



## 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 15: Topic 3: Media streaming](#) / [Bài mock exam #3](#)

**Bắt đầu vào lúc** Thursday, 30 November 2023, 2:31 PM

**Trạng thái** Đã xong

**Kết thúc lúc** Thursday, 30 November 2023, 3:01 PM

**Thời gian thực hiện** 30 phút 7 giây

**Điểm** 9,75 trên 10,00 (98%)

### Câu hỏi 1

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:

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

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

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

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

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

- ☐ a. 4
- ☐ b. 8
- ☐ c. 2
- ☒ d. 5



Your answer is correct.

## Câu hỏi 2

Đúng

Đạt điểm 0,25 trên 0,25

Khi nói về hai cách ghi dữ liệu gián tiếp và trực tiếp, nút nào dưới đây là điểm khác biệt giữa hai mô hình?

- ☐ a. Tầng message queue
- ☐ b. Nút lưu trữ dài hạn
- ☒ c. Batch loader
- ☐ d. Tầng Analysis



Your answer is correct.

## Câu hỏi 3

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:

$f_0(\text{"ktdl"}) = 2$

$f_1(\text{"ktdl"}) = 1$

$f_2(\text{"ktdl"}) = 1$

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

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

- ☐ a. 2
- ☐ b. 5
- ☐ c. 23
- ☒ d. 8



Your answer is correct.

**Câu hỏi 4**

Đúng

Đạt điểm 0,25 trên 0,25

Trong Spark Streaming, lớp trừu tượng thể hiện stream dữ liệu là?

- ☒ a. DStream
- ☐ b. DataStream
- ☐ c. DtStream
- ☐ d. Stream



Your answer is correct.

**Câu hỏi 5**

Đúng

Đạt điểm 0,25 trên 0,25

Mô tả nào sau đây khớp với sender-based message logging?

- ☐ a. Tất cả đều sai.
- ☐ b. Gói tin được lưu ngay sau khi nhận.
- ☒ c. Gói tin được lưu ngay trước khi gửi.
- ☐ d. Gói tin được lưu ngay trước khi xử lý.



Your answer is correct.



**Câu hỏi 6**

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  $f_i$  (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. 11
- ☐ b. 17
- ☐ c. 20
- ☐ d. 10



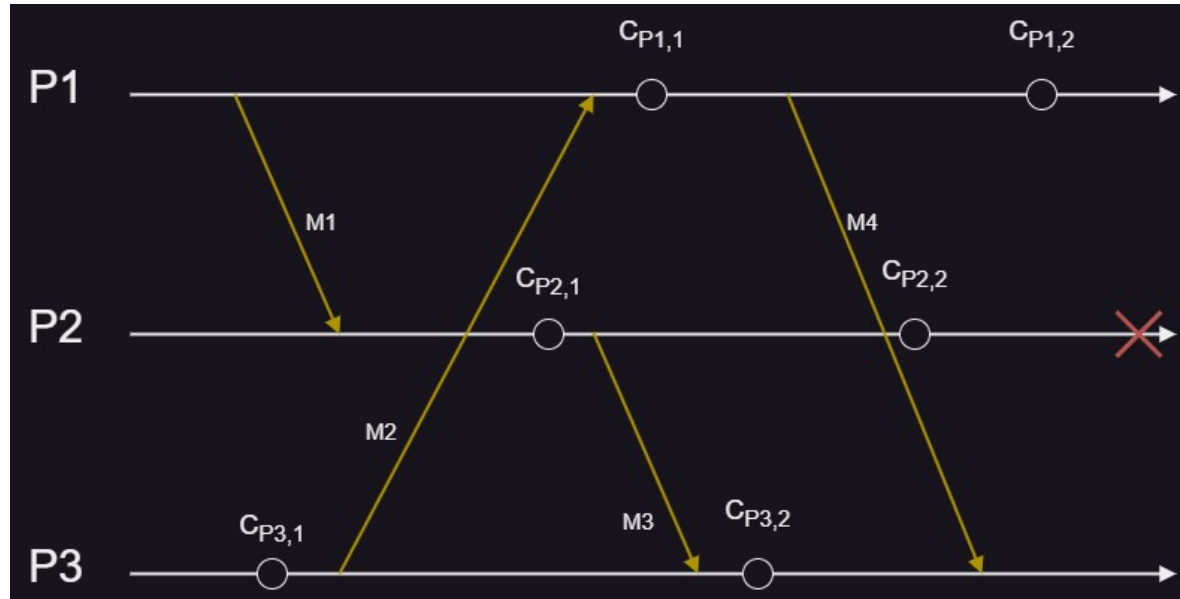
Your answer is correct.

**Câu hỏi 7**

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 khiến trạng thái toàn cục KHÔNG nhất quán?



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



Your answer is correct.

**Câu hỏi 8**

Đúng

Đạt điểm 0,25 trên 0,25

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

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



Your answer is correct.

**Câu hỏi 9**

Đúng

Đạt điểm 0,25 trên 0,25

Cho ma trận Count-min sketch như hình sau.

A	B	C	D
35	42	42	12
23	85	55	31
75	76	33	1
77	84	35	75

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

f0("xinchao") = 2

f1("xinchao") = 1

f2("xinchao") = 1

f3("xinchao") = 0

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

- ☐ a. 31
- ☐ b. 75
- ☒ c. 42
- ☐ d. 35



Your answer is correct.

**Câu hỏi 10**

Đúng

Đạt điểm 0,25 trên 0,25

Trong DNS-based discovery, giả sử máy A cần tìm host của máy B, IP trung gian có vai trò?

- ☐ a. Là IP của máy B.
- ☐ b. Là IP của máy A.
- ☐ c. Hỗ trợ truy vấn host của máy A.
- ☒ d. Hỗ trợ truy vấn host của máy B.



Your answer is correct.

## Câu hỏi 11

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. Giá trị nhỏ nhất
- ☐ d. Cả 3 câu đều sai



Your answer is correct.

## Câu hỏi 12

Đúng

Đạt điểm 0,25 trên 0,25

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

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



Your answer is correct.

## Câu hỏi 13

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là bất đẳng thức chính xác nhất khi so sánh stream time và event time?

- ☐ a. Stream time < event time
- ☒ b. Stream time  $\geq$  event time
- ☐ c. Stream time  $\leq$  event time
- ☐ d. Stream time > event time



Your answer is correct.



**Câu hỏi 14**

Đúng

Đạt điểm 0,25 trên 0,25

Cho ma trận Count-min sketch như hình sau.

29	14	29	43
25	38	32	9
46	26	3	8
41	0	17	29
31	24	49	10

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

 $f_0(\text{"fit"}) = 0$  $f_1(\text{"fit"}) = 3$  $f_2(\text{"fit"}) = 2$  $f_3(\text{"fit"}) = 1$ 

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

- ☐ a. 9
- ☐ b. 26
- ☒ c. 0
- ☐ d. Không thể xác định



Your answer is correct.

**Câu hỏi 15**

Đúng

Đạt điểm 0,25 trên 0,25

Ưu thế của phương pháp caching refresh-ahead?

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



Your answer is correct.



## Câu hỏi 16

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



Your answer is correct.

## Câu hỏi 17

Đúng

Đạt điểm 0,25 trên 0,25

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

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



Your answer is correct.

## Câu hỏi 18

Đúng

Đạt điểm 0,25 trên 0,25

Trong thuật toán HyperLogLog, số lượng bit lấy ban đầu trong dãy bit dữ liệu được xác định bằng công thức nào dưới đây?

- ☐ a.  $\log_2(m^2)$
- ☒ b.  $\log_2(m)$
- ☐ c.  $\log_{10}(m^2)$
- ☐ d.  $\log_{10}(m)$



Your answer is correct.



## Câu hỏi 19

Đúng

Đạt điểm 0,25 trên 0,25

Trong RabbitMQ, để kết nối máy A đến máy B tạo thành cụm bằng cách dùng câu lệnh, các thông tin của máy B RabbitMQ cần để thực hiện gom cụm là?

- ☐ a. IP, host, cookie
- ☒ b. Tên nút, host, cookie
- ☐ c. Tên nút, cookie
- ☐ d. Tên nút, IP, cookie



Your answer is correct.

## Câu hỏi 20

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 32 cột?

- ☐ a. 7
- ☐ b. 6
- ☐ c. 8
- ☒ d. 5



Your answer is correct.

## Câu hỏi 21

Đúng

Đạt điểm 0,25 trên 0,25

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

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



Your answer is correct.



## Câu hỏi 22

Đúng

Đạt điểm 0,25 trên 0,25

Một topic trong Kafka bao gồm các hàng dữ liệu được gọi là?

- ☒ a. Partition
- ☐ b. Bin
- ☐ c. Stack
- ☐ d. Dataframe



Your answer is correct.

## Câu hỏi 23

Đúng

Đạt điểm 0,25 trên 0,25

Trong RabbitMQ, Persistent Delivery Mode và Durable lần lượt miêu tả tính chất của?

- ☐ a. Hàng đợi và broker
- ☒ b. Gói tin và hàng đợi
- ☐ c. Broker và hàng đợi
- ☐ d. Hàng đợi và gói tin



Your answer is correct.

## Câu hỏi 24

Sai

Đạt điểm 0,00 trên 0,25

Cấu hình cửa sổ trượt cho tăng 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 nằm gọn trong sự kiện.

- ☐ a. 9
- ☐ b. 14
- ☐ c. 15
- ☒ d. 6



Your answer is incorrect.

**Câu hỏi 25**

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



Your answer is correct.

**Câu hỏi 26**

Đúng

Đạt điểm 0,25 trên 0,25

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

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



Your answer is correct.

**Câu hỏi 27**

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là đặc trưng của phương pháp logging?

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



Your answer is correct.



**Câu hỏi 28**

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là mệnh đề đúng về publish/subscribe pattern?

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



Your answer is correct.

**Câu hỏi 29**

Đúng

Đạt điểm 0,25 trên 0,25

Write-back là phương pháp caching nâng cấp từ phương pháp?

- ☒ a. Tất cả đều sai
- ☐ b. Write-around
- ☐ c. Refresh-ahead
- ☐ d. Read-through



Your answer is correct.

**Câu hỏi 30**

Đúng

Đạt điểm 0,25 trên 0,25

Tác vụ có rủi ro cao cần hệ thống thời gian thực đáp ứng milli giây. Rủi ro cao ở đây là?

- ☐ a. Máy chủ quá tải
- ☐ b. Rò rỉ dữ liệu
- ☐ c. Mất mát dữ liệu
- ☒ d. Rủi ro tính mạng con người



Your answer is correct.



## Câu hỏi 31

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là chức năng của cân bằng tải?

- ☒ a. Kháng lỗi
- ☐ b. Phân tích dữ liệu
- ☐ c. Tất cả đều sai
- ☐ d. Nén dữ liệu



Your answer is correct.

## Câu hỏi 32

Đúng

Đạt điểm 0,25 trên 0,25

Một hàm băm cho ra kết quả băm với độ dài không đổi là 1 byte. 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. 256
- ☐ b. 1023
- ☐ c. 1024
- ☒ d. 255



Your answer is correct.

## Câu hỏi 33

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là lợi ích của publish/subscribe pattern?

- ☐ a. Giảm kích thước gói tin
- ☒ b. Tăng khả năng mở rộng
- ☐ c. Tất cả đều sai
- ☐ d. Giảm chi phí điện năng của máy chủ



Your answer is correct.



**Câu hỏi 34**

Đúng

Đạt điểm 0,25 trên 0,25

Trong thuật toán lấy mẫu ngẫu nhiên, người ta tính xác suất lựa chọn dữ liệu vào mẫu theo một giá trị xác suất có tính chất?

- ☐ a. Không đổi
- ☐ b. Ngẫu nhiên
- ☒ c. Giảm dần
- ☐ d. Tăng dần



Your answer is correct.

**Câu hỏi 35**

Đúng

Đạt điểm 0,25 trên 0,25

Trong thuật toán Count-min sketch (CMS), ma trận có số cột là?

- ☒ a. Giới hạn giá trị băm của hàm băm lớn nhất
- ☐ b. Số lượng hàm băm
- ☐ c. Tần suất dữ liệu
- ☐ d. Giới hạn giá trị băm



Your answer is correct.

**Câu hỏi 36**

Đúng

Đạt điểm 0,25 trên 0,25

Trong kiến trúc publish/subscribe pattern, gói tin gửi đến client là đầu ra của ...

- ☐ a. Slave
- ☐ b. Topic
- ☐ c. Producer
- ☒ d. Subscription



Your answer is correct.



## Câu hỏi 37

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



Your answer is correct.

## Câu hỏi 38

Đúng

Đạt điểm 0,25 trên 0,25

Cấu hình cửa sổ trượt cho tầng phân tích như sau: Kích thước cửa sổ là 15 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  $t + 30$  giây. Tính số cửa sổ nằm gọn trong khoảng sự kiện.

- ☐ a. 1
- ☒ b. 4
- ☐ c. 8
- ☐ d. 3



Your answer is correct.

## Câu hỏi 39

Đúng

Đạt điểm 0,25 trên 0,25

Đâu là mệnh đề đúng về publish/subscribe pattern?

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



Your answer is correct.



**Câu hỏi 40**

Đúng

Đạt điểm 0,25 trên 0,25

Xe thông minh gửi thông tin vị trí cho hệ thống để nhận về tình trạng giao thông tuyến đường gần đó và thể hiện lên màn hình dữ liệu vừa nhận. Hệ thống trên thuộc pattern nào?

- ☒ a. Request/response pattern
- ☐ b. One-way pattern
- ☐ c. Request/acknowledge pattern
- ☐ d. Publish/subscribe pattern



Your answer is correct.

**◀ Short essay exercise (group)**

Chuyển tới...