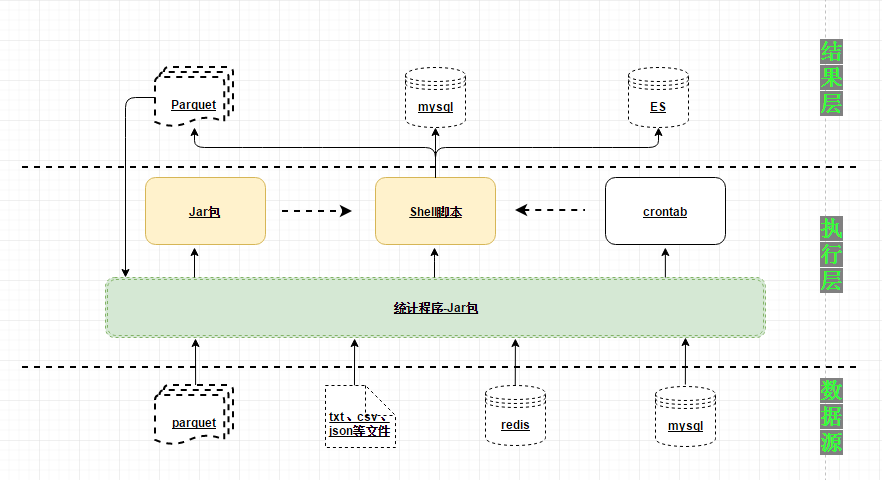
**VR业务线BI组件迁移说明书**

1. **组件描述**

* 组件逻辑结构



* 组件清单

|  |  |  |  |
| --- | --- | --- | --- |
| 层级 | 类别 | 文件清单 | 备注 |
| 数据源 | Parquet文件 | /log/vrapp/parquet/\*  /log/vrapp/\* |  |
| txt/csv/json等 | /log/vrapp/temple/\*  /log/ipLocationData/\*  /test/config.json |  |
| 执行层 | Jar包与shell脚本 | Ip: 2-17  用户：spark  目录：  /script/bi/vrapp/\* |  |
| 结果层 | Parquet | /log/vrapp/temple/\* |  |
| MySQL | 2-15:  vr\_bi库 |  |

1. **迁移步骤**

* 前置条件

1. HDFS环境部署完成
2. Spark环境部署完成
3. ES服务部署完成
4. MySQL环境部署完成
5. Redis环境部署完成
6. 源数据准备OK
7. 日志流服务正常工作
8. BI可视化后台接口和前端服务接口正常工作

* 迁移步骤

1. rawlog/parquet日志迁移；

迁移前文所述的HDFS中的parquet文件

1. txt/csv/json等文件迁移；

迁移前文所述的HDFS中的重要非日志型文件

1. ES数据迁移；

迁移前文所述的Elastic Search中的历史结果数据

1. MySQL与Redis数据迁移；

迁移前文所述的MySQL和Redis数据库中的所有数据

1. Jar包与shell脚本迁移；

迁移前文所述的执行程序的jar包以及相关的shell脚本

1. crontab配置信息迁移;

迁移前文所述的crontab的定时调度脚本配置

1. 杭州机房的结果数据同步至北京机房db1与db2；

完成杭州机房和北京机房的数据库同步，杭州机房为主库，北京机房为从库

1. 切断杭州机房数据同步至db1，启动服务，定时写入数据至db1；

切断杭州机房和北京机房的数据库同步机制

1. 新机房BI服务与杭州BI服务并行运行；
2. 对比分析两套BI环境产生的数据，对比db1与db2中的数据；
3. 前端配置
4. 服务正式切换至新机房的BI环境，切断杭州服务。
5. **检查清单**

Check List

|  |  |  |
| --- | --- | --- |
| 阶段 | 检测点 | 检测步骤 |
| rawlog/parquet日志迁移 | 1. 日志是否有丢失； 2. 日志的组织目录结构是否与原来的目录结构一致； | 1. 对比两端文件大小是否一致； 2. 对比两端文件的路径是否一致 |
| txt/csv/json等文件迁移 | 1、日志是否有丢失；  2、日志的组织目录结构是否与原来的目录结构一致； | 1、对比两端文件大小是否一致；  2、对比两端文件的路径是否一致 |
| ES数据迁移 | 1、日志是否有丢失；  2、日志的组织目录结构是否与原来的目录结构一致； | 1、对比两端文件大小是否一致；  2、对比两端文件的路径是否一致 |
| MySQL与Redis数据迁移 | 1. 数据是否丢失； 2. 数据库用户名以及密码是否一直； 3. 数据库中的表名是否一致 |  |
| Jar包与shell脚本迁移 | 1. 是否完成迁移； 2. 目录结构是否一致； 3. 权限是否一致 |  |
| crontab配置信息迁移 | 1. crontab配置与shell脚本的部署是否一致； 2. crontab的配置是否完整 |  |
| 杭州机房的结果数据同步至北京机房db1与db2 | 是否数据一致 | 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 |
| 切断杭州机房数据同步至db1，启动服务，定时写入数据至db1 | 1. 新服务是否能够正常写入数据； 2. 站点树是否同步； 3. 是否正常运行； 4. 是否产生数据； 5. crontab调度是否可以定时调度； 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。  运行程序中观察邮件告警服务情况。 |
| 新机房BI服务与杭州BI服务并行运行 |  |  |
| 对比分析两套BI环境产生的数据 | 检测两端db1与db2的数据是否完全一致 | 1. 检测重要数据在两个数据库中的是否一致； 2. 检测北京服务的BI展示是否正常服务 |
| 前端展示配置 | 1. 检测BI服务是否可以OK 2. 检测前端BI服务是否可以访问 | 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。 |
| 服务正式切换至新机房的BI环境 | 检测前端展示是否可以正常展示数据 | 1、对外服务域名切换至北京机房的服务，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的同步 |