**poseidon安装文档**

|  |  |  |
| --- | --- | --- |
| 文件状态：  [] 草稿  [√] 正式发布  [ ]正在修改 | 文件标识： | poseidon安装文档 |
| 当前版本： | 1.0 |
| 作者： | 安培 |
| 完成日期： |  |

修订历史

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 版本 | 编写/修订说明 | 修订人 | 修订日期 | 备注 |
| 1 |  | 创建文档 | 安培 | 2018/06/19 |  |

目录

[一、 准备yarn集群环境 4](#_Toc517278378)

[二、 准备poseidon数据库环境 5](#_Toc517278379)

[三、 准备poseidon服务器环境 6](#_Toc517278380)

[四、 准备任务执行引擎 9](#_Toc517278381)

# 准备yarn集群环境

1. **准备机器**

准备yarn环境的机器，并确定能正常访问每台机器。

从yarn环境上拿到3个yarn环境的配置文件 core-site.xml，hdfs-site.xml，yarn-site.xml，以备部署波塞冬环境时候使用。

1. **安装软件包**

在所有机器安装 hadoop-2.6.3 和 jdk1.7.0\_79，并配置好以下环境变量

HADOOP\_HOME=/usr/local/hadoop-2.6.3

HADOOP\_CONF\_DIR=${HADOOP\_HOME}/etc/hadoop

JAVA\_HOME=/usr/java/jdk1.7.0\_79/

请务必使用指定目录和软件包版本，否则可能存在兼容性问题。

1. **创建任务日志路径**

在所有 NodeManager 机器创建 /data/log/hadoop/ 目录, 为jstorm任务运行时存放日志的位置

请务必使用此目录位置，否则可能存在兼容性问题。

1. **创建用户和用户组**

所有机器配置 hadoop用户用户组，且hadoop用户可访问 HADOOP\_HOME JAVA\_HOME。可访问 /data/log/hadoop/ 目录

1. **准备hbase环境**

poseidon平台的统计监控功能都依赖于hbase环境，因此务必提前搭建好一套hbase环境。并且获得 hbase的zkRoot，zkPort，zkHost等信息。

# 准备poseidon数据库环境

1. **创建数据库**

使用github下载poseidon 源码，使用doc 目录中streamsuit.sql 文件创建mysql数据库。

脚本中会创建poseidon项目所需表，并初始化配置数据。并会初始化几个管理员账号，用户可根据需要修改初始化的管理员账号信息。

1. **配置数据库连接信息**

将数据库连接信息配置于poseidon 服务器的usr\local\startboot\datasource.properties 文件中。

例如：

spring.datasource.url=jdbc:mysql://10.101.23.101:3306/poseidon?useUnicode=true&amp;characterEncoding=UTF8

spring.datasource.username=ucar\_presto

spring.datasource.password=ucar\_prestotest

你也可以通过修改源码中spring boot的spring.datasource.location 参数修改此位置。

# 准备poseidon服务器环境

1. **准备机器**

准备poseidon环境的机器，并确定能正常访问每台机器，服务器可以有多台。支持高可用部署。服务器会通过 yarn集群配置的 zk进行leader. select。

1. **安装软件包**

在所有机器安装 hadoop-2.6.3， jdk1.7.0\_79，jstorm-2.2.1，flink-1.3.2，python

并配置好以下环境变量

HADOOP\_HOME=/usr/local/hadoop-2.6.3

HADOOP\_CONF\_DIR=${HADOOP\_HOME}/etc/hadoop

JAVA\_HOME=/usr/java/jdk1.7.0\_79/

PYTHON\_HOME=/usr/bin/python

JSTORM\_HOME = /usr/local/jstorm/jstorm-2.2.1/

FLINK\_HOME =/usr/local/flink-1.3.2

请务必使用指定目录和软件包版本，否则可能存在兼容性问题。

1. **修改hadoop软件包配置**

使用yarn环境的配置文件core-site.xml，hdfs-site.xml，yarn-site.xml，覆盖${HADOOP\_HOME}/etc/hadoop 下core-site.xml，hdfs-site.xml，yarn-site.xml 文件。

1. **修改jstorm软件包配置**

修改usr\local\jstorm\jstorm-2.2.1\conf 下storm.yaml文件内容，修改参数

storm.zookeeper.port: 5181，确保此参数值为5181

使用github下载streamjstorm 源码，使用maven打包生成jstorm-core 子项目生成jstorm-core-1.0.jar 的jar包。将此jar包覆盖usr\local\jstorm\jstorm-2.2.1\lib 目录下 jstorm-core-1.0.jar软件包

1. **修改flink软件包配置**

根据集群实际环境修改usr\local\flink-1.3.2\conf 下 flink-conf.yaml文件内容

主要是zk高可用配置和hadoop配置文件位置的配置。

env.HADOOP\_CONF\_DIR: /usr/local/hadoop-2.6.3/etc/hadoop

fs.hdfs.hadoopconf: /usr/local/hadoop-2.6.3/etc/hadoop

high-availability:zookeeper

high-availability.storageDir: hdfs:///hadoop2yarn/flink/ha/

high-availability.zookeeper.quorum: 10.204.245.43:5181,10.204.245.44:5181,10.204.245.45:5181

high-availability.zookeeper.path.root: /flink

high-availability.zookeeper.path.namespace: /flink\_yarn

使用yarn环境的配置文件yarn-site.xml，覆盖usr\local\flink-1.3.2\conf 下yarn-site.xml文件。

1. **创建任务文件临时目录**

在所有机器创建 /usr/local/tmpjar/ 目录, 为任务文件版本文件的暂存路径。

1. **创建用户和用户组**

所有机器配置 hadoop用户用户组，且hadoop用户可访问 以上步骤中所有目录。

1. **打包**poseidon**项目war包**

使用github下载的streamsuit 源码打包poseidon项目，

根据使用的不同环境修改maven的filter文件。Maven打包后相关参数会替换core-site.xml，hdfs-site.xml，yarn-site.xml 文件中的参数。

maven.hadoop.address=hdfs://hadoop2yarn/

maven.hadoop.tmp.dir=/data/

maven.hadoop.hadoopuser=hadoop

maven.hadoop.zookeeper.quorum=hadoop1:2181,hadoop2:2181,hadoop3:2181

maven.hadoop.zookeeper.parent-znode=/hadoop-ha-yarn-jstorm

maven.hadoop.replication=2

maven.hadoop.nameservices=hadoop2yarn

maven.hadoop.host1=jstormonyarn.dmz.test01.tj1-1

maven.hadoop.host2=jstormonyarn.dmz.test01.tj1-2

maven.hadoop.rpcport=9001

maven.hadoop.httpport=50070

maven.hadoop.shared.edits.dir=qjournal://jstormonyarn.dmz.test01.tj1-1:8485;jstormonyarn.dmz.test01.tj1-2:8485/hadoop2yarn

maven.hadoop.dfs.journalnode.edits.dir=/usr/local/hadoop-2.6.3/data

maven.hadoop.resourcemanager.cluster-id=rmcluster-jstorm

maven.hadoop.rmport=8088

maven.hadoop.resource.memory-mb=7168

maven.hadoop.minimum-allocation-mb=1024

maven.hadoop.nodemanager.resource.cpu-vcores=4

maven.hadoop.scheduler.maximum-allocation-mb=3172

1. **启动**poseidon**项目**

分别在每台服务器使用命令行启动spring boot的java项目。（注意必须以hadoop用户启动）

/usr/java/jdk1.7.0\_79/bin/java -jar /usr/local/startboot/streamsuite.war &

控制台输出 "StreamSuiteMain ---started" 后证明项目已启动成功

# 准备任务执行引擎

想要成功在poseidon平台提交 jstorm任务。必须在平台的“实时引擎版本配置”功能中上传所需要的引擎包。具体上传方式请参考 <<用户使用手册.doc >>

想要成功在poseidon平台提交cql相关任务。必须在“任务文件管理”功能中进行上传cql引擎包。具体上传方式请参考 <<用户使用手册doc>>

因此需要进行如下准备工作，准备需要上传的引擎包。

1. **准备jstormAm 引擎包**

使用github下载的streamjstorm 源码打包jstormAm 引擎包,

首先需要使用yarn环境的配置文件core-site.xml，hdfs-site.xml，yarn-site.xml，覆盖jstorm-on-yarn 子项目下core-site.xml，hdfs-site.xml，yarn-site.xml 文件。

使用maven打包生成jstorm-on-yarn，target目录下会生成 jstorm-yarn-1.0.0-jar-with-dependencies.jar 文件, 此文件为jstormAm 引擎包。

1. **准备 jstorm引擎包**

使用github下载streamjstorm 源码，将软件包目录下的streamjstorm子目录下的所有文件压缩为zip文件包。（注意压缩后的ZIP文件不是以streamjstorm为根目录，而是直接包含streamjstorm目录下的所有文件）。

压缩后的ZIP文件即为jstorm引擎包。

1. **准备 cql引擎包**

打包好的poseidon项目war包，自身就是cql引擎包。