UDH MapReduce开发环境搭建

目录

[一、概述 2](#_Toc426470563)

[二、需要哪些东西 2](#_Toc426470564)

[三、搭建客户端jdk环境 2](#_Toc426470565)

[四、下载UDH-SDK 2](#_Toc426470566)

[五、修改远端UDH集群配置 3](#_Toc426470567)

[六、下载远端UDH集群配置 3](#_Toc426470568)

[七、修改客户端开发环境配置 4](#_Toc426470569)

[八、打包Map及Reduce类 4](#_Toc426470570)

[九、客户端调度执行MapReduce 4](#_Toc426470571)

# 一、概述

在使用UDH过程中，难免需要开发MapReduce程序，鉴于平时大家都已经习惯使用Eclipse，本文就介绍如何搭建基于Eclipse的UDH MapReduce的开发环境。

# 二、需要哪些东西

1. UDH集群

这个可以是已经搭建好的远端UDH集群，也可以是本地伪集群。

本文就以使用远端UDH集群为例进行说明。

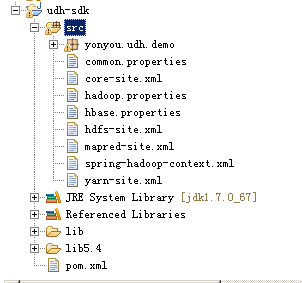
1. 客户端开发环境

# 三、搭建客户端jdk环境

这个比较简单，没什么可说的，只对JDK版本要求做下说明：

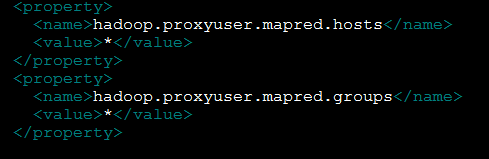
|  |  |  |
| --- | --- | --- |
| UDH | JDK最小版本 | 推荐JDK版本 |
| UDH-1.0.4+ | 1.7.0\_55 | 1.7.0\_67+ |
| 1.8.0\_40 | 1.8.0\_40+ |
| < UDH-1.0.4 | 1.7.0\_45 | 1.7.0\_45+ |

# 四、下载UDH-SDK



# 五、修改远端UDH集群配置

1、在core-site.xml添加下面配置项（非必须，如果不用mapred，需要自己在UDH集群中建立用户，并赋权），建议通过UDH Manager管理界面添加，不然重启服务后，会覆盖已修改配置。



2、在HDFS上建用户目录

$sudo –u hdfs Hadoop fs –mkdir /user/mapred

$ sudo –u hdfs Hadoop fs –chown –R mapred:hadoop /user/mapred

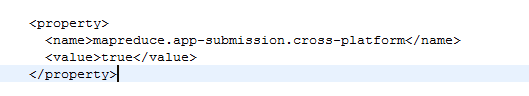
# 六、下载远端UDH集群配置

从UDH集群下载hadoop的相关配置文件（hdfs-site.xml，mapred-site.xml，yarn-site.xml，core-site.xml），并覆盖UDH-SDK中的相关文件。

# 七、修改客户端开发环境配置

如果开发坏境是windows，需要调整以下配置项：

在mapred-site.xml里添加：



# 八、打包Map及Reduce类

把开发的Map及Reduce类打包，并上传的远端UDH集群的HDFS上，比如：

hdfs://namenodehost:8020/user/mapred/lib/mapreducedemo.jar

# 九、客户端调度执行MapReduce

1. 把mapreduce包加载到执行上下文中：

job.addFileToClassPath(new Path("/user/mapred/lib/mapreducedemo.jar"));

1. 指定相关用户执行，这里使用前面配置的mapred。

具体见下面代码片段：

