1. Take given json file and create schema file, which can be standard json schema/xml schema formats or define your own schema.



Note: In future there may be schema evaluation or schema changes. So, define schema format which supports all the future changes with minimalistic modification

1. Read the given json file using the above schema and flatten the data and write to csv. Maintain relation between each flatten output. So that whenever we want to combine using the relationship key we can combine and form entire record.

Example for monitoring we will have one csv, for campaign will have one csv.

Note: There are couple array objects and struct objects in json.

1. Write Kafka and spark streaming application to find top seller for every 10 min. Output should be in csv file.

criteria for considering top seller:

Total items sales should be >=5

Total amount of items should be >=500

1. 4. create spark application to join two files of each of size 100Gb and 250Gb efficiently.

Note: Consider all expected issues and ensure the code should handle the data provided without any issues.