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

2. Tác vụ có rủi ro cao cần hệ thống thời gian thực đáp ứng milli giây. Rủi ro cao ở đây là? **Rủi ro tính mạng con người**

3. Đâu là một kiểu async trong request/response pattern? **Full-async**

4. Đâu là miêu tả đúng về gói tin acknowledgement trong request/acknowledgement pattern? **Chứa dữ liệu dùng trong tương lai**

5. Khác biệt giữa request/response và request/acknowledge là? **Mục đích gói tin hồi đáp**

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

7. Đâu là lợi ích của publish/subscribe pattern? **Tăng khả năng mở rộng**

8. Đâu là hạn chế của publish/subscribe pattern? **Thứ tự các gói tin không đảm bảo**

9. Đâu là mệnh đề đúng về publish/subscribe pattern? **Tất cả đều sai (1 topic chỉ có thể kết nối với 1 subscription, Bước lọc gói tin được thực hiện ở topic, Một gói tin chỉ ứng với 1 topic)**

10. Tên gọi khác của one-way pattern là? **Fire and forget**

11. Ví dụ nào dưới đây là streaming pattern? **Nghe nhạc trên Spotify**

12. Mở rộng hệ thống bằng cách thêm RAM và CPU thuộc dạng? **Mở rộng đứng**

13. 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? **Request/response pattern**

14. Đâu là chức năng của cân bằng tải? **Kháng lỗi**

15. Trạng thái toàn cục nhất quán có thể hiểu là? **Trạng thái toàn hệ thống có tính khả thi nếu không có lỗi.**

16. Hiện tượng nào sau đây thuộc nhóm lỗi byzantine? **Lỗi từ hệ quả tấn công khai thác zero-day vào một nút trong hệ thống.**

17. Hiện tượng nào sau đây thuộc nhóm lỗi response failure? **Gói tin bị mất một phần khi đến nơi.**

18. Đâu là phương pháp kháng lỗi cho tầng thu thập dữ liệu? **Logging**

19. Đâu là đặc trưng của phương pháp logging? **Gói tin bị lỗi được gửi lại.**

20. Lợi thế của phương pháp concurrent checkpointing? **Loại bỏ sự phụ thuộc vào checkpoint bất kỳ**

21. Một trạng thái toàn cục được xem là không nhất quán nếu … **Bên nhận đã nhận gói tin, nhưng bên gửi chưa gửi.**

22. Đâu là một mô hình kiến trúc logging? **SBML**

23. Trong tầng phân tích, chính sách kích hoạt dựa trên yếu tố nào? **Thời gian hoặc lượng dữ liệu**

24. 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? **Giảm dần**

25. Giá trị thống kê nào phù hợp với kỹ thuật tổng hợp lấy mẫu? **Phương sai**

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

27. Dịch vụ nào dưới đây hỗ trợ Zero-configuration networking? **Avahi**

28. Trong RabbitMQ, Persistent Delivery Mode và Durable lần lượt miêu tả tính chất của? **Gói tin và hàng đợi**

29. Hai cổng nào của RabbitMQ server được khởi chạy mặc định? **5672, 25672**

30. Thuật toán nào dưới đây có chức năng băm (hash)? **MD5**

31. Giao thức lớp vận chuyển (transport) mà RabbitMQ dùng là? **TCP**

32. Khác biệt giữa phương pháp tumbling và cửa sổ trượt trong tầng phân tích? **Cửa sổ trong tumbling không chồng lên nhau.**

33. Trong RabbitMQ, hàng đợi tạm thời được sinh ra và xóa mất trước và sau khi sử dụng? Cụ thể, nó sẽ bị xóa ngay sau khi? **Dừng kết nối consumer**.

34. Trong Spark Streaming, lớp trừu tượng thể hiện stream dữ liệu là? **DStream**

35. Dữ liệu được lưu trong thiết bị lưu trữ dài hạn được gọi là? **Data at rest**

36. Trong DNS-based discovery, giả sử máy A cần tìm host của máy B, IP trung gian có vai trò? **Hỗ trợ truy vấn host của máy B.**

37. 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? **Tên nút, host, cookie**

38. Trong thuật toán HyperLogLog, giả sử bins[i] == n và giá trị vừa tính cho bins[i] là m. Giá trị cần gán cho bins[i] là? **Max(n, m)**

39. Dữ liệu đang trong dòng dữ liệu thời gian thực được gọi là? **In-flight data**

40. Câu nào dưới đây là đúng về RabbitMQ? **Có thể tồn tại cùng lúc nhiều hàng đợi.**

41. Lựa chọn nào dưới đây KHÔNG phải là đặc điểm của phương pháp caching read-through? **Cache ở trạng thái bị động.**

42. Đâu là tên một công nghệ in-memory database (IMDB)? **Apache Geode**

43. Trong thuật toán Count-min sketch (CMS), ma trận có số cột là? **Giới hạn giá trị băm của hàm băm lớn nhất**

44. IMDG khác IMDB ở yếu tố nào? **Hỗ trợ phân tán**

45. 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? **Batch loader**

46. Write-back là phương pháp caching nâng cấp từ phương pháp? **Tất cả đều sai (Refresh-ahead, Write-around, Read-through)**

47. Nếu một gói tin không thể bị mất và được đọc một lần duy nhất bởi consumer thì đó là kiểu đảm bảo gì? **Exactly once**

48. Nếu một gói tin không thể bị mất và có thể được đọc nhiều lần bởi consumer thì đó là kiểu đảm bảo gì? **At least once**

49. Mô tả nào sau đây khớp với sender-based message logging? **Gói tin được lưu ngay trước khi gửi.**

50. Điều nào sau đây là đúng với checkpointing? **Mỗi nút có thể có số checkpoint khác nhau trong cùng khoảng thời gian.**

51. Trong logging, thuật ngữ tiếng Anh nào ám chỉ việc gửi lại gói tin? **Replay**

52. Kiểu đảm bảo nào dưới đây có hiệu năng trung bình? **At least once**

53. Một topic trong Kafka bao gồm các hàng dữ liệu được gọi là? **Partition**

54. Trong thuật toán HyperLogLog, số lượng bit lấy ban đầu trong dãy bit dữ liệu được xác định bằng công thức nào dưới đây? **log2(m)**

55. Trong phương pháp tumbling của tầng phân tích, chính sách loại bỏ là? **Khi đã đầy dữ liệu**

56. Hai kiểu phương pháp tumbling của tầng phân tích là? **Count-based, temporal**

57. Đâu là bất đẳng thức chính xác nhất khi so sánh stream time và event time? **Stream time >= event time**

58. Ví dụ nào dưới đây liên quan đến ràng buộc concept drift? **Hệ thống gợi ý không còn chính xác do mô hình bị lỗi thời.**

59. Trong tầng phân tích, công cụ thực hiện chính sách kích hoạt của phương pháp cửa sổ trượt là? **Cửa sổ trượt**

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

61. Thiết bị Universal Serial Bus thuộc kiểu nào dưới đây? **Lưu trữ dài hạn**

62. Đâu là tên của một công nghệ caching? **Hazelcast**

63. Trong phương pháp cửa sổ trượt của tầng phân tích, số cửa số nằm gọn hay nằm một phần trong tầng phân tích được gọi là? **Cửa sổ được quan tâm**

64. Ví dụ nào dưới đây liên quan đến ràng buộc one-pass? **Dữ liệu đã xử lý không thể được xử lý lại.**

65. Đâu là hạn chế của phương pháp caching write-behind? **Tăng khả năng mất mát dữ liệu**

66. Đâu là vấn đề của phương pháp incremental checkpointing? **Xung đột hash.**

67. Trong RabbitMQ, phương pháp basic QOS với tham số prefetch\_count=1 hoạt động như thế nào? **Gửi tối đa một gói tin đến worker, chỉ gửi tiếp khi nhận được ACK.**

68. Đâu là đặc trưng của phương pháp incremental checkpointing? **Deduplication**

69. Ưu thế của phương pháp caching refresh-ahead? **Cả 3 câu đều sai. (Độ tin cậy cao, Hiệu quả với dữ liệu caching lớn, Quản lý cache theo khóa - giá trị)**

70. Trong Kafka, cấu hình batch.size là của? **Producer**

71. Nếu một nút trong hệ thống bị lỗi thì … **Khôi phục toàn bộ hệ thống sao cho trạng thái toàn cục nhất quán.**

Sử dụng thuật toán Count-min sketch, cần hàm băm tối đa bao nhiêu bits để dung với ma trận n cột? **Tìm m sao cho 2^m <= n**

1 bytes = 8 bit

Một hàm băm cho ra kết quả băm với độ dài không đổi là n 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? **2^n-1**

Một hàm băm cho ra kết quả băm với độ dài không đổi là n bytes. Khi thể hiện kết quả đó dưới dạng số nhị phân thì ta sẽ có bao nhiêu chữ số? **n\*8**