

TRẮC NGHIỆM

Câu 1: Trong NameNode, lượng bộ nhớ cần thiết là

- A. Tương tự như của node chính**
- B. Phải có ít nhất một nửa node chính
- C. Phải gấp đôi node chính
- D. Chỉ phụ thuộc vào số lượng node dữ liệu mà nó sẽ xử lý

Câu 2: Khi một node dự phòng được sử dụng trong một cụm thì không cần

- A. Node kiểm tra (Check point node)
- B. Node tên phụ (Secondary name node)
- C. DataNode phụ (Secondary data node)**
- D. Nhận thức về giá đỡ (Rack awareness)

Câu 3: Nhận thức về giá trong NameNode có nghĩa là (Rack awareness in name node means)

- A. Nó biết có bao nhiêu giá đỡ có sẵn trong cụm**
- B. Nó nhận thức được ánh xạ giữa nút và giá đỡ
- C. Nó nhận biết được số lượng nút trong mỗi rack
- D. Nó biết những DataNode nào không có sẵn trong cụm.

Câu 4: Khi một máy được khai báo là datanode, dung lượng ổ đĩa trong đó (When a machine is declared as a datanode, the disk space in it)

- A. Chỉ có thể được sử dụng cho lưu trữ HDFS
- B. Có thể được sử dụng cho cả lưu trữ HDFS và không phải HDFS**
- C. Không thể truy cập bằng các lệnh không phải hadoop
- D. không thể lưu trữ các tệp văn bản.

Câu 5: Mục đích của nút checkpoint trong cụm Hadoop là (The purpose of checkpoint node in a Hadoop cluster is to)

- A. Kiểm tra xem NameNode có hoạt động không
- B. Kiểm tra xem tệp hình ảnh có đồng bộ giữa NodeName và NameNode phụ hay không**
- C. Hợp nhất hình ảnh và chỉnh sửa nhật ký và tải nó trở lại NameNode đang hoạt động.
- D. Kiểm tra xem các DataNode nào không thể truy cập được.

Câu 6: Khi một tệp trong HDFS bị người dùng xóa (When a file in HDFS is deleted by a user)

- A. nó đã mất vĩnh viễn
- B. Nó sẽ đi vào thùng rác nếu được định cấu hình.**

- C. Nó bị ẩn khỏi người dùng nhưng vẫn ở trong hệ thống tệp
- D. File sin HDFS không thể bị xóa

Câu 7: Nguồn của kiến trúc HDFS trong Hadoop có nguồn gốc là (The source of HDFS architecture in Hadoop originated as)

- A. Hệ thống tệp phân phối của Google**
- B. Hệ thống tệp phân tán của Yahoo
- C. Hệ thống tệp phân tán của Facebook
- D. Hệ thống tệp phân tán Azure

Câu 8: Giao tiếp giữa các quá trình giữa các nút khác nhau trong Hadoop sử dụng (The inter process communication between different nodes in Hadoop uses)

- A. REST API
- B. RPC**
- C. RMI
- D. IP Exchange

Câu 9: Loại dữ liệu mà Hadoop có thể xử lý là (The type of data Hadoop can deal with is)

- A. Structred (Có cấu trúc)
- B. Semi-structured (Bán cấu trúc)
- C. Unstructured (Không có cấu trúc)
- D. All of the above (Tất cả những điều trên)**

Câu 10: YARN là viết tắt của

- A. Yahoo's another resource name
- B. Yet another resource negotiator**
- C. Yahoo's archived Resource names
- D. Yet another resource need.

Câu 11: Chế độ cài đặt phân phối hoàn toàn (không ảo hóa) cần tối thiểu (The fully distributed mode of installation(without virtualization) needs a minimum of)

- A. 2 Máy vật lý**

- B. 3 Máy vật lý
- C. 4 Máy vật lý
- D. 1 Máy vật lý

Câu 12: Chạy Start-dfs.sh kết quả là

- A. Bắt đầu NameNode và DataNode**
- B. Chỉ NameNode bắt đầu
- C. Chỉ bắt đầu datanode
- D. Khởi động NameNode và trình quản lý tài nguyên

Câu 13: Điều nào sau đây không phải là mục tiêu của HDFS?

- A. Phát hiện lỗi và khôi phục
- B. Xử lý tập dữ liệu khổng lồ
- C. Ngăn chặn việc xóa dữ liệu**
- D. Cung cấp băng thông mạng cao để di chuyển dữ liệu

Câu 14: Lệnh "hadoop fs -test -z URI" cho kết quả 0 nếu

- A. nếu đường dẫn là một thư mục
- B. nếu đường dẫn là một tệp
- C. nếu đường dẫn không trống
- D. nếu tệp có độ dài bằng 0**

Câu 15: Trong HDFS, các tệp không thể

- A. Đọc
- B. Xóa
- C. Thực thi**
- D. Lưu trữ (Archived)

Câu 16: hadoop fs –expunge

- A. Cung cấp danh sách các DataNode
- B. Được sử dụng để xóa một tệp
- C. Được sử dụng để trao đổi một tệp giữa hai DataNode.
- D. Dọn sạch thùng rác.**

Câu 17: Tất cả các tệp trong một thư mục trong HDFS có thể được hợp nhất với nhau bằng cách sử dụng

A. Getmerge

B. putmerge

C. remerge

D. mergeall

Câu 18: Hệ số sao chép của tệp trong HDFS có thể được thay đổi bằng cách sử dụng

A. changerep

B. rerep

C. setrep

D. xrep

Câu 19: Dấu phẩy được sử dụng để sao chép một dạng thư mục từ node này sang node khác trong HDFS là

A. rcp

B. dcp

C. drcp

D. distcp

Câu 20: Tệp lưu trữ được tạo trong Hadoop luôn có phần mở rộng là

A. hrc

B. Har

C. Hrh

D. Hrar

Câu 21: Để hủy lưu trữ một tệp đã được lưu trữ trong haddop, hãy sử dụng lệnh

A. Unrar

B. Unhar

C. Cp

D. Cphar

Câu 22: Dữ liệu từ một cụm hadoop từ xa có thể

A. không được đọc bởi một cụm hadoop khác

B. được đọc bằng http

C. được đọc bằng hhttp

D. được đọc suing hftp

Câu 23: Mục đích của việc khởi động NameNode trong chế độ khôi phục là để

A. Khôi phục NameNode không thành công

B. Khôi phục một DataNode bị lỗi

C. Khôi phục dữ liệu từ một trong những vị trí lưu trữ siêu dữ liệu

D. Khôi phục dữ liệu khi chỉ có một vị trí lưu trữ siêu dữ liệu

Câu 24: Khi bạn tăng số lượng tệp được lưu trữ trong HDFS, Bộ nhớ được yêu cầu bởi NameNode

A. Tăng

B. Giảm

C. Vẫn không thay đổi

D. Có thể tăng hoặc giảm

Câu 25: Nếu chúng tôi tăng kích thước tệp được lưu trữ trong HDFS mà không tăng số tệp, thì bộ nhớ được yêu cầu bởi NameNode

A. Tăng

B. Giảm

C. Vẫn không thay đổi

D. Có thể tăng hoặc giảm

Câu 26: Yếu tố giới hạn hiện tại đối với kích thước của một cụm hadoop là

A. Nhiệt lượng dư thừa tạo ra trong trung tâm dữ liệu

B. Giới hạn trên của băng thông mạng

C. Giới hạn trên của RAM trong NameNode

D. 4000 datanode

Câu 27: Tính năng decommission trong hadoop được sử dụng cho

A. Hủy cấp phép NameNode

B. Hủy khai thác các DataNode

C. Hủy cấp phép NameNode phụ.

D. Giải nén toàn bộ cụm Hadoop.

Câu 28: Bạn có thể dự trữ lượng sử dụng đĩa trong một DataNode bằng cách định cấu hình dfs.datanode.du.reserved trong tệp nào sau đây

A. Hdfs-site.xml

B. Hdfs-default.xml

C. Core-site.xml

D. Mapred-site.xml

Câu 29: NameNode mất bản sao duy nhất của tệp fsimage. Chúng tôi có thể khôi phục điều này từ

A. Datanode

B. Secondary namenode

C. Checkpoint node

D. Never

Câu 30: Trong hệ thống HDFS với kích thước khối 64MB, chúng tôi lưu trữ một tệp nhỏ hơn 64MB. Điều nào sau đây là đúng?

A. Tệp sẽ tiêu tốn 64MB

B. Tệp sẽ tiêu tốn hơn 64MB

C. Tệp sẽ tiêu tốn ít hơn 64MB.

D. Không thể đoán trước được.

Câu 31: Một công việc đang chạy trong lon hadoop

A. Bị giết bằng lệnh

- B. Không bao giờ có thể bị giết bằng một lệnh
- C. Chỉ có thể bị giết bằng cách tắt NameNode
- D. Được tạm dừng và chạy lại

Câu 32: Số lượng nhiệm vụ mà trình theo dõi tác vụ có thể chấp nhận phụ thuộc vào

- A. Bộ nhớ tối đa có sẵn trong nút
- B. Không giới hạn
- C. Số lượng khe cắm được định cấu hình trong đó**
- D. Theo quyết định của jobTracker

33. Khái niệm sử dụng nhiều máy để xử lý dữ liệu được lưu trữ trong hệ thống phân tán không phải là mới.

Máy tính hiệu suất cao (HPC) sử dụng nhiều máy tính để xử lý khối lượng lớn dữ liệu được lưu trữ trong mạng vùng lưu trữ (SAN). So với HPC, Hadoop

- A. Có thể xử lý khối lượng dữ liệu lớn hơn.
- B. Có thể chạy trên một số lượng máy lớn hơn HPC cluster.
- C. Có thể xử lý dữ liệu nhanh hơn với cùng băng thông mạng so với HPC.**
- D. Không thể chạy các công việc tính toán chuyên sâu.

34. Hadoop khác với máy tính tình nguyện ở chỗ

- A. Tình nguyện viên đóng góp thời gian CPU chứ không phải băng thông mạng.**
- B. Tình nguyện viên đóng góp băng thông mạng chứ không phải thời gian CPU.
- C. Hadoop không thể tìm kiếm các số nguyên tố lớn.
- D. Chỉ Hadoop mới có thể sử dụng mapreduce.

35. So với RDBMS, Hadoop

- A. Có tính toàn vẹn dữ liệu cao hơn.
- B. Có giao dịch ACID không
- C. IS thích hợp để đọc và viết nhiều lần
- D. Hoạt động tốt hơn trên dữ liệu phi cấu trúc và bán cấu trúc.**

36. Vấn đề chính gặp phải khi đọc và ghi dữ liệu song song từ nhiều đĩa là gì?

- A - Xử lý khối lượng lớn dữ liệu nhanh hơn.
- B - Kết hợp dữ liệu từ nhiều đĩa.**
- C - Phần mềm cần thiết để thực hiện nhiệm vụ này là cực kỳ tốn kém.
- D - Phần cứng cần thiết để thực hiện tác vụ này là cực kỳ tốn kém.

37. Điều nào sau đây đúng với ổ đĩa trong một khoảng thời gian?

A - Thời gian tìm kiếm dữ liệu đang cải thiện nhanh hơn tốc độ truyền dữ liệu.

B - Thời gian tìm kiếm dữ liệu đang cải thiện chậm hơn tốc độ truyền dữ liệu.

C - Thời gian tìm kiếm dữ liệu và tốc độ truyền dữ liệu đều đang tăng tương ứng.

D - Chỉ tăng dung lượng lưu trữ mà không tăng tốc độ truyền dữ liệu.

38. Tính năng định vị dữ liệu trong Hadoop có nghĩa là

A - lưu trữ cùng một dữ liệu trên nhiều nút.

B - chuyển vị trí dữ liệu từ nút này sang nút khác.

C - đồng định vị dữ liệu với các nút tính toán.

D - Phân phối dữ liệu trên nhiều nút.

39. Cái nào trong số này cung cấp hệ thống xử lý Luồng được sử dụng trong hệ sinh thái Hadoop?

A - Solr

B - Tez

C - Tia lửa

D - Hive

40. Các tệp HDFS được thiết kế cho

A - Nhiều người viết và sửa đổi ở các hiệu số tùy ý.

B - Chỉ nối vào cuối tệp

C - Chỉ ghi thành tệp một lần.

D - Truy cập dữ liệu có độ trễ thấp.

41. Tệp HDFS nhỏ hơn kích thước một khối

A - Không thể lưu trữ trong HDFS.

B - Chiếm toàn bộ kích thước của khối.

C - Chỉ chiếm kích thước mà nó cần chứ không phải toàn khối.

D - Có thể trải dài trên nhiều khối.

42. Kích thước khối HDFS lớn hơn so với kích thước của các khối đĩa để

A - Chỉ các tệp HDFS có thể được lưu trữ trong đĩa được sử dụng.

B - Thời gian tìm kiếm là tối đa

C - Không thể chuyển một tệp lớn được tạo từ nhiều khối đĩa.

D - Một tệp duy nhất lớn hơn kích thước đĩa có thể được lưu trữ trên nhiều đĩa trong cụm.

43. Trong một cụm Hadoop, điều gì đúng với khối HDFS không còn khả dụng do hỏng đĩa hoặc lỗi máy?

A - Nó bị mất vĩnh viễn

B - Nó có thể được sao chép ở các vị trí thay thế của nó cho các máy sống khác.

C - NameNode cho phép yêu cầu của khách hàng mới tiếp tục cố gắng đọc nó.

D - Tiến trình công việc Mapreduce chạy bỏ qua khối và dữ liệu được lưu trữ trong đó.

44. Tiện ích nào được sử dụng để kiểm tra tình trạng của hệ thống tệp HDFS?

A - fchk

B - fsck

C - fsch

D - fcks

45. Lệnh nào liệt kê các khối tạo nên mỗi tệp trong hệ thống tệp.

A - hdfs fsck / -files -blocks

B - hdfs fsck / -blocks -files

C - hdfs fchk / -blocks -files

D - hdfs fchk / -files -block

46. DataNode và NameNode là tương ứng

A - Nút chính và nút công nhân

B - Nút công nhân và nút chính

C - Cả hai đều là các nút công nhân

D - Không có

47. Trong đĩa cục bộ của NameNode, các tệp được lưu trữ liên tục là:

A - hình ảnh vùng tên và nhật ký chỉnh sửa

B - vị trí khối và hình ảnh vùng tên

C - chỉnh sửa nhật ký và chặn vị trí

D - Hình ảnh không gian tên, chỉnh sửa vị trí nhật ký và chặn.

48. Khi khách hàng giao tiếp với hệ thống tệp HDFS, nó cần giao tiếp với

A - chỉ NameNode

B - chỉ DataNode

C - cả NameNode và DataNode

D - Không có

49. Hadoop sử dụng những cơ chế nào để làm cho namenode có khả năng chống lại sự cố.

A - Sao lưu siêu dữ liệu hệ thống tệp vào đĩa cục bộ và gắn kết NFS từ xa.

B - Lưu trữ siêu dữ liệu hệ thống tệp trên đám mây.

C - Sử dụng máy có ít nhất 12 CPU

D - Sử dụng phần cứng đắt tiền và đáng tin cậy.

50. Vai trò chính của NameNode phụ là

A - Sao chép siêu dữ liệu hệ thống tệp từ NameNode chính.

B - Sao chép siêu dữ liệu hệ thống tệp từ NFS được lưu trữ bởi NameNode chính

C - Theo dõi xem NameNode chính có đang hoạt động hay không.

D - Định kỳ hợp nhất hình ảnh vùng tên với nhật ký chỉnh sửa.

51. Đối với các tệp HDFS được truy cập thường xuyên, các khối được lưu vào bộ nhớ đệm

A - bộ nhớ của DataNode

B - trong bộ nhớ của NameNode

C - Cả A&B

D - Trong bộ nhớ của ứng dụng khách đã yêu cầu quyền truy cập vào các tệp này.

52. Các ứng dụng người dùng có thể hướng dẫn NameNode để lưu vào bộ đệm các tệp bằng cách

A - thêm tên tệp bộ đệm vào nhóm bộ đệm

B - thêm cấu hình bộ đệm vào nhóm bộ đệm

C - thêm chỉ thị bộ nhớ cache vào nhóm bộ nhớ cache

D - chuyển tên tệp làm tham số cho nhóm bộ nhớ cache

53. Trong Hadoop 2.x, liên kết HDFS phát hành có nghĩa là

A - Cho phép các NameNode giao tiếp với nhau.

B - Cho phép một cụm mở rộng quy mô bằng cách thêm nhiều DataNode dưới một NameNode.

C - Cho phép một cụm mở rộng quy mô bằng cách thêm nhiều NameNode hơn.

D - Thêm nhiều bộ nhớ vật lý hơn cho cả NameNode và DataNode.

54. Theo liên kết HDFS

A - Mỗi NameNode quản lý siêu dữ liệu của toàn bộ hệ thống tệp.

B - Mỗi NameNode quản lý siêu dữ liệu của một phần hệ thống tệp.

C - Lỗi một NameNode làm mất một số tính khả dụng của siêu dữ liệu từ toàn bộ hệ thống tệp.

D - Mỗi DataNode đăng ký với mỗi NameNode.

55. Mục tiêu chính của HDFS Tính sẵn sàng cao là

A - Tạo bản sao của NameNode chính nhanh hơn.

B - Để giảm thời gian chu kỳ cần thiết để khôi phục lại NameNode chính mới sau khi nút chính hiện có bị lỗi.

C - Ngăn chặn việc mất dữ liệu do lỗi của NameNode chính.

D - Ngăn chặn biểu mẫu tên chính trở thành điểm lỗi duy nhất.

56. Là một phần của tính khả dụng cao HDFS, một cặp NameNode chính được cấu hình. Điều gì là đúng với họ?

A - Khi một yêu cầu của khách hàng đến, một trong số họ được chọn ngẫu nhiên sẽ phục vụ yêu cầu đó.

B - Một trong số chúng đang hoạt động trong khi cái còn lại vẫn tắt.

C - Các DataNode chỉ gửi báo cáo khối đến một trong các NameNode.

D - Nút chờ nhận các điểm kiểm tra định kỳ của không gian tên của NameNode đang hoạt động.

57. Zookeeper đảm bảo rằng

A - Tất cả các NameNode đang tích cực phục vụ các yêu cầu của khách hàng

B - Chỉ có một NameNode đang tích cực phục vụ các yêu cầu của khách hàng

C - Chuyển đổi dự phòng được kích hoạt khi bất kỳ DataNode nào bị lỗi.

D - Quản trị viên hadoop không thể bắt đầu chuyển đổi dự phòng.

58. Theo Tính khả dụng cao của Hadoop, nghĩa là Hàng rào

A - Ngăn NameNode hoạt động trước đó bắt đầu chạy lại.

B - Ngăn chặn việc bắt đầu chuyển đổi dự phòng trong trường hợp mạng bị lỗi với NameNode hoạt động.

C - Ngăn chặn sự cố sập nguồn đối với NameNode đã hoạt động trước đó.

D - Ngăn không cho NameNode đã hoạt động trước đó ghi vào nhật ký chỉnh sửa.

59. Cơ chế nào sau đây không phải là cơ chế hàng rào cho NameNode đã hoạt động trước đó?

A - Tắt cổng mạng của nó thông qua lệnh quản lý từ xa.

B - Thu hồi quyền truy cập của nó vào thư mục lưu trữ được chia sẻ.

C - Định dạng ổ đĩa của nó.

D - STONITH

60. Thuộc tính được sử dụng để đặt hệ thống tệp mặc định cho Hadoop trong core-site.xml là-

A - filesystem.default

B - fs.default

C - fs.defaultFS

D - hdfs.default

61. Yếu tố sao chép mặc định cho hệ thống tệp HDFS trong hadoop là

A - 1

B - 2

C - 3

D - 4

62. Khi chạy trên chế độ phân phối giả lập, hệ số sao chép được đặt thành

A - 2

B - 1

C - 0

D - 3

63. Đối với thư mục HDFS, hệ số sao chép (RF) là

A - giống như RF của các tệp trong thư mục đó

B - Không

C - 3

D - Không áp dụng.

64. Điều sau không được phép trên các tệp HDFS

A - Xóa

B - Đổi tên

C - Di chuyển

D - Đang thực hiện.

65. HDFS có thể được truy cập qua HTTP bằng cách sử dụng

A - lược đồ URI viewfs

B - lược đồ URI webhdfs

Lược đồ URI C - wasb

D - HDFS ftp

66. Điều gì là đúng về HDFS?

A - Hệ thống tệp HDFS có thể được gắn trên Hệ thống tệp của máy khách cục bộ bằng NFS.

B - Hệ thống tệp HDFS không bao giờ có thể được gắn vào Hệ thống tệp của máy khách cục bộ.

C - Bạn có thể chỉnh sửa bản ghi hiện có trong tệp HDFS đã được gắn kết bằng NFS.

D - Bạn không thể thêm vào tệp HDFS được gắn bằng NFS.

67. Máy khách đọc dữ liệu từ hệ thống tệp HDFS trong Hadoop

A - lấy dữ liệu từ NameNode

B - lấy vị trí khối từ datanode

C - chỉ lấy các vị trí khối tạo thành NameNode

D - lấy cả dữ liệu và vị trí khối từ NameNode

68. Kịch bản nào yêu cầu băng thông cao nhất để truyền dữ liệu giữa các nút trong Hadoop?

A - Các nút khác nhau trên cùng một giá đỡ

B - Các nút trên các giá đỡ khác nhau trong cùng một trung tâm dữ liệu.

C - Các mã trong các trung tâm dữ liệu khác nhau

D - Dữ liệu trên cùng một nút.

69. Vị trí khối hiện tại của HDFS nơi dữ liệu đang được ghi vào,

A - hiển thị cho khách hàng yêu cầu nó.

B - Vị trí khối không bao giờ hiển thị đối với các yêu cầu của khách hàng.

C - Người đọc có thể nhìn thấy hoặc không.

D - chỉ hiển thị sau khi dữ liệu được lưu trong bộ đệm được cam kết.

70. Tùy chọn nào trong số này không phải là tùy chọn lập lịch có sẵn với YARN?

A - Bộ lập lịch tối ưu

B - Bộ lập lịch FIFO

C - Bộ lập lịch dung lượng

D - Bộ lập lịch công bằng

71. Chế độ nào sau đây không phải là chế độ hoạt động của Hadoop?

A - Chế độ phân phối giả

B - Chế độ phân phối toàn cầu

C - Chế độ độc lập

D - Chế độ phân phối hoàn toàn

72. Sự khác biệt giữa chế độ độc lập và giả phân phối là

A - Đứng một mình không thể sử dụng bản đồ giảm

B - Đứng riêng có một quy trình java duy nhất chạy trong đó.

C - Chế độ phân tán giả không sử dụng HDFS

D - Chế độ phân tán giả cần hai hoặc nhiều máy vật lý.

73. Công việc khung hadoop được viết bằng

A - C ++

B - Python

C - Java

D – Go

74. Lệnh hdfs để tạo bản sao của tệp từ hệ thống cục bộ là

A - CopyFromLocal

B - copyfromlocal

C - CopyLocal

D - copyFromLocal

75. Lệnh hadfs được sử dụng để

A - Sao chép tệp từ hệ thống tệp cục bộ sang HDFS.

B - Sao chép tệp hoặc thư mục từ hệ thống tệp cục bộ sang HDFS.

C - Sao chép các tệp từ HDFS sang hệ thống tệp cục bộ.

D - Sao chép tệp hoặc thư mục từ HDFS sang hệ thống tệp cục bộ.

76. Sao chép thiếu trong HDFS có nghĩa là-

A - Không có sự sao chép nào diễn ra trong các DataNode.

B - Quá trình sao chép rất chậm trong các DataNode.

C - Tần suất sao chép trong các DataNode rất thấp.

D - Số lượng bản sao được tái tạo ít hơn so với quy định của hệ số sao chép.

77. Khi NameNode nhận thấy rằng một số khối được sao chép quá mức, nó

A - Dừng công việc sao chép trong toàn bộ hệ thống tệp hdfs.

B - Nó làm chậm quá trình nhân bản cho các khối đó

C - Nó xóa các khối thừa.

D - Nó để lại các khối thừa như nó vốn có.

78. Thuộc tính nào dưới đây được định cấu hình trên core-site.xml?

A - Yếu tố nhân rộng

B - Tên thư mục để lưu trữ tệp hdfs.

C - Máy chủ và cổng nơi tác vụ MapReduce chạy.

D - Các biến môi trường Java.

79. Thuộc tính nào dưới đây được định cấu hình trên hdfs-site.xml?

A - Yếu tố nhân rộng

B - Tên thư mục để lưu trữ tệp hdfs.

C - Máy chủ và cổng nơi tác vụ MapReduce chạy.

D - Các biến môi trường Java.

80. Thuộc tính nào dưới đây được định cấu hình trên mapred-site.xml?

A - Yếu tố nhân rộng

B - Tên thư mục để lưu trữ tệp hdfs.

C - Máy chủ và cổng nơi tác vụ MapReduce chạy.

D - Các biến môi trường Java.

81. Thuộc tính nào dưới đây được định cấu hình trên hadoop-env.sh?

A - Yếu tố nhân rộng

B - Tên thư mục để lưu trữ tệp hdfs

C - Máy chủ và cổng nơi tác vụ MapReduce chạy

D - Các biến môi trường Java.

82. Lệnh để kiểm tra xem Hadoop có hoạt động hay không là:

A - Jsp

B - Jps

C - Hadoop fs -test

D - Không có

83. Các khối dữ liệu ánh xạ thông tin với các tệp tương ứng của chúng được lưu trữ trong

A - DataNode

B - Trình theo dõi công việc

C - Trình theo dõi tác vụ

D - NameNode

84. Tệp trong Namenode lưu trữ thông tin ánh xạ vị trí khối dữ liệu với tên tệp là

A - dfsimage

B - nameimage

C - fsimage

D - image

85. NameNode biết rằng DataNode đang hoạt động bằng cách sử dụng một cơ chế được gọi là

A - heartbeats

B - datapulse

C - h-signal

D - Active-pulse

86. Bản chất của phần cứng cho NameNode phải là

A - Cao cấp hơn loại hàng hóa

B - Loại hàng hóa

C - Không thành vấn đề

D - Chỉ cần có nhiều Ram hơn mỗi DataNode

87. Trong Hadoop, Snappy và LZ4 là những ví dụ về

A - Cơ chế vận chuyển tệp giữa các DataNode

B - Cơ chế nén dữ liệu

C - Cơ chế sao chép dữ liệu

D - Cơ chế đồng bộ hóa dữ liệu

88. Hệ thống apache nào dưới đây giải quyết việc nhập dữ liệu phát trực tuyến vào hadoop

A - Ozie

B - Kafka

C - Flume

D - Hive

89. Sự phân chia đầu vào được sử dụng trong MapReduce cho biết

A - Kích thước trung bình của các khối dữ liệu được sử dụng làm đầu vào cho chương trình

B - Chi tiết vị trí nơi bắt đầu của toàn bộ bản ghi đầu tiên trong một khối và toàn bộ bản ghi cuối cùng trong khối kết thúc.

C - Tách dữ liệu đầu vào cho chương trình MapReduce thành kích thước đã được định cấu hình trong mapred-site.xml

D - Không có

90. Đầu ra của một nhiệm vụ ánh xạ là

A - Cặp khóa-giá trị của tất cả các bản ghi của tập dữ liệu.

B - Cặp khóa-giá trị của tất cả các bản ghi từ phân tách đầu vào được trình ánh xạ xử lý

C - Chỉ các phím được sắp xếp từ phân tách đầu vào

D - Số hàng được xử lý bởi tác vụ ánh xạ.

91. Vai trò của nút Nhặt ký là

A - Báo cáo vị trí của các khối trong một DataNode

B - Báo cáo thông tin nhặt ký chỉnh sửa của các khối trong DataNode.

C - Báo cáo lịch trình khi công việc sẽ chạy

D - Báo cáo hoạt động của các thành phần khác nhau do người quản lý tài nguyên xử lý

92. Người giữ vườn thú (zookeeper)

A - Phát hiện lỗi của NameNode và chọn NameNode mới.

B - Phát hiện lỗi của các DataNode và chọn một DataNode mới.

C - Ngăn phần cứng quá nóng bằng cách tắt chúng.

D - Duy trì danh sách tất cả các thành phần địa chỉ IP của cụm Hadoop.

93. Nếu địa chỉ IP hoặc tên máy chủ của DataNode thay đổi

A - NameNode cập nhật ánh xạ giữa tên tệp và tên khối

B - NameNode không cần cập nhật ánh xạ giữa tên tệp và tên khối

C - Dữ liệu trong DataNode đó sẽ bị mất vĩnh viễn

D - Có NameNode phải được khởi động lại

94. Khi một ứng dụng khách liên hệ với NameNode để truy cập tệp, NameNode phản hồi với

A - Kích thước của tệp được yêu cầu.

B - ID khối của tệp được yêu cầu.

C - ID khối và tên máy chủ của bất kỳ DataNode nào chứa khối đó.

D - Block ID và tên máy chủ của tất cả các DataNode chứa khối đó.

95. HDFS là viết tắt của

A - Hệ thống tệp phân tán cao. ([Highly distributed file system.](#))

B - Hệ thống tệp được hướng dẫn Hadoop ([Hadoop directed file system](#))

C - Vỏ tệp phân tán cao ([Highly distributed file shell](#))

D - Hệ thống tệp phân tán Hadoop. ([Hadoop distributed file system.](#))

96. Công cụ Hadoop được sử dụng để phân tán dữ liệu một cách đồng nhất trên các DataNode được đặt tên là:

A - Bộ lập lịch

B - Cân bằng

C - Máy rải

D - PV

97. Khi một jobTracker lên lịch, một công việc sẽ được tìm kiếm đầu tiên

A - Một nút có vị trí trống trong cùng giá đỡ với DataNode

B - Bất kỳ nút nào trên cùng giá đỡ với DataNode

C - Bất kỳ nút nào trên rack liền kề với rack của datanode

D - Chỉ bất kỳ nút nào trong cụm

98. Tín hiệu nhip tim được gửi từ

A - JObtracker thành Tasktracker

B - Tasktracker to Job tracker

C - Trình theo dõi công việc đến NameNode

D - Trình theo dõi tác vụ đến NameNode

99. Job tracker runs on

A - Namenode

B - Datanode

C - Secondary namenode

D - Secondary datanode

100. Tùy chọn nào sau đây không phải là tùy chọn lập lịch có sẵn trong YARN

A - Bộ lập lịch cân bằng

B - Lập lịch trình công bằng

C - Bộ lập lịch dung lượng

D - Máy quét FiFO.

101. Định dạng đầu vào mặc định là gì?

A - Định dạng đầu vào mặc định là xml. Nhà phát triển có thể chỉ định các định dạng đầu vào khác nếu thích hợp nếu xml không phải là đầu vào chính xác.

B - Không có định dạng nhập mặc định. Định dạng đầu vào luôn phải được chỉ định.

C - Định dạng đầu vào mặc định là định dạng tệp tuần tự. Dữ liệu cần được xử lý trước trước khi sử dụng định dạng đầu vào mặc định.

D - Định dạng đầu vào mặc định là TextInputFormat với phân bù byte làm khóa và toàn bộ dòng dưới dạng giá trị.

102. Cái nào không phải là một trong những tính năng dữ liệu lớn?

A - Vận tốc [Velocity](#)

B - Tính xác thực [Veracity](#)

C - âm lượng [volume](#)

D - đa dạng [variety](#)

103. Công nghệ nào được sử dụng để lưu trữ dữ liệu trong Hadoop?

[A - HBase](#)

[B - Avro](#)

[C - Sqoop](#)

[D - Zookeeper](#)

104. Công nghệ nào được sử dụng để tuần tự hóa dữ liệu trong Hadoop?

[A - HBase](#)

[B - Avro](#)

[C - Sqoop](#)

[D - Zookeeper](#)

105. Công nghệ nào được sử dụng để nhập và xuất dữ liệu trong Hadoop?

A - HBase

B - Avro

C - Sqoop

D - Người giữ vườn thú

106. Công nghệ nào sau đây là cơ sở dữ liệu lưu trữ tài liệu?

A - HBase

B - Hive

C - Cassandra

D - CouchDB

107. Điều nào sau đây không đúng đối với Hadoop?

A - Đây là một khung phân tán.

B - Thuật toán chính được sử dụng trong đó là Map Reduce

C - Nó chạy với đồ cứng hàng hóa

D - Tất cả đều đúng

108. Cái nào sau đây lưu trữ dữ liệu?

A - Name node

B - Data node

C - Master node

D - None of these

109. Nút nào sau đây quản lý các nút khác?

A - Name node

B - Data node

C - slave node

D - None of these

110. AVRO là gì?

A - Avro là một thư viện tuần tự hóa java.

B - Avro là một thư viện nén java.

C - Avro là một thư viện java tạo các tệp bảng chia nhỏ.

D - Không câu trả lời nào đúng.

111. Bạn có thể chạy Bản đồ - Giảm công việc trực tiếp trên dữ liệu Avro không?

A - Có, Avro được thiết kế đặc biệt để xử lý dữ liệu qua Map-Reduce.

B - Có, nhưng cần có mã hóa mở rộng bổ sung.

C - Không, Avro được thiết kế đặc biệt chỉ để lưu trữ dữ liệu.

D - Avro chỉ định siêu dữ liệu cho phép truy cập dữ liệu dễ dàng hơn. Dữ liệu này không thể được sử dụng như một phần của quá trình thực thi thu nhỏ bản đồ, thay vì chỉ đặc tả đầu vào.

112. Bộ nhớ đệm phân tán là gì?

A - Bộ đệm phân tán là thành phần đặc biệt trên NameNode sẽ lưu vào bộ đệm dữ liệu được sử dụng thường xuyên để phản hồi máy khách nhanh hơn. Nó được sử dụng trong bước giảm.

B - Bộ nhớ đệm phân tán là thành phần đặc biệt trên DataNode sẽ lưu vào bộ đệm dữ liệu được sử dụng thường xuyên để phản hồi máy khách nhanh hơn. Nó được sử dụng trong bước bản đồ.

C - Bộ đệm phân tán là một thành phần lưu trữ các đối tượng java.

D - Bộ nhớ đệm phân tán là một thành phần cho phép các nhà phát triển triển khai các chum để xử lý Map-Reduce.

113. Có thể ghi là gì?

A - W ghi là một giao diện java cần được triển khai để truyền dữ liệu trực tuyến đến các máy chủ từ xa.

B - W ghi là một giao diện java cần được thực hiện để ghi HDFS.

C - Writes là một giao diện java cần được triển khai để xử lý MapReduce.

D - Không câu trả lời nào đúng.

114. HBASE là gì?

A - Hbase là bộ Java API riêng biệt cho cụm Hadoop.

B - Hbase là một phần của dự án Apache Hadoop cung cấp giao diện để quét một lượng lớn dữ liệu bằng cơ sở hạ tầng Hadoop.

C - Hbase là một "cơ sở dữ liệu" giống như giao diện với dữ liệu cụm Hadoop.

D - HBase là một phần của dự án Apache Hadoop cung cấp giao diện giống SQL để xử lý dữ liệu.

115. Hadoop xử lý khối lượng lớn dữ liệu như thế nào?

A - Hadoop sử dụng song song rất nhiều máy. Điều này tối ưu hóa việc xử lý dữ liệu.

B - Hadoop được thiết kế đặc biệt để xử lý lượng lớn dữ liệu bằng cách tận dụng phân cứng MPP.

C - Hadoop gửi mã đến dữ liệu thay vì gửi dữ liệu đến mã.

D - Hadoop sử dụng các kỹ thuật bộ nhớ đệm phức tạp trên NameNode để tăng tốc độ xử lý dữ liệu.

116. Khi sử dụng HDFS, điều gì xảy ra khi tệp bị xóa khỏi dòng lệnh?

A - Nó sẽ bị xóa vĩnh viễn nếu thùng rác được bật.

B - Nó được đặt vào một thư mục thùng rác chung cho tất cả người dùng cho cụm đó.

C - Nó bị xóa vĩnh viễn và các thuộc tính tệp được ghi lại trong tệp nhật ký.

D - Nó được chuyển vào thư mục thùng rác của người dùng đã xóa nó nếu thùng rác được bật.

117. Khi lưu trữ tệp Hadoop, câu nào sau đây là đúng? (Chọn hai câu trả lời)

Các tệp đã lưu trữ sẽ hiển thị với phần mở rộng .arc.

Nhiều tệp nhỏ sẽ trở thành ít tệp lớn hơn.

MapReduce xử lý tên tệp gốc ngay cả sau khi tệp được lưu trữ.

Các tệp đã lưu trữ phải được lưu trữ tại Liên hợp quốc cho HDFS và MapReduce để truy cập các tệp nhỏ, gốc.

Lưu trữ dành cho các tệp cần được lưu nhưng HDFS không còn truy cập được nữa.

A - 1 & 3

B - 2 & 3

C - 2 & 4

D - 3 & 4

118. Khi ghi dữ liệu vào HDFS điều gì là đúng nếu hệ số sao chép là ba? (Chọn 2 câu trả lời)

Dữ liệu được ghi vào DataNodes trên ba giá đỡ riêng biệt (nếu Rack Aware).

Dữ liệu được lưu trữ trên mỗi Mã dữ liệu bằng một tệp riêng biệt chứa giá trị tổng kiểm tra.

Dữ liệu được ghi vào các khối trên ba DataNodes khác nhau.

Khách hàng được trả lại thành công khi ghi thành công khối đầu tiên và kiểm tra tổng kiểm tra.

A - 1 & 3

B - 2 & 3

C - 3 & 4

D - 1 & 4

119. Nhiệm vụ nào sau đây là trong số các nhiệm vụ của các DataNode trong HDFS?

A - Duy trì cây hệ thống tệp và siêu dữ liệu cho tất cả các tệp và thư mục.

B - Không có phương án nào đúng.

C - Kiểm soát việc thực hiện một tác vụ bản đồ riêng lẻ hoặc một tác vụ thu gọn.

D - Lưu trữ và truy xuất các khối khi được khách hàng hoặc NameNode yêu cầu.

E - Quản lý không gian tên hệ thống tệp.

120. Thành phần nào sau đây truy xuất các phân tách đầu vào trực tiếp từ HDFS để xác định số tác vụ bản đồ?

A - Mã Tên.

B - TaskTrackers.

C - JobClient.

D - JobTracker.

E - Không có lựa chọn nào đúng.

121. Giao diện org.apache.hadoop.io.Writable khai báo hai phương thức nào? (Chọn 2 câu trả lời.)

public void readFields (DataInput).

công khai khoảng trống đọc (DataInput).

public void writeFields (DataOutput).

ghi khoảng trống công khai (DataOutput).

A - 1 & 4

B - 2 & 3

C - 3 & 4

D - 2 & 4

122. Câu nào sau đây là đúng đối với các cặp <key, value> của một công việc MapReduce?

A - Một lớp khóa phải triển khai Writable.

B - Một lớp khóa phải triển khai WritableComparable.

C - Một lớp giá trị phải triển khai WritableComparable.

D - Một lớp giá trị phải mở rộng khả năng so sánh được.

123. Phát biểu nào sau đây là sai về Bộ nhớ đệm phân tán?

A - Khung công tác Hadoop sẽ đảm bảo rằng bất kỳ tệp nào trong Bộ đệm phân tán được phân phối cho tất cả các tác vụ bản đồ và giảm bớt.

B - Các tệp trong bộ đệm có thể là tệp văn bản hoặc chúng có thể là tệp lưu trữ như tệp zip và JAR.

C - Disk I / O bị tránh vì dữ liệu trong bộ đệm được lưu trong bộ nhớ.

D - Khung công tác Hadoop sẽ sao chép các tệp trong Bộ đệm ẩn phân tán vào nút phụ trước khi bất kỳ nhiệm vụ nào cho công việc được thực thi trên nút đó.

124. Thành phần nào sau đây không phải là thành phần chính của HBase?

A - Máy chủ Vùng.

B - Nagios.

C - ZooKeeper.

D - Máy chủ chính.

125. Điều nào sau đây là sai về RawComparator?

A - So sánh các khóa theo byte.

B - Hiệu suất có thể được cải thiện trong giai đoạn sắp xếp và đủ bằng cách sử dụng RawComparator.

C - Các khóa trung gian được giải mã hóa để thực hiện so sánh.

126. Con quỷ (demon) nào chịu trách nhiệm sao chép dữ liệu trong Hadoop?

A - HDFS.

B - Trình theo dõi tác vụ.

C - Trình theo dõi công việc.

D - NameNode.

E - DataNode.

127. Các phép từ đầu ra của xáo trộn và sắp xếp thực hiện giao diện nào sau đây?

A - Viết được.

B - Có thể so sánh được.

C - Có thể cấu hình.

D - Có thể so sánh được.

E - Có thể so sánh được.

128. Để áp dụng một bộ kết hợp, một thuộc tính phải được thỏa mãn bởi các giá trị được phát ra từ bộ ánh xạ là gì?

A - Combiner luôn có thể được áp dụng cho mọi dữ liệu

B - Đầu ra của bộ ánh xạ và đầu ra của bộ kết hợp phải cùng một cặp giá trị khóa và chúng có thể không đồng nhất

C - Đầu ra của bộ ánh xạ và đầu ra của bộ kết hợp phải cùng một cặp giá trị khóa. Chỉ khi các giá trị thỏa mãn thuộc tính liên kết và giao hoán thì nó mới có thể được thực hiện.