

Quiz 9

01 Consider the characteristic of having a very-high dimensional space in the context of finding similar items in a dataset. Which of Vs (listed below) of Big Data is more closely related to this characteristic?

- ☐ Variety
- ☐ Visibility
- ☐ Velocity

X Volume

02 Which of the tasks below can be considered a data enrichment task in the context of Process mining?

X Annotating each event with the Case ID (process instance identifier) it belongs to

- ☐ Filtering out database exceptions stored in the log
- ☐ Mapping the original representation of an event to MXML (representation used for storing event logs)
- ☐ None of the above

03 Consider the Min-Hashing example shown in Slide #29 of the lecture on Finding Similar Items. What's the signature matrix for the permutation [1,7,2,4,3,6,5] (read this as a column vector where the element on top is 1).

- ☐ [3,1,1,2]

X [1,2,1,2]

- ☐ [1,5,3,2]
- ☐ [2,2,1,3]

04 The problem of not being able to access relevant events that are needed to perform process mining tasks can be considered as related to the Visibility dimension of Big Data.

X True

- ☐ False

05 In the architecture example discussed in the lecture on how to leverage on Apache Hadoop to perform event correlation (in process mining), which of the options below best characterizes the type of processing used in the solution:

X Batch processing

- ☐ Interactive processing
- ☐ Graph-based processing
- ☐ None of the above