**微鲸/虎鲸业务线机房迁移计划书**

1. 组件描述
2. 组件逻辑结构：

组件的输入和输出服务依赖见下图：



2、组件清单：

|  |  |  |
| --- | --- | --- |
| 层级 | 类别 | 文件清单 |
| 输入层 | HDFS输入目录 | /log/whaley/  /log/ipLocationData  /test/config.json |
| Mysql输入数据库 | 2-18（whaleybi用户）  terminal\_upgrade库  2-20（bislave用户）  ucenter库  2-21（bislave用户）  tvservice库  2-22（bislave用户）  mtv\_cms库  1-8（bislave用户）  dolphin\_terminal库 |
| Redis输入数据库（不迁移） | 1-3  db：4  2-16  db：0 |
| 执行层 | Shell执行脚本 | 2-17（spark用户）  /script/bi/whaley |
| Crontab执行脚本 | 2-17（spark用户）  crontab脚本  2-18（playbi用户）  crontab脚本 |
| 输出层 | Mysql输出数据库 | 2-18（whaleybi用户）  helios\_bi库  whaley\_bi库  orca\_bi库 |
| ES输出数据库 | 1-9  helios库  orca库 |

1. 迁移步骤
2. 前置条件：
3. HDFS服务
4. Spark服务
5. ES服务
6. Mysql服务
7. Redis服务
8. 历史日志备份数据
9. 日志流服务正常工作
10. BI可视化部分迁移完成，接口工作正常
11. BI管理平台部分迁移完成
12. 邮件告警服务
13. 迁移步骤：

计划前7步由运维执行，后4步由大数据组执行。

1. 日志迁移：

全量迁移HDFS上/log/whaley/目录下的全部目录及文件。

1. 一些重要文件迁移：

迁移HDFS上/test/config.json文件和/log/ipLocationData的目录和文件。

1. Mysql历史数据库迁移：

迁移输入层和输出层的数据库到北京机房的数据库（设定人为编号为DB1）。

1. ES历史数据库迁移：

迁移输出层ES的数据库。

1. 执行程序迁移：

迁移2-17节点的/script/bi/whaley全量目录及文件。

1. crontab迁移：

迁移2-17节点（spark用户）和2-18节点（playbi用户）的crontab脚本。

1. 杭州机房的Mysql数据库建立主从同步到北京机房的数据库（设定人为编号为DB2）
2. 启动北京机房的运算服务，产生新的Redis数据，执行统计程序，程序的运行结果存入DB1。（此处执行时将其与杭州机房相差的几天数据补充完整）。
3. 启动BI前端可视化展示，从北京机房DB1提取数据展示。
4. 两地并行，观察DB1和DB2的数据是否一致。并行期，对外服务使用杭州机房的BI前端。
5. 切断杭州机房mysql到北京机房DB2的主从同步，结束两地并行期。对外服务改为北京机房的BI前端。

三、检查清单

前7步由运维执行，后4步由大数据组执行。

|  |  |  |
| --- | --- | --- |
| 迁移阶段 | 检测点 | 检测方法 |
| 1、日志迁移 | 1、日志是否丢失  2、HDFS文件是否目录层级完全一致 | 1、检测截止到某一天的两端的日志文件大小  2、检测两端的HDFS目录的文件树结构、权限 |
| 2、重要文件迁移 | 1、文件是否有损坏  2、文件的所属位置是否一致 | 1、对比两侧的文件大小  2、检测两端的文件所属的HDFS位置、文件权限 |
| 3、Mysql数据库迁移 | 1、数据库是否全量迁移  2、数据库层级是否一致 | 1、检测数据库的库、表层级是否一致  2、检测截止到某一刻的数据库的各表大小是否一致  3、各数据库的访问用户名、密码是否一致 |
| 4、ES数据库迁移 | 1、数据库是否全量迁移  2、库名是否一致 | 1、检测截止到某一天的各库数据量大小是否一致  2、检测ES数据库的层级结构 |
| 5、执行程序迁移 | 1、检测目录及文件是否一致 | 1、检测/script/bi/whaley目录的文件树结构、权限 |
| 6、crontab迁移 | 1、检测crontab文件是否完全一致 | 1、 完全对比两侧的crontab文件 |
| 7、建立数据库主从同步 | 1、检测同步是否成功建立  2、检测同步的一致性 | 1、检测北京机房第二套数据库的库、表层级是否与杭州机房一致  2、观察杭州机房数据产生后，北京机房的db2是否同时产生数据  3、监测主从同步的情况 |
| 8、启动北京机房运算任务 | 1、redis库是否重新生成  2、站点树是否可以正常同步  3、程序是否可以通过脚本正常运行  4、DB1中是否可以有运算结果正常产生  5、crontab脚本执行是否有报错  6、邮件告警服务是否正常 | 1. 运行redis同步的部分程序，检测redis中各db的生成情况 2. 运行站点树同步程序，检测站点树更新时的同步情况。 3. 运行统计程序，检测程序的执行情况和spark集群的运行情况。 4. 运行统计程序，补充杭州机房与北京机房db1相差的几天数据，观察数据产生情况。 5. 全量执行一天数据，检测crontab脚本情况。 6. 运行程序中观察邮件告警服务情况。 |
| 9、前端展示 | 1、检测BI服务是否可以完全跑通  2、检测BI前端是否可以正常访问 | 1、配置BI的后端和前端可视化展示，连接北京机房db1作为数据源。  2、进入BI展示页面，检测各报表项的执行情况。  3、观察几天，检测各报表项的新数据产生情况。 |
| 10、两地并行期 | 1、观察北京机房的db1和db2是否一致 | 1、检测重要数据在两个库的表中是否完全一致。  2、持续检测北京机房的BI展示是否对外服务正常。 |
| 11、切换服务 | 1、检测北京机房对外服务是否正常 | 1、将对外服务的域名切换到北京服务，连接数据源为北京机房的db1。  2、切断杭州机房与北京机房db2的主从同步，废弃北京机房db2。 |