日志收集系统部署手册

1. 程序包结构说明：

|  |  |
| --- | --- |
| **nginx/** | **nginx相关目录，nginx对外暴露日志数据接收接口** |
| nginx/conf/nginx.conf | nginx主配置文件 |
| nginx/conf/online | 各个日志接收接口的具体配置文件，包含多个文件 |
| **logcenter/** | **logcenter相关目录，logcenter负责日志数据的初步解析和存储** |
| logcenter/bin | 程序启动脚本，包含启动、关闭、重启脚本 |
| logcenter/conf | 程序配置文件目录 |
| logcenter/lib | 程序依赖jar包目录 |
| logcenter/log | 日志文件输出目录 |
| **parquet/** | **转parquet程序，运行于spark环境中的作业** |
| parquet/MoreTVBISpark.jar | 电视猫转parquet程序 |
| parquet/WhaleyBISpark.jar | 微鲸转parquet程序 |
| **/script** | **日志文件备份脚本** |
| /script/ WeixinLogUpload.sh |  |
| /script/ MedusaLogUpload.sh |  |
| /script/ MedusaPostLogUpload.sh |  |
| /script/ MoreTVLogUpload.sh |  |
| /script/ WeixinLogUpload.sh |  |

2. 术语表

|  |  |
| --- | --- |
| SPAKR\_MASTER\_URL | spark集群master地址 |
| logpath\_logcenter | logcenter日志文件输出位置，默认/var/log/logcenter |
| logpath\_nginx | nginx日志文件输出位置，默认/var/log/nginx |
| hdfspath\_logcenter | logcenter日志文件HDFS备份路径，现网环境位置参见附录《HDFS文件目录》 |
| hdfspath\_ngxin | nginx日志文件HDFS备份路径，现网环境位置参见附录《HDFS文件目录》 |
| hdfspath\_parquet | 转parquet程序输出路径，现网环境位置参见附录《HDFS文件目录》 |

3. 环境要求

|  |  |
| --- | --- |
| 软件环境 | JDK8、scala2.11、spark1.6 standalone |
| 运行用户 | nginx：默认为其安装目录的owner，遵循运维规划  logcenter：默认为其安装目录的owner，遵循运维规划  日志文件备份脚本：默认为hadoop，遵循运维规划  转parquet程序：默认为spark，遵循运维规划 |
| 文件访问权限说明 | |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | nginx  启动用户 | logcenter  启动用户 | 日志文件备份脚本  启动用户 | 转parquet程序启动用户 | | logpath\_nginx | 读写 | N/A | 读 | N/A | | logpath\_logcenter | N/A | 读写 | 读 | N/A | | hdfspath\_ngxin | N/A | N/A | 读写 | 读 | | hdfspath\_logcenter | N/A | N/A | 读写 | 读 | | hdfspath\_parquet | N/A | N/A | N/A | 读写 | |

4. 部署说明

nginx、logcenter为不间断运行的服务程序，转parquet程序、日志备份脚本为crontab调度的脚本，部署依次分为3个步骤执行：logcenter部署、nginx部署、crontab脚本任务部署（转parquet程序及日志备份脚本）

4.1 logcenter部署

1. 程序包部署：将logcenter子目录拷贝到部署目录，如有必要并修改目录owner，具体位置遵循运维规划
2. 配置修改：

* 监听端口修改：conf/application.ini文件，修改配置项-Dhttp.port，如-Dhttp.port=8081

1. 启动程序：

执行bin/startup.sh启动程序，查看log/application.log,正常情况下日志显示应用已经启动且监听2中所配置的端口，可通过ps、netstat指令进一步确认进程是否启动和端口是否被监听

1. 程序停止、重启：

执行/bin/shutdown.sh停止程序；执行/bin/restart.sh重启程序

4.2 nginx部署

1. 程序包部署：nginx作为基础组件，程序包部署遵循运维标准，目前采用openrestry发行包
2. 配置修改：

* server监听端口修改：online目录下所有的conf文件，修改server配置节中的listen属性
* logcenter端口修改：nginx.conf文件，修改logcenter upstream配置节，指向实际的logcenter服务地址

1. 启动程序：

执行nginx目录下的/sbin/nginx启动程序；可通过ps、netstat指令进一步确认进程是否启动和端口是否被监听

1. 配置热加载：

执行nginx目录下的/sbin/nginx –s reload 重新加载配置

4.3 crontab脚本任务部署

1. 程序包部署：将转parquet程序及日志备份脚本拷贝到部署目录，如有必要并修改目录owner，具体位置遵循运维规划
2. 配置修改：

* 日志备份脚本：HDFS日志备份目录按照运维规划做统一修改，hadoop相关环境引用更具机房环境做统一修改
* 转parquet程序：

1. 提交crontab任务：

按照预定调度策略提交crontab脚本，调度策略参见《crontab调度策略列表》

4.4 部署确认和服务检查点

* 1. 通过如下几个点确认部署后服务运行正常
  2. 通过tail检查nginx日志文件，看日志是否正常输出；
  3. 通过tail检查logcenter日志文件，看日志是否正常输出；
  4. 通过crontab -l 确认脚本任务已经正确部署；
  5. 在预定的调度时间点，检查日志文件是否被备份；
  6. 在预定的调度时间点，检查转parquet任务是否被调度，是否正常产出parquet文件

5. 附录

5.1 HDFS文件目录

|  |  |  |
| --- | --- | --- |
| **相关术语** | **路径** | **备注** |
| hdfspath\_ngxin  hdfspath\_logcenter | /backup | 日志备份数据 |
| /bak | 日志备份数据 |
| hdfspath\_parquet | /mbi/parquet | 电视猫2.x用户行为日志 |
| /log/medusa/parquet | 电视猫3.x用户行为日志 |
| /log/activity/parquet | 活动日志 |
| /log/danmu/parquet | 弹幕日志 |
| /log/dbsnapshot/parquet | 电视猫和微鲸用户库快照 |
| /log/medusaAndMoretvMerger | 电视猫2.x和3.x合并日志 |
| /log/metis/parquet | metis用户行为日志 |
| /log/moretvloginlog/parquet | 电视猫用户登录日志 |
| /log/mtvkidsloginlog/parquet | 少儿颗粒化应用用户登录日志 |
| /log/weixin/parquet | 微信公众号操作日志 |
| /log/whaley/parquet | 微鲸用户行为日志 |

5.2 crontab调度策略列表

|  |  |  |
| --- | --- | --- |
| **应用** | **脚本&指令** | **调度策略** |
| 转parquet程序 |  |  |
|  |  |
| 日志备份脚本 | WhaleyLogUpload2.sh | 30 0 \* \* \* |
|  |  |