Streaming Assignment

The problem

At TaxFix, the data revolves all around events that a user triggers on the mobile or web applications. Often it is required to have these events aggregated in a meaningful way. For example: our Customer Success team would like to see the last 5 events triggered by a user.

Your challenge
The objective is a simple streaming application that does the following:

- Receives input text lines, from **netcat**, with a JSON object like:
  `{ "Id": 1, "Url": "http://foobar.com", "Time": 123456789 }`

- Aggregates in real-time by "Id" field, and keeps the last 5 URLs seen for that Id

- An ElasticSearch index needs to be updated with that information(Id + Last 5 URLs) on every streaming iteration.

Resources
- Please take the example from Spark Streaming (the one that uses netcat as input) as a starting point:
  http://spark.apache.org/docs/latest/streaming-programming-guide.html#a-quick-example

- Elastic Search Integration:
  https://www.elastic.co/guide/en/elasticsearch/hadoop/master/spark.html

Technology
- The solution would ideally use the following technologies (but you are free to switch them with your preference):
  - Netcat (for streaming input)
  - Spark Streaming (for processing and aggregations)
  - ElasticSearch (for storing the final result)
  - Your choice of programming language (Python / Java / Scala)

- Please containerize your solution using Docker, so that we are able to run it

Time Advice
Don't spend more than 3-4 hours on implementing this. If the ElasticSearch integration is not working or some other problems arise, just do something else interesting.