## 安装

1. Hadoop的各个组件均采用XML文件进行配置，这些配置文件都在 etc/Hadoop 子目录中
   * core-site.xml 文件用于配置通用属性
   * hdfs-site.xml文件用于配置HDFS属性
   * mapred-site.xml文件用于配置MapReduce属性
   * yarn-site.xml用于配置Yarn属性

Hadoop MapReduce Next Generation – cluster setup

## 目的

本文描述了安装、配置和管理Hadoop集群，集群可以是几个节点到成千上万个节点的集群。

为了使用Hadoop，你可能需要首先在单个节点上安装

## 安装

安装Hadoop集群通常包括两种方法：在每台集群的机器上解压软件；安装RPMs

在集群中，设置两台机器为管理机，一台机器设置成NameNode，一台机器设置成ResourceManager。

集群中其余的机器即是DataNode，也是NodeManger。这些是slave节点。

## 在非安全模式运行Hadoop

下面的部分描述了怎么配置Hadoop集群

配置文件：

Hadoop的配置包含两个重要的配置文件：

* 默认的只读配置 ： core-default.xml, hafs-default.xml, yarn-default.xml和mapred-default.xml
* 指定站点配置：conf/core-sit.xml, conf/hdfs-sit.xml, conf/yarn-site.xml和conf/mapred-site.xml

另外，可以通过在conf/hadoop-env.sh和yarn-env.sh中设置指定站点的值来控制bin/目录下的Hadoop脚本。

站点配置

需要配置Hadoop守护进程的执行环境和参数来配置Hadoop集群

Hadoop的守护进程是 NameNode/DataNode 和 ResourceManager/NodeManager

配置Hadoop守护进程环境

管理员应该使用conf/hadoop-env.sh 和 conf/yarn-env.sh脚本来定制Hadoop守护进程的运行环境。最少你要正确配置每个远程节点的JAVA\_HOME。

大多数情况下，你要指定HADOOP\_PID\_DIR和HADOOP\_SECURE\_DN\_PID\_DIR的目录，该目录只能被运行Hadoop守护进程的用户写入。否则可能遭到符号链接攻击