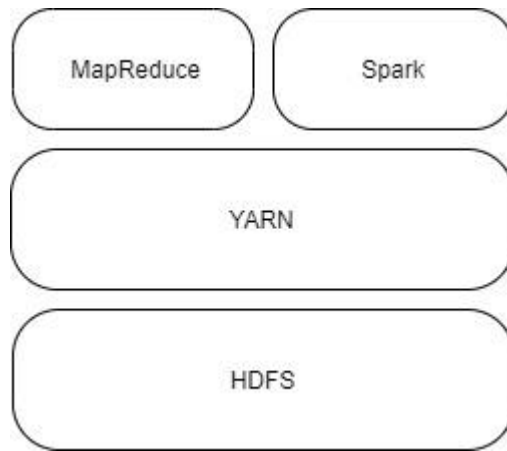
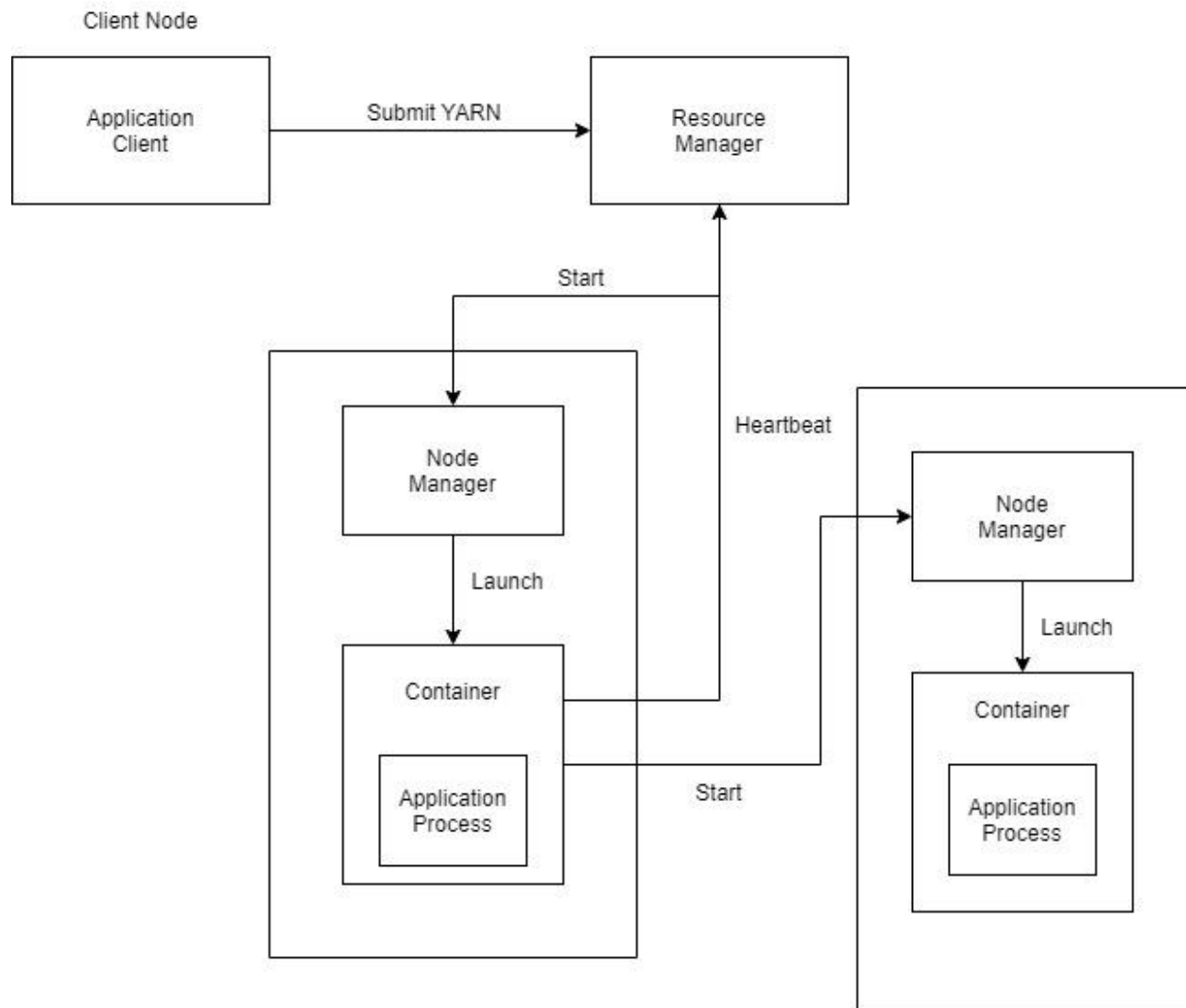


Kiến trúc HDFS

1. Client tương tác với HDFS đưa ra lệnh đọc hoặc viết file.
2. HDFS gọi tới name node đưa ra yêu cầu việc lấy ra vị trí file hoặc tạo file mới.
3. HDFS trả về FSDataInput/Output stream cho client đọc dữ liệu hoặc viết dữ liệu mới.
4. Dữ liệu sẽ được đưa ra cho client đọc. Sau khi viết, dữ liệu sẽ được đóng gói thành các packets nhỏ và đưa dần vào các data nodes.
5. Sau khi hoàn thành client sẽ đưa ra lệnh đóng.
6. Với quá trình viết dữ liệu, HDFS sẽ báo cho name node việc viết đã hoàn thành và đưa vị trí.



YARN application



Kiến trúc YARN Hadoop

1. Client liên lạc với resource manager và yêu cầu chạy application master process.
2. Resource manager sau đó tìm 1 node manager có thể chạy application master trong container.
3. Application master có thể chạy theo yêu cầu client hoặc là báo với resource manager để yêu cầu thêm containers.

Name node : quản lý hệ thống file, lưu trữ thông tin của dữ liệu trong các datanode

Data node : lưu trữ thông tin trong các file dựa theo chỉ dẫn của namenode và cho phép clients đọc và viết lên file.

Resource manager : quản lý tài nguyên (CPU và bộ nhớ) của các nodes.

Node manager : khởi chạy và theo dõi việc thực hiện trong các container của các nodes.

Application process : phối hợp thực hiện công việc.

Container : quản lý các tài nguyên hệ thống để thực hiện application process, được lập lịch bởi resource manager và quản lý bởi node managers.

Các lệnh cơ bản thao tác với HDFS :

hdfs fsck / : kiểm tra tất cả hệ thống (fsck : filesystem check)

hdfs dfsadmin : đưa ra thông tin hệ thống

start-balancer.sh : phân bổ các block 1 cách đồng đều

start-all.sh : khởi chạy tất cả.

stop-all.sh : dừng mọi thứ

hdfs namenode -format : format lại name node (sẽ làm mất dữ liệu trong hdfs)

hadoop fs -mkdir /user/username : tạo directory

hadoop fs -chown -R username /user/username : cấp quyền cho user cho directory

hadoop fs -put file <path> : copy files từ ngoài local file system vào hdfs

hadoop fs -get (-f) <src> <localdest> : copy file từ hdfs ra ngoài local file system

hadoop fs -touch <path> : update thời gian truy cập và chỉnh sửa file, nếu chưa có file thì sẽ tạo file trống với mốc thời gian là mốc tạo file.

hadoop fs -touchz <path> : tạo file trống.

`hadoop fs -rm (-R) <path>` : xóa file, -R để xóa directory và mọi thứ trong đó.

`hadoop fs -mv <path> <dest>` : di chuyển file giữa 2 địa điểm.

`hadoop fs -tail <path>` : stdout kilobyte cuối của file.