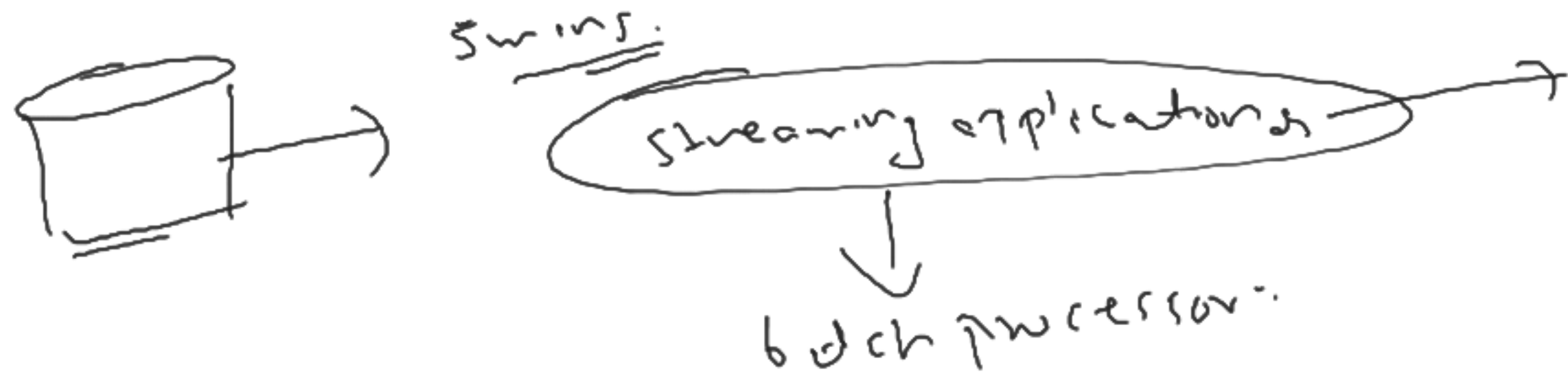


Streaming is not subject for batch processing.
↓
Superset.

batchwise processing using spark streaming.



Microbatch Streaming Processing

① hard work/stress.

1. Automatic looping between microbatches.
2. Batch start and partition management.
3. Intermediate state management.
4. combining result to the previous batch results.
5. Fault tolerant restart management.

- ① Spark Streaming (DataStream) API (Unstructured Streaming) (legacy)
↓
evolution
1. RDD based streaming API.
 - ② Lacks Spark SQL engine optimisations.
 - ③ No support for event time semantics.
 - ④ No futures upgrades and enhancements expected.
-

RDD -> DStream (Structured Streaming)

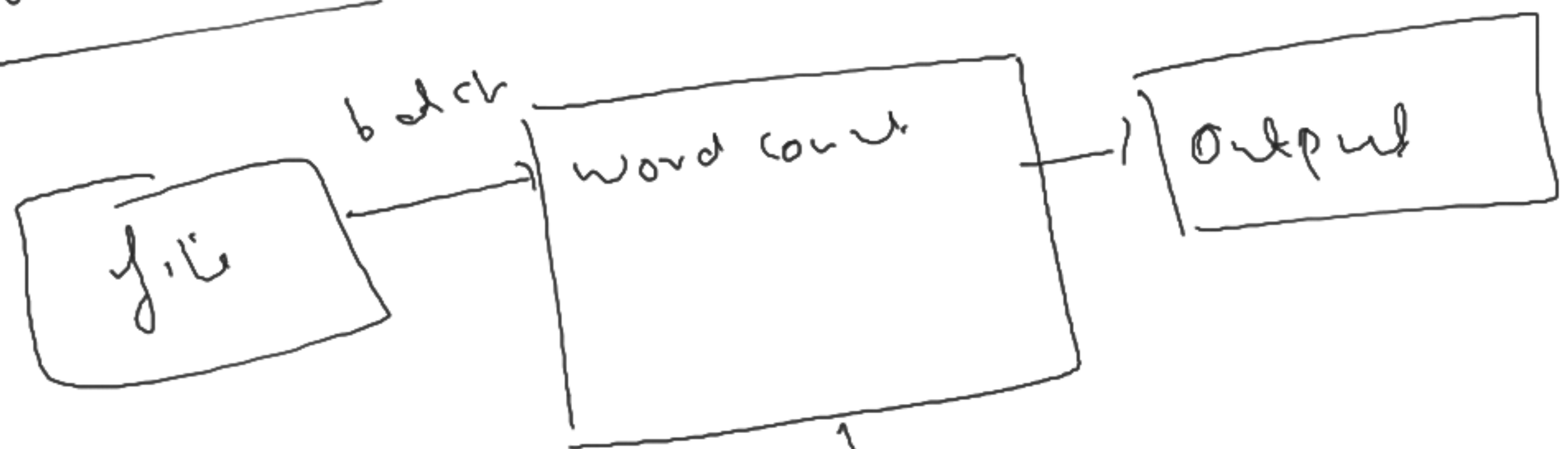
- ① Dataframe based streaming API.
- ① SQL engine optimisation.
- ③ supports event time semantics.
- ④ Enabled further enhancements and new features.

- ① Read a streaming source.
- ② Transform \rightarrow Output DF.
- ③ Write the output.

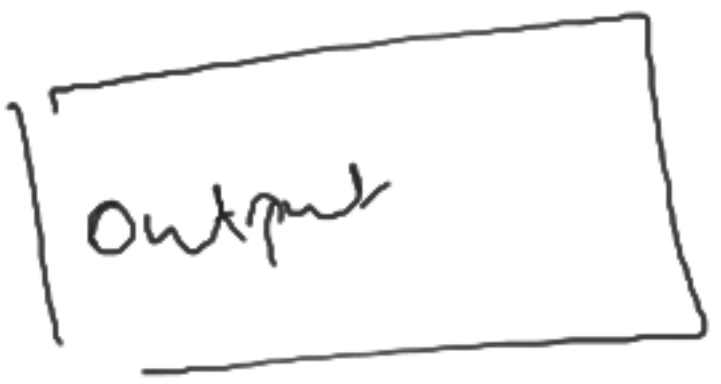
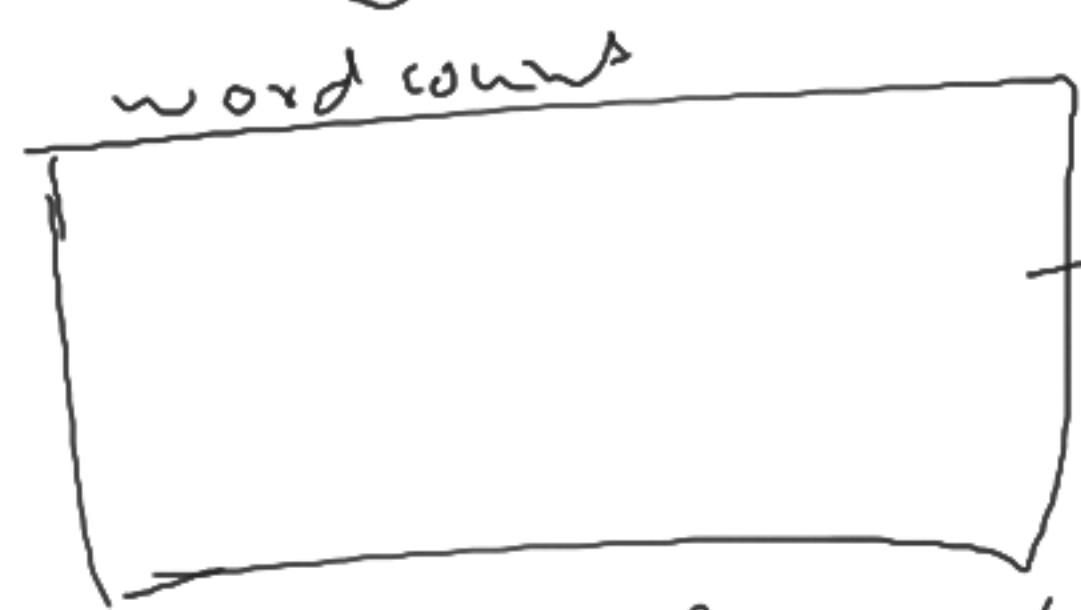
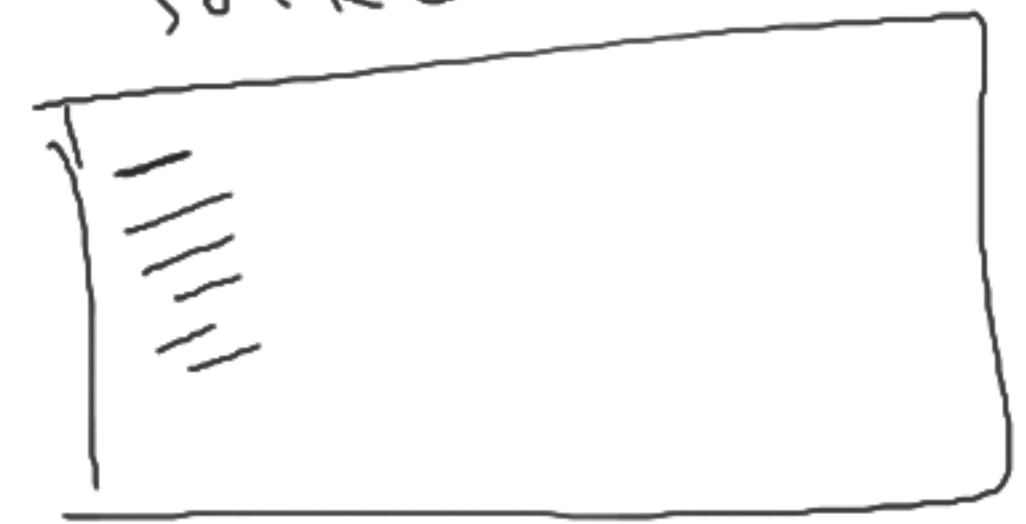
Transformations (batch processing) will be available to streaming processing.

Action can't be performed on streaming jobs.

Word Count - (File)



socket (7777).



Port - 80,443,

(Reserved,
(for priv.)

(01)

(Internet)