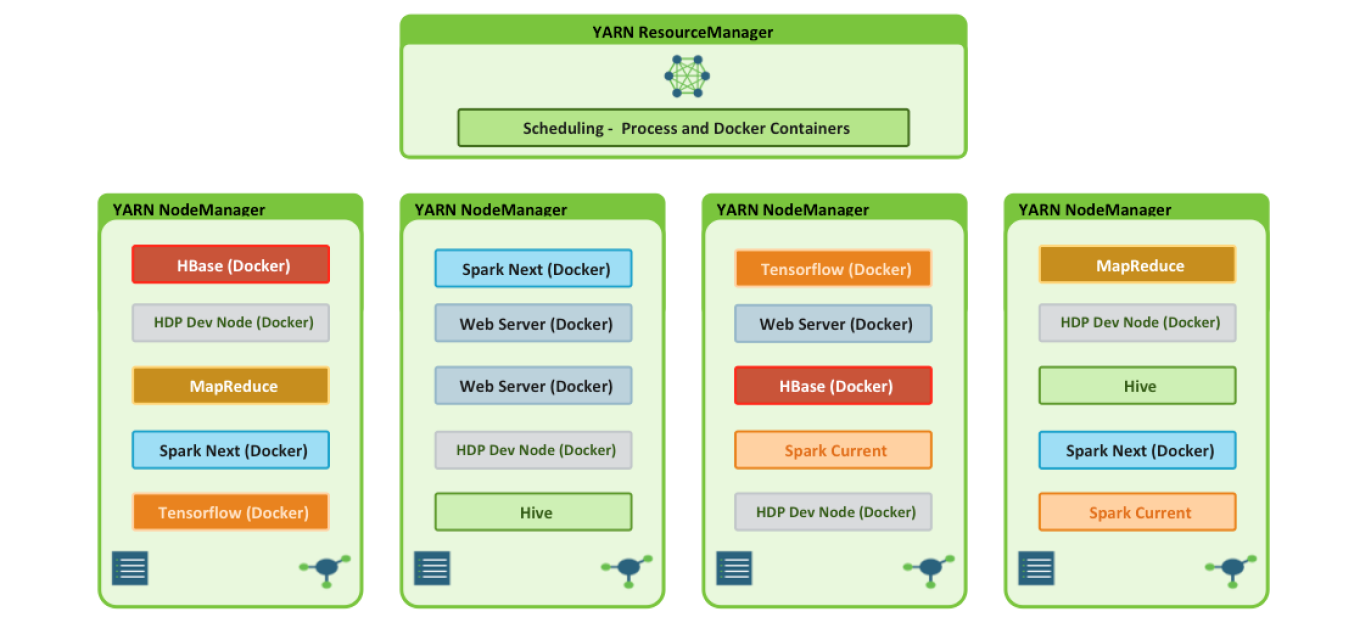
Hadoop 3.0新特性

Hadoop 3.0是一个大版本，具有很多新特性，

1. YARN Service framework
2. Erasure coding
3. 扩展性及有用性，支持多Standby NameNodes
4. 支持GPU、FPGA
5. Intra-node disk balancing
6. Hadoop 3支持基于用户limit及应用优先级实现队列内资源竞争
7. 放置约束，用于在集群中部署微服务

https://zh.hortonworks.com/blog/apache-hadoop-3-1-giant-leap-big-data/

为了支持多样化的数据密集型负载，使用容器技术是主要途径



使用Docker将封装lib和依赖文件中及在同一个集群可以支持不同版本的服务

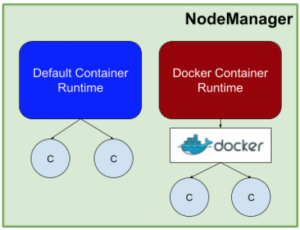
要实现以下功能：

* 编排无状态的分布式应用
* 打包Spark应用的lib，减少部署遇到的依赖冲突问题
* 在YARN中启动Container Cloud
* 在不同OS上运行测试，而且OS升级不影响应用的运行

为了在YARN中支持Docker，需要两个关键技术：

1. Docker runtime support
2. YARN Service

YARN Service在一个服务定义文件中用于定义组件的layout及服务执行参数（重试policies及配置）。用户向集群提交spec file，使用Docker runtime来通过NM启动Docker。



Blog 1~5:

https://zh.hortonworks.com/blog/first-class-support-long-running-services-apache-hadoop-yarn/

https://issues.apache.org/jira/browse/YARN-5079

https://issues.apache.org/jira/browse/YARN-4793

https://issues.apache.org/jira/browse/YARN-4757

https://issues.apache.org/jira/browse/YARN-6419

https://zh.hortonworks.com/blog/data-lake-3-0-part-2-multi-colored-yarn/

https://zh.hortonworks.com/blog/data-lake-3-0-deploy-minutes-cut-tco-half/

https://zh.hortonworks.com/blog/trying-containerized-applications-apache-hadoop-yarn-3-1/

<https://zh.hortonworks.com/blog/apache-hadoop-3-1-0-released-and-a-look-back/>