

2022 年 1 季度工作总结

潘重宇

2022/4/18

在去年 11 月份从数据分析岗转岗实验室大数据平台运维后，经过一个多月的学习与探索，我基本熟悉了大数据平台的原理、操作与架构，了解了岗位的工作内容与职责。在 2022 年第一季度的时间里，我主要完成了以下工作内容。

- 4 个大数据平台运维工具编写
 - HDFS 与本地服务器的文件压缩、解压缩互传工具编写。
 - Hive 表格数据文件的自动快照创建与清理工具编写。
 - HDFS 文件本地备份工具编写。
 - Hive 日志文件抓取工具编写。
- 2 个大数据平台配置运行项目优化
 - MySQL 配置项优化。
 - Hive 表格数据碎片文件的合并与自动合并配置。
- UAT 测试集群相关工作
 - UAT 测试集群的搭建与测试。
 - UAT 测试集群与 HDS 集群的 Kerberos 互信建立与测试。
 - livy/Spark3 组件在 HUE 页面集成。
- 2 项集群 bug 修复
 - 集群 log4j2 组件远程代码注入漏洞修复。
 - HDFS 文件系统高权限问题的研究与解决。
- 大数据平台日常运维项目
 - 天网、资金清算与数据提取项目组 30 多张数据表的导入加载。
 - 每日加载错误表格手动修复。

通过实际项目工作的进行，我更好的学习与掌握了大数据平台的相关知识，尤其是 HDFS 系统的运行原理和部分 API 的使用方法，并通过这编写了自动化工具，解决了一些运维工作严重依赖于手动操作的问题，提高了自动化水平。同时对大数据平台的运行做了些许优化，能够提高平台用户的数据查询、保存的效率。