

Question 1

Correct Noté sur 1,00

The discretized streams model sets epochs at small time intervals and compute a Spark job for each epoch.

(Bareme: -12.5% if incorrect selection)

Veuillez choisir une réponse.

- ☐ a. I don't know (no point, no penalty)
- ☐ b. False
- ☒ c. True ✓

Votre réponse est correcte.

La réponse correcte est : True




Question 2

Correct Noté sur 1,00

In the context of a streaming system such as Spark streaming, "event time" refers to the time at which the data event is received by Spark's processor event engine: a record with a lower event time will thus be processed before records with higher event times.

(Bareme : from -16.67% if incorrect selection).

Veillez choisir une réponse.

- ☐ a. I don't know (no point, no penalty)
- ☐ b. True
- ☒ c. No  Event time is to be distinguished from processing time. Event time typically plays a role on how data is grouped for agregation, but it's processing time is which indicate when data is processed

☐ d.

Votre réponse est correcte.

La réponse correcte est :

No

Question 3

Correct Noté sur 1,00

the discretized streams model has lower latency than the continuous processing mode.

(Bareme: -12.5% if incorrect selection)

Veuillez choisir une réponse.

- ☒ a. False ✓
- ☐ b. I don't know (no point, no penalty)
- ☐ c. True

Votre réponse est correcte.

La réponse correcte est : False



Question 4

Incorrect Noté sur 1,00

the continuous processing mode does not set epochs

(Bareme: -12.5% if incorrect selection)

Veillez choisir une réponse.

- ☐ a. I don't know (no point, no penalty)
- ☒ b. True ✖
- ☐ c. False

Votre réponse est incorrecte.

La réponse correcte est : False

Question 5

Incorrect Noté sur 1,00

In spark structured streaming, data that arrives late might be automatically discarded instead of being processed in the window that contains it's event time.

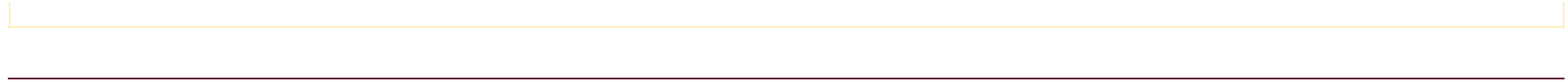
(Bareme: -12.5% if incorrect selection)

Veillez choisir une réponse.

- ☐ a. True
- ☒ b. False ✖
- ☐ c. I don't know (no point, no penalty)

Votre réponse est incorrecte.

La réponse correcte est : True



Question 6

Correct Noté sur 1,00

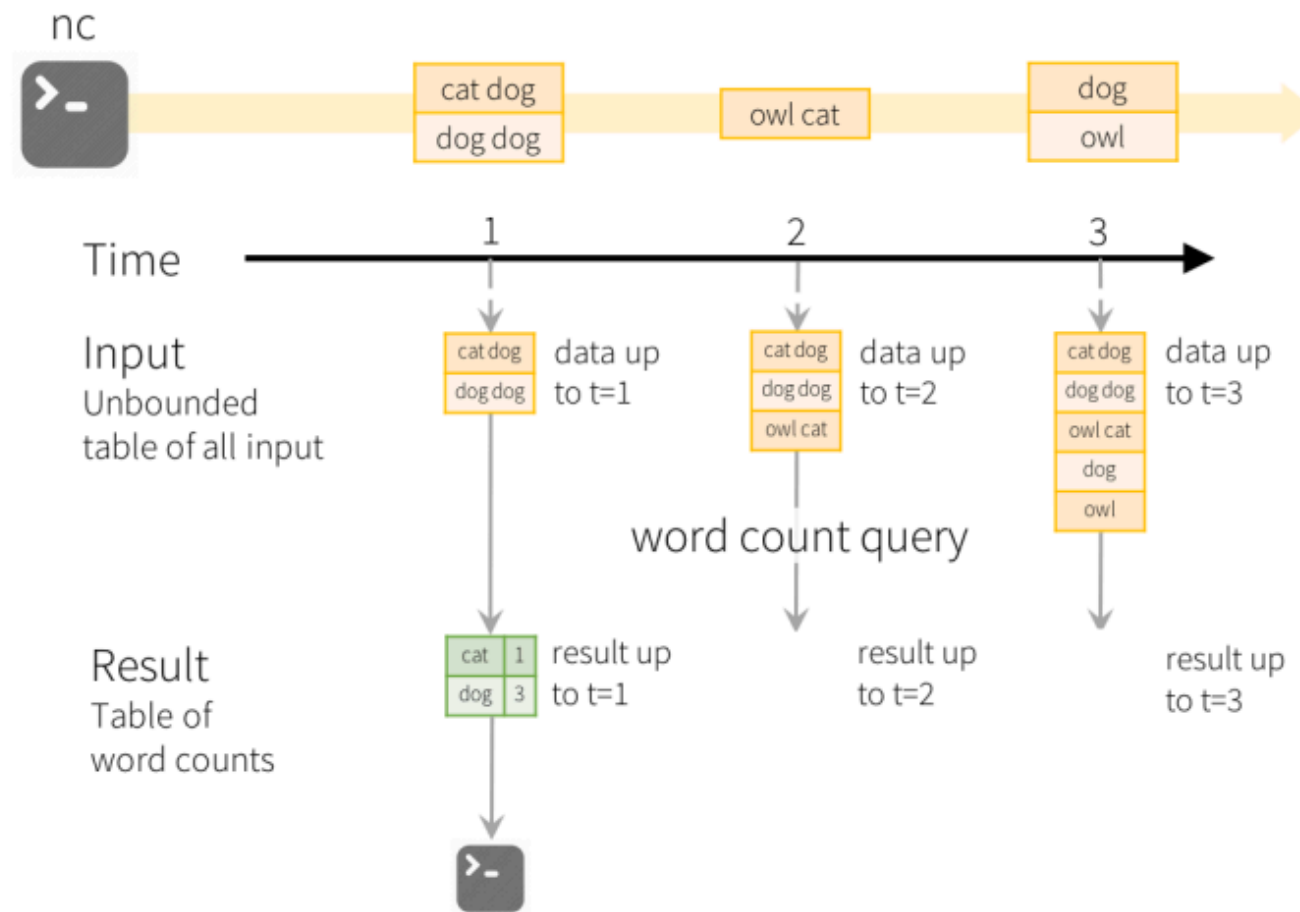
We recall that in spark structured streaming there are 3 output modes: complete mode, append mode, and update mode.

We consider a wordcount example where for the data read up to $t=1$ contains 2 records 'cat dog' and 'dog dog'.

In that case, the result of word counts up to $t=1$ contains 2 records= (cat, 1) and (dog,3).


Next, we suppose that between $t=1$ and $t=2$ we receive a new input record 'owl cat'.

Select the correct statement below



(Bareme: -20% if incorrect selection)

Veillez choisir une réponse.

- ☐ a. I don't know (no penalty, no point)
- ☐ b. In complete mode, after receiving the new data for the result up to $t=2$, spark streaming should write 2 lines.
- ☐ c. In complete mode, after receiving the new data for the result up to $t=2$, spark streaming should write 0 line.
- ☒ d. In complete mode, after receiving the new data for the result up to $t=2$, spark streaming should write 3 lines. 
- ☐ e. In complete mode, after receiving the new data for the result up to $t=2$, spark streaming should write 1 line.

Votre réponse est correcte.

La réponse correcte est : In complete mode, after receiving the new data for the result up to $t=2$, spark streaming should write 3 lines.

Question 7

Correct Noté sur 1,00

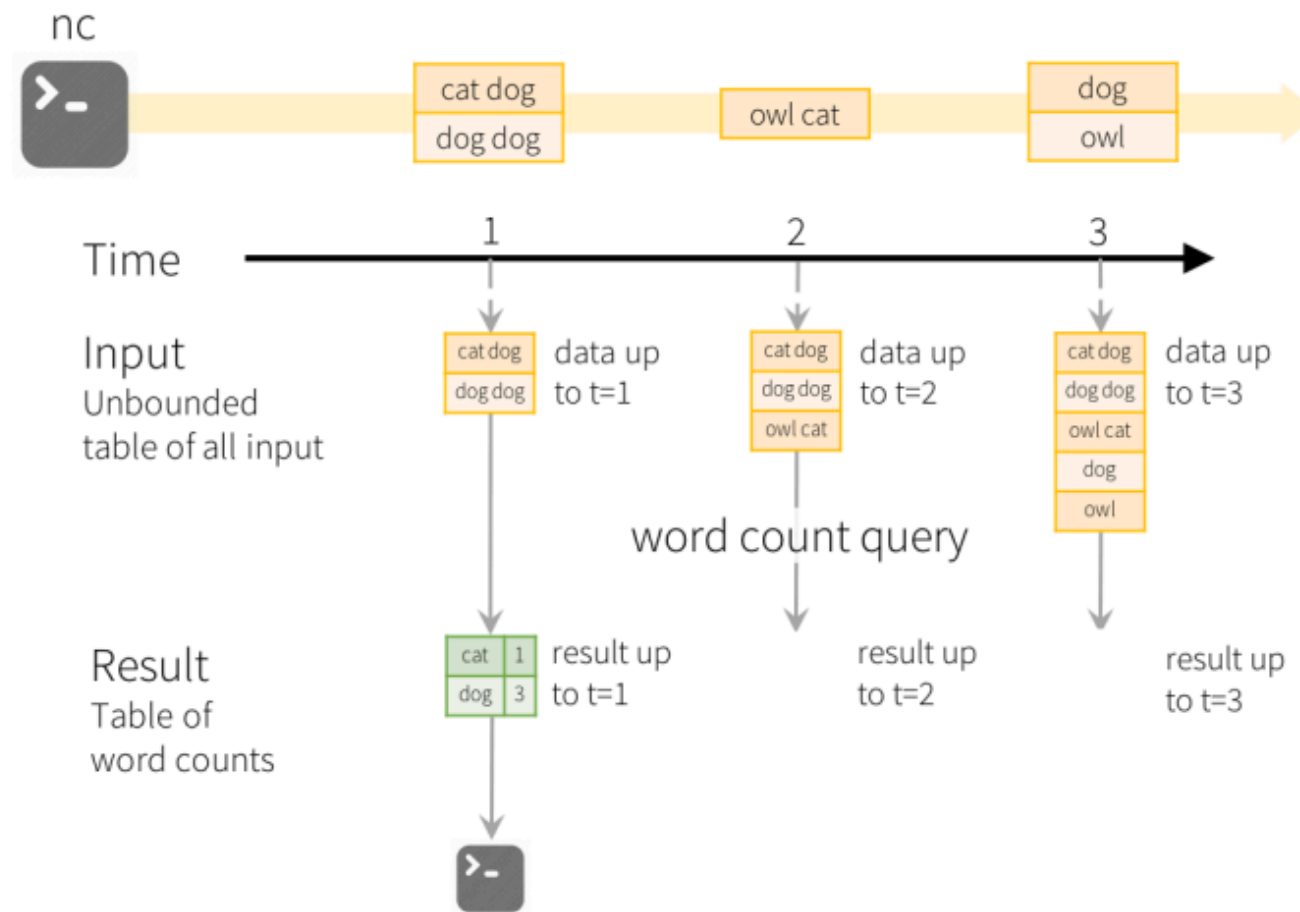
We recall that in spark structured streaming there are 3 output modes: complete mode, append mode, and update mode.

We consider a wordcount example where for the data read up to $t=1$ contains 2 records 'cat dog' and 'dog dog'.

In that case, the result of word counts up to $t=1$ contains 2 records= (cat, 1) and (dog,3).


Next, we suppose that between $t=1$ and $t=2$ we receive a new input record 'owl cat'.

Select the correct statement below



(Bareme: -20% if incorrect selection)

Veillez choisir une réponse.

- ☐ a. In update mode, after receiving the new data for the result up to $t=2$, spark streaming should write 0 line .
- ☐ b. In update mode, after receiving the new data for the result up to $t=2$, spark streaming should write 1 line .
- ☒ c. In update mode, after receiving the new data for the result up to $t=2$, spark streaming should write 2 lines . 
- ☐ d. I don't know (no penalty, no point)
- ☐ e. In update mode, after receiving the new data for the result up to $t=2$, spark streaming should write 3 lines .

Votre réponse est correcte.

La réponse correcte est : In update mode, after receiving the new data for the result up to $t=2$, spark streaming should write 2 lines .

Question 8

Incorrect Noté sur 1,00

In Spark structured streaming, Session windows typically overlap

(Bareme: -16.67% if incorrect selection)

Veillez choisir une réponse.

- ☒ a. True ✖
- ☐ b. I don't know (no point, no penalty)
- ☐ c. False

Votre réponse est incorrecte.

La réponse correcte est : False

Question 9

Correct Noté sur 1,00

Update mode requires the whole aggregate data to be preserved for aggregates so watermarking cannot be used in update mode to clean the state.

(Bareme: -16.67% if incorrect selection)

Veillez choisir une réponse.

- ☒ a. False ✓
- ☐ b. True
- ☐ c. I don't know (no point, no penalty)

Votre réponse est correcte.

La réponse correcte est : False

Question 10

Correct Noté sur 1,00

In spark structured streaming, the purpose of watermarking is to manage checkpointing for fault tolerance.

(Bareme: -16.67% if incorrect selection)

Veillez choisir une réponse.

- ☐ a. I don't know (no point, no penalty)
- ☐ b. True
- ☒ c. False ✓

Votre réponse est correcte.

La réponse correcte est : False