Rebalancing HDFS Cluster

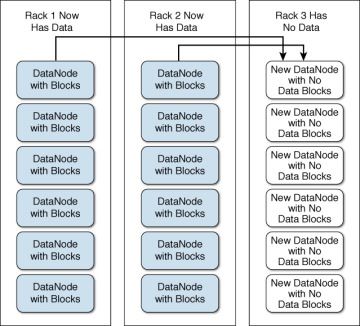
[ĐỨC ANH LÊ](https://www.facebook.com/duc.anh.lelele)·[THỨ SÁU, 6 THÁNG 3, 2020](https://www.facebook.com/notes/c%E1%BB%99ng-%C4%91%E1%BB%93ng-big-data-vi%E1%BB%87t-nam/rebalancing-hdfs-cluster/488770332001822/)·

Hadoop là hệ thống lưu trữ dữ liệu lớn phân tán, phân tán nghĩa là dữ liệu sẽ được lưu trữ trên nhiều máy tính khác nhau. Chính vì điều này mà Hadoop có khả năng mở rộng theo chiều ngang rất tốt, giả sử muốn tăng khả năng lưu trữ của Hadoop, bạn chỉ cần thêm một vài DataNode vào cụm đó. Lợi ích là vậy nhưng điều này cũng vô tình gây ra sự mất cân bằng về dữ liệu giữa các DataNode. Rõ ràng những DataNode mới thêm vào lúc đầu sẽ không có dữ liệu trong khi các DataNode cũ thì phải lưu trữ quá nhiều, việc đọc dữ liệu cũng bận rộn hơn. Hadoop không tự động di chuyển dữ liệu giữa các DataNode để cân bằng, nếu bạn sử dụng Ambari để quản lý cụm thì sẽ nhận được cảnh báo như sau:



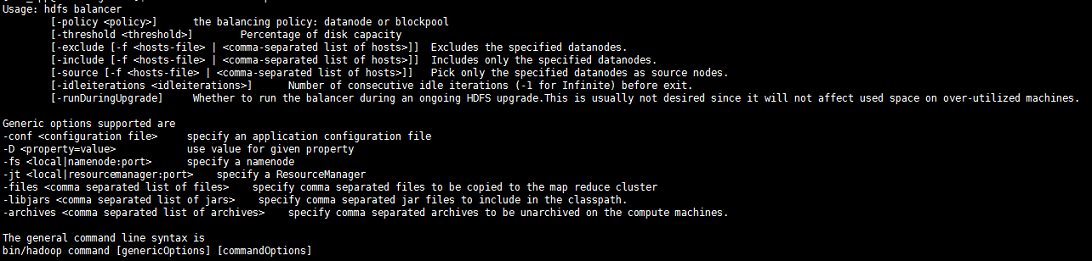
Cảnh báo dung lượng DataNode bị đầy trên Ambari

Hadoop cung cấp công cụ balancer cho phép cân bằng dữ liệu giữa các DataNode, chuyển block từ node đang bị chiếm dụng nhiều sang node chiếm dụng ít. Hình dưới mô tả việc rebalance trong cụm, với Rack3 ban đầu không có dữ liệu và Rack1 Rack2 với dữ liệu bị chiếm nhiều. Các block từ Rack1 và Rack2 sẽ được chuyển sang Rack3 tới một ngưỡng threshold xác định.



Mô tả cách thức dữ liệu balancer

Hadoop balancer có thể được chạy bằng lệnh balancer. Đây là cách sử dụng của lệnh này:



Cách dùng lệnh hdfs balancer

Bởi vì các DataNode có thể có cấu hình khác nhau, một node có thể lưu được nhiều hơn node khác, có node lưu được vài trăm TB, node khác chỉ có vài chục TB. Hadoop dùng 1 tham số threshold tính theo phần trăm dung lượng của node để coi cụm là cân bằng.

Tham số threshold biểu thị phần trăm sai lệch của DataNode so với trung bình toàn cụm. Ví dụ toàn cụm đang có phần trăm cân bằng là 70% với threshold = 10%, một node sẽ coi là cân bằng nếu dung lượng chiếm dụng của nó nằm trong khoảng 60 – 80%. Giá trị mặc định của threshold là 10%.

Thời gian chạy balancer của cụm là nhanh hay chậm phụ thuộc vào kích cỡ của cụm và độ chêch lệch dữ liệu. Nếu bạn mới chạy balance cho cụm lần đầu hoặc sau khi mới thêm nhiều DataNode, thời gian chạy sẽ rất lâu để dữ liệu có thể cân bằng giữa các node, có thể vài ngày.

Khi chạy balancer có 2 khái niệm cần chú ý:

⦁ Average DFS Used: Phần trăm dung lượng sử dụng của toàn cụm .

Average DFS Used = DFS Used \* 100 / Present Capacity

⦁ Phần trăm sử dụng DataNode: Phần trăm dung lượng sử dụng của DataNode.

Balancer sẽ chạy nếu dung lượng sử dụng của DataNode so với dung lượng sử dụng toàn cụm lớn hơn hoặc nhỏ hơn giá trị threshold. Nếu không nằm trong khoảng này, balancer sẽ không chạy trên node đó.

Chạy balancer trên toàn cụm sẽ ảnh hưởng đến các công việc đang chạy trên cụm, để tránh điều này bạn nên:

⦁ Lựa chọn ra một vài DataNode có dung lượng lớn để cân bằng với một vài DataNode có dung lượng nhỏ hơn, tránh ảnh hưởng đến toàn cụm.

⦁ Tinh chỉnh tham số như băng thông, số luồng chạy cũng như dung lượng block chạy trong mỗi lần chuyển dữ liệu, từ đó tăng tốc độ balance.

⦁ Chạy balancer thường xuyên trong cụm, tránh để tình trạng chêch lệch dữ liệu quá lớn giữa các DataNode.

⦁ Bắt đầu chạy balancer với threshold lớn, sau đó giảm dần để tránh gây cao tải cho toàn cụm.

Link bài 2: <https://www.facebook.com/notes/c%E1%BB%99ng-%C4%91%E1%BB%93ng-big-data-vi%E1%BB%87t-nam/how-to-optimize-hdfs-balancer/492400398305482/>

[#ApacheHadoop](https://www.facebook.com/hashtag/apachehadoop?source=note&epa=HASHTAG) [#Rebalance](https://www.facebook.com/hashtag/rebalance?source=note&epa=HASHTAG) [#Hadoop](https://www.facebook.com/hashtag/hadoop?source=note&epa=HASHTAG) [#Optimize](https://www.facebook.com/hashtag/optimize?source=note&epa=HASHTAG) [#HDFS](https://www.facebook.com/hashtag/hdfs?source=note&epa=HASHTAG)