Consumer Poll Behavior

* Kafka consumers have a poll model that means they will ask data from kafka.

push model - Server pushes data & consumer waits.

poll model - allows consumers to control where in the log they want to consume, how fast and gives them the ability to replay events.

* fetch.min.bytes (default 1):
  Controls how much data you want to pull at least on each request.
  - helps improving throughput and decreasing request number.
  - At the cost of latency.

* Max.poll.records (default: 500):
  - controls how many records to receive per poll request
  - Increase if your messages are very small and have a lot of available RAM.
  - Good to monitor how many records are polled per request.

Max.partitions.fetch.bytes (default 1MB):
- max data returned by the broker per partition.
- If you read from 100 partitions, you'll need a lot of RAM

fetch.max.bytes (default 50 MB)
- max data returned for each fetch request (covers multiple partitions).
- the consumer performs multiple fetches in parallel.

change these settings only if your consumer maxes out on throughput already