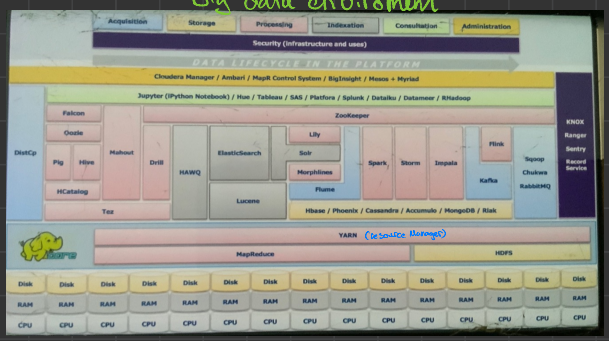
Hadoop is an open source platform that allows for distributed processing of large data sets of across clusters of computers using simple programming models.

Based on: Map Reduce algorithm.

Big data environment



Authentication: by default there is no authentication mechanism is enforced on hadoop cluster.

Authorization: HDFS
Hadoop Data files support posix permissions.

How it process the data?
It have two component 1. MapReduce
2. Yarn

Data protection in transit: no encryption is applied on data by default.

What things to search for a Hadoop vuln:

1. Name Node (HDFS metadata port: 8020
   Yarn job submission port: 8030
   HDFS Name Node Web GUI port: 50070 / 50470
   MapReduce Jobhistory web GUI port: 19888 / 19890
   Yarn Resource Manager Web UI port: 8088 / 8090

2. You have to find all the services on the hadoop cluster.

```
Nmap for Hadoop Hunting

nmap --script hadoop-datanode-info ip_address

nmap --script hadoop-namenode-info ip_address

nmap --script hadoop-jobtracker-info ip_address

nmap --script hadoop-secondary-namenode-info
ip_address

nmap --script hadoop-tasktracker-info ip_address
```

Hunting for hadoop :

* Browsing the HDFS data leak :

> nmap -p 50075 IP_target

↳ if port is open use hdfsbrowser.py tool to exploit.

> python hdfsbrowser.py IP_target

↳ if there is service running you will get infos.

* Hadoopsnooper tool :