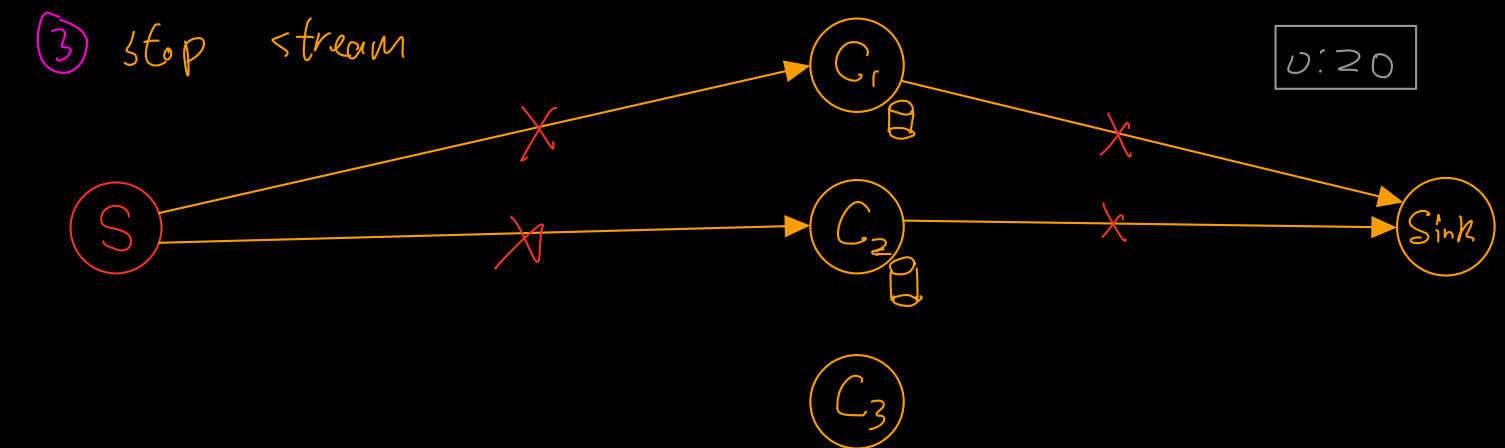
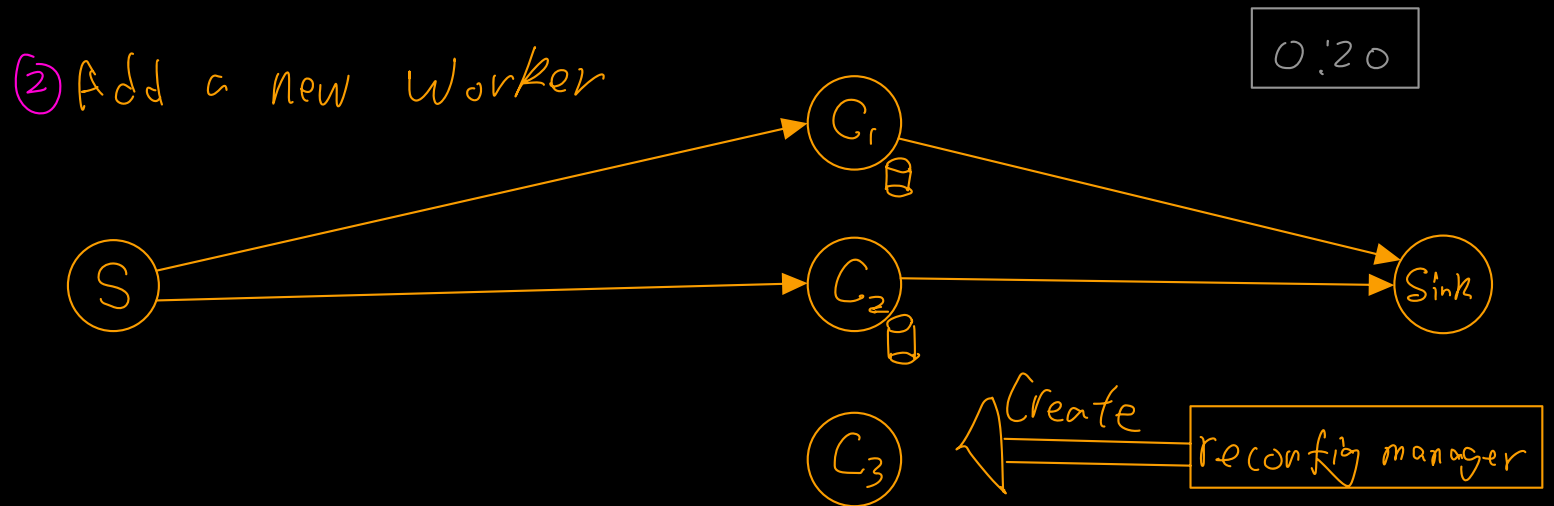
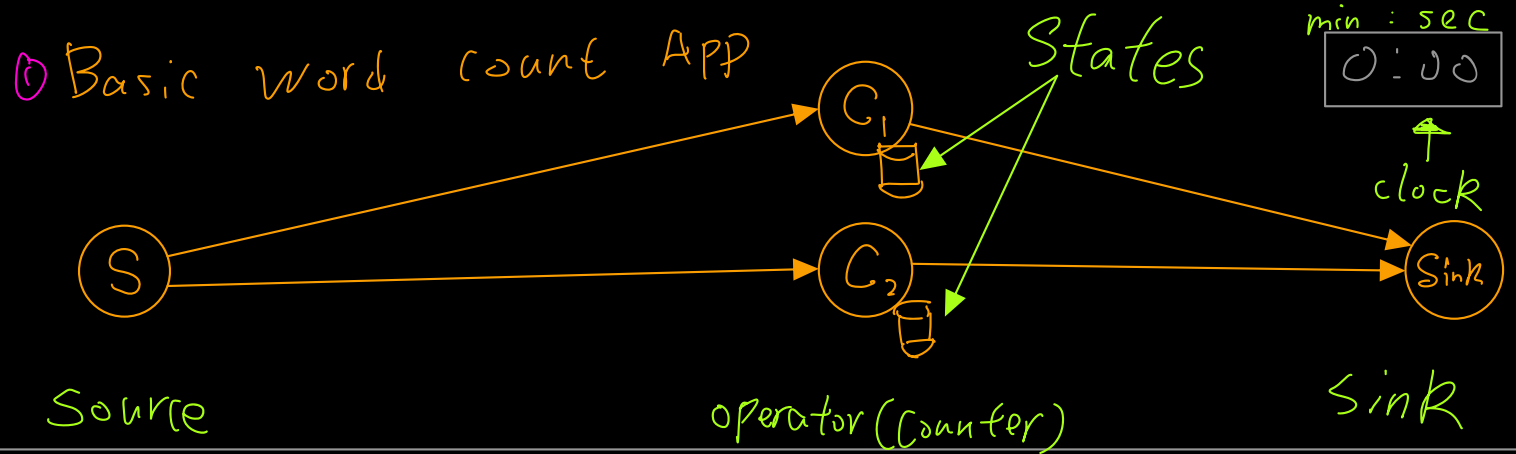
1. Implement a time based tumbling window to calculate the through put, and dump it into a log file and using another program to plot the through put change, or port them into a realtime visualization app with throughput stats plotted lively.

G

1. Update the word counter (value of the key-value pair) and pass it to the downflow(sink)

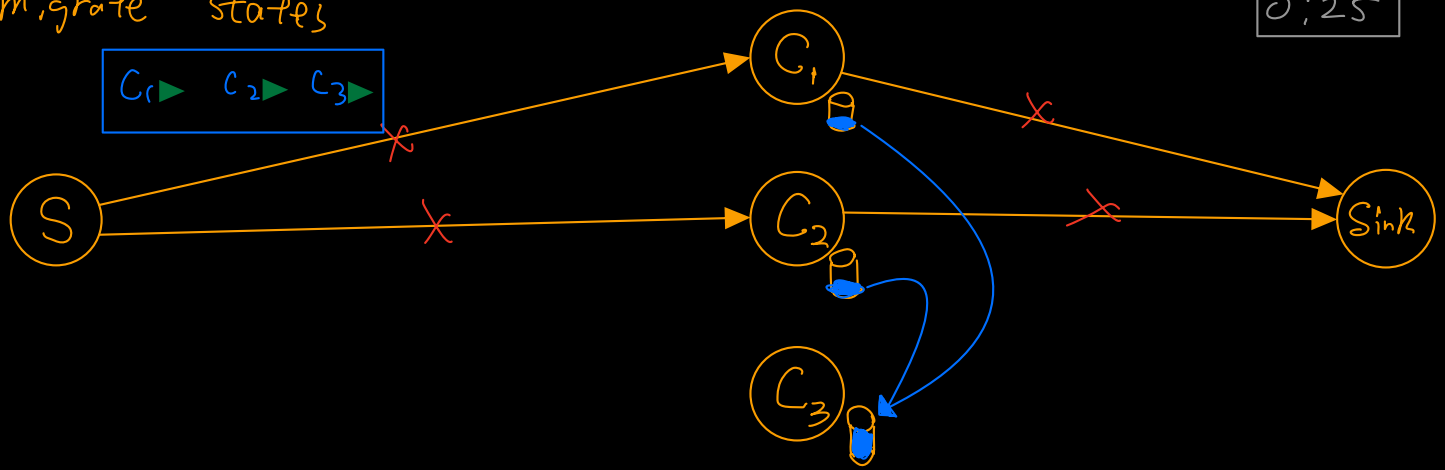
Reconfig Controller

1. A process that start timing once instantiated.
2. Notify and trigger reconfiguration
3. Instantiate new worker

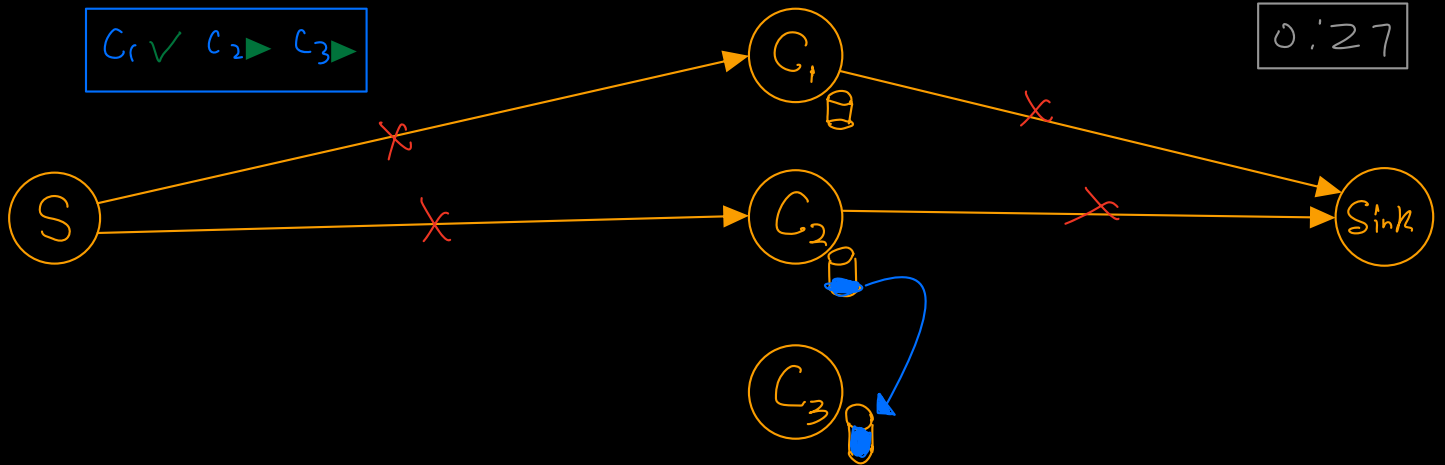


5 migrate states

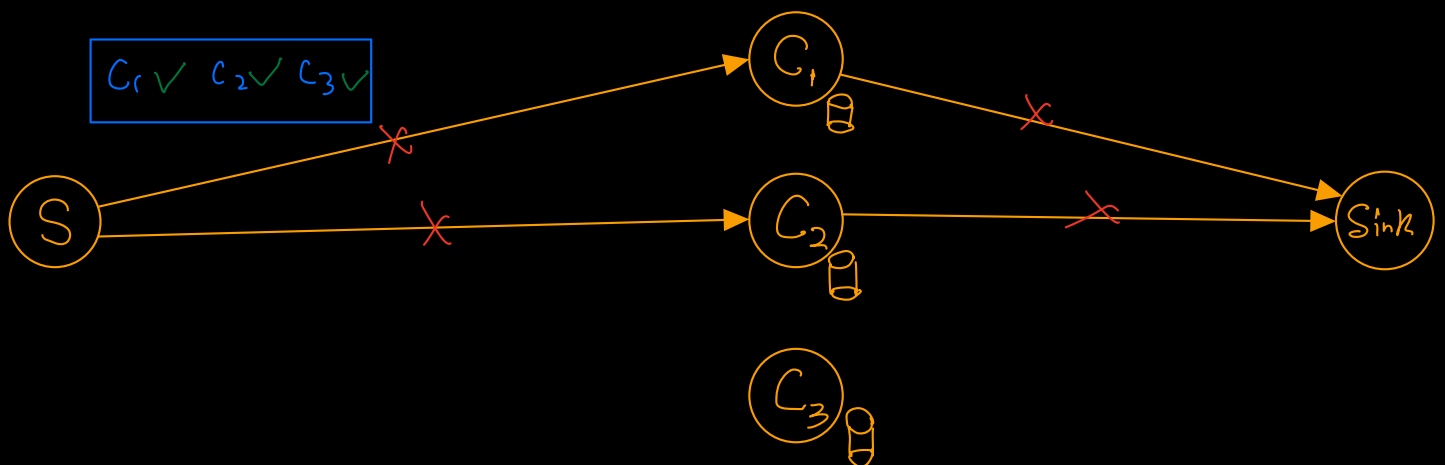
0:25



0:27



0:30



⑥ resume data stream

0:31

