Design and Describe a system to process a file/stream of JSON content in a non-serial manner, persist the records to a transactional data store, and log a “Finished” event upon persistence of the final record. The goal is to minimize processing time and actively recognize processing completion, without losing data. Communicate your design using any mediums you deem necessary or effective. (eg. Descriptive English language, Illustrations, Pseudocode, etc.)

1. Assume you have control of the file/stream.
2. Assume that the JSON total size is on the order of multiple gigabytes or larger, with variable record schema and size,
3. Assume that the Transactional persistence mechanism may be intermittently unavailable and also when available has the potential to reject individual records.
4. Individual records can fail persistence. If they do fail, the process should continue and the failure facts should not be lost, in order to attempt remediation.
5. Please include a description of how the remediation process could work.
6. Be sure to describe stack component expectations and feel free to make specific component recommendations that meet your needs. eg.

"This component should be an in-memory store. Redis would be a good choice."

Process:

- process a file/stream of JSON content in a non-serial manner
- persist the records to a transactional data store
- log a “Finished” event upon persistence of the final record.

Goal:

- minimize processing time
- actively recognize processing completion
- Data should not be lost

Given Assumptions:

- have control of the file/stream.
- total size is on the order of multiple gigabytes or larger, with variable record schema and size

Known Problems:

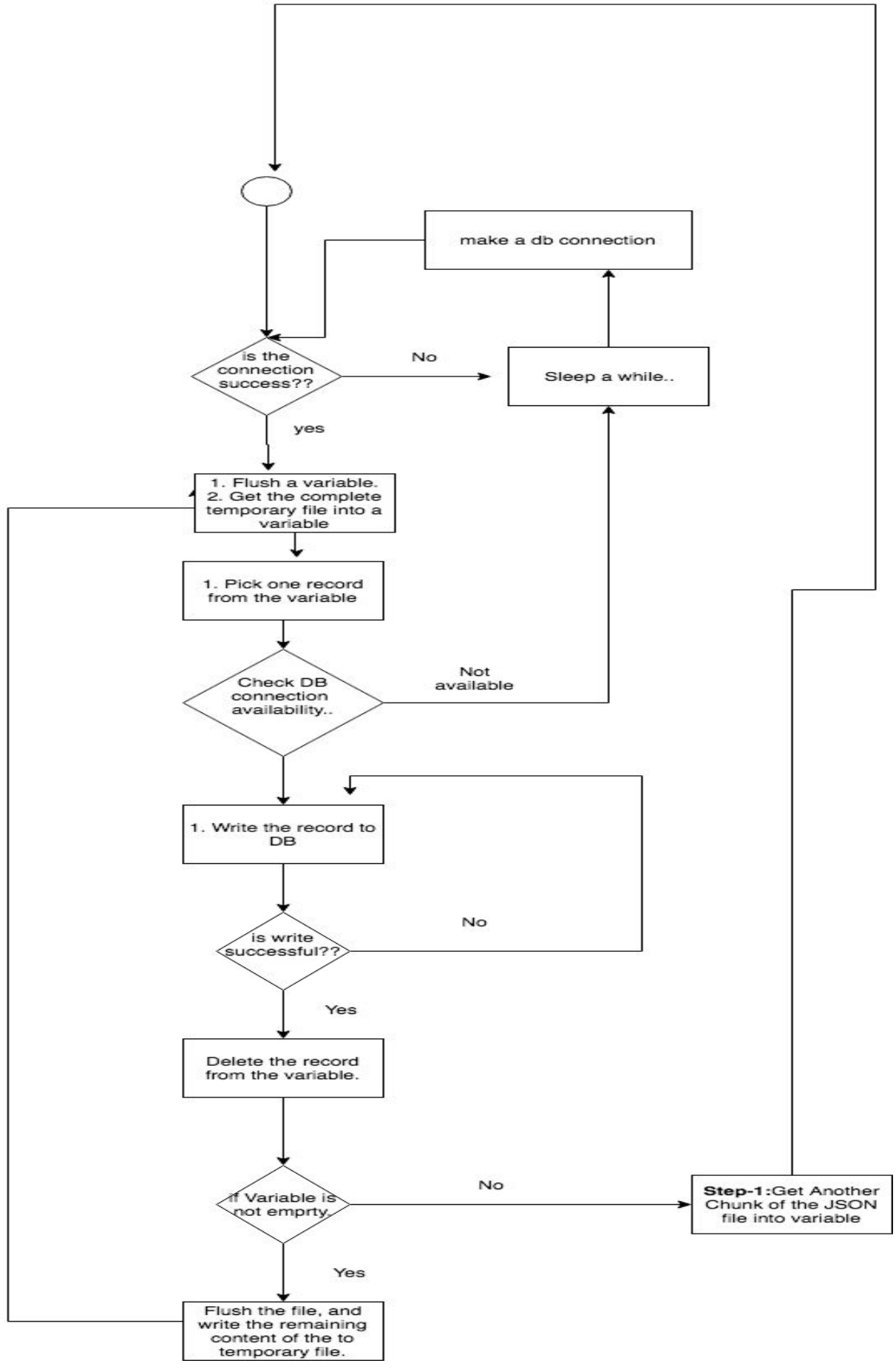
- Transactional persistence mechanism may be intermittently unavailable
- and also when available has the potential to reject individual records.
- Individual records can fail persistence.
- If they do fail, the process should continue and the failure facts should not be lost, in order to attempt remediation.

Taken Assumptions:

- I am assuming that the definition of process is writing the content of json to mongodb/redis db.

Design:

- Choose a best python in build module the process a big JSON file as chunks using **generators** concept.
- Make sure that the size of the chunk should be < 0.5 of the hardware(RAM).
- **Step-1:** Read a first chunk from the JSON into a variable.
- **Step-2:** Write it into a temporary persistent file. Make sure that the temporary file will never be lost even there is an unexpected error in the logic.
- **Step-3:** After the successful writing of the content to a file, delete the read chunk from the actual file.
- **Step-4: Process the complete chunk**(Defined below) at a time. Delete the temporary file after processing all the chunk successfully.
- Repeat **Step-1** to **Step-4.**
- **Process the complete chunk implementation has been mentioned in the below flow diagram.**


make a db connection
is the connection success??
No
Sleep a while..
yes
1. Flush a variable.
2. Get the complete temporary file into a variable
1. Pick one record from the variable
Check DB connection availability..
Not available
1. Write the record to DB
is write successful??
No
Yes
Delete the record from the variable.
if Variable is not emprty.
No
Step-1:Get Another Chunk of the JSON file into variable
Yes
Flush the file, and write the remaining content of the to temporary file.