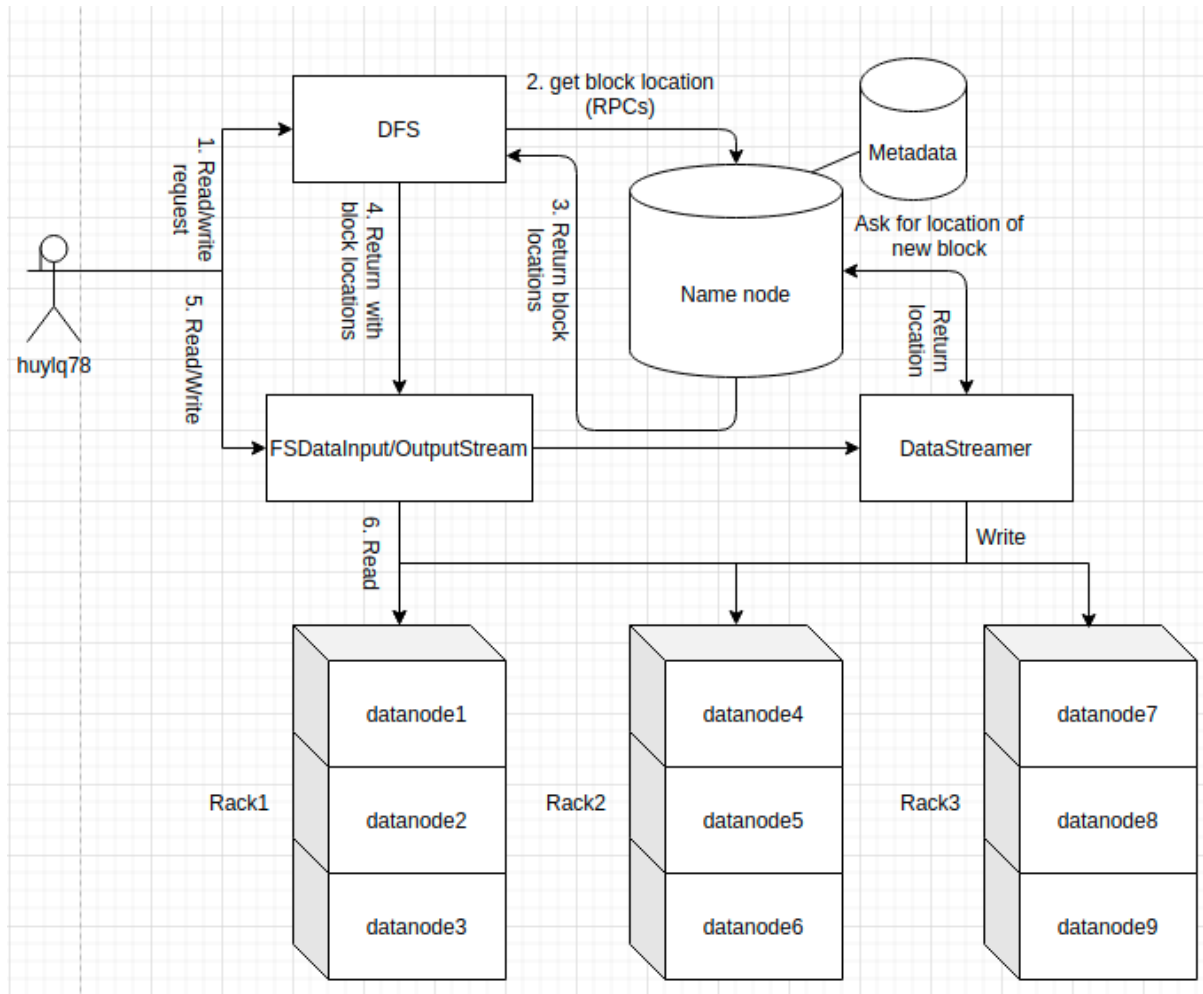


1.Kiến trúc HDFS



1. Client thực hiện request đọc/ghi file vào distributed file system (DFS)
 - 2.1 read: DFS gọi Name node để lấy địa chỉ của các block của file mà client request qua cơ chế RPCs.
 - 2.2 write: DFS gọi Name node để kiểm tra xem file đã tồn tại hoặc người dùng có quyền viết không.
 - 3.1 read: Name node trả về địa chỉ của các block
 - 3.2 write: Name node tạo 1 bản ghi của file mới
 4. DFS trả về Input/output stream để client có thể đọc/ghi
 5. Client thực hiện đọc ghi thông qua i/o stream
 - 6.1 read: Client đọc với địa chỉ lưu trong InPutStream
 - 6.1 write: OutPutStream chia data được ghi thành các gói nhỏ và đưa đến DataStreamer
- DataStreamer có nhiệm vụ nhận địa chỉ các block mới để ghi vào node trong cụm

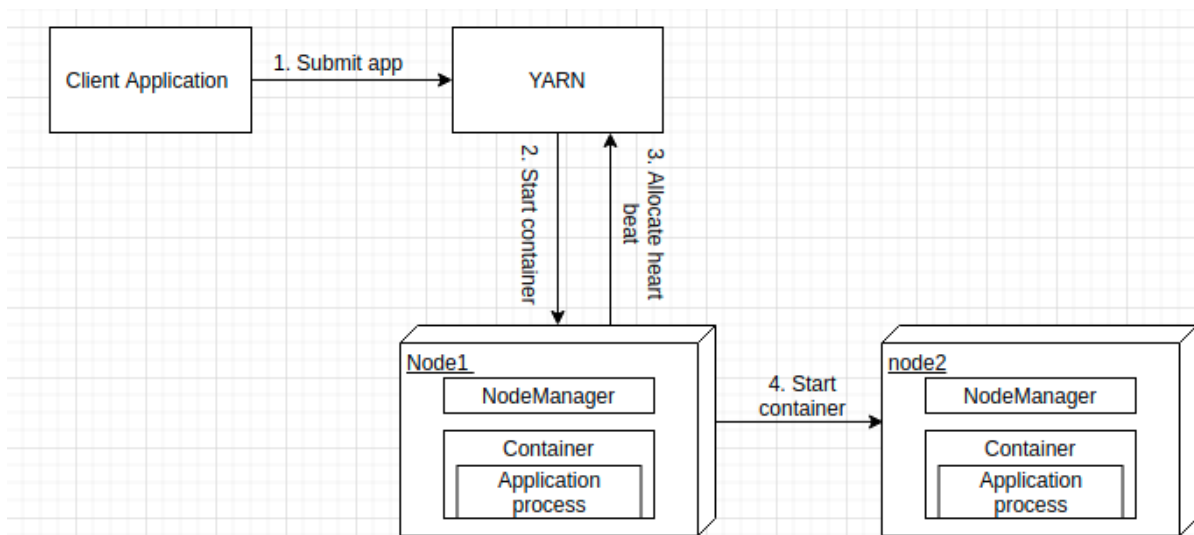
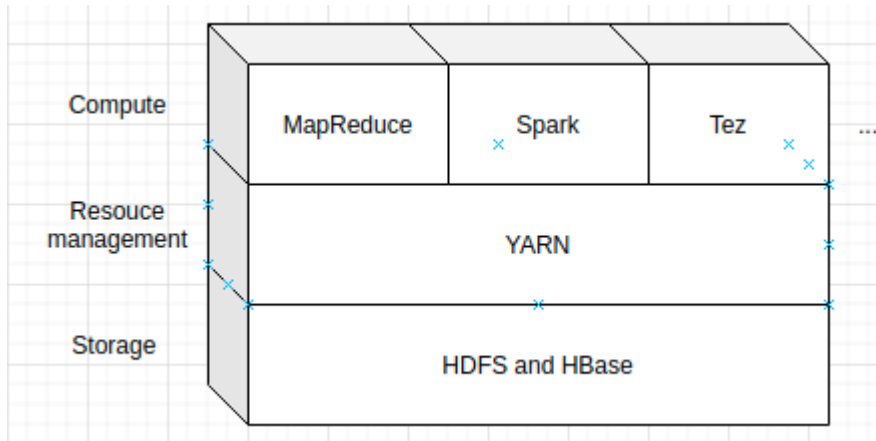
-Name node:

- + Quản lý cấu trúc thư mục trong hdfs (FSImage)
- + Lưu trữ quyền của client với mỗi file (FSImage)
- + Lưu trữ vị trí của các block trong file và kích thước của chúng (FSImage)
- + Lưu trữ các hoạt động ghi của client (EditLog)

-Data node:

- + Lưu trữ File
- + Thực hiện đọc ghi theo yêu cầu client
- + Thực hiện tạo, xóa, tạo bản sao theo yêu cầu name node

2.kiến trúc YARN



1. Client gửi yêu cầu chạy app đến YARN
2. YARN tìm node đủ điều kiện để chạy app thông qua node manager
3. Node1 gửi yêu cầu thêm node để chạy tính toán phân tán
4. Node2 chạy app cùng node1

3.Lệnh trên hdfs

- appendToFile <localsrc> ... <dst>: thêm nội dung file từ <localsrc> đến <dst>. Nếu <localsrc> = - lệnh sẽ lấy từ stdin
- cat [-ignoreCrc] <src>: đọc file ra stdout
- checksum <src>: Trả về checksum
- chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH: thay đổi quyền
- chown [-R] [OWNER][:[GROUP]] PATH: thay đổi người sở hữu file
- cp [-f] [-p | -p[topax]] [-d] <src> ... <dst>: copy từ src đến dst
- find <path> ... <expression>: tìm find match expression

- ls: liệt kê file
- mkdir: tạo folder
- mv: di chuyển file
- put: cop file
- rm: xóa file/folder
- touchz: tạo file
- fsck: check hệ thống file
- job: submit job
- hadoop dfsadmin -report: trả về báo cáo về filesystem
- hadoop dfsadmin -safemode: vào/thoát safe mode(không được chỉnh sửa hệ thống file)