



Ung dung Big Data: Truyen du lieu trong thoi gian thuc_ Nhom 01

[Nhà của tôi](#) / [Các khoá học của tôi](#) / [BDAS436177_23_1_01](#) / [Day 14: Topic 3: Streaming binary files](#) / [Bài mock exam #2](#)

Bắt đầu vào lúc	Thursday, 23 November 2023, 3:04 PM
Trạng thái	Đã xong
Kết thúc lúc	Thursday, 23 November 2023, 3:20 PM
Thời gian thực hiện	15 phút 28 giây
Điểm	10,00 trên 10,00 (100%)

Câu hỏi 1

Đúng

Đạt điểm 0,25 trên 0,25

Dữ liệu đang trong dòng dữ liệu thời gian thực được gọi là?

- ☐ a. Streaming data
- ☒ b. In-flight data
- ☐ c. Analysed data
- ☐ d. Online data



Your answer is correct.

Câu hỏi 2

Đúng

Đạt điểm 0,25 trên 0,25

Cho ma trận Count-min sketch như hình sau.

41	44	28	23	43
40	11	17	36	45
49	43	26	33	20
30	42	40	40	19
28	10	39	17	44

Giả sử kết quả các hàm băm fi (với i là chỉ mục hàng trong ma trận, bắt đầu từ 0) như sau:

$f_0(\text{"bigdata"}) = 1$

$f_1(\text{"bigdata"}) = 2$

$f_2(\text{"bigdata"}) = 2$

$f_3(\text{"bigdata"}) = 0$

$f_4(\text{"bigdata"}) = 3$

Tần suất ước lượng của dữ liệu chuỗi "bigdata" là?

- ☐ a. 26
- ☒ b. 17
- ☐ c. 10
- ☐ d. 30



Your answer is correct.

Câu hỏi 3

Đúng

Đạt điểm 0,25 trên 0,25

Trong phương pháp cửa sổ trượt của tầng phân tích, số cửa sổ nằm gọn hay nằm một phần trong tầng phân tích được gọi là?

- ☐ a. Cửa sổ chồng
- ☐ b. Cả 3 câu đều sai
- ☒ c. Cửa sổ được quan tâm
- ☐ d. Cửa sổ được chứa



Your answer is correct.

Câu hỏi 4

Đúng

Đạt điểm 0,25 trên 0,25

Ví dụ nào dưới đây liên quan đến ràng buộc one-pass?

- ☐ a. Vì một sự kiện giảm giá, lượng dữ liệu vào hệ thống streaming tăng đột biến.
- ☐ b. Hệ thống gợi ý không còn chính xác do mô hình bị lỗi thời.
- ☐ c. Số lượng người dùng thường xuyên truy cập trang web ở mức 500-1000 người/giờ.
- ☒ d. Dữ liệu đã xử lý không thể được xử lý lại.



Your answer is correct.

Câu hỏi 5

Đúng

Đạt điểm 0,25 trên 0,25

Cho ma trận Count-min sketch như hình sau.

18	44	47	77
15	24	5	57
67	13	31	2
8	4	5	43

Giả sử kết quả các hàm băm fi (với i là chỉ mục hàng trong ma trận, bắt đầu từ 0) như sau:

f0("bigdata") = 1

f1("bigdata") = 2

f2("bigdata") = 2

f3("bigdata") = 0

Tần suất ước lượng của dữ liệu chuỗi "bigdata" là?

- ☐ a. 8
- ☒ b. 5
- ☐ c. 2
- ☐ d. 4



Your answer is correct.

Câu hỏi 6

Đúng

Đạt điểm 0,25 trên 0,25

Tên gọi khác của one-way pattern là?

- ☐ a. Fire then forget
- ☒ b. Fire and forget
- ☐ c. Shot and forget
- ☐ d. Shot then forget



Your answer is correct.

Câu hỏi 7

Đúng

Đạt điểm 0,25 trên 0,25

Cấu hình cửa sổ trượt cho tầng phân tích như sau: Kích thước cửa sổ là 20 giây, thanh trượt là 5 giây. Khi chạy thử hệ thống, người ta nhận thấy thời gian dữ liệu vào hệ thống sớm nhất là t giây thì dữ liệu ra trễ nhất là $t + 60$ giây. Tính số cửa sổ **cần quan tâm**.

- ☐ a. 14
- ☒ b. 15
- ☐ c. 6
- ☐ d. 9



Your answer is correct.

Câu hỏi 8

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là hạn chế của phương pháp caching write-behind?

- ☐ a. Kiến trúc phức tạp.
- ☐ b. Độ trễ cao.
- ☒ c. Tăng khả năng mất mát dữ liệu.
- ☐ d. Chi phí tăng cao.



Your answer is correct.

Câu hỏi 9

Đúng

Đạt điểm 0,25 trên 0,25

Trạng thái toàn cục nhất quán có thể hiểu là?

- ☒ a. Trạng thái toàn hệ thống có tính khả thi nếu không có lỗi.
- ☐ b. Trạng thái toàn hệ thống sau khi đã khôi phục từ checkpoints.
- ☐ c. Trạng thái ban đầu của toàn hệ thống.
- ☐ d. Trạng thái hệ thống lấy trạng thái một nút làm đại diện.



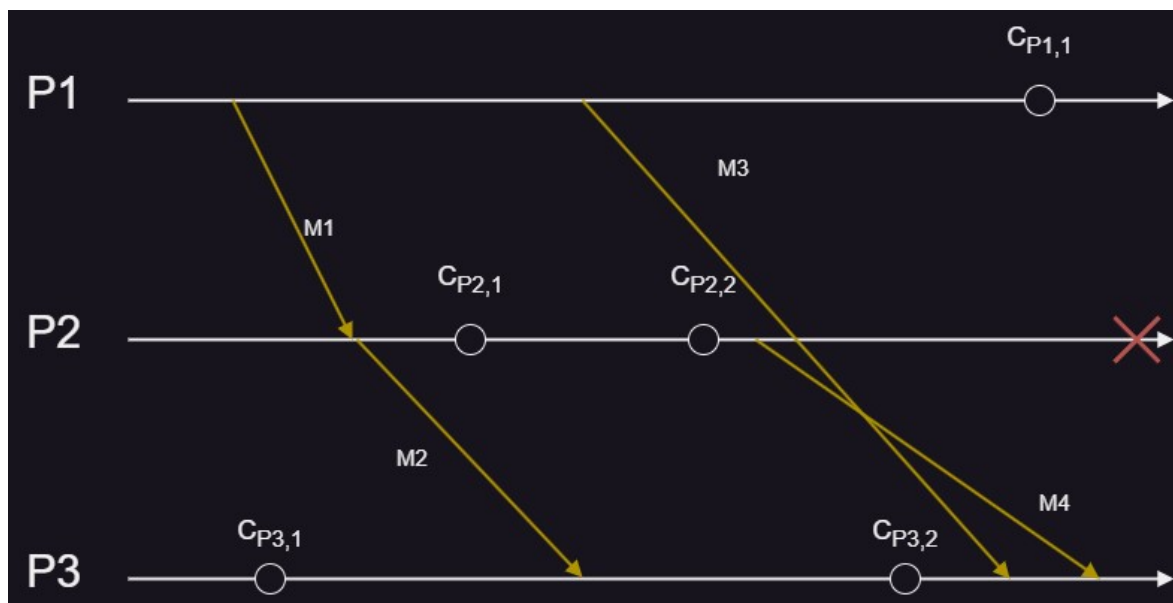
Your answer is correct.

Câu hỏi 10

Đúng

Đạt điểm 0,25 trên 0,25

Cho biểu đồ thể hiện quá trình trao đổi gói tin giữa các tiến trình như sau.



Trong trường hợp một trong các hệ thống bị lỗi tại X đỏ, tập checkpoint nào sau đây giúp trạng thái toàn cục nhất quán?

- ☐ a. $\{CP_{1,1}, CP_{2,2}, CP_{3,1}\}$
- ☐ b. Tất cả đều sai
- ☒ c. $\{CP_{1,1}, CP_{2,2}, CP_{3,2}\}$
- ☐ d. $\{CP_{1,1}, CP_{2,1}, CP_{3,1}\}$



Your answer is correct.

Câu hỏi 11

Đúng

Đạt điểm 0,25 trên 0,25

Cho ma trận Count-min sketch như hình sau.

41	44	28	23	43
40	11	17	36	45
49	43	26	33	20
30	42	40	40	19
28	10	39	17	44

Giả sử kết quả các hàm băm fi (với i là chỉ mục hàng trong ma trận, bắt đầu từ 0) như sau:

 $f_0(\text{"cntt"}) = 3$ $f_1(\text{"cntt"}) = 1$ $f_2(\text{"cntt"}) = 0$ $f_3(\text{"cntt"}) = 1$ $f_4(\text{"cntt"}) = 4$

Tần suất ước lượng của dữ liệu chuỗi "cntt" là?

- ☐ a. 20
- ☒ b. 11
- ☐ c. 17
- ☐ d. 10



Your answer is correct.

Câu hỏi 12

Đúng

Đạt điểm 0,25 trên 0,25

Đây là vấn đề của phương pháp incremental checkpointing?

- ☐ a. Chi phí ổ lưu trữ cao.
- ☐ b. Khôi phục chậm so với phương pháp khác.
- ☒ c. Xung đột hash.
- ☐ d. Cấu hình phức tạp.



Your answer is correct.

Câu hỏi 13

Đúng

Đạt điểm 0,25 trên 0,25

Trong RabbitMQ, phương pháp basic QOS với tham số prefetch_count=1 hoạt động như thế nào?

- ☐ a. Gửi nhiều gói tin đến worker, chỉ gửi tiếp khi nhận được ACK.
- ☐ b. Gửi tối đa một gói tin đến worker, chỉ gửi tiếp khi nhận được N-ACK.
- ☐ c. Gửi nhiều gói tin đến worker, chỉ gửi tiếp khi nhận được N-ACK.
- ☒ d. Gửi tối đa một gói tin đến worker, chỉ gửi tiếp khi nhận được ACK.



Your answer is correct.

Câu hỏi 14

Đúng

Đạt điểm 0,25 trên 0,25

Một hàm băm cho ra kết quả băm với độ dài không đổi là 8 bits. Giả sử ta giới hạn giá trị nhỏ nhất của kết quả là 0, giá trị số nguyên tối đa mà hàm có thể thể hiện là?

- ☒ a. 255
- ☐ b. 63
- ☐ c. 64
- ☐ d. 256



Your answer is correct.

Câu hỏi 15

Đúng

Đạt điểm 0,25 trên 0,25

Khác biệt giữa phương pháp tumbling và cửa sổ trượt trong tăng phân tích?

- ☐ a. Tumbling kém hơn về mặt hiệu năng.
- ☐ b. Tumbling không xét yếu tố thời gian.
- ☒ c. Cửa sổ trong tumbling không chồng lên nhau.
- ☐ d. Tumbling cần lượng dữ liệu vào không đổi.



Your answer is correct.



Câu hỏi 16

Đúng

Đạt điểm 0,25 trên 0,25

Đây là đặc trưng của phương pháp incremental checkpointing?

- ☐ a. Encryption
- ☒ b. Deduplication
- ☐ c. Replication
- ☐ d. RAID



Your answer is correct.

Câu hỏi 17

Đúng

Đạt điểm 0,25 trên 0,25

Cho biểu đồ thể hiện quá trình trao đổi gói tin giữa các tiến trình như sau. Trạng thái toàn cục tại thời điểm checkpoint $C_{P3,2}$ được khởi tạo là?

- ☐ a. Nhận M1, nhận M2, nhận M3, gửi M4.
- ☒ b. Nhận M1, nhận M2, nhận M3.
- ☐ c. Nhận M1, nhận M2, gửi M3, gửi M4.
- ☐ d. Tất cả đều sai.



Your answer is correct.

Câu hỏi 18

Đúng

Đạt điểm 0,25 trên 0,25

Ưu thế của phương pháp caching refresh-ahead?

- ☐ a. Độ tin cậy cao.
- ☐ b. Hiệu quả với dữ liệu caching lớn.
- ☒ c. Cả 3 câu đều sai.
- ☐ d. Quản lý cache theo khóa - giá trị.



Your answer is correct.



Câu hỏi 19

Đúng

Đạt điểm 0,25 trên 0,25

Trong Kafka, cấu hình batch.size là của?

- ☐ a. Consumer
- ☐ b. Zookeeper
- ☐ c. Connect
- ☒ d. Producer



Your answer is correct.

Câu hỏi 20

Đúng

Đạt điểm 0,25 trên 0,25

Mở rộng hệ thống bằng cách thêm RAM và CPU thuộc dạng?

- ☒ a. Mở rộng đứng
- ☐ b. Mở rộng đứng và ngang
- ☐ c. Tất cả đều sai
- ☐ d. Mở rộng ngang



Your answer is correct.

Câu hỏi 21

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là tên một công nghệ in-memory database (IMDB)?

- ☐ a. Redis
- ☐ b. IndexedDB
- ☒ c. Apache Geode
- ☐ d. SQLite



Your answer is correct.

Câu hỏi 22

Đúng

Đạt điểm 0,25 trên 0,25

Một hàm băm cho ra kết quả băm với độ dài không đổi là 20 bytes. Khi thể hiện kết quả đó dưới dạng số nhị phân thì ta sẽ có bao nhiêu chữ số?

- ☐ a. 140
- ☒ b. 160
- ☐ c. 200
- ☐ d. 120



Your answer is correct.

Câu hỏi 23

Đúng

Đạt điểm 0,25 trên 0,25

Sử dụng thuật toán Count-min sketch, cần hàm băm tối đa bao nhiêu bits để dùng với ma trận 15 cột?

- ☒ a. 3
- ☐ b. 5
- ☐ c. 2
- ☐ d. 4



Your answer is correct.

Câu hỏi 24

Đúng

Đạt điểm 0,25 trên 0,25

Hiện tượng nào sau đây thuộc nhóm lỗi response failure?

- ☒ a. Gói tin bị mất một phần khi đến nơi.
- ☐ b. Gói tin bị mất hoàn toàn trong khi gửi.
- ☐ c. Hệ thống không phản hồi ACK sau khi nhận gói tin.
- ☐ d. Tất cả đều sai.



Your answer is correct.

Câu hỏi 25

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là đặc trưng của phương pháp logging?

- ☐ a. Chống trùng.
- ☐ b. Lưu dữ liệu từ RAM.
- ☒ c. Gói tin bị lỗi được gửi lại.
- ☐ d. Phụ thuộc vào các phiên bản sao lưu trước.



Your answer is correct.

Câu hỏi 26

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là mệnh đề đúng về publish/subscribe pattern?

- ☒ a. Tất cả đều sai
- ☐ b. Bước lọc gói tin được thực hiện ở topic
- ☐ c. Một gói tin chỉ ứng với 1 topic
- ☐ d. 1 topic chỉ có thể kết nối với 1 subscription



Your answer is correct.

Câu hỏi 27

Đúng

Đạt điểm 0,25 trên 0,25

Dữ liệu được lưu trong thiết bị lưu trữ dài hạn được gọi là?

- ☐ a. In-process data
- ☐ b. Idle data
- ☒ c. Data at rest
- ☐ d. Local data



Your answer is correct.

Câu hỏi 28

Đúng

Đạt điểm 0,25 trên 0,25

Cấu hình cửa sổ trượt cho tầng phân tích như sau: Kích thước cửa sổ là 20 giây, thanh trượt là 5 giây. Khi chạy thử hệ thống, người ta nhận thấy thời gian dữ liệu vào hệ thống sớm nhất là t giây thì dữ liệu ra trễ nhất là $t + 60$ giây. Tính số cửa sổ bị chồng chất.

- ☒ a. 6
- ☐ b. 14
- ☐ c. 15
- ☐ d. 9



Your answer is correct.

Câu hỏi 29

Đúng

Đạt điểm 0,25 trên 0,25

Nếu một nút trong hệ thống bị lỗi thì ...

- ☐ a. Khôi phục nút bị lỗi sao cho trạng thái nút đó nhất quán.
- ☐ b. Khôi phục toàn bộ hệ thống sao cho trạng thái nút bị lỗi trước đó nhất quán.
- ☒ c. Khôi phục toàn bộ hệ thống sao cho trạng thái toàn cục nhất quán.
- ☐ d. Khôi phục nút bị lỗi sao cho trạng thái toàn cục nhất quán.



Your answer is correct.

Câu hỏi 30

Đúng

Đạt điểm 0,25 trên 0,25

IMDG khác IMDB ở yếu tố nào?

- ☒ a. Hỗ trợ phân tán.
- ☐ b. Hỗ trợ đảm bảo.
- ☐ c. Hỗ trợ phân tầng.
- ☐ d. Hỗ trợ đa nền tảng.



Your answer is correct.

Câu hỏi 31

Đúng

Đạt điểm 0,25 trên 0,25

Một trạng thái toàn cục được xem là không nhất quán nếu ...

- ☐ a. Gói tin chưa được gửi hoặc nhận.
- ☐ b. Bên gửi đã gửi gói tin, nhưng bên nhận chưa nhận.
- ☒ c. Bên nhận đã nhận gói tin, nhưng bên gửi chưa gửi.
- ☐ d. Tất cả đều sai.



Your answer is correct.

Câu hỏi 32

Đúng

Đạt điểm 0,25 trên 0,25

Lựa chọn nào không phải phân loại hệ thống thời gian thực theo thời gian đáp ứng?

- ☐ a. Soft
- ☒ b. Quick
- ☐ c. Near
- ☐ d. Hard



Your answer is correct.

Câu hỏi 33

Đúng

Đạt điểm 0,25 trên 0,25

Giá trị thống kê nào phù hợp với kỹ thuật tổng hợp lấy mẫu?

- ☒ a. Phương sai
- ☐ b. Tổng
- ☐ c. Cả 3 câu đều sai
- ☐ d. Giá trị nhỏ nhất



Your answer is correct.

Câu hỏi 34

Đúng

Đạt điểm 0,25 trên 0,25

Hiện tượng nào sau đây thuộc nhóm lỗi byzantine?

- ☐ a. Đường cáp mạng bị vấn đề, khiến gói tin không thể tới kịp thời.
- ☐ b. Nút trong hệ thống không phản hồi yêu cầu.
- ☒ c. Lỗi từ hệ quả tấn công khai thác zero-day vào một nút trong hệ thống.
- ☐ d. Gói tin bị mất một phần do bị thay đổi từ tin tặc bên ngoài.



Your answer is correct.

Câu hỏi 35

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là một mô hình kiến trúc logging?

- ☐ a. BSML
- ☐ b. HBML
- ☐ c. BRML
- ☒ d. SBML



Your answer is correct.

Câu hỏi 36

Đúng

Đạt điểm 0,25 trên 0,25

Trong hệ thống dữ liệu tĩnh, dữ liệu có n bảng ghi vào thì số lượng bảng ghi được phân tích là?

- ☒ a. Bảng n
- ☐ b. Lớn hơn n
- ☐ c. Bảng 0
- ☐ d. Nhỏ hơn n



Your answer is correct.



Câu hỏi 37

Đúng

Đạt điểm 0,25 trên 0,25

Câu nào dưới đây là đúng về RabbitMQ?

- ☐ a. RabbitMQ không thể chạy nếu thiếu producer hoặc consumer.
- ☒ b. Có thể tồn tại cùng lúc nhiều hàng đợi.
- ☐ c. Các hàng đợi phân biệt thông qua số thứ tự khởi tạo.
- ☐ d. Chỉ một consumer được kéo dữ liệu từ một hàng đợi.



Your answer is correct.

Câu hỏi 38

Đúng

Đạt điểm 0,25 trên 0,25

Giao thức lớp vận chuyển (transport) mà RabbitMQ dùng là?

- ☐ a. MQTT
- ☒ b. TCP
- ☐ c. UDP
- ☐ d. AMQP



Your answer is correct.

Câu hỏi 39

Đúng

Đạt điểm 0,25 trên 0,25

Sử dụng thuật toán Count-min sketch, cần hàm băm tối đa bao nhiêu bits để dùng với ma trận 10 cột?

- ☐ a. 4
- ☐ b. 2
- ☒ c. 3
- ☐ d. 5



Your answer is correct.



Câu hỏi 40

Đúng

Đạt điểm 0,25 trên 0,25

Trong tầng phân tích, chính sách kích hoạt dựa trên yếu tố nào?

- ☐ a. Lượng dữ liệu
- ☐ b. Thời gian và lượng dữ liệu
- ☐ c. Thời gian
- ☒ d. Thời gian hoặc lượng dữ liệu



Your answer is correct.

◀ Short essay exercise (group)

Chuyển tới...