Functional Requirements
topK(k, startTime, endTime)

Non-Functional Requirements
- Scalabale
- Highly Available
- Highly Performant (few 10's milliseconds to return top K list)

Estimates
-
    K ~<= 1000
    Network Bandwidth:

Design
-
    Data Structures:
    - MinHeap,
    - HashTable,
    - Count-Min Sketch
      (Alternatives Counter based Algorithms like Loosy Counting, Space Saving, Sticky Sampling)

    Components
    - API Gateway
        - Single entry point for all clients
        - Aggregates data on fly
        - Data is flushed based on time 7 size
        - Serializes data in compact format
        - Buffering & Batching
    - Fast processor
        - Creates Count-Min Sketch for short periods of time
        - Because memory is no longer a problem, no need to partition
        - Data replication may not be required
    - Storage
        - NoSQL DB
        - Final Count-Min Sketch
        - Saves for several minutes, for each minute
        - Flush old data to DFS
    - Data Partitioner
        - parses batches of events into individual
    - Partition Processor
        - Aggregates im-memory for several minutes
        - Generates files of specified size
    - MapReduce job for TopK: for each hour

Retrieval

- Last 5 Min Query: Combine results for each minute
- Last 2 hours list: Combine results for each 1 hour