YARN

<https://www.cloudera.com/documentation/enterprise/5-4-x/topics/cdh_ig_running_spark_on_yarn.html>

<https://www.cloudera.com/documentation/enterprise/5-4-x/topics/cdh_ig_yarn_tuning.html>

<http://blog.csdn.net/lw_ghy/article/details/50859346>

1 yarn的集群进程：ResourceManager、NodeManager。

2 每个应用程序都对应一个ApplicationMaster的进程。是yarn启动的第一个Container。

3 ApplicationMaster同ResourceManager进行交互。

如果是Spark程序，则Spark ApplicationMaster向ResourceManager申请资源，ResourceManager 接受申请资源的请求，并根据集群资源使用情况，分配资源（给一个NodeManager）。

Spark ApplicationMaster获得请求响应后，则与对应的NodeManager交互。NodeManager启动一个Container，Spark的Executor则在这个Container上运行。

4 如果是cluster模式，则driver运行在ApplicationMaster的同一个Container中。这个Container既要负责申请资源，又负责任务的调度。

如果是client模式，则driver运行在提交程序的节点。则ApplicationMaster只负责申请资源，而driver在所在的节点进行任务调度。

5 cluster模式和client模式的区别在于：Driver所在的位置不同，ApplicationMaster的工作量不同。