Ans 1.

YARN ARCHITECTURE

COMPONENETS OF YARN:

1. RESOURCE MANAGER: master daemon of the hdfs system.

2. NODE MANAGER: slave daemon of the hdfs system.

3. APPLICATION MASTER: controls all the application manager of the yarn system

4. APPLICATION MANAGER: starts when job starts and dies when job dies, make use of container.

5.CONTAINER: contain four elements :

6. SCHEDULER: it has some default scheduler. Such as FIFO, capacity scheduler, Fair Scheduler.

FIFO: First come first served, whatever job comes first the memory get assigned to that job. It doesn’t give any priority. Requests for the first application in the queue are allocated first; once its requests have been satisfied, the next application in the queue is served, and so on.

Capacity: With the Capacity Scheduler, a separate dedicated queue allows the small job to start as soon as it is submitted. Small job get the memory space first and then the long job will be done. Due to this the job gets priority it doesn’t has to wait. Job can be run parallel.

Fair: With the Fair Scheduler, there is no need to reserve a set amount of capacity, since it will dynamically balance resources between all running jobs. Just after the first (large) job starts, it is the only job running, so it gets all the resources in the cluster. When the second (small) job starts, it is allocated half of the cluster resources, so that each job is using its fair share of resources.